

Viewing CyVerse Hosted Data at UCSC

Host data at CyVerse

Binary indexed files:

twoBitPath, **bigDataUrl**

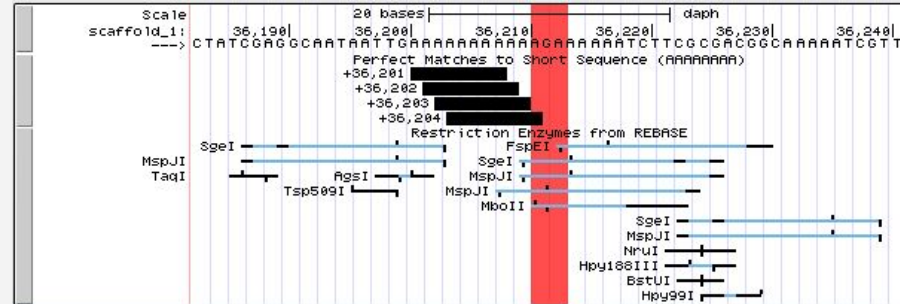
Text files: **hub.txt**,

genomes.txt, **trackDb.txt**

```
>scaffold_1
GTTGTAATACTCTATTCTACAATAAAACCAA
TCATAGGTTGAATTGGCGTTGAAGTAAAACAAA
...
>scaffold_2
AGTTATGACAAACTATAAAAAGTCGGTAGAGACAAAAG
TCGTTTCGTGGACGAAGCGACCAAAACTGAGCACAAGATA
...
>scaffold_3
CATAAATTCATAAATCAATTCATGAAGAATAATT
TAGAAAATTTCCAGGAAGTTTGAAGTTGCTA
...
```

```
hubDirectory
├__hub.txt
├__genomes.txt
├__hg19
└__trackDb.txt
```

Visualize at UCSC

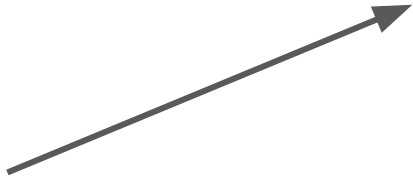


CyVerse’s “*Send to Genome Browser*” option creates links accepting byte-range requests for binary index data.


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
```



```
PRESLEY_0030:6:5:16900:3432#0/2 65 chr1 10047 254 67M9S * 0 0 CCTAACCCCTAACCCCTAACCCCTAACCC
faJhfafe]a_cfaddfcaffw_edfabfcdcf^cace^c\d\aaawaJZZZ\K^VZTabBBBBBBBB
PRESLEY_0030:6:26:1717:9490#0/2 65 chr1 10053 254 61M15S * 0 0 CCTAACCCCTAACCCCTAACCCCTAACCC
f_fcdeffhfaecffdf]ffdcf]ffcff]cb'bb[d]db^W^_^_J_Y^BBBBBBBBBBBBBBBB
PRESLEY_0030:5:69:17803:15567#0/2 65 chr1 10060 255 2573M1S * 0 0 AACTAACCCCTAACCCCTAA
hgfgghhhhfhhhhhhhhhghghhhhhghghhhhhhhfhhgdhhchghfchfhhadbhhea]_cfca[
PRESLEY_0030:5:75:7248:15014#0/2 65 chr1 10060 255 2573M1S * 0 0 AACTAACCCCTAACCCCTAA
hfhfgghghfhhhffhghhdghcagfhhhfchhfhc fhcgcedfaff_hdfeeheaacc[Rchhh]egbd[bb
PRESLEY_0030:5:101:10596:3305#0/2 65 chr1 10060 255 2573M1S * 0 0 AACTAACCCCTAACCCCTAA
hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhghhhhhhhahhhghd_dhgcgded^a^
PRESLEY_0030:7:79:7804:15262#0/2 65 chr1 10060 255 2573M1S * 0 0 AACTAACCCCTAACCCCTAA
hhhhhhhhghfhghghghfhhhhhhhhhhhhh_dhhffchhdadhfhf[hhhhghhhc]ccRffccchdg]gR
PRESLEY_0030:7:91:10770:19281#0/2 65 chr1 10060 255 2572M2S * 0 0 AACTAACCCCTAACCCCTAA
hhghhhhhhhhhhhhhhhhhfhghhhhhhhhhhhhhghhhghhhghhhhehehfad[cfhd_fbd`BB
PRESLEY_0030:7:101:18643:21267#0/2 65 chr1 10060 255 2573M1S * 0 0 AACTAACCCCTAACCCCTAA
hhfhhhhhhhghghhhghfhghhhfhghhhffgghhhffaghhhhchcdcafaffffaff[ff]ccef_ddegegacacaR
[brianlee@hgwdev ~]$ head -n 40 tempDelete
PRESLEY_0030:7:82:16832:9421#0/2 73 chr1 10042 255 68M8S * 0 0 CTAAACCCCTAACCCCTAACCC
hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhghhhghfhghhhchghfgfhghccQLcPUWJYa
PRESLEY_0030:6:7:18371:16699#0/2 65 chr1 10043 254 71M5S * 0 0 TAACCCCTAACCCCTAACCC
Yfffhgeghghghghghhhgh_ffffchfehghffhfhchffff]faccfc[fb[e]ebT]_`^QaY'
PRESLEY_0030:6:5:16900:3432#0/2 65 chr1 10047 254 67M9S * 0 0 CCTAACCCCTAACCCCTAACCCCTAACCC
```



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
GTTGTAAATACTCTATTCTACAATAAAACCAAAGATAACTCTTTATCAG
TCATAGGTTGAATTGGCGTTGAAGTAAACAAAAATACTGCTCAAAGG
...
>scaffold_2
AGTTATGACAACTATAAAAAGTCGGTAGAGACAAAAGCGAAAAGGATCT
TCGTTTCGTGGACGAAGCGACCAAACCTGAGCACAAGATAAATCCCGAATA
...
>scaffold_3
CATAAATTCATAAATCAATTCATGAAGAATAATTTTAGAAAAATGGTTCA
TAGAAAAATTTCCAGGAAGTTTGAAGTTGCTATAATGATTATTTCTCTTG
...
>scaffold_4
ATACTGCCCCATGCCTCATCTACTCCCTTCTTAACCTTTTCATCAACAAC
TGCTTAAGTTGGTCATCACTGACTGAGTGAGCCCAAACCTGGAATAGCTGT
...
```

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 chr1 10047 254 67M9S * 0 0 CCTAACCTAACCTAACCTAACCC
fa]hfafa]a_cfaddfcaffW_edfabfcdfcfc^cace^c\d\aaWaJYZZ\K^VZTaBBB888888
PRESLEY_0030:6:26:1717:9490#0/2 65 chr1 10053 254 61M15S * 0 0 CCTAACCTAACCTAACCTAACCC
f_fcdefhfhaecffdf]ffdcf]ffcff]cb`bb[d]db*W`^^_Y^BBB8888888888888888
PRESLEY_0030:5:69:17883:15567#0/2 65 chr1 10060 255 2573M1S * 0 0 AACTAACCTAACCTTAA
hgfgghhhfhhhhhhhhghghhhhhghghhhhhhhghghhhghghhhchghfchfhhadbhhea]_cfca[
PRESLEY_0030:5:75:7248:15014#0/2 65 chr1 10060 255 2573M1S * 0 0 AACTAACCTAACCTTAA
hfhfgghhgfhhhffhghhdghcagfhhhfchhfhc fhgcedfaff_hdfeeeheacc[Rchhh]egbd[bb
PRESLEY_0030:5:101:10596:3305#0/2 65 chr1 10060 255 2573M1S * 0 0 AACTAACCTAACCTTAA
hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhghghghhhhhahhhghd_dnhgceded^a^
.....
```

twoBitToFa -seq=chr1 -start=1499900 -end=1500055

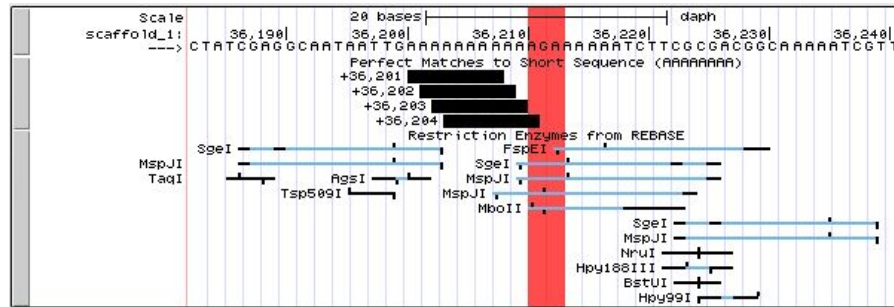
<http://yourGenome/your.2bit> output.fa



```
>chr1:1499900-1500055
GCTACCATCACCCAAAAAGCTGAGGAGTTTGAATTCACCTTCAGCACAACT
ATCATTAAATTAATTTTTGAACCTCTGAGCCTGGAAGAGAAAAACAGGTTTTG
GTTCAACATGAAGAATACTGTGATTTGACCCGTGACAGAGCTTTTCTGTTA
```

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



Discovery Environment

The Discovery Environment integrates powerful, community-recommended software tools into a system that:

- Makes big data management easy. Upload, organize, edit, view and search with ease!
- Has 500+ scientific apps that utilize compute clusters and HPC resources as needed.
- Hides the complexity needed to do these tasks.

Log in with your CyVerse ID

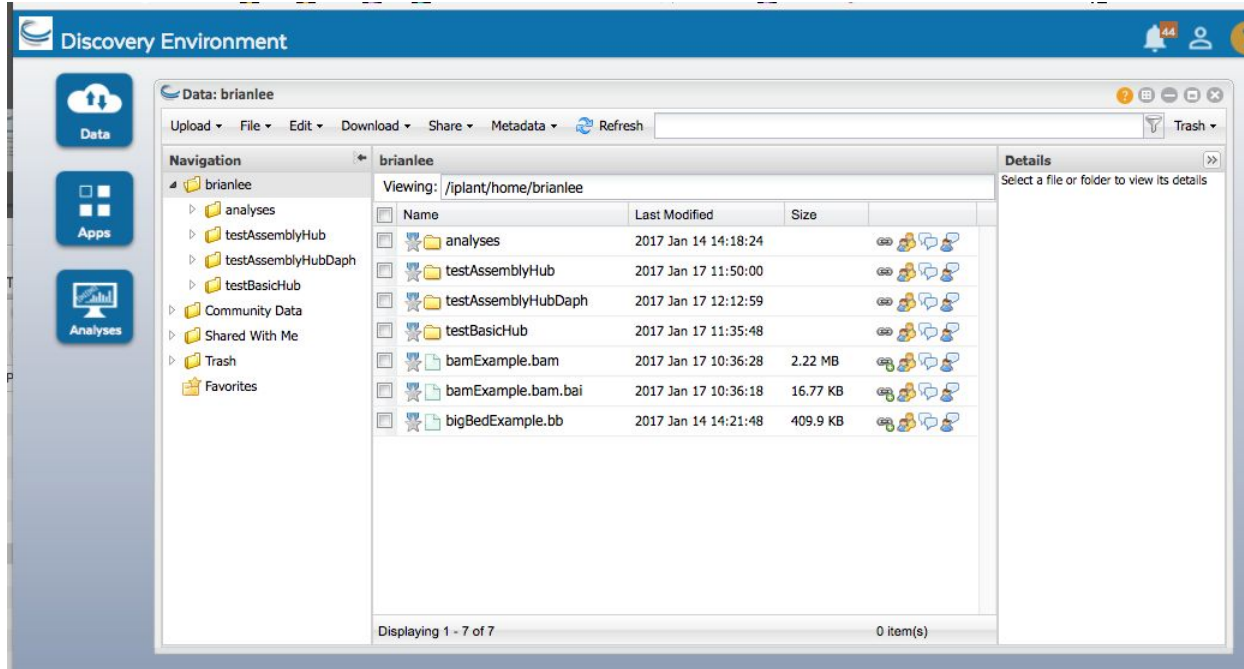
[Forgot Password?](#) [Register Now](#)

Minimum screen resolution supported: 1024 x 768

Hosting your data at CyVerse

CyVerse Storage Solution

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



The screenshot displays the CyVerse Discovery Environment interface. The top navigation bar includes the CyVerse logo, the text "Discovery Environment", and icons for notifications (44), user profile, and help. The main content area is titled "Data: brianlee" and features a menu with options: Upload, File, Edit, Download, Share, Metadata, Refresh, and Trash. On the left, there are three vertical buttons: "Data", "Apps", and "Analyses". The central pane shows a file browser for the user "brianlee" at the path "/iplant/home/brianlee". The left sidebar lists folders: analyses, testAssemblyHub, testAssemblyHubDaph, testBasicHub, Community Data, Shared With Me, Trash, and Favorites. The main table displays a list of files and folders with columns for Name, Last Modified, and Size. The files listed are:

Name	Last Modified	Size
analyses	2017 Jan 14 14:18:24	
testAssemblyHub	2017 Jan 17 11:50:00	
testAssemblyHubDaph	2017 Jan 17 12:12:59	
testBasicHub	2017 Jan 17 11:35:48	
bamExample.bam	2017 Jan 17 10:36:28	2.22 MB
bamExample.bam.bai	2017 Jan 17 10:36:18	16.77 KB
bigBedExample.bb	2017 Jan 14 14:21:48	409.9 KB

The bottom status bar indicates "Displaying 1 - 7 of 7" and "0 item(s)".

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**

The screenshot shows the CyVerse storage interface for user 'brianlee'. A share menu is open over the file 'bamExample.bam'. The menu options are: Share with Collaborators..., Create a Public Link..., Share Folder Location..., Send to CoGe, Send to Genome Browser (highlighted), and Send to Tree Viewer. The file list below shows 'bamExample.bam' selected, with a size of 2.22 MB. The details panel on the right shows the file's metadata, including 'Info-Type: bam' and 'Send to: Genome Browser'.

Name	Last Modified	Size	Actions
bamExample.bam	2017 Jan 17 10:36:28	2.22 MB	[Share] [Info] [Download]
bamExample.bam.bai	2017 Jan 17 10:36:18	16.77 KB	[Share] [Info] [Download]
bigBedExample.bb	2017 Jan 14 14:21:48	409.9 KB	[Share] [Info] [Download]

Property	Value
Last Modified:	2017 Jan 17
Date Submitted:	2017 Jan 17
Permissions:	own
Share:	1
Size:	2.22 MB
Type:	application/octet-stream
Info-Type:	bam
md5 Checksum:	c4fdcc3...
Send to:	Genome Browser

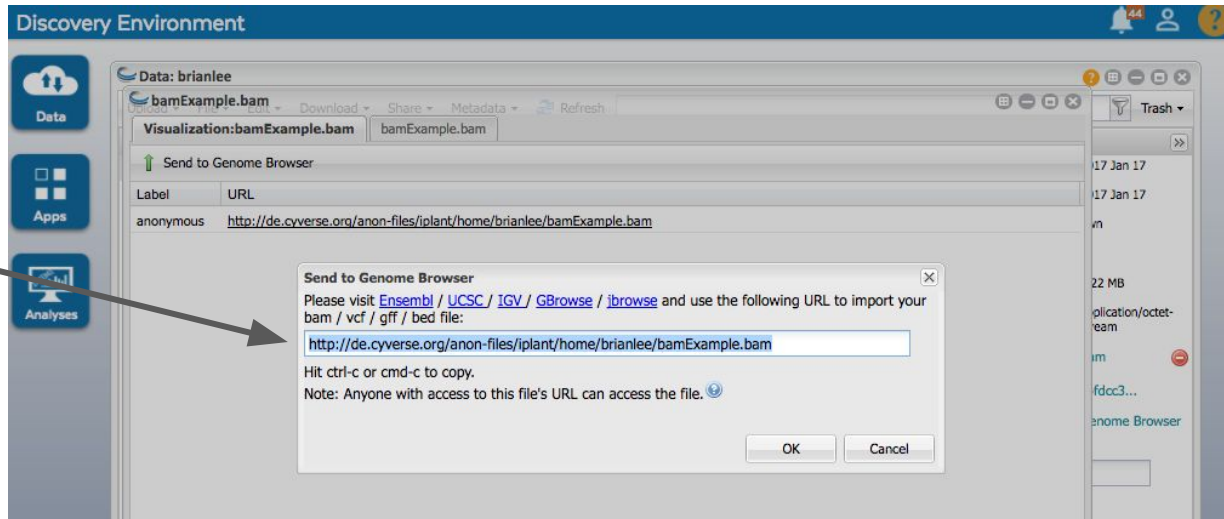
**Info-Type:
bam**

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***



<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.

The screenshot shows the CyVerse web interface. The main window displays a file list for the folder 'testAssemblyHubDaph'. The file 'hub.txt' is selected, and a context menu is open over it. The menu options include 'Share with Collaborators...', 'Create a Public Link...', 'Share Folder Location...', 'Send to CoGe', 'Send to Genome Browser', and 'Send to Tree Viewer'. The 'Create a Public Link...' option is highlighted. The details pane on the right shows the following information for the selected file:

Details	
Last Modified:	2017 Jan 17
Date Submitted:	2017 Jan 17
Permissions:	own
Share:	Begin sharing
Size:	181 bytes
Type:	text/plain
Info-Type:	Select
md5 Checksum:	101332b...
Send to:	-
Tags:	

At the bottom of the interface, it says 'Displaying 1 - 5 of 5' and '1 item(s)'.

<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
```

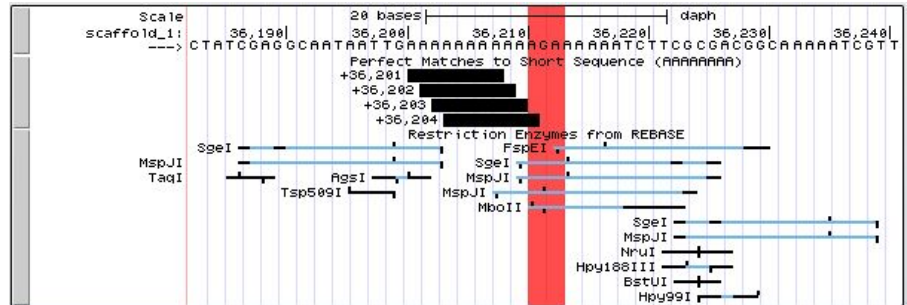


```
>chr1:1499900-1500055  
GCTACCATCACCCAAAAAGCTGAGGAGTTTGAATTCCTTCAGCACAACT  
ATCATTAATTAATTTTTGAACCTCTGAGCCTGGAAGAGAAAACAGGTTTG  
GTTCAACATGAAGAATACTGTGATTTGACCCGTGACAGAGCTTTCTGTTA
```

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 *AGCT*s for only the window currently viewing
(chr1:1,499,900-1500,055)



move start

< 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in.
Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. Press "?" for keyboard shortcuts.

move end

< 2.0 >

default tracks

default order

hide all

add custom tracks

track hubs

configure

multi-region

reverse

resize

refresh

collapse all

Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

expand all

Mapping

refresh

[Base Position](#)

dense ▾

[Assembly](#)

hide ▾

[Gap](#)

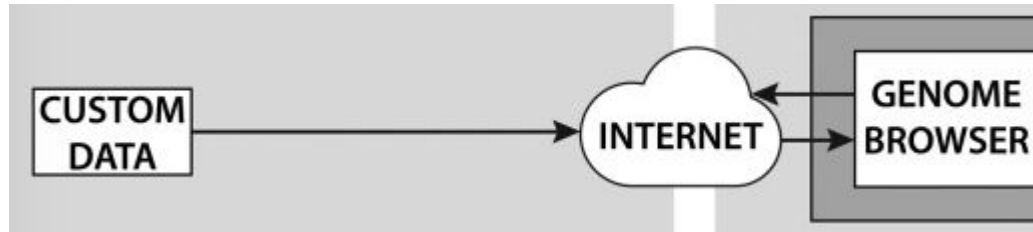
hide ▾

[Restr Enzymes](#)

full ▾

[Short Match](#)

squish ▾



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 *AGCT*s for only the window currently viewing
(chr1:1,499,900-1500,055)



move start

< 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in.
Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. Press "?" for keyboard shortcuts.

move end

< 2.0 >

default tracks

default order

hide all

add custom tracks

track hubs

configure

multi-region

reverse

resize

refresh

collapse all

Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.

expand all

Mapping

refresh

[Base Position](#)

dense ▾

[Assembly](#)

hide ▾

[Gap](#)

hide ▾

[Restr Enzymes](#)

full ▾

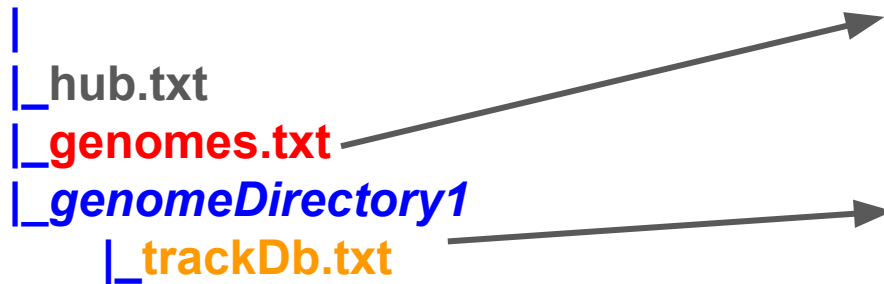
[Short Match](#)

squish ▾

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**)

hubDirectory



Informs location of genome 2bit file(s): **twoBitPath**

Informs location of track data file(s): **bigDataUrl** (BAM)

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

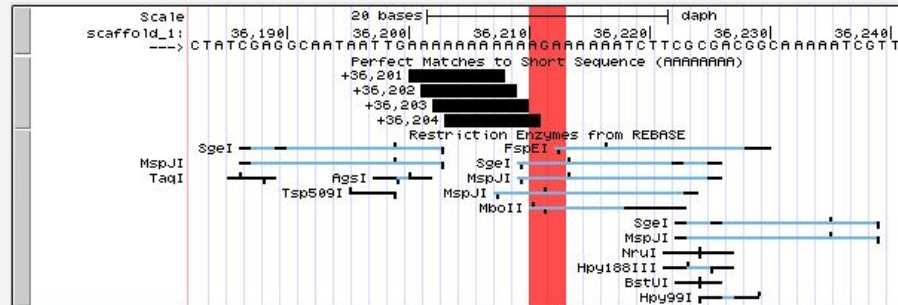
Text files: **hub.txt,**

genomes.txt,
trackDb.txt

```
>scaffold_1
GTTGTAATACTCTATTCTACAATAAAACCAA
TCATAGGTTGAATTGGCGTTGAAGTAAAACAAA
...
>scaffold_2
AGTTATGACAACTATAAAAAAGTCGGTAGAGACAAAAG
TCGTTTCGTGGACGAAGCGACCAAAACTGAGCACAAGATA
...
>scaffold_3
CATAAATTCATAAATCAATTCATGAAGAATAATT
TAGAAAATTTCCAGGAAGTTTGAAGTTGCTA
...
```

```
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 `irsync`) 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.

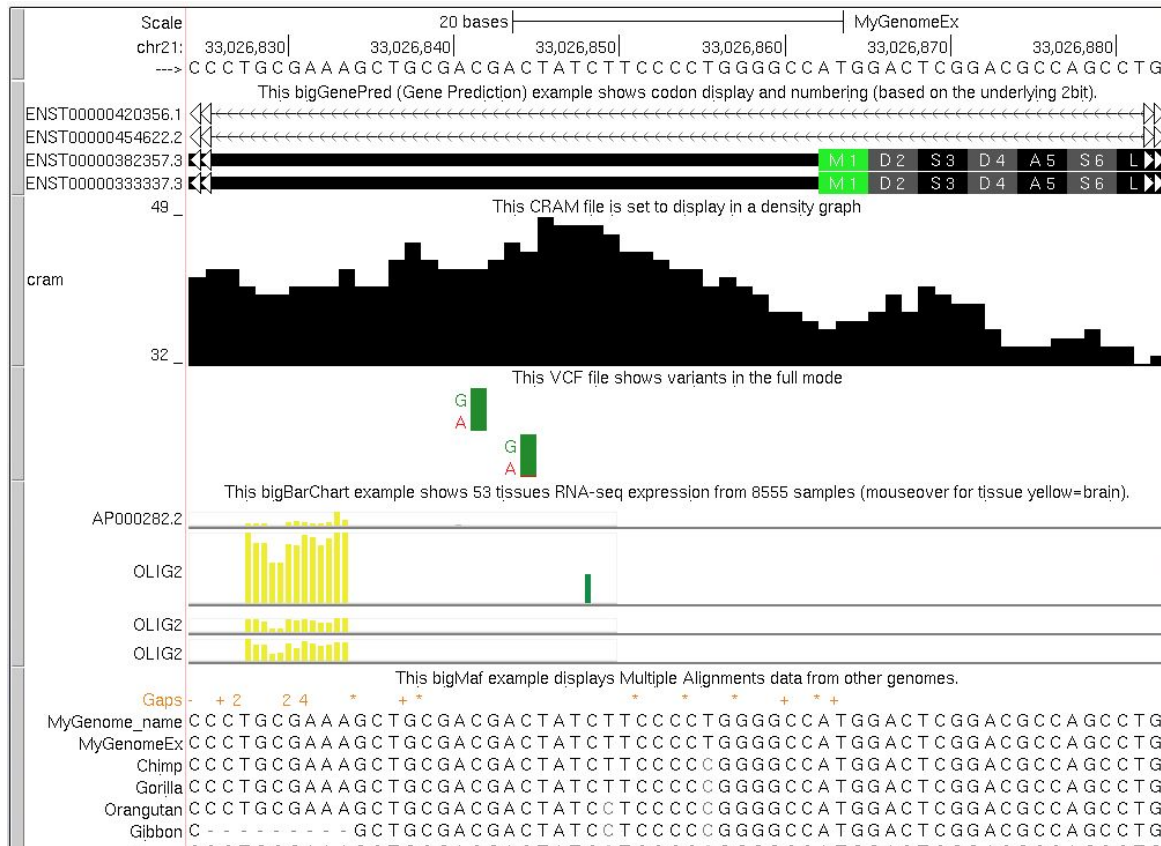


<https://de.cyverse.org/anon-files/>

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

```
$ 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

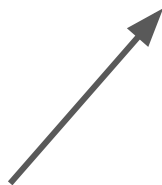
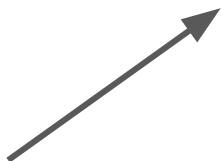
chr21 33050000 33060000

...

Additional Track Types

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

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

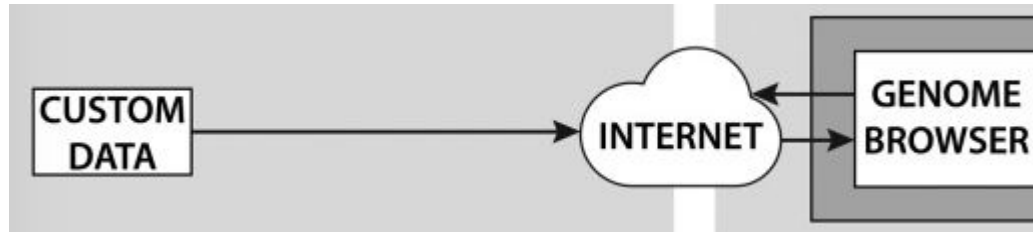


```
chr21 33030000 33040000
chr21 33050000 33060000
...
...
...
...
```

```
chr1 248956422
...
chr21 46709983
chr22 50818468
...
```

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

[bigDataUrl=http://CyVerse.path/out.bb](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:

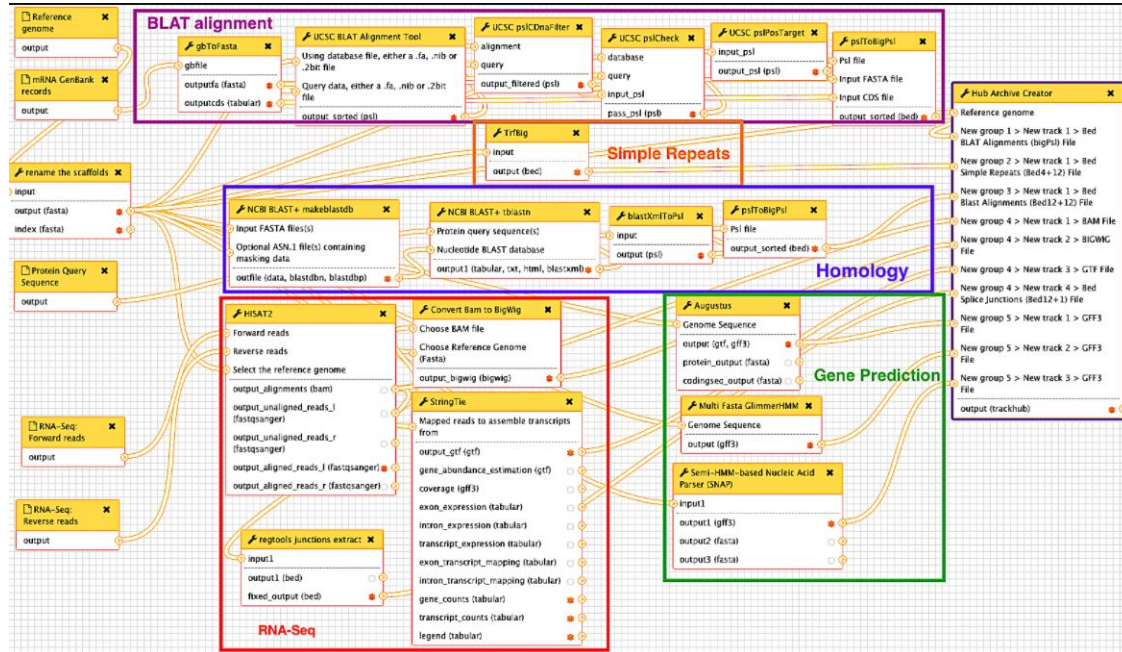
- BAM, CRAM, VCF
- bigGenePred, bigBarChart
- bigPsl, bigChain, bigMaf,
- bigNarrowPeak, halSnake
- bigBed, bigWig

Assembly Track Hub Resources

G-OnRamp

Galaxy workflow turning data like RNA-Seq into Assembly Hubs

<http://gonramp.wustl.edu/>





Thank You!



UNIVERSITY OF CALIFORNIA
SANTA CRUZ | Genomics
Institute

