



BIONUMERICS®

version 8 - PLUGINS



MLST for WGS plugin

Contents

1	Starting and setting up BIONUMERICS	5
1.1	Introduction	5
1.2	Startup program	5
1.3	Installation of the MLST for WGS plugin	6
2	Managing MLST schemes	9
2.1	Introduction	9
2.2	Adding an MLST scheme	10
2.3	Viewing an MLST scheme	11
2.4	Exporting an MLST scheme	12
2.5	Updating an MLST scheme	13
2.6	Removing an MLST scheme	13
3	MLST settings	15
3.1	Accessing the MLST settings	15
3.2	General settings	15
3.3	MLST scheme-specific settings	15
4	MLST analysis	19
4.1	Selecting entries	19
4.2	Starting an analysis	19
5	MLST reports	21
5.1	Opening MLST reports	21
5.2	Report styles	21
5.3	Details section	21
5.4	Info section	23
6	Working with a custom MLST scheme	25
6.1	Introduction	25
6.2	Registering new alleles and sequence types	25
6.3	Typical workflow	25

NOTES

SUPPORT BY APPLIED MATHS, A BIOMÉRIEUX COMPANY

While the best efforts have been made in preparing this manuscript, no liability is assumed by the authors with respect to the use of the information provided.

Applied Maths, a bioMérieux company, will provide support to research laboratories in developing new and highly specialized applications, as well as to diagnostic laboratories where speed, efficiency and continuity are of primary importance. Our software thanks its current status for a part to the response of many customers worldwide. Please contact us if you have any problems or questions concerning the use of BIONUMERICS[®], or suggestions for improvement, refinement or extension of the software to your specific applications:

Applied Maths NV

Keistraat 120
9830 Sint-Martens-Latem
Belgium
PHONE: +32 9 2222 100
FAX: +32 9 2222 102
E-MAIL: BE-DAU-INFO@biomerieux.com
URL: <https://www.bionumerics.com>

Applied Maths, Inc.

11940 Jollyville Road, Suite 115N
Austin, Texas 78759
U.S.A.
PHONE: +1 512-482-9700
FAX: +1 512-482-9708
E-MAIL: US-DAU-INFO@biomerieux.com

LIMITATIONS ON USE

The BIONUMERICS[®] software, its plugin tools and their accompanying guides are subject to the terms and conditions outlined in the License Agreement. The support, entitlement to upgrades and the right to use the software automatically terminate if the user fails to comply with any of the statements of the License Agreement. No part of this guide may be reproduced by any means without prior written permission of the authors.

Copyright ©1998-2022, Applied Maths NV. All rights reserved.

BIONUMERICS[®] is a registered trademark of Applied Maths NV. All other product names or trademarks are the property of their respective owners.

BIONUMERICS® uses following third-party software tools and libraries:

- Python 3.8 release from the Python Software Foundation, <https://www.python.org/>
- Xerces library for XML input and output from the Apache Software Foundation, <https://xerces.apache.org/>
- NCBI toolkit version 2.11.0, <https://www.ncbi.nlm.nih.gov/BLAST/>
- SRA Toolkit, <https://ncbi.github.io/sra-tools/>
- Boost c++ libraries, <https://www.boost.org/>
- Samtools for interacting with SAM / BAM files, <https://www.htslib.org/download/>
- 7-Zip (7za.exe), <https://www.7-zip.org/>
- Zlib library, <https://zlib.net/>
- Pigz for parallel gzip compression, <https://zlib.net/pigz/>
- Cairo 2D graphics library version 1.12.14, <https://cairographics.org/>
- Crypto++ library version 5.5.2, <https://www.cryptopp.com/>
- OpenSSL library, <https://www.openssl.org/>
- libSVM library for Support Vector Machines, <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- SQLite version 3.7.17, <https://www.sqlite.org/>
- pymzML Python module version 2.4.7, <https://github.com/pymzml/pymzML>
- NumPy Python library version 1.19.1, <https://www.numpy.org/>
- BioPython Python library version 1.78, <https://www.biopython.org/>
- pyodbc Python module version 4.0.30, <https://pypi.org/project/pyodbc/>
- Jinja2 Python library version 2.11.2, <https://pypi.org/project/Jinja2/>
- MarkupSafe Python library version 1.1.1, <https://pypi.org/project/MarkupSafe/>
- regex Python library version 2.5.91, <https://pypi.org/project/regex/>
- Chromium Embedded Framework, <https://bitbucket.org/chromiumembedded/cef/wiki/Home>
- SPAdes genome assembler version 3.15.3, <https://bioinf.spbau.ru/spades> *
- SKESA version 2.3.0, <https://github.com/ncbi/SKESA/releases>
- Unicycler version 0.5.0, <https://github.com/rrwick/Unicycler/releases> *
- Velvet for Windows, source code can be downloaded from <https://www.bionumerics.com/download/open-source>
- Bowtie2 version 2.2.5 (<https://bowtie-bio.sourceforge.net/bowtie2/index.shtml>)*
- SNAP version 2.0.0, <https://www.microsoft.com/en-us/research/project/snap/>
- RAxML version 8.2.11, <https://github.com/stamatak/standard-RAxML/releases>

- FastTree version 2.1.10, <https://www.microbesonline.org/fasttree/>
- CFSAN SNP pipeline version 2.2.0, <https://github.com/CFSAN-Biostatistics/snp-pipeline>
*
- Prokka version 1.14.5, <https://github.com/tseemann/prokka> *
- sourmash version 4.1.0, <https://github.com/dib-lab/sourmash> **
- SeqSero2 for Windows, source code can be downloaded from <https://www.bionumerics.com/download/open-source>
- Fastp version 0.22.0, <https://github.com/OpenGene/fastp>

*: On Calculation Engine only **: See license conditions below

Sourmash license conditions:

Copyright: 2016, The Regents of the University of California. License: BSD-3-Clause

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of The Regents of the University of California, nor the names of contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Chapter 1

Starting and setting up BIONUMERICS

1.1 Introduction


The *MLST for WGS plugin* is specifically designed for performing Multi Locus Sequence Typing (MLST) starting from Whole Genome Sequences (WGS). The plugin is based on the genotyping framework and uses an extended BLAST approach to identify and optionally extract MLST alleles from genome sequences.


In contrast to its predecessor the *MLST for Sanger sequencing plugin*, the *MLST for WGS plugin* allows the installation of more than one MLST scheme per BIONUMERICS database. MLST schemes can be selected from public online repositories (e.g. PubMLST, BIGSdb Pasteur, Enterobase, ...) or custom schemes can be created and maintained.


The *MLST for WGS plugin* is supported in the **BIONUMERICS-SEQ** and **BIONUMERICS-SUITE** configurations.

1.2 Startup program

Make sure the latest version of BIONUMERICS is installed (<https://www.bionumerics.com/download/software>). The installation manual can be downloaded from <https://www.bionumerics.com/download/manuals>.

When BIONUMERICS is launched from the Windows start panel or when the BIONUMERICS shortcut () on your computer's desktop is double-clicked, the **Startup program** is run. This program shows the *BIONUMERICS Startup* window (see Figure 1.1).

A new BIONUMERICS database is created from the Startup program by pressing the  button.

An existing database is opened in BIONUMERICS with  or by simply double-clicking on a database name in the list.

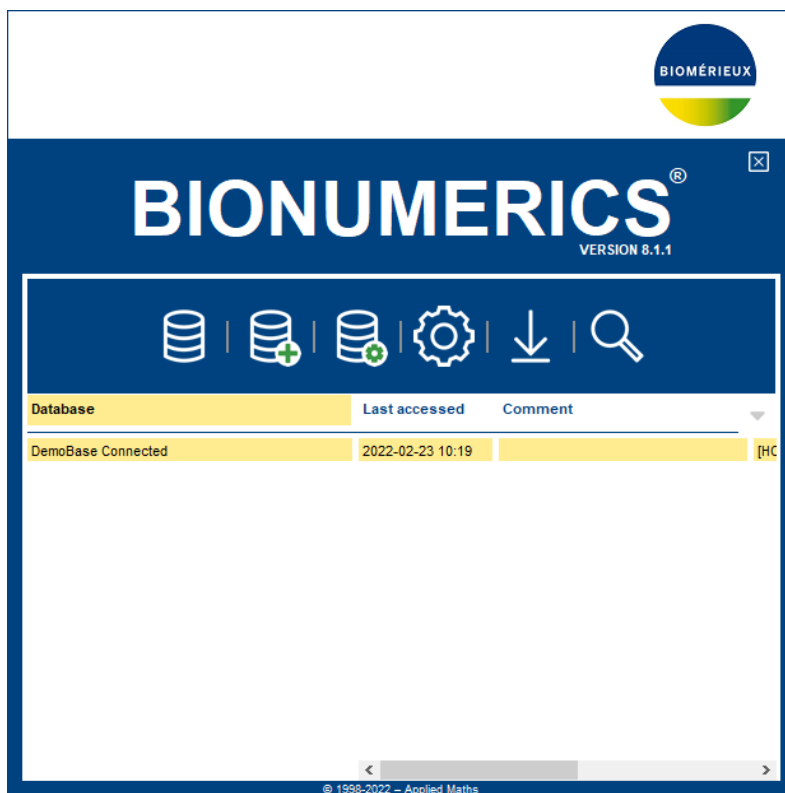


Figure 1.1: The *BIONUMERICS* Startup window.

1.3 Installation of the MLST for WGS plugin

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

- 3.1 Call the *Plugins and Scripts* dialog box from the *Main* window with **File > Install / remove plugins...** (🔧).
- 3.2 Select the *MLST for WGS plugin* from the list and press the **<Install>** button.
- 3.3 Confirm the installation of the plugin.
- 3.4 Press **<Close>** to close the *Plugins and Scripts* dialog box and to continue to the *Main* window.
- 3.5 Close and reopen the database to activate the features of the *MLST for WGS plugin*.
- 3.6 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 1.2).

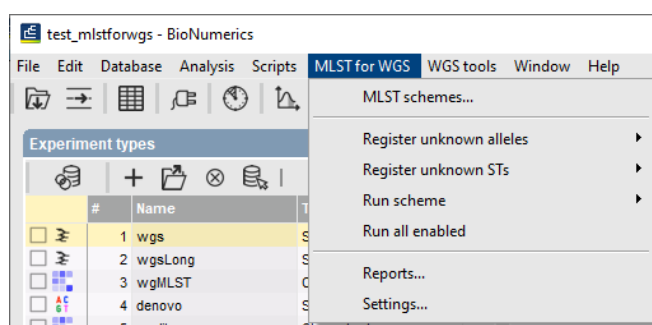


Figure 1.2: New menu items after installation of the *MLST for WGS plugin*

Chapter 2

Managing MLST schemes

2.1 Introduction

The *MLST for WGS plugin* allows multiple MLST schemes to be used in a single BIONUMERICS database.

MLST schemes can be managed via the *Configure MLST schemes* dialog box, which opens after selecting **MLST for WGS > MLST schemes...** (see Figure 2.1).

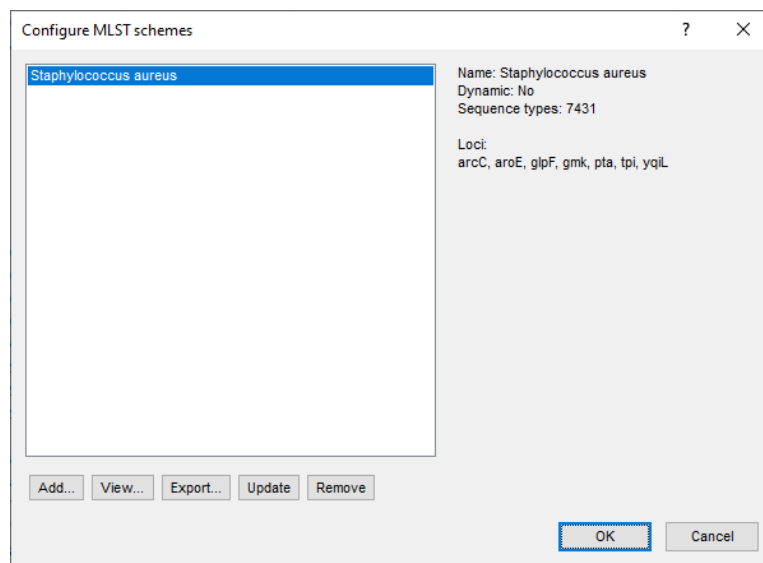


Figure 2.1: The *Configure MLST schemes* dialog box.

The *Configure MLST schemes* dialog box lists all MLST schemes that are installed in the database (initially, this dialog shows up empty). Details for the highlighted MLST scheme are displayed on the right.

From this dialog, MLST schemes can be added, viewed, exported to text files, updated and removed.



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.

2.2 Adding an MLST scheme

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 2.2).

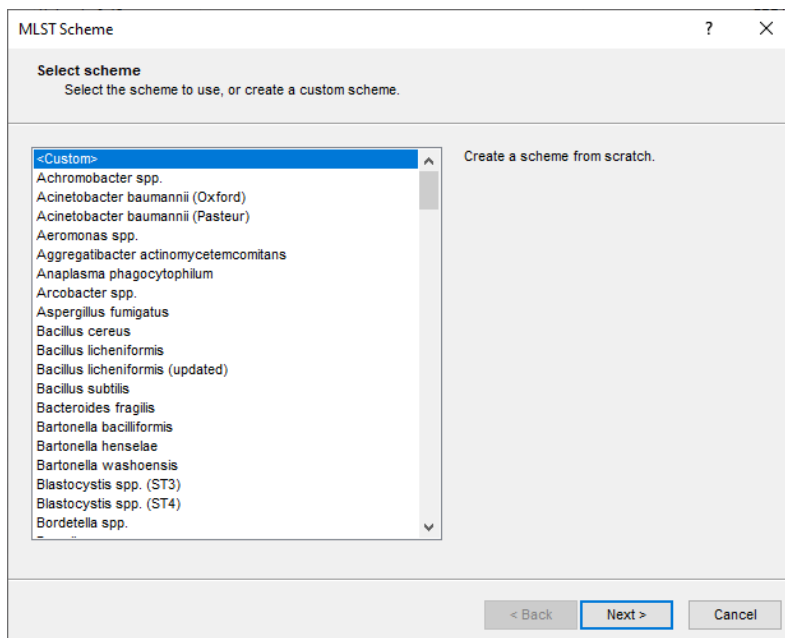


Figure 2.2: The *MLST scheme wizard*: select MLST scheme page.

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.

2.1 Select the most suitable scheme from the list and press **<Next>** to proceed to the next page in the *MLST scheme wizard* (see Figure 2.3).

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.

The **Name** of the MLST scheme can be freely edited, without effect on the actual scheme definition.

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).

To add a locus, press **<Add...>**. This action opens the *MLST Scheme* dialog box (see Figure 2.4).

Enter a **Name** for the locus and a **Location**. The latter can be an URL or the path to a file on your own computer or a network drive.

Press **<Edit...>** to edit the highlighted locus in the list. This opens the same *MLST Scheme* dialog

MLST Scheme ? X

Configure scheme
Configure the MLST scheme name, loci and sequence types.

Name:

Sequence types:

Locus	Location
arcC	http://rest.pubmist.org/db/pubmist_saureus_seqdef/loci/arc...
aroE	http://rest.pubmist.org/db/pubmist_saureus_seqdef/loci/aro...
glpF	http://rest.pubmist.org/db/pubmist_saureus_seqdef/loci/glpF...
gmk	http://rest.pubmist.org/db/pubmist_saureus_seqdef/loci/gmk...
pta	http://rest.pubmist.org/db/pubmist_saureus_seqdef/loci/pta...
tpi	http://rest.pubmist.org/db/pubmist_saureus_seqdef/loci/tpi/a...
yqil	http://rest.pubmist.org/db/pubmist_saureus_seqdef/loci/yqil...

Add... Edit... Remove

☐ Expand scheme dynamically

< Back Finish Cancel

Figure 2.3: The *MLST scheme* wizard: configure scheme page.

MLST Scheme ? X

Name:

Location:

OK Cancel

Figure 2.4: The *MLST Scheme* dialog box.

box as discussed above.

Press <**Remove**> to delete the highlighted locus in the list.

Checking the option **Expand scheme dynamically** allows you to add new alleles and sequence types as they are found. This option should only be used by scheme curators in the "master" database (see 6 for more information)

When the <**Finish**> button is pressed, the MLST scheme data (alleles, profiles, sequence types, etc.) is downloaded and stored in dedicated tables 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.

2.3 Viewing an MLST scheme

To see all settings of the highlighted MLST scheme in the *Configure MLST schemes* dialog box, press <**View...**>. This action opens the *View MLST scheme* dialog box (see Figure 2.5).

Following information is displayed about the MLST scheme: the **Name** of the scheme, whether or not the scheme is **Dynamic**, the number of sequence types currently available in the scheme (**Sequence type count**) and the location where the sequence types are defined (**Sequence types source**). Furthermore, information about each locus defined in the scheme is shown in a table

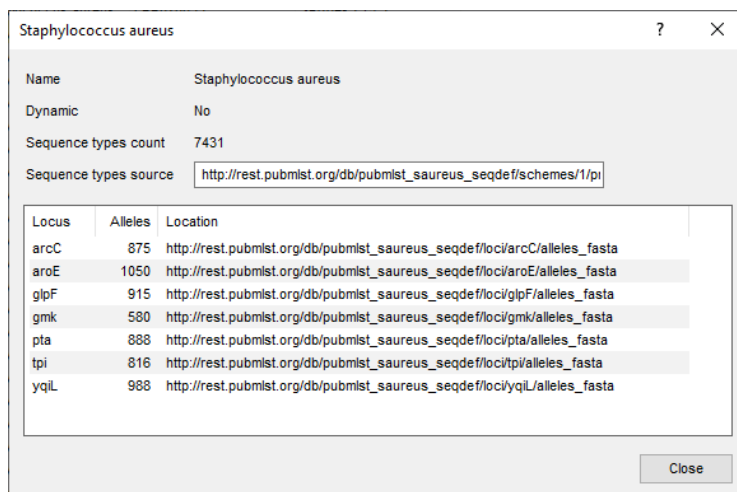


Figure 2.5: The *View MLST scheme* dialog box, here displaying the information of the *Staphylococcus aureus* MLST scheme on PubMLST.

with following columns:

- 'Locus': unique name for the locus within the scheme
- 'Alleles': number of alleles available for the locus
- 'Location': location (file path or URL) where the alleles are defined



All information in the *View MLST scheme* dialog box is static (i.e. read-only). In case one or more settings need to change, you should create a new MLST scheme with the correct settings.

2.4 Exporting an MLST scheme

To export the scheme definition of the highlighted MLST scheme in the *Configure MLST schemes* dialog box to text files, press <**Export...**>. This action opens the *Export MLST scheme* dialog box (see Figure 2.6).

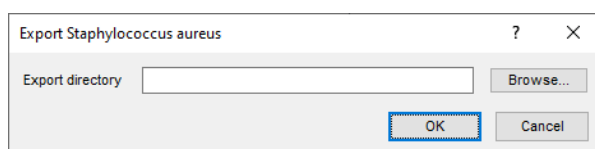


Figure 2.6: The *Export MLST scheme* dialog box.

Browse for an **Export directory** and press <**OK**> to export the MLST scheme definition to files.

For each locus, BIONUMERICS creates a separate FASTA file, containing the allele sequences that are locally stored in the database. The sequence types, clonal complex information (if available) and allelic profiles are stored in the tab-delimited `Sequence_types.tsv` file.

2.5 Updating an MLST scheme

To update the highlighted MLST scheme in the *Configure MLST schemes* dialog box, press <**Update**>. This action will update the MLST scheme data in the BIONUMERICS database with the information available at the stored location of the loci and sequence types.

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.

2.6 Removing an MLST scheme

To remove the highlighted MLST scheme in the *Configure MLST schemes* dialog box, press <**Remove**>. This action will delete all data associated with the MLST scheme from the BIONUMERICS database.

Chapter 3

MLST settings

3.1 Accessing the MLST settings

Settings for the *MLST for WGS plugin* can be accessed via **MLST for WGS > Settings...** in the *Main* window.

3.2 General settings

The *General* tab of the *MLST genotyping settings* dialog box (see Figure 3.1) holds settings for the MLST reports and for general processing.

Under **Reporting**, the entry information fields that will be displayed in the genotyping reports can be specified in the **Included info fields** list. Simply check the ballot box next to an information field name to include the field in the report.

The **Exports directory** can be specified for all exports from the genotyping reports. By default, the exported files are stored in a subdirectory of the database directory, but a different location can be selected via the **<Browse>** button or entered directly in the text box.

The **Input sequence experiment**, i.e. the sequence experiment containing the whole genome sequences to be screened, should be selected from the corresponding drop-down list. Select the **<Create>** option in case you wish to create a new sequence experiment type. In the latter case, make sure to import whole genome sequences in this experiment type before running the plugin.



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.

3.3 MLST scheme-specific settings

For each MLST scheme in the database (see 2), an additional tab is displayed in the *MLST genotyping settings* dialog box. See Figure 3.2 for an example with two installed MLST schemes.

MLST genotyping settings

General Staphylococcus aureus Staphylococcus aureus (DLST)

BLAST-based MLST detection.

Reporting

Included info fields

- ☐ Organism name
- ☐ Instrument
- ☐ Study accession
- ☐ Study title
- ☐ ST info
- ☐ outbreak

Exports directory C:\Users\Public\Documents\BioNumerics\homedir8\W Browse...

Processing

Input sequence experiment denovo

Enabled features

- ☒ Staphylococcus aureus
- ☐ Staphylococcus aureus (DLST)

OK Cancel

Figure 3.1: The *MLST genotyping settings* dialog box, *General* tab.

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. Use the drop-down menu to choose an existing experiment type or information field or select the **<Create>** option to create a new experiment type or information field. A default name is suggested, but you can adjust this if you want to.

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.

The column 'Name' in the dialog contains all targets available in the knowledge base. Via the drop-down list in the 'Sequence experiment' column, an existing sequence experiment type can be specified in which the extracted sequence will be stored. When the **<Create>** option is selected, a dialog will pop up asking for the name of the sequence experiment type. By default, the name of the target will be suggested.



An indication of how many loci are being extracted is given between brackets next to **Locus extraction experiments** in the *MLST genotyping settings* dialog box (see Figure 3.2).

MLST genotyping settings

General Staphylococcus aureus Staphylococcus aureus (DLST)

Find MLST sequence types from assembly or loci sequence experiments. BLAST-based extended alignments are used to find and correct locus sequences.

Scheme

Name: Staphylococcus aureus

Sequence types: 7431

Locus	Alleles
arcC	875
aroE	1050
glpF	915
gmK	580
pta	888
tpi	816
yqiL	988

Results

MLST experiment: MLST (2)

Sequence type field: MLST_ST (2)

Clonal complex field: MLST_CC (2)

Locus extraction experiments (0/7): Change...

OK Cancel

Figure 3.2: The *MLST genotyping settings* dialog box, *MLST scheme* tab.

Change PCR extraction Sequence experiment

Name	Sequence experiment
stx1-det	<None>
stx1a	<None>
stx1c	<None>
stx1d	<None>
stx2-det_F4_R1	<None>
stx2-det_F4-f_R1-e/f	<None>
stx2a_F2_R3	<None>
stx2a_F2_R2	<None>
stx2b	<None>
stx2c	<None>
stx2d_F1_R1	<None>
stx2d_F1_O55-R	<None>
stx2d_F1_R2	<None>
stx2e	<None>
stx2f	<None>

OK Cancel

Figure 3.3: The *Change sequence experiment* dialog box.

Chapter 4

MLST analysis

4.1 Selecting entries

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.

Analyses are performed on the selected entries in the database. For example, to select a single entry, hold the **Ctrl**-key and click 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.

More options for selecting entries can be found in the BIONUMERICS reference manual (see the Reference manual, Chapter Database entries).

4.2 Starting an analysis

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.

When the analysis is finished, the progress bar disappears. The MLST results are stored in the MLST character experiments and information fields which you have defined in the *MLST genotyping settings* dialog box. The settings can always be consulted or adjusted using **MLST for WGS > Settings...** (see [3](#) for details). If the settings are changed, the analysis should be re-run to reflect these changes.

Chapter 5

MLST reports

5.1 Opening MLST reports

Since the *MLST for WGS plugin* is based on the genotyping framework, it generates reports similar to the genotyping plugin reports. An MLST report (see Figure 5.1) can be opened for the selected entries with **MLST for WGS > Reports...**

Clicking on an entry in the *Entries* panel of the *Genotyping MLST reports* window (or using the up and down arrow keys on the keyboard) shows the report for the highlighted entry.

At the top of each report the creation date of the report (**Date**), the Key (**Name**), and information fields that were checked in the *General* tab of the MLST settings are displayed, followed by a summary of the results of all analyzed MLST schemes.

Selecting **File > Exit** closes the *Genotyping MLST reports* window.

5.2 Report styles

In the *Genotyping report* window, three different report styles can be applied from the drop-down list in the panel header or via the menu (**Report > Report styles**):

1. **Summary**: only a summary of the results is shown.
2. **Default**: the summarized results and most details are shown in a tabular format. In this report style, all columns of the results tables can be sorted alphabetically or numerically by clicking on their headers.
3. **Complete**: the summarized results and all available details are shown. More exhaustive information is presented in an additional row, for example descriptions of the detected genes, decision trees, etc.. Result tables cannot be sorted in this report style.

5.3 Details section

The Details section in the MLST report displays the **Sequence type**, **Clonal complex** and a result table for each analyzed MLST scheme.

All loci in the MLST scheme are listed in the table, with following fields:

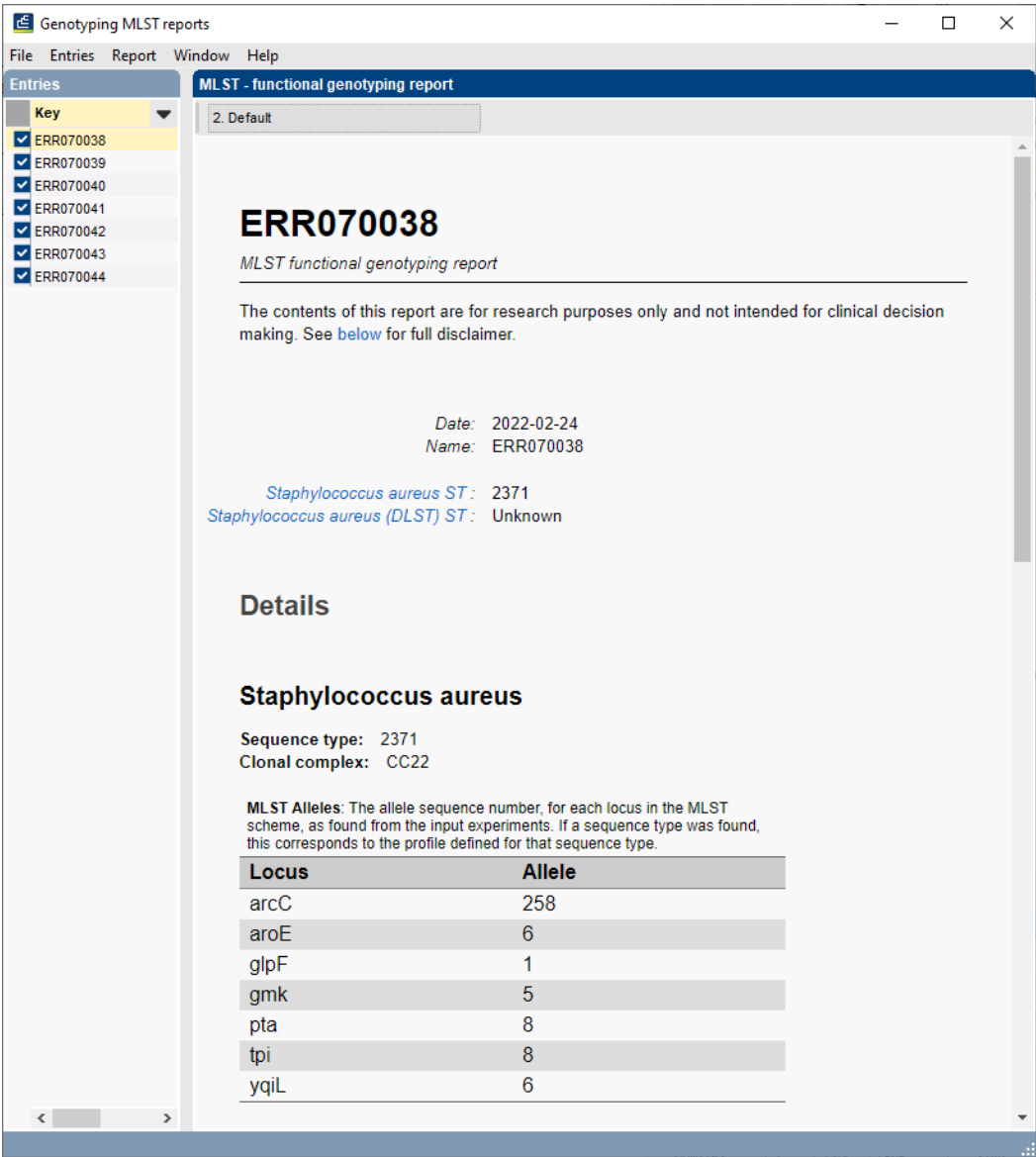


Figure 5.1: The *Genotyping MLST reports* window.

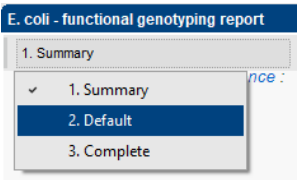


Figure 5.2: Drop-down list to select report styles in the *Genotyping report* window.

- **Locus:** Name of the MLST locus.
- **Allele:** Allele number detected for this locus. 'Unknown' means that an allele is detected for which there is no allele number in the MLST scheme definition yet. If no allele is detected for this locus, the text 'Missing' is displayed.

5.4 Info section

At the bottom of each report, an *Info section* is shown which contains information regarding the analysis date, plugin version, and settings of each MLST analysis.

Chapter 6

Working with a custom MLST scheme

6.1 Introduction

In case no MLST scheme has been published yet for the organism of your interest or if you want to start your own MLST scheme, a custom MLST scheme that grows dynamically (see [2.2](#)) should be used.

6.2 Registering new alleles and sequence types

When starting an MLST scheme from scratch, it frequently occurs that new alleles and sequence types are found. New alleles are indicated as 'Unknown' in the MLST reports and show up as missing values in the MLST character experiments. New sequence types are indicated as 'Unknown' in reports and information fields. New alleles and sequence types can be added to the MLST scheme stored in the BIONUMERICS database. For maximum flexibility, registering unknown alleles and unknown sequence types can be done separately. Both commands work on the selected entries and are only available for MLST schemes with the option **Expand scheme dynamically** checked (see [2.2](#)).

To add new alleles for the selected entries, use **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.

To add new MLST profiles and sequence types for the selected entries, use **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.

6.3 Typical workflow

Following is a typical workflow with a new MLST scheme:

1. Set up a custom scheme with at least one allele per locus and at least one sequence type.

In this stage, the locus and sequence type files are typically stored locally or on a network drive. Make sure to check the **Expand scheme dynamically** option (see 2.2).

2. Run the MLST scheme (see 4).
3. Register unknown alleles and sequence types in the BIONUMERICS database (see 6.2).
4. Run the scheme again to assign the new alleles and sequence types.

The above procedure can be repeated as long as the MLST scheme is used in a single BIONUMERICS database. When working with the MLST scheme from more than one BIONUMERICS database, following steps should be taken on regular intervals:

1. Update the MLST scheme definitions files with the data in the BIONUMERICS database using the export functionality (see 2.4). Note that clonal complex information should be added manually.
2. Update the MLST scheme in all databases (see 2.5).



To ensure a consistent nomenclature, the option **Expand scheme dynamically** can only be checked in a single BIONUMERICS database.

To make the MLST scheme available for a broader audience, it is recommended to host the scheme with e.g. PubMLST.

