



CLC **Server**

End User

USER MANUAL

Manual for
CLC Server 8.0
Windows, Mac OS X and Linux

April 1, 2016

This software is for research purposes only.

CLC bio, a QIAGEN Company
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark



Contents

1	Introduction	4
2	Using the server from a CLC Workbench	6
2.1	Installing relevant plugins in the Workbench	6
2.2	Log into the server	6
2.3	Browsing and searching data from the workbench	7
2.3.1	Deleting data	8
2.4	Running analyses on the server	8
2.5	Accessing files on, and writing to, areas of the server filesystem	10
2.6	Monitoring processes	12
3	Using a web browser as client	13
3.1	Browsing and searching data	13
3.2	Import and export	14
4	Appendix	16
4.1	CLC Genomics Server	16
4.2	Biomedical Genomics Server Extension	19

Chapter 1

Introduction

Welcome to *CLC Server 8.0*, a central element of the CLC product line enterprise solutions.

The latest version of the user manual can also be found in pdf format at <http://www.clcbio.com/usermanuals>.

You can get an overview of the server solution in figure 1.1. The software depicted here, including *CLC Server* is for research purposes only.

Using a server means that data can be stored centrally and analyses run on a central machine rather than a personal computer. Please see section 4.1 and section 4.2 for a listing of tools shipped with CLC Servers.

After logging into the CLC Server from a Workbench, data on the server will be listed in the Workbench navigation area and analyses can be started as usual. The key difference is that when you are logged into a CLC Server from a Workbench, you will be get the choice of where to run the analysis: on the Workbench or on the CLC Server.

This manual describes how to use a CLC Server as a Workbench user. For information about administrating the server, please see the Administrator Manual.

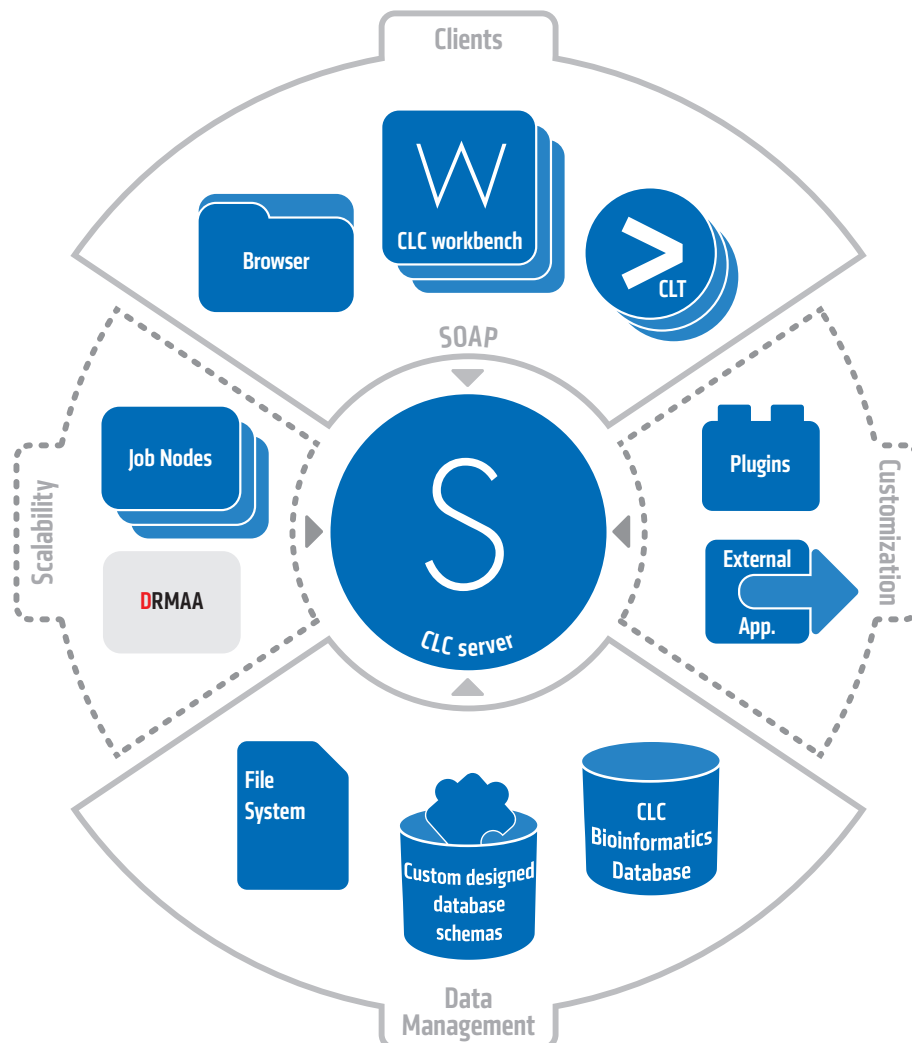


Figure 1.1: An overview of the server solution from CLC bio. Note that not all features are included with all license models.

Chapter 2


Using the server from a CLC Workbench

2.1 Installing relevant plugins in the Workbench

To access the CLC Server from a CLC Workbench, the CLC Workbench Client Plugin must be installed in the Workbench. This will allow you to log into the CLC Server, access data from CLC Server file and data locations and submit analyses to your CLC Server.

Plugins are installed in a Workbench using the Plugins Manager¹, which can be accessed via the menu in the Workbench

Help | Plugins... ()

or via the **Plugins** () button on the Toolbar.

From within the Plugins manager, choose the Download Plugins tab and click on the relevant plugin or module. Then click on the button labeled **Download and Install**.

If you are working on a system not connected to the internet, then you can also install the plugin or module by downloading the cpa file from the plugins page of our website

<http://www.clcbio.com/clc-plugin/>

Then start up the Plugin manager within the Workbench, and click on the button at the bottom of the Plugin manager labeled **Install from File**.

You need to restart the Workbench before the plugin is ready for use.

2.2 Log into the server

Once the plugin is installed, log into the server:

File | CLC Server Login ()

That will bring up a login dialog as shown in figure 2.1).

The first time you log into the server, you have to expand the dialog by clicking **Advanced**. That will allow you to enter the host and port for the server as shown in figure 2.2).

¹In order to install plugins on many systems, the Workbench must be run in administrator mode. On Windows Vista and Windows 7, you can do this by right-clicking the program shortcut and choosing "Run as Administrator".

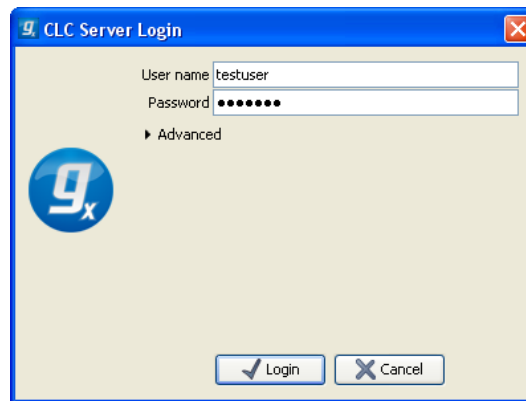


Figure 2.1: Expand the login dialog by clicking Advanced.

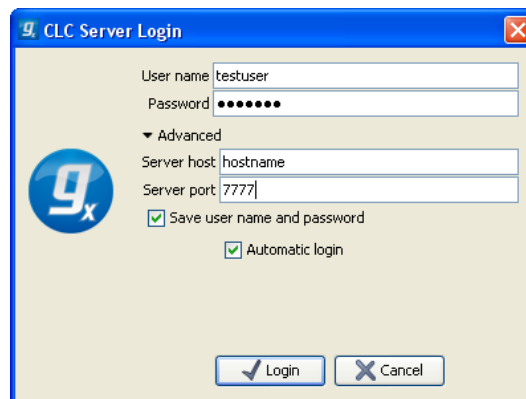


Figure 2.2: Specifying host and port.

In addition you can choose to save user name and password and automatically log into the server when the Workbench starts.

Note that you need to get the login information from your server administrator.

When you press **Login**, the Workbench connects to the server. You will see a progress bar in the login dialog. If the login is successful, the dialog will disappear, and you will be able to use the server as described below.

2.3 Browsing and searching data from the workbench

Once you are logged in, the data locations on the server are shown in the **Navigation Area** (Figure 2.3).

Once logged in, the server data locations can be used in the same way as local data locations. We refer to the user manual of the Workbench for information about using the **Navigation Area** (click the location and press **F1** on the keyboard to get context help).

You may also wish to have a look at the search chapter, *Searching your data*, in the workbench user manual (press **F1** and look for *Searching your data*).

It is possible to have both local and server locations added at the same time. This means that you can work on e.g. temporary sequences located on your own computer and then when you have more complete results, you can drag the elements to a folder in the server location.

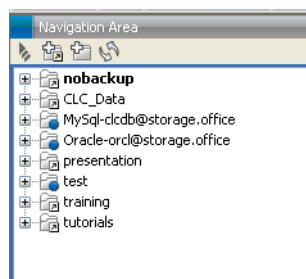


Figure 2.3: Three server locations on the server appears in the Navigation Area (marked with blue dots). The remaining five local folders are file locations in the Workbench that are independent of the server.

Note that when logged in a CLC Server with a *Biomedical Genomics Server Extension*, you will have in your Navigation Area two folders called CLC_References. The blue dot indicates which of these repositories is installed on the server.

2.3.1 Deleting data

When you delete data located on the server, it will be placed in the **Recycle bin** (🗑️) in the same way as when you delete data located on the Workbench. The data in the recycle bin can only be accessed by you and the server administrator.

Please note that the server administrator might have configured the recycle bin to be automatically emptied at regular intervals.

2.4 Running analyses on the server

The tools available on the different types of CLC Server are listed in section 4.1 and section 4.2.

For more information on the tools, please see the manual for the workbench at <http://www.clcbio.com/usermanuals> or click the **Help** (?) button in the dialogs that are shown when you run the tools.

When you run the analyses you will be faced with an initial dialog asking you where you wish the analysis to be run:

- **Workbench.** Run the analysis on the computer the CLC Workbench is running on.
- **Server.** Run the analysis using the CLC Server. For job node setups, analyses will be run on the job nodes.
- **Grid.** Submit the job to the CLC Server such that the job is then sent to be run on grid nodes.

An example of such a dialog is shown in figure 2.4.

You can check the **Remember setting and skip this step** option if you wish to always use the selected option when submitting analyses. If you change your mind later on and want to switch, click **Previous** in the dialog when you start up an analysis. You will then be taken back to the dialog where you can choose where the analysis should be run.

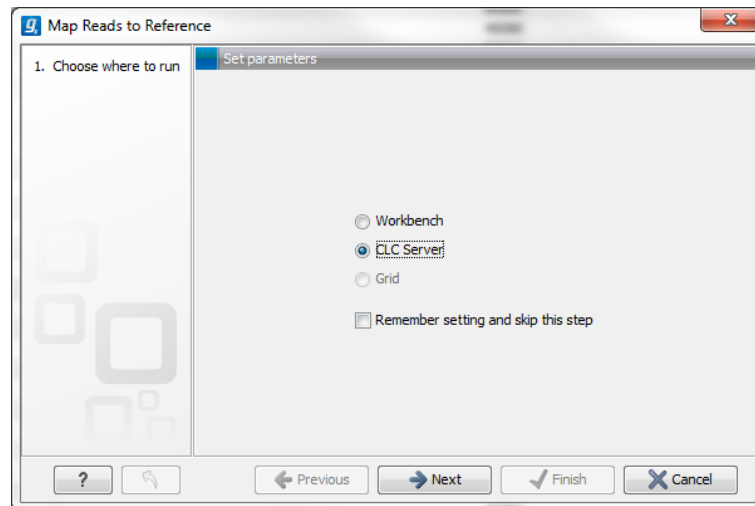


Figure 2.4: Selecting where to run the analysis.

Launching an analyses to run on a CLC Server, or on grid nodes, is identical to launching the same analyses to run on a Workbench, but there are three things to be aware of:

- You can only select data stored in locations configured on the server, with the exception of running data import. This means that when you are given the option to choose the data to use in an analysis, only server locations are shown. Further information about the exceptional case of data import is provided below.
- You have to save the result. For single analyses run on the Workbench, you can normally choose how to handle the results: **Open** or **Save**. Results from analyses performed on the server must be saved, so the option to open the results instead of saving them will be unavailable.
- When you click **Finish**, the analysis is submitted to the server to be handled. This means that you can close the Workbench or disconnect from the server without affecting the analysis. (See the notes on import below for an exception to this.) If an analysis has been completed when your Workbench is closed, or while it is not connected to the server, you will see a notification about this when you next log into the server from the Workbench.

Important note about data import to a server from a Workbench When using an import tool (Figure 2.5) you are offered the option of importing data from a local area, that is, an area accessible from the machine the CLC Workbench is running on, or from an area the CLC Server has access to. When importing data into the server from an area the CLC Server has access to, you can close your Workbench or disconnect it, right after submitting the job. However, when importing data from your local system to a server, the first part of the import involves uploading the data from the local system to the server system. During the upload phase, the Workbench must maintain its connection to the server. If you try to close the Workbench during this phase, you will see a warning dialog. You can see what stage tasks are at by opening the **Processes** tab in the lower left corner of the Workbench. Data upload from the Workbench to the server runs as a local process in the **Processes** tab. When the upload is done, a new process for the import is started. This will have a server icon. At this point, you can disconnect or close your Workbench without affecting the import.

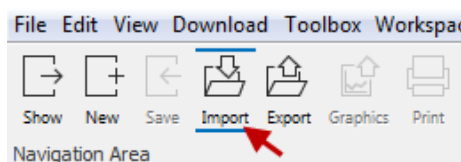


Figure 2.5: Import tools are found under this menu in the Workbench.

2.5 Accessing files on, and writing to, areas of the server filesystem

There are situations when it is beneficial to be able to interact with (non-CLC) files directly on your server filesystem.

A common use case would be importing high-throughput sequencing data or large molecule libraries from folders where it is stored on the same system that your *CLC Server* is running on. This could eliminate the need for each user to copy large data files to the machine the CLC Workbench is running on before importing the data into a *CLC Server* data area.

Another example is if you wish to export data from CLC format to other formats and save those files on your server machine's filesystem (as opposed to saving the files in the system your Workbench is running on).

From the administrator's point of view, this is about configuring folders that are safe for the *CLC Server* to read and write to on the server machine system.

This means that users logged into the *CLC Server* from their Workbench will be able to access files in that area, and potentially write files to that area. Note that the *CLC Server* will be accessing the file system as the *user running the server process* - not as the user logged into the Workbench. This means that you should be careful when opening access to the server filesystem in this way. Thus, only folders that do not contain sensitive information should be added.

Folders to be added for this type of access are configured in the web administration interface **Admin** tab. Under **Main configuration**, open the **Import/export directories** (Figure 2.6) to list and/or add directories.

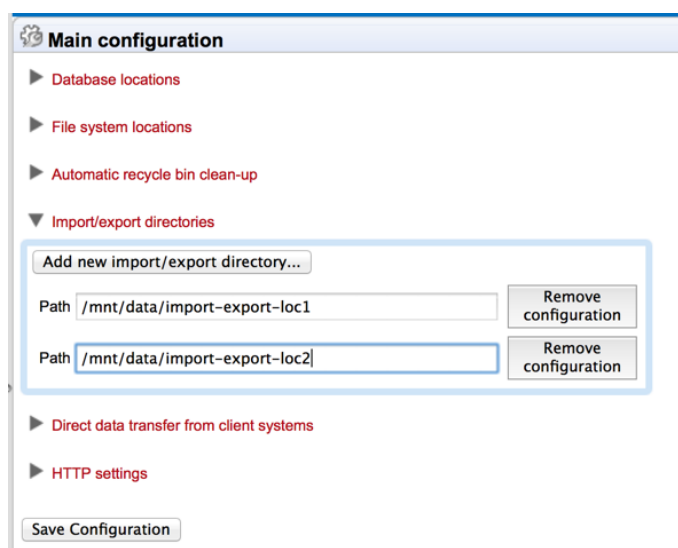


Figure 2.6: Folders on the server that should be available from the Workbench for browsing, importing data from and exporting data to.

Press the **Add new import/export directory** button to specify a path to a folder on the server. This folder and all its subfolders will then be available for browsing in the Workbench for certain activities (e.g. importing data functions).

The import/export directories can be accessed from the Workbench via the Import function in the Workbench. If a user, that is logged into the CLC Server via their CLC Workbench, wishes to import e.g. high throughput sequencing data, an option like the one shown in figure 2.7 will appear.

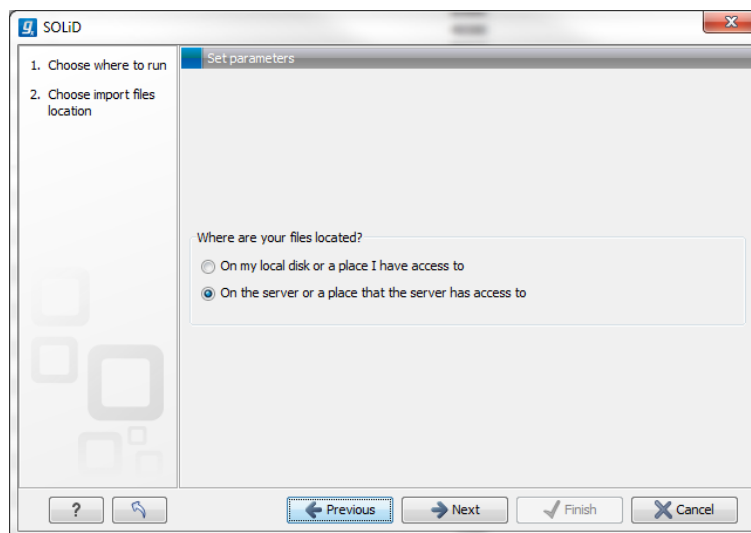


Figure 2.7: Deciding source for e.g. high-throughput sequencing data files.

On my local disk or a place I have access to means that the user will be able to select files from the file system of the machine their CLC Workbench is installed on. These files will then be transferred over the network to the server and placed as temporary files for importing. If the user chooses instead the option *On the server or a place the server has access to*, the user is presented with a file browser for the selected parts of the server file system that the administrator has configured as an Import/export location (an example is shown in figure 2.8).

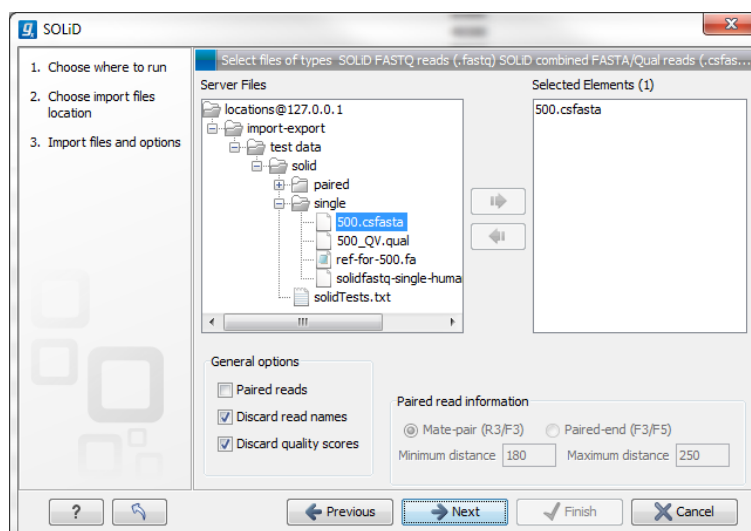


Figure 2.8: Selecting files on server file system.

Note: Import/Export locations should NOT be set to subfolders of any defined CLC file or data

location. CLC file and data locations should be used for CLC data, and data should only be added or removed from these areas by CLC tools. By definition, an Import/Export folder is meant for holding non-CLC data, for example, sequencing data that will be imported, data that you export from the CLC Server, or BLAST databases.

Note that your server administrator needs to configure the server to import files directly from the server file system.

2.6 Monitoring processes

You can monitor processes running on the server or the local Workbench by opening the **Processes** tab at the bottom left of the Workbench. This tab is next to the **Toolbox** tab. A list of submitted and running processes (see figure 2.9) will be visible there.

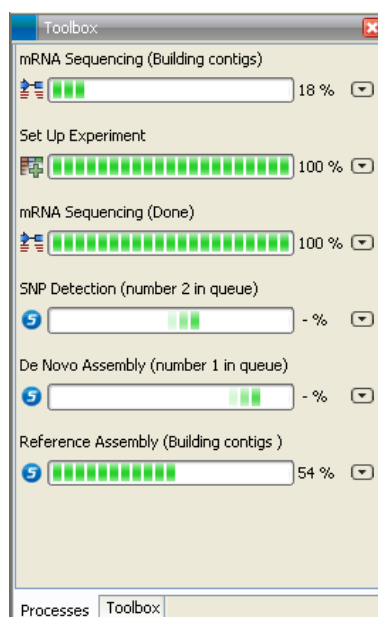


Figure 2.9: Monitoring processes.

Processes running on the server will have a server icon (S) whereas processes running locally have icons specific to the analysis being run, for example (mRNA) and (De Novo). In the image, you can also see that two of the server processes are queued.

Server processes that are queued or running will reappear in the Workbench processes tab if you restart the Workbench (and log into the server). Server processes that are finished when you close the Workbench will not be shown again in the processes tab when you restart your Workbench.

Chapter 3

Using a web browser as client

Besides using the CLC Workbench as a client, you also have access to the server from the web interface. The web interface gives you access to browsing and searching data and importing and exporting data.

Simply type the address of the server into your browser, followed by the port number, and you will see a login dialog similar to what is shown in figure 3.1 (you will need to get information from your server administrator about the server address).

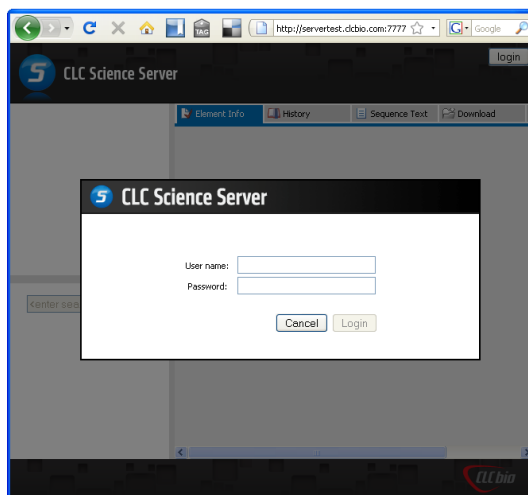


Figure 3.1: The web interface of the server.

3.1 Browsing and searching data

To the left, you will see the data locations connected to the server. You can browse the folder hierarchy of the data location. When you click an element in the tree, a number of options are available:

- Click the **Element Info** (📄) tab to see the properties of this element. Note that you can edit the information in this view.
- Click the **History** (🕒) tab to see the history of this element.

- Click the **Sequence Text** (📄) tab to see a textual representation of this element (only works for sequences)

An example of a protein sequence in the text view is shown in figure 3.2.

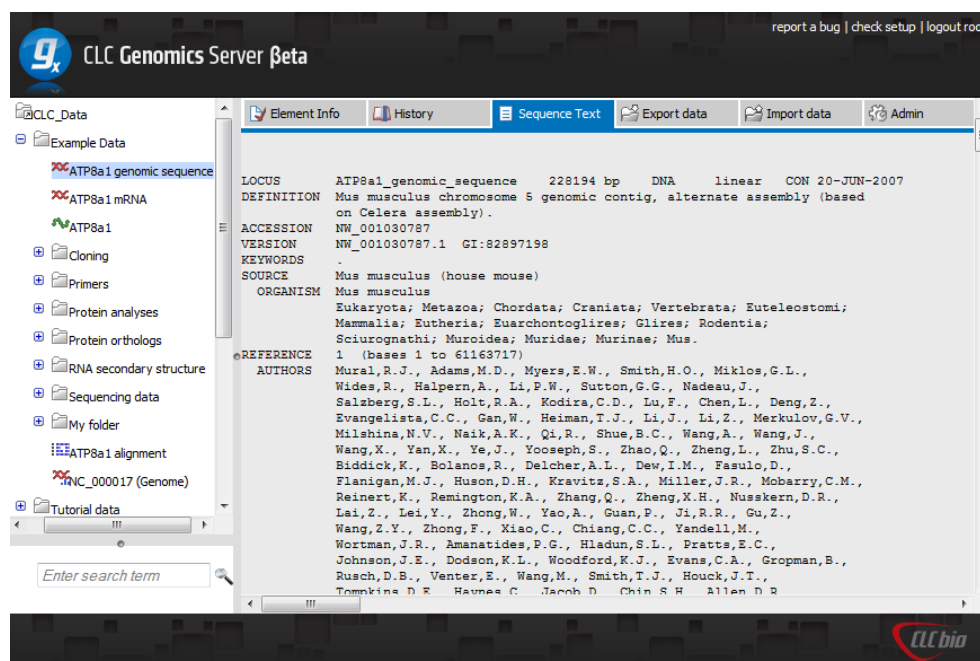


Figure 3.2: Inspecting the text view of a protein sequence.

Note that these views are a subset of the views that you find in the CLC Workbench.

3.2 Import and export

It is possible to import from and export to the server. If you wish to import data from the server, click **Import** (📁) and select the relevant data. Leave the file import format to 'Automatic' and press the 'Import File' button (figure 3.3).

You can also put data into the import/export directories: Select the data you wish to export and click **Export** (📁). Next, tick 'Save on server' and select the folder where you want the data to be (see figure 3.4).

Click the button labeled **Export**. The server will automatically recognize the file format and interpret the file.

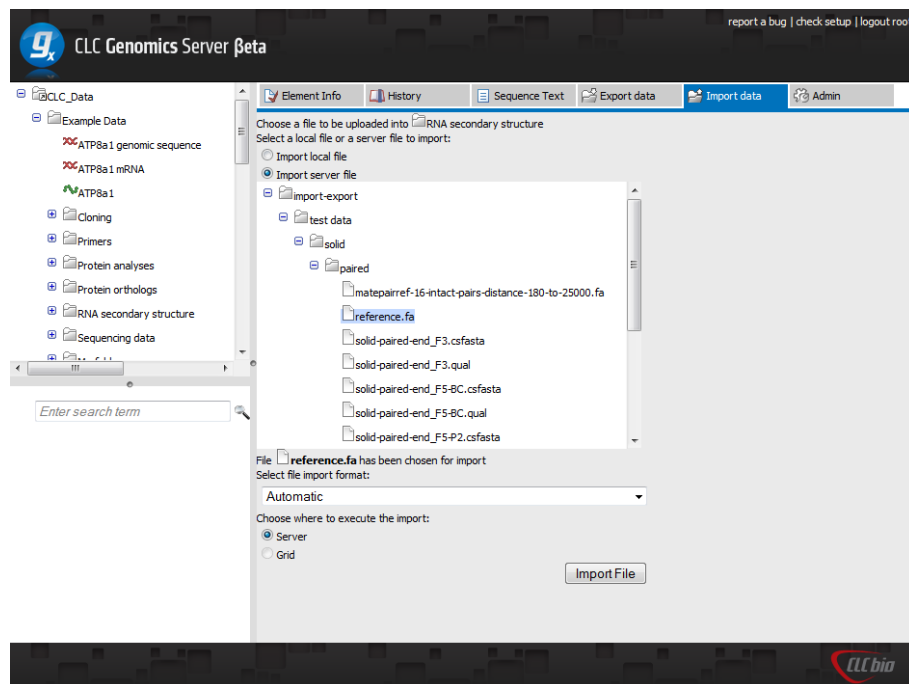


Figure 3.3: Importing a sequence from the server.

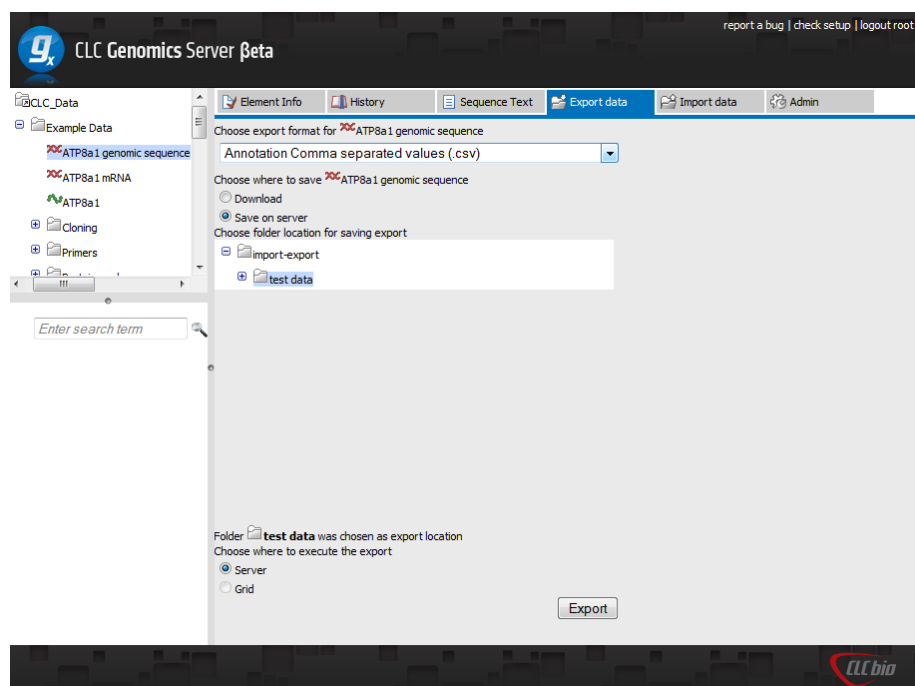


Figure 3.4: Exporting sequences to the server.

Chapter 4

Appendix

4.1 CLC Genomics Server

The *CLC Genomics Server* is shipped with the following tools and analyses that can all be started from *CLC Genomics Workbench* and *CLC Server Command Line Tools*:

- Import
- Export
- Download Reference Genome Data
- Classical Sequence Analysis
 - Create Alignment (Alignments and Trees)
 - K-mer Based Tree Construction (Alignments and Trees)
 - Create Tree (Alignments and Trees)
 - Model Testing (Alignments and Trees)
 - Maximum Likelihood Phylogeny (Alignments and Trees)
 - Extract Annotations (General Sequence Analysis)
 - Extract Sequences (General Sequence Analysis)
 - Motif Search (General Sequence Analysis)
 - Translate to Protein (Nucleotide Analysis)
 - Convert DNA to RNA (Nucleotide Analysis)
 - Convert RNA to DNA (Nucleotide Analysis)
 - Reverse Complement Sequence (Nucleotide Analysis)
 - Reverse Sequence (Nucleotide Analysis)
 - Find Open Reading Frames (Nucleotide Analysis)
 - Download Pfam Database (Protein Analysis)
 - Pfam Domain Search (Protein Analysis)
- Molecular Biology Tools

- Assemble Sequences (Sequencing Data Analysis)
 - Assemble Sequences to Reference (Sequencing Data Analysis)
 - Secondary Peak Calling (Sequencing Data Analysis)
 - Find Binding Sites and Create Fragments (Primers and Probes)
 - Add attB Sites (Cloning and Restriction Sites - Gateway Cloning)
 - Create Entry clone (BP) (Cloning and Restriction Sites - Gateway Cloning)
 - Create Expression clone (LR) (Cloning and Restriction Sites - Gateway Cloning)
- BLAST
 - BLAST
 - BLAST at NCBI
 - Download BLAST Databases
 - Create BLAST Database
- NGS Core Tools
 - Sample Reads
 - Create Sequencing QC Report
 - Merge Overlapping Pairs
 - Trim Sequences
 - Demultiplex Reads
 - Map Reads to Reference
 - Local Realignment
 - Create Detailed Mapping Report
 - Merge Read Mappings
 - Remove Duplicate Mapped Reads
 - Extract Consensus Sequence
- Track Tools
 - Convert to Tracks
 - Convert from Tracks
 - Merge Annotation Tracks
 - Annotate with Overlap Information (Annotate and Filter)
 - Extract Reads Based on Overlap (Annotate and Filter)
 - Filter Annotations on Name (Annotate and Filter)
 - Filter Based on Overlap (Annotate and Filter)
 - Create GC Content Graph Tracks (Graphs)
 - Create Mapping Graph Tracks (Graphs)
 - Identify Graph Threshold Areas(Graphs)
- Resequencing Analysis

- Create Statistics for Target Regions
- InDels and Structural Variants
- Coverage Analysis
- Basic Variant Detection (Variant Detectors)
- Fixed Ploidy Variant Detection (Variant Detectors)
- Low Frequency Variant Detection (Variant Detectors)
- Annotate from Known Variants (Annotate and Filter Variants)
- Filter against Known Variants (Annotate and Filter Variants)
- Annotate with Exon Numbers (Annotate and Filter Variants)
- Annotate with Flanking Sequences (Annotate and Filter Variants)
- Filter Marginal Variant Calls (Annotate and Filter Variants)
- Filter Reference Variants (Annotate and Filter Variants)
- Compare Sample Variant Tracks (Compare Variants)
- Compare Variants within Group (Compare Variants)
- Fisher Exact Test (Compare Variants)
- Trio Analysis (Compare Variants)
- Filter against Control Reads (Compare Variants)
- GO Enrichment Analysis (Functional Consequences)
- Amino Acid Changes (Functional Consequences)
- Annotate with Conservation Score (Functional Consequences)
- Predict Splice Site Effect (Functional Consequences)
- Link Variants to 3D Protein Structure (Functional Consequences)
- Download 3D Protein Structure Database (Functional Consequences)
- Transcriptomics Analysis/ Expression Analysis
 - Create Track from Experiment
 - RNA-Seq Analysis (RNA-Seq Analysis)
 - Extract and Count (Small RNA Analysis)
 - Annotate and Merge Counts (Small RNA Analysis)
 - Create Box Plot (Quality Control)
 - Hierarchical Clustering of Samples (Quality Control)
 - Principal Component Analysis (Quality Control)
 - Empirical Analysis of DGE (Statistical Analysis)
 - Proportion-based Statistical Analysis (Statistical Analysis)
 - Gaussian Statistical Analysis (Statistical Analysis)
 - Create Histogram (General Plots)
- Epigenomics Analysis
 - Transcription Factor ChIP-Seq

- Annotate with Nearby Gene Information
- De Novo Sequencing
 - De Novo Assembly
 - Map Reads to Contigs
- Legacy Tools
 - Probabilistic Variant Detection (Legacy)
 - Quality-based Variant Detection (Legacy)

The functionality of the *CLC Genomics Server* can be extended by installation of Server plugins. The available plugins can be found at http://www.clcbio.com/server_plugins.

Latest improvements

CLC Genomics Server is under constant development and improvement. A detailed list that includes a description of new features, improvements, bugfixes, and changes for the current version of *CLC Genomics Server* can be found at:

<http://www.clcbio.com/products/clc-genomics-server-latest-improvements/>.

4.2 Biomedical Genomics Server Extension

A *CLC Genomics Server* with the *Biomedical Genomics Server Extension* is shipped with the tools and analyses listed below. These can all be started from the *Biomedical Genomics Workbench* or by using the *CLC Server Command Line Tools*.

- Import
- Export
- Download Reference Genome Data
- Genome Browser
 - Create GC Content Graph (Graphs)
 - Create Mapping Graph (Graphs)
 - Identify Graph Threshold Area (Graphs)
- Quality Control
 - QC for Sequencing Reads
 - QC for Target Sequencing
 - QC for Read Mapping
- Preparing Raw Data
 - Merge Overlapping Pairs
 - Trim Sequences

- Demultiplex reads
- Resequencing Analysis
 - Identify Known Mutations from Sample Mappings
 - Trim Primers of Mapped Reads
 - Extract Reads Based on Overlap
 - Map Reads to Reference
 - Local Realignment
 - Merge Read Mappings
 - Copy Number Variant Detection
 - Remove Duplicate Mapped Reads
 - Indels and Structural Variants
 - Whole Genome Coverage Analysis
 - Basic Variant Detection (Variant Detectors)
 - Fixed Ploidy Variant Detection (Variant Detectors)
 - Low Frequency Variant Detection (Variant Detectors)
- Add Information to Variants
 - Add Information from Variant Databases
 - Add Conservation Scores
 - Add Exon Number
 - Add Flanking Sequence
 - Add Fold Changes
 - Add information about Amino Acids Changes
 - Add Information from Genomic Regions
 - Add Information from Overlapping Genes
 - Link Variants to 3D Protein Structure
 - Download 3D Protein Structure Database
 - Add Information from 1000 Genomes Project (From Databases)
 - Add Information from COSMIC (From Databases)
 - Add Information from Clinvar (From Databases)
 - Add Information from Common dbSNP (From Databases)
 - Add Information from Hapmap (From Databases)
 - Add Information from dbSNP (From Databases)
- Remove Variants
 - Remove Variants Found in External Databases
 - Remove Variants Not Found in External Databases
 - Remove False Positive
 - Remove Germline Variants

- Remove Reference Variants
 - Remove Variants Inside Genome Regions
 - Remove Variants Outside Genome Regions
 - Remove Variants Outside Targeted Regions
 - Remove Variants Found in 1000 Genomes Project (From Databases)
 - Remove Variants Found in Common dbSNP (From Databases)
 - Remove Variants Found in Hapmap (From Databases)
- Add Information to Genes
 - Add Information from Overlapping Variants
- Compare Samples
 - Compare Shared Variants Within a Group of Samples
 - identify Enriched Variants in Case vs Control Group
 - Trio Analysis
- Identify Candidate Variants
 - Identify Candidate Variants
 - Remove Information from Variants
 - Identify Variants with Effect on Splicing
- Identify Candidate Genes
 - Identify Differentially Expressed Gene Groups and Pathways
 - Identify Highly Mutated Gene Groups and Pathways
 - Identify Mutated Genes
 - Select Genes by Name
- Expression Analysis
 - Extract Differentially Expressed Genes
 - RNA-Seq Analysis (RNA-Seq Analysis)
 - Create Fold Change Track (RNA-Seq Analysis)
 - Extract and Count (Small RNA Analysis)
 - Annotate and Merge Counts (Small RNA Analysis)
 - Create Box Plot (Quality Control)
 - Hierarchical Clustering of Samples (Quality Control)
 - Principal Component Analysis (Quality Control)
 - Empirical Analysis of DGE (Statistical Analysis)
 - Proportion-based Statistical Analysis (Statistical Analysis)
 - Gaussian Statistical Analysis (Statistical Analysis)
 - Create Histogram (General Plots)

- Helper Tools
 - Extract Sequences
 - Filter Based on Overlap
- Cloning and Restriction Sites
 - Add attB Sites (Gateway Cloning)
 - Create Entry clone (BP) (Gateway Cloning)
 - Create Expression clone (LR) (Gateway Cloning)
- Sanger Sequencing
 - Assemble Sequences (Sequencing Data Analysis)
 - Assemble Sequences to Reference (Sequencing Data Analysis)
 - Secondary Peak Calling (Sequencing Data Analysis)
 - Find Binding Sites and Create Fragments (Primers and Probes)
- Epigenomics Analysis
 - Transcription Factor ChIP-Seq
 - Annotate with Nearby Gene Information
- Legacy Tools
 - Probabilistic Variant Detection (Legacy)
 - Quality-based Variant Detection (Legacy)

The functionality of the *CLC Server* can be extended by installation of Server plugins. The available plugins can be found at http://www.clcbio.com/server_plugins.