

COMPUTATIONAL BIOLOGY SERVICE UNIT, 3CPG

RNA-Seq

CBSU Computational Resources for the Workshop

Robert Bukowski (bukowski@cornell.edu); Jaroslaw Pillardy (jp86@cornell.edu)

6/27/2011

In this edition of the workshop, the hands-on exercises will be hosted on dedicated workstations of the CBSU/3CPG BioHPC Lab (<http://cbsu.tc.cornell.edu/lab/lab.aspx>). Essentially, the Lab is a collection of 36 powerful Linux (CentOS 5) workstations with a common authentication system and shared user home directories. Three of the workstations may be accessed via high-end graphical consoles located in 625 Rhodes (contact us for door code if you need access to this room). All workstations may be accessed remotely using any SSH client (see “Install prerequisite software on your PC” section for details). In order to use the workstations, you need to have an account at CBSU Lab and have an ssh client available on your laptop. The workstations can be also access via a VNC client, providing “remote desktop” capabilities.

If connecting from outside of the Cornell network, you will first need to establish a VPN connection. See section “Install Prerequisite Software” for more details.

Below we describe steps you have to take to use the Lab resources during the workshop.

Confirm your intent to perform the hands-on exercises

Send e-mail to cbsu@cornell.edu expressing your interest in completing the hands-on part of the workshop. If you do not yet have an account on CBSU/3CPG Lab machines, we will create one for you (once your new account is created, you will receive a “welcome e-mail” with your user ID and temporary password, as well as instructions on how to make a reservation – you can ignore the reservation part of the letter, as the special reservation will be created for you for the workshop). Also, we will assign a machine for you to work on. Please check the workshop web page <http://cbsu.tc.cornell.edu/ww/1/session.aspx?wid=9&sid=12> for machine allocations.

Install prerequisite software on your PC

These prerequisites are described at http://cbsu.tc.cornell.edu/lab/doc/Remote_access.pdf. In short, to access the Lab machines, you will need the following software:

SSH client

The primary way of accessing the Lab machines is to log in by opening an SSH (Secure Shell) terminal window using an SSH client. An ssh client we recommend for Windows is **Putty** (<http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>). Just download the putty.exe file and put it anywhere on your hard drive.

If you are connecting from a MAC or a Linux desktop, you can just use the native ssh client launched from within your machine’s terminal window.

Java (optional)

Java is required for the default VNC client.

Xming (optional)

The ssh window is just a text terminal – you can run all workshop commands in it, but to run any graphical software, such as IGV viewer, you will need an X-windows manager (such as **Xming**) running on your windows PC. **Xming** can be downloaded from <http://www.straightrunning.com/XmingNotes/> - Download and install **xming - mesa** (click the Xming - mesa link under the Public Domain Releases table).

If you are connecting from a MAC or a Linux desktop, any additional software is not needed; simply use the “-Y” option while opening an ssh connection in machine’s terminal window.

SFTP client for file transfers (optional)

Any data transfer between your PC and the workshop machine is not necessary to complete the workshop. However, you may wish to transfer some result files to your PC to analyze them later, open them using your favorite spreadsheet program, etc. In such a case, you will need an SFTP client installed on your PC. For Windows PCs, we recommend **Filezilla** (<http://filezilla-project.org/>) or **WinScp** (<http://winscp.net/eng/index.php>).

On a MAC, you can use the file transfer program called **Fetch** or just use the command-line sftp client from within your MAC’s terminal window. The native command-line sftp client is also available on Linux desktop computers.

VPN (optional)

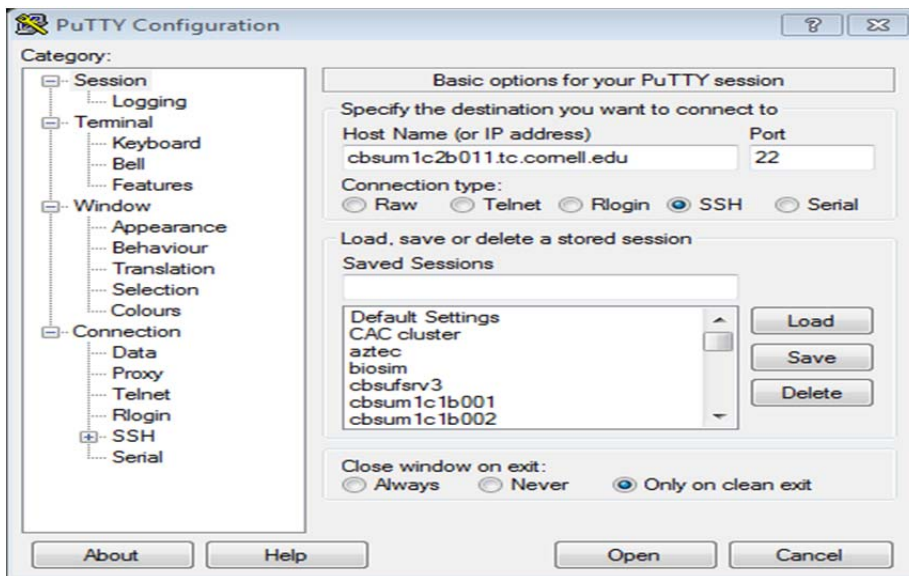
If you anticipate connecting to the Lab machines from outside of the Cornell network (e.g., from home), you will need to first establish a VPN (Virtual Private Network) connection – this effectively makes your computer a part of Cornell network regardless of where you are. CIT offers compatible VPN clients for all platforms (<http://www.cit.cornell.edu/services/vpn/howto/install.cfm>).

Use your Lab workstation

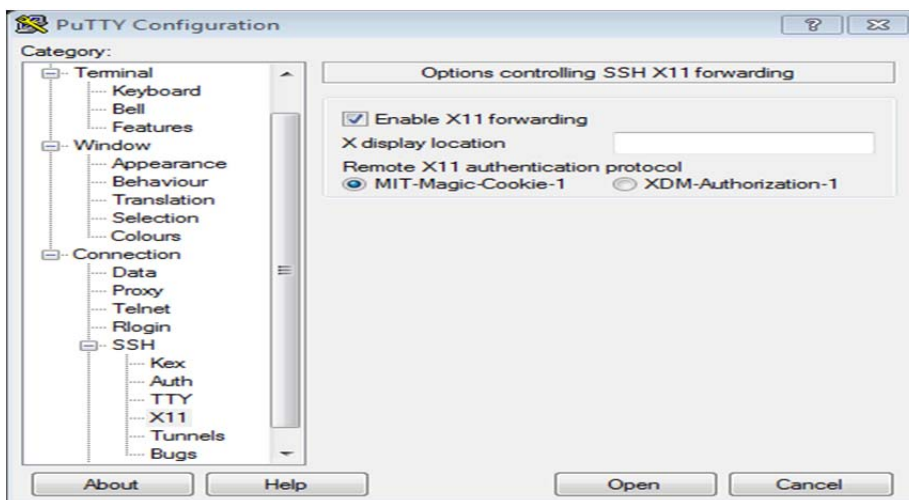
In this Section, we explain how to connect to your assigned machine from your PC/MAC, how to configure your machine for the workshop assignment, how to work on your machine, and how to release it when no longer needed.

Connect with SSH

On a PC, launch the SSH client by double-clicking on its icon. If you are using **PuTTY**, you will see a window similar to:



In the Host Name box, enter the name of your assigned workstation (in the example above, it is cbsum1c2b011.tc.cornell.edu; yours will be different!). Expand the SSH entry (bottom right), select “X11”, and check “X11 forwarding”:



Click "Open". If you see a warning message, click "OK". You should then see a terminal window. Enter your user ID on the login prompt, then password. You can open additional SSH terminal windows by repeating the procedure. See http://cbsu.tc.cornell.edu/lab/doc/Remote_access.pdf for information on how to configure details of the **PuTTY** client.

If you installed **Xming** on your machine (this is an X-Windows manager for Windows PC), start it now (by clicking on its icon). The program will start in the background. You will need **Xming** running if you want to use the IGV viewer to visualize alignments.

If you are connecting from a MAC or from a Linux machine, simply open a terminal window on that machine and type

```
ssh -Y <your_id>@cbsum1c2b011.tc.cornell.edu
```

replacing <your id> with your own Lab user name and replacing 'cbsum1c2b011' with the name of the machine that has been assigned to you.

Connect with VNC

You can access your Linux workstations using VNC protocol and operate remotely in a graphical desktop environment. You need to go to "[My Reservations](#)" page (<http://cbsu.tc.cornell.edu/lab/labresman.aspx>), and click on "Connect VNC" for a workstation you want to connect to. It will launch a new browser window that will load a VNC Java applet - your browser and computer must support Java. If your password is not yet stored in VNC you will be asked first to enter it before the applet loads. The applet opens a Java application window, please click "OK" on "Connection Details" form and then type in your password in "VNC authentication form" (user name is blank, and it should be!). If your password is correct a Linux desktop will open in the Java applet window.

Configure your reserved machine for the workshop

If you are connecting to your reserved machine for the first time, you will need to prepare it for the workshop by first creating your own subdirectory on the machine's local scratch disk.

In the terminal window, type the following commands (replacing rb299 with your own user ID and hitting "Enter" after each command):

```
cd /workdir
mkdir rb299
```

These commands create the scratch directory `/workdir/rb299`. This directory is located on a hard drive physically attached to the machine and thus suitable for doing calculations on it thanks to fast access (as opposed to your home directory – see below).

Working on your Lab workstation

The workshop exercises are performed by typing commands in the SSH terminal window (you can open multiple terminal windows simply by logging in multiple times using your SSH client). Here are several hints to remember:

- During the exercise, you can close your SSH window at any time (by entering ‘exit’ command or pressing Ctrl-D). Doing so will not terminate any of your running programs provided that these programs were launched in the background (i.e., with the ‘&’ at the end of the command). You can return to your machine at any time simply by launching the SSH window again. Once you reconnect, you should see all your programs still running, unless they finished in the meantime while you were gone.
- Similarly you can close your VNC session window at any time, you can reconnect to the session at any time the same way you launched it.
- Important note about storage on the workstations. The machine features two kinds of storage:
 - **Your home directory (`/home/<your_id>`)** and its subdirectories are physically located on a different machine (a file server) and **network-mounted** on all workstations. Thus, whichever workstation you log in to, you always see the same home directory and the data stored there is preserved. However, because it is network-mounted, access to the home directory is slow, so you should not run any calculations there.
 - the local scratch located in `/workdir`. This space is physical attached to the machine and therefore access to it is fast. Any calculations should be run there (rather than in the home directory). Create your own subdirectory (e.g., `/workdir/<your_id>`) to work in.
 - Remember: `/workdir` should be **treated strictly as scratch space**, i.e, files created there are **not guaranteed** to be preserved after the workshop (more generally, when your reservation expires). After completing the calculations, you will have to copy any important result files you wish to save to your **home directory**, where they will be preserved and accessible from other workstations.
- The workshop assignments do not require any file transfer between your PC and the Lab machine. However, if you would like to examine some output files using your own computer, you can fetch these files using any SFTP client. Please consult the 3CPG website for advice on SFTP clients. Use your Lab machine’s name (the same one used in the ssh client/command) with your Lab user ID and password to connect.

Terminate your Lab workstation session

Once your calculations are completed, remember to copy any important result or intermediate files to your home directory. Later, when you make another reservation and log in to another workstation (possibly different from the one you worked on before), you will see these files and will be able to continue working on them (first, you will need to create the `/workdir/<your_id>` directory on this new workstation and copy your files from your home to that directory).

After saving all the important files in home directory, clean up your scratch directory `/workdir` to make space for other users. For example, you can remove it completely with the following commands:

```
cd /workdir
```

```
rm -Rf /workdir/<your_id>
```

(replace `<your_id>` with your actual Lab user name). Then simply close all SSH terminal windows (for example, by typing 'exit' in each of them).