

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 text 'Release 55 2 Dec 2021' is next to it. A search bar in the center contains the word 'kinase', with a red arrow pointing to it. To the right of the search bar is a search icon (a magnifying glass inside a square), which is highlighted with a red square. 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 section titled 'Search for...' on the left, which includes a filter dropdown and a list of search categories: Genes, Organisms, Popset Isolate Sequences, Genomic Sequences, Genomic Segments, SNPs, SNPs (from Array), ESTs, and Metabolic Pathways. To the right of this is a section titled 'Overview of Resources and Tools' with icons for various tools: Take a Tour, Getting Started (highlighted with a red box), Search Strategies, Genome Browser, Transcriptomic Resources, Phenotypic Data, Analyze My Data, Downloads, How to Submit Data, and Cura Ann. Below this is a 'Getting Started' section with text explaining the site search functionality and a 'Read More' link.

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 the PlasmoDB website search results for the term 'kinase'. The top navigation bar includes links for 'My Strategies', 'Searches', 'Tools', 'My Workspace', 'Data', 'About', 'Help', and 'Contact Us'. A search bar at the top right contains the word 'kinase'. Below the navigation bar, there is a section for 'All results matching kinase' with a button to 'Export as a Search Strategy'. The results are displayed in a table with columns for 'Gene', 'Organism', and 'Fields matched'. The first three results are: 1. Gene - PCYB_132500, Organism: Plasmodium cynomolgi strain B, Fields matched: GO terms; InterPro domains; Product descriptions. 2. Gene - PKNOH_S07456300, Organism: Plasmodium knowlesi strain Malayan Strain Pk1 A, Fields matched: GO terms; InterPro domains; Orthologs; Product descriptions. 3. Gene - PKNOH_S140234600, Organism: Plasmodium knowlesi strain Malayan Strain Pk1 A, Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions. On the left side, there is a 'Filter results' section with a table showing counts for various categories: Genome (16,147), Population biology (352), Metabolism (328), Data access (1), and Data sets (3). A red arrow points to the 'Genes' link in the 'Filter results' section. Below the 'Filter results' section is a 'Filter fields' section with a dropdown menu. Below the 'Filter fields' section is a 'Filter organisms' section with a search bar and a list of organisms: Plasmodiidae (16,148), Hepatocystis sp. ex Piliocolobus tephrosceles 2019 (292), and Plasmodium (15,856). A red bracket is drawn under the 'Filter results' and 'Filter fields' sections, and another red bracket is drawn under the 'Filter organisms' section.

Category	Count
Genome	16,147
Population biology	352
Metabolism	328
Data access	1
Data sets	3

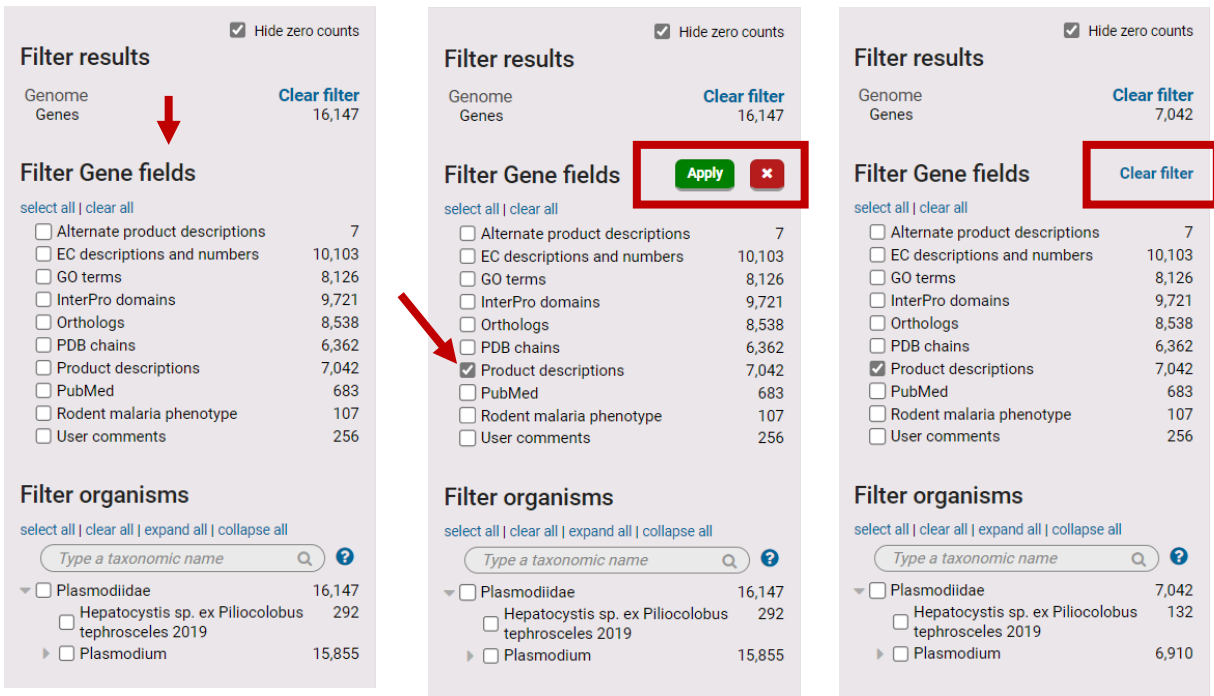
Gene	Organism	Fields matched
PCYB_132500	Plasmodium cynomolgi strain B	GO terms; InterPro domains; Product descriptions
PKNOH_S07456300	Plasmodium knowlesi strain Malayan Strain Pk1 A	GO terms; InterPro domains; Orthologs; Product descriptions
PKNOH_S140234600	Plasmodium knowlesi strain Malayan Strain Pk1 A	EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions
Pf7G8-2_000214800	Plasmodium falciparum 7G8 2019	EC descriptions and numbers; GO terms; InterPro domains; Orthologs; Product descriptions
PfNF135_070013800	Plasmodium falciparum NF135, C10	EC descriptions and numbers; GO terms; InterPro domains; Orthologs; Product descriptions
PfNF54_070035200	Plasmodium falciparum NF54	EC descriptions and numbers; GO terms; InterPro domains; Orthologs; Product descriptions

Results are summarized by category

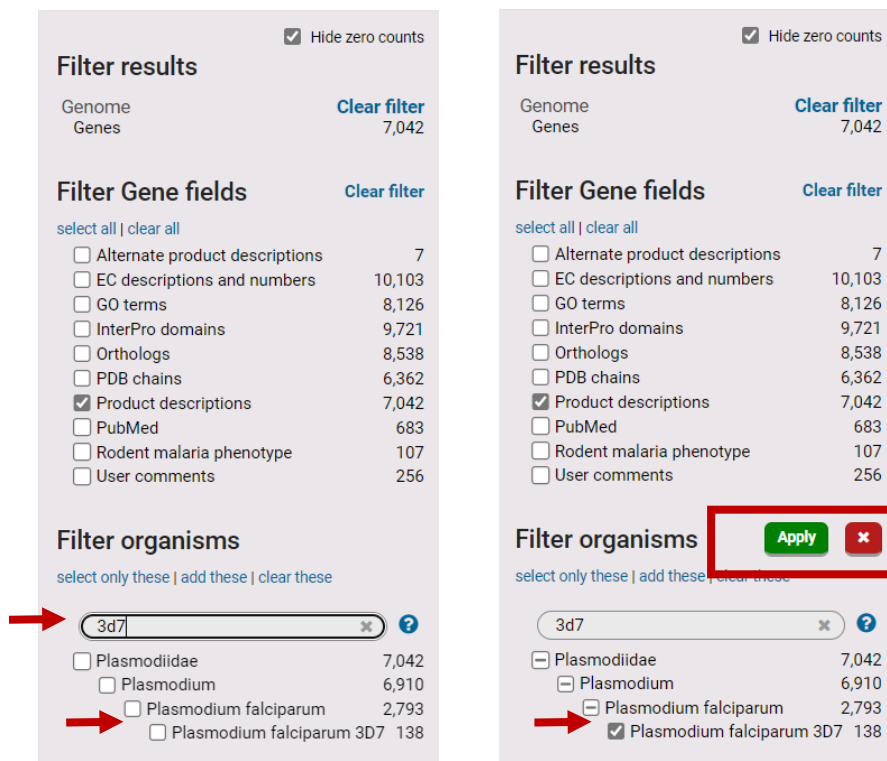
Details panel with information about each item returned

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.

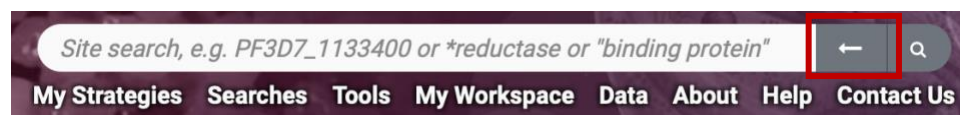


5. **Export the results to a search strategy.** Use the blue *Export as a search strategy* button at the top right-hand side of the results. Once exported you will be able to take advantage of over 100 specialized searches using the Add Step button. We will learn more about this in a future exercise.

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 titled "Unnamed Search Strategy *". The interface includes a "Text" step with "138 Genes" and an "Add a step" button. Below this, it shows "138 Genes (114 ortholog groups)" and a "Revise this search" button. The main area displays a table of gene results for *Plasmodium falciparum* 3D7, with columns for Gene ID, Transcript ID, Organism, Genomic Location, and Product Description. A sidebar on the left shows an "Organism Filter" with options for "Hepatocystis sp. ex Piliocolobus tephrosceles 2019" and "Plasmodium".

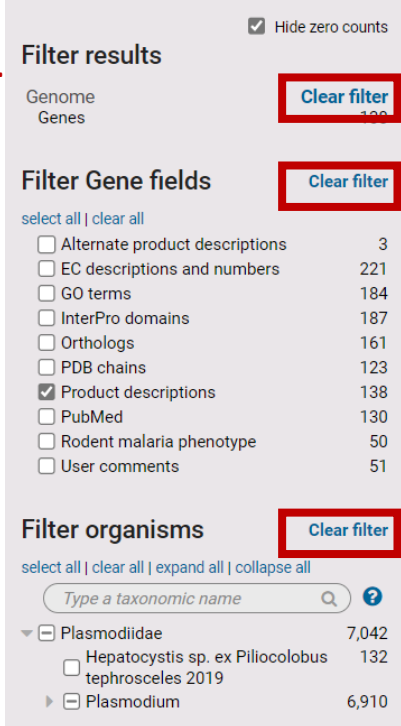
Gene ID	Transcript ID	Organism	Genomic Location (Gene)	Product Description
PF3D7_0102600	PF3D7_0102600.1	<i>Plasmodium falciparum</i> 3D7	Pf3D7_01_v3:118,812..122,534(-)	serine/threonine protein kinase, FIKK family
PF3D7_0103700	PF3D7_0103700.1	<i>Plasmodium falciparum</i> 3D7	Pf3D7_01_v3:166,497..168,687(+)	L-seryl-tRNA(Sec) kinase, putative
PF3D7_0107600	PF3D7_0107600.1	<i>Plasmodium falciparum</i> 3D7	Pf3D7_01_v3:313,824..319,525(+)	eukaryotic translation initiation factor 2-alpha kinase 2, putative
PF3D7_0110900	PF3D7_0110900.1	<i>Plasmodium falciparum</i> 3D7	Pf3D7_01_v3:419,727..420,942(-)	adenylate kinase-like protein 1
PF3D7_0111500	PF3D7_0111500.1	<i>Plasmodium falciparum</i> 3D7	Pf3D7_01_v3:439,395..442,195(-)	UMP-CMP kinase, putative
PF3D7_0203100	PF3D7_0203100.1	<i>Plasmodium falciparum</i> 3D7	Pf3D7_02_v3:148,914..157,296(-)	protein kinase, putative
PF3D7_0211700	PF3D7_0211700.1	<i>Plasmodium falciparum</i> 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.



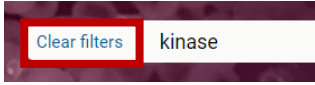
7. **Clear all filters.** You can achieve this in two ways: 1. You can click on each of the clear filter options in the filter results panel (boxes below). 2. You can click on the *clear filters* option in the site search window, which serves to Clear All filters.

1



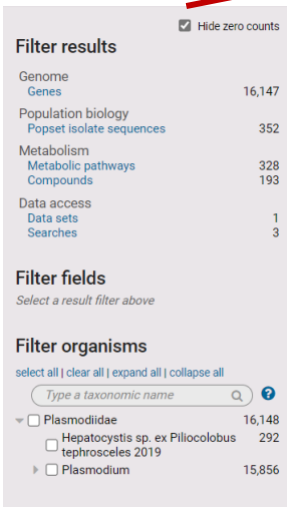
The 'Filter results' panel shows three sections: 'Genome' (with a 'Clear filter' button), 'Filter Gene fields' (with a 'Clear filter' button), and 'Filter organisms' (with a 'Clear filter' button). Each button is highlighted with a red box.

2

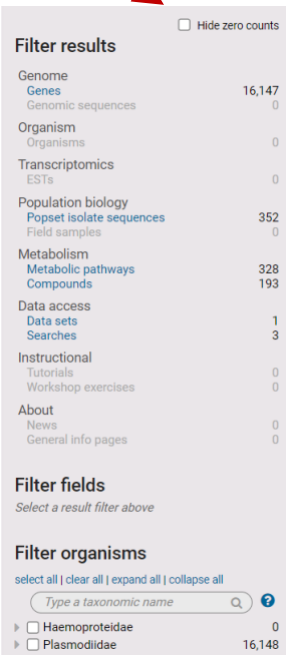


The site search window shows a search bar with the text 'kinase' and a 'Clear filters' button highlighted with a red box.

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



The 'Filter results' panel shows the 'Hide zero counts' checkbox checked. The results list includes: Genome (16,147), Genes (16,147), Population biology (352), Popset isolate sequences (352), Metabolism (328), Metabolic pathways (328), Compounds (193), Data access (1), Data sets (3), Searches (3), Filter fields (Select a result filter above), Filter organisms (select all | clear all | expand all | collapse all), Type a taxonomic name (search bar), Plasmodiidae (16,148), Hepatocystis sp. ex Piliocolobus tephrosceles 2019 (292), Plasmodium (15,856).



The 'Filter results' panel shows the 'Hide zero counts' checkbox unchecked. The results list includes: Genome (16,147), Genes (16,147), Genomic sequences (0), Organism (0), Organisms (0), Transcriptomics (0), ESTs (0), Population biology (352), Popset isolate sequences (352), Field samples (0), Metabolism (328), Metabolic pathways (328), Compounds (193), Data access (1), Data sets (3), Searches (3), Instructional (0), Tutorials (0), Workshop exercises (0), About (0), News (0), General info pages (0), Filter fields (Select a result filter above), Filter organisms (select all | clear all | expand all | collapse all), Type a taxonomic name (search bar), Haemoproteidae (0), Plasmodiidae (16,148).

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.

The screenshot shows the Plasmadb website interface. At the top, there is a navigation bar with links: My Strategies, Searches, Tools, My Workspace, Data, About, Help, and Contact Us. A search bar is located in the top right corner. Below the navigation bar, there is a section titled "All results matching *" with a sub-header "1 - 20 of 939,975". On the left side, there is a "Filter results" panel with various categories and counts: Genome (319,221), Genes (22,089), Genomic sequences (22,089), Organism (58), Organisms (58), Transcriptomics (287,336), ESTs (287,336), Population biology (152,489), Popset isolate sequences (0), Field samples (0), Metabolism (3,267), Metabolic pathways (154,803), Compounds (154,803), Data access (294), Data sets (379), Searches (379), Instructional (15), Tutorials (4), Workshop exercises (4), About (2), News (2), and General info pages (18). Below the "Filter results" panel, there is a "Filter fields" section with a link "Select a result filter above". Under "Filter organisms", there is a search bar "Type a taxonomic name" and two checkboxes: "Haemoproteidae" (2,988) and "Plasmodiidae" (338,720). The main content area displays a list of search results, each with a compound name, its CHEBI ID, and a brief description. The results are paginated, showing 1 to 20 of 939,975 results.

Plasmadb
Plasmodium Informatics Resources

My Strategies Searches Tools My Workspace Data About Help Contact Us

My Organism Preferences ()

All results matching *

Export as to download

1 - 20 of 939,975

Hide zero counts

Filter results

Genome 319,221
Genes 22,089
Genomic sequences 22,089
Organism 58
Organisms 58
Transcriptomics 287,336
ESTs 287,336
Population biology 152,489
Popset isolate sequences 0
Field samples 0
Metabolism 3,267
Metabolic pathways 154,803
Compounds 154,803
Data access 294
Data sets 379
Searches 379
Instructional 15
Tutorials 4
Workshop exercises 4
About 2
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 2,988
Plasmodiidae 338,720

Compound - CHEBI:100000 (2S,3S,4R)-3-[4-(3-cyclopentylprop-1-ynyl)phenyl]-4-(hydroxymethyl)-1-(2-methoxyazetidine)carboxamide
Compound - CHEBI:100001 N-[(2R,3S,6R)-2-(hydroxymethyl)-6-[2-[[oxo-4-(trifluoromethyl)anilino]methyl]aminopyridine]carboxamide
Compound - CHEBI:100002 3-chloro-N-[(5S,6S,9S)-5-methoxy-3,6,9-trimethyl-2-oxo-11-oxa-3,8-diazabicyclo[1.1.0]tetra-1,3,15-trien-14-yl]benzenesulfonamide
Compound - CHEBI:100003 (4R,7S,8R)-8-methoxy-4,7,10-trimethyl-11-oxo-14-(1-oxobutylamino)-N-propyl-2-oxo-1,4-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-pyridyl)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-tetrahydrofuran-6-yl]-3-methoxybenzenesulfonamide
Compound - CHEBI:100006 N-(1,3-benzodioxol-5-ylmethyl)-2-[(2R,3R,6S)-3-[[2,5-difluoroanilino]-oxomethyl]aryl(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-azetidinecarboxamide
Compound - CHEBI:100009 (2R,3S,4S)-1-(4-fluorophenyl)sulfonyl-4-(hydroxymethyl)-3-phenyl-2-azetidinecarboxamide
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[1.1.0]tetra-1,3,15-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-1,5-benzoxazin-8-yl]-1-(1-piperidinyl)ethanone
Compound - CHEBI:100012 N-[(2R,4aR,12aR)-2-[2-(cyclohexylmethylamino)-2-oxoethyl]-5-methyl-6-oxo-2,3,4,4-tetrahydro-1,5-benzoxazin-8-yl]-1,5-benzoxazin-2-yl]-2-pyrazinecarboxamide
Compound - CHEBI:100013 2-[(2S,4aR,12aS)-8-(ethylcarbamoylamino)-5-methyl-6-oxo-2,3,4,4a,12,12a-hexahydro-1,5-benzoxazin-2-yl]-N-[(1S)-1-phenylethyl]acetamide

PlasmoDB Plasmodium Informatics Resources

My Strategies Searches Tools My Workspace Data About Help Contact Us

My Organism Preferences (0)

All results matching ***kinase**

Export as to download

1 - 20 of 19,523

Filter results

☐ Hide zero counts

Genome
Genes 17,589
Genomic sequences 0

Organism
Organisms 0

Transcriptomics
ESTs 0

Population biology
Popset isolate sequences 1,273
Field samples 0

Metabolism
Metabolic pathways 453
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
select all | clear all | expand all | collapse all

Type a taxonomic name

☐ Haemoproteidae 0
☐ Plasmodiidae 17,590

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
Organism: Plasmodium fragile strain nilgiri
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions

Gene - AK88_00657 CAMK/CDPK protein kinase
Organism: Plasmodium fragile strain nilgiri

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.

PlasmoDB Plasmodium Informatics Resources

Release 52 20 May 2021

My Strategies Searches Tools My Workspace Data About Help Contact Us

PF3D7_0310100

Export as a Search Strategy to download or mine your results

Genes matching **PF3D7_0310100**

1 - 2 of 2

Filter results

☒ Hide zero counts

Genome
Genes 2

Filter Gene fields
select all | clear all

☐ External links 1
☐ Gene ID 1
☐ Notes from annotators 1

Filter organisms
select all | clear all | expand all | collapse all

Type a taxonomic name

☐ Plasmodiidae 2
☐ Plasmodium 2

Gene - PF3D7_0310100 calcium-dependent protein kinase 3
Gene name or symbol: CDPK3
Organism: Plasmodium falciparum 3D7
Fields matched: External links; Gene ID

Gene - PF3D7_0310100 calcium-dependent protein kinase 3
Gene name or symbol: CDPK3
Organism: Plasmodium falciparum 3D7
Fields matched: External links; Gene ID

Gene - PGSY75_0310100 calcium-dependent protein kinase 3
Organism: Plasmodium gaboni strain SY75
Fields matched: Notes from annotators