UCSC Genome Browser in the Cloud (GBiC)

Jairo Navarro Gonzalez, Galt Barber, Anna 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 another researcher. In rare occasions:

- Users would like to view sensitive information--such as patient data--or extremely large files
  - Track and assembly hubs are not a feasible option to view the data (cannot send data over the internet)
- Slow internet connections can also impede genomic analysis

In these instances, sending data over the internet is not feasible. The Genome Browser in a Box (GBiB) is a great alternative to the GBiC script when you do not have a web server.

Why Choose GBiC?

- Can be completely offline
- IT admin cannot accept a virtual machine image (GBiB)
- Manually edit the script for customization options
- Best performance when rendering local BAM, bigWig, bigBed files (similar to GBiB)
- Relatively simple to install on a bare metal server, virtual machine or cloud instance -- just run the script
- Free for academic and personal use -- License required for commercial use

Supported Operating Systems

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

Manual Source Installation

- Not easy to setup, even for experienced Unix administrators (manual installation instructions)
- Install linux packages and update the linux distribution yourself
- Apply UCSC data updates yourself using rsync or MySQL table loads
- Configuration changes on our side may break your setup

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. A **minimal installation** of hg38 uses 5.8Gb of disk space. A **full mirror** of hg38 could require 5-6TB to store all genome annotation tracks.

```
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 /gbdb

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 + /gbdb files for a space-separated
                      list of genomes
```

- Features are always being added to the GBiC script
- For a more detailed tutorial, visit: https://bit.ly/3AJVRFN
- To download the GBiC script from the Genome Browser store, visit: https://genome-store.ucsc.edu/

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
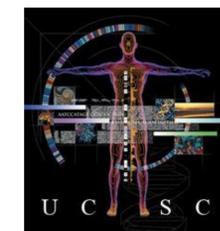
Acknowledgements

More Information

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

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

UNIVERSITY OF CALIFORNIA
SANTA CRUZ | Genomics Institute