


RNA sequence data analysis

(Part 2: Loading data generated by the pathogen portal's RNAseq pipeline in the Genome Browser)

Exercise 11

For this exercise we will be using:

<http://pathogenportal.org>
<http://microsporidiadb.org>

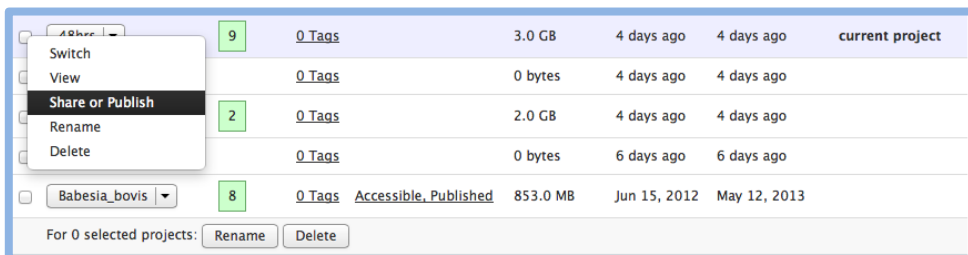
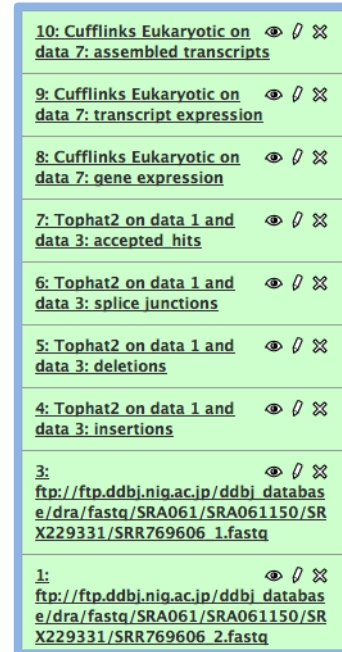
1. Explore the results of the RNA-sequence pipeline. What files were generated? To view contents of any of the results, click on the eye icon () next to the file name.

!!! important note – do not click on the icon next to the file called “Tophat2 on data 1 and data 3: accepted_hits” – this file is huge and will not display but rather will download the contents to your computer.

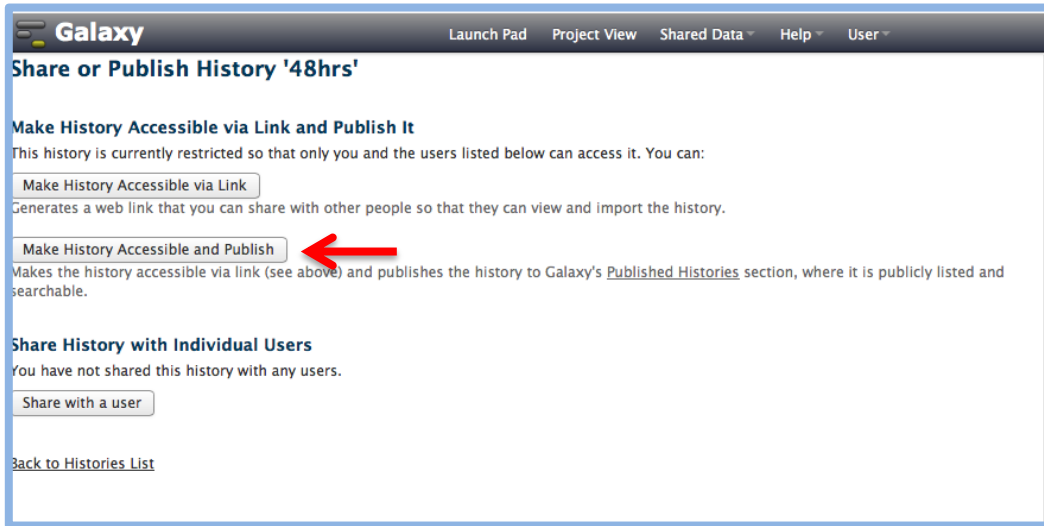
TopHat generates four files: insertions, deletions, splice junctions and accepted hits. The accepted hits file is the BAM file (binary alignment map). Note that many alignment programs will generate a file called a SAM file (sequence alignment map) which is a table including text of the alignment and mapping. However, for viewing results in a sequence browser like GBrowse, the file needs to be converted into the binary formatted (BAM) – you do not have to worry about this for this exercise.

Cufflinks generates three files: gene expression, transcript expression and assembled transcripts. The gene expression and transcript expression files for our purposes should be identical since EuPathDB genomes do not have separate genes and transcripts. These files include the FPKM values for each gene in the genome analyzed – in this case *Encephalitozoon cuniculi* ECII.

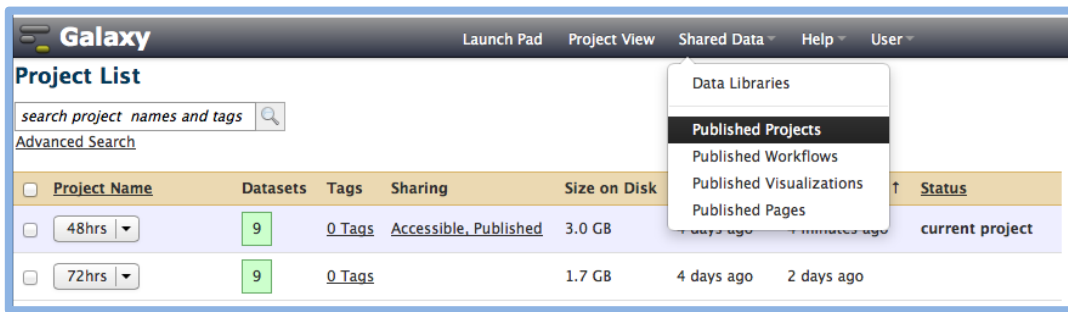
2. Share your accepted hits files. Click on the drop down menu for your project and select the option “share or publish”.



On the next page select the option: Make History Accessible and Publish



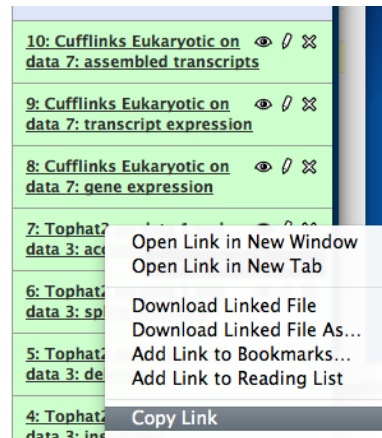
Once your project is published other people can access it by going to “Published Projects” section under the Shared data menu option in the Galaxy menu bar.



3. Load your BAM data into GBrowse. Navigate to the genome browser in MicrosporidiaDB and choose a landmark for *Encephalitozoon cuniculi* ECII you can just cut and paste the following into the “landmark or region” box:
ECII_CH11:98,571..148,570

Next, do the following to copy the link to the tophat accepted hits in pathogenportal to GBrowse:

- Control click (same as right click on a windows machine) on the eye icon for the tophat accepted hits.
- In GBrowse click on the “Custom Tracks” tab.



Browser | Select Tracks | Snapshots | **Custom Tracks** | Preferences

Search

Landmark or Region :
 Search

Examples : AL590443:85000-115000, ECI_CH11:115000..135000.

- c. Click on the “From a URL” link and paste the link you copied from pathogenportal.

Browser | Select Tracks | Snapshots | **Custom Tracks** | Preferences

Custom Tracks

[\[Help with uploading custom tracks\]](#)
 There are no tracks yet.
 Add custom tracks : [\[From text\]](#) [\[From a URL\]](#) [\[From a file\]](#)

Fetch track file from this URL
 Import Remove

- d. Delete the last portion of the URL: display/?preview=True

Browser | Select Tracks | Snapshots | **Custom Tracks** | Preferences

Custom Tracks

[\[Help with uploading custom tracks\]](#)
 There are no tracks yet.
 Add custom tracks : [\[From text\]](#) [\[From a URL\]](#) [\[From a file\]](#)

Fetch track file from this URL
 Import Remove

- e. Click on import.....and be patient.

Browser | Select Tracks | Snapshots | **Custom Tracks** | Preferences

Custom Tracks

[\[Help with uploading custom tracks\]](#)

http_rnaseq.pathogenportal.org_datasets_...
[\[http_rnaseq.pathogenportal.org_datasets_b8b5c2c15db111b6_\]](#) Imported
[Click to add a description](#)

Source files:	Created	Size	
http_rnaseq.pathogenportal.org_datasets_b8b5c2c15db111b6_	Tue Jun 4 01:06:27 2013	579480617 bytes	
Configuration	Tue Jun 4 01:09:06 2013	1425 bytes	[edit]

Add custom tracks : [\[From text\]](#) [\[From a URL\]](#) [\[From a file\]](#)

- f. Once the data has loaded click on the Browser tab to view your data.

4. Load the assembled transcript data. Cufflinks generates this file in a format called GFF. This format is not accepted by GBrowse so you have to convert it to another format called BED. To do this click on the pencil icon next to the file. Click on “Covert Format” then click on convert. A new file will be generated in BED format. You can not copy the link to the file and load it into GBrowse the same way you loaded the BAM file.

The image shows a sequence of steps in the GBrowse interface to convert a GFF file to BED format. It starts with a list of datasets on the left, including '10: Cufflinks Eukaryotic on data 7: assembled transcripts'. A red arrow points from the pencil icon next to this dataset to the 'Edit Attributes' tab of its detail view. In the 'Edit Attributes' view, another red arrow points from the 'Convert Format' tab to a 'Convert to new format' dialog box. This dialog box has a dropdown menu set to 'Convert GFF to BED' and a 'Convert' button. A red arrow points from the 'Convert' button to a new dataset entry '12: Convert GFF to BED on data 10' in the dataset list. Below the dialog box, there are 'Auto-detect' and 'Permissions' tabs, with a note: 'This will inspect the dataset and attempt to correct the above column values if they are not accurate.'