

VEuPathDB: A bioinformatics resource for facilitating data exploration, analysis and integration for vectors and eukaryotic pathogens

Guided workshop BioMalPar XXI: biology and pathology of the malaria parasite May 22, 2025 12-13:30 PM

Workshop objectives:

- Introduction to VEuPathDB resources focusing on PlasmoDB and VectorBase.org
- Use site search
- Build a search strategy
- Explore gene pages (optional)

Site Search

The site search is in the header of any VEuPathDB site and is available on every page. The site search queries the databases for your term or ID and returns a list of pages and documents that contain your query term.

1. **Search for a keyword.** Enter the word *kinase* in the site search window. Then click enter on your keyboard or click on the search icon.



Site Search result format: The site search returns a categorized list of pages and documents that contain your term. Site search results are summarized on the left with a details panel on the right. Changing the panel on the left will populate the details panel with that result. What is the total number of results with the word kinase? Are all the results genes?

All results matching kinase

Category	Count
Genomes	19,596
Population biology	352
Metabolism	352
Data access	1

Results are summarized by category.

Details panel with information about each item returned.

2. **Filter the site search result by category:** How many genes included the word kinase in their product descriptions?

Filter the results so that you can only view gene results, and the Filter Fields section expands to reveal additional filtering options. Select the *Product descriptions* field and choose *Apply*. Once a filter is applied, it can be removed by clicking on *Clear filter* (right panel below).

Filter results	Filter Gene fields	Filter results	Filter Gene fields	Filter results	Filter Gene fields
Genome Genes	<input checked="" type="checkbox"/> EC descriptions and numbers <input type="checkbox"/> GO terms <input type="checkbox"/> InterPro domains <input type="checkbox"/> Notes from annotators <input type="checkbox"/> Orthologs <input type="checkbox"/> PDB chains <input checked="" type="checkbox"/> Product description <input type="checkbox"/> Product descriptions (all) <input type="checkbox"/> PubMed <input type="checkbox"/> Rodent malaria phenotype <input type="checkbox"/> User comments	<input checked="" type="checkbox"/> EC descriptions and numbers <input type="checkbox"/> GO terms <input type="checkbox"/> InterPro domains <input type="checkbox"/> Notes from annotators <input type="checkbox"/> Orthologs <input type="checkbox"/> PDB chains <input checked="" type="checkbox"/> Product description <input type="checkbox"/> Product descriptions (all) <input type="checkbox"/> PubMed <input type="checkbox"/> Rodent malaria phenotype <input type="checkbox"/> User comments	<input checked="" type="checkbox"/> EC descriptions and numbers <input type="checkbox"/> GO terms <input type="checkbox"/> InterPro domains <input type="checkbox"/> Notes from annotators <input type="checkbox"/> Orthologs <input type="checkbox"/> PDB chains <input checked="" type="checkbox"/> Product description <input type="checkbox"/> Product descriptions (all) <input type="checkbox"/> PubMed <input type="checkbox"/> Rodent malaria phenotype <input type="checkbox"/> User comments	Genome Genes	<input checked="" type="checkbox"/> EC descriptions and numbers <input type="checkbox"/> GO terms <input type="checkbox"/> InterPro domains <input type="checkbox"/> Notes from annotators <input type="checkbox"/> Orthologs <input type="checkbox"/> PDB chains <input checked="" type="checkbox"/> Product description <input type="checkbox"/> Product descriptions (all) <input type="checkbox"/> PubMed <input type="checkbox"/> Rodent malaria phenotype <input type="checkbox"/> User comments
<input type="checkbox"/> select all <input type="checkbox"/> clear all	<input type="checkbox"/> select all <input type="checkbox"/> clear all	<input type="checkbox"/> select all <input type="checkbox"/> clear all	<input type="checkbox"/> select all <input type="checkbox"/> clear all	<input type="checkbox"/> select all <input type="checkbox"/> clear all	<input type="checkbox"/> select all <input type="checkbox"/> clear all
<input type="checkbox"/> Hide zero counts		<input type="checkbox"/> Hide zero counts		<input type="checkbox"/> Hide zero counts	
<input type="button" value="Clear filter"/>	19,596	<input type="button" value="Clear filter"/>	19,596	<input type="button" value="Clear filter"/>	7,630

3. **Filter the site search result by organism:** How many of the above genes are found in *Plasmodium falciparum* 3D7? Explore the *Filter organisms* section of the results filter and use the search filter to navigate the tree.

Filter results	Filter Gene fields	Filter results	Filter Gene fields
Genome Genes	<input checked="" type="checkbox"/> EC descriptions and numbers <input type="checkbox"/> GO terms <input type="checkbox"/> InterPro domains <input type="checkbox"/> Notes from annotators <input type="checkbox"/> Orthologs <input type="checkbox"/> PDB chains <input checked="" type="checkbox"/> Product description <input type="checkbox"/> Product descriptions (all) <input type="checkbox"/> PubMed <input type="checkbox"/> Rodent malaria phenotype <input type="checkbox"/> User comments	<input checked="" type="checkbox"/> EC descriptions and numbers <input type="checkbox"/> GO terms <input type="checkbox"/> InterPro domains <input type="checkbox"/> Notes from annotators <input type="checkbox"/> Orthologs <input type="checkbox"/> PDB chains <input checked="" type="checkbox"/> Product description <input type="checkbox"/> Product descriptions (all) <input type="checkbox"/> PubMed <input type="checkbox"/> Rodent malaria phenotype <input type="checkbox"/> User comments	<input checked="" type="checkbox"/> EC descriptions and numbers <input type="checkbox"/> GO terms <input type="checkbox"/> InterPro domains <input type="checkbox"/> Notes from annotators <input type="checkbox"/> Orthologs <input type="checkbox"/> PDB chains <input checked="" type="checkbox"/> Product description <input type="checkbox"/> Product descriptions (all) <input type="checkbox"/> PubMed <input type="checkbox"/> Rodent malaria phenotype <input type="checkbox"/> User comments
<input type="checkbox"/> select all <input type="checkbox"/> clear all	<input type="checkbox"/> select all <input type="checkbox"/> clear all	<input type="checkbox"/> select all <input type="checkbox"/> clear all	<input type="checkbox"/> select all <input type="checkbox"/> clear all
<input type="checkbox"/> Hide zero counts		<input type="checkbox"/> Hide zero counts	
<input type="button" value="Clear filter"/>	7,630	<input type="button" value="Clear filter"/>	7,630
Filter organisms	Filter organisms	Filter organisms	Filter organisms
<input type="checkbox"/> select all <input type="checkbox"/> clear all <input type="checkbox"/> expand all <input type="checkbox"/> collapse all	<input type="checkbox"/> select all <input type="checkbox"/> clear all <input type="checkbox"/> expand all <input type="checkbox"/> collapse all	<input type="checkbox"/> select all <input type="checkbox"/> clear all <input type="checkbox"/> expand all <input type="checkbox"/> collapse all	<input type="checkbox"/> select all <input type="checkbox"/> clear all <input type="checkbox"/> expand all <input type="checkbox"/> collapse all
Type a taxonomic name	Type a taxonomic name	Type a taxonomic name	Type a taxonomic name
<input type="checkbox"/> Reference only	<input type="checkbox"/> Reference only	<input type="checkbox"/> Reference only	<input type="checkbox"/> Reference only
▶ <input type="checkbox"/> Haemoproteidae 331	▶ <input type="checkbox"/> Haemoproteidae 331	▶ <input type="checkbox"/> Haemoproteidae 103	▶ <input type="checkbox"/> Haemoproteidae 103
▶ <input type="checkbox"/> Plasmodiidae 19,265	▶ <input type="checkbox"/> Plasmodiidae 19,265	▶ <input type="checkbox"/> Plasmodiidae 7,527	▶ <input type="checkbox"/> Plasmodiidae 7,527
<input type="checkbox"/> 3d7	<input type="checkbox"/> 3d7	<input type="checkbox"/> 3d7	<input type="checkbox"/> 3d7
<input type="checkbox"/> Reference only	<input type="checkbox"/> Reference only	<input type="checkbox"/> Reference only	<input type="checkbox"/> Reference only
<input type="checkbox"/> select only these <input type="checkbox"/> add these <input type="checkbox"/> clear these	<input type="checkbox"/> select only these <input type="checkbox"/> add these <input type="checkbox"/> clear these	<input type="checkbox"/> select only these <input type="checkbox"/> add these <input type="checkbox"/> clear these	<input type="checkbox"/> select only these <input type="checkbox"/> add these <input type="checkbox"/> clear these
<input type="checkbox"/> Plasmodiidae 7,527	<input type="checkbox"/> Plasmodiidae 7,527	<input type="checkbox"/> Plasmodiidae 7,527	<input type="checkbox"/> Plasmodiidae 7,527
<input type="checkbox"/> Plasmodium 7,395	<input type="checkbox"/> Plasmodium 7,395	<input type="checkbox"/> Plasmodium 7,395	<input type="checkbox"/> Plasmodium 7,395
<input type="checkbox"/> Plasmodium falciparum 2,792	<input type="checkbox"/> Plasmodium falciparum 2,792	<input type="checkbox"/> Plasmodium falciparum 2,792	<input type="checkbox"/> Plasmodium falciparum 2,792
<input checked="" type="checkbox"/> Plasmodium falciparum 3D7 [Ref] 137	<input checked="" type="checkbox"/> Plasmodium falciparum 3D7 [Ref] 137	<input checked="" type="checkbox"/> Plasmodium falciparum 3D7 [Ref] 137	<input checked="" type="checkbox"/> Plasmodium falciparum 3D7 [Ref] 137

4. **Export the results to a search strategy.** Use the blue *Export as a Search Strategy* button at the top right-hand side of the results. Once exported, you can take advantage of over 100 specialized searches using the Add Step button.

Genes matching kinase (filtered by fields and organisms)

1 - 20 of 137

Filter results

Genome Genes Hide zero counts Clear filter 137

Gene - PF3D7_0616000 pyridoxal kinase
Gene name or symbol: PDXK
Gene type: protein coding gene
Organism: Plasmodium falciparum 3D7

Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all); PubMed

Export as a Search Strategy ►

My Search Strategies

Opened (1) All (261) Public (45) Help

Unnamed Search Strategy * ↗

Text 137 Genes + Add a step Step 1

137 Genes (113 ortholog groups) Revise this search

Gene Results | Genome View | Analyze Results

Organism Filter

Genes: 137 Transcripts: 138 Show Only One Transcript Per Gene Show only the Genes in my basket.

Rows per page: 50 Download Send to... Add Columns

Gene ID	Transcript ID	Organism	Interpro Description
PF3D7_0102600	PF3D7_0102600.1	Plasmodium falciparum 3D7	Protein kinase domain;Tyrosine-protein kinase, active site;Protein kinase-like dom
PF3D7_0103700	PF3D7_0103700.1	Plasmodium falciparum 3D7	N/A

5. **Identify all the orthologs of the *P. falciparum* genes in all organisms in PlasmoDB.** 1. Click on the “Add a step” button to reveal a popup. 2. Select the “Transform in related Records” option. 3. then click the “Orthologs” button. 4. In the next popup, select all organisms and click on “Run Step” button.

My Search Strategy ↗

Add a step to your search strategy ↗

Opened (1) All (169) Public (45) Help

Unnamed Search Strategy *

Text 137 Genes + Add a step Step 1

137 Genes (113 ortholog groups)

Organism Filter

Combine with other Genes

Transform into related records

Use Genomic Colocation to combine with other features

Transform 137 Genes into...

Orthologs
Metabolic Pathways
Compounds

Add a step to your search strategy ↗

Your Genes from Step 1 will be converted into Orthologs

Configure Search Learn More View Data Sets Used

Organism Haemoproteidae Plasmodidae 62 selected, out of 62

Synteny Orthologs Only? Run Step

- Notice you now have a second step with the results of all orthologs in PlasmoDB.
- The step that is highlighted in yellow indicates what results you are viewing.
- You can expand the organisms on the left to see the distribution of orthologs across organisms.
- The central portion includes a list of all the gene results.

Unnamed Search Strategy *

The screenshot shows the PlasmoDB search interface. At the top, there's a search bar labeled "Unnamed Search Strategy *". Below it, a search history section shows "Text 137 Genes" and "Orthologs 7,325 Genes" (highlighted in yellow), with a "+ Add a step" button. Step 1 is "Text 137 Genes" and Step 2 is "Orthologs 7,325 Genes".

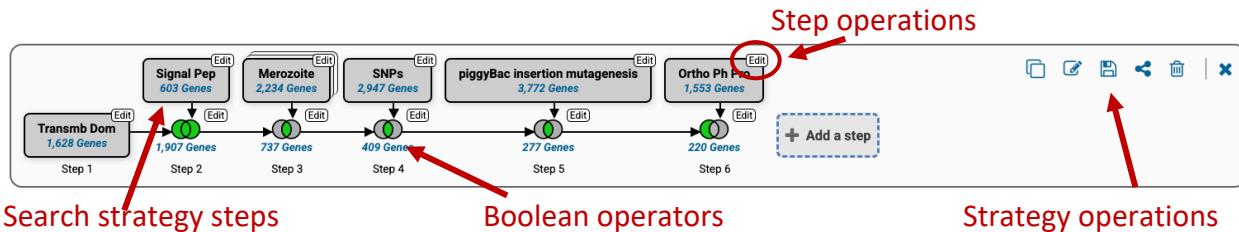
The main area displays search results for "7,325 Genes (112 ortholog groups)". There are three tabs: "Gene Results" (selected), "Genome View", and "Analyze Results". The results table has columns for Gene ID, Transcript ID, Organism, and Genomic Location (Gene). The table lists various Plasmodium species and their orthologs, such as Plasmodium fragile strain nilgiri, Plasmodium inui San Antonio 1, and Plasmodium brasiliannum strain Bolivian I. The results are paginated from 1 to 74, with 100 rows per page. A "Download" button and a "Send to..." dropdown are also present.

Organism Filter:

- select all | clear all | expand all | collapse all
- Hide zero counts | Reference only
- Search organisms...
- Haemoproteidae 82
 - Haemoproteus tartakovskyi strain SISKIN1 [Ref]
- Plasmodiidae 7,243
 - Hepatophilus sp. ex *Plasmodium tephrocytis* 2019 [Ref]
 - Plasmodium 7,129
 - Plasmodium adleri G01 [Ref] 139
 - Plasmodium berghei ANKA [Ref] 104
 - Plasmodium bilicollinsi G01 [Ref] 129
 - Plasmodium blacklocki G01 [Ref] 131
 - Plasmodium brasiliannum strain Bolivian I [Ref]
 - Plasmodium chabaudii chabaudi [Ref] 105
 - Plasmodium coatneyi Hacketti [Ref] 110
 - Plasmodium cynomolgi 228
 - Plasmodium falciparum 2,621
 - Plasmodium fragile strain nilgiri [Ref] 113
 - Plasmodium gaboni 265
 - Plasmodium gallinaceum 8A [Ref]
 - Plasmodium inui San Antonio 1 [Ref] 109
 - Plasmodium knowlesi 331
 - Plasmodium malariae UG01 [Ref] 107
 - Plasmodium ovale 213
 - Plasmodium praefalciparum strain G01 [Ref] 134
 - Hide Organism Filter

Search Strategies

Search strategies in VEuPathDB resources allow you to combine results from different datatype searches using Boolean operators (e.g. intersect, union, minus). Search strategies enable you to develop in silico experiments based on data from the species of interest or from other species (or strains) by leveraging orthology.



Getting started with your first search strategy

There are a few things to consider before developing a search strategy:

1. What is your question? Or what are you trying to find out? (overall strategy)
2. Can you break down your question into smaller components? (strategy steps)
3. What data or analyses can be used to answer the various components of your main question?
4. How will you combine the different components of your question? Ie. Which Boolean operators.

Example question

Big question: I would like to identify possible *Plasmodium* vaccine candidates.

Let's break it down:

1. What are the characteristics of a *Plasmodium* protein that might be a good vaccine candidate?
 - a. Secreted protein (proteins with a predicted secretory signal peptide).
 - b. Membrane proteins (proteins with transmembrane domains).
 - c. Expressed at a specific developmental stage (transcriptomics and/or proteomics).
 - d. What about homology with host proteins? (proteins with specific phyletic patterns).

Let's break it down:

1. Explore the types of searches available in PlasmoDB. What searches can identify genes that are secreted or membrane-bound? Expand the menu on the left-hand side of the home page and look at the available searches. Hint: try filtering the searches with a keyword like “signal peptide,” “secreted,” or “transmembrane.”

The screenshot shows the PlasmoDB homepage. On the left, a search bar contains the query "signal". Below it, a expanded search menu under "Genes" lists various categories such as Annotation, curation and identifiers, Epigenomics, Function prediction, Gene models, Genetic variation, Genomic Location, Immunology, Orthology and synteny, Pathways and interactions, Phenotype, Protein features and properties, Protein targeting and localization, Proteomics, Sequence analysis, Structure analysis, Taxonomy, Text, and Transcriptomics. In the center, a section titled "Overview of Resources and Tools" includes links for Take a Tour, Subscription, User Survey Results, Funding, Getting Started, Search Strategies, Genome Browser, Transcriptomic Resources, Phenotypic Data, and Downloads. A "Subscription" section discusses the implementation of a subscription service starting in March 2025. On the right, a "Tutorials and Exercises" section offers links to Apollo: Manual gene, Finding old Gene IDs, Gene Pages, and Genetic Variation.

2. Start by searching for all genes in *P. falciparum* 3D7 with a predicted secretory signal peptide.

Identify Genes based on Predicted Signal Peptide

The screenshot shows the "Configure Search" interface. At the top, there are links for "Learn More" and "View Data Sets Used". Below is a "Reset values to default" button. The "Organism" section shows "1 selected, out of 62" with options to "select only these", "add these", or "clear these". A dropdown menu for "3D7" is open, showing categories: Plasmodiidae, Plasmodium, Plasmodium falciparum, and Plasmodium falciparum 3D7 [Reference]. The "Reference only" checkbox is unchecked. To the right, a search strategy panel titled "Unnamed Search Strategy *" displays a single step: "Signal Pep 603 Genes". A red arrow points from this step to a large red oval around the "Get Answer" button below. The "Get Answer" button is highlighted with a red circle. At the bottom, there are links for "Build a Web Services URL from this Search >>" and two optional fields: "Give this search a name (optional)" and "Give this search a weight (optional)".

3. Expand your strategy by adding a step of all genes with at least one transmembrane domain. Think about how you will combine the results of the transmembrane domain search and your previous signal peptide results.

The screenshot shows the PlasmoDB interface with the following steps:

- Top Right:** A modal window titled "Add a step to your search strategy" is open. It has two tabs: "Combine with other Genes" (selected) and "Transform into related records". Under "Combine with other Genes", there are three options: "1 INTERSECT 2" (radio button), "1 UNION 2" (radio button, circled in red), "1 MINUS 2", and "2 MINUS 1". Below this, it says "Choose which Genes to combine. From..." with radio buttons for "A new search" (selected), "An existing strategy", and "My basket". A search bar contains "transm" and "Protein targeting and localization Transmembrane Domain Count".
- Bottom Left:** A separate search interface for "Search for Genes by Transmembrane Domain Count" is shown. It displays "603 Genes" and a search bar with "transm" and "Protein targeting and localization Transmembrane Domain Count".
- Bottom Center:** The main search strategy interface titled "Unnamed Search Strategy *". It shows a flowchart: "Signal Pep 603 Genes" (Step 1) → "Transmb Dom 1,628 Genes" (Step 2) → "1,907 Genes" (Step 3). Step 2 has an "Edit" button above it. Step 3 has an "Add a step" button to its right. Below the flowchart, there are fields for "Organism" (selected: Plasmodium), "Minimum Number of Transmembrane Domains" (1), and "Maximum Number of Transmembrane Domains" (99). A "Run Step" button is highlighted with a red box.

4. Add another step and search for the RNAseq data.

The screenshot shows the PlasmoDB interface with the following steps:

- Top Right:** A modal window titled "Add a step to your search strategy" is open. It has two tabs: "Combine with other Genes" (selected) and "Transform into related records". Under "Combine with other Genes", there are four options: "2 INTERSECT 3" (radio button, circled in blue), "2 UNION 3", "2 MINUS 3", and "3 MINUS 2". Below this, it says "Choose which Genes to combine. From..." with radio buttons for "A new search" (selected), "An existing strategy", and "My basket". A search bar contains "RNase" and "Gene models Unannotated Intron Junctions Transcriptomics RNA-Seq Evidence".
- Bottom Left:** Three search strategies are listed:
 - Combine with other Genes:** Shows "Transmb Dom 1,628 Genes" (Step 2) → "1,907 Genes" (Step 3).
 - Transform into related records:** Shows "Transmb Dom 1,628 Genes" (Step 2) → "1,907 Genes" (Step 3).
 - Use Genomic Colocation to combine with other features:** Shows "Transmb Dom 1,628 Genes" (Step 2) → "1,907 Genes" (Step 3).

Notice that there are different data sets and different types of searches for each dataset:

FC: Fold change search allows you to find genes with expression profiles by comparing experimental samples.

DE: Differential expression is available for experiments with biological replicates, enabling comparisons with statistical power.

P: Percentile search allows you to identify genes expressed at the desired expression level compared to all other genes in the genome.

SA: Sense anti-sense searches are available for strand-specific data, enabling exploration of antisense expression compared to sense expression.

For this step, select the percentile search for the *P. falciparum* experiment: “Intraerythrocytic cycle transcriptome (3D7)”.

Configure this search to find all genes highly expressed in the merozoite stages (40-5 hours post-invasion).

Add a step to your search strategy 

Search for Genes by RNA-Seq Evidence

The results will be  intersected with  the results of Step 2.

Legend:  Similarity  Differential Expression  Fold Change  WGCNA  Percentile  SenseAntisense

Filter Data Sets:  14 results (filtered from a total of 58)

Organism	Data Set	Choose a Search
Plasmodium chabaudi chabaudi	IDC in SR10KO and WT (2020) (Subudhi et al. 2020)	   
Plasmodium falciparum 3D7	Intraerythrocytic development cycle transcriptome (2019) (Wichers et al. 2019)	  
Plasmodium falciparum 3D7	Intraerythrocytic development cycle transcriptome (2018) (Toenhaeke et al.)	 
Plasmodium falciparum 3D7	Transcriptome during intraerythrocytic development (Bartfai et al.)	 
Plasmodium falciparum 3D7	Blood stage transcriptome (3D7) (Otto et al.)	 
Plasmodium falciparum 3D7	Intraerythrocytic development cycle transcriptome by UTR-Seq (2020) (Chappell et al. 2020)	  
Plasmodium falciparum 3D7	IDC in constant temperature and darkness (2020) (Subudhi et al. 2020)	   
Plasmodium falciparum 3D7	Intraerythrocytic cycle transcriptome (3D7) (Hoeljmakers et al.)	   
Plasmodium falciparum 3D7	Intraerythrocytic development cycle transcriptome by DAFT-Seq (3D7, HB3, IT, 2020) (Chappell et al. 2020)	   

Add a step to your search strategy 

Fold Change **Percentile** SenseAntisense

Configure Search Learn More View Data Sets Used

Reset values to default

Experiment

Intraerythrocytic cycle transcriptome (3D7) - Sense
 Intraerythrocytic cycle transcriptome (3D7) - Antisense

Samples

40-5 hours post-invasion
 2-10 hours post-invasion
 7-15 hours post-invasion
 12-20 hours post-invasion
 17-25 hours post-invasion
 22-30 hours post-invasion
 27-35 hours post-invasion
 32-40 hours post-invasion
 select all clear all

Minimum expression percentile

Maximum expression percentile

Matches Any or All Selected Samples? 

Protein Coding Only: 

Run Step



5. How many of your genes do not have orthologs in mammals? To find out, add a step, then run the “Orthology Phylogenetic Profile” search. Configure this search to find all genes in PlasmoDB that do not have orthologs in mammals.

Find genes in these organisms

1 selected, out of 62
 select only these | add these | clear these

3D7 Reference only

- Plasmodiidae
- Plasmodium
- Plasmodium falciparum
- Plasmodium falciparum 3D7 [Reference]

Select orthology profile

Click on to determine which organisms to include or exclude in the orthology profile.
 (= no constraints | = must be in group | = must not be in group | = mixture of constraints)

mam

All Organisms

- * Eukaryota (EUKA)
- * Metazoa (META)
 - * Chordata (CHOR)
 - Mammalia (MAMM)
 - Bos taurus breed Hereford (btau)
 - Canis lupus familiaris isolate S1007034 (clup)
 - Cavia porcellus 2N (cpor)
 - Equus caballus (Horse) (ecab)
 - Gorilla gorilla gorilla (Western lowland gorilla) (ggor)
 - Homo sapiens REF (hsap)
 - Loxodonta africana (African elephant) (lafr)
 - Monodelphis domestica (Gray short-tailed opossum) (mdom)
 - Macaca fascicularis REF (mfas)
 - Myotis lucifugus (Little brown bat) (mluc)
 - Macaca mulatta isolate AG07107 (mmua)
 - Macaca mulatta isolate 17573 (mmu1)
 - Mus musculus C57BL6J (mmus)
 - Myotis myotis mMyoMyo1 (mmyo)
 - Ornithorhynchus anatinus (Duckbill platypus) (oana)
 - Phyletor macrocephalus (Sperm whale) (Physeter catodon) (phma)
 - Pan troglodytes (Chimpanzee) (ptrc)
 - Rattus norvegicus BN/NHsdMcwi (rnor)

Run Step

My Search Strategies

Opened (1) All (170) Public (40) Help

Possible Pf3D7 vaccine candidates *

Step 1: Translmb Dom (1,628 Genes)
 Step 2: Signal Pept (603 Genes) 1,907 Genes
 Step 3: PF3D7 IRBC cycle RNA-Seq (%ile) (1,109 Genes) 378 Genes
 Ortho Ph Pro (3,163 Genes) 272 Genes

+ Add a step

272 Genes (241 ortholog groups)

Analyze Results

Genes: 272 Transcripts: 276 Show Only One Transcript Per Gene Show only the Genes in my basket.

Gene ID	Transcript ID	Genomic Location (Gene)	Product Description	Gene Name or Symbol
PF3D7_0102500	PF3D7_0102500.1	PF3D7_01_v3:109,005..117,516(-)	erythrocyte binding antigen-181	EBA181
PF3D7_0102700	PF3D7_0102700.1	PF3D7_01_v3:124,394..126,213(+)	merozoite-associated tryptophan-rich antigen	MaTrA

Analyze your Gene results with a tool below.

Gene Ontology Enrichment

Metabolic Pathway Enrichment

kinase phosphatase exported membrane

Word Enrichment

Explore these results – do these make sense? Are any functions enriched in these results? You can find this out by running a GO enrichment analysis:

- Click on the Analyze Results tab.
- Click on the GO enrichment option.
- Select a GO category to enrich for (e.g., Biological process, cellular component, or molecular function).

Gene Ontology Enrichment

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

▼ Parameters

<p>Organism ? <input type="text" value="Plasmodium falciparum 3D7"/></p> <p>Ontology ? <input checked="" type="radio"/> Biological Process <input type="radio"/> Cellular Component <input type="radio"/> Molecular Function</p> <p>Evidence ? <input checked="" type="checkbox"/> Computed <input checked="" type="checkbox"/> Curated select all clear all</p> <p>Limit to GO Slim terms ? <input checked="" type="radio"/> No <input type="radio"/> Yes</p> <p>P-Value cutoff ? <input type="text" value="0.05"/> (0 - 1)</p>	<p>Analysis Results:</p> <p>19 rows Open in Revigo Show Word Cloud Download</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">GO ID ?</th> <th style="width: 20%;">GO Term ?</th> <th style="width: 10%;">Genes in the bgd with this term ?</th> <th style="width: 10%;">Genes in your result with this term ?</th> <th style="width: 10%;">Percent of bgd genes in your result ?</th> <th style="width: 10%;">Fold enrichment ?</th> <th style="width: 10%;">Odds ratio ?</th> <th style="width: 10%;">P-value ?</th> <th style="width: 10%;">Benjamini ?</th> <th style="width: 10%;">Bonferroni ?</th> </tr> </thead> <tbody> <tr> <td>G0:0044409</td> <td>entry into host</td> <td>108</td> <td>38</td> <td>35.2</td> <td>6.13</td> <td>10.30</td> <td>2.90e-21</td> <td>8.87e-19</td> <td>8.87e-19</td> </tr> <tr> <td>G0:0052126</td> <td>movement in host environment</td> <td>127</td> <td>40</td> <td>31.5</td> <td>5.49</td> <td>8.77</td> <td>2.57e-20</td> <td>3.93e-18</td> <td>7.85e-18</td> </tr> <tr> <td>G0:0044403</td> <td>biological process involved in symbiotic interaction</td> <td>407</td> <td>53</td> <td>13.0</td> <td>2.27</td> <td>2.84</td> <td>3.87e-9</td> <td>3.54e-7</td> <td>1.19e-6</td> </tr> <tr> <td>G0:0044419</td> <td>biological process involved in interspecies interaction between organisms</td> <td>409</td> <td>53</td> <td>13.0</td> <td>2.26</td> <td>2.82</td> <td>4.63e-9</td> <td>3.54e-7</td> <td>1.42e-6</td> </tr> </tbody> </table>	GO ID ?	GO Term ?	Genes in the bgd with this term ?	Genes in your result with this term ?	Percent of bgd genes in your result ?	Fold enrichment ?	Odds ratio ?	P-value ?	Benjamini ?	Bonferroni ?	G0:0044409	entry into host	108	38	35.2	6.13	10.30	2.90e-21	8.87e-19	8.87e-19	G0:0052126	movement in host environment	127	40	31.5	5.49	8.77	2.57e-20	3.93e-18	7.85e-18	G0:0044403	biological process involved in symbiotic interaction	407	53	13.0	2.27	2.84	3.87e-9	3.54e-7	1.19e-6	G0:0044419	biological process involved in interspecies interaction between organisms	409	53	13.0	2.26	2.82	4.63e-9	3.54e-7	1.42e-6
GO ID ?	GO Term ?	Genes in the bgd with this term ?	Genes in your result with this term ?	Percent of bgd genes in your result ?	Fold enrichment ?	Odds ratio ?	P-value ?	Benjamini ?	Bonferroni ?																																										
G0:0044409	entry into host	108	38	35.2	6.13	10.30	2.90e-21	8.87e-19	8.87e-19																																										
G0:0052126	movement in host environment	127	40	31.5	5.49	8.77	2.57e-20	3.93e-18	7.85e-18																																										
G0:0044403	biological process involved in symbiotic interaction	407	53	13.0	2.27	2.84	3.87e-9	3.54e-7	1.19e-6																																										
G0:0044419	biological process involved in interspecies interaction between organisms	409	53	13.0	2.26	2.82	4.63e-9	3.54e-7	1.42e-6																																										

The strategy developed in this exercise may be accessed using this link:

<https://plasmodb.org/plasmo/app/workspace/strategies/import/df16787b3f9044db>

A variation of the above strategy that leverages additional data may be accessed using this link:

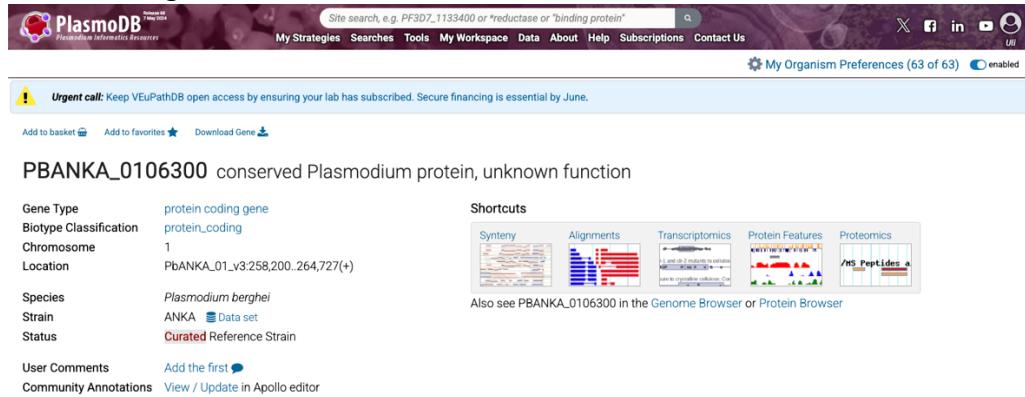
<https://plasmodb.org/plasmo/app/workspace/strategies/import/66b5340817f0b0ed>

Gene Pages

Gene pages provide an encyclopedic view of all the information available in the database on a gene-by-gene basis.

1. Visit the gene page for PBANKA_0106300:
https://plasmodb.org/plasmo/app/record/gene/PBANKA_0106300

2. What is the gene called?



PBANKA_0106300 conserved Plasmodium protein, unknown function

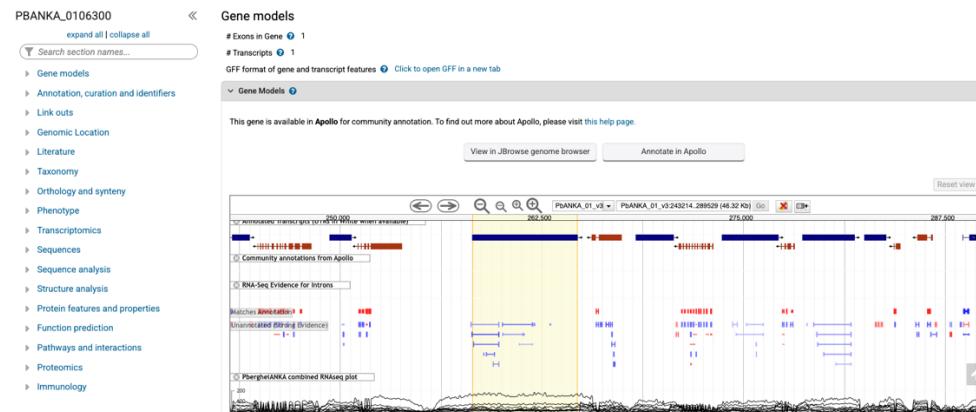
Gene Type: protein coding gene
Biotype Classification: protein_coding
Chromosome: 1
Location: PbANKA_01_v3:258,200-264,727(+)
Species: *Plasmodium berghei*
Strain: ANKA (Data set)
Status: Curated Reference Strain
User Comments: Add the first...
Community Annotations: View / Update in Apollo editor

Shortcuts

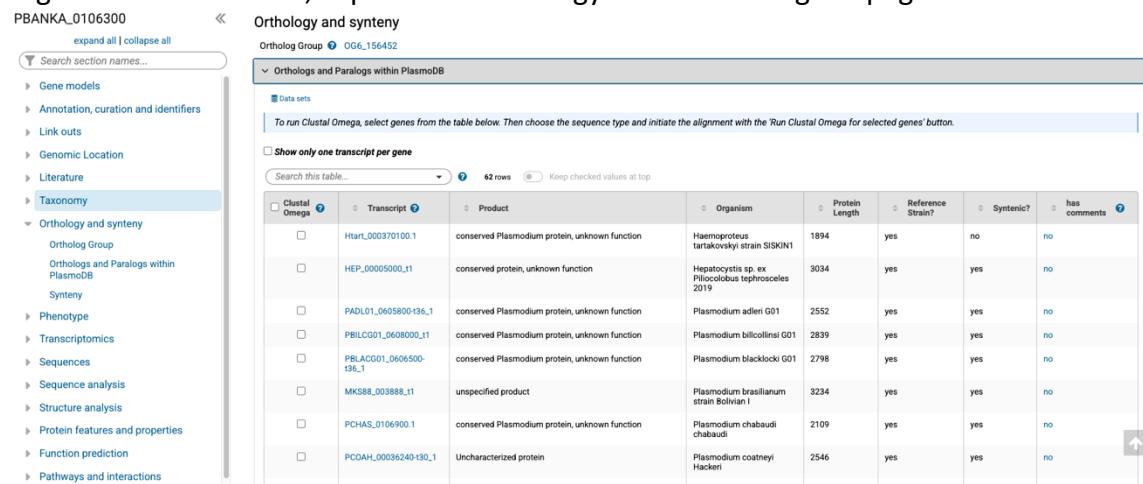
Synteny Alignments Transcriptomics Protein Features Proteomics MS Peptides a

Also see PBANKA_0106300 in the Genome Browser or Protein Browser

3. How many exons does this gene have? To find out, explore the gene model section.



4. Does this gene have orthologs across organisms in PlasmoDB? What is it called in other organisms? To find out, explore the orthology section of the gene page.



5. Explore the protein features and properties section of the gene page. Does this gene have any InterPro domains? Do you think the product description of this gene should be changed?

PBANKA_0106300

[expand all](#) | [collapse all](#)

[Search section names...](#)

- [PlasmoDB](#)
- [Synteny](#)
- [Phenotype](#)
- [Transcriptomics](#)
- [Sequences](#)
- [Sequence analysis](#)
- [Structure analysis](#)
- [Protein features and properties](#)
 - [Attributes and Protein Browser](#)
 - [BLASTP \(protein-protein BLAST\)](#)
 - [GPI anchor prediction: big-PI Predictor](#)
 - [InterPro Domains](#)
 - [InterProScan: Run on EBI site](#)
 - [MitoProt](#)
 - [STRING: functional protein association networks](#)
 - [WoLF PSORT](#)

Protein features and properties

[Attributes and Protein Browser](#)

[BLASTP \(protein-protein BLAST\)](#)

[GPI anchor prediction: big-PI Predictor](#)

[InterPro Domains](#)

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

The InterPro family of databases includes CATH, CCD, HAMAP, PANTHER, Pfam, PIRSF, PRINTS, PRODOM, PROFILE, PROSITE, SFLD, SMART, SUPERFAMILY, TIGRFAMS. The InterPro database used may not be the latest version; to view the version used, press 'Data sets'. To run an analysis using the current version, see the InterProScan section below.

[Search this table...](#) 5 rows

Name	Interpro ID	Primary ID	Secondary ID	Description	Start Min	End Min	E-value	Transcript ID(s)
CDD	N/A	cd08824	LOTUS	N/A	1142	1195	N/A	PBANKA_0106300.1
InterProScan	N/A	IPR036465	vWFA_dom_sf	von Willebrand factor A-like domain superfamily	1048	1133	N/A	PBANKA_0106300.1
InterProScan	N/A	IPR025677	OSTHTH-assoc_dom	OSTHTH associated domain	1055	1125	N/A	PBANKA_0106300.1
Pfam	IPR025677	PF14418	OHA	OSTHTH associated domain	1055	1125	2.3E-19	PBANKA_0106300.1
SUPERFAMILY	IPR036465	SSF53300	vWA-like	von Willebrand factor A-like domain superfamily	1048	1133	4.0E-05	PBANKA_0106300.1

6. Explore the transcriptomics section of the gene page. Notice there are two types of transcriptomic data sets available: bulk and single-cell RNASeq.

PBANKA_0106300

[expand all](#) | [collapse all](#)

[Search section names...](#)

- [Annotation, curation and identifiers](#)
- [Link outs](#)
- [Genomic Location](#)
- [Literature](#)
- [Taxonomy](#)
- [Orthology and synteny](#)
- [Phenotype](#)
- [Transcriptomics](#)
 - [RNA-Seq Transcription Summary](#)
 - [Transcript Expression](#)
 - [User Dataset Transcriptomics Graphs](#)
 - [Single Cell RNA-Seq \(scRNA-Seq\)](#)

Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells)

RNA-seq analyses of various exo-erythrocytic stages of *Plasmodium berghei* including sporozoites, a time-course of liver stage development and detached cells

[Caldelari et al.](#) [RNA-Seq](#)

tpm - PBANKA_0106300

X-axis
Y-axis
Transcript abundance in Transcripts per Million (TPM). The percentile graph shows the ranking of expression for this gene compared to all others in this experiment.

Choose gene for which to display graph
 PBANKA_0106300

Choose graph(s) to display
 tpm percentile

Graph options
 Show log Scale (not applicable for log(ratio) graphs, percentile graphs or data tables)

Full Dataset Description

View in genome browser

[Coverage](#)

PBANKA_0106300

[expand all](#) | [collapse all](#)

[Search section names...](#)

- [Annotation, curation and identifiers](#)
- [Link outs](#)
- [Genomic Location](#)
- [Literature](#)
- [Taxonomy](#)
- [Orthology and synteny](#)
- [Phenotype](#)
- [Transcriptomics](#)
 - [RNA-Seq Transcription Summary](#)
 - [Transcript Expression](#)
 - [User Dataset Transcriptomics Graphs](#)
 - [Single Cell RNA-Seq \(scRNA-Seq\)](#)
- [Sequences](#)
- [Sequence analysis](#)
- [Structure analysis](#)

Single Cell RNA-Seq (scRNA-Seq)

[Data sets](#)

[Search this table...](#) 2 rows

Name	Summary	Attribution	Assay Type
A spatiotemporally resolved single cell atlas of the Plasmodium liver stage - Parasite Only	A spatiotemporally resolved single cell atlas of the Plasmodium liver stage: To study the liver stage of the rodent malaria parasite <i>Plasmodium berghei</i> , we molecularly characterized thousands of infected and uninfected hepatocytes in different time points and inferred their spatial coordinates, thus enabling us to characterize the host's and parasite's temporal expression programs in a zonally-controlled manner. This dataset contains only cells from <i>Plasmodium berghei</i> , subset from the host cells and reclustered.	Afriat et al.	scRNA-Seq