

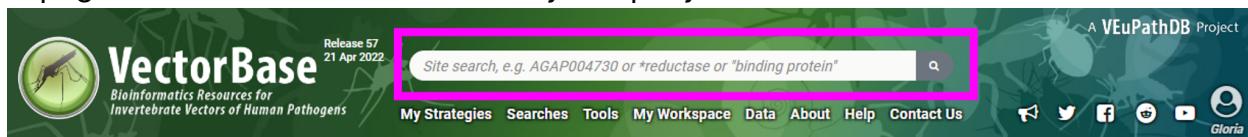
Site Search

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

Learning objectives:

- Use keywords in site search
- Explore site search results
- Filter site search results by categories
- Filter site search results by organisms
- Filter site search results by category fields
- Export results to a search strategy
- Find a specific gene using its ID in 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. Enter the word *kinase* (singular) 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).

A screenshot of the VectorBase.org homepage after performing a search for "kinase". The search results are displayed in a sidebar on the left under the heading "Search for...". The results include "Genes", "Organisms", "Genomic Sequences", "Genomic Segments", "ESTs", "Metabolic Pathways", and "Compounds". The "Genes" result is currently selected. The main content area shows an "Overview of Resources and Tools" with icons for "Take a Tour", "Getting Started" (which is highlighted), "Search Strategies", "Genome Browser", "Transcriptomic Resources", "MapVEu", and "Analyze My Data". Below this is a "Getting Started" section with a brief introduction to the site's resources and a "SITE SEARCH" section with instructions for using the search function.

2. The site search returns a categorized list of pages and documents that contain your term. Site search results are summarized by category, with a details panel on the right. Changing the panel on the left will populate the details panel with that list. What is the total number of results with the word kinase? Are all the results genes? Explore the filter panel on the left side of the webpage.

The screenshot shows the VectorBase search results for the query "kinase". The top navigation bar includes the VectorBase logo, release information (Release 57, 21 Apr 2022), a search bar containing "kinase", and various menu options like My Strategies, Searches, Tools, My Workspace, Data, About, Help, Contact Us, and social media links. A user profile icon for "Gloria" is also present.

The main content area displays the results under the heading "All results matching kinase". It shows a total of 1 - 20 of 72,615 results. The results are categorized into "Filter results" and "Filter fields".

Filter results:

Category	Count
Genome	72,615
Genes	72,090
Metabolism	328
Metabolic pathways	193
Data access	1
Data sets	3
Searches	

A pink arrow points from the "Genes" category to the "Genes" section in the "Filter fields" area.

Filter fields:

Select a result filter above

Filter organisms:

select all | clear all | expand all | collapse all

Taxonomic name	Count
Arthropoda	70,455
Mollusca	1,636

A pink bracket highlights the "Filter fields" and "Filter organisms" sections.

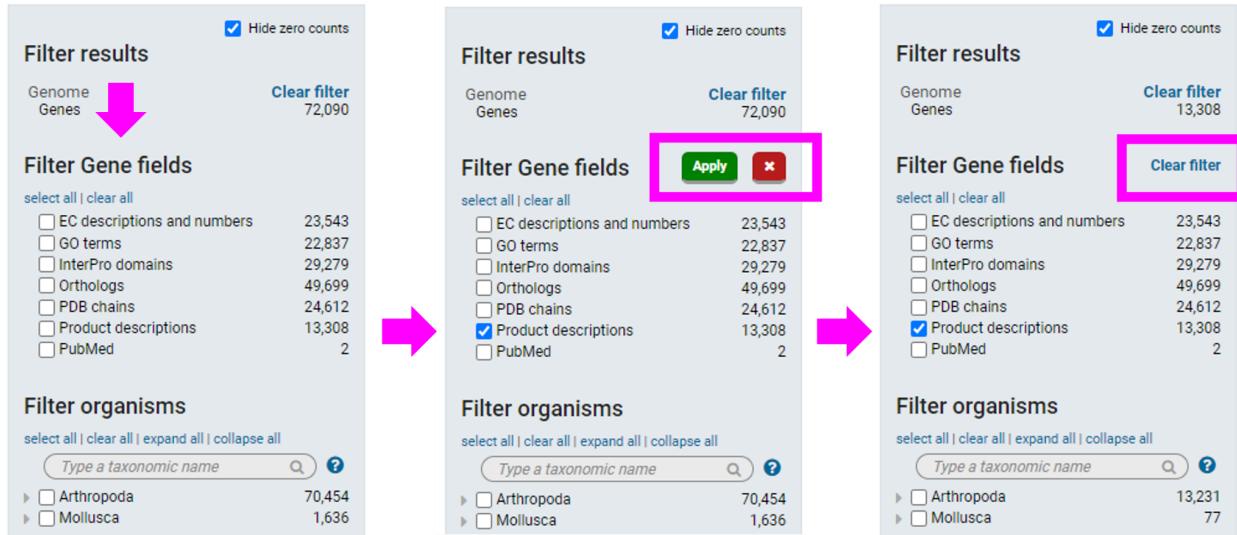
Results summarized in categories

Details panel with information about each item returned

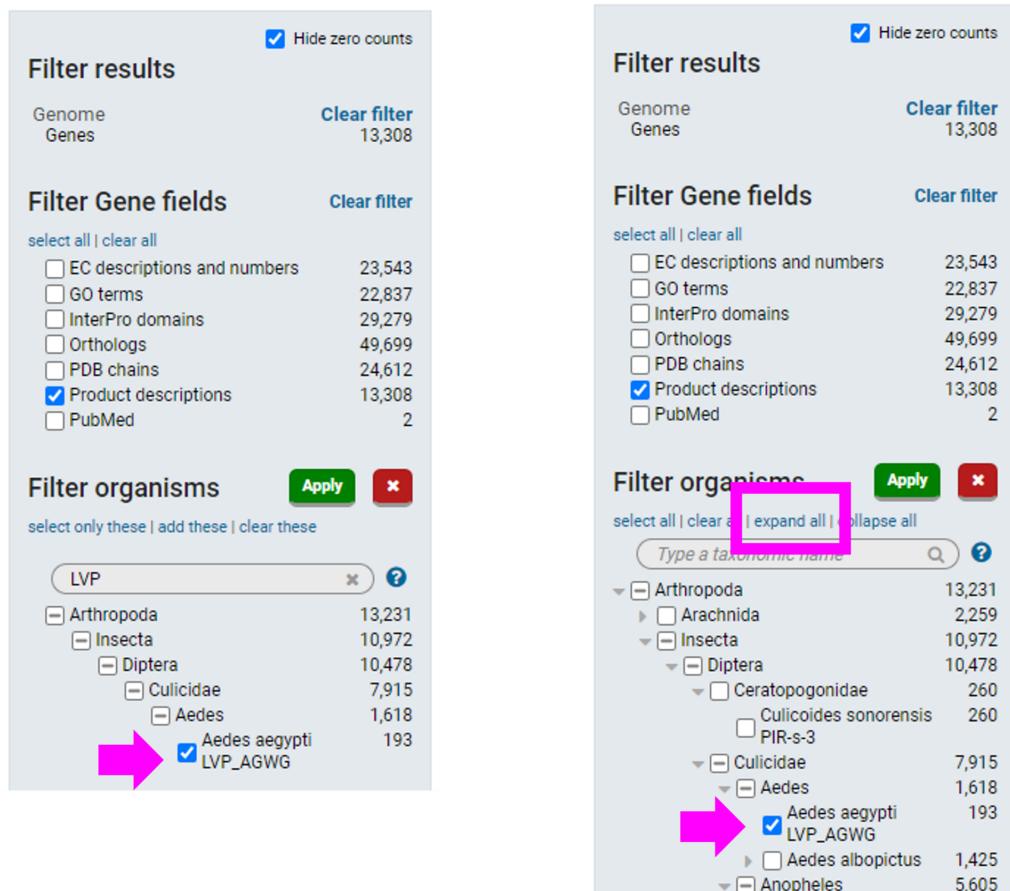
The results list shows several entries, each with a gene ID, name, organism, and a list of matched fields (e.g., GO terms, InterPro domains, Orthologs, PDB chains, Product descriptions). A pink bracket highlights the first few results in the list.

3. 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). How many of the genes included the word kinase in their product descriptions?

Notice that once you filter the result by genes (click on the *Genes* filter), the *Filter Fields* section expands to reveal additional filtering options. Select the *Product descriptions* field and Choose *Apply* this filter or cancel it (box middle panel below). Once a filter is applied it can be cleared by clicking on *Clear filter* (box left panel below).



4. How many of the above genes are found in *Aedes aegypti* LVP_AGWG? How did you find this number? Hint: explore the *Filter organisms* section of the results filter. There is a search option to aid navigation through the organism tree (left) or the tree can be expanded to find the organism of interest (right). Select the correct organism and apply the filter.

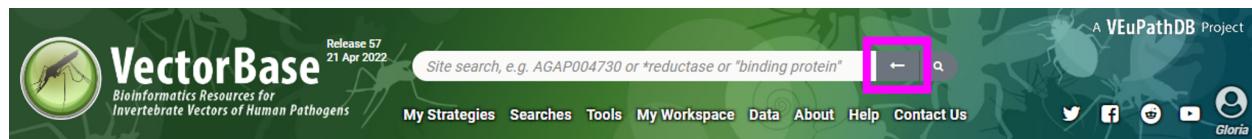


5. Export the results to a search strategy. Use the blue *Export as a search strategy* button at the top right-hand side of the results.

The screenshot shows a search interface for '193 Genes (145 ortholog groups)'. At the top, there's a blue button labeled 'Export as a Search Strategy' with a downward arrow icon. Below it is a pink arrow pointing down to the main content area. The content area includes an 'Organism Filter' sidebar on the left and a 'Gene Results' table on the right. The table has columns for Gene ID, Transcript ID, Gene Name or Symbol, Product Description, and Previous ID(s). The first four rows of the table are:

	Gene ID	Transcript ID	Gene Name or Symbol	Product Description	Previous ID(s)
	AAEL012094	AAEL012094-RA	N/A	casein kinase ii, alpha chain (cmgc group iv)	N/A
	AAEL009765	AAEL009765-RA	N/A	choline/ethanolamine kinase	N/A
	AAEL007662	AAEL007662-RM	N/A	casein kinase	N/A
	AAEL007651	AAEL007651-RC	N/A	phosphorylase b kinase	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

The screenshot shows the 'Filter results' panel with three 'Clear filter' buttons highlighted by a pink box. The first button is under 'Genome Genes'. The second button is under 'Filter Gene fields'. The third button is under 'Filter organisms'.

Filter results

Hide zero counts

Genome Genes Clear filter

Filter Gene fields Clear filter

select all | clear all

<input type="checkbox"/> EC descriptions and numbers	436
<input type="checkbox"/> GO terms	381
<input type="checkbox"/> InterPro domains	487
<input type="checkbox"/> Orthologs	815
<input type="checkbox"/> PDB chains	407
<input checked="" type="checkbox"/> Product descriptions	193
<input type="checkbox"/> PubMed	1

Filter organisms Clear filter

select all | clear all | expand all | collapse all

Type a taxonomic name Clear filters ?

▶ Arthropoda 13,231

▶ Mollusca 77

2

The screenshot shows the site search bar with the word 'kinase' entered. The 'Clear filters' button is highlighted by a pink box.

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

Filter results	
Genome Genes	72,090
Metabolism Metabolic pathways	328
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	
<input type="text" value="Type a taxonomic name"/> <input type="button" value="🔍"/> <input type="button" value="❓"/>	
▶ <input type="checkbox"/> Arthropoda	70,455
▶ <input type="checkbox"/> Mollusca	1,636

Filter results	
Genome Genes Genomic sequences	72,090 0
Organism Organisms	0
Transcriptomics ESTs	0
Population biology Popset isolate sequences	0
Field samples	0
Metabolism Metabolic pathways	328
Compounds	193
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	
<input type="text" value="Type a taxonomic name"/> <input type="button" value="🔍"/> <input type="button" value="❓"/>	
▶ <input type="checkbox"/> Arthropoda	70,455
▶ <input type="checkbox"/> Mollusca	1,636

- Try running a search with a wild card. The wild card is denoted by an asterisk *. The wild card can be used alone to retrieve all results available to the site search 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.

VectorBase Bioinformatics Resources for Invertebrate Vectors of Human Pathogens

Release 57
21 Apr 2022

My Strategies Searches Tools My Workspace Data About Help Contact Us Gloria

My Organism Preferences (58 of 58) enabled

All results matching *

Export as a Search Strategy to download or mine your results

1 - 20 of 6,308,870

Filter results	<input type="checkbox"/> Hide zero counts
Genome	
Genes	993,315
Genomic sequences	1,383,067
Organism	
Organisms	58
Transcriptomics	
ESTs	2,006,462
Population biology	
Popset isolate sequences	0
Field samples	1,766,004
Metabolism	
Metabolic pathways	3,267
Compounds	154,803
Data access	
Data sets	1,244
Searches	613
Instructional	
Tutorials	14
Workshop exercises	3
About	
News	2
General info pages	18
Filter fields	Select a result filter above

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-oxanil]-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-[(5S,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:100000 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-pyran-2,3-c-[1,5]oxazocin-8-yl]-1-(1-piperidinyl)ethanone

VectorBase Bioinformatics Resources for Invertebrate Vectors of Human Pathogens

Release 57
21 Apr 2022

My Strategies Searches Tools My Workspace Data About Help Contact Us Gloria

My Organism Preferences (58 of 58) enabled

All results matching *kinase

Export as a Search Strategy to download or mine your results

1 - 20 of 76,693

Filter results	<input type="checkbox"/> Hide zero counts
Genome	
Genes	76,032
Genomic sequences	0
Organism	
Organisms	0
Transcriptomics	
ESTs	0
Population biology	
Popset isolate sequences	0
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

Gene - AAEL000006 phosphoenolpyruvate carboxykinase

Organism: Aedes aegypti LVP_AWG

▶ Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions

Gene - AAEL000025 phosphoenolpyruvate carboxykinase

Organism: Aedes aegypti LVP_AWG

▶ Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions

Gene - AAEL000080 phosphoenolpyruvate carboxykinase

Organism: Aedes aegypti LVP_AWG

▶ Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions

Gene - AAEL000194 phosphatidylinositol 4-kinase

Organism: Aedes aegypti LVP_AWG

▶ Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions

Gene - AAEL000217 serine/threonine protein kinase

Organism: Aedes aegypti LVP_AWG

▶ Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions

Gene - AAEL000278 poly(p)/ATP NAD kinase

Organism: Aedes aegypti LVP_AWG

▶ Fields matched: EC descriptions and numbers; External links; GO terms; InterPro domains; Orthologs; PDB chains; Product descriptions

10. Try searching for a specific gene ID. Enter the gene ID below in the site search window: AGAP004707

The screenshot shows the VectorBase homepage with a green header. In the top right corner, there is a search bar containing the text "AGAP004707". A pink rectangular box highlights this search term. Below the header, there are navigation links: My Strategies, Searches, Tools, My Workspace, Data, About, Help, Contact Us, and social media icons for Twitter, Facebook, and YouTube, along with a "Gloria" logo. A blue button labeled "My Organism Preferences (58 of 58) enabled" is also visible.

The main content area displays the results for "Genes matching AGAP004707". It shows a single result card for "Gene - AGAP004707 voltage-gated sodium channel". The card includes the gene name or symbol ("para"), organism ("Anopheles gambiae PEST"), and a note about fields matched ("Gene ID; Names, IDs, and aliases"). To the left of the results, there are filter panels for "Filter results", "Filter Gene fields", and "Filter organisms". The "Filter results" panel shows a count of 1 for "Genome Genes". The "Filter Gene fields" panel shows filters for "Gene ID" and "Names, IDs, and aliases", both with a count of 1. The "Filter organisms" panel shows a taxonomy search field with "Type a taxonomic name" and a dropdown menu with "Arthropoda" and "Insecta", both with a count of 1.

When the query ID has an exact match in the database, the site search returns a card at the top of the details panel for easy access to the gene page. The site search also returns other pages that contain the query ID. Click on the Gene ID to go to the gene page.

Note - In VectorBase, every gene for every genome, uses a nomenclature system for the gene IDs based on each organism scientific name, strain and a consecutive number. Examples:

	gene ID	Transcript ID (use the gene stable id + "R" + a number or letter)	Protein ID (use the same id as their transcript, substituting "P" for the "R")
<u>Aedes albopictus</u> Foshan <u>FPA</u>	<u>AALFPA</u> _058096	AALFPA_058096.R 24198	AALFPA_058096.P 24198
<u>Anopheles gambiae</u> <u>PEST</u>	<u>AGAP</u> 004707	AGAP004707-RA	AGAP004707-PA