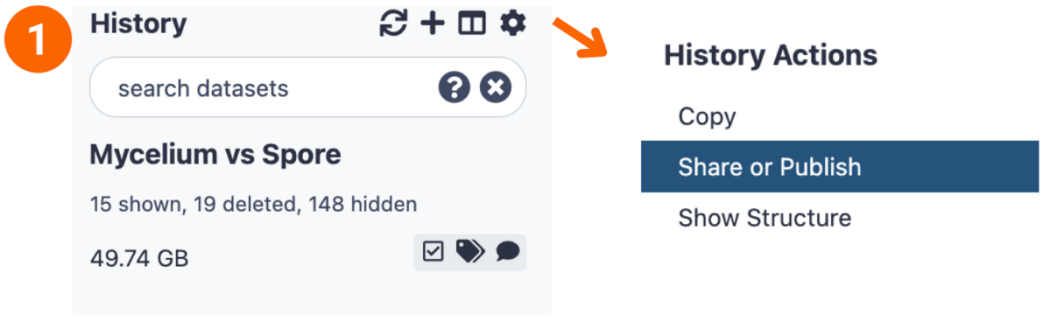
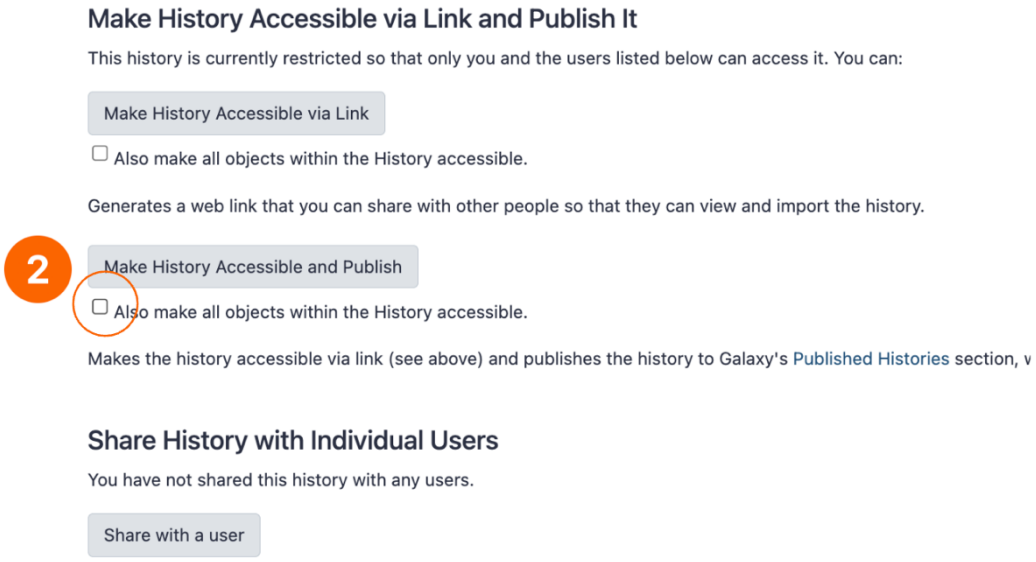
RNA sequence data analysis via Galaxy, Part 2

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.

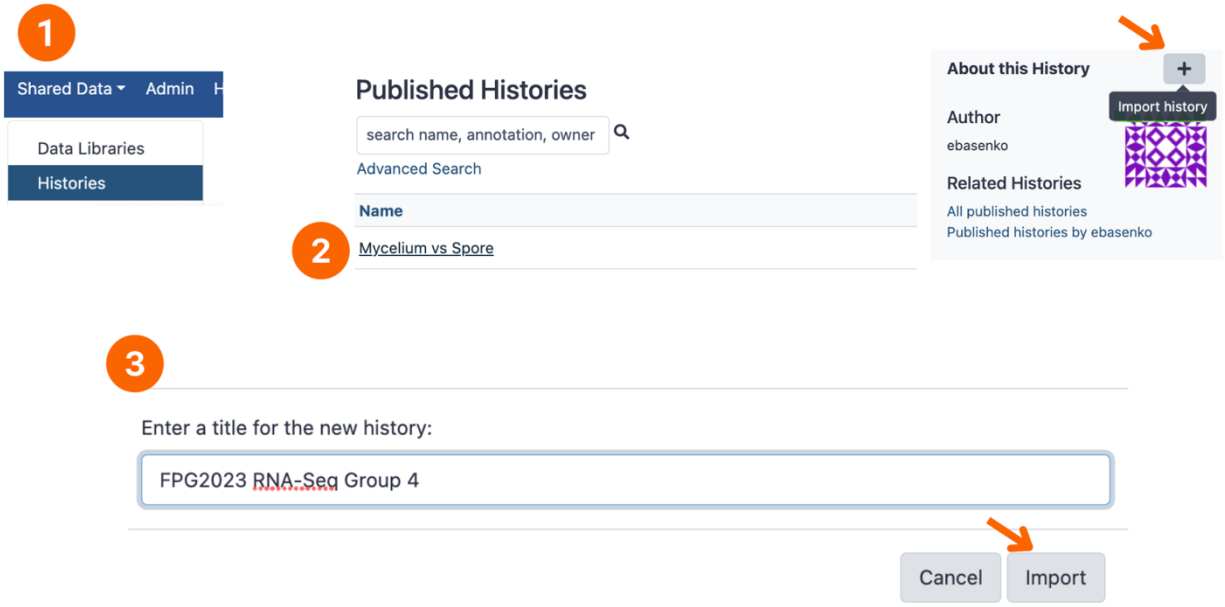
- **Sharing workflow histories with others.**
    1. Make sure your history has a useful name (e.g, Mycelium vs Spore, RNA Group3, etc.) and click on the history action menu icon.
    2. Select the "Make History Accessible and Publish" option and check to make sure that all objects within History are accessible.





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

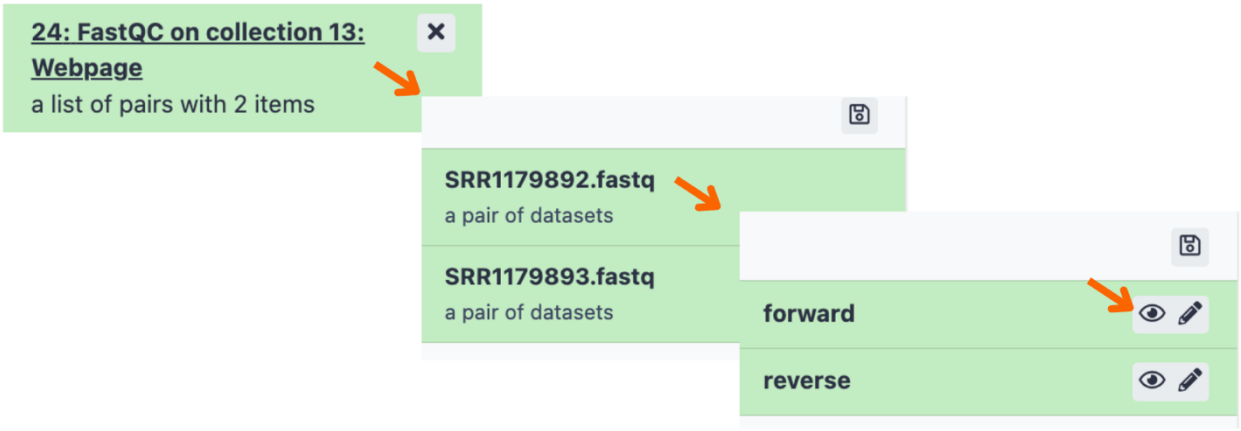
1. Click on “Shared Data” at the top and select “Histories”.
2. Click on the history shared by your colleague, click on the plus icon on the far right and choose to import the history.
3. You can give it a descriptive name if you prefer or leave it as is.

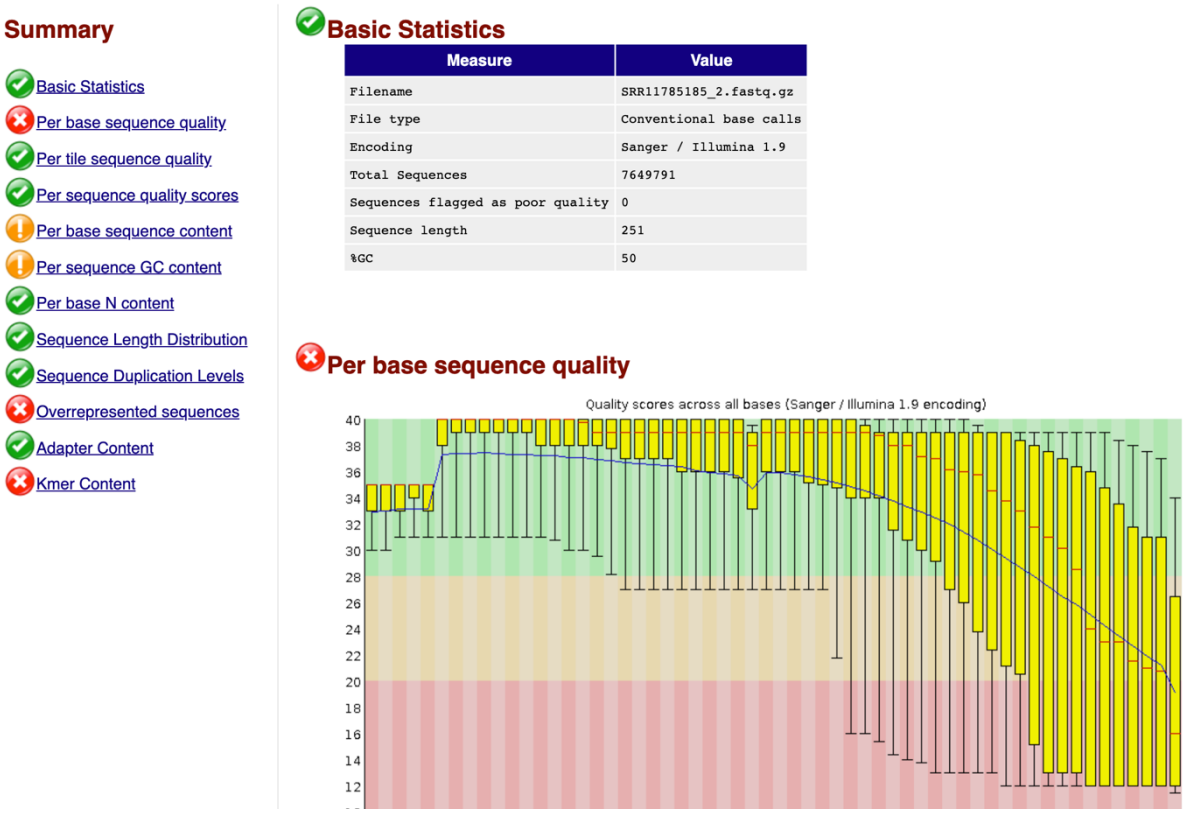


If everything worked out, you should see a list of completed workflow steps highlighted in green. The workflow generates many output files, however not all of the output files are visible. You can explore all the hidden files clicking on the word “hidden” (orange circle) – this will reveal all hidden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

- **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.





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.

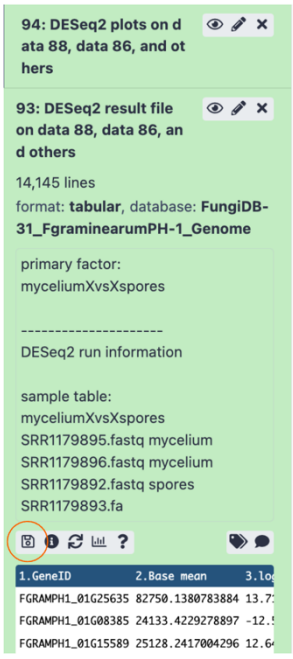
B. **DESeq2 results file** – this is a table which contains the actual differential expression results. These can be viewed within galaxy but it will be more useful to download this table and open in Excel so you can sort results and big genes of interest.

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)

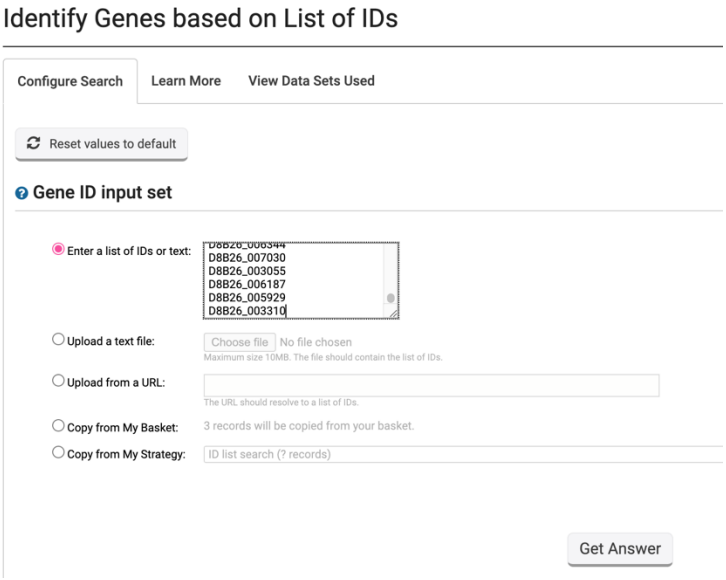
- **Download DESeq2 results (tabular format) by clicking on the floppy disk save icon.**

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



- **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



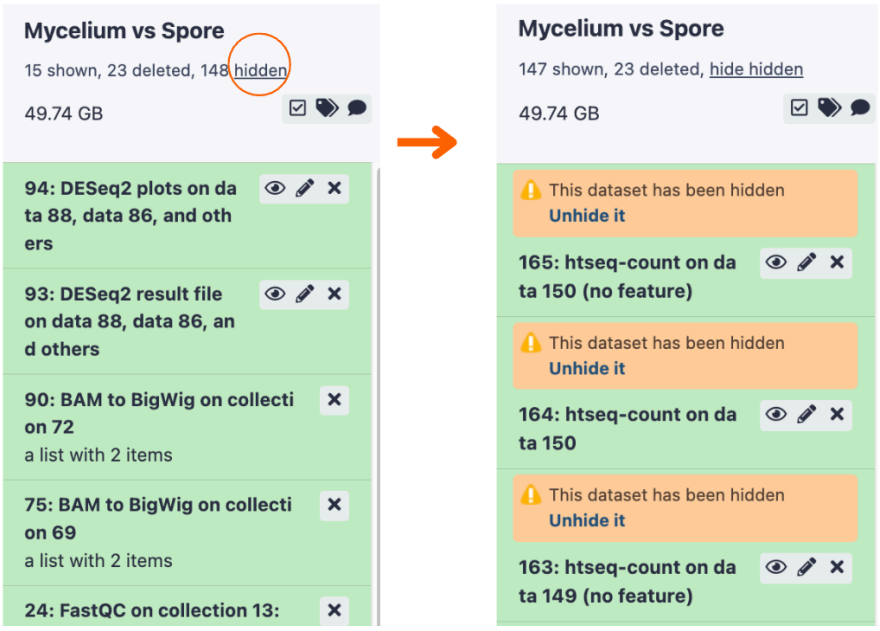
  3. Next, analyze results with GO or metabolic enrichment tools. Note: you can do the same for down-regulated genes.

Exporting data to VEuPathDB

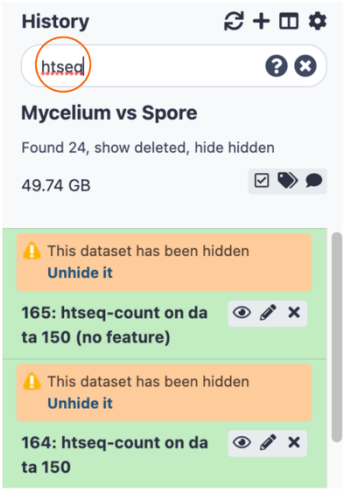
The VEuPathDB RNAseq export tool provides a mechanism to export your RNAseq results (TPM values) and BigWig RNAseq coverage files. The advantage of doing this is that it allows you to search the TPM data using the RNAseq search in VEuPathDB and view the BigWig files in the genome browser.

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.

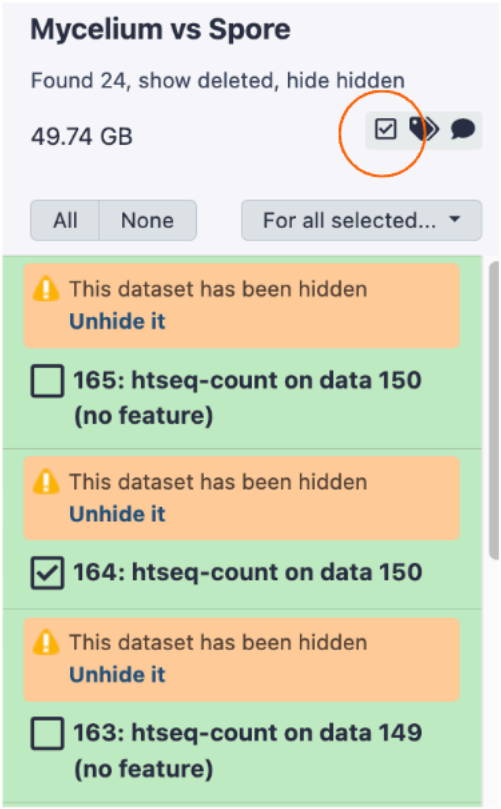


  2. **Search for htseq-count files.**
     Use the search datasets box at the top of your history to find any file in your history with the work "htseq-count". To do this, type "htseq" and click the "Enter" key on your keyboard.

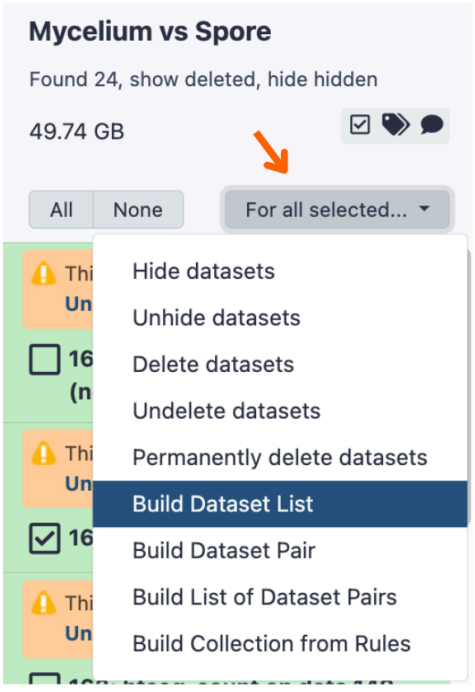


3. **Select "htseq-count on data" files.** Click on the "operation on multiple datasets" tool and select the individual htseq-count files. These should look something like this: **htseq-count on data** xx. Do not select "no feature" or "..on collection" files.

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

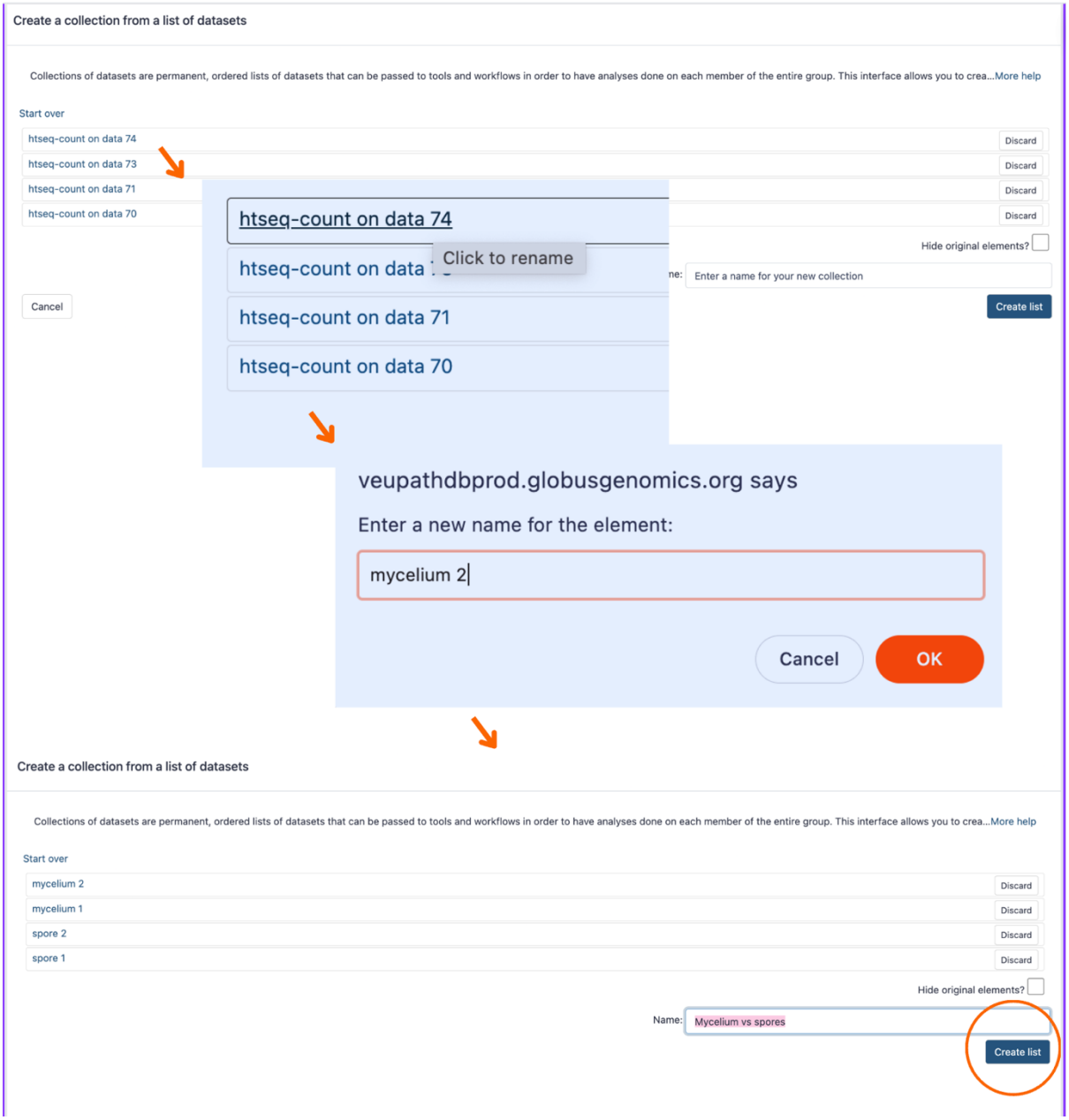


4. **"Build dataset list".** Click on the "For all selected" button and choose the "Build dataset list" option.



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.



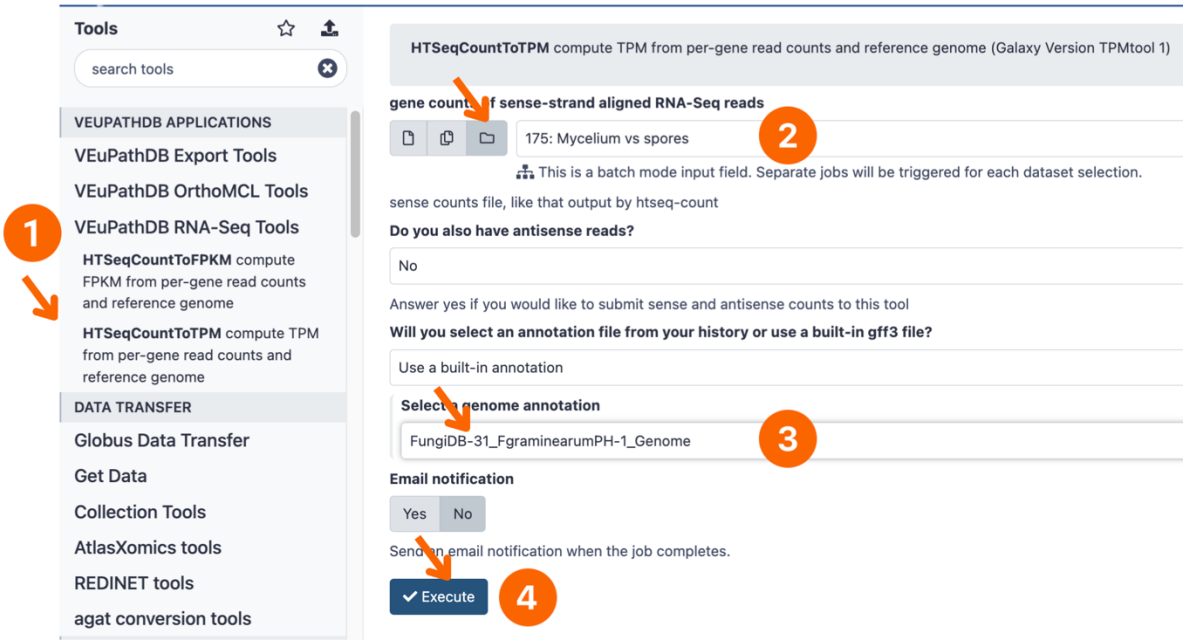
- **Create a Dataset List with "BAM to BigWig on data" files.**

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.

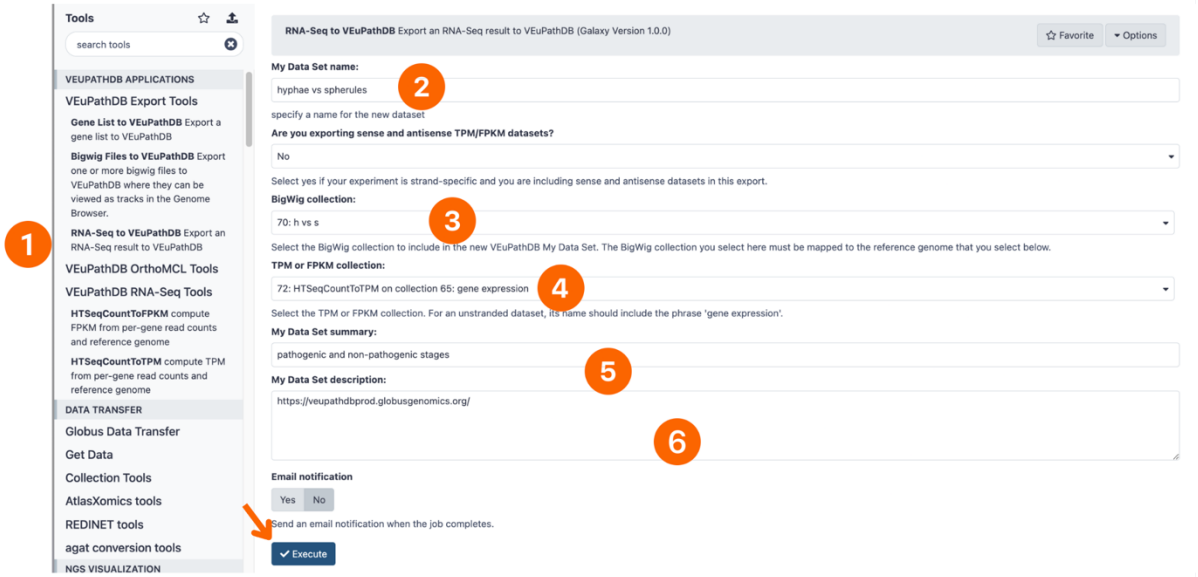
- **Use the HTSeqCountToTPM tool to convert counts to TPM**

    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.

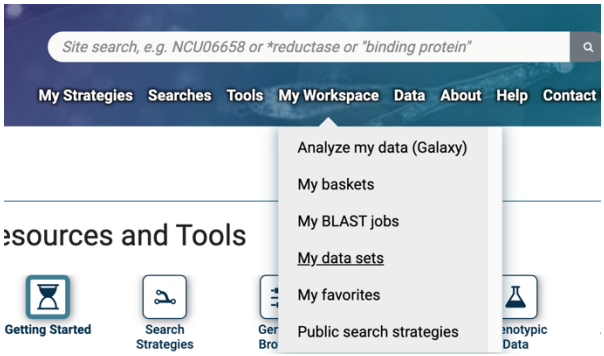


- **Export TPM counts and BigWig data to VEuPathDB/FungiDB workspace.**

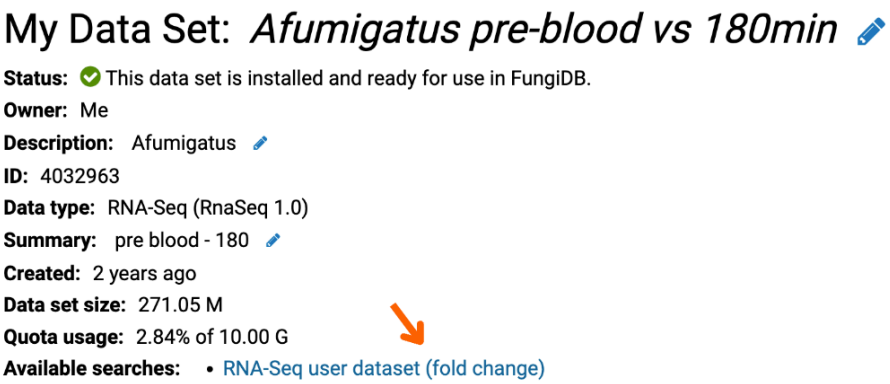
    1. Click on “VEuPathDB Export Tools” > “RNA-Seq to VEuPathDB”
    2. Enter a Data Set name.
    3. Choose, if not already selected, the correct BigWig collection.
    4. Choose, if not already selected, the correct TPM collection.
    5. Provide a data set summary.
    6. Provide a data set description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



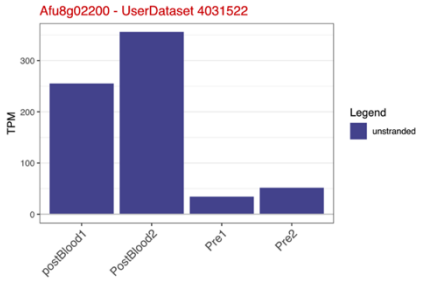
- **Explore your data in FungiDB**
  1. Click on the "My Workspace" link in the grey menu bar. Then select "My data sets" from the list.



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



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.