UCSC Genome Browser

Acknowledgements

funding:

National Human Genome Research Institute (NHGRI)

California Institute for Regenerative Medicine (CIRM)

QB3 (UCBerkeley, UCSF, UCSC)

Chan Zuckerberg Initiative

Howard Hughes Medical Institute

Viewing CyVerse Hosted Data at UCSC

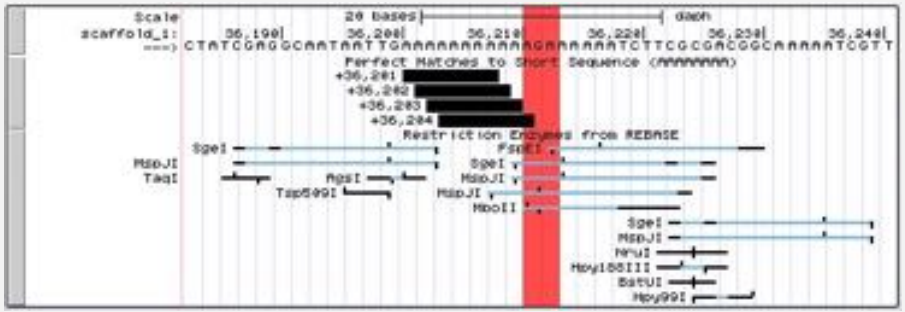
Host data at CyVerse

Binary indexed files:
twoBitPath, bigDataUrl

Text files: ***hub.txt***

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

Visualize at UCSC

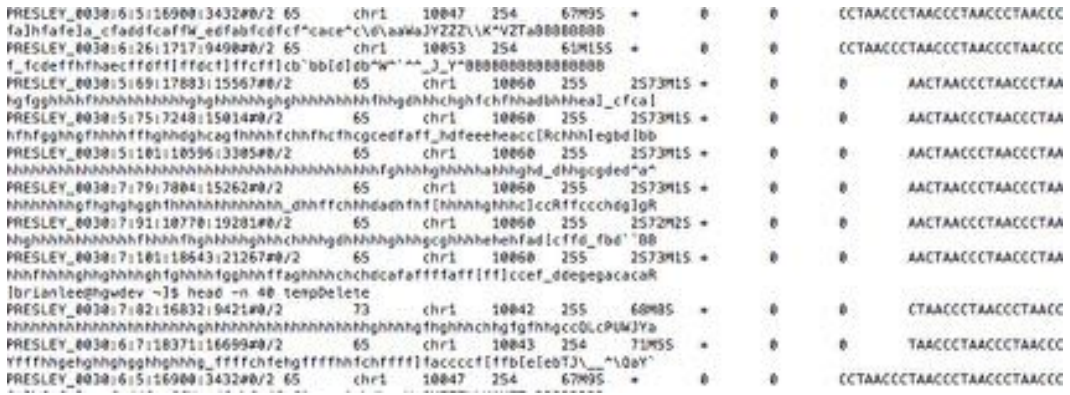


CyVerse's *"Send to Genome Browser"* option creates links accepting byte-range requests for binary index data.

Examples of Visualizing binary indexed data in the UCSC Genome Browser

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

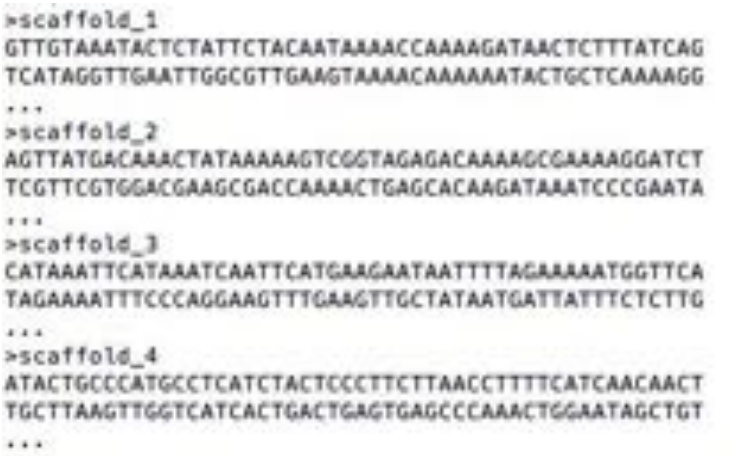
```
samtools view -S -b sample.sam > sample.bam
```



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)

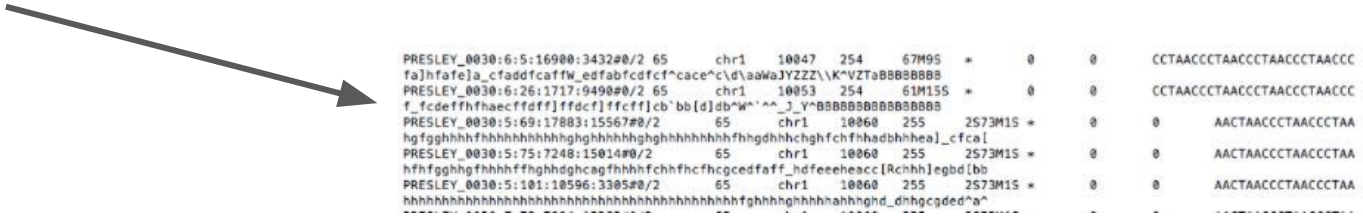
```
faToTwoBit input.fasta output.2bit
```



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
```

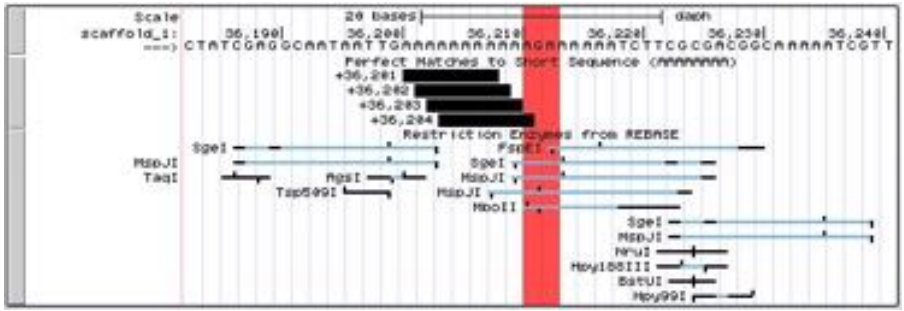



```
twoBitToFa -seq=chr1 -start=1499900 -end=1500055
http://yourGenome/your.2bit output.fa
```

```
>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

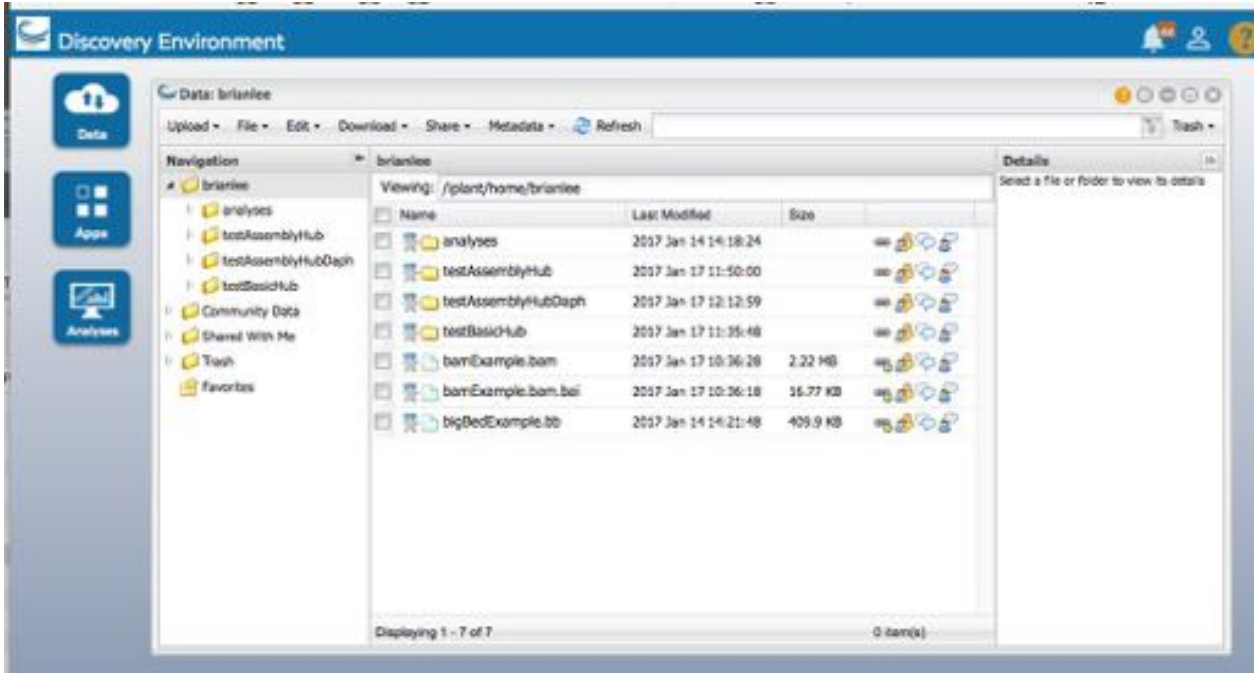
CyVerse Storage Solution

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

Hosting your data at CyVerse

CyVerse Storage Solution

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



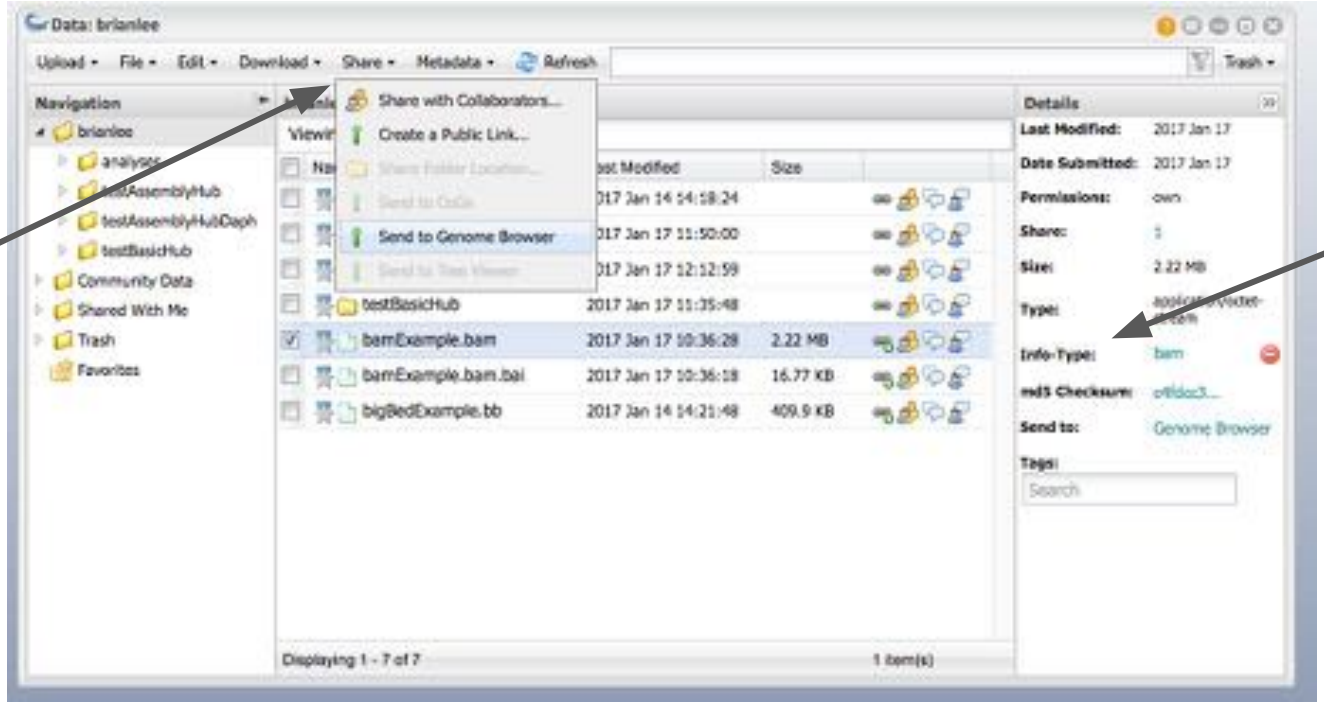
Hosting your data at CyVerse

CyVerse Storage Solution

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

Share Menu:

Send to Genome Browser



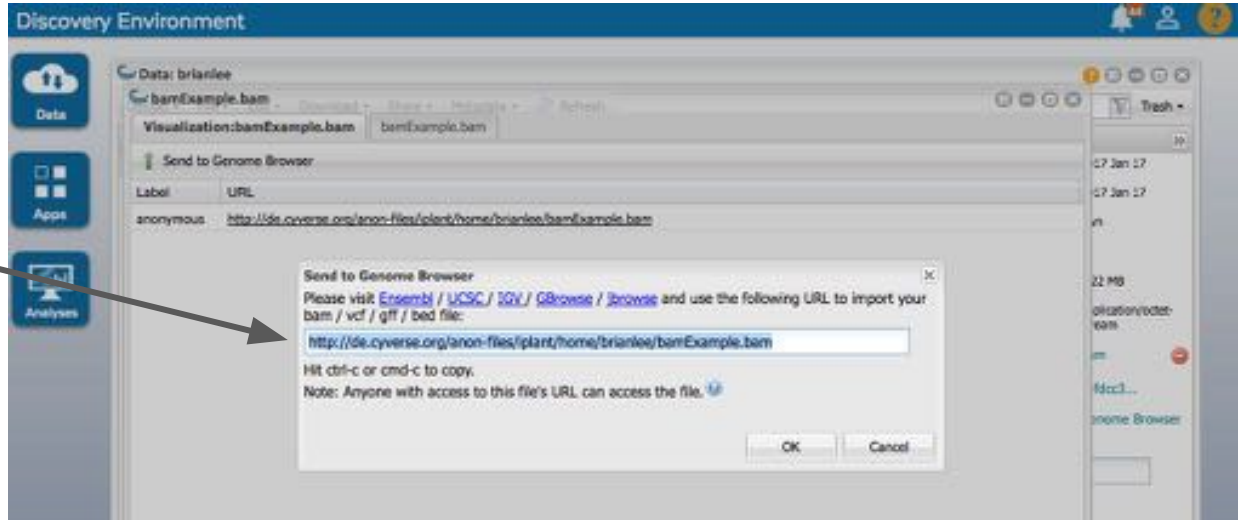
Info-Type: bam

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 visualization: ***bigDataUrl***



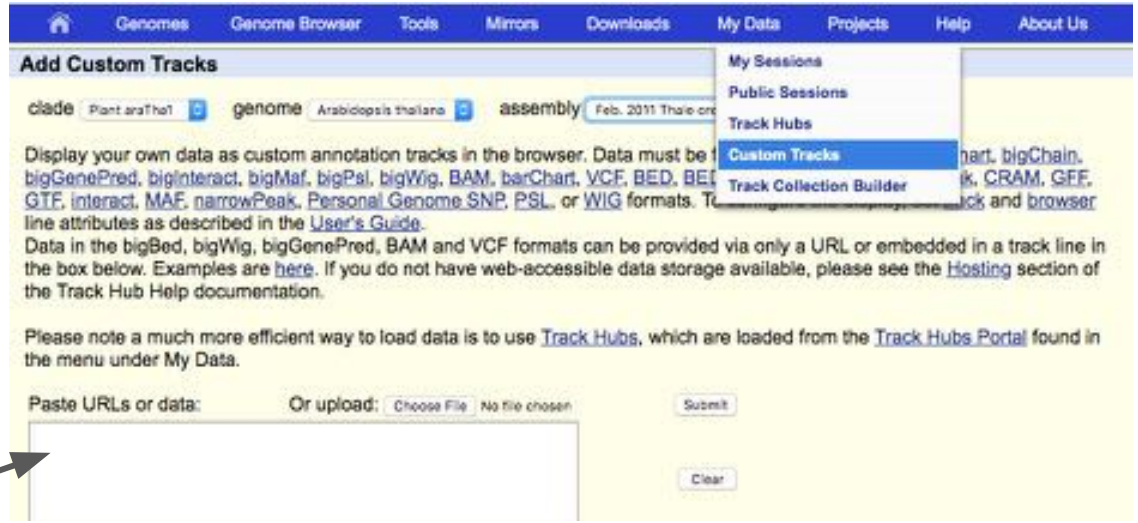
https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/Bam_Ex1/DNase_example.bam

Hosting your data at CyVerse

CyVerse Storage Solution

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

Paste the link into the BAM as Custom Track on any Assembly or Assembly Hub



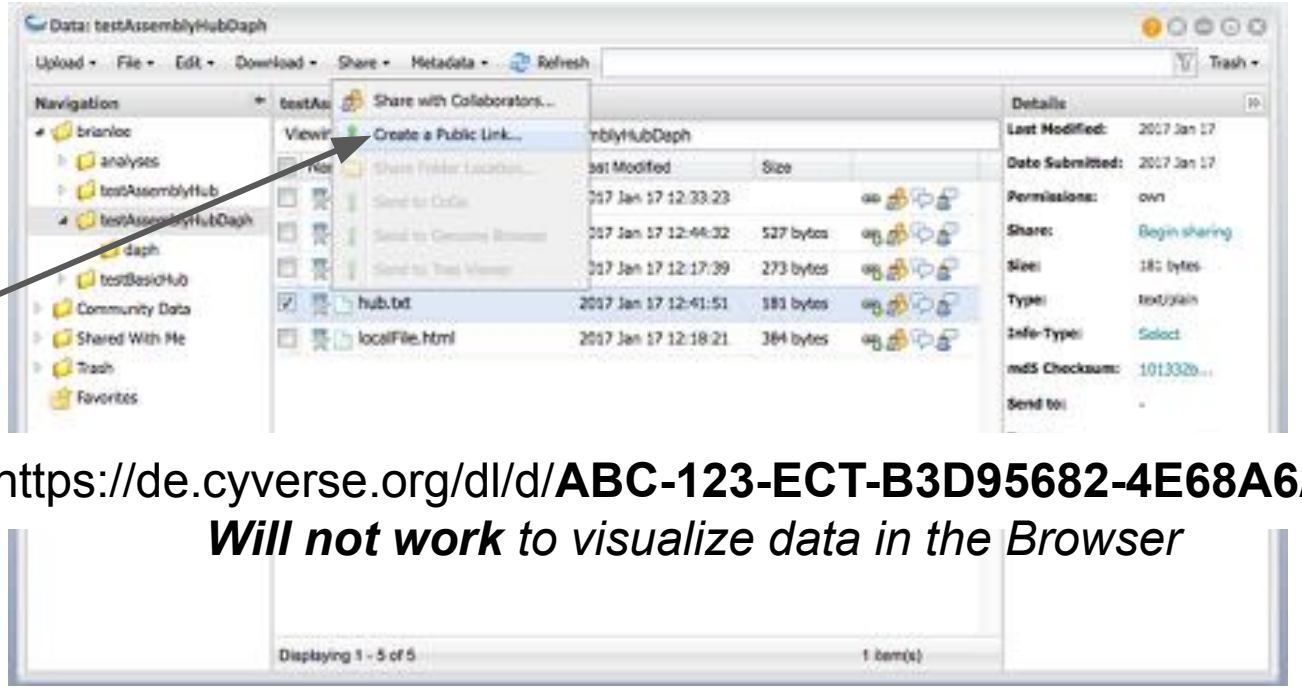
https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/Bam_Ex1/DNase_example.bam

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/data 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

***Will not work** to visualize data in the Browser*

Viewing the CyVerse hosted 2bit at UCSC

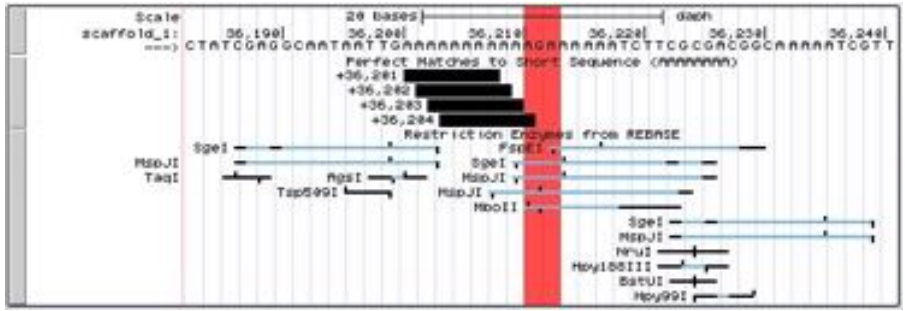
```
twoBitToFa -seq=chr1 -start=1499900 -end=1500055
https://data.cyverse.org/dav-anon/iplant/home/your.2bit output.fa
```

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

Viewing the CyVerse hosted 2bit at UCSC

twoBitPath

https://data.cyverse.org/dav-anon/iplant/home/your.2bit



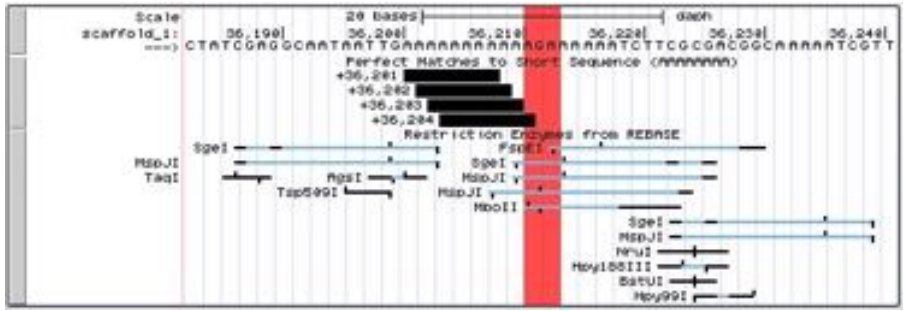
Viewing the CyVerse hosted 2bit at UCSC

hub plantAraTha1
useOneFile on
shortLabel Plant araTha1
longLabel Plant araTha1 Hub
email contact@email.com

genome araTha1
description Feb. 2011 Thale cress
twoBitPath araTha1.2bit
organism Arabidopsis thaliana
defaultPos chr1:1000000-2000000
scientificName Arabidopsis thaliana

track cytoBandIdeo
longLabel Chromosome ideogram
shortLabel cytoBandIdeo
bigDataUrl cytoBandIdeo.bigBed
type bigBed

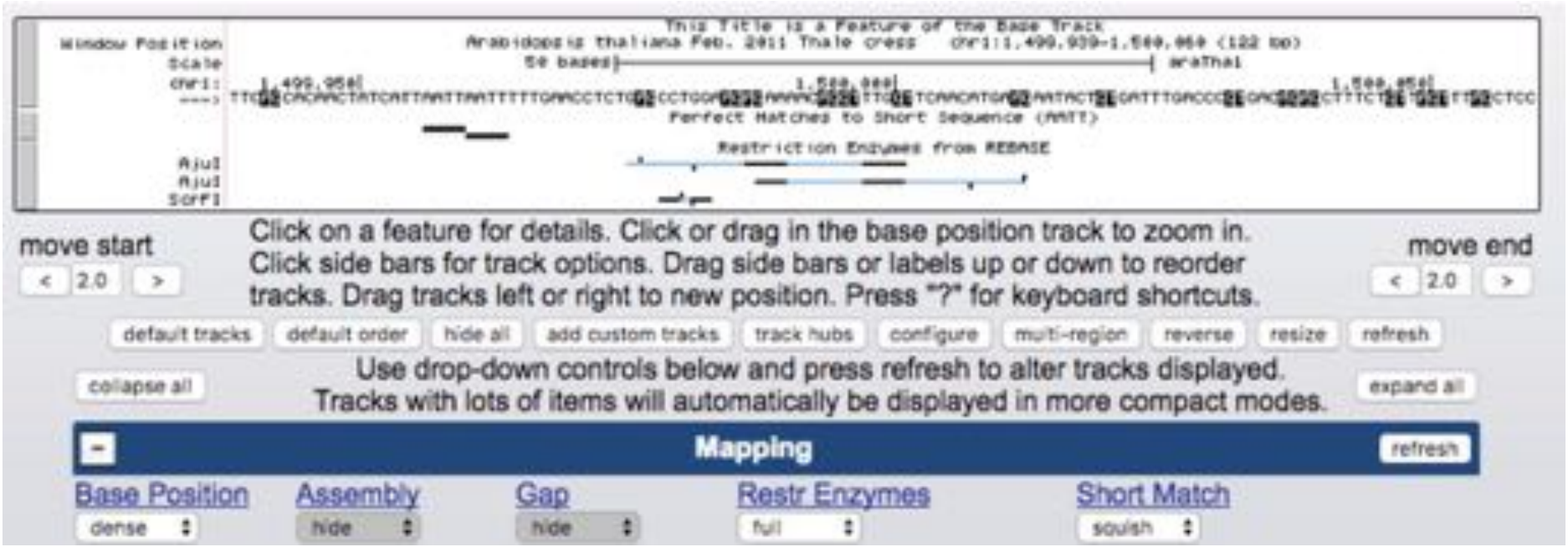
hub.txt



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

hub plantAraTha1
useOneFile on
shortLabel Plant araTha1
longLabel Plant araTha1 Hub
email contact@email.com

The hub stanza sets **useOneFile on**, limiting hub to only one genome

genome araTha1
description Feb. 2011 Thale cress
twoBitPath araTha1.2bit
organism Arabidopsis thaliana
defaultPos chr1:1000000-2000000
scientificName Arabidopsis thaliana

track cytoBandIdeo
longLabel Chromosome ideogram with cytogenetic bands
shortLabel cytoBandIdeo
bigDataUrl cytoBandIdeo.bigBed
type bigBed

Viewing the CyVerse hosted 2bit at UCSC

hub plantAraTha1
useOneFile on
shortLabel Plant araTha1
longLabel Plant araTha1 Hub
email contact@email.com

genome araTha1
description Feb. 2011 Thale cress
twoBitPath araTha1.2bit
organism Arabidopsis thaliana
defaultPos chr1:1000000-2000000
scientificName Arabidopsis thaliana

The genome stanza shares where to find the 2bit (and what to call the new genome)

track cytoBandIdeo
longLabel Chromosome ideogram with cytogenetic bands
shortLabel cytoBandIdeo
bigDataUrl cytoBandIdeo.bigBed
type bigBed

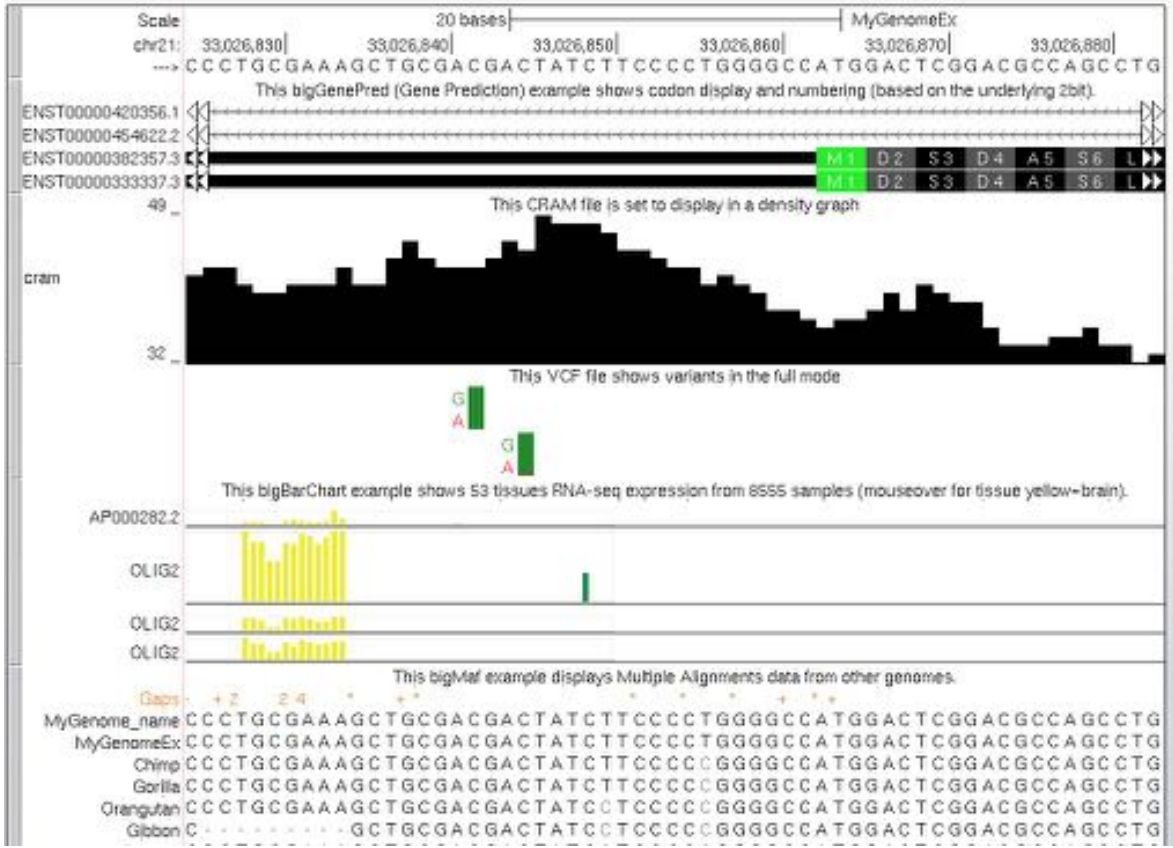
Viewing the CyVerse hosted 2bit at UCSC

```
hub plantAraTha1
useOneFile on
shortLabel Plant araTha1
longLabel Plant araTha1 Hub
email contact@email.com

genome araTha1
description Feb. 2011 Thale cress
twoBitPath araTha1.2bit
organism Arabidopsis thaliana
defaultPos chr1:1000000-2000000
scientificName Arabidopsis thaliana

track cytoBandIdeo
longLabel Chromosome ideogram with cytogenetic bands
shortLabel cytoBandIdeo
bigDataUrl cytoBandIdeo.bigBed
type bigBed
```

The multiple track stanzas share where to find the annotation tracks and what to name them and the data type (BAM, CRAM, VCF, bigGenePred, bigBarChart, bigPsl, bigChain, bigMaf, bigNarrowPeak,bigWig, bigBed, others).

The bigDataUrl must point to an online location of the binary-indexed data that can accept byte-range requests.

Additional Track Types



Viewing the CyVerse hosted 2bit at UCSC

Host all data at CyVerse

Binary indexed files:

twoBitPath,

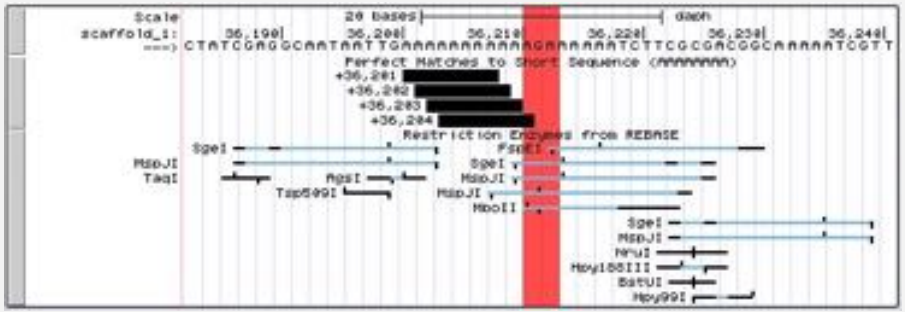
bigDataUrl

Text files: **hub.txt**

```
>scaffold_1
GTTGTAAATACTCTATTCTACAATAAAACCAA
TCATAGGTTGAATTGGCGTTGAAGTAAAACAAA
...
>scaffold_2
AGTTATGACAAACTATAAAAAGTCGGTAGAGACAAAAG
TCGTTCGTGGACGAAGCGACCAAAACTGAGCACAAGAT
...
>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 any recursive directory structures:

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

hubDirectory
|
|_hub.txt
|_genomeFile.2bit
|_trackAnnotation.bigBed

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

https://data.cyverse.org/dav-anon/

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

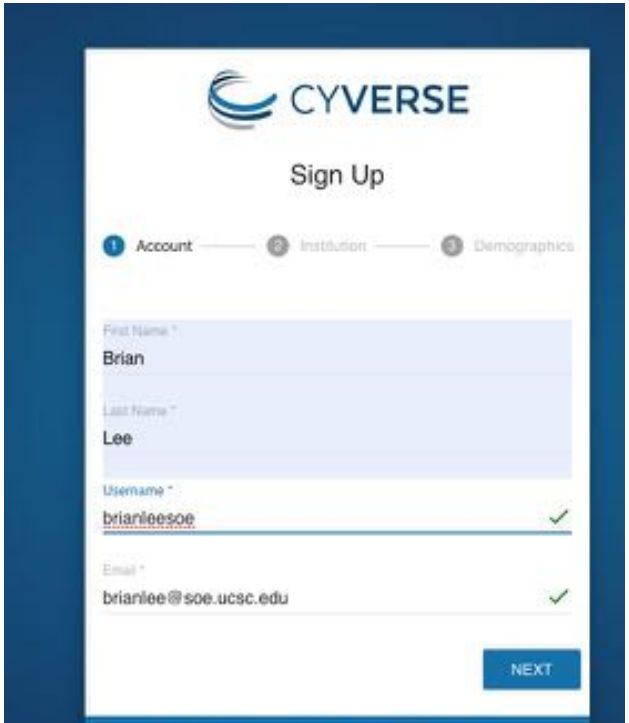
```
$ ichmod read anonymous data_store_directory_name
```

Hosting your data at CyVerse

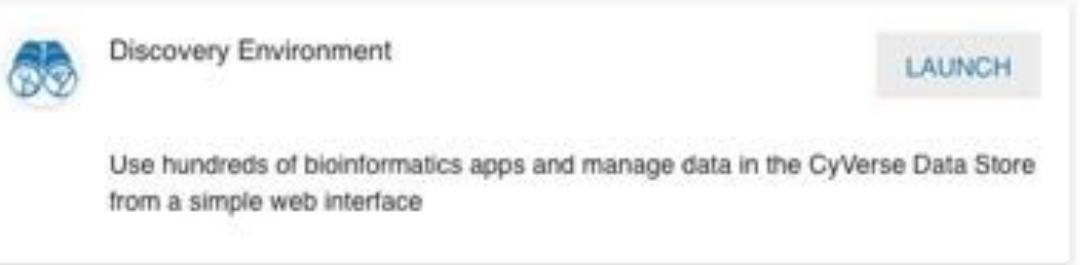
CYVERSE

About Products Learning Events News Launch Log In | Register

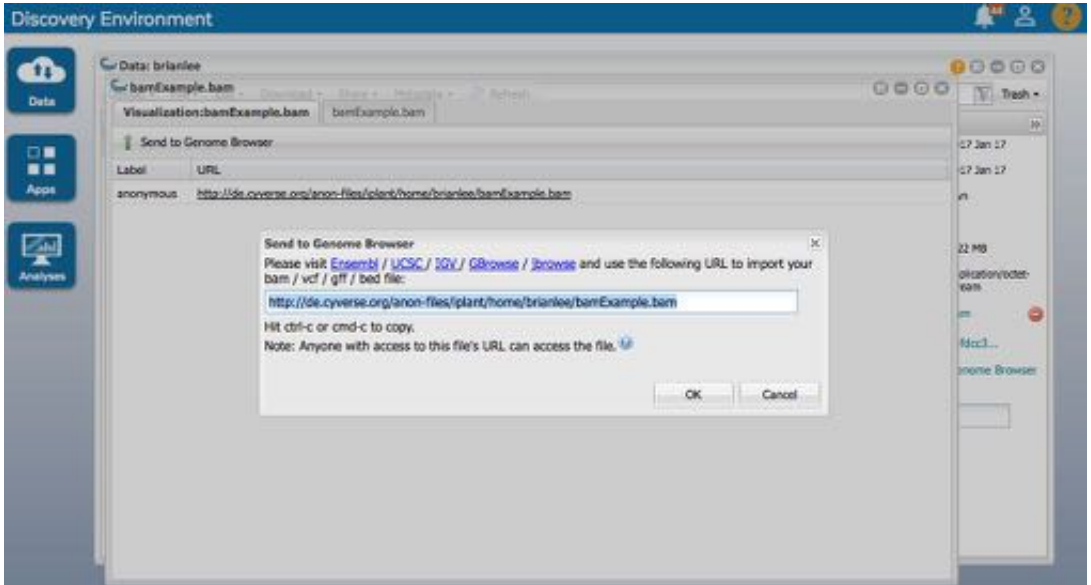
Sign-up for a CyVerse Account



Hosting your data at CyVerse

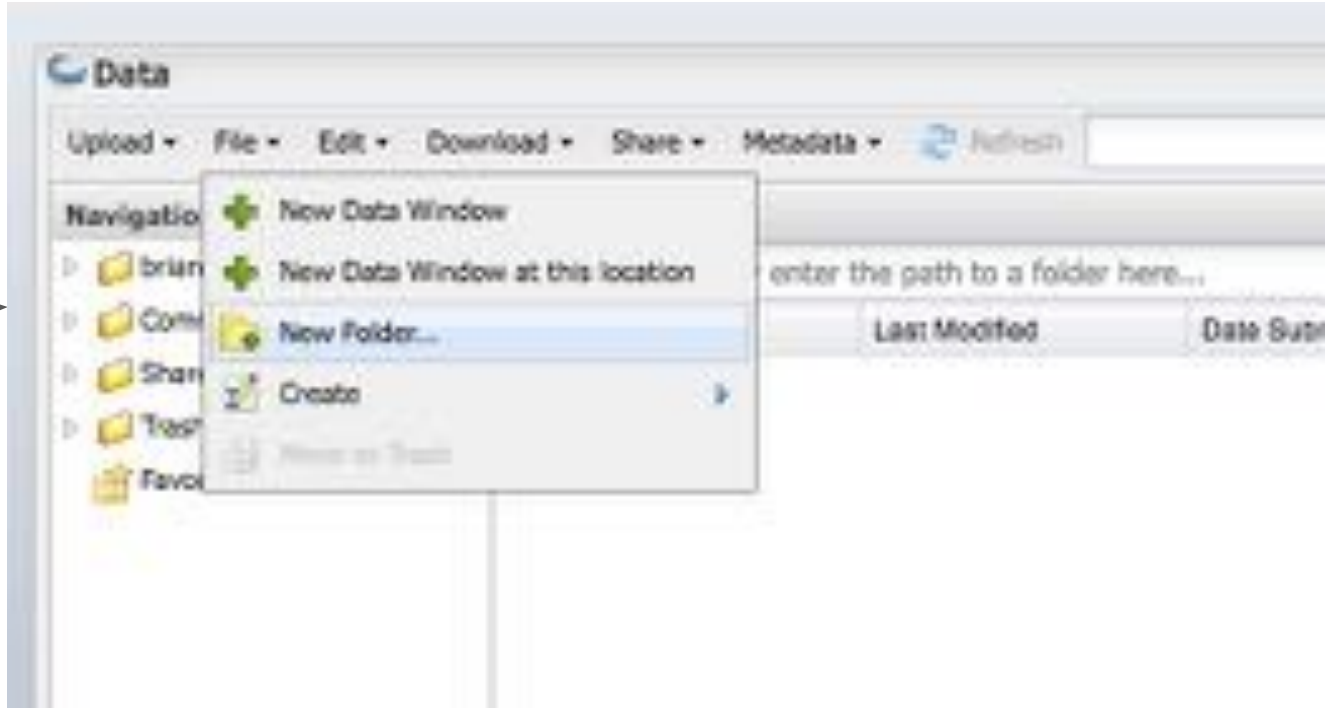


Launch the Discovery Environment and click Data



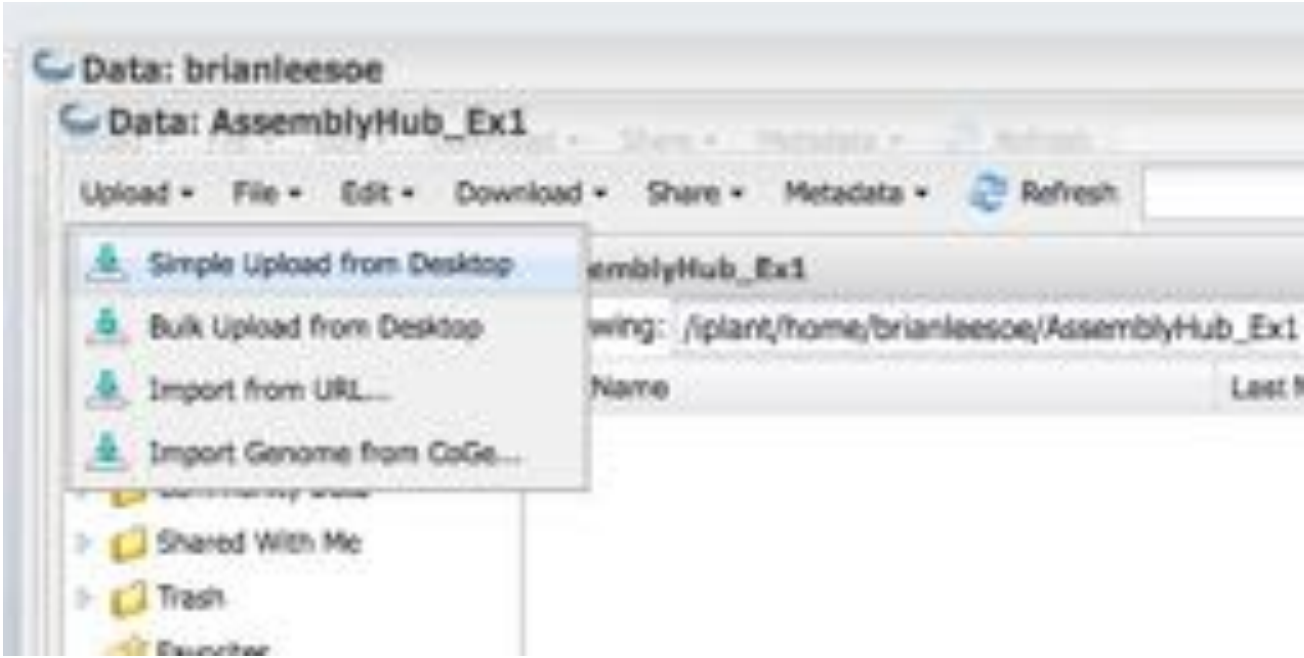
Hosting your data at CyVerse

Make a New Folder for Your Genome Data



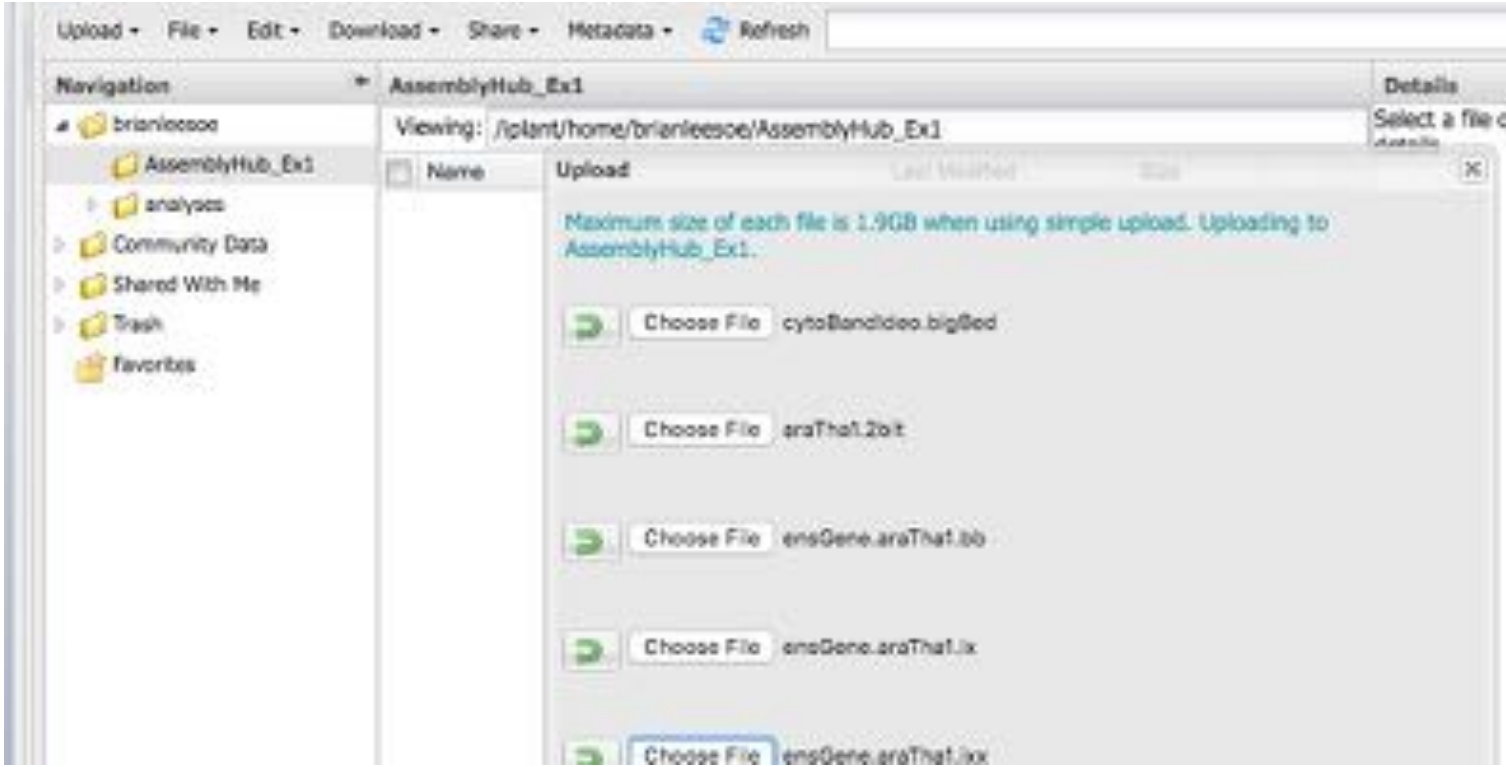
Hosting your data at CyVerse

Upload your files from your created 2bit files from your Desktop to CyVerse



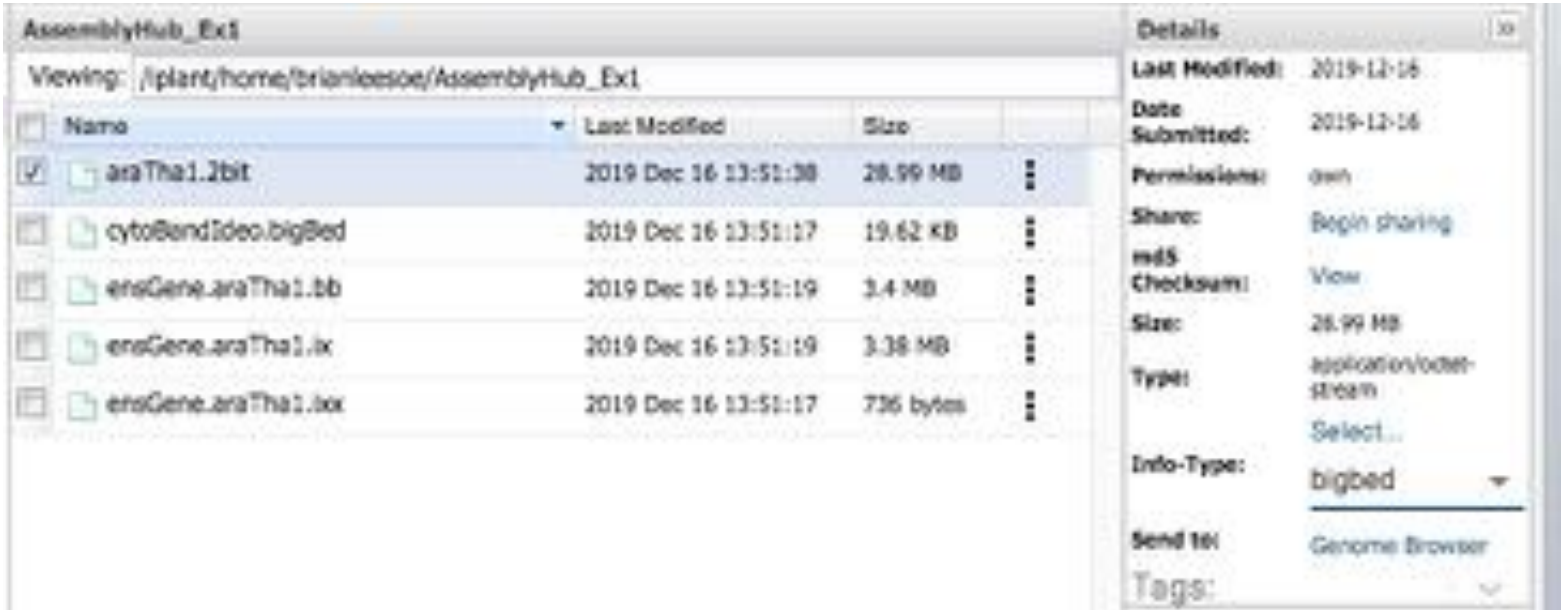
Hosting your data at CyVerse

Choose the Files to Upload (supporting hub files here)



Hosting your data at CyVerse

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



Select each file and set the **Type** to **bigBed** and then click the **Send to: Genome Browser**

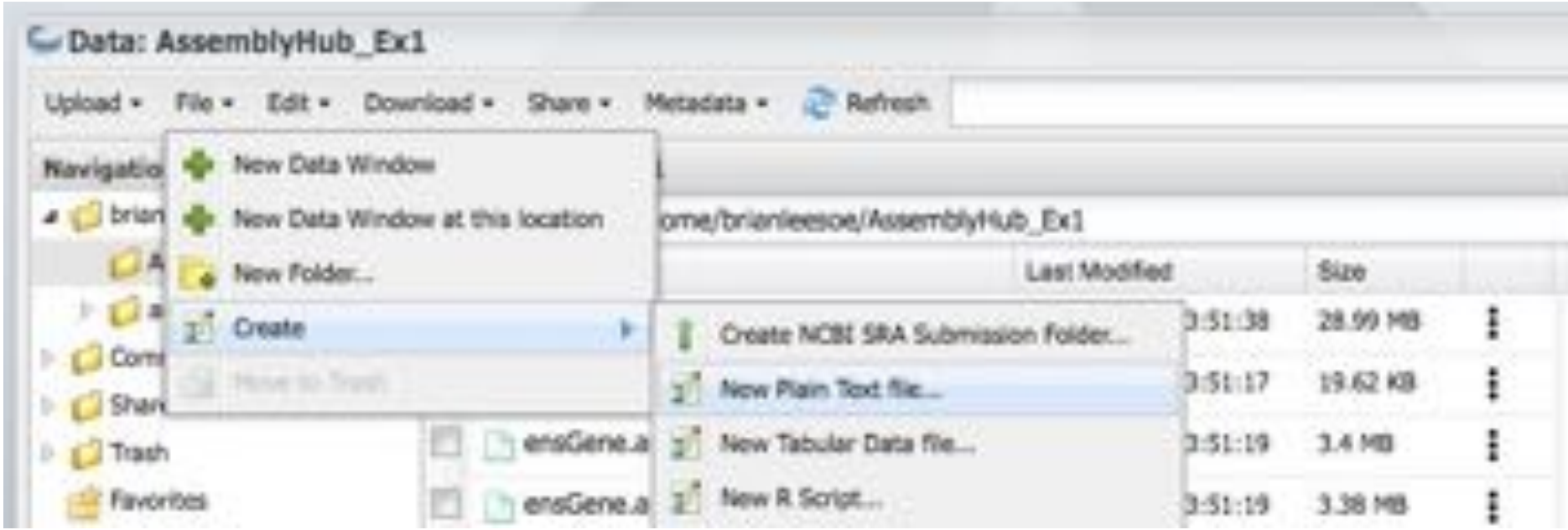
Hosting your data at CyVerse

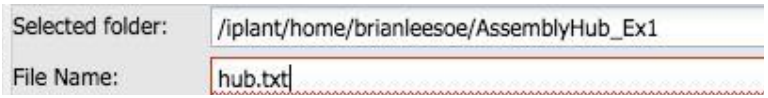
Creating a Link that Accepts Byte-Ranges: “Send to Genome Browser”

These generated **View in Genome Browser** links, which allow byte-range access to the data.

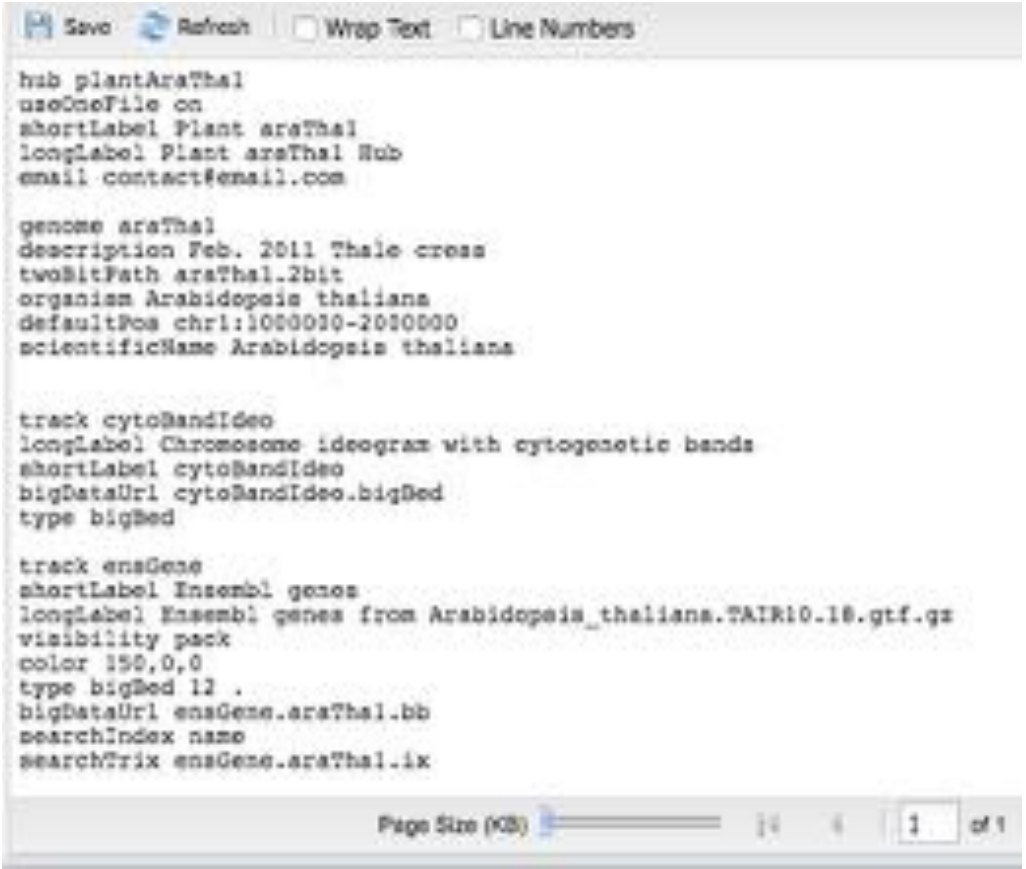
Hosting your data at CyVerse

Create a new Text File (or just upload one) for the **hub.txt**

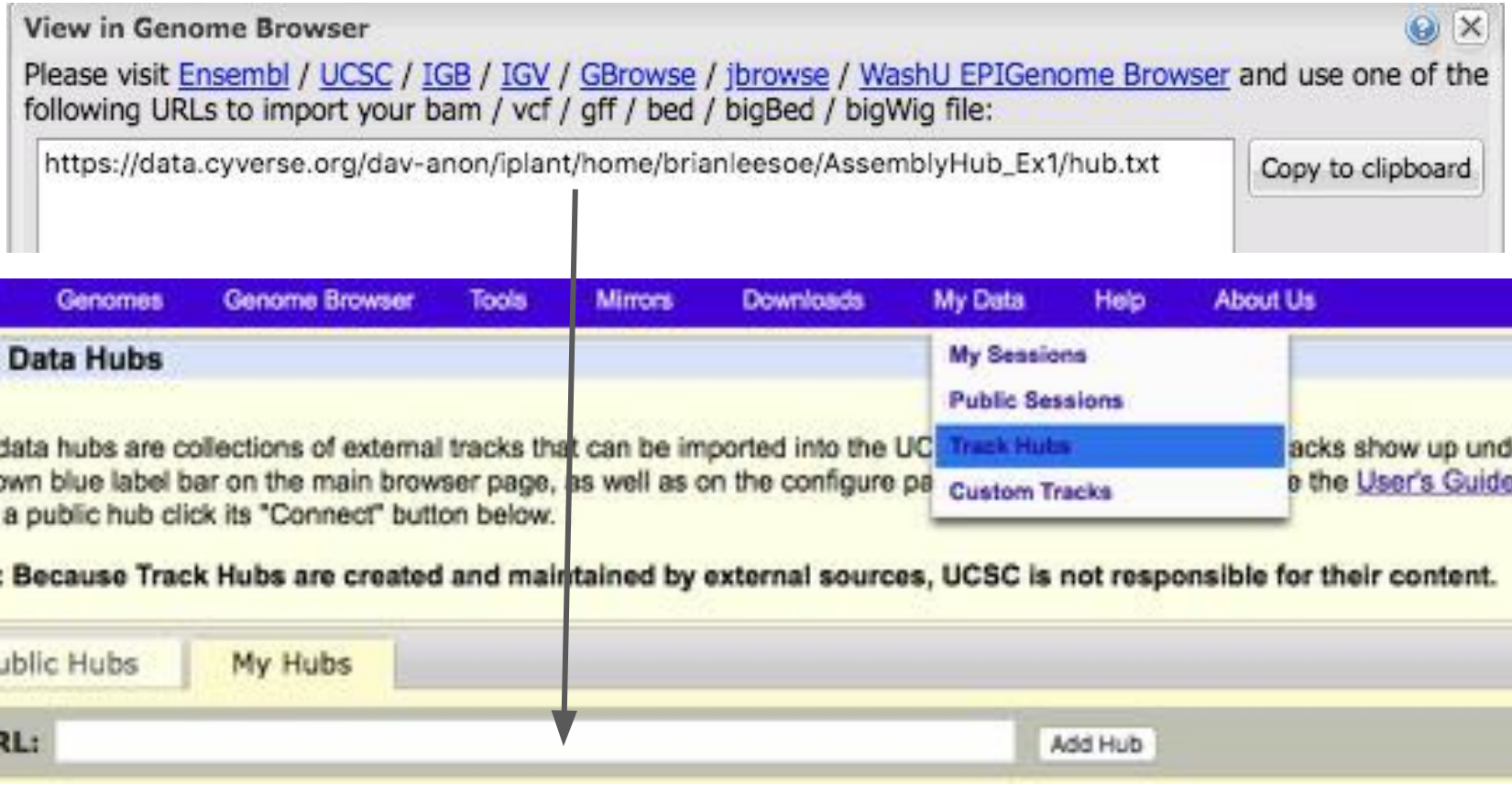




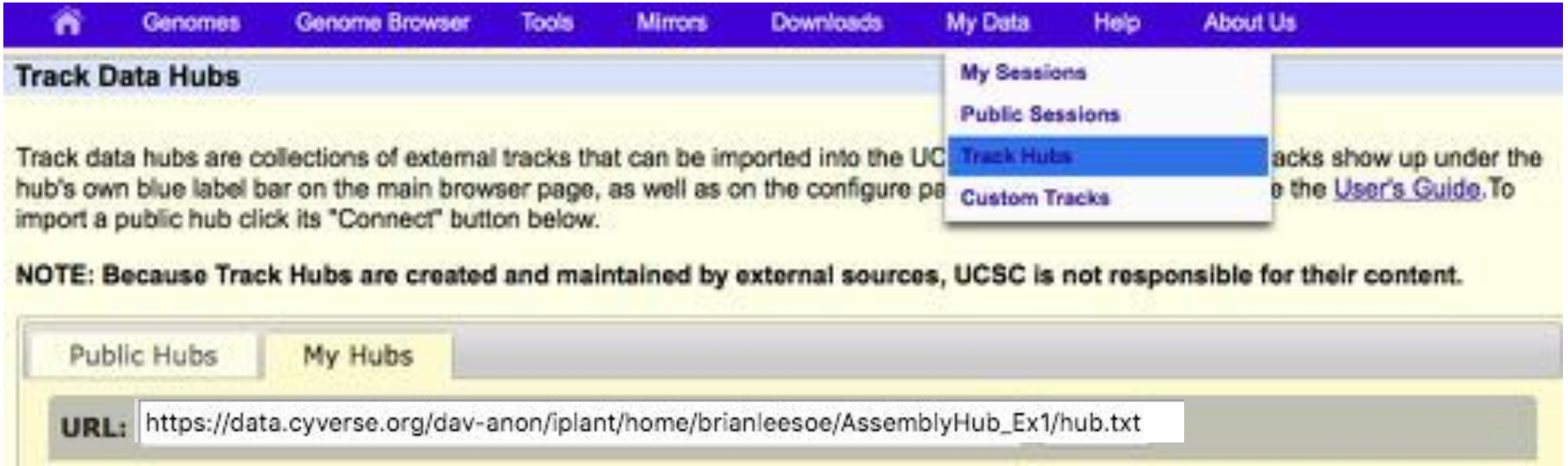
Hosting your data at CyVerse



Hosting your data at CyVerse

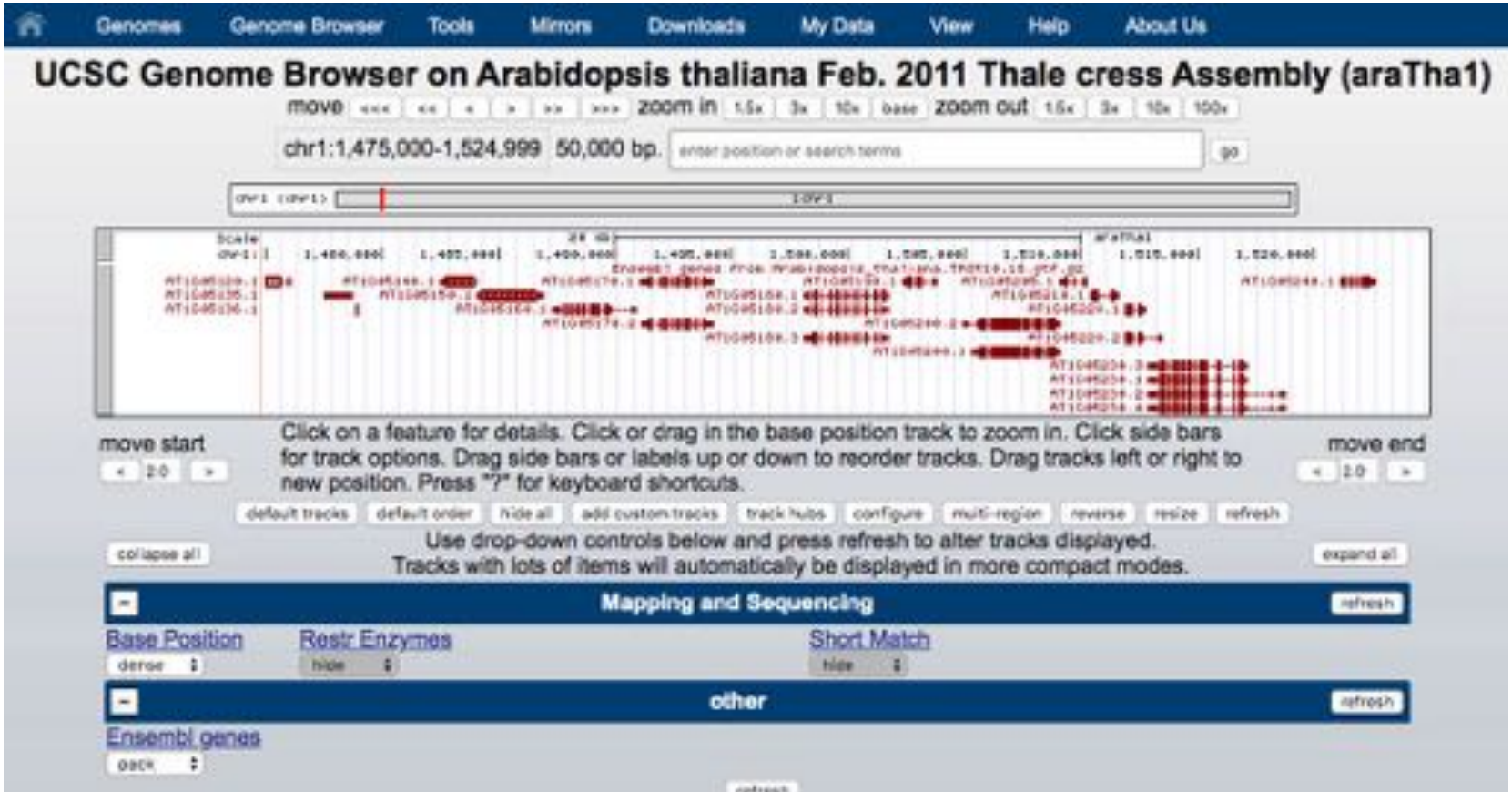


Hosting your data at CyVerse



https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/AssemblyHub_Ex1/hub.txt

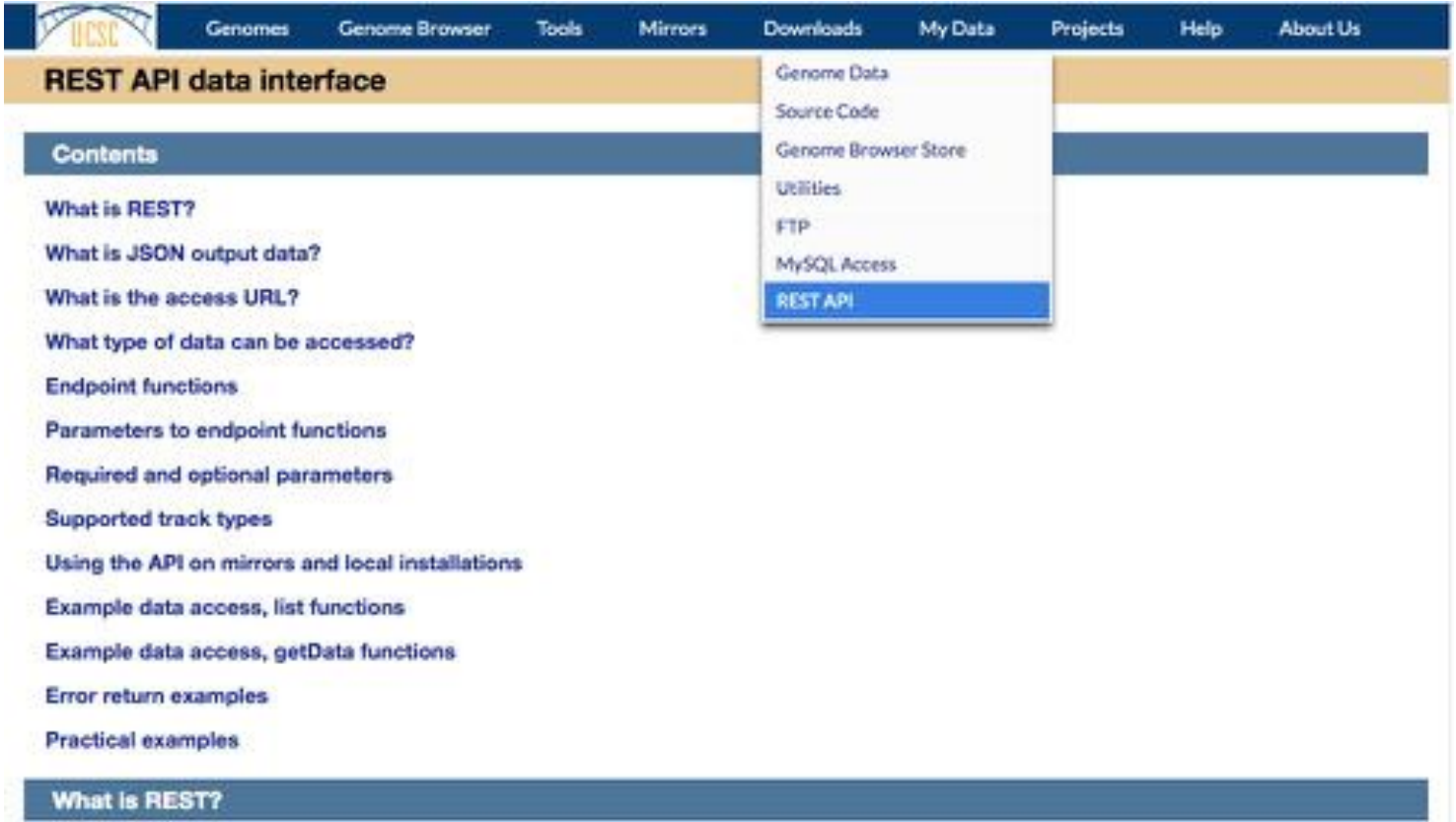
Hosting your data at CyVerse



https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/AssemblyHub_Ex1/hub.txt

Hosting your data at CyVerse

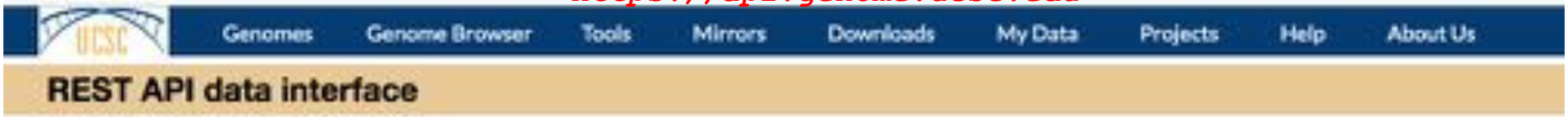
https://api.genome.ucsc.edu



https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/AssemblyHub_Ex1/hub.txt

Hosting your data at CyVerse

https://api.genome.ucsc.edu

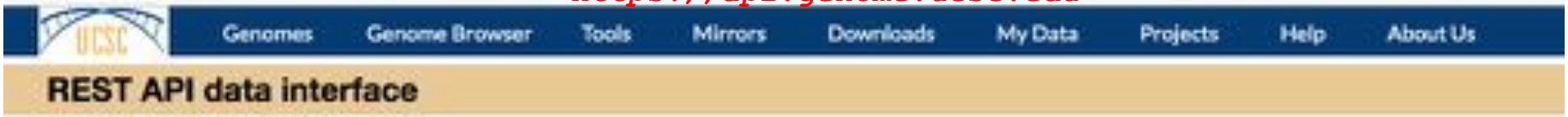


http://api.genome.ucsc.edu/getData/sequence?genome=araTha1&chrom=chr1&start=4321&end=4333&hubUrl=https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/AssemblyHub_Ex1/hub.txt

https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/AssemblyHub_Ex1/hub.txt

Hosting your data at CyVerse

https://api.genome.ucsc.edu



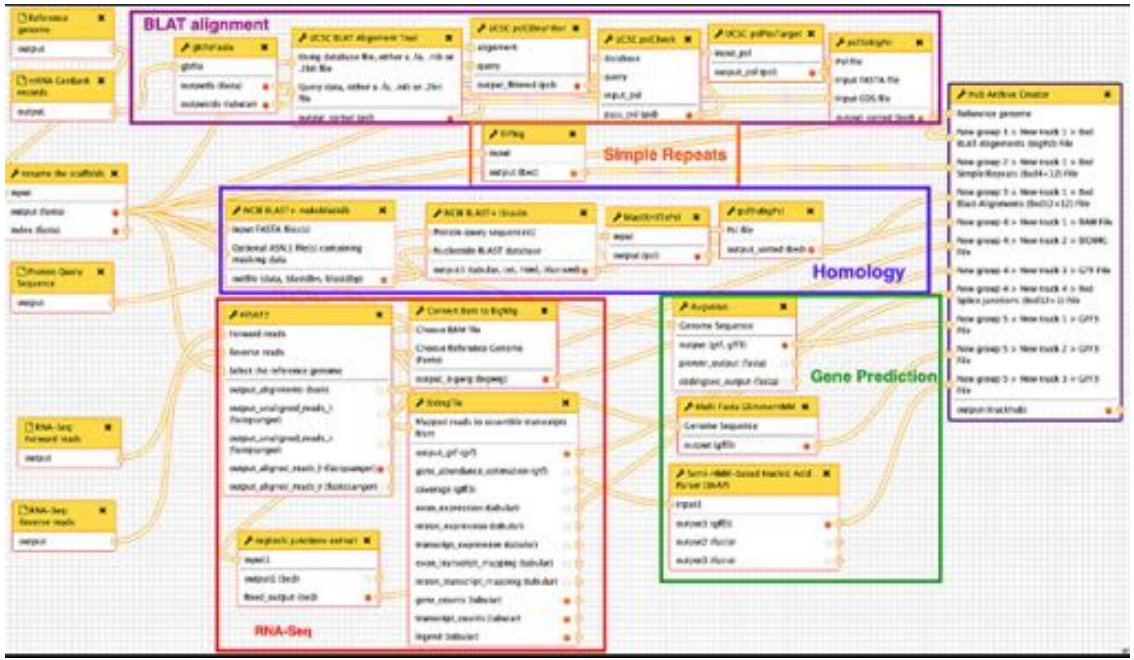
http://api.genome.ucsc.edu/getData/track?genome=araTha1&chrom=chr1&start=4321&end=4333&track=ensGene&hubUrl=https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/AssemblyHub_Ex1/hub.txt


```
hubUrl: "https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/AssemblyHub_Ex1/hub.txt"
genome: "araTha1"
chromSize: 30427671
start: 4321
end: 4333
bigDataUrl: "https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/AssemblyHub_Ex1/ensGene.araTha1.bb"
trackType: "bigBed 12 ."
ensGene:
  0:
    chrom: "chr1"
    chromStart: 3630
    chromEnd: 5899
    name: "AT1G01010.1"
    score: 0
    strand: "+"
    thickStart: 3759
    thickEnd: 5630
    reserved: "0"
    blockCount: 6
    blockSizes: "283,201,120,390,153,461,"
    chromStarts: "0,365,855,1075,1543,1808,"
itemsReturned: 1
```

Assembly Track Hub Creation Resources

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

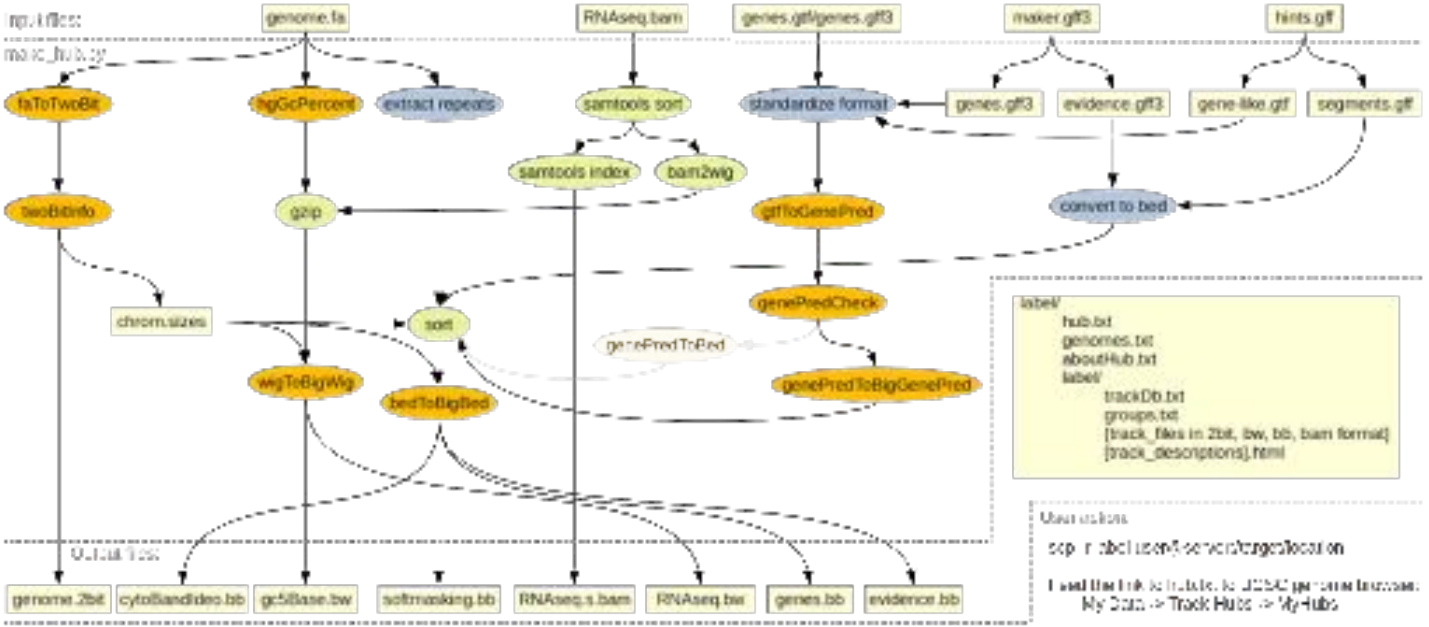
http://gonramp.wustl.edu/



Assembly Track Hub Creation Resources

MakeHub Command line tool for the fully automatic generation of assembly hubs

https://github.com/Gaius-Augustus/MakeHub



Thank You!