

## VEuPathDB RNA-Seq BigWig & expression value export & integration tool

**Note:** using VEuPathDB ***My Datasets*** and the VEuPathDB Galaxy workspace requires you to log in to the site since the data is stored privately under your account.

Currently this tool works on workflows that use Cufflinks to generate expression values. It is recommended that all data files used are grouped using the collections method in Galaxy. For more information about collections please see the [Galaxy Collections help pages](#). An example workflow is available on the VEuPathDB Galaxy front page for you to use and try out.

Once a workflow has run successfully you can export and integrate your data privately into VEuPathDB by following these steps:

1. Click on the “VEuPathDB Export Tools” menu item in the left-hand panel, then click on the “RNA-Seq to VEuPathDB” export tool.

The screenshot shows the VEuPathDB Galaxy Site interface. On the left, there is a sidebar with various tools and applications. A red arrow points to the "VEuPathDB Export Tools" section, which contains two items: "Bigwig Files to VEuPathDB" and "RNA-Seq to VEuPathDB". The "RNA-Seq to VEuPathDB" item is highlighted with a red box. The main content area displays the "Welcome to the VEuPathDB Galaxy Site" page, which includes a brief introduction, a list of things you can do with VEuPathDB, and a "Get started with VEuPathDB pre-configured workflows" section. The right side of the screen shows a "History" panel listing several completed workflows, each with a green status bar and a delete icon.

Workflow ID	Description	Status
165	Cufflinks on collection 130: total map mass	Completed
164	Cufflinks on collection 130: assembled transcripts	Completed
163	Cufflinks on collection 130: transcript expression	Completed
162	Cufflinks on collection 130: gene expression	Completed
161	Cufflinks on data 12 9: total map mass	Completed
160	Cufflinks on data 12 9: assembled transcripts	Completed
159	Cufflinks on data 12 9: transcript expression	Completed

- The RNA-Seq export tool parameters will appear in the middle section of the Galaxy page. Fill out the parameters completely:
  - Enter a meaningful name for your dataset (you can always modify this).
  - Select whether your RNA-Seq dataset is strand specific or not.
  - Select the BigWig collection from the list.
  - Select the Cufflinks **gene** expression collection.
  - Enter a summary for your dataset (you can always modify this).
  - Enter a description for your dataset (you can always modify this).
  - Click on the “Execute” button.

**RNA-Seq to VEuPathDB Export an RNA-Seq result to VEuPathDB (Galaxy Version 1.0.0)** ▼ Options

**My Data Set name:**

**A** Name of my dataset  
specify a name for the new dataset

**Is your dataset strand-specific?**

**B** No  
Is this a strand-specific dataset?

**BigWig collection:**

**C** 172: BAM to BigWig on collection 130  
Select the BigWig collection to include in the new VEuPathDB My Data Set. The bigwig collection you select here must be mapped to the reference genome that you select below.

**FPKM collection:**

**D** 162: Cufflinks on collection 130: gene expression  
Select the FPKM collection. For an unstranded dataset, its name should include the phrase 'gene expression'.

**My Data Set summary:**

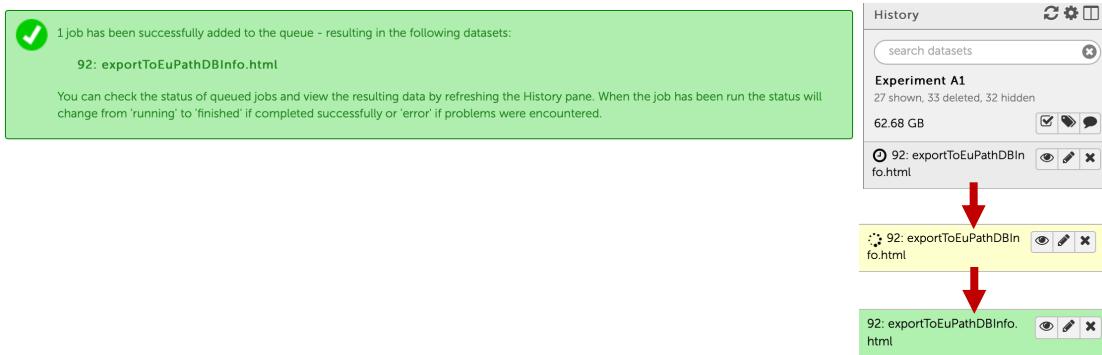
**E** A short summary of my dataset

**My Data Set description:**

**F** A more detailed description of my dataset

**G**  Execute

- After you click on the execute button a new step is initiated in the right hand history panel: ExportToVEuPathDBInfo.html. The step will initially be grey, then yellow when running and green when completed. This should not take a long time.



- Once the export is complete, go to the VEuPathDB site where your data will be visualized. For example, if your data is *Plasmodium berghei* RNA-Seq data go to PlasmoDB.org
- Click on the “My Workspace” menu in the header then select the *My data sets* link.

The screenshot shows the ToxoDB beta homepage. At the top, there's a search bar with placeholder text "E.g., \* or TGME49\_239250 or synth\* or 'oxo group\*'". Below the search bar is a navigation menu with links: My Strategies, Searches, Tools, My Workspace, Data, About, Help, Contact Us, and a Guest account link. A red arrow points to the "My Data Sets" link under the "My Workspace" menu. The main content area features a "Search for..." sidebar on the left with categories like Genes, Organisms, Popset Isolate Sequences, etc. The central area is titled "Overview of Resources and Tools" and includes sections for "Getting Started" and "Transcriptomic Resources".

- The “My Datasets” page includes a list of all datasets that you have exported to VEuPathDB with the most recent export at the top of the list.

The screenshot shows the "My Data Sets" page. At the top, there's a search bar and filter options. The main table lists one dataset:

	Name / ID	Summary	Type	EuPathDB Websites	Status	Owner	Created	File Count	Size	Quota Usage
<input type="checkbox"/>	Experiment A1 (12640)	Experiment A1	RNASeq (1.0)	ToxoDB	<span>Installing</span>	Me	7 minutes ago	9	295.20 M	3.09%

- Click on the experiment name in the Name/ID column to enter that dataset’s page. If the experiment is still being installed into VEuPathDB you should notice the status line indicating that the dataset is currently being installed.

The screenshot shows the "My Dataset: Experiment A1" page. At the top, there's a status message: "Status: Installing This data set is currently being installed in ToxoDB. Please check again soon." Below this, there are detailed dataset statistics:

- Owner: Me
- Description: samples from experiment A1
- ID: 12640
- Data Type: RNASeq (RNASeq 1.0)
- Summary: Experiment A1
- Created: 7 minutes ago
- Dataset Size: 295.20 M
- Quota Usage: 3.09% of 10.00 G

#### Use This Dataset in ToxoDB

The screenshot shows the "Compatibility Information" section. It displays a table comparing EuPathDB Website, Required Resource, Required Resource Release, and Installed Resource Release for ToxoDB and TgondIME49 Genome. Both rows show values of 29 for all four columns.

EuPathDB Website	Required Resource	Required Resource Release	Installed Resource Release
ToxoDB	TgondIME49 Genome	29	29

This dataset is compatible with the current release, build 40, of ToxoDB. It is installed for use.

#### Data Files

- Once the dataset is installed, it is ready for use (note the status line indicates it is ready with a green checkmark). You can access the results in two ways:

- Access available searches against your data
- Access GBrowse tracks representing your data

**My Dataset: Experiment\_A1**

→ **Status:** This data set is installed and ready for use in ToxoDB.

**Owner:** Me

**Description:** samples from experiment A1

**ID:** 12993

**Data Type:** RNASeq (RnaSeq 1.0)

**Summary:** Experiment A1

**Created:** an hour ago

**Dataset Size:** 295.20 M

**Quota Usage:** 3.09% of 10.00 G

→ **Available Searches:** • [genes by RNA-Seq user dataset \(fold change\)](#)

## Use This Dataset in ToxoDB

### Compatibility Information

EuPathDB Website	Required Resource	Required Resource Release	Installed Resource Release
ToxoDB	TgondiiME49 Genome	29	29

This dataset is compatible with the current release, build 40, of ToxoDB. It is installed for use.

→ **GBrowse Tracks**

Filename	GBrowse Status
day3_rep2.bw	<input type="radio"/> This file has not been added to GBrowse. <a href="#">Send To GBrowse</a>
day3_rep1.bw	<input type="radio"/> This file has not been added to GBrowse. <a href="#">Send To GBrowse</a>

- To run a search against your data, click on the “genes by RNA-Seq user dataset (fold change)” link.

10. The next page provides you with customizable search parameters with your analyzed samples available for selection.

Identify Genes based on genes by RNA-Seq user dataset (fold change)

Your RNA-Seq Dataset

Experiment A1

For the Experiment **unstranded** return **protein coding** **Genes** that are **Up or down regulated** with a Fold change  $\geq$  **2**

between each gene's expression value in the following **Reference Samples**

**day3\_rep1**  
**day3\_rep2**  
**day4\_rep1**  
**day4\_rep2**

**select all** | **clear all**

and its expression value in the following **Comparison Samples**

**day3\_rep1**  
**day3\_rep2**  
**day4\_rep1**  
**day4\_rep2**

**select all** | **clear all**

Example showing one gene that would meet search criteria  
(Dots represent this gene's expression values for selected samples)

Up or down regulated

This graphic will help you visualize the parameter choices you make at the left; it will begin to display when you choose a Reference Sample or a Comparison Sample.

See the [detailed help for this search](#).

Get Answer

11. Once you configure the search parameters, you can click on the “Get Answer” button to get your results.

Identify Genes based on genes by RNA-Seq user dataset (fold change)

Your RNA-Seq Dataset

Experiment A1

For the Experiment **unstranded** return **protein coding** **Genes** that are **Up or down regulated** with a Fold change  $\geq$  **2**

between each gene's expression value in the following **Reference Samples**

**day3\_rep1**  
**day3\_rep2**  
**day4\_rep1**  
**day4\_rep2**

**select all** | **clear all**

and its **minimum** expression value in the following **Comparison Samples**

**day3\_rep1**  
**day3\_rep2**  
**day4\_rep1**  
**day4\_rep2**

**select all** | **clear all**

Example showing one gene that would meet search criteria  
(Dots represent this gene's expression values for selected samples)

Up-regulated

You are searching for genes that are up-regulated between two reference samples and two comparison samples.

For each gene, the fold change is calculated:

**Aff. change = minimum expression level in comparison / maximum expression level in reference**

And returns genes when fold change  $\geq$  2. This is the average of the minimum expression value in which to look for genes and the maximum expression value in which to look for genes that meet your fold change cutoff. To broaden the search, you can set the minimum reference value, or average of the maximum comparison value.

See the [detailed help for this search](#).

183 Genes (166 ortholog groups) Review this search

Gene Results | Generic View | Analyze Results

Rows per page: 50

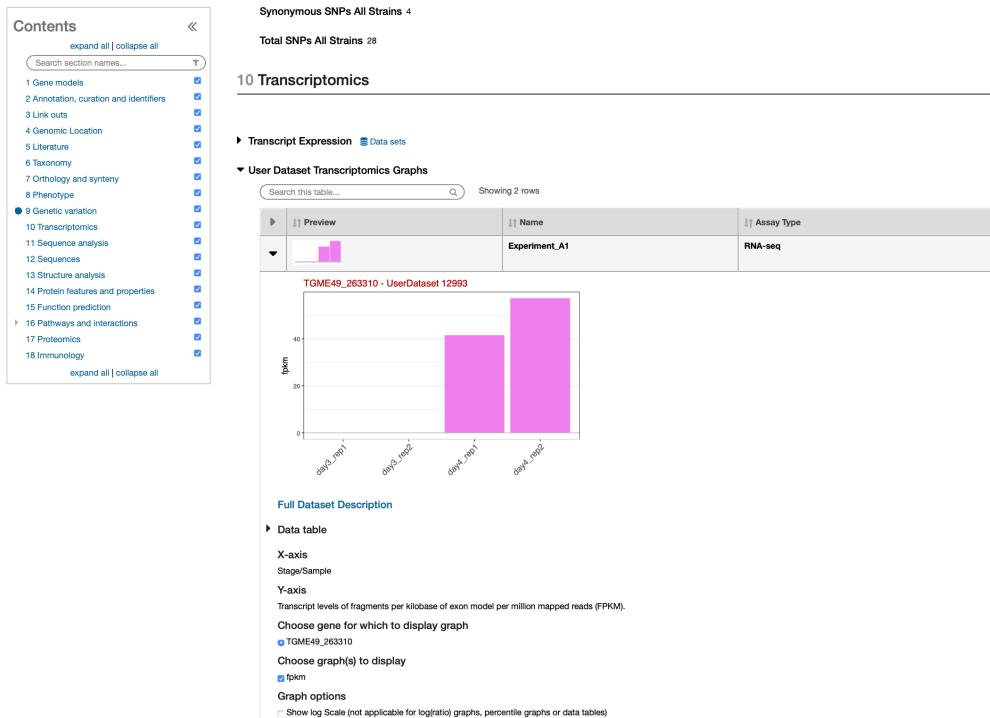
Gene ID	Transcript ID	Product Description	Fold Change	Chosen Ref	Chosen Comp	Profile
TOME49_27590	TOME49_27590.056.1	hypothetical protein	369.5	1.31	482.15	
TOME49_220560	TOME49_220560.056.1	hypothetical protein	393.1*	0	399.11	

Download | Add to Dossier | Add Column

Get Answer

12. The results page includes custom columns based on your data including: fold change, expression values for the reference and comparator (FPKM) and a profile graph. The results can be further intersected with results from other searches in VEuPathDB or analyzed using one of the available analysis tools such as GO enrichment.

13. The data is also represented on gene pages in the Transcriptomics section under the heading “User Dataset Transcriptomics Graphs”.



14. To view RNA-Seq coverage graphs in the genome browser, go back to the “My Datasets” section, select the dataset you want to view and scroll down to the Genome Browser Tracks section.

#### Genome Browser Tracks

Filename	Genome Browser Link
Cyst.bw	<a href="#">View in Genome Browser</a>
Tachy.bw	<a href="#">View in Genome Browser</a>

15. Now you can click on the “View in Genome Browser” button to visualize your data in the genome browser.

