BioNumerics®

MICROBIAL DATA ANALYSIS SOFTWARE

MLPA plugin

PLUGINS

VERSION 7.6

www.applied-maths.com

Contents

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.

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

This guide is designed as a tutorial for the *MLPA plugin* of BioNumerics. This plugin facilitates data analysis for Multiplex Ligation-dependent Probe Amplification (MLPA) [1] and related techniques, such as Multiplex Amplicon Quantification (MAQ), in BioNumerics. The minimal configuration for the installation of the *MLPA plugin* includes the Fingerprint data module (import and pre-processing of fingerprints) and the Character data module (storage of quantification values and scores).

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***>.

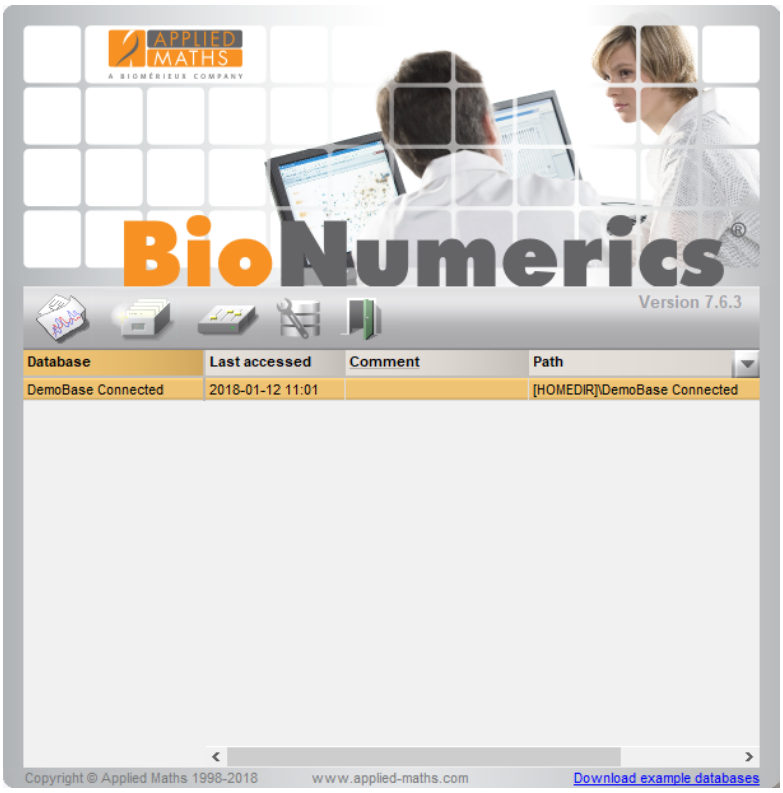


Figure 1.1: The *BioNumerics Startup* window.

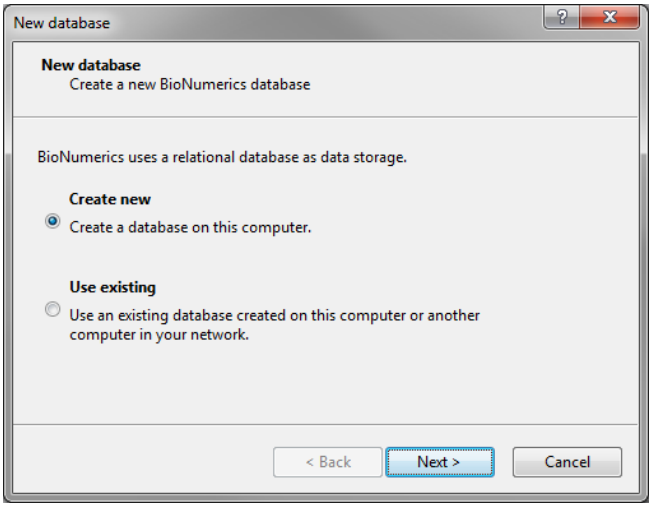


Figure 1.2: The *New database* wizard page.

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

The *Plugins* dialog box appears.

1.4 Installing the MLPA 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*

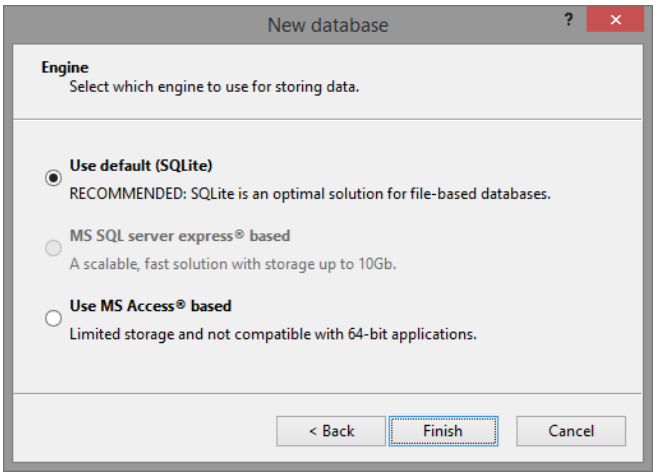


Figure 1.3: The *Database engine* wizard page.

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

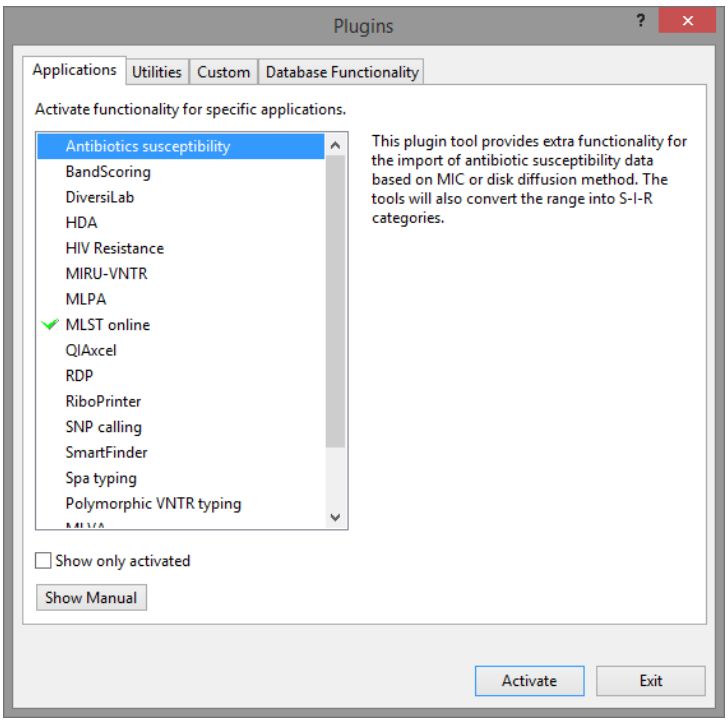


Figure 1.4: The *Plugins* dialog box.

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 *MLPA* from the list in the *Applications tab* and press the <***Activate***> button.

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

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

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

The *MLPA plugin* installs extra buttons and menu items in the *Main* window, *Comparison* window and *Fingerprint* window of the software (see Figure 1.5).

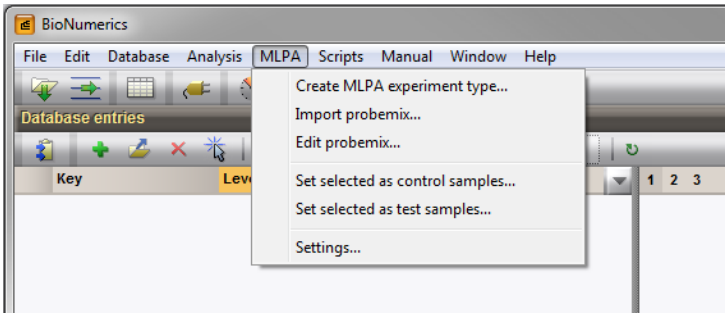


Figure 1.5: MLPA menu in the *Main* window.

In addition, a fingerprint experiment type **MLPA_**, which can be used as base experiment type for the actual MLPA fingerprint types, and two auxiliary character experiment types **MLPA_ctrls** and **MLPA_stats** are created after installation of the *MLPA plugin*. Finally, two experiment information fields 'MLPA probemix' and 'MLPA description' are created, which allow for easy sorting of the experiment types. These fields can be edited in the *MLPA probes and fragments* dialog box (see 3.3).

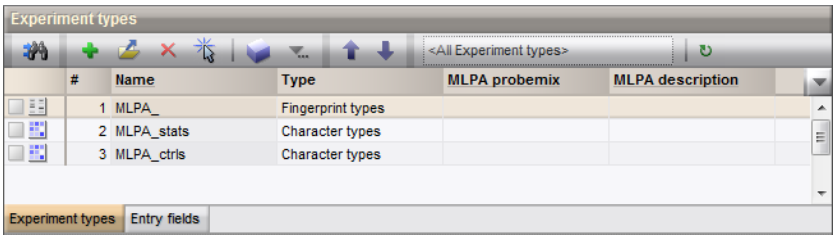


Figure 1.6: MLPA experiment types and information fields.

The two auxiliary character types are intended for internal use by the software only and should not be modified or deleted by the user. **MLPA_ctrls** is an open character set with binary values, of which the characters are the MLPA fingerprint types. It is used to keep track of entries that are control samples for probe mixes. **MLPA_stats** is a character set used to store statistical results for all probes and probe mixes.

Chapter 2

MLPA data import in BioNumerics

2.1 Introduction

MLPA amplification products are typically analyzed via a capillary electrophoresis-type automated sequencer.

When fingerprints are run on a capillary sequencer, the resulting data can have two fundamentally different formats:

- **Curve files** (also referred to as electropherograms, chromatogram files or trace files): The binary encoded raw data as produced by the capillary electrophoresis equipment. Since it is the raw data, it contains the complete information but some fingerprint preprocessing (e.g. normalization, band assignment, ...) in BioNumerics is required.
- **Peak tables**: Text files containing a listing of peaks from the chromatograms, with their corresponding metrics (sizes in base pairs) and peak height and/or peak area. This type of data has been processed by the software which controls the capillary electrophoresis equipment.

In contrast to the `.fsa` files generated by Applied Biosystems sequencers, the `.scf` raw curve files generated by Beckman-Coulter equipment are *not* corrected for spectral overlap of the fluorescent dyes. This is generally not a problem when a single data channel is used. However, in experimental setups where all four dyes are employed (i.e. three data channels and one reference channel), it is advised to import `.crv` files in BioNumerics. The latter files can be exported from the Beckman-Coulter software and contain curves which are corrected for cross talk.

For more information about the import of fingerprint data from capillary electrophoresis equipment, see the Reference manual, Chapter Setting up fingerprint type experiments.

When importing peak tables directly into the database (not creating a synthetic gel file), the peaks will be read-only. This means that no editing is possible in the *Comparison* window. For MLPA analysis, make sure that the option ***Create synthetic profiles*** in the *Import peak table dialog box* is checked.

2.2 MLPA example data set

The example data that will be used to illustrate the functionality of the *MLPA plugin* are publicly available sample data from the MLPA validation study by EuroGenTest (http://www.eurogentest.org/web/info/public/unit1/MLPAvalidation.xhtml). They represent one run of Applied Biosystems se-

quencer curve files, containing information of 16 different patient samples and one blank, tested for the Duchenne Muscular Dystrophy (DMD) gene by the P034 MLPA kit from MRC-Holland.

The example data set contains a 'no template control' (BLANCO), two negative controls (samples D1.03.02235 DMD and D1.03.02609 DMD), for which all exons score as 'normal' and two positive controls (D1.03.01831 DMD and D1.04.04280 DMD) that have following known aberrations:

- D1.03.01831 DMD: heterozygous gain of exon 2 to 10.
- D1.04.04280 DMD: heterozygous loss of exon 8 to 10.

The MLPA example data, including probe definitions, can be downloaded from the Applied Maths website (`http://www.applied-maths.com/download/sample-data`, click on "MLPA sample data").

2.3 MLPA base fingerprint type settings

During installation of the *MLPA plugin*, a base fingerprint type experiment called **MLPA_** is created. This experiment will contain the default settings for import and processing of the dyes, which will be copied to all fingerprint types that are created when sequencer fingerprint data is imported.

The **MLPA_** settings can be changed after installation of the plugin in the *Fingerprint conversion settings dialog box*.

3.1 In the *Main* window, double-click on the experiment **MLPA_** in the *Experiment types* panel.

3.2 In the *Fingerprint type* window, select ***Settings*** > ***General settings...*** ( ).

3.3 For the example data, change the ***Data OD range*** in the *Raw data tab* to "4096" (= 12-bit). Leave all other settings unaltered, press <***OK***> to accept the settings and close the dialog boxes.

2.4 Import example

The necessary steps to import the example MLPA data will be given below. For a more detailed description of the import of capillary sequencer fingerprints we refer to the Reference manual, Chapter Setting up fingerprint type experiments.

4.1 Select ***File*** > ***Import...*** (, **Ctrl+I**) to call the *Import* dialog box, choose ***Import curves*** under ***Fingerprint type data*** and press <***Import***>.

4.2 Browse to the downloaded and unzipped example MLPA data folder and select all Applied Biosystems curves files (extension `.fsa`) found in `MLPA data \DMDFiles \DMD-P034-1`. Press <***Open***>.

The files are displayed in the *Input* wizard page and the default suggested ***Fingerprint file name*** is the folder name, ***DMD-P034-1*** (see Figure 2.1).

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

The way the information should be imported in the database can be specified with an import template. In the example data set, the dye name can be parsed from the file information and the unique sample information can be parsed from the file names. A new import template needs to be defined:

4.4 Press the ***Create new*** button to call the *Import rules* dialog box (see Figure 2.2).

The *Import rules* dialog box (see Figure 2.2) lists the information present in the selected files as **Source**, their linked **Source type** and the **Destination** component they are associated with (currently all set to <None>).

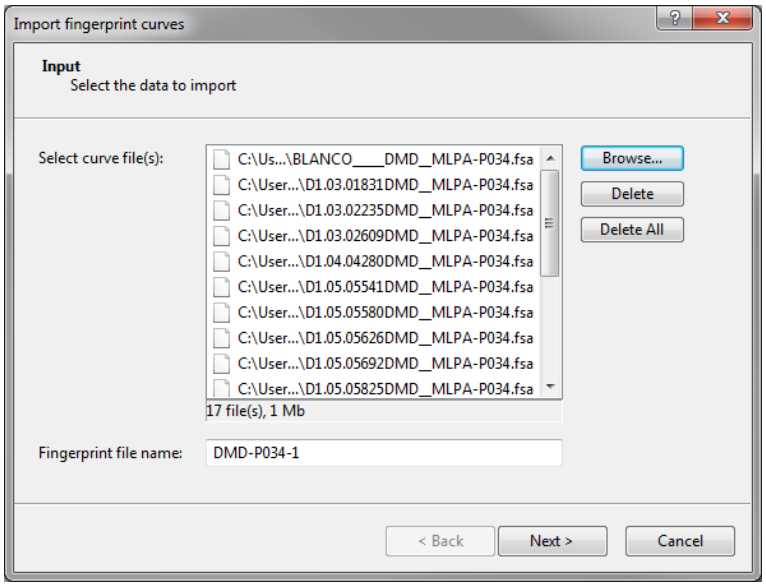


Figure 2.1: Selected curve files.

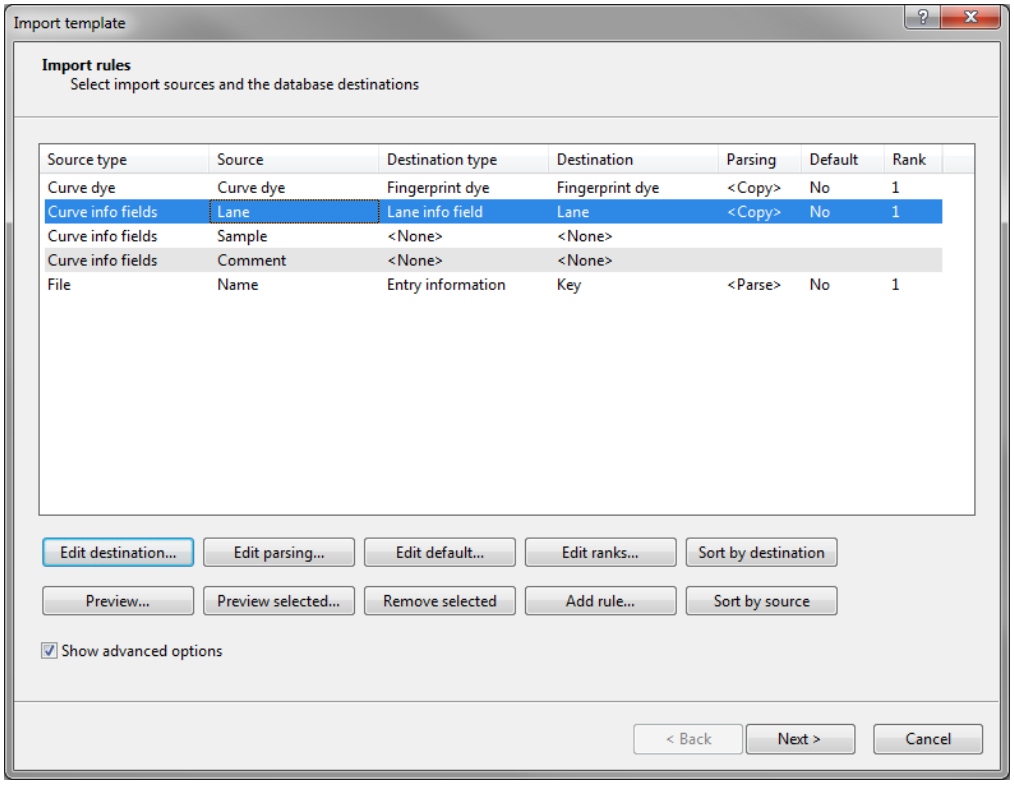


Figure 2.2: Import rules.

4.5 Select ***Curve dye*** from the list, select <***Edit destination***> and select ***Fingerprint dye*** as corresponding field. Press <***OK***>.

Next, we will specify a new rule that links the part of the file name appearing before the underscore to the ***Key*** field.

4.6 Select ***Name*** from the list, select <***Edit destination***> and select ***Key*** as corresponding field. Press <***OK***>.

4.7 Check the option ***Show advanced options***, make sure the last row is selected in the grid panel and press the <***Edit parsing***> button.

4.8 In the *Data parsing* dialog box, fill in following data parsing string: "[DATA]_*". The asterisk will serve as wildcard.

4.9 Press the <***Preview***> button and press <***OK***> when the parsing is correct (see Figure 2.3).

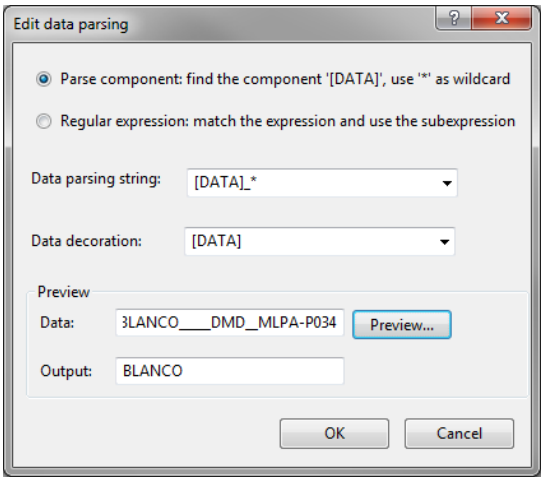


Figure 2.3: Data parsing string.

In the example files, "Lane" information is available. We will import this information and store it as fingerprint information.

4.10 Select ***Lane*** from the list and select <***Edit destination***>. Select ***Create new*** under ***Lane info field*** and press <***OK***> twice and confirm the creation of the new field.

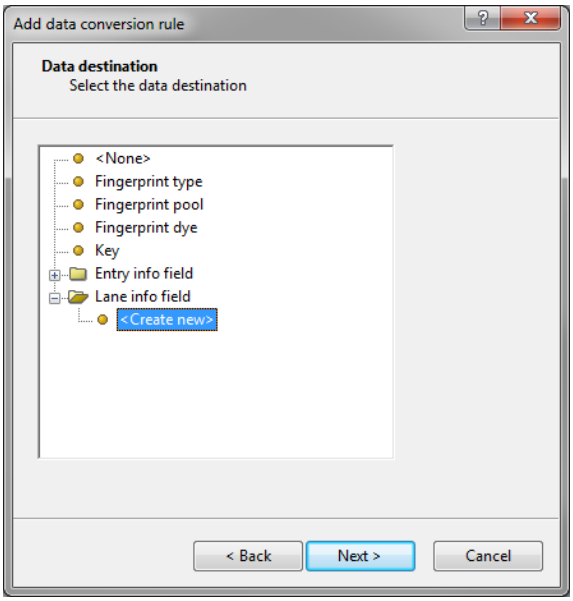


Figure 2.4: Link to lane information field.

The *Import rules* dialog box should now look like Figure 2.2.

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

In the example data set, the Rox channel contains the size standard (GeneScan 500 ROX) and the Joe channel represents the actual MLPA peaks, that correspond to the probe mix P034.

4.12 Make sure ***Rox*** is selected as reference dye and uncheck the dyes ***Fam*** and ***Tamra*** to prevent the import of these channels (see Figure 2.5).

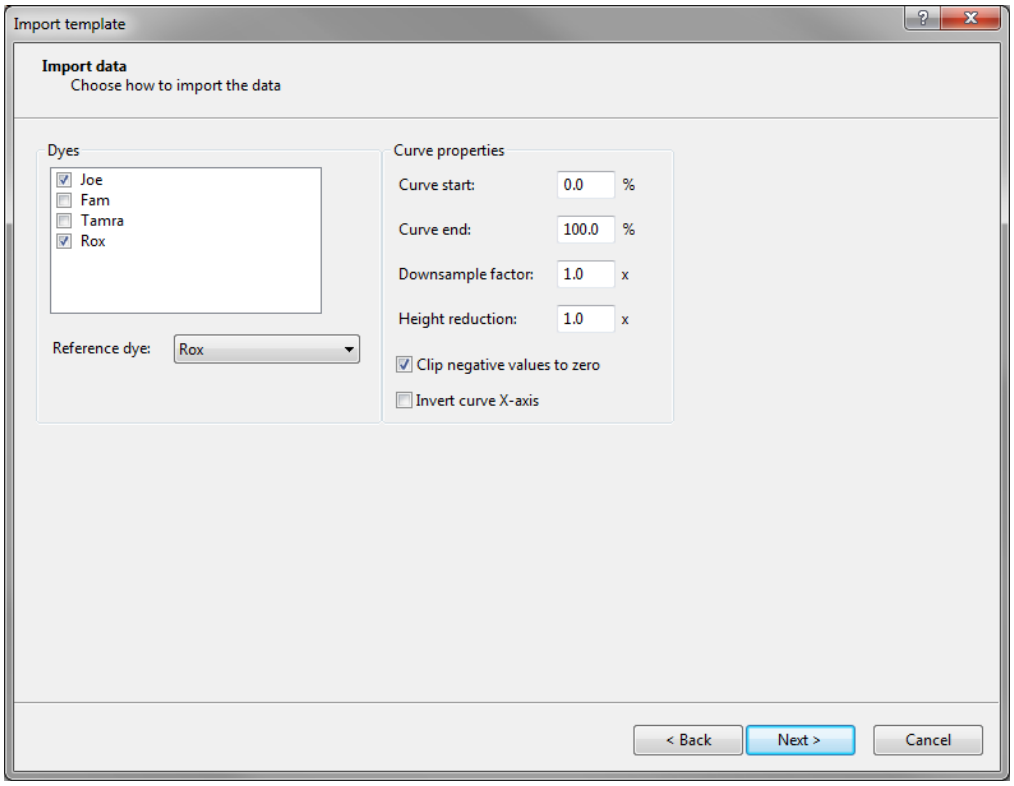


Figure 2.5: Specify how to import the data.

4.13 Press <***Next***> and <***Finish***>.

4.14 Specify a template name, e.g. **Import MLPA data** and press <***OK***>.

4.15 Make sure the newly created template is selected and press the <***Preview***> button.

The preview should now look like Figure 2.6.

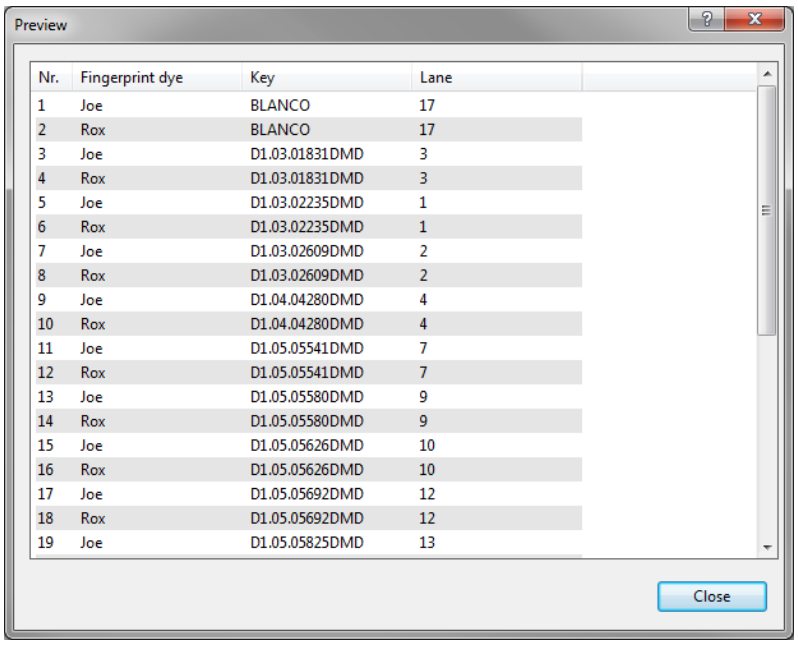


Figure 2.6: Preview of the parsing.

4.16 Close the preview.

4.17 Make sure the **MLPA_** experiment and created template are selected and press <***Next***>.

A fingerprint type needs to be present in the database for each pool (if present) and dye combination. The names of these fingerprint types are composed of the base fingerprint type name, followed by the pool name (if present), and the name of the dye (e.g. BaseFpr_ROX). If one or more of these fingerprint types are not present in the database, a new dialog box pops up, listing all missing fingerprint types (see Figure 2.7).

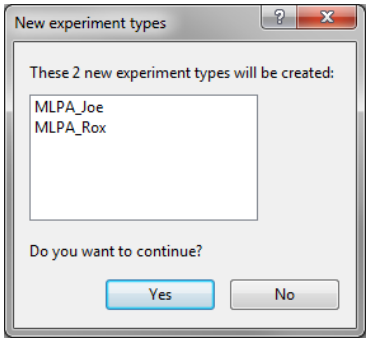


Figure 2.7: Confirm creation of experiments.

4.18 Confirm the creation of the two missing fingerprint type experiments.

4.19 Press <***Next***> to confirm the creation of 17 new entries (see Figure 2.8).

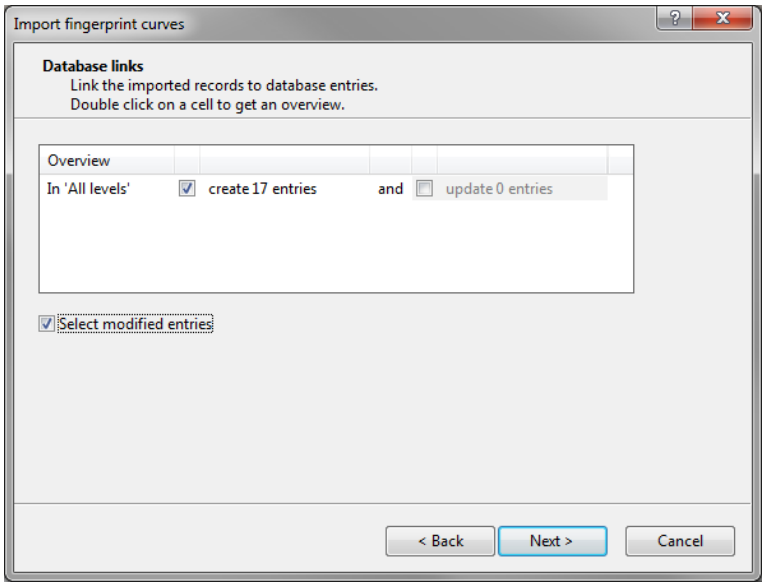


Figure 2.8: Database links.

4.20 Make sure ***Open curve preprocessing window*** is checked in the last step and press <***Finish***>.

BioNumerics splits the curve files into separate fingerprint files per imported dye. Each batch results in one fingerprint file per imported dye. The fingerprint file is composed of the file name plus a suffix referring to the dye name (e.g. "DMD-P034-1_Rox"). After data import, the *Main* window looks as in Figure 2.9.

The *Fingerprint curve processing* window opens. The two channels from the run are automatically loaded and displayed in the *Fingerprint curve processing* window.

4.21 Click on the icon left of the **DMD-P034-1_Joe** channel in the *Channels* panel.

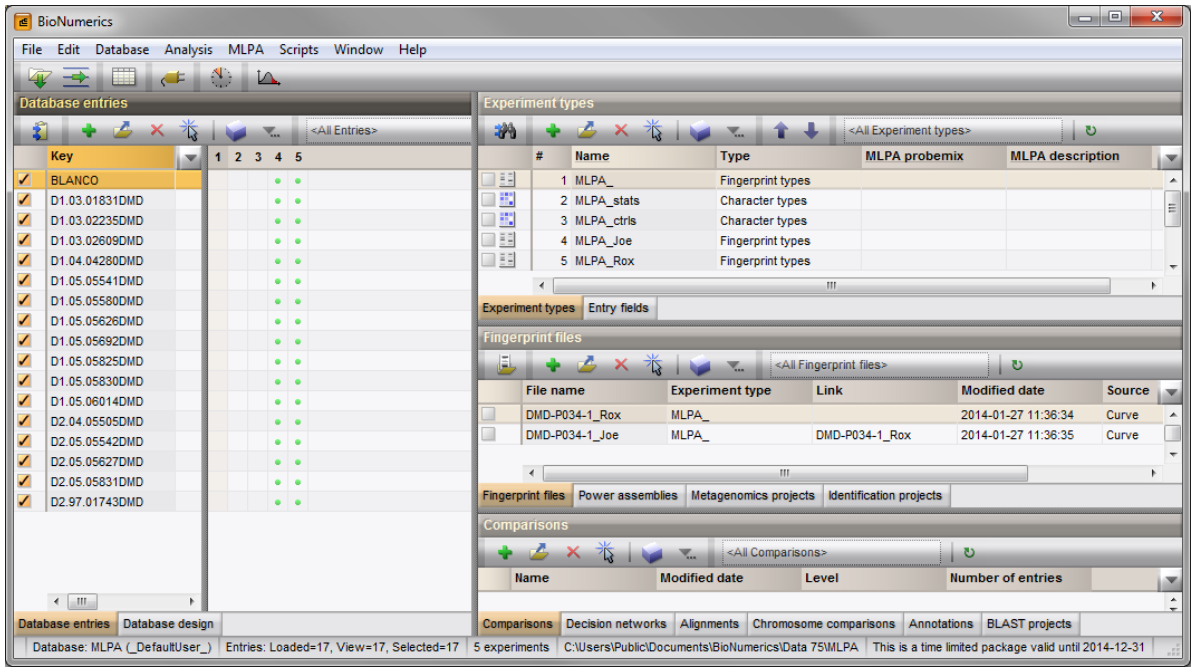


Figure 2.9: The *Main* window after import of the example MLPA data.

The data channel is now hidden from the view and its icon is displayed as (see Figure 2.10).

4.22 Use the zoom sliders on the left and on top to optimize the display of the fingerprint curves.

Since the raw chromatogram files have not undergone any preprocessing, normalization will have to be performed. This requires a *reference system* to be defined, based upon the marker peaks available in the reference dye.

4.23 Make sure the reference dye is the only dye visible in the upper panel.

4.24 Select ***Bands*** > ***Search reference bands...*** (, **Ctrl+F**) to call the *Search reference bands* dialog box.

4.25 Specify a peak detection ***OD range*** of **2** (in %) and a peak detection ***Curve range*** of **1** (in %). Press <***OK***>.

The bands that fall within the specified criteria are marked with a solid line at the band's position (see Figure 2.10).

4.26 To have a reference system automatically created based on a lane containing commercial size marker, first highlight a suitable lane and then select ***References*** > ***Define size standard...***.

This will display the *Size standard* dialog box, from which a size marker can be selected (see Figure 2.11).

In the example curve files, the GeneScan 500 ROX size standard is used as reference, containing 16 bands with known molecular weight.

4.27 Select **GeneScan 500 ROX** from the list, select **Pattern match** and press <***OK***> twice.

4.28 Save the data to the database with ***File*** > ***Save*** (, **Ctrl+S**).

The software will automatically create the reference system and calibration curve for each of the fingerprint types. Since this allows the calculation of metrics information, a metrics scale now becomes available in the upper part of the *Fingerprints* panel.

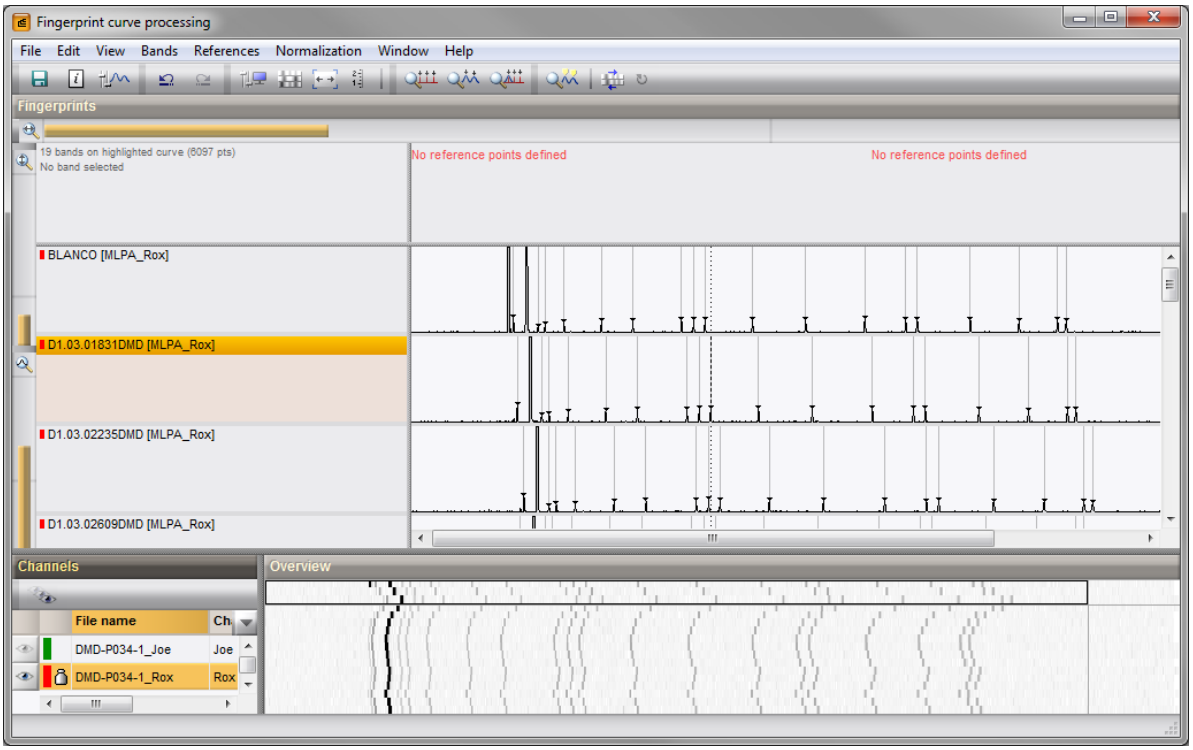


Figure 2.10: The *Fingerprint curve processing* window only displaying the reference dye.

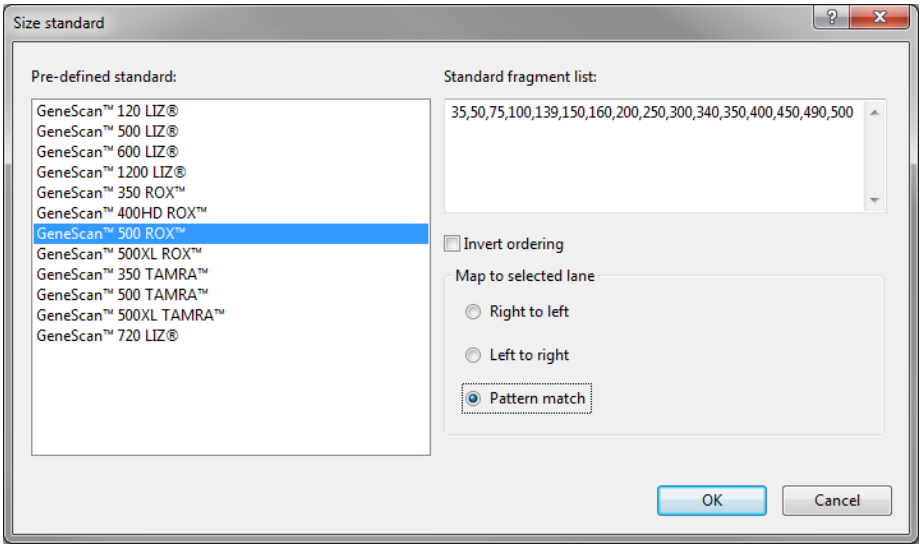


Figure 2.11: Choose a size standard from the list.

Normalization is achieved by assigning bands in the reference channel to external reference positions.

4.29 To normalize a complete run at once, select ***Normalization*** > ***Auto assign reference positions (all lanes)...*** (, **Ctrl+A**), leave all settings unaltered and press <***OK***>.

4.30 When the assignment of the marker bands to reference positions is made, the data can be shown in normalized mode with ***Normalization*** > ***Show normalized view*** (, **Shift+N**).

4.31 Click on the icon left of the **Rox** and **Joe** channels in the *Channels* panel.

The data channel is now shown and the reference channel is hidden from the view.

4.32 Select ***Bands*** > ***Search data bands...*** (, **Ctrl+Shift+F**) to call the *Search data bands* dialog box.

4.33 Specify a peak detection ***OD range*** of **2** (in %) and a peak detection ***Curve range*** of **1** (in %). Press <***OK***>.

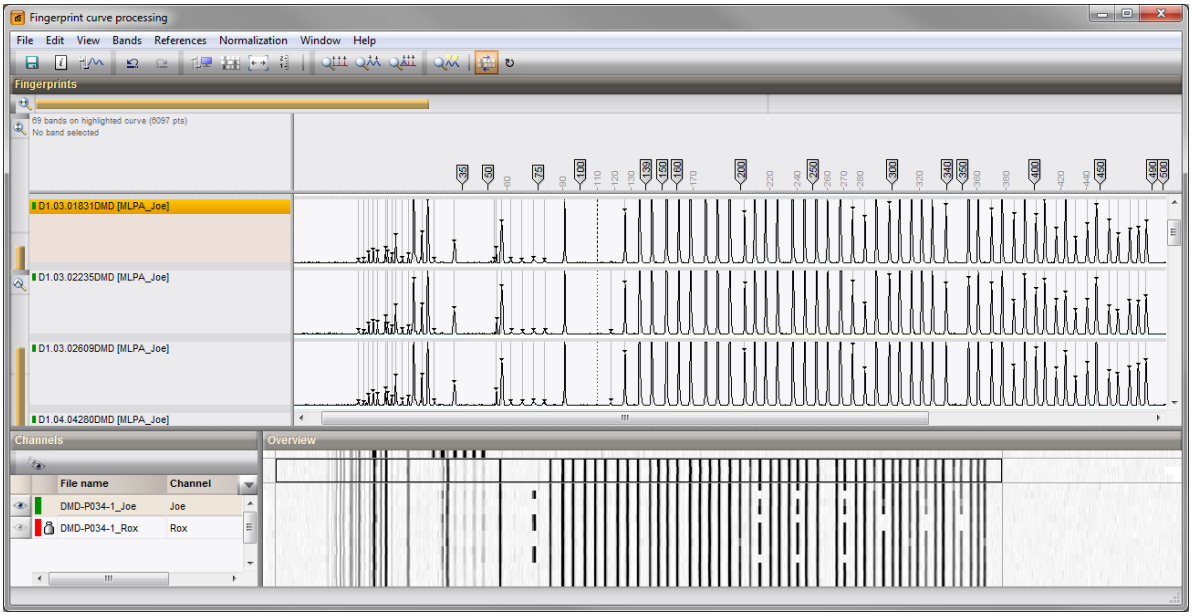


Figure 2.12: Normalized view - data channel.

4.34 Save the changes and close the *Fingerprint curve processing* window.

2.5 Calibration curves

When saving the data to the database with ***File*** > ***Save*** (, **Ctrl+S**) in the *Fingerprint curve processing* window, the software will automatically create the reference system and calibration curve for each of the fingerprint types.

5.1 Double-click on the base fingerprint type (**MLPA_**) in the *Experiments* panel to open the *Fingerprint type* window.

5.2 In the *Fingerprint type* window, call ***Settings*** > ***Edit reference system***, or double-click in the *R01 panel* to call the *Reference system window*.

A calibration curve for the reference system **R01** of the base fingerprint type **MLPA_** is displayed.

5.3 Close the *Reference system window* and the *Fingerprint type* window.

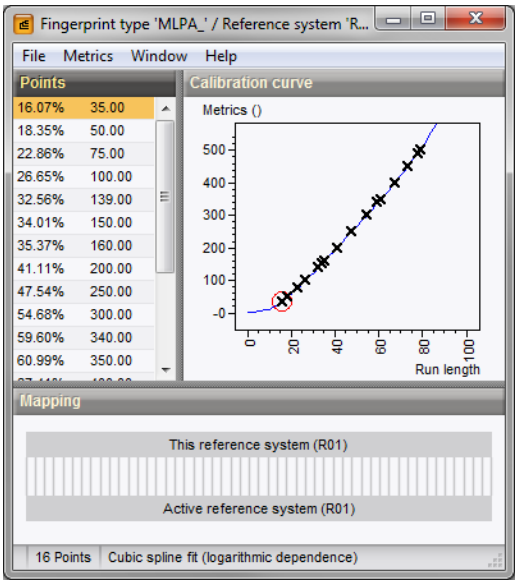


Figure 2.13: The *Reference system window*, with a calibration curve displayed.

Chapter 3

The MLPA experiment type

3.1 Work flow for analyzing MLPA data

In BioNumerics, an MLPA experiment type consists of a *fingerprint type* plus two auxiliary (or "helper") *character types*. A *fingerprint* is thereby defined as a densitometric record, seen as a one-dimensional profile of peaks or bands. Examples are gel and capillary electrophoresis patterns. A *character* is a certain feature of a sample that can be assigned a value. This value can be binary, multi-state or numerical.

Since MLPA data in principle consist of a number of runs on an automated sequencer, they are imported in BioNumerics as fingerprint type data (see 2). The imported format can be the native binary format as produced by the capillary electrophoresis equipment (raw curve files or electropherograms) or can be data that is preprocessed by the equipment's software (peak tables in text files). Before curves can be further analyzed, they need to be *normalized* against an external reference system. Since a size calling has already been performed before data input, this step can be omitted for imported peak tables.

An MLPA experiment type is characterized by the *probe mix* used (see 3.3). Probe mix definitions contain probe names, probe lengths and some additional probe information. Based on the probe mix definitions, *band classes* are created and stored with the fingerprint. To identify a peak in an MLPA profile as being amplified from a certain probe, the peak is matched to a band class corresponding to a probe (see 4.2). This process is called *band matching* in BioNumerics. For more information about band matching, see the Reference manual, Chapter Band matching and polymorphism analysis.

The two auxiliary character types in an MLPA experiment type are called "*FPR*_quant" and "*FPR*_score", with "*FPR*" the name of the MLPA fingerprint type they correspond to. "*FPR*_quant", further referred to as the quantification character type, is used to store the relative quantification values of all probes in the probe mix. The relative quantifications are calculated on normalized peak sizes of matched peaks. "*FPR*_score", further referred to as the scores character type, is used to store the scores (e.g. heterozygous loss, duplication, etc.) for each of the probes. The automated scoring is based on the values stored in the quantification character type, but can be manually overruled.

The auxiliary character types are required by the software to store information in. There will never be any need to manually alter information stored in these character sets. In fact, any manual change to the characters or character values could result in malfunction of the *MLPA plugin*.

A schematic overview of the MLPA analysis work flow is given in Figure 3.1.

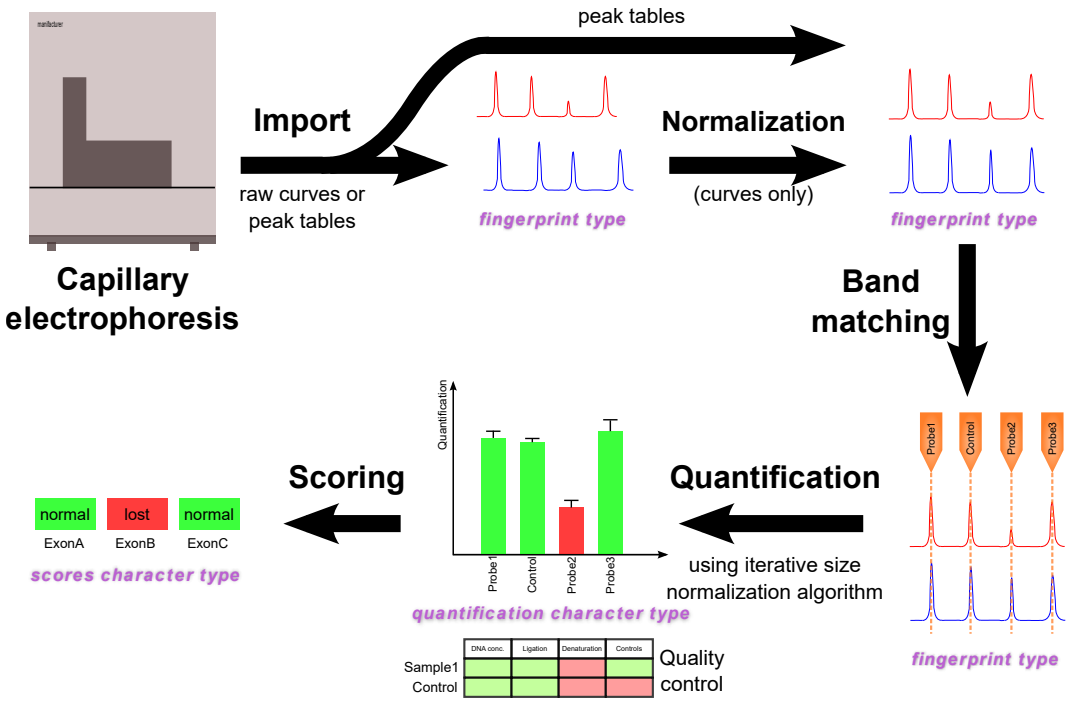


Figure 3.1: Work flow for MLPA analysis in BioNumerics. The experiment type in which the corresponding information is saved, is indicated in purple.

3.2 Creating a new MLPA experiment type

Before an MLPA analysis can be started, an MLPA experiment type needs to be created first:

2.1 Select ***MLPA*** > ***Create MLPA experiment type***.

This pops up the *Create MLPA experiment type* dialog box, prompting for the source fingerprint type (see Figure 3.2).

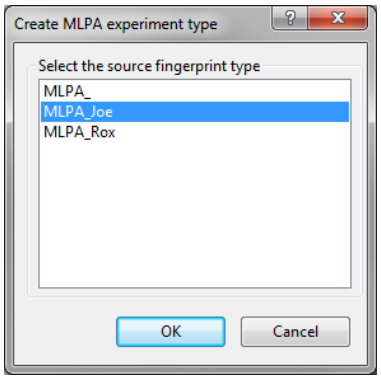


Figure 3.2: The *Create MLPA experiment type* dialog box.

Select the MLPA experiment containing the MLPA peak data. Two additional character type experiments will be created.

2.2 In this example, highlight **MLPA_Joe** from the list (since this is the fingerprint type that actually contains the MLPA peaks; see 2.4) and press <***OK***>.

Two additional character types are created: **MLPA_Joe_score** and **MLPA_Joe_quant**. These character types will be used to store the scores and the relative peak quantifications, respectively. Initially, the characters

types are empty.

3.3 MLPA probe mixes

The *MLPA plugin* provides functionality for creating, importing and editing probe mix definitions. For import, the expected file format is that from the probe mix definition files of MRC-Holland, as available from their website (http://www.mrc-holland.com).

3.1 Select ***MLPA*** > ***Import probe mix***.

This opens the *Import MLPA probemix* dialog box (see Figure 3.3).

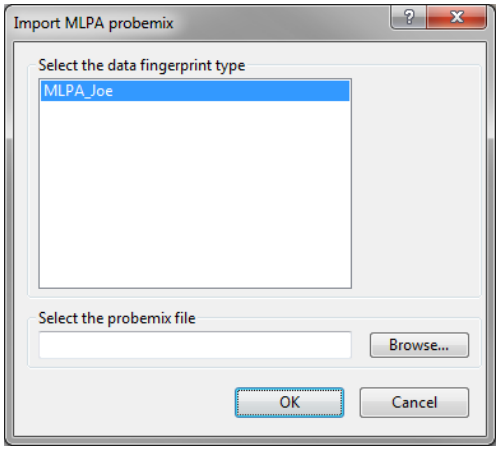


Figure 3.3: The *Import MLPA probemix* dialog box.

Under ***Select the data fingerprint type***, all fingerprint types that belong to a MLPA experiment are listed (currently just one).

By pressing <***Browse***>, a MLPA probe mix file can be selected.

3.2 Press <***Browse***> and navigate to the downloaded `MLPA data files \MLPA data \MLPA probes` folder.

3.3 Select the file `P034 MLPA probemix lot 0707,0906.txt` and press <***Open***>.

3.4 Back in the *Import MLPA probe mix dialog box*, press <***OK***> to import the P034 probe mix file.

The *MLPA probes and fragments* dialog box appears (see Figure 3.4), which consists of three tabs: the *Probes*, *QC fragments* and *Probe mix info tab*.

Probes tab:

The *Probes tab* lists all the probes in the mix, with their 'Name', 'Type' (test or control probe), 'Length' (nominal length in bases), 'Actual length' (observed length in bases), and any additional (present in the probe definition file or user-defined) information fields. Any content of information fields in this list, except in the 'Name' and 'Type' fields, can be edited by clicking twice and typing directly into the field.

A probe can be added to the list by pressing <***Add probe***>. This pops up the *Add* dialog box (see Figure 3.5).

The *Add* dialog box prompts for the 'Name', 'Type', 'Length' and optionally the 'Actual length' of the probe.

To delete a probe or a selection of probes from the list, highlight the probe(s) and press <***Delete probe(s)***>. The program will ask for confirmation before actually deleting the highlighted probe(s). Selections of

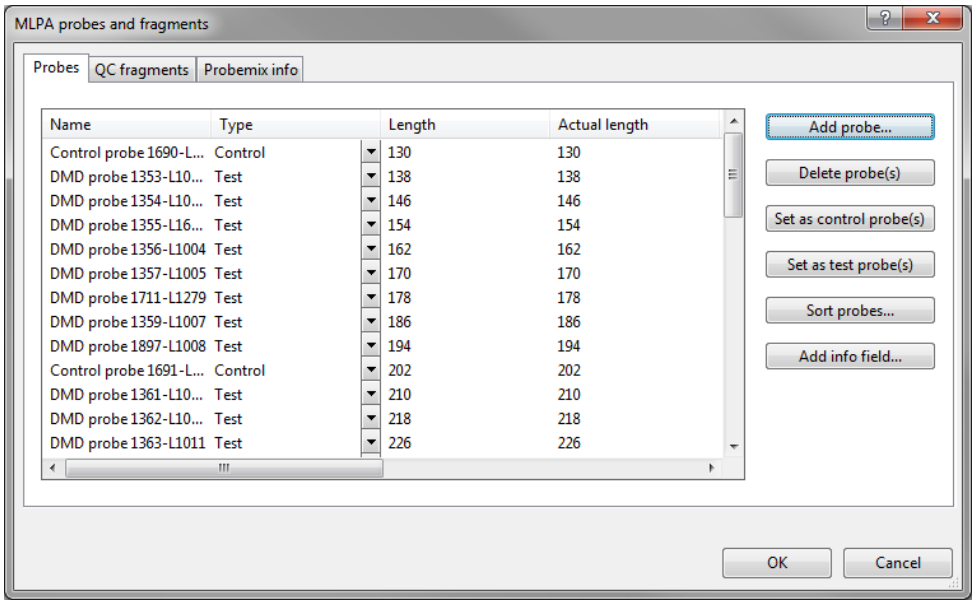


Figure 3.4: The *MLPA probes and fragments* dialog box, with the *Probes tab* displayed.

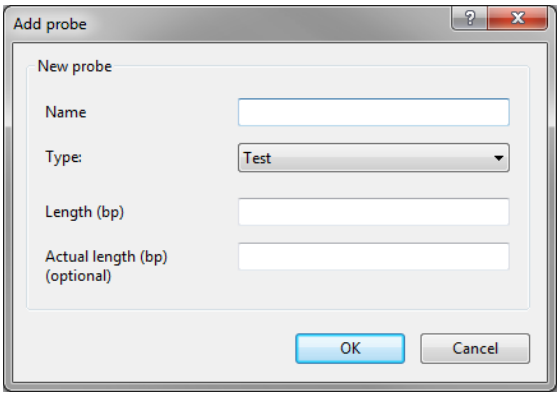


Figure 3.5: The *Add* dialog box: add probe.

individual probes can be made by **Ctrl+clicking** the probes to be selected or a whole range can be selected at once by clicking the first probe and, while holding the **Shift**-key, clicking the last probe in the range.

The type of a probe can be changed by clicking the 'Type' field for that probe twice and selecting "Control" or "Test" from the drop-down list that appears. For selections of probes, use the <***Set as control probe(s)***> or <***Set as test probe(s)***> buttons.

Probes can be sorted according to any of the available information fields. Pressing the <***Sort probes***> button displays the *Sort probes* dialog box (see Figure 3.6).

Here, the information field to sort on and the order (ascending or descending) can be specified. The option ***Sort numerically*** is available for any field that contains numeric information only ('Length' and 'Actual length').

An additional information field can be added by pressing <***Add info field***> (see Figure 3.7).

The *Add* dialog box will prompt for the name of the information field to add. The field will only be added after pressing <***OK***>.

QC fragments tab:

The *QC fragments tab* lists the quality control fragments in the probe mix, with their 'Name', 'Type',

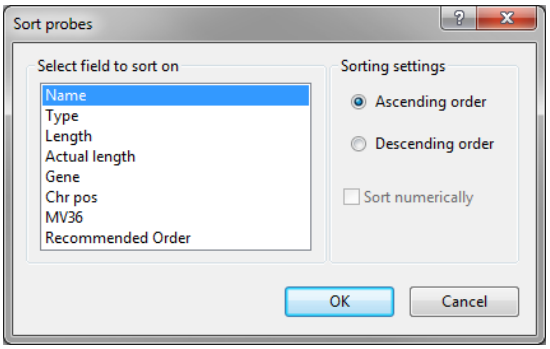


Figure 3.6: The *Sort probes* dialog box.

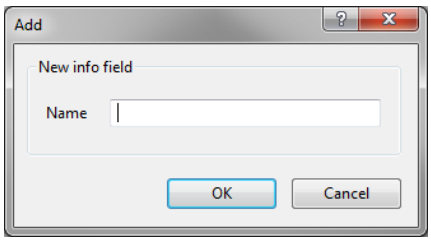


Figure 3.7: The *Add* dialog box: add a new information field.

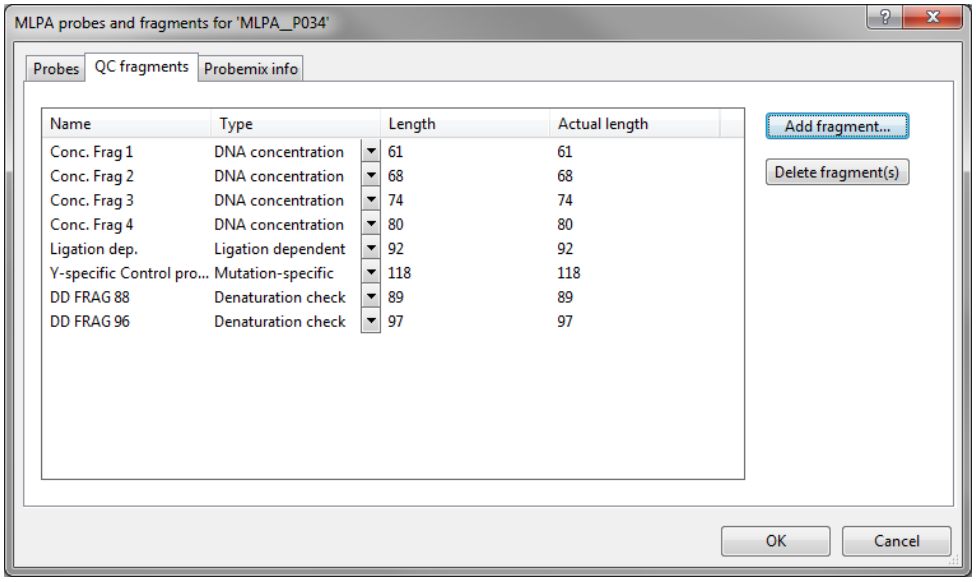


Figure 3.8: The *MLPA probes and fragments* dialog box, with the *QC fragments tab* displayed.

'Length' (nominal length in bases), and 'Actual length' (observed length in bases). The 'Type of quality control fragment' can be either "DNA concentration", "Ligation dependent", "Denaturation check", "Mutation-specific" or "Unknown" and can be set from the drop-down list that appears. The 'Length' and 'Actual length' fields of a quality control fragment can be edited by clicking twice and typing directly into the field.

A quality control fragment can be added to the list by pressing <***Add fragment***>. This pops up the *Add* dialog box (see Figure 3.9).

This window prompts for the ***Name***, ***Type***, ***Length*** and optionally the ***Actual length*** of the quality control

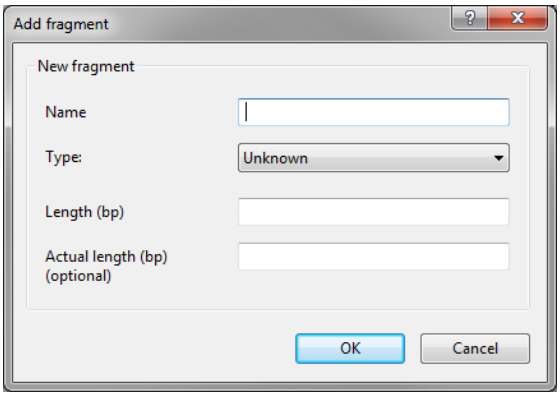


Figure 3.9: The *Add* dialog box to add a quality control fragment to the MLPA probe mix.

fragment.

To delete a fragment or a selection of fragments from the list, highlight the fragment(s) and press <***Delete fragment(s)***>. The program will ask for confirmation before actually deleting the highlighted fragment(s). Selections of individual fragments can be made by **Ctrl+clicking** the fragments to be selected or a whole range can be selected at once by clicking the first fragment and, while holding the **Shift**-key, clicking the last fragment in the range.

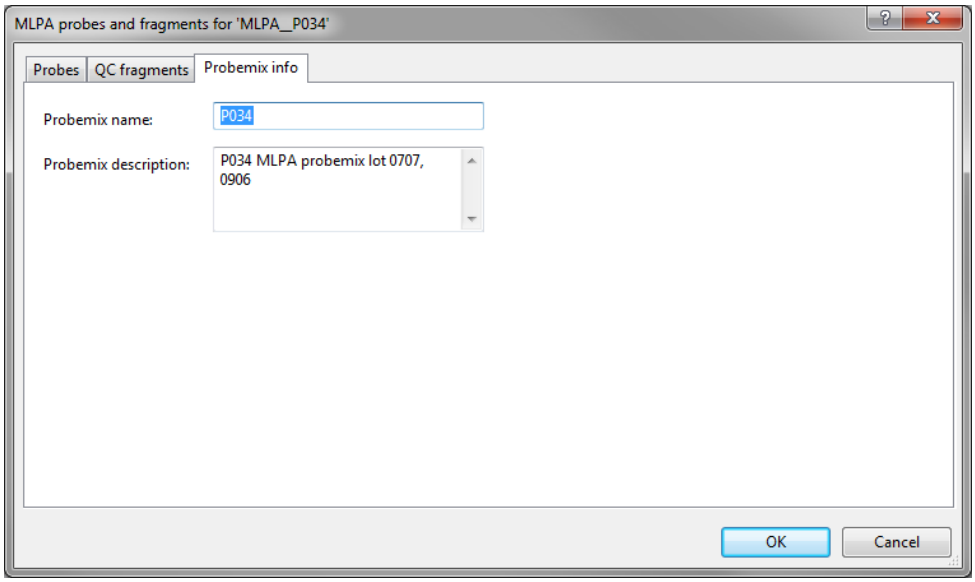


Figure 3.10: The *MLPA probes and fragments* dialog box, with the *Probe mix info tab* displayed.

Probe mix info tab:

The *Probe mix info tab* contains the 'Probe mix name' and 'Probe mix description', which can both be edited if desired.

Pressing <***OK***> in the *MLPA probes and fragments* dialog box saves the probe mix to the database:

- For each of the probes and quality control fragments, a band class is created in the source fingerprint type.
- For each of the probes, a character is created in the quantification and scores character types, and the additional information is saved in character information fields.

- The probe mix name and probe mix description are copied to the corresponding experiment information fields of source fingerprint type and character types.

After the import of a probe mix has completed, the *MLPA probes and fragments* dialog box can be called again with ***MLPA*** > ***Edit probemix***. If no probe mix has been imported yet, an empty *MLPA probes and fragments* dialog box appears and probes can be entered manually.

3.5 For the MLPA example data, press <***Sort probes***> in the *Probes tab* to call the *Sort probes* dialog box (see Figure 3.6).

3.6 Highlight the "Recommended Order" field, check ***Ascending order*** and press <***OK***>.

3.7 Press <***OK***> in the *MLPA probes and fragments* dialog box to save the probe mix.

In the *Experiment types* panel of the *Main* window, the 'MLPA probemix' and 'MLPA description' fields for the MLPA experiments now contain the probe mix name ("P034") and the probe mix description ("P034 MLPA probemix lot 0707, 0906"), respectively.

3.8 Double-click on **MLPA_Joe** in the *Experiment types* panel to open its *Fingerprint type* window.

3.9 Click on the *Band Classes* panel to display the band classes defined for **MLPA_Joe**.

In this panel (see Figure 3.11), it can be seen that band classes were generated that correspond to the probes and fragments in the probe mix.

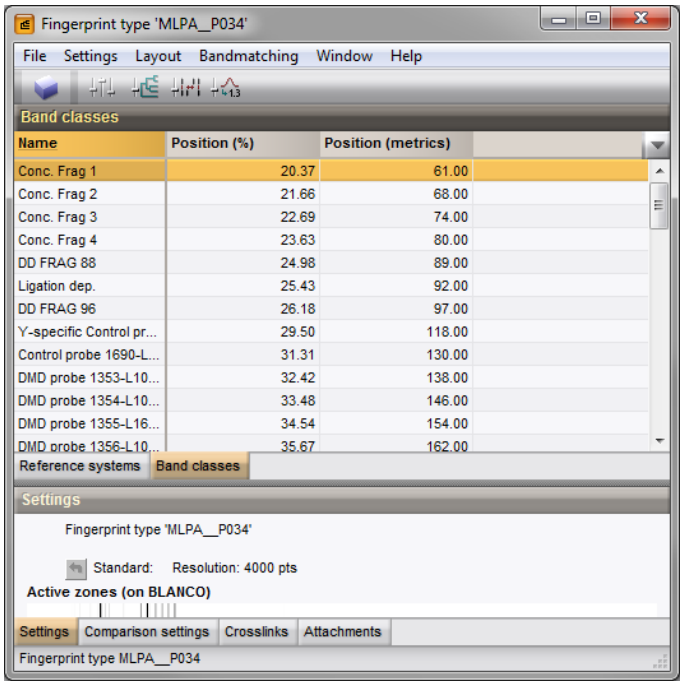


Figure 3.11: The *Band Classes* panel of the *Fingerprint type* window of **MLPA_Joe**, displaying the band classes that were generated upon import of the P034 probe mix.

3.10 Close the *Fingerprint type* window.

3.11 Double-click on **MLPA_Joe _score** or **MLPA_Joe _quant** in the *Experiment types* panel to open its *Character type* window.

In this window (see Figure 3.12 for **MLPA_Joe _score**), it can be seen that characters are added to the character types that correspond to the probes in the probe mix.

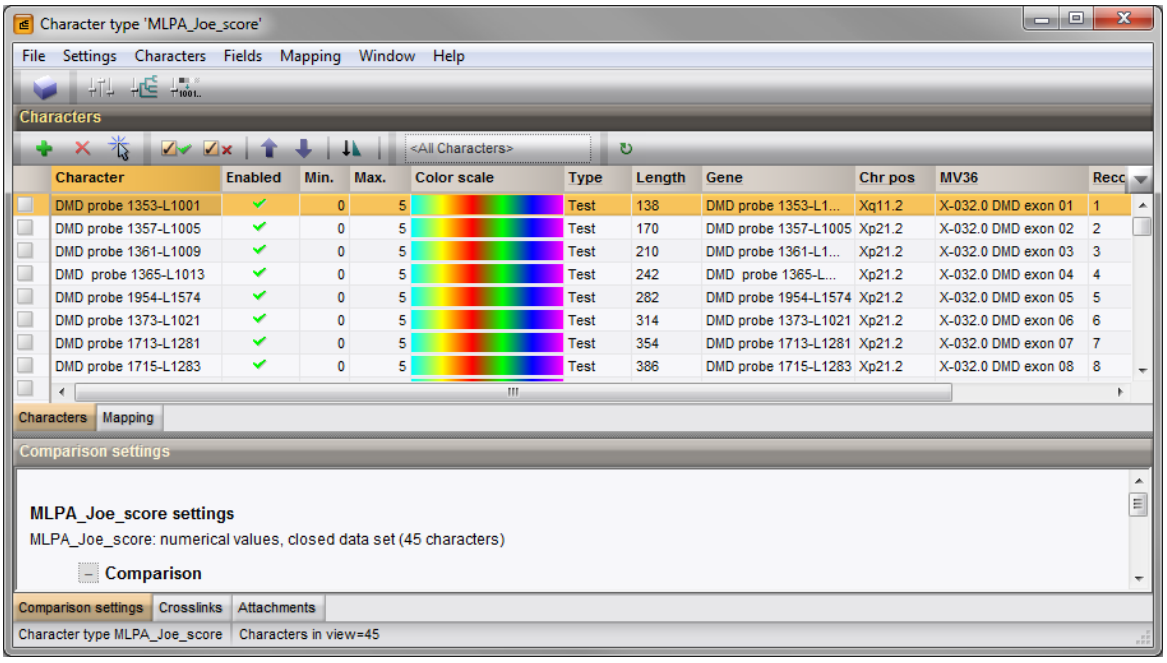


Figure 3.12: The *Character type* window for **MLPA_Joe _score**, with a character created for each MLPA probe in the probe mix.

The color scale for character types **MLPA_Joe _score** and **MLPA_Joe _quant** (see the *Character type* window of **MLPA_Joe _score** in Figure 3.12) and the mapping for **MLPA_Joe _score** is set according to the default MLPA experiment settings. These settings can be adjusted if needed (see 3.4).

3.4 MLPA experiment settings

The mapping and colors of the scoring categories of a MLPA experiment can be modified by the user.

4.1 Select ***MLPA*** > ***Settings*** to display the *MLPA settings* dialog box (see Figure 3.13).

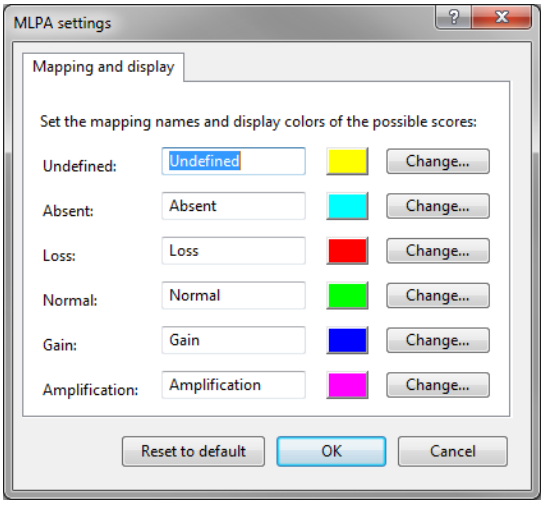


Figure 3.13: The *MLPA settings* dialog box.

The default mapping names of the scoring categories are ”Undefined”, ”Absent”, ”Loss”, ”Normal”, ”Gain” and ”Amplification”. They can be edited by typing in the corresponding text boxes. The mapping names are

stored with the scores character type and will be printed in the MLPA report (see 4.9). The corresponding color is shown next to the text box and can be edited after pressing the <***Change***> button. This action calls the *Color* dialog box, from which a color can be picked (see Figure 3.14).

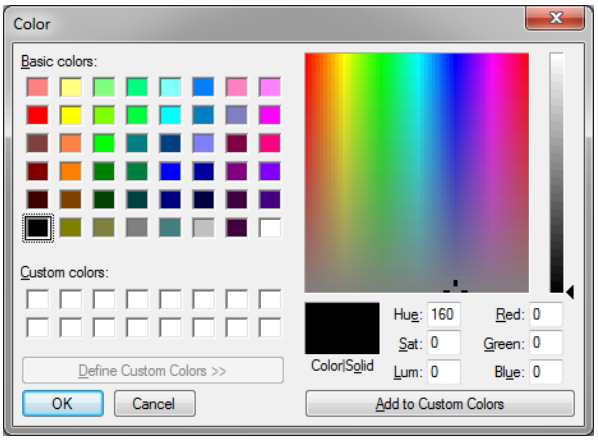


Figure 3.14: The *Color* dialog box.

Pressing <***OK***> in the *Color* dialog box will set the selected color in the *MLPA settings* dialog box.

The colors are used in the fingerprint band symbols and in the color scales of the auxiliary character types. Pressing the <***Reset to default***> button reverts the mapping names and colors to their default values.

3.5 Control and test samples

In a MLPA analysis, it is essential to include the appropriate control samples.

The example data set (see 2.2) contains a ”no template” control (BLANCO), two negative (samples D1.03.02235 DMD and D1.03.02609 DMD) controls, for which all exons will score as ”normal” and two positive controls (D1.03.01831 DMD and D1.04.04280 DMD) that have a number of known aberrations.

For easy reference later on, create an information field to hold the information whether a sample is a control and which type of control (no template, negative or positive) it is.

5.1 Make sure the *Database entries* panel is the active panel in the *Main* window and select ***Edit*** > ***Information fields*** > ***Add information field...***.

5.2 Enter the name of the new database field,e.g. ”Control” and press <***OK***> to create the field.

5.3 For entry BLANCO, enter ”no template” in the ’Control’ information field.

5.4 For entries D1.03.02235 DMD and D1.03.02609 DMD, enter ”negative” in the ’Control’ field.

5.5 For entries D1.03.01831 DMD and D1.04.04280 DMD, enter ”positive” in the ’Control’ field.

Set the negative and positive controls as MLPA control samples:

5.6 Press the **F4**-key to unselect all entries in the database.

5.7 Using **Ctrl+click**, select all samples that are indicated as negative or positive.

Selected entries are marked by a checked ballot box (☑)

5.8 Select ***MLPA*** > ***Set selected as control samples***.

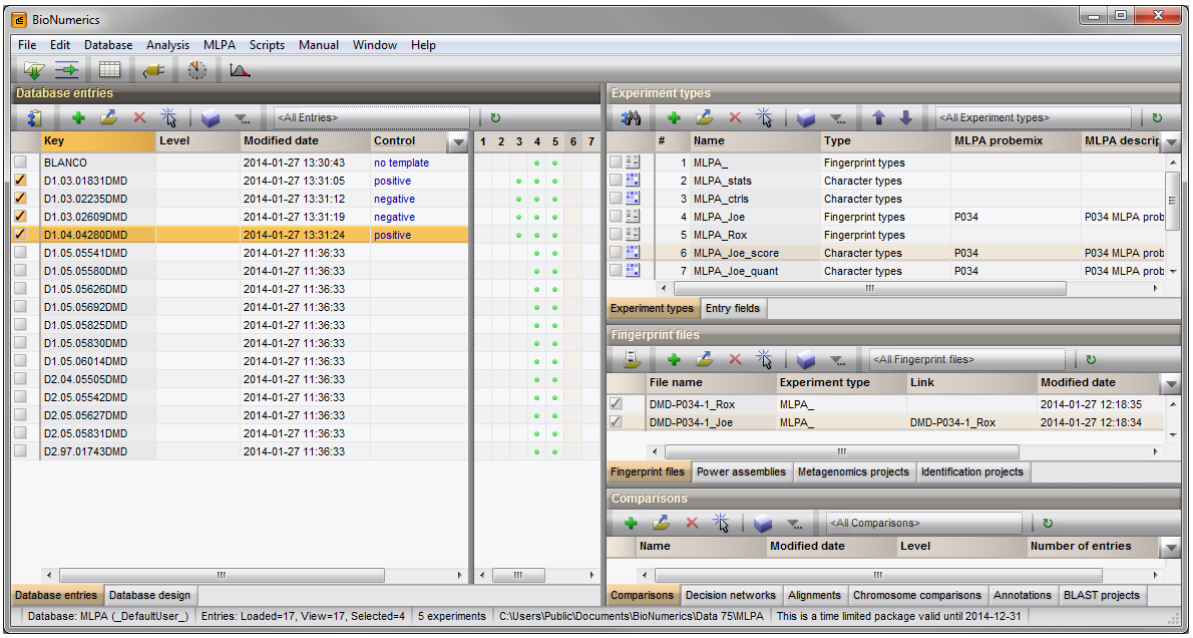


Figure 3.15: Negative and positive controls.

In the *Experiment presence* panel, a colored dot will appear for the selected entries in the **MLPA_ctrls** character type column.

Optionally, the remaining samples could be selected and set as test samples via ***MLPA*** $>$ ***Set selected as test samples***. Note that it is not required to set an entry as test sample.

The option ***MLPA*** $>$ ***Set selected as test samples*** can be used to revert erroneous control sample assignments.

Chapter 4

MLPA data analysis

4.1 Creating a comparison for analysis of MLPA data

A MLPA analysis in BioNumerics is always done in the *Comparison* window. A comparison can in principle contain any selection of entries. These entries could be contained in a single sequencer run, but they could also originate from different runs under condition that the runs were properly normalized (see the Reference manual, Chapter Setting up fingerprint type experiments for more information).

The steps involved in a MLPA analysis will be illustrated using the example MLPA database. This database currently contains the results of a single run and we will perform a MLPA analysis on all samples included in this run.

1.1 In the *Main* window select all entries in the *Database entries* panel, e.g. using the **Ctrl+A** keyboard shortcut.

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

1.3 Select ***File*** > ***Save*** ( , **Ctrl+S**) to save the comparison. When prompted for a name, enter e.g. "MLPA-run1".

For MLPA analysis it may be useful to change the configuration of the *Comparison* window:

1.4 Click in the caption of the *Dendrogram* panel, drag the panel over the *Experiment data* panel and release it in the center of the appearing docking guide to dock it as a tab of the *Experiment data* panel.

1.5 Repeat the previous step for the *Similarities* panel.

1.6 Drag the vertical separator between the *Experiments* panel, *Analyses* panel and *Groups* panel on the one hand and the *Experiment data* panel on the other hand to the left (the experiment names should still be visible).

1.7 Drag the vertical separator between the *Experiments* panel and the *Information fields* panel to the right (make sure that the 'Key' and 'Control' fields are still visible).

The above actions hide the *Dendrogram* panel and *Similarities* panel (which are generally not needed for a MLPA analysis) and reserve as much space as possible for the *Experiment data* panel, in which the MLPA fingerprint patterns will be displayed. For more display options of panels and windows, see the Reference manual, Chapter The BioNumerics user interface.

1.8 In the *Experiments* panel of the *Comparison* window, click on the eye button () that proceeds **MLPA_Joe** to display the MLPA fingerprints.

1.9 Optionally, select ***Fingerprints*** > ***Show densitometric curves*** () to display the densitometric curves.

1.10 Optionally, select ***Fingerprints*** > ***Show bands*** ( ) to display the band positions.

1.11 Zoom in vertically and horizontally using the zoom sliders until the curves are clearly visible.

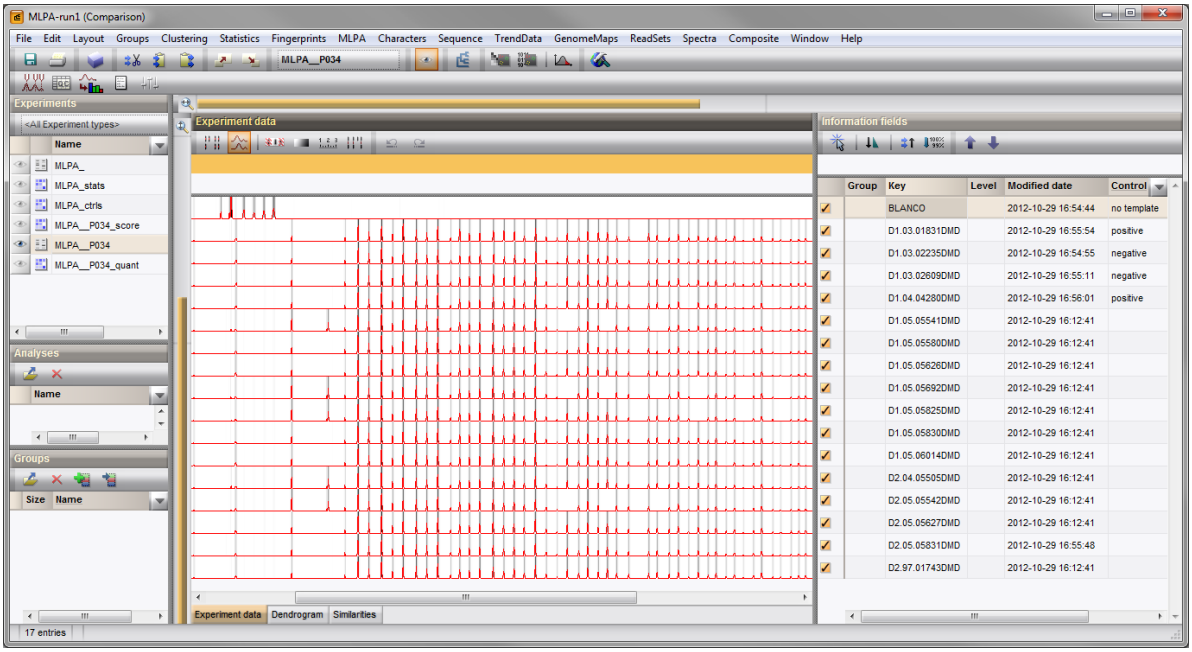


Figure 4.1: The *Comparison* window: displaying the MLPA fingerprints

4.2 Matching bands to MLPA probes and fragments

Before MLPA amplification products can be quantified, the software needs to know which band corresponds to which probe or fragment. Since the probes and fragments from the probe mix are saved as band classes with the corresponding fingerprint type (see 3.2), this can be achieved by a process called *band matching* in BioNumerics.

2.1 Select ***MLPA*** > ***Match band classes***.

This calls the *Match band classes* dialog box (see Figure 4.2).

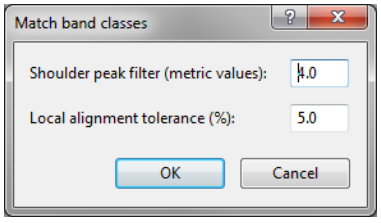


Figure 4.2: The *Match band classes* dialog box.

The ***Shoulder peak filter (metric values)*** is the threshold below which a band is considered a shoulder peak: the larger the value, the more restrictive the filter. ***Local alignment tolerance (%)*** is the tolerance, expressed as a percentage of the run length, that a peak can deviate from its expected position (as stored with the band classes) in order to still be assigned to that band class.

2.2 Leave the default settings unaltered and press <***OK***>.

This command will load the band classes that were saved with the fingerprint type (corresponding to probes and quality control fragments) and assign the highest peak in a window around each band class to that class. In addition, it will display the curves and bands of the fingerprint type in the *Experiment data* panel (if not already displayed).

The more generic band matching tool ***Fingerprints*** > ***Perform band matching...*** () can be used as well. However, the band matching tool provided with the plugin will often perform better for MLPA fingerprint data.

A number of bands are not matched to their corresponding probe or quality control fragment. Since this is the first band matching for the P034 probe mix, the actual probe/fragment sizes are different from the theoretical probe/fragment sizes.

The detected probe and fragment sizes depend on the capillary electrophoresis equipment, size standard, size calling or normalization, etc. If any of these parameters change, the detected sizes are expected to change too.

To correct the issue, we will assign the correct bands to their corresponding band classes for one or a few entries and re-center the MLPA band class positions:

2.3 If any bands were wrongly assigned, first remove the band matching with ***Fingerprints*** > ***Undo*** ( , **Ctrl+Z**).

2.4 For the first test sample in the comparison (not the BLANCO), locate the right-most band and assign it to band class "Control probe 1692-L1531". Assignment is done by clicking on the band, which is indicated by a red diamond, and drag it to the band class position. Using default settings, a band that is assigned to a band class will be indicated with a yellow cross.

2.5 Proceed from right to left through the pattern, assigning each band to its corresponding band class.

2.6 Select ***MLPA*** > ***Center all band class positions***.

The band class positions are now re-centered based on the assignments made and automatically saved to the fingerprint type.

2.7 Select ***MLPA*** > ***Match band classes*** or press the  button again.

The band matching will now be much more accurate. Carefully review and manually adjust the assignment if needed.

Optionally, bands that are not matched to a probe can be deleted with ***MLPA*** > ***Remove bands not associated with a band class***. We will not remove the unmatched bands for the example data, to be able to calculate the number of shoulder peaks later on (see 4.3).

For a detailed description on how to select, add or delete bands in a comparison, we refer to the Reference manual, Chapter Band matching and polymorphism analysis.

When the band matching is completed, save the comparison:

2.8 Select ***File*** > ***Save*** ( , **Ctrl+S**).

When a comparison is saved using ***File*** > ***Save*** ( , **Ctrl+S**) or ***File*** > ***Save as...***, any modified band information (added or deleted bands) is automatically saved to the corresponding gels. Modified band information can also be stored without making any modification to the comparison itself by selecting ***Fingerprints*** > ***Save modified band information...***.

4.3 Quality control

A MLPA probe mix typically contains a number of quality control fragments:

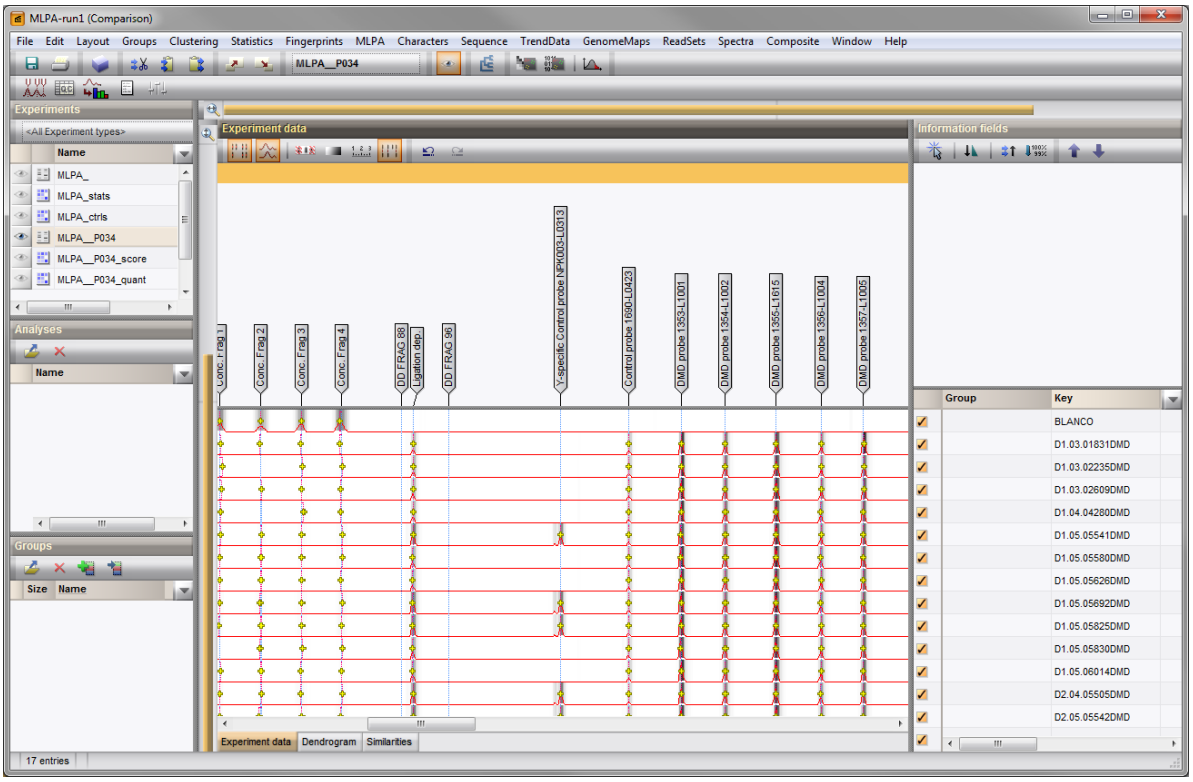


Figure 4.3: Comparison after MLPA band matching.

- **DNA concentration fragments**: The peak size of these fragments is inversely proportional to the concentration of template DNA in the PCR reaction; they will only be visible in samples with a very low DNA concentration.
- **Ligation-dependent fragment**: This peak is only present when sufficient template DNA was added and when ligation occurred. Its peak size should be similar to that of other MLPA amplification products.
- **Denaturation fragments**: These fragments have a high %G+C and are therefore difficult to denature. Their peak sizes should be comparable to that of other MLPA amplification products.

The (relative) sizes of these quality control fragment peaks are therefore important to assess the quality of the MLPA reaction.

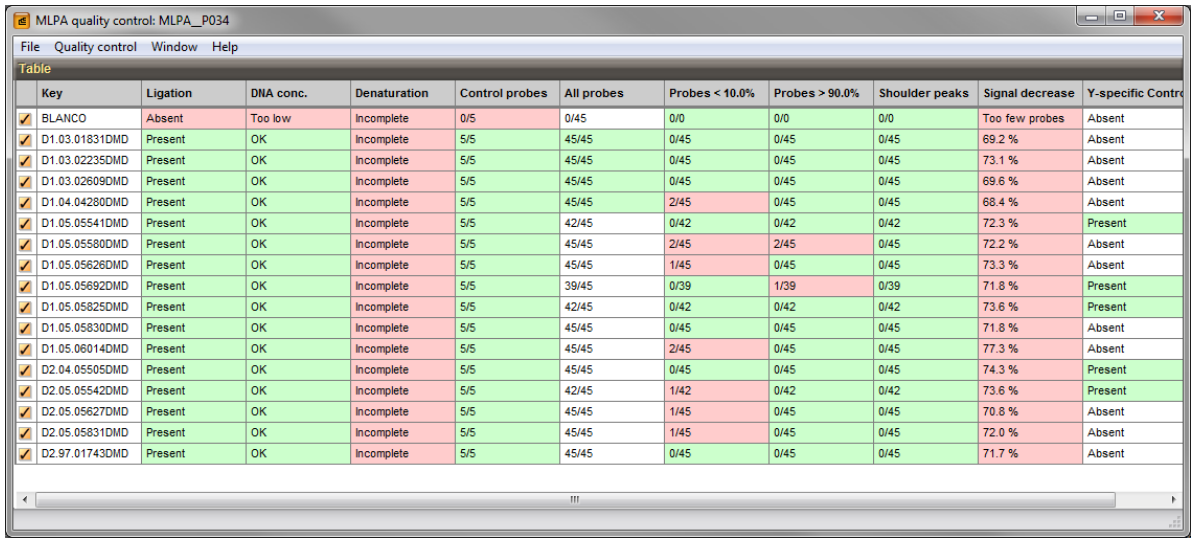
Furthermore, for an accurate quantification, peak sizes should be within the linear range of the total optical density: very large peaks will be ”topped off” since the signal is higher than the dynamic range of the detector, while very small peaks will be influenced in an unacceptable way by the noise that is inherent to the analysis.

Finally, there are two parameters that are an indication of a lower quality of the capillary electrophoresis: a strong signal decrease along the running distance and the presence of shoulder peaks in the profile.

The quality control in the *MLPA plugin* checks the presence of the quality control fragments and control probes. Probes above or below a certain threshold are flagged and the signal decrease and presence of shoulder peaks are reported.

3.1 Select ***MLPA*** > ***Quality control*** or press the  button to display the *MLPA quality control window* (see Figure 4.4).

For each sample in the comparison, the table reports a number of quality control parameters:



Key	Ligation	DNA conc.	Denaturation	Control probes	All probes	Probes < 10.0%	Probes > 90.0%	Shoulder peaks	Signal decrease	Y-specific Contr
BLANCO	Absent	Too low	Incomplete	0/5	0/45	0/0	0/0	0/0	Too few probes	Absent
D1.03.01831DMD	Present	OK	Incomplete	5/5	45/45	0/45	0/45	0/45	69.2 %	Absent
D1.03.02235DMD	Present	OK	Incomplete	5/5	45/45	0/45	0/45	0/45	73.1 %	Absent
D1.03.02609DMD	Present	OK	Incomplete	5/5	45/45	0/45	0/45	0/45	69.6 %	Absent
D1.04.04280DMD	Present	OK	Incomplete	5/5	45/45	2/45	0/45	0/45	68.4 %	Absent
D1.05.05541DMD	Present	OK	Incomplete	5/5	42/45	0/42	0/42	0/42	72.3 %	Present
D1.05.05580DMD	Present	OK	Incomplete	5/5	45/45	2/45	2/45	0/45	72.2 %	Absent
D1.05.05626DMD	Present	OK	Incomplete	5/5	45/45	1/45	0/45	0/45	73.3 %	Absent
D1.05.05692DMD	Present	OK	Incomplete	5/5	39/45	0/39	1/39	0/39	71.8 %	Present
D1.05.05825DMD	Present	OK	Incomplete	5/5	42/45	0/42	0/42	0/42	73.6 %	Present
D1.05.05830DMD	Present	OK	Incomplete	5/5	45/45	0/45	0/45	0/45	71.8 %	Absent
D1.05.06014DMD	Present	OK	Incomplete	5/5	45/45	2/45	0/45	0/45	77.3 %	Absent
D2.04.05505DMD	Present	OK	Incomplete	5/5	45/45	0/45	0/45	0/45	74.3 %	Present
D2.05.05542DMD	Present	OK	Incomplete	5/5	42/45	1/42	0/42	0/42	73.6 %	Present
D2.05.05627DMD	Present	OK	Incomplete	5/5	45/45	1/45	0/45	0/45	70.8 %	Absent
D2.05.05831DMD	Present	OK	Incomplete	5/5	45/45	1/45	0/45	0/45	72.0 %	Absent
D2.97.01743DMD	Present	OK	Incomplete	5/5	45/45	0/45	0/45	0/45	71.7 %	Absent

Figure 4.4: The *MLPA quality control window* for the example data set.

- **Ligation**: Whether or not a peak was found for the ligation-dependent quality control fragment.
- **DNA conc.**: Whether or not sufficient template DNA was used in the PCR reaction.
- **Denaturation**: Whether or not the denaturation was complete.
- **Control probes**: The number of control probe peaks observed, over the expected number of control probe peaks, based on the probe mix used.
- **All probes**: The number of test probe peaks observed, over the expected number of test probe peaks, based on the probe mix used.
- **Probes $<$ "min%OD"**: The number of probes that have a peak size below the specified minimum percentage of the OD range, over the total number of probes observed.
- **Probes $>$ "max%OD"**: The number of probes that have a peak size above the specified maximum percentage of the OD range, over the total number of probes observed.
- **Shoulder peaks**: The number of probes for which shoulder peaks were detected, over the total number of probes observed.
- **Signal decrease**: The relative decrease in signal (expressed in %) towards the end of the fingerprint, calculated based on the medians of the five first and five last probe peak sizes.
- **Mutation-specific control peaks**: The presence or absence of any specific control peaks present in the probe mix, e.g. the Y-specific control probe in the P034 probe mix.

Individual cells in the table are colored green if the parameter passes the quality control settings or red if it fails for some reason. The quality control settings can be adjusted and the table updated with the new settings:

3.2 Select ***Quality control*** $>$ ***Settings*** to call the *Quality control settings* dialog box (see Figure 4.5).

Under QC fragments, the thresholds can be entered to flag a sample as either reliable or unreliable based on the sizes of the quality control fragment peaks. The default thresholds correspond to the guidelines proposed by MRC-Holland (http://www.mrc-holland.com):

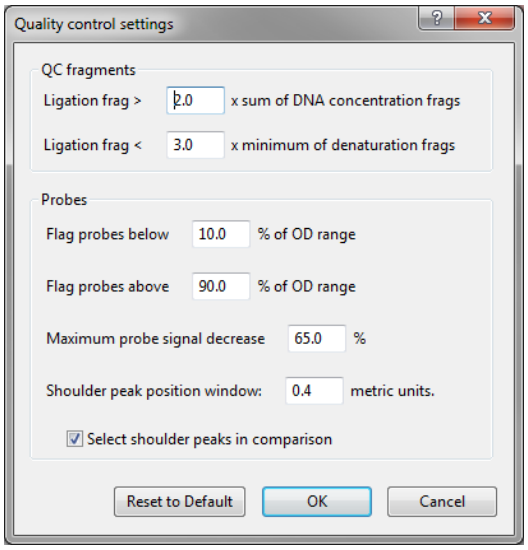


Figure 4.5: The *Quality control settings* dialog box.

- The ligation fragment peak size should be larger than twice the sum of the DNA concentration fragment peak sizes.
- The ligation fragment peak should be smaller than three times the size of the smallest peak produced by the denaturation fragments.

If the first condition is not met, the DNA concentration of the sample will be flagged as ”too low” in the *MLPA quality control window*. If the second conditions is not fulfilled, the denaturation is expected to be incomplete and the sample flagged accordingly. In case no peak is found for the ligation-dependent fragment, this will be indicated in the report as well.

Under ***Probes***, the option is provided to ***Flag probes below*** or ***Flag probes above*** a certain percentage of the OD range. ***Maximum probe signal decrease*** refers to the decrease in intensity of peaks in function of the run length (so-called ”ski-sloping”), a typical phenomenon in capillary electrophoresis. Furthermore, a window (the ***Shoulder peak position window***) can be entered in metric units (e.g. base pairs) in which the program will search for shoulder peaks. When ***Select shoulder peaks in comparison*** is checked, the shoulder peaks will be selected in the comparison, ready for deletion.

A shoulder peak will only be found if a band is actually assigned at that position. For example, if previously ***MLPA*** > ***Remove bands not associated with a band class*** is selected, no shoulder peaks will be detected.

Clicking the <***Reset to Default***> button will revert all settings to their default values.

3.3 As an illustration, change one or a few settings, e.g. set the ***Maximum probe signal decrease*** to 70% and press <***OK***>.

The *MLPA quality control window* will now be updated with the new settings.

The report table can be sorted according to the information in any of the columns:

3.4 For example, to sort the MLPA profiles according to loss in signal intensity, click in the header of the ’Signal decrease’ column and select ***Quality control*** > ***Sort by column***.

In the *MLPA quality control window*, samples that do not meet the quality criteria can be selected for subsequent removal from the comparison.

For example, since the no template control (BLANCO) does not contain any peaks, we could remove it from

the analysis:

3.5 **Ctrl+click** the BLANCO sample; it is now selected ( ).

3.6 In the underlying *Comparison* window, select ***Edit*** > ***Cut selection*** (, **Ctrl+X**) to remove the BLANCO from the comparison.

When changes are made to the underlying comparison (e.g. adding or removing entries), to the band definitions or band matching, the *MLPA quality control window* can be updated by selecting ***Quality control*** > ***Update***.

3.7 Close the *MLPA quality control window* with ***File*** > ***Exit***.

4.4 Band selection tools

A number of band assignment functions (see 4.5) in the *MLPA plugin* work only on the currently selected band(s). Bands can be selected in several ways:

4.1 To select a single band, just click on it with the mouse.

4.2 To select a number of adjacent bands, press the **Shift**-key on the keyboard while dragging a rectangle with the mouse.

4.3 To select all bands belonging to a certain band class, double-click on any band in that band class.

4.4 To select all bands that belong to the currently highlighted entry, use ***MLPA*** > ***Select bands (current entry)*** or press **Shift+E** on the keyboard.

4.5 To select all bands that belong to the currently selected entries, use ***MLPA*** > ***Select bands(selected entries)*** or press **Shift+S** on the keyboard.

4.6 To select all bands in a comparison at once, use ***MLPA*** > ***Select bands (all entries)*** or press **Shift+A** on the keyboard. Only bands that are associated with a band class are selected this way.

4.7 To select all bands that correspond to MLPA control probes at once, use ***MLPA*** > ***Select bands (control probes)*** or press **Shift+P** on the keyboard.

4.8 To select all bands from samples that are defined as control samples (see 3.5), use ***MLPA*** > ***Select bands (control samples)*** or press **Shift+C** on the keyboard.

Any selection of bands is cleared when clicking on a random position in the fingerprint image.

4.5 User–defined and automated band assignment

Since MLPA amplification products are quantified relative to the size of control probes, we can manually assign the control probes in all samples as normal:

5.1 Select ***MLPA*** > ***Select bands (control probes)*** or press **Shift+P** on the keyboard to select the control probes in all samples.

5.2 Select ***MLPA*** > ***Assign as normal (selected bands)*** or press **Shift+F7** on the keyboard to assign the control probes as normal.

The bands that correspond to control probes are now displayed as green diamonds.

For the control samples, the band assignments are also known: all bands for the negative controls are normal and most bands (except the known deletions and duplications; see 2.2 for the controls included with the example data) for the positive controls are normal.

5.3 Select ***MLPA*** > ***Select bands (control samples)*** or press **Shift+C** on the keyboard to select all bands in the control samples.

5.4 Select ***MLPA*** > ***Assign as normal (selected bands)*** or press **Shift+F7** on the keyboard to assign the bands in the control samples as normal.

All bands in the control samples (negative and positive) are now displayed as green diamonds. Since the positive control samples contain a number of deletions or duplications, some manual adjustments are needed:

5.5 In control sample D1.03.01831 DMD, select the band that is assigned to DMD probe 1357-L1005 and select ***MLPA*** > ***Assign as gain (selected bands)*** or press **Shift+F8** on the keyboard. The band is indicated with a blue diamond.

5.6 For the same entry, assign the bands corresponding to DMD probes 1361-L1009, 1365-L1013, 1954-L1574, 1373-L1021, 1713-L1281,1715-L1283, 1385-L1033, and 1718-L1286 as gain, using the procedure described in the previous step.

5.7 In control sample D1.04.04280 DMD, select the band that is assigned to DMD probe 1715-L1283 and select ***MLPA*** > ***Assign as loss (selected bands)*** or press **Shift+F6** on the keyboard. The band is indicated with a red diamond.

5.8 For the same entry, assign the bands corresponding to DMD probes 1385-L1033 and 1718-L1286 as loss, using the procedure described in the previous step.

If a mistake was made during band assignment, the assignment can be reset to undefined:

5.9 Select the band(s) for which you want to reset the assignment and select ***MLPA*** > ***Reset assignments (selected bands)*** or press **Shift+F5** on the keyboard.

Now that all known bands are assigned, the remaining bands can be automatically assigned:

5.10 Select ***MLPA*** > ***Assign automatically (all bands)*** or press the  button.

When an automatic band assignment is executed, a number of actions occur: band sizes are calibrated within and between samples, the quantifications are saved in the quantification character type, quantified bands are scored, the scores saved in the score character type, and the character statistics are calculated. After the calculations, the quantification character type is displayed.

In case the automatic assignment conflicts with manual assignments, conflicts are listed in a separate *MLPA assignment warnings window* (see Figure 4.6).

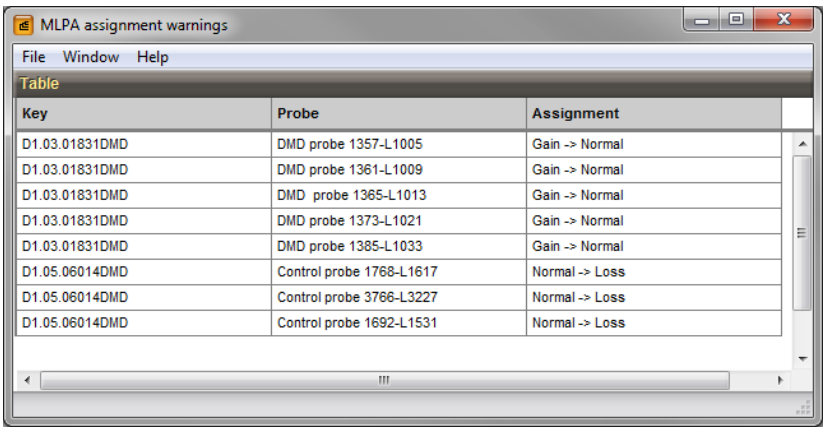


Figure 4.6: The *MLPA assignment warnings window*.

The quantifications and scores that were calculated are discussed in the next paragraph (4.6).

The automated band assignment which we just performed is based on the default normalization, quantification and scoring settings. See 4.7 on how to modify these settings.

4.6 MLPA quantifications and scores

After automatic band assignment, the quantification character type is displayed and put in focus. It can be called any other time by clicking the eye button ( ) that precedes the quantification character type in the *Experiments* panel of the *Comparison* window (see Figure 4.7).

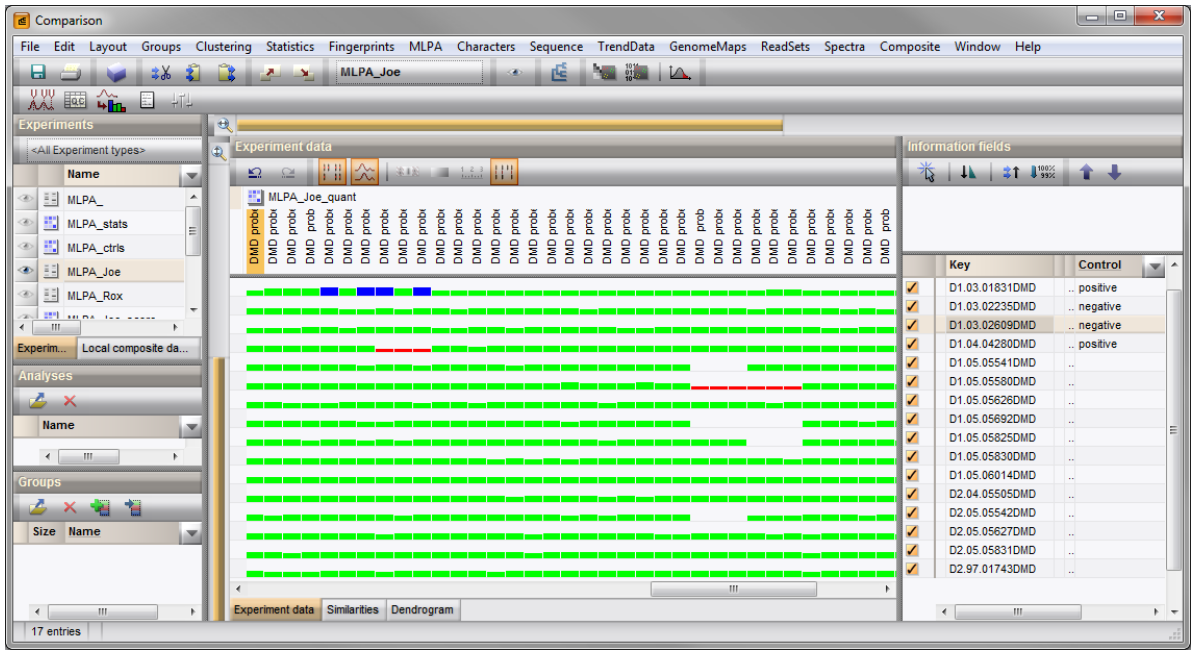


Figure 4.7: The *Comparison* window for the example data, with **MLPA_Joe _quant** displayed.

By default, the quantifications are displayed as bar graphs, colored according to their scoring.

The scores character type is used to store the scores of all probes in the probe mix. It can be displayed pressing the eye button ( ) next to the experiment name in the *Experiments* panel.

The characters (= probes) in the quantification and scores character types are sorted according to the sort order specified for the probe mix (see 3.3). To order the probes differently, use following procedure:

6.1 Select ***MLPA*** > ***Edit probe mix*** in the *Main* window.

6.2 If more than one MLPA experiment type is present, choose the correct one from the list and press <***OK***>.

6.3 In the *MLPA probes and fragments dialog box*, press <***Sort probes***>.

6.4 Select a different sort field and/or sort order in the *Sort probes dialog box* and press <***OK***>.

6.5 Press <***OK***> to close the *MLPA probes and fragments dialog box*.

In the *Comparison* window, the character experiment(s) need to be re-loaded in order to reflect the current sort order:

6.6 Click the eye button ( ) that precedes the character type in the *Experiments* panel of the *Comparison* window to hide the experiment data and then to load the data again.

The probes are now displayed in the specified order.

The information that is displayed in the caption of the character type's experimental data (by default the character name) can be changed to any of the character information fields that were imported from the probe mix by setting a default field for the character type. For example, display the exon number for **MLPA_Joe _quant** as follows:

6.7 In the *Experiment types* panel of the *Main* window, double-click **MLPA_Joe _quant** to open its *Character type* window.

6.8 Click on the header of character field 'MV36' to highlight it and select ***Fields*** > ***Use a default field***.

In the *Comparison* window, the exon information is now displayed instead of the character names for **MLPA_Joe _quant**.

Automatic scores can be manually overruled in the fingerprint type (see 4.5). To update the scores character type with the manual scores, select ***MLPA*** > ***Export scoring to character type***.

4.7 MLPA analysis settings

All MLPA analysis settings, regarding peak size normalization, peak quantification and scoring are grouped in one dialog box.

7.1 Select ***MLPA*** > ***Settings*** or press the button to call the *MLPA analysis settings* dialog box. By default, the *Normalization tab* is displayed (see Figure 4.8).

The *MLPA analysis settings* dialog box consists of four tabs: *Normalization*, *Quantification*, *Scoring*, and *Analysis templates*.

Normalization tab:

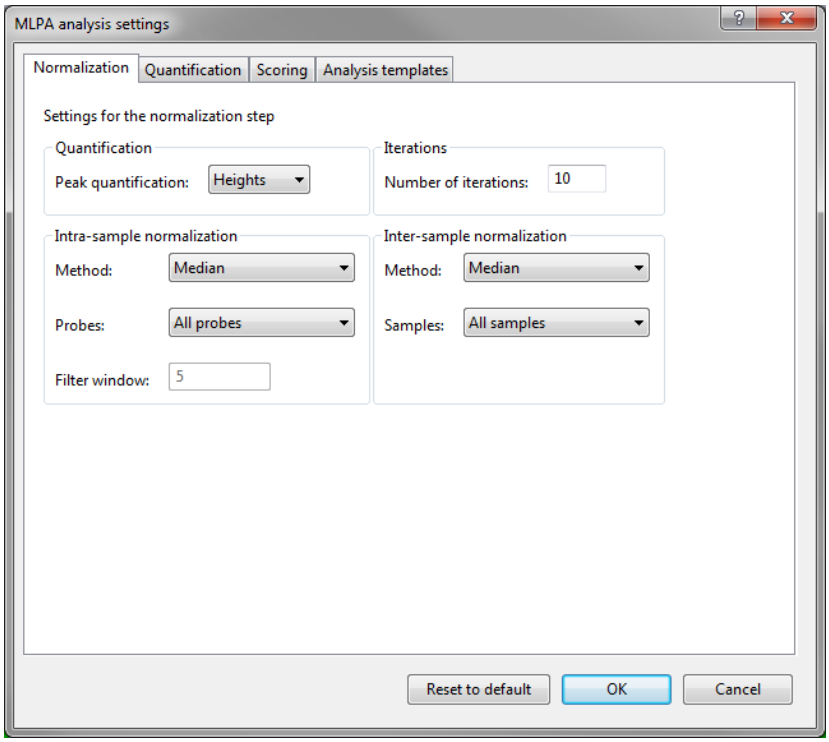


Figure 4.8: The *MLPA analysis settings* dialog box, *Normalization tab* displayed.

The *MLPA plugin* uses an iterative normalization algorithm that normalizes peak sizes within and between

samples. The settings for this algorithm are displayed in the *Normalization tab*.

As parameter for ***Peak quantification***, either the peak ”Heights” or ”Areas” can be selected, meaning that either the peak height or the surface under the peak is used in the calculations.

The ***Number of iterations*** that the algorithm goes through can be specified as a whole number between 0 and 100. In nearly all circumstances, the default value of 10 iterations will be more than sufficient. Obviously, if ”0” is entered, no size normalization will be performed.

For ***Intra-sample normalization***, the default setting is to scale all peaks in a sample with the ***Median*** value of ***All probes*** in that sample. Using this setting, an acceptable normalization can be obtained in nearly all situations. An second scaling method that scales peaks with the same factor, irrespective of their position in the sample, is the ***Mean*** value. However, this method is more sensitive to outliers compared to ***Median***. Instead of using a fixed scaling factor, a curve (either ***Linear*** or ***Exponential***) can be fitted through the sizes. This is only useful if the slope of the signal decrease varies between samples. If all samples have the same sloping, then the probe-specific scaling will correct for it. The probes that are used to calculate a median or mean value from, or that are used to fit a curve through, can be ***All probes***, the ***Control probes*** only or ***Filtered probes***. If ***Filtered probes*** is selected, a ***Filter window*** can be specified as an odd number, representing the number of peaks that will be taken into account.

For ***Inter-sample normalization***, the default setting is to scale samples with the ***Median*** of ***All samples***. Alternative methods are ***Mean*** and ***Minimum***, referring to the average peak size or minimum peak size, respectively, of the samples used to normalize against. The latter can be ***Control samples*** (see 3.5 on how to assign control samples), ***All samples*** or the ***Selected samples***.

Quantification tab:

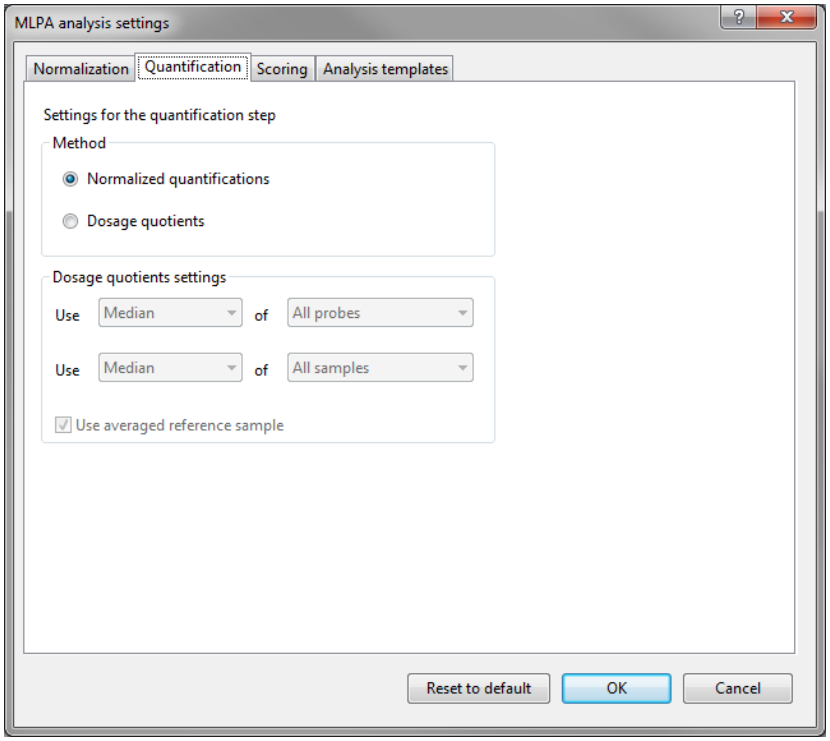


Figure 4.9: The *MLPA analysis settings* dialog box, *Quantification tab* displayed.

In the *Quantification tab* of the *MLPA analysis settings* dialog box (see Figure 4.9), the settings for relative quantification of the MLPA amplification products are grouped. The choice is offered between ***Normalized quantifications***, i.e. using the intra- and inter-sample normalization procedure for which the settings are discussed in the *Normalization tab*, or calculating ***Dosage quotients*** on the normalized quantifications.

The dosage quotient (D_Q)is calculated as: $D_Q = \frac{a/b}{A/B}$, with a the peak size of the test probe in the test sample, A the (averaged) peak size of the corresponding probe in the control sample(s), b the (averaged) peak size of the control probe(s) in the test sample and B the (averaged) peak size of the control probe(s) in the control sample(s).

While a in the formula above is always unambiguously defined, b, A and B can be calculated in different ways when more than one control probe and/or control sample exist. This can be specified in the ***Dosage quotient settings***. b and B can be calculated as either the ***Median*** or ***Mean*** of either the ***Control Probes*** or ***All probes*** in the fingerprint. A and B can be calculated as either the ***Median*** or ***Mean*** of the ***Control samples***, ***All samples*** or the ***Selected samples***. When ***Use average reference sample*** is checked (default), an average reference sample is first calculated based on all reference samples in the comparison. When this option is unchecked, the A/B ratios are first calculated and then averaged.

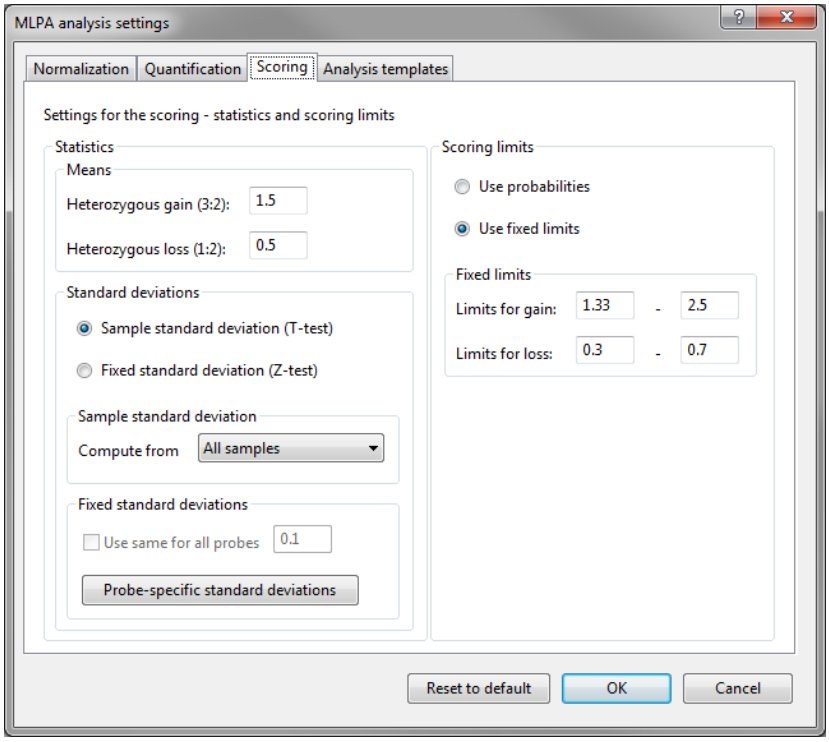


Figure 4.10: The *MLPA analysis settings* dialog box, *Scoring tab* displayed.

Scoring tab:

The settings in the *Scoring tab* of the *MLPA analysis settings* dialog box (see Figure 4.10) determine how the bands are scored, based on the quantifications obtained.

For calculation of the odds ratios, the program needs to know the distribution of the peak quantifications. Therefore, under ***Statistics***, the expected ***Means*** for a ***Heterozygous gain (3:2)*** and ***Heterozygous loss (1:2)*** can be specified as a factor, relative to a normal sample. As the standard deviation of the distribution, the ***Sample standard deviation*** can be used, calculated on the ***Control samples***, ***All samples*** or on the ***Selected samples***. Alternatively, a ***Fixed standard deviation*** can be used. This is specified as a percentage for all probes (when ***Use same for all probes*** is checked), or entered for each probe individually after pressing <***Probe-specific standard deviations***>.

As ***Scoring limits***, either the odds ratios (***Use probabilities***) or fixed cut-offs (***Use fixed limits***) can be employed. In the latter case, the ***Limits for gain*** and ***Limits for loss*** can be entered in the corresponding text boxes.

Analysis templates tab:

The *Analysis templates tab* (see Figure 4.11) allows the user to save the current settings as an analysis

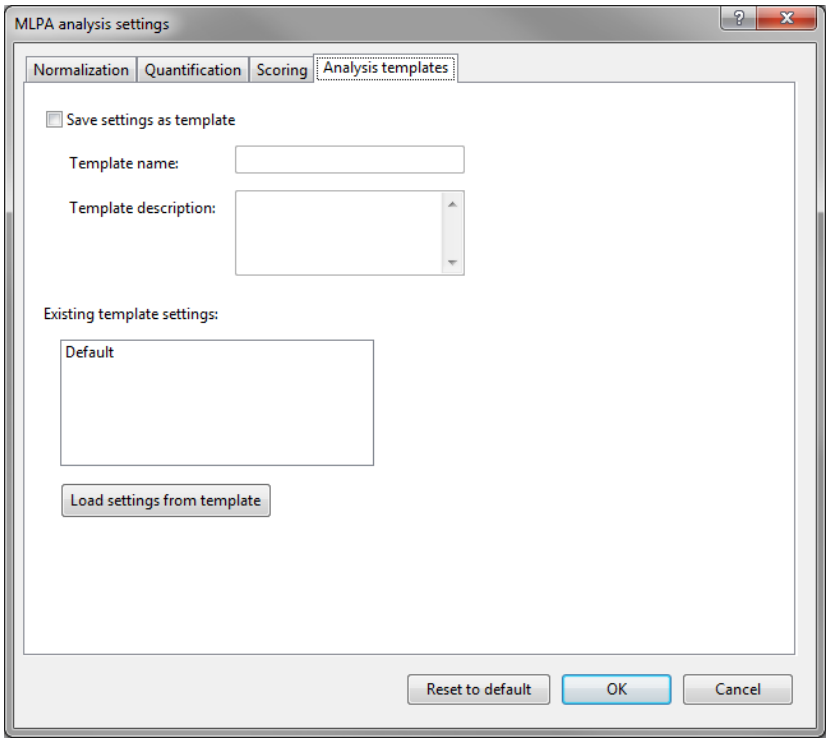


Figure 4.11: The *MLPA analysis settings* dialog box, *Analysis templates tab* displayed.

template by checking ***Save settings as template***. A ***Template name*** should be entered and optionally a ***Template description***.

Saved templates can be applied by selecting the template from the ***Existing template settings*** list and pressing <***Load settings from template***>. An analysis template is automatically created for each MLPA experiment type after the first time the *MLPA analysis settings* dialog box has been approved. These settings can then be loaded to be used in other MLPA experiment types.

Since analysis templates are objects, they can be shared, locked, queried, and their use can be restricted just like for any other object in a BioNumerics database (see the Reference manual, Chapter Database objects).

4.8 Probe distributions

Plots of the quantification distribution of the MLPA probes can be shown.

8.1 With a MLPA fingerprint type or auxiliary character type highlighted in the *Experiments* panel, select ***MLPA*** > ***Probe distributions***. This pops up the *MLPA probe distribution window* (see Figure 4.12).

For each probe in the probe mix and additionally for all probes together, a distribution plot is displayed. This plot shows a histogram of the number of values in relation to their quantification. Bars in the histogram are colored according to their score (see 3.4 on how to set these colors). On top of the histogram, a normal distribution is plotted for each scoring category. Furthermore, a number of statistics are reported with each plot: the number of characters in this scoring category, the mean quantification and the standard deviation.

The probe distribution plots are arranged in the order as specified for the probe mix (see 3.3).

From the probe distributions of the example data, it can be seen that a number of probes that were automatically assigned as loss, are manually assigned as normal (see the ”All probes” plot in Figure 4.12).

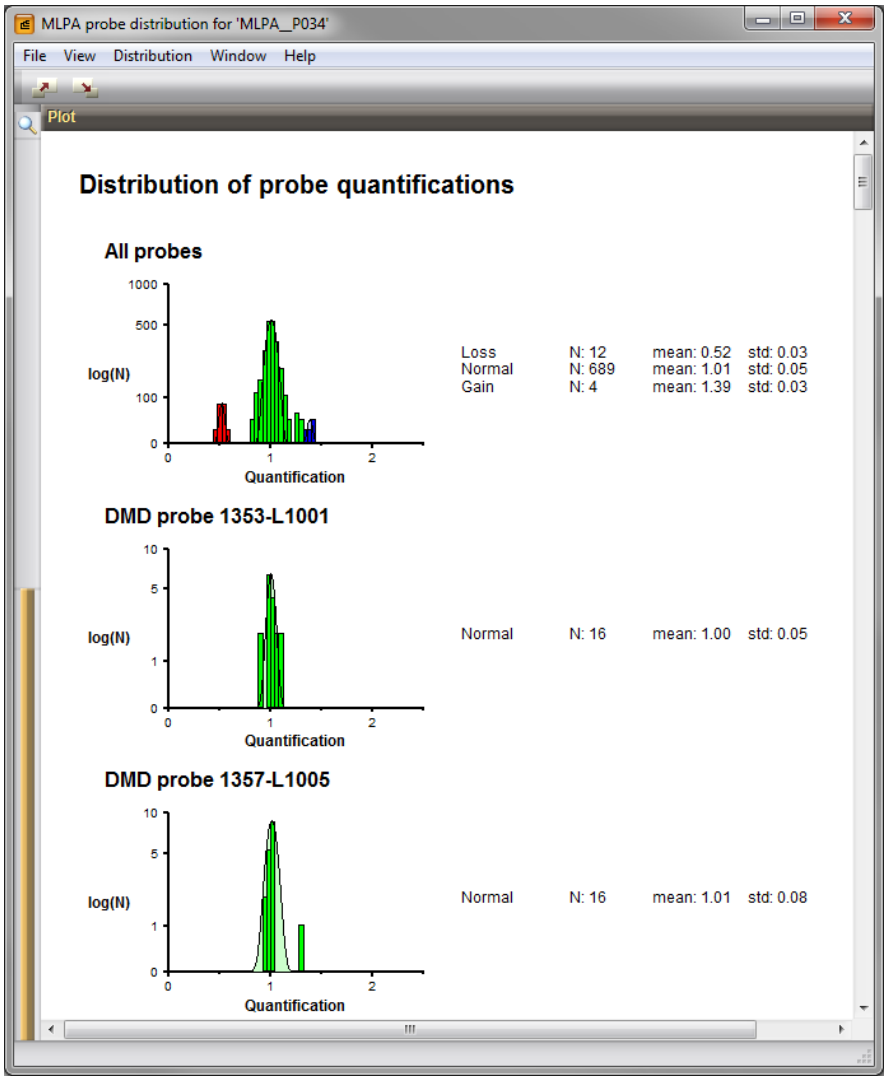


Figure 4.12: The *MLPA probe distribution window*.

Instead of plotting all probes, a selection can be made of probes which should be plotted. For example, to display only the plots of the control probes in the example data set:

8.2 Select ***Distribution*** > ***Select probes to plot***.

This calls the *Select probes to plot* dialog box (see Figure 4.13).

Select the probes you wish to plot.

8.3 From the list that appears, select the five control probes by clicking on the first control and, while holding the **Shift**-key, clicking on the last control.

8.4 Press <***OK***>. The *MLPA probe distribution window* now only displays the plots for the five control probes.

For the control probes, which were all manually assigned as normal, it can be seen that some have a normalized size that corresponds to a loss in the automatic assignment.

The plots can be printed via ***File*** > ***Print*** or copied to the clipboard via ***File*** > ***Copy to clipboard***. From the clipboard, the copied information can then be pasted in applications such as Adobe Photoshop, Microsoft PowerPoint, Microsoft Word, etc.

The overall means for bands scored as loss and gain can be saved to the MLPA analysis settings (see 4.7),

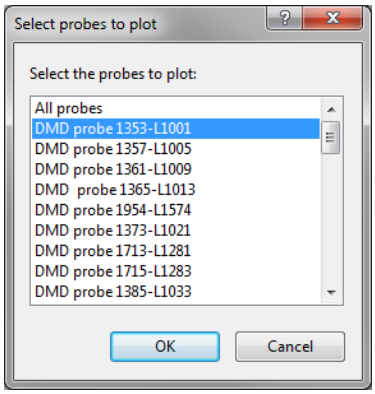


Figure 4.13: The *Select probes to plot* dialog box: select probes.

as well as the standard deviation for the normal bands of each probe. We will do this for the example data:

8.5 Select ***Distribution*** > ***Save statistics to settings***. A confirmation dialog box appears, asking "Do you want to save the quantification means and the probe standard deviations to the analysis settings?".

8.6 Press <***Yes***> to confirm.

8.7 In the *Comparison* window, select ***MLPA*** > ***Settings*** to call the *MLPA analysis settings* dialog box.

8.8 In the *Scoring tab*, it can be seen that the means for ***Heterozygous gain (3:2)*** and ***Heterozygous loss (1:2)*** have been updated.

8.9 Press <***Probe-specific standard deviations***> to display a *MLPA statistics dialog box* with standard deviations used for each of the probes: these have been updated as well.

8.10 Press <***Cancel***> to close the statistics dialog box.

8.11 Check ***Fixed standard deviation*** under ***Standard deviations*** and ***Use probabilities*** under ***Scoring limits*** in the *MLPA analysis settings* dialog box.

8.12 Press <***OK***> to accept the new settings.

In the *Experiment data* panel, the histogram colors for a few character values in the **MLPA_Joe _quant** character type are updated.

8.13 Select ***MLPA*** > ***Assign automatically (all bands)*** or press the button to perform an assignment with the new settings.

4.9 MLPA reports

For each entry, a customizable report can be shown, including database information on the sample, probe information, quality control, analysis settings, and probe statistics.

9.1 Click on a sample in the *Information fields* panel to highlight it (for example, the positive control D1.04.04280 DMD) and select ***MLPA*** > ***Report (current entry)***.

The first time this command is executed for a given MLPA experiment type, the *MLPA report settings* dialog box appears (see Figure 4.14).

Under ***Sample info fields to include***, all database information fields of the samples are listed. Fields to be displayed in the report can be highlighted using **Ctrl+click** or a complete range highlighted by clicking on the first field and, while holding the **Shift**-key, clicking on the last field to be included. The ***Sample info***

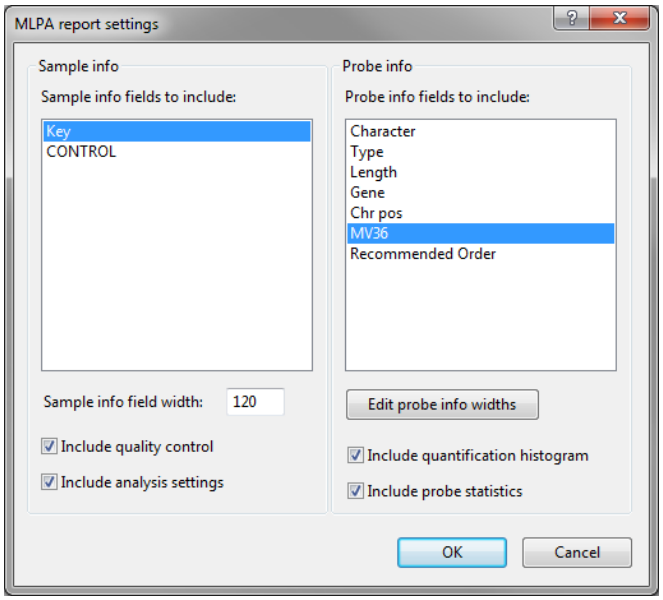


Figure 4.14: The *MLPA report settings* dialog box.

field width (in points) can be entered. Quality control parameters and analysis settings can be displayed in the report by checking ***Include quality control*** and ***Include analysis settings***, respectively.

The *current* analysis settings will be reported, i.e. when the analysis settings are changed *after* the automatic assignment is performed, the reported settings might not correspond to the settings that were used to perform the analysis.

Under ***Probe info fields to include***, all experiment information fields of the probes are listed. Fields to be displayed in the report can be highlighted using **Ctrl+click** or a complete range highlighted by clicking on the first field and, while holding the **Shift**-key, clicking on the last field to be included. The widths of individual probe information fields can be set by clicking the <***Edit probe info widths***> button. This pops up a new dialog (see Figure 4.15), in which the 'Field width' column can be edited.

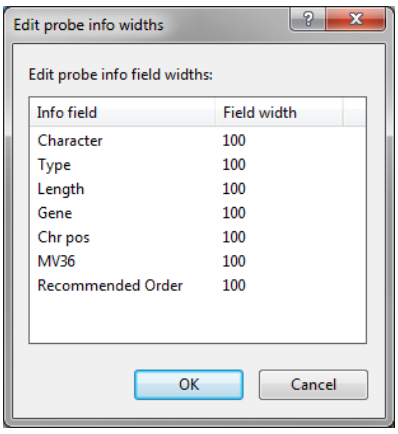


Figure 4.15: Edit probe info fields widths.

A colored bar for each probe, sized proportionally to the band quantification, can be shown or hidden by checking or unchecking ***Include quantification histogram***. Probe statistics, such as standard deviation and odds ratios can be displayed by checking ***Include probe statistics***.

The scoring and color of the histogram are taken from the scores character type, and reflect any manual overruling of the automatic assignment.

9.2 For an optimal display of the example data, highlight the 'MV36' probe info field and set its field width to "120".

9.3 Press <***OK***>. This shows a *MLPA report window* as depicted in Figure 4.16.

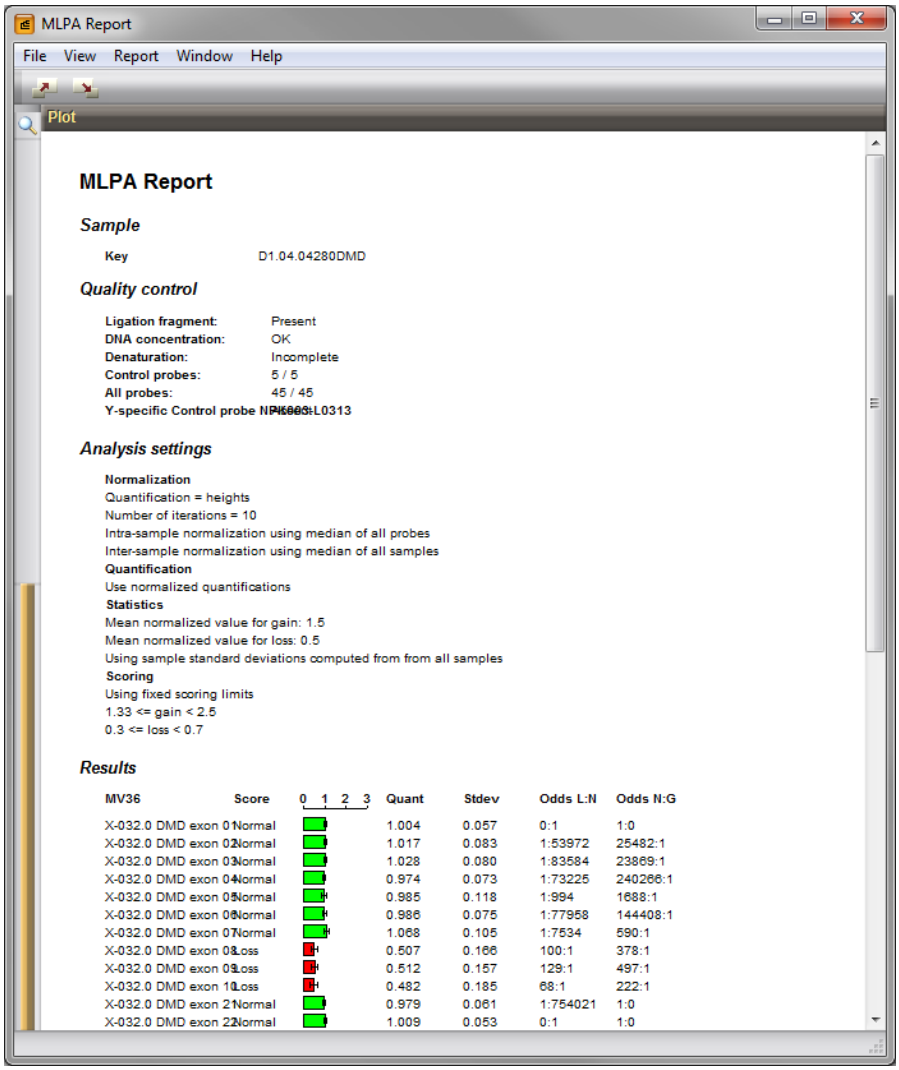


Figure 4.16: The *MLPA report window*.

9.4 To make any modification to the report settings, select ***Report*** > ***Settings***. This will call the *MLPA report settings* dialog box again (see Figure 4.14).

The report can be printed via ***File*** > ***Print*** or copied to the clipboard via ***File*** > ***Copy to clipboard***. From the clipboard, the copied information can then be pasted in applications such as Adobe Photoshop, Microsoft PowerPoint, Microsoft Word, etc.

Bibliography

[1] J.P. Schouten, C.J. McElgunn, R. Waaijer, D. Zwijnenburg, F. Diepvens, and G. Pals. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic acids research*, 30(12):e57, 2002.

APPLIED MATHS

A BIOMÉRIEUX COMPANY

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