

Site Search

Note: this exercise uses VectorBase 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

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).

The screenshot shows the VectorBase beta website. At the top, there is a search bar with the word "kinase" typed in. An arrow points to the search bar from the left, and a red square highlights the search icon (magnifying glass) on the right. Below the search bar, the navigation menu includes "My Strategies", "Searches", "Tools", "My Workspace", "Data", "About", "Help", and "Contact Us". On the far right, there are social media icons and a "Guest" link. The main content area features a "Search for..." panel on the left with categories like Genes, Organisms, Genomic Sequences, etc., and an "Overview of Resources and Tools" section on the right with various links such as "Take a Tour", "What is New" (which is highlighted in blue), "Getting Started", "Search Strategies", "Genome Browser", "Translational Resources", "MapVEU", "Analyze My Data", and "Downloads". A sidebar on the right lists "New Tools" and "Recent Changes".

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

The screenshot shows the search results page for "kinase". At the top, it says "All results matching kinase" and "1 - 20 of 43,876". There is a "Filter results" section with a dropdown menu set to "Genome". Other options include "Metabolism", "Metabolic pathways", and "Compounds". Below this is a "Filter fields" section with a dropdown menu set to "Select a result filter above". Under "Filter organisms", there are checkboxes for "Arthropoda" and "Mollusca". The main content area lists several gene entries, each with a link to its details page. The first entry is "Gene - AALB009782 Mitogen-activated protein kinase kinase kinase kinase [Source:UniProtKB/TrEMBL;Acc:A0A182FTA2]". The last entry shown is "Gene - AEPI000864 Mitogen-activated protein kinase kinase kinase kinase [Source:UniProtKB/TrEMBL;Acc:A0A182P1T0]".

only view gene results (hint: click on the word *genes* in the *Filter results* section; arrow in image below).

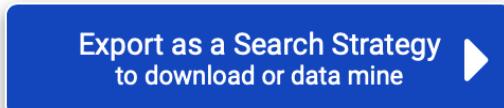
- How many of the genes included the word kinase in their product descriptions? Notice that once you filter the result by genes (click on the *Genes* filter), the fields section expands to reveal additional filtering options. Once you select the *Product descriptions* field you are provided the option 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).

Panel	Initial State	After Applying 'Product descriptions'	After Clearing Filter
Filter results	Genome Genes Clear filter 43,494	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	Genome Genes Clear filter 8,085 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	select all clear all expand all collapse all Arthropoda 42,156 Mollusca 1,338	select all clear all expand all collapse all Arthropoda 8,003 Mollusca 82

- 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.

Organism	Count
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
str. PEST	244
Anopheles maculatus	130
Anopheles melas	201
Anopheles merus	212
Anopheles minimus	177
Anopheles quadriannulatus	200
Anopheles sinensis	402
Anopheles stephensi	204
Culex	389

5. 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).



My Search Strategies

Opened (1) All (1) Public (3) Help

Unnamed Search Strategy * [Edit](#)

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

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?

The image shows two side-by-side 'Filter results' panels. The left panel has the 'Hide zero counts' checkbox checked (indicated by a blue checkmark), while the right panel has it unchecked (indicated by a grey square). A red arrow points from the left panel to the right panel, highlighting the difference in the data displayed.

Filter results (Left Panel, Hide zero counts checked):

Category	Sub-categories	Count
Genome	Genes	43,494
Metabolism	Metabolic pathways	293
	Compounds	85
Data access	Data sets	1
	Searches	3

Filter fields
Select a result filter above

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

Organism Group	Count
Arthropoda	42,156
Mollusca	1,338

Filter results (Right Panel, Hide zero counts unchecked):

Category	Sub-categories	Count
Genome	Genes	43,494
	Genomic sequences	85
Organism	Organisms	293
Transcriptomics	ESTs	85
Population biology	Popset isolate sequences	1
	Field samples	3
Metabolism	Metabolic pathways	293
	Compounds	85
Data access	Data sets	1
	Searches	3
Instructional	Tutorials	293
	Workshop exercises	85
About	News	1
	General info pages	3

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 *

1 - 20 of 4,457,608

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, 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

All results matching *kinase

1 - 20 of 45,121

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

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

AAEL007018

The screenshot shows the VectorBase beta search results for the gene ID AAEL007018. The search bar at the top contains "AAEL007018". Below the search bar, there are several tabs: My Strategies, Searches, Tools, My Workspace, Data, About, Help, and Contact Us. On the right side, there are social media links for Twitter, Facebook, YouTube, and a "Guest" link. A blue button on the right says "Export as a Search Strategy" with a "▶" icon. The main content area displays the results for "Genes matching AAEL007018 (filtered by organisms)". It shows one result: "Gene - AAEL007018 udp-glucose 4-epimerase [Source:VB Community Annotation]". Below this, it specifies the organism: "Organism: Aedes aegypti LVP_AGWG". A note indicates "Fields matched: Gene ID; Transcripts". To the left of the results, there are three filter panels: "Filter results" (Genome Genes), "Filter Gene fields" (Gene ID, Transcripts), and "Filter organisms" (Anthropoda, Insecta). The "Clear filter" button is located at the top of each panel.

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.