

Additional Alignments Plugin

USER MANUAL

User manual for Additional Alignments Plugin 1.9

Windows, macOS and Linux

December 3, 2018

This software is for research purposes only.

QIAGEN Aarhus
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark




Contents

1 Introduction	4
2 Alignment methods	5
2.1 ClustalO	5
2.2 ClustalW	6
2.3 MUSCLE	7
3 Install and uninstall plugins	8
3.1 Install	8
3.2 Uninstall	9
Bibliography	11

Chapter 1

Introduction

This module allows for use of three additional alignment methods, which are otherwise not distributed with the CLC Workbench. These methods will very briefly be mentioned below and we urge the users of this plugin to read the manuals for the individual alignment methods. These are generally easy to find by running an internet search.

The alignment methods provided with the Additional Alignments plugin can be found in the **Toolbox** in **Alignments and Trees**  and work with CLC Genomic Workbench 5.0.0 and higher. Note that all 3 alignments programs are developed by third parties. They are provided as 64-bit versions for Linux and Mac but are only available in 32-bit version for Windows, which limits the memory usage at 4Gb regardless of the total memory available. Would you encounter memory issues on Windows (most likely with ClustalO and MUSCLE), try using the programs on another operating system or reduce your original input dataset.

Chapter 2

Alignment methods

Three different alignment methods are included in this module: ClustalO, ClustalW, and MUSCLE. The included alignment methods and how they are used are very briefly described below. For more detailed information on each of the alignment methods we encourage to read the original research papers and the respective websites of the different methods. The papers [Edgar and Batzoglou, 2006] and [Plyusnin and Holm, 2012] describe the below mentioned algorithms and some of their strengths.

When the plugin is installed it can be found in the toolbox as shown in figure 2.1.

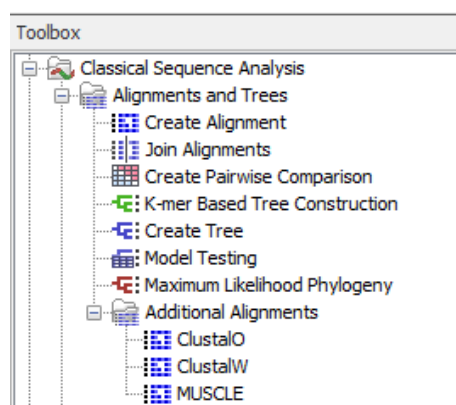


Figure 2.1: The Additional Alignments plugin can be found in the toolbox under "Alignments and Trees".

To create an alignment using an alignment method in the plugin, select sequences to align and go to

Classical Sequence Analysis (📁) | **Alignments and Trees** (📁) | **Additional Alignments** (📁)

Choose the desired alignment (🔧). This opens the dialog shown in figure 2.2.

2.1 ClustalO

Clustal Omega or ClustalO (see <http://www.clustal.org/omega>) method has been developed to overcome the bottleneck that occurs in some pipelines when presented with data sets consisting of many thousands of sequences. The strength of ClustalO is that it can align virtually any number of sequences at high speed and with accurate alignments as a result [Sievers et al.,

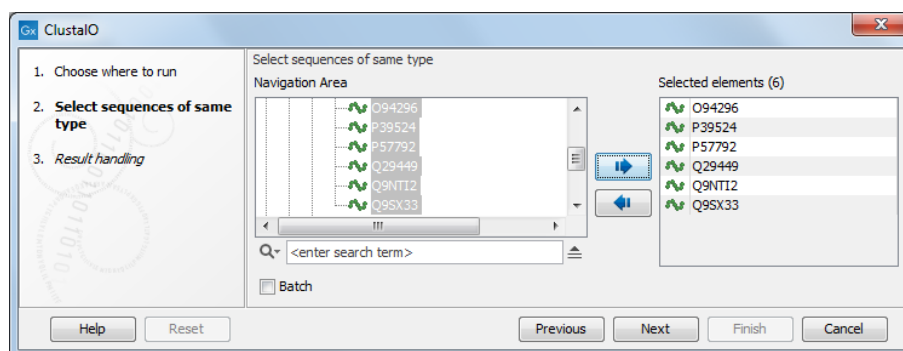


Figure 2.2: Selecting sequences for the alignment.

2011].

The ClustalO program in the Additional Alignments plugin has no user defined settings. Hence, to run the ClustalO program, you simply have to select the sequences to align and decide whether to open or save the result. Please note that the result consists of two output files: an alignment file and a phylogenetic tree file.

2.2 ClustalW

The ClustalW alignment method was in the mid nineties improved over previous progressive alignment methods [Thompson et al., 1994] (see http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=7984417). Even though this alignment method is more than ten years old it is still a highly popular alignment method and has become the method of choice to many researchers. One of the reasons for the popularity is the availability for most computer platforms and the easy integration on websites.

There is still one problem for ClustalW as for all other stand alone alignment programs. None of them can visualize annotations on the aligned sequences. Most researches align their own sequence to some reference sequences, which are often retrieved from one or more of the large public databases, e.g. GenBank. As input these programs require a Fasta file and they typically export in their own alignment format. The ClustalW program, when used in a CLC Workbench, uses ClustalW for the alignment calculation but still retains annotations on the sequences when the alignment is displayed (see figure 2.4).

One of the advantages of ClustalW is that it uses very little computer memory on rather large sequences so you can align large sequences without having a state-of-the-art computer.

ClustalW parameter settings

ClustalW has a single parameter to set: *Guide Tree Algorithm*: Select Normal or Quick. Quick uses a fast but not as accurate algorithm for the alignment guide tree.

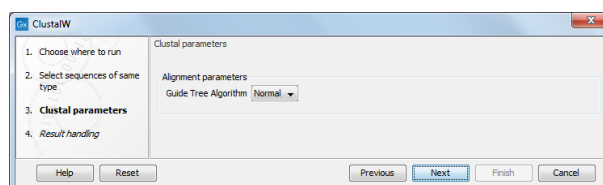


Figure 2.3: Setting ClustalW alignment parameters

The output of the ClustalW alignment can be seen in figure 2.4. In addition to the alignment output

file, a phylogenetic tree file is also generated.

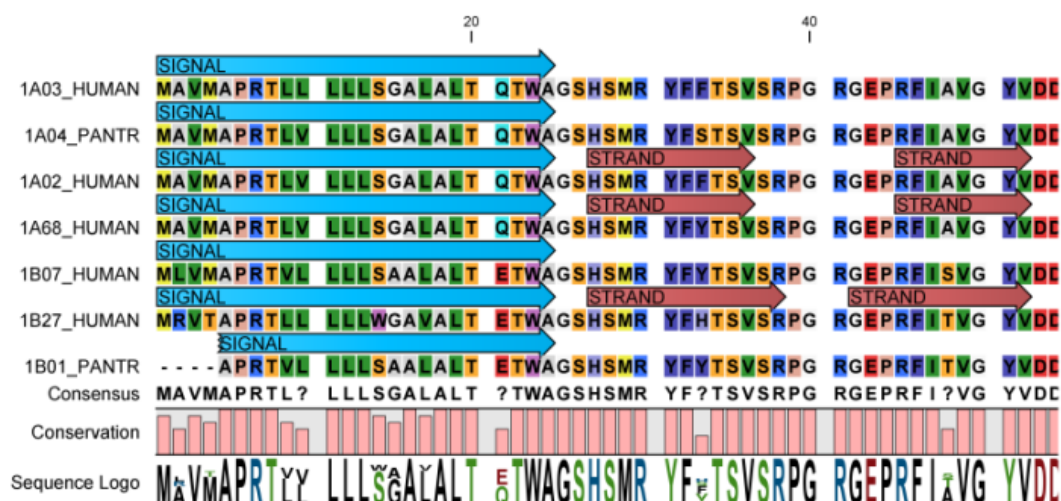


Figure 2.4: Output of ClustalW alignment. Annotations from the original sequence is retained on the output, and is shown when setting the Annotation layout to Show annotation.

2.3 MUSCLE

MUSCLE was described in [Edgar, 2004b] and [Edgar, 2004a]. See <http://www.drive5.com/muscle/> to find the various command line options.

MUSCLE parameter settings

- *Find Diagonals (faster for similar diagonals)*
- *Maximum iterations* If you are aligning a large number of sequences or if the sequences are very long, it is advisable to limit the number of iterations. A good number in such cases is 2.
- *Maximum hours* Maximum time to run in hours. The actual time may exceed the requested limit by a few minutes. Decimals are allowed, so 1.5 means one hour and 30 minutes. When the maximum hours limit is met, MUSCLE will be terminated without producing any alignment.
- *Maximum Memory in mb* Set maximum memory to be allocated to run. It is not recommended to exceed the default value (1000 mb).

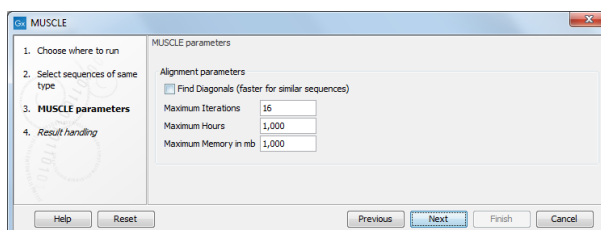


Figure 2.5: Setting MUSCLE alignment parameters

When using the MUSCLE program, there is only one output, that is an alignment file.



Chapter 3

Install and uninstall plugins

Additional Alignments plugin is installed as a plugin.

Note: In order to install plugins and modules, the Workbench must be run in administrator mode. On Linux and Mac, it means you must be logged in as an administrator. On Windows, you can do this by right-clicking the program shortcut and choosing "Run as Administrator".

Plugins are installed and uninstalled using the plugin manager.

Help in the Menu Bar | Plugins... () or Plugins () in the Toolbar

The plugin manager has two tabs at the top:

- **Manage Plugins.** This is an overview of plugins that are installed.
- **Download Plugins.** This is an overview of available plugins on QIAGEN Aarhus server.

3.1 Install

To install a plugin, click the **Download Plugins** tab. This will display an overview of the plugins that are available for download and installation (see figure 3.1).

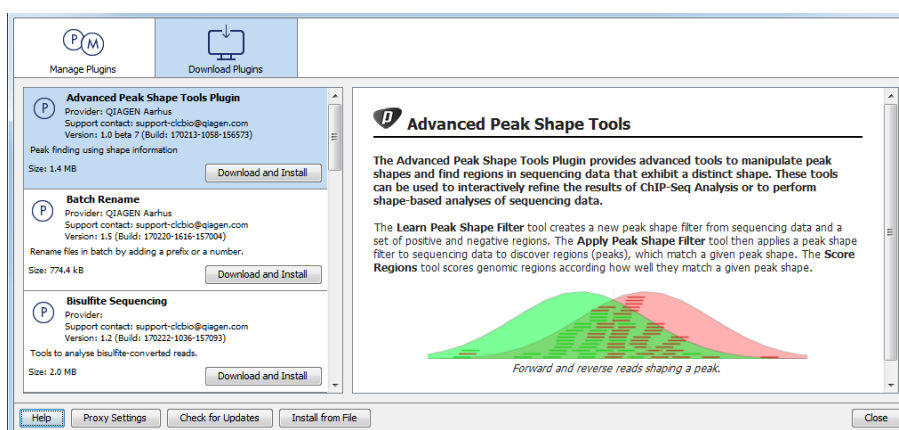


Figure 3.1: The plugins that are available for download.

Select Additional Alignments plugin to display additional information about the plugin on the right

side of the dialog. Click **Download and Install** to add the plugin functionalities to your workbench.

Accepting the license agreement

The end user license agreement (EULA) must be read and accepted as part of the installation process. figure 3.2.

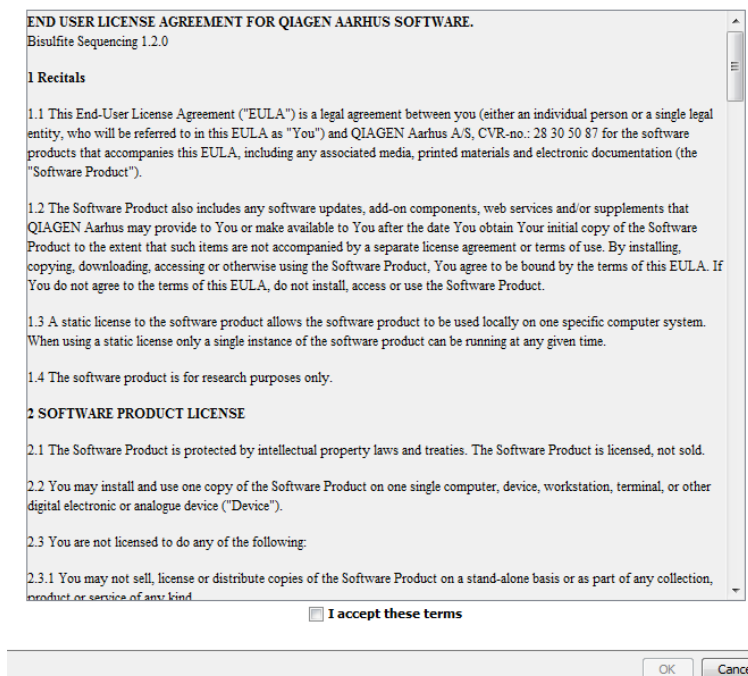


Figure 3.2: The End User License Agreement is presented during the installation process.



Please read the EULA text carefully, and if you agree to it, check the box next to the text **I accept these terms**. If further information is requested from you, please fill this in before clicking on the **Finish** button.

If Additional Alignments plugin is not shown on the server but you have the installer file on your computer (for example if you have downloaded it from our website), you can install the plugin by clicking the **Install from File** button at the bottom of the dialog and specifying the plugin *.cpa file saved on your computer.

When you close the dialog, you will be asked whether you wish to restart the workbench. The plugin will not be ready for use until you have restarted.

3.2 Uninstall

Plugins are uninstalled using the plugin manager:

Help in the Menu Bar | **Plugins...** () or **Plugins** () in the **Toolbar**

This will open the dialog shown in figure 3.3.

The installed plugins are shown in the **Manage plugins** tab of the plugin manager. To uninstall, select Additional Alignments plugin and click **Uninstall**.

If you do not wish to completely uninstall the plugin, but you do not want it to be used next time you start the Workbench, click the **Disable** button.

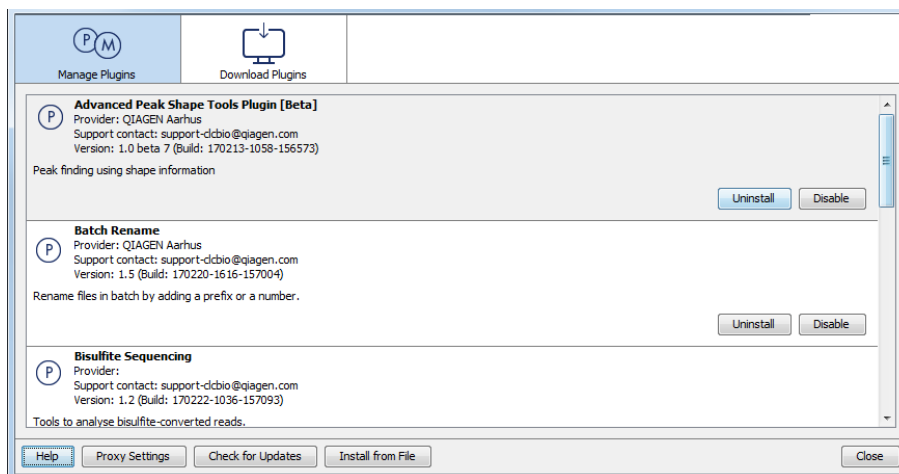


Figure 3.3: *The plugin manager with plugins installed.*

When you close the dialog, you will be asked whether you wish to restart the workbench. The plugin will not be uninstalled until the workbench is restarted.

Bibliography

- [Edgar, 2004a] Edgar, R. C. (2004a). Muscle: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5:113.
- [Edgar, 2004b] Edgar, R. C. (2004b). Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32(5):1792–1797.
- [Edgar and Batzoglou, 2006] Edgar, R. C. and Batzoglou, S. (2006). Multiple sequence alignment. *Curr Opin Struct Biol*, 16(3):368–373.
- [Plyusnin and Holm, 2012] Plyusnin, I. and Holm, L. (2012). Comprehensive comparison of graph based multiple protein sequence alignment strategies. *BMC Bioinformatics*, 13:64.
- [Sievers et al., 2011] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soeding, J., Thompson, J. D., and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol*, 7:539.
- [Thompson et al., 1994] Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22):4673–4680.