Using JBrowse with large amounts of data

Mitchell Skinner
Ian Holmes Lab
UC Berkeley

“Large amounts of data”

=

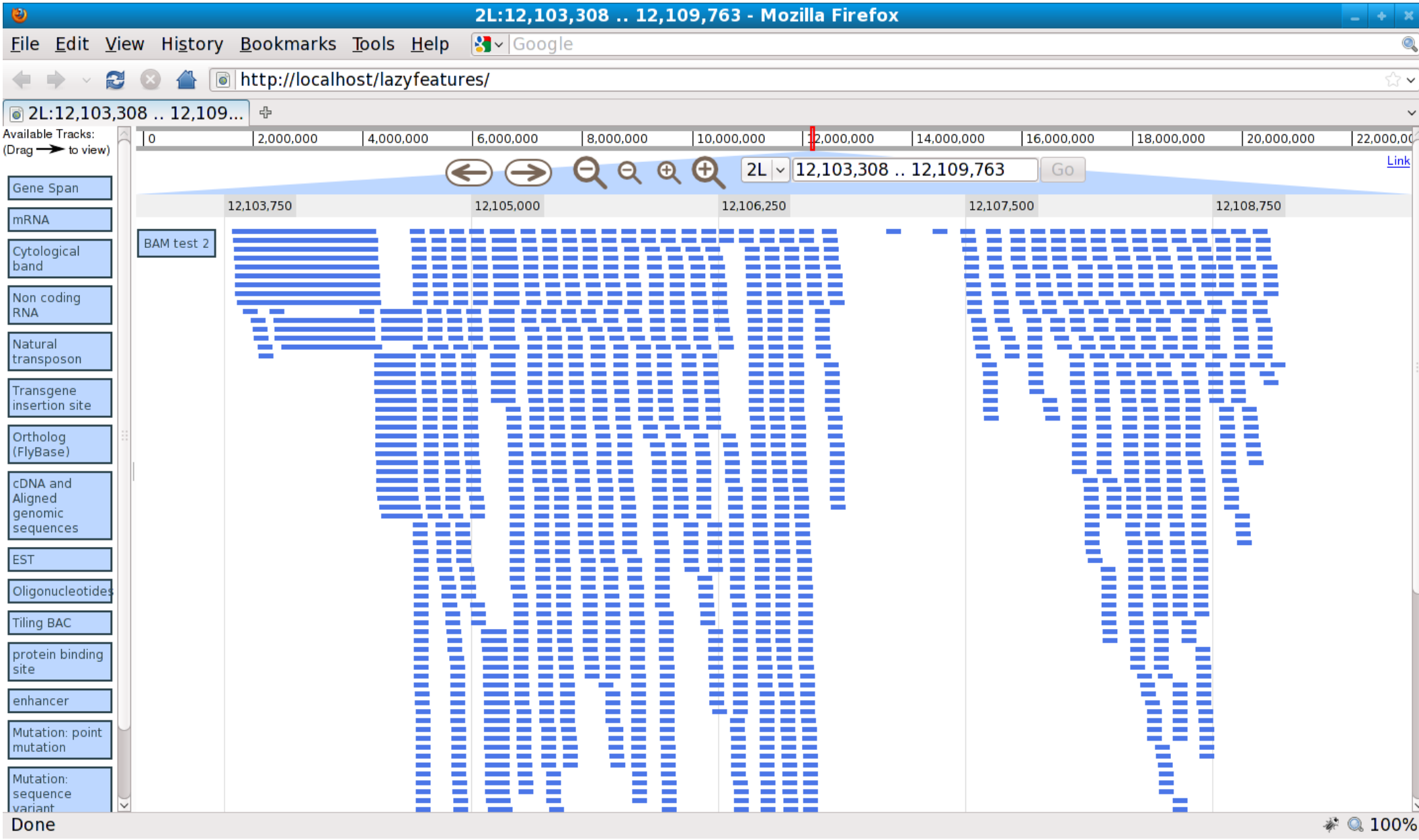
next-gen sequencing

JBrowse

- Moves work from web server to web browser
- Web browsers are limited
  - Have to be careful not to overload them
- Have to break up the data into digestible chunks

JBrowse

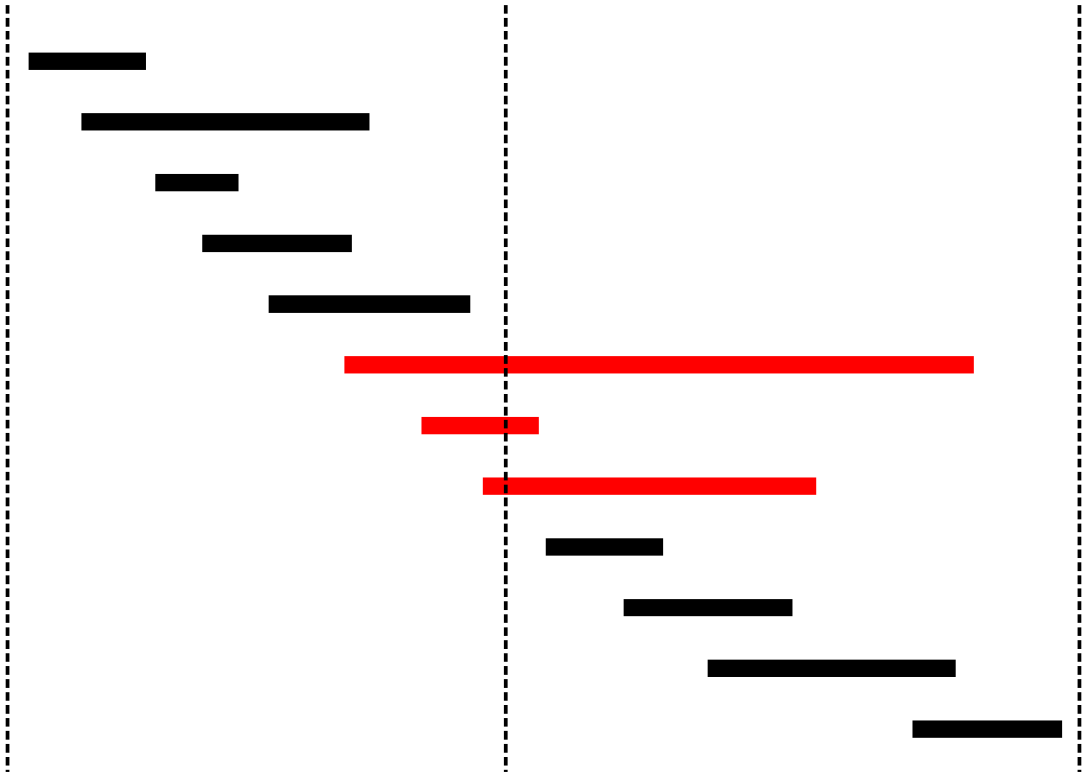
- Assumes that reads are much more common than writes
- Moves work from read-time to write-time
- Have to break up the data ahead of time


2L:12,103,308 .. 12,109,763 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Google
http://localhost/lazyfeatures/
2L:12,103,308 .. 12,109...
Available Tracks:
(Drag → to view)
Gene Span
mRNA
Cytological band
Non coding RNA
Natural transposon
Transgene insertion site
Ortholog (FlyBase)
cDNA and Aligned genomic sequences
EST
Oligonucleotides
Tiling BAC
protein binding site
enhancer
Mutation: point mutation
Mutation: sequence variant
0
2,000,000
4,000,000
6,000,000
8,000,000
10,000,000
12,000,000
14,000,000
16,000,000
18,000,000
20,000,000
22,000,00
2L
12,103,308 .. 12,109,763
Go
Link
12,103,750
12,105,000
12,106,250
12,107,500
12,108,750
BAM test 2
Done
100%

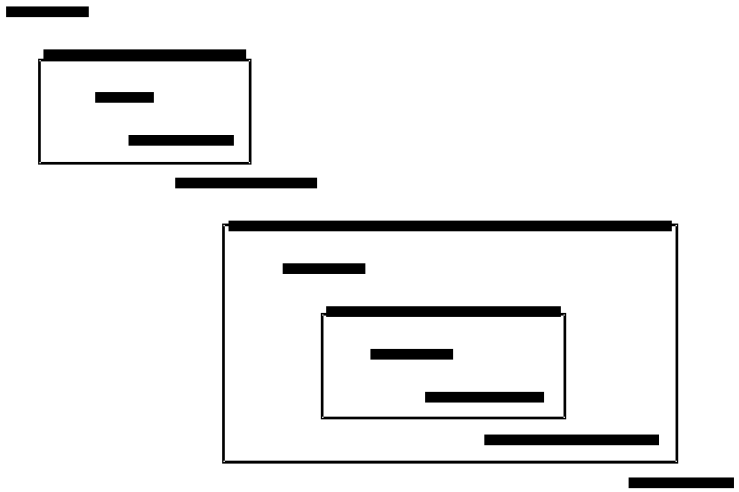
Results

- On one test data set:
  - 4.4 million features
  - 8 minutes to process
    - From 242 megabyte BAM file
    - Not paired-end
  - Used 400 megabytes of RAM
  - 330 megabytes on disk (without sequence)
  - Compresses down to 80 megabytes

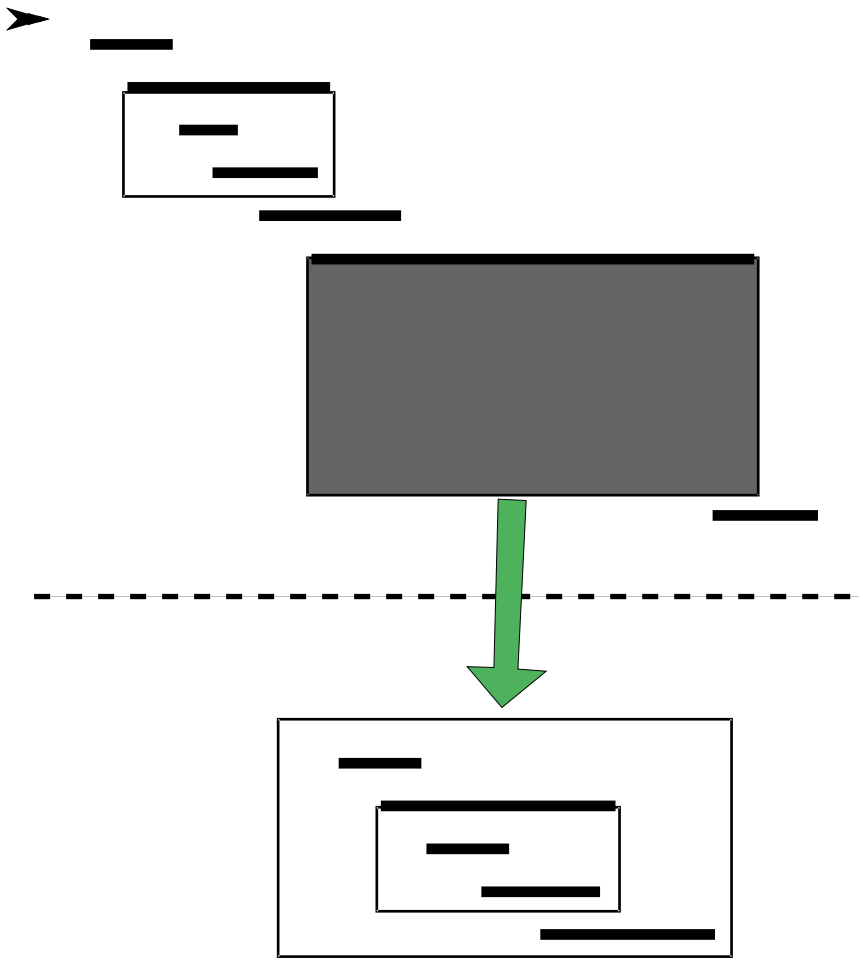
How to break up the data?



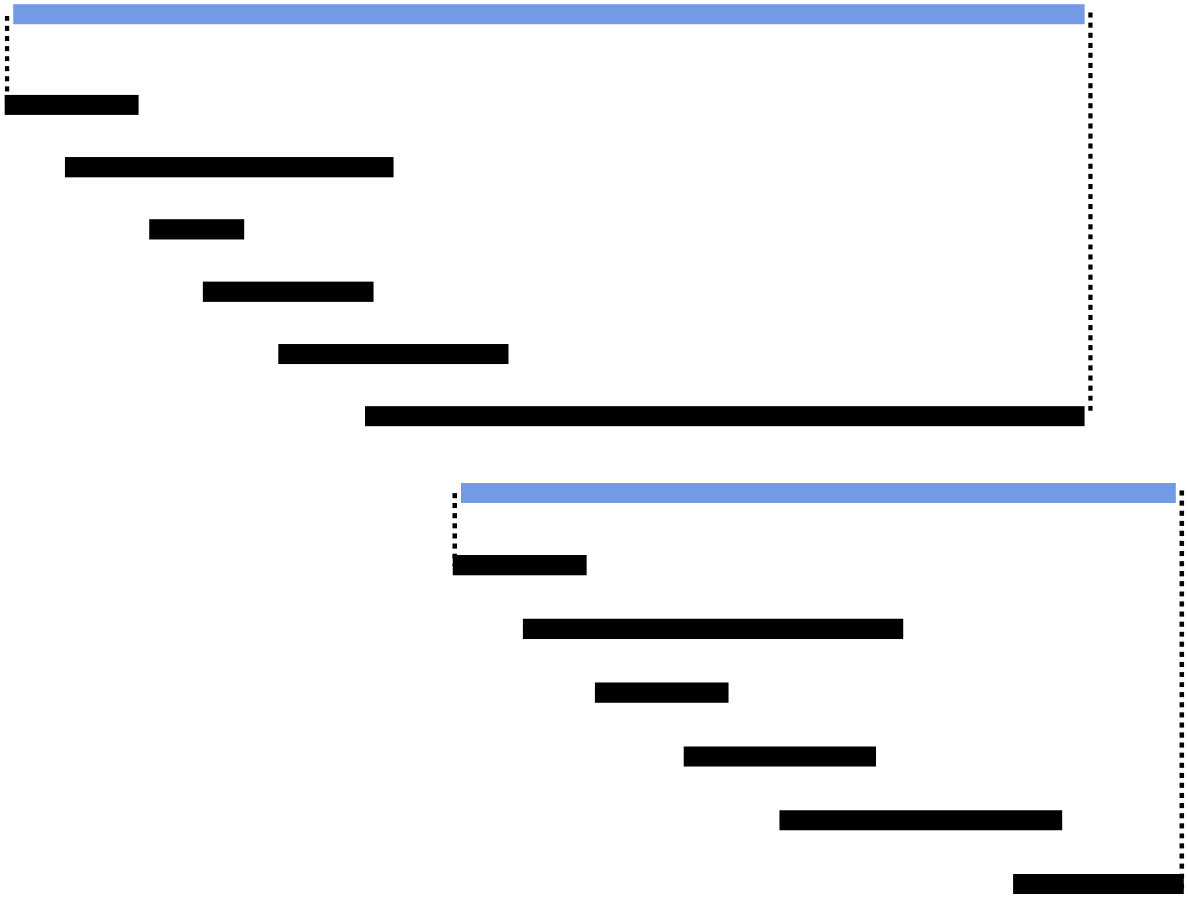
JBrowse uses NCLists



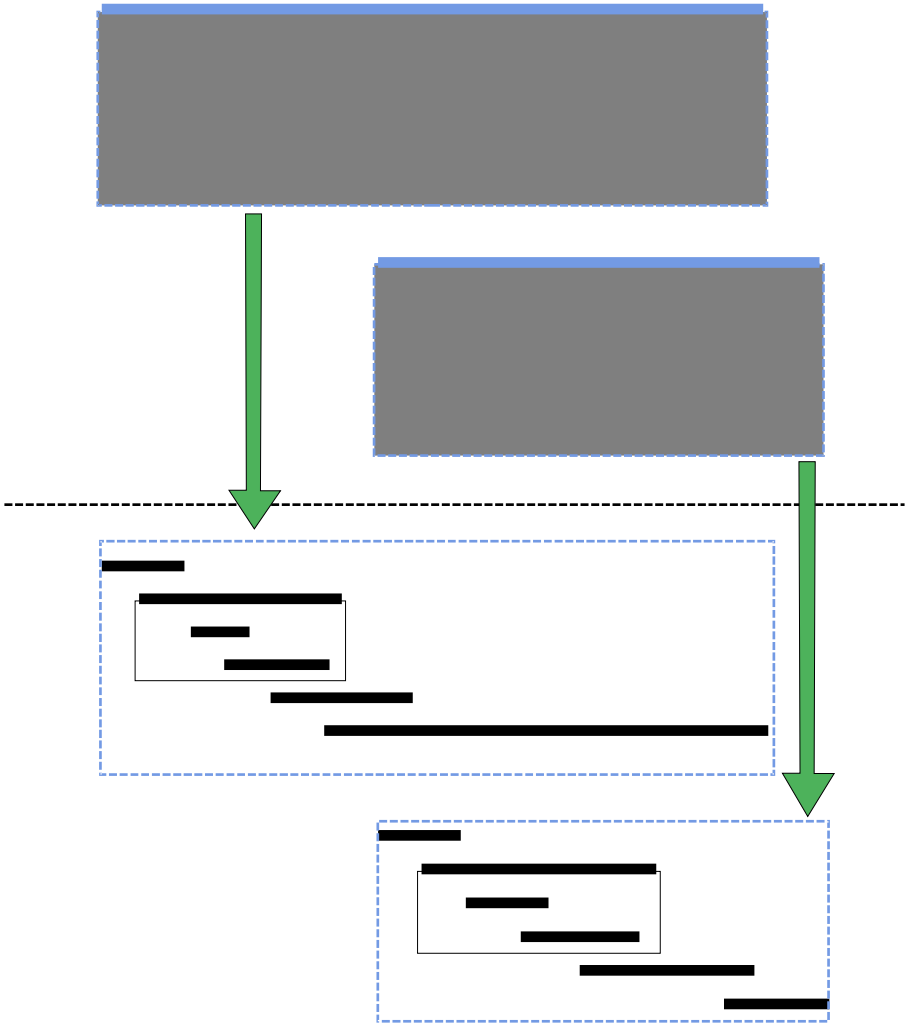
Lazy NCLists?



“fake” features



Lazy Loading in Jbrowse



Other approaches to lazy loading

- Heng Li (SAMTools)
  - Binning, linear index
- Jim Kent (BigBed/BigWig)
  - R-Trees
- JBrowse javascript client can't use them directly
  - But Jbrowse could access them through a proxy

Thanks

- Ian Holmes
- Andrew Uzilov
- Chris Mungall
- Lincoln Stein
- GMOD
- NHGRI

More:
http://jbrowse.org