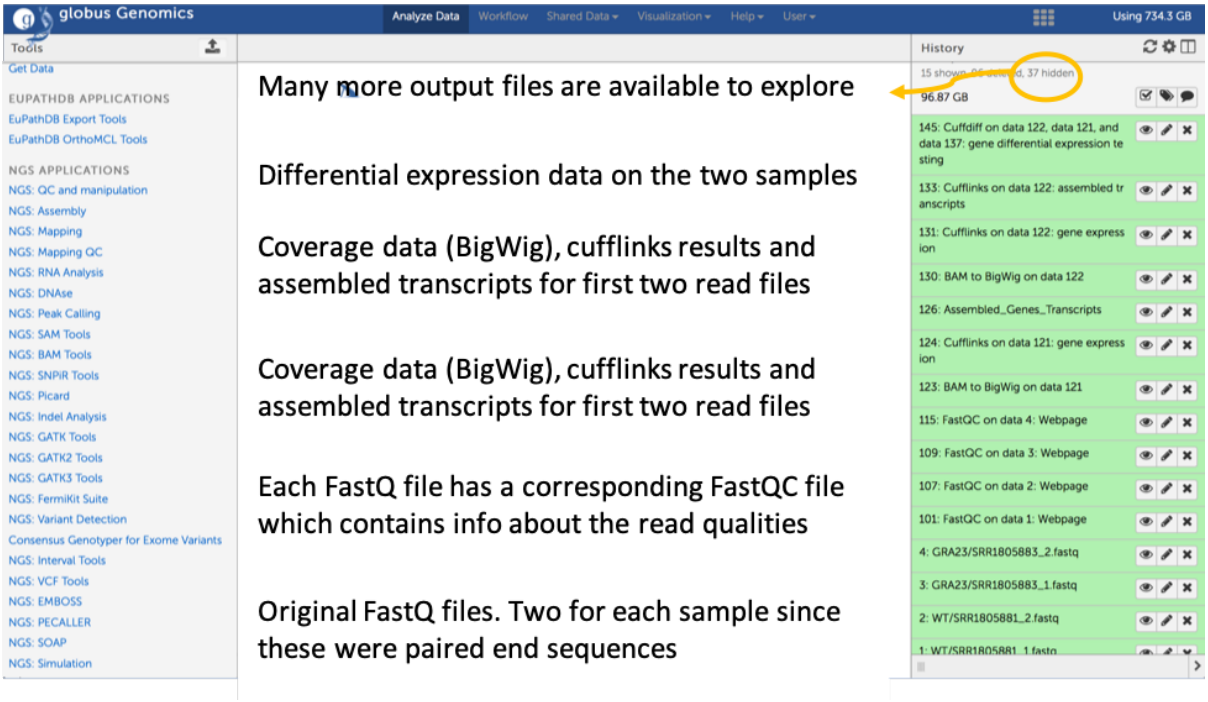
FungiDB: Data analysis via EuPathDB Galaxy.
RNA-Seq Part II: Data analysis (Group Exercise)

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.

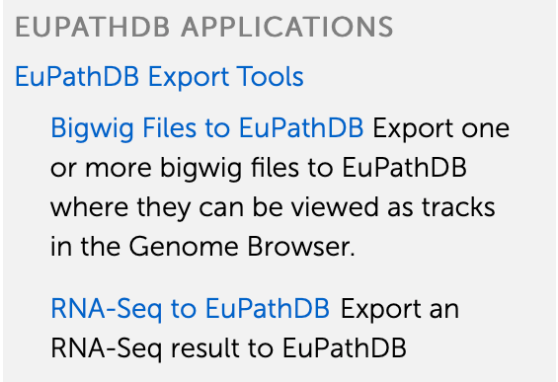
Step 1: Explore the FastQC results.
You can do this by clicking on the view data icon
(). An explanation of each of the FastQC results is provided as a link in class.



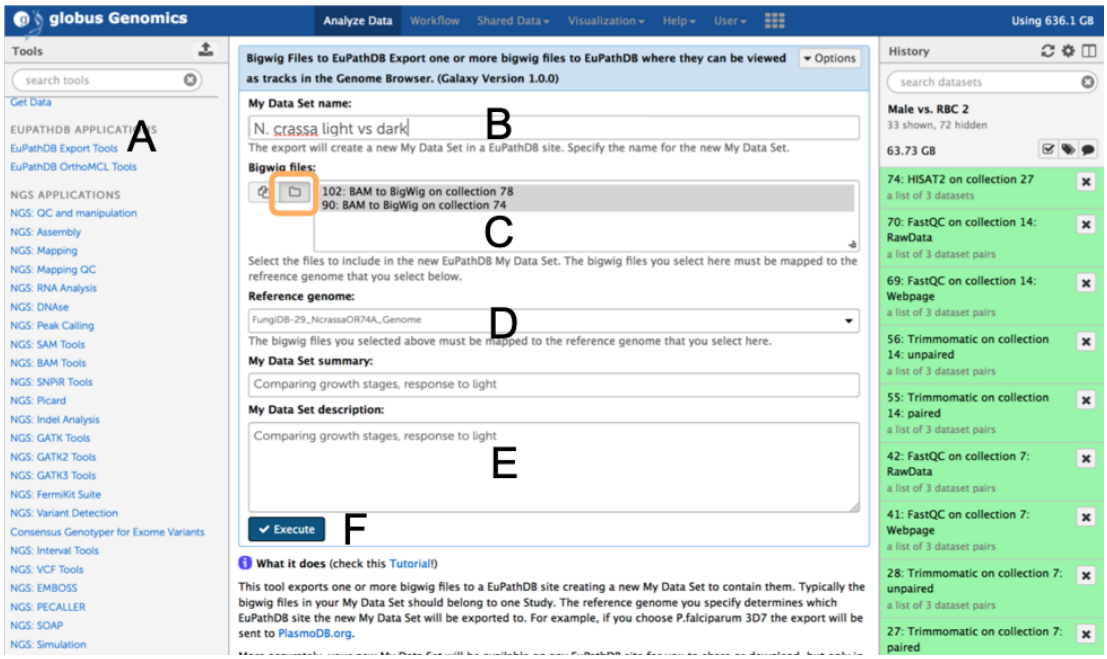
More about FastQC analysis and troubleshooting:

Step 2: Export Galaxy results to FungiDB/EuPathDB
Click on the EuPathDB Export Tools (left menu) and choose the RNA-Seq to EuPathDB option.

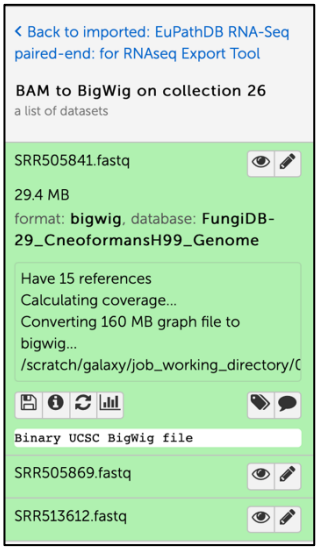
A. Click on EuPathDB Export Tools, then click on BigWig Files to EuPathDB (left Tools panel). The export tool will appear in the central portion.
B. Give your dataset a name.
C. Select "Dataset Collections" (icon looks like a folder). Then select all the BigWig collections that appear (Shift click).



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.

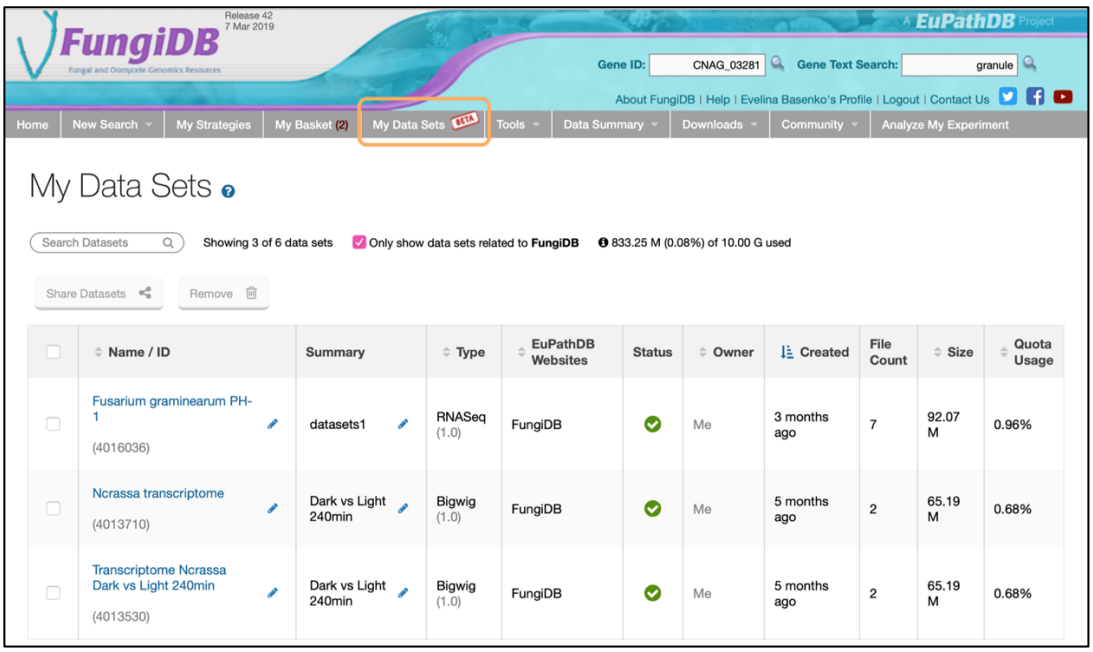


Note: You can also download the file by clicking on the floppy disk icon ( ), as shown below:

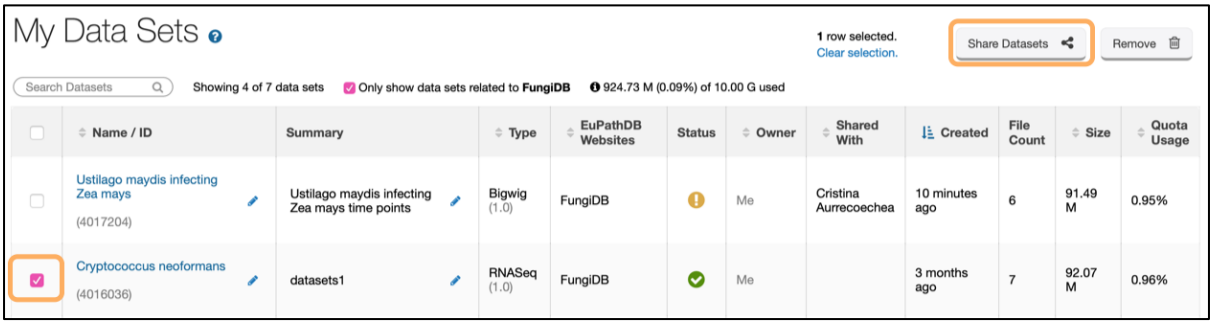


Step 3: Navigate to FungiDB/EuPathDB My Data Sets

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:



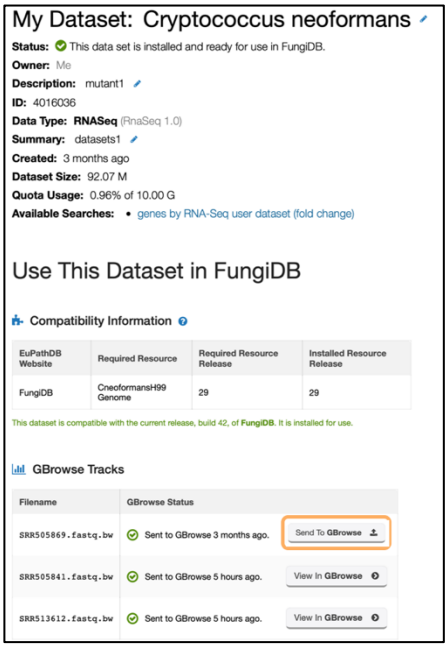
Note: You can share your datasets with collegues by selecting a dataset you wish to share and then click on the Share button at the top right:



Step 4: Send BigWig files in GBrowse

a. Click on the name of the dataset and examine the data set record page

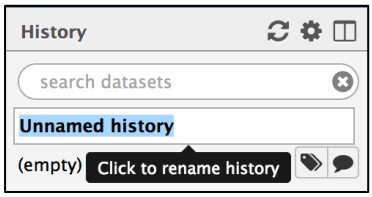
b. Scroll down to the GBrowse tracks section and click on the "Send to GBrowse" buttons for <u>each of the files in the list</u>**.**



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.

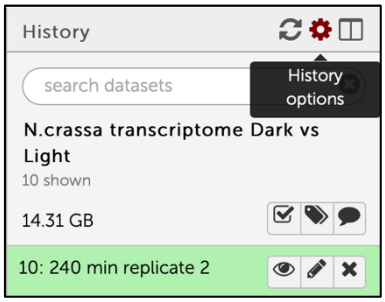
Step 5: Sharing histories with others

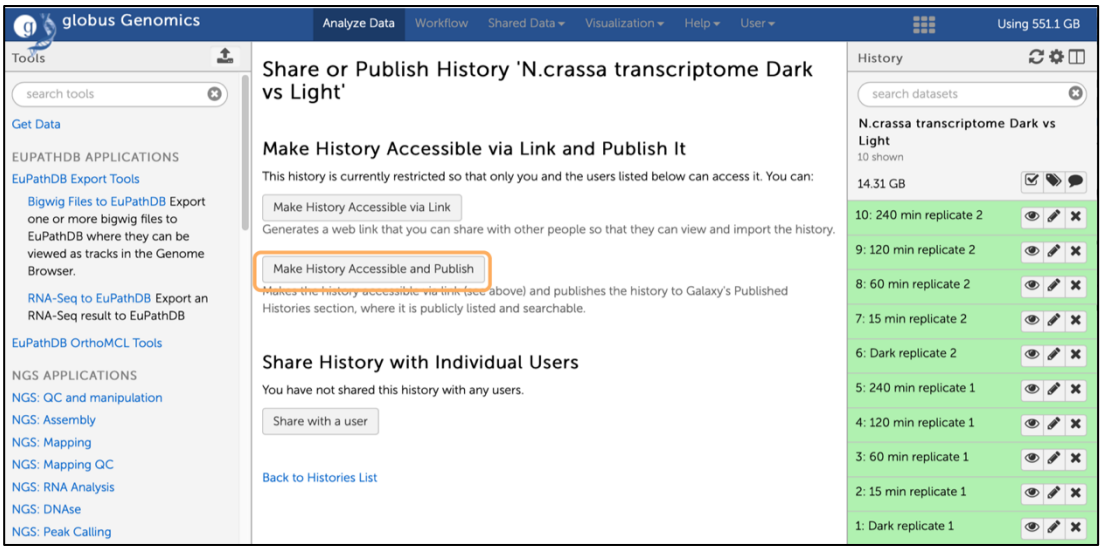
a. Make sure your history has a useful name – you can change the name by clicking on "unnamed history"



b. Click on the history options menu icon

c. Select the "Share or Publish" option, the click on the "Make History Accessible and Publish" button in the center section.

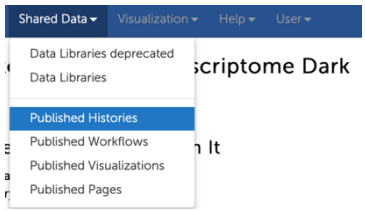


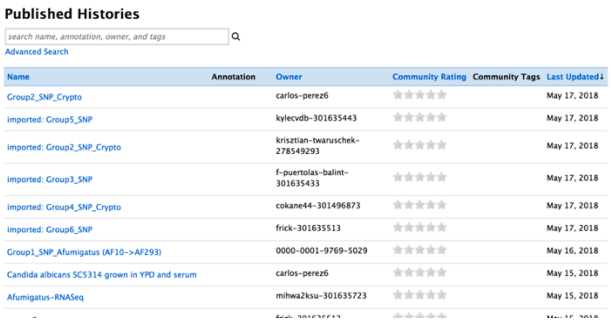


d. Select the To import a shared history, go to the "histories" section (under the shared data menu item).

e. Find the history you would like to import and

f. Click on the import link

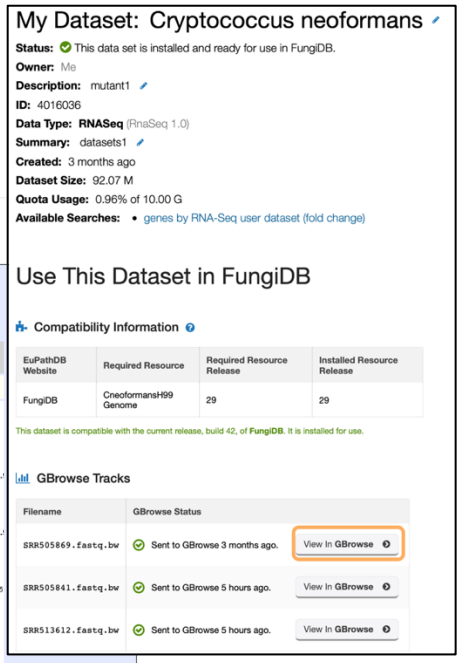


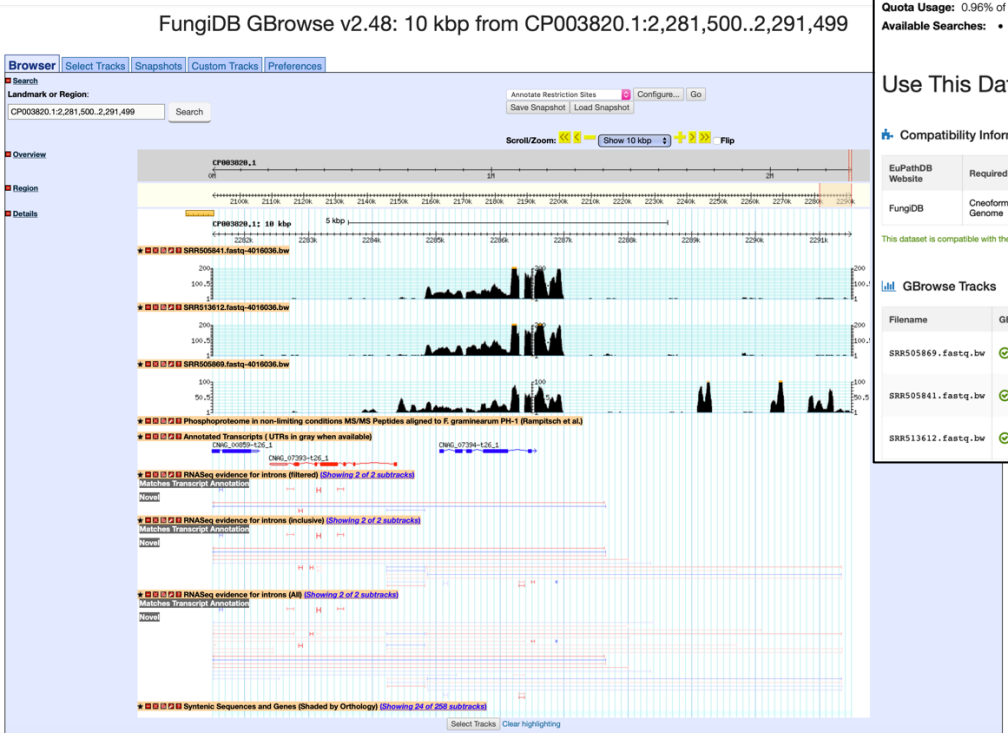


Note: This action will share your results with the rest of group. When working with your own data, click on the Make History Accessible via Link to get a link to share with selected individuals and not EuPathDB Galaxy users.

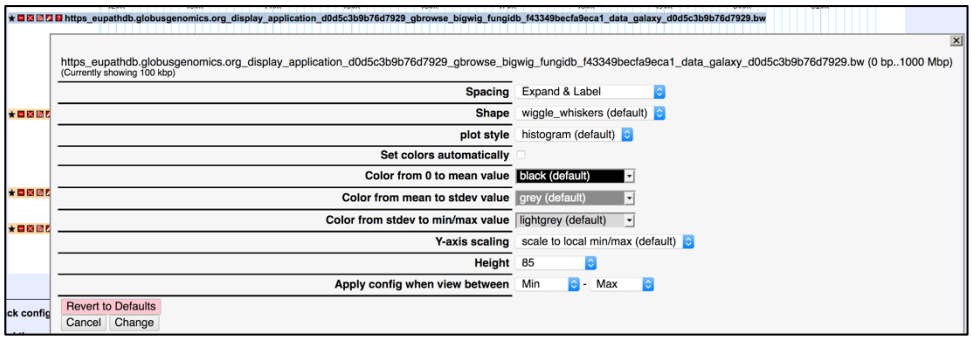
Step 6: Display coverage results in GBrowse.

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





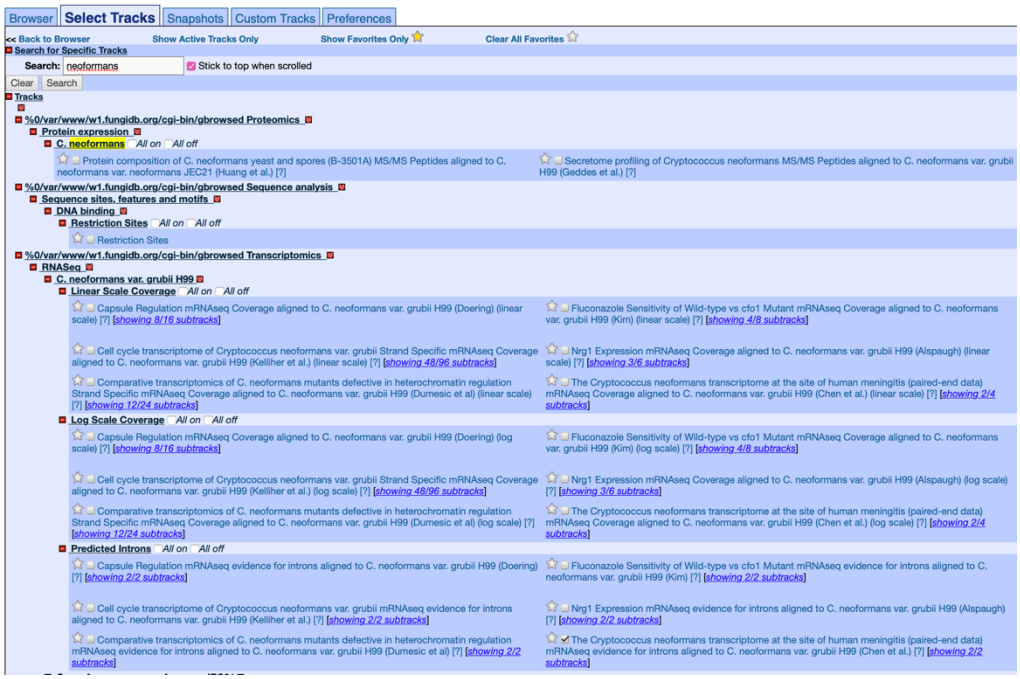
Note: You can modify the display of the tracks. For example, to adjust the Y axis click on the "configure this track icon" - *. You can adjust various aspects of the display, including Y-axis scaling, height, and color of tracks.*



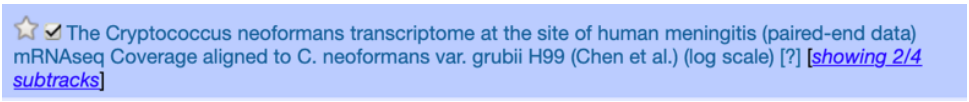
- Explore your results – zoom in or zoom out.

- 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)

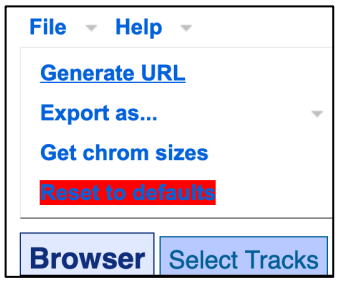


- Select a dataset for your organism (shown here for *C. neoformans*) and click on Browser tab to return to viewing tracks.



- 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 form your browser to share it with others.

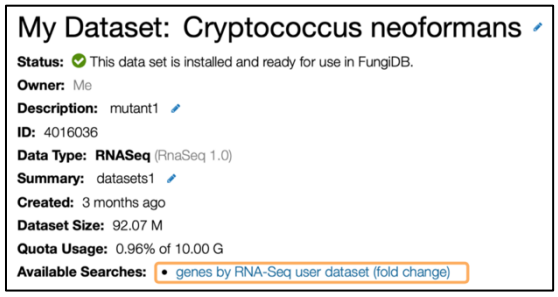
This will quite a long url, e.g.: https://fungidb.org/cgi-bin/gbrowse/fungidb/?start=2256500;stop=2276499;ref=CP003820.1;width=1280;version=100;flip=0;grid=1;id=4f1b72865ffcb507843f766649ad169a;l=track_SRR505869.fastq-4016036.bw_1%1EcneoH99_Chen_HumanMeningitis_Paired_End_rnaSeq_RSRCCoverage%1Etrack_SRR513612.fastq-4016036.bw_1%1Etrack_SRR505841.fastq-4016036.bw_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

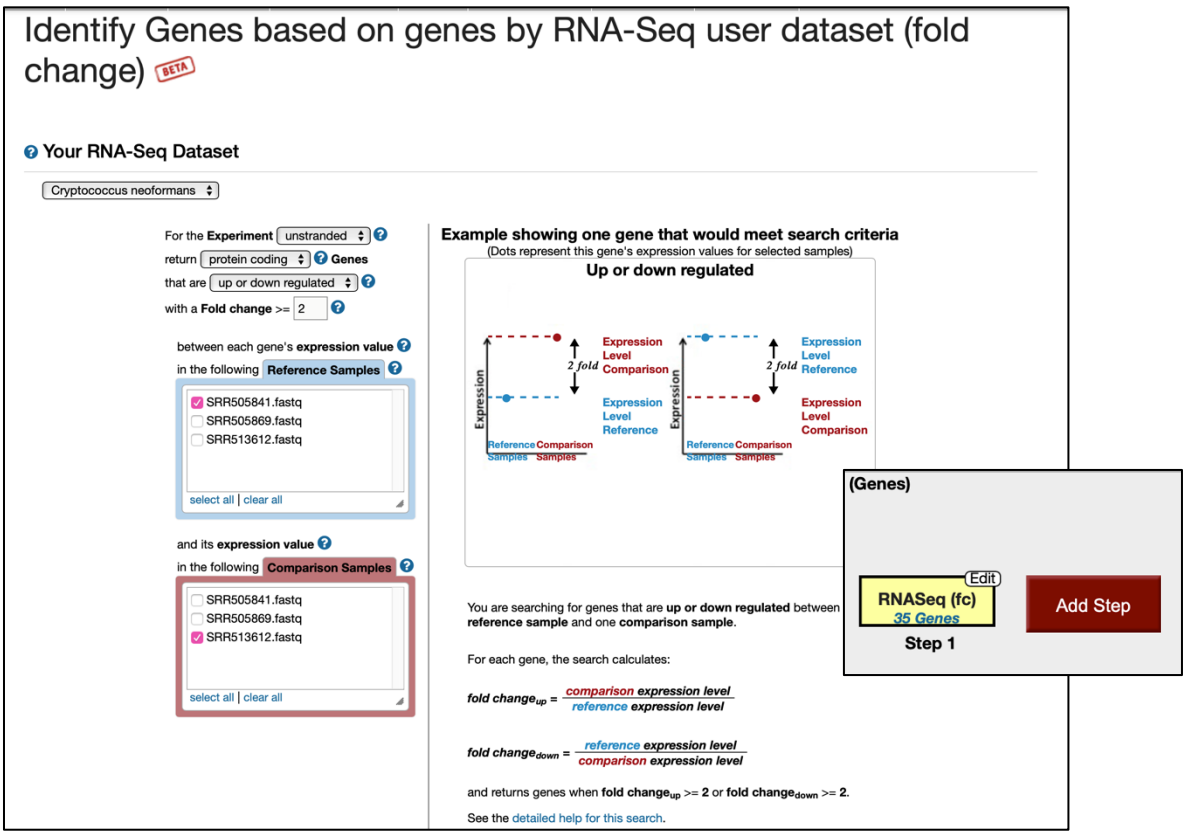
Cufflinks.cuffdiff finds significant changes in transcript expression, splicing, and promoter use. The Cufflinks.cuffdiff module takes a GTF file of transcripts as input, along with two or more SAM or BAM files containing the fragment alignments for two or more samples.

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 with 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.



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 gene 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").*