

Cornell University
Life Sciences
Core Laboratories Center

BioHPC Lab at Cornell

Robert Bukowski

Computational Biology Service Unit

Life Sciences Core Laboratories Center
Cornell University

<http://cbsu.tc.cornell.edu/lab/doc/BioHPCLabIntro20120910.pdf>

Computational Biology Service Unit

- Biotech Core Facility providing services for a fee
- Computational biology and bioinformatics research services (including next generation sequencing support).
- Software and database development
- Hosting Cornell software
- Education and outreach
- Computational infrastructure
 - BioHPC (this is where the hands-on part of the workshop will take place)**

BioHPC Laboratory is a computational resource configured and optimized for the needs of computational biology

- Available to Cornell and external users
- Charged per hour of usage (free for GBS workshop exercises)
- For details and fee structure, visit

<http://cbsu.tc.cornell.edu/lab/lab.aspx>



Hardware/OS:

31 Linux remote workstations.

Dell PowerEdge M600 blade server, two quad-core Intel Xeon E5420 2.5GHz CPUs, 16GB RAM, 1TB HD local storage.

3 Linux interactive workstations (large monitors).

Dell Precision T3500n, Quad Core Intel Xeon W3550 3.0GHz CPU, HDD SATA controller with 2 2TB HDD, 24GB RAM, NVIDIA Quadro NVS 295 Graphics Card, 24in LCD monitor.

5 Linux large RAM remote workstation.

Dell PowerEdge R710 with four quad-core Intel Xeon E5520 2.27GHz, 1TB HD, 64GB RAM.
Red Barn Server with four 12-core AMD Opteron 6172 SKT G34 2.1GHz, 13TB HD, 512GB RAM.
3 Red Barn Servers with four 16-core AMD Opteron 6272 2.1GHz, 13TB HD, 512GB RAM.

File server.

Dell Precision T110, Intel Xeon X3430, 2.4 GHz, HDD SATA controller with 4 2TB HDD, 36 TB SATA enclosure, 8GB RAM.

1 Windows interactive workstation.

Dell Precision PWS T5400, Quad Core Intel Xeon E5410 2.33GHz CPU, HDD SATA controller with total of 10TB HDD storage, 16GB RAM, NVIDIA Quadro FX570 Graphics Card, 24in LCD monitor.

Software (partial list):

- Linux: 454 gsAssembler, ABySS , AllPaths, Annovar, ATSAS , bamtools, BEDtools, BioPerl, BioPython, BLAST, BLAT, Bowtie, Bowtie2, BWA, CBSU RNAseq, CLUSTALW, Cufflinks, FASTX, GATK, HyPhy, iAssembler, IGV, LUCY, MACS, MACS2, MAQ, MMSEQ, Netbeans, Novoalign, NovoalignCS, Oases, Picard, plink, prinseq, QIIME, R, RSEM, Samtools, SOAPdenovo , SOAPdenovo-Trans, stacks, tabix, TASSEL, TopHat, Trinity, vcftools, Velvet
- Windows: NextGENe (SoftGenetics), NGen (DNASTar), GenomeStudio (SoftGenetics).

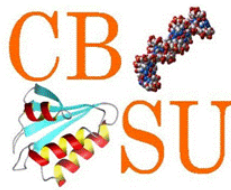
Up to date software list with additional information (version etc.) is online

<http://cbsu.tc.cornell.edu/lab/labsoftware.aspx>

GBS-related software used during the workshop is pre-installed on BioHPC machines.

To install on your own machine at home, visit
[*www.maizegenetics.net*](http://www.maizegenetics.net)

BioHPC Laboratory



All (but one) BioHPC Lab machines run Linux (CentOS 5 and 6)

Slides from previous Linux workshops are available online:

http://cbsu.tc.cornell.edu/lab/doc/Linux_workshop_Part1.pdf

http://cbsu.tc.cornell.edu/lab/doc/Linux_workshop_Part2.pdf

CBSU organizes “Linux for Biologists” workshops (two parts)
check our website for next installment or sign up for CBSU mail list

Workstations feature two kinds of storage:

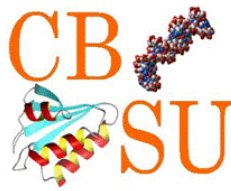
Network storage (storage of essential files only, **no computations!**)

/home/<your_id>
/programs
/shared_data

Local storage (all computations occur in directory
/workdir/<your_id>)

/workdir
/local_data

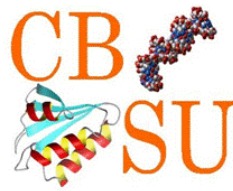
BioHPC Laboratory



Normally needed to access BioHPC workstations:

- **userID and password** (sent by e-mail upon account creation)
- **Credits** in Lab Credit Account (available for purchase)
- **Reserved machine(s)** – reservations handled via BioHPC Lab website <http://cbsu.tc.cornell.edu/lab/lab.aspx>

BioHPC Laboratory



Computational Biology
Service Unit

BioHPC Lab website allows for managing your account, make and manage reservations and connect to the workstations.

The screenshot shows a web browser window displaying the Cornell University BioHPC Laboratory website. The browser's address bar shows the URL <http://cbsu.tc.cornell.edu/lab/lab.aspx>. The website header features the Cornell University logo and the text "Cornell University Life Sciences Core Laboratories Center Computational Biology Service Unit". A search bar is present with the text "SEARCH CORNELL:" and a "go" button. Below the header is a navigation menu with items: HOME, Mission and Services, BioHPC, Staff, Publications and Outreach, CBSU/3CPG BioHPC Lab, Contact us, Forum, and Mail list. The main content area is titled "CBSU / 3CPG BioHPC Laboratory (625 Rhodes)" and contains a description of the lab as a joint facility established by CBSU and 3CPG. It mentions that the lab is targeted for biologists who want to learn Linux operations and that workstations will also serve as development and test machines for clusters and/or commercial cloud services. A limited number of commercial software licenses are maintained by the Life Sciences Core Laboratory Center (CLC). The lab is located in 625 Rhodes and its hardware infrastructure includes:

- 3 Linux interactive workstations: Dell Precision PWS T5400, Quad Core Intel Xeon E5410 2.33GHz CPU, HDD SATA controller with total of 10TB HDD storage, 16GB RAM, NVIDIA Quadro FX570 Graphics Card, 24in LCD monitor.
- 1 Windows interactive workstation: Dell Precision PWS T5400, Quad Core Intel Xeon E5410 2.33GHz CPU, HDD SATA controller with total of 10TB HDD storage, 16GB RAM, NVIDIA Quadro FX570 Graphics Card, 24in LCD monitor.
- 31 Linux remote workstations, Dell PowerEdge M600 blade server, two quad-core Intel Xeon E5420 2.5GHz CPUs, 16GB RAM, 1TB HD local storage. Most of these workstations are available in "remote workstations" category, but some may be placed in "restricted" category if needed for collaborative projects or workshops.
- 2 large RAM memory workstations: Dell PowerEdge R710 with four quad-core Intel Xeon E5520 2.27GHz, 1TB HD, 64GB RAM (cbsulm01 in "restricted" category). Red Barn Server with four 12-core AMD Opteron 6172 SKT G34 2.1GHz, 13TB HD, 512GB RAM (cbsulm02 in "Restricted category").
- File server: Dell Precision T110, Intel Xeon X3430, 2.4 GHz, HDD SATA controller with total of 36TB HDD, 8GB RAM.

The software installed on the workstations is described in detail on our [lab software page](#).

The browser window also shows several open tabs: "10 Day We...", "Ithaca NY...", "Ithaca Cle...", "CBSU @ C...", and "CBSU ...". The status bar at the bottom of the browser shows the URL <http://cbsu.tc.cornell.edu/lab/lab.aspx> and a zoom level of 100%.

Reservations for this workshop have been pre-arranged, workstation assignments are published on the workshop web pages:

http://cbsu.tc.cornell.edu/lab/doc/workstations_20120913.htm

No fee for workshop exercises

Connecting to workstations

Detailed access instructions:

http://cbsu.tc.cornell.edu/lab/doc/wrkshp_CBSUlab_access.pdf

In a nutshell:

1. **ssh client (use your user ID and password)**

i. **on Windows: PuTTY**

(<http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html> - just download the **putty.exe** file, put it anywhere on your hard drive, and double-click to launch)

ii. **on Mac or Linux:** native command-line ssh client: open terminal window and type **ssh my_id@my_machine**

2. **VNC** – for access to Linux graphical desktop (we will **NOT** use this during workshop).

Connecting from outside Cornell network (hotel room):

VPN (Cornell ID required)

OR (no Cornell ID required)

1. Connect to **cbsuss02.tc.cornell.edu** (same userID and password)
2. once on cbsuss02, ssh to your assigned machine from command line, e.g., :
ssh cbsum1c1b005

Transferring files to and from workstations: **sftp**

1. **FileZilla** (<http://filezilla-project.org>)
2. **WinScp** (<http://winscp.net/eng/index.php>)
3. **sftp** (native command-line clients on Linux and Mac)
4. **Fetch** (Mac)

Connect to your assigned workstation using your user ID and password an one of the sftp clients.