

Site Search in VEuPathDB

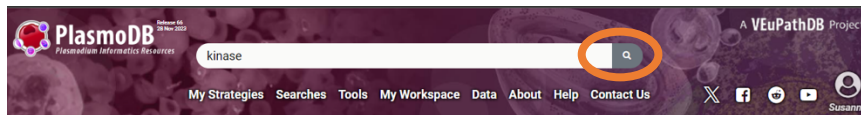
Learning objectives

- 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 (*)

Introduction

The site search is located 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. Go to PlasmoDB.org¹ and 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 number² of results with the word kinase?

¹ Note that this exercise uses PlasmoDB.org as an example database, but the same functionality is available on all VEuPathDB resources.

² VEuPathDB is updated regularly, so the numbers you see while completing the exercise may differ slightly from those shown in the screenshots.

Are all the results genes?

All results matching kinase

The screenshot shows the search results for 'kinase'. The left sidebar has a 'Filter results' section with a red arrow pointing to the 'Genes' link, which has a count of 19,596. Below this is the 'Filter fields' section with a red bracket underneath. The 'Filter organisms' section shows 'Haemoproteidae' with 331 results and 'Plasmodiidae' with 19,266 results. The main results area shows a list of genes, with the first three highlighted: 'Gene - PCYB_132500', 'Gene - PKNOH_S07456300', and 'Gene - PKNOH_S140234600'. Each gene entry includes its name, type, organism, and a list of fields matched.

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).

The three panels show the process of filtering the search results. The first panel shows the 'Filter Gene fields' section with a red arrow pointing to the 'Product description' checkbox. The second panel shows the 'Apply' button highlighted with a red box. The third panel shows the 'Clear filter' button highlighted with a red box. The 'Filter organisms' section shows 'Haemoproteidae' with 331 results and 'Plasmodiidae' with 19,265 results.

- Filter the site search result by organism: [How many of the above genes are found in *Plasmodium falciparum* 3D7?](#)

(Hint: Explore the *Filter organisms* section of the results filter and use the search filter to navigate the tree.

The left screenshot shows the 'Filter results' interface with the 'Filter organisms' section. The search box contains '3d7'. The list of organisms includes Plasmodium falciparum 3D7 [Ref] with 137 genes. The right screenshot shows the same interface with the 'Apply' button highlighted in a red box. Red arrows point to the search box and the 'Apply' button.

- 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. We will learn more about this in a future exercise.

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

The screenshot shows the search results page for 'kinase'. A red circle highlights the 'Export as a Search Strategy' button in the top right corner. A red arrow points from this button to the 'Add a step' button in the 'My Search Strategies' section below.

My Search Strategies

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

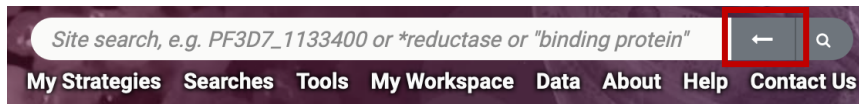
The screenshot shows the 'My Search Strategies' section with a search strategy named 'Unnamed Search Strategy'. A red arrow points from the 'Export as a Search Strategy' button in the previous screenshot to the 'Add a step' button in this section.

137 Genes (113 ortholog groups) [Revise this search](#)

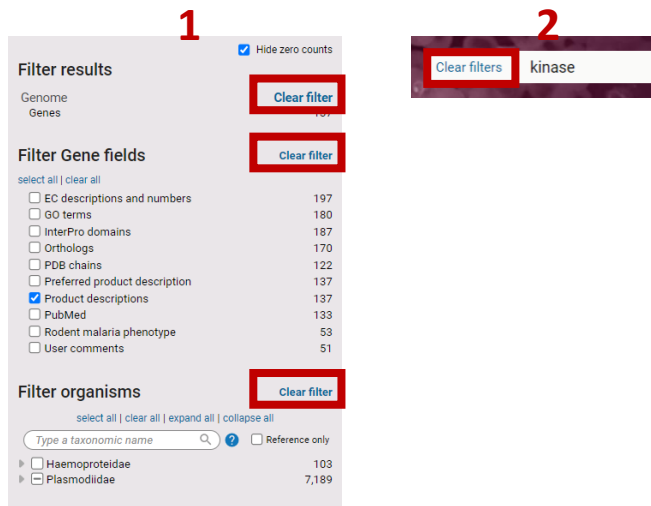
The screenshot shows the 'Gene Results' table with the following data:

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

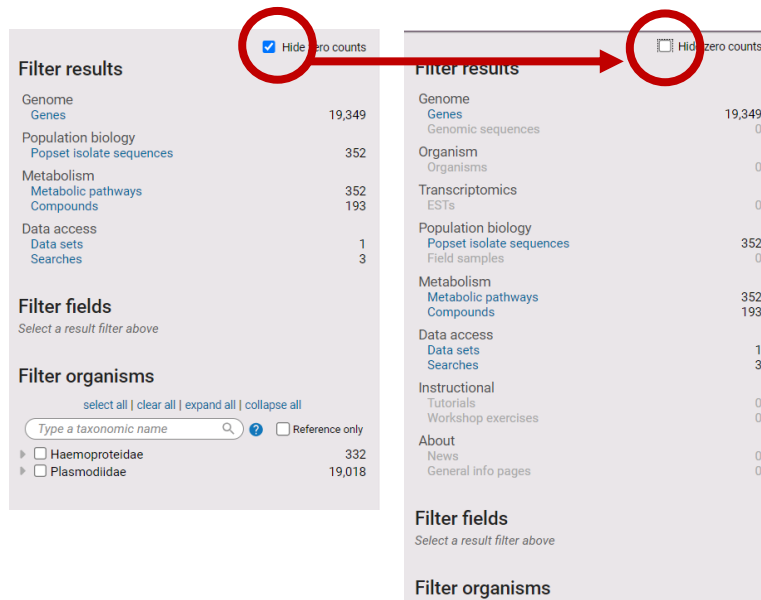
- Return to the site search results page. You can achieve this in two ways: 1. Your previous results and filter settings were preserved and can be accessed by clicking on the 'back to results' arrow in the site search window. 2. Click on your browser's back arrow.



- Clear all filters. You can achieve this in two ways: 1. click on each clear filter option in the filter results panel. 2. You can click on the *clear filters option* in the site search window to Clear All filters.



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



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 kinase, like phosphofructokinase. As usual, results can then be explored using the filters in the *Results filter* on the left side of the website. Feel free to compare the results when you run a search or the word *kinase* to a search with a wild card **kinase* or **kinase**.

All results matching *

1 - 20 of 978,719

Export as a Search Strategy to download or mine your results

Filter results ☒ Hide zero counts

Genome
Genes 356,359
Genomic sequences 22,633
Organism
Organisms 63
Transcriptomics
ESTs 287,336
Population biology
Popset isolate sequences 153,109

Compound - CHEBI:100000 (2S,3S,4R)-3-[4-(3-cyclopentylprop-1-ynyl)phenyl]-4-(hydroxymethyl)-1-(2-methoxy-1-oxoethyl)-2-azetidinecarboxylic acid
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-2-yl]propan-2-amine
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]hexa-2,4,9-trien-3-amine
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-2-yl]propan-2-amine
Compound - CHEBI:100005 N-[(1S,3S,4aS,9aR)-1-(hydroxymethyl)-3-[2-oxo-2-(1-piperidinyl)ethyl]-3,4,4a,9a-tetrahydro-1H-pyran-3,4-benzofuran-2-yl]propan-2-amine
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-4-one
Compound - CHEBI:100007 LSM-11386

All results matching *kinase

1 - 20 of 23,170

Export as a Search Strategy to download or mine your results

Filter results ☒ Hide zero counts

Genome
Genes 21,205
Population biology
Popset isolate sequences 1,273
Metabolism
Metabolic pathways 484
Compounds 204
Data access
Data sets 1
Searches 3

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)

9. 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. Although your search for *PF3D7_0310100* does return a direct link to the gene in *P. falciparum* 3D7, it also returns a link to the *P. gaboni* strain gene. Why?

PF3D7_0310100

My Organism Preferences (60 of 60) enabled

Genes matching **PF3D7_0310100**

Export as a Search Strategy to download or mine your results

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
☐ Names, IDs, and aliases 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
Gene type: protein coding gene
Organism: Plasmodium falciparum 3D7
Fields matched: External links; Gene ID; Names, IDs, and aliases

Gene - PGSV75_0310100 calcium-dependent protein kinase 3
Gene type: protein coding gene
Organism: Plasmodium gaboni strain SY75
Fields matched: Notes from annotators: gap found within coding sequence--ort

1 - 2 of 2

Direct link to PF3D7_0310100

Why is this gene returned by a search for PF3D7_0310100?