



BIONUMERIC Tutorial:

Setup and analysis using a public and a custom MLST scheme

1 Introduction

In this tutorial we will demonstrate how to perform Multi Locus Sequence Typing (MLST) starting from Whole Genome Sequences (WGS) by using the *MLST for WGS plugin*. The plugin allows the installation of multiple public and custom MLST schemes per BIONUMERIC database and uses an extended BLAST approach to identify and optionally extract MLST alleles from genome sequences.

2 Preparing the database

2.1 Introduction to the demonstration database

As the *MLST for WGS plugin* uses an extended BLAST approach to identify MLST alleles from genome sequences, we first need to prepare a BIONUMERIC database which contains genomic sequences. How to import genomes into a BIONUMERIC database or how to obtain a genome assembly from raw sequence reads is explained in dedicated tutorials and will not be repeated here. In this tutorial we will make use of the genome sequences already available in our **WGS demo database** for *Escherichia coli*.

The **WGS_demo_database_for_Escherichia_coli** can be downloaded directly from the *BIONUMERIC Startup* window (see [2.2](#)), or restored from the back-up file available on our website (see [2.3](#)).

2.2 Option 1: Download demo database from the Startup Screen

1. Click the  button, located in the toolbar in the *BIONUMERIC Startup* window.

This calls the *Tutorial databases* window (see [Figure 1](#)).

2. Select **WGS_demo_database_for_Escherichia_coli** from the list and select **Database > Download** (.
3. Confirm the installation of the database and press **<OK>** after successful installation of the database.

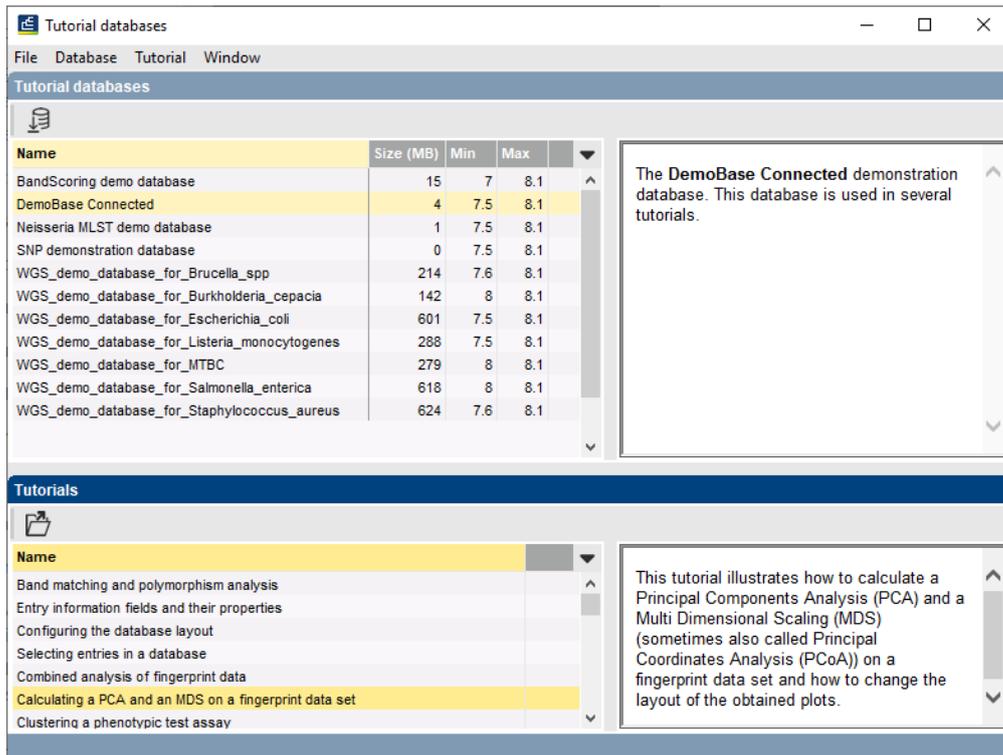


Figure 1: The *Tutorial databases* window, used to download the demonstration database.

4. Close the *Tutorial databases* window with **File > Exit**.

The **WGS_demo_database_for_Escherichia_coli** appears in the *BIONUMERICS Startup* window.

5. Double-click the **WGS_demo_database_for_Escherichia_coli** in the *BIONUMERICS Startup* window to open the database.

2.3 Option 2: Restore demo database from back-up file

A BIONUMERICS back-up file of the demo database for *Escherichia coli* is also available on our website. This backup can be restored to a functional database in BIONUMERICS.

6. Download the file `WGS_EC.bnbk` file from <https://www.bionumerics.com/download/sample-data>, under 'WGS_demo_database_for_Escherichia_coli'.



In contrast to other browsers, some versions of Internet Explorer rename the `WGS_EC.bnbk` database backup file into `WGS_EC.zip`. If this happens, you should manually remove the `.zip` file extension and replace with `.bnbk`. A warning will appear ("If you change a file name extension, the file might become unusable."), but you can safely confirm this action. Keep in mind that Windows might not display the `.zip` file extension if the option "Hide extensions for known file types" is checked in your Windows folder options.

7. In the *BIONUMERICS Startup* window, press the  button. From the menu that appears, select **Restore database....**
8. Browse for the downloaded file and select **Create copy**. Note that, if **Overwrite** is selected, an existing database will be overwritten.

- Specify a new name for this demonstration database, e.g. "WGS_Ecoli_demobase".
- Click <**OK**> to start restoring the database from the backup file.
- Once the process is complete, click <**Yes**> to open the database.

The *Main* window is displayed.

3 Installing the MLST for WGS plugin

Proceed as follows to install the *MLST for WGS plugin*:

- Call the *Plugins and Scripts* dialog box from the *Main* window with **File > Install / remove plugins...** (⌘P).
- Select the *MLST for WGS plugin* from the list and press the <**Install**> button.
- Confirm the installation of the plugin.
- Press <**Close**> to close the *Plugins and Scripts* dialog box and to continue to the *Main* window.
- Close and reopen the database to activate the features of the *MLST for WGS plugin*.
- Press <**Yes**> to allow the creation of additional tables required by the *MLST for WGS plugin*.

The *MLST for WGS plugin* installs menu items in the main menu of the software under **MLST for WGS** (see Figure 2).

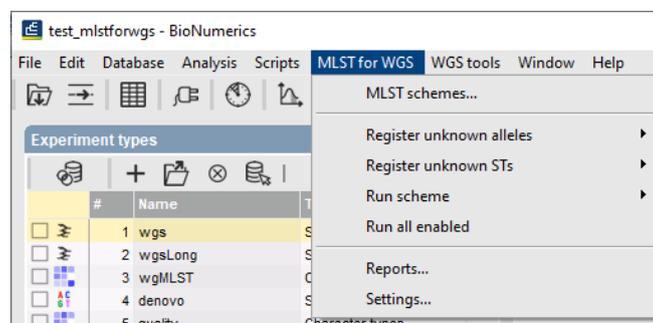


Figure 2: New menu items after installation of the *MLST for WGS plugin*.

4 Adding MLST schemes

The *MLST for WGS plugin* allows multiple MLST schemes to be used in a single BIONUMERICs database. In this tutorial we will install two public schemes and a custom scheme for *E. coli*.

- Select **MLST for WGS > MLST schemes...** (see Figure 3) to open the *Configure MLST schemes* dialog box.

The *Configure MLST schemes* dialog box lists all MLST schemes that are installed in the database (initially, this dialog shows up empty). From this dialog, MLST schemes can be added, viewed, exported to text files, updated and removed.

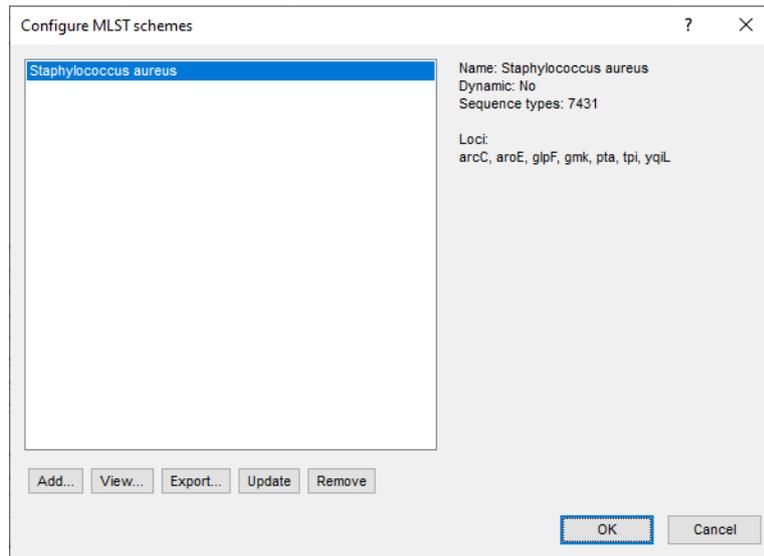


Figure 3: The *Configure MLST schemes* dialog box.



Once an MLST scheme is added, it cannot be modified anymore. It is only possible to remove the scheme and re-create the scheme with other parameters.

2. Press **<Add...>** in the *Configure MLST schemes* dialog box to start adding a new MLST scheme. This action opens the *MLST scheme wizard* (see Figure 4).

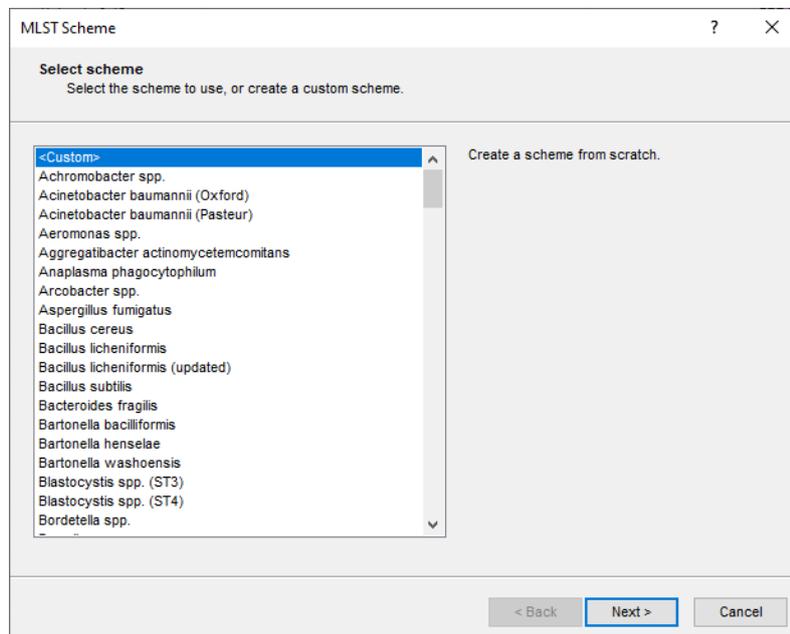


Figure 4: The *MLST scheme wizard*: select MLST scheme page.

The list displays all organisms for which an MLST repository is available online and an additional "<Custom>" option. For the highlighted organism in the list, the organism name, number of available sequence types, abbreviations of the MLST loci and the URL where the database is located are listed in the right panel.

4.1 Adding a public MLST scheme

We will start with the installation of two public schemes in the BIONUMERICS database.

3. Select the ***Escherichia coli (Achtman)*** scheme from the list and press **<Next>** to proceed to the next page in the *MLST scheme wizard* (see Figure 5).

Locus	Location
adk	http://rest.pubmlst.org/db/pubmlst_escherichia_seqdef/loci/...
fumC	http://rest.pubmlst.org/db/pubmlst_escherichia_seqdef/loci/...
gyrB	http://rest.pubmlst.org/db/pubmlst_escherichia_seqdef/loci/...
icd	http://rest.pubmlst.org/db/pubmlst_escherichia_seqdef/loci/...
mdh	http://rest.pubmlst.org/db/pubmlst_escherichia_seqdef/loci/...
purA	http://rest.pubmlst.org/db/pubmlst_escherichia_seqdef/loci/...
recA	http://rest.pubmlst.org/db/pubmlst_escherichia_seqdef/loci/...

Figure 5: The *MLST scheme wizard*: configure scheme page.

In this page of the wizard, the MLST scheme can be further configured and modified, if necessary. Most users will want to simply use an MLST scheme from the online list for typing their isolates, without making any modifications to the MLST scheme. This ensures a consistent nomenclature regarding allele numbers, sequence types and clonal complexes (if available).



Editing parameters from an existing MLST scheme (other than the scheme name) in this step will make it incompatible with the public MLST scheme.

4. Press **<Finish>** to download and store the MLST scheme data (alleles, profiles, sequence types, etc.) in the BIONUMERICS connected database.



In case of large MLST databases available online, the installation may take several minutes, depending on the speed of your internet connection.

5. Press **<Add...>** in the *Configure MLST schemes* dialog box and repeat the previous steps to install the ***Escherichia coli (Pasteur)*** scheme.

The *Configure MLST schemes* dialog box will now list the two installed schemes (see Figure 6).

It is recommended to update an online MLST scheme on regular intervals to make sure that the latest alleles and sequence type definitions become available locally in the BIONUMERICS database. To update the highlighted MLST scheme in the *Configure MLST schemes* dialog box, press **<Update>**.

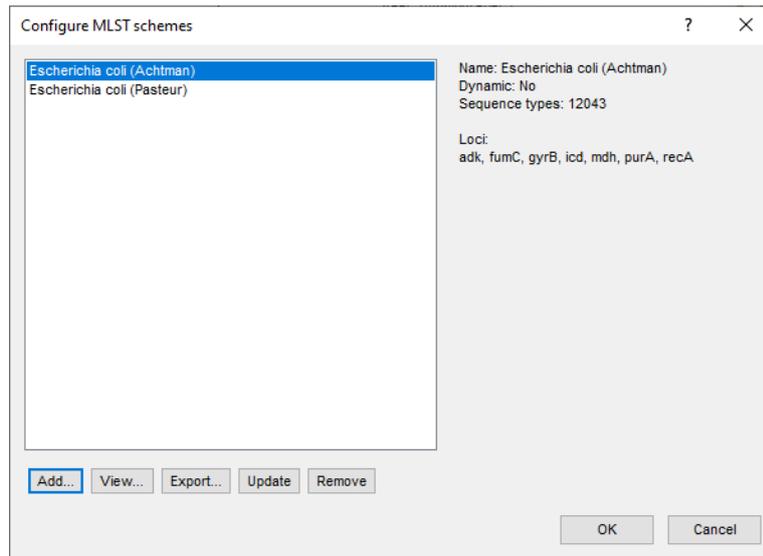


Figure 6: The *Configure MLST schemes* dialog box listing the two installed schemes.

4.2 Adding a custom MLST scheme

We will now add our own MLST scheme to the BIONUMERICS database. This can be useful when no public scheme exists yet for the organism being studied. The locus and sequence type files of the custom MLST scheme are typically stored locally or on a network drive. The locus fasta files should at least contain one allele sequence per locus and the sequence type file should at least contain one sequence type.

The custom scheme which will be used in this tutorial to explain the work flow can be downloaded from the BIONUMERICS website (<https://www.bionumerics.com/download/sample-data>, click on "Custom E. coli MLST scheme").

6. Press <**Add...**> in the *Configure MLST schemes* dialog box, select the <**Custom**> option from the list and press <**Next**> to proceed to the next page in the *MLST scheme* wizard (see Figure 7).
7. Enter a name for your custom scheme e.g. "Custom *E. coli* scheme".

The location for the **Sequence types** definition can be entered as an URL or a path to a file on your own computer or a network drive. The expected format is either Comma Separated Values (CSV) or Tab Separated Values (TSV). The field can be left empty if no sequence types are used (e.g. in case of single-locus schemes).

8. Enter the path to the sequence type tsv file on your computer.
9. Press <**Add...**> to add the first locus. This action opens the *MLST Scheme* dialog box (see Figure 8).

The custom scheme used in this example consists of three loci (i.e. locus1, locus2 and locus3).

10. Enter the name of the first locus in our custom scheme i.e. "locus1" and enter the path to the corresponding fasta file on your computer.
11. Repeat the previous step for the other two loci in our custom scheme i.e. "locus2" and "locus3".

When starting an MLST scheme from scratch, it frequently occurs that new alleles and sequence types are found. New alleles and sequence types are indicated as 'Unknown' in the MLST reports

The screenshot shows a dialog box titled "MLST Scheme" with a subtitle "Configure scheme" and the instruction "Configure the MLST scheme name, loci and sequence types." It contains a "Name" text box, a "Sequence types" text box, and a table with columns "Locus" and "Location". Below the table are "Add...", "Edit...", and "Remove" buttons, and a checkbox labeled "Expand scheme dynamically". At the bottom right are "< Back", "Finish", and "Cancel" buttons.

Figure 7: Configure a custom scheme.

The screenshot shows the same dialog box with "Name" set to "arcC" and "Location" set to "http://rest.pubmist.org/db/pubmist_saureus_seqdef/loci/arcC/allele". The "OK" button is highlighted.

Figure 8: The *MLST Scheme* dialog box.

and show up as missing values in the MLST character experiments and information fields. Checking the option **Expand scheme dynamically** allows you to add new alleles and sequence types as they are found (see 7).

12. Check the option **Expand scheme dynamically**.

The *MLST scheme* wizard now looks like Figure 9.

13. Press <**Finish**> to download and store the MLST scheme data (alleles, profiles, sequence types, etc.) in the BIONUMERICs connected database.

The *Configure MLST schemes* dialog box will now list the two installed public schemes and the installed custom scheme (see Figure 10).

14. Press <**OK**> to close the *Configure MLST schemes* dialog box.

5 Manage the MLST genotyping settings

Before we can run the MLST schemes installed in our BIONUMERICs database, we have to manage the MLST genotyping settings.

1. Access the settings for the *MLST for WGS plugin* by pressing **MLST for WGS > Settings...** in the *Main* window.

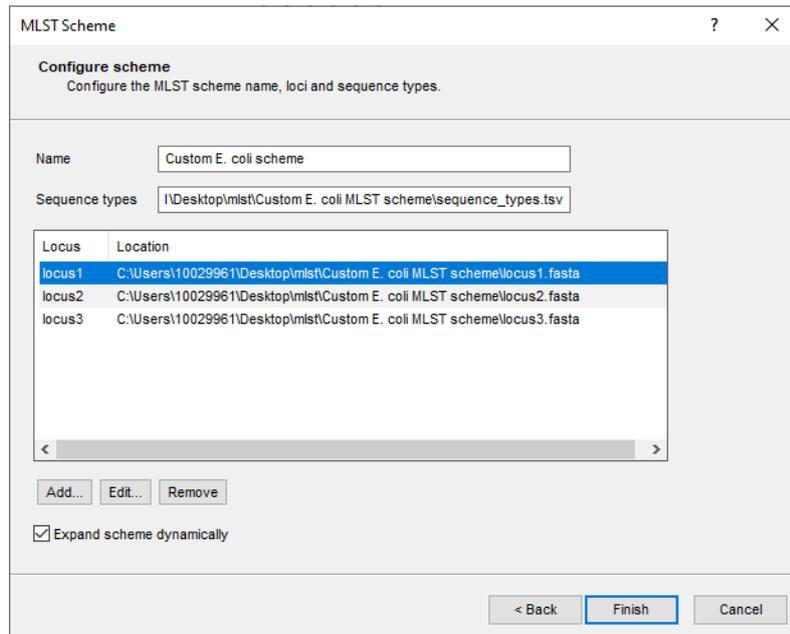


Figure 9: Configure a custom scheme.

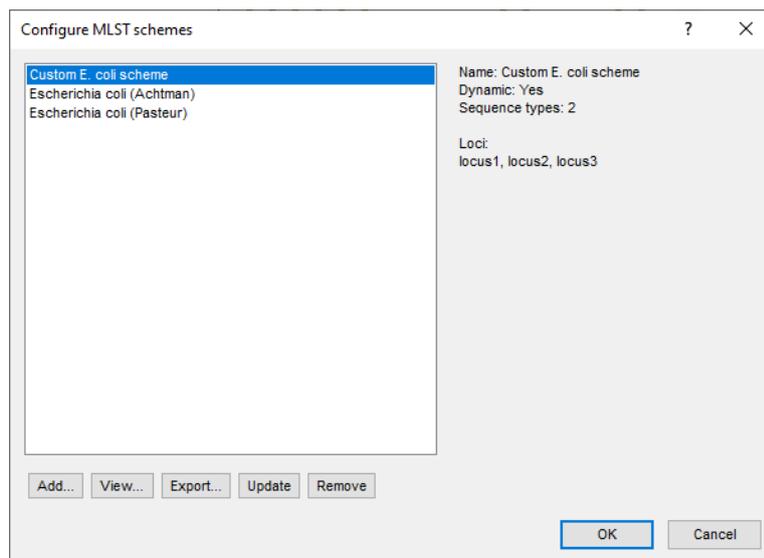


Figure 10: The installed MLST schemes.

The *General* tab of the *MLST genotyping settings* dialog box (see Figure 11) holds settings for the MLST reports and for general processing.



It is crucial to specify at least the **input sequence experiment** in the settings. If not specified, the error message "The input sequence experiment must be set to process entries." will be generated when the plugin is run.

In the **Enabled features** list, all features offered by the plugin are listed and enabled by default. If specific analyses are not required, you can uncheck them here to save on processing time and to omit the corresponding sections from the reports.

2. Select **BioProject** as information field to include in the MLST report, select the sequence experiment which contains the genomic sequences and make sure all installed schemes are enabled.

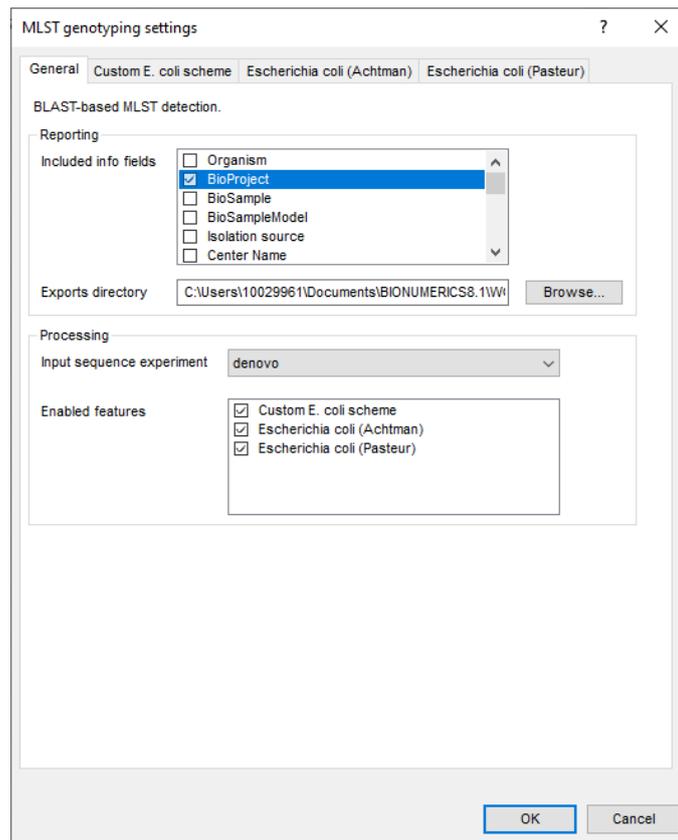


Figure 11: The *General* tab of the *MLST genotyping settings* dialog box.

For each MLST scheme in the database, an additional tab is displayed in the *MLST genotyping settings* dialog box (see Figure 12).

3. Select the tab for the custom *E. coli* scheme.

Information about the **Scheme** is displayed as static text: the **Name** of the scheme, the number of **Sequence types** and for each locus its name and allele count.

In the **Results** panel, the experiment types and entry information fields to which the screening results will be written can be dictated. The **MLST experiment** is a character type experiment in which the MLST allelic profiles will be stored for this MLST scheme. Typically, this experiment type is used for cluster analysis when comparing isolates based on MLST. The **Sequence type field** and **Clonal complex field** are entry information fields in which the sequence type and clonal complex are stored, respectively. Optionally, the MLST alleles detected on the genome sequence can be extracted and stored in separate sequence experiments. To do so, press the **<Change...>** button next to **Locus extraction experiments**. This action will open the *Change sequence experiment* dialog box.

4. Use the drop-down menus to choose an existing experiment type or information field or select the **<Create>** option to create a new experiment type or information field (see Figure 12). A default name is suggested, but you can adjust this if you want to.
5. Repeat the previous step for the other two installed MLST schemes.
6. Press **<OK>** to close the *MLST genotyping settings* dialog box.

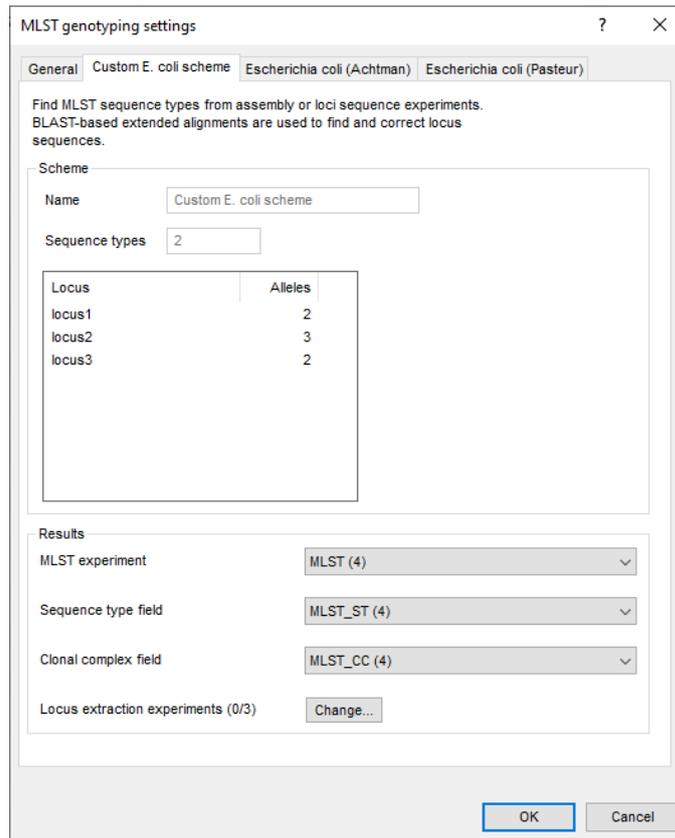


Figure 12: A scheme specific tab.

6 Run MLST analyses

Once the plugin is installed and the settings have been specified, the actual screening of the genome sequences of the selected entries is an easy process.

1. Select a single entry by holding the **Ctrl**-key and clicking on the entry in the *Database entries* panel. Alternatively, use the **space bar** or click the ballot box next to the entry. In order to select a range of entries, hold the **Shift**-key and click on the last entry in the range.

MLST analyses can be run for all schemes checked in the *MLST genotyping settings* dialog box at once (using **MLST for WGS > Run all enabled**) or for each scheme separately with the corresponding command (**MLST for WGS > Run scheme > scheme_name**).

The analysis time increases proportionally with the number of selected entries and the number of enabled MLST schemes. It also depends on the number of loci in the schemes and whether or not allele sequences are extracted. A complete analysis may take up to several minutes or even hours.

2. Select **MLST for WGS > Run all enabled**.

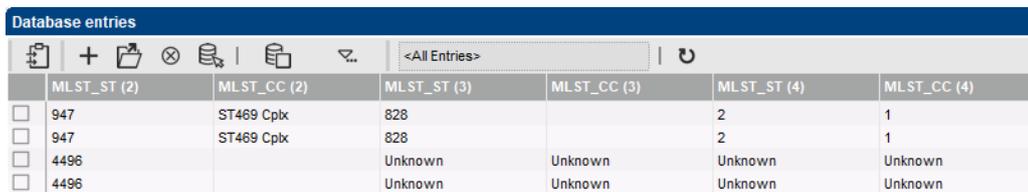
When the analysis is finished, the progress bar disappears. The MLST results are stored in the MLST character experiments (see Figure 13) and information fields (see Figure 14) which you have defined in the *MLST genotyping settings* dialog box.



Character	Value	Mapping
adk	6	<+>
fumC	65	<+>
gyrB	5	<+>
icd	1	<+>
mdh	9	<+>
purA	13	<+>
recA	4	<+>

Press insert to add character

Figure 13: A MLST character experiment.



MLST_ST (2)	MLST_CC (2)	MLST_ST (3)	MLST_CC (3)	MLST_ST (4)	MLST_CC (4)
<input type="checkbox"/> 947	ST469 Cplx	828		2	1
<input type="checkbox"/> 947	ST469 Cplx	828		2	1
<input type="checkbox"/> 4496		Unknown	Unknown	Unknown	Unknown
<input type="checkbox"/> 4496		Unknown	Unknown	Unknown	Unknown

Figure 14: MLST information fields.

7 Register new alleles and sequence types

New alleles and sequence types can be added easily to a custom scheme.

1. Select an entry or multiple entries in the database for which at least one allele is new (i.e. indicated as "unknown" in the MLST report).
2. Register unknown alleles in the BIONUMERICs database for the selected entries by selecting **MLST for WGS > Register unknown alleles > scheme_name** (with scheme_name the name of your custom MLST scheme).

The software determines how many new alleles can be registered and ask for confirmation to add these to the MLST scheme in the database. Once the alleles are added, a message indicates that the MLST analysis should be re-run before the allele numbers are assigned.

3. Run the scheme again to assign the new alleles.

Now the new alleles are registered in the database, new sequence types can be registered as well.

4. Register unknown profiles and sequence types in the BIONUMERICs database for the selected entries by selecting **MLST for WGS > Register unknown STs > scheme_name**.

The software determines how many new sequence types can be registered and ask for confirmation to add these to the MLST scheme in the database. Once the sequence types are added, a message indicates that the MLST analysis should be re-run before the sequence types are assigned.

5. Run the scheme again to assign the new sequence types.

8 Reports

1. Open an MLST report for the selected entries by selecting **MLST for WGS > Reports....**

The *Report* window contains an MLST report for each of the selected entries (see Figure 15).

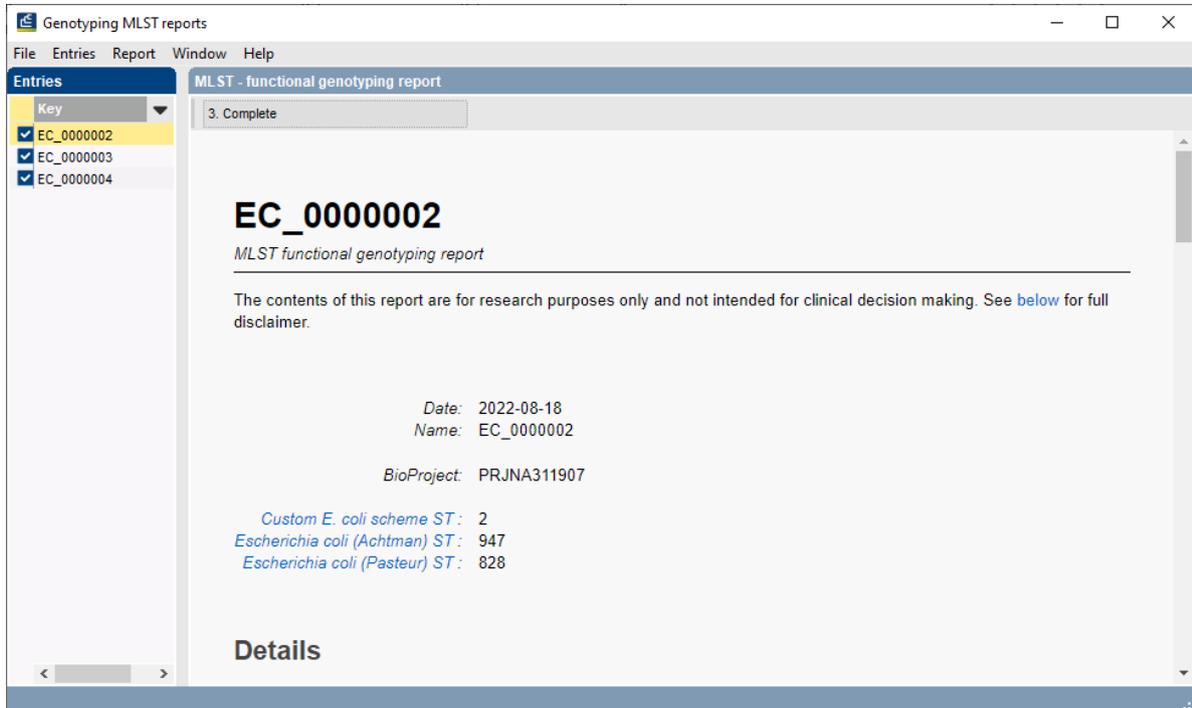


Figure 15: The MLST report.

2. Select another entry in the *Entries* panel to update the results in the *MLST report* panel.

The creation date of the report (**Date**), the Key (**Name**), and information fields checked in the *MLST genotyping settings* dialog box are displayed in the *MLST report* panel.

3. Select **Report > Report styles** in the *Report* window and make sure the option **Summary** is selected.

A summary of the results is displayed in the *Report* window.

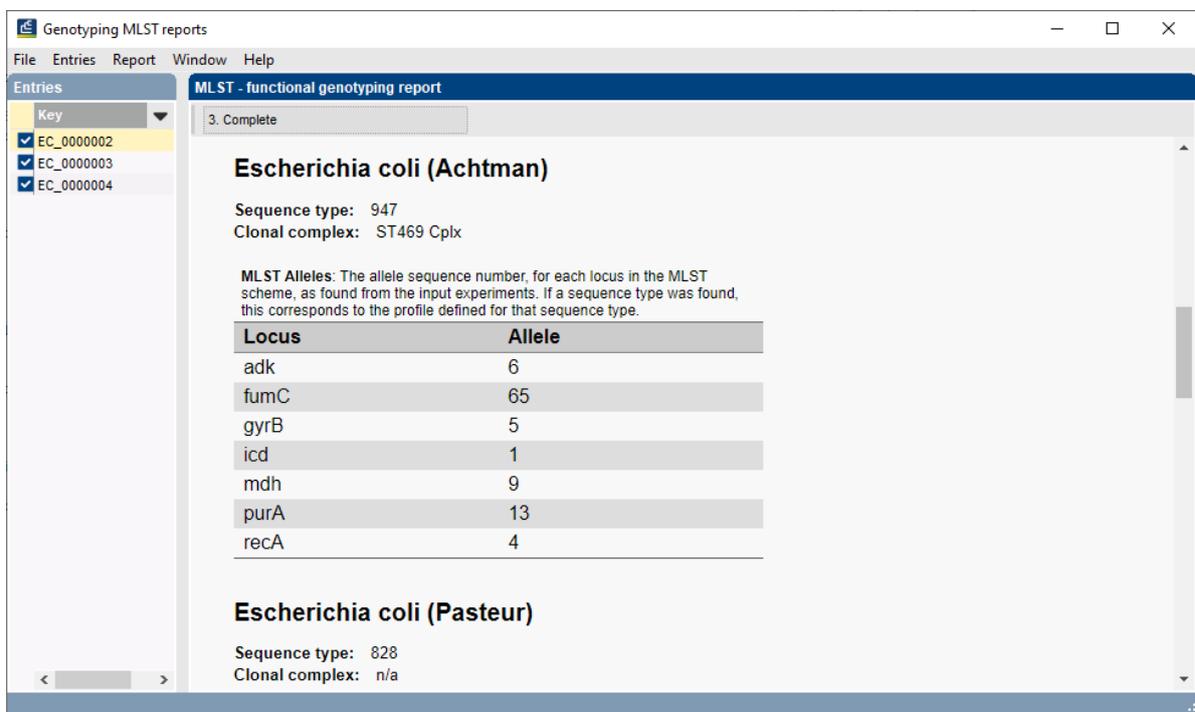
4. Select **Report > Report styles** in the *Report* window and select the option **Complete** (see Figure 15).

In the **Complete** view, the summarized results as well as all available details are shown.

5. Click on a hyperlink of one of the enabled schemes to display the detailed results in the *MLST report* panel (see Figure 16).

6. Select **File > Exit** to close the *Report* window.

For more detailed information on the MLST analysis and interpretation of the reported results, please check the MLST for WGS plugin manual.



The screenshot displays the 'Genotyping MLST reports' application window. The main panel shows the 'MLST - functional genotyping report' for '3. Complete'. The report details for **Escherichia coli (Achtman)** include a sequence type of 947 and a clonal complex of ST469 Cplx. A table lists the MLST alleles for various loci. Below this, the report for **Escherichia coli (Pasteur)** shows a sequence type of 828 and a clonal complex of n/a. The left sidebar shows a list of entries with keys EC_0000002, EC_0000003, and EC_0000004.

Locus	Allele
adk	6
fumC	65
gyrB	5
icd	1
mdh	9
purA	13
recA	4

Figure 16: Detailed results section of the MLST report.