

Site Search

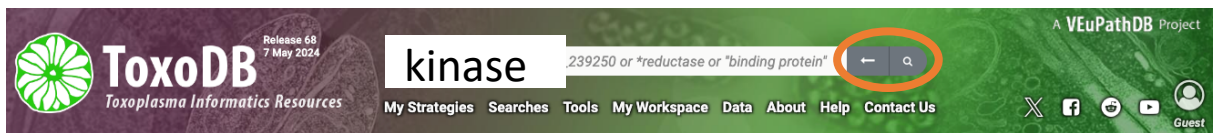
Note: this exercise uses ToxoDB.org as an example database, but the same functionality is available on all VEuPathDB resources.

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 (*) and quotation marks (" ")

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. **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.



2. **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 number of results with the word kinase? Are all the results genes?

All results matching kinase

Export as a Search Strategy
to download or mine your results

1 - 20 of 16,569

Filter results

Genome

Genes

Population biology

Popset isolate sequences

Metabolism

Metabolic pathways

Compounds

Data access

Data sets

Searches

15,909

110

350

193

3

4

Filter fields

Select a result filter above

Filter organisms

select all | clear all | expand all | collapse all

Type a taxonomic name

☐ Eimeriidae

4,643

☐ Sarcocystidae

11,269

Gene - LOC34620361

mitogen-activated protein kinase kinase kinase 3

Gene type: protein coding gene

Organism: Cyclospora cayentanensis isolate NF1_C8

Fields matched: GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)

Gene - BESB_003070

protein kinase

Gene type: protein coding gene

Organism: Besnoitia besnoiti strain Bb-Ger1

Fields matched: EC descriptions and numbers; GO terms; Orthologs; Product description; Product descriptions (all)

Gene - BESB_007440

Wee kinase

Gene type: protein coding gene

Organism: Besnoitia besnoiti strain Bb-Ger1

Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)

Gene - BESB_010200

protein kinase

Gene type: protein coding gene

Organism: Besnoitia besnoiti strain Bb-Ger1

Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; Product description; Product descriptions (all)

Gene - BESB_012260

CMGC kinase

Gene type: protein coding gene

Organism: Besnoitia besnoiti strain Bb-Ger1

Results are summarized by category

Details panel with information about each item returned

- Filter the site search result by category:** How many of the genes included the word kinase in their product descriptions?

Filter the results so that you 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).

Filter results

Genome

Genes

Clear filter

6,374

Filter Gene fields

select all | clear all

☐ Apollo Annotations

1

☐ EC descriptions and numbers

6,921

☐ GO terms

8,593

☐ InterPro domains

9,434

☐ Notes from annotators

192

☐ Orthologs

11,303

☐ PDB chains

6,106

☐ Product description

6,374

☐ Product descriptions (all)

6,386

☐ PubMed

36

☐ User comments

526

☐ Apollo Annotations

1

☐ EC descriptions and numbers

6,921

☐ GO terms

8,593

☐ InterPro domains

9,434

☐ Notes from annotators

192

☐ Orthologs

11,303

☐ PDB chains

6,106

☒ Product description

6,374

☐ Product descriptions (all)

6,386

☐ PubMed

36

☐ User comments

526

Filter organisms

select all | clear all | expand all | collapse all

Type a taxonomic name

☐ Eimeriidae

1,416

☐ Sarcocystidae

4,958

Filter results

Genome

Genes

Clear filter

15,909

Filter Gene fields

select all | clear all

☐ Apollo Annotations

1

☐ EC descriptions and numbers

6,921

☐ GO terms

8,593

☐ InterPro domains

9,434

☐ Notes from annotators

192

☐ Orthologs

11,303

☐ PDB chains

6,106

☒ Product description

6,374

☐ Product descriptions (all)

6,386

☐ PubMed

36

☐ User comments

526

☐ Apollo Annotations

1

☐ EC descriptions and numbers

6,921

☐ GO terms

8,593

☐ InterPro domains

9,434

☐ Notes from annotators

192

☐ Orthologs

11,303

☐ PDB chains

6,106

☐ Product description

6,374

☐ Product descriptions (all)

6,386

☐ PubMed

36

☐ User comments

526

Filter organisms

select all | clear all | expand all | collapse all

Type a taxonomic name

☐ Eimeriidae

4,643

☐ Sarcocystidae

11,266

Filter results

Genome

Genes

Clear filter

6,374

Filter Gene fields

select all | clear all

☐ Apollo Annotations

1

☐ EC descriptions and numbers

6,921

☐ GO terms

8,593

☐ InterPro domains

9,434

☐ Notes from annotators

192

☐ Orthologs

11,303

☐ PDB chains

6,106

☒ Product description

6,374

☐ Product descriptions (all)

6,386

☐ PubMed

36

☐ User comments

526

☐ Apollo Annotations

1

☐ EC descriptions and numbers

6,921

☐ GO terms

8,593

☐ InterPro domains

9,434

☐ Notes from annotators

192

☐ Orthologs

11,303

☐ PDB chains

6,106

☐ Product description

6,374

☐ Product descriptions (all)

6,386

☐ PubMed

36

☐ User comments

526

Filter organisms

select all | clear all | expand all | collapse all

Type a taxonomic name

☐ Eimeriidae

1,416

☐ Sarcocystidae

4,958

Did you notice that there are two product description search options? The search option *Product description* searches for the preferred product displayed on the top of the gene record page.

TGME49_259830 diacylglycerol kinase DGK2

There are genes that have more than one product description, i.e. from different sources. The product descriptions are all listed in the *All Product Descriptions* table on the gene record page. Use the search option *Product descriptions (all)* to search all product descriptions listed in the table.

2 Annotation, curation and identifiers

▼ All Product Descriptions **Data sets**

This table provides a list of product descriptions from all relevant sources. The most commonly accepted product description also appears at the top of this gene page.

Search this table... 2 rows

Product	Alternate?	Evidence Code	Reference	Support for Evidence Code Assignment	Transcript ID(s)	Assigned By
basal complex component BCC2	No	IDA	PMID:36310866	N/A	TGME49_231070.R1401	VEuPathDB
protein kinase	Yes	N/A	N/A	N/A	TGME49_231070.R1401	GenBank

4. **Filter the site search result by organism:** How many of the above genes are found in *Toxoplasma gondii* ME49? Explore the *Filter organisms* section of the results filter and use the search filter to navigate the tree.

The image shows two side-by-side screenshots of a web interface for filtering search results. Both screenshots have a 'Filter results' header with a 'Hide zero counts' checkbox checked. Below the header, there are sections for 'Filter Gene fields' and 'Filter organisms'.

Filter Gene fields: This section contains a list of gene fields with checkboxes and counts. The 'Product description' field is checked in both screenshots, showing a count of 6,374. Other fields include Apollo Annotations (1), EC descriptions and numbers (6,921), GO terms (8,593), InterPro domains (9,434), Notes from annotators (192), Orthologs (11,303), PDB chains (6,106), Product descriptions (all) (6,386), PubMed (36), and User comments (526).

Filter organisms: This section contains a search box with 'ME49' entered and a 'Reference only' checkbox. Below the search box, there is a list of organisms with checkboxes and counts. In the left screenshot, 'Toxoplasma gondii ME49 [Ref]' is selected. In the right screenshot, the 'Apply' button is highlighted in a red box.

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. Once exported you will be able to 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)

1 - 20 of 228

Filter results

Hide zero counts ☒ Clear filter 228

Gene: TGME49_205550 AGC kinase
Gene type: protein coding gene
Organism: Toxoplasma gondii ME49
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologous groups; Product description; Product descriptions (all)

Export as a Search Strategy
to download & reuse your results

Unnamed Search Strategy *

Text 228 Genes Step 1 + Add a step

228 Genes (188 ortholog groups) [Revise this search](#)

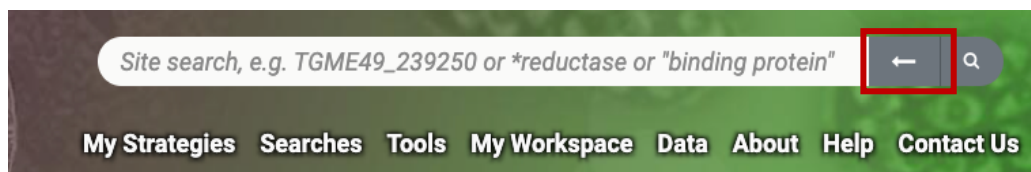
Gene Results **Genome View** **Analyze Results**

Genes: 228 Transcripts: 229 ☐ Show Only One Transcript Per Gene

Rows per page: 20

Gene ID	Transcript ID	Organism	Genomic Location (Gene)	Product Description
TGME49_322000	TGME49_322000.t26_1	Toxoplasma gondii ME49	KE138866.6,374..8,567(+)	myosin-light-chain kinase
TGME49_322010	TGME49_322010.t26_1	Toxoplasma gondii ME49	KE138866.8,748..10,926(+)	myosin-light-chain kinase
TGME49_322100	TGME49_322100.t26_1	Toxoplasma gondii ME49	KE138868.2,030..3,655(-)	myosin-light-chain kinase
TGME49_322120	TGME49_322120.t26_1	Toxoplasma gondii ME49	KE138868.11,137..11,931(-)	rho-kinase family protein
TGME49_323300	TGME49_323300.t26_1	Toxoplasma gondii ME49	KE139140.1..600(+)	myosin-light-chain kinase
TGME49_293260	TGME49_293260.R3284	Toxoplasma gondii ME49	TGME49_chrla:218,415..226,262(+)	ATPase/histidine kinase/DNA gyrase B/HSP90 dom

- 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. You can click on each of the *clear filter options* in the filter results panel. 2. You can click on the *clear filters option* in the site search window, which serves to Clear All filters.

1

2

Filter results

☒ Hide zero counts

Genome
Genes 228

Filter Gene fields

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

- ☐ EC descriptions and numbers 262
- ☐ GO terms 276
- ☐ InterPro domains 318
- ☐ Orthologs 322
- ☐ PDB chains 224
- ☒ Product description 228
- ☐ Product descriptions (all) 240
- ☐ PubMed 36
- ☐ User comments 135

Filter organisms

[select only these](#) | [add these](#) | [clear these](#)

ME49 x ? ☐ Reference only

- ☐ Sarcocystidae 4,958
- ☐ Toxoplasma 3,774
- ☒ Toxoplasma gondii ME49 [Ref] 228

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

1

2

Filter results

☒ Hide zero counts

Genome
Genes 15,909

Population biology
Popset isolate sequences 110

Metabolism
Metabolic pathways 350
Compounds 193

Data access
Data sets 3
Searches 4

Filter fields

Select a result filter above

Filter organisms

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

Type a taxonomic name ? ☐ Reference only

- ☐ Eimeriidae 4,643
- ☐ Sarcocystidae 11,269

Filter results

☐ Hide zero counts

Genome
Genes 15,909
Genomic sequences 0

Organism
Organisms 0

Transcriptomics
ESTs 0

Population biology
Popset isolate sequences 110
Field samples 0

Metabolism
Metabolic pathways 350
Compounds 193

Data access
Data sets 3
Searches 4

Instructional
Tutorials 0
Workshop exercises 0

About
News 0
General info pages 0

Filter fields

Select a result filter above

Filter organisms

9. **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 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. Feel free to compare the results you get when you run a search of the word *kinase* to a search with a wild card **kinase* or **kinase**.

All results matching *

1 - 20 of 901,526

Export as a Search Strategy to download or mine your results

☐ Hide zero counts

Filter results	
Genome	
Genes	285,644
Genomic sequences	147,712
Organism	
Organisms	37
Transcriptomics	
ESTs	288,077

Compound - CHEBI:100000	(2S,3S,4R)-3-[4-(3-cyclopentylprop-1-ynyl)phenyl]-4-(hydroxymethyl)-1-(2-met...
Compound - CHEBI:100001	N-[(2R,3S,6R)-2-(hydroxymethyl)-6-[2-[[oxo-[4-(trifluoromethyl)anilino]methyl]a...
Compound - CHEBI:100002	3-chloro-N-[(5S,6S,9S)-5-methoxy-3,6,9-trimethyl-2-oxo-11-oxa-3,8-diazabicycl...
Compound - CHEBI:100003	(4R,7S,8R)-8-methoxy-4,7,10-trimethyl-11-oxo-14-(1-oxobutylamino)-N-propyl...
Compound - CHEBI:100004	1-(2,5-difluorophenyl)-3-[(5S,6S,9S)-5-methoxy-3,6,9-trimethyl-2-oxo-8-oxo(2-...
Compound - CHEBI:100005	N-[(1S,3S,4aS,9aR)-1-(hydroxymethyl)-3-[2-oxo-2-(1-piperidinyl)ethyl]-3,4,4a,9...

All results matching *kinase

1 - 20 of 17,910

Export as a Search Strategy to download or mine your results

☒ Hide zero counts

Filter results	
Genome	
Genes	17,100
Population biology	
Popset isolate sequences	117
Metabolism	
Metabolic pathways	482
Compounds	204
Data access	
Data sets	3
Searches	4

Filter fields
Select a result filter above

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

Type a taxonomic name ☐ Reference only

<input type="checkbox"/> Eimeriidae	4,960
<input type="checkbox"/> Sarcocystidae	12,143

Gene - BESB_000930	rhoptry kinase family protein ROP21
Gene type:	protein coding gene
Organism:	Besnoitia besnoiti strain Bb-Ger1
Fields matched:	EC descriptions and numbers; GO terms; InterPro domains; Orthologs; Product description; Product descriptions (all)
Gene - BESB_001190	CMGC kinase, CK2 family
Gene type:	protein coding gene
Organism:	Besnoitia besnoiti strain Bb-Ger1
Fields matched:	GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)
Gene - BESB_001990	rhoptry kinase family protein ROP39
Gene type:	protein coding gene
Organism:	Besnoitia besnoiti strain Bb-Ger1
Fields matched:	GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)
Gene - BESB_002440	rhoptry kinase family protein ROP31
Gene type:	protein coding gene
Organism:	Besnoitia besnoiti strain Bb-Ger1
Fields matched:	EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)
Gene - BESB_003070	protein kinase
Gene type:	protein coding gene
Organism:	Besnoitia besnoiti strain Bb-Ger1

10. **Search for an exact term.** Quotation marks around your search term can make your search results more precise. This ensures only results containing the exact phrase you are looking for are retrieved. Search for the term *adenosine kinase*. Now run the same search with quotation marks: "*adenosine kinase*". Can you observe a difference?

All results matching adenosine kinase

Export as a Search Strategy to download or mine your results

1 - 20 of 19,110

Filter results

☒ Hide zero counts

Genome
Genes 16,779

Population biology
Popset isolate sequences 112

Metabolism
Metabolic pathways 1,132
Compounds 1,080

Data access
Data sets 3
Searches 4

Filter fields
Select a result filter above

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

Type a taxonomic name ☐ Reference only

☐ Eimeriidae 4,853
☐ Sarcocystidae 11,929

Gene - LOC34620506 adenosine kinase
Gene type: protein coding gene
Organism: Cyclospora cayetanensis isolate NF1_C8
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Metabolic pathways; Orthologs; PDB chains; Product description; Product descriptions (all)

Gene - TGME49_250880 adenosine kinase
Gene type: protein coding gene
Organism: Toxoplasma gondii ME49
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Metabolic pathways; Orthologs; PDB chains; Product description; Product descriptions (all); PubMed; User comments

Gene - EAH_00063030 adenosine kinase, putative
Gene type: protein coding gene
Organism: Eimeria acervulina Houghton
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Metabolic pathways; Orthologs; PDB chains; Product description; Product descriptions (all)

Gene - ENH_00053770 adenosine kinase, putative
Gene type: protein coding gene
Organism: Eimeria necatrix Houghton
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Metabolic pathways; Orthologs; PDB chains; Product description; Product descriptions (all)

All results matching "adenosine kinase"

Export as a Search Strategy to download or mine your results

1 - 20 of 37

Filter results

☒ Hide zero counts

Genome
Genes 31

Population biology
Popset isolate sequences 2

Metabolism
Metabolic pathways 4

Filter fields
Select a result filter above

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

Type a taxonomic name ☐ Reference only

☐ Eimeriidae 7
☐ Sarcocystidae 24

Gene - LOC34620506 adenosine kinase
Gene type: protein coding gene
Organism: Cyclospora cayetanensis isolate NF1_C8
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)

Gene - TGME49_250880 adenosine kinase
Gene type: protein coding gene
Organism: Toxoplasma gondii ME49
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all); PubMed; User comments

Gene - EAH_00063030 adenosine kinase, putative
Gene type: protein coding gene
Organism: Eimeria acervulina Houghton
Fields matched: EC descriptions and numbers; GO terms; InterPro domains; Orthologs; PDB chains; Product description; Product descriptions (all)

Gene - ENH_00053770 adenosine kinase, putative
Gene type: protein coding gene
Organism: Eimeria necatrix Houghton
Fields matched: EC descriptions and numbers; Orthologs; PDB chains; Product description; Product descriptions (all)

11. **Search for a specific gene ID.** Enter the gene ID in the site search window: *TGME49_311100*. 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 *TGME49_311100* does return a direct link to the gene in *T. gondii* ME49, it also returns links to two other genes. Why?

Genes matching TGME49_311100

Export as a Search Strategy
to download or mine your results

1 - 3 of 3

Filter results

Genome

Genes

3

Filter Gene fields

select all | clear all

☐ Gene ID

☐ Names, IDs, and aliases

☐ User comments

1

1

3

Filter organisms

select all | clear all | expand all | collapse all

Type a taxonomic name

☐ Reference only

☐ Sarcocystidae

☐ Toxoplasma

3

3

Gene - TGME49_311100 bradyzoite formation deficient protein BFD2

Gene name or symbol: BFD2
Gene type: protein coding gene
Organism: Toxoplasma gondii ME49

Direct link to TGME49_311100

Fields matched: Gene ID; Names, IDs, and aliases; User comments

Gene - TGME49_311100 bradyzoite formation deficient protein BFD2

Gene name or symbol: BFD2
Gene type: protein coding gene
Organism: Toxoplasma gondii ME49

Fields matched: Gene ID; Names, IDs, and aliases; User comments

Gene - TGVEG_311100 zinc finger (CCCH type) motif-containing protein

Gene type: protein coding gene
Organism: Toxoplasma gondii VEG

Fields matched: User comments

Gene - TGME49_200385 bradyzoite-formation deficient protein BFD1

Gene name or symbol: BFD1
Gene type: protein coding gene
Organism: Toxoplasma gondii ME49

Fields matched: User comments

1 - 3 of 3

Why are these two genes
returned by a search for
TGME49_311100?