

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

Table of Contents

1. [Network analysis of local vs. iPlant BAM files.](#) [October 16, 2015](#)
2. [Network analysis of UCSC BAM files.](#) [October 21, 2015](#)
- 3.

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.

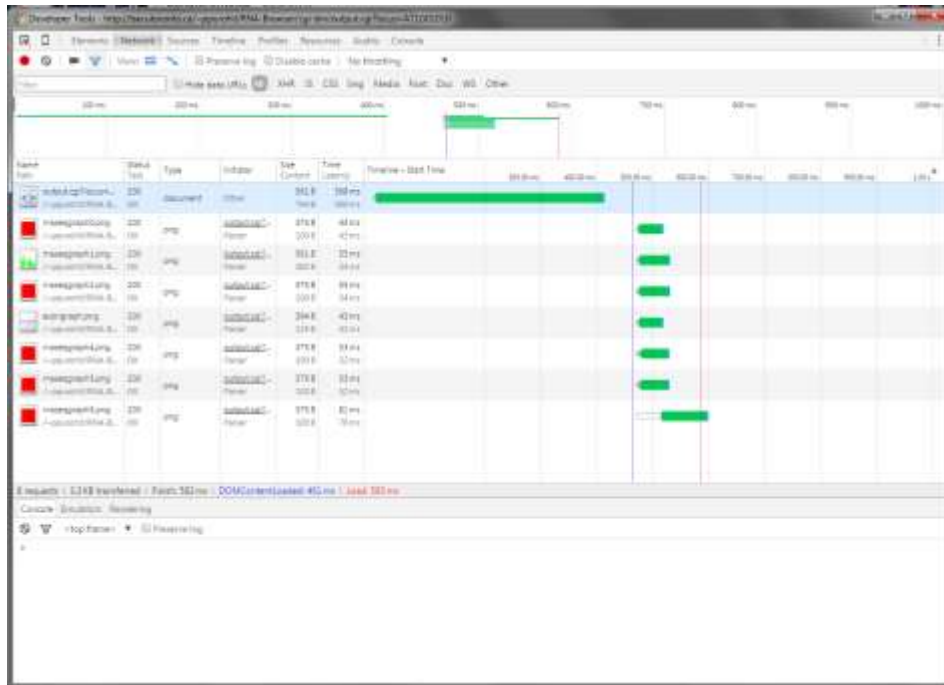


Figure 1: ~600ms for local BAM file.

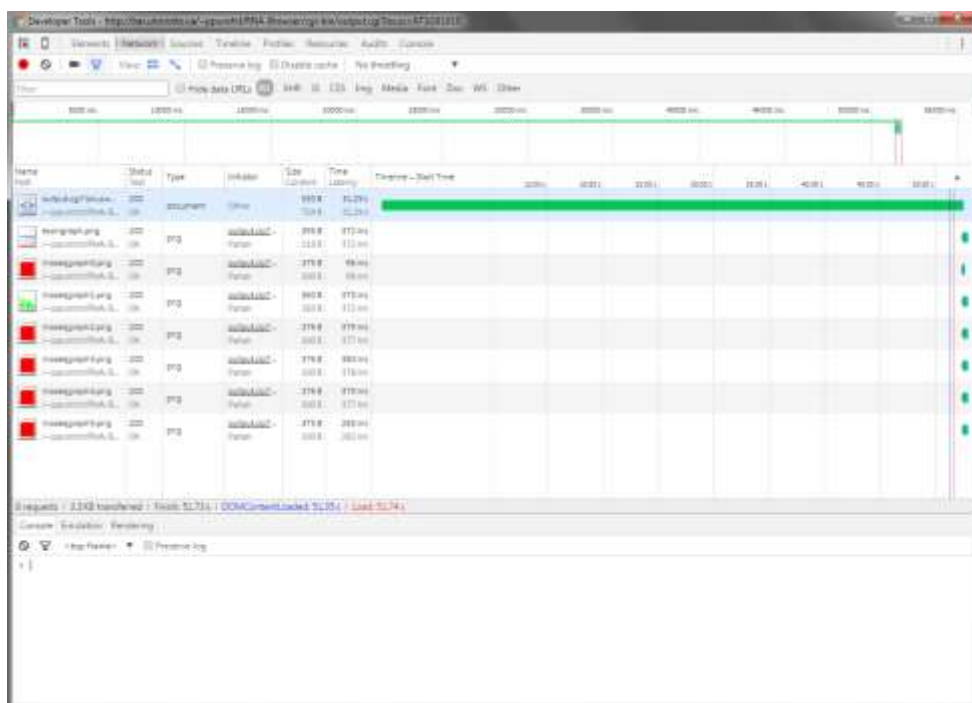


Figure 2: ~60 seconds for iPlant BAM file.

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.
 - o ... and it's still fast!