

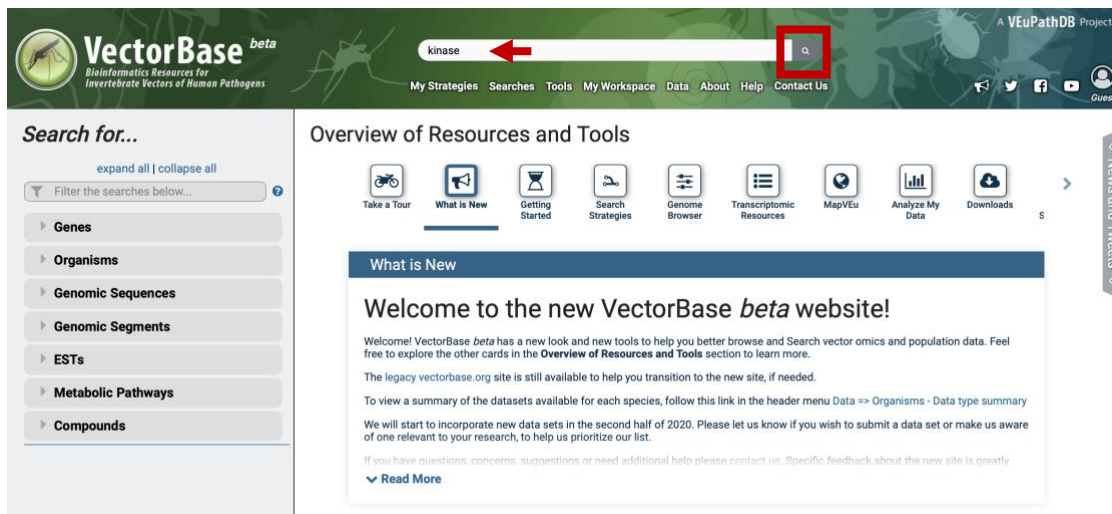
Site Search

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

Learning objectives:

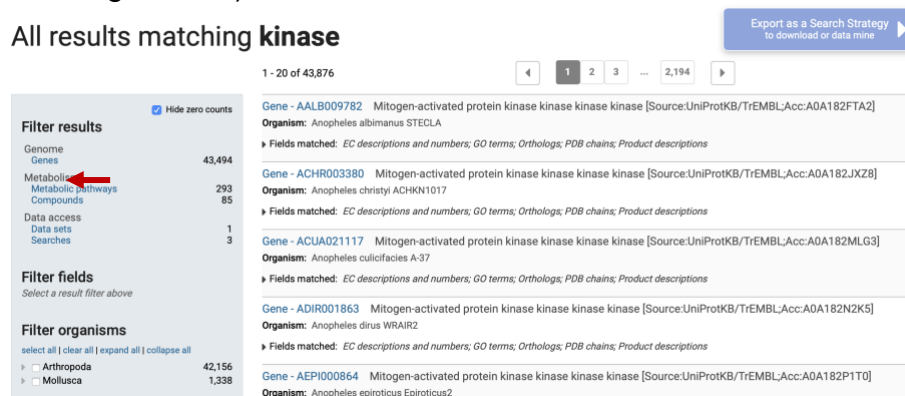
- Learn how to run a site search
- Learn how to explore site search results
- Learn how to filter results by categories
- Learn how to filter results by organisms
- Learn how to filter results by category fields
- Learn how to export results to a search strategy
- Learn how to find a specific gene using its ID

1. Enter the word *kinase* in the site search window (top center of the page, arrow in the image below). Then click enter on your keyboard or click on the search icon (square in the image below).



2. How many results with the word *kinase* did you get? Are all the results genes? Explore the filter panel on the left side of the webpage. Filter the results so that you only view gene results (hint: click on the word *genes* in the *Filter results* section; arrow in image below).

All results matching **kinase**



- 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 fields section expands to reveal additional filtering options. Once you select the *Product descriptions* field you are provided the option to *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).

Filter results ☒ Hide zero counts

Genome **Genes** [Clear filter](#) 43,494

Filter Gene fields

[select all](#) | [clear all](#)

- ☐ EC descriptions and numbers 24,315
- ☐ GO terms 16,978
- ☐ Orthologs 25,638
- ☐ PDB chains 17,900
- ☐ Product descriptions 8,085
- ☐ PubMed 3

Filter organisms

[select all](#) | [clear all](#) | [expand all](#) | [collapse all](#)

- ☐ Arthropoda 42,156
- ☐ Mollusca 1,338

- How many of the above genes are found in *Anopheles gambiae* str. PEST? How did you find this number? (hint: explore the *Filter organisms* section of the results filter). Select the correct organism and apply the filter.
- Export the results to a search strategy. (hint: to achieve this click on the blue *Export as a search strategy* button at the top right-hand side of the results).

Export as a Search Strategy

to download or data mine

Filter organisms [Apply](#) [x](#)

[select all](#) | [clear all](#) | [expand all](#) | [collapse all](#)

- ☒ Arthropoda 8,003
 - ☐ Arachnida 1,072
 - ☒ Insecta 6,931
 - ☒ Diptera 6,542
 - ☐ Culicidae 4,698
 - ☐ Aedes 386
 - ☒ Anopheles 3,923
 - ☐ Anopheles albimanus 173
 - ☐ Anopheles arabiensis 221
 - ☐ Anopheles atroparvus 203
 - ☐ Anopheles christyi 157
 - ☐ Anopheles coluzzii 192
 - ☐ Anopheles culicifacies 201
 - ☐ Anopheles darlingi 243
 - ☐ Anopheles dirus 184
 - ☐ Anopheles epiroticus 168
 - ☐ Anopheles farauti 212
 - ☐ Anopheles funestus 199
 - ☒ Anopheles gambiae 244
 - ☒ Anopheles gambiae str. PEST
 - ☐ Anopheles maculatus 130
 - ☐ Anopheles melas 201
 - ☐ Anopheles merus 212
 - ☐ Anopheles minimus 177
 - ☐ Anopheles quadriannulatus 200
 - ☐ Anopheles sinensis 402
 - ☐ Anopheles stephensi 204
 - ☐ Culex 389

My Search Strategies

[Opened \(1\)](#) [All \(1\)](#) [Public \(3\)](#) [Help](#)

Unnamed Search Strategy *

Text [Add a step](#)

Step 1

244 Genes (219 ortholog groups) [Revise this search](#)

Gene Results **Genome View** **Analyze Results**

Genes: 244 Transcripts: 310 ☐ Show Only One Transcript Per Gene

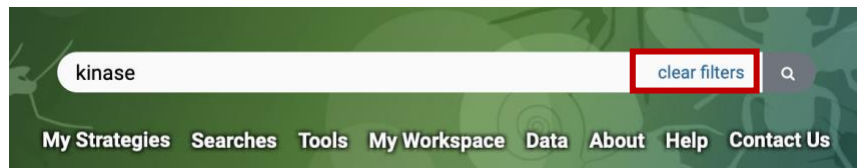
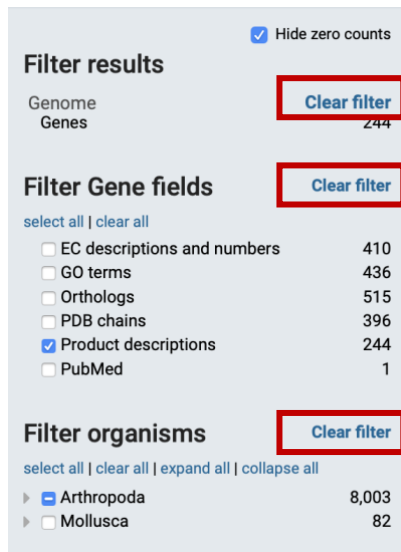
Rows per page: 20 [Download](#) [Add to Basket](#) [Add Columns](#)

Gene ID	Transcript ID	Organism	Genomic Location (Gene)	Product Description
AGAP004699	RA	Anopheles gambiae str. PEST	AgamP4_2L:1,973,601..1,976,987(+)	RAF proto-oncogene serine/protein kinase [Source:VB G. Annotation]

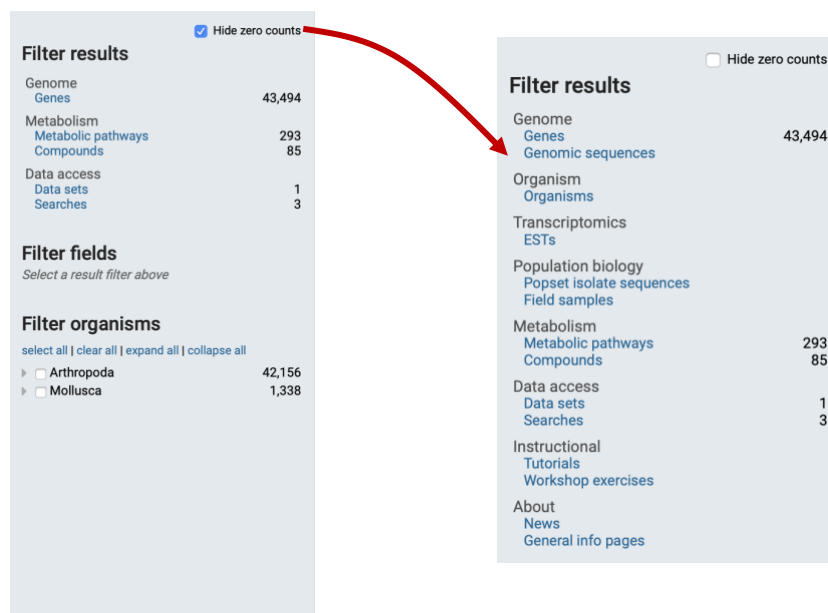
6. Return to the site search results page. How did you do this? (hint: you can achieve this in two ways: 1. Click on your browser's back arrow. 2. Click on the back to results arrow in the site search window. Notice that your previous results and filter settings were preserved.



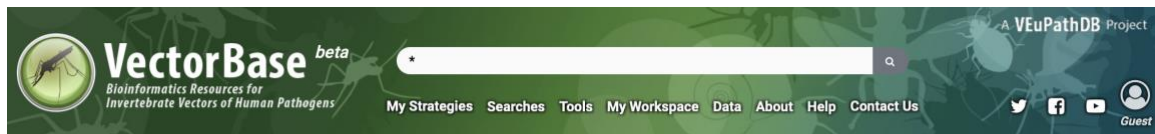
7. Clear all filters. How did you do this? (hint: you can achieve this in two ways: 1. You can click on each of the clear filter options in the filter results panel on the left (boxes below). 2. You can click on the single *clear filters* option in the site search window.



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



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.



All results matching *

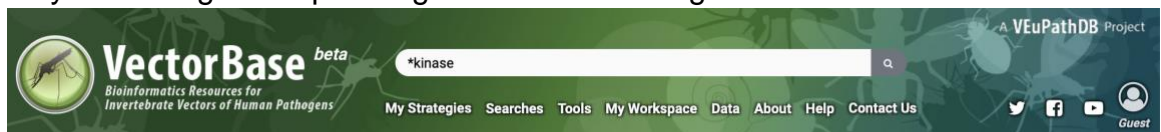
Export as a Search Strategy
to download or data mine

1 - 20 of 4,457,608

Filter results	
<input checked="" type="checkbox"/> Hide zero counts	
Genome	
Genes	671,737
Genomic sequences	1,334,334
Organism	
Organisms	43
Transcriptomics	
ESTs	1,118,822
Population biology	
Field samples	1,266,855
Metabolism	
Metabolic pathways	3,045
Compounds	61,998
Data access	
Data sets	419
Searches	331
Instructional	
Tutorials	13
About	
News	1
General info pages	10
Filter fields	
Select a result filter above	

Compound - CHEBI:10000	Vismione D
Compound - CHEBI:10001	Visnadin
Compound - CHEBI:10002	Visnagin
Compound - CHEBI:10003	ribostamycin sulfate
Definition: An aminoglycoside sulfate salt resulting from the reaction of ribostamycin with sulfuric acid.	
Compound - CHEBI:100147	nalidixic acid
Definition: A monocarboxylic acid comprising 1,8-naphthyridin-4-one substituted by carboxylic acid, ethyl and methyl groups at positions 3, 1, and 7, respectively.	
Compound - CHEBI:10014	Voacamine
Compound - CHEBI:10015	vobasine
Definition: An indole alkaloid that is vobasan in which the bridgehead methyl group is substituted by a methoxycarbonyl group and an additional oxo substituent is present in the 3-position.	
Compound - CHEBI:10016	vobtusine
Compound - CHEBI:10017	volemitol
Definition: A heptitol that is heptane-1,2,3,4,5,6,7-heptol that has R-configuration at positions 2, 3, 5 and 6.	
Compound - CHEBI:10018	volkenin
Definition: A cyanogenic glycoside that is (4R)-4-hydroxycyclopent-2-ene-1-carbonitrile attached to a beta-D-glucopyranosyloxy at position 1.	
Compound - CHEBI:10019	Vomicine
Compound - CHEBI:10022	Vomitoxin
Compound - CHEBI:10023	voriconazole

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



All results matching *kinase

Export as a Search Strategy
to download or data mine

1 - 20 of 45,121

Filter results	
<input checked="" type="checkbox"/> Hide zero counts	
Genome	
Genes	44,659
Metabolism	
Metabolic pathways	367
Compounds	91
Data access	
Data sets	1
Searches	3
Filter fields	
Select a result filter above	
Filter organisms	
select all clear all expand all collapse all	
Arthropoda	43,291
Mollusca	1,368

Gene - AAEL000006	phosphoenolpyruvate carboxykinase [Source:VB Community Annotation]
Organism: Aedes aegypti LVP_AGWG	
Fields matched: EC descriptions and numbers; GO terms; Orthologs; PDB chains; Product descriptions	
Gene - AAEL000025	phosphoenolpyruvate carboxykinase [Source:VB Community Annotation]
Organism: Aedes aegypti LVP_AGWG	
Fields matched: EC descriptions and numbers; GO terms; Orthologs; PDB chains; Product descriptions	
Gene - AAEL000080	phosphoenolpyruvate carboxykinase [Source:VB Community Annotation]
Organism: Aedes aegypti LVP_AGWG	
Fields matched: EC descriptions and numbers; GO terms; Orthologs; PDB chains; Product descriptions	
Gene - AAEL000194	phosphatidylinositol 4-kinase [Source:VB Community Annotation]
Organism: Aedes aegypti LVP_AGWG	
Fields matched: EC descriptions and numbers; GO terms; Orthologs; PDB chains; Product descriptions	
Gene - AAEL000217	serine/threonine protein kinase [Source:VB Community Annotation]
Organism: Aedes aegypti LVP_AGWG	

window:

AAEL007018

The screenshot shows the VectorBase website interface. At the top, there is a green header with the VectorBase logo and navigation links. A search bar contains the text 'AAEL007018'. Below the header, the main content area displays 'Genes matching AAEL007018 (filtered by organisms)' with a button to 'Export as a Search Strategy'. On the left, there are filter panels for 'Filter results', 'Filter Gene fields', and 'Filter organisms'. The 'Filter results' panel shows 'Genome' and 'Genes' with a 'Clear filter' button. The 'Filter Gene fields' panel shows 'Gene ID' and 'Transcripts' with a 'Clear filter' button. The 'Filter organisms' panel shows 'Arthropoda' and 'Insecta' with a 'Clear filter' button. The main results area shows a single result for 'Gene - AAEL007018' with the description 'udp-glucose 4-epimerase [Source:VB Community Annotation]' and 'Organism: Aedes aegypti LVP_AGWG'. The result is highlighted with a blue border.

VectorBase *beta*
Bioinformatics Resources for
Invertebrate Vectors of Human Pathogens

AAEL007018 clear filters

My Strategies Searches Tools My Workspace Data About Help Contact Us

VEuPathDB Project

Genes matching **AAEL007018** (filtered by organisms) Export as a Search Strategy to download or data mine

1 - 1 of 1

Filter results Hide zero counts
Genome Clear filter 1
Genes

Filter Gene fields
select all | clear all
Gene ID 1
Transcripts 1

Filter organisms Clear filter
select all | clear all | expand all | collapse all
Arthropoda 1
Insecta 1

Gene - AAEL007018 udp-glucose 4-epimerase [Source:VB Community Annotation]
Organism: Aedes aegypti LVP_AGWG
Fields matched: Gene ID; Transcripts

Gene - AAEL007018 udp-glucose 4-epimerase [Source:VB Community Annotation]
Organism: Aedes aegypti LVP_AGWG
Fields matched: Gene ID; Transcripts

Notice that the gene of interest appears at the top for easy access. You can click on the Gene ID to go the gene page.