

Interpreting RNAseq Mapping results

(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://plasmodb.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.

- a. TopHat in RNArocket generates five files:
Align_summary: this includes a summary of how the alignment went (ie. the number of reads that were aligned).
- b. *Insertions*: reported insertions.
- c. *Deletions*: reported deletions.
- d. *Splice junctions*: reported junctions. Each junction consists of two connected BED blocks, where each block is as long as the maximal overhang of any read spanning the junction. The score is the number of alignments spanning the junction.
- e. *Accepted hits*: 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.

14: Tophat2 on data 2 and data 1: accepted hits (Genome Coverage BedGraph)  
13: Tophat2 on data 2 and data 1: accepted hits (- BigWig)  
12: Tophat2 on data 2 and data 1: accepted hits (+ BigWig)  
10: Cufflinks on data 7: assembled transcripts  
9: Cufflinks on data 7: transcript expression  
8: Cufflinks on data 7: gene expression  
7: Tophat2 on data 2 and data 1: accepted hits  
6: Tophat2 on data 2 and data 1: splice junctions  
5: Tophat2 on data 2 and data 1: deletions  
4: Tophat2 on data 2 and data 1: insertions  
3: Tophat2 on data 2 and data 1: align_summary  
2: EBI SRA: SRX129648 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR445/SRR445171/SRR445171_2.fast  

Cufflinks generates three files:

gene expression, transcript expression and assembled transcripts. The gene

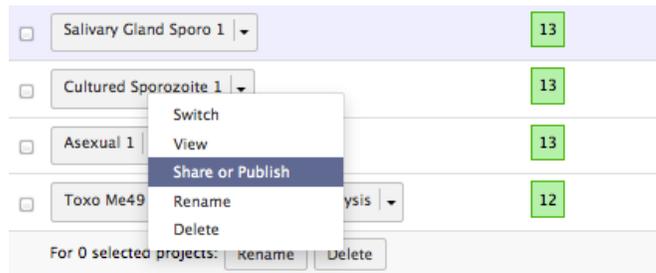
expression and transcript expression files for our purposes should be identical since EuPathDB genomes do not have separate genes and transcripts. These files include the FPKM values (Fragments Per Kilobase of transcript per Million mapped reads) for each gene in the genome analyzed - in this case *Giardia* assemblages.

Additional files include files of the format BigWig and BedGraph. You can read more about these file formats here:

<http://genome.ucsc.edu/goldenPath/help/bigWig.html>

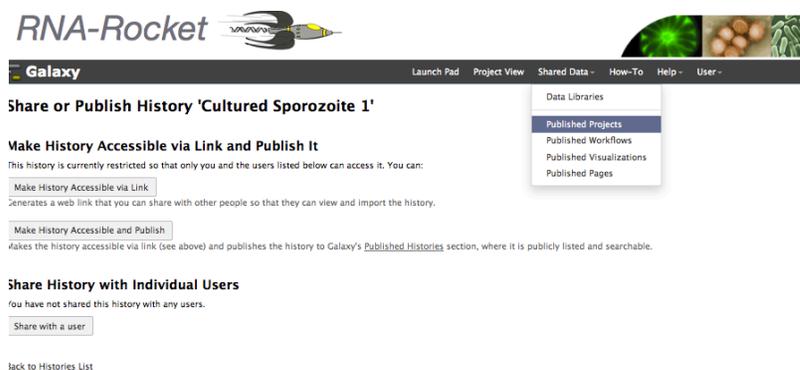
In a nutshell, these are file formats created from large binary files like BAM files and makes it possible to load these data in a genome browser.

Note: to share your data with the rest of the workshop, select “Share or Publish” from the drop down menu on project you want to share. On the next page click on “Make History Accessible and Publish”.



Project Results can be accessed via the "Project" panel to the right.

To import a history, select *Published Projects* from the *Shared data* menu item. Select the project you want to import then click on *import history* in the upper right hand side of the screen.



2. Load your BAM data (accepted hits) into GBrowse. Click on your "Tophat2 on data 2 and data 1: accepted_hits" in your project history panel. This will show you information about the file including a link to display data in GiardiaDB - click on the link.
3. Load the assembled transcript data. Essentially use a similar procedure as above.
4. Wait a couple of minutes for GBrowse to load your data.
5. Once data has been loaded, you can configure the track display settings. For example, you can adjust the Y-axis scaling to a fixed axis.

14: Tophat2 on data 1 and data 2: accepted_hits (Genome Coverage BedGraph)

13: Tophat2 on data 1 and data 2: accepted_hits (- BigWig)

12: Tophat2 on data 1 and data 2: accepted_hits (+ BigWig)

10: Cufflinks on data 7: assembled transcripts

9: Cufflinks on data 7: transcript expression

8: Cufflinks on data 7: gene expression

7: Tophat2 on data 1 and data 2: accepted_hits
 1.5 GB
 format: bam, database: pfal3D7
 Log: tool progress Log: tool progress [2014-06-14 08:55:03] Beginning TopHat run (v2.0.10)
 ----- [2014-06-14 08:55:03] Checking for Bowtie Bowtie version: 2.1.0.0 [2014-06-14 08:55:03] Checking for Samtools
 display at EupathDB [plasmodb](#)
 Binary bam alignments file

6: Tophat2 on data 1 and data 2: splice junctions

5: Tophat2 on data 1 and

To save or share track configurations, select 'Generate URL' from File menu (above) and cut and paste resulting URL.

If you load tracks and they appear empty, please try two things:
 1. Make sure you are viewing the correct species/strain to which the data was mapped.
 2. Reset your GBrowse and try again.

EuPathDB GBrowse Tutorial

Browser | Select Tracks | Snapshots | Custom Tracks | Preferences

Search
 M76611

Annotate Restriction Sites | Configure... | Go
 Save Snapshot | Load Snapshot

Examples: Pf3D7_11_v3:1205700..1305700, AABL01000674:7300..57800.

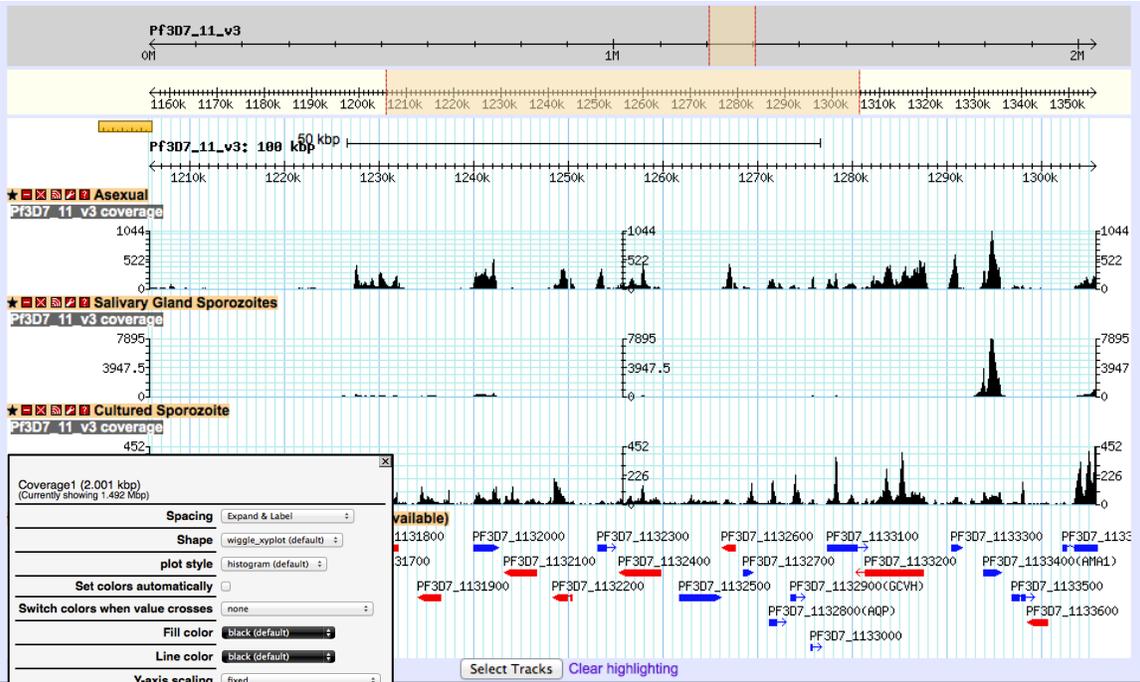
Data Source
 PlasmoDB GBrowse v2.48

ScrollZoom: << < > >> Show 31 bp < > < > Flip

The following 26 region match your request.

The default region is not that useful so click on the example landmark

Name	Type	Description	Position	Match Score
Pf_M7661100100	gene-annotation		PfIT_mito_v3:1550..1580	n/a
Pf_M7661100200	gene-annotation		PfIT_mito_v3:1581..1618	n/a
Pf_M7661100300	gene-annotation		PfIT_mito_v3:1619..1672	n/a
Pf_M7661100400	gene-annotation		PfIT_mito_v3:1673..1712	n/a
Pf_M7661100500	gene-annotation		PfIT_mito_v3:1751..1773	n/a
Pf_M7661100600	gene-annotation		PfIT_mito_v3:1830..1936	n/a
Pf_M7661100700	gene-annotation		PfIT_mito_v3:1937..2052	n/a
Pf_M7661100800	gene-annotation		PfIT_mito_v3:2053..2140	n/a



Coverage1 (2.001 kbp)
(Currently showing 1.482 Mbp)

Spacing Expand & Label

Shape wiggle_xyplot (default)

plot style histogram (default)

Set colors automatically

Switch colors when value crosses none

Fill color black (default)

Line color black (default)

Y-axis scaling fixed

Fixed Y-axis range 0 - 200

Show variance band

Height 50 (default)

Apply config when view between Min - Max

Revert to Defaults

Cancel Change