

## Genome Browsers for interacting with Genomic data

Many different browser environments

1. [NCBI Genome Maize Maize chromosome](#)
2. [Ensembl](#)
3. [FungiDB](#)
4. [Mycocosm](#)
5. [Wormbase](#)
6. [FlyBase](#)
7. [Saccharomyces Genome Database](#)
8. [TAIR](#)
9. [Gramene](#) - Plant Comparative Resources
10. [Phytozome](#) - Plant Comparative Genomics portal
11. [UCSC Genome Browser](#)

## Setting up your own - JBrowse - Genome Browser

(some of this was first written by: Tania Kurbessoian [[@tania-k](#)](<https://github.com/tania-k>) )

To visualize genome annotation combined with Epigenomic, Transcriptomic, or Variant data you want to visualize them onto a genome browser. [JBrowse](#) provides an easy to setup tool for this visualization.

There is substantial [documentation](#) that describes installation and add-on features. The [FAQ](#) is also incredibly helpful.

## Setup JBrowse on HPCC

These steps will show you how to setup JBrowse on HPCC with some already installed systems to make it easier for you

### Configure your HPCC account to be able to share via HTTP / Web

First you need to configure your account to be able to share data via the web.

Follow the [directions on the HPCC manual](#) so that you can configure your home folder `~/.html` to be able to serve up data. If you do not want to make everything in this folder public you can use some simple strategies to enable a password protected space by [creating a .htaccess](#) file or you can use a web form of security - security through obscurity.

```
mkdir ~/.html/private
chmod 711 ~/.html/private
mkdir ~/.html/private/MYSEKR1TPROJECT
# install jbrowse or data or others in ~/.html/private/MYSEKR1TPROJECT
```

On the web a user browsing will not have permission to see <https://cluster.hpcc.ucr.edu/~YOURUSERNAME/private/MYSEKR1TPROJECT> but the URL <https://cluster.hpcc.ucr.edu/~YOURUSERNAME/private/MYSEKR1TPROJECT> will be visible. This is called security by obscurity, if you generate a long random string instead of MYSEKR1TPROJECT it would be hard to guess it (though note this not really secure since anyone reading network traffic and see the string would now know the the URL to go to).

Generally if you want to protect the data, setup a `.htaccess/.htpasswd` to require logging in. [Making an htpasswd file](#)

## Setting up your own copy of JBrowse software

The next directions are specific to the UCR HPCC. These instructions use an already build conda environment which you can link to:

```
# startup an interactive job on the cluster
srun -p short -N 1 -n 4 --mem 16gb --pty bash -l
mkdir -p ~/bigdata/jbrowse2
cd ~/.html/private
ln -s ~/bigdata/jbrowse2 .
cd ~/.html/private/jbrowse2
```

```
module load jbrowse/2
jbrowse create SARS-CoV-2
```

If you are going to support multiple JBrowse environments you only need to have a custom data folder. So you can symlink to all the files within the jbrowse checkout and then make a separate data folder too. Otherwise you need to make sure you have a separate custom jbrowse checkout for each project you are supporting.

To download the SARS genome and annotation from NCBI - this can be either in the folder you want to put the data or you can later symlink or copy from this folder

```
curl -o NC_045512.fna.gz https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/009/858/895/GCF_009858.1_SARS-CoV-2.fna.gz
curl -o NC_045512.gff.gz https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/009/858/895/GCF_009858.1_SARS-CoV-2.gff.gz
gunzip NC_045512.fna.gz
module load samtools
samtools faidx NC_045512.fna
# to load GFF we need to ensure it is sorted
module load bcftools
zgrep "^#" NC_045512.gff.gz > header
```

```

zgrep -v "^#" NC_045512.gff.gz | sort -k1,1 -k4,4n > NC_045512.sorted.gff
bgzip -i NC_045512.sorted.gff
tabix NC_045512.sorted.gff.gz

```

To load genome you have already put in the SARS-CoV-2 folder - you need to  
*GO INTO THE SARS-CoV-2 folder*

```

cd SARS-CoV-2
# if downloaded the data into the folder
jbrowse add-assembly NC_045512.fna --load inplace
# if you had a different folder for this
jbrowse add-assembly ../path/to/NC_045512.fna --load symlink

# if you forgot to create the index it will give you a message
# then you need to do
# samtools faidx NC_045512.fna
# if you created the gff and ran bgzip and tabix in this folder
jbrowse add-track NC_045512.sorted.gff.gz --load inplace
# if you had put this in another folder
# jbrowse add-track ../path/to/NC_045512.sorted.gff.gz --load symlink

```

To load VCF files (SNPs and variants)

```

jbrowse add-track ../path/to/SARS-CoV-2.vcf.gz --load symlink
# if there are warnings you need to build an index you can srn
# module load bcftools
# tabix ../path/to/SARS-CoV-2.vcf.gz
# then re-run the add-track
# if VCF file is in this directory
# jbrowse add-track SARS-CoV-2.vcf.gz --load inplace

```

To load BAM files, WIG files, or other gFF you can use same add-track. For  
 BAM, CRAM, files they need to have been indexed

```

module load samtools
samtools index SRR11140748.bam

```

To add this file you can do

```

jbrowse add-track ../path/to/SRR11140748.bam --load symlink
# or if the file was made IN this directory
jbrowse add-track SRR11140748.bam --load inplace
# if it gives you a warning about index file run
# samtools index BAMFILE

```

Now navigate to the web with you link based on your username and  
 folder - this would look like this but you need to put your username  
<http://cluster.hpcc.ucr.edu/~USERNAME/private/jbrowse2/SARS-CoV-2>

To see my example go to <http://cluster.hpcc.ucr.edu/~jstajich/private/jbrowse2/SARS-CoV-2/>