

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)

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](#)

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
```

```
[10:47 PM] jairo@Macbook ubuntu20/ $ vagrant ssh
Welcome to Ubuntu 20.04.2 LTS (GNU/Linux 5.4.0-80-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of Tue 28 Sep 2021 09:48:41 PM UTC

System load:  0.41          Processes:      117
Usage of /:   2.3% of 61.31GB  Users logged in:  0
Memory usage: 14%          IPv4 address for eth0: 10.0.2.15
Swap usage:   0%

This system is built by the Bento project by Chef Software
More information can be found at https://github.com/chef/bento
vagrant@vagrant:~$
```

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
```

```
vagrant@vagrant:~$ sudo bash browserSetup.sh -b install
-----
UCSC Genome Browser installation
-----
Detected OS: linux/debian, bullseye/sid

This script will go through three steps:
1 - setup apache and mysql, open port 80, deactivate SELinux
2 - copy CGI binaries into /usr/local/apache/cgi-bin, html files into /usr/local/apache/htdocs
3 - optional: download genome assembly databases into mysql and /gdb

This script will now install and configure Mysql and Apache if they are not yet installed.
Your distribution's package manager will be used for this.
If Mysql is not installed yet, it will be installed, secured and a root password defined.

This script will also deactivate SELinux if active and open port 80/http.
```

```
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 + /gdb 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>

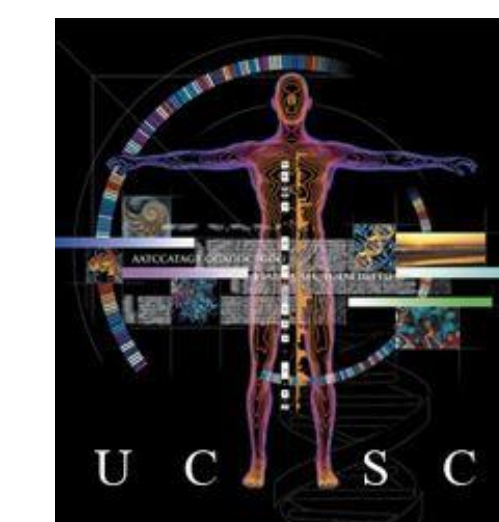
Acknowledgements

This work was funded by NHGRI award 4U41HG002371 to the UCSC Center for Genomic Science. We would like to acknowledge the work of the UCSC Genome Bioinformatics technical staff (<http://genome.ucsc.edu/staff.html>), our many collaborators, and our users for their feedback and support.

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>



CSHL Genome Informatics 2021 September 29th, 2021

Digital copy of this poster available at:
http://hgwdev.soe.ucsc.edu/~jairo/posters/CSHL_2021.pdf