UCSC Genome Browser in the Cloud (GBiC)

Jairo Navarro Gonzalez, Galt Barber, Ana Benet-Pagès, Jonathan Casper, Hiram Clawson, Angie Hinrichs, Brian Lee, Chris Lee, Luis Nassar, Gerardo Perez, Brian Raney, Kate Rosenbloom, Daniel Schmelter, Ann Zweig, David Haussler, Robert Kuhn, Max Haeussler, Jim Kent

General Overview

The UCSC Genome Browser is a valuable resource for viewing and sharing genomic data with other researchers.

- On occasion, users would like to view sensitive information--such as patient data--or extremely large files.
- Slow internet connections can also impede genomic analysis.

In these instances, sending data over the internet is not feasible.

Why GBiC?

- Can be completely offline
  - No data sent over the internet
- Customization options
- Best performance when rendering local BAM/bigWig/bigBed files
- Relatively simple to install on a virtual machine or cloud instance --- just run the script

Supported Operating Systems

- Ubuntu 18/20 LTS
- Centos 7.2/8
- Fedora 20
- Amazon EC2 (Centos)
- Microsoft Azure (Ubuntu)

Installing a virtual box

Best to be used on a fresh install of an operating system. **Vagrant boxes** can quickly setup a virtualbox for the linux operating system.

```
vagrant init box/name    # this installs the vm and makes a Vagrantfile
vim Vagrantfile          # if need be
vagrant up               # starts the VM
```


Running the script

The first argument of the script is called the **command**. The first command that you will need to run is **install**, which installs the Genome Browser dependencies, binary files and basic MySQL (MariaDB) infrastructure:

```
sudo bash browserSetup.sh -b install
```



```
browserSetup.sh [options] [command] [assemblyList] – UCSC genome browser install script

command is one of:
   install    - install the genome browser on this machine. This is usually
                required before any other commands are run.
   minimal    - download only a minimal set of tables. Missing tables are
                downloaded on-the-fly from UCSC.
   mirror     - download a full assembly (also see the -t option below).
                No data is downloaded on-the-fly from UCSC.
   update     - update the genome browser software and data, updates
                all tables of an assembly, like "mirror"
   cgiUpdate  - update only the genome browser software, not the data. Not
                recommended, see documentation.
   clean      - remove temporary files of the genome browser older than one
                day, but do not delete any uploaded custom tracks
   addTools   - copy the UCSC User Tools, e.g. blat, featureBits, overlapSelect,
                bedToBigBed, pslCDnaFilter, twoBitToFa, gff3ToGenePred,
                bedSort, ... to /usr/local/bin

parameters for 'minimal', 'mirror' and 'update':
   <assemblyList>      - download MySQL + /gbdb files for a space-separated
                         list of genomes
```

Features are always being added to the GBiC installation script. For a more detailed tutorial, visit:
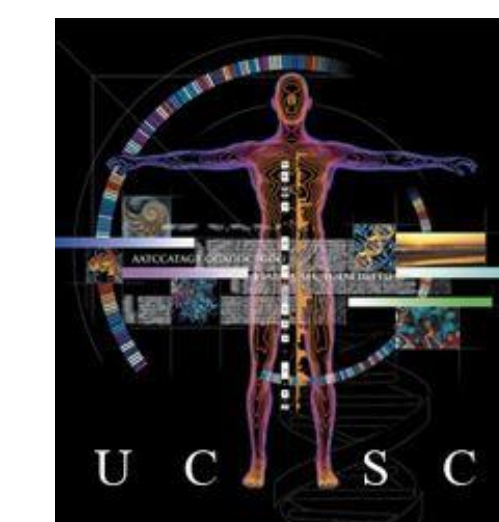https://github.com/ucscGenomeBrowser/kent/tree/master/src/product/installer

References

The UCSC Genome Browser database: 2021 update: Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, Powell CC, Nassar LR, Maulding ND, Lee CM, Lee BT, Hinrichs AS, Fyfe AC, Fernandes JD, Diekhans M, Clawson H, Casper J, Benet-Pagès A, Barber GP, Haussler D, Kuhn RM, Haeussler M, Kent WJ *Nucleic Acids Res*. 2021 Jan 8;49(D1):D1046–D1057. PMID: 33221922; PMC: PMC7779060

Acknowledgements

More Information

Have questions?
Send us a question on our public mailing list:
genome@soe.ucsc.edu

Genome Browser tutorials and videos:
http://genome.ucsc.edu/training/index.html

UNIVERSITY OF CALIFORNIA
SANTA CRUZ | Genomics Institute

CSHL Genome Informatics 2021   September 29st, 2021
Digital copy of this poster available at:
http://hgwdev.soe.ucsc.edu/~jairo/posters/CSHL_2021.pdf