BIOMÉRIEUX

BIONUMERICS®

version 8 - PLUGINS

QIAxcel plugin

Contents

1 Starting and setting up BIONUMERICS **5**
- 1.1 Introduction . . . 5
- 1.2 Startup program . . . 5
- 1.3 Creating a new database . . . 5
- 1.4 Installing the QIAxcel plugin . . . 7

2 Importing ScreenGel data **9**
- 2.1 Supported ScreenGel data formats . . . 9
- 2.2 The Import tree . . . 10
- 2.3 Create a ScreenGel experiment type . . . 10
- 2.4 The ScreenGel import wizard . . . 11
- 2.5 Normalization . . . 14

3 Analyzing ScreenGel data **19**
- 3.1 Selections in BIONUMERICS . . . 19
- 3.2 The Comparison window . . . 19
- 3.3 Cluster analysis . . . 20

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 QIAxcel Advanced System from Qiagen (http://www.qiagen.com) performs fully automated separation of DNA and RNA fragments according to their molecular weight.

The QIAxcel ScreenGel software supplied with the QIAxcel Advanced System provides both electropherograms and gel images of the nucleic acid separation and can be used to analyze the size and concentration of the fragments.

From the QIAxcel ScreenGel software the raw data can be exported as densitometric curves and the analyzed data as peak tables. These densitometric curves and peak tables can be imported into BIONUMERICS using the *QIAxcel plugin*.

The *QIAxcel plugin* is supported in the **BIONUMERICS-GEL**, **BIONUMERICS-MALDI** 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 Creating a new database

3.1 Press the  button in the BIONUMERICS *BIONUMERICS Startup* window to enter the *New database* wizard.

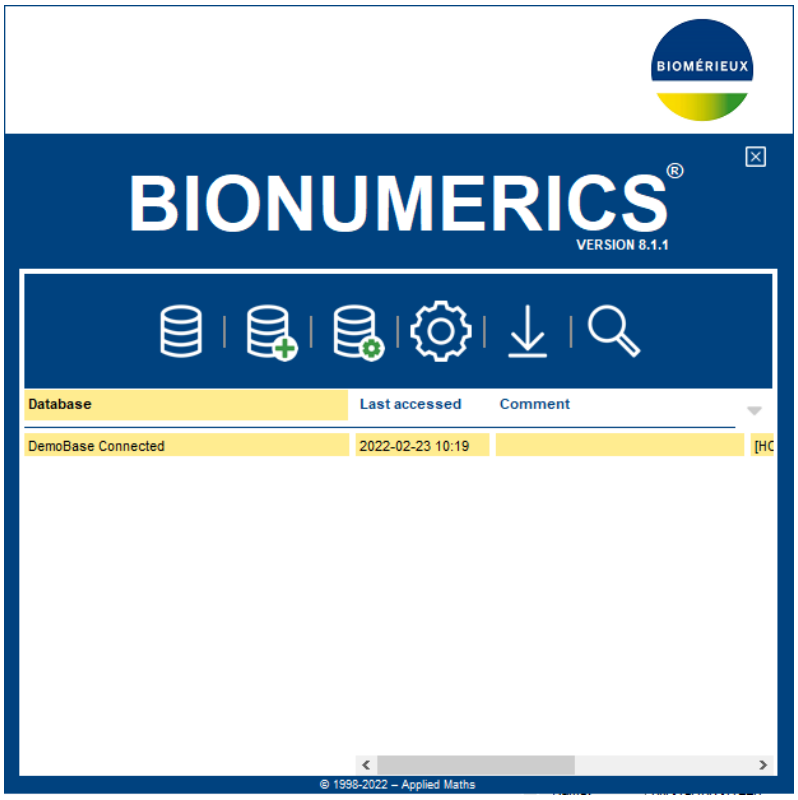

Figure 1.1: The *BIONUMERICS Startup* window.

3.2 Enter a name for the database, and press <***Next***>.

A new dialog box pops up, prompting for the type of database (see Figure 1.2).

3.3 Leave the default option selected and press <***Next***>.

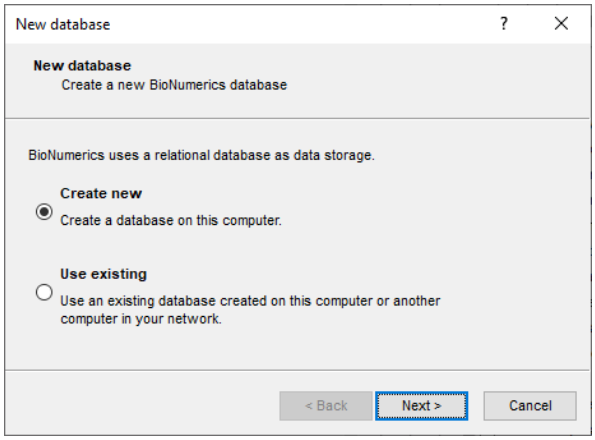

Figure 1.2: The *New database* wizard page.

A new dialog box pops up, prompting for the database engine (see Figure 1.3).

3.4 Leave the default option selected and press <***Finish***> to complete the setup of the new database.

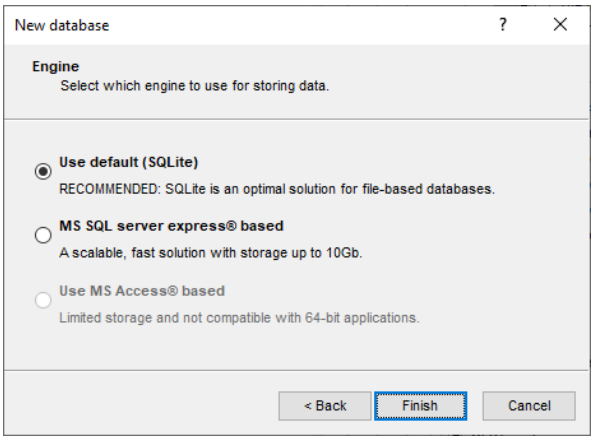

Figure 1.3: The *Engine* wizard page.

1.4 Installing the QIAxcel plugin

The *Plugins and Scripts* dialog box can be called from the *Main* window by selecting ***File*** > ***Install / remove plugins...*** ( ) (see Figure 1.4).

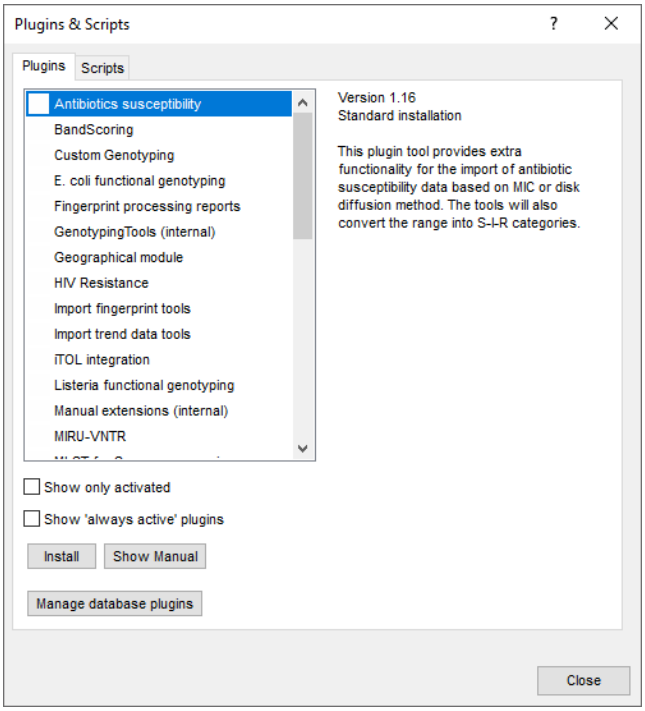

Figure 1.4: The *Plugins and Scripts* dialog box.

When a particular plugin is selected from the list of plugins, a short description appears in the right panel.

A selected plugin can be installed with the <***Install***> button. The software will ask for confirmation before installation. Some plugins are only supported in specific BIONUMERICS configurations. If the plugin is not supported by your BIONUMERICS configuration, it cannot be installed and an error message will be generated.

Once a plugin is installed, it is marked with a green V-sign. It can be removed again with the <***Uninstall***> button.

If the selected plugin is documented, pressing <***Show Manual***> will open its manual in the *Help* window.

4.1 Select the *QIAxcel plugin* from the list and press the <***Install***> button.

4.2 The program will ask to confirm the installation of the plugin. Press <***OK***> to confirm the installation.

4.3 Press <***Close***> to close the *Plugins and Scripts* dialog box and to continue to the *Main* window.

4.4 Close and reopen the database to activate the features of the *QIAxcel plugin*.

4.5 Select ***File*** > ***Import...*** (, **Ctrl+I**) to call the *Import data* wizard and press <***Next***> with the <***Manual selection***> option highlighted to proceed to the second page of the wizard.

Upon installation of the *QIAxcel plugin*, the ***Import ScreenGel data*** import item is activated in the *Import data* wizard (see Figure 1.5).

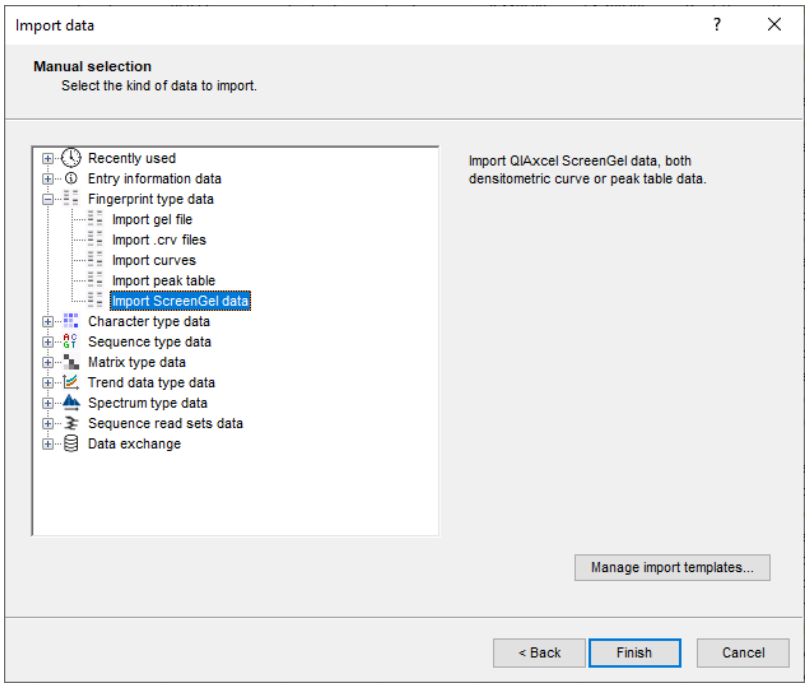

Figure 1.5: Import ScreenGel data menu item in the *Import data* wizard.

4.6 Close the *Import data* wizard.

Chapter 2

Importing ScreenGel data

2.1 Supported ScreenGel data formats

BIONUMERICS supports the import of raw QIAxcel data as densitometric curves and the import of analyzed data as peak tables.

Densitometric curves can be exported from the QIAxcel ScreenGel software in two formats: as an XML file or as a CSV file. Both formats contain the same information, but the CSV format is less verbose and will therefore import a bit faster.

- To export an XML file containing the raw data, check "XML Raw Data Export" in the Export options. This will produce a `*_Rw.xml` file in the default export directory `ProgramData\QIAGEN\QIAxcel\ScreenGel\Data\Export`.
- For a CSV file export, select "Rawdata_CSV_export" under "Additional Output formats" (see Figure 2.1).

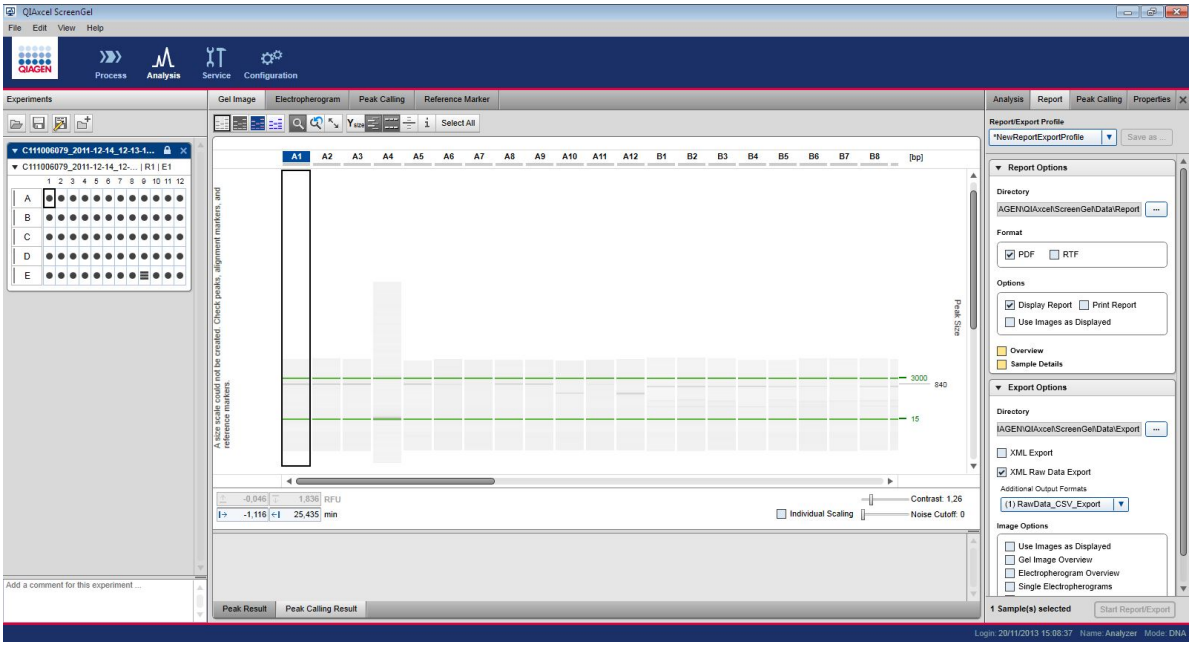

Figure 2.1: The ScreenGel software.

Peak tables represent analyzed data: they contain only positions and sizes of called peaks. Peak tables can be exported from the ScreenGel software as files with the extension *.xdr or *.xdrx.

2.2 The Import tree

2.1 Select ***File*** > ***Import...*** (, **Ctrl+I**) to call the *Import data* wizard and press <***Next***>.

2.2 Select ***Import ScreenGel data*** under ***Fingerprint type data*** and press <***Import***> button to start with the import of the data.

If one or more fingerprint types have already been created in the database, the import wizard opens (see Figure 2.3).

When no fingerprint type experiment is present in the database the *Create a default ScreenGel experiment type* dialog box opens (see Figure 2.2).

2.3 Create a ScreenGel experiment type

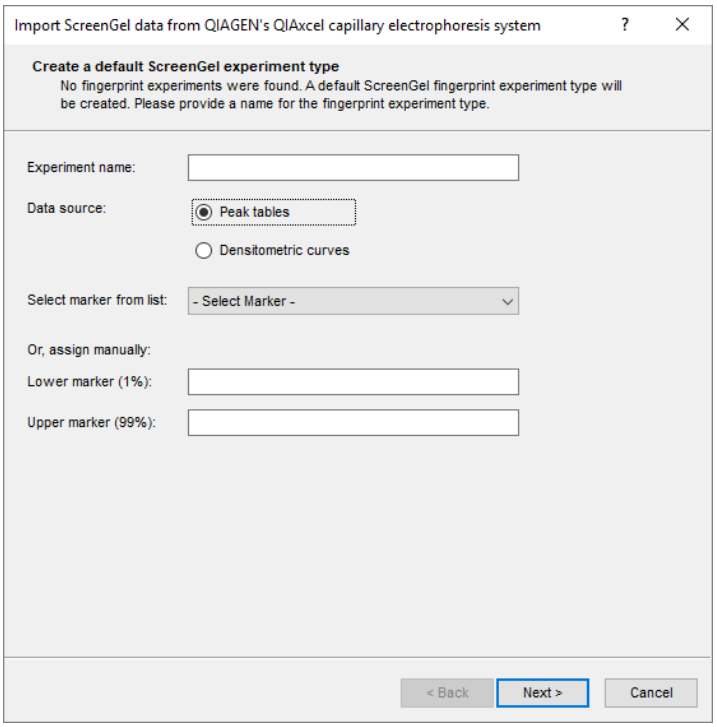

Figure 2.2: The *Create a default ScreenGel experiment type* dialog box.

If you want to import ***Densitometric curves*** you just need to provide a name for the experiment (***Experiment name***). An experiment with an OD range of 100000 points and normalized tracks with a resolution of 2000 points will be created.

If you want to import ***Peak tables*** you need to provide a name for the experiment (***Experiment name***), and you need to provide the alignment marker that you used in the experiment. You can select this marker from the drop down list, or fill in the values for the ***lower*** and ***upper marker*** manually. An experiment will be created with the same OD range and normalized tracks as for the

densitometric curves. The selected/entered alignment marker will be used to create a reference system with a calibration curve that has a linear dependence between the lower and upper marker.

The OD range of the experiment may need to be decreased to visualize the bands/peaks.

2.4 The ScreenGel import wizard

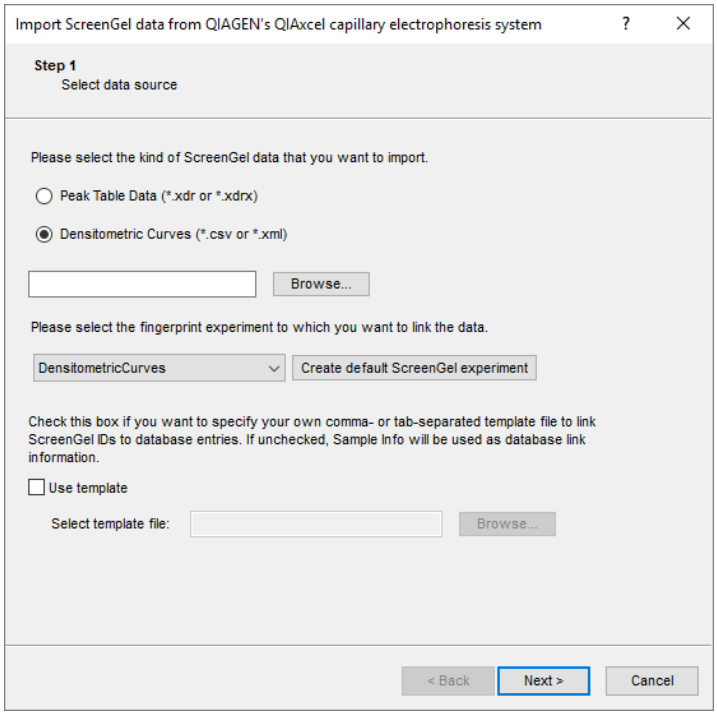

Figure 2.3: *Step 1* dialog box.

From the QIAxcel ScreenGel software the raw data can be exported as densitometric curves and the analyzed data as peak tables. In the first step the file format needs to be selected:

- Densitometric curves can be exported from the QIAxcel ScreenGel software in two formats: as an XML file or as an CSV file. Both formats contain the same information and can be imported into BIONUMERICS (check ***Densitometric Curves (*.csv or *.xml)***) but the CSV format will be imported the fastest.
- Peak tables with the extension *.xdr and *.xdrx can be imported into BIONUMERICS when checking the option ***Peak Table Data (*.xdr or *xdrx)***.

From the drop down list, the fingerprint type experiment type should be selected to which you want to link the data. A new fingerprint type experiment can be created with <***Create default ScreenGel experiment***>. This calls the *Create experiment* dialog box (see Figure 2.4).

If ***Densitometric curves*** is checked you just need to provide a name for the experiment (***Experiment name***). An experiment with an OD range of 100000 points and normalized tracks with a resolution of 2000 points will be created.

If ***Peak tables*** is checked, you need to provide a name for the experiment (***Experiment name***), and you need to provide the alignment marker that you used in the experiment. You can select this marker from the drop down list, or fill in the values for the ***lower*** and ***upper marker*** manually. An

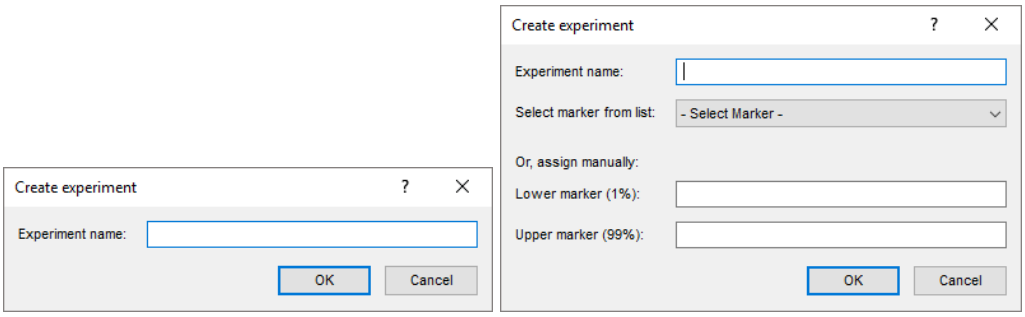

Figure 2.4: The *Create experiment* dialog box.

experiment will be created with the same OD range and normalized tracks as for the densitometric curves. The selected/entered alignment marker will be used to create a reference system with a calibration curve that has a linear dependence between the lower and upper marker.

During the set-up of your ScreenGel experiment, you can choose to add *Sample Info* in the QI-Axcel ScreenGel software. This *Sample Info* gets exported with the XML and CSV files. If you have provided *Sample Info* for your curves, this information can be used to link the curves to existing entries in your database. In the last step of the wizard the corresponding BIONUMERICS information field can be selected.

If you do not have any *Sample Info* provided, but want to have control over how the curves get linked to existing entries, you can use a *template file*. This is a simple tab- or comma delimited file with two columns (no headers). The first column should contain the *ScreenGel IDs* and the second column should contain the database entry information (see Figure 2.5). In the last step of the wizard the corresponding BIONUMERICS information field can be selected.

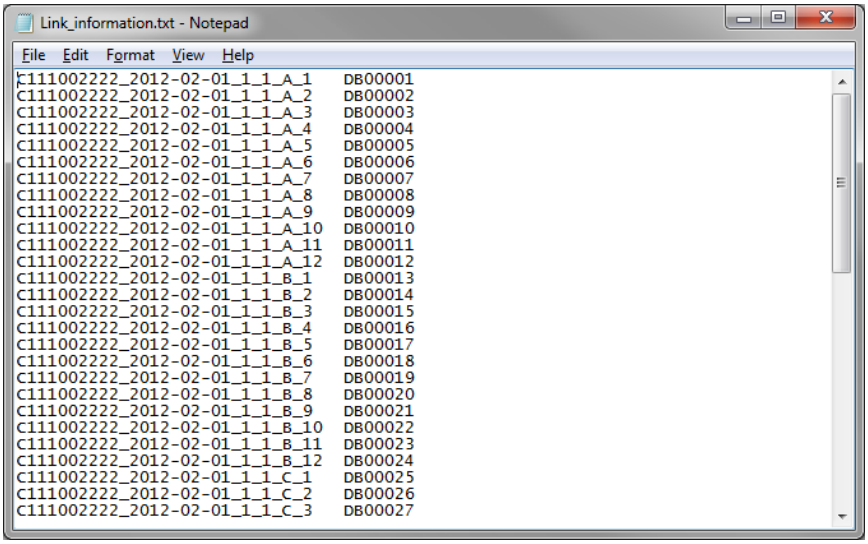

Figure 2.5: A template file: ScreenGel IDs (left) and database information (right).

If no *Sample Info* is detected and no template file is selected (for example when importing in an empty database), the *ScreenGel IDs* can be imported and stored in an information field selected by the user. In the last step of the wizard this information field can be selected.

4.1 Browse for the file, select the fingerprint type experiment to which you want to link the data, and optionally select a template file. Press <***Next***>.

The plugin will read the information present in the selected file(s). If no *Sample Info* was provided

in the ScreenGel file and no template file was selected, a message box will appear, asking if the *ScreenGel IDs* should be imported into the database (see Figure 2.6). When you do not want to link the data to entries in the database press <***No***>.

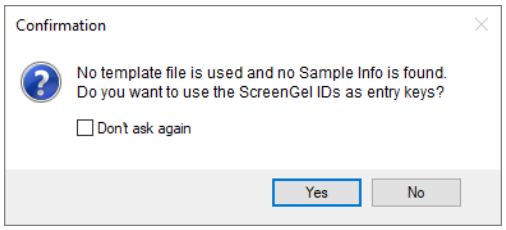

Figure 2.6: No *Sample Info* detected and no template file selected.

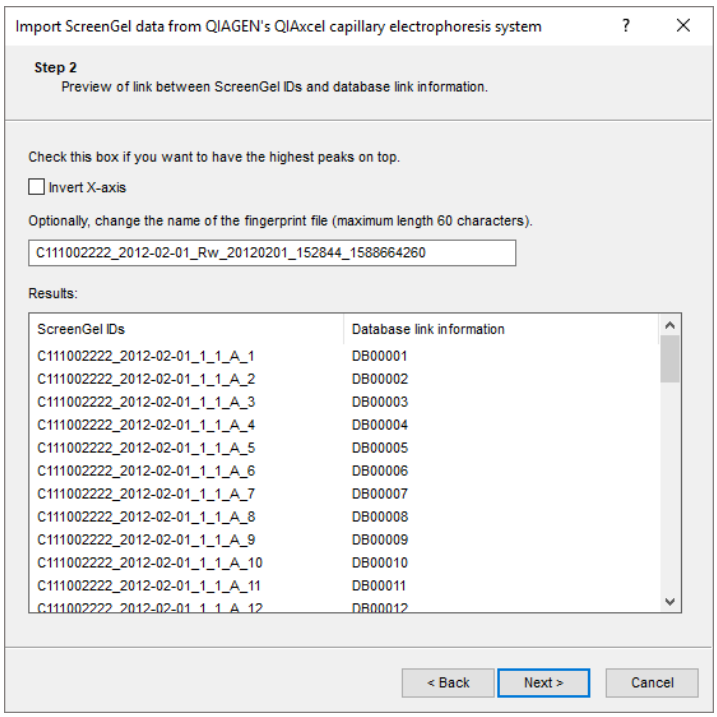

Figure 2.7: *Step 2* dialog box.

In the second step of the import wizard, you will find a check box to invert the X-axis when importing ***Densitometric Curves***. Checking this option will show the highest peaks on top (i.e. as is the case in slab gel electrophoresis).

Optionally, you can also change the name of the fingerprint file to something more meaningful (the default is a combination of the selected file and a time stamp). The total length of the fingerprint file name cannot exceed 60 characters.

The first column in the *Results panel* contains the *ScreenGel IDs* parsed from the ScreenGel file. The second column is editable and contains the information that will be imported into the BIONUMERICS database. If *SampleInfo* was provided with your file, this *SampleInfo* is automatically displayed in the second column. In case a template file was selected, the information of the template file is displayed.

4.2 Click on <***Next***> to go to the final step.

The last step prompts for some final settings (see Figure 2.8):

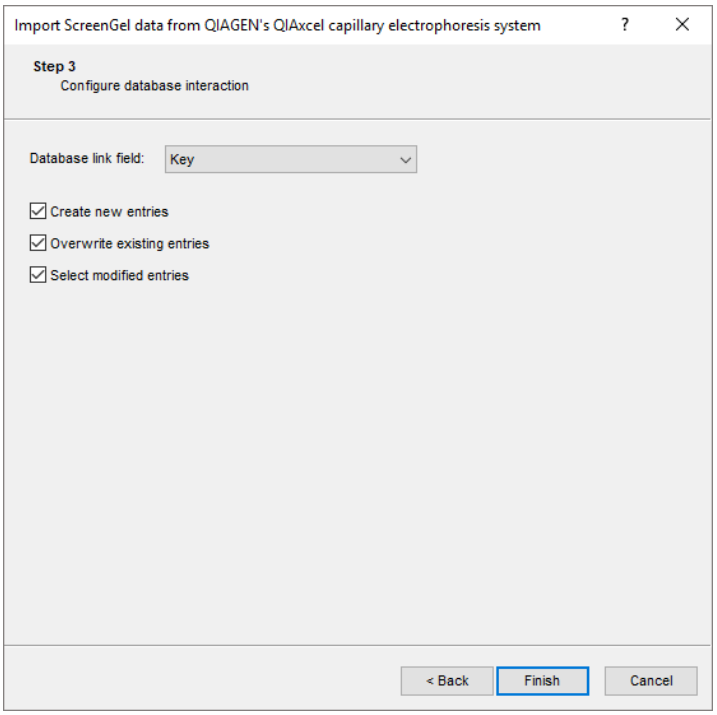

Figure 2.8: *Step 3* dialog box.

- ***Database link field***: will link the data based on the information present in this information field. If entries are already present in the database with the same information, the import routine will link the data to these entries. If the entries are not yet present in the database, the data will be linked to new entries in the database (if the option ***Create new entries*** is checked).
- ***Create new entries***: will create new entries when the linked information is not found in the database.
- ***Overwrite existing entries***: will overwrite the entry if the entry is already found in the database.
- ***Select modified entries***: will select all entries for which data is imported. This can be useful if you have a lot of entries already in your database and you want to make e.g. a comparison of only the entries you just imported.

4.3 Click on <***Finish***> to start with the import of the data.

The data is imported into the database and linked to the selected fingerprint type experiment.

2.5 Normalization

After import of the densitometric curves, the data still needs to be normalized. For an in depth treatment of the normalization, we refer to the Reference manual, Chapter Setting up fingerprint type experiments.

5.1 Make sure the ***Normalization*** tab is selected in the *Fingerprint processing* window.

5.2 Select a lane and right click on the lowest peak of the alignment markers.

5.3 Select ***Add external reference position***.

5.4 A confirmation window will be shown asking if you want to switch to normalized mode. Click <***OK***> to confirm.

5.5 Enter the size (in bp) of the smallest alignment marker fragment.

5.6 Repeat this action for the highest peak of the alignment markers.

The name of the reference position should only contain numerical values.

The bands of the alignment markers should be the highest and lowest bands of your lane.

5.7 To let the program assign the bands and reference positions automatically, select ***Normalization*** > ***Auto assign...*** ( ), keep ***Using bands*** as search method and click <***OK***> to confirm. The band assignment can be corrected if necessary.

The presence of primer dimer pollution on the curves might interfere with the automatic assignment of the bands. The *Primer dimer polution correction* dialog box can be called with ***Normalization*** > ***Auto assign ScreenGel data*** (see Figure 2.9).

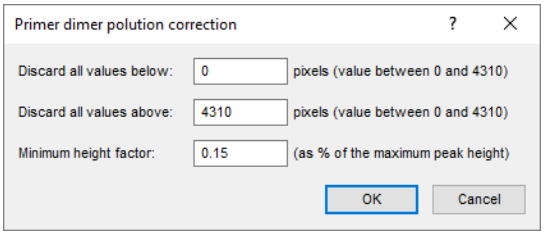

Figure 2.9: The *Primer dimer polution correction* dialog box.

A correction can be applied on the curves, discarding all peaks below or above a certain peak height.

5.8 Toggle between the normalized and original view with the menu item ***Normalization*** > ***Show normalized view*** (, **Shift+N**) (see Figure 2.10).

5.9 Switch to the *Bands* tab and select ***Bands*** > ***Auto search bands*** ( ). Click on the <***Search on all lanes***> button to start the search for bands. The band assignment can be corrected if necessary.

If you have run a ladder lane with your experiment, the bands of the ladder lane can be used as calibration for the reference system. This will normally be done only once.

5.10 Select the ladder lane and select ***Bands*** > ***Use ladder lane for calibration*** (see Figure 2.11).

5.11 The fingerprint preprocessing information needs to be saved before creating the calibration curve. Confirm by clicking <***Yes***>.

5.12 If the reference system that is used to process the data already contains a calibration curve, that calibration curve will be discarded. Confirm by clicking <***Yes***>.

Select the DNA size marker from the list or manually assign the values to the peaks found in the ladder lane. If the reference system position names are numbers, they can be optionally used as calibration points. Press <***OK***>.

The bands of the ladder lane will now be used for calibration.

5.13 Select ***File*** > ***Save*** and close the *Fingerprint processing* window with ***File*** > ***Exit***.

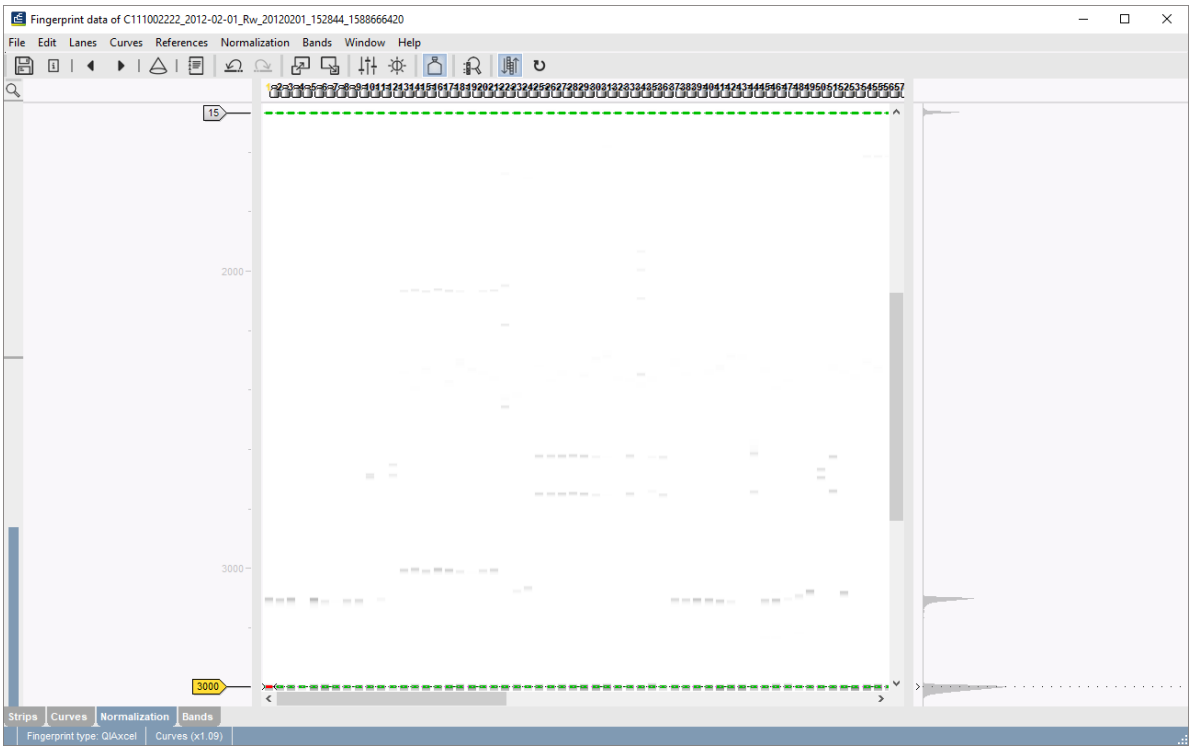

Figure 2.10: View of the imported ScreenGel after normalization.

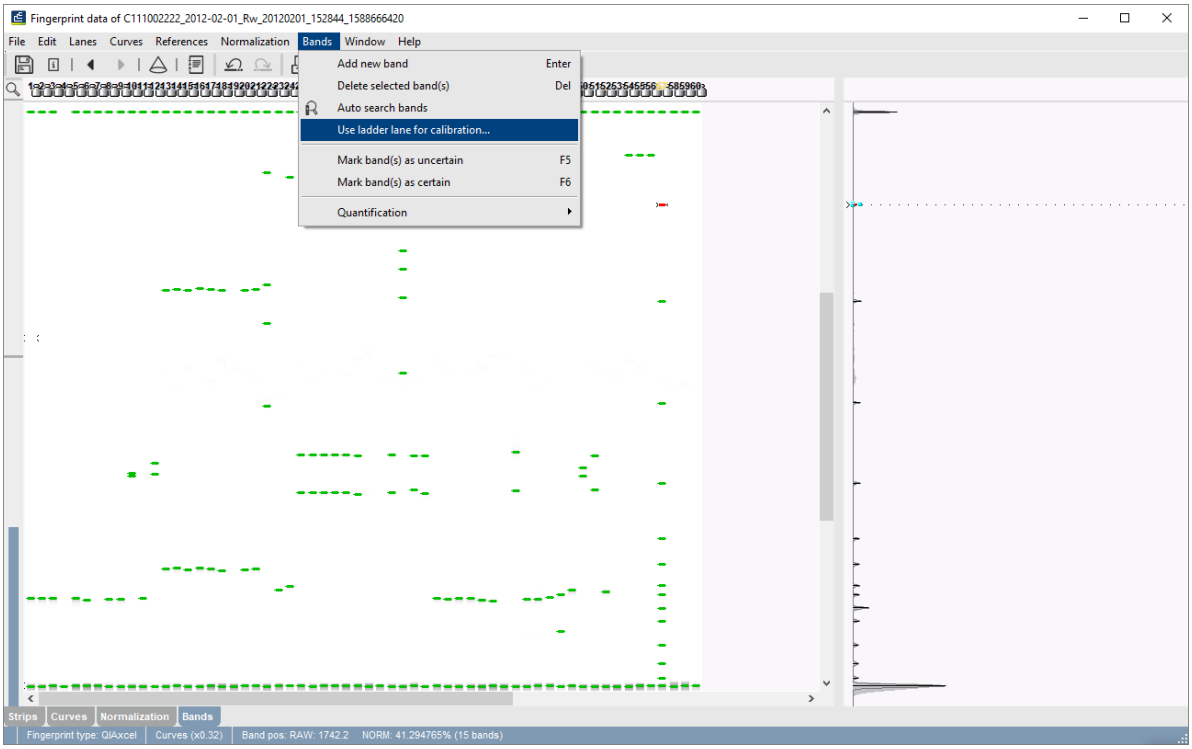

Figure 2.11: Use the bands of the ladder lane as calibration for the reference system.

5.14 Double-click on the experiment in the *Experiment types* panel of the *Main* window.

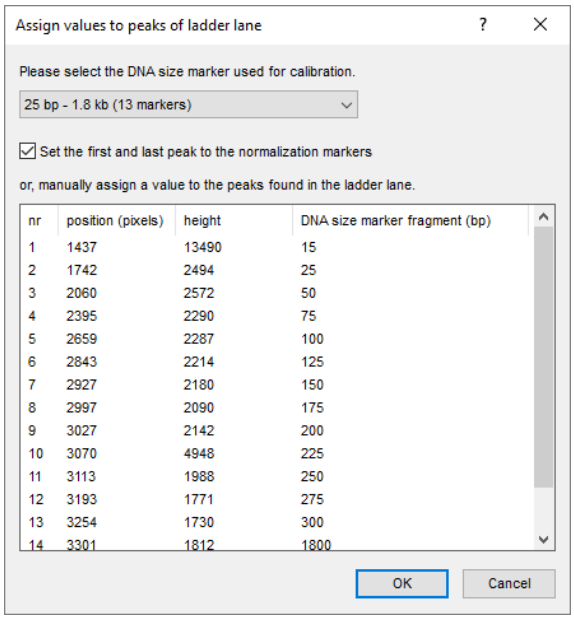

Figure 2.12: The *Assign values to peaks of ladder lane* dialog box: Select the DNA size marker or assign the values manually.

The reference system is displayed in the *Reference systems* panel of the *Fingerprint type* window.

5.15 Double-click on reference system in the *Reference systems* panel to call the *Fingerprint Reference system* window.

If a ladder lane was used for calibration (see Figure 2.11), the metrics and run lengths of the bands of the ladder lane are plotted and a cubic spline (default) is fitted through these points.

Chapter 3

Analyzing ScreenGel data

Some useful features in the context of ScreenGel data analysis will be highlighted in this chapter. More detailed information about the fingerprint analysis possibilities in the software can be found in the Reference manual, Chapter Setting up fingerprint type experiments.

3.1 Selections in BIONUMERICS

1.1 Select a single entry in the *Database entries* panel by holding the **Ctrl**-key and left-clicking on the entry. Alternatively, use the **space bar** to select a highlighted entry or click the ballot box next to the entry.

Selected entries are marked by a checked ballot box ( ) and can be unselected in the same way.

1.2 In order to select a group of entries, hold the **Shift**-key and click on another entry.

A group of entries can be unselected the same way.

1.3 All entries can be selected at once with ***Edit*** $>$ ***Select all*** (**Ctrl+A**).

1.4 Clear all selected entries with ***Database*** $>$ ***Entries*** $>$ ***Unselect all entries (all levels)*** (**F4**).

3.2 The Comparison window

2.1 Make a selection in the *Database entries* panel.

2.2 Highlight the *Comparisons* panel in the *Main* window and select ***Edit*** $>$ ***Create new object...*** (+) to create a new comparison for the selected entries.

2.3 You can drag the vertical separator lines between the panels to the left or to the right, in order to divide the space among the panels optimally.

2.4 Click on the next to the fingerprint experiment name in the *Experiments* panel to display the patterns in the *Experiment data* panel.

2.5 Press to show the band positions in the *Experiment data* panel.

3.3 Cluster analysis

3.1 In the *Experiments* panel of the *Comparison* window, make sure the correct fingerprint type experiment is selected.

3.2 Select ***Clustering*** $>$ ***Calculate*** $>$ ***Cluster analysis (similarity matrix)...***.

The coefficients are subdivided in two categories: ***Curve based*** and ***Band based***. The curve based coefficients calculate the similarities based upon the densitometric curves; the binary coefficients measure the similarity based upon common and different bands.

3.3 Select a similarity coefficient from the list and press $<$***Next***$>$.

3.4 In the second step, choose a clustering method (e.g. **UPGMA**) and press $<$***Finish***$>$.

When finished, the dendrogram and the similarity matrix are shown in the *Comparison* window.

3.5 Select ***File*** $>$ ***Save*** (, **Ctrl+S**) to save the comparison.

All calculations done on the data is stored along. This includes similarity matrices in all experiment types where they have been calculated and any dendrogram that has been calculated.

3.6 Enter a name, e.g. “MyComp” and press $<$***OK***$>$.

3.7 Close the comparison with ***File*** $>$ ***Exit***. The comparison **MyComp** is listed in the *Comparisons panel* of the *Main* window.

3.8 To open an existing comparison, highlight the comparison in the *Comparisons* panel and select ***Edit*** $>$ ***Open highlighted object...*** (, **Enter**). Alternatively, just double-click on the comparison name.

More information about the clustering of fingerprint data can be found in the Reference manual, Chapter Cluster analysis of fingerprints.

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