BIOMÉRIEUX

BIONUMERICS®

version 8 - PLUGINS

Import trend data tools 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

This guide is designed as a tutorial for the *Import trend data tools plugin* of BIONUMERICS. This plugin allows you to install additional import routines in the *Import data* wizard which can be accessed with ***File*** > ***Import...*** (, **Ctrl+I**):

- Import of ***BioScreen trend data files*** (see 2.1)
- Import of ***RisoSmart trend data files*** (see 2.2)

The *Import trend data tools plugin* is supported in all BIONUMERICS 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 Installing the Import trend data tools plugin

Installing a plugin in a BIONUMERICS database is done from the *Plugins and Scripts* dialog box (see Figure 1.2), which can be called from the *Main* window by selecting ***File*** > ***Install / remove plugins...*** ( ).

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

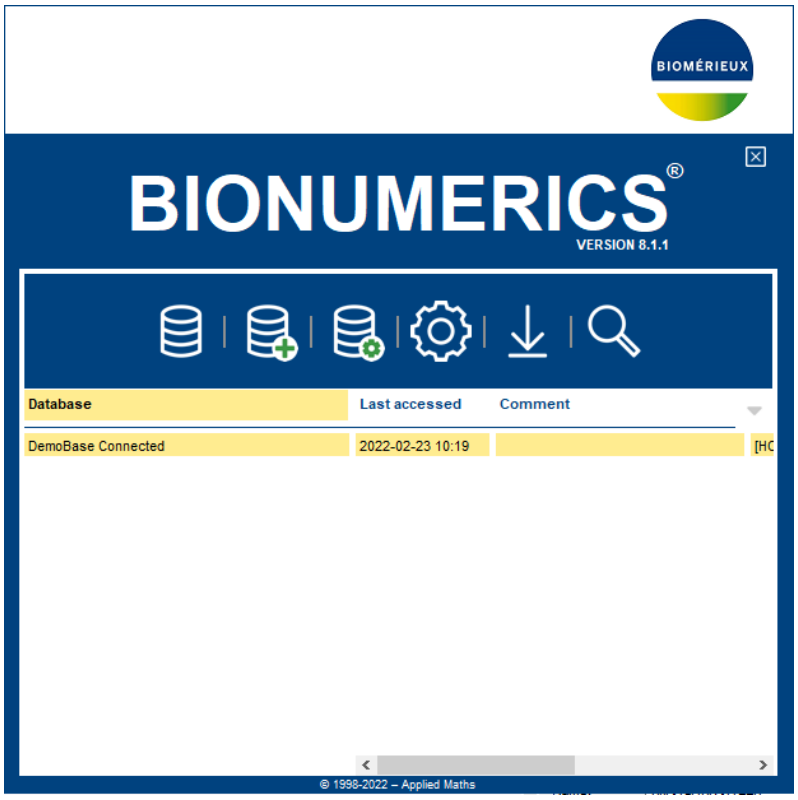

Figure 1.1: The *BIONUMERICS Startup* window.

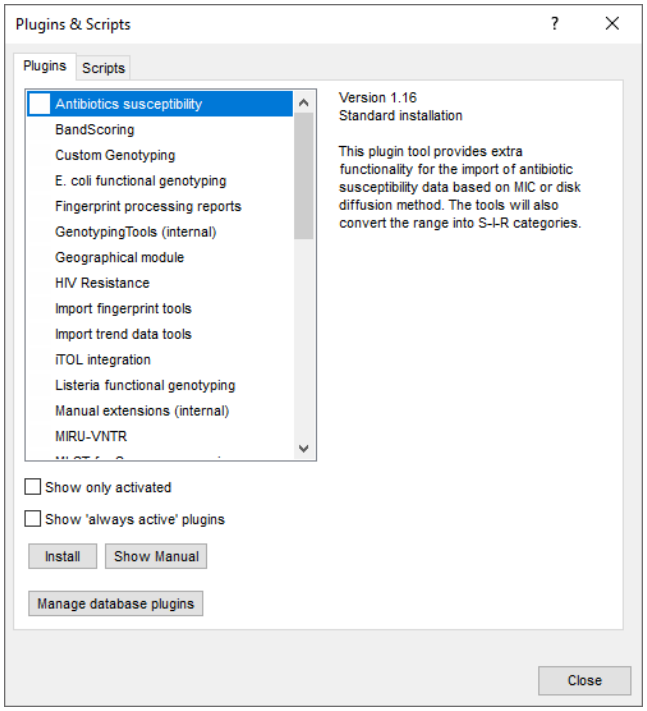

Figure 1.2: The *Plugins and Scripts* dialog box.

If the selected plugin is documented, pressing <***Show Manual***> will open its manual in the *Help* window. Proceed as follows to install the *Import trend data tools plugin*, starting from the *Plugins*

and Scripts dialog box:

3.1 Select the *Import trend data tools plugin* in the list and press the <***Install***> button.

3.2 Confirm the installation of the plugin (see Figure 1.3).

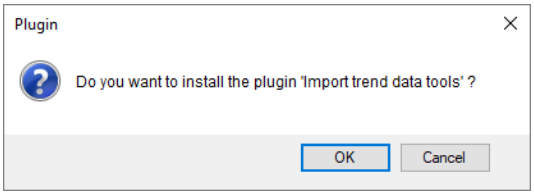

Figure 1.3: Confirm installation of the *Import trend data tools plugin*.

3.3 Confirm the installation of the plugin and press <***OK***>.

Once the plugin is successfully installed, it is marked with a green V-sign in the *Plugins and Scripts* dialog box (see Figure 1.4).

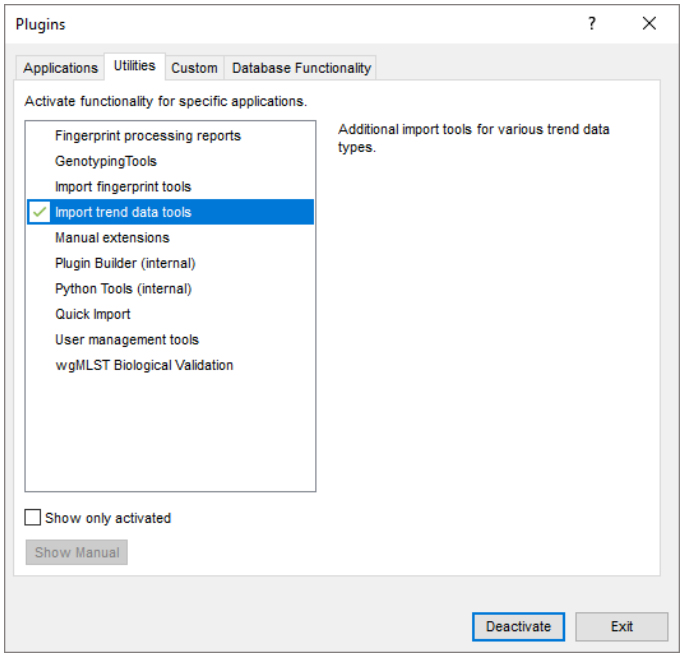

Figure 1.4: Installed plugin.

3.4 Close the *Plugins and Scripts* dialog box.

3.5 Close and reopen the database to activate the features of the *Import trend data tools plugin*.

Chapter 2

Import routines

2.1 Import of BioScreen trend data files

1.1 Select ***File*** > ***Import...*** (, **Ctrl+I**) to call the *Import data* wizard.

1.2 With <***Manual selection***> highlighted, press <***Next***> to show all import options in a tree view (see Figure 2.1).

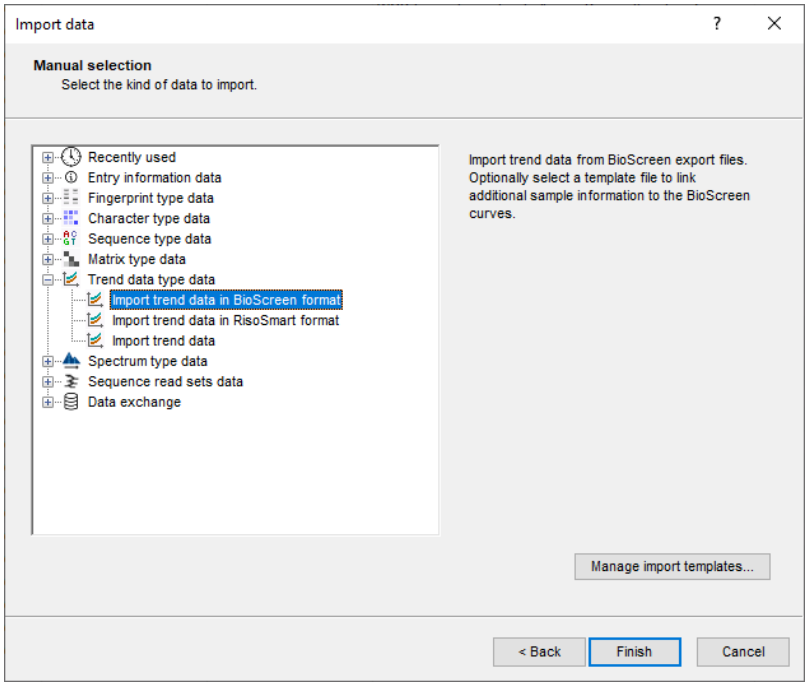

Figure 2.1: The second page of the *Import data* wizard.

1.3 Select ***Import trend data in BioScreen format*** under ***Trend data type data*** and press <***Finish***> button to start with the import of the data. Confirm the action.

The *Import BioScreen trend data* wizard page appears (see Figure 2.2).

The BioScreen import routine accepts BioScreen *csv files* and *txt files*.

Pressing the <***Browse***> button allows you to select the csv or text file(s) that you want to import, located on your computer, external drive or on a network location. Alternatively, files can be added to the import list through drag and drop. The number of files and total size is displayed below

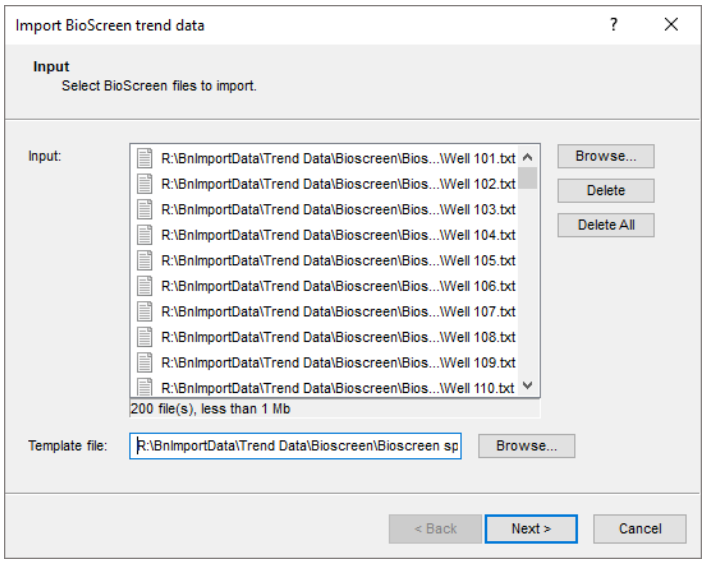

Figure 2.2: The *Import BioScreen trend data* wizard page.

the list. With the <***Delete***> button all selected files are removed from the import list. All files are deleted at once from the import list when pressing <***Delete All***>.

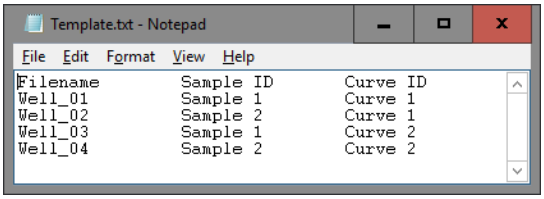

Figure 2.3: Template file.

Using a *template file* additional sample and curve information can be imported. A *template file* is a simple tab- or comma delimited file. The first column should contain the *file names* of the selected BioScreen files and the other columns should contain additional sample and curve information (see Figure 2.3 for an example).

1.4 Browse for the BioScreen file(s), optionally select a template file, and press <***Next***>.

The *Import rules* dialog box is displayed (see Figure 2.4).

When no template file is selected, only two rows are present in the grid. Double-clicking on the row opens a new dialog, where the data destination can be selected (see Figure 2.5). Using the last row in the grid, the (parsed) file name of the selected file(s) can be used. The text ***File*** is specified in the ***Source type*** column and the text ***Name*** is displayed in the ***Source*** column.

When a template file was selected, additional rows appear in the grid that can be linked to a destination in the database (see Figure 2.6 for an example).

1.5 Specify a *destination* for one or more selected rows by pressing the <***Edit destination***> button or by double-clicking. Make sure the correct row is selected as ***Trend data curve***.

1.6 Press <***Preview***> to check the defined rules (see Figure 2.7 for an example). Close the preview.

1.7 Press <***Next***> to go to the next step.

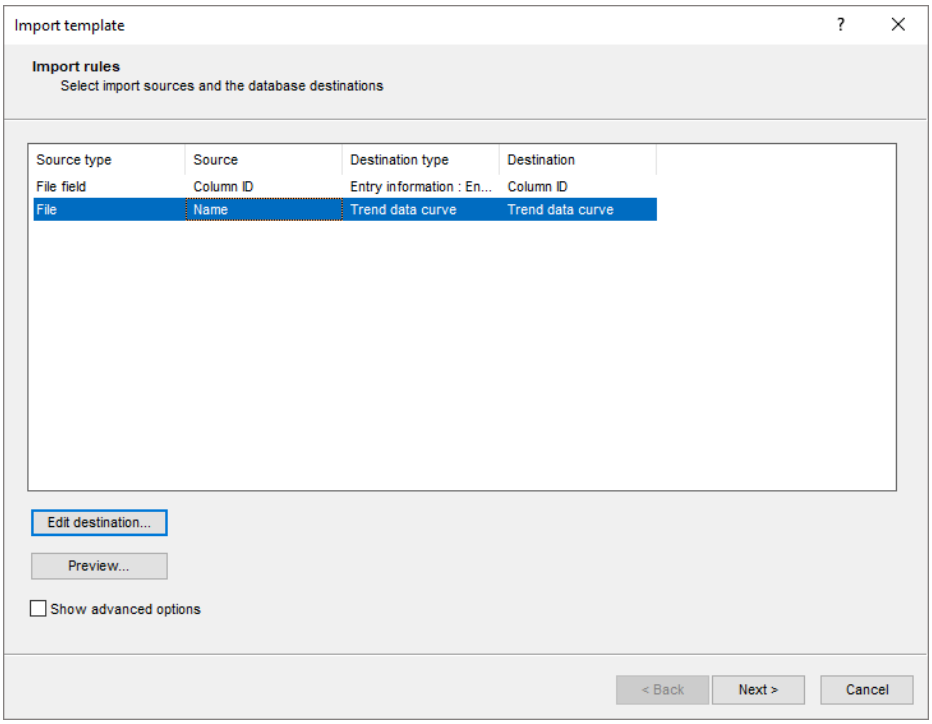

Figure 2.4: The *Import rules* dialog box.

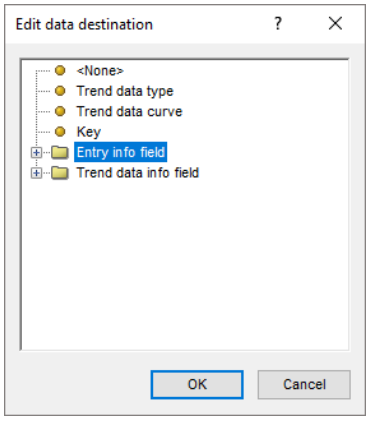

Figure 2.5: Edit data destination.

- If a row in the grid is linked to the ***Key*** field in the database, ***Key*** is automatically selected as the entry link field. If entries are already present in the database with the same (parsed) key information, the import tool will link the data to these entries.

- If no row entry in the grid is linked to the ***Key*** field, but one or more rows are linked to an entry information field in the database, these fields can be selected from the list. If entries are already present in the database with this linked information, the import tool 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 x entries*** is checked in the last step of the wizard).

- If no fields are selected from the list, no check for existing entries will be performed, and all data will be linked to new entries in the database (if the option ***Create x entries*** is checked in the last step of the wizard). New keys are automatically generated during import.

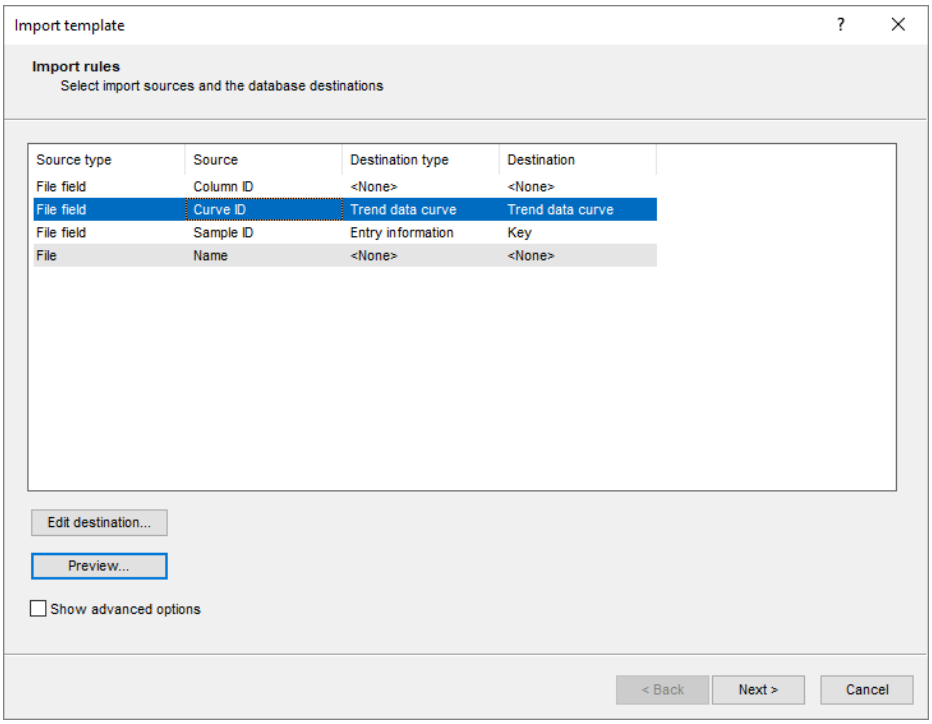

Figure 2.6: The *Import rules* dialog box with a linked template file.

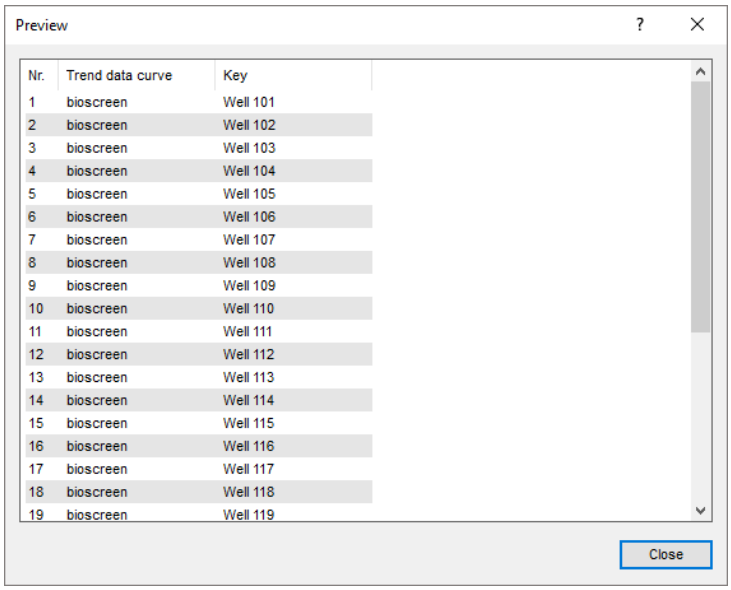

Figure 2.7: Preview.

1.8 Press <***Finish***> to go to the final step.

Each import template has its own unique ***Name***. Optionally, a descriptive text string can be entered in the ***Description*** input field.

1.9 Specify a template name (e.g. **Import BioScreen data**) and press <***OK***> to save all template settings to the database (see Figure 2.9).

When a template has been created and saved, the template ***Name*** is shown in the *Import templates panel* and is automatically selected (see Figure 2.10). The template ***Description*** is shown in panel on the right.

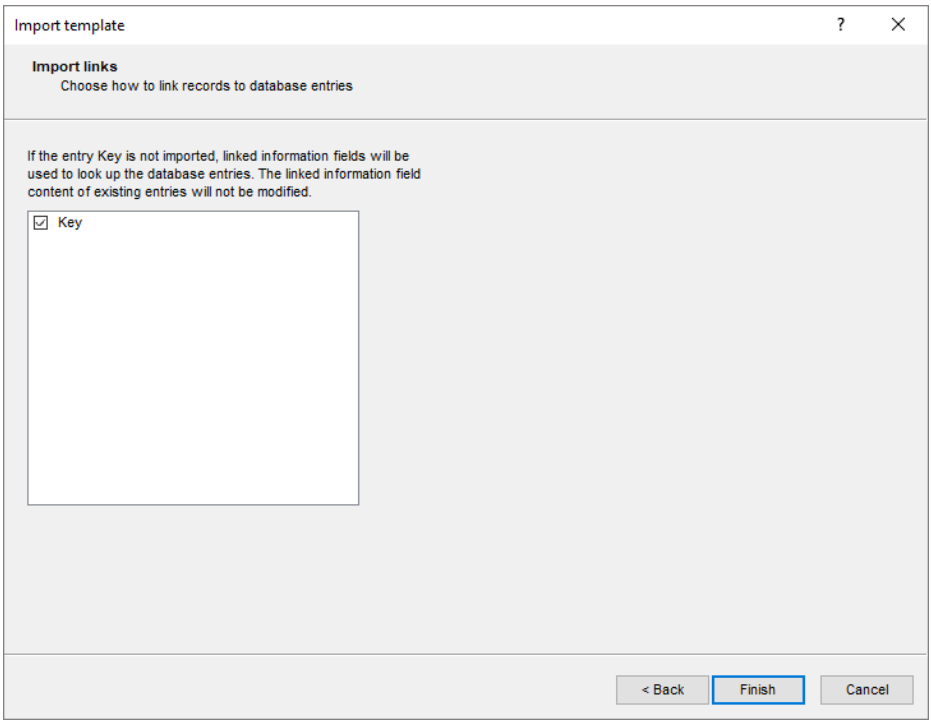

Figure 2.8: The *Import links* dialog box.

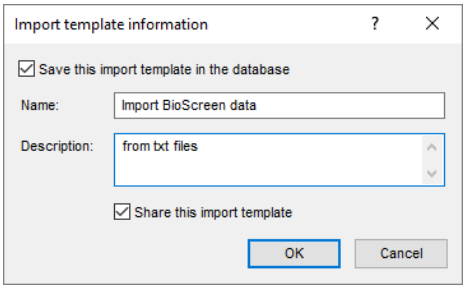

Figure 2.9: The *Import template information* dialog box.

The trend data can be linked to an existing trend type experiment or to a new trend type experiment (<***Create New***>). When the data is linked to a new trend type experiment, the next dialog will prompt for the trend type name (see Figure 2.11). The creation of the new experiment needs to be confirmed (twice).

When new trend curves need to be created, a new dialog will pop asking to confirm the creation of the curve(s) (see Figure 2.12).

1.10 Press <***Next***> to go to the next step.

The last step prompts for some final settings (see Figure 2.13).

- When ***Create x entries*** is checked, the import tool is allowed to create the new entries in the database.
- Check the option ***Update x entries*** if you want the software to be able to update the information for existing entries.
- If the option ***Select modified entries*** is checked, entries in the database that were modified during the import routine will be selected after import.

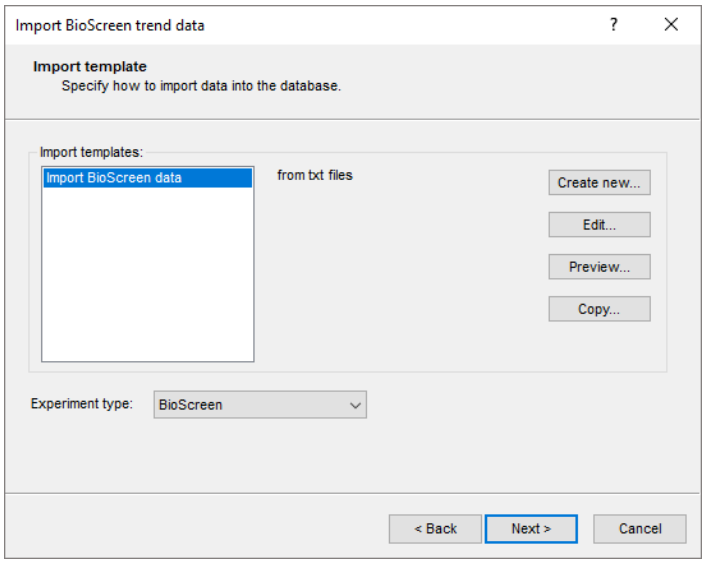

Figure 2.10: Import templates.

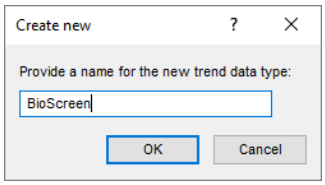

Figure 2.11: Create a new trend data type experiment.

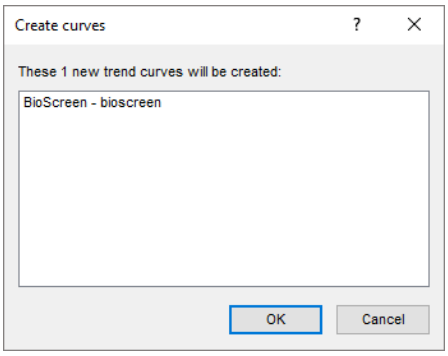

Figure 2.12: Create new trend curve(s).

1.11 Press <***Finish***> to start the import.

Entries are created/updated and are displayed in the *Database entries* panel of the *Main* window (see Figure 2.14). Linked sample information - if defined - is stored in the corresponding entry fields. When the option ***Select modified entries*** was checked, the new/updated entries are marked by a checked ballot box (☑).

The import trend data is linked to the appointed trend type experiment in the database. The presence of data for an entry/experiment combination is indicated with a green colored dot in the *Experiment presence* panel.

Clicking on a green colored dot opens the trend curve card for that entry (see Figure 2.15 for an example).

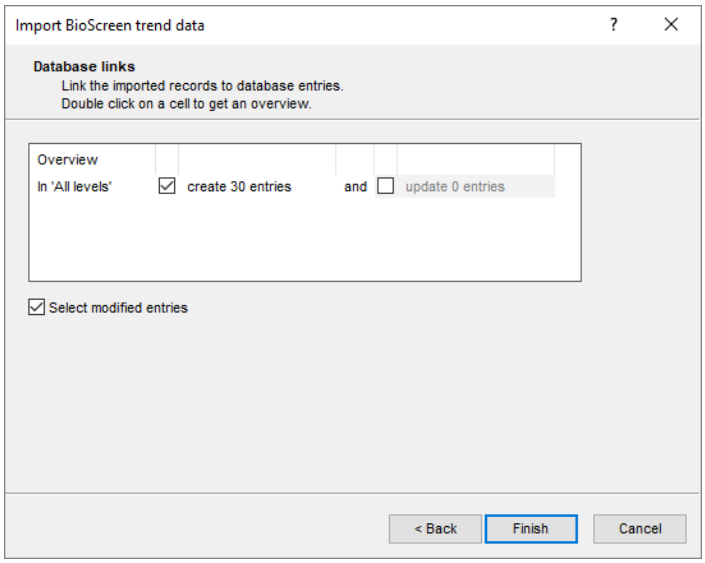

Figure 2.13: Database links.

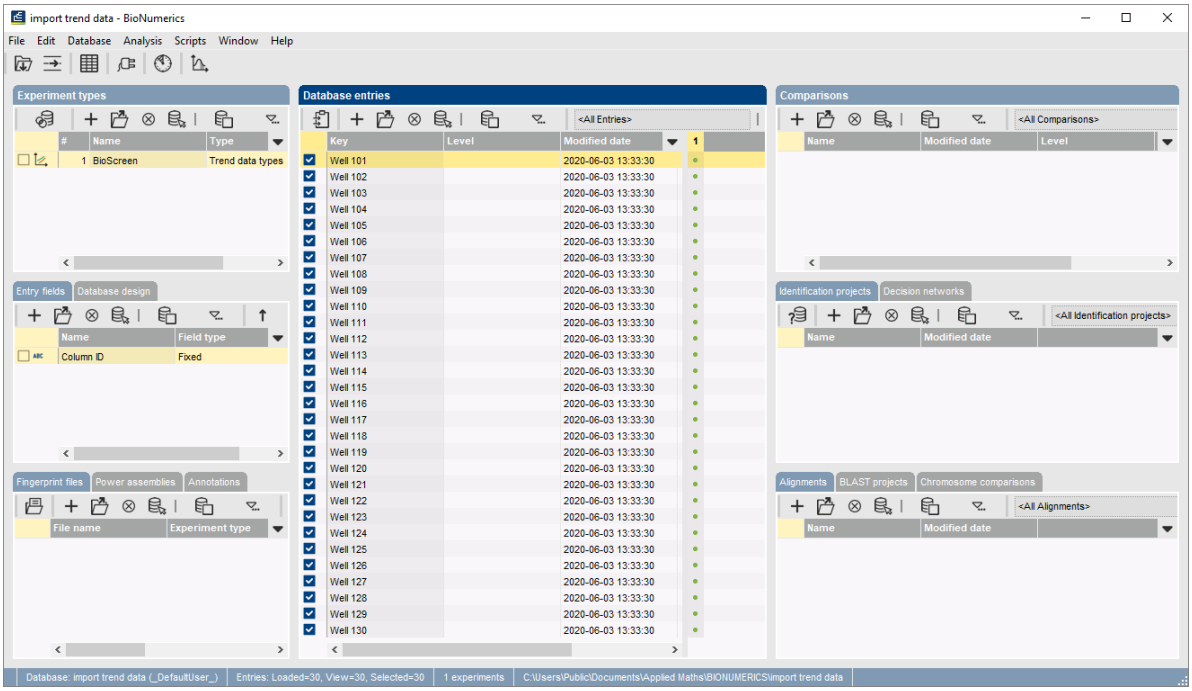

Figure 2.14: The *Main* window after import of BioScreen data.

2.2 Import of RisoSmart trend data files

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

2.2 With <***Manual selection***> highlighted, press <***Next***> to show all import options in a tree view (see Figure 2.16).

2.3 Select ***Import trend data in RisoSmart format*** under ***Trend data type data*** and press <***Finish***> button to start with the import of the data. Confirm the action.

The *Import RisoSmart trend data* wizard page appears (see Figure 2.17).

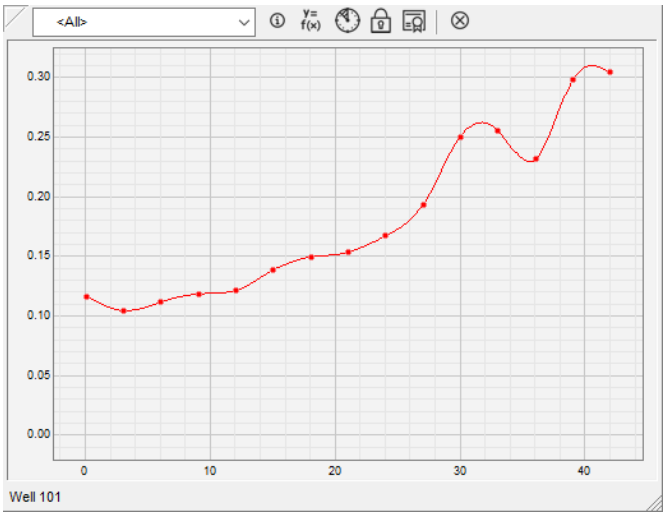

Figure 2.15: Experiment card.

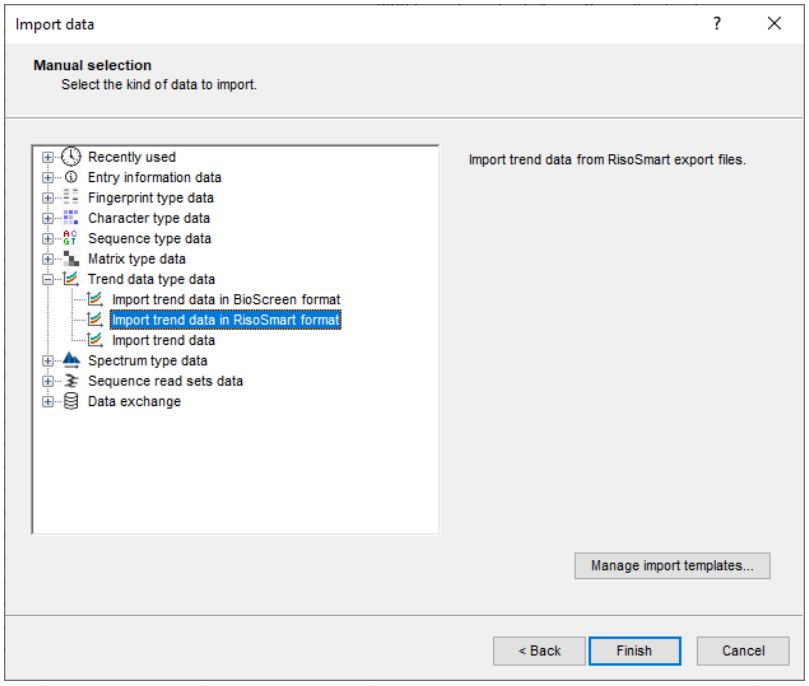

Figure 2.16: The second page of the *Import data* wizard.

The RisoSmart import routine accepts RisoSmart *csv files* and *txt files*.

Pressing the <***Browse***> button allows you to select the csv or text file(s) that you want to import, located on your computer, external drive or on a network location. Alternatively, files can be added to the import list through drag and drop. The number of files and total size is displayed below the list. With the <***Delete***> button all selected files are removed from the import list. All files are deleted at once from the import list when pressing <***Delete All***>.

2.4 Browse for the RisoSmart file(s) and press <***Next***>.

The *Import rules* dialog box is displayed (see Figure 2.18).

For every header field detected in the selected file(s), a row is present in the grid. Double-clicking on the row opens a new dialog, where the data destination can be selected (see Figure 2.19).

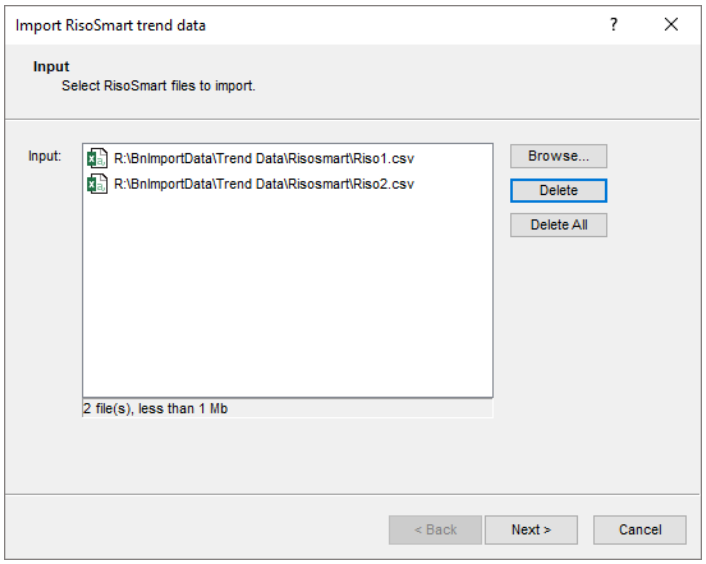

Figure 2.17: The *Import RisoSmart trend data* wizard page.

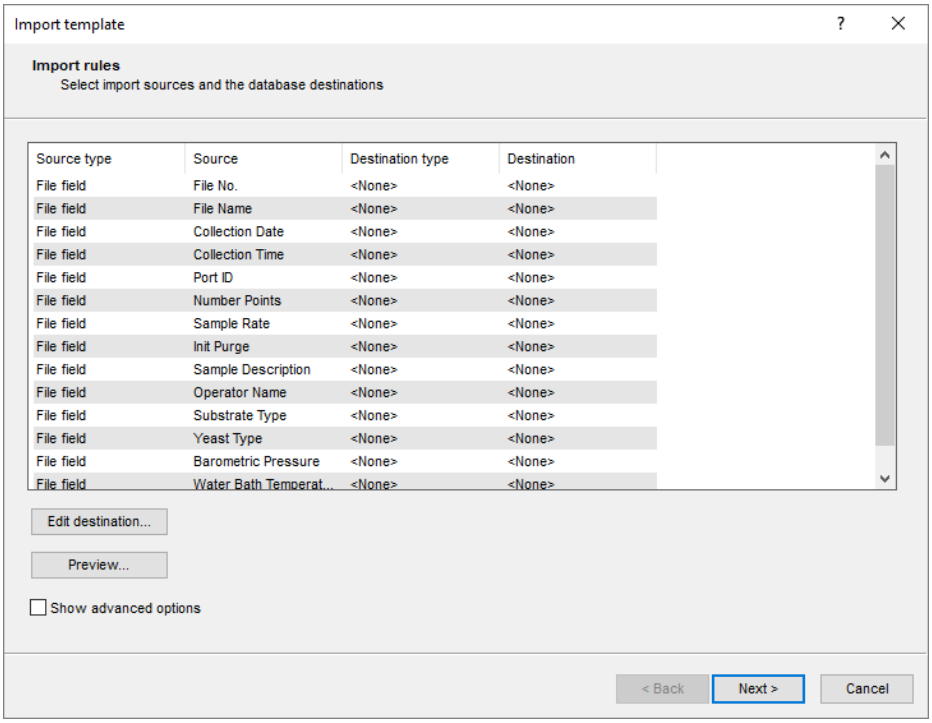

Figure 2.18: The *Import rules* dialog box.

Using the last row in the grid, the (parsed) file name of the selected file(s) can be used. The text ***File*** is specified in the ***Source type*** column and the text ***Name*** is displayed in the ***Source*** column.

2.5 Specify a *destination* for one or more selected rows by pressing the <***Edit destination***> button or by double-clicking. Make sure the correct row is selected as ***Trend data curve***.

2.6 Press <***Preview***> to check the defined rules. Close the preview.

2.7 Press <***Next***> to go to the next step (see Figure 2.20).

- If a row in the grid is linked to the ***Key*** field in the database, ***Key*** is automatically selected

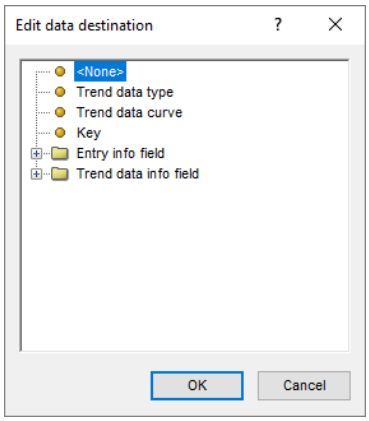

Figure 2.19: Edit data destination.

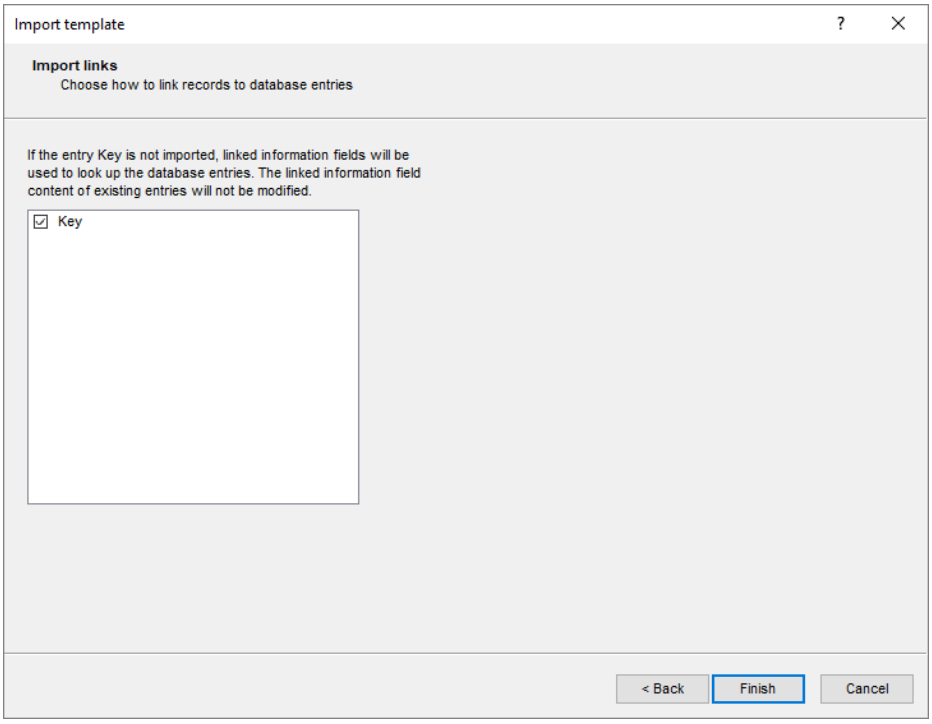

Figure 2.20: The *Import links* dialog box.

as the entry link field. If entries are already present in the database with the same (parsed) key information, the import tool will link the data to these entries.

- If no row entry in the grid is linked to the ***Key*** field, but one or more rows are linked to an entry information field in the database, these fields can be selected from the list. If entries are already present in the database with this linked information, the import tool 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 x entries*** is checked in the last step of the wizard).
- If no fields are selected from the list, no check for existing entries will be performed, and all data will be linked to new entries in the database (if the option ***Create x entries*** is checked in the last step of the wizard). New keys are automatically generated during import.

2.8 Press <***Finish***> to go to the final step.

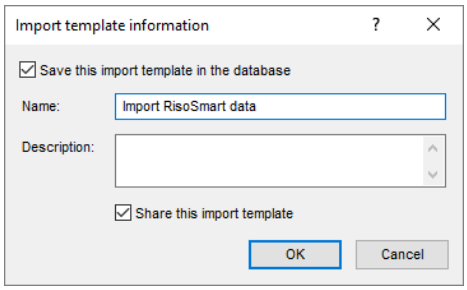

Figure 2.21: The *Import template information* dialog box.

Each import template has its own unique ***Name***. Optionally, a descriptive text string can be entered in the ***Description*** input field.

2.9 Specify a template name (e.g. **Import RisoSmart data**) and press <***OK***> to save all template settings to the database (see Figure 2.21).

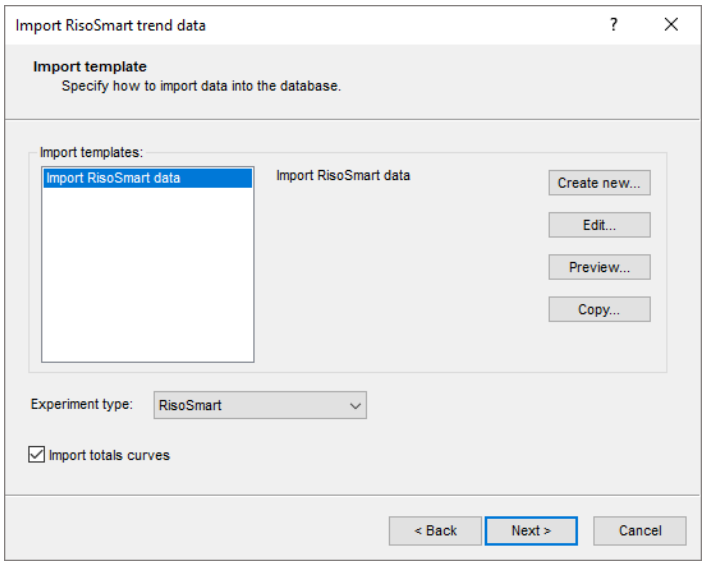

Figure 2.22: Import templates.

When a template has been created and saved, the template ***Name*** is shown in the *Import templates panel* and is automatically selected (see Figure 2.22). The template ***Description*** is shown in panel on the right.

The trend data can be linked to an existing trend type experiment or to a new trend type experiment (<***Create New***>). When the data is linked to a new trend type experiment, the next dialog will prompt for the trend type name (see Figure 2.23). The creation of the new experiment needs to be confirmed.

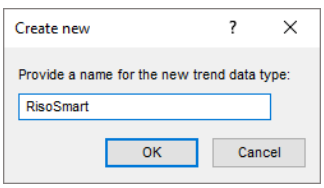

Figure 2.23: Create new trend data type experiment.

Typically the RisoSmart file(s) also contain the cumulative data of the measurements. This data

will be stored in trend data type experiment, composed of the name of the linked trend data type experiment followed by the text "_totals". If this experiment is not present in the database, the creation of this experiment needs to be confirmed by the user.

When new trend curves need to be created, a new dialog will pop asking to confirm the creation of the curve(s).

2.10 Press <***Next***> to go to the next step.

The last step prompts for some final settings (see Figure 2.24).

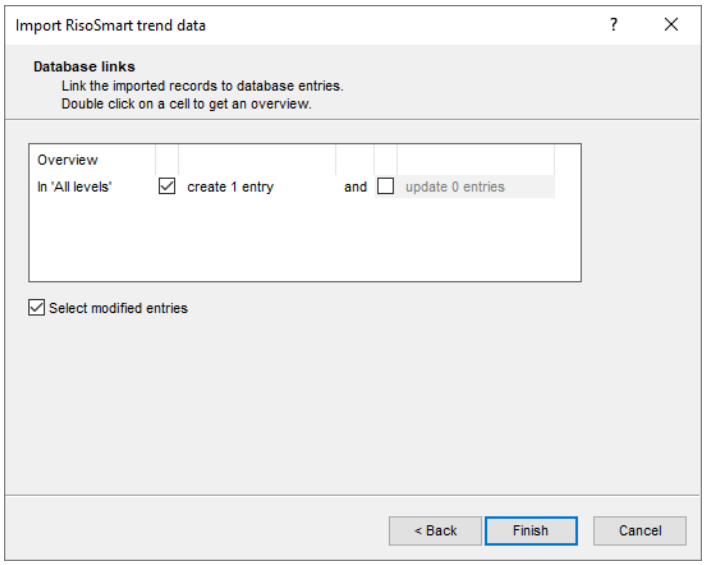

Figure 2.24: Database links.

- When ***Create x entries*** is checked, the import tool is allowed to create the new entries in the database.
- Check the option ***Update x entries*** if you want the software to be able to update the information for existing entries.
- If the option ***Select modified entries*** is checked, entries in the database that were modified during the import routine will be selected after import.

2.11 Press <***Finish***> to start the import.

Entries are created/updated and are displayed in the *Database entries* panel of the *Main* window (see Figure 2.25). Linked sample information - if defined - is stored in the corresponding entry fields. When the option ***Select modified entries*** was checked, the new/updated entries are marked by a checked ballot box (☑).

The import trend data is linked to the appointed trend type experiments in the database. The presence of data for an entry/experiment combination is indicated with a green colored dot in the *Experiment presence* panel.

Clicking on a green colored dot opens the trend curve card for that entry (see Figure 2.26 for an example).

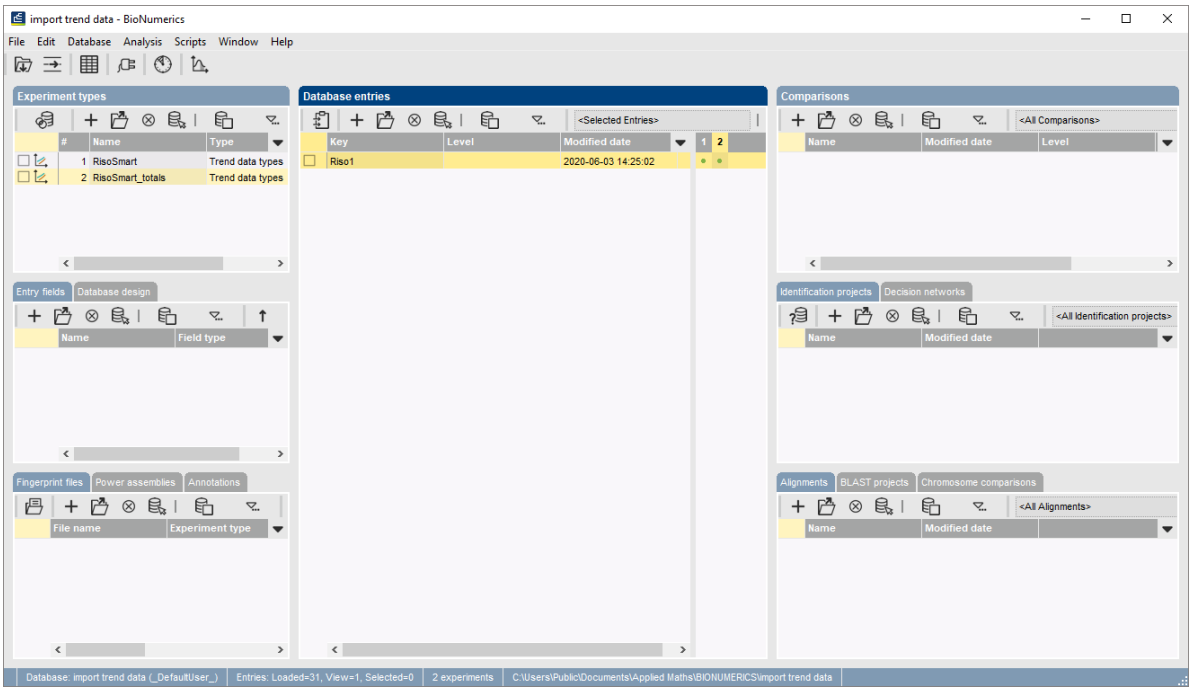

Figure 2.25: The *Main* window after import of the data.

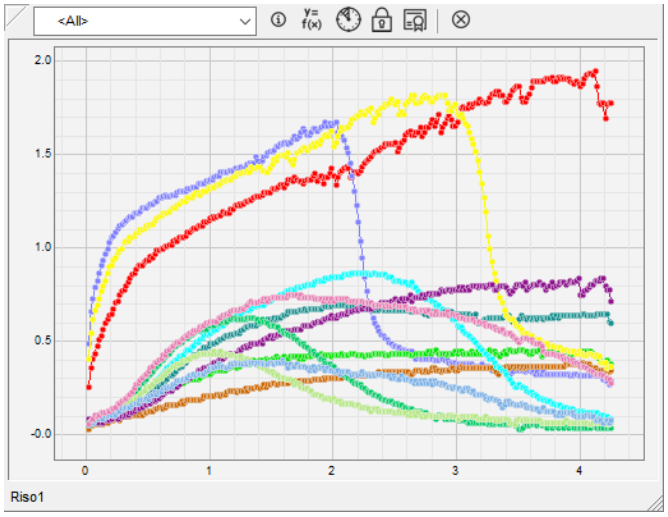

Figure 2.26: Trend card experiment.

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