

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

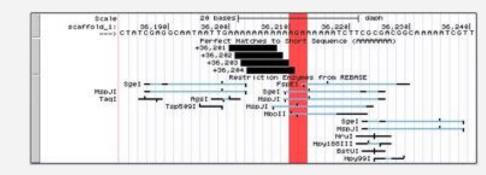
Visualize at UCSC

Binary indexed files:

twoBitPath, bigDataUrl

Text files: hub.txt

```
>Scoffold_1
GTTGTAAATACTCTATTCTACAATAAAACCAA
TCATAGGTTGAATTGGCGTTGAAGTAAAACCAA
...
>Bcaffold_2
AGTTATGACAAACTATAAAAAGTCGGTAGAGACAAAAC.
TCGTTCGTGGACGAAGCGACCAAAACTGAGCACAAGAT/
...
>Scoffold_3
CATAAATTCATAAATCAATTCATGAAGAATAATT
TAGAAAAATTCCCCAGGAAGTTGGAAGTTGCTA
```

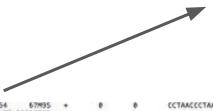


CyVerse's "Send to Genome Browser" option creates links accepting byte-range requests for binary index data.

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



PRESLEY_0030:6:5:16900:343200/2 65 falhfafela_cfaddfcaffW_edfabfcdfcf*ca	chri	10847 Wally2221	254 K*V2Ta91	67/95	*	0		CCTAACCCTAACCCTAACCCTAACCC	
PRESLEY_8030:6:26:1717:949000/2 65 f_fcdeffhfhaecffdfffffcffffcb'bb	chrl	10053	254	61M155	*	0		CCTAACCCTAACCCTAACCCTAACCC	
PRESLEY_8030:5:69:17883:15567#0/2 hgfgghhhhfhhhhhhhhhhhhghghhhhhhghghhhh	65	chri	10050	255	2573M15				AACTAACCCTAACCCTAA
PRESLEY_8030:5:75:7248:15014#0/2 hthfoghhoffhhhiffhohidghcag1hhhhfchhfh	65	chri	10060	255	2573415			0	AACTAACCCTAACCCTAA
PRESLEY_803815110111059613305#8/2 NARABANNARABANNARABANNARA	65	chri	10050	255	2573915	*			AACTAACCCTAACCCTAA
PRESLEY_8030:7:79:7804:15262#0/2 hhibbahagfaghagan hhababhahana dhaf	65	chrt	10050	255	2573915	+		0	AACTAACCCTAACCCTAA
PRESLEY_0030:7:91:10770:19201#0/2 Nighhahahahahahfhhahfhgahhahgahachah	65	chri	10050	255	2572925				AACTAACCCTAACCCTAA
PRESLEY_8038:7:101:18643:21267#0/2 Nhhfhhhighhghhhhghfghhhhfgghhhffaghhh	65	chri	10060	255	2573415			0	AACTAACCCTAACCCTAA
brianleeghgwdev - s head -n 40 temp0	elete	1.55		1220			2	50	**************
PRESLEY_8030: 7:02:16032:9421#0/2 Nananahahahahahahahahahahahahahahahah						•			CTAACCCTAACCCTAACC
PRESLEY_8030:6:7:18371:16699#0/2 Yfffhhgefghhghgghhghhhg_ffffchfehgfff	65 fhatchfff	tlfacecer	10843	254	71MSS	•			TAACCCTAACCCTAACCC
PRESLEY_0030:6:5:16900:3432#0/2 65	chet	10847	254	67/495		0	. 0	CCTAJ	ACCCTARCCCTARCCCTARCCC



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
                                                PRESLEY 0030:6:5:16900:3432#0/2 65
                                                                                                      CCTAACCCTAACCCTAACCC
                                                falhfafela cfaddfcaffW edfabfcdfcf^cace^c\d\aaWaJYZZZ\\K^VZTaB
                                                PRESLEY 0030:6:26:1717:9490#0/2 65
                                                                      chr1
                                                                                                      CCTAACCCTAACCCTAACCC
                                                AACTAACCCTAACCCTAA
                                                PRESLEY 0030:5:75:7248:15014#0/2
                                                                                                          AACTAACCCTAACCCTAA
                                                hfhfgghhafhhhhffhghhdghcagfhhhhfchhfhcfhcqcedfaff hdfeeeheacc[Rchhh]egbd[bb
                                                PRESLEY 0030:5:101:10596:3305#0/2
                                                                      65
                                                                           chr1 10060 255
                                                                                                          AACTAACCCTAACCCTAA
```

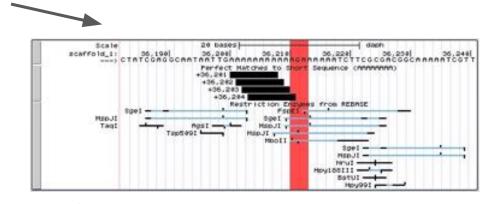
```
twoBitToFa -seq=chr1 -start=1499900 -end=1500055
http://yourGenome/your.2bit output.fa
```



>chr1:1499900-1500055
GCTACCATCACCCAAAAAGCTGAGGAGTTTGAATTCACTTCAGCACAACT
ATCATTAATTAATTTTTGAACCTCTGAGCCTGGAAGAGAAAACAGGTTTG
GTTCAACATGAAGAATACTGTGATTTGACCCGTGACAGAGCTTTCTGTTA

Viewing Data at UCSC

bigDataUrl http://location_of/file.bam



twoBitPath http://yourGenome/your.2bit

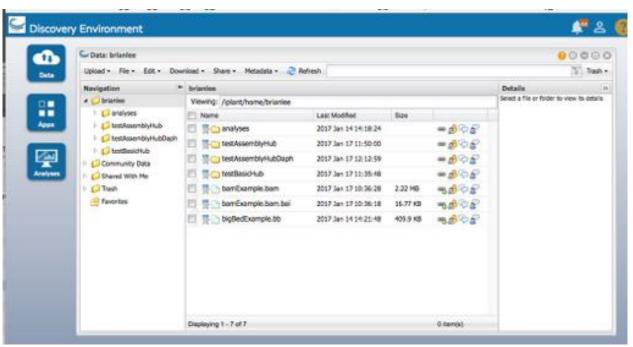
CyVerse Storage Solution

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



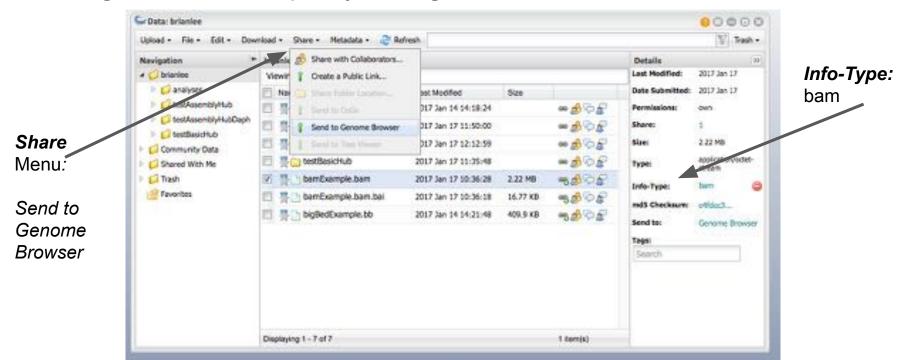
CyVerse Storage Solution

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



CyVerse Storage Solution

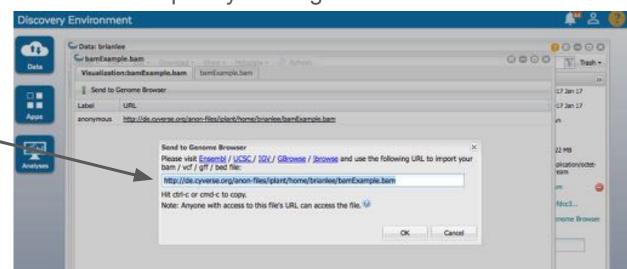
Creating a Link that Accepts Byte-Ranges: "Send to Genome Browser"



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

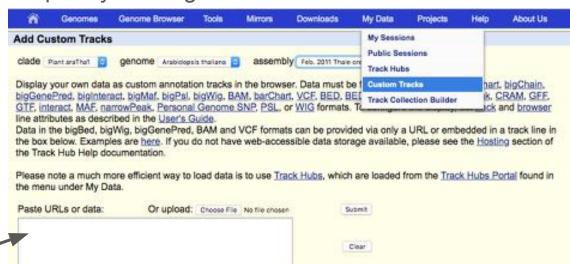


https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/Bam_Ex1/DNase_example.bam

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



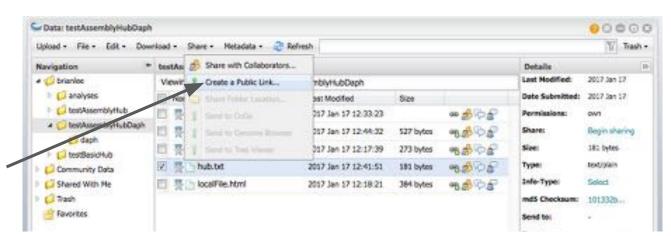
https://data.cyverse.org/dav-anon/iplant/home/brianleesoe/Bam_Ex1/DNase_example.bam

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

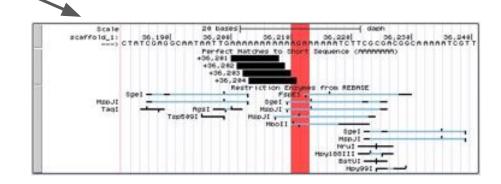
Will not work to visualize data in the Browser

```
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

twoBitPath

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

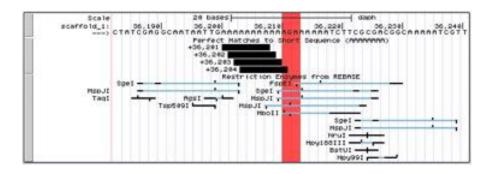


hub plantAraTha1 useOneFile on shortLabel Plant araTha1 longLabel Plant araTha1 Hub email contact@email.com

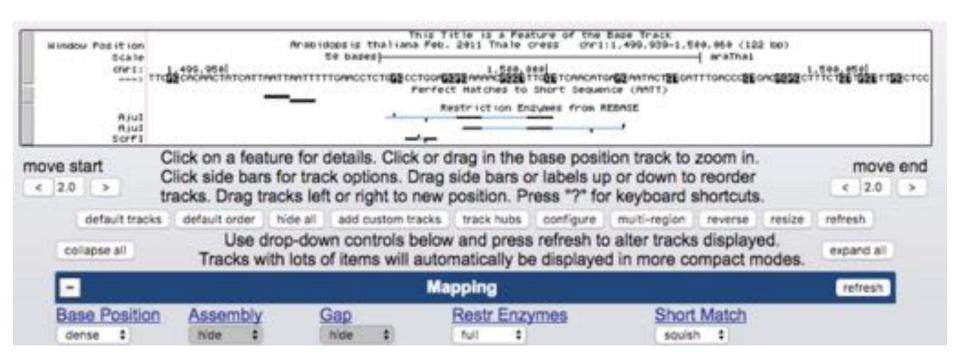
genome araTha1
description Feb. 2011 Thale cress
twoBitPath araTha1.2bit
organism Arabidopsis thaliana
defaultPos chr1:1000000-2000000
scientificName Arabidopsis thaliana

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

hub.txt



Browser requests *AGCTs* for only the window currently viewing (chr1:1,499,900-1500,055)



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

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

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

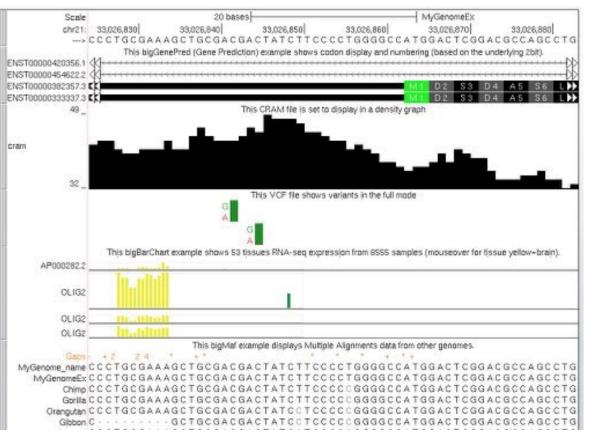
genome araTha1
description Feb. 2011 Thale cress
twoBitPath araTha1.2bit
organism Arabidopsis thaliana
defaultPos chr1:1000000-2000000
scientificName Arabidopsis thaliana

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



Host all data at CyVerse

Visualize at UCSC

Binary indexed files:

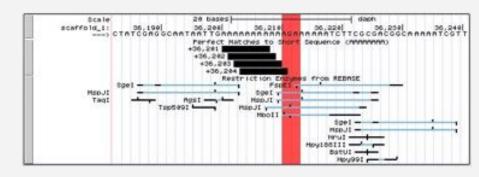
twoBitPath,

bigDataUrl

```
>scaffold_1
GTTGTAAATACTCTATTCTACAATAAAACCAA
TCATAGGTTGAATTGGCGTTGAAGTAAAACCAA
...
>scaffold_2
AGTTATGACAAACTATAAAAAGTCGGTAGAGACAAAAGA.
TCGTTCGTGGACGAAGCGACCAAAACTGAGCACAAGAT/
...
>scaffold_3
CATAAATTCATAAATCAATTCATGAAGAATAAT/
TAGAAAATTTCCCAGGAAGTTTGAAGTTGCTA/
```

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

Text files: hub.txt



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

```
hubDirectory
| Move all your local BAMS, VCFs, and related Assembly Hub files with the single irsync command. https://data.cyverse.org/dav-anon/
| trackAnnotation.bigBed Then make the files publicly accessible with the ichmod command.
```

ichmod read anonymous data store directory name



About

Products +

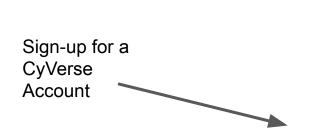
Learning

Events

News

Launch +

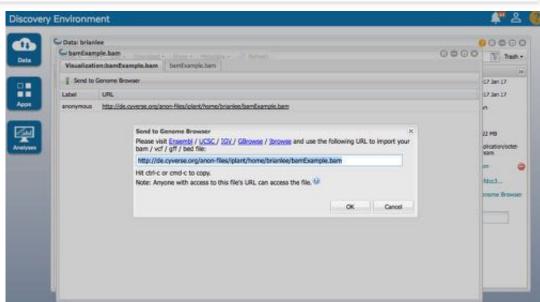
Log In | Register

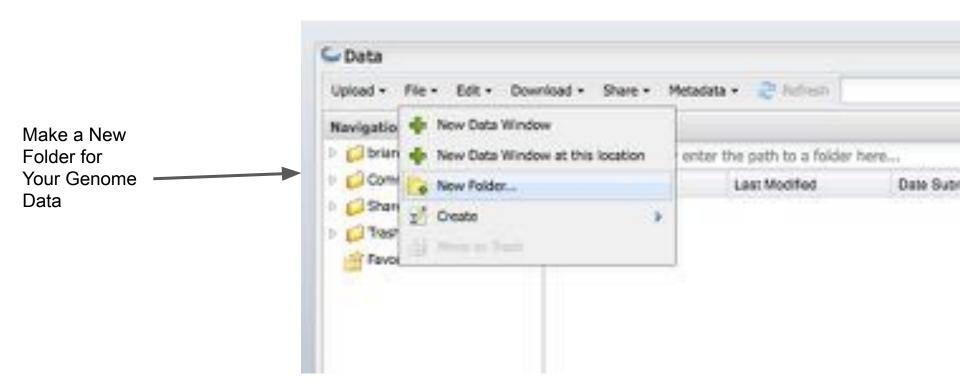


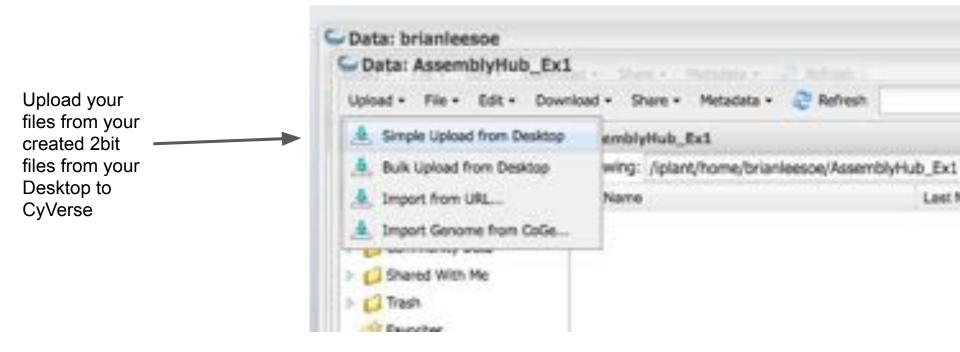


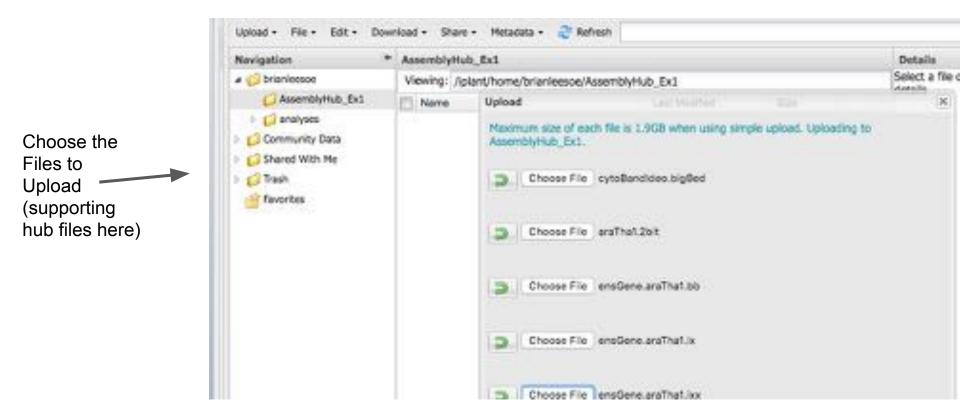


Launch the Discovery Environment and click Data

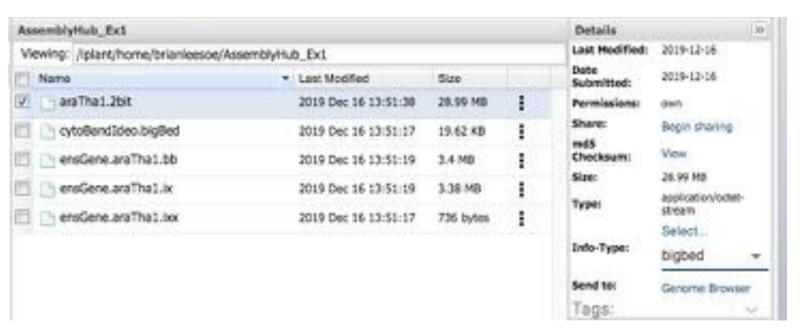








Creating a Link that Accepts Byte-Ranges: "Send to Genome Browser"



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

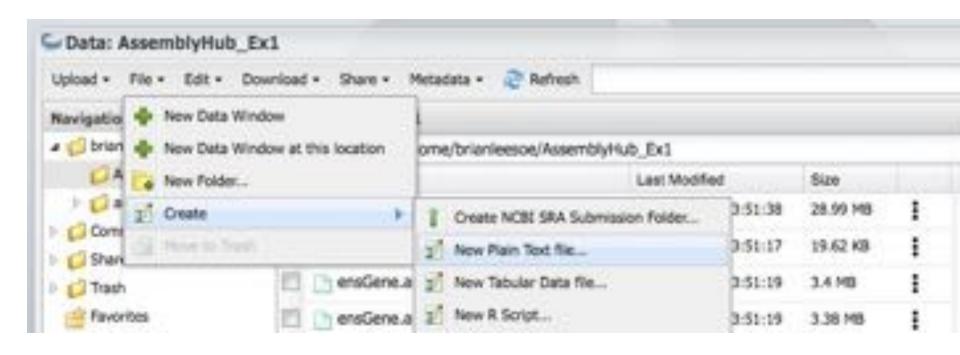


Creating a Link that Accepts Byte-Ranges: "Send to Genome Browser"



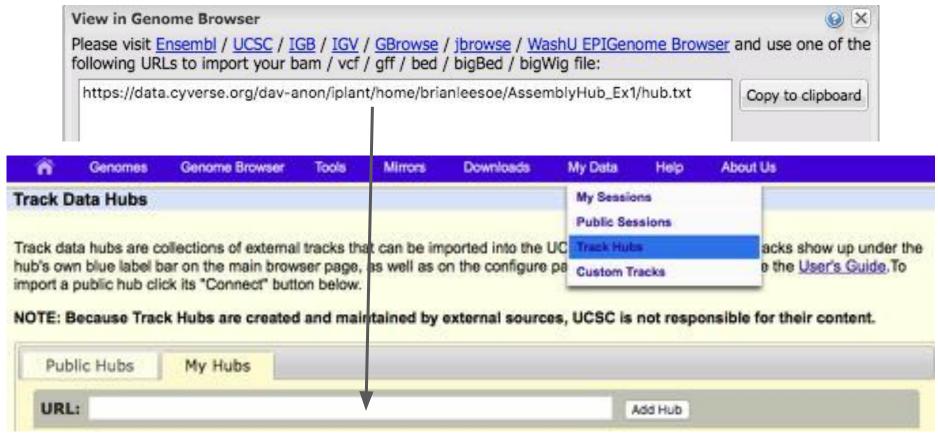
These generated View in Genome Browser links, which allow byte-range access to the data.

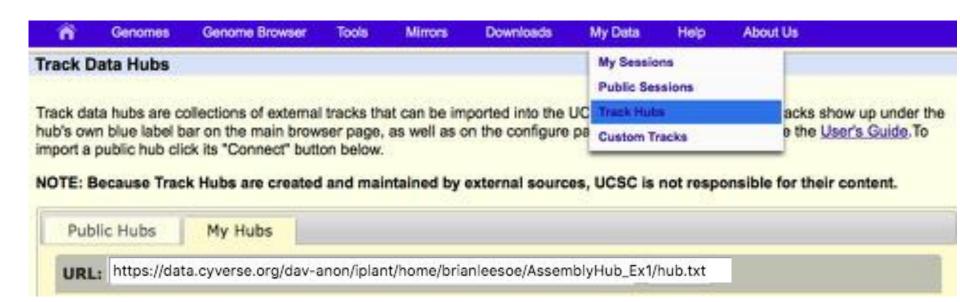
Create a new Text File (or just upload one) for the hub.txt

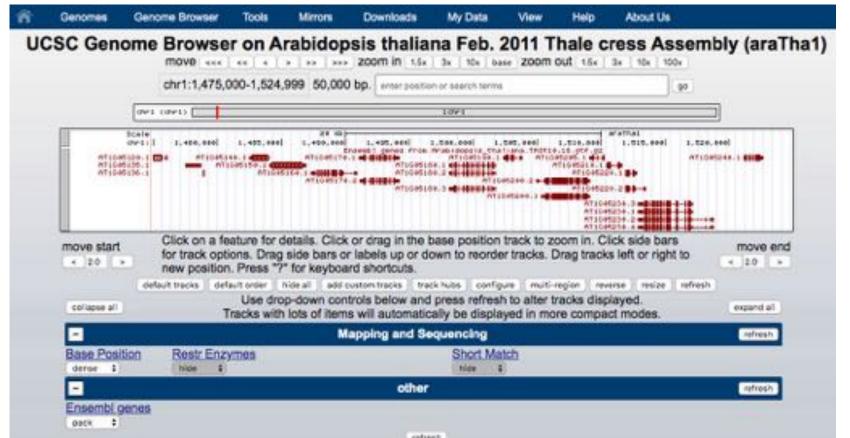


Selected folder:	/iplant/home/brianleesoe/AssemblyHub_Ex1				
File Name:	hub.txt				

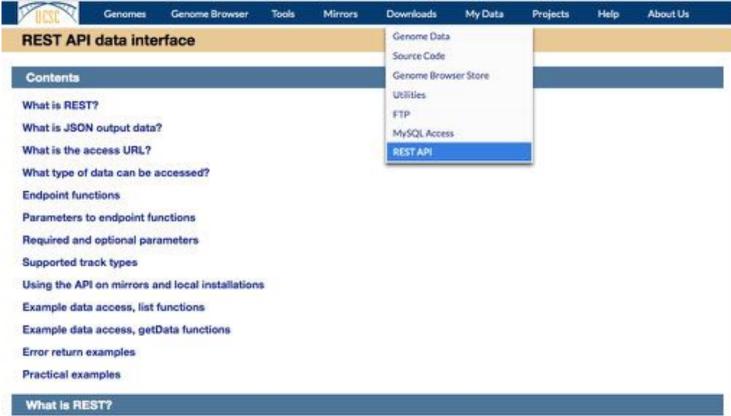
Savo Refresh Wrap Text Line Numbers				
hub plantAraThal useOneFile on shortLabel Plant araThal longLabel Plant araThal Hub enail contact#enail.com				
genome araThal description Feb. 2011 Thale cress twoBitFath araThal.2bit organism Arabidopsis thalians defaultPos chrl:1000010-2010000 scientificName Arabidopsis thalians				
track cytoBandIdeo longLabel Chromosome ideogram with cytogenetic bends shortLabel cytoBandIdeo bigDataUrl cytoBandIdeo.bigBed type bigBed				
track ensGene shortLabel Ensembl genes longLabel Ensembl genes from Arabidopeis_thalians.TAI visibility pack color 150,0,0 type bigBed 12 . bigDataUrl ensGene.araThal.bb searchTrix ensGene.araThal.ix	R10.	10.5	gtf.gs	
Page Size (KB)	4	6	1	of 1







https://api.genome.ucsc.edu



https://api.genome.ucsc.edu



Genomes

Genome Browser

Tools

Mirrors

Downloads

My Data

Projects

Help

About Us

REST API data interface

```
downloadTimer:
                       "2019:32:39720:12:302"
 downloadTuneStamp:
                       1576600038
- hubbirls
                       "https://data.cyverse.org/dav-amon/iglant/home/brienlocase/AssemblyHub_Ex1/hub.txt"
 genome:
                       "araThal"
 chronic
                        "chr3"
 start:
                        4321
 ಕಾರ್ಡ್
                       4333
 dea:
                       "totateterser"
```

https://api.genome.ucsc.edu



Genomes

Genome Browser

Tools:

Mirrors

Downloads

My Data

Projects

Help

About Us

REST API data interface

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

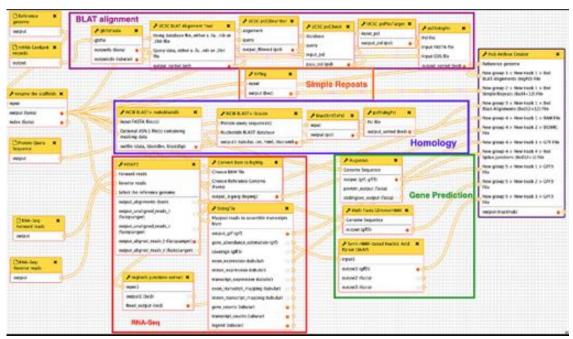
"Mitsi//data.cyverse.org/dav-anon/jplant/home/brianleosoe/AssemblyHub Ex1/hub.txt"

```
genones
                       "araThal"
 chron5lze)
                        38427675
                        4321
 starti
w bioDataUrli
                        "https://data.cyverse.org/dav-anon/iplant/home/brianleesee/AssemblyHub_Ex1/ensGene.araThal.bb"
                       "bigled 12 ."
 trackType
# ensGenel
  w 01
                        "chra"
       chronStart:
                        3638
       chronEnds
                        5899
                       "AT1081818.1"
       names
       scores
       strandi
       thickStart:
                       3759
       thickEndo
                        5638
                       1180
       reserveds
       blockCount!
                       "283, 281, 128, 398, 153, 461,"
       block5trest
       chromStarts:
                       "0,365,855,1075,1543,1888,"
  ItemsReturned:
```

Assembly Track Hub Creation Resources

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

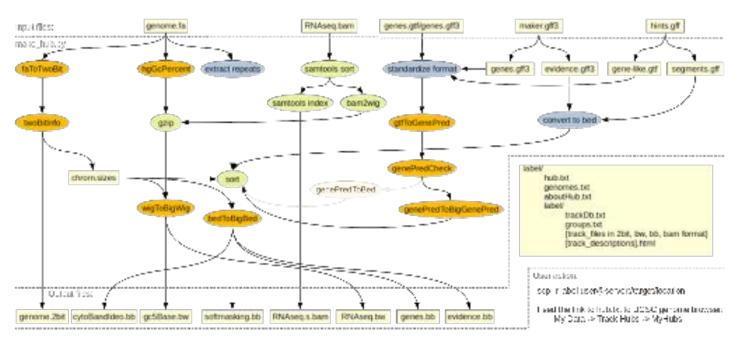
http://gonramp.wustl.edu/



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!



