BioNumerics®

MICROBIAL DATA ANALYSIS SOFTWARE

MLVA 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 *MLVA plugin* of BioNumerics. MLVA stands for Multi-Locus VNTR Analysis and the *MLVA plugin* allows the determination of VNTR copy numbers based on fingerprints, generated on capillary electrophoresis systems. See 2.1 for more information on MLVA principles and used terminology. One or more MLVA schemas can be set up per database and the plugin makes it possible to manage and assign MLVA types based on VNTR profiles.

The minimal BioNumerics configuration for the installation of the *MLVA plugin* includes the Fingerprint data module (import of sequence trace files or peak tables) and the Character data module (storage of fragment sizes and copy numbers). For clustering and population genetics, the Tree and Network inference module is also required.

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***>.

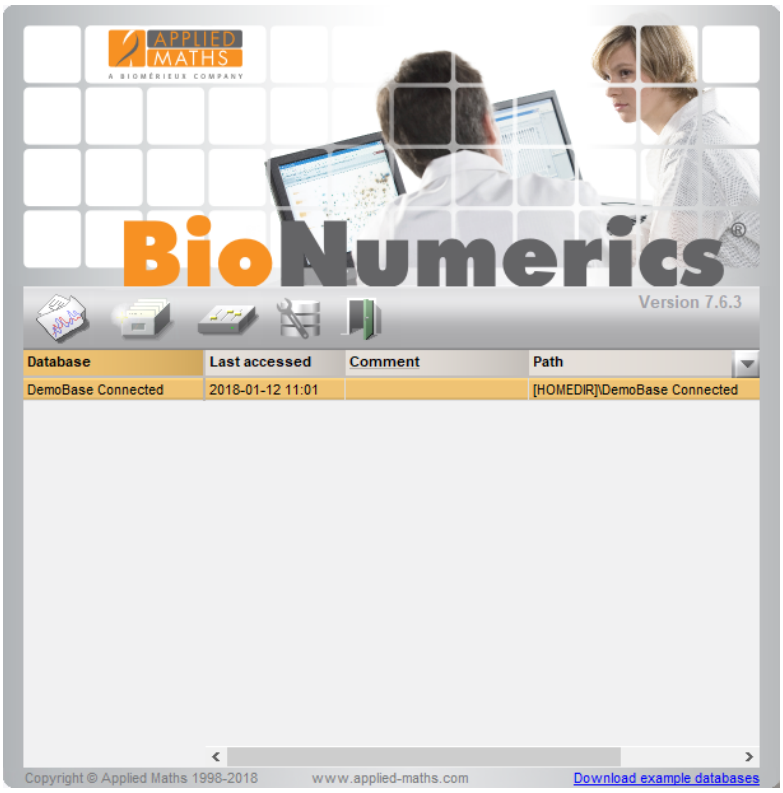


Figure 1.1: The *BioNumerics Startup* window.

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***>.

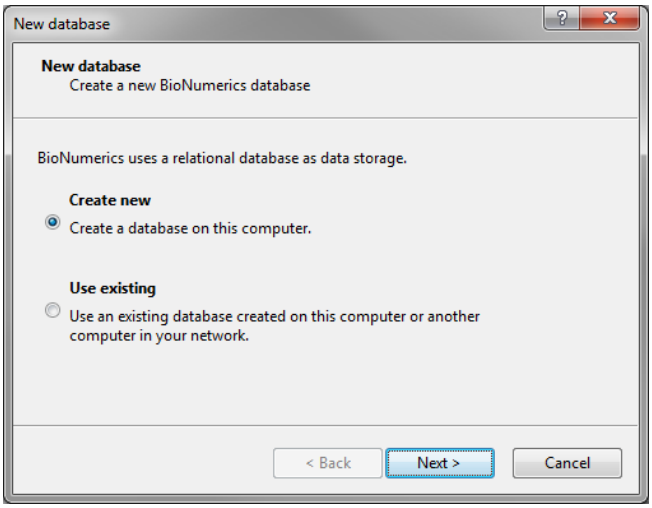


Figure 1.2: The *New database* wizard page.

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

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

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

The *Plugins* dialog box appears.

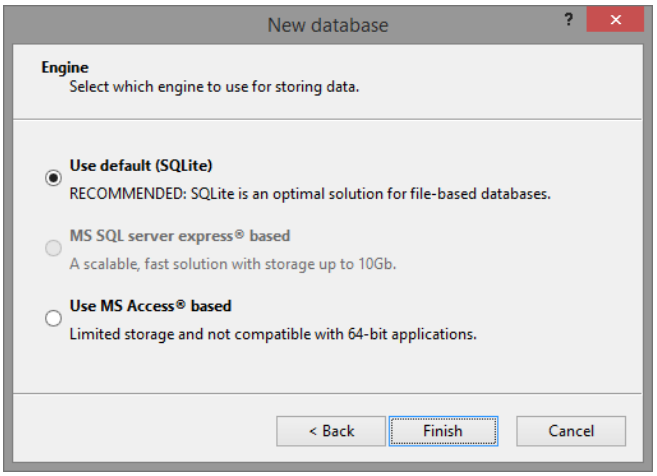


Figure 1.3: The *Database engine* wizard page.

1.4 Installing the MLVA 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* window by selecting ***File*** > ***Install / remove plugins...*** ( ).

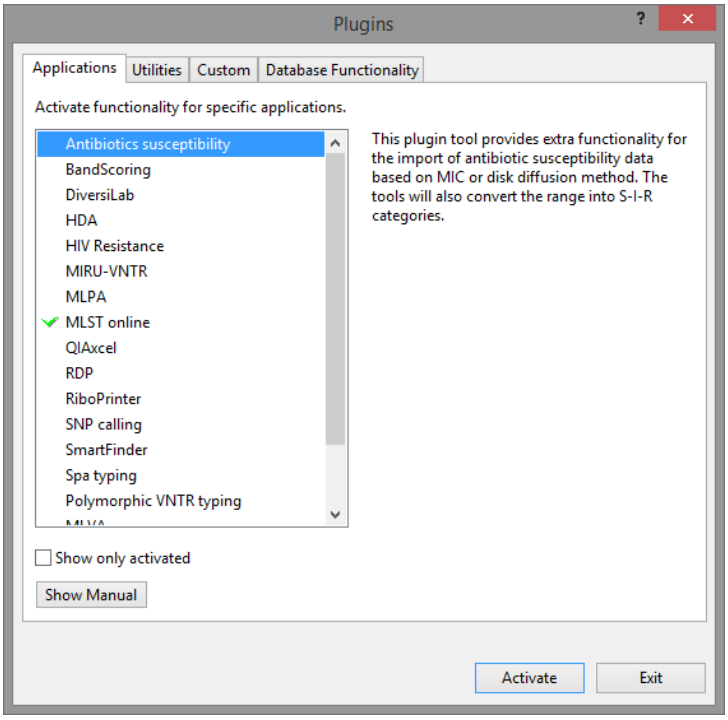


Figure 1.4: The *Plugins* dialog box.

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

The program will ask to confirm the installation of the plugin.

4.2 Press <***Yes***> to confirm the installation of the *MLVA plugin*.

A message appears, stating that the MLVA database components need to be updated.

4.3 Press <***Yes***> to continue.

Updating the MLVA database components will involve creating additional tables and columns in the database to hold MLVA-specific information. Therefore, the current user requires CREATE TABLE and ALTER TABLE privileges on the relational database. This will always be the case for a MS Access, MS SQL Server Express or SQLite database created by BioNumerics. For a shared database on a central database server, please check with your database administrator before proceeding.

When the database update is finished, the *MLVA plugin* is completely installed.

4.4 You can now close the *Plugins* dialog box by pressing the <***Exit***> button.

The *MLVA plugin* installs several menu items and buttons in the *Main* window (see Figure 1.5). In order to visualize the buttons, it might be necessary to select ***Window*** > ***Restore default configuration*** from the menu or to restart the software.

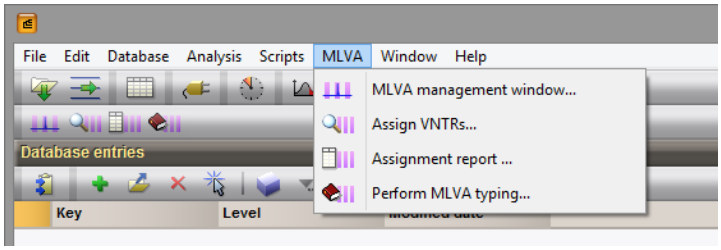


Figure 1.5: Menu and buttons generated by the *MLVA plugin* in the *Main* window.

Before the database can actually be used to determine VNTR copy numbers, an MLVA schema should be set up first (see 2).

Chapter 2

Setting up an MLVA schema

2.1 MLVA principles

A **VNTR (Variable Number of Tandem Repeats)** locus typically exhibits a large range of copy numbers, even among highly related bacterial strains. For a selected set of tandem repeats, comparison of the **copy numbers** between bacterial strains can be used to obtain insight about the relationships at a micro-evolutionary level.

In VNTR copy number analysis, a VNTR locus is PCR-amplified using conserved primers, located outside the tandem repeat. The size in base pairs (bp) of each PCR product therefore corresponds to the sum of the size of the tandem repeat plus the offsets at both ends (see Figure 2.1).

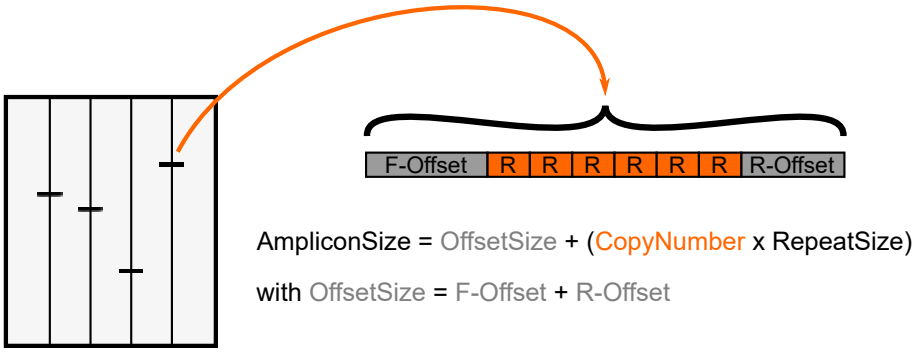


Figure 2.1: Calculation of repeat copy numbers from fragment sizes.

Knowing the repeat size, the copy number can theoretically be calculated as follows:

$$CopyNumber = \frac{AmpliconSize - OffsetSize}{RepeatSize}$$

Sizing of the obtained PCR products is typically done via capillary electrophoresis. In practice, observed metrics for a VNTR locus may differ from the theoretically calculated metrics. Defining a VNTR **mapping** is a method to deal with these size deviations. A mapping essentially consists of a series of VNTR **bins**, namely one bin for each possible VNTR copy number. A bin corresponds to a size range (defined by a minimum and maximum size), that the VNTR copy number is known to exhibit based on previous experience.

Obviously, sizing differences are expected to be larger between capillary electrophoresis equipment from different vendors or from separate product lines than between two apparatuses from the same type. Laboratories that own more than one type of automated sequencer can define a **machine type** for each type of

sequencer, not only to accommodate for the observed sizing differences but possibly also for different dye sets used.

In a typical **MLVA (Multi-Locus VNTR Analysis)** experimental setup, a number of VNTR loci (i.e. target VNTR regions on the genome) are selected that are sufficiently and complementary discriminatory for the organisms studied.

With **MLVA typing**, we mean the process of assigning types (typically numbers) to VNTR copy number profiles. Similar to MLST typing, this greatly improves the communication in scientific reports. An **MLVA typing schema** therefore consists of a list of types, with for each type the corresponding VNTR copy numbers for all VNTRs examined in the typing schema and optionally the clonal complex that the type belongs too.

An **MLVA schema** contains all the above information, i.e. theoretical VNTR definitions, mappings, machine types (if defined) and MLVA typing schemas. The *MLVA plugin* allows you to create more than one MLVA schema, which is very useful e.g. in case the database contains strains belonging to more than one species.

2.2 The MLVA management window

In the *Main* window, select ***MLVA > MLVA management window...*** (). Initially, when there is no MLVA schema present yet, the *Import schemas* dialog box will pop up (see Figure 2.2).

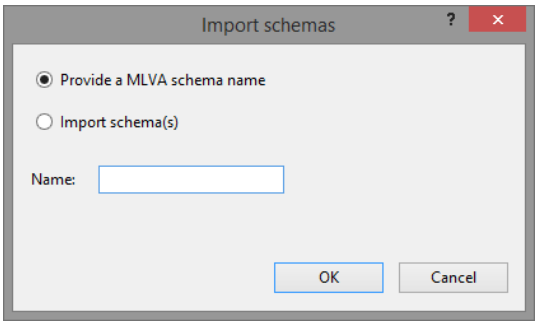


Figure 2.2: The *Import schemas* dialog box.

There are two options for adding a new MLVA schema:

- ***Provide a MLVA schema name***: This option allows you to create a new MLVA schema from scratch. Enter a ***Name*** for the MLVA schema and define the schema in the *MLVA management* window.
- ***Import schema(s)***: Import one or more schemas that were defined in another BioNumerics database, either by yourself or by somebody else. Check this option and browse for the XML file containing the MLVA schema. This file can be located on your local hard drive, on a network drive or on the internet. In the latter case, an URL should be provided. See 2.7 for more information about exporting an MLVA schema to an XML file. Check ***Create fingerprint types*** to let the software create the fingerprint type experiments that are associated with the VNTRs as defined in the schema. This is the recommended option when importing an existing MLVA schema in an empty database. If you already imported one or more capillary electrophoresis runs, uncheck ***Create fingerprint types*** and link the VNTRs manually to the corresponding fingerprint type.

When the <***OK***> button is pressed, the character types **[SCHEMA_NAME]_vals** and **[SCHEMA_NAME]_frags** are created automatically. They will be used, respectively, to store the VNTR copy numbers and the molecular sizes (in bp) for each VNTR defined in the schema.

Next, the *MLVA management* window will open. Figure 2.3 shows the *MLVA management* window when only a MLVA schema name was provided.

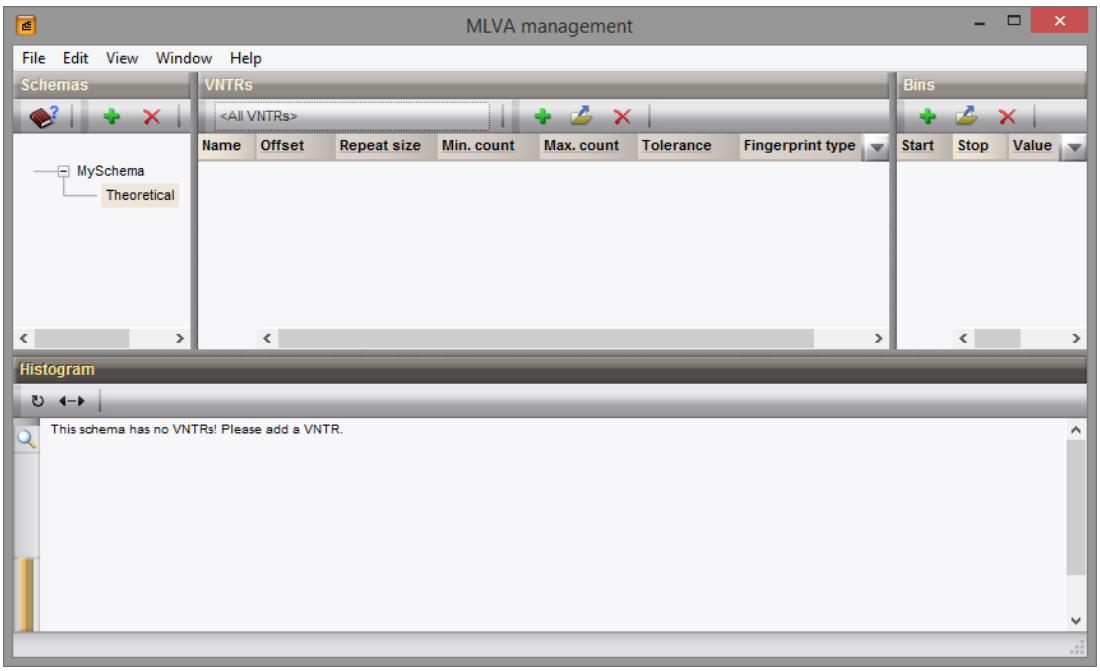


Figure 2.3: The *MLVA management* window with an empty MLVA schema "MySchema" created, ready to start defining VNTRs.

The *MLVA management* window consists of four dockable panels:

- The *Schemas* panel displays a tree-like overview with all MLVA schemas defined in this database and their corresponding mappings, typing schemas and instrument types (if defined). Each schema has at least one mapping ('Theoretical') that corresponds to the theoretical prediction of the VNTR bins based on the VNTR definitions (see 2.3).
- In the *VNTRs* panel, all defined VNTRs for the highlighted schema in the *Schemas* panel are shown in a grid with their Name, Offset, Repeat size, Minimum count, Maximum count, Tolerance and the associated Fingerprint type.
- The *Bins* panel shows the VNTR bins for the highlighted VNTR in the *VNTRs* panel in a grid with their Start and Stop metrics and the Value (i.e. the copy number).
- Electropherograms for the selected database entries and the corresponding VNTR are plotted *Histogram* panel, against a backdrop of the VNTR bins for the highlighted VNTR.

2.3 VNTR definitions

The copy number of a VNTR can be theoretically calculated from its metrics values when its offset and repeat size are known (see 2.1).

To add a VNTR to a schema, select ***Edit*** > ***VNTRs*** > ***Add VNTR...*** ( ). This action calls the *Add VNTR* dialog box (see Figure 2.4).

This dialog box prompts for following parameters:

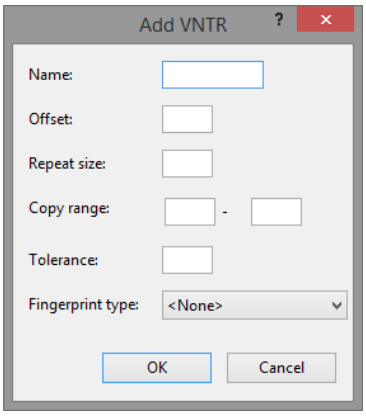


Figure 2.4: The *Add VNTR* dialog box, prompting for all parameters needed to define a VNTR.

- ***Name***: This name should be unique and will be used for further reference. For each VNTR, a character with the same name will be automatically created in the **[SCHEMA_NAME]_vals** and **[SCHEMA_NAME]_frags** character types.

- ***Offset***: Each fragment consists of a repeat portion and a constant portion due to the fact that the primers do not occur exactly at the start and the end of the repeat region. This parameter specifies the size (in base pairs) of the constant portion.

- ***Repeat size***: This parameter specifies the size of a unit repeat block (in base pairs).

- ***Copy range***: Specifies the minimum and maximum number of copies that the plugin will consider during the copy number determination. Within the same lane and the same dye, more than one VNTR can be loaded if there is a sufficient difference in fragment size. Imposing a limit on the copy range allows the software to distinguish which band corresponds to which VNTR.

- ***Tolerance***: This parameter specifies the maximum difference in base pairs between the expected fragment length and the actual length.

- ***Fingerprint type***: This drop-down list contains all fingerprint types defined in the database and is used to indicate in which fingerprint type this VNTR was run. The fingerprint type is determined by the dye and – if several pools are used – by the pool tag (see 3).

Pressing <***OK***> will add the VNTR to the schema. The VNTR will be listed in the *VNTRs* panel of the *MLVA management* window and the bins that were calculated based on the theoretical values provided are shown in the *Bins* panel.

A VNTR can be edited at any time by highlighting it in the *VNTRs* panel and selecting ***Edit*** > ***VNTRs*** > ***Edit VNTR...*** (). This action opens the *Edit VNTR* dialog box for that VNTR.

This dialog offers the same set of parameters as the *Add VNTR* dialog box does. However, once a VNTR is defined, its ***Name*** cannot be edited anymore.

As an alternative to using the *Edit VNTR* dialog box, the VNTRs can be edited directly in the *VNTRs* panel by clicking twice on a cell in this grid panel. Please note that the associated fingerprint type cannot be edited this way.

A VNTR can be removed from a schema by highlighting it and selecting ***Edit*** > ***VNTRs*** > ***Remove VNTR*** (). The software will ask for confirmation before actually deleting the VNTR.

2.4 Mapping definitions

2.4.1 Creating a mapping

In practice, often small deviations from the theoretical metrics are observed. To accommodate for this, the user can define one or more custom *mappings* in an MLVA schema. Such a mapping is essentially a set of VNTR bin definitions, i.e. for each possible VNTR copy number a minimum and a maximum size in base pairs.

To create a new mapping for the highlighted schema in the *Schemas* panel, select ***Edit*** > ***Mappings*** > ***Add mapping...*** (). The *Add mapping* dialog box pops up (see Figure 2.5).

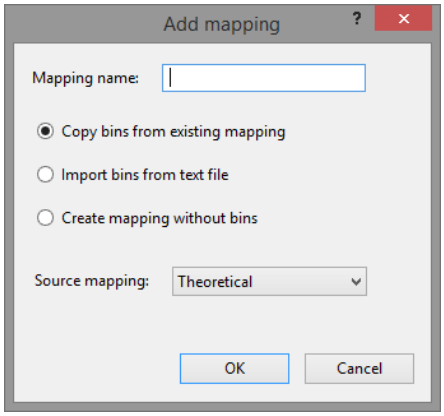


Figure 2.5: The *Add mapping* dialog box.

A unique ***Mapping name*** should be provided for the new mapping.

Three options are available to create a new mapping:

- ***Copy bins from existing mapping***: With this option selected, a ***Source mapping*** can be selected from a drop-down list, from which the VNTR bins are copied. This copy can then be the starting point for a new mapping by editing the VNTR bins (see 2.4.2).
- ***Import bins from text file***: This option allows you to browse for a ***Mapping file***, from where the bin information is imported. See 2.4.3 for more information on how to create a ***Mapping file***.
- ***Create mapping without bins***: This option simply creates a new mapping, without any VNTR bin definitions. Since the bins should be defined manually one by one (see 2.4.2), this is generally not the recommended option.

Pressing <***OK***> will add the new mapping to the MLVA schema. It will be shown in the tree overview in the *Schemas* panel of the *MLVA management* window. When ***Copy bins from existing mapping*** or ***Import bins from text file*** was selected, the bins from the existing mapping or the file, respectively, will be shown in the *Bins* panel for the highlighted VNTR in the *VNTRs* panel.

2.4.2 Editing a mapping

A mapping is edited by adding, deleting or editing the VNTR bin definitions. A bin is added to the highlighted mapping in the *Schemas* panel by selecting ***Edit*** > ***Bins*** > ***Add bin...*** (). The *Add bin* dialog box pops up (see Figure 2.6).

Theoretical mappings cannot be edited and a warning message will be displayed when trying to add, delete or edit a VNTR bin definition. The bins in a theoretical mapping are solely determined by the VNTR definitions (Offset, Repeat size, Tolerance, etc.; see 2.3).

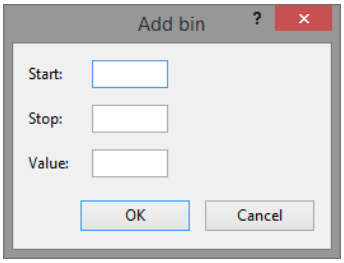


Figure 2.6: The *Add bin* dialog box to add a VNTR bin to a mapping.

Start and ***Stop*** are respectively the minimum and the maximum metrics value (in base pairs) for the fragment to be assigned to this VNTR bin. Values containing decimal digits (e.g. "174.27") are allowed.

Value corresponds to the name of this bin. Typically, this will be an integer (i.e. the VNTR copy number), but other values are allowed as well (e.g. "4s").

When the <***OK***> is pressed, BioNumerics will check the entered values (***Start*** and ***Stop*** should be decimal numbers, with ***Stop*** larger than ***Start***, the bin should not overlap with any existing bin and the entered ***Value*** should not correspond to an existing bin) and add the bin to the mapping if OK.

For each bin created, a corresponding character mapping with the same name as the bin's ***Value*** will be created in the **[SCHEMA_NAME]_vals** character type.

A bin can be deleted from a mapping with ***Edit*** > ***Bins*** > ***Remove bin*** ( ). The software will ask for confirmation before actually deleting the VNTR bin.

There are basically two ways to edit an existing bin:

1. By editing the ***Start***, ***Stop*** and ***Value*** for the bin, much in the same way as when creating a new bin.
2. Graphically, by drag-and-drop of the bins in the *Histogram* panel.

For the first method, highlight the bin in the *Bins* panel and select ***Edit*** > ***Bins*** > ***Edit bin...*** ( ). This action displays the *Edit bin* dialog box for the highlighted bin (see Figure 2.7).

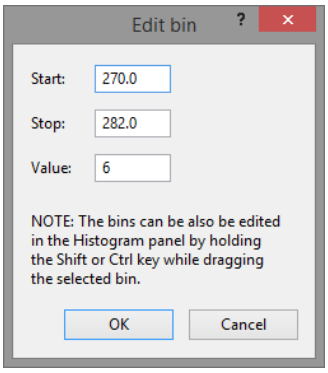


Figure 2.7: The *Edit bin* dialog box to edit a VNTR bin.

This dialog box allows to edit the ***Start***, ***Stop*** and ***Value*** for the bin in the same way as described for the *Add bin* dialog box (see higher).

As an alternative to using this dialog box, the values can be edited via drag-and-drop in the *Bins* panel.

For editing VNTR bins in a graphical fashion, fingerprint curves should be displayed in the *Histogram* panel (see Figure 2.8). Hence, electropherograms should be imported first (see 3) and fingerprint types should be assigned to the VNTRs in the VNTR definitions (see 2.3).

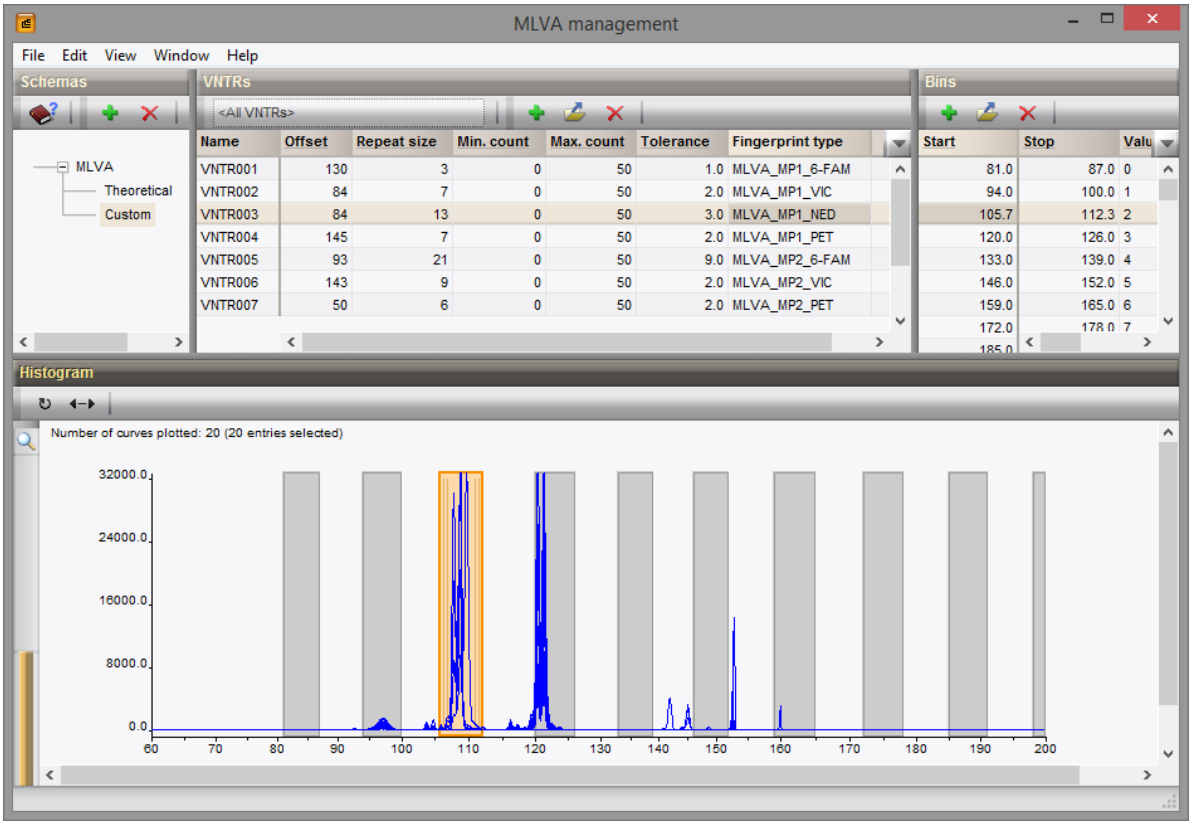


Figure 2.8: The *MLVA management* window with fingerprint curves and VNTR bins displayed in the *Histogram* panel.

In the *Histogram* panel, click on the bin which needs to be edited. The bin will now be highlighted in orange. To set the Start value of a bin, **Ctrl+click** on the *left* side of the bin and drag it to the desired position. Similarly, to set the Stop value of a bin, **Ctrl+click** on the *right* side of the bin and drag it to the desired position. To reposition a bin without changing its width, use **Shift+click**.

2.4.3 Exporting a mapping

An existing mapping can be exported with ***Edit*** > ***Mappings*** > ***Export mapping...***. The resulting `export.csv` file will open in your system's default CSV editor (often MS Excel). This file can be used in the *Add mapping* dialog box to import a new mapping from (see 2.4.1).

2.5 Instrument types

Only in cases where the same MLVA experiments are run on different types of capillary electrophoresis equipment (possibly also using a different dye set), it will be necessary to define more than one *instrument type*.

Select ***Edit*** > ***Instrument types*** > ***Add instrument type...*** to open the *Add instrument type* dialog box (see Figure 2.9).

Enter the ***Instrument type name*** and press <***OK***>.

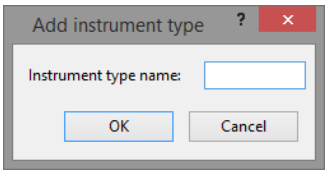


Figure 2.9: The *Add instrument type* dialog box.

As soon as an instrument type is added, the ***Instrument type name*** will be displayed between brackets next to each mapping as a reminder that these are specific for a certain type of automated sequencer (see Figure 2.10).

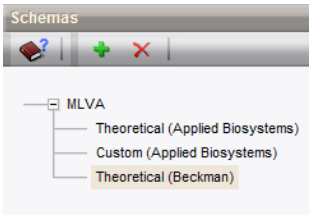


Figure 2.10: Detail of the *Schemas* panel, showing an MLVA schema with two instrument types.

When a second, third, etc. instrument type is added, only the theoretical VNTR definitions are taken over, since custom mappings are apparatus-specific and will need to be redefined anyways. Furthermore, since most likely the fingerprint data from the new instrument type will be stored in a different set of fingerprint types, the VNTRs are not linked yet to any fingerprint type.

To rename an existing instrument type, select ***Edit*** > ***Instrument types*** > ***Rename instrument type...***. The *Rename instrument type* dialog box pops up (see Figure 2.11).

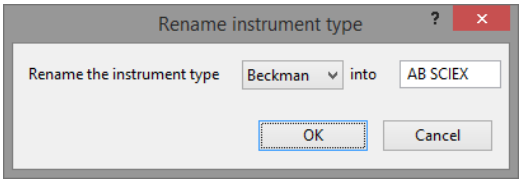


Figure 2.11: The *Rename instrument type* dialog box.

The instrument type to be renamed can be selected from the drop-down list and the new name should be provided in the text field. Pressing <***OK***> will rename the instrument type.

An instrument type can be deleted by highlighting one of the mappings that belong to the instrument type in the *Schemas* panel and selecting ***Edit*** > ***Instrument types*** > ***Remove instrument type***. The software will ask for confirmation before removing the instrument type.

2.6 MLVA typing

To create a new MLVA typing schema, select ***Edit*** > ***Typing schemas*** > ***Add typing schema...***. This action opens the *Add typing schema* dialog box (see Figure 2.12).

This dialog box prompts for a ***Name*** for the MLVA typing schema. In the ***VNTRs*** list below, all VNTRs from the MLVA schema are listed. Using **Ctrl+click** and **Shift+click**, those VNTRs on which the typing should be performed, can be highlighted in the list.

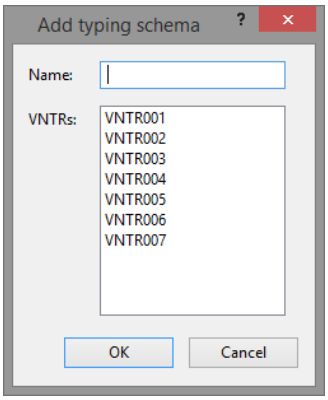


Figure 2.12: The *Add typing schema* dialog box.

Pressing <***OK***> will create an MLVA typing schema with the provided name. It will be available from the drop-down list in the header of the *VNTRs* panel.

At this stage, the MLVA typing schema is created but it consists only of a subset of the VNTRs from the MLVA schema, i.e. no actual typing information is available yet.

To edit the MLVA typing schema and specify the additional information, select ***File*** > ***Typing...*** (). This opens the *Select typing* dialog box (see Figure 2.13).

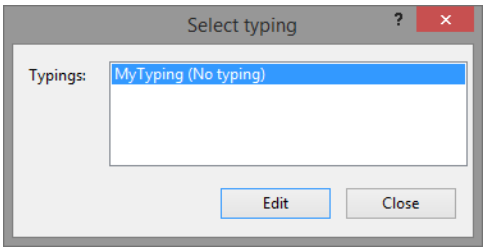


Figure 2.13: The *Select typing* dialog box.

Highlight the MLVA typing schema that you want to edit in the list and press <***Edit***>. This action will display the *Typing settings* dialog box (see Figure 2.14).

For MLVA typing schemas that lack typing information, the text 'No typing' is mentioned between brackets next to the typing schema name. This is the case for the **MyTyping** in Figure 2.13.

In the upper part of the dialog box, the ***Database information*** for the MLVA typing should be specified:

- From the drop-down list next to ***Type information field***, the entry information field to store the MLVA type in needs to be provided. The <***Create new***> option allows you to create a new entry information field, for which you can specify a name in the *Create new information field* dialog box when the <***OK***> button is pressed.
- Specifying an entry information field to hold the clonal complex information (***CC information field***) is optional; select <***None***> if no clonal complex information is available.
- The ***Qualifier for unknown types*** is the text that will be displayed in the ***Type information field*** when the MLVA profile does not correspond to a known type in the database. The default value is "Unknown".

Two options are available for providing the typing data, i.e. the information about which combination of

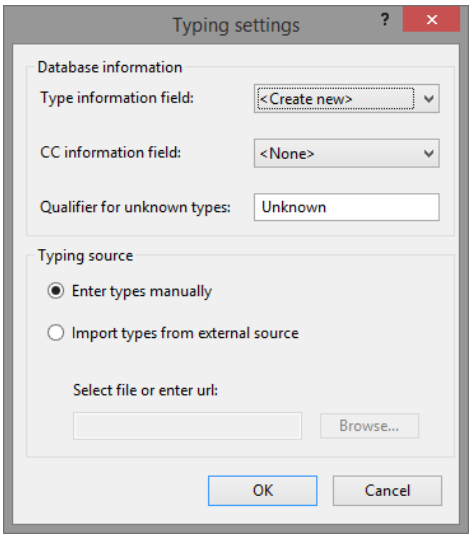


Figure 2.14: The *Typing settings* dialog box.

VNTR copy numbers (for VNTRs included in the typing schema) corresponds to which MLVA type (under ***Typing source***):

- ***Enter types manually***: Choose this option to enter the type and VNTR profile for each type individually in the *Typing management* window.
- The option ***Import types from external source*** allows you to import a CSV or tab-delimited text file containing the typing data, either from a local or network drive or from a URL.

The required format for ***Import types from external source*** is a comma or tab-separated file with a header consisting of "Type", a list of all VNTRs included in the typing schema and "Clonal complex". Each row represents an MLVA type, with the corresponding VNTR copy numbers and optionally the clonal complex. Such a file can be exported from the *Typing management* window by clicking the column properties button and selecting ***Save content to file*** (see the Reference manual, Chapter Database objects for more information).

If the <***OK***> button is pressed with ***Import types from external source*** selected, the types from the external file will be imported and displayed in the *Typing management* window (see Figure 2.15). Alternatively, if ***Enter types manually*** was selected, the software will ask "Do you want to scan the selected entries for types?". If VNTR copy numbers were determined already (see 4), the software will scan the selected entries for new VNTR types and propose to add these to the database (see further for the description of the *Update* dialog box). Next, the *Typing management* window will open (see Figure 2.15).

The *Typing management* window provides an overview of the types that are currently present in the MLVA typing schema, with the clonal complex they belong to and the corresponding copy numbers for each of the VNTRs that are included in the typing schema.

The settings for this MLVA typing schema can be accessed via ***File*** > ***Settings...***. This action displays the *Typing settings* dialog box, as discussed earlier.

A new MLVA type can be manually added with ***Types*** > ***Add type...*** ( ). This opens the *Add new type* dialog box (see Figure 2.16).

In this dialog, you can enter the ***Type***, ***Clonal complex*** and copy numbers of each of the VNTRs (under ***Profile***). Pressing <***OK***> will add the MLVA type to the list in the *Typing management* window.

A highlighted type in the *Typing management* window can be edited with ***Types*** > ***Edit type...*** ( ). This action will open the *Edit type* dialog box (see Figure 2.17).

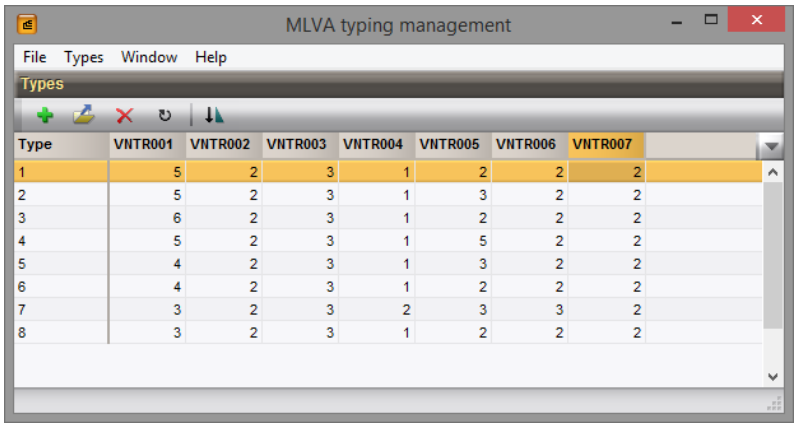


Figure 2.15: The *Typing management* window.

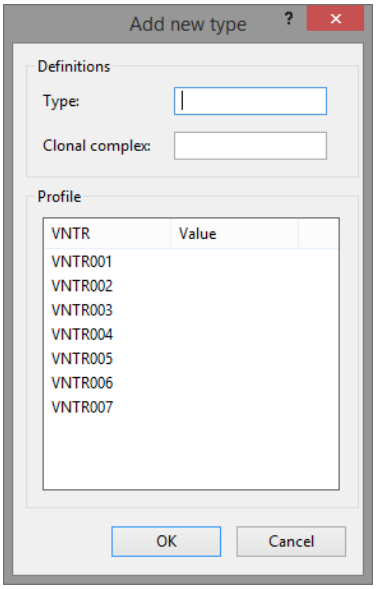


Figure 2.16: The *Add new type* dialog box, to manually add a type to the MLVA typing schema.

This dialog allows you to edit the ***Type***, ***Clonal complex*** and copy numbers of each of the VNTRs (under ***Profile***).

A type can be deleted by highlighting it in the *Typing management* window and selecting ***Types*** > ***Remove type*** ( ). The software will ask for confirmation before actually removing the type from the database.

The types in the *Typing management* window can be sorted on any column by first highlighting the column and then selecting ***Types*** > ***Sort types***.

The action that is triggered by ***Types*** > ***Update types*** ( ) depends on whether the VNTR profiles are stored locally (option ***Enter types manually*** in the *Typing settings* dialog box) or when they are taken from an external source (option ***Import types from external source*** in the *Typing settings* dialog box):

- With **locally stored VNTR profiles**, selecting ***Types*** > ***Update types*** ( ) will check for new types on the currently selected entries. When no entry selection is present, the software will offer to scan the complete database. In case entries are found with MLVA types that are not yet defined in the MLVA typing schema, the new types will be reported in the *Update* dialog box (see Figure 2.18).
- When **VNTR profiles are stored in an external file**, selecting ***Types*** > ***Update types*** ( ) will update

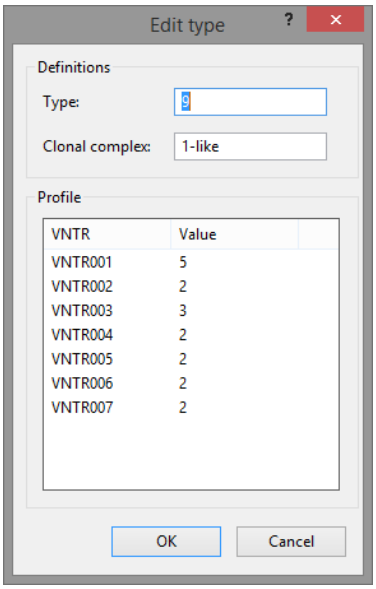


Figure 2.17: The *Edit type* dialog box.

the BioNumerics database with the profiles from the external source.

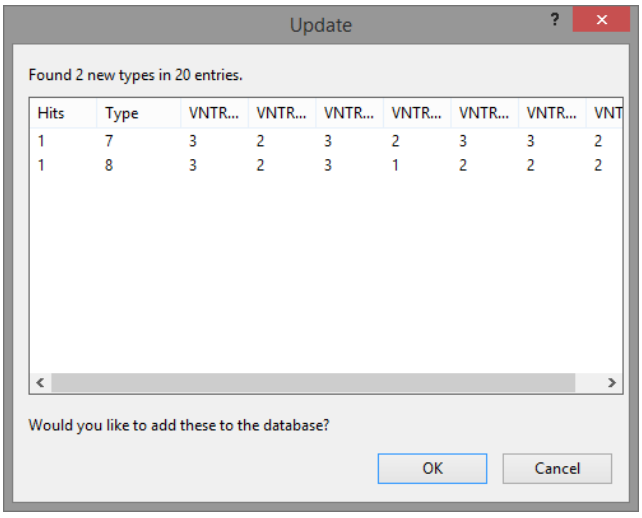


Figure 2.18: The *Update* dialog box.

This dialog box will report any new types that were found among the selected entries. Pressing <***OK***> will add these types to the MLVA typing schema. Pressing <***Cancel***> will keep the MLVA typing schema in its current state.

A complete MLVA typing schema can be removed by selecting it from the drop-down list in the toolbar of the *VNTRs* panel and using ***Edit*** > ***Typing schemas*** > ***Remove typing schema***. The software will ask for confirmation before actually deleting the MLVA typing schema.

2.7 Exporting and importing an MLVA schema

A complete MLVA schema, including VNTR definitions, mappings, instrument types and MLVA typing schemas can be exported with ***File*** > ***Export to XML...***. This action opens the *Export schema* dialog box

(see Figure 2.19).

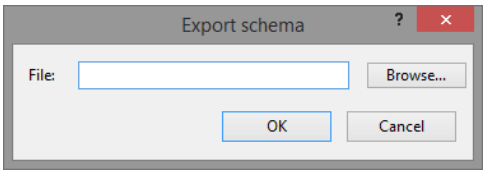


Figure 2.19: The *Export schema* dialog box.

This dialog allows you to enter (or browse for) a file name under which to save the XML file containing the MLVA schema definitions.

An XML file containing an MLVA schema, created as described above, can be imported in another database via ***File*** > ***Import from XML...***. The *Import schemas* dialog box pops up (see Figure 2.20).

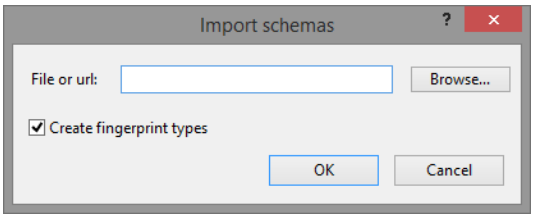


Figure 2.20: The *Import schemas* dialog box.

In the ***File or URL*** text field, enter or browse for the XML file containing the MLVA schema definitions. If the file is located on a web server, enter the URL that points to this file.

Check ***Create fingerprint types*** to let the software create the fingerprint type experiments that are associated with the VNTRs as defined in the schema. This is the recommended option when importing an existing MLVA schema in an empty database. If you already imported one or more capillary electrophoresis runs, uncheck ***Create fingerprint types*** and link the VNTRs manually to the corresponding fingerprint type.

Press <***OK***> to import the MLVA schema definitions from the XML file. If an MLVA schema with the same name already exists in the database, the software will ask whether or not to overwrite the schema.

Chapter 3

Importing and processing capillary electrophoresis data

3.1 Pooling strategies

VNTR amplification products are usually analyzed on automated sequencers using capillary electrophoresis. These instruments allow different VNTRs to be *pooled* using different color dyes and run together on the same capillary column. A *pool* (sometimes called *panel*) is a mixture of usually 4 or 5 dyes, one being a *reference sample* for fragment size calculation, the others containing each a VNTR for a given strain or sample. The pooling can happen through multiplex PCR or the amplification products can be mixed after PCR. The number of PCR amplicons that can be pooled together depends on (1) the number of color dyes used: if 5 color dyes are used, 4 differently labeled fragments can be pooled (one dye contains the reference sample), and (2) the possibility to combine PCR-amplicons with significantly different lengths.

Figure 3.1 shows a typical setup where 4 PCR products (*target VNTRs*) are mixed in one pool, using 5 color dyes. Since in this example in total 8 VNTR targets are examined per sample, two pools are generated, each containing 4 target PCR products.

For economy reasons, several VNTRs are sometimes marked with the same dye and loaded as a mixture in the same column of a capillary sequencer. A condition is that the mixed VNTR PCR products have size ranges that do not overlap. Using the above example, an alternative set-up is given in Figure 3.2, where PCR has been carefully designed so that each time a short and a long amplicon can be labeled with the same dye and mixed in the same pool.

Thus, each VNTR in the MLVA experimental setup is defined by a **pool**, a **dye** and a **size range**. The size range is defined by the repeat length, the offset and the copy number range (see 2.1). Note that the size range is only important in case different VNTRs are marked with the same dye.

3.2 Data formats

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

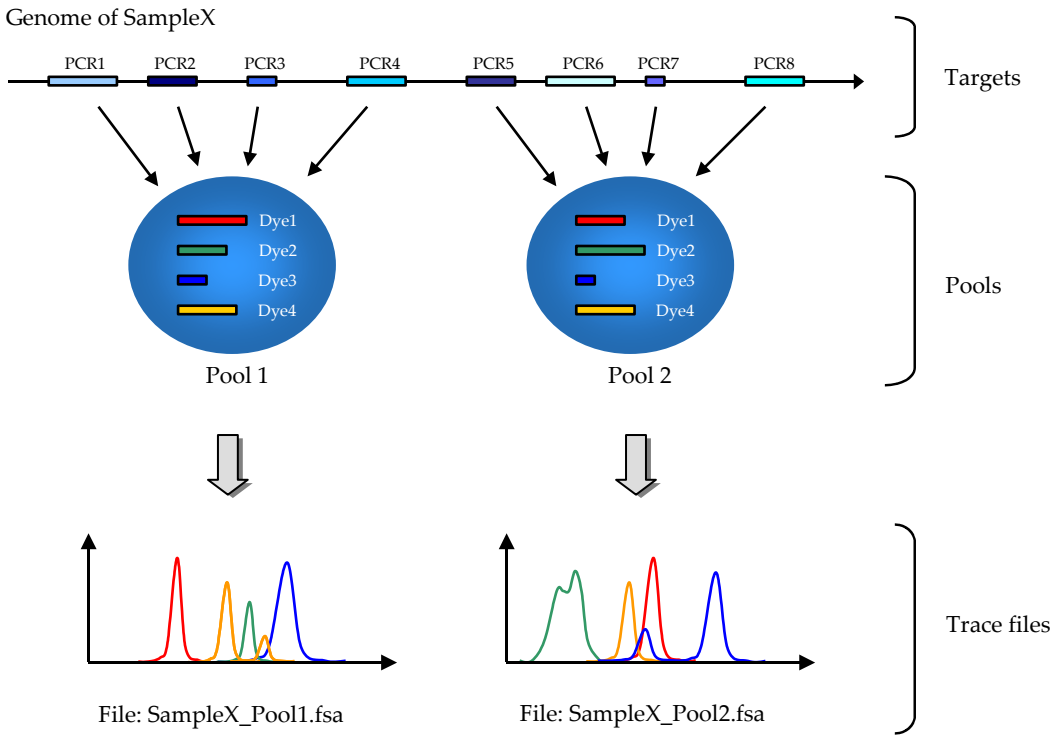


Figure 3.1: Schematic overview of the relation between VNTR loci (= PCR targets), pools and curve files in a typical MLVA experimental setup. In this example, only one target gene is used per dye in the same pool.

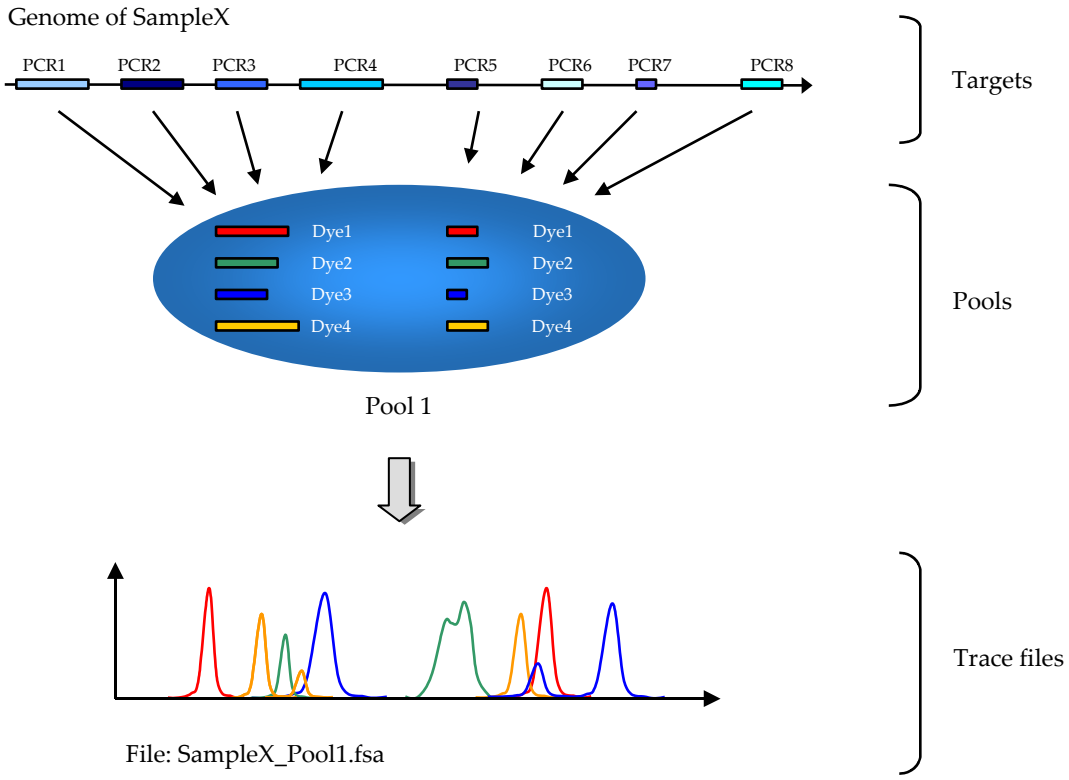


Figure 3.2: Schematic overview of the relation between VNTR loci (= PCR targets), pools and curve files in an alternative MLVA experimental setup. In this example, two compatible targets (i.e. with different lengths) are labeled with the same dye and mixed in one pool.

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.

Both formats can be imported into BioNumerics using the import routines available in the *Import* dialog box. The *Import* dialog box is called with the command ***File*** > ***Import...*** (, **Ctrl+I**).

- Curve files (i.e. electropherograms) can be imported in the database with the option ***Import curves*** listed under ***Fingerprint type data*** in the import tree.
- Peak files can be imported in the database with the option ***Import peak table*** listed under ***Fingerprint type data*** in the import tree.

Detailed information on each of these import routines can be found in the Reference manual, Chapter Setting up fingerprint type experiments.

The following chapter illustrates the import and processing procedures using example data sets for curve files (3.3.1) and peak tables (3.3.2).

3.3 Importing and processing sequencer data: examples

3.3.1 Importing and processing curve files

3.3.1.1 Sample data

Raw curve files from Applied Biosystems and Beckman (now AB SCIEX) sequencers can be imported directly in BioNumerics. A batch of Applied Biosystems curve files can be downloaded from the Applied Maths website (`http://www.applied-maths.com/download/sample-data`, click on "VNTR sequencer trace files"). These example files will be used to illustrate the import steps in this tutorial. Eight VNTR targets per sample were amplified: two pools were generated (**MP1** and **MP2**), each containing four PCR products, using four color dyes (**6-FAM**, **VIC**, **NED**, and **PET**).

3.3.1.2 Importing curves

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

3.2 Browse to the downloaded and unzipped example data folder `VNTR sequencer trace files` and select all Applied Biosystems curves files (extension `.fsa`). Press <***Open***>.

The files are displayed in the *Input* wizard page and the default suggested ***Fingerprint file name*** is the folder name.

3.3 Change the name to **Batch1** and press <***Next***> (see Figure 3.3).

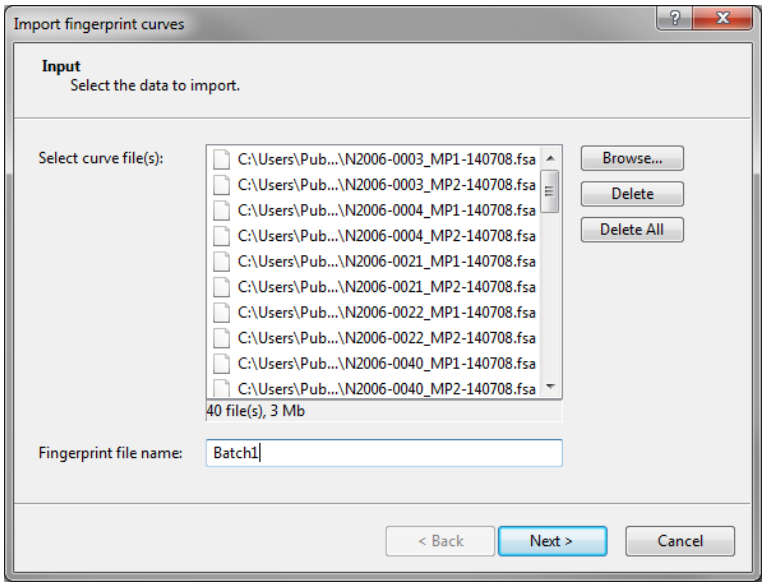


Figure 3.3: Selected curve files and fingerprint file name.

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 sample and pool information can be parsed from the file names. A new import template needs to be defined:

3.4 Press the <***Create new***> button to call the *Import rules* dialog box (see Figure 3.4).

The *Import rules* dialog box 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>).

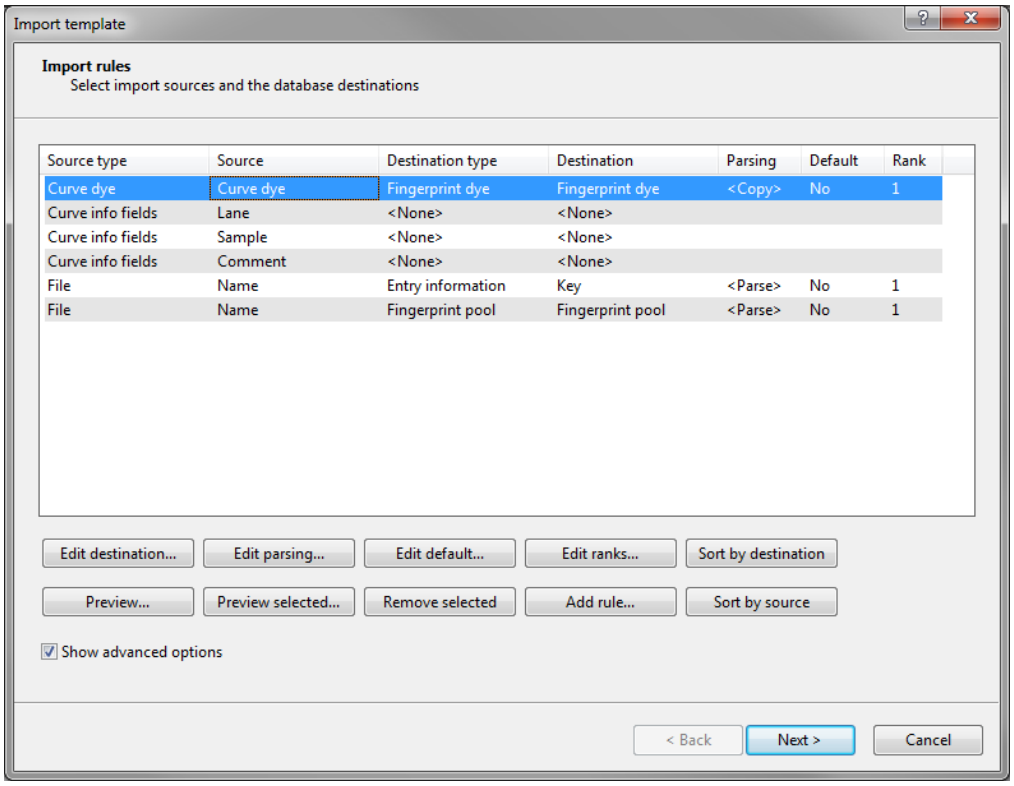


Figure 3.4: Import rules.

3.5 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.

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

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

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

3.9 Press the <***Preview***> button and press <***OK***> when the parsing is correct (see Figure 3.5).

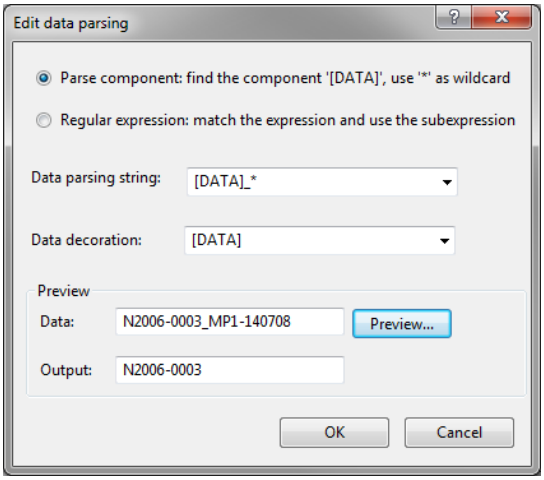


Figure 3.5: Parsing string.

Next, we will specify a new rule that links the part of the file name appearing after the underscore and before the hyphen ("-") to the ***Pool*** field.

3.10 Select <***Add rule***>, select ***Name*** under ***File*** as data source, press <***Next***>, and select ***Fingerprint pool*** as data destination. Press <***Next***> once more.

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

3.12 Press the <***Preview***> button and press <***Next***> and <***Finish***> when the parsing is correct (see Figure 3.6).

The *Import rules* dialog box should now look like Figure 3.4.

3.13 Press the <***Preview***> button to check the outcome of all defined import rules (see Figure 3.7).

3.14 Close the preview and press <***Next***> to go to the next step.

In the example data set, the LIZ channel contains the size standard (GeneScan 500 LIZ).

3.15 Make sure ***LIZ*** is selected as ***Reference dye*** and press <***Next***> and <***Finish***>.

3.16 Specify a template name, e.g. **Import VNTR curve files** and press <***OK***>.

3.17 Make sure <***Create new***> is selected as fingerprint type experiment, select the new template and press <***Next***>.

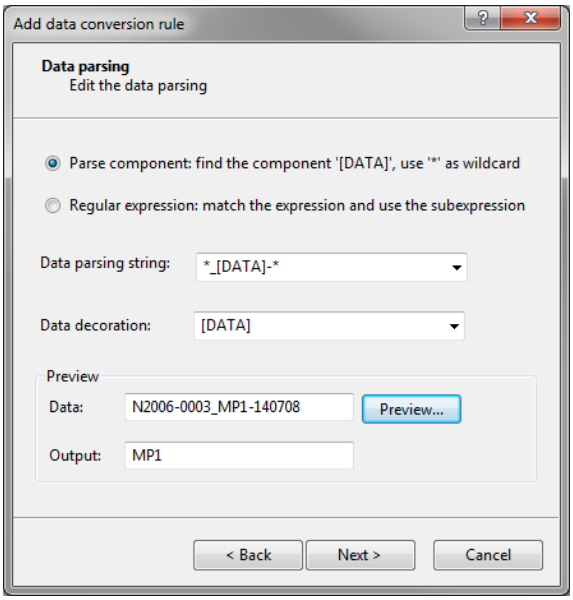


Figure 3.6: Parsing string.

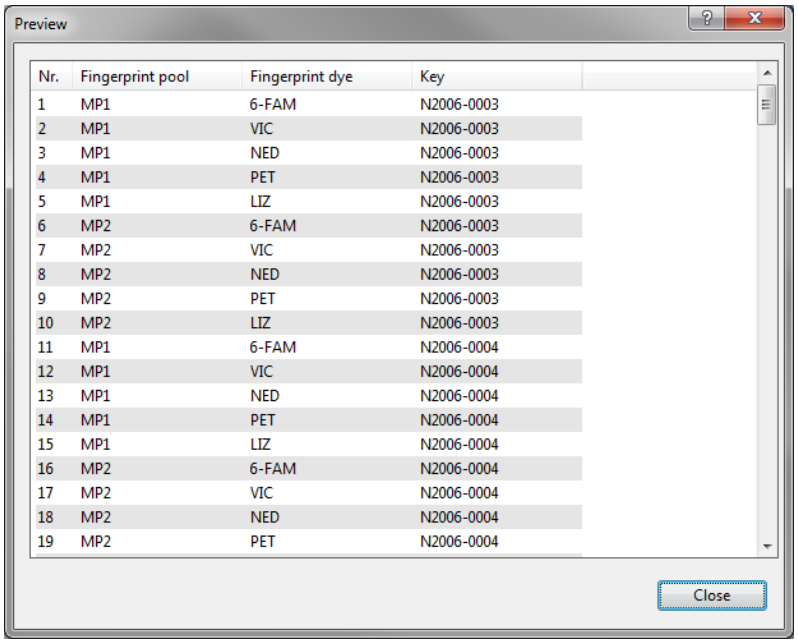


Figure 3.7: Preview of the import rules.

3.18 Specify a name for the new base fingerprint type experiment (e.g. **MLVA**) and press <***OK***>. Confirm the creation of the new experiment in the database.

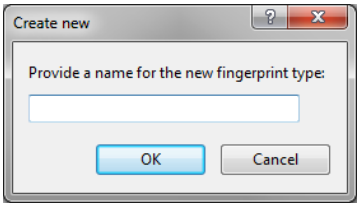


Figure 3.8: New base fingerprint type experiment.

A fingerprint type needs to be present in the database for each pool and dye combination. The names of these fingerprint types are composed of the base fingerprint type name, followed by the pool name, and the name of the dye. If one or more of these fingerprint types are not present in the database, a new dialog box pops up, listing the missing fingerprint types (see Figure 3.9).

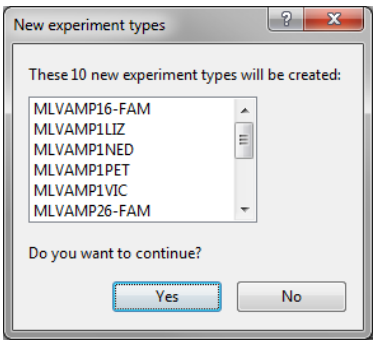


Figure 3.9: Missing fingerprint experiments.

3.19 Confirm the creation of the missing fingerprint type experiments.

3.20 Press <***Next***> to confirm the creation of new entries in the database.

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

For each dye checked in the *Dyes panel* of the *Import data* dialog box, a new fingerprint file is created, composed of the file name specified and the name of the dye (e.g. Batch1_LIZ). These files are displayed in the *Fingerprint files* panel. The reference file is shown in the **Link** column. Double-clicking on a fingerprint file opens the *Fingerprint* window. If lane information was imported with the individual lanes, this information is displayed in the *Fingerprint information panel*.

The imported fingerprint lanes are linked to new entries in the database. The lanes are linked to the corresponding fingerprint "dye" type. The names of these fingerprint types are composed of the base fingerprint type name, followed by the pool name, and the name of the dye. The fingerprint type experiments are displayed in the *Experiment types* panel.

Entries for which fingerprint data was imported are selected in the database.

After data import, the *Main* window looks as in Figure 3.10.

3.3.1.3 Processing curves

When the option ***Open curve preprocessing window*** was checked in the last step of the import routine, the *Fingerprint curve processing* window opens when pressing the <***Finish***> button. All channels from the run are automatically loaded and displayed in the *Fingerprint curve processing* window.

The *Fingerprint curve processing* window can also be called from the *Main* window by highlighting one of the channels in the *Fingerprint files* panel and selecting ***Open fingerprint data...*** (). Alternatively, you can first open the *Fingerprint* window with ***Edit*** > ***Open highlighted object...*** (, **Enter**) and then select ***File*** > ***Edit fingerprint data...*** ().

3.22 Click on the icon left of the data channels in the *Channels* panel.

The data channels are now hidden from the view and its icons are displayed as .

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

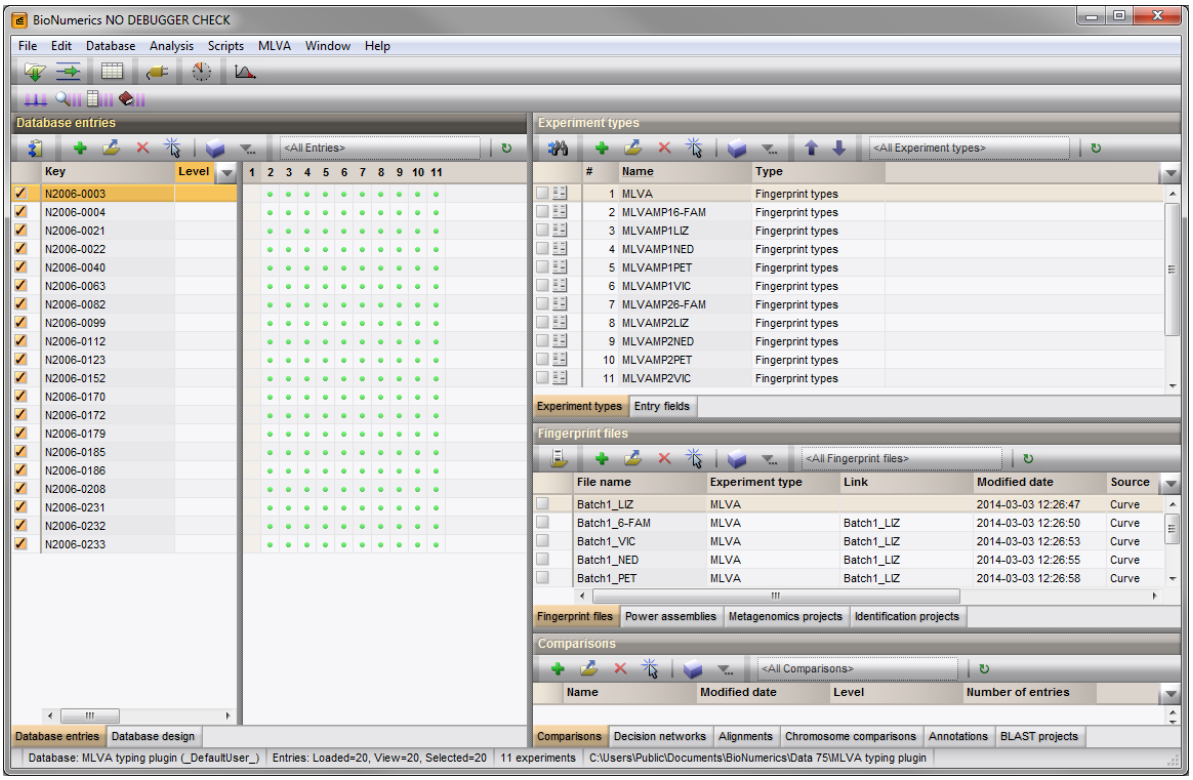


Figure 3.10: The main window after import of the data.

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.

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

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

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

The bands that fall within the specified criteria are marked with a solid line at the band's position (see Figure 3.11).

3.27 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 3.12).

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

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

3.29 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.

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

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

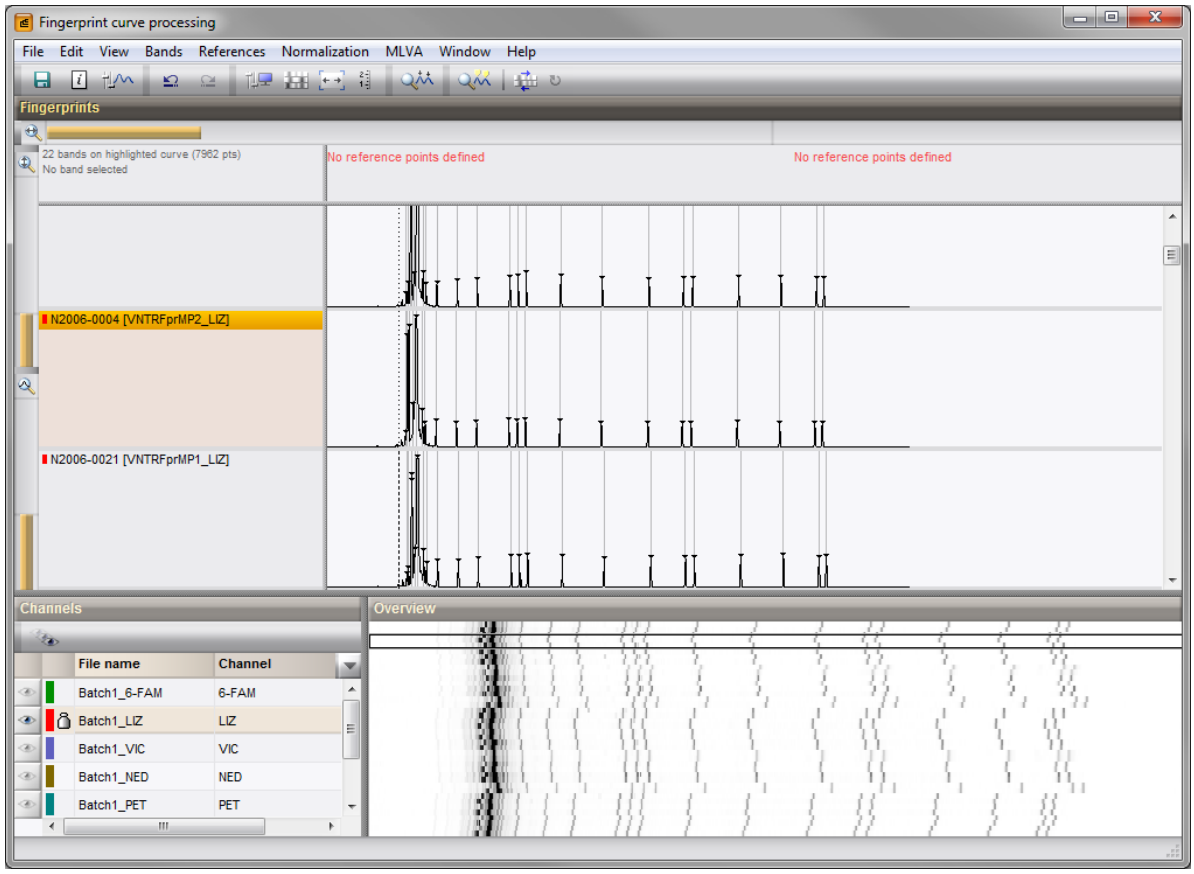


Figure 3.11: The *Fingerprint curve processing* window only displaying the reference dye.

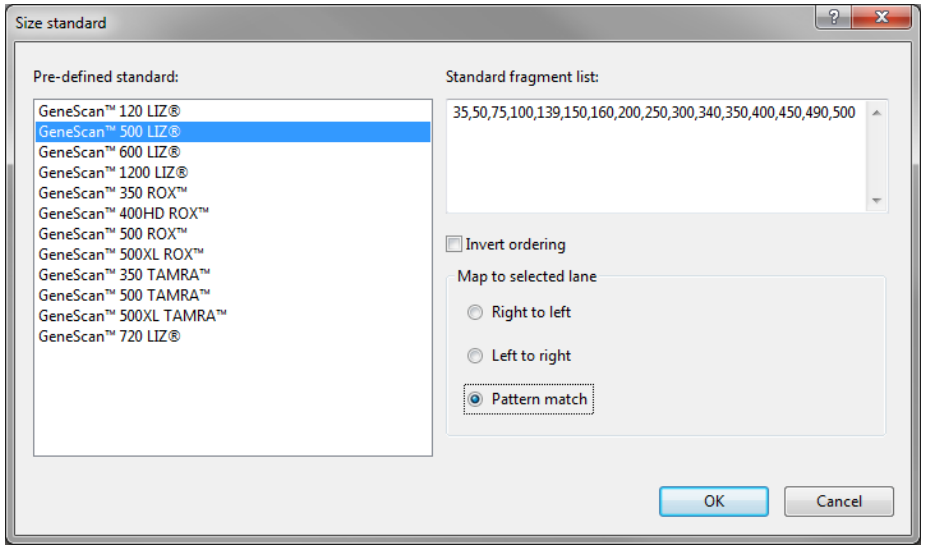


Figure 3.12: Select marker from the list.

3.31 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**).

3.32 Click on the icon left of the data channels in the *Channels* panel.

3.33 Click on the icon left of the LIZ reference channel.

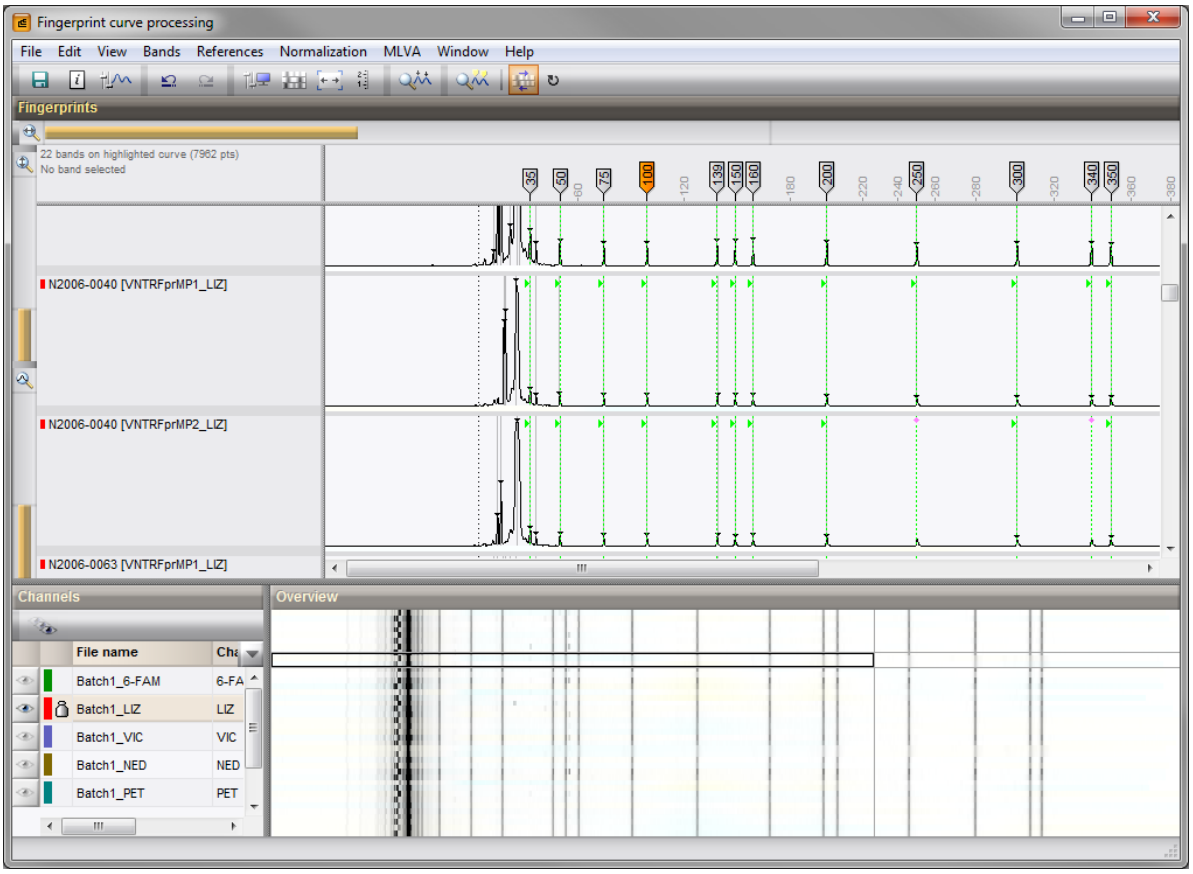


Figure 3.13: Normalized view - reference dye.

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

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

3.35 Check ***Remove doublets***, ***Remove shadow bands***, ***Filter by fragment length*** and specify a minimum length of 50. Press <***OK***>.

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

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. We can check this from the *Main* window:

3.37 Double-click on the base fingerprint type in the *Experiments* panel to open the *Fingerprint type* window.

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

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

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

3.3.2 Importing peak data from peak table files

Text files containing a listing of peaks can be imported in BioNumerics. An example peak file can be downloaded from the Applied Maths website (http://www.applied-maths.com/download/sample-data,

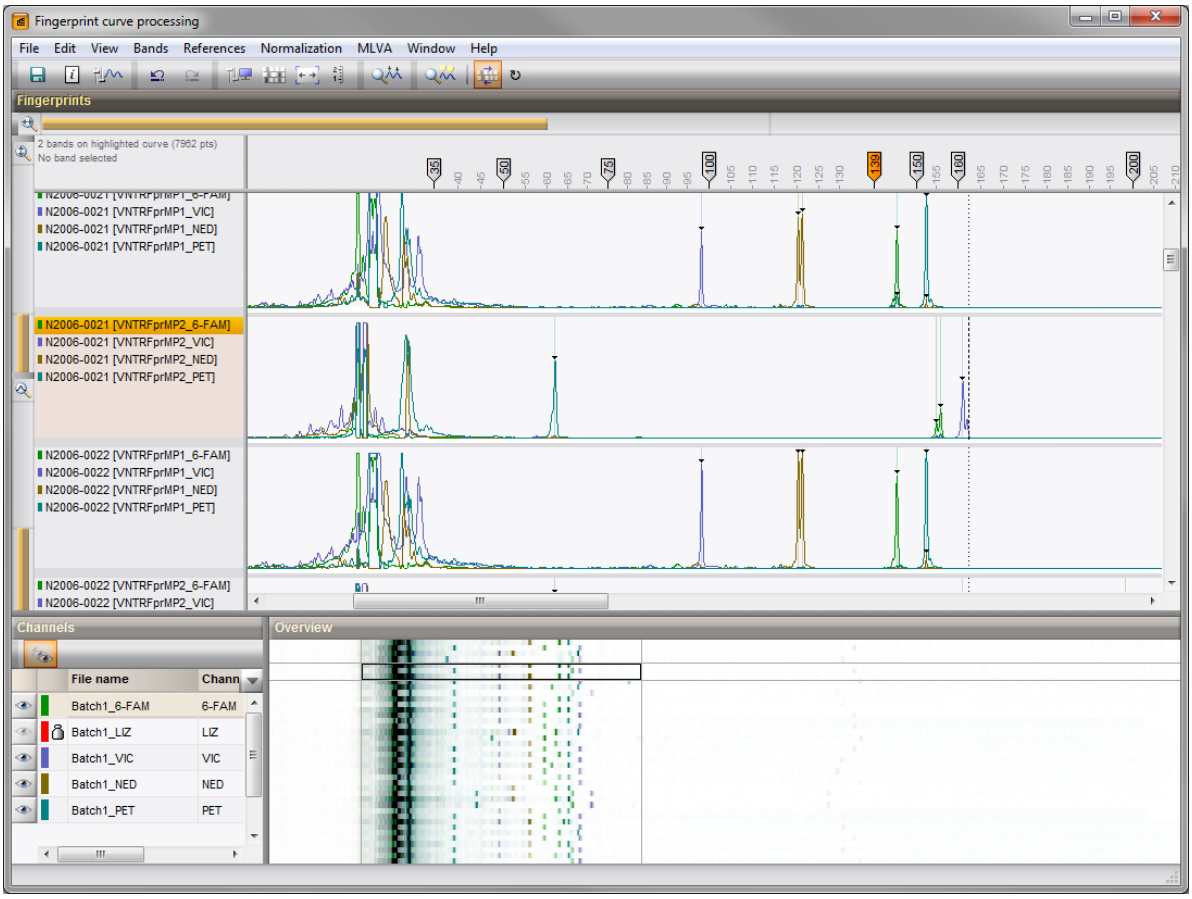


Figure 3.14: Normalized view - data dyes.

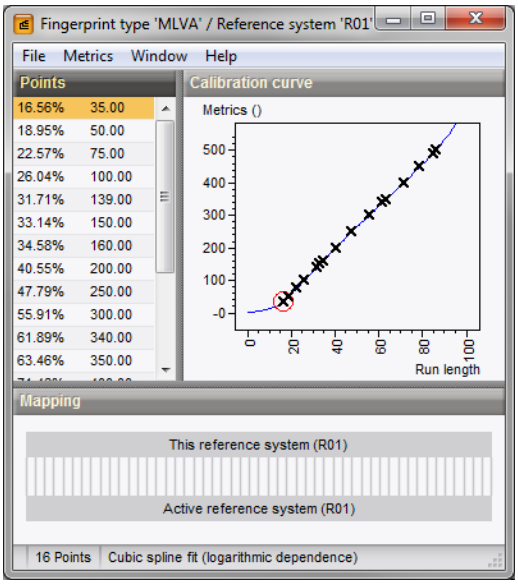


Figure 3.15: The *Fingerprint Reference system* window, with a calibration curve displayed.

click on ”VNTR sample peak table”). This example file will be used to illustrate the import steps in this tutorial.

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

3.41 Browse to the downloaded and unzipped example data folder `VNTR Peak table` and select the `Beckman_Sample.txt` sample file. Press <***Open***>.

The path is displayed in the *Input* wizard page.

3.42 Check ***Import as fingerprint file***, specify a name (e.g. **Beckman1**) and press <***Next***>.

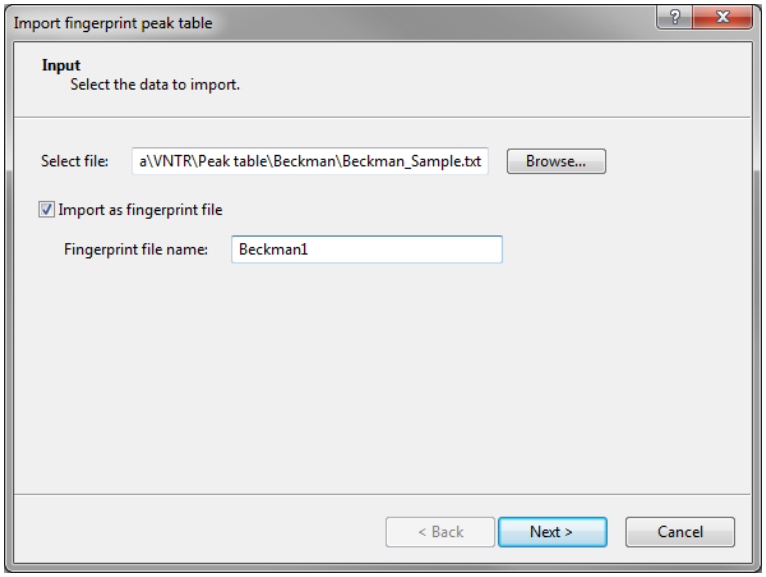


Figure 3.16: Selected peak table file.

The way the information should be imported in the database can be specified with an import template. In the example Beckman peak file, the dye, sample and pool information is provided.

3.43 Select the predefined template **Beckman with pools** and press <***Edit***> to call the *Import rules* dialog box.

The *Import rules* dialog box lists the import rules defined for the import template **Beckman with pools**.

3.44 Press the <***Preview***> button to check the parsing of the file information based on the rules defined for the template.

From the preview (Figure 3.17), it can be seen that all information from the example file is parsed correctly.

3.45 Close the preview and press <***Next***> twice and <***Finish***>.

3.46 Make sure ***Create new*** is selected as base fingerprint type experiment, select the **Beckman with pools** template and press <***Next***>.

3.47 Specify a name for the new base fingerprint type experiment (e.g. **MLVA**) and press <***OK***>. Confirm the creation of the new experiment in the database.

Since a new fingerprint type experiment is created and added to the database, the *Experiment settings* wizard page pops up prompting for some settings.

3.48 For this exercise, enter an ***Intensity range*** of "65536" (= 16-bit), a ***Min. fragment length*** of "10", and a ***Max. fragment length*** of "700" (see Figure 3.19). Press <***OK***>.

A fingerprint type needs to be present in the database for each pool and dye combination. The names of these fingerprint types are composed of the base fingerprint type name, followed by the pool name, and the name of the dye. A new dialog box pops up, listing all missing fingerprint types (see Figure 3.20).

3.49 Confirm the creation of the missing fingerprint type experiments.

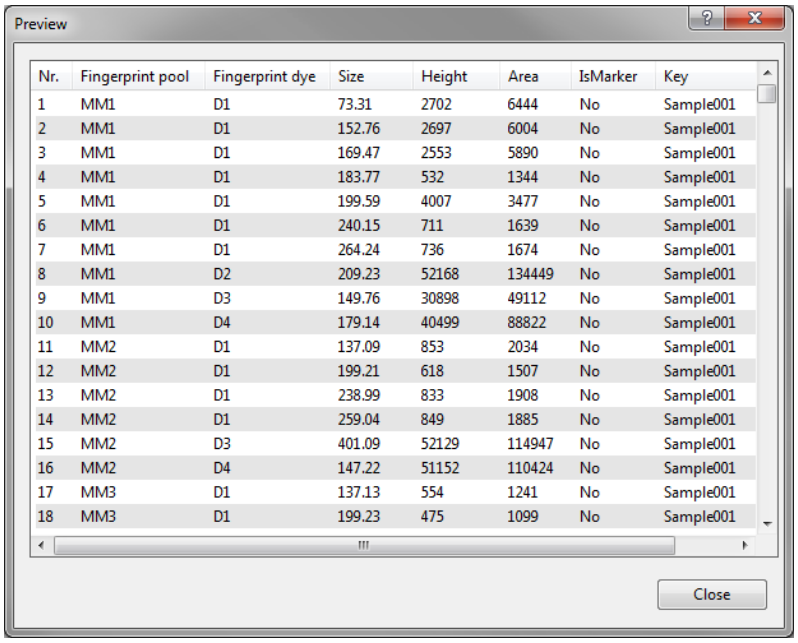


Figure 3.17: Preview.

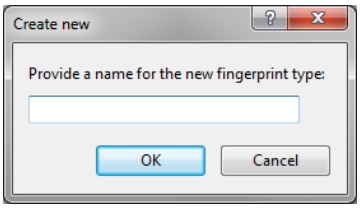


Figure 3.18: New base fingerprint type experiment.

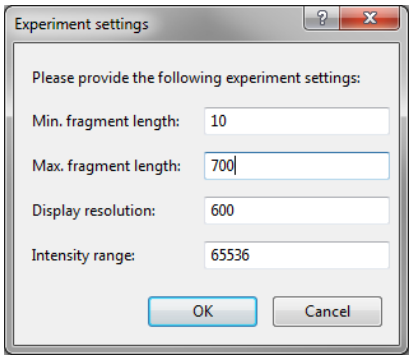


Figure 3.19: Experiment settings.

3.50 Press <***Finish***> to confirm the creation of new entries in the database.

For each import dye, a new fingerprint file is created, composed of the file name specified and the name of the dye (e.g. Beckman1_D1). These files are displayed in the *Fingerprint files* panel.

BioNumerics reads the band positions from the mapped "SIZE" column, the peak heights from the mapped "HEIGHT" column, the area information from the mapped "AREA" column and generates densitometric curves using this information. The imported fingerprint lanes are linked to new entries in the database. The lanes are linked to the corresponding fingerprint "dye" type. The names of these fingerprint types are composed of the base fingerprint type name, followed by the pool name, and the name of the dye. The

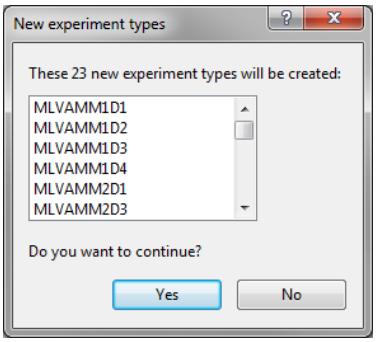


Figure 3.20: Missing fingerprint types.

fingerprint type experiments are displayed in the *Experiment types* panel.

If no reference system has been specified for the base fingerprint type, BioNumerics creates for all missing fingerprint types and the base fingerprint type, a linear reference system between the user-defined ***Minimum*** and ***Maximum fragment length*** positions, and copies the reference system to a calibration system.

Entries for which fingerprint data was imported are selected in the database.

After data import, the *Main* window looks as in Figure 3.21.

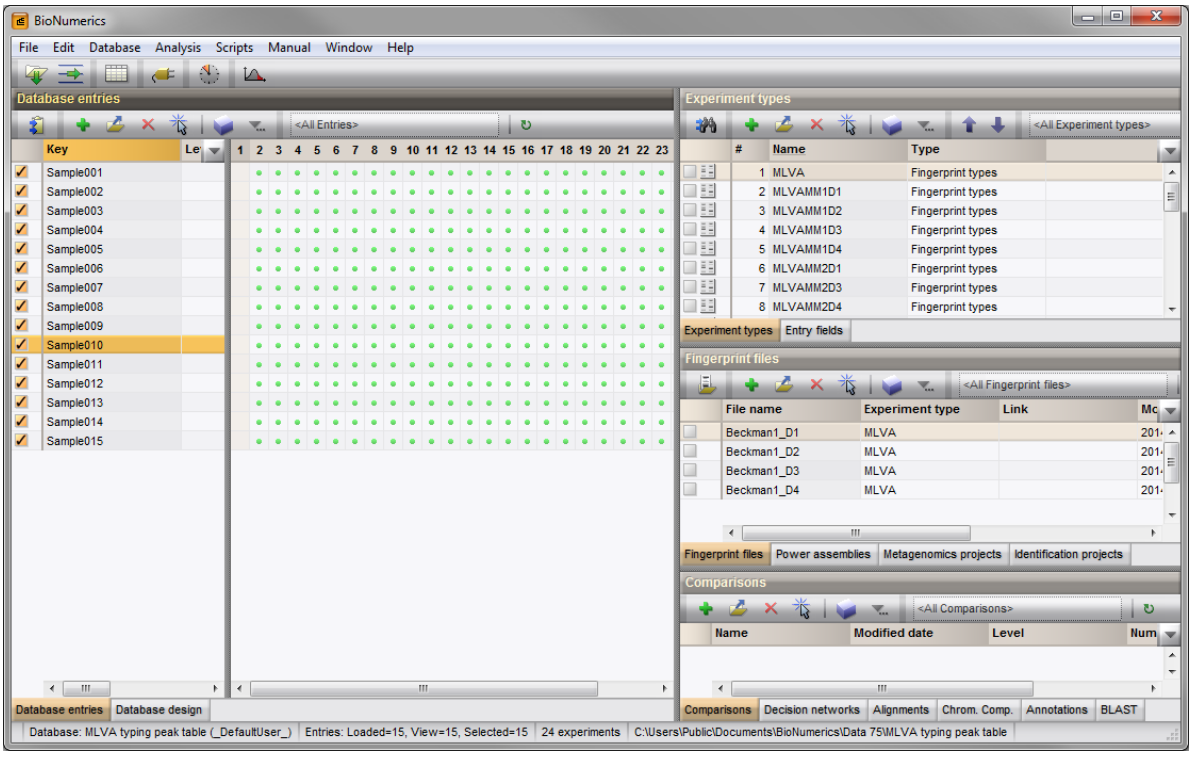


Figure 3.21: The *Main* window after import of the data.

Chapter 4

Calculating and assigning VNTR copy numbers

4.1 Automatic assignment

VNTR copy numbers can be determined for selected entries in the database.

1.1 Select the entries in the database for which you want to calculate the copy numbers.

1.2 Select ***MLVA*** > ***Assign VNTRs...*** ().

The *Assign VNTRs* dialog box pops up, listing all VNTRs defined in the database (see Figure 4.1).

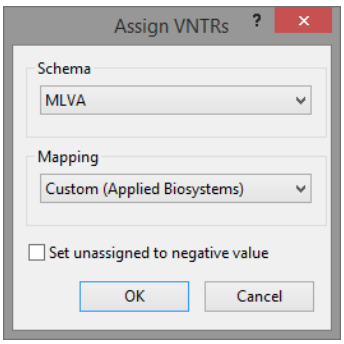


Figure 4.1: The *Assign VNTRs* dialog box, to determine VNTR copy numbers for the selected entries in the database.

This dialog box allows you to select an MLVA ***Schema***, since more than one MLVA schema may be defined in the database. It also allows you to select a ***Mapping*** from any of the mappings (theoretical or custom mappings) defined in the selected MLVA schema.

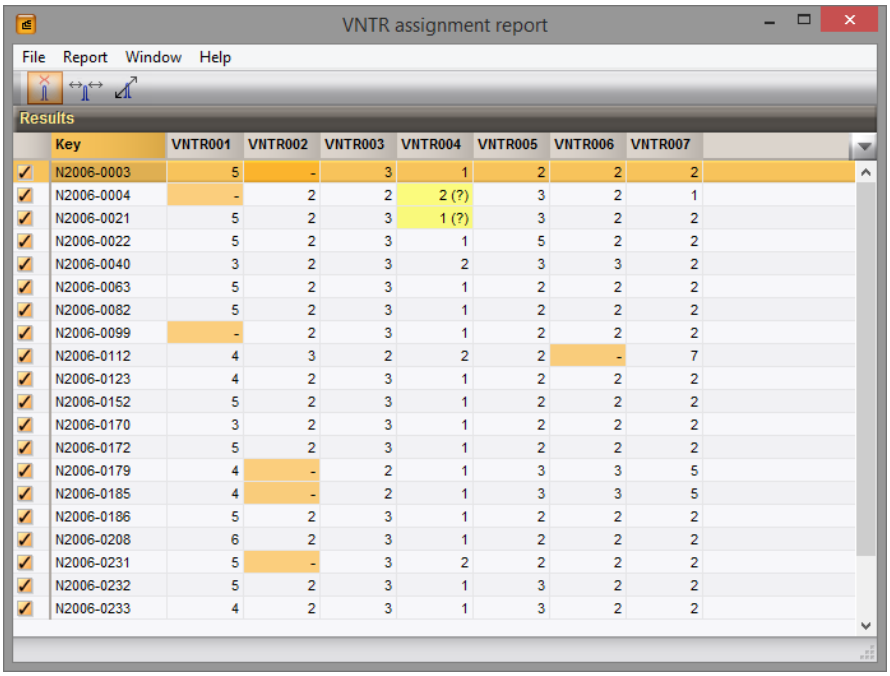
Enable the option ***Set unassigned to negative value*** to score the absence of a VNTR as negative, rather than absent. All absent VNTRs will then be scored with the character value "-2" in the character type **[SCHEMA_NAME]_vals**. The absent VNTRs, all having the score "-2", will be considered as a separate category for each VNTR when clustering the character information in the **[SCHEMA_NAME]_vals** experiment using the categorical coefficient.

Press <***OK***> to look for VNTRs in the fingerprints of the selected entries.

When the assignment is complete, the *VNTR reporting* window will pop up (see 4.2).

4.2 Assignment report

After assignment of the VNTRs in the database, the *VNTR reporting* window is displayed (see Figure 4.2). This window basically contains a table, listing all selected entries (keys) as rows and all VNTRs from the MLVA schema as columns. Note that this interactive report window can be called at any time from the *Main* window with ***MLVA*** > ***Assignment report ...*** ( ).



Key	VNTR001	VNTR002	VNTR003	VNTR004	VNTR005	VNTR006	VNTR007
N2006-0003	5	-	3	1	2	2	2
N2006-0004	-	2	2	2 (?)	3	2	1
N2006-0021	5	2	3	1 (?)	3	2	2
N2006-0022	5	2	3	1	5	2	2
N2006-0040	3	2	3	2	3	3	2
N2006-0063	5	2	3	1	2	2	2
N2006-0082	5	2	3	1	2	2	2
N2006-0099	-	2	3	1	2	2	2
N2006-0112	4	3	2	2	2	-	7
N2006-0123	4	2	3	1	2	2	2
N2006-0152	5	2	3	1	2	2	2
N2006-0170	3	2	3	1	2	2	2
N2006-0172	5	2	3	1	2	2	2
N2006-0179	4	-	2	1	3	3	5
N2006-0185	4	-	2	1	3	3	5
N2006-0186	5	2	3	1	2	2	2
N2006-0208	6	2	3	1	2	2	2
N2006-0231	5	-	3	2	2	2	2
N2006-0232	5	2	3	1	3	2	2
N2006-0233	4	2	3	1	3	2	2

Figure 4.2: The *VNTR reporting* window.

Each cell, corresponding to an entry/VNTR combination, provides information about the copy number assignment.

- Any number displayed in a cell corresponds to the copy number determined for the corresponding entry/VNTR combination. When *only* a number is shown, this means that only a single band falls within the expected size ranges (bins) for the selected VNTR mapping and hence that the assignment is unambiguous.
- A number followed by a question mark between brackets (e.g. ”5 (?)”) means that more than one band in the fingerprint falls within the expected size ranges for the VNTR. The VNTR assignment algorithm selects by default the copy number that corresponds to the band with the largest peak height.
- In case a copy number was manually assigned (see 4.3), the copy number will be followed by ”(manual)” to indicate this fact. Please note that double assignments can be only be made manually (e.g. ”3/5 (manual)”).
- A hyphen (”-”) means that no feasible band was found, i.e. none of the bands present in the fingerprint fall within the expected size ranges for the VNTR.
- The text ”No data” will be displayed when no fingerprint experiment exists for this entry/VNTR combination.
- The text ”N/A” will be displayed when a VNTR assignment has not been performed yet for this entry/VNTR combination.

Alternative to copy numbers, the peak heights can be displayed with ***Report*** > ***View peak heights***. This command is a toggle, so select it again to return to the view with copy numbers.

Additional information is visualized via color scales:

- The default view (which can be called any time with ***Report*** > ***Show errors*** ( )) highlights possible errors: orange when no feasible band is detected and yellow when more than one feasible band is found.
- With ***Report*** > ***Show absolute deviation*** ( ), the cell colors provide an indication of the *absolute deviation* of the observed fragment size from the expected fragment size. The cell color ranges between green-yellow-orange-red according to increasing absolute deviation. This makes it easy to detect the magnitude of differences between the observed and expected sizes at a glance.
- With ***Report*** > ***Show deviation*** ( ), the cell colors express the *deviation* of the observed size from the expected size. The cell color ranges from blue (when the observed size is smaller than the expected size) over white to red (observed size is larger than the expected size).

2.1 Double-click on a key/VNTR cell to open the *Experiment card* window.

Any cell in the *VNTR reporting* window for which there is VNTR assignment present, can be double-clicked to open the fingerprint profile in the *Fingerprint curve processing* window (see 4.3).

4.3 Checking and manually assigning VNTR copy numbers

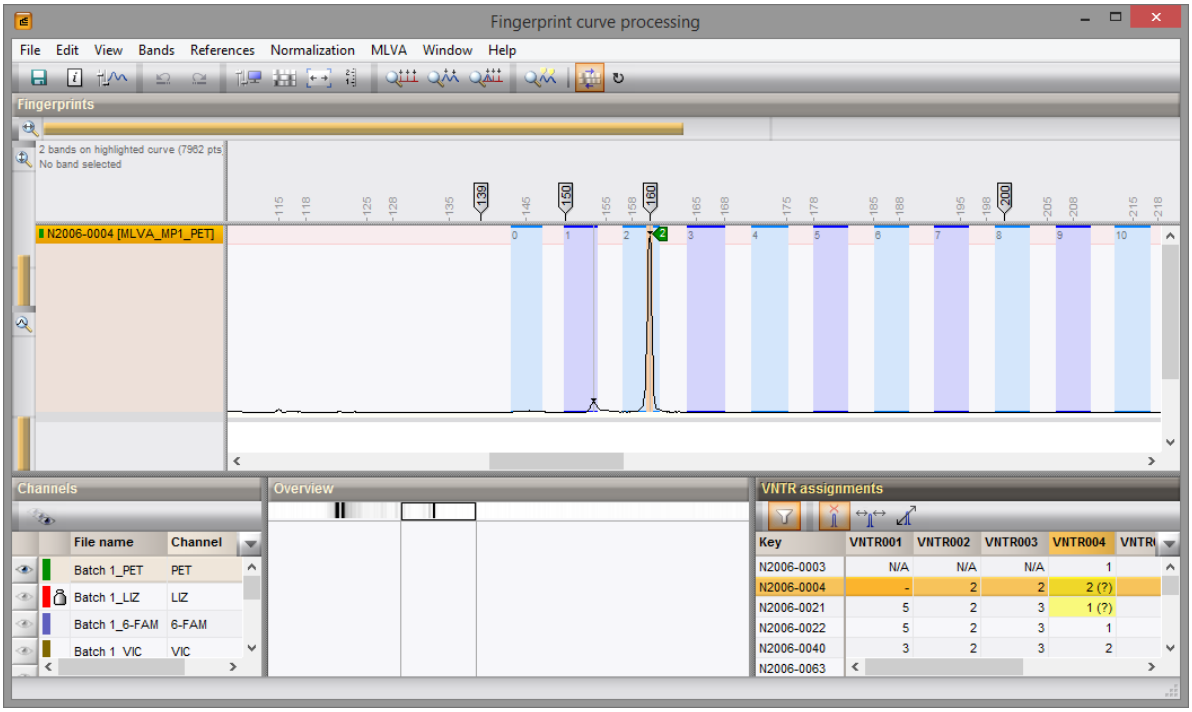


Figure 4.3: The *Fingerprint curve processing* window for a single electropherogram with VNTR copy number assignment.

When the *MLVA plugin* is installed in the database, the *Fingerprint curve processing* window will contain an additional *VNTR assignments panel* (see bottom right in Figure 4.3). This panel contains a grid with VNTR copy number assignments, identical to the *VNTR reporting* window and with the same display options

(***MLVA*** > ***Show errors*** (), ***MLVA*** > ***Show absolute deviation*** () and ***MLVA*** > ***Show deviation*** (); see 4.2 for an explanation). In addition, VNTR bins and labels with the assigned copy numbers are drawn on the *Fingerprints* panel, together with the electropherograms.

To switch to normalized view, select ***Normalization*** > ***Show normalized view*** (, **Shift+N**).

The automatic VNTR copy number assignment can be ran again with ***MLVA*** > ***Automatic assignment...***, e.g. to undo any manual assignments to use another mapping. This action opens the *Assign VNTRs* dialog box (see Figure 4.4).

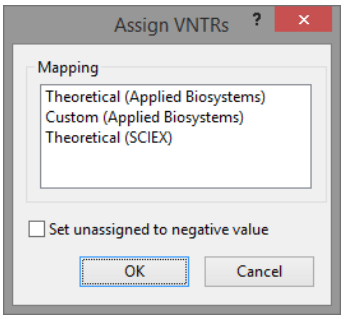


Figure 4.4: The *Assign VNTRs* dialog box.

This dialog box allows you to select a ***Mapping*** from any of the mappings (theoretical or custom mappings) defined in the MLVA schema.

Enable the option ***Set unassigned to negative value*** to score the absence of a VNTR as negative, rather than absent. All absent VNTRs will then be scored with the character value ”-2” in the character type **[SCHEMA_NAME]_vals**. The absent VNTRs, all having the score ”-2”, will be considered as a separate category for each VNTR when clustering the character information in the **[SCHEMA_NAME]_vals** experiment using the categorical coefficient.

Press <***OK***> to start the automatic VNTR copy number assignment.

To assign a VNTR copy number for a single band, first select the band using **Ctrl+click** and use ***MLVA*** > ***Manual assignment selected peak (single)*** (**Ctrl+M**). A copy number will be assigned that corresponds to the nearest VNTR bin.

In some profiles, more than one VNTR peak might occur. The *MLVA plugin* allows you to make double assignments by selecting two bands with **Ctrl+click** and ***MLVA*** > ***Manual assignment selected peak (double)***.

A VNTR copy number assignment for a selected band can be undone by ***MLVA*** > ***Delete assignment selected peak*** (**Ctrl+Shift+Del**).

Deleting a band with ***Bands*** > ***Delete highlighted bands*** (**Del**) does not automatically removes the VNTR copy number assignment!

With ***MLVA*** > ***Lane view filter*** (), one can toggle to a view mode where all lanes in the fingerprint file are shown or a view mode in which only the relevant lane for the highlighted entry/VNTR combination in the *VNTR assignments panel*.

Chapter 5

Importing repeat numbers from external files

5.1 Introduction

When VNTR copy numbers are already stored in external text, Excel, or other ODBC-compatible file, these copy numbers can be imported in BioNumerics and stored in the character type experiment **MIRU-VNTR_vals** using import routines available in the *Import* dialog box.

5.2 Import routines

The *Import* dialog box is called with the command ***File*** > ***Import...*** ( , **Ctrl+I**). The import tree options are organized in groups based upon the type of data (see Figure 5.1).

- With the ***Import fields and characters (text file)*** option, listed under the topic ***Character type data*** in the Import tree, the copy numbers can be imported from text files in the database and linked to new or existing database entries.
- With the ***Import fields and characters (Excel file)*** option, listed under the topic ***Character type data*** in the Import tree, the copy numbers can be imported from an Excel file in the database and linked to new or existing database entries.
- With the ***Import fields and characters (ODBC)*** option, listed under the topic ***Character type data*** in the Import tree, the copy numbers can be imported from ODBC-compatible files in the database and linked to new or existing database entries.

The import routines will not be covered in detail in this manual. For more detailed information, see the Reference manual, Chapter Setting up character type experiments.

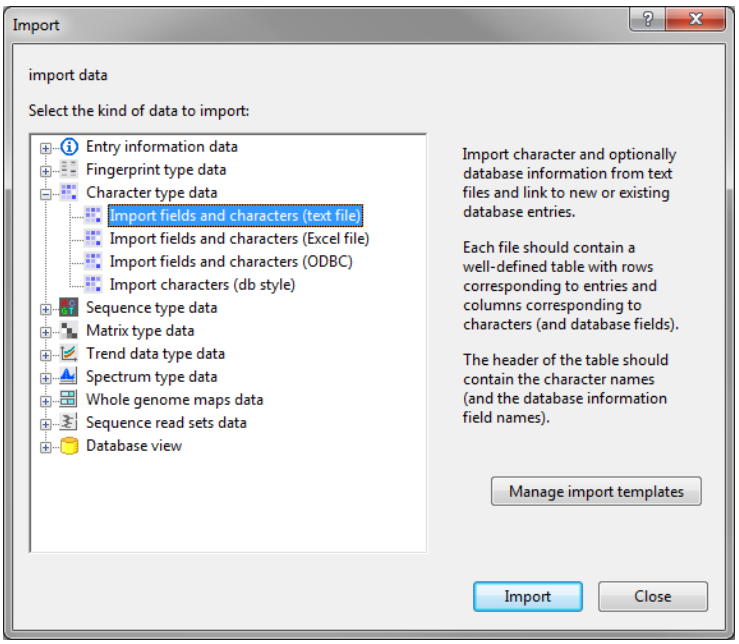


Figure 5.1: The character import routines in the *Import* dialog box.

Chapter 6

MLVA typing

6.1 Setting up an MLVA typing

Before a typing can be performed based on MLVA profiles, obviously an MLVA typing schema should be set up first. This is discussed in 2.6.

6.2 Assigning types

To assign an MLVA type to a selection of entries, use ***MLVA*** > ***Perform MLVA typing...*** (). This action opens the *Perform typing* dialog box (see Figure 6.1).

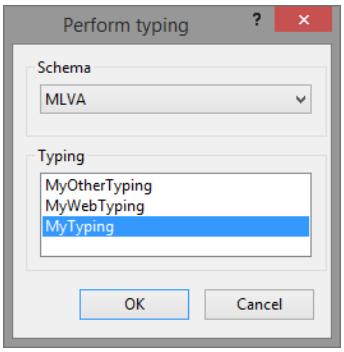


Figure 6.1: The *Perform typing* dialog box.

This dialog box allows you to select an MLVA ***Schema***, since more than one MLVA schema may be defined in the database. It also allows you to select one or more MLVA typing schemas (***Typing***) from any of the MLVA typing schemas defined in the selected MLVA schema (see 2.6).

Upon pressing <***OK***> in the *Perform typing* dialog box, the software will determine the MLVA types and will add these to the entry information field that was specified for the typing information (see 2.6).

When no **[SCHEMA_NAME]_frags** character experiment is present, the text ”No data” is entered. The text ”Incomplete profile” will be filled in when copy numbers are not available for all VNTRs as defined in the MLVA typing schema. For new profiles, i.e. VNTR copy number combinations that are not in the database yet, the text ”Unknown” will be filled in.

When new VNTR profiles were found, the software will display the *Update* dialog box (see 2.6). In case the MLVA types are locally stored in the database (option ***Enter types manually*** checked in the *Typing settings* dialog box), this dialog will allow you to add the new profiles to the database. In case the MLVA types are

obtained from an external file or URL (option ***Import types from external source*** checked in the *Typing settings* dialog box), the software will offer to export the new types to a text file. If <***OK***> is clicked, the *Select file to export* dialog box appears, which allows you to save the text file containing the new VNTR profiles to your hard drive. This file can then be sent to the MLVA database curators.

The software will automatically change the entry selection to samples that require further examination, i.e. entries with an unknown MLVA type. This allows you to use e.g. ***MLVA*** > ***Assignment report ...*** () on the selection, to check the VNTR copy number assignments.

Chapter 7

MLVA data analysis

Some useful features in the context of MLVA/VNTR data analysis will be highlighted in this chapter. More detailed information about the analysis possibilities in the software can be found in the BioNumerics manual.

7.1 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**).

7.2 The Comparison window

2.1 Make a selection in the *Database entries* panel (see 7.1).

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.

A *Comparison* window is created, with the selected database entries. The *Comparison* window is divided in six main panels: the *Dendrogram* panel, which shows the dendrogram if calculated, the *Experiment data* panel, showing the images of the experiments, the *Information fields* panel, which shows the database fields in the same layout as in the database, the *Similarities* panel, which shows the similarity values, the *Experiments* panel, which shows the available experiment types and the *Groups* panel, showing the groups, if defined.

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 experiment name **[SCHEMA_NAME]_vals** in the *Experiments* panel and select ***Characters*** > ***Show values*** () to display the repeat numbers in the *Experiment data* panel (see Figure 7.1).

2.5 In the *Comparison* window, groups can be created based on a selection (***Groups*** > ***Create new group from selection*** (, **Ctrl+G**)) or based on the content of an information field: right-click in the header of a field and select ***Create groups from database field*** from the menu.

7.3 Cluster analysis

3.1 In the *Experiments* panel of the *Comparison* window, make sure the **[SCHEMA_NAME]_vals** experiment is selected.

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

For VNTR data, the coefficients that make most sense are:

- ***Categorical***: preferred if differences in copy numbers should be treated in a qualitative way.
- ***Euclidean distance***: preferred if differences in copy numbers should be treated in a quantitative way (larger difference means more distant organisms).

3.3 Select the ***Categorical (values)*** coefficient from the list and press <***Next***>.

The categorical coefficient compares the repeat numbers to see if they are the same or different but does not quantify the difference.

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 7.1 for an example). More information about the *Comparison* window can be found in the manual.

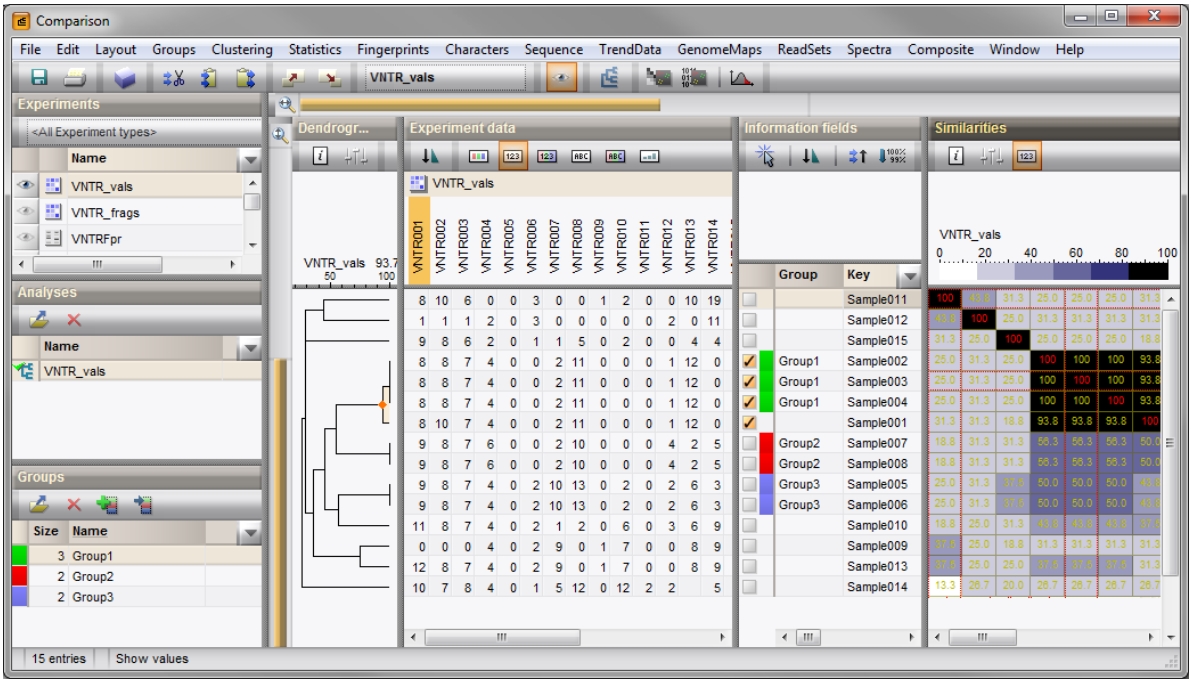


Figure 7.1: The *Comparison* window with a UPGMA tree displayed.

A minimum spanning tree in BioNumerics is calculated the *Advanced cluster analysis* window. This window can be launched from the *Comparison* window:

3.5 Select ***Clustering*** > ***Calculate*** > ***Advanced cluster analysis...*** or press the button and select ***Advanced cluster analysis*** to launch the *Create network* wizard.

The predefined template ***MST for categorical data*** uses the categorical coefficient for the calculation of the similarity matrix, and will calculate a standard minimum spanning tree with single and double locus variance priority rules.

3.6 Specify an analysis name (for example **VNTR1**), make sure **[SCHEMA_NAME]_vals** is selected, select ***MST for categorical data***, and press <***Next***>.

The *Advanced cluster analysis* window pops up (see Figure 7.2 for an example). The *Network panel* displays the minimum spanning tree, the upper right panel (*Entry list*) displays the entries that are present in the tree. The *Cluster analysis method panel* displays the settings used, in this example the priority rules that result in the displayed network. The colors of the comparison groups (if defined) are automatically shown as node colors.

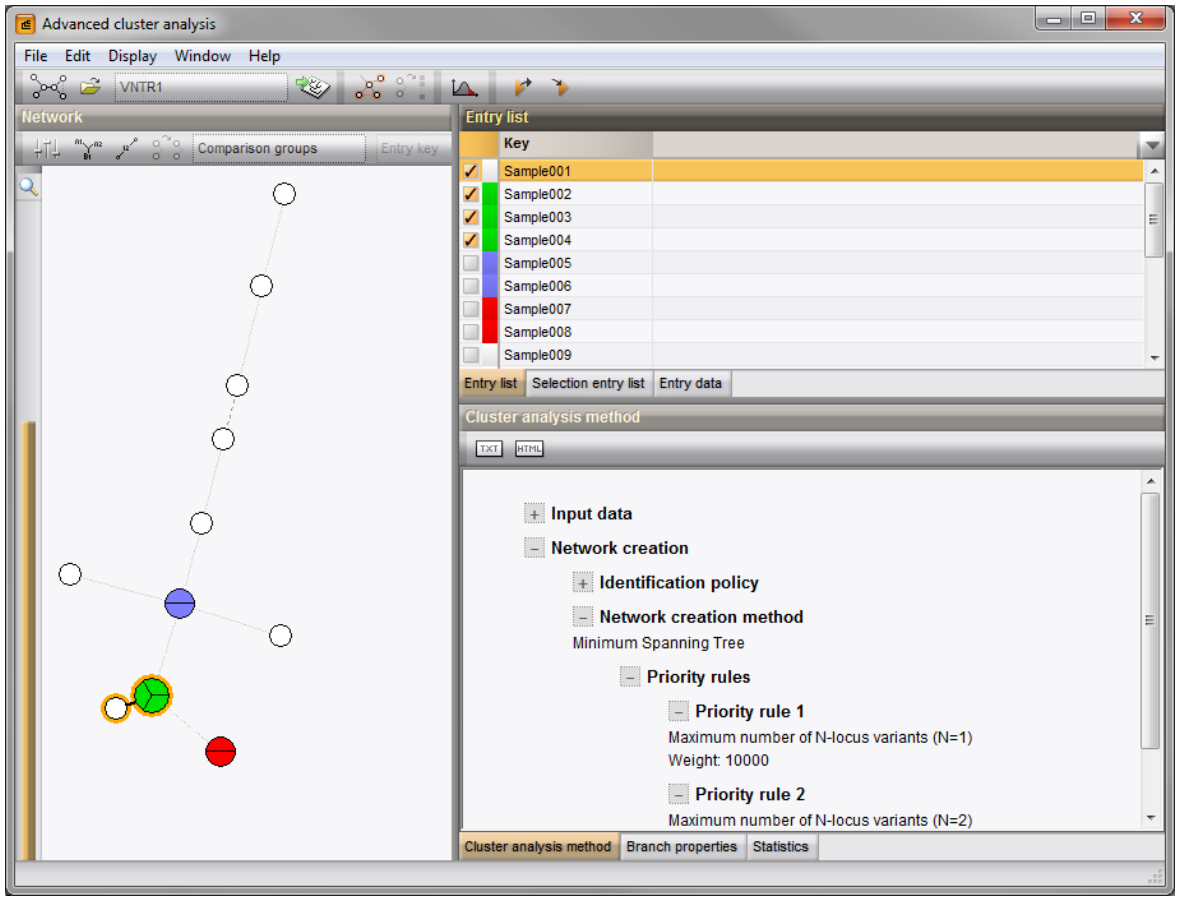


Figure 7.2: The *Advanced cluster analysis* window.

3.7 To change the display settings of the network (node/branch labels, node/branch colors, etc.) press or choose ***Display*** > ***Display settings***.

Detailed information about the *Advanced cluster analysis* window can be found in the manual.

3.8 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.

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

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

3.11 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.

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