

RiboPrinter plugin

PLUGINS
VERSION 7.6



Contents

1	Starting and setting up BioNumerics	3
1.1	Introduction	3
1.2	Startup program	3
1.3	Creating a new database	3
1.4	Installing the RiboPrinter plugin	4
2	Importing RiboPrinter patterns	7
2.1	The Import tree	7
2.2	The Import wizard	7
3	Analyzing RiboPrinter patterns	19
3.1	Selections in BioNumerics	19
3.2	The Comparison window	19
3.3	Cluster analysis	19

NOTES

SUPPORT BY APPLIED MATHS

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 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: info@applied-maths.com
URL: <http://www.applied-maths.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: info-US@applied-maths.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, 2018, 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:

- The Python[®] 2.7.4 release from the Python Software Foundation (<http://www.python.org/>).
- A library for XML input and output from the Apache Software Foundation (<http://www.apache.org>).
- NCBI toolkit version 2.2.10 (<http://www.ncbi.nlm.nih.gov/BLAST/>).
- The Boost c++ libraries (<http://www.boost.org/>).
- Samtools for interacting with SAM / BAM files (<http://www.htslib.org/download/>)
- The 7-Zip command line version (7za.exe) from 7-Zip, copyright 1999-2010 Igor Pavlov. <http://www.7-zip.org/>
- Velvet for Windows, source code can be downloaded from <http://www.applied-maths.com/download/open-source>.
- Ray for Windows, source code can be downloaded from <http://www.applied-maths.com/download/open-source>.
- Mothur for Windows, source code can be downloaded from <http://www.applied-maths.com/download/open-source>.
- Cairo 2D graphics library version 1.12.14 (<http://cairographics.org/>).
- Crypto++ Library version 5.5.2 (<http://www.cryptopp.com/>).
- libSVM library for Support Vector Machines (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>).
- SQLite version 3.7.17 (<http://www.sqlite.org/>).
- Gecko engine version 21 (<https://developer.mozilla.org/en-US/docs/Mozilla/Gecko>).
- pymzML Python[®] module for high throughput bioinformatics on mass spectrometry data (<https://github.com/pymzml/pymzML>).
- Numpy Python[®] library version 1.8.1 (<http://www.numpy.org/>).
- BioPython Python[®] library version 1.64 (<http://www.biopython.org/>).
- PIL Python library[®] version 1.1.7 (<http://www.pythonware.com/products/pil/>).
- The SPAdes genome assembler version 3.7.1 (<http://bioinf.spbau.ru/spades>).

Chapter 1

Starting and setting up BioNumerics


1.1 Introduction

The RiboPrinter System from DuPont Qualicon (<http://www2.dupont.com>) is an automated system that is used for the characterization of bacterial samples. The RiboPrinter system generates a DNA fingerprint of regions of ribosomal RNA genes that is unique for the organism at the strain level. These DNA fingerprints are known as RiboPrint patterns and can be exported from the RiboPrinter system in xml format and text file format. Both formats can be imported in BioNumerics with the *RiboPrinter plugin*.


1.2 Startup program

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*.

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 Since we want to create a new database to demonstrate the features of the plugin, leave the default option selected and press <Next>.

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

3.4 Leave the default option selected and press <Next>.

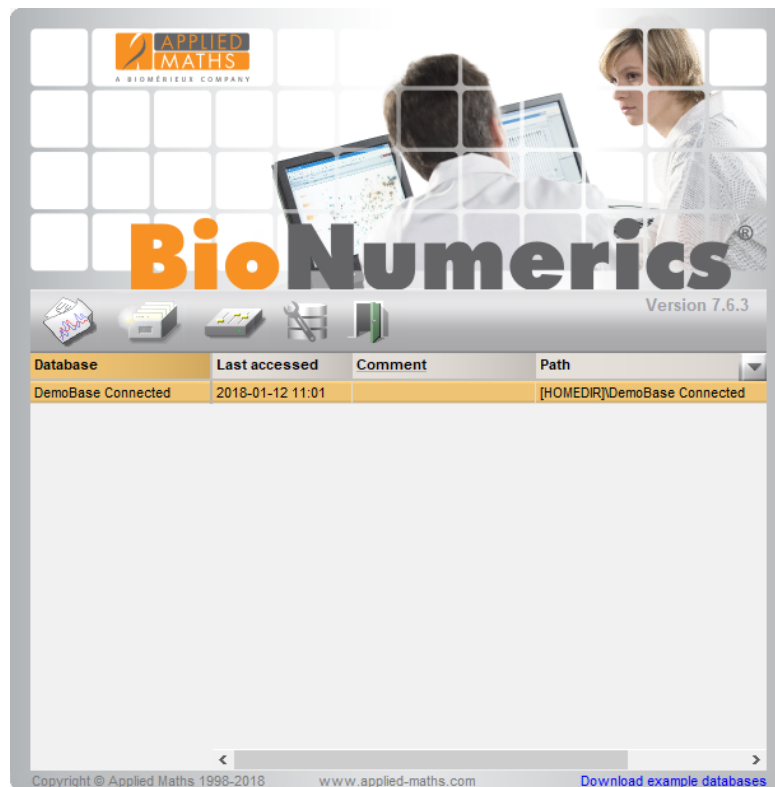


Figure 1.1: The *BioNumerics* Startup window.

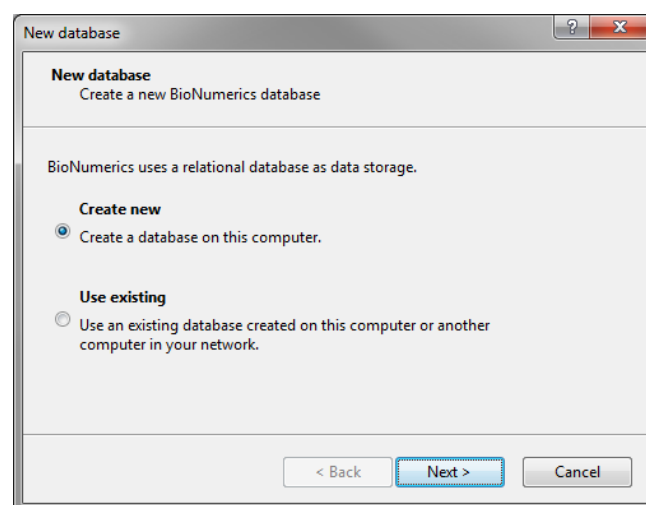


Figure 1.2: The *New database* wizard page.

3.5 Press <*Finish*> to complete the setup of the new database.

The *Plugins* dialog box appears.

1.4 Installing the RiboPrinter plugin

If a database is opened for the first time, the *Plugins* dialog box will appear by default (see Figure 1.4).

If the database has already been opened previously, the *Plugins* dialog box can be called from the *Main*

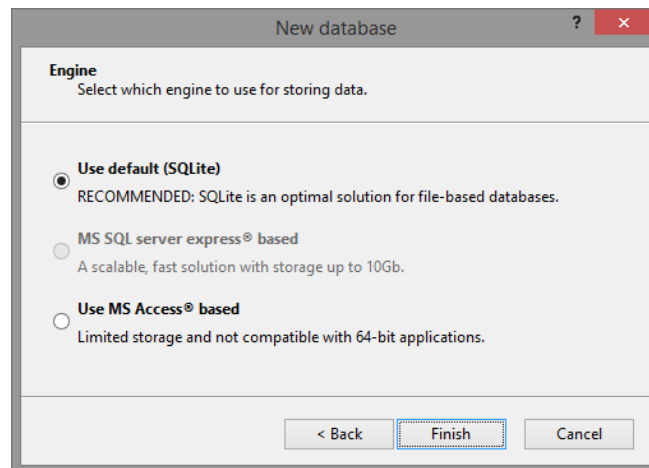


Figure 1.3: The *Database engine* wizard page.

window by selecting **File > Install / remove plugins...** (🔧).

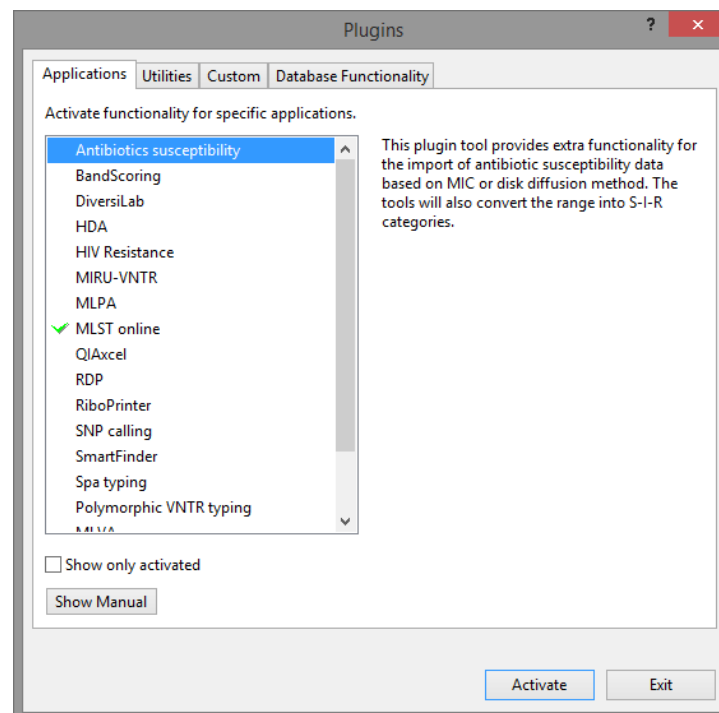


Figure 1.4: The *Plugins* dialog box.

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 **<Activate>** button. The software will ask for confirmation before installation. Some plugins depend on functionality offered by specific BioNumerics modules. If a required module is missing, the plugin 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 **<Deactivate>** button.

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

4.1 Select the *RiboPrinter* plugin from the list in the *Applications* tab and press the **<Activate>** button.

4.2 Press **<Proceed>** (or **<Exit>**) to close the *Plugins* dialog box and to continue to the *Main* window.

4.3 Close and reopen the database to activate the features of the *RiboPrinter* plugin.

4.4 Select **File > Import...** (📁, **Ctrl+I**) to call the *Import* dialog box.

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

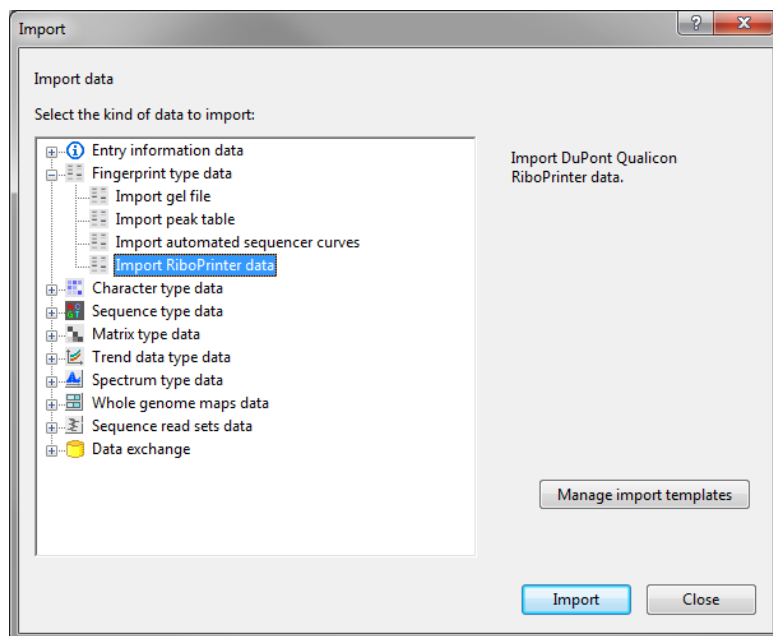


Figure 1.5: Import RiboPrinter data menu item in the Import plugin tree.

4.5 Close the *Import* dialog box.

Chapter 2

Importing RiboPrinter patterns

2.1 The Import tree

1.1 Select **File > Import...** (📁➡️), **Ctrl+I** to call the *Import* dialog box.

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

The *Input* dialog box appears (see Figure 2.1).

2.2 The Import wizard

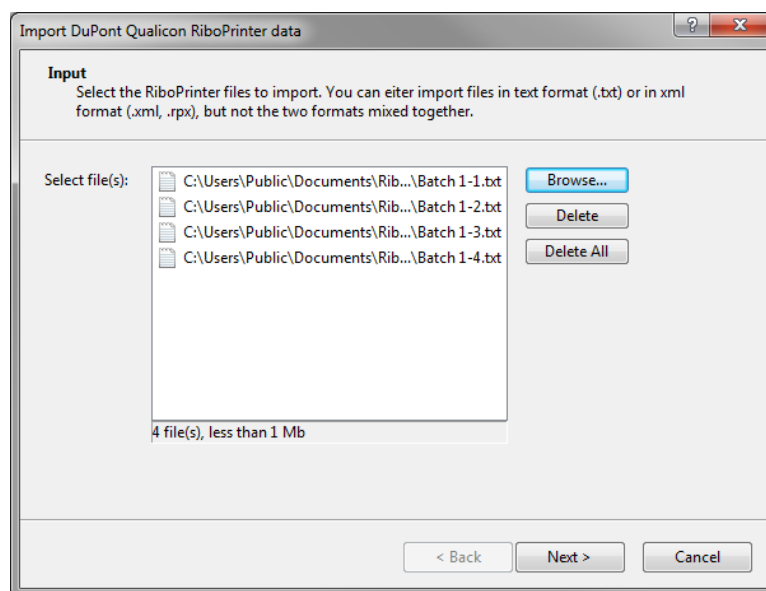


Figure 2.1: The *Input* dialog box.

The RiboPrinter import routine accepts RiboPrinter *xml files* with the extension ***.rpx** and ***.xml** and RiboPrinter *text files* with the extension ***.txt** (also known as the *qnx file* format).

Pressing the **<Browse>** button allows you to select the xml 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>**.

2.1 Browse for the RiboPrinter xml or text file(s) and press **<Next>**.

The *Import template* wizard page is displayed (see Figure 2.2).

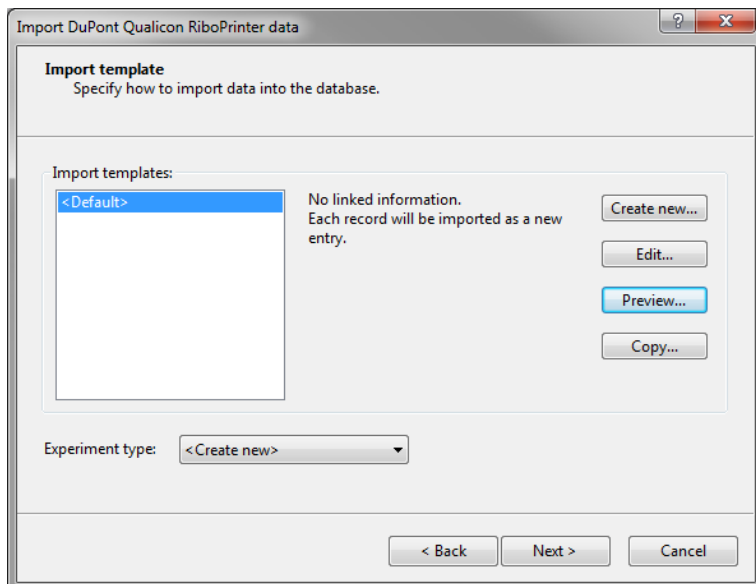


Figure 2.2: The *Import template* wizard page.

The way the fingerprint information should be imported and stored in the database should be specified with an import template. The *Import templates panel* lists all RiboPrinter templates that have been created and stored in the database. Initially only the **Default** template is listed. This template will link the RiboPrinter patterns to the selected **Experiment type** experiment and will create a new entry for each new pattern (if the option **Create new entries** is checked in the final step). The keys are automatically created by the import routine and no additional sample and batch information is imported.

Because the RiboPrinter files contain additional sample and batch information, it is recommended to create a custom import template (**<Create new>**) where the links between the file information and the database can be specified.

2.2 Press **<Create new>** to call the *Import rules* dialog box (see Figure 2.3 and Figure 2.4).

The batch and sample tags, parsed from the selected file(s), correspond to rows in the grid. The text **Batch** and **Sample** is specified in the **Source type** column and the name of the tag is displayed in the **Source** column:

General sample tags:

- **Sample number:** unique identifier for the sample; composed of the batch number (= combination of the instrument and accession number), letter S (sample) or M (marker), and the lane number, e.g. **200-220-S-5**
- **Genus:** suggested Genus name entered by the user
- **Species:** suggested Species name entered by the user
- **Sample Comment:** comment entered by the user
- **Strain number:** strain number specified by the user

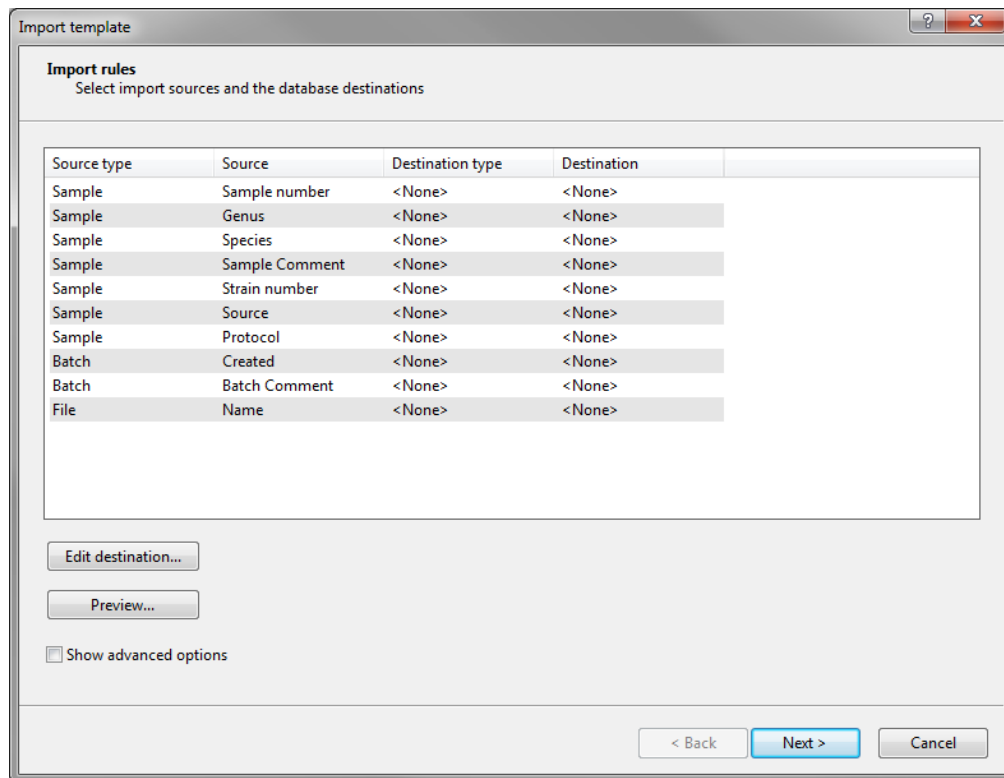


Figure 2.3: The *Import rules* dialog box: RiboPrinter xml file tags.

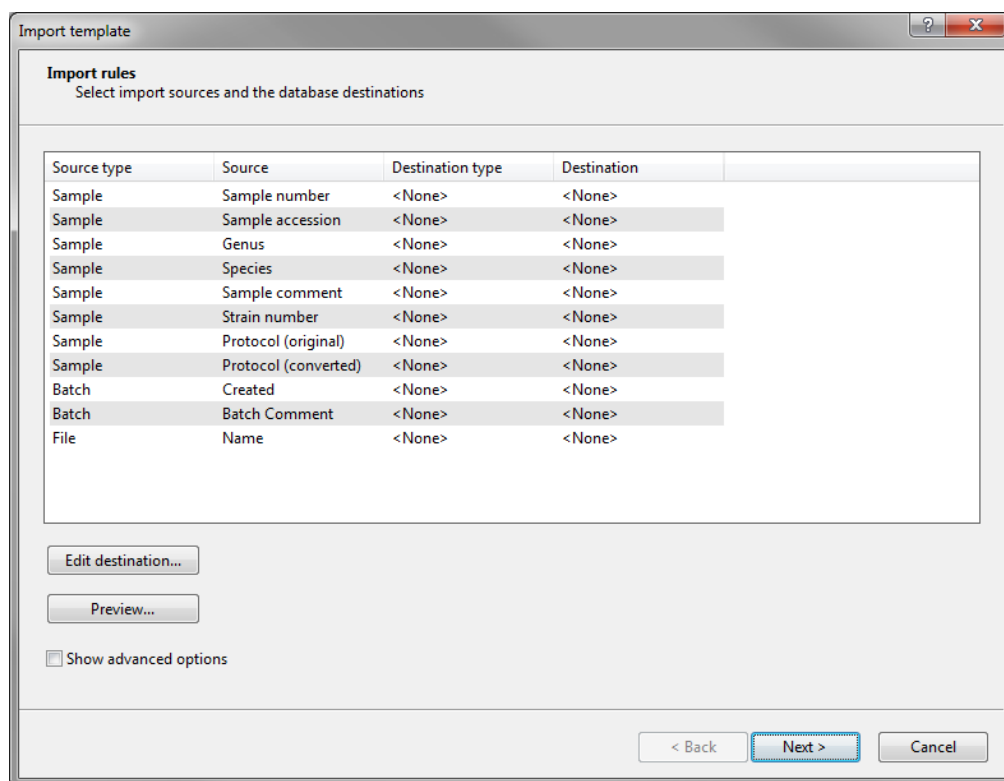


Figure 2.4: The *Import rules* dialog box: RiboPrinter text file tags.

Sample tags specific for RiboPrinter xml files (see Figure 2.3):

- **Source:** GUID (can be resolved to a country name in a future version of the plugin)
- **Protocol:** restriction enzyme used in the protocol (e.g. EcoRI, PstI, PVuII)

Sample tags specific for RiboPrinter text files (see Figure 2.4):

- **Sample accession:** composed of the batch number (= combination of the instrument and accession number) and the lane number
- **Protocol (original):** protocol name (e.g. VCA, VCB, VCC)
- **Protocol (converted):** restriction enzyme used in the protocol (e.g. EcoRI, PstI, PVuII)

Batch tags:

- **Created:** creation date of the batch
- **Batch Comment:** comment entered by the user

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

2.3 Specify a *destination* for one or more selected rows by pressing the **<Edit destination>** button or by double-clicking.



To select multiple rows, hold the **Ctrl**-key and click the rows to be selected. To select a complete range of rows at once, select the first row, hold the **Shift**-key and click the last row.

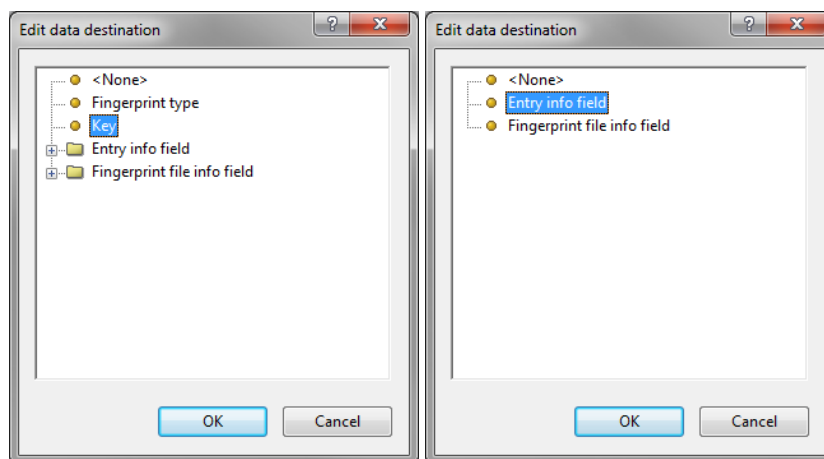


Figure 2.5: Edit data destination for one (left) or multiple (right) rows.

When only one row is selected in the grid, the information of this row can be linked to:

- The default information field **Key**.
- A new or existing non-default entry information field (select the **<Create new>** option or an existing field under the topic **Entry info field**, respectively).
- A new or existing fingerprint file information field (select **<Create new>** or select an existing field under the topic **Fingerprint file info field**, respectively).
- A **Fingerprint type** experiment.

When multiple rows are selected in the grid, the information of these rows can be linked to:

- Non-default entry information fields (select the **Entry info field** option).
- Fingerprint information fields (select the **Fingerprint file info field** option).

When pressing the **<OK>** button, the plugin checks if the selected row(s) can automatically be mapped to field(s) in the database. If no field(s) exist(s) with the same name, a new dialog box pops up prompting for the name(s).

2.4 Specify a data destination for some/all **Batch** and **Sample** tags (examples: Figure 2.6 and Figure 2.7).

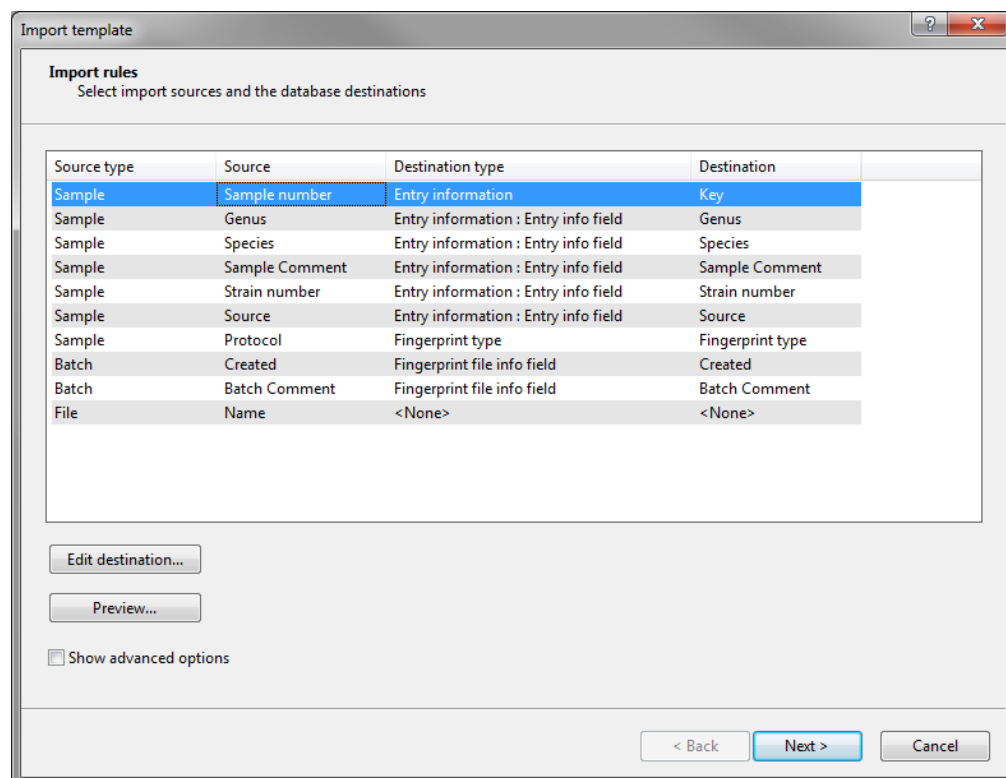


Figure 2.6: Import rules defined for the RiboPrinter xml file tags: the unique **Sample number** is linked to the **Key** field; the **Batch** fields are linked to **Fingerprint file information fields**; the **Sample** fields (except **Protocol**) are linked to **Entry information fields**; the **Protocol** information is used as **Fingerprint type** name.

If no row entry in the grid is linked to a **Fingerprint type** destination, the patterns will be linked to the fingerprint type experiment selected from the list in the *Import template* window (see Figure 2.2).

2.5 Press the **<Preview>** button to open the *Preview* dialog box.

This dialog displays the parsed information using the template settings.

2.6 Close the preview with the **<Close>** button.

2.7 Press the **<Next>** button to call the dialog where the entry link field needs to be defined (see Figure 2.8).

- 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.

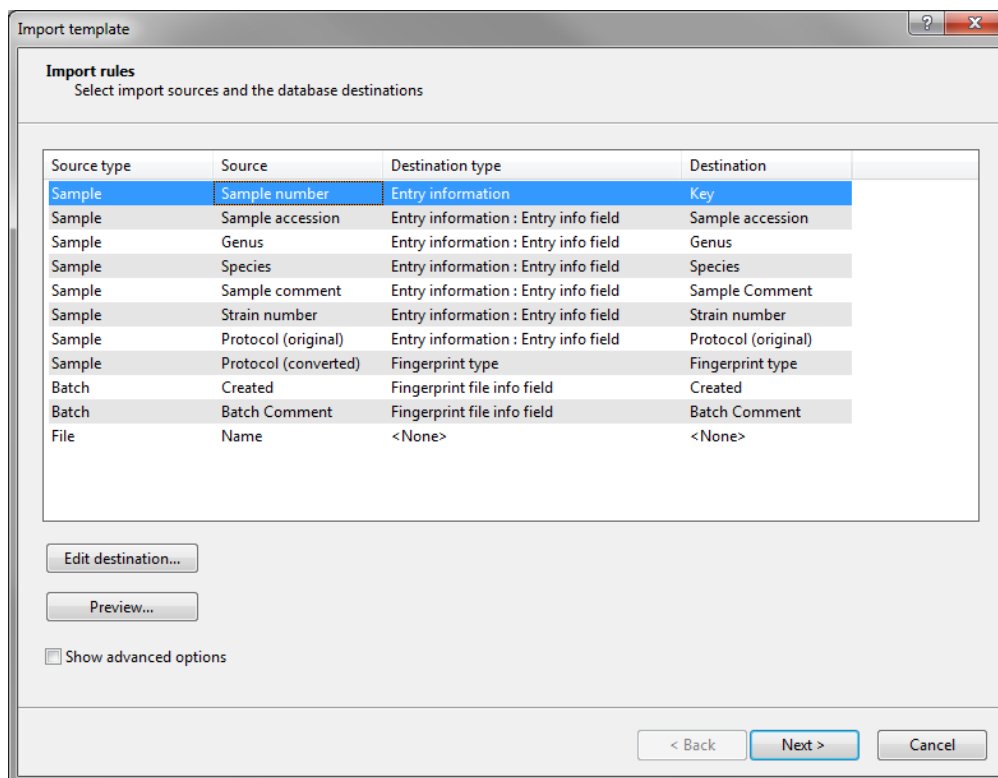


Figure 2.7: Import rules defined for the RiboPrinter text file tags: the unique *Sample number* is linked to the *Key* field; the *Batch* fields are linked to *Fingerprint file information fields*; the *Sample* fields (except *Protocol (converted)*) are linked to *Entry information fields*; the *Protocol (converted)* information is used as *Fingerprint type* name.

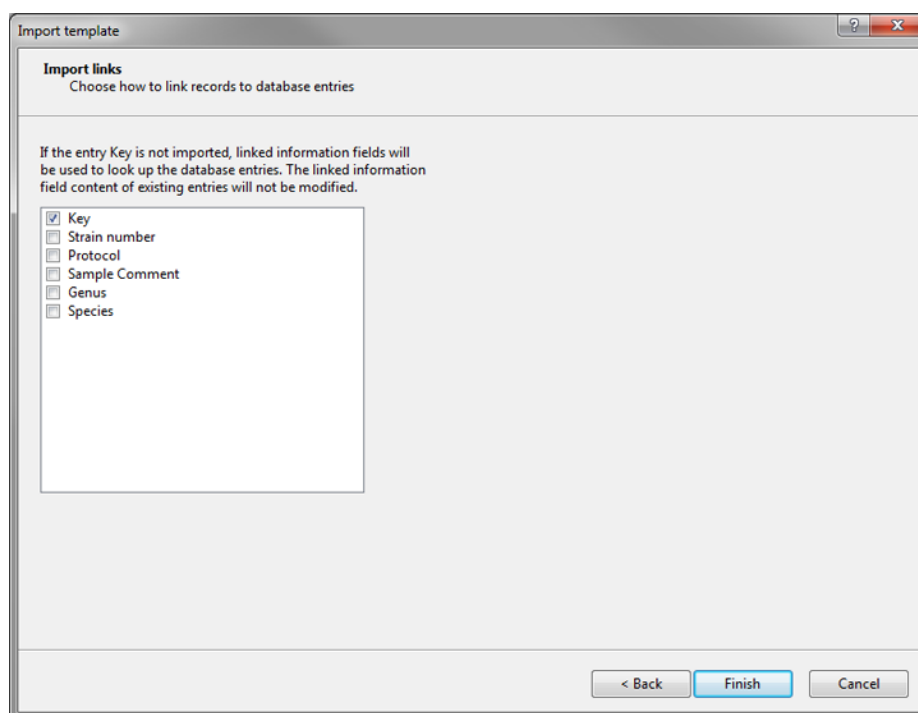


Figure 2.8: The *Import links* dialog box.

- 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.

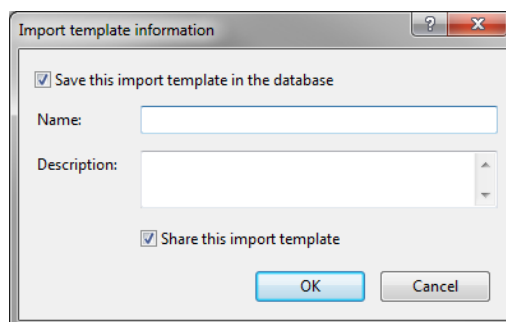


Figure 2.9: The *Import template information* dialog box.

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 RiboPrinter XML files**) and press **<OK>** to save all template settings to the database.

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.

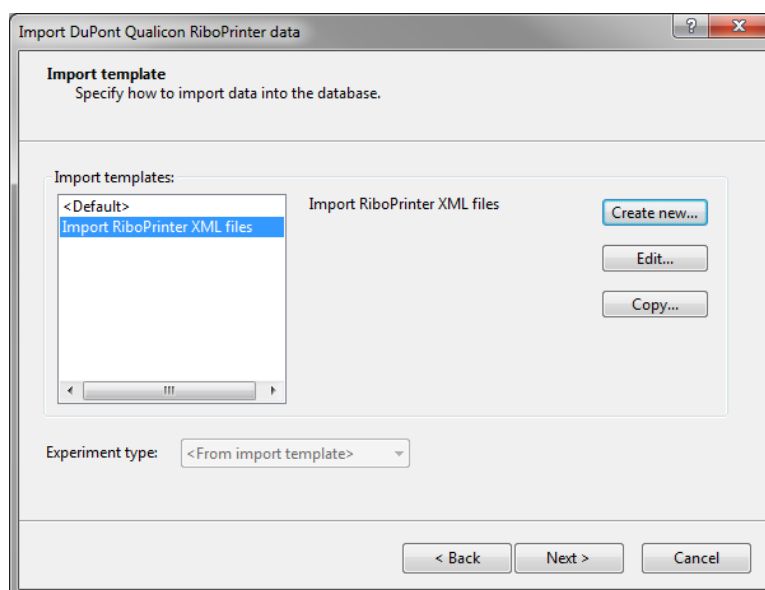


Figure 2.10: Import template added to the list.

If a row in the grid is linked to the **Fingerprint type name** destination, the text **From import template** is automatically selected in the **Experiment type** text box. The import tool will link the patterns to the

corresponding fingerprint type. If one or more fingerprint types are not present in the database, the next step will ask for the confirmation of these fingerprints in the database (see Figure 2.11).

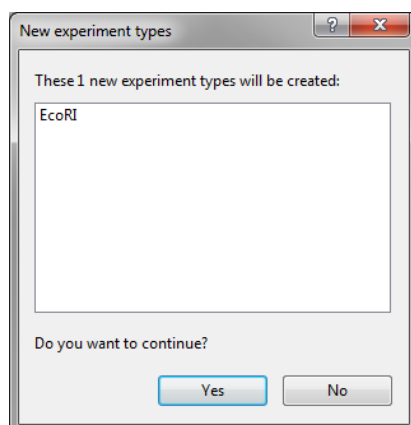


Figure 2.11: Missing fingerprint(s) in the database.

If no row entry in the grid is linked to a **Fingerprint type name** destination, the patterns can be linked to an existing fingerprint type experiment or to a new fingerprint type experiment (<**Create New**>). When the fingerprints are linked to a new fingerprint type experiment, the next dialog will prompt for the fingerprint type name (see Figure 2.12).

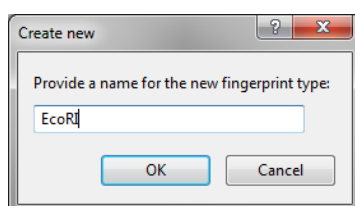


Figure 2.12: Specify a fingerprint type name.

2.10 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.

When mapped **Key** information exceeds the maximum number of allowed characters (i.e. 60 characters), the **Create *x* entries** option will have a red background. The entries with a **Key** exceeding the maximum number of allowed characters will not be created in the database. Double-clicking on the red menu item opens the **Entry key import** dialog box listing all errors.

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

Entries are created/updated and are displayed in the **Database entries** panel of the **Main** window. The linked sample information is stored in the corresponding entry fields (see Figure 2.16). When the option **Select modified entries** was checked, the new/updated entries are marked by a checked ballot box (☑).

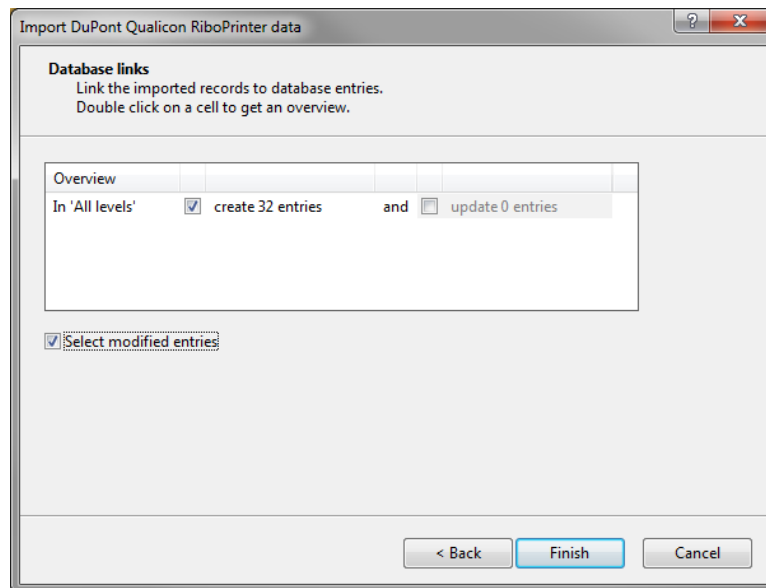


Figure 2.13: Specify the database links.

Each batch results in one separate fingerprint file containing the corresponding fingerprint patterns. The fingerprint files are displayed in the *Fingerprint files* panel. The batch number – composed of the instrument and accession number – is used as file name and the linked batch information is stored in the corresponding fingerprint fields.



Mapped fields, exceeding the maximum number of allowed characters (i.e. 79 characters) will be truncated to 79 characters during import. A message pops up asking the user to confirm the import action (see Figure 2.14).



An error message will pop up when trying to add a file to the database that already exists (see Figure 2.15). The patterns of this batch will not be imported. To import the batch, remove the existing fingerprint file from the *Fingerprint files* panel before importing the data.

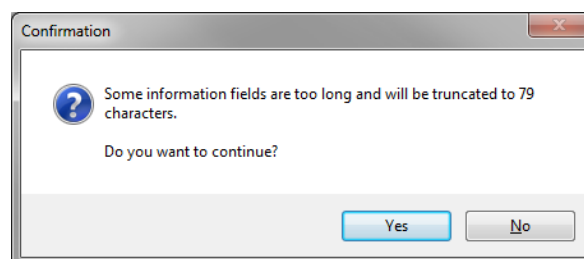


Figure 2.14: Warning message.

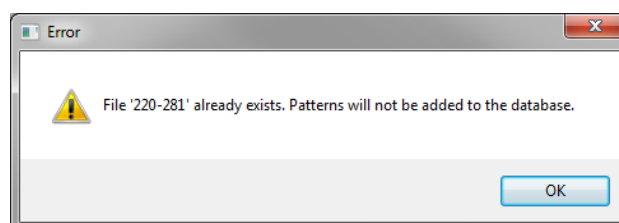


Figure 2.15: Error message.

The patterns are linked to the appointed fingerprint type experiment in the database. The presence of a

pattern for an entry/experiment combination is indicated with a green colored dot in the *Experiment presence* panel.

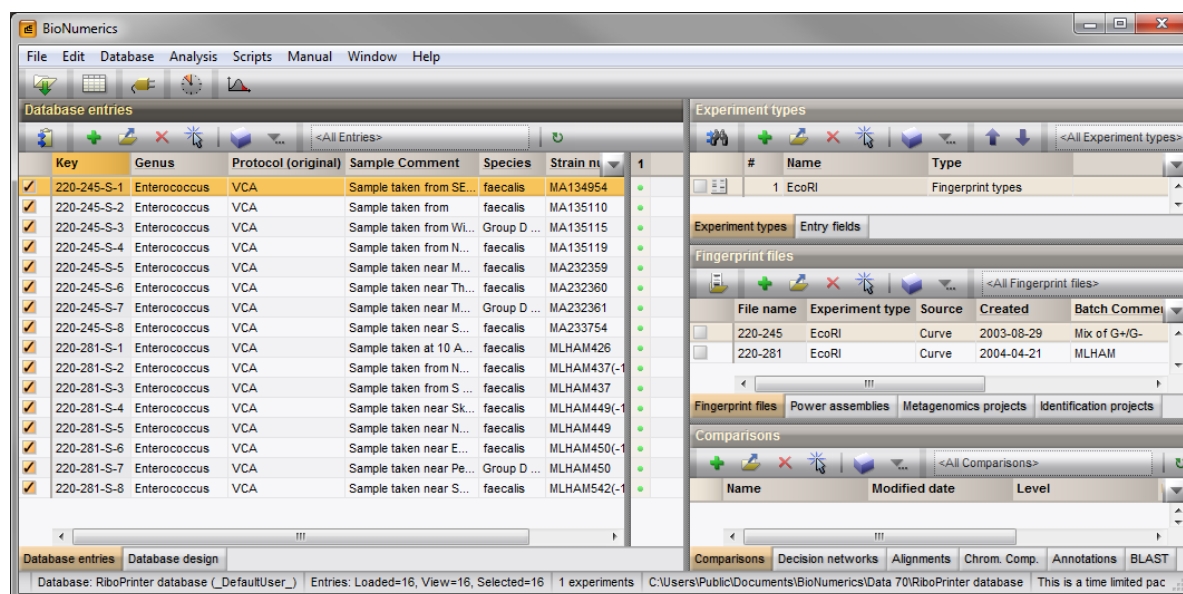


Figure 2.16: The *Main* window after import of RiboPrinter data.

An automatic band detection is performed on the imported patterns using the band search settings of the linked experiment. To view these settings select **Settings > General settings...** (⚙️) in the *Fingerprint type* window (see Figure 2.17). When patterns are linked to a fingerprint experiment that is created during import, a **Minimum profiling** of 1% is used and the option **Rel. to max. val** is unchecked while other band search settings are left at their defaults.

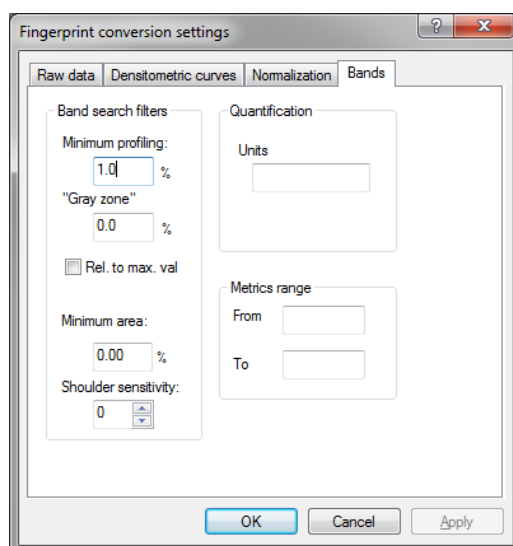


Figure 2.17: The band search settings in the *Fingerprint processing settings* dialog box.

The band assignments can be inspected in the *Fingerprint processing* window (double-click on the file in the *Fingerprint files* panel and choose **File > Edit fingerprint data...** (✎)). Changes to the band assignments can be made here or in the *Comparison* window (see the Reference manual, Chapter Setting up fingerprint type experiments for more information).

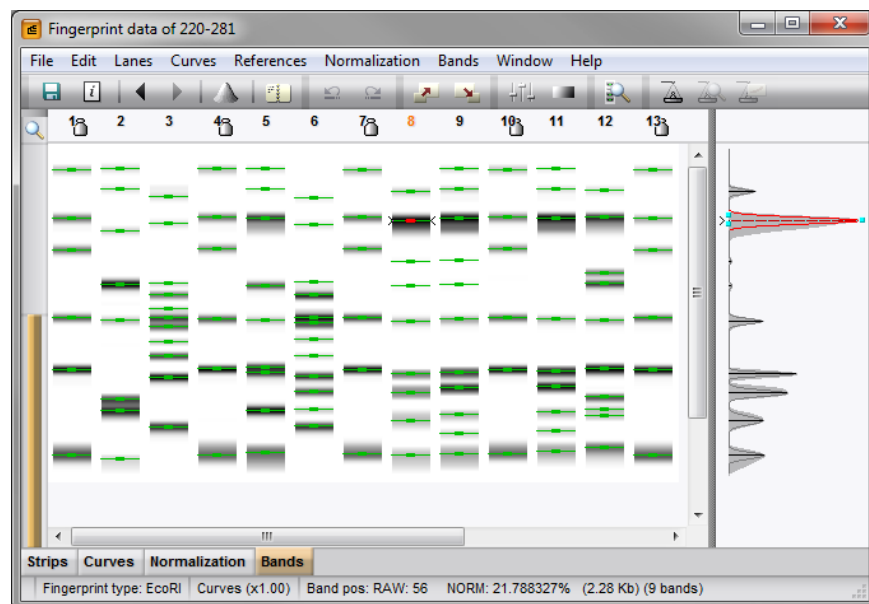


Figure 2.18: The *Fingerprint processing* window with the *Bands* tab selected.

Chapter 3

Analyzing RiboPrinter patterns

Some useful features in the context of RiboPrinter 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, Part Fingerprint types.


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.

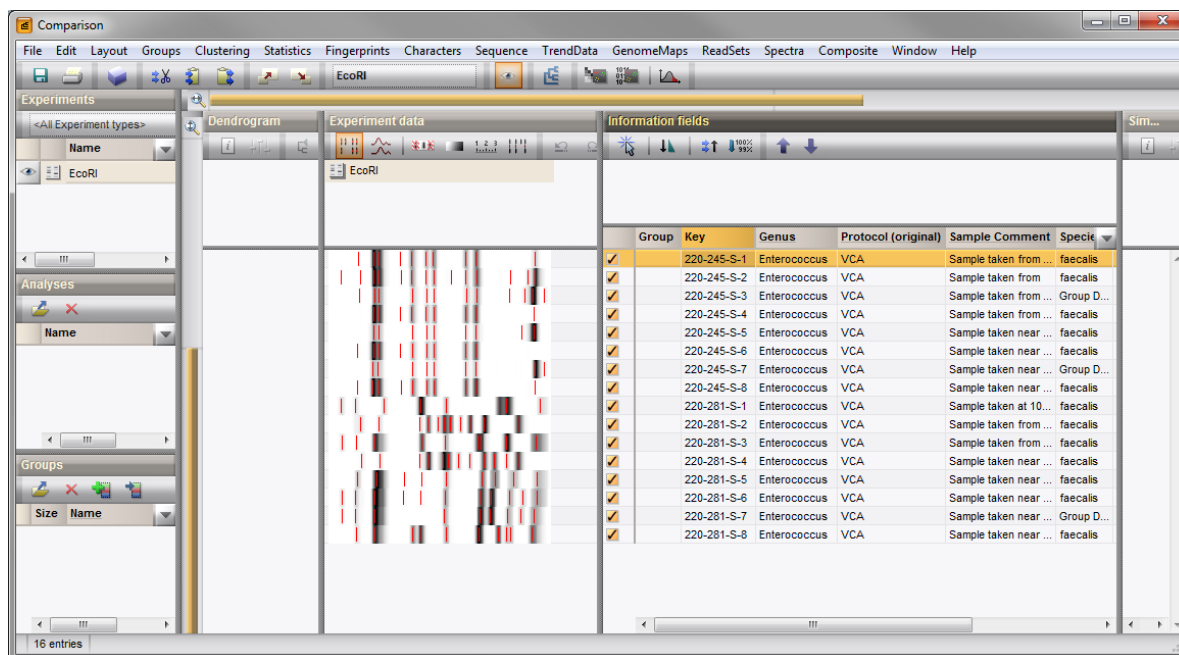


Figure 3.1: The *Comparison* window.

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 (see Figure 3.2 for an example). Detailed information about the *Comparison* window can be found in the Reference manual, Chapter Comparisons in BioNumerics.

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.

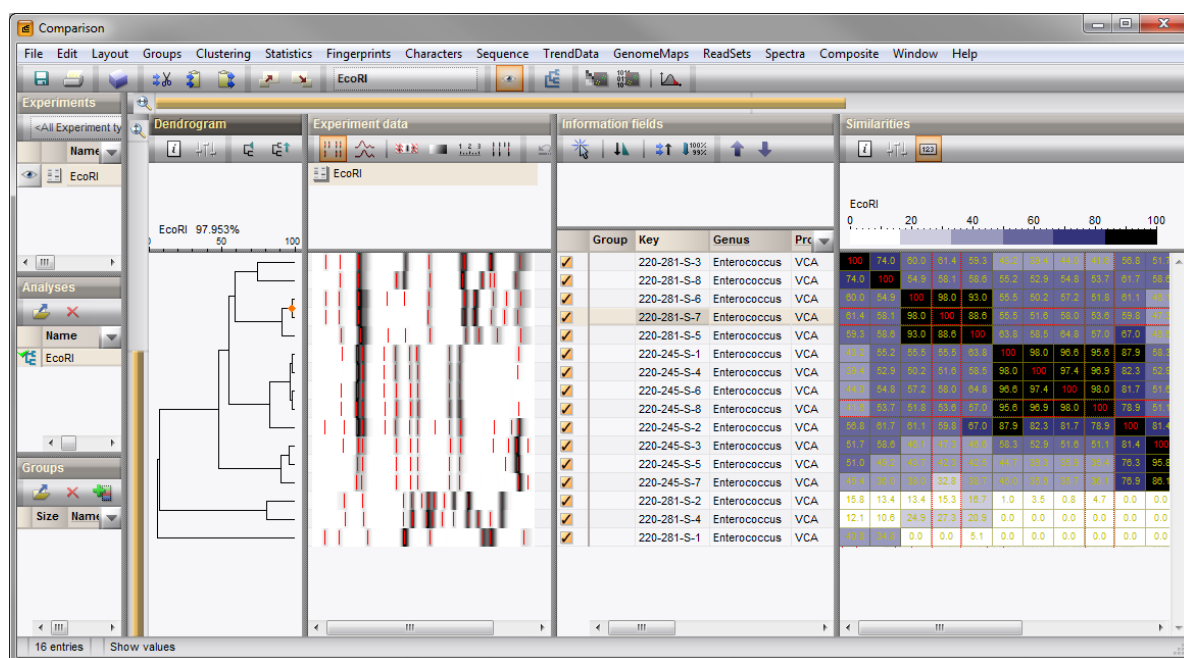


Figure 3.2: The *Comparison* window with dendrogram, image, entry names and similarity matrix.



A B I O M É R I E U X C O M P A N Y

Copyright 1998-2018, Applied Maths NV. All rights reserved.

Please contact us for any additional information you might require, we will gladly help you!

Headquarters

📍 Keistraat 120 • 9830 Sint-Martens-Latem • Belgium
☎ +32 922 22 100 ✉ info@applied-maths.com

USA and Canada

📍 11940 Jollyville Rd., Suite 115N • Austin, TX 78750 USA
☎ +1 512 482 9700 ✉ info-us@applied-maths.com