

FungiDB: Data analysis via EuPathDB Galaxy.

RNA-Seq Part II: Data analysis (Group Exercise)

The goal of this exercise is to examine the results from the Galaxy RNA-seq analysis workflow that ran overnight. If everything worked out you should see a list of completed workflow steps (Green) in the history panel. The workflow generates many output files, however not all are visible. You can explore all the hidden files clicking on the word “hidden” (red circle) – this will reveal all hidden files.

Step 1: Explore the FastQC results.

You can do this by clicking on the view data icon



(eye). An explanation of each of the FastQC results is provided as a link in class.

Many more output files are available to explore

Differential expression data on the two samples

Coverage data (BigWig), cufflinks results and assembled transcripts for first two read files

Coverage data (BigWig), cufflinks results and assembled transcripts for first two read files

Each FastQ file has a corresponding FastQC file which contains info about the read qualities

Original FastQ files. Two for each sample since these were paired end sequences

Workflow Step	Description
145: Cuffdiff on data 122, data 121, and data 137: gene differential expression testing	Cuffdiff on data 122, data 121, and data 137: gene differential expression testing
133: Cufflinks on data 122: assembled transcripts	Cufflinks on data 122: assembled transcripts
131: Cufflinks on data 122: gene expression	Cufflinks on data 122: gene expression
130: BAM to BigWig on data 122	BAM to BigWig on data 122
126: Assembled_Genes_Transcripts	Assembled_Genes_Transcripts
124: Cufflinks on data 121: gene expression	Cufflinks on data 121: gene expression
123: BAM to BigWig on data 121	BAM to BigWig on data 121
115: FastQC on data 4: Webpage	FastQC on data 4: Webpage
109: FastQC on data 3: Webpage	FastQC on data 3: Webpage
107: FastQC on data 2: Webpage	FastQC on data 2: Webpage
101: FastQC on data 1: Webpage	FastQC on data 1: Webpage
4: GRA23/SRR1805883_2.fastq	GRA23/SRR1805883_2.fastq
3: GRA23/SRR1805883_1.fastq	GRA23/SRR1805883_1.fastq
2: WT/SRR1805881_2.fastq	WT/SRR1805881_2.fastq
1: WT/SRR1805881_1.fastq	WT/SRR1805881_1.fastq

More about FastQC analysis and troubleshooting:

Step 2: Export Galaxy results to FungiDB/EuPathDB

Click on the EuPathDB Export Tools (left menu) and choose the RNA-Seq to EuPathDB option.

- Click on EuPathDB Export Tools, then click on BigWig Files to EuPathDB (left Tools panel). The export tool will appear in the central portion.
- Give your dataset a name.
- Select “Dataset Collections” (icon looks like a folder). Then select all the BigWig collections that appear (Shift click).

EUPATHDB APPLICATIONS

EuPathDB Export Tools

Bigwig Files to EuPathDB Export one or more bigwig files to EuPathDB where they can be viewed as tracks in the Genome Browser.

RNA-Seq to EuPathDB Export an RNA-Seq result to EuPathDB

- D. Select the reference genome for your experiment.
- E. Provide a short summary and dataset description – these could be the same for the purpose of this exercise.
- F. Click on the Execute button. This will initiate a new step in your history which will indicate the transfer progress.

A

B

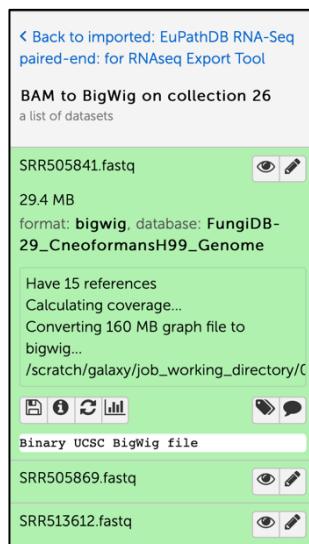
C

D

E

F

Note: You can also download the file by clicking on the floppy disk icon (), as shown below:



Step 3: Navigate to FungiDB/EuPathDB My Data Sets

Once your dataset is successfully uploaded into EuPathDB, navigate to FungiDB and sign in to your account. Your imported data will be located under **My Data Sets (BETA)** tab:

The screenshot shows the FungiDB interface with the 'My Data Sets' tab selected. The page displays a table of datasets with columns for Name / ID, Summary, Type, EuPathDB Websites, Status, Owner, Created, File Count, Size, and Quota Usage. Three datasets are listed:

Name / ID	Summary	Type	EuPathDB Websites	Status	Owner	Created	File Count	Size	Quota Usage
Fusarium graminearum PH-1 (4016036)	datasets1	RNASeq (1.0)	FungiDB	✓	Me	3 months ago	7	92.07 M	0.96%
Ncrrassa transcriptome (4013710)	Dark vs Light 240min	Bigwig (1.0)	FungiDB	✓	Me	5 months ago	2	65.19 M	0.68%
Transcripotome Ncrrassa Dark vs Light 240min (4013530)	Dark vs Light 240min	Bigwig (1.0)	FungiDB	✓	Me	5 months ago	2	65.19 M	0.68%

Note: You can share your datasets with colleagues by selecting a dataset you wish to share and then click on the Share button at the top right:

The screenshot shows the FungiDB 'My Data Sets' page with two datasets listed. The second dataset, 'Cryptococcus neoformans', has a 'Share' button highlighted with an orange box.

Name / ID	Summary	Type	EuPathDB Websites	Status	Owner	Shared With	Created	File Count	Size	Quota Usage
Ustilago maydis infecting Zea mays (4017204)	Ustilago maydis infecting Zea mays time points	Bigwig (1.0)	FungiDB	!	Me	Cristina Aurrecochea	10 minutes ago	6	91.49 M	0.95%
Cryptococcus neoformans (4016036)	datasets1	RNASeq (1.0)	FungiDB	✓	Me		3 months ago	7	92.07 M	0.96%

Step 4: Send BigWig files in GBrowse

- Click on the name of the dataset and examine the data set record page
- Scroll down to the GBrowse tracks section and click on the “Send to GBrowse” buttons for each of the files in the list.

The send to GBrowse button will change to “View in GBrowse”. The import may take a few moments so move on to the next step - “Sharing histories with others”. We will come back to it afterwards.

This screenshot shows the detailed view of the 'Cryptococcus neoformans' dataset. It includes sections for Compatibility Information, GBrowse Tracks, and a note about the dataset being installed and ready for use.

Compatibility Information:

EuPathDB Website	Required Resource	Required Resource Release	Installed Resource Release
FungiDB	Cryptococcus neoformans H99 Genome	29	29

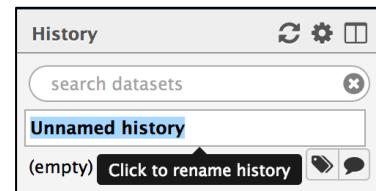
GBrowse Tracks:

Filename	GBrowse Status	Action
ERRS505869.fastq.bw	Sent to GBrowse 3 months ago	Send To GBrowse
ERRS505841.fastq.bw	Sent to GBrowse 5 hours ago	View In GBrowse
ERRS513612.fastq.bw	Sent to GBrowse 5 hours ago	View In GBrowse

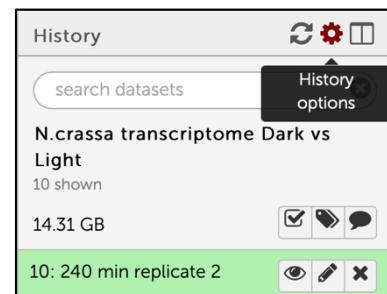
Step 5: Sharing histories with others

- Make sure your history has a useful name – you can change the name by clicking on “unnamed history”

- Click on the history options menu icon

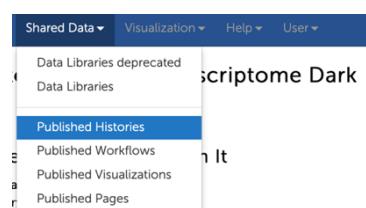


- Select the “Share or Publish” option, the click on the “Make History Accessible and Publish” button in the center section.



This screenshot shows the 'Share or Publish History' dialog from the globus Genomics interface. It displays the history title 'N.crassa transcriptome Dark vs Light'. Under 'Make History Accessible via Link and Publish It', there is a note that the history is currently restricted. Two buttons are shown: 'Make History Accessible via Link' (disabled) and 'Make History Accessible and Publish' (highlighted with a red box). Below these buttons is a note explaining the difference between the two options. The right side of the dialog shows a list of history entries with their details and sharing options. The entire dialog is set against a background of the main globus Genomics interface.

- Select the 'To import a shared history' option, go to the “histories” section (under the shared data menu item).
- Find the history you would like to import and
- Click on the import link



Name	Annotation	Owner	Community Rating	Community Tags	Last Updated
Group2_SNP_Crypto		carlos-perez6	★★★★★		May 17, 2018
imported: Group5_SNP		kytecvdb-301635443	★★★★★		May 17, 2018
imported: Group2_SNP_Crypto		kristjan-twarschek-278549293	★★★★★		May 17, 2018
imported: Group3_SNP		f-puertolas-balint-301635433	★★★★★		May 17, 2018
imported: Group4_SNP_Crypto		cokane44-301496873	★★★★★		May 17, 2018
imported: Group6_SNP		frick-301635513	★★★★★		May 17, 2018
Group1_SNP_Afumigatus (AF10->AF293)		0000-0001-9769-5029	★★★★★		May 16, 2018
Candida albicans SC5314 grown in YPD and serum		carlos-perez6	★★★★★		May 15, 2018
Afumigatus-RNASeq		milwea2ksu-301635723	★★★★★		May 15, 2018

Note: This action will share your results with the rest of group. When working with your own data, click on the Make History Accessible via Link to get a link to share with selected individuals and not EuPathDB Galaxy users.

Step 6: Display coverage results in GBrowse.

Return to My Data Sets in FungiDB page and click on View in GBrowse button for each track.

My Dataset: Cryptococcus neoformans

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

Owner: Me

Description: mutant1

ID: 4016036

Data type: RNASeq (RnaSeq 1.0)

Summary: datasets1

Created: 3 months ago

Dataset Size: 92.07 M

Quota Usage: 0.96% of 10.00 G

Available Searches: • genes by RNA-Seq user dataset (fold change)

Use This Dataset in FungiDB

Compatibility Information

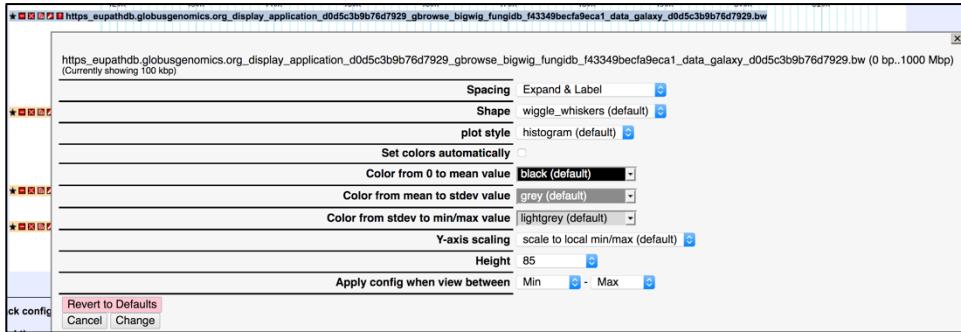
EuPathDB Website	Required Resource	Required Resource Release	Installed Resource Release
FungiDB	CryptococcusH99 Genome	29	29

This dataset is compatible with the current release, build 42, of FungiDB. It is installed for use.

GBrowse Tracks

Filename	GBrowse Status	Action
SRR505869.fastq.bw	Sent to GBrowse 3 months ago.	View In GBrowse
SRR513612.fastq.bw	Sent to GBrowse 5 hours ago.	View In GBrowse
SRR513612.fastq.bw	Sent to GBrowse 5 hours ago.	View In GBrowse

Note: You can modify the display of the tracks. For example, to adjust the Y axis click on the “configure this track icon” - 🔍. You can adjust various aspects of the display, including Y-axis scaling, height, and color of tracks.



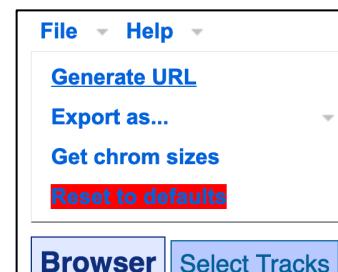
- Explore your results – zoom in or zoom out.
- Find a region of interest to examine or activate other transcriptomics tracks (integrated dataset for *Cryptococcus neoformans*) to see how your data match up.

For example, let's examine transcriptomics data for the current region. To select additional tracks navigate to the **Select Tracks** tab and search for an organism you are working with (e.g. shown for “*neoformans*” below)

- Select a dataset for your organism (shown here for *C. neoformans*) and click on Browser tab to return to viewing tracks.

The Cryptococcus neoformans transcriptome at the site of human meningitis (paired-end data) mRNASeq Coverage aligned to *C. neoformans* var. *grubii* H99 (Chen et al.) (log scale) [?] [showing 2/4 subtracks]

- Display other data available in FungiDB, if applicable.
- To share the current view of GBrowse session with others you would need to generate a session-specific url. To do this click on File, Generate URL at the top of GBrowse window:



Now copy and paste the url from your browser to share it with others.

This will quite a long url, e.g.: https://fungidb.org/cgi-bin/gbrowse/fungidb/?start=2256500;stop=2276499;ref=CP003820.1;width=1280;version=100;flip=0;grid=1;id=4f1b72865ffcb507843f766649ad169a;l=track_SRR505869.fastq-4016036.bw_1%1EcneoH99_ChennHumanMeningitis_Paired_End_rnaSeq_RSRCCoverage%1Etrack_SRR513612.fastq-4016036.bw_1%1Etrack_SRR505841.fastq-4016036.bw_1%1EGene%1EGSNAPUnifiedIntronJunctionRefined%1EGSNAPUnifiedIntronJunctionInclusive%1EGSNAPUnifiedIntronJunctionAll%1ESynteny;h_region=ChrI_A_nidulans_FGSC_A4%3A3484265..3488187%40wheat;h_feat=fgramph1_01t03307%40yellow

Step 7: Exploring the CuffDiff results

Cufflinks.cuffdiff finds significant changes in transcript expression, splicing, and promoter use. The Cufflinks.cuffdiff module takes a GTF file of transcripts as input, along with two or more SAM or BAM files containing the fragment alignments for two or more samples.

Cufflinks.cuffdiff produces a number of output files that contain test results for changes in expression at the level of transcripts, primary transcripts, and genes. It also tracks changes in the relative abundance of transcripts sharing a common transcription start site, and in the relative abundances of the primary transcripts of each gene. Tracking the former shows changes in splicing, and the latter shows changes in relative promoter use within a gene.

- To explore gene expression in your dataset, click on the **genes by RNA-Seq user dataset (fold change)** link

This action will initiate a transcriptomics evidence search that uses underlying FungiDB tools and automatically generated data to create a custom analysis of your data. This analysis is private and remains associated with your account. You can choose to share or delete this data as shown above.

My Dataset: Cryptococcus neoformans

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

Owner: Me

Description: mutant1

ID: 4016036

Data Type: RNASeq (RnaSeq 1.0)

Summary: datasets1

Created: 3 months ago

Dataset Size: 92.07 M

Quota Usage: 0.96% of 10.00 G

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

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

>Your RNA-Seq Dataset

Cryptococcus neoformans

For the Experiment unstranded return protein coding Genes that are up or down regulated with a Fold change >= 2 between each gene's expression value in the following Reference Samples

SRR505841.fastq
 SRR505869.fastq
 SRR513612.fastq

select all | clear all

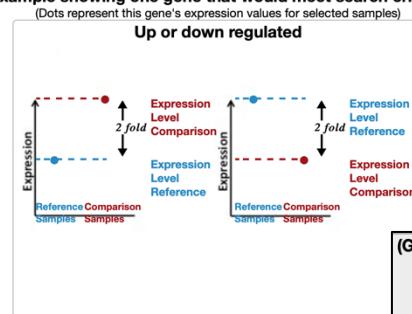
and its expression value in the following Comparison Samples

SRR505841.fastq
 SRR505869.fastq
 SRR513612.fastq

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



You are searching for genes that are up or down regulated between reference sample and one comparison sample.

For each gene, the search calculates:

$$fold\ change_{up} = \frac{comparison\ expression\ level}{reference\ expression\ level}$$

$$fold\ change_{down} = \frac{reference\ expression\ level}{comparison\ expression\ level}$$

and returns genes when $fold\ change_{up} \geq 2$ or $fold\ change_{down} \geq 2$.

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

(Genes)

Edit RNASEq (fc)
35 Genes Add Step

Step 1

Note: Each group is working with a different dataset so group results will be different from screenshots shown below.

- Examine your results. Create additional query (e.g. look for up-regulated or down-regulated genes choosing a default or custom Fold change values).

- How many genes were up-regulated or down-regulated?
 - Do the results make sense?
- Can you think of ways to integrate other FungiDB-based tools to create a comprehensive analysis of your data?
- Create a 3-4 step query using various tools in FungiDB by clicking Add Step button

Hint: You can look for secreted peptides or genes with transmembrane domains, or even cross reference your data with phenotype records (if available). You can even transform the data into orthologs in other species by Using Add Step > Transform by orthology.

Note: You can also download the CuffDiff results in tabular format directly from Galaxy instead of importing into My Data Sets space. This may be useful if you are interested in downloading the files to your computer or carrying out command line approach to further data analysis.

**** important: the file name ends with the extension.tabular – change this to .txt then open the file in Excel. Once the file is opened you can explore the results in Excel (e.g. sort the results based on the log2 fold change or by the columns called “value1” or “value2”).*

39: Cufflinks on data 26:

assembled transcripts
41,910 lines
format: gtf, database: FungiDB-
29_AfumigatusAf293_Genome

cufflinks v2.1.1
cufflinks -q --no-update-check -l
4000 -F 0.100000 -j 0.150000 -p 4
-G /mnt/galaxy/indices2/genomes
/FungiDB-
29_AfumigatusAf293_Genome
/annotation/FungiDB-
29_AfumigatusAf293.gff

display with IGV local

1.Seqname	2.Source	3.
Chr8_A_fumigatus_Af293	Cufflinks	tr
Chr8_A_fumigatus_Af293	Cufflinks	ex
Chr8_A_fumigatus_Af293	Cufflinks	tr