

## Enrichments and Transformations: Tools for investigating your search results.

EuPathDB sites offer enrichment and transformation tools that will help you interpret and investigate your search results. Our enrichment analyses provide a statistical assessment to help you identify classes of genes that are overrepresented or a metabolic pathway that is common to the genes in your result set. They use the Fisher's exact test to compare frequency of the terms and pathways associated with your search result to the full set of annotated genes for your organism. Another powerful tool is Transform by Orthology which will change your list of genes into a list of their orthologs in the species of your choice.

1. Find the overrepresented Biological Process in your search result from this morning: kinase genes that have a signal peptide and/or a transmembrane domain.
  - a. Focus on the **\*kinase\* step (step 1)** of your strategy from this morning or use the one that we saved in the Public Strategies tab of the My Strategies section.

**My Strategies:** New Opened (2) All (75) Basket **Public Strategies (38)**

(Genes)

"kinase" 1467 Genes Step 1 Add Step

**My Strategies:** New Opened (2) All (75) Basket **Public Strategies (38)** Help

**Examples & Public Strategies**

To make one of your strategies visible to the community, go to the All tab and click its Public checkbox

Search:

Strategies (38)	Returns	Description
Kinases with Signal Peptides and/or TM Domains	Genes	From the Finding Genes workshop exercise 1. search for 'kinase' in
Genes that were antigens in infected children but not in infected adults and that had non-syn SNPs	Genes	In step 1, samples were first sorted by infection status then by ad...
SNPs comparing all strains from Amambua Ngwa et al.	SNPs	S

Signal Pep 23346 Genes

"kinase" product 1467 Genes Step 1 → 196 Genes Step 2 Add Step

**1467 Genes from Step 1**  
Strategy: Kinases with Signal Peptides and/or TM

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

All Results	Ortholog Groups	<i>P.berghei</i>	<i>P.chabaudi</i>	<i>P.cynomolgi</i>	<i>P.falciparum</i>
		ANKA	chabaudi	strain B	3D7
1467	164	131	132	124	156

Gene Results Genome View Analyze Results **BET**

- b. Look at the results for *P. falciparum* 3D7. Are you able to identify types of proteins just by looking at the product descriptions?
  - Yes!! These genes are kinases. Trick question since we searched for genes that have the word kinase in their product description.

- Can you tell anything about their subcellular location? The metabolic pathways they are involved in?

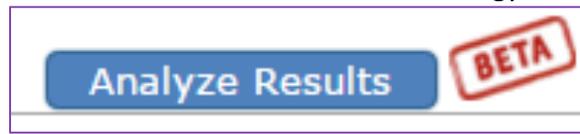
c. **Perform a Gene Ontology Enrichment analysis for the Biological Processes associated with our *P. falciparum* kinase genes.**

- At EuPathDB we get GO terms from the official annotation and by transferring GO terms associated with InterPro domains (using InterPro scan). Here are some useful links with additional information:

<http://geneontology.org/page/documentation>

<http://geneontology.org/page/go-enrichment-analysis>

- Make sure your strategy is focused on Step 1 and click the blue Analyze Results tab next to the result table and then choose the Gene Ontology Enrichment analysis.



Organism => Plasmodium falciparum 3D7

Ontology => Biological Process

GO Sources => Select All

P-value => 0.05

- What biological process are enriched in your set of kinases? Notice the P-values are very low (increased statistical significance). Do these enrichments make sense considering what we already know about these genes?

Gene Results | Genome View | Gene Ontology Enrichment | **Analyze Results**

[ Rename This Analysis | Copy This Analysis ]

### Gene Ontology Enrichment

Find Gene Ontology terms that are enriched in your gene result. [Read More](#)

▸ **Parameters**

**Analysis Results:** [Download](#)

Got a total of 132 results    Filter :

This analysis result may be lost if you change the browser. To save this analysis result, click the Download button.

GO ID	GO Term	Genes in the bkgd with this term	Genes in your result with this term	Percent of bkgd Genes in your result	Fold enrichment	Odds ratio	P-value	Benjamini-Hochberg
GO:0016310	phosphorylation	129	110	85.3	25.22	84.42	4.09e-117	5.40e-115
GO:0006796	phosphate-containing compound metabolic process	164	112	68.3	20.2	70.2	5.73e-112	2.52e-110

- d. Find the genes in your result set that are annotated with **GO:0016310, phosphorylation**. This is the most highly enriched GO term in your result set and has an unadjusted p-value of  $4 \times 10^{-117}$ . To do this we will need to intersect your \*kinase\* results with GO term search.
- Focus on the Signal Peptide search and edit the strategy choosing Insert Step Before.
  - Choose the GO Term search under Putative Function and cut and paste the GO term into the GO Term or GO ID parameter. See screenshots below.
  - How many *P. falciparum* genes in your \*kinase\* result set are annotated with the Phosphorylation GO Term?

**STEP 2 : Signal Pep**  
The nested strategy gets opened below.  
Results: 23346 Genes

**Insert Step**

- Run a new Search for
- Transform by Orthology
- Add contents of Basket
- Add existing Strategy
- Filter by assigned Weight
- Genes
- Genomic Segments
- SNPs
- SNPs (from Chips)
- ORFs
- Text, IDs, Organism
- Genomic Position
- Gene Attributes
- Protein Attributes
- Protein Features
- Similarity/Pattern
- Transcript Expression
- Protein Expression
- GO Term
- Metabolic Pathway **BETA**
- Y2H Protein Interaction
- Predicted Functional Interaction

**Insert Step 2 : GO Term**

Organism

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

GO Term or GO ID

Begin typing to see suggestions to choose from (CTRL or CMD click)  
Or paste a list of IDs separated by a comma, new-line, white-space.

Free Text (use "\*" for wildcard)

**Combine Genes in Step 1 with Genes in Step 2:**

- 1 Intersect 2
- 1 Union 2
- 1 Relative to 2, using genomic colocation
- 1 Minus 2
- 2 Minus 1

**(Genes)**

GO Term 1077 Genes  
Signal Pep 23346 Genes  
"kinase" product 1467 Genes  
110 Genes from Step 2  
25 Genes

**110 Genes from Step 2**  
Strategy: Kinases with Signal Peptides and/or TM Domains

All Results	Ortholog Groups	<i>P. falciparum</i> (n=110)			
		<i>P. berghei</i> ANKA	<i>P. chabaudi</i> chabaudi	<i>P. cynomolgi</i> strain B	<i>P. falciparum</i> 3D7
948	113	84	83	69	110

Gene Results   **BETA**

First 1 2 3 4 5 Next Last

Gene ID	Genomic Location	Product Description
PF3D7_0731400	PF3D7_07_v3: 1,353,758 - 1,355,482 (+)	serine/threonine protein kinase family, pseudogene (FIKX)

- e. OPTIONAL - Explore the other enrichments analyses.

- Is there a particular cellular component that is enriched for the *P. falciparum* \*kinase\* genes? Run another enrichment analysis with the ontology parameter set to Cellular Component.
- What metabolic pathways are associated with the *P. berghei* ANKA kinase genes? What genes contribute to this enrichment?

2. Find *P. berghei* ANKA orthologs of the *P. falciparum* 3D7 \*kinase\* genes. For this exercise you will use <http://plasmodb.org>

Our Transform by Orthology function allows you to create a list of orthologs from any gene set. Using the Transform function within PlasmoDB allows you to transform to orthologs of closely related species that are available in PlasmoDB. Using the Transform function in EuPathDB.org you can transform into distantly related species that are available from EuPathDB.org. These are valuable tools and allow you to take advantage of expression data and annotation associated with well-studied genomes.

- a. Focus on the \*kinase\* search and use the filter table to limit the result set to *P. falciparum* genes. You can use your search from this morning, open the Public Strategy called \*kinase\*, or use the shared strategy link below.

<http://plasmodb.org/plasmo/im.do?s=3975e694d0bf0f69>

**My Strategies:** New Opened (4) All (77) Basket

(Genes)

\*kinase\*  
156 Genes  
Step 1

Add Step

**156 Genes from Step 1**  
Strategy: \*kinase\*(2)

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

All Results	Ortholog Groups	<i>P.berghei</i>	<i>P.chabaudi</i>	<i>P.cynomolgi</i>	<i>P.falciparum</i>	( nr Genes: 156)	
		ANKA	chabaudi	strain B	3D7	IT	
1467	164	131	132	124	156		155

Gene Results Genome View Analyze Results BETA

First 1 2 3 4 5 Next Last Advanced Paging

Gene ID	Organism	Genomic Location
PF3D7_0102600	<i>P. falciparum</i> 3D7	PF3D7_01_v3: 119,041 - 121,249 (-)

- b. Transform the *P. falciparum* genes into their *P. berghei* orthologs.
- Click Add Step, choose Transform by Orthology from the first menu and set the organism parameter to *P. berghei* ANKA.

The image shows a workflow interface. At the top left, a box labeled 'Step 1' contains a search for '\*kinase\*' resulting in 156 genes. A red arrow points from the 'Add Step' button to a dropdown menu. The menu is open, showing options like 'Run a new Search for', 'Transform by Orthology', 'Add contents of Basket', etc. A red circle highlights 'Transform by Orthology'. Another red arrow points from this option to a 'Add Step' dialog box. The dialog box is titled 'Add Step 2 : Transform by Orthology' and shows a tree view of organisms. 'Plasmodium berghei ANKA' is selected under the 'Plasmodium berghei' category. A 'Run Step' button is at the bottom right of the dialog. A final red arrow points from the 'Run Step' button to a summary diagram. The diagram shows 'Step 1' with '\*kinase\*' (156 Genes) and 'Step 2' with 'Orthologs' (135 Genes), with an 'Add Step' button between them.

- c. Our Step 1 results returned 131 *P. berghei* kinases but transforming the *P. falciparum* genes to their *P. berghei* orthologs returned 135 genes. Do you think the product names of these orthologs contain the word 'kinase'? Can you devise a strategy to retrieve the extra genes?
- d. OPTIONAL – Explore other transformations.
- Transform to pathways
  - Transform to compounds

3. Find *Cryptosporidium* and *Leishmania* orthologs of the *P. falciparum* 3D7 \*kinase\* genes.

For this exercise we will use <http://eupathdb.org>

Since EuPathDB.org access all integrated genomes, we can transform our *P. falciparum* genes into orthologs of distantly related species.

- a. Retrieve gene IDs for the list of *P. falciparum* 3D7 \*kinase\* genes. We can use the download function to show the list of gene IDs in a new browser tab.
- Click Download 156 genes on the right side of the screen, above the result table.

- Choose Tab delimited Excel: choose from columns.
- Click Clear all to remove extra columns. Gene IDs are retrieved by default.
- Click Get Report and a browser tab will open and contain your gene IDs.

**Download 156 Genes from the search:**  
Text (product name, notes, etc.)

Please select a format from the dropdown list to create the download report.  
Tab delimited (Excel): choose from columns

\*\*Note: IDs will automatically be included in the report and the report will be sorted by ID.

Below select type and format for download

You may include additional columns in the report  
clear all | expand all | collapse all  
reset to current | reset to default

- Gene ID
- Search-Specific
- Text, IDs, Species
- Genomic Position
- Gene Attributes
- Protein Attributes
- Protein Features
- Transcript Expression
- Putative Function
- Evolution
- Population Biology
- Predicted Protein Sequence
- Predicted RNA/mRNA Sequence (introns spliced out)
- No. of Pathways
- Coding Sequence
- Updated Sequence
- Search Weight

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

Column names:  include  exclude

Download type and format:  Text File  Excel File\*\*  Show in Browser

Get Report

Browser window content:

```

[Gene ID]
PF3D7_0102600
PF3D7_0103700
PF3D7_0107600
PF3D7_0110600
PF3D7_0110900
PF3D7_0111500
PF3D7_0203100
PF3D7_0211700
PF3D7_0213400
PF3D7_0214600
PF3D7_0217500
PF3D7_0301200
PF3D7_0302100
PF3D7_0309200
PF3D7_0310100
PF3D7_0311300
  
```

- b. Enter the list of *P. falciparum* genes into EuPathDB using the Gene ID search.
- Open a new browser page to <http://eupathdb.org> and navigate to the Gene ID search.
  - Cut and paste your *P. falciparum* IDs into the text box for the 'Gene ID Input set' parameter and click Get Answer.

AmoebaDB CryptoDB FungiDB GiardiaDB

Identify Genes by:

Expand All | Collapse All

- Text, IDs, Organism
- List of IDs
- Organism
- User Comments
- Having Updated Annotation at GeneDB

Identify Genes based on List of IDs

Gene ID input set  Enter a list of IDs or text: PF3D7\_1133400

Upload a text file: Choose File No file chosen  
Maximum size: 10MB. The file should contain the list of IDs.

Copy from My Basket: 0 Genes will be copied from your Basket

Copy from My Strategy: Choose a Gene strategy: IDs List(2) (70 Genes)

Get Answer

Give this search a name (optional)

Give this search a weight (optional)

Cut and Paste your IDs here.

- c. Use the Transform by Orthology function to convert the list of *P. falciparum* genes into their *Cryptosporidium* and *Leishmania* orthologs.
- Just as in 2b, click Add Step and choose Transform by Orthology.
  - At the Add Step 2: Transform by Orthology popup, choose *Cryptosporidium* and *Leishmania* for the organism parameter.
  - How many orthologs did you get? You can use the filter parameter to look at different species subsets. Are all of these annotated as Kinases? You will be surprised how many are not annotated as kinase.

