Metadata Import Plugin
User manual

User manual for
Metadata Import Plugin 1.0

Windows, Mac OS X and Linux

August 30, 2013

This software is for research purposes only.

CLC bio
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark

Contents

Chapter 1

Introduction to the Metadata Import Plugin

1.1 Introduction to the Metadata Import Plugin

The Metadata Import Plugin has been developed to facilitate the import, storage and handling of sequences and their metadata. So far it has only been possible to manually import tabular sequence data to the CLC Workbench one entry at the time. With the development of the Metadata Import Plugin, simultaneous import of multiple tabular data has greatly been facilitated. In addition, the Metadata Import Plugin provides a platform that enables handling of stored sequence properties.

The Metadata Import Plugin is an add-on module to the CLC Workbench consisting of the two individual tools that are listed below and described in detail in the following chapters.

- **Sequences in Table Format**. Imports sequences from a tabular data source.

- **Associate with Metadata**. Makes it possible to add metadata to existing objects.

Currently, the tools support two types of files:

- Plain text: .txt, .csv

- Excel spreadsheet (Excel 97 and newer.): .xls, .xlsx

Chapter 2

System requirements and installation of the Metadata Import Plugin

2.1 System requirements

The system requirements of the Metadata Import Plugin are:

- Windows XP, Windows Vista, or Windows 7, Windows Server 2003 or Windows Server 2008

- Mac OS X 10.6 or later. However, Mac OS X 10.5.8 is supported on 64-bit Intel systems.

- Linux: Red Hat 5.0 or later. SUSE 10.2 or later. Fedora 6 or later.

- 32 or 64 bit

- 1 GB RAM required

- 2 GB RAM recommended

- 1024 x 768 display recommended

- CLC Genomics Workbench

2.2 How to install a plug-in

Plug-ins are installed using the plug-in manager[1]:

> **Help in the Menu Bar | Plug-ins and Resources... (⊡)**

> or **Plug-ins (⊡) in the Toolbar**

The plug-in manager has four tabs at the top:

- **Manage Plug-ins.** This is an overview of plug-ins that are installed.

- **Download Plug-ins.** This is an overview of available plug-ins on CLC bio's server.

---

[1] In order to install plug-ins on Windows, the Workbench must be run in administrator mode: Right-click the program shortcut and choose "Run as Administrator". Then follow the procedure described below.

- **Manage Resources.** This is an overview of resources that are installed.

- **Download Resources.** This is an overview of available resources on CLC bio's server.

To install a plug-in, click the **Download Plug-ins** tab. This will display an overview of the plug-ins that are available for download and installation (see figure 2.1).
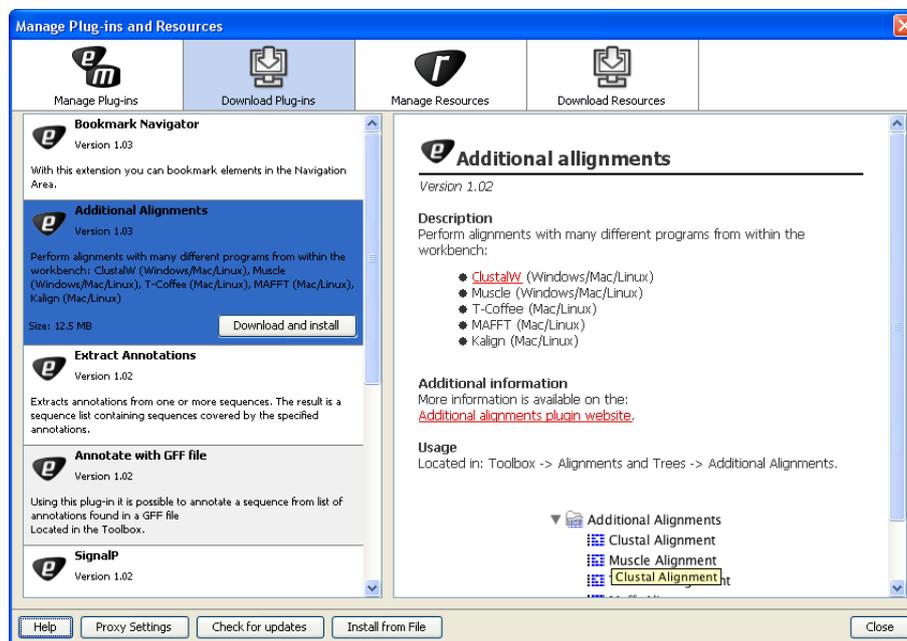

Figure 2.1: *The plug-ins that are available for download.*

Clicking a plug-in will display additional information at the right side of the dialog. This will also display a button: **Download and Install**.

Click the Metadata Import Plugin and press **Download and Install**. A dialog displaying progress is now shown, and the plug-in is downloaded and installed.

If the Metadata Import Plugin Plug-in is not shown on the server, and you have it on your computer (e.g. if you have downloaded it from our web-site), you can install it by clicking the **Install from File** button at the bottom of the dialog. This will open a dialog where you can browse for the plug-in. The plug-in file should be a file of the type ".cpa".

When you close the dialog, you will be asked whether you wish to restart the CLC Genomics Workbench. The plug-in will not be ready for use before you have restarted.

2.3 How to uninstall a plug-in

Plug-ins are uninstalled using the plug-in manager:

> **Help in the Menu Bar** | **Plug-ins and Resources... (**🖳**)**

> or **Plug-ins (**🖳**) in the Toolbar**
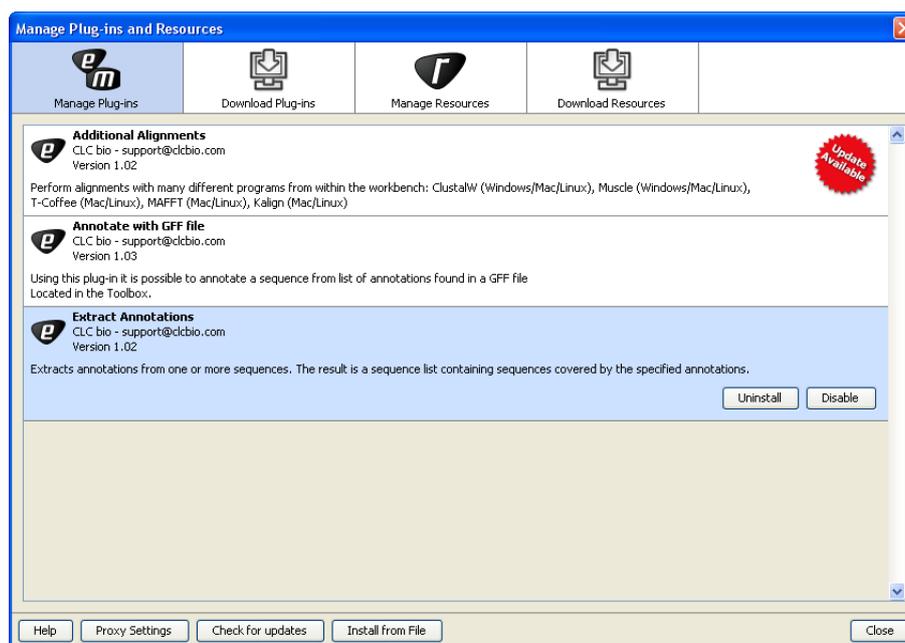
This will open the dialog shown in figure 2.2.

Figure 2.2: *The plug-in manager with plug-ins installed.*

The installed plug-ins are shown in this dialog. To uninstall:

Click the Metadata Import Plugin | Uninstall

If you do not wish to completely uninstall the plug-in but you don't want it to be used next time you start the Workbench, click the **Disable** button.

When you close the dialog, you will be asked whether you wish to restart the workbench. The plug-in will not be uninstalled before the workbench is restarted.

Chapter 3

Import Sequences

3.1   What is the Import Sequences in Table Format tool?

The Import Sequences in Table Format tool enables import of sequences stored in tabular format. Such data can be imported as individual sequences or as a list of sequences.

When using the Import Sequences in Table Format tool it is necessary to define input information using the Attribute Manager to create the relevant attributes. How to find and use the Attribute Manager is described in detail in section 3.2. Attributes can be defined prior to using the Import Sequences in Table Format tool or after the sequences have been imported. If the attributes are defined after the sequences have been imported, they can be assigned at a later point using the Associate with Metadata tool.

3.2   Customized attributes on data locations

The CLC Genomics Workbench makes it possible to define location-specific attributes on all elements stored in a data location. This could be company-specific information such as LIMS id, freezer position etc. Note that the attributes scheme belongs to a location, so if you have added multiple locations, they will have their own separate set of attributes.

Note! A Metadata Import Plugin is available. The plugin consists of two tools: "Import Sequences in Table Format" and "Associate with metadata". These tools allow sequences to be imported from a tabular data source and make it possible to add metadata to existing objects.

3.2.1   Configuring which fields should be available

To configure which fields that should be available[1]:

> **right-click the data location | Location | Attribute Manager**

This will display the dialog shown in figure 3.1.

Click the **Add Attribute** ( ) button to create a new attribute. This will display the dialog shown in figure 3.2.

---

[1]If the data location is a server location, you need to be a server administrator to do this
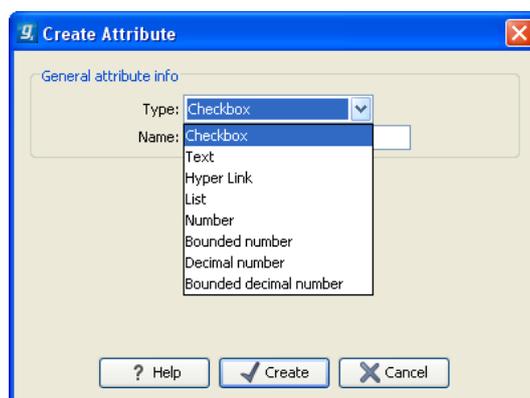
Figure 3.1: *Adding attributes.*

Figure 3.2: *The list of attribute types.*

First, select what kind of attribute you wish to create. This affects the type of information that can be entered by the end users, and it also affects the way the data can be searched. The following types are available:

- **Checkbox**. This is used for attributes that are binary (e.g. true/false, checked/unchecked and yes/no).

- **Text**. For simple text with no constraints on what can be entered.

- **Hyper Link**. This can be used if the attribute is a reference to a web page. A value of this type will appear to the end user as a hyper link that can be clicked. Note that this attribute can only contain one hyper link. If you need more, you will have to create additional attributes.

- **List**. Lets you define a list of items that can be selected (explained in further detail below).

- **Number**. Any positive or negative integer.

- **Bounded number**. Same as number, but you can define the minimum and maximum values that should be accepted. If you designate some kind of ID to your sequences, you can use the bounded number to define that it should be at least 1 and max 99999 if that is the range of your IDs.

- **Decimal number**. Same as number, but it will also accept decimal numbers.

- **Bounded decimal number**. Same as bounded number, but it will also accept decimal numbers.

When you click **OK**, the attribute will appear in the list to the left. Clicking the attribute will allow you to see information on its type in the panel to the right.

3.2.2 Editing lists

Lists are a little special, since you have to define the items in the list. When you click a list in the left side of the dialog, you can define the items of the list in the panel to the right by clicking **Add Item** ( ) (see figure 3.3).
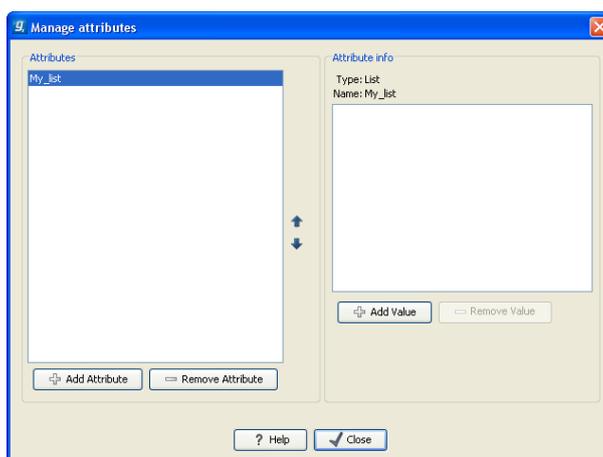
Figure 3.3: *Defining items in a list.*

Remove items in the list by pressing **Remove Item** ( ).

3.2.3 Removing attributes

To remove an attribute, select the attribute in the list and click **Remove Attribute** ( ). This can be done without any further implications if the attribute has just been created, but if you remove an attribute where values have already been given for elements in the data location, it will have implications for these elements: The values will not be removed, but they will become static, which means that they cannot be edited anymore. They can only be removed (see more about how this looks in the user interface below).

If you accidentally removed an attribute and wish to restore it, this can be done by creating a new attribute of exactly the same name and type as the one you removed. All the "static" values will now become editable again.

When you remove an attribute, it will no longer be possible to search for it, even if there is "static" information on elements in the data location.

Renaming and changing the type of an attribute is not possible - you will have to create a new one.

3.2.4   Changing the order of the attributes

You can change the order of the attributes by selecting an attribute and click the **Up** and **Down** arrows in the dialog. This will affect the way the attributes are presented for the user as described below.

3.3   How to run the Import Sequences in Table Format tool

The Import Sequences in Table Format tool can be accessed in the following two ways:

If you are working with the *CLC Genomics Workbench*, the Import Sequences in Table Format tool can be accessed from the toolbar:

> **Import | Sequences in Table Format (🗒)**

To open the Import Sequences in Table Format tool from the *CLC Main Workbench*:

> **File | Sequences in Table Format (🗒)**

This opens the dialog shown in figure 3.4.
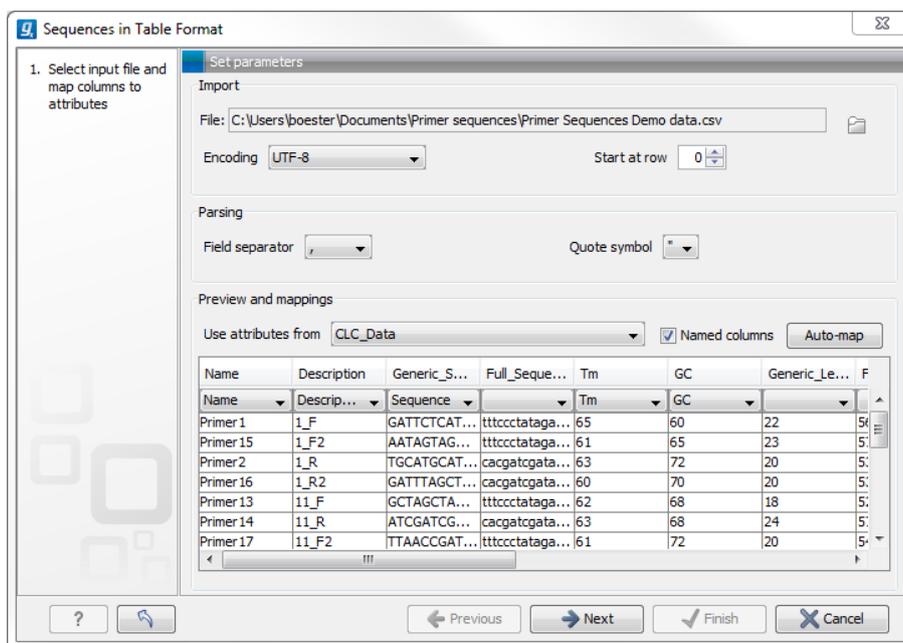
Figure 3.4: *Import Sequences in Table Format wizard step 1. Select the tabular data to import, specify parsing parameters when handling plain text files, and assign columns to the relevant attributes. In this example a primer list is imported. The list contains different information about the individual primers.*

Under **Import**, start out by selecting the tabular data to import by clicking on the folder (🗁). The selected data will now be shown in the wizard table.

Set the *Encoding* to UTF-8 (already selected as default option) or change it to the relevant encoding, if necessary.

Select the number of lines to skip before parsing. This can be done with the *Start at row* function

(figure 3.5).

Next, set the **Parsing** parameters (if handling plain text files) using the *Field separator* to specify the character used to separate fields/cells and the *Quote symbol* to specify the type of quote used in the file.
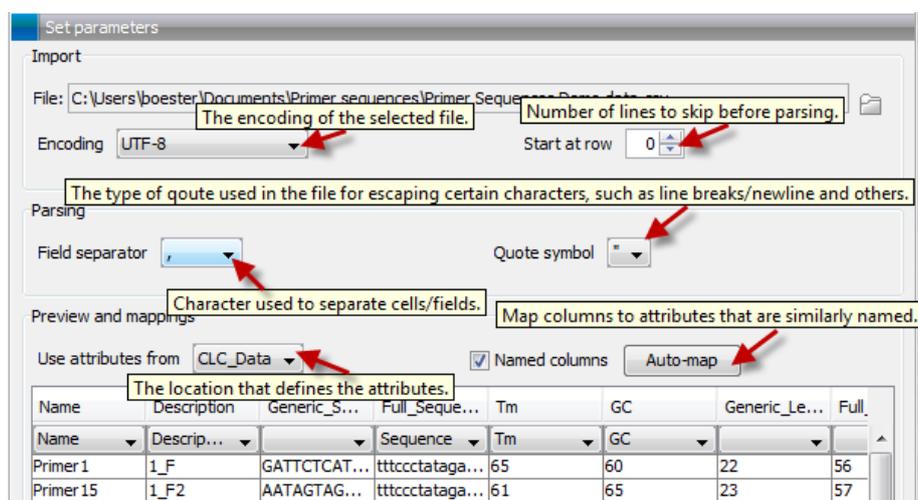

Figure 3.5: *Import Sequences in Table Format wizard step 1. Specify the separator character and quote type as well as the number of lines to skip before parsing. Tick* Named columns *if relevant and assign attributes to the columns.*

Next, under **Preview and mappings** and *Use attributes from* function, specify which attribute location to be used. Tick *Named columns* if the first row contains the name of the columns. Finally, columns can be assigned to the relevant attributes. This can be done manually by clicking on the drop-down lists located over each column or, when *Named columns* is ticked, automatically using the Auto-map button. If using the Auto-map button, manual adjustments can afterwards be made with the drop-down lists.

Note! The only column heading that must be assigned is "Sequence". Assignment of all other attributes is optional.

Click **Next**. This leads to the **Results handling** step shown in figure 3.6.

In this step, the parameters for output handling must be specified. Ticking "Output sequences separately" will import sequences as individual objects. Ticking "Output sequences in a list" will result in the tabular data being imported as a list of sequences.

After selecting whether the imported data should be opened or saved, click **Finish**.

The primers are now imported in the CLC Workbench. When the imported data are opened in the CLC Workbench, they will per default be opened in the **Graphical Sequence List** (▤) view (figure 3.7).

Clicking on the icon **Show Table** (▦) will show the imported primers in a table view (figure 3.8 top). To create a new sequence list containing only a subset of the sequences, click on the relevant rows in the Sequence Table and the "Create New Sequence List" button (figure 3.8 bottom).
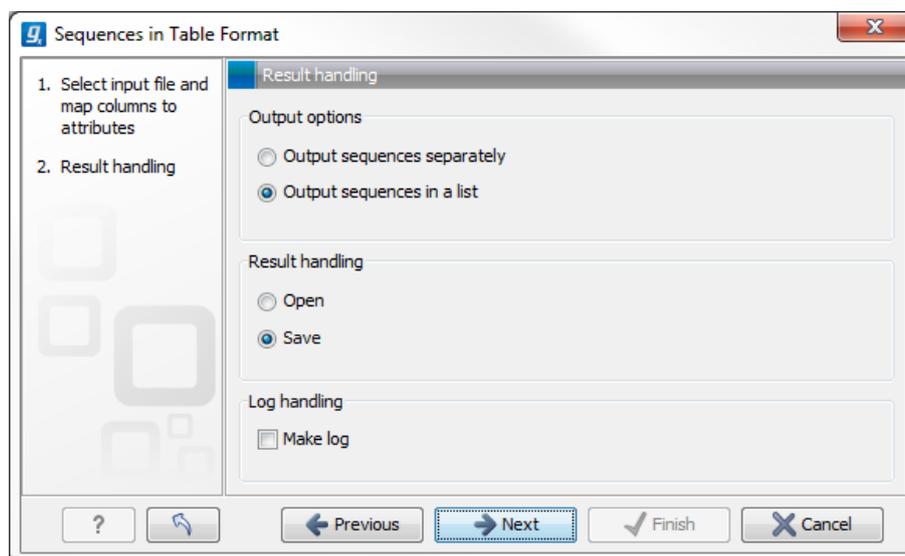
Figure 3.6: *Import Sequences in Table Format wizard step 2. Select whether imported tabular data should be presented as individual sequences or as a list of sequences.*
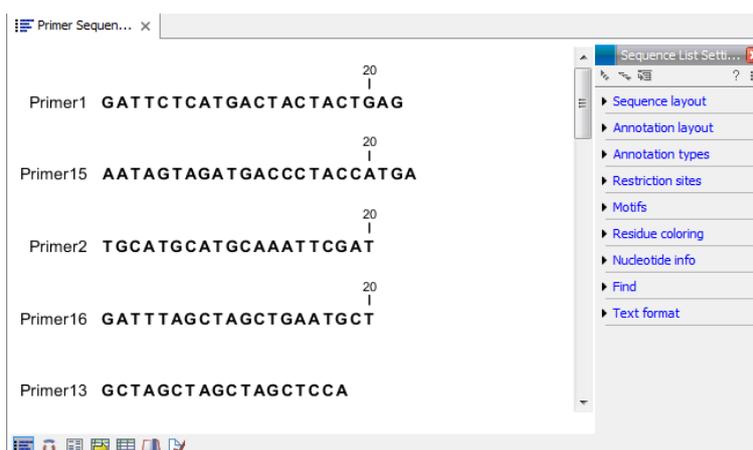

Figure 3.7: *Import Sequences output. The imported tabular data opens per default in the Grafical Sequence List mode.*

Figure 3.8: *Import Sequences in Table Format output. The imported tabular data presented in the Sequence Table (top). A sequence list containing the selected subset of the sequences (bottom).*

Chapter 4

Associate with Metadata

4.1   What is the Associate with Metadata tool?

The Associate with Metadata tool enables addition of metadata to existing objects. In order to associate metadata to existing objects, identical names are required in the two datasets.

4.2   How to run the Associate with Metadata tool

The Associate with Metadata tool is found under **File** (see Figure 4.1).
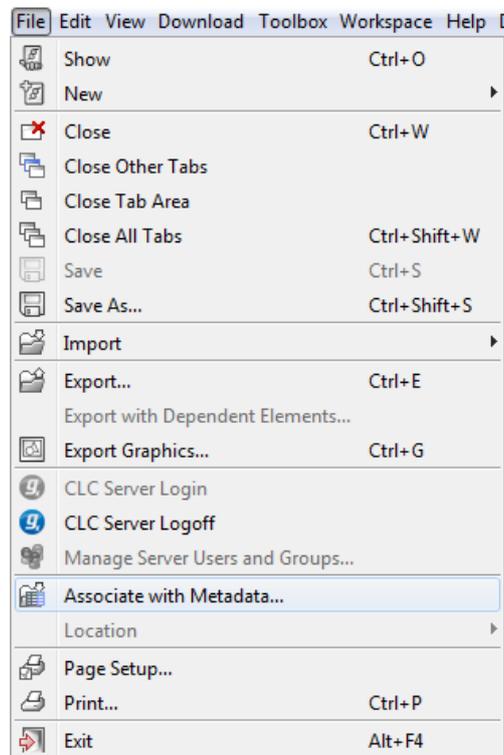
File | Associate with Metadata (▦)

Figure 4.1: *Click on "Associate with metadata".*

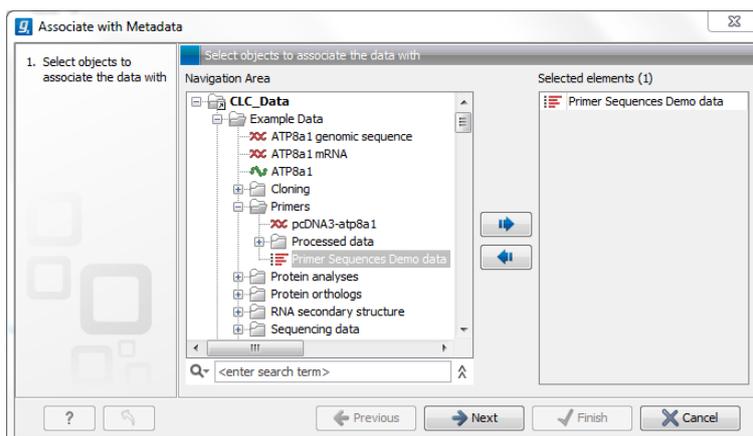This opens the dialog shown in figure 4.2.

Figure 4.2: *Select the object to associate data with. In this example we wish to add information to a list of primers.*

Select the relevant object(s) to associate data with and click **Next**. This leads to the step shown in figure 4.3.

Note! More than one object can be selected, however, due to the fact that attributes are assigned specifically to the individual data locations, the selected objects must be from the same data location to be in concordance with the relevant attributes.
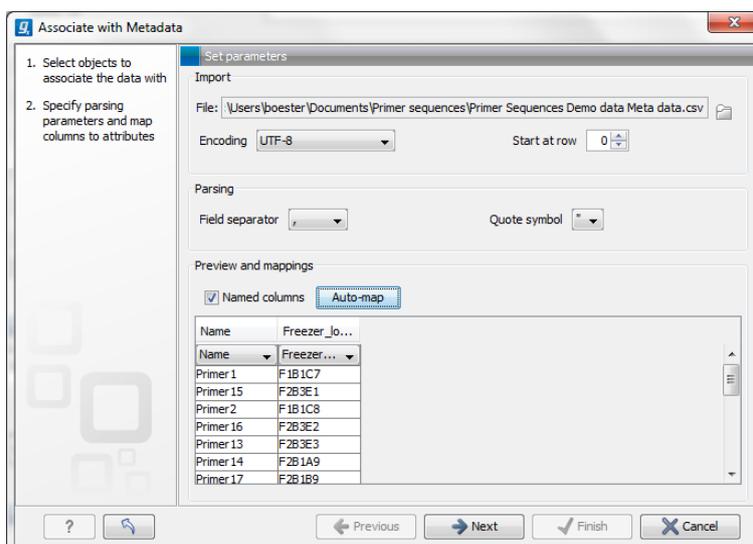

Figure 4.3: *Select the file containing the relevant metadata. In this example we wish to add information about the freezer location. In order to be able to associate the freezer location information with the existing information in the primer table, we need to have an identical (and unique) identifier that is found in both the original data and the metadata. In this case the name is used as identifier.*

Under **Import**, start out by selecting the relevant metadata to import by clicking on the folder (📁). The selected data will now be shown in the wizard table.

Set the *Encoding* to UTF-8 (already selected as default option) or change it to the relevant

encoding, if necessary.

Select the number of lines to skip before parsing using the *Start at row* function.

Next, set the **Parsing** parameters using the *Field separator*, when handling plain text files, to specify the character used to separate fields/cells and the *Quote symbol* to specify the type of quote used in the file to escape certain characters, such as line breaks, newline and others.

Assign attributes to the relevant columns. This can be done manually by clicking on the dropdown lists located over each column or automatically using the Auto-map button. If using the Auto-map button manual adjustments can be made with the dropdown buttons.

Click **Next** and after the **Result handling** step, click **Finish**.

The metadata has now been added to the primer sequences. Note that the metadata is added directly to the existing primer sequences without generating a new object. The added metadata can be viewed for each individual sequence by opening it and then clicking the "Show Element Info" button in the lower right corner in the view area (see figure 4.4).

Figure 4.4: *Select the table format and double click on one of the sequences (top). The attributes and the added metadata can now be viewed for the sequence of interest by clicking on the "Show Element Info" icon in the lover right corner of the view area (buttom).*