Viewing CyVerse Hosted Data at UCSC

Host data at CyVerse

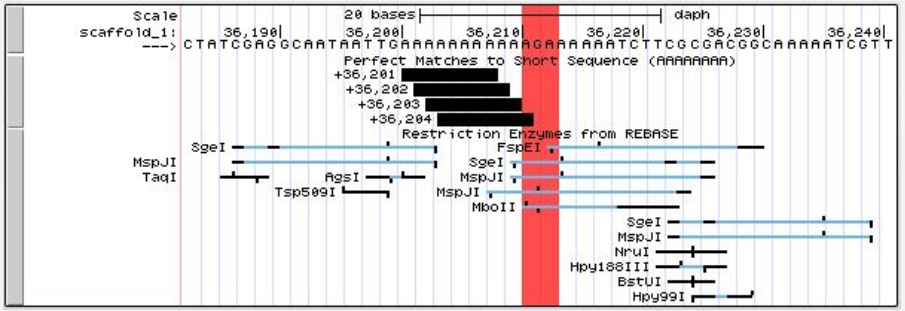
Binary indexed files:
twoBitPath, bigDataUrl

Text files: ***hub.txt,***
genomes.txt, trackDb.txt

```
>scaffold_1
GTTGTAAATACTCTATTCTACAATAAAACCAA
TCATAGGTTGAATTGGCGTTGAAGTAAAACAAA
...
>scaffold_2
AGTTATGACAAACTATAAAAAGTCGGTAGAGACAAAAG
TCGTTCGTGGACGAAGCGACCAAAACTGAGCACAAGATA
...
>scaffold_3
CATAAATTCATAAATCAATTCATTCGAAGAATAATT
TAGAAAATTTCCCAGGAAGTTTGAAGTTGCTA
...
```

```
hubDirectory
|__hub.txt
|__genomes.txt
|__hg19
      |__trackDb.txt
```

Visualize at UCSC

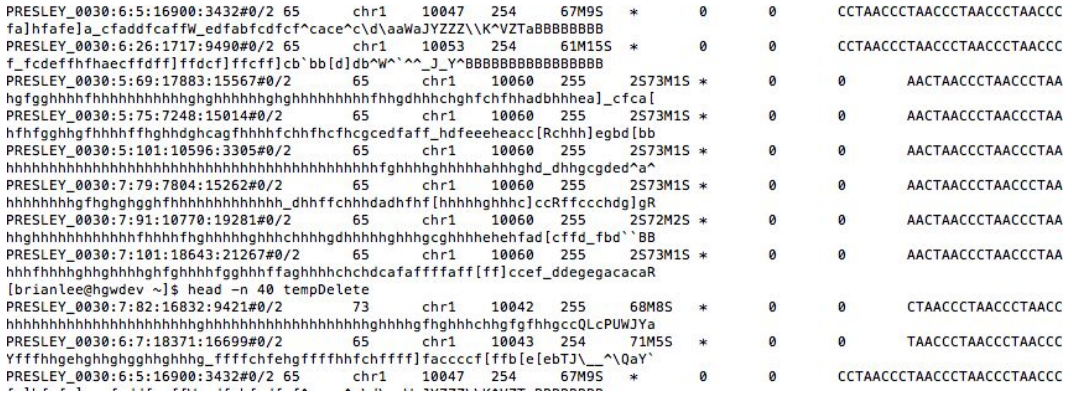


CyVerse's *"Send to Genome Browser"* option creates links accepting byte-range requests for binary index data.

Examples of Visualizing binary indexed data in the UCSC Genome Browser

A BAM file (.bam) is the binary version of a SAM file.

```
samtools view -S -b sample.sam > sample.bam
```



The resulting binary file **sample.bam** (with an additional accompanied index file *sample.bam.bai*) can have data more easily extracted and can also be viewed in Genome Browsers.

A 2bit file is a binary indexed version of a FASTA file (stores sequence ACGT as 00 01 11 10)

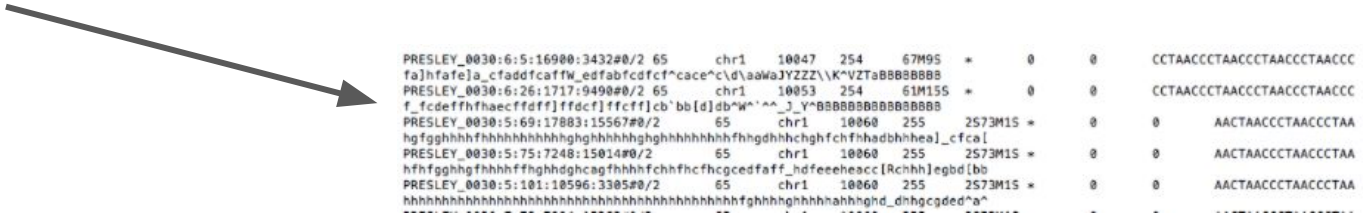
```
faToTwoBit input.fasta output.2bit
```

```
>scaffold_1
GTTGTAAATACTCTATTCTACAATAAAACCAAAAGATAACTCTTTATCAG
TCATAGGTTGAATTGGCGTTGAAGTAAAACAAAAAATACTGCTCAAAAGG
...
>scaffold_2
AGTTATGACAAACTATAAAAAGTCGGTAGAGACAAAAGCGAAAAGGATCT
TCGTTCGTGGACGAAGCGACCAAAACTGAGCACAAGATAAATCCCGAATA
...
>scaffold_3
CATAAATTCATAAATCAATTCATGAAGAATAATTTTAGAAAAATGGTTCA
TAGAAAATTTCCCAGGAAGTTTGAAGTTGCTATAATGATTATTTCTCTTG
...
>scaffold_4
ATACTGCCCATGCCTCATCTACTCCCTTCTTAACCTTTTCATCAACAACT
TGCTTAAGTTGGTCATCACTGACTGAGTGAGCCCAAACTGGAATAGCTGT
...
```

The resulting indexed binary file **output.2bit** can have data more easily extracted and can also be viewed in the UCSC Genome Browser.

Extracting a specific window location of data from a BAM and 2bit file

```
samtools view http://location_of/file.bam  "chr1:1499900-1500055" >
output.sam
```

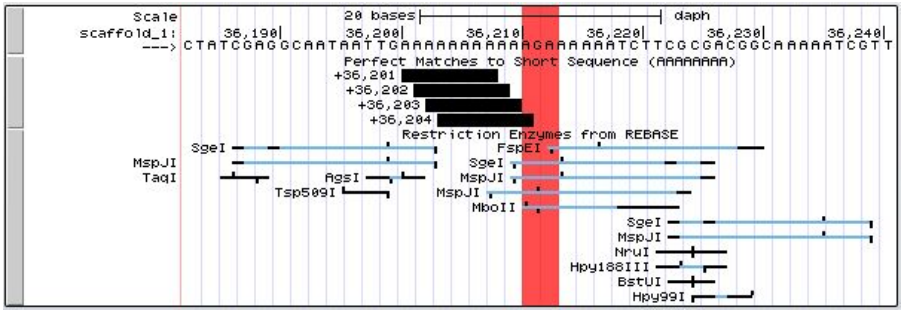



```
twoBitToFa -seq=chr1 -start=1499900 -end=1500055
http://yourGenome/your.2bit output.fa
```

```
>chr1:1499900-1500055
GCTACCATCACCCAAAAAGCTGAGGAGTTTGAATTCACTTCAGCACAACT
ATCATTAATTAATTTTTGAACCTCTGAGCCTGGAAGAGAAAACAGGTTTG
GTTCAACATGAAGAATACTGTGATTTGACCCGTGACAGAGCTTTCTGTTA
```

Viewing Data at UCSC

bigDataUrl *http://location_of/file.bam*

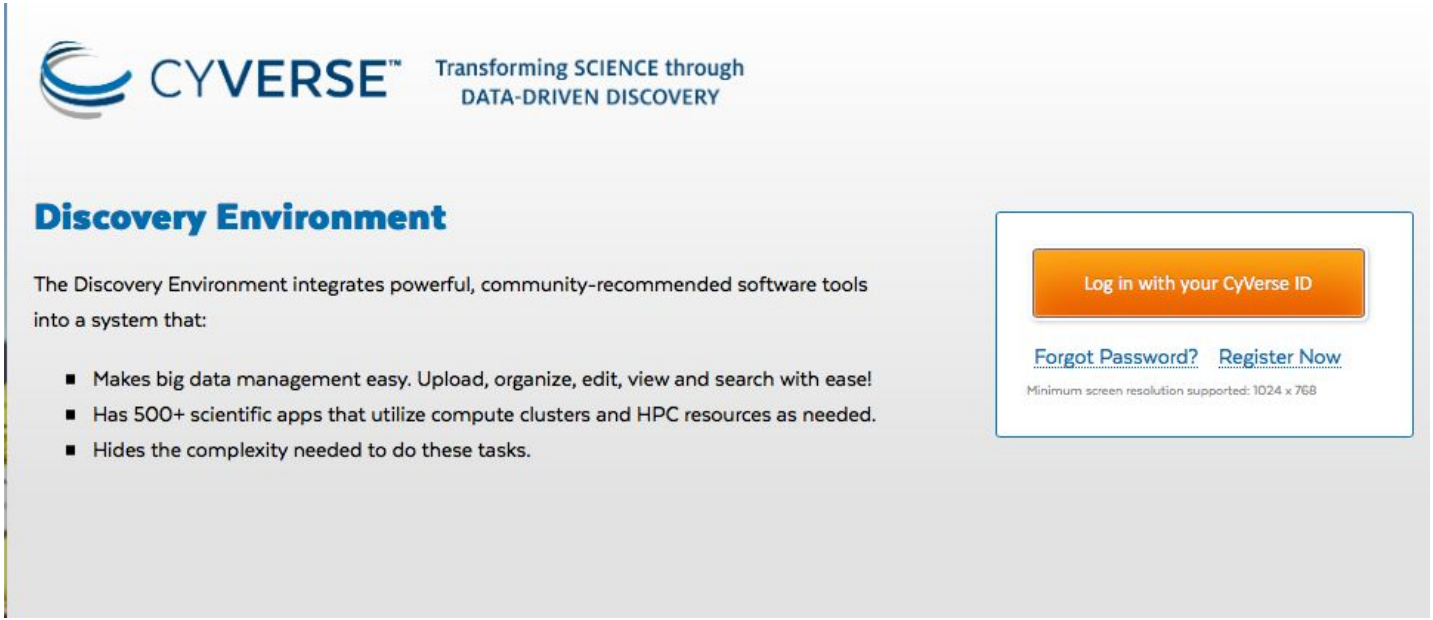


twoBitPath *http://yourGenome/your.2bit*

Hosting your data at CyVerse

CyVerse Storage Solution

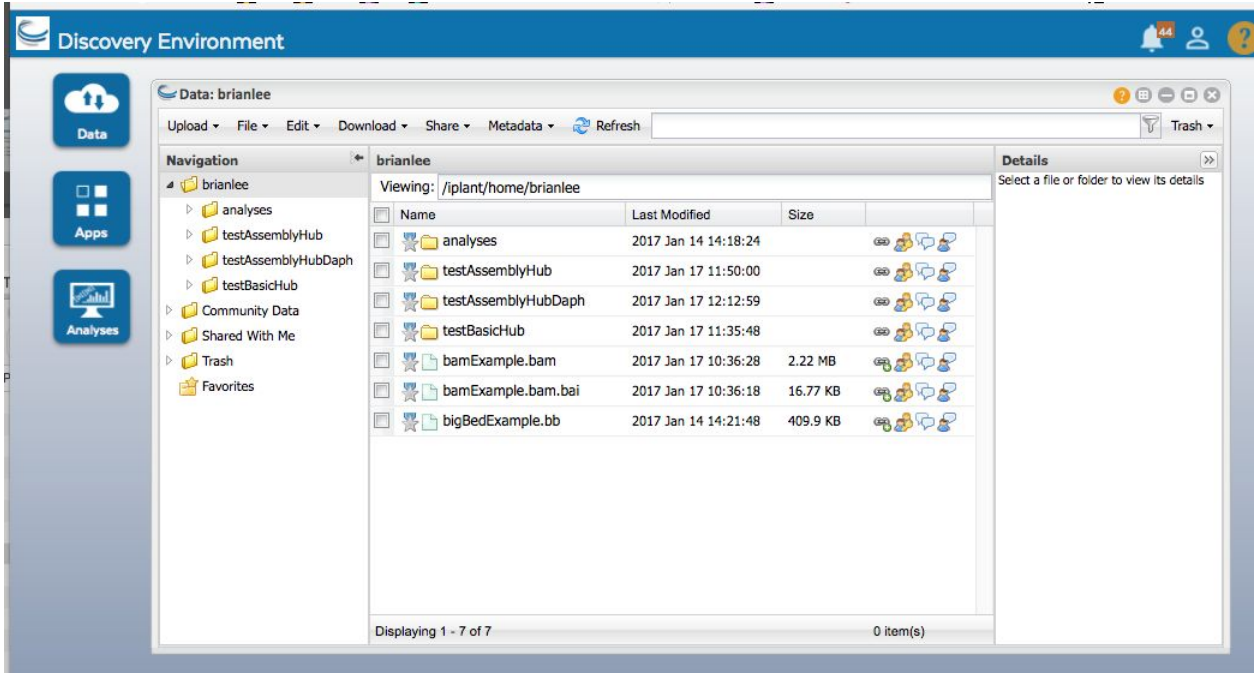
https://de.cyverse.org/ Discovery Environment



Hosting your data at CyVerse

CyVerse Storage Solution

https://de.cyverse.org/ Discovery Environment



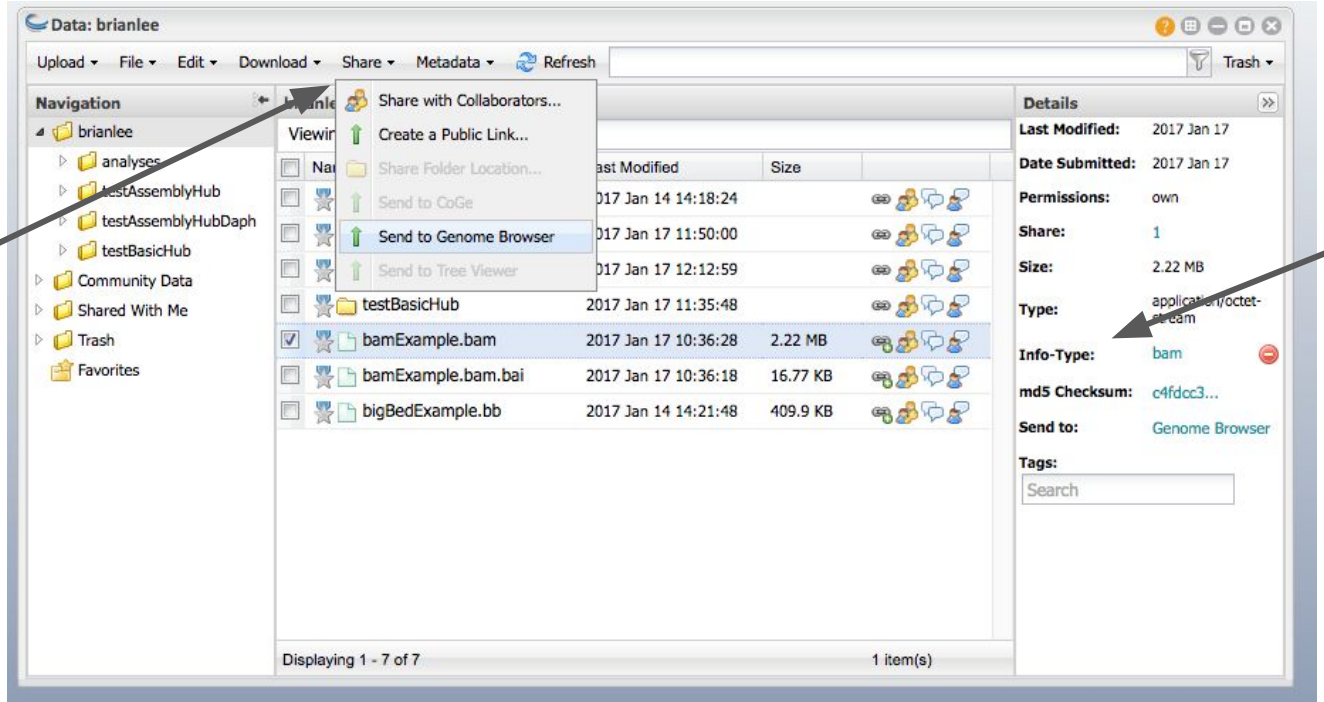
Hosting your data at CyVerse

CyVerse Storage Solution

Creating a Link that Accepts Byte-Ranges: “Send to Genome Browser”

Share Menu:

Send to Genome Browser



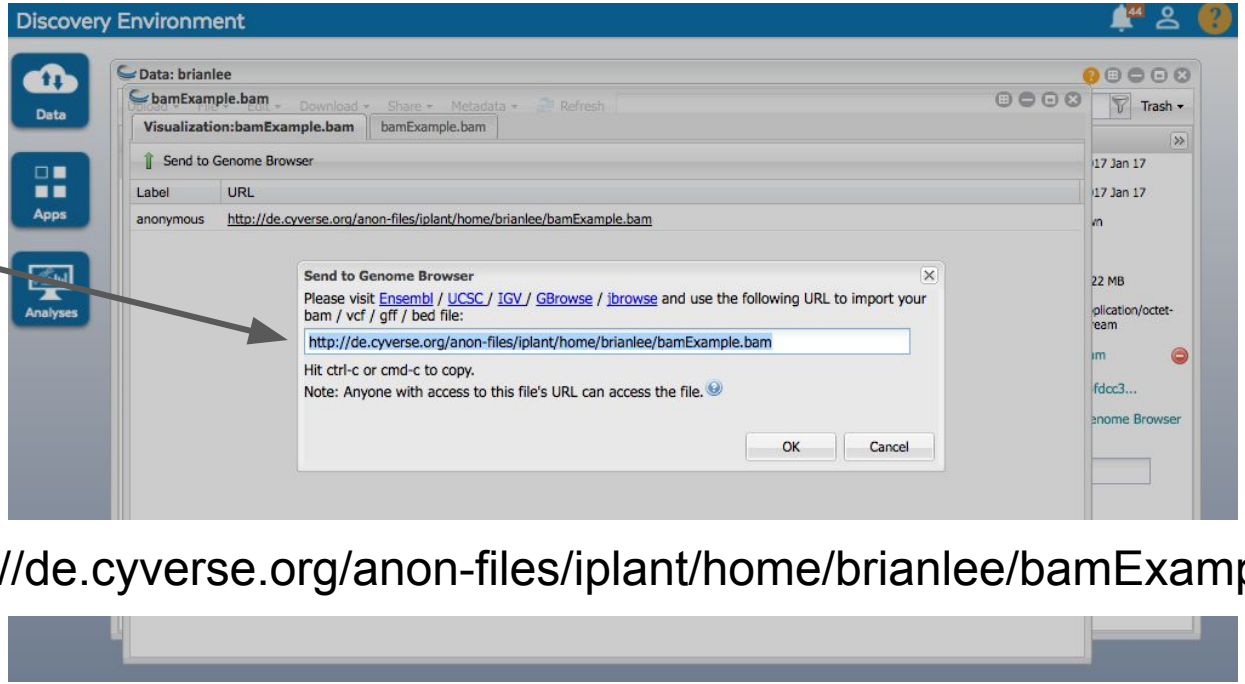
Info-Type: bam

Hosting your data at CyVerse

CyVerse Storage Solution

Creating a Link that Accepts Byte-Ranges: **“Send to Genome Browser”**

Results in a link you can use in later visualizaiton: ***bigDataUrl***



http://de.cyverse.org/anon-files/iplant/home/brianlee/bamExample.bam

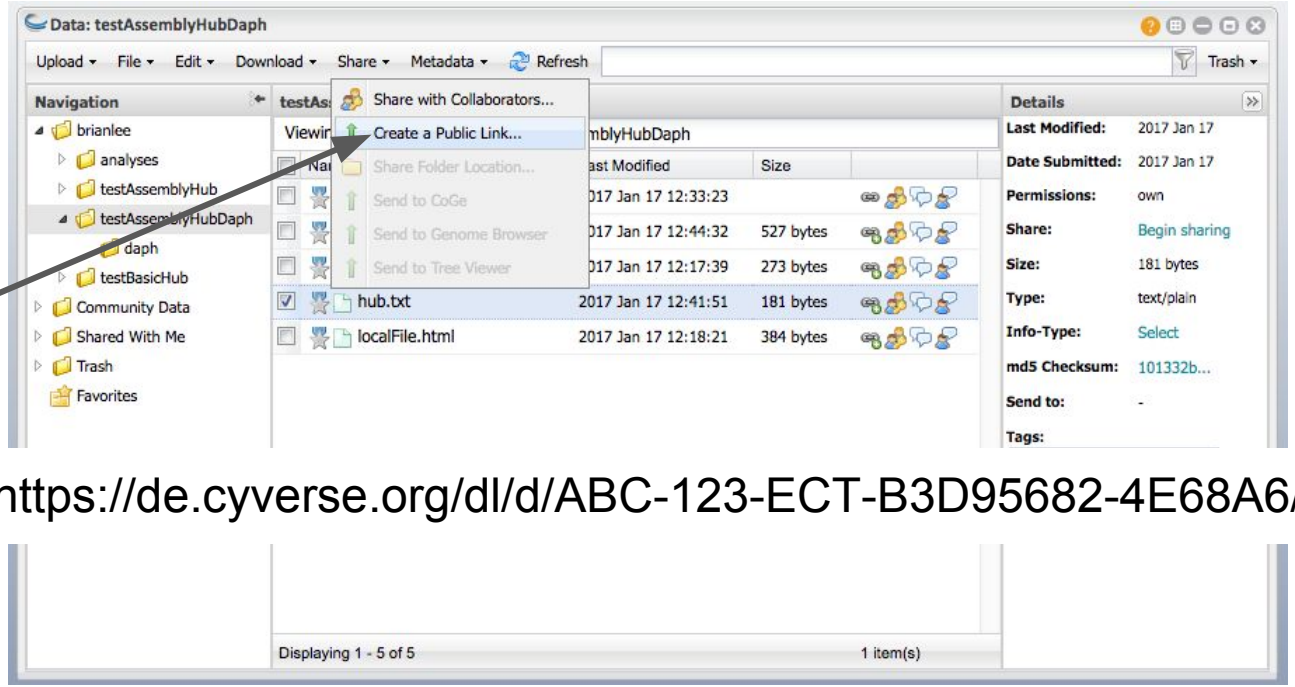
Hosting your data at CyVerse

CyVerse Storage Solution

NOTE: The "Create a Public Link" is not the same as "Send to Genome Browser"

The **"Create a Public Link"** option will work for static interactions, like downloading text files.

It will not work for data byte-range requests needed for visualization.



https://de.cyverse.org/dl/d/ABC-123-ECT-B3D95682-4E68A6/fileName

Viewing the CyVerse hosted 2bit at UCSC

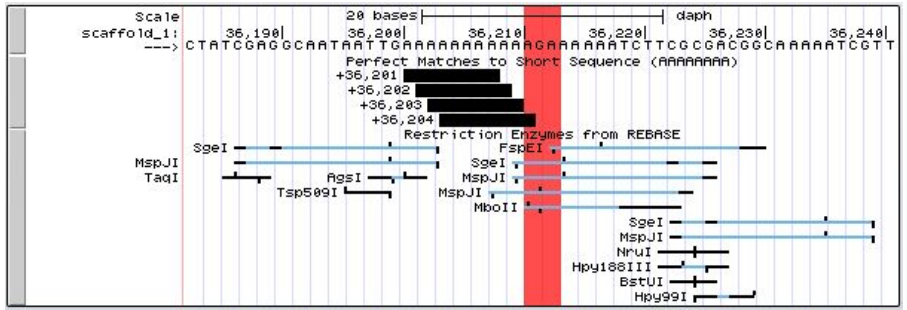
```
twoBitToFa -seq=chr1 -start=1499900 -end=1500055
https://de.cyverse.org/anon-files/iplant/home/your.2bit output.fa
```

```
>chr1:1499900-1500055
GCTACCATCACCCAAAAAGCTGAGGAGTTTGAATTCACTTCAGCACAACT
ATCATTAATTAATTTTTGAACCTCTGAGCCTGGAAGAGAAAACAGGTTTG
GTTCAACATGAAGAATACTGTGATTTGACCCGTGACAGAGCTTTCTGTTA
```

Viewing the CyVerse hosted 2bit at UCSC

twoBitPath

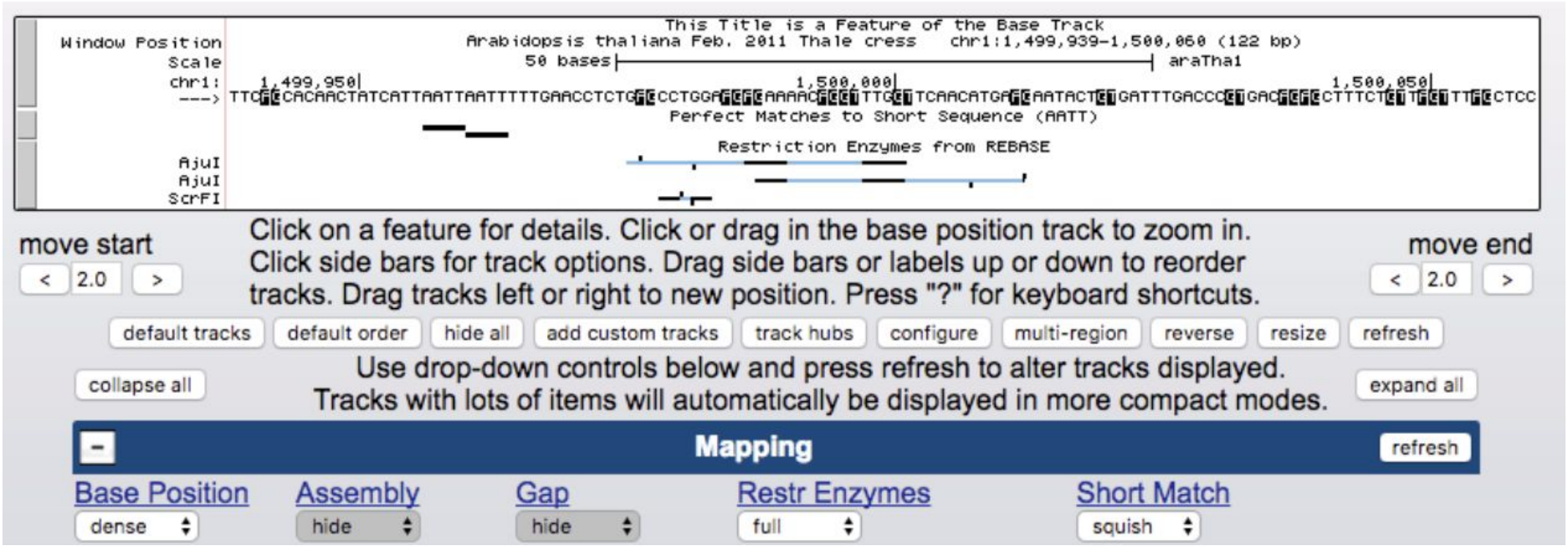
https://de.cyverse.org/anon-files/iplant/home/your.2bit

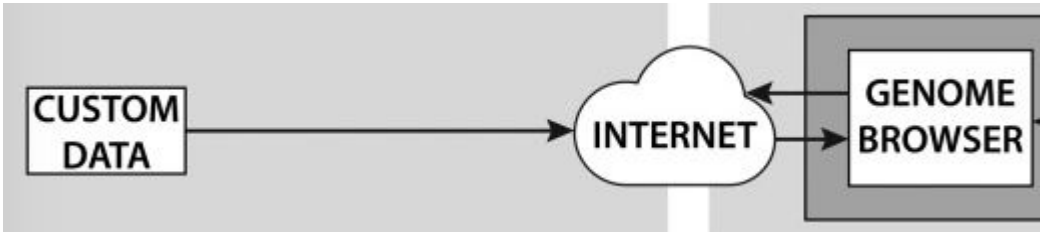


Viewing the CyVerse hosted 2bit at UCSC

Browser requests *AGCTs* for only the window currently viewing

(chr1:1,499,900-1500,055)





twoBitPath http://CyVerse_location_of/output.2bit

genomes.txt shares where to find the 2bit (and what to call new genome)

genome *yourGenome*
trackDb *http://location_of/trackDb.txt*
scientificName *Your Genome*
description *Feb. 2017 Assembly*
organism *Your organism*
defaultPos *chr1:1000000-2000000*
groups *groups.txt*
htmlPath *http://yourGenome/description.html*
twoBitPath *http://CyVerse_location_of/output.2bit*

Window of Sequence viewed at UCSC

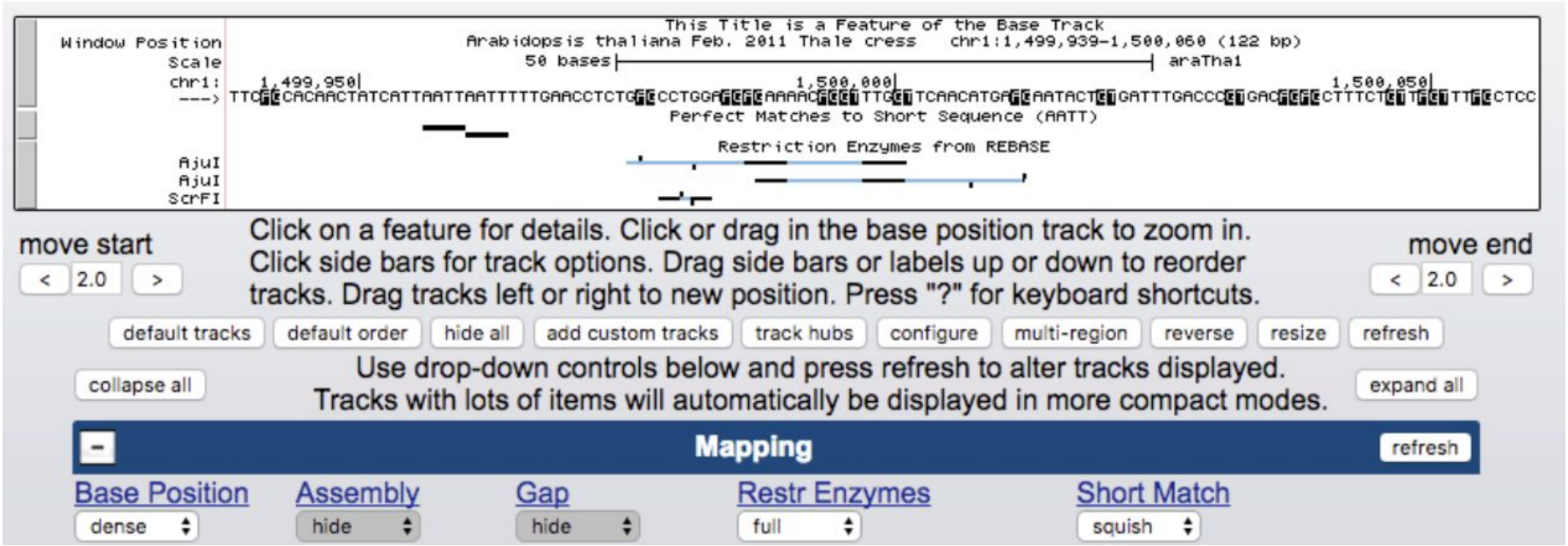
Three tracks are generated from the 2bit on the fly:

- Base Position
- Restriction Enzymes
- Short Match Track (30 bases)

Viewing the CyVerse hosted 2bit at UCSC

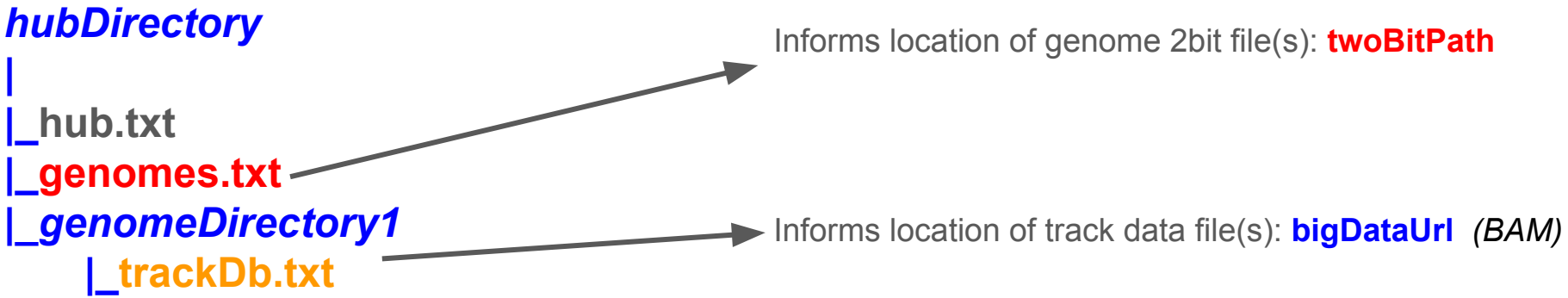
Browser requests *AGCTs* for only the window currently viewing

(chr1:1,499,900-1500,055)



Viewing the CyVerse hosted 2bit at UCSC

Track Hubs are text files (**hub.txt, genomes.txt, trackDb.txt**) that describe and point to the location of binary indexed data files (**twoBitPath, bigDataUrl**)



Viewing the CyVerse hosted 2bit at UCSC

Track Hubs are text files (**hub.txt, genomes.txt, trackDb.txt**) that describe and point to the location of binary indexed data files (**twoBitPath, bigDataUrl**)

hub.txt

hub *MyHubsNameWithoutSpaces*
shortLabel *My Hub's Name*
longLabel *Longer label about my hub.*
email *myEmail@address*
genomesFile *http://location_of/genomes.txt*

genomes.txt

genome *yourGenome*
trackDb *http://location_of/trackDb.txt*
twoBitPath *http://location_of/output.2bit*
...

trackDb.txt

track *bam1*
type *bam*
shortLabel *BAM example*
longLabel *This BAM file is a RNA-seq example from the ENCODE project*
visibility *dense*
bigDataUrl *http://location_of/file.bam*
...

track *bam2*
...

track *bam3*
...

Each new track gets a new stanza
bigDataUrl lines point to location of binary indexed data
More trackDb parameters can be defined too (color, ect.)

Viewing the CyVerse hosted 2bit at UCSC

Host all data at CyVerse

Binary indexed files:

twoBitPath,

bigDataUrl

Text files: **hub.txt,**

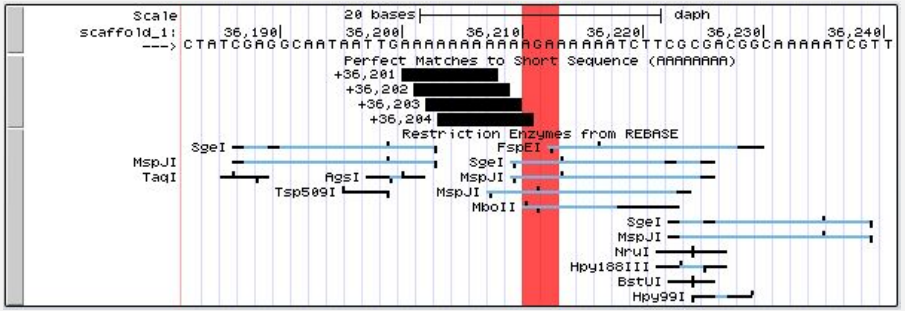
genomes.txt,

trackDb.txt

```
>scaffold_1
GTTGTAAATACTCTATTCTACAATAAAACCAA
TCATAGGTTGAATTGGCGTTGAAGTAAAACAAA
...
>scaffold_2
AGTTATGACAAACTATAAAAAGTCGGTAGAGACAAAAG
TCGTTCGTGGACGAAGCGACCAAAACTGAGCACAAGATA
...
>scaffold_3
CATAAATTCATAAATCAATTCATGAAGAATAATT
TAGAAAATTTCCCAGGAAGTTTGAAGTTGCTA
...
```

```
hubDirectory
|__hub.txt
|__genomes.txt
|__hg19
      |__trackDb.txt
```

Visualize at UCSC



Viewing the CyVerse hosted 2bit at UCSC

Easily move all data with CyVerse `iCommands` (`rsync` becomes `irysnc`) that allows transfer of 2-100GB files and recursive hub directory structures:

```
$ irsync -r local_directory_name i:data_store_directory_name
```

hubDirectory
|
|_hub.txt
|_genomes.txt
|_*genomeDirectory1*
|_trackDb.txt

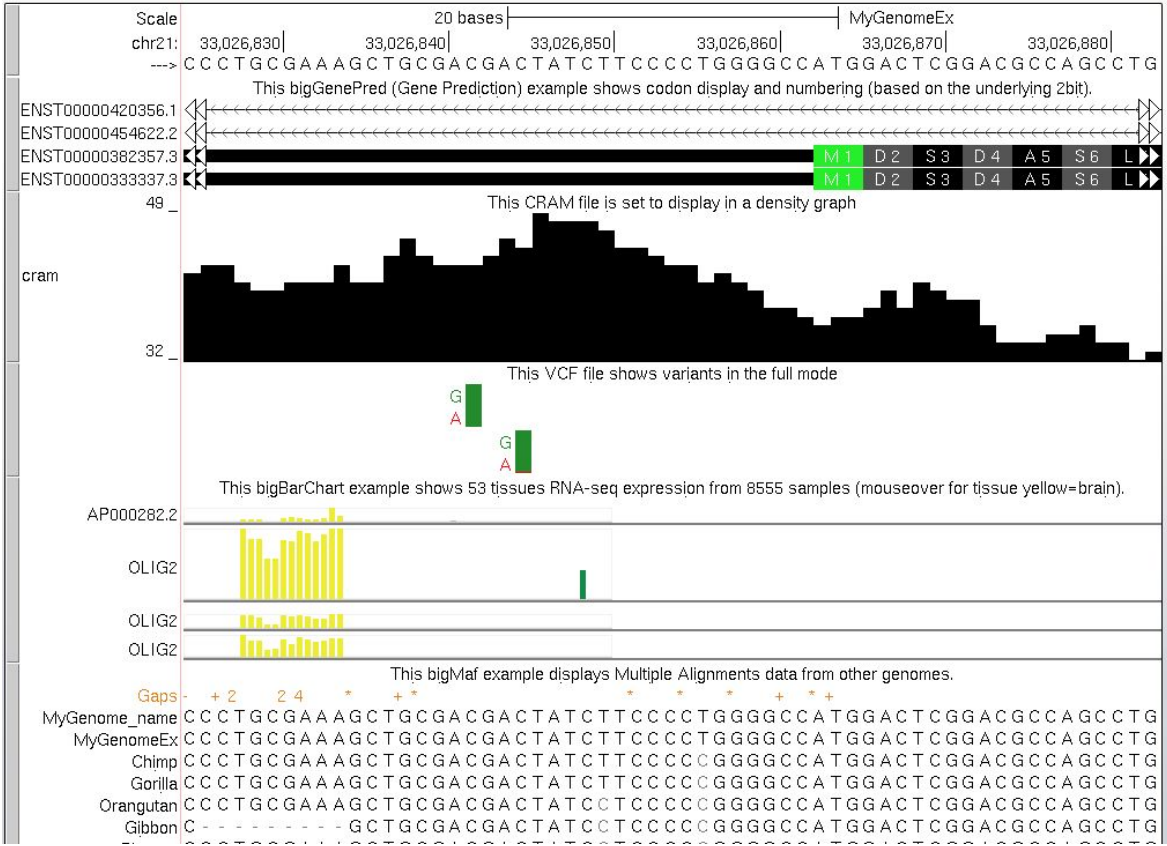
Move all your local BAMS, VCFs, and related Assembly Hub files with the single **`irsync`** command.

Then make the files publicly accessible with the **`ichmod`** command.

https://de.cyverse.org/anon-files/

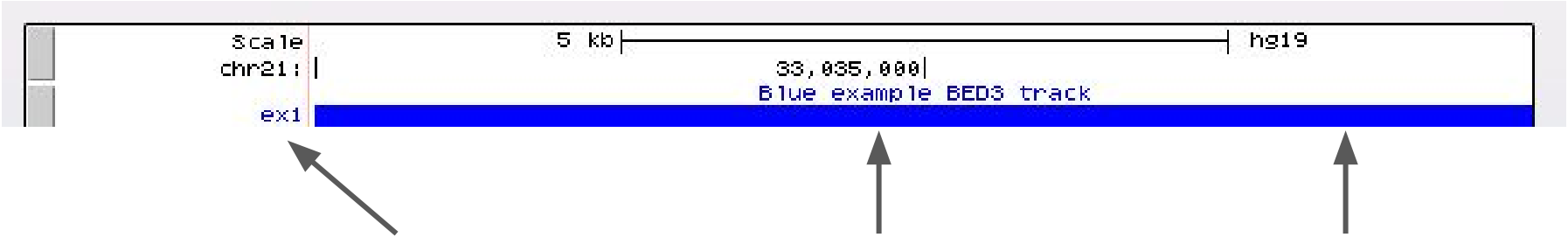
```
$ ichmod read anonymous data_store_directory_name
```

Additional Track Types



Additional Track Types

simple as adding text-based custom tracks



track name=ex1 description="Blue example BED3 track" color=0,0,255

chr21 33030000 33040000

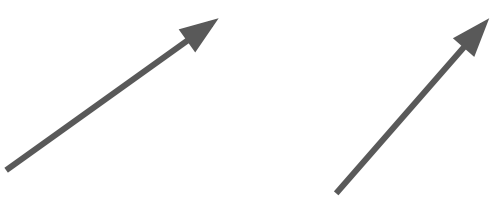
chr21 33050000 33060000

...

Additional Track Types

create binary indexed versions of text files (`bedToBigBed`)

```
bedToBigBed  in.bed  chrom.sizes out.bb
```

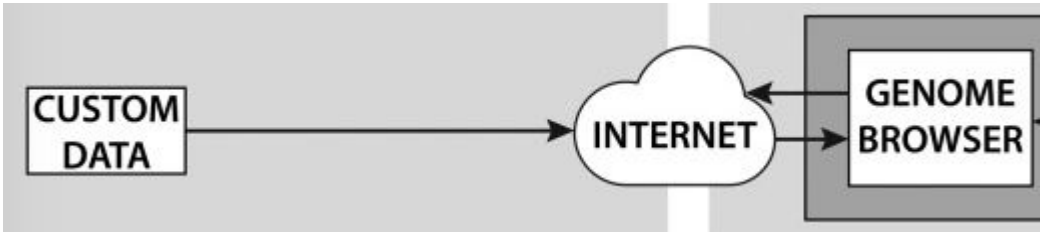


chr21 33030000 33040000
chr21 33050000 33060000
…
…
…
…

chr1 248956422
…
chr21 46709983
chr22 50818468
…

The resulting binary file **out.bb** (bigBed) can be hosted online:

*bigDataUrl=http://CyVerse.path/**out.bb***



bigDataUrl http://CyVerse.path/to/out.bb

The **trackDb.txt** shares descriptions and where to find binary indexed tracks

View Results at UCSC

track *ex1bb*
type *bigBed*
shortLabel *ex1*
longLabel *Blue example BED3 track*
color *0,0,255*
bigDataUrl *http://CyVerse.path/to/out.bb*

track *track2*
bigDataUrl *http://CyVerse.path/to/out2.bb*
...

Additional Track Types

Other binary indexed formats visualizable at UCSC:

- BAM, CRAM, VCF
- bigGenePred, bigBarChart
- bigPsl, bigChain, bigMaf,
- bigNarrowPeak, halSnake
- bigBed, bigWig

Additional Track Types

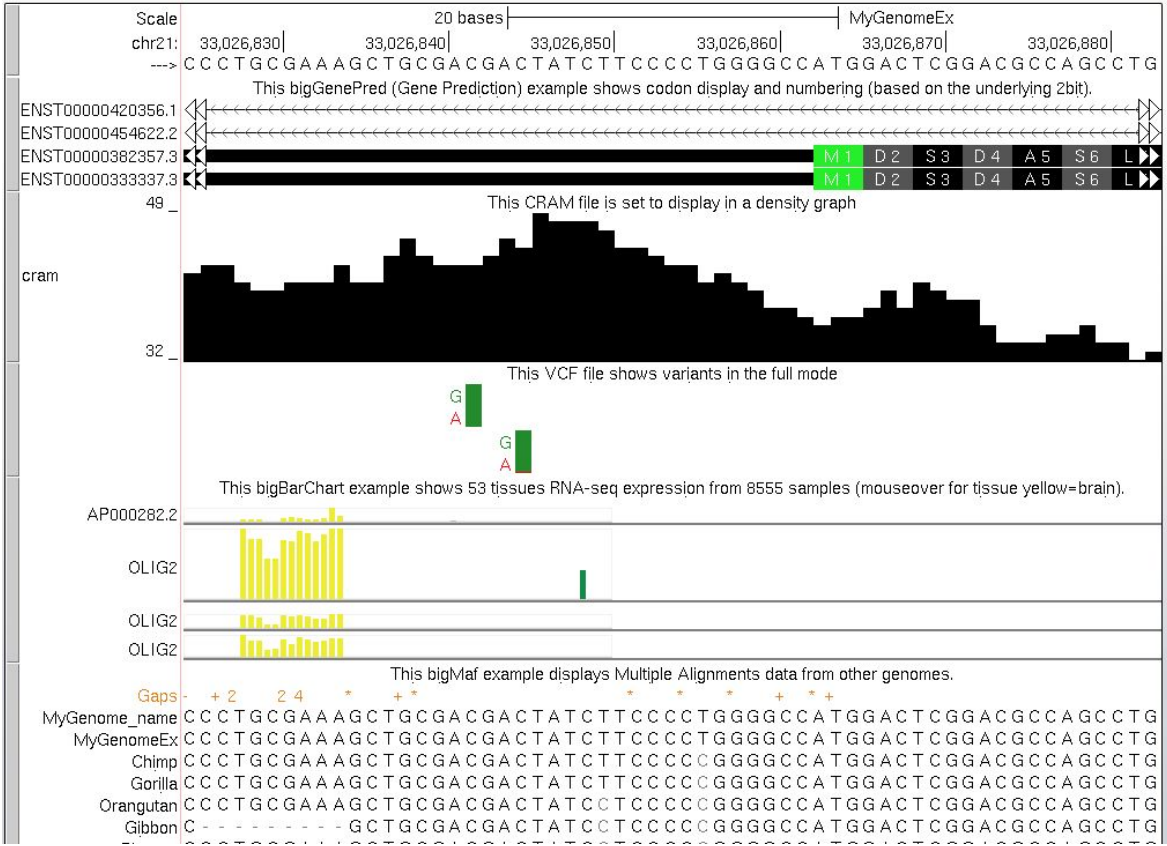
bigGenePred

CRAM

VCF

bigBarChart

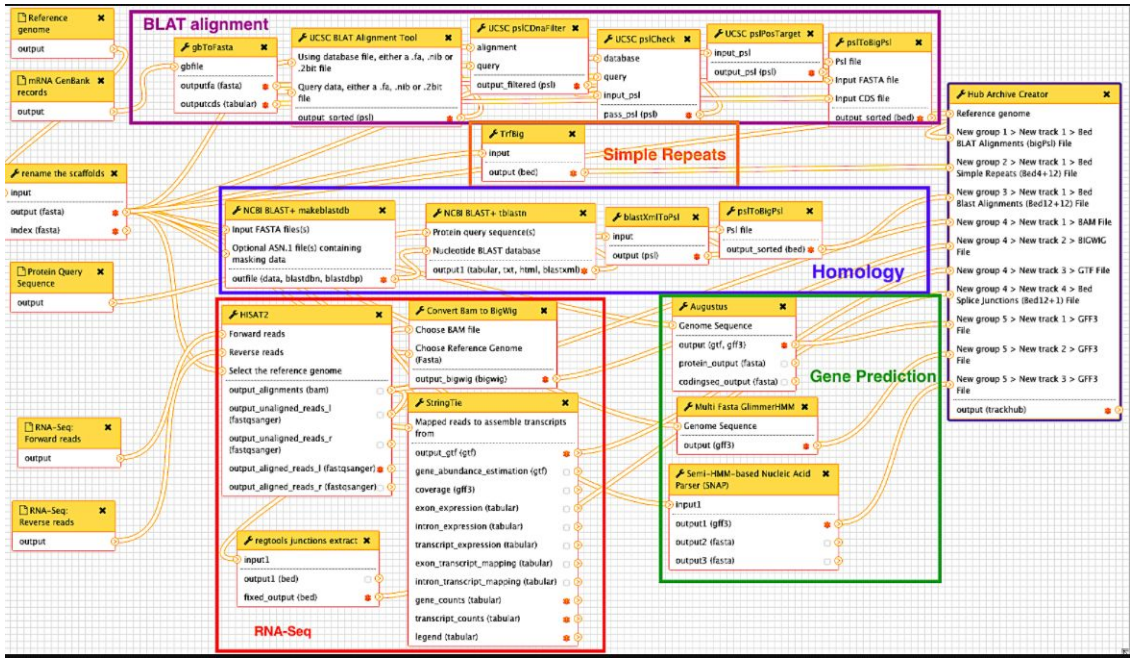
bigMaf



Assembly Track Hub Resources

G-OnRamp Galaxy workflow turning data like RNA-Seq into Assembly Hubs

http://gonramp.wustl.edu/



Thank You!