

Getting Up and Running with CLC Genomics Workbench at NC State University

Introduction: This resource guide is intended to help NC State users of CLC Genomics Workbench access and utilize this powerful software suite. CLC Genomics was acquired with funds provided by ORIED as well as participating groups of faculty that lobbied for its acquisition. As such, the purchase of CLC Genomics is the result of a partnership between ORIED and NC State faculty and this software is licensed only to these users. *Laboratories lacking a site license may negotiate limited, trial use of CLC Genomics with the Director of the GSL or participating faculty willing to sponsor availability of the software.* The intention of such negotiations for limited, trial use is to enable non-licensed users to determine if they will purchase a license for their laboratory.

Getting Up and Running: This document details how to contact a network administrator, setup and install CLC Genomics software, and utilize a CLC Genomics Server managed by the NC State Office of Information Technology (OIT) for more powerful analysis. It is important to note that only a single network user can use the CLC Genomics Workbench license in the working mode (i.e. non-viewing), regardless of whether you are connected to the larger Genomics Server or not. Users should be aware of their use time and adhere to scheduling rules specified by your network contact. *Use of the CLC Genomics server is strongly advised, as it will reduce the time each user will tie-up the network license,* as once are jobs are submitted to the Server, the user is free to log off and close CLC Genomics workbench, while their jobs continue to run.

Step 1. Determine your Network and Contact Person

Each CLC Genomics network is locally managed and administered by participating laboratories, Departments, or Centers that dedicated funds to the license purchase. **Eligibility to use these Networks, and how they are administered, is at their discretion.** The GSL hosts support resources for CLC Genomics on its website, *but is not directly involved in managing these networks.* All queries about CLC use should be directed to the Network Contact Person (see below).

- 1a)** Visit the GSL resource webpage for CLC Genomics Workbench to identify participating NC State CLC network groups.

[Link: <https://research.ncsu.edu/gsl/bioinformatic-resources/clc/>]

- 1b)** Select a network. This will provide a contact person and email, a list of participating labs, and any additional Plugins that are available (see Section 3).
- 1c)** Contact the person managing the network. If you are eligible to use it, the contact person will email you an IP address for access. Save this IP address for Step 2.

Step 2. Download and install CLC Genomics Workbench software.

CLC Genomics Workbench is freely downloadable from Qiagen Bioinformatics, however connection to a licensed network or a license key (permanent or temporary) is necessary to fully run the software. Without license or network access, the software runs in limited mode only.

2a) CLC software can be downloaded from the following website:

<http://download.clcbio.com/browse.php>

[scroll to **clcgenomicswb** and select **22.0.2** from the bottom]

It is recommended to only use CLC Genomics Workbench v23, as it is the version compatible with the CLC Genomics Server installed on the HPC. Specifically, CLC WB version 23.0.2 has been tested and validated.

2b) Select the appropriate download link for your computer's operating system. Fill out the respective trial information form (name, email, etc.). You may need to accept cookies from Qiagen in your web browser to download the installer.

2c) Ensure that you are logged into your desktop or laptop computer as an Administrator and that you open CLC Genomics or Biomedical Workbench as Administrator (on a Windows PC, right click on the CLC Genomics Workbench icon and select "Run as administrator").

If you own the computer you are logging into or have specifically been given ADMIN rights to your computer, then you can login without any further permissions.

If NCSU owns your computer then you will need to either get the IT department to give you permanent ADMIN access or give you temporary admin access while you install.

Note: If you DO NOT have ADMIN rights, you will be prompted with the following picture when running the installer. Contrast this picture with the one in Section 2e.



You need a license...
In order to use this application you need a valid license.
Please choose how you would like to obtain a license for your workstation.



You do not have sufficient privileges to request a license.
Please start the Workbench from an administrator account to request your license.

2d) Click on the downloaded CLC Genomics Installer package and follow instructions for the Setup Wizard.

2e) Once the software is installed, a License Manager Window will open. To use a Network License, select the option "Configure License Server Connection."

You need a license...

In order to use this application you need a valid license.
Please choose how you would like to obtain a license for your workbench.

- ☐ **Request an Evaluation License**
Choose this option if you would like to try out the application for 30 days.
Please note that only a single evaluation license will be allowed for each computer.
- ☐ **Download a License**
Choose this option if you have a License Order ID and would like to download a license.
- ☐ **Import a License from a File**
Choose this option if you have a License File on your computer and would like to import it.
- ☐ **Upgrade from an existing Workbench**
Choose this option if you have an older version of this workbench with a commercial license, and would like to upgrade your license.
- ☒ **Configure License Server Connection**
Choose this option if your company or institution is using a central CLC License Server. This option also enables you to disable a license server connection.

2f) Check “Enable license server connection”. Then, select the “Manually specify license server” radio button option.

Configure License Server connection...

Please choose how you would like to connect to your CLC License server.

☒ Enable license server connection

☐ Automatically detect license server.

☒ Manually specify license server:

Hostname/IP-address:

Port:

☐ Use custom username when requesting a license

Username:

☐ Disable license borrowing

If you choose this option, users of this computer will not be able to borrow licenses from the License Server.

2g) Paste the IP address supplied by your Network Contact Person, and select the appropriate Port number. The default Port 6200 will likely be the correct one, unless specified differently by the network contact person.

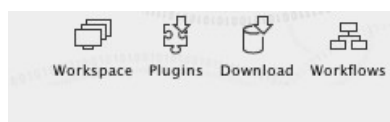
2h) Select “Finish.”

2i) To verify successful network connection select “Help” from the top drop-down menu and select “License Manager.” A license key should now be visible in the License Overview window.

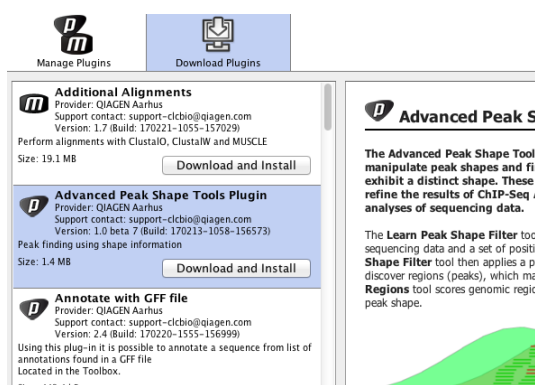
Step 3. Installing Additional Plug-ins for CLC Genomics

CLC Genomics Workbench has enhanced utility through Plug-in modules that allow the user to tailor the software to suit their particular application needs. Many Plug-ins are freely available (e.g., Advance RNAseq analysis, Bisulfite Sequencing, etc.), however some do require extra licenses that may and may not be available for your network. Check your Network Contact info on the GSL CLC support webpage (link: <https://research.ncsu.edu/gsl/bioinformatic-resources/clc/>) to see what additional “cost extra” Plug-ins may be available.

3a) Open CLC Genomics Workbench and select the “Plugins” icon from the top right corner of the viewing Window.



3b) Select the Download Plugins tab. Then select the desired Plugin.



3c) Click Download and Install. You may be prompted to fill out a registry form.

3d) Accept the Licensing Agreement Terms.

3e) If the network you are using have these Plug-in licenses available, no further steps are needed. The Plug-in will be available after restarting CLC Genomics. If the Plug-in licenses are not available for your network, a separate license key, purchased from Qiagen, will be required. *Note: many Plug-ins are available for a free 14-day trial.* You can edit your Plug-in through the Manage Plug-ins tab.

Step 4. Connecting to the CLC Genomics Server for Job Submissions

This function allows users to submit jobs to the computing cluster managed by the NC State Office of Information Technology for faster processing of sequence analysis jobs. **It is highly recommended that users utilize the CLC Genomics Server for processing jobs in order to minimize network use.** Currently, only **ONE USER CAN USE THE NETWORK LICENSE AT ANY ONE TIME**, however users can quickly set up a job run, submit it to the server for processing, and then log off the network by closing the desktop application, thus freeing the license for another user. CLC Genomics Server jobs are stand alone and will run independently, and users can check back on their job status at a later date.

- 4a) To get set up for using the CLC Genomics Server, send an email to oit_hpc@help.ncsu.edu to request a folder and account setup. **In the email, please carbon copy (CC:) your CLC Network contact – they will need to verify you are a member of their network.** Once verified, the CLC Genomics Server administrator will set up an account for you with a folder on the home directory for data storage and running your jobs – *only you have access to your folder.*
- 4b) To Login to the CLC Genomics server, you must be connected to the internet **and** connect to the NCSU VPN first. Only users who are on campus with a hardline ethernet connection to the NCSU network can connect to the Genomics Server without VPN. For what software to install and how to connect to the NCSU VPN, follow the instructions provided on this website:

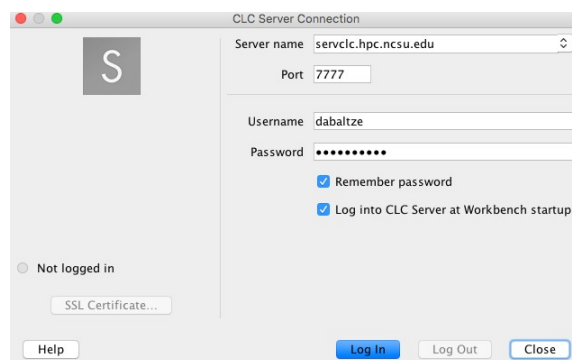
<https://oit.ncsu.edu/campus-it/campus-data-network/vpn/>

You will need your Unity ID/Password and duo authentication code to connect to NCSU VPN.

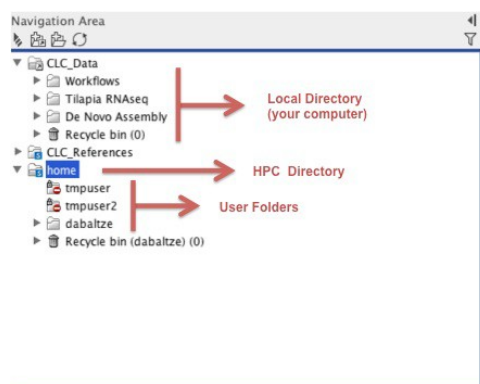
- 4c) After you have connected to NCSU VPN, you can then connect to the CLC Genomics server in your CLC Genomics Workbench. In the Workbench, click on “File” at the top left corner and select “CLC Server Connection.” Enter your NC State Unity username and password. Enter the following information for Server name and Port:

Server name: servclhpc.ncsu.edu

Port : 7777



- 4d) Click “Login”. This should log you into the CLC Genomics Server, which can be seen in the Navigation Area in the top left panel (home folder with small blue “s”). Double clicking this icon in the Navigation Area will open a list of folders, with one containing your username. You will have access to this folder only, as noted by the “lock” icon present on the other folders.”

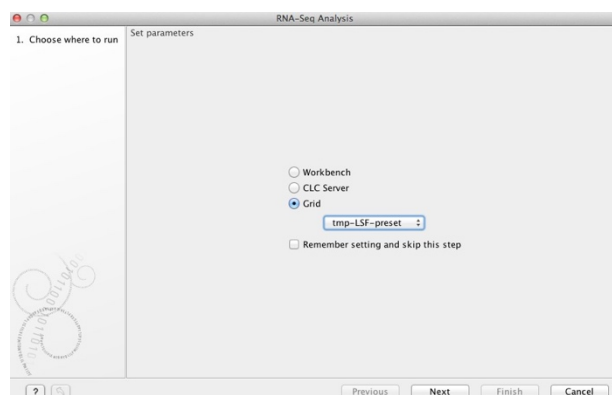


4e) Additional Considerations:

Choose Where to Run options

When submitting jobs on the HPC, please select the “Grid” radio button when prompted to choose where to run the job. Additional Grid preset selections (e.g., 128GB) may be available to optimize your run parameters based on the size of your data.

If “Workbench” is selected, the run will be performed on your local computer, tying up the network license. Additionally, several CLC Genomics Workbench analyses write temporary data files to the default temporary directory that use a lot of disk space. If you do intend to use the workbench on your local computer and would like more information on these temporary files see the “Changing the temp directory” section of the CLC Genomics Workbench manual linked below (see: 5c below). Additional Grid preset selections may be available to optimize your run parameters.



When running jobs on the server, you will need to select the file locations for references and or data files that have the small server icon on them (references files and data files locations with the same name will have duplicate locations, one for your local client file structure and another for your servers file structure). You may need to copy or “drag over” these files to your server folder from your local directory.

Backend Data Transfer to the HPC System

Depending on the size of your data, you may need to import/export data to/from your CLC Server folder from the backend (i.e., without using the CLC Workbench on your computer).

Note: Data transferred by Backend Data Transfer is accessible to all CLC users and therefore confidential data should not be transferred in this way.

Backend data transfer is only needed for transferring very large datasets. If you are not sure whether backend data transfer is appropriate for you, then you can send an email to Qiagen Tech Support (ts-bioinformatics@qiagen.com) to request recommendations. If you need to do backend data transfer then you will need to acquire an HPC account first and then request to be added to the “clcdata” group. Please visit the link below to see how to

acquire an HPC account.

<https://projects.ncsu.edu/hpc/Accounts/GetAccess.php>

After you have acquired an HPC account you can send email to oit_hpc@help.ncsu.edu to request to be added to the “clcddata” group. Upon addition to the “clcddata” group a subfolder for you will also be created under the location `/roi_share/clcbio/import-export` on HPC system.

(1) Backend Data Movement (SFTP protocol for PC, scp protocol for Mac)

On a Windows PC:

Download WinSCP

(link: <https://oit.ncsu.edu/campus-it/web-services/unity-file-space/winscp/>)

or MobaXterm

(link: <https://projects.ncsu.edu/hpc/Documents/mobaxterm.php>)

through NCSU) and follow a SFTP file protocol to transfer data.

For example, using WinSCP:

Login to `login.hpc.ncsu.edu`
Using NCSU Unity ID and NCSU Unity password

Highlight all files you want to upload, right click, and upload to:
`/roi_share/clcbio/import-export/<your username>/`

On a Mac:

Upload:

```
scp foo [Unity ID]@login.hpc.ncsu.edu:/roi_share/clcbio/import-export/<your
username>/foo
```

Download:

```
scp [Unity ID]@login.hpc.ncsu.edu:/roi_share/clcbio/import-export/<your
username>/foo foo
```

After files have been uploaded to `/roi_share/clcbio/import-export/<your username>` then login to HPC login node “`login.hpc.ncsu.edu`” and then do the following commands so that the data is available for import into the CLC Server folder:

```
chmod -R g+rwX /roi_share/clcbio/import-export/$USER/*
chgrp -R clcddata /roi_share/clcbio/import-export/$USER/*
```

(2) Importing/Exporting data to/from the CLC Server

Import:

Launch CLC Workbench -> Login to CLC Server -> Click on the "Import" icon near the top left -> Select the import scheme that you need -> Select the "Grid" button and use the first or any LSF preset -> Click "Next" -> Select the "On the server or a place the server has access to" button -> Click "Next" -> In the left-side "Server files" window, go to folder import-export -> select your [Unity ID] -> Select the files you want to import to your CLC folder and click on the right arrow to place them in the right-side window -> Click "Next" -> Select a folder in your CLC folders to import the files into the folder.

Once data import is done you can use the data to run jobs in the CLC Server using the files.

Export:

Launch your CLC Workbench -> Login to CLC Server > Click "Export" -> Select the appropriate export format (e.g, "Text") -> Click "Select" -> Select "Workbench" -> Click "Next" -> In the left-side window, go to one of your CLC folders from which you want to export files from -> Select the files you want to export -> Click the right arrow to place them in the right-side window -> Click "Next" -> Choose the directory under import-export to export files to in your own directory (your [Unity ID]).

After the export is done then you can use WinSCP on your Windows PC or use the command scp on your MAC to download the data from HPC system to your personal computer.

Step 5. CLC Genomics Tutorials and Support

5a) **Tutorials** - To get familiar with the software and analysis steps, it is recommended to first try a couple tutorials to learn how to utilize the Navigation Area, Toolbox, and Processes tabs. A list of these tutorials, along with sample sequence data, is available at <http://www.clcbio.com/support/tutorials/>.

5b) **Support** – For user help and support, the following resources are available:

NC State CLC Network Access – Contact your network administrator. A list of CLC networks and the contact persons are available at the link below:

<https://research.ncsu.edu/gsl/bioinformatic-resources/clc/>

CLC Genomics Software Support – For help in using CLC Genomics software, contact Qiagen Technical Support at <http://www.clcbio.com/support>. This support can also be utilized with CLC Genomics under the "Help" section.

NC State High Performance Computing – OIT support staff can address processing issues for the CLC Genomics server only. Information on how to contact this team can be found at <https://projects.ncsu.edu/hpc/About/Contact.php>. The contact email at this link will generate a Help ticketing system request.

5c) **Manuals** – For user help and support, the following resources are available

CLC Genomics Workbench Manual

[http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/754/index.php?manual=Introduction CLC Genomics Workbench.html](http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/754/index.php?manual=Introduction%20CLC%20Genomics%20Workbench.html)