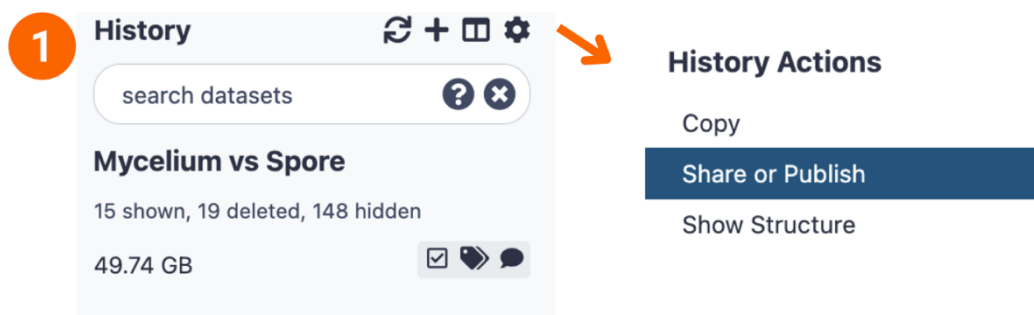


## RNA sequence data analysis via Galaxy, Part 2

### Learning objectives:

- Examine RNA-Seq analysis workflow and outputs.
  - Import data from Galaxy to FungiDB My Workspace.
  - Analyze the results using FungiDB interface and tools.
- **Sharing workflow histories with others.**
    1. Make sure your history has a useful name (e.g, Mycelium vs Spore, RNA Group3, etc.) and click on the history action menu icon.
    2. Select the “Make History Accessible and Publish” option and check to make sure that all objects within History are accessible.



### Make History Accessible via Link and Publish It

This history is currently restricted so that only you and the users listed below can access it. You can:

[Make History Accessible via Link](#)

☐ Also make all objects within the History accessible.

Generates a web link that you can share with other people so that they can view and import the history.

**2** [Make History Accessible and Publish](#)

☐ Also make all objects within the History accessible.

Makes the history accessible via link (see above) and publishes the history to Galaxy's [Published Histories](#) section, v

### Share History with Individual Users

You have not shared this history with any users.

[Share with a user](#)

- **Importing workflow histories and output files into your own Galaxy workspace.**

1. Click on “Shared Data” at the top and select “Histories”.
2. Click on the history shared by your colleague, click on the plus icon on the far right and choose to import the history.
3. You can give it a descriptive name if you prefer or leave it as is.

The screenshot illustrates the Galaxy web interface for importing workflow histories. It is divided into three numbered steps:

- Step 1:** The user navigates to the "Histories" section under the "Shared Data" menu.
- Step 2:** The user searches for a specific history. The search results show a history titled "Mycelium vs Spore" by the user "ebasenko".
- Step 3:** The user clicks the "Import history" button (indicated by an orange arrow) next to the search result. Below the search results, a form prompts the user to "Enter a title for the new history:". The title entered is "FPG2023 RNA-Seq Group 4". At the bottom right, there are "Cancel" and "Import" buttons, with an orange arrow pointing to the "Import" button.

If everything worked out, you should see a list of completed workflow steps highlighted in green. The workflow generates many output files, however not all of the output files are visible. You can explore all the hidden files clicking on the word “hidden” (orange circle) – this will reveal all hidden files.

Many more output files are available to explore →

Differential expression data on the two collection →

Coverage data in BigWig format →

FastQC results (one per each file submitted) →

### Mycelium vs Spore

16 shown, 18 deleted, 148 hidden

49.74 GB

☒

94: DESeq2 plots on data 88, data 86, and others	
93: DESeq2 result file on data 88, data 86, and others	
90: BAM to BigWig on collection 72 a list with 2 items	
75: BAM to BigWig on collection 69 a list with 2 items	
39: FastQC on collection 18: Webpage a list of pairs with 2 items	
24: FastQC on collection 13: Webpage a list of pairs with 2 items	
18: mycelium a list of pairs with 2 items	
13: spores a list of pairs with 2 items	
8: SRR1179896_2.fastq.gz	
7: SRR1179896_1.fastq.gz	
6: SRR1179895_2.fastq.gz	

- Explore the FastQC results.

To do this find the step called “FastQC on collection ##: Webpage”. Click on the name this will open up the FastQ pairs, click on one of them then click on view data icon (👁) on either forward or reverse. Note that each FastQ file will have its own FastQC results.

**24: FastQC on collection 13: Webpage** ✕  
 a list of pairs with 2 items

**SRR1179892.fastq**  
 a pair of datasets

**SRR1179893.fastq**  
 a pair of datasets

**forward**

**reverse**

👁 ✎

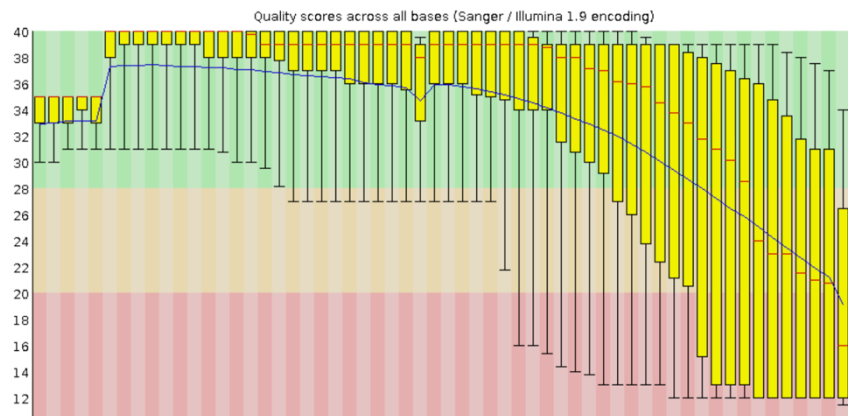
## Summary

- ✅ [Basic Statistics](#)
- ❌ [Per base sequence quality](#)
- ✅ [Per tile sequence quality](#)
- ✅ [Per sequence quality scores](#)
- ⚠️ [Per base sequence content](#)
- ⚠️ [Per sequence GC content](#)
- ✅ [Per base N content](#)
- ✅ [Sequence Length Distribution](#)
- ✅ [Sequence Duplication Levels](#)
- ❌ [Overrepresented sequences](#)
- ✅ [Adapter Content](#)
- ❌ [Kmer Content](#)

## Basic Statistics

Measure	Value
Filename	SRR11785185_2.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	7649791
Sequences flagged as poor quality	0
Sequence length	251
%GC	50

## ❌ Per base sequence quality



## Explore the differential expression results.

We will explore two output files:

- A. **DESeq2 Plots** – you can view these directly in galaxy by clicking on the view icon. These plots give you an idea about the quality of the experiment. The link above includes a detailed description of the graphs.
- B. **DESeq2 results file** – this is a table which contains the actual differential expression results. These can be viewed within galaxy but it will be more useful to download this table and open in Excel so you can sort results and big genes of interest.

The tabular file contains 7 columns:

COLUMN	DESCRIPTION
1	Gene Identifiers
2	mean normalized counts, averaged over all samples from both conditions
3	the logarithm (to basis 2) of the fold change (See the note in inputs section)
4	standard error estimate for the log2 fold change estimate
5	Wald statistic
6	p value for the statistical significance of this change
7	p value adjusted for multiple testing with the Benjamini-Hochberg procedure which controls false discovery rate (FDR)

- **Download DESeq2 results (tabular format) by clicking on the floppy disk save icon.**

\*\*\* **Important:** the file name ends with the extension “. tabular  
“ change this to .txt and then open the file in Excel.

94: DESeq2 plots on data 88, data 86, and others

93: DESeq2 result file on data 88, data 86, and others

14,145 lines

format: tabular, database: FungiDB-31\_FgraminearumPH-1\_Genome

primary factor: myceliumXvsXspores

DESeq2 run information

sample table:

myceliumXvsXspores

SRR1179895.fastq mycelium

SRR1179896.fastq mycelium

SRR1179892.fastq spores

SRR1179893.fa

1. GeneID	2. Base mean	3. Log2 fold change	4. Standard error	5. Wald statistic	6. P value	7. Adjusted P value
FGRAMPH1_01G25635	82750.1380783884	13.7				
FGRAMPH1_01G08385	24133.4229278897	-12.1				
FGRAMPH1_01G15589	25128.2417004296	12.6				

- **Explore the results in Excel.**

1. Sort them based on the log2 fold change – column 3.

2. Pick a list of gene IDs from column 3 that are upregulated with a good corrected P value (column 7) and load then into FungiDB using the “List of IDs” search.

	A	B	C	D	E	F	G
8	D8B26_0010	1432.94686	4.14837844	0.21276917	19.4970844	1.16E-84	7.25E-82
9	D8B26_0041	1459.15095	4.12515507	0.21288538	19.3773525	1.20E-83	6.95E-81
0	D8B26_0047	149.884174	4.11535522	0.34755905	11.8407366	2.40E-32	1.93E-30
1	D8B26_0025	12524.2357	4.09249452	0.17888678	22.8775683	7.77E-116	7.88E-113
2	D8B26_0065	297.307163	4.03853435	0.2783354	14.5095963	1.05E-47	1.64E-45
3	D8B26_0033	1682.63609	4.03468031	0.22941812	17.5865811	3.12E-69	1.33E-66
4	D8B26_0069	242.253822	4.01567422	0.29254924	13.72649	7.05E-43	9.53E-41
5	D8B26_0024	1129.38482	3.97988586	0.26221324	15.1780507	4.94E-52	1.00E-49
6	D8B26_0079	401.277324	3.9579969	0.27562766	14.3599407	9.23E-47	1.39E-44
7	D8B26_0055	242.517663	3.85630043	0.28041015	13.3287365	3.08E-40	3.73E-38

### Identify Genes based on List of IDs

Configure Search
Learn More
View Data Sets Used

Reset values to default

Gene ID input set

☒ Enter a list of IDs or text:

D8B26\_006344  
D8B26\_007030  
D8B26\_003055  
D8B26\_006187  
D8B26\_005929  
D8B26\_003310

☐ Upload a text file:

Choose file
No file chosen
Maximum size 10MB. The file should contain the list of IDs.

☐ Upload from a URL:

The URL should resolve to a list of IDs.

☐ Copy from My Basket:

3 records will be copied from your basket.

☐ Copy from My Strategy:

ID list search (? records)

Get Answer

3. Next, analyze results with GO or metabolic enrichment tools. Note: you can do the same for down-regulated genes.

## Exporting data to VEuPathDB

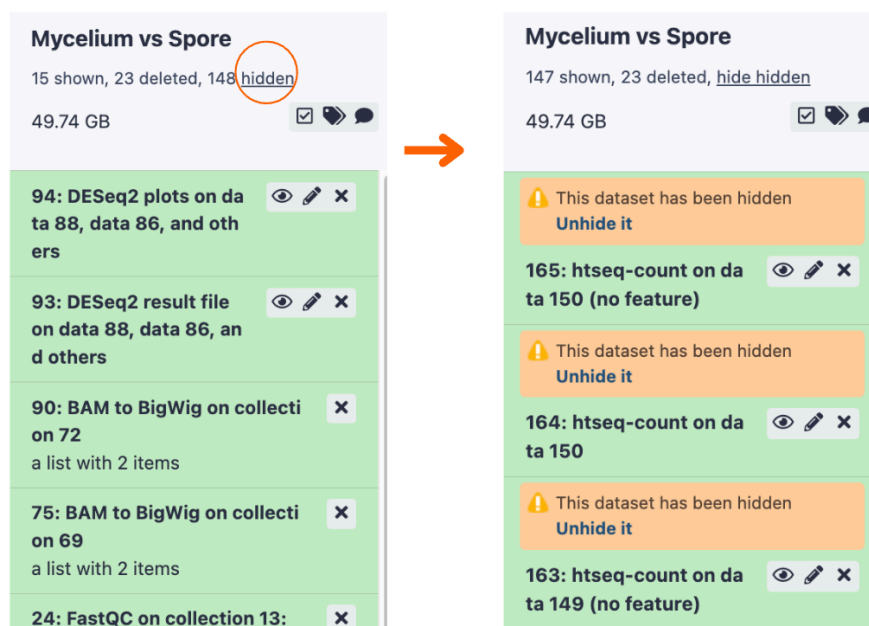
The VEuPathDB RNAseq export tool provides a mechanism to export your RNAseq results (TPM values) and BigWig RNAseq coverage files. The advantage of doing this is that it allows you to search the TPM data using the RNAseq search in VEuPathDB and view the BigWig files in the genome browser.

However, to use this feature you need to generate TPM values for genes in your datasets and organize your results into two collections, one for the TPMs and one for the BigWigs.

- **Create a Dataset List with “htseq-count on data” files.**

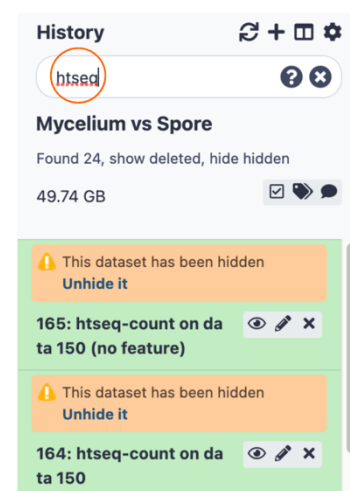
### 1. Reveal hidden files.

Click on the link at the top of your history that says “## hidden”. This will show all hidden files.



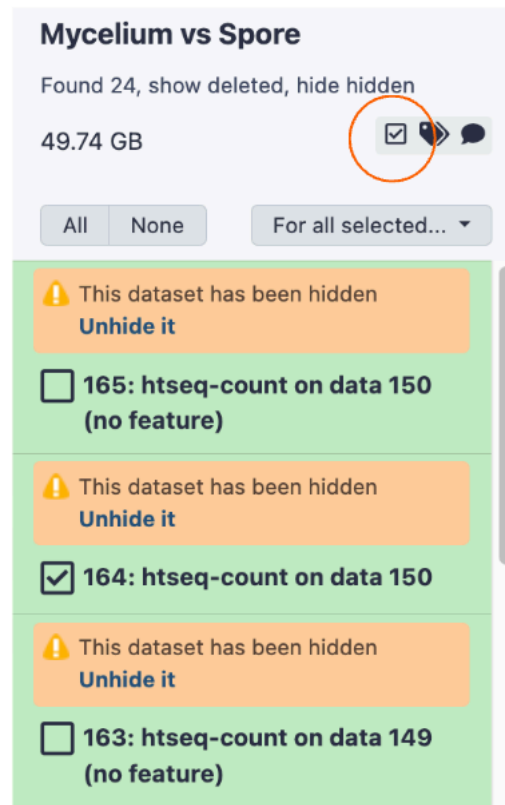
### 2. Search for htseq-count files.

Use the search datasets box at the top of your history to find any file in your history with the work “htseq-count”. To do this, type “htseq” and click the “Enter” key on your keyboard.



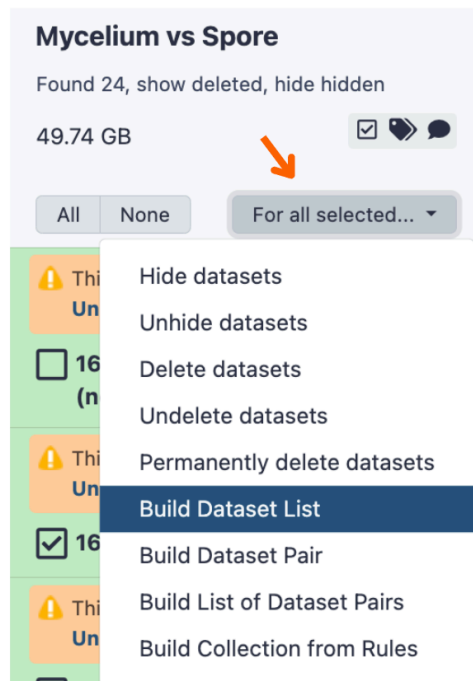
3. Select “htseq-count on data” files. Click on the “operation on multiple datasets” tool and select the individual htseq-count files. These should look something like this: **htseq-count on data xx**. Do not select “no feature” or “..on collection” files.

Note: if you are comparing two conditions each done in duplicate then you should have selected 4 files.



4. “Build dataset list”.

Click on the “For all selected” button and choose the “Build dataset list” option.





## 5. Rename each htseq-count sample, give the collection a name and create a dataset list.

Note: the htseq-count files will in the same order as the raw files loaded into the history. Use the “Guide to FPG2023 RNA-Seq histories and file organisation” in Part 1 for more info.

The screenshot shows the 'Create a collection from a list of datasets' interface. It features a list of datasets on the left, a central panel for renaming, and a right panel for collection details. The datasets listed are 'htseq-count on data 74', 'htseq-count on data 73', 'htseq-count on data 71', and 'htseq-count on data 70'. The central panel shows the selected dataset 'htseq-count on data 74' being renamed to 'mycelium 2'. The right panel shows the collection name 'Mycelium vs spores' and a 'Create list' button. Orange arrows indicate the flow from the dataset list to the renaming dialog and then to the final collection list.

Create a collection from a list of datasets

Collections of datasets are permanent, ordered lists of datasets that can be passed to tools and workflows in order to have analyses done on each member of the entire group. This interface allows you to crea...More help

Start over

htseq-count on data 74 Discard

htseq-count on data 73 Discard

htseq-count on data 71 Discard

htseq-count on data 70 Discard

Cancel

Click to rename

Enter a new name for the element:

mycelium 2

Cancel OK

Create a collection from a list of datasets

Collections of datasets are permanent, ordered lists of datasets that can be passed to tools and workflows in order to have analyses done on each member of the entire group. This interface allows you to crea...More help

Start over

mycelium 2 Discard

mycelium 1 Discard

spore 2 Discard

spore 1 Discard

Hide original elements? ☐

Name: Mycelium vs spores

Create list

- **Create a Dataset List with “BAM to BigWig on data” files.**

Use the tutorial for htseq-count files to create a dataset list with BigWig files. Do not use “BAM to BigWig on collection” files.

Now that your count and bigwig files are nice and organized, the next step is to convert the counts into TPMs.

- **Use the HTSeqCountToTPM tool to convert counts to TPM**

1. Select the HTSeqCountToTPM tool (under the VEuPathDB RNAseq tools in the left menu).
2. Make sure the list of count files is selected.
3. Select the reference organism.
4. Click on the “Execute” button.

**Tools** ☆ ⬆

search tools ✕

**VEUPATHDB APPLICATIONS**

- VEuPathDB Export Tools
- VEuPathDB OrthoMCL Tools
- VEuPathDB RNA-Seq Tools**
- HTSeqCountToFPKM compute FPKM from per-gene read counts and reference genome
- HTSeqCountToTPM compute TPM from per-gene read counts and reference genome

**DATA TRANSFER**

- Globus Data Transfer
- Get Data
- Collection Tools
- AtlasXomics tools
- REDINET tools
- agat conversion tools

**HTSeqCountToTPM** compute TPM from per-gene read counts and reference genome (Galaxy Version TPMtool 1)

**gene counts of sense-strand aligned RNA-Seq reads**

⚙️ This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

sense counts file, like that output by htseq-count

**Do you also have antisense reads?**

Answer yes if you would like to submit sense and antisense counts to this tool

**Will you select an annotation file from your history or use a built-in gff3 file?**

**Select genome annotation**

**Email notification**

Send an email notification when the job completes.

- **Export TPM counts and BigWig data to VEuPathDB/FungiDB workspace.**

1. Click on “VEuPathDB Export Tools” > “RNA-Seq to VEuPathDB”
2. Enter a Data Set name.
3. Choose, if not already selected, the correct BigWig collection.
4. Choose, if not already selected, the correct TPM collection.
5. Provide a data set summary.
6. Provide a data set description and click on the “Execute” button.

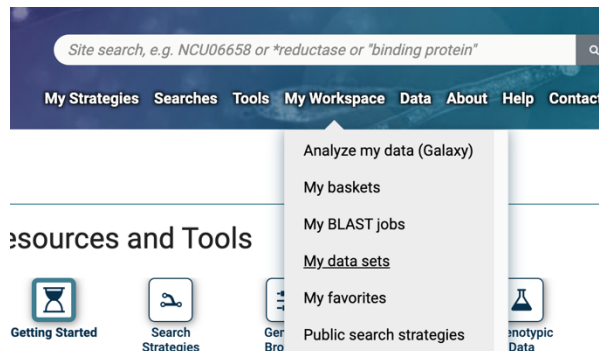
The screenshot shows the 'RNA-Seq to VEuPathDB Export an RNA-Seq result to VEuPathDB (Galaxy Version 1.0.0)' tool interface. The left sidebar contains a 'Tools' menu with 'VEuPathDB APPLICATIONS' expanded, showing 'VEuPathDB Export Tools' and 'RNA-Seq to VEuPathDB' (annotated with a red circle 1). The main panel has the following fields and options:

- My Data Set name:** A text input field containing 'hyphae vs spherules' (annotated with a red circle 2).
- Are you exporting sense and antisense TPM/FPKM datasets?** A dropdown menu set to 'No'.
- BigWig collection:** A dropdown menu set to '70: h vs s' (annotated with a red circle 3).
- TPM or FPKM collection:** A dropdown menu set to '72: HTSeqCountToTPM on collection 65: gene expression' (annotated with a red circle 4).
- My Data Set summary:** A text input field containing 'pathogenic and non-pathogenic stages' (annotated with a red circle 5).
- My Data Set description:** A text input field containing 'https://veupathdbprod.globusgenomics.org/' (annotated with a red circle 6).
- Email notification:** Radio buttons for 'Yes' and 'No', with 'No' selected.
- Execute:** A blue button with a checkmark and the text 'Execute'.

An orange arrow points to the 'Execute' button.

- **Explore your data in FungiDB**

1. Click on the “My Workspace” link in the grey menu bar. Then select “My data sets” from the list.



2. Explore the RNA-Seq dataset via the fold-change search in FungiDB.

## My Data Set: *Afumigatus pre-blood vs 180min*

**Status:** ✔ This data set is installed and ready for use in FungiDB.

**Owner:** Me

**Description:** Afumigatus

**ID:** 4032963

**Data type:** RNA-Seq (RnaSeq 1.0)

**Summary:** pre blood - 180

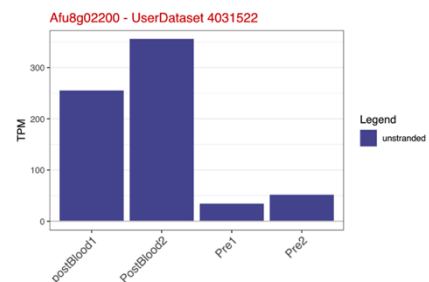
**Created:** 2 years ago

**Data set size:** 271.05 M

**Quota usage:** 2.84% of 10.00 G

**Available searches:** • RNA-Seq user dataset (fold change)

Note that custom graphs are generated for your data in the results table so you can easily visualize the results for each gene.



3. Explore the coverage plots in the genome browser.

