BIOMÉRIEUX

BIONUMERICS®

version 8 - PLUGINS

Resistance detection plugin

Contents

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.

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:

Chapter 1

Starting and setting up BIONUMERICS

1.1 Introduction

The *Resistance detection plugin* allows you to screen whole genome sequences to predict antibiotic resistance. The genome sequences can be imported in BIONUMERICS using one of the import routines available in the software or can be the result of a de novo assembly performed in BIONUMERICS on a sequence read set.

The resistance knowledge base used by this plugin (see 3 for more information) is not organism-specific and contains resistance determinants for a wide range of species.

All analyses are performed locally in the BIONUMERICS database itself.

The *Resistance detection 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.

1.3 Installation of the Resistance detection plugin

Proceed as follows to install the *Resistance detection plugin*:

3.1 Call the *Plugins and Scripts* dialog box from the *Main* window with ***File*** > ***Install / remove plugins...*** ().

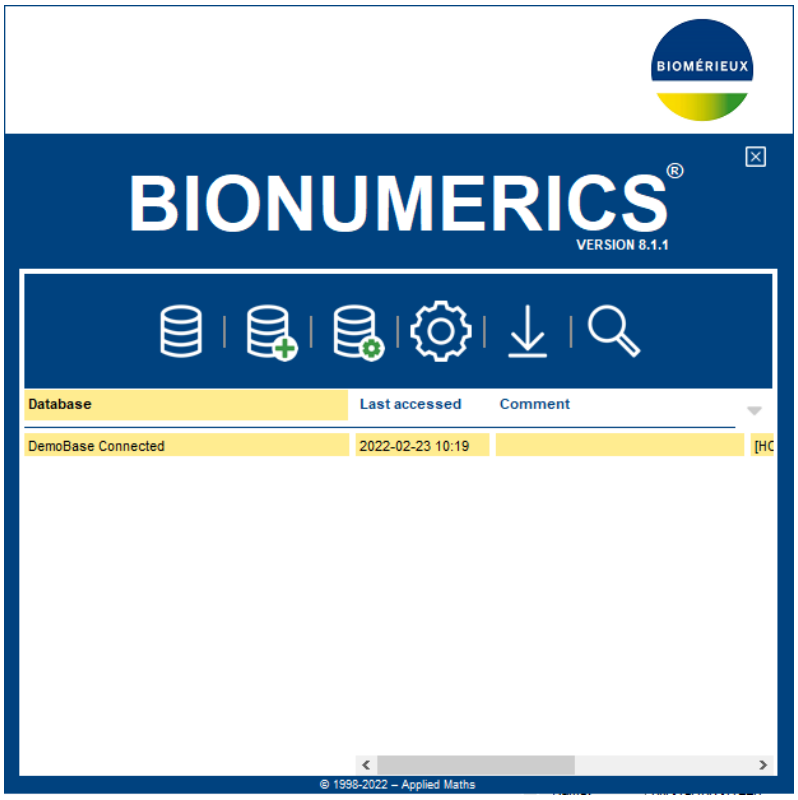

Figure 1.1: The *BIONUMERICS Startup* window.

3.2 Select the *Resistance detection plugin* from the list and press the <***Install***> button.

3.3 Confirm the installation of the plugin.

During installation, the plugin downloads its online resistance detection knowledge base (see 3) from `https://www.bionumerics.com`, which requires a connection to the internet. Depending on the bandwidth of your connection, this process can take up to several minutes. When the download is complete, the question 'Do you want to change the Resistance detection genotyping settings now?' pops up.

3.4 Press <***Yes***> to confirm that you want to change the plugin settings.

The *Resistance detection settings* dialog box appears, which will be discussed in detail in 2. For the plugin to work, at the very least we need to specify the sequence experiment containing the (de novo) genome sequences on which the screening will be performed.

3.5 If already available, select the sequence experiment in your database that contains genome sequences as ***Input sequence experiment*** and press <***OK***>.

A message appears, confirming the installation of the plugin and prompting you to restart BIONUMERICS.

3.6 Press <***OK***> in the confirmation message.

3.7 Press <***Close***> to close the *Plugins and Scripts* dialog box.

3.8 Close and re-open the database to complete the installation of the plugin.

The *Resistance detection plugin* installs menu items in the main menu of the software under ***Resistance detection*** (see Figure 1.2).

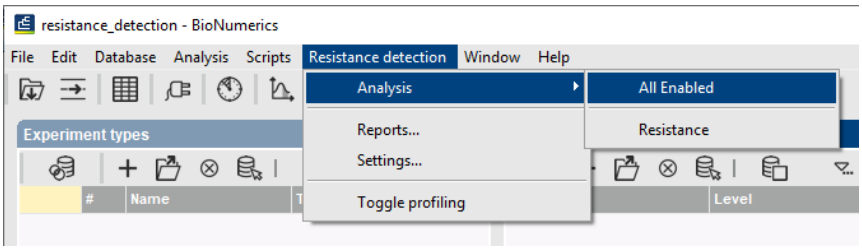

Figure 1.2: New menu items, available after installation of the *Resistance detection plugin*.

More information regarding the *Resistance detection plugin* settings is available under 2.

Chapter 2

Settings of the Resistance detection plugin

2.1 Accessing the resistance detection settings

Settings for the *Resistance detection plugin* can be accessed via ***Resistance detection*** > ***Settings...*** in the *Main* window.

2.2 General settings

The *General* tab of the *Resistance detection settings* dialog box (see Figure 2.1) holds settings for the genotyping reports and for general processing.

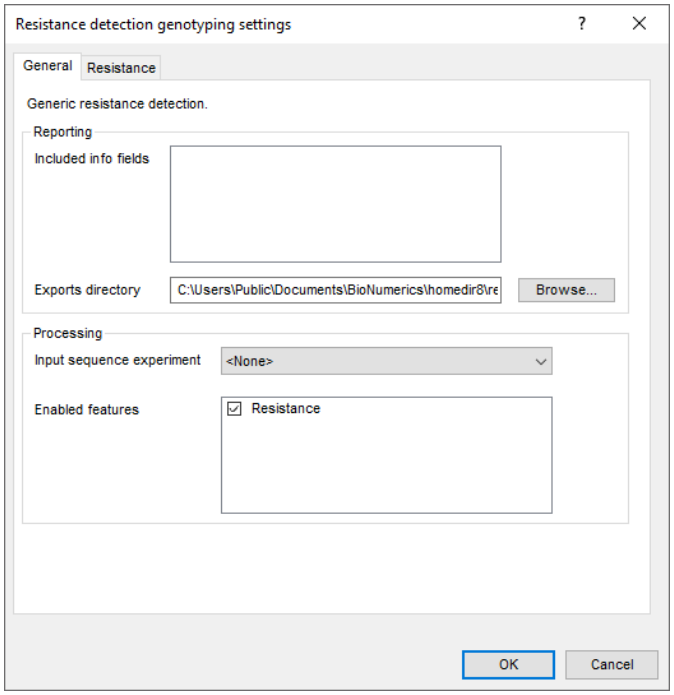

Figure 2.1: The *Resistance detection settings* dialog box, *General* tab.

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.

2.3 Resistance settings

The *Resistance* tab in the *Resistance detection settings* dialog box (see Figure 2.2) groups the settings specific for the detection of acquired resistance.

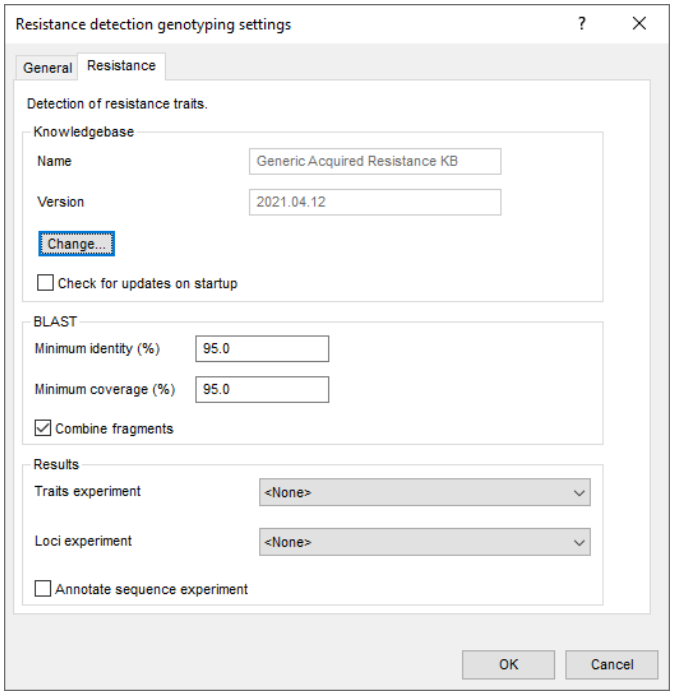

Figure 2.2: The *Resistance detection settings* dialog box, *Resistance* tab.

Resistance prediction is based on the absence or presence of certain antibiotic resistance related genes and alleles in assembled genome sequence. Presence or absence is detected based on a BLAST approach using the list of antibiotic resistance related genes as query and the de novo assembled genome as target.

Under ***Knowledgebase***, the ***Name*** and ***Version*** of the specified knowledge base version for this

feature is shown. When no knowledge base is specified yet, "<None>" will be indicated in both fields. When a specified knowledge base version cannot be found (e.g. because it is deleted), "<Missing>" is shown in both fields. A different knowledge base version can be selected by pressing the <***Change...***> button.

With ***Check for updates on startup*** checked, BIONUMERICS will check if a newer knowledge base version is available online for this feature each time the database is opened. This requires an internet connection.

See 3 for more information on *Resistance detection plugin* knowledge bases.

In the ***BLAST*** panel, two settings for the BLAST algorithm can be specified:

- ***Minimum identity (%)*** is the minimum sequence identity (as percentage) of the query sequence against the knowledge base's reference sequences.
- ***Minimum length for coverage*** specifies the minimum overlap (as percentage) between the subsequence found in the target assembly sequence and the reference sequence from the knowledge base.

If the option ***Combine fragments*** is checked, genes that occur fragmented in the genome (i.e. split over two contigs) can still be detected.

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.

Following character experiments can optionally be specified for acquired resistance:

- ***Traits experiment***: contains the results for each antibiotic group: 0 = not detected (sensitive), 1 = detected (resistant). The default name is **Resistance_traits**.
- ***Loci experiment***: contains the results for each resistance gene: 0 = not detected (sensitive), when detected (resistant) the % identity of the best BLAST hit is shown. The default name is **Resistance_loci**.

The characters in the characters experiments are displayed in the same order they are listed in their knowledge base. However, it might be more convenient for interpretation to have them displayed alphabetically. How to rearrange characters in a character type experiment is described in the Reference manual, Chapter Setting up character type experiments.

For the resistance character experiments, it can be convenient to map the character values (-1, 0, 1) to categorical names (indecisive (-), sensitive (S), resistant (R)) by creating a character mapping. How to do this is explained in the Reference manual, Chapter Setting up character type experiments.

Check ***Annotate sequence experiment*** to annotate the input sequence with the detected genotyping features.

Chapter 3

Resistance detection knowledge base

3.1 Introduction

The *Resistance detection plugin* makes use of a resistance detection knowledge base that contains a set of target sequences associated with antimicrobial resistance. These target sequences are not limited to a specific organism. By providing the resistance detection knowledge base online from https://www.bionumerics.com, it can easily be updated without the need to install a new plugin version.

The resistance knowledge base is based on curated public repositories (e.g. ResFinder https://cge.cbs.dtu.dk/services/ResFinder/) and converted to the specific format required by the plugin. It has a change log with detailed information on the changes in each version.

The *Resistance detection plugin* can only use an *online* knowledge base. The downloaded knowledge base is encrypted and cannot be modified by the user. For using your own, custom knowledge bases we recommend the *Custom genotyping plugin*, which was specifically designed for this purpose.

By default, the plugin uses the most recent knowledge base version available at the time of installation. When more than one knowledge base version is available online, users can specify which version to use.

3.2 Specifying a different knowledge base version

The resistance knowledge base is specified in the resistance detection settings.

2.1 In the *Main* window, select ***Resistance detection*** > ***Settings...*** to open the *Resistance detection settings* dialog box.

2.2 Click on the *Resistance tab* and press <***Change...***>.

This action opens the *Change knowledge base* dialog box (see Figure 3.1).

The *Change knowledge base* dialog box shows the downloaded knowledge bases for this feature. The currently used knowledge base (if specified) will be highlighted by default. Following information is displayed about the knowledge bases:

- 'Type': either Online or Local. The knowledge base type will always be Online for all organism-specific genotyping plugins and for the *Resistance detection plugin*.

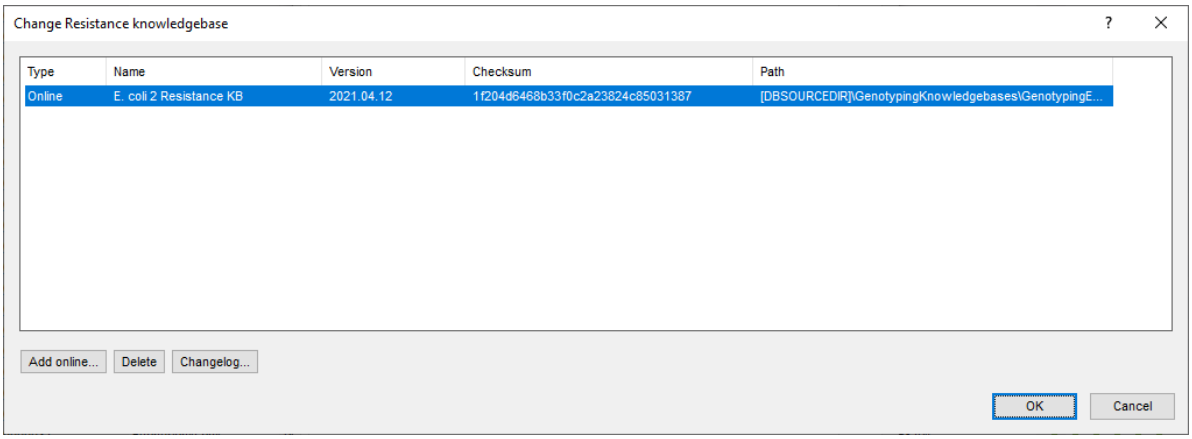

Figure 3.1: The *Change knowledge base* dialog box.

- 'Name': the name of the knowledge base.
- 'Version': the knowledge base version, typically a last modified date formatted as YYYY.MM.DD.
- 'Checksum': MD5 checksum, used to verify the integrity of the downloaded knowledge base.
- 'Path': file path, relative to the source files directory (indicated with the token `DBSOURCEDIR`), where the knowledge base is downloaded to.

To manually check if other versions (newer or older) of the knowledge base are available online, press <***Add online...***>. This opens the *Download online knowledge base* dialog box (see Figure 3.2).

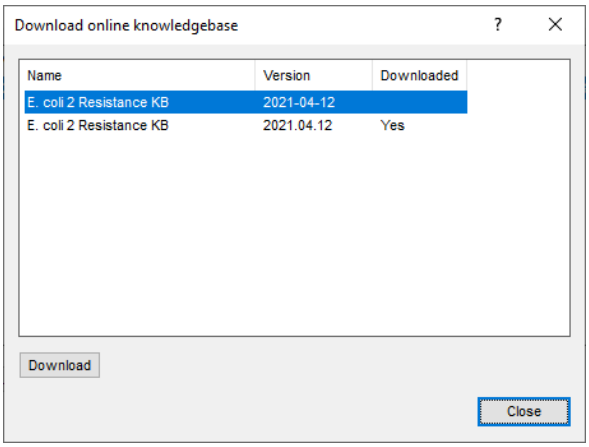

Figure 3.2: The *Download online knowledge base* dialog box.

Opening this dialog will check the BIONUMERICS website (`https://www.bionumerics.com`) for the latest knowledge base version for this feature. When the latest version is already downloaded (i.e. the knowledge base is up to date), the 'Downloaded' field will indicate "Yes". If this is not the case, you can press the <***Download***> button to download the latest version.

Close the dialog with <***Close***>.

A highlighted knowledge base can be deleted by pressing <***Delete***>. The software will ask for confirmation before actually removing the knowledge base.

By pressing <***Changelog...***> you are referred to the change log page on the BIONUMERICS website of the online knowledge base you have selected. Here you can find information regarding the source(s) used to create the knowledge base and the changes made in respect to prior versions of the knowledge base. This information should help you decide which knowledge base version to use.

To specify a different knowledge base for the feature, click on the preferred knowledge base to highlight it and press <***OK***>. This action will close the *Change knowledge base* dialog box and will show the genotyping settings again with the newly specified knowledge base.

3.3 Automated check for knowledge base updates

For each feature that uses an online knowledge base, there is a check box ***Check for updates on startup***. When this is checked, BIONUMERICS will automatically connect to `https://www.bionumerics.com` to check for knowledge base updates each time the database is opened. When updates are available, the *Update knowledge bases* dialog box pops up (see Figure 3.3).

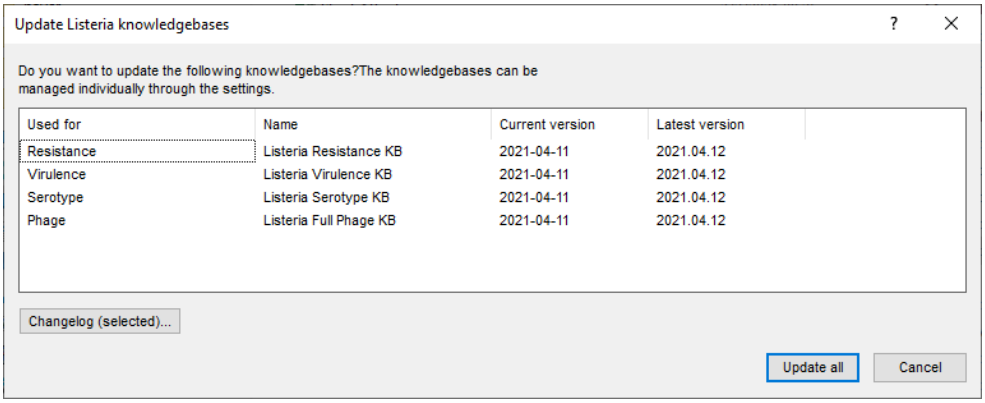

Figure 3.3: The *Update knowledge bases* dialog box pops up at startup when one or more new knowledge bases are available online.

By pressing <***Changelog (selected)***> you are referred to the change log page on the BIONUMERICS website of the highlighted online knowledge base.

Press <***Update all***> to apply all knowledge base updates. The updated knowledge bases will be downloaded and set as default in the genotyping settings, meaning that the updated knowledge base will be used the next time an analysis is run for the corresponding feature.

Knowledge base updates can only be done in bulk from this dialog box, i.e. for all available updates at once. If you need to update one or more knowledge bases selectively, press <***Cancel***> in the *Update knowledge bases* dialog box and update the knowledge bases one by one via the genotyping settings (see **??**).

Chapter 4

Resistance detection 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

Screening for acquired antibiotic resistance traits can be done using ***E. coli*** $>$ ***Analysis*** $>$ ***All Enabled***.

The analysis time depends on the number of selected entries and may take up to several minutes or even hours.

When the analysis is finished, the progress bar disappears. The screening results are stored in the database experiments and information fields which you have defined in the *Resistance detection settings* dialog box. The settings can always be consulted or adjusted using ***Resistance detection*** $>$ ***Settings...*** (see 2 for details).

Chapter 5

Resistance detection reports

5.1 Opening resistance detection reports

A resistance detection report (see Figure 5.1) can be opened for the selected entries with ***Resistance detection*** > ***Reports...***.

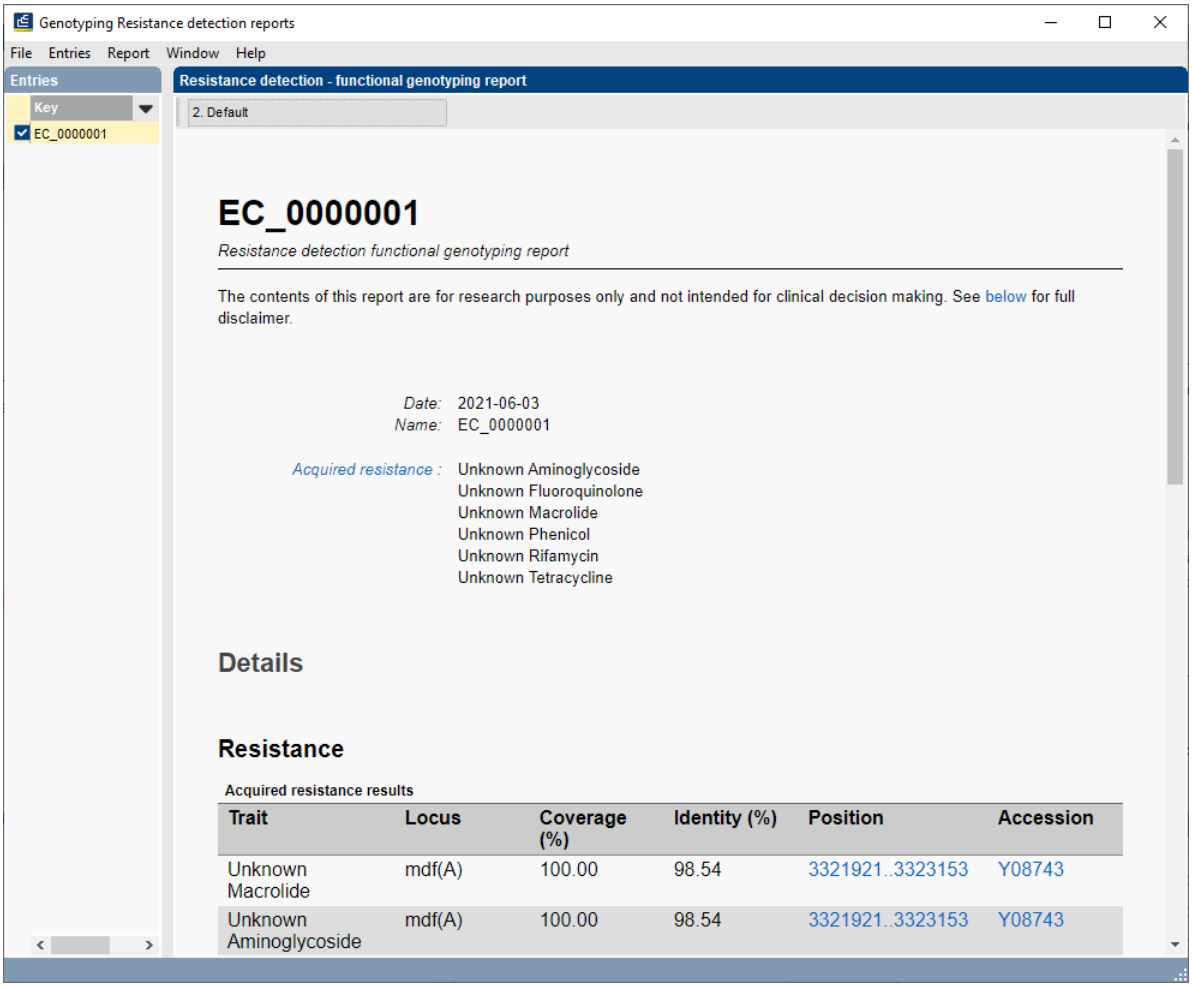

Figure 5.1: The *Genotyping report* window, showing a report for sample EC_0000001.

Clicking on an entry in the *Entries* panel of the *Genotyping report* 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 genotyping settings are displayed, followed by a summary of the results of all analyzed traits.

Selecting ***File*** > ***Exit*** closes the *Genotyping report* 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.

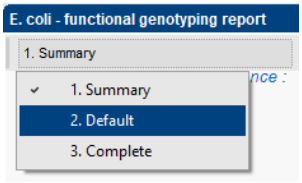

Figure 5.2: Drop-down list to select report styles in the *Genotyping report* window.

5.3 Details section

5.3.1 Introduction

The Details section in the genotyping report contains a detailed result table for each analyzed genotyping feature.

Some result tables contain hyperlinks. These hyperlinks open in the default web browser, except for the 'Position' field. The latter links open the *Sequence editor* window in BIONUMERICS with the positions highlighted on the sequence.

5.3.2 Resistance

All detected acquired resistance traits are listed in the table, with following fields:

- **Trait:** Name of the resistance trait.

- **Locus:** Detected locus that confers the trait.
- **Coverage (%):** Percent overlap between the query and the target sequence.
- **Identity (%):** BLAST identity, expressed as a percentage.
- **Position:** Position(s) of the BLAST match on the input genome. Multiple positions are possible when ***Combine fragments*** was checked in the genotyping settings.
- **Accession:** Link to the GenBank accession of the detected locus.
- **Description:** Description for the detected locus (in ***Complete*** template only).
- **Publication:** Link to the publication in PubMed describing the detected locus (in ***Complete*** template only).

5.4 Info section

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

5.5 Exporting report information

The genotype information for all selected entries in the *Genotyping report* window can be exported with ***Entries*** $>$ ***Export selected***. The results for the currently shown report can be exported with ***Report*** $>$ ***Export current***.

A Tab Separated Values (`*.tsv`) file is created for each functionality and stored in the report export directory as specified in the genotyping settings. The location of the files is opened after the export.

The displayed report can be printed directly to a printer using ***Report*** $>$ ***Print...***.

The dialog box that appears is the standard Windows Print dialog box, allowing you to choose a printer and change the properties. Depending on your system, this also allows printing to a PDF file to create an export of the displayed report. Please note that background colors, such as those in lists and mutational decision trees, may be lost in this step regardless of printer settings.

bioMérieux S.A. • 69280 Marcy l'Étoile • France • Tel.: + 33 (0)4 78 87 20 00 • Fax: +33 (0)4 78 87 20 90
www.biomerieux.com