A multi-track RNA-seq browser for visualization of Arabidopsis thaliana transcription patterns from different growth states and conditions.

Table of Contents

Date: October 16, 2015

Agenda:

1. Check if having local BAM files can speed up data retrieval using samtools' mpileup () call.

Protocol:

1. Check if local BAM files can speed up data retrieval using samtools' mpileup () call.
    a. Downloaded the BAM file for experiment SRR547531 using wget ().
    b. Executed mpileup () through SSH.
        i. Used local BAM file and compare to iPlant BAM file
    c. Ran the output.cgi script with the two BAM files (local vs. iPlant).
        i. Used Chrome Dev Tools to analyze the TTFB (time to first byte).

Results:

1. Check if local BAM files can speed up data retrieval using samtools' mpileup () call.
    a. Done, file size = 662MB.
    b. Local BAM file mpileup call returns data very quickly (< 1 second), iPlant BAM file takes 10-20 seconds.
    c. Local BAM file returns data in ~600ms! The iPlant BAM file takes ~60 seconds!
        i. Only calling the generate_rnaseq_graph() function once to produce a single image.

Notes/Questions:

- Figure out why a URL to the BAM file doesn't work. Currently it works with a relative path..!
- See if BAM files hosted on Drop box result in data retrieval that is just as fast as local BAM files. This might implicate the iPlant server as the bottleneck and prove the HTTP request's innocence.

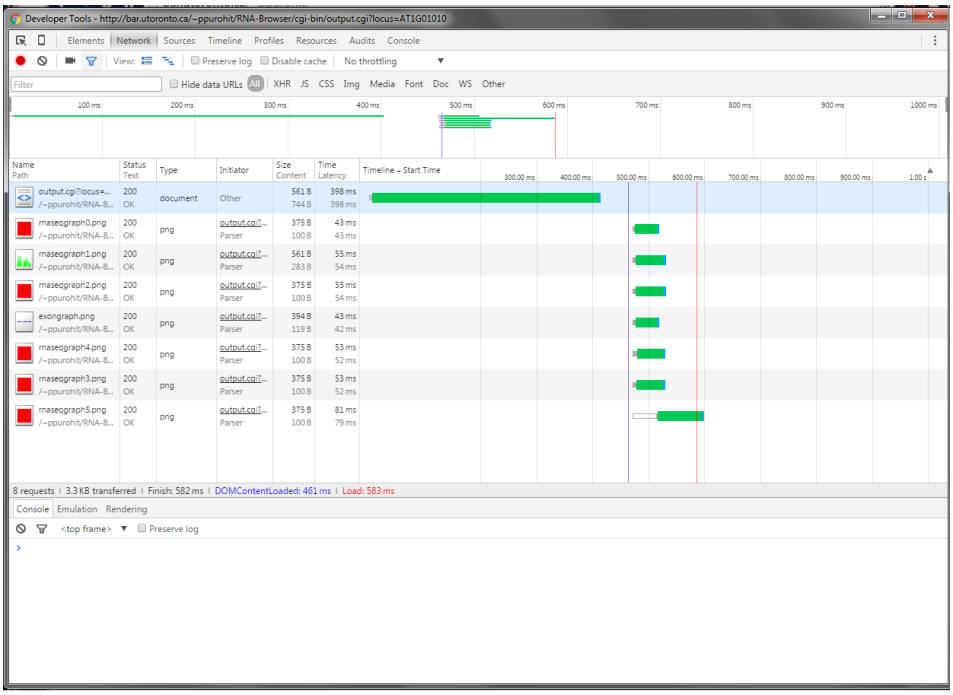

Figure 1: ~600ms for local BAM file.

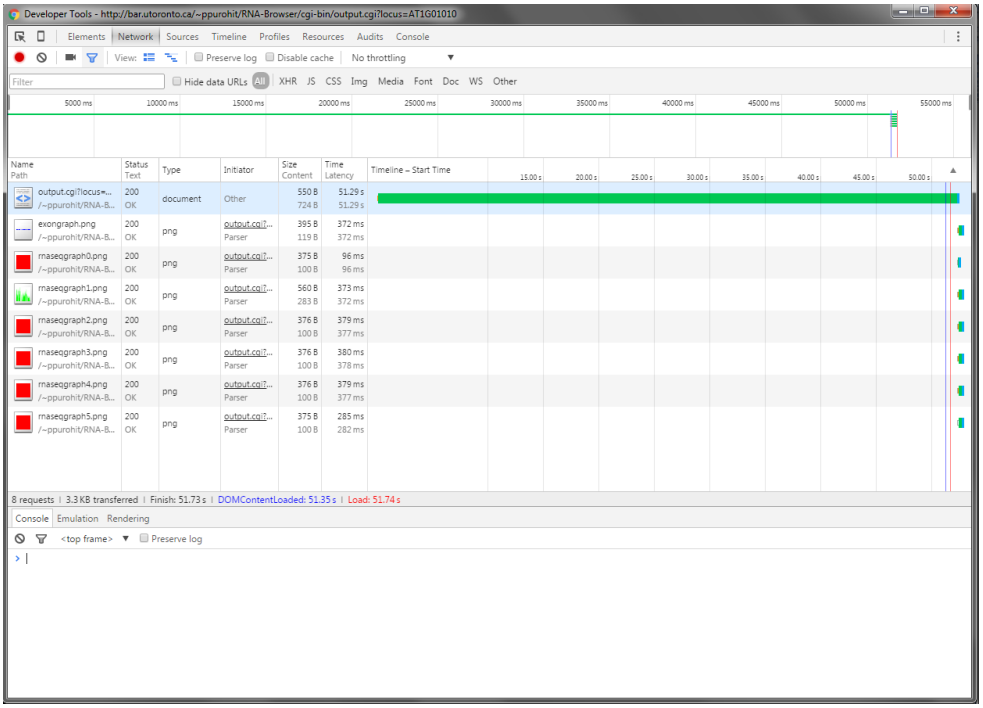

Figure 2: ~60 seconds for iPlant BAM file.

Date: October 21, 2015

Agenda:

1. Check if BAM files stored on other servers are just as fast as local BAM for the mpileup() call.

Protocol:

1. Executed 3 mpileup () commands through SSH.
    a. `time samtools mpileup -r chr2:8032000-10329941 http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeSydhRnaSeq/wgEncodeSydhRnaSeqK562Ifna6hPolyaAln.bam > ucsc.txt`
    b. `time samtools mpileup -r Chr2:10327050-10329941 http://vision.iplantcollaborative.org/iplant/home/araport/rnaseq_bam/aerial/SRR547531/accepted_hits.bam > iplant.txt`
    c. `time samtools mpileup -r Chr2:10327050-10329941 http://bar.utoronto.ca/~ppurohit/RNA-Browser/cgi-bin/data/iplant/home/araport/rnaseq_bam/aerial/SRR547531/accepted_hits.bam > bar.txt`

Results:

1. Done.
2. Executed mpileup () through SSH.
    a. UCSC.edu BAM file:
        i. 1.322 s for 6 260 175 bytes
    b. iPlant BAM file:
        i. 15.294 s for 1 248 931 bytes
    c. bar.utoronto.ca BAM file:
        i. 0.060 s for 1 248 931 bytes

Notes/Questions:

- The UCSC BAM file returned only ~1500 bytes for the `Chr2:10327050-10329941` region. This smaller size could be the reason why the command is fast. The query region was therefore increased to get 6x more data.
    - … and it's still fast!

Date: October 23, 2015

To-do List:

1. Research proposal (Oct 29)
2. XML File Updates
    a. Change the Newland links to their new Vision links
    b. Add colours as discussed in Oct 22 meeting w/ NP
        i. Try to make them web safe colours without changing the shade too much
    c. Add the missing information in the XML file (some pictures missing)
3. Download the mpileup data for the default locus from all of the BAM files on iPlant
    a. Spread this out over 3-4 days
4. RNA Browser:
    a. Start working with locally stored mpileup data
    b. Change the image dimensions for RNA-Seq graph to ~250 x 50.
    c. Make the exon graph slim and add a horizontal line through the middle
    d. Start making dynamic requests for RNA-Seq graphs
        i. Sample flow:
            1. User comes on our app
            2. The RNA-Seq graphs of the default gene are pre-made and loaded on page load
            3. When the user enters a particular locus, the app will load 3 graphs, the rest are dynamically generated after page load ...
    e. Image read map heights:
        i. Start with default height of 1000 reads
        ii. Have a button that allows the user to re-generate all images such that the max height is the maximum read for any base pair
    f. FPKM calculations:
        i. Genie is doing this, make sure to have the information she needs for these calculations

Date: October 25, 2015

Agenda:

1. Get Richard's new BAM Locator XML file and update the `newland` links to `vision` links.
2. Change the `displayxml.cgi` to account for changes made in the attribute names.

Protocol:

1. Got Richard's new BAM Locator XML file and updated the `newland` links to `vision` links.
    a. Wrote a Java program to go through an XML file, find the correct new link and replace it.
2. The attribute name changed to `svgname` from `subunitname` (correct name is there now).
    a. Changed the correct IF statement in displayxml.cgi to look for `svgname` as opposed to `subunitname`.

Results:

1. Success. The code replaced all files correctly.
2. Success. The `displayxml.cgi` script works correctly with the new BAM locator XML file.

Date: October 29, 2015

Agenda:

1. Add colours picked out by Dr. Provart to the BAM locator XML file's foreground column.
2. Update the BAM file links to the new Amazon S3 links.

Protocol:

1. Add colours picked out …
    a. Manually copy-pasted the new HEX colour codes.
    b. Fixed cases where the colour attribute was missing.
2. Update BAM file links to Amazon …
    a. Ran the Java code from Oct 25, 2015 w/ Amazon S3 prefix instead of the iPlant prefix

Results:

1. Add colours picked out…
    a. Successful.
    b. Live on BAR @ http://bar.utoronto.ca/~ppurohit/RNA-Browser/cgi-bin/displayxml.cgi
2. Update BAM file links to Amazon
    a. Success. It is live on BAR at the link shown above.

Notes:

1. The images still look incorrect for some of the experiments (i.e. the image should be root but is not)
    a. Look into this and fix it.