



Deployment Manual

CLC Workbenches

Manual for
CLC Workbenches: deployment and technical information, version 1.2
Windows, Mac OS X and Linux

August 10, 2009

CLC bio
Finlandsgade 10-12
DK-8200 Aarhus N
Denmark



Contents

1 Introduction	5
1.1 Deployment strategies	6
2 Installation	7
2.1 System requirements	7
2.2 Available installers	8
2.2.1 Java	8
Java on Mac	8
2.2.2 Overview of available installers	9
2.3 What does it do?	9
2.3.1 Extracting and copying files to the installation directory	9
2.3.2 Setting the amount of memory available for the JVM	9
2.3.3 Shortcuts and file associations	10
2.4 Silent installation	10
3 License	11
3.1 License server set-up on clients	11
4 Plug-ins and resources	13
5 Storing and backing up data	14
5.1 Data structure	14
5.2 Back-up of data	15
5.3 Special configurations for large amounts of data	15
5.3.1 Temporary data	15
5.3.2 Disk space requirements	15

6 Overview - where do we put things?	17
6.1 Computer-level information	17
6.2 User-level information	17
Index	18

Chapter 1

Introduction

If you are in charge of installing and maintaining CLC Workbenches in your organization, you probably have a lot of questions about installers, licenses, and where do we put files for this and that.

This manual is written to answer these questions. It is primarily aimed towards client installations (i.e. the CLC Workbenches). For installing the *CLC Bioinformatics Database* and *CLC Genomics Server*, please refer to the installation chapter in the respective user manuals (see <http://www.clcio.com/usermanuals>).

At the moment, the CLC Workbenches are:

- *CLC Main Workbench*
- *CLC Protein Workbench*
- *CLC DNA Workbench*
- *CLC RNA Workbench*
- *CLC Genomics Workbench*

In addition, there is the *CLC Sequence Viewer* which is very similar to the other Workbenches except when it comes to licensing (it is free) and plug-ins (some plug-ins cannot be installed in the *CLC Sequence Viewer*).

In the first part of this manual, we take a closer look at the installer: what does it do, what are the options etc. Then we go into details of the license system, followed by an explanation of the concept of plug-ins. Finally there is an overview chapter which tells you where to find all the different files in a client installation. This is useful when deploying the Workbench but also for back-up purposes.

We will not go into details of Linux installations, but if you need these details, please contact us.

Questions, comments and feedback on this manual are very welcome at support@clcbio.com.

1.1 Deployment strategies

Deployment strategies should be developed so that they fit your existing IT set-up. However, the following solution has proven to work in many cases. Note that this is a brief summary of the steps to take, and that the rest of this manual is devoted to more detailed information.

1. On a single computer with the same OS as the target computers, complete a full installation including:
 - Running the installer
 - Set up license server connection ¹
 - Install plug-ins and resources
 - Specify proxy server information if needed
2. Run the installer in silent mode (-q) on all the target computers
3. Copy the following files from the installation in 1) to the target computers
 - `license.properties` and `proxy.properties` from the `settings` folder in the installation directory
 - `plugins` and `resource` folder from the installation directory

We recommend you tailor this strategy to your own organization - this is just an example of how it can be done.

Creating a GHOST image or similar to copy to all the target computers is also a possibility, but there are a few routines performed by the installer that need to be taken into account, especially allocating memory (see section 2.3.2) and the creation of shortcuts and file associations.

¹If you do not use a license server, you will have to activate licenses on each computer

Chapter 2

Installation

This chapter deals with the installer and related information about the installation process.

2.1 System requirements

The system requirements of the CLC workbenches except the *CLC Genomics Workbench* are these:

- Windows 2000, Windows XP or Windows Vista
- Mac OS X 10.4 or newer
- Linux: Redhat or SuSE
- 32 or 64 bit
- 256 MB RAM required
- 512 MB RAM recommended
- 1024 x 768 display recommended

The requirements for the *CLC Genomics Workbench* are:

- Windows 2000, Windows XP or Windows Vista
- Mac OS X 10.4 or newer
- Linux: Redhat or SuSE
- 256 MB RAM required
- 512 MB RAM recommended
- 1024 x 768 display recommended
- Intel or AMD CPU required
- Assembly and analysis of genomes up to 10 mega-bases:

- 2 GB RAM required,
 - 4 GB RAM recommended
- Assembly and analysis of larger genomes
 - 2 GB RAM required,
 - 8 GB RAM recommended
- 64 bit computer and operating system required to use more than 2GB RAM

2.2 Available installers

There are installers available for each platform (Windows, Mac OS X and Linux). Each of these installers is available in a 32-bit and a 64-bit version, except for the Mac OS X installer. For Linux, there is both a `.sh` installer and an `.rpm` package.

2.2.1 Java

The Workbenches are based on Java, and this means that there has to be a Java Runtime Environment (JRE) on the computer to run the Workbench. For both Linux and Windows, the installers have a built-in JRE that will be installed in the installation directory of the Workbench. The advantage of this is twofold:

1. For computers who do not already have a JRE installed, the need for downloading and installing a JRE is eliminated.
2. for computers who already have a JRE installed, there will never be compatibility problems because the Workbench always uses its own JRE.

The built-in JRE is the latest Java 6 JRE from Sun Microsystems (<http://java.sun.com>).

The JRE used for running the CLC Workbench will not interfere with existing JREs on the computer.

Java on Mac

Since the Workbench uses Apple's JRE, there is no JRE included in the installer. When running the Workbench on 64-bit systems, please make sure that the 64-bit Java is used for launching applications:

Go to `/Applications/Utilities/Java` and double click on Java Preferences. In the **Java application versions**, reorder the list to have JRE6/64 bit at the top. Note that this may change the behavior of other Java-based programs on the computer.

Note! After setting the Java preference, you can either uninstall and reinstall the workbench and the memory will be adjusted automatically, or you can set the memory yourself (see section [2.3.2](#)).

2.2.2 Overview of available installers

The table below shows an overview of the installers that are available.

Platform	JRE included	Special 64-bit version
Windows (2000, XP and Vista)	Yes	Yes
Mac OS X (10.4 and 10.5)	No	No
Linux installer	Yes	Yes
Linux package	Yes	Yes

2.3 What does it do?

The installer performs the following tasks:

2.3.1 Extracting and copying files to the installation directory

The Workbench is installed into the following directory per default (we use *CLC Main Workbench 5* as example):

Windows C:\Program files\Main Workbench 5

Mac OS X Applications/Main Workbench 5

Note that each major version of a Workbench has its own installation directory. This means that when upgrading from e.g. *CLC Main Workbench 5* to *CLC Main Workbench 6*, the old installation directory of version 4 will be left untouched when you install *CLC Main Workbench 6*.

If you wish to remove the old installation, please run the Uninstall program.

Minor updates will use the existing installation directory of the Workbench.

The installation directory can be defined during installation - the above are the default installation directories (see section 2.4 for more information on how to define the installation directory).

2.3.2 Setting the amount of memory available for the JVM

When running the Workbench, the Java Virtual Machine (JVM) needs to know how much memory it can use. This depends on the amount of physical memory (RAM) and can thus be different from computer to computer. Therefore, the installer investigates the amount of RAM during installation and sets the amount of memory that the JVM can use when running the Workbench.

On Windows and Linux, this value is stored in a property file called `workbenchname.vmoptions` (e.g. `clcmainwb.vmoptions`) which contains a text like this:

```
-Xmx1400m
```

The number (1400) is the amount of memory the Workbench is allowed to use.

On Mac OS X, the `-Xmx` value is stored in `Info.plist` in the application bundle (Control-click the application and choose "Show Package Contents").

The value is set to 75% of the computers RAM per default, and at a maximum of 1400 MB (for 64-bit systems there is no maximum). If you do not wish to use the installer on each computer and use an image instead, either make sure all computers have the same amount of RAM, or set the number to 75% of the computer with the smallest amount of RAM (this value should not be lower than 200 MB, and for genomics-scale data, it should be significantly higher).

2.3.3 Shortcuts and file associations

The installer also creates shortcuts for starting the Workbench, and it creates file associations so that `.clc` files will be opened by the Workbench.

2.4 Silent installation

The installer also has a silent installation mode which is activated by the `-q` parameter when running the installer from a command line, e.g.

```
CLCMainWorkbench_5_0_1.exe -q
```

On Windows, if you wish to have console output, `-console` can be appended as *the second parameter* (this is only needed when running on Windows where there is no output per default):

```
CLCMainWorkbench_5_0_1.exe -q -console
```

You can also in silent mode define a different installation directory: `-dir`.

```
CLCMainWorkbench_5_0_1.exe -q -console  
-dir "c:\program files\bioinformatics\clc\clcmainwb"
```

Note! Both the `-console` and the `-dir` options only work when the installer is run in silent mode.

The `-q` and the `-console` options work for the Uninstall program as well.

Chapter 3

License

There are fundamentally two kinds of licenses for the Workbenches:

Fixed license A license order ID has to be activated against our server for each computer. The license will then be fixed to this computer. This requires manual intervention for each activation.

Floating license A license server is installed in your organization. It hosts a number of licenses which can be shared among all computers. Note that the license server is available for both Linux, Windows and Mac OS X.

For large installations, the floating license is by far the best option, since all the license administration takes place on the server (find the manual for the license server together with the server distribution). The fixed license requires manual work during installation and also if the licenses need to be updated.

For information on how to use the floating license, please refer to the user manual of the Workbenches (see <http://www.clcio.com/usermanuals>).

Plug-ins use the same licensing system as the Workbenches, so all the concepts described here also apply to the plug-in licenses.

3.1 License server set-up on clients

The connection to the license server can be set up as described in the Workbench user manual (see also figure 3.1).

The license server information is stored in a file called `license.properties` in the `settings` folder in the Workbench installation directory. This means, that you need write access to the installation directory (with the default installation directory, you need to be an administrator to have this write access) in order to set up a connection to the license server. The file contains the following:

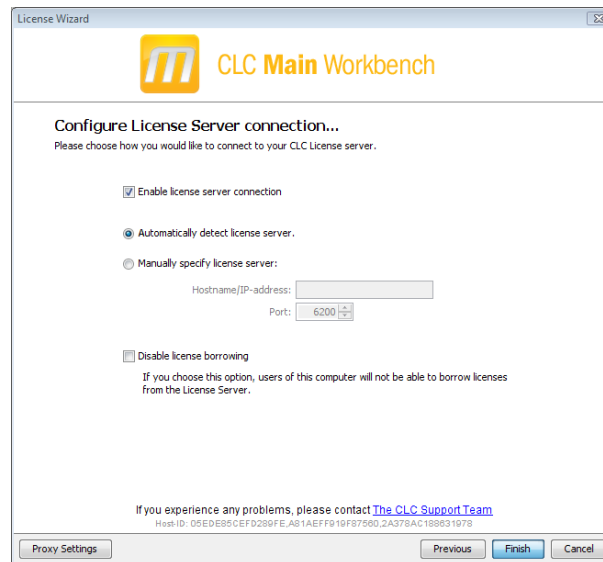


Figure 3.1: Connecting to a license server.

```
#License Settings
#Mon Aug 04 09:57:26 CEST 2008
serverip=
serverport=6200
disableborrow=false
autodiscover=true
useserver=true
```

Since all this information is stored in a file in the installation directory, it can easily be copied to all clients, and the license configuration is completed. When the Workbench is started, it will look in this file, and if `useserver=true` then it will try to connect to the license server, and no license dialogs will be shown to the user.

Chapter 4

Plug-ins and resources

There is a graphical user interface to install plug-ins called the **Plug-in Manager** (📦) which is invoked in the **Help** menu (see figure 4.1).

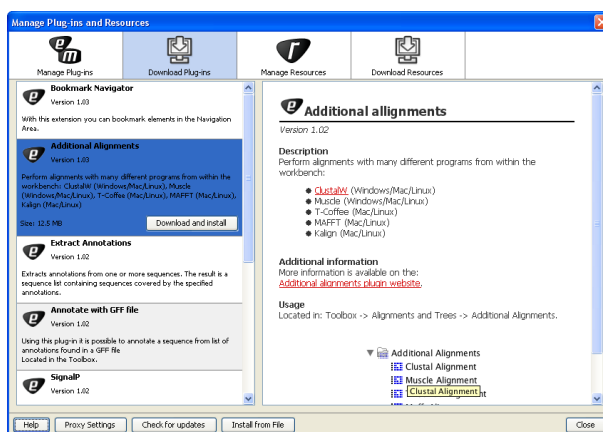


Figure 4.1: The Plug-in Manager.

Plug-ins are either general modules or extensions provided by CLC bio (see <http://www.clcbio.com/plugins>) or can be custom-made plug-ins specific to your organization.

Plug-ins can either be downloaded and installed directly in the **Plug-in Manager**, or they can be installed from a file Install from File button at the bottom of the **Plug-in Manager**.

Resources are installed in the same way as plug-ins. Resources can be e.g. PFAM databases used by the Workbench's PFAM Domain Search (→).

Installing a plug-in is basically just a matter of putting files in the right folder. All plug-in files are put in `plugins` and all resources in `resource` in the installation directory. This means that the contents of these folders can be copied to other computers, and they will have the plug-ins installed.

Licenses for the plug-ins are handled the same way as the Workbench licenses, see section 3.

Chapter 5

Storing and backing up data

This chapter explains how data is stored, gives general guidance on size of data, and outlines configurations needed for running analyses on large amounts of data.

5.1 Data structure

The data in the **Navigation Area** is organized into a number of **Locations**. When the the CLC workbenches except the *CLC Genomics Workbench* is started for the first time, there is one location called *CLC_Data*.

A location represents a folder on the computer: The data shown under a location in the **Navigation Area** is stored on the computer in the folder which the location points to.

This is explained visually in figure 5.1.

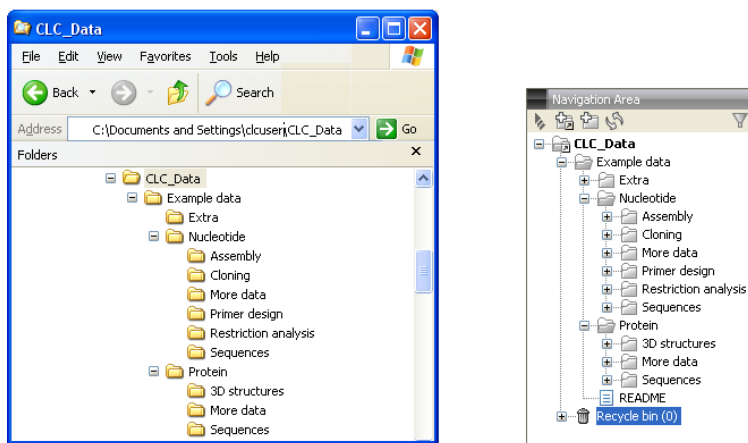


Figure 5.1: In this example the location called 'CLC_Data' points to the folder at *C:\Documents and settings\clcuser\CLC_Data*.

If the database plug-in is installed, you will be able to add database locations in a similar way as described in the manual for the *CLC Bioinformatics Database* (see <http://www.clcio.com/usermanuals>).

The list of locations is stored in a file called *model_settings_300.xml* in the *settings* folder in the user home (see section 6.2). We do not recommend manual editing of this file,

although it is standard xml.

5.2 Back-up of data

Since all data used in the Workbench is stored as files in the locations specified, a back-up procedure has to include all the locations. If the data needs to be restored from a back-up, simply copy the files back into the folder locations and start the Workbench.

Database locations needs a different back-up procedure.

Besides the data itself, user-level settings should also be included in the back-up (see section 6.2).

5.3 Special configurations for large amounts of data

Especially the *CLC Genomics Workbench* is often used with large amounts of data. This means that special configurations often need to be made. This concerns locations for temporary data and disk space in general.

5.3.1 Temporary data

The Workbench has a built-in cache system that intends to make sure that the Workbench does not run out of memory even for large data sets. During various processes such as assembly and RNA-Seq analysis, the Workbench often writes temporary files to the disk. Depending on the data set, these temporary files can take up a lot of disk space.

If there is not enough space in the default tmp directory, the tmp directory can be re-directed:

Create a text file called `path.properties` and save it in the `settings` folder in the Workbench installation directory. The file should include one line like this:

```
tmpdir = /path-to-temp
```

Instead of `path-to-temp` you write the absolute path to the new tmp directory. When the Workbench is restarted, it will then use the new directory for storing temporary data.

Note! It is imperative for acceptable performance that data transfer to the temp directory is not over a network connection. Since the Workbench will spend a lot of time writing and reading these files, disk speed has a great impact on overall performance when working with large data sets.

5.3.2 Disk space requirements

It is hard to give general guidance on disk space requirements, but we have made an example of a typical work flow for *CLC Genomics Workbench* to illustrate.

For calculating disk space for next-generation sequencing data you need to consider the following:

- Reads are imported and take up space as raw reads (see details below). Once imported, you can delete the original sequence file if you do not need it for other purposes.

- When the data has been assembled, either de-novo or against a reference, they take up space once again (this time more space since there is also information about where they map etc).
- Reference sequences also take up space.
- The computer doing the analysis needs space for tmp files. Once the assembly is done, the temporary files are deleted. The temporary files usually do not take up more space than the final result

The formulae giving the disk space usage:

Bytes per read: $28 + (\text{length of read name}) + 0.25 \times (\text{length of read})$

Note that you can discard read names during import.

If quality scores are present, add: $6 + (\text{length of read})$

If color space encoding is present, add: 7

As an example, a data set of 5.2 million 35 bp reads imported by *CLC Genomics Workbench* 3.6.1 using the **Discard sequence names** option including quality scores takes up:

$5,244,764 \times ((28 + 0 + 0.25 \times 35) + (6 + 35)) = 389 \text{ MB}$

When assembled to a 4.7 Mbp annotated reference sequence, the contig takes up 511 MB.

Chapter 6

Overview - where do we put things?

This part gives you an overview of all the parts of a CLC Workbench installation. Some parts are at the level of the computer, i.e. shared by all users, whereas other parts are at the user level.

6.1 Computer-level information

In the *Workbench installation directory*, you will find the following:

Licenses The license information depends on what kind of license you use:

Fixed licenses Stored in the `licenses` folder (they are unique for each computer)

Floating licenses Information about the license server connection is stored in the `license.properties` file in the `settings` folder.

Plug-ins are stored in the `plug-ins` folder and can be copied to other computers running the same version of the Workbench.

Resources are stored in the `resource` folder and can be copied to other computers running the same version of the Workbench.

Memory allocation for the VM is stored in the `workbenchname.vmoptions` file.

Error logs to be sent to support@clcbio.com for use in case of program errors are `output.log` and `error.log`.

Proxy settings Information about proxy server (when the Workbench needs access to online services) is stored in the `proxy.properties` file in the `settings` folder (the file will only be created if a proxy server has been specified).

In addition, file associations for `.clc` files are stored in the registry database on Windows.

6.2 User-level information

The user-level information is found in the application data folder:

Windows 2000 and XP `C:\Documents and settings\username\Application data\CLC bio`

Windows Vista C:\Users\username\AppData\Roaming\CLC bio

Mac OS X User home/Library/Application Support/CLC bio

In this folder, the following information may be useful to you:

User settings The user settings file stores information such as view settings, parameters, workspaces, log-in information to database locations and other settings customized by the user. The user settings file is found in the application data folder under `settings/workbench name/version name`. It is recommended to back up the user settings file. When upgrading to a new version of the Workbench, the user's old user settings file is copied by the Workbench the first time it is run.

Locations Information about which locations the user has added in the **Navigation Area** is stored in the `model_settings_300.xml` file in the `settings` folder.

Index

- CLC Bioinformatics Database, 5
- CLC Genomics Server, 5
- 32-bit installer, 8
- 64 bit installer, 8
- Back up
 - user settings, 18
- Command-line installation, 10
- Data structure, 14
- Database
 - local, 14
- Error log, 17
- Floating license, 11
- GHOST image, 6
- Image, copy, 6
- Install
 - plug-ins, 13
 - resources, 13
- Installers, overview, 9
- Introduction, 5
- Java, 8
- JRE, Java Runtime Environment, 8
- JVM, Java Virtual Machine, 8
- License order ID, 11
- License server, 11
 - set-up on clients, 11
- Licenses, 11
 - for plug-ins, 11
- Linux, 5
- Linux installer vs. package, 8
- Locations, 18
- Memory allocation, 9
- Output log, 17
- Plug-in licenses, 11
- Plug-in Manager, 13
- Plug-ins, 13
- Proxy server, 17
- Quiet installation, 10
- RAM, 9
- Resources, 13
 - .rpm, Linux package, 8
 - .sh, Linux installer, 8
- Silent installation, 10
- System requirements, 7
- User settings, 18
- VM, Virtual Machine, 8
 - .vmoptions, memory allocation, 9
- Xmx argument, 9