

## Site Search

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

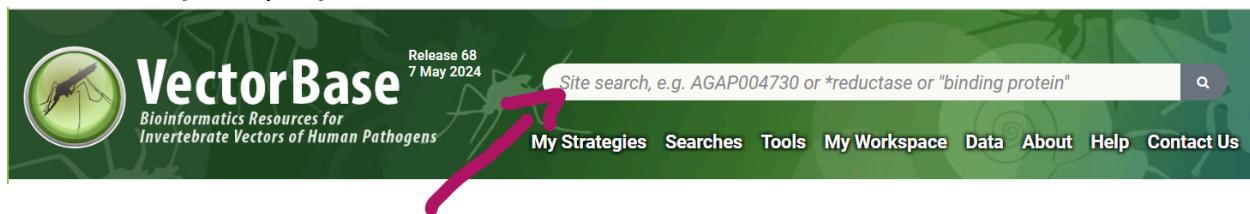
### What is the VectorBase Site Search?

- The site search queries the database for your search term (keyword or identifier) and returns a list of pages and documents that contain your query term
- The site search is a convenient starting point for mining the vast amount of data available in VectorBase and identifying and filtering information relevant to you
- The site search box is located in the header of the VectorBase site and is available on every page

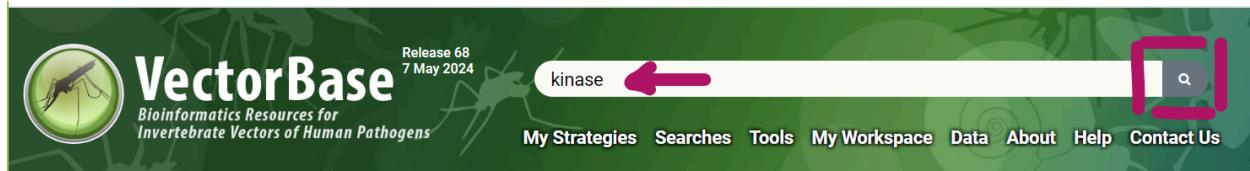
### Learning objectives:

- Search by keywords (e.g., gene name/symbol or function/description) or identifiers (e.g., gene IDs)
- Filter site search results by categories and fields
- Export results to a search strategy
- Find a specific gene using the gene ID in site search
- Navigate to and from the site search result
- Explore searches using wildcards (\*)

The site search is located in the header 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 **VectorBase.org** and search for a keyword. 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).



2. **Site Search result format:** The site search returns a categorized list of pages and documents that contain your search term. Site search results are presented with a **summary panel on the left** and a **details panel on the right**. Clicking on a particular result type on the left will populate the details panel with that result.

- a. What is the total number of results with the word kinase?
- b. Are all the results genes? (Red circle in image below)

The screenshot shows the VectorBase website interface. At the top, there is a navigation bar with links: My Strategies, Searches, Tools, My Workspace, Data, About, Help, Contact Us, and a search bar containing the word "kinase". Below the navigation bar, a green banner says "All results matching kinase". The main content area displays a table of search results. The first row shows "Gene - LLOJ007004 MAP kinase kinase kinase Mekk1" with a count of 110,983. A red circle highlights the number "110,983". To the left of the table, under "Filter results", there is a section for "Genes" which is highlighted with a red arrow. The table also includes rows for "Metabolism", "Compounds", "Data access", "Data sets", and "Searches". On the right side, a "Details panel with information about item returned" is shown for the first gene entry, displaying its name, type, organism, and fields matched. A red bracket groups the "Details panel" and the "Summary panel with categories of results" (which is located at the bottom left of the main content area).

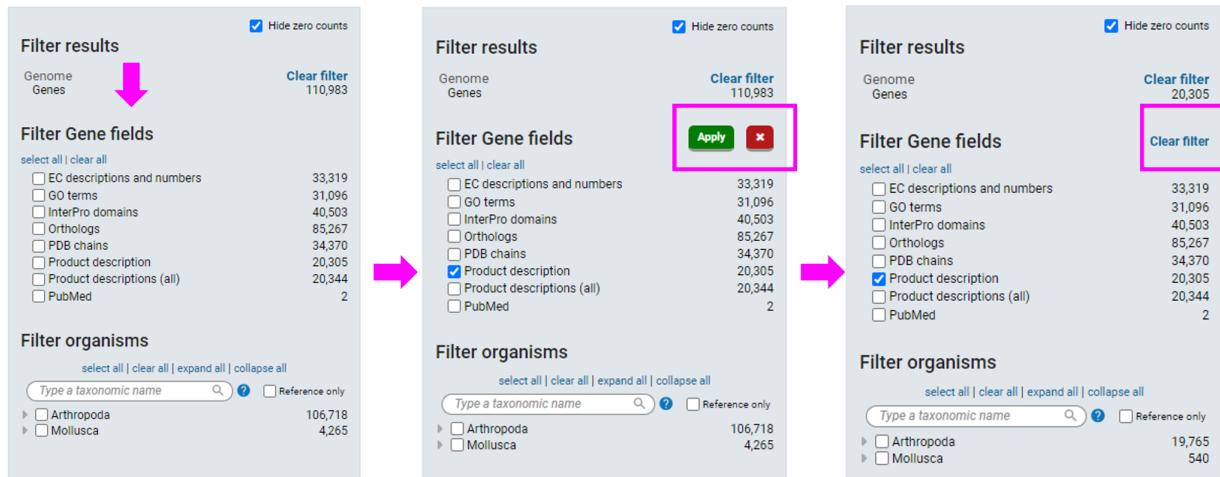
All results matching kinase	
1 - 20 of 111,508	
<b>Filter results</b> <input checked="" type="checkbox"/> Hide zero counts Genome <b>Genes</b> Metabolism Compounds Data access Data sets Searches  <b>Filter fields</b> Select a result filter above  <b>Filter organisms</b> select all   clear all   expand all   collapse all Type a taxonomic name <input type="text"/> <input type="checkbox"/> Reference only Arthropoda 106,719 Mollusca 4,265	
<b>Gene - LLOJ007004</b> MAP kinase kinase kinase Mekk1 Gene name or symbol: Mekk1 Gene type: protein coding gene Organism: <i>Lutzomyia longipalpis</i> Jacobina Fields matched: GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptor	
Gene - AALC636_010967 mitogen-activated protein kinase kinase kinase 4-like Gene type: protein coding gene Organism: <i>Aedes albopictus</i> C6/36 cell line Fields matched: Orthologs; Product description; Product descriptions (all)	
Gene - AALFPA_056150 mitogen-activated protein kinase kinase kinase 3-like Gene type: protein coding gene Organism: <i>Aedes albopictus</i> Foshan FPA Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptor	
Gene - AALFPA_058996 mitogen-activated protein kinase kinase kinase 4 Gene type: protein coding gene Organism: <i>Aedes albopictus</i> Foshan FPA Fields matched: EC descriptions and numbers; InterPro domains; Orthologs; Product description; Product descriptions (all)	

**Summary panel with categories of results**

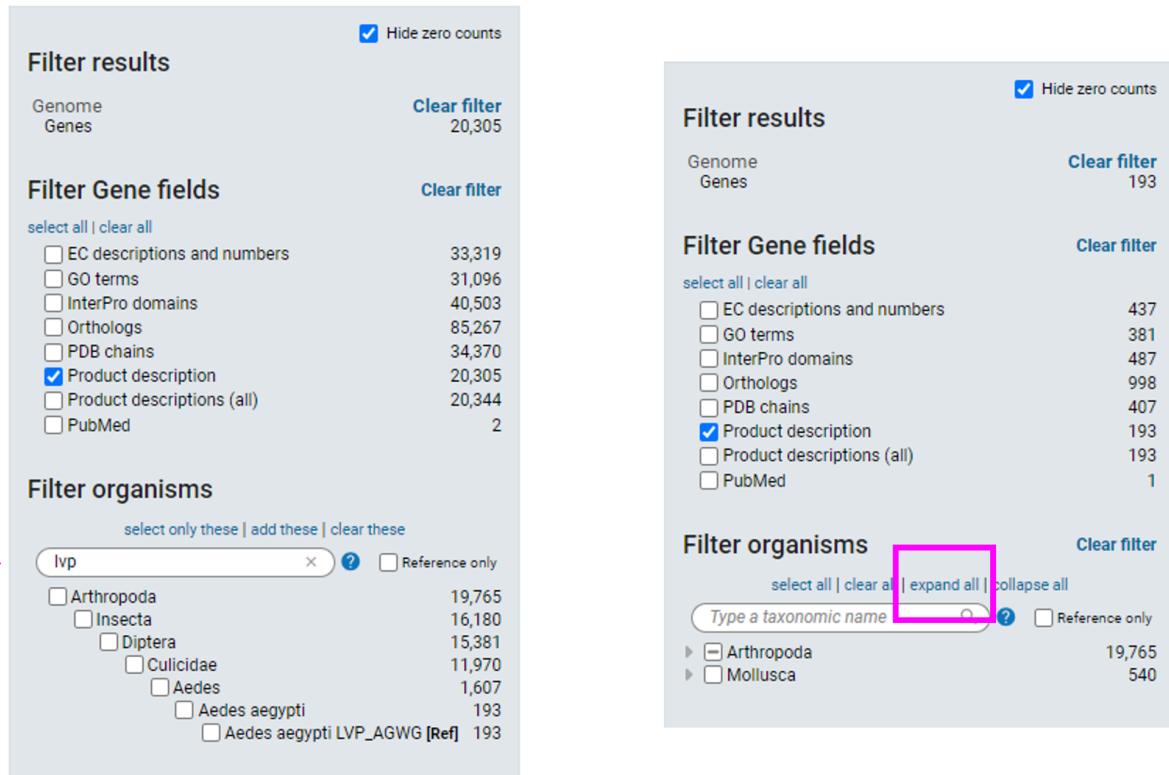
**Details panel with information about item returned**

3. **Filter the site search result by category:** Click on “genes” (arrow in image above) to view only results from that category. Notice that the **Filter Fields** section expands to reveal additional filtering options (categories).

- a. How many of the genes included the word kinase in their product descriptions? To find out, select the “Product description” field (arrow, left panel in image below) and choose Apply (arrow, middle panel in image below)
- b. Select the *Product Descriptions* field and choose *Apply*. Once a filter is applied it can be removed by clicking on *Clear filter* (circle, right panel in image 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.

Export as a Search Strategy ➔

↓

Unnamed Search Strategy \* [Edit](#)

**Step 1**

**Text** **193 Genes** [Add a step](#)

**193 Genes (145 ortholog groups)** [Revise this search](#)

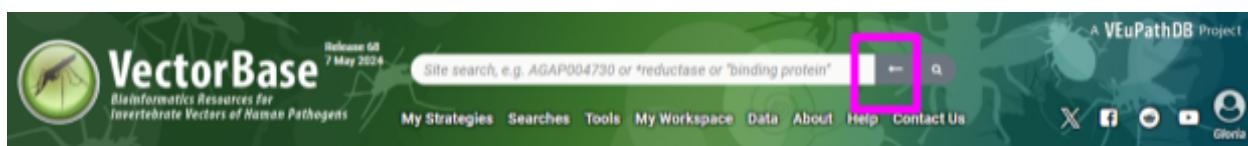
**Gene Results** [Genome View](#) [Analyze Results](#)

Genes: 193 Transcripts: 535  Show Only One Transcript Per Gene  Show only the Genes in my basket.

Rows per page: 1000 [Download](#) [Send to...](#) [Add Columns](#)

Gene ID	Transcript ID	Gene Name or Symbol	Product Description
AAEL002026	AAEL002026-RA	N/A	protein serine/threonine kinase, putative
AAEL005682	AAEL005682-RA	N/A	protein serine/threonine kinase, putative
AAEL005687	AAEL005687-RA	N/A	protein serine/threonine kinase, putative
AAEL005691	AAEL005691-RA	N/A	protein serine/threonine kinase, putative
AAEL005710	AAEL005710-RB	N/A	protein serine/threonine kinase, putative
AAEL005711	AAEL005711-RA	N/A	protein serine/threonine kinase, putative
AAEL001963	AAEL001963-RA	N/A	protein serine/threonine kinase, 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.

**Filter results**

Genome Genes  Hide zero counts

**Filter Gene fields**

select all | clear all

<input type="checkbox"/> EC descriptions and numbers	437
<input type="checkbox"/> GO terms	381
<input type="checkbox"/> InterPro domains	487
<input type="checkbox"/> Orthologs	998
<input type="checkbox"/> PDB chains	407
<input checked="" type="checkbox"/> Product description	193
<input type="checkbox"/> Product descriptions (all)	193
<input type="checkbox"/> PubMed	1

**Filter organisms**

select all | clear all | expand all | collapse all

Type a taxonomic name    Reference only

Arthropoda	19,765
Mollusca	540

option  
1



option  
2

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

The screenshot shows two identical-looking filter panels side-by-side. Each panel has a header 'Filter results' at the top. Below the header, there are several sections: 'Genome', 'Metabolism', 'Data access', 'Filter fields', 'Filter organisms', and 'About'. Each section contains a list of items with counts. A pink box highlights the 'Hide zero counts' checkbox in both the left panel and the top right corner of the right panel.

Category	Item	Count
Genome	Genes	110,983
	Genomic sequences	0
Metabolism	Metabolic pathways	328
	Compounds	193
Data access	Data sets	1
	Searches	3
<b>Filter fields</b>		
Select a result filter above		
<b>Filter organisms</b>		
select all   clear all   expand all   collapse all		
<input type="text" value="Type a taxonomic name"/> <input type="button" value="🔍"/> <input type="checkbox"/> Reference only		
> <input type="checkbox"/> Arthropoda 106,719 > <input type="checkbox"/> Mollusca 4,265		
<b>Filter results</b>		
<input type="checkbox"/> Hide zero counts		
Genome	Genes	110,983
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
<b>Filter fields</b>		
Select a result filter above		
<b>Filter organisms</b>		
select all   clear all   expand all   collapse all		
<input type="text" value="Type a taxonomic name"/> <input type="button" value="🔍"/> <input type="checkbox"/> Reference only		
> <input type="checkbox"/> Arthropoda 106,719 > <input type="checkbox"/> Mollusca 4,265		

9. 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-[(1S,3S,4S,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-[(1S,3S,4S,9aR)-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 search bar containing "AGAP004707". A pink box highlights the search term in the bar. Below the search bar, there's a navigation menu with links like "My Strategies", "Searches", "Tools", "My Workspace", "Data", "About", "Help", and "Contact Us". On the right side, there are social media icons for X, Facebook, Twitter, and YouTube, and a link to "Gloria". The main content area displays the results for "Genes matching AGAP004707". It shows one result card for "Gene - AGAP004707 voltage-gated sodium channel". The card includes details: Gene name or symbol: para; Gene type: protein coding gene; Organism: Anopheles gambiae PEST. Below the card, there are two more collapsed sections for other matches. At the bottom left, there are filter results for "Genome Genes" (1 result), "Filter Gene fields" (with options for "Gene ID" and "Names, IDs, and aliases"), and "Filter organisms" (with a dropdown for "Arthropoda" and "Insecta"). At the bottom right, there's a button to "Export as a Search Strategy".

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

