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

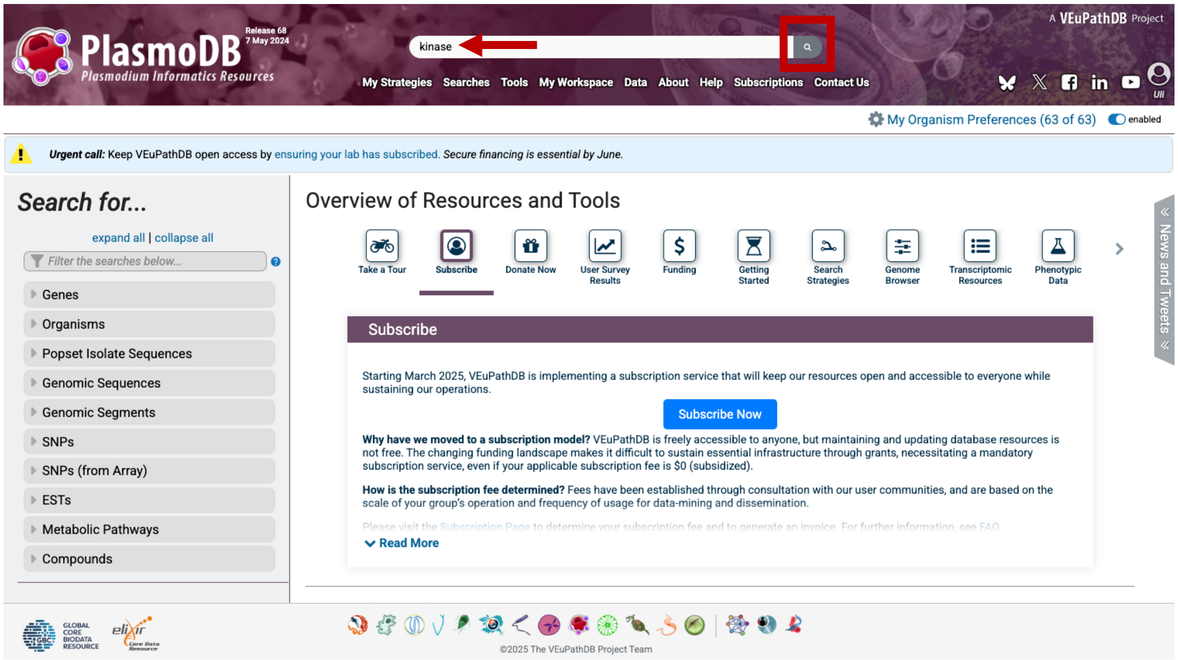
Workshop objectives:

- 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 located in the header of any VEuPathDB site and is available from 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 (arrow in the image below). Then click enter on your keyboard or click on the search icon (square in the image below).



Urgent call: Keep VEuPathDB open access by ensuring your lab has subscribed. *Secure financing is essential by June.*

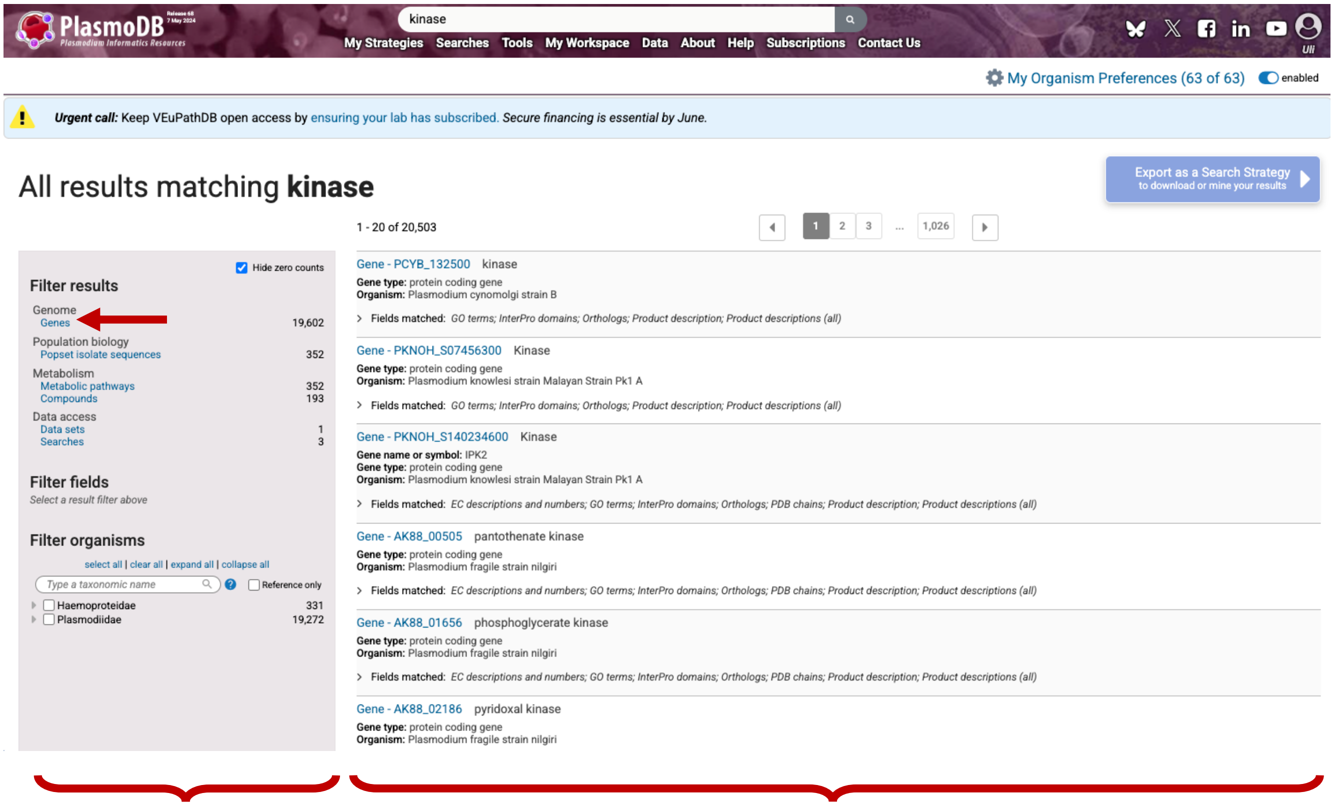
Starting March 2025, VEuPathDB is implementing a subscription service that will keep our resources open and accessible to everyone while sustaining our operations.

Why have we moved to a subscription model? VEuPathDB is freely accessible to anyone, but maintaining and updating database resources is not free. The changing funding landscape makes it difficult to sustain essential infrastructure through grants, necessitating a mandatory subscription service, even if your applicable subscription fee is $0 (subsidized).

How is the subscription fee determined? Fees have been established through consultation with our user communities, and are based on the scale of your group's operation and frequency of usage for data-mining and dissemination.

Please visit the Subscription Page to determine your subscription fee and to generate an invoice. For further information, see FAQ.

2. **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?



Urgent call: Keep VEuPathDB open access by ensuring your lab has subscribed. *Secure financing is essential by June.*

All results matching **kinase**

☑ Hide zero counts

Filter results

Genome	
Genes	19,602
Population biology	
Popset isolate sequences	352
Metabolism	
Metabolic pathways	352
Compounds	193
Data access	
Data sets	1
Searches	3

Filter fields

Select a result filter above

Filter organisms

select all | clear all | expand all | collapse all

Type a taxonomic name ☐ Reference only

☐ Haemoproteidae	331
☐ Plasmodiidae	19,272

1 - 20 of 20,503

Gene - PCYB_132500 kinase
Gene type: protein coding gene
Organism: Plasmodium cynomolgi strain B
> **Fields matched:** *GO terms; InterPro domains; Orthologs; Product description; Product descriptions (all)*

Gene - PKNOH_S07456300 Kinase
Gene type: protein coding gene
Organism: Plasmodium knowlesi strain Malayan Strain Pk1 A
> **Fields matched:** *GO terms; InterPro domains; Orthologs; Product description; Product descriptions (all)*

Gene - PKNOH_S140234600 Kinase
Gene name or symbol: IPK2
Gene type: protein coding gene
Organism: Plasmodium knowlesi strain Malayan Strain Pk1 A
> **Fields matched:** *EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)*

Gene - AK88_00505 pantothenate kinase
Gene type: protein coding gene
Organism: Plasmodium fragile strain nilgiri
> **Fields matched:** *EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)*

Gene - AK88_01656 phosphoglycerate kinase
Gene type: protein coding gene
Organism: Plasmodium fragile strain nilgiri
> **Fields matched:** *EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)*

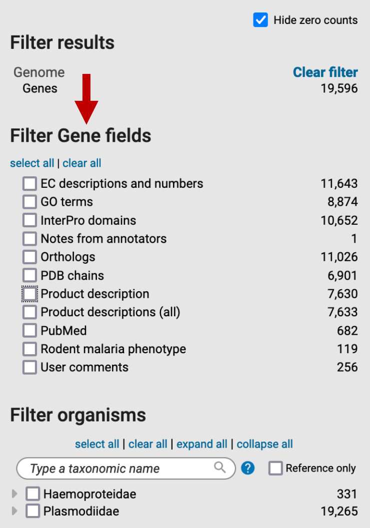
Gene - AK88_02186 pyridoxal kinase
Gene type: protein coding gene
Organism: Plasmodium fragile strain nilgiri

Results are summarized by category

Details panel with information about each item returned

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

   Filter the results so that you only view gene results (hint: click on the word *genes* in the *Filter results* section; arrow in image above) and the Filter Fields section expands to reveal additional filtering options. Select the *Product description* field and choose *Apply* (middle panel below). Once a filter is applied it can be removed by clicking on *Clear filter* (right panel below).



☑ Hide zero counts

Filter results

Genome	Clear filter
Genes	19,596

Filter Gene fields

select all | clear all

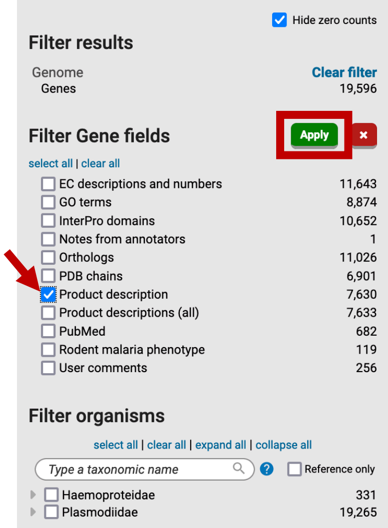
☐ EC descriptions and numbers	11,643
☐ GO terms	8,874
☐ InterPro domains	10,652
☐ Notes from annotators	1
☐ Orthologs	11,026
☐ PDB chains	6,901
☐ Product description	7,630
☐ Product descriptions (all)	7,633
☐ PubMed	682
☐ Rodent malaria phenotype	119
☐ User comments	256

Filter organisms

select all | clear all | expand all | collapse all

Type a taxonomic name ? ☐ Reference only

☐ Haemoproteidae	331
☐ Plasmodiidae	19,265



☑ Hide zero counts

Filter results

Genome	Clear filter
Genes	19,596

Filter Gene fields Apply ×

select all | clear all

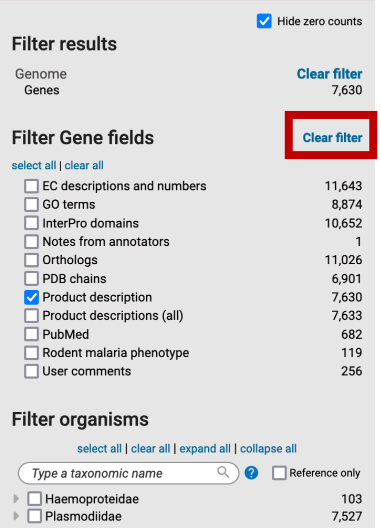
☐ EC descriptions and numbers	11,643
☐ GO terms	8,874
☐ InterPro domains	10,652
☐ Notes from annotators	1
☐ Orthologs	11,026
☐ PDB chains	6,901
☑ Product description	7,630
☐ Product descriptions (all)	7,633
☐ PubMed	682
☐ Rodent malaria phenotype	119
☐ User comments	256

Filter organisms

select all | clear all | expand all | collapse all

Type a taxonomic name ? ☐ Reference only

☐ Haemoproteidae	331
☐ Plasmodiidae	19,265



☑ Hide zero counts

Filter results

Genome	Clear filter
Genes	7,630

Filter Gene fields Clear filter

select all | clear all

☐ EC descriptions and numbers	11,643
☐ GO terms	8,874
☐ InterPro domains	10,652
☐ Notes from annotators	1
☐ Orthologs	11,026
☐ PDB chains	6,901
☑ Product description	7,630
☐ Product descriptions (all)	7,633
☐ PubMed	682
☐ Rodent malaria phenotype	119
☐ User comments	256

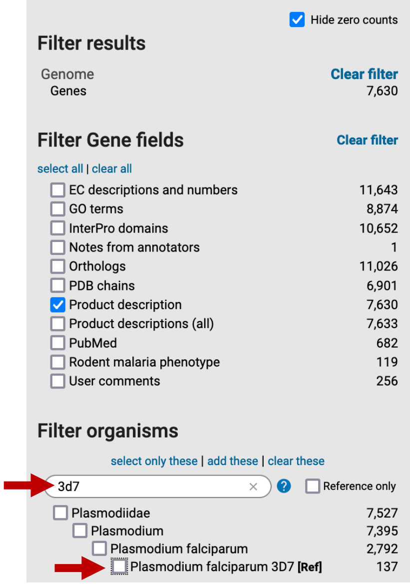
Filter organisms

select all | clear all | expand all | collapse all

Type a taxonomic name ? ☐ Reference only

☐ Haemoproteidae	103
☐ Plasmodiidae	7,527

4. **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.



☑ Hide zero counts

Filter results

Genome	Clear filter
Genes	7,630

Filter Gene fields Clear filter

select all | clear all

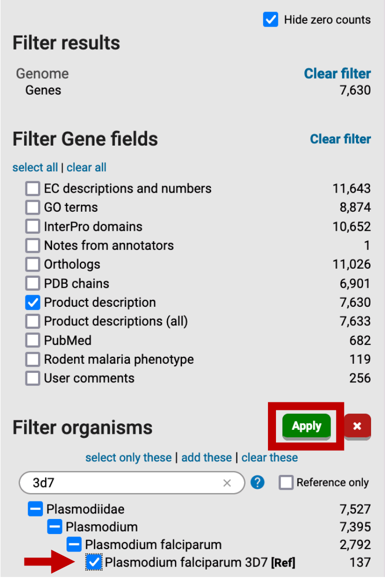
☐ EC descriptions and numbers	11,643
☐ GO terms	8,874
☐ InterPro domains	10,652
☐ Notes from annotators	1
☐ Orthologs	11,026
☐ PDB chains	6,901
☑ Product description	7,630
☐ Product descriptions (all)	7,633
☐ PubMed	682
☐ Rodent malaria phenotype	119
☐ User comments	256

Filter organisms

select only these | add these | clear these

3d7 × ? ☐ Reference only

☐ Plasmodiidae	7,527
☐ Plasmodium	7,395
☐ Plasmodium falciparum	2,792
☐ Plasmodium falciparum 3D7 **[Ref]**	137



☑ Hide zero counts

Filter results

Genome	Clear filter
Genes	7,630

Filter Gene fields Clear filter

select all | clear all

☐ EC descriptions and numbers	11,643
☐ GO terms	8,874
☐ InterPro domains	10,652
☐ Notes from annotators	1
☐ Orthologs	11,026
☐ PDB chains	6,901
☑ Product description	7,630
☐ Product descriptions (all)	7,633
☐ PubMed	682
☐ Rodent malaria phenotype	119
☐ User comments	256

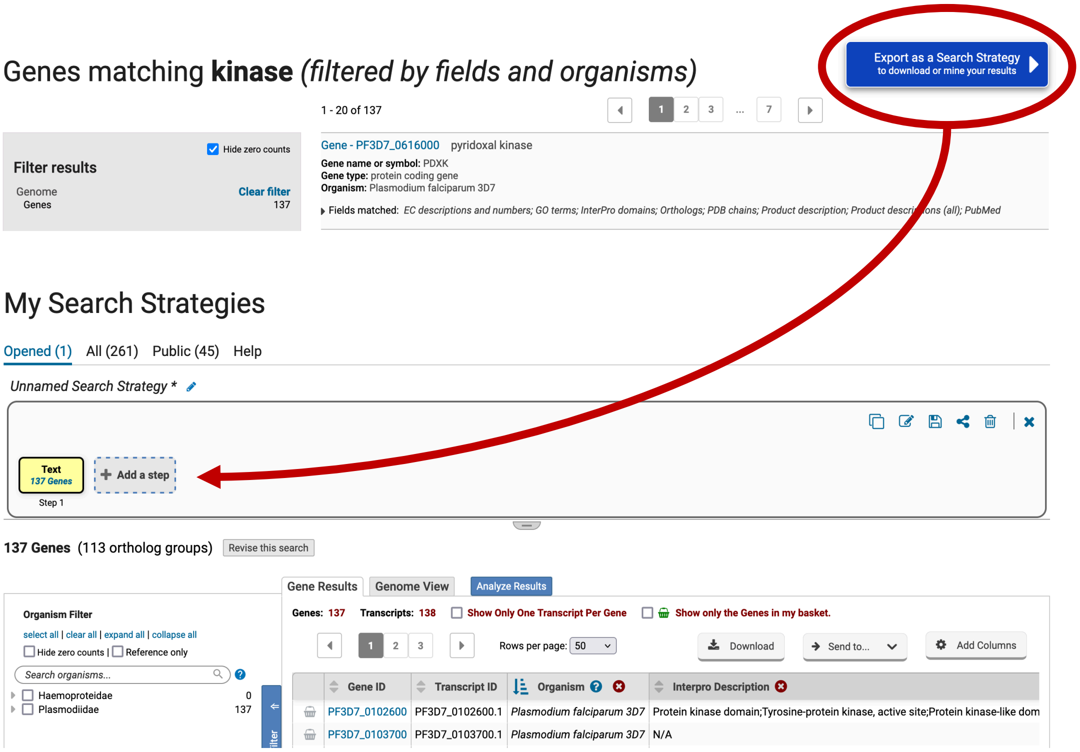
Filter organisms Apply ×

select only these | add these | clear these

3d7 × ? ☐ Reference only

☐ Plasmodiidae	7,527
☐ Plasmodium	7,395
☐ Plasmodium falciparum	2,792
☑ Plasmodium falciparum 3D7 **[Ref]**	137

5. **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 will be able to take advantage of over 100 specialized searches using the Add Step button. We will learn more about this in the next exercise/demo.



Genes matching **kinase** *(filtered by fields and organisms)*

Filter results

☑ Hide zero counts

Genome Genes 137 Clear filter

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*

My Search Strategies

*Unnamed Search Strategy **

137 Genes (113 ortholog groups)

Organism Filter

☐ Hide zero counts | ☐ Reference only

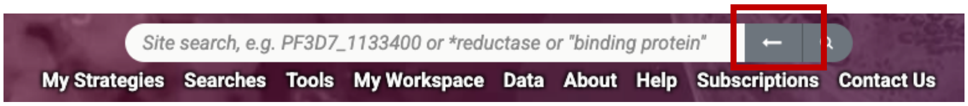
☐ Haemoproteidae 0
☐ Plasmodiidae 137

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

Rows per page: 50

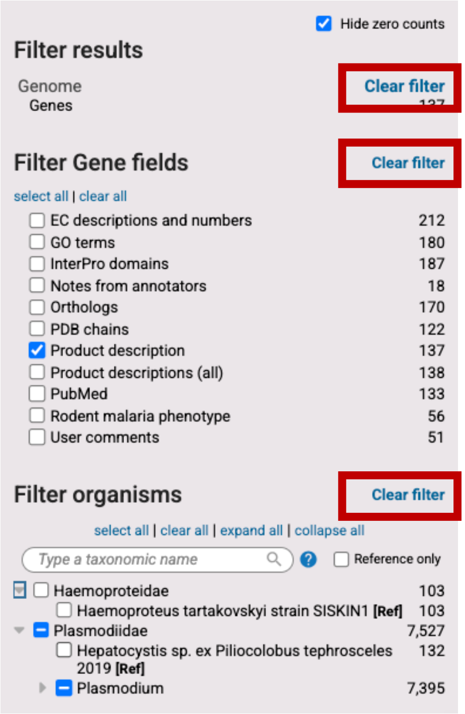
Gene ID	Transcript ID	Organism	Interpro Description
PF3D7_0102600	PF3D7_0102600.1	*Plasmodium falciparum 3D7*	Protein kinase domain;Tyrosine-protein kinase, active site;Protein kinase-like dom
PF3D7_0103700	PF3D7_0103700.1	*Plasmodium falciparum 3D7*	N/A

6. Return to the site search results page. You can achieve this in two ways: 1. Your previous results and filter settings were preserved and can be accessed by clicking on the ‘back to results’ arrow in the site search window. 2. Click on your browser’s back arrow.



7. **Clear all filters.** You can achieve this in two ways: 1. You can click on each of the clear filter options in the filter results panel (boxes below). 2. You can click on the *clear filters option* in the site search window, which serves to Clear All filters.

1



Filter results

Genome
Genes

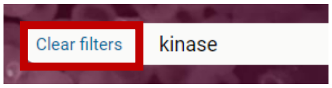
Filter Gene fields

☐ EC descriptions and numbers	212
☐ GO terms	180
☐ InterPro domains	187
☐ Notes from annotators	18
☐ Orthologs	170
☐ PDB chains	122
☑ Product description	137
☐ Product descriptions (all)	138
☐ PubMed	133
☐ Rodent malaria phenotype	56
☐ User comments	51

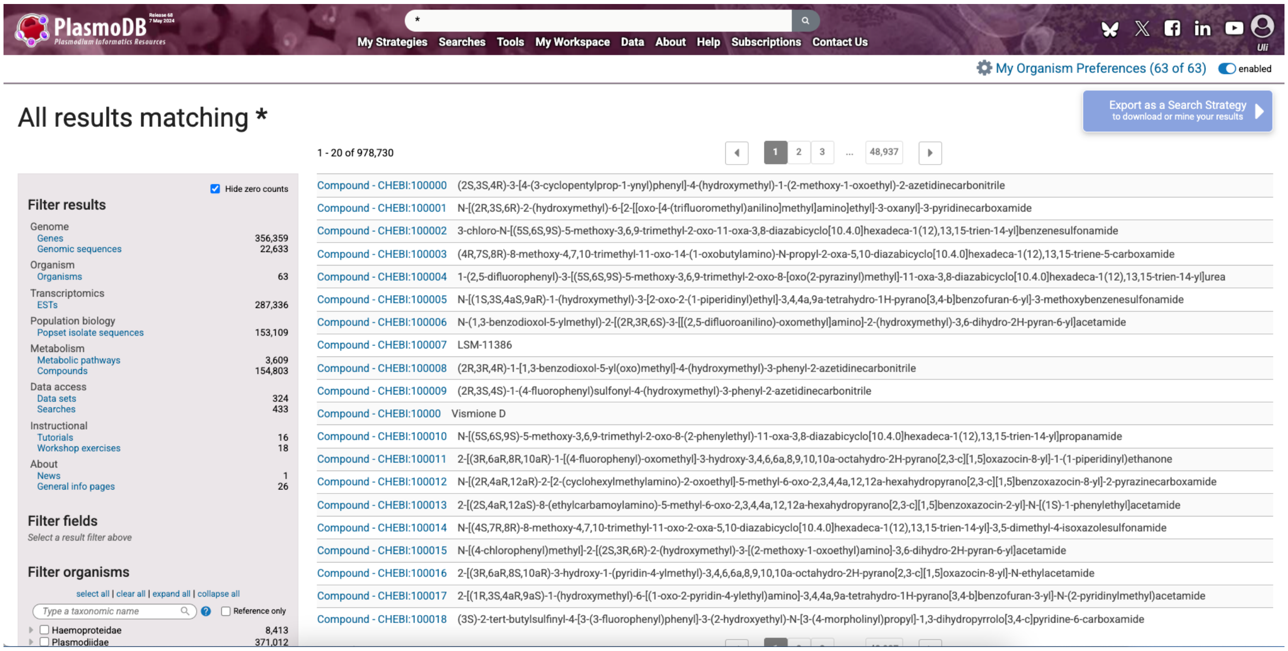
Filter organisms

☐ Haemoproteidae	103
☐ Haemoproteus tartakovskyi strain SISKIN1 **[Ref]**	103
☐ Plasmodiidae	7,527
☐ Hepatocystis sp. ex Piliocolobus tephrosceles 2019 **[Ref]**	132
☐ Plasmodium	7,395

2



8. **Run a wild card search.** The wild card (denoted by an asterisk *) can be used alone to retrieve all site search results or combined with a word such as **kinase* to retrieve compound words ending with the word kinase like phosphofructokinase. As usual results can then be explored using the filters in the *Results filter* on the left side of the website.



All results matching *

☑ Hide zero counts

Filter results

Genome	
Genes	356,359
Genomic sequences	22,633
Organism	
Organisms	63
Transcriptomics	
ESTs	287,336
Population biology	
Popset isolate sequences	153,109
Metabolism	
Metabolic pathways	3,609
Compounds	154,803
Data access	
Data sets	324
Searches	433
Instructional	
Tutorials	16
Workshop exercises	18
About	
News	1
General info pages	26

Filter fields

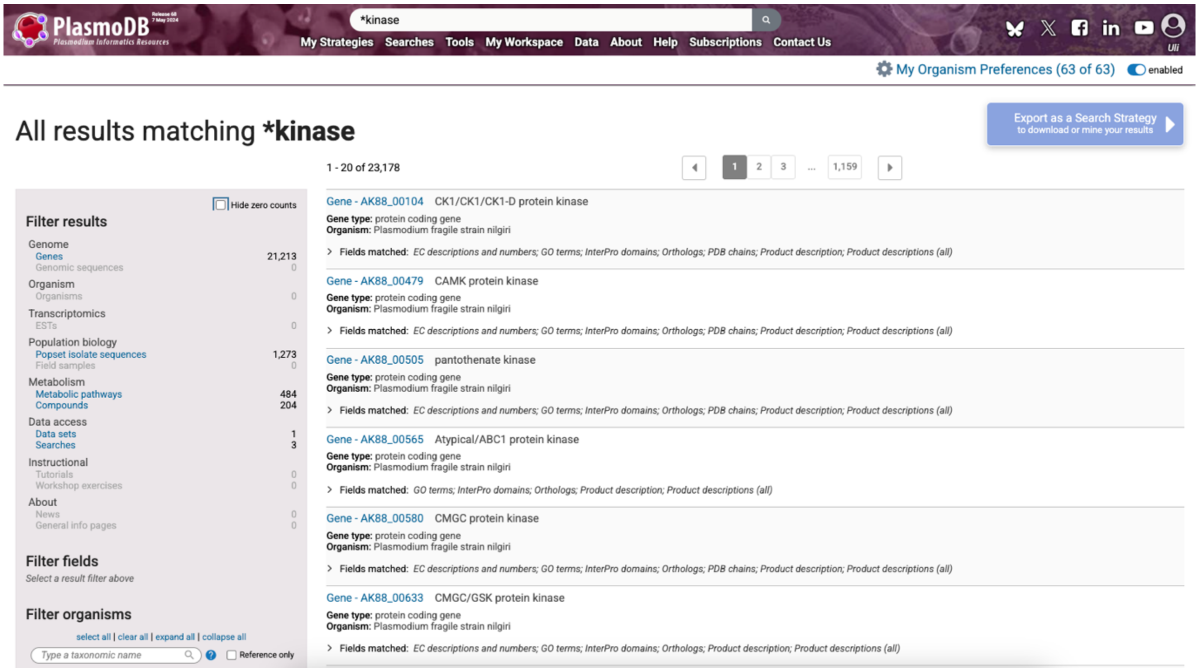
Select a result filter above

Filter organisms

☐ Haemoproteidae	8,413
☐ Plasmodiidae	371,012

1 - 20 of 978,730

Compound - CHEBI:100000	(2S,3S,4R)-3-[4-(3-cyclopentylprop-1-ynyl)phenyl]-4-(hydroxymethyl)-1-(2-methoxy-1-oxoethyl)-2-azetidinecarbonitrile
Compound - CHEBI:100001	N-[(2R,3S,6R)-2-(hydroxymethyl)-6-[2-[[oxo-[4-(trifluoromethyl)anilino]methyl]amino]ethyl]-3-oxanyl]-3-pyridinecarboxamide
Compound - CHEBI:100002	3-chloro-N-[(5S,6S,9S)-5-methoxy-3,6,9-trimethyl-2-oxo-11-oxa-3,8-diazabicyclo[10.4.0]hexadeca-1(12),13,15-trien-14-yl]benzenesulfonamide
Compound - CHEBI:100003	(4R,7S,8R)-8-methoxy-4,7,10-trimethyl-11-oxo-14-(1-oxobutylamino)-N-propyl-2-oxa-5,10-diazabicyclo[10.4.0]hexadeca-1(12),13,15-triene-5-carboxamide
Compound - CHEBI:100004	1-(2,5-difluorophenyl)-3-[(5S,6S,9S)-5-methoxy-3,6,9-trimethyl-2-oxo-8-[oxo(2-pyrazinyl)methyl]-11-oxa-3,8-diazabicyclo[10.4.0]hexadeca-1(12),13,15-trien-14-yl]urea
Compound - CHEBI:100005	N-[(1S,3S,4aS,9aR)-1-(hydroxymethyl)-3-[2-oxo-2-(1-piperidinyl)ethyl]-3,4,4a,9a-tetrahydro-1H-pyrano[3,4-b]benzofuran-6-yl]-3-methoxybenzenesulfonamide
Compound - CHEBI:100006	N-(1,3-benzodioxol-5-ylmethyl)-2-[(2R,3R,6S)-3-[[(2,5-difluoroanilino)-oxomethyl]amino]-2-(hydroxymethyl)-3,6-dihydro-2H-pyran-6-yl]acetamide
Compound - CHEBI:100007	LSM-11386
Compound - CHEBI:100008	(2R,3R,4R)-1-[1,3-benzodioxol-5-yl(oxo)methyl]-4-(hydroxymethyl)-3-phenyl-2-azetidinecarbonitrile
Compound - CHEBI:100009	(2R,3S,4S)-1-(4-fluorophenyl)sulfonyl-4-(hydroxymethyl)-3-phenyl-2-azetidinecarbonitrile
Compound - CHEBI:10000	Vismione D
Compound - CHEBI:100010	N-[(5S,6S,9S)-5-methoxy-3,6,9-trimethyl-2-oxo-8-(2-phenylethyl)-11-oxa-3,8-diazabicyclo[10.4.0]hexadeca-1(12),13,15-trien-14-yl]propanamide
Compound - CHEBI:100011	2-[(3R,6aR,8R,10aR)-1-[(4-fluorophenyl)-oxomethyl]-3-hydroxy-3,4,6,6a,8,9,10,10a-octahydro-2H-pyrano[2,3-c][1,5]oxazocin-8-yl]-1-(1-piperidinyl)ethanone
Compound - CHEBI:100012	N-[(2R,4aR,12aR)-2-[2-(cyclohexylmethylamino)-2-oxoethyl]-5-methyl-6-oxo-2,3,4,4a,12,12a-hexahydropyrano[2,3-c][1,5]benzoxazocin-8-yl]-2-pyrazinecarboxamide
Compound - CHEBI:100013	2-[(2S,4aR,12aS)-8-(ethylcarbamoylamino)-5-methyl-6-oxo-2,3,4,4a,12,12a-hexahydropyrano[2,3-c][1,5]benzoxazocin-2-yl]-N-[(1S)-1-phenylethyl]acetamide
Compound - CHEBI:100014	N-[(4S,7R,8R)-8-methoxy-4,7,10-trimethyl-11-oxo-2-oxa-5,10-diazabicyclo[10.4.0]hexadeca-1(12),13,15-trien-14-yl]-3,5-dimethyl-4-isoxazolesulfonamide
Compound - CHEBI:100015	N-[(4-chlorophenyl)methyl]-2-[(2S,3R,6R)-2-(hydroxymethyl)-3-[(2-methoxy-1-oxoethyl)amino]-3,6-dihydro-2H-pyran-6-yl]acetamide
Compound - CHEBI:100016	2-[(3R,6aR,8S,10aR)-3-hydroxy-1-(pyridin-4-ylmethyl)-3,4,6,6a,8,9,10,10a-octahydro-2H-pyrano[2,3-c][1,5]oxazocin-8-yl]-N-ethylacetamide
Compound - CHEBI:100017	2-[(1R,3S,4aR,9aS)-1-(hydroxymethyl)-6-[(1-oxo-2-pyridin-4-ylethyl)amino]-3,4,4a,9a-tetrahydro-1H-pyrano[3,4-b]benzofuran-3-yl]-N-(2-pyridinylmethyl)acetamide
Compound - CHEBI:100018	(3S)-2-tert-butylsulfinyl-4-[3-(3-fluorophenyl)phenyl]-3-(2-hydroxyethyl)-N-[3-(4-morpholinyl)propyl]-1,3-dihydropyrrolo[3,4-c]pyridine-6-carboxamide



All results matching ***kinase**

1 - 20 of 23,178

Filter results

Genome
Genes 21,213
Genomic sequences 0
Organism
Organisms 0
Transcriptomics
ESTs 0
Population biology
Popset isolate sequences 1,273
Field samples 0
Metabolism
Metabolic pathways 484
Compounds 204
Data access
Data sets 1
Searches 3
Instructional
Tutorials 0
Workshop exercises 0
About
News 0
General info pages 0

Filter fields

Select a result filter above

Filter organisms

Gene - AK88_00104 CK1/CK1/CK1-D protein kinase
Gene type: protein coding gene
Organism: Plasmodium fragile strain nilgiri
› **Fields matched:** *EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)*

Gene - AK88_00479 CAMK protein kinase
Gene type: protein coding gene
Organism: Plasmodium fragile strain nilgiri
› **Fields matched:** *EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)*

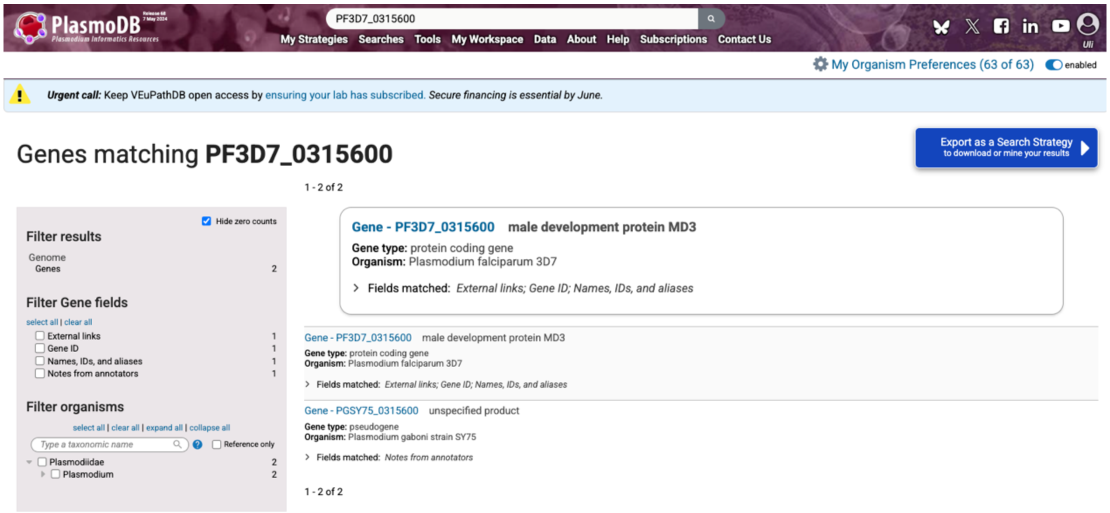
Gene - AK88_00505 pantothenate kinase
Gene type: protein coding gene
Organism: Plasmodium fragile strain nilgiri
› **Fields matched:** *EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)*

Gene - AK88_00565 Atypical/ABC1 protein kinase
Gene type: protein coding gene
Organism: Plasmodium fragile strain nilgiri
› **Fields matched:** *GO terms; InterPro domains; Orthologs; Product description; Product descriptions (all)*

Gene - AK88_00580 CMGC protein kinase
Gene type: protein coding gene
Organism: Plasmodium fragile strain nilgiri
› **Fields matched:** *EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)*

Gene - AK88_00633 CMGC/GSK protein kinase
Gene type: protein coding gene
Organism: Plasmodium fragile strain nilgiri
› **Fields matched:** *EC descriptions and numbers; GO terms; InterPro domains; Orthologs; Product description; Product descriptions (all)*

9. **Search for a specific gene ID.** Enter the gene ID in the site search window: *PF3D7_0315600*. When there is an exact match for an ID in the database, the site search offers a card in the details panel to draw attention to the direct link to the gene page.



Urgent call: Keep VEuPathDB open access by ensuring your lab has subscribed. *Secure financing is essential by June.*

Genes matching **PF3D7_0315600**

1 - 2 of 2

Filter results

Genome
Genes 2

Filter Gene fields

☐ External links 1
☐ Gene ID 1
☐ Names, IDs, and aliases 1
☐ Notes from annotators 1

Filter organisms

☐ Plasmodiidae 2
☐ Plasmodium 2

Gene - PF3D7_0315600 male development protein MD3
Gene type: protein coding gene
Organism: Plasmodium falciparum 3D7
› Fields matched: *External links; Gene ID; Names, IDs, and aliases*

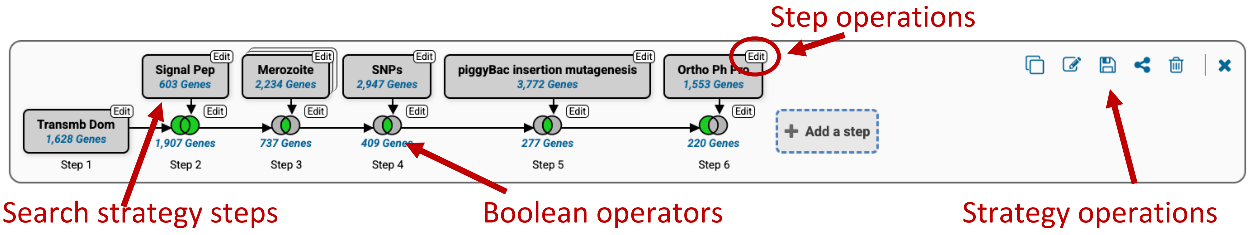
Gene - PF3D7_0315600 male development protein MD3
Gene type: protein coding gene
Organism: Plasmodium falciparum 3D7
› **Fields matched:** *External links; Gene ID; Names, IDs, and aliases*

Gene - PGSY75_0315600 unspecified product
Gene type: pseudogene
Organism: Plasmodium gaboni strain SY75
› **Fields matched:** *Notes from annotators*

1 - 2 of 2

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.



Step operations

Search strategy steps

Boolean operators

Strategy operations

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

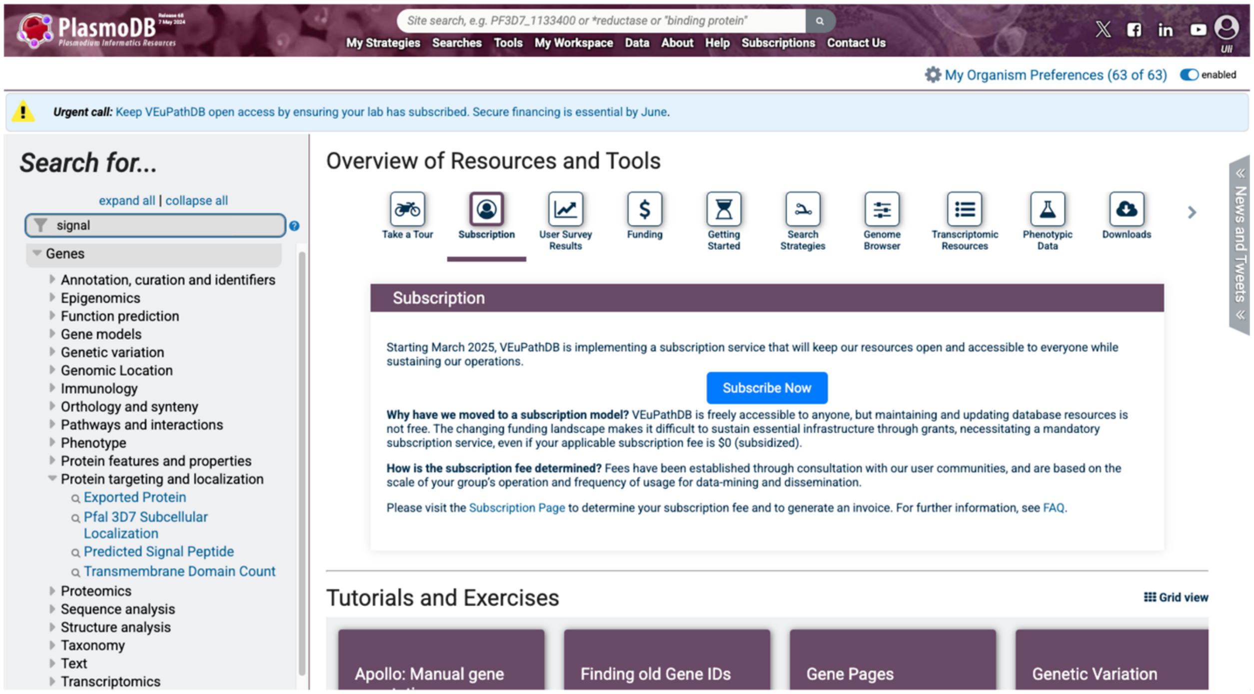
Big question: I would like to identify possible *Plasmodium* vaccine candidates.

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.”



Urgent call: Keep VEuPathDB open access by ensuring your lab has subscribed. Secure financing is essential by June.

Subscription

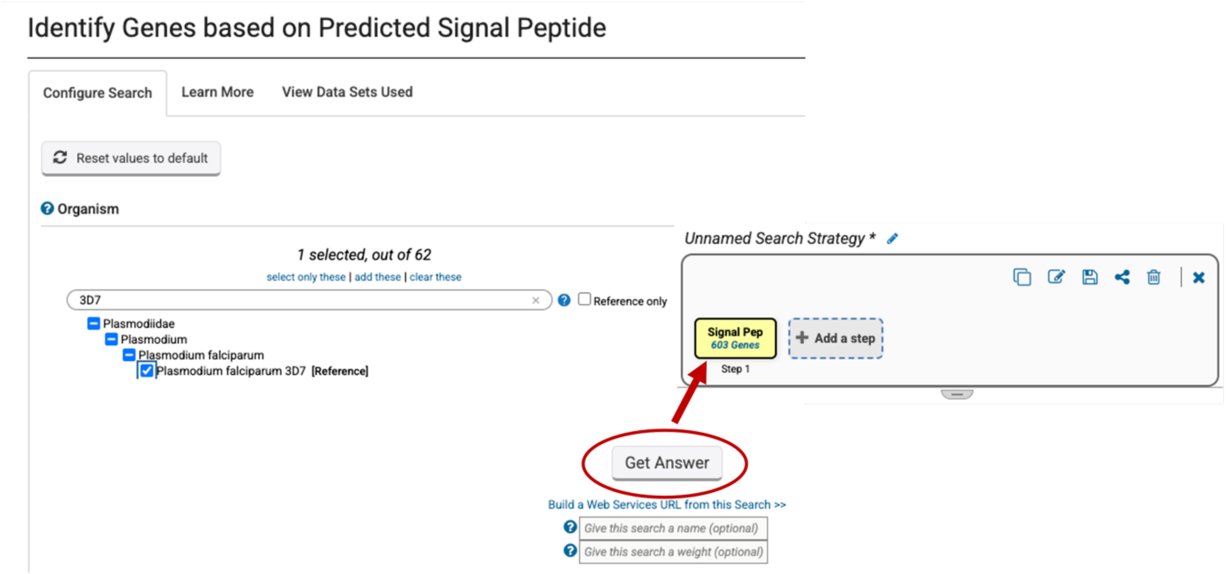
Starting March 2025, VEuPathDB is implementing a subscription service that will keep our resources open and accessible to everyone while sustaining our operations.

Why have we moved to a subscription model? VEuPathDB is freely accessible to anyone, but maintaining and updating database resources is not free. The changing funding landscape makes it difficult to sustain essential infrastructure through grants, necessitating a mandatory subscription service, even if your applicable subscription fee is $0 (subsidized).

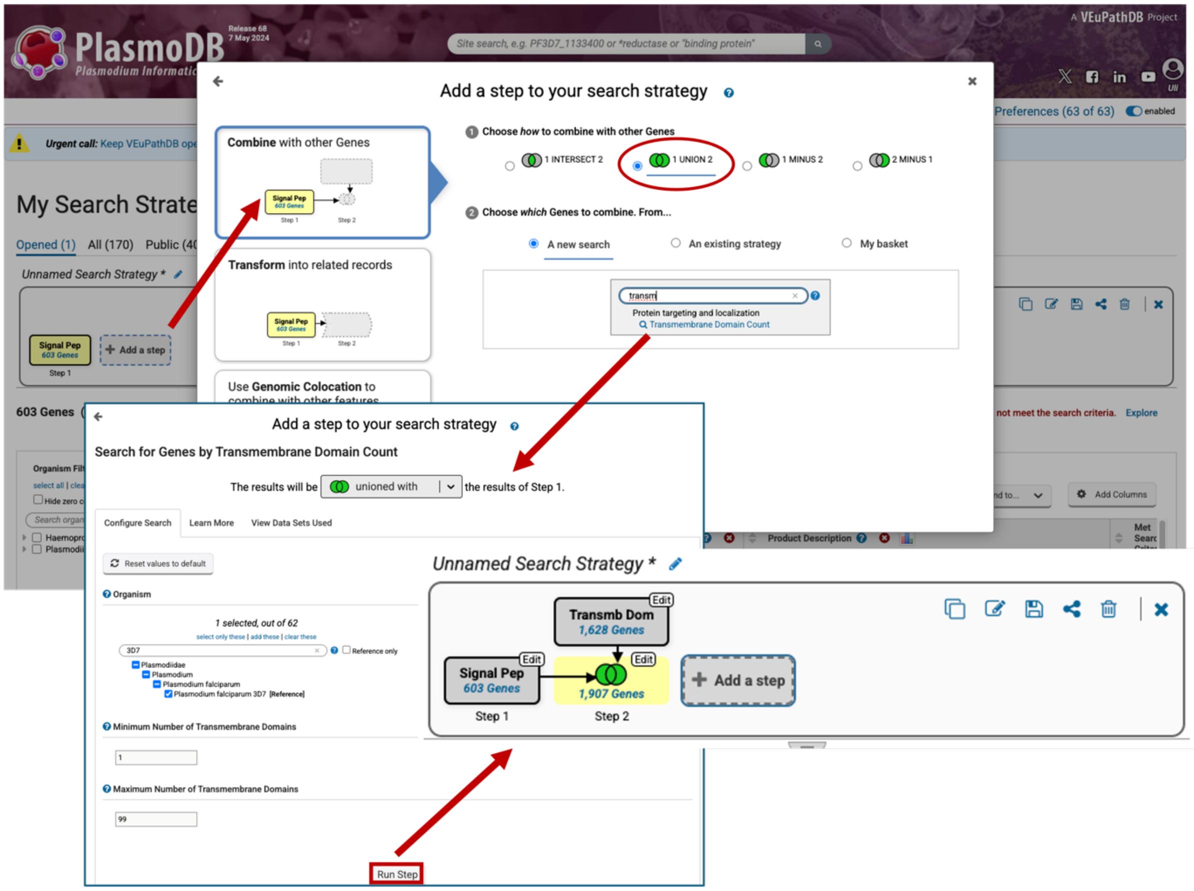
How is the subscription fee determined? Fees have been established through consultation with our user communities, and are based on the scale of your group’s operation and frequency of usage for data-mining and dissemination.

Please visit the Subscription Page to determine your subscription fee and to generate an invoice. For further information, see FAQ.

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



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.



Add a step to your search strategy

1. Choose *how* to combine with other Genes

○ 1 INTERSECT 2 ● 1 UNION 2 ○ 1 MINUS 2 ○ 2 MINUS 1

2. Choose *which* Genes to combine. From...

● A new search ○ An existing strategy ○ My basket

transm

Protein targeting and localization
Transmembrane Domain Count

Add a step to your search strategy

Search for Genes by Transmembrane Domain Count

The results will be unioned with the results of Step 1.

Organism

1 selected, out of 62

3D7 ☐ Reference only

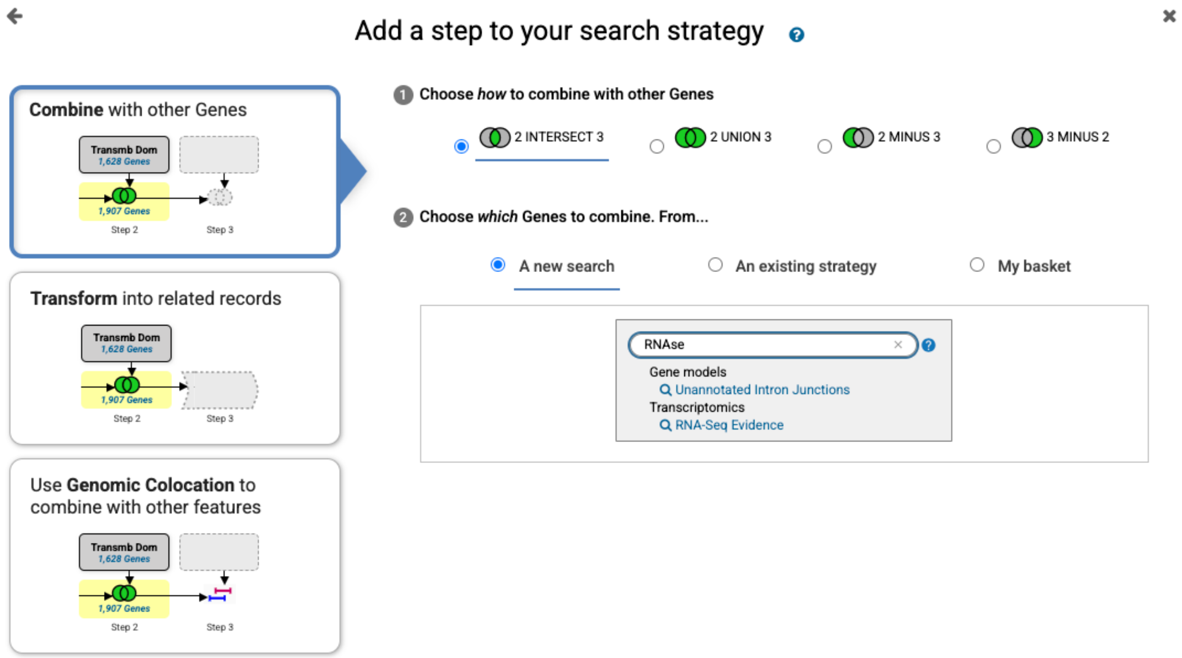
- Plasmodiidae
  - Plasmodium
    - Plasmodium falciparum
      - ☑ Plasmodium falciparum 3D7 [Reference]

Minimum Number of Transmembrane Domains: 1

Maximum Number of Transmembrane Domains: 99

*Unnamed Search Strategy **

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


Add a step to your search strategy

1. Choose *how* to combine with other Genes

● 2 INTERSECT 3 ○ 2 UNION 3 ○ 2 MINUS 3 ○ 3 MINUS 2

2. Choose *which* Genes to combine. From...

● A new search ○ An existing strategy ○ My basket

RNAse

Gene models
Unannotated Intron Junctions
Transcriptomics
RNA-Seq Evidence

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.

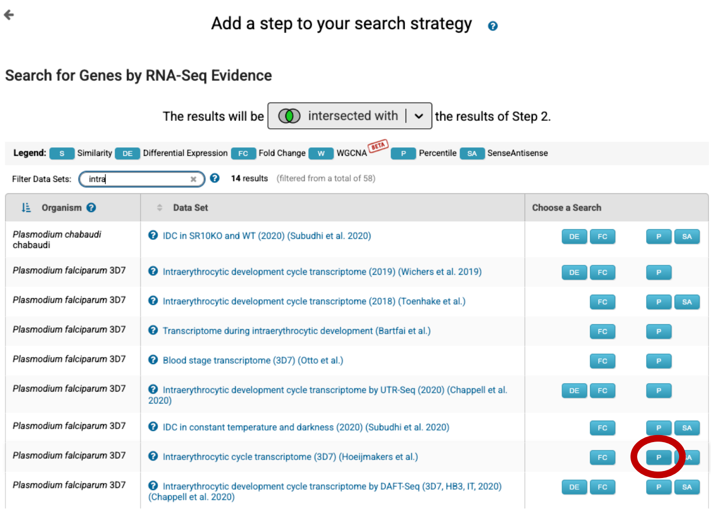
DE: Differential expression is available for experiments with biological replicates, enabling comparisons with statistical power.

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

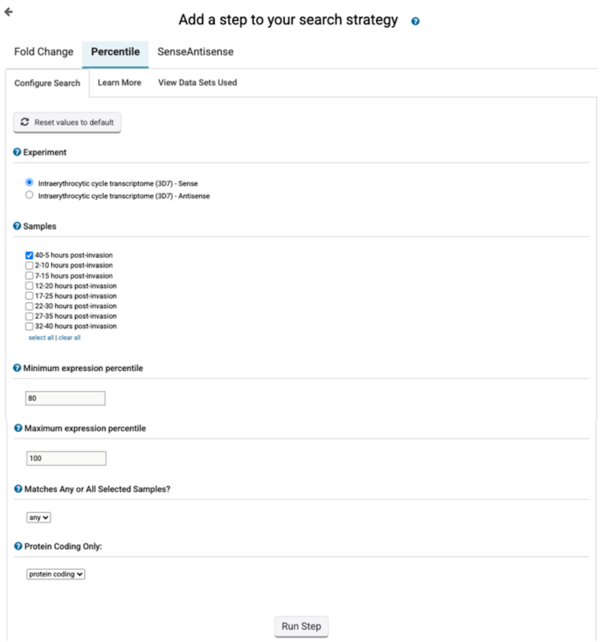
The results will be intersected with the results of Step 2.

Legend: S Similarity DE Differential Expression FC Fold Change W WGCNA BETA P Percentile SA SenseAntisense

Filter Data Sets: intra

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)	DE FC P SA
Plasmodium falciparum 3D7	Intraerythrocytic development cycle transcriptome (2019) (Wichers et al. 2019)	DE FC P
Plasmodium falciparum 3D7	Intraerythrocytic development cycle transcriptome (2018) (Toenhake et al.)	FC P SA
Plasmodium falciparum 3D7	Transcriptome during intraerythrocytic development (Bartfai et al.)	FC P
Plasmodium falciparum 3D7	Blood stage transcriptome (3D7) (Otto et al.)	FC P
Plasmodium falciparum 3D7	Intraerythrocytic development cycle transcriptome by UTR-Seq (2020) (Chappell et al. 2020)	DE FC P
Plasmodium falciparum 3D7	IDC in constant temperature and darkness (2020) (Subudhi et al. 2020)	FC P SA
Plasmodium falciparum 3D7	Intraerythrocytic cycle transcriptome (3D7) (Hoeijmakers et al.)	FC P SA
Plasmodium falciparum 3D7	Intraerythrocytic development cycle transcriptome by DAFT-Seq (3D7, HB3, IT, 2020) (Chappell et al. 2020)	DE FC P SA



Add a step to your search strategy

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

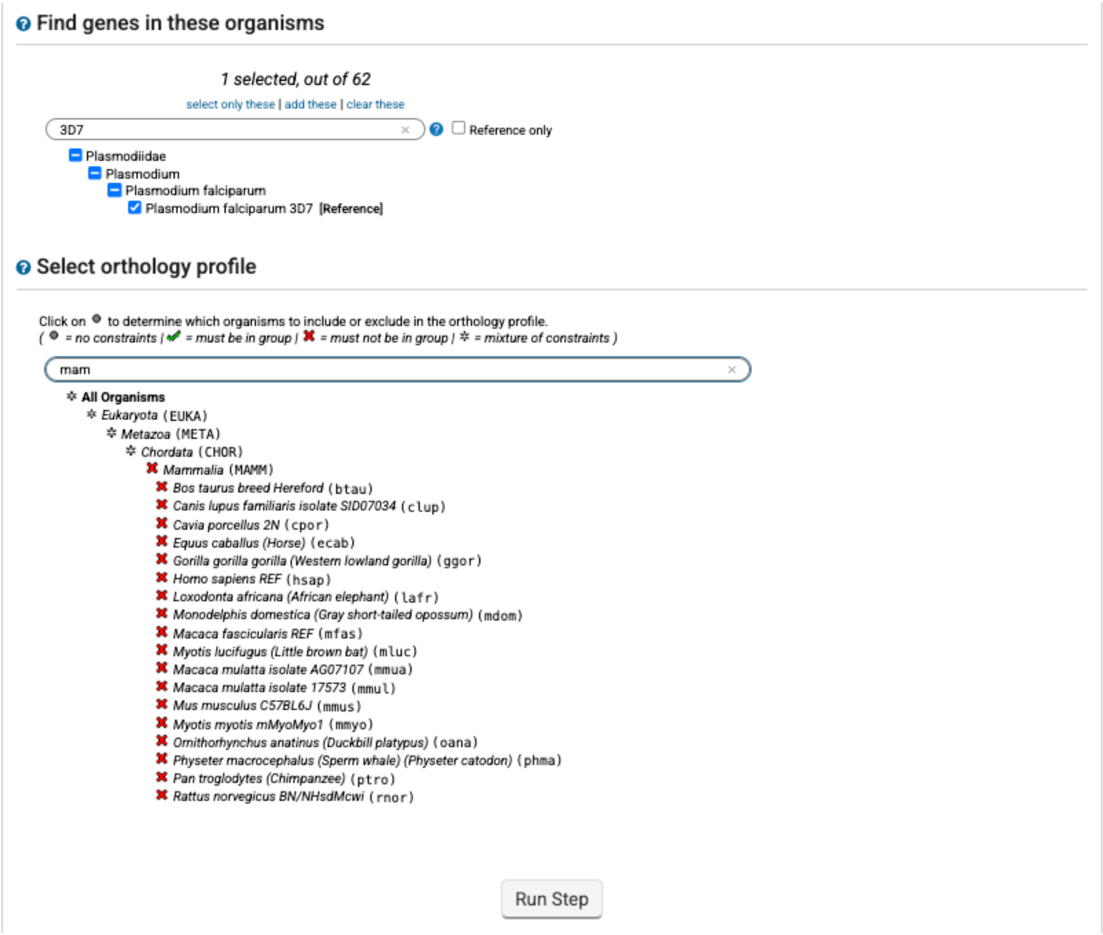
Minimum expression percentile: 80

Maximum expression percentile: 100

Matches Any or All Selected Samples?: any

Protein Coding Only: protein coding

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

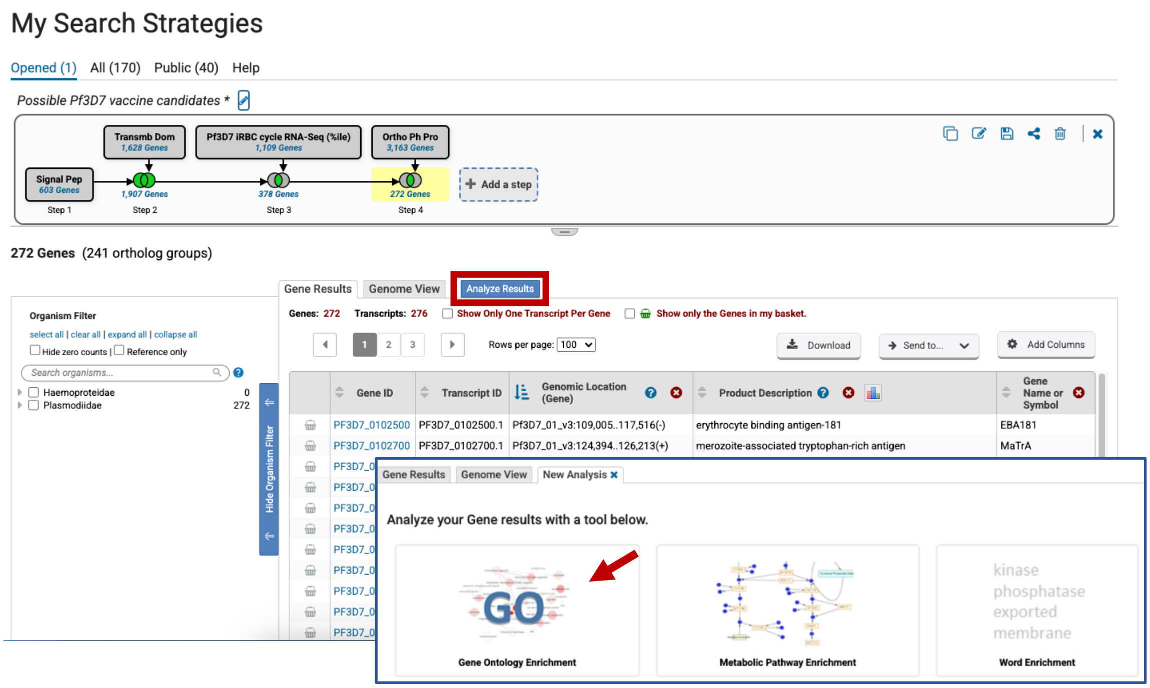
- ☐ Plasmodiidae
  - ☐ Plasmodium
    - ☐ Plasmodium falciparum
      - ☑ Plasmodium falciparum 3D7 [Reference]

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 SID07034* (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* (mmul)
          - ✗ *Mus musculus C57BL6J* (mmus)
          - ✗ *Myotis myotis mMyoMyo1* (mmyo)
          - ✗ *Ornithorhynchus anatinus (Duckbill platypus)* (oana)
          - ✗ *Physeter macrocephalus (Sperm whale) (Physeter catodon)* (phma)
          - ✗ *Pan troglodytes (Chimpanzee)* (ptro)
          - ✗ *Rattus norvegicus BN/NHsdMcwi* (rnor)



My Search Strategies

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

*Possible Pf3D7 vaccine candidates **

Step 1: Signal Pep — *603 Genes*
Step 2: Transmb Dom — *1,628 Genes* → *1,907 Genes*
Step 3: Pf3D7 iRBC cycle RNA-Seq (%ile) — *1,109 Genes* → *378 Genes*
Step 4: Ortho Ph Pro — *3,163 Genes* → *272 Genes*

272 Genes (241 ortholog groups)

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

Rows per page: 100

Organism Filter

select all | clear all | expand all | collapse all

☐ Hide zero counts | ☐ Reference only

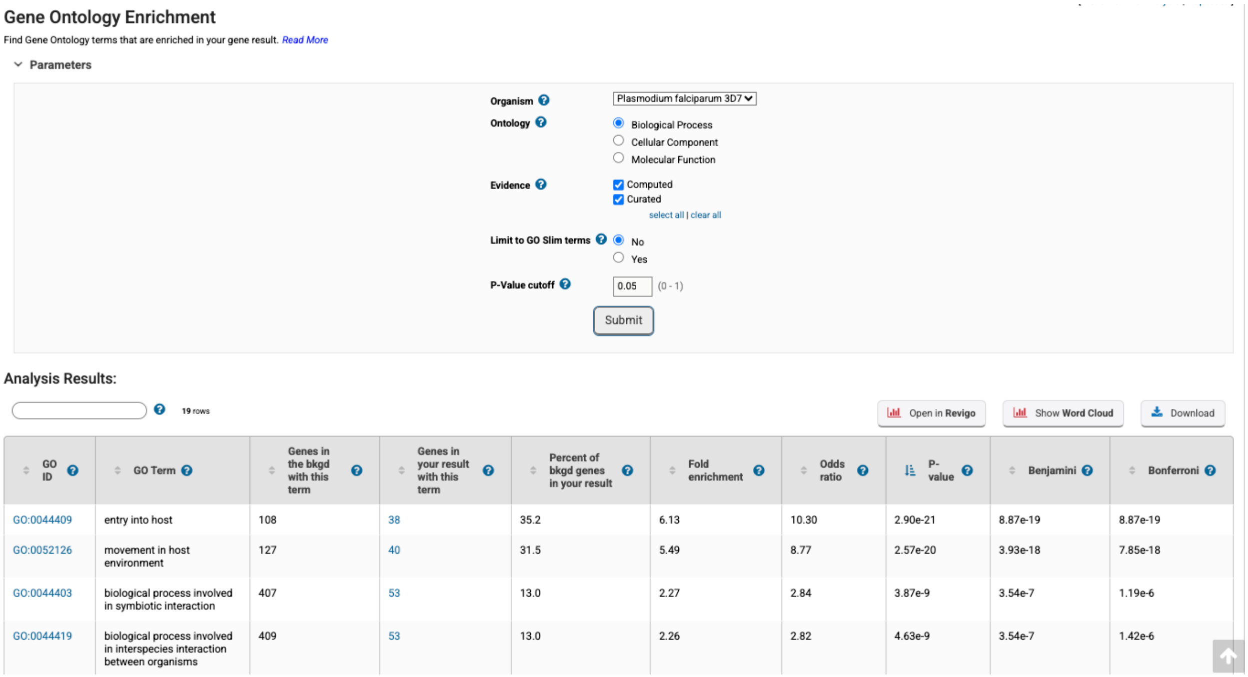
Organism	Count
☐ Haemoproteidae	0
☐ Plasmodiidae	272

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
PF3D7_0				
PF3D7_0				
PF3D7_0				
PF3D7_0				
PF3D7_0				
PF3D7_0				
PF3D7_0				
PF3D7_0				
PF3D7_0				

Analyze your Gene results with a tool below.

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*

Organism: Plasmodium falciparum 3D7

Ontology: ● Biological Process ○ Cellular Component ○ Molecular Function

Evidence: ☑ Computed ☑ Curated

select all | clear all

Limit to GO Slim terms: ● No ○ Yes

P-Value cutoff: 0.05 (0 - 1)

Analysis Results:

19 rows

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	Bonferroni
GO:0044409	entry into host	108	38	35.2	6.13	10.30	2.90e-21	8.87e-19	8.87e-19
GO:0052126	movement in host environment	127	40	31.5	5.49	8.77	2.57e-20	3.93e-18	7.85e-18
GO:0044403	biological process involved in symbiotic interaction	407	53	13.0	2.27	2.84	3.87e-9	3.54e-7	1.19e-6
GO: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

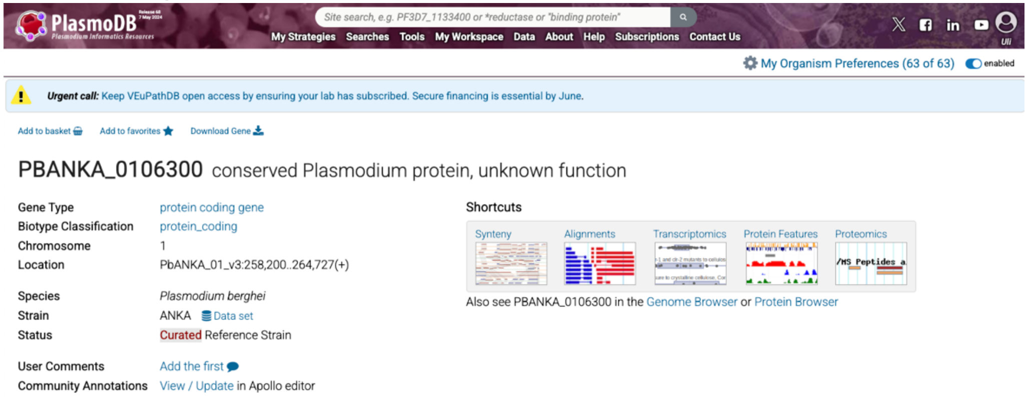
A variation of the above strategy that leverages additional data may be accessed using this link:
https://plasmodb.org/plasmo/app/workspace/strategies/import/66b5340817f0b0ed

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?



Urgent call: Keep VEuPathDB open access by ensuring your lab has subscribed. Secure financing is essential by June.

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

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.

Gene models

\# Exons in Gene 1

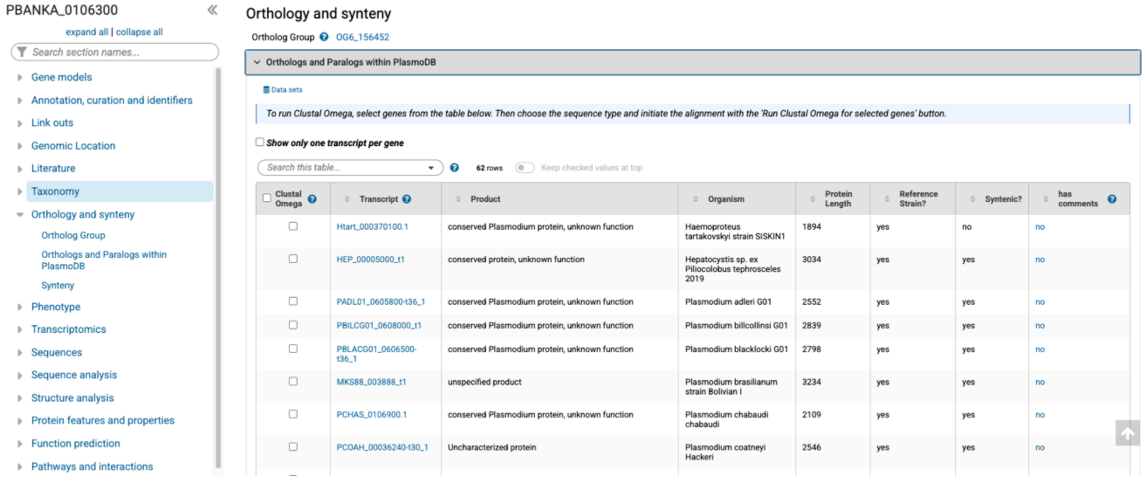
\# Transcripts 1

GFF format of gene and transcript features Click to open GFF in a new tab

Gene Models

This gene is available in **Apollo** for community annotation. To find out more about Apollo, please visit this help page.

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.



Orthology and synteny

Ortholog Group OG6_156452

Orthologs and Paralogs within PlasmoDB

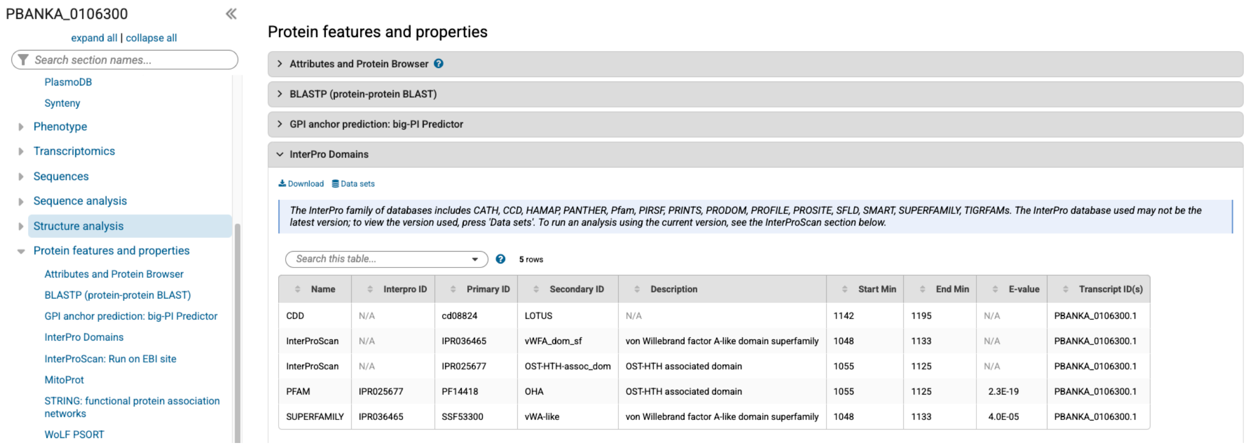
To run Clustal Omega, select genes from the table below. Then choose the sequence type and initiate the alignment with the 'Run Clustal Omega for selected genes' button.

☐ ***Show only one transcript per gene***

62 rows

☐ Clustal Omega	Transcript	Product	Organism	Protein Length	Reference Strain?	Syntenic?	has comments
☐	Htart_003700100.1	conserved Plasmodium protein, unknown function	Haemoproteus tartakovskyi strain SISKIN1	1894	yes	no	no
☐	HEP_00005000_t1	conserved protein, unknown function	Hepatocystis sp. ex Piliocolobus tephrosceles 2019	3034	yes	yes	no
☐	PADL01_0605800-t36_1	conserved Plasmodium protein, unknown function	Plasmodium adleri G01	2552	yes	yes	no
☐	PBILCG01_0608000_t1	conserved Plasmodium protein, unknown function	Plasmodium billcollinsi G01	2839	yes	yes	no
☐	PBLACG01_0606500-t36_1	conserved Plasmodium protein, unknown function	Plasmodium blacklocki G01	2798	yes	yes	no
☐	MKS88_003888_t1	unspecified product	Plasmodium brasilianum strain Bolivian I	3234	yes	yes	no
☐	PCHAS_0106900.1	conserved Plasmodium protein, unknown function	Plasmodium chabaudi chabaudi	2109	yes	yes	no
☐	PCOAH_00036240-t30_1	Uncharacterized protein	Plasmodium coatneyi Hackeri	2546	yes	yes	no

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?



Protein features and properties

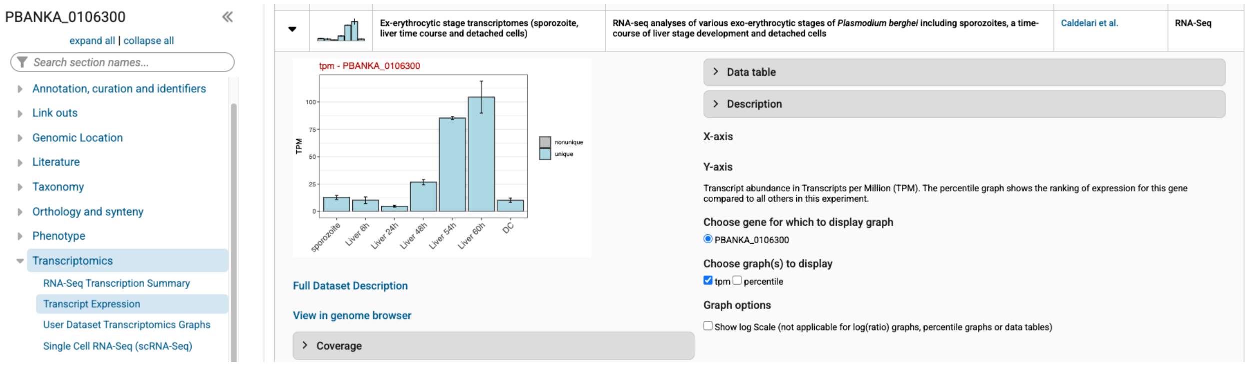
InterPro Domains

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.

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	OST-HTH-assoc_dom	OST-HTH associated domain	1055	1125	N/A	PBANKA_0106300.1
PFAM	IPR025677	PF14418	OHA	OST-HTH 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.



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

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

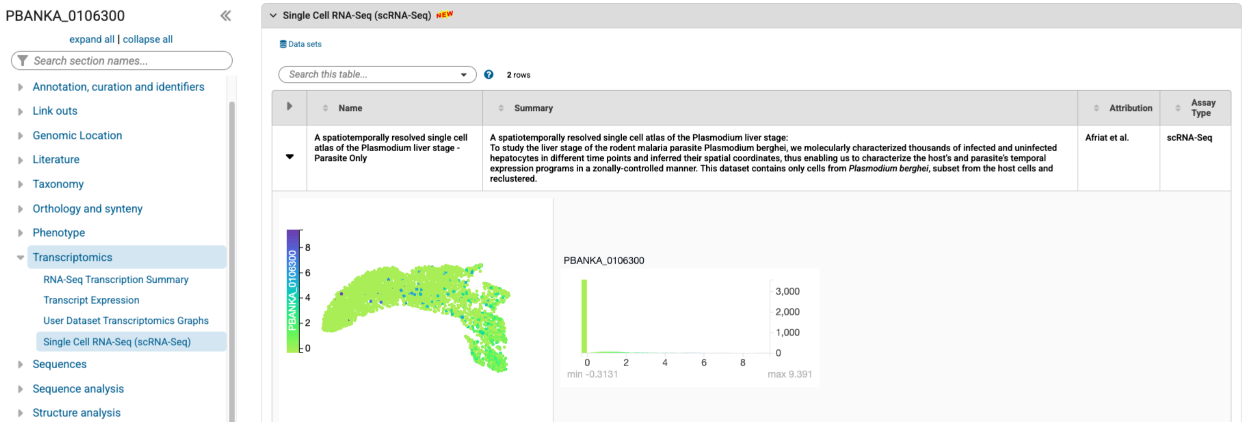
● PBANKA_0106300

Choose graph(s) to display

☑ tpm ☐ percentile

Graph options

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



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 Plasmodium berghei, 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 *Plasmodium berghei*, subset from the host cells and reclustered.	Afriat et al.	scRNA-Seq