

## Site Search

**Note:** this exercise uses PlasmoDB.org as an example database, but the same functionality is available on all VEuPathDB resources.

### Learning objectives:

- Search by keywords or identifiers
- Filter site search results by categories and fields
- Export results to a search strategy
- Find a specific gene using its ID in site search
- Navigate to and from the site search result
- Explore searches using wild cards (\*)

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).

The screenshot displays the PlasmoDB website interface. At the top, the PlasmoDB logo is on the left, and the release information 'Release 55 2 Dec 2021' is next to it. A search bar in the center contains the text 'kinase', with a red arrow pointing to it. To the right of the search bar is a red square button with a magnifying glass icon, representing the search icon. Below the search bar is a navigation menu with links: 'My Strategies', 'Searches', 'Tools', 'My Workspace', 'Data', 'About', 'Help', and 'Contact Us'. On the far right of the header, there are social media icons and a user profile icon labeled 'Omar'. Below the header, there is a link to 'My Organism Preferences (58 of 58)' and a toggle switch labeled 'enabled'. The main content area is divided into two sections. On the left, under the heading 'Search for...', there is a dropdown menu with options: 'Genes', 'Organisms', 'Popset Isolate Sequences', 'Genomic Sequences', 'Genomic Segments', 'SNPs', 'SNPs (from Array)', 'ESTs', and 'Metabolic Pathways'. On the right, under the heading 'Overview of Resources and Tools', there is a row of icons for various tools: 'Take a Tour', 'Getting Started', 'Search Strategies', 'Genome Browser', 'Transcriptomic Resources', 'Phenotypic Data', 'Analyze My Data', 'Downloads', 'How to Submit Data', and 'Cura Ann'. Below this row, there is a section titled 'Getting Started' which contains text about VEuPathDB and a link to 'Read More'.

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?

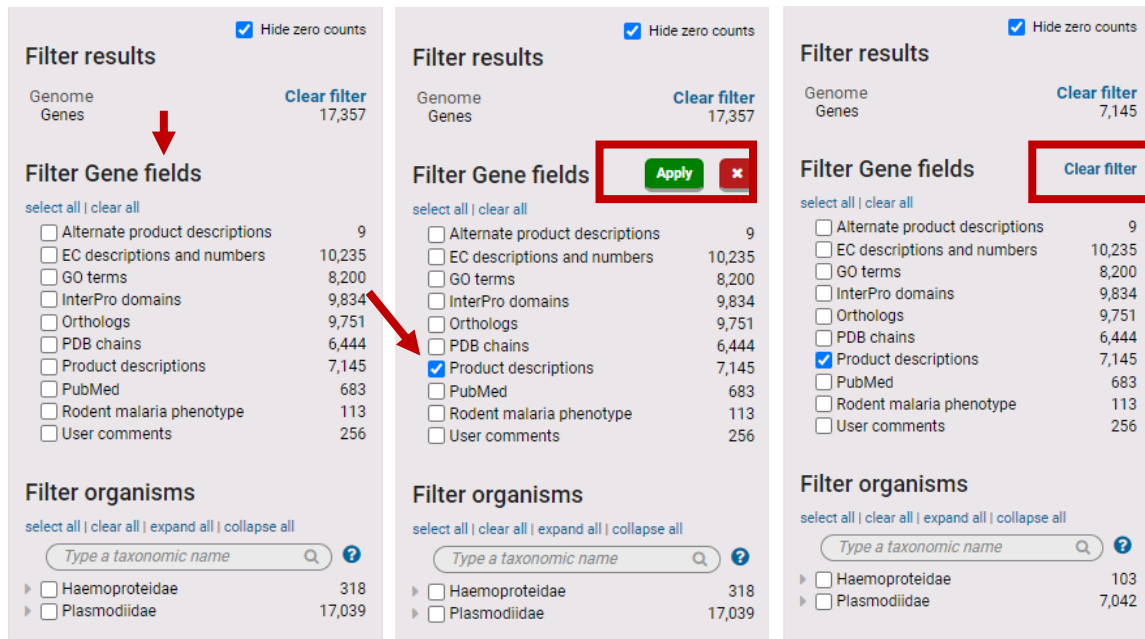
The screenshot shows a search results page for the term 'kinase'. At the top right, there is a link to 'My Organism Preferences (24 of 58)' and a 'disabled' toggle. Below this, the main heading is 'All results matching kinase'. To the right of the heading is a button 'Export as a Search Strategy to download or mine your results'. Below the heading is a pagination bar showing '1 - 20 of 18,234' and a set of page numbers (1, 2, 3, ..., 912). On the left side, there is a 'Filter results' section with a 'Hide zero counts' checkbox. It lists various categories with their respective counts: Genome (17,357), Genes (318), Population biology (352), Popset isolate sequences (193), Metabolism (328), Metabolic pathways (193), Compounds (1), Data access (3), Data sets (1), and Searches (3). Below this is a 'Filter fields' section with the text 'Select a result filter above'. Underneath is a 'Filter organisms' section with a search bar and a list of organisms: Haemoproteidae (318) and Plasmodiidae (17,040). On the right side, there is a details panel showing the first few results. Each result includes the gene name, organism, and fields matched. For example, the first result is 'Gene - PCYB\_132500 kinase' from 'Plasmodium cynomolgi strain B', with fields matched: 'GO terms; InterPro domains; Product descriptions'. The second result is 'Gene - PKNOH\_S07456300 Kinase' from 'Plasmodium knowlesi strain Malayan Strain Pk1 A', with fields matched: 'GO terms; InterPro domains; Orthologs; Product descriptions'. The third result is 'Gene - PKNOH\_S140234600 Kinase' from 'Plasmodium knowlesi strain Malayan Strain Pk1 A', with fields matched: 'EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions'. The fourth result is 'Gene - Htart\_000239600 Protein kinase domain/Protein tyrosine kinase/Kinase-like, putative' from 'Haemoproteus tartakovskyi strain SISKIN1', with fields matched: 'EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions'. The fifth result is 'Gene - Pf7G8-2\_000214800 Protein kinase domain/Protein tyrosine kinase/Kinase-like, putative' from 'Plasmodium falciparum 7G8 2019', with fields matched: 'EC descriptions and numbers; GO terms; InterPro domains; Orthologs; Product descriptions'. The sixth result is 'Gene - PfNF135\_070013800 Protein kinase domain/Protein tyrosine kinase/Kinase-like, putative' from 'Plasmodium falciparum NF135.C10', with fields matched: 'EC descriptions and numbers; GO terms; InterPro domains; Orthologs; Product descriptions'.

Results are summarized by category

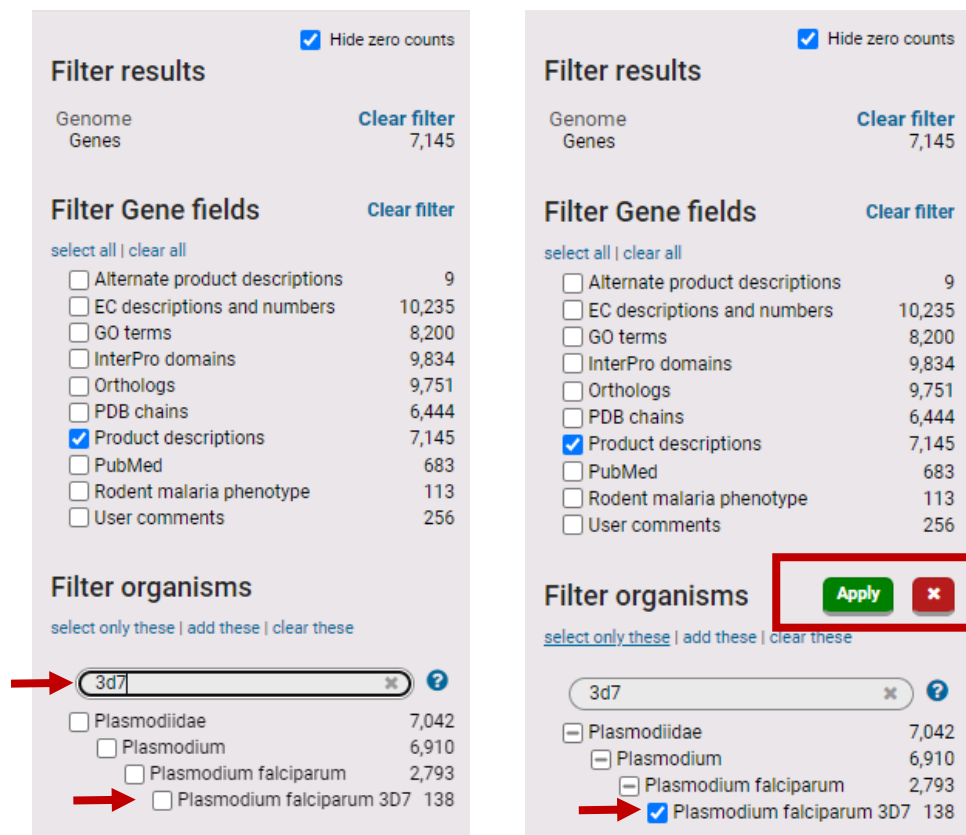
Details panel with information about each item returned

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

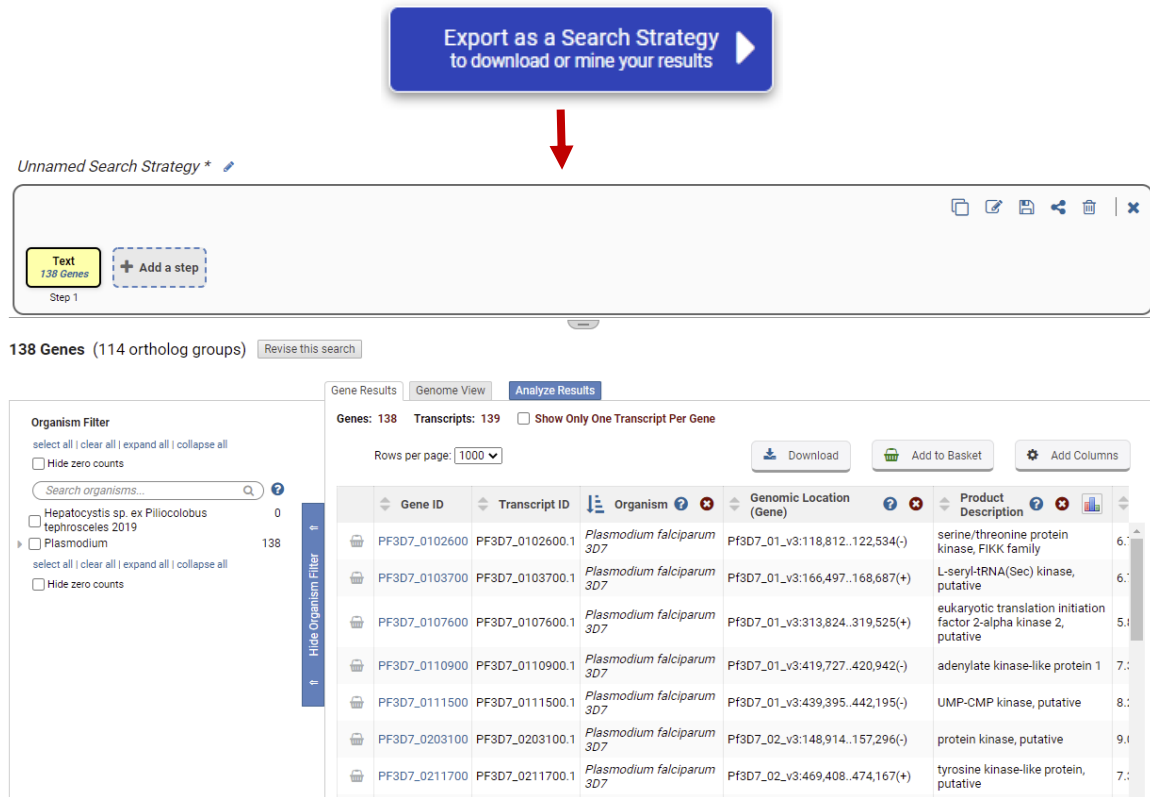
Filter the results so that you only view gene results (hint: click on the word *genes* in the *Filter results* section; arrow in image above) and the Filter Fields section expands to reveal additional filtering options. Select the *Product descriptions* field and choose *Apply* (middle panel below).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



4. **Filter the site search result by field:** 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.



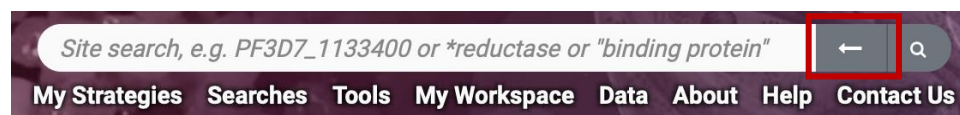
5. **Export the results to a search strategy.** Use the blue *Export as a search strategy* button at the top right-hand side of the results. Once exported you will be able to take advantage of over 100 specialized searches using the Add Step button. We will learn more about this in a future exercise.



The screenshot shows a blue button labeled "Export as a Search Strategy to download or mine your results" with a right-pointing arrow. A red arrow points down from this button to a search strategy interface. The interface has a title bar "Unnamed Search Strategy \*" and a toolbar with icons for file operations. Below the toolbar, there is a "Text" button labeled "138 Genes" and an "Add a step" button. The main content area shows "138 Genes (114 ortholog groups)" and a "Revise this search" button. On the left, there is an "Organism Filter" section with a search bar and a list of organisms: "Hepatocystis sp. ex Piliocolobus tephrosceles 2019" (0) and "Plasmodium" (138). The main table displays gene results for "Plasmodium falciparum 3D7". The table has columns: Gene ID, Transcript ID, Organism, Genomic Location (Gene), and Product Description. The rows show various genes and transcripts, such as PF3D7\_0102600, PF3D7\_0103700, PF3D7\_0107600, PF3D7\_0110900, PF3D7\_0111500, PF3D7\_0203100, and PF3D7\_0211700. The product descriptions include "serine/threonine protein kinase, FIKK family", "L-seryl-tRNA(Sec) kinase, putative", "eukaryotic translation initiation factor 2-alpha kinase 2, putative", "adenylate kinase-like protein 1", "UMP-CMP kinase, putative", "protein kinase, putative", and "tyrosine kinase-like protein, putative".

Gene ID	Transcript ID	Organism	Genomic Location (Gene)	Product Description
PF3D7_0102600	PF3D7_0102600.1	Plasmodium falciparum 3D7	PF3D7_01_v3:118,812..122,534(-)	serine/threonine protein kinase, FIKK family
PF3D7_0103700	PF3D7_0103700.1	Plasmodium falciparum 3D7	PF3D7_01_v3:166,497..168,687(+)	L-seryl-tRNA(Sec) kinase, putative
PF3D7_0107600	PF3D7_0107600.1	Plasmodium falciparum 3D7	PF3D7_01_v3:313,824..319,525(+)	eukaryotic translation initiation factor 2-alpha kinase 2, putative
PF3D7_0110900	PF3D7_0110900.1	Plasmodium falciparum 3D7	PF3D7_01_v3:419,727..420,942(-)	adenylate kinase-like protein 1
PF3D7_0111500	PF3D7_0111500.1	Plasmodium falciparum 3D7	PF3D7_01_v3:439,395..442,195(-)	UMP-CMP kinase, putative
PF3D7_0203100	PF3D7_0203100.1	Plasmodium falciparum 3D7	PF3D7_02_v3:148,914..157,296(-)	protein kinase, putative
PF3D7_0211700	PF3D7_0211700.1	Plasmodium falciparum 3D7	PF3D7_02_v3:469,408..474,167(+)	tyrosine kinase-like protein, putative

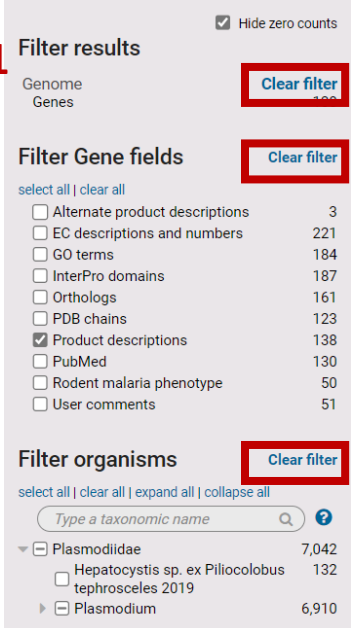
6. Return to the site search results page. You can achieve this in two ways: 1. Your previous results and filter settings were preserved and can be accessed by clicking on the 'back to results' arrow in the site search window. 2. Click on your browser's back arrow.



The screenshot shows a site search bar with the text "Site search, e.g. PF3D7\_1133400 or \*reductase or 'binding protein'". A red box highlights a back arrow icon in the search bar. Below the search bar is a navigation menu with links: "My Strategies", "Searches", "Tools", "My Workspace", "Data", "About", "Help", and "Contact Us".

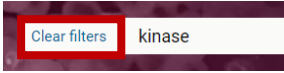
7. **Clear all filters.** You can achieve this in two ways: 1. You can click on each of the clear filter options in the filter results panel (boxes below). 2. You can click on the *clear filters* option in the site search window, which serves to Clear All filters.

**1**



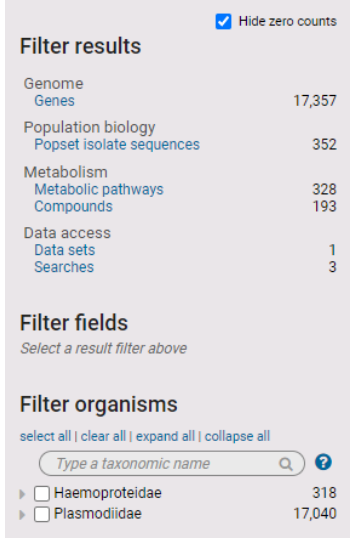
The screenshot shows the 'Filter results' panel with three 'Clear filter' buttons highlighted in red boxes: one for the main filter results, one for 'Filter Gene fields', and one for 'Filter organisms'.

**2**

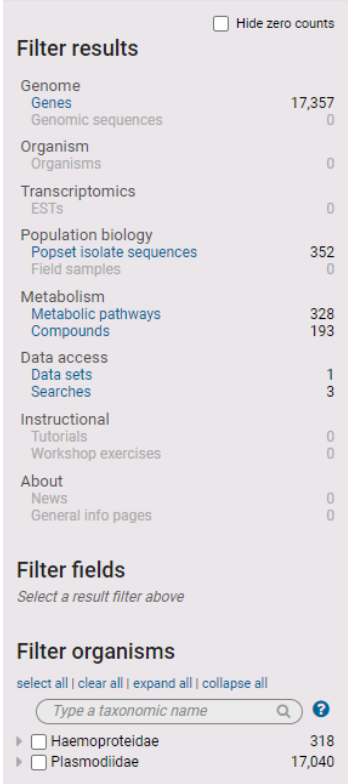


The screenshot shows the site search window with the 'Clear filters' button highlighted in a red box.

8. Click the *Hide zero counts* check box in the *Filter results* panel. What does this do?



The screenshot shows the 'Filter results' panel with the 'Hide zero counts' checkbox checked. The results are filtered to show only items with non-zero counts.



The screenshot shows the 'Filter results' panel with the 'Hide zero counts' checkbox unchecked. The results include items with zero counts, which are highlighted in blue.

9. **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.

PlasmoDB Database 2.0 19 Aug 2022

My Strategies Searches Tools My Workspace Data About Help Contact Us

My Organism Preferences (24 of 58) disabled

All results matching \*

1 - 20 of 946,043

**Filter results**

☐ Hide zero counts

Genome  
Genes 324,646  
Genomic sequences 22,089

Organism  
Organisms 58

Transcriptomics  
ESTs 287,336

Population biology  
Popset isolate sequences 153,109  
Field samples 0

Metabolism  
Metabolic pathways 3,267  
Compounds 154,803

Data access  
Data sets 317  
Searches 379

Instructional  
Tutorials 15  
Workshop exercises 4

About  
News 2  
General info pages 18

**Filter fields**  
Select a result filter above

**Filter organisms**  
select all | clear all | expand all | collapse all  
Type a taxonomic name

☐ Haemoproteidae 8,415  
☐ Plasmodiidae 338,910

Compound - CHEBI:100000 (2S,3S,4R)-3-[4-(3-cyclopentylprop-1-ynyl)phenyl]-4-(hydroxymethyl)-1-(2-methoxy-1-oxoethyl)-2-azetidinecarboxitrile

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-yl)methyl)-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-azetidinecarboxitrile

Compound - CHEBI:100009 (2R,3S,4S)-1-(4-fluorophenyl)sulfonyl-4-(hydroxymethyl)-3-phenyl-2-azetidinecarboxitrile

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-pyranol[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

PlasmoDB Database 2.0 19 Aug 2022

\*kinase

My Strategies Searches Tools My Workspace Data About Help Contact Us

My Organism Preferences (24 of 58) disabled

All results matching \*kinase

1 - 20 of 20,768

**Filter results**

☒ Hide zero counts

Genome  
Genes 18,834

Population biology  
Popset isolate sequences 1,273

Metabolism  
Metabolic pathways 453  
Compounds 204

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

☐ Haemoproteidae 353  
☐ Plasmodiidae 18,482

Gene - AK88\_00104 CK1/CK1/CK1-D protein kinase  
Organism: Plasmodium fragile strain nilgiri  
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions

Gene - AK88\_00479 CAMK protein kinase  
Organism: Plasmodium fragile strain nilgiri  
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions

Gene - AK88\_00505 pantothenate kinase  
Organism: Plasmodium fragile strain nilgiri  
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions

Gene - AK88\_00565 Atypical/ABC1 protein kinase  
Organism: Plasmodium fragile strain nilgiri  
Fields matched: GO terms; InterPro domains; Orthologs; Product descriptions

Gene - AK88\_00580 CMGC protein kinase  
Organism: Plasmodium fragile strain nilgiri  
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions

Gene - AK88\_00633 CMGC/GSK protein kinase  
Organism: Plasmodium fragile strain nilgiri  
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; Product descriptions

Gene - AK88\_00642 serine/threonine protein kinase

10. **Search for a specific gene ID.** Enter the gene ID in the site search window: *PF3D7\_0310100*. 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.

The screenshot displays the PlasmoDB website interface. At the top, the PlasmoDB logo is on the left, and the release information 'Release 52 29 May 2021' is on the right. A search bar in the center contains the text 'PF3D7\_0310100'. Below the search bar, a navigation menu includes links for 'My Strategies', 'Searches', 'Tools', 'My Workspace', 'Data', 'About', 'Help', and 'Contact Us'. On the far right, there are social media icons for Twitter, Facebook, and YouTube, along with a user profile icon labeled 'Susanne'.

Below the navigation bar, the main heading reads 'Genes matching PF3D7\_0310100'. To the right of this heading is a blue button labeled 'Export as a Search Strategy to download or mine your results'. Below the heading, the text '1 - 2 of 2' indicates the number of results.

On the left side, there is a 'Filter results' panel. It includes a 'Hide zero counts' checkbox, which is checked. Under 'Filter results', there are sections for 'Genome' (with 'Genes' listed and a count of 2), 'Filter Gene fields' (with options for 'External links', 'Gene ID', and 'Notes from annotators', each with a count of 1), and 'Filter organisms' (with a search bar and a list of 'Plasmodiidae' and 'Plasmodium', each with a count of 2).

The main content area displays three search results. The first result is highlighted with a blue border and contains the following information: 'Gene - PF3D7\_0310100 calcium-dependent protein kinase 3', 'Gene name or symbol: CDPK3', 'Organism: Plasmodium falciparum 3D7', and 'Fields matched: External links; Gene ID'. The second result is 'Gene - PF3D7\_0310100 calcium-dependent protein kinase 3', 'Gene name or symbol: CDPK3', 'Organism: Plasmodium falciparum 3D7', and 'Fields matched: External links; Gene ID'. The third result is 'Gene - PGSY75\_0310100 calcium-dependent protein kinase 3', 'Organism: Plasmodium gaboni strain SY75', and 'Fields matched: Notes from annotators'.