BIONUMERICS Tutorial:

Importing spoligo typing data

1 Introduction

Spacer oligonucleotide typing (spoligotyping) is widely used for differentation of *Mycobacterium tuberculosis* bacteria. With this method the DR region is investigated: this region consists of multiple copies of a conserved directly repeated 36 bp sequence (DR), separated by multiple non-repetitive (43) spacer sequences. Different *Mycobacterium tuberculosis* strains have various complements of these 43 spacers and form the basis of the spoligo typing assay.

Different systems are in use for the designation of spoligo type patterns. In this tutorial the import of following spoligo information is covered:

- Import of the spoligo type patterns as a 43-digit binary code stored in an external file (see 2).
- Import of the octal codes from an external file and conversion to the 43-digit binary code (see 3).

The data files used in this tutorial can be downloaded from https://www.applied-maths.com/download/sample-data (click on "Spoligotyping data").

2 Import of 43 digit Spoligo codes

2.1 Preparing the database

1. Create a new database (see tutorial "Creating a new database") or open an existing database.

Since we will be importing the spoligo data as character data, we will first create a character type to hold this data. The steps below can be skipped if a suitable character type is already present in the database.

2. In the *Main* window, click on + in the toolbar of the *Experiment types* panel and select ***Character type*** from the list. Press <***OK***>.

The *New character type* wizard prompts you to enter a name for the new character type.

3. Enter a name, for example "Spoligo" and press <***Next***>.

In the next step of the wizard, the choice is offered between ***Numerical values*** and ***Binary data***.

4. Choose ***Binary data***.

After pressing <***Next***> again, the wizard asks if the character type has an open (***Yes***) or closed (***No***) character set.

5. Make sure ***No*** is selected and press the <***Finish***> button to complete the setup of the new character type.

The *Experiment types* panel now lists the new character type **Spoligo**.

2.2 Import wizard

In this section, we will import data from the text file `SpoligoBinary.txt` (see Figure 1). This text file contains a unique strain identifier in the first column (***Key***) and the binary spoligo data in the 43 other columns.

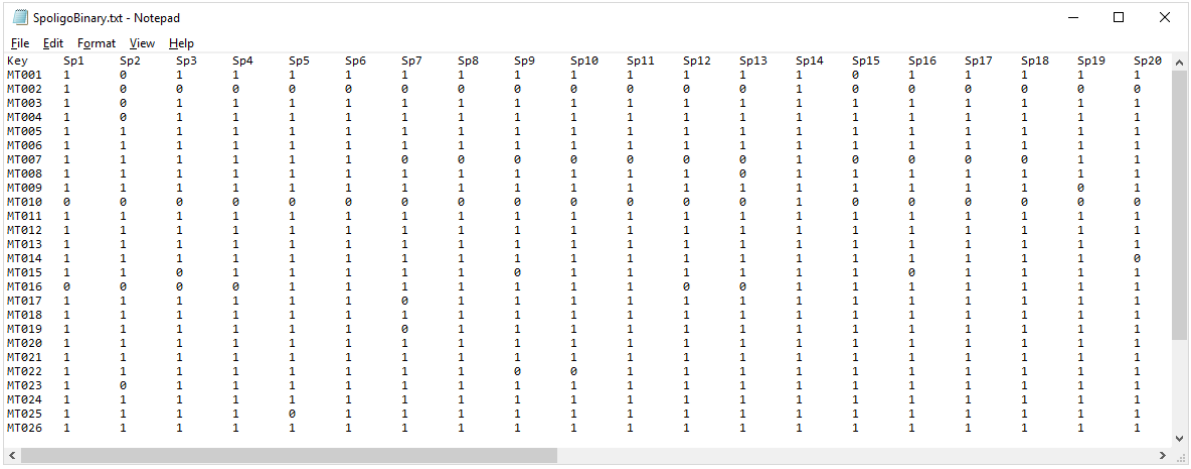

Figure 1: Text file containing binary spoligo data.

6. Select ***File*** > ***Import...*** (, **Ctrl+I**) to open the *Import* dialog box.

7. Choose the option ***Import fields and characters (text file)*** under the ***Character type data*** item in the tree and press <***Import***> (see Figure 2).

8. Press <***Browse***> and browse for the downloaded `SpoligoBinary.txt` file. Next, press <***Open***> and press <***Next***>.

As this is the first time we import character data from a text file into the database, we need to create a new import template by specifying ***Import rules***.

9. Select "Key" in the list and click <***Edit destination***> or simply double-click on "Key". Select "Key" as the BIONUMERICS destination field in the *Edit data destination* dialog box and press <***OK***> (see Figure 3).

The information in the grid is updated.

10. Make a multiple selection for all 43 rows. Do this by selecting "Sp1" and while holding the **Shift**-key, click on "Sp43". Press <***Edit destination***>, select the character type **Spoligo** as destination under ***Character value*** and click <***OK***> (see Figure 4).

11. Press <***OK***> and then <***Yes***> to confirm the creation of new characters.

The grid panel is updated (see Figure 5).

12. Press <***Preview***> to see what you are about to import (see Figure 6).

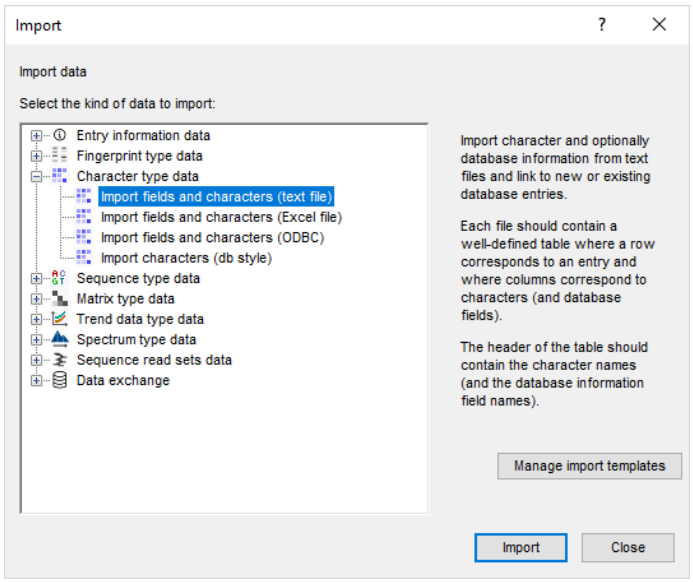

Figure 2: The *Import* dialog box.

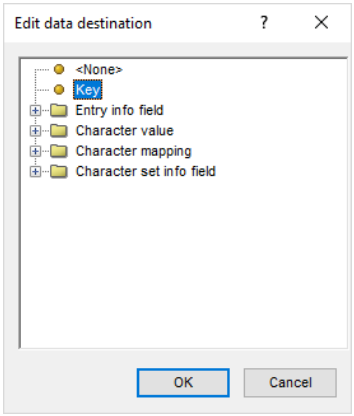

Figure 3: Edit data destination.

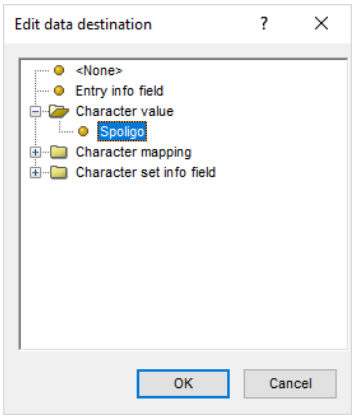

Figure 4: Link to Spoligo character experiment.

13. Press the <***Close***> button to close the preview and press <***Next***> to proceed to the *Import links* dialog box.

14. Make sure ”Key” is checked in the ***Import links*** step and press <***Finish***>.

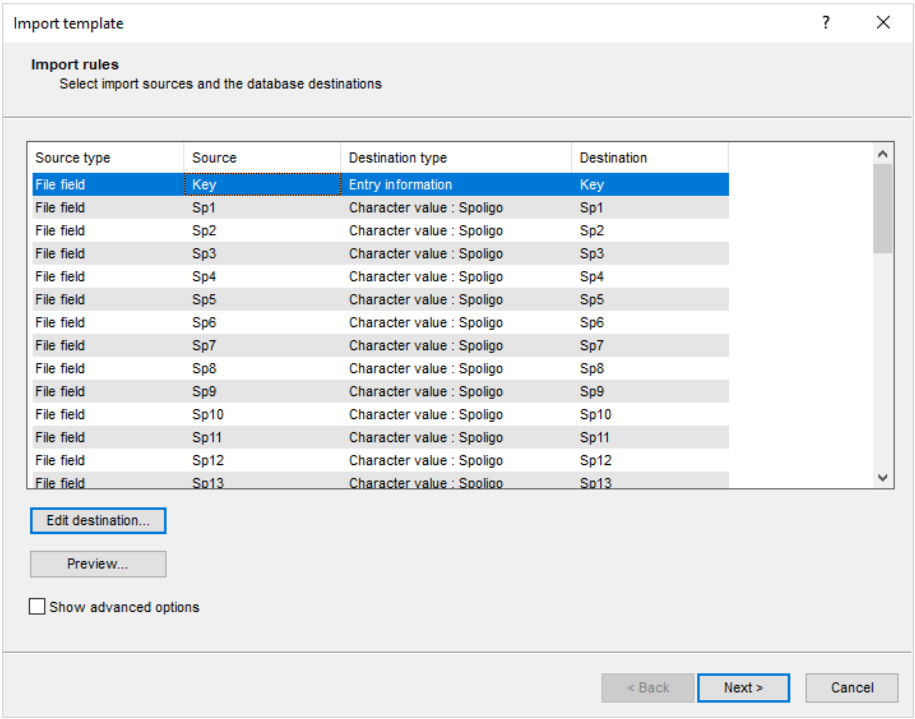

Figure 5: Import rules.

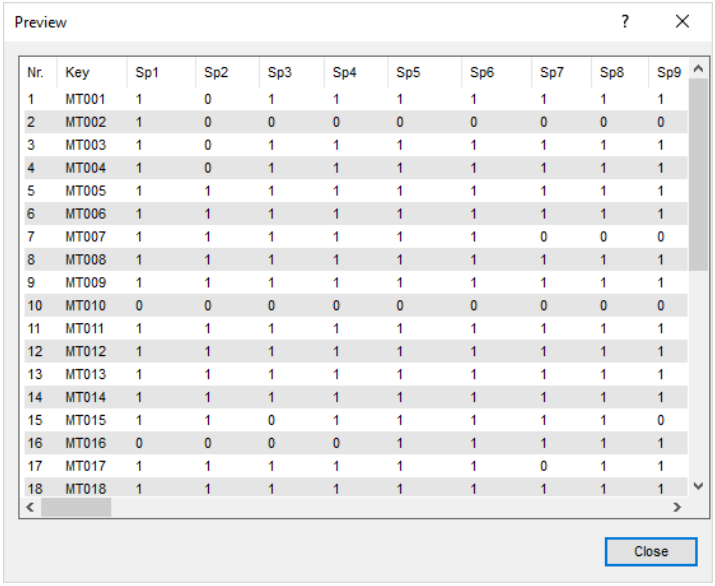

Figure 6: Preview.

The import template needs to be saved to be able to use it again later on.

15. Enter a ***Name*** for the import template (e.g. “Spoligo binary data”) and optionally a ***Description***. Next, press <***OK***>.

16. In the *Import template* wizard page, the new template is added and is automatically selected (see Figure 7). Click <***Next***>.

In case there are no entries present with the same key as in the external file, the *Database links* wizard page will indicate that 35 new entries will be created during import (see Figure 8).

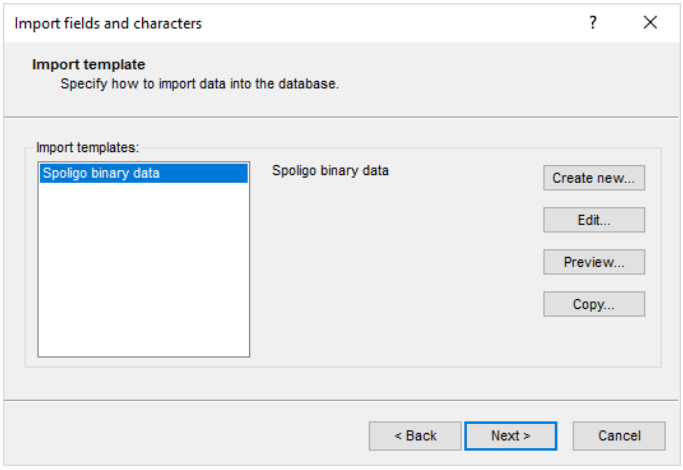

Figure 7: Import template.

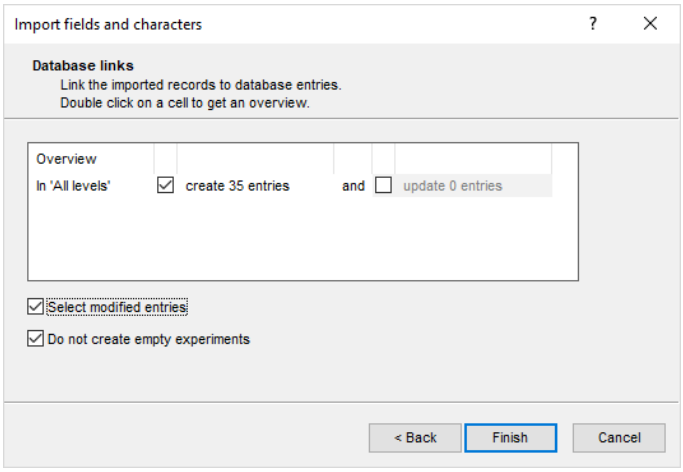

Figure 8: Add 35 new entries.

17. Press <***Finish***> to start the actual import. The progress of the import is shown while database information is added to the BIONUMERICS database.

The entries are displayed in the *Database entries* panel and all entries are automatically selected. The character data is stored in the character type **Spoligo** (see Figure 9).

18. Click on a green colored dot in the *Experiment presence* panel to open the experiment card for an entry.

The values are displayed as a ”plate” with one column. The ”0” values have a white color, the ”1” values are colored in black.

19. Hover over the experiment card with the mouse.

The key of the entry is shown and the name and value of the character being pointed at.

20. Click in the upper left corner of the card to close the card.

21. To view the values in a list, double-click on the experiment **Spoligo** in the *Experiment types* panel, select ***Settings*** > ***General settings...*** (), select the *Experiment card* tab and change the representation to ***List***. Close the two windows.

22. Click on a green colored dot in the *Experiment presence* panel to open the experiment card for an entry.

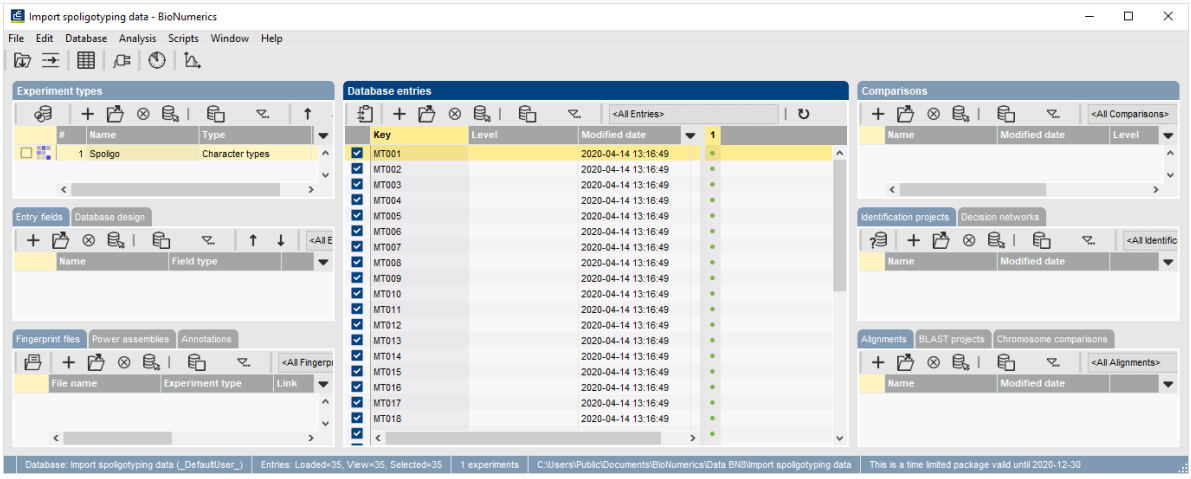

Figure 9: The *Main* window after import of the data.

The imported values are now displayed in a list (see Figure 10).

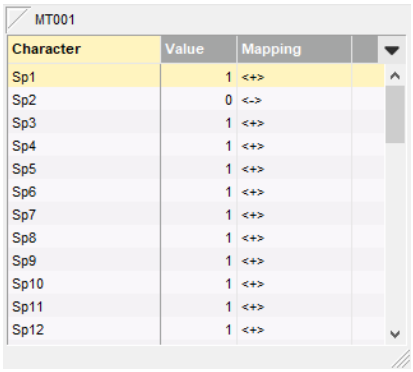

Figure 10: Character experiment card.

23. Close the experiment card by clicking in the left upper corner of the card.

Information about the follow-up analysis of binary data sets in BIONUMERICS can be found in the tutorial ”Clustering a binary data set”, available on the Applied Maths website.

3 Import and conversion of octal Spoligo codes

3.1 Preparing the database

1. Create a new database (see tutorial ”Creating a new database”) or open an existing database.

Octal spoligo codes will be imported and stored in an information field (see 3.2), and a conversion script will convert the octal spoligo codes to the original binary patterns and store these patterns in a character type experiment (see 3.3).

We will first create a character type to hold this data. The steps below can be skipped if a suitable character type is already present in the database.

2. In the *Main* window, click on + in the toolbar of the *Experiment types* panel and select ***Character type*** from the list. Press <***OK***>.

The *New character type* wizard prompts you to enter a name for the new character type.

3. Enter a name, for example “Spoligo” and press <***Next***>.

In the next step of the wizard, the choice is offered between ***Numerical values*** and ***Binary data***.

4. Choose ***Binary data***.

After pressing <***Next***> again, the wizard asks if the character type has an open (***Yes***) or closed (***No***) character set.

5. Make sure ***No*** is selected and press the <***Finish***> button to complete the setup of the new character type.

The *Experiment types* panel now lists the new character type **Spoligo**.

6. Double-click on the **Spoligo** experiment in the *Experiment types* panel to call the *Character type* window.

7. Select ***Characters*** > ***Add array of characters...*** and specify **43** as the number of rows and **1** as the number of columns.

43 characters are now listed in the *Character type* window (see Figure 11).

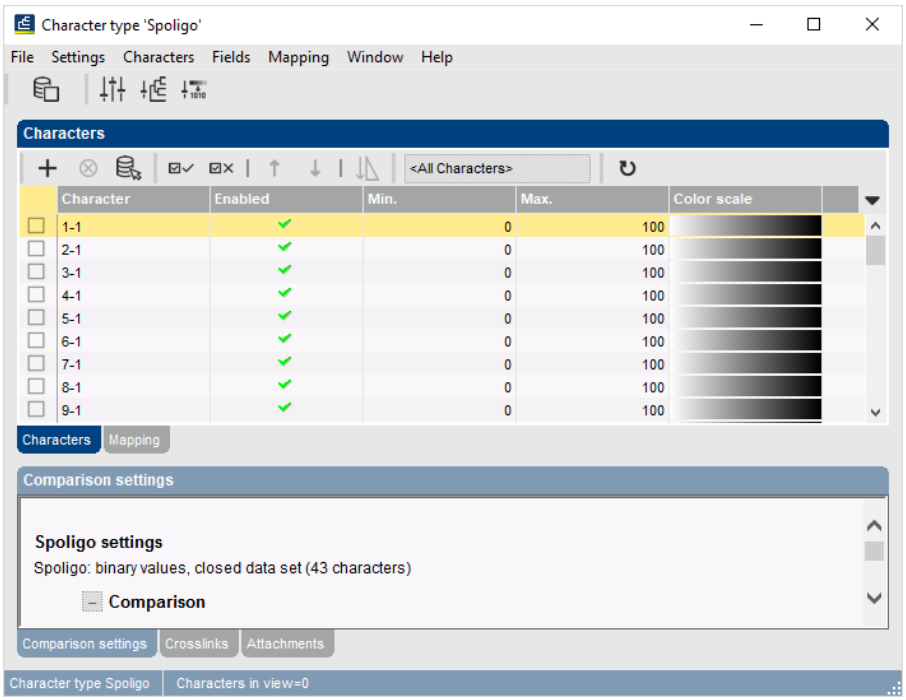

Figure 11: The *Character type* window.

8. Close the *Character type* window.

3.2 Import wizard

In this section, we will import data from the text file `SpoligoOctal.txt` (see Figure 12). This text file contains a unique strain identifier in the first column (***Key***) and the octal spoligo code in the second column (***Octal***).

9. Select ***File*** > ***Import...*** (, **Ctrl+I**) to open the *Import* dialog box.

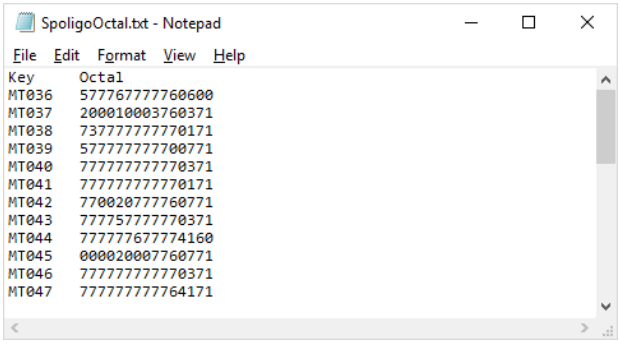

Figure 12: Text file containing octal spoligo codes.

10. Choose the option ***Import fields (text file)*** under the ***Entry information data*** item in the tree and press <***Import***> (see Figure 13).

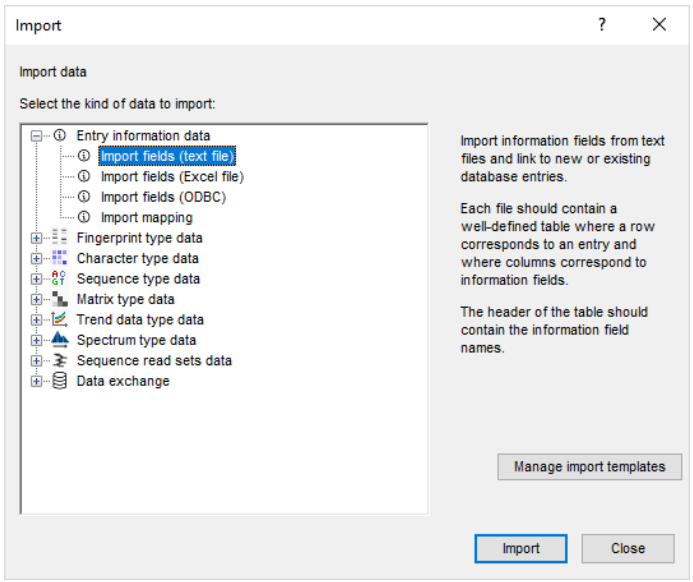

Figure 13: The *Import* dialog box.

11. Press <***Browse***> and browse for the downloaded `SpoligoOctal.txt` file. Next, press <***Open***> and press <***Next***>.

As this is the first time we import entry information from a text file into the database, we need to create a new import template by specifying ***Import rules***.

12. Select "Key" in the list and click <***Edit destination***> or simply double-click on "Key". Select "Key" as the BIONUMERICS destination field in the *Edit data destination* dialog box and press <***OK***> (see Figure 14).

The information in the grid is updated.

13. Select "Octal" from the list, press <***Edit destination***>, select "Create new" as destination under ***Entry info field*** and click <***OK***>.

14. Press <***OK***> and then <***Yes***> to confirm the creation of the new information field.

The grid panel is updated (see Figure 15).

15. Press <***Preview***> to see what you are about to import (see Figure 16).

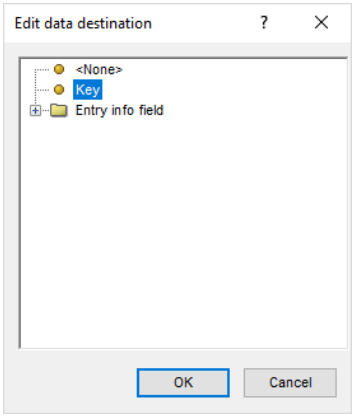

Figure 14: Edit data destination.

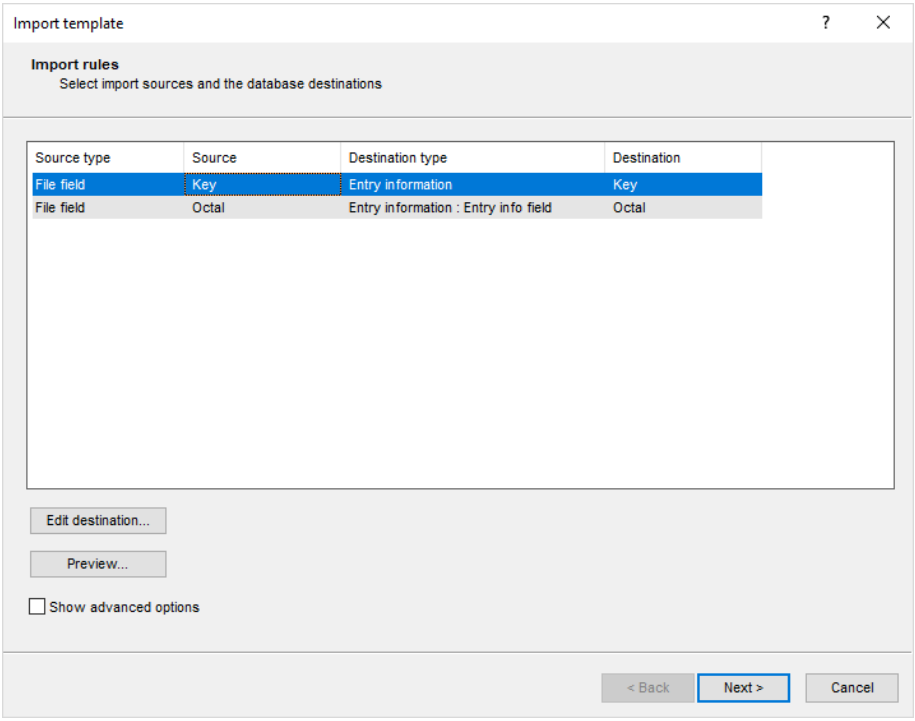

Figure 15: Import rules.

16. Press the <***Close***> button to close the preview and press <***Next***> to proceed to the *Import links* dialog box.

17. Make sure ”Key” is checked in the ***Import links*** step and press <***Finish***>.

The import template needs to be saved to be able to use it again later on.

18. Enter a ***Name*** for the import template (e.g. “Spoligo octal code”) and optionally a ***Description***. Next, press <***OK***>.

19. In the *Import template* wizard page, the new template is added and is automatically selected (see Figure 17). Click <***Next***>.

In case there are no entries present with the same key as in the external file, the *Database links* wizard page will indicate that 35 new entries will be created during import (see Figure 18).

20. Press <***Finish***> to start the actual import. The progress of the import is shown while database information is added to the BIONUMERICS database.

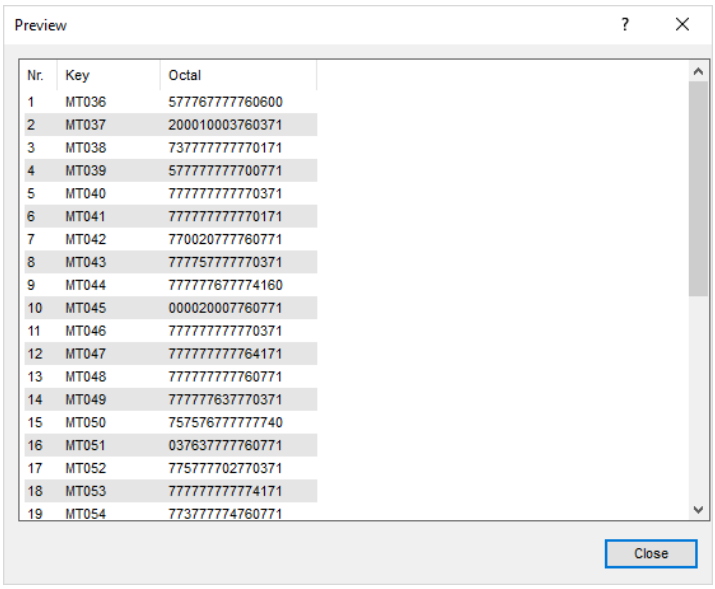

Figure 16: Preview.

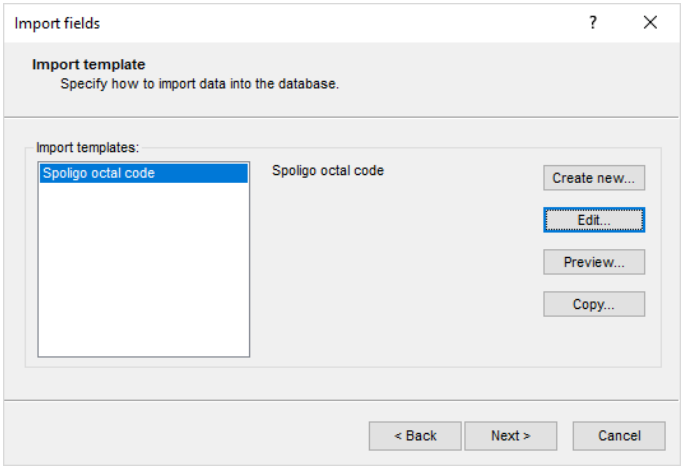

Figure 17: Import template.

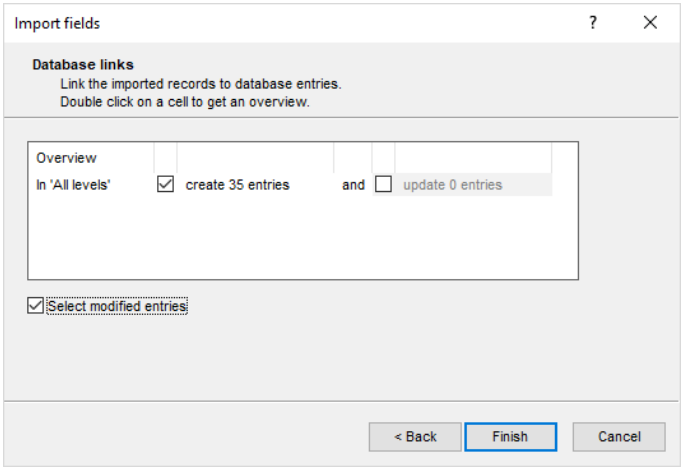

Figure 18: Add 35 new entries.

The entries are displayed in the *Database entries* panel and all entries are automatically selected. The octal codes are stored in the ***Octal*** entry information field (see Figure 19).

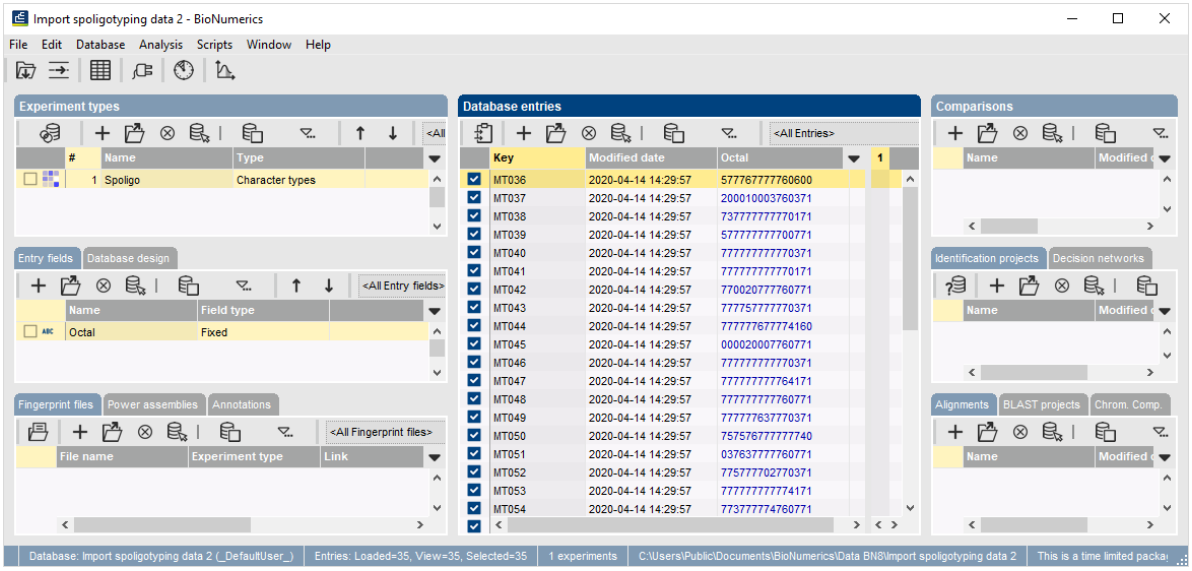

Figure 19: The *Main* window after import of the data.

3.3 Conversion

21. In the *Database entries* panel of the *Main* window select all entries containing an octal code that you wish to convert to the 43 digit spoligo code. To select all entries at once, make sure the *Database entries* panel is the active panel and select ***Edit*** > ***Select all*** (**Ctrl+A**).

22. Select ***Scripts*** > ***Run script from file...***, navigate to the downloaded and unzipped `Spoligotyping data` folder and select the script `ConvertSPOLIGOcode.BNS`.

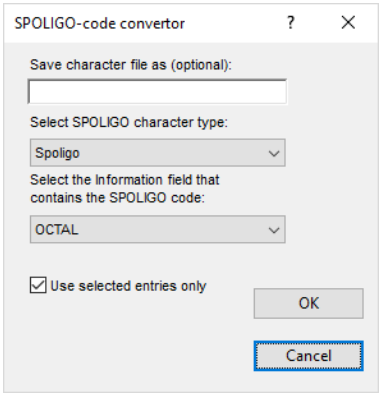

Figure 20: Conversion dialog.

23. Select the ***Character type*** that will hold the 43 digit code after conversion of the data. In our example, make sure ***Spoligo*** is selected.

24. Select the ***Information field*** that contains the octal code. In our example, make sure ***Octal*** is selected.

25. Check the option ***Use selected entries only*** to apply the script only on the selected entries in the database. When this option is unchecked, all entries in the database will be included in the conversion action.

26. Press <***OK***> to run the script.

The 43 digit codes are stored in the character type **Spoligo** (see Figure 21).

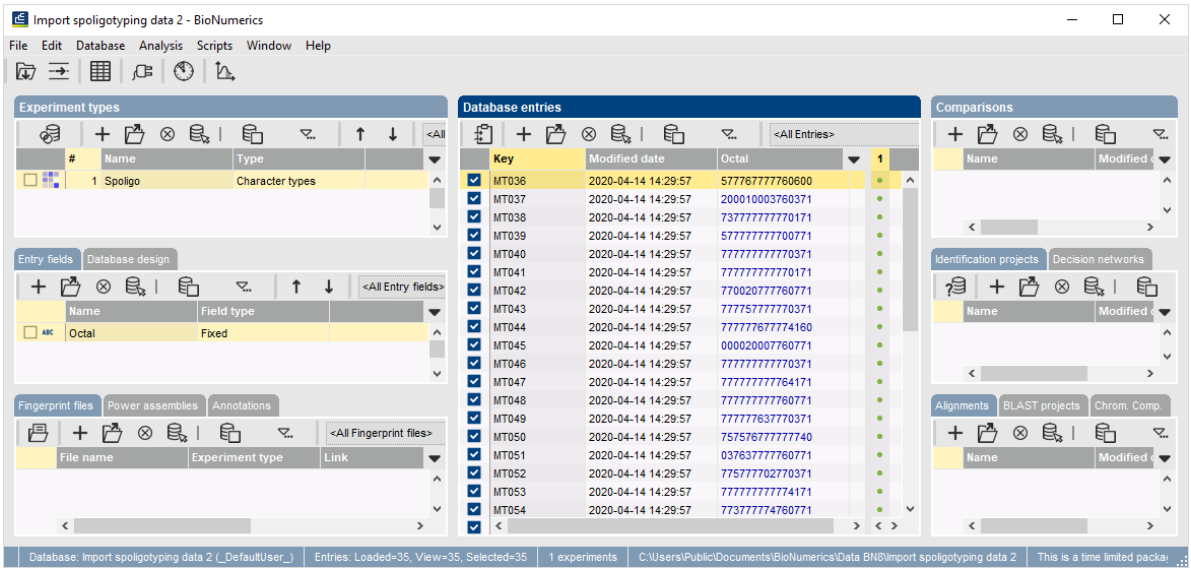

Figure 21: The *Main* window after import of the data.

27. Click on a green colored dot in the *Experiment presence* panel to open the experiment card for an entry.

The values are displayed as a ”plate” with one column. The ”0” values have a white color, the ”1” values are colored in black.

28. Hover over the experiment card with the mouse.

The entry Key is shown and the name and value of the character being pointed at.

29. Click in the upper left corner of the card to close the card.

30. To view the values in a list, double-click on the experiment **Spoligo** in the *Experiment types* panel, select ***Settings*** > ***General settings...*** (), select the *Experiment card* tab and change the representation to ***List***. Close the two windows.

31. Click on a green colored dot in the *Experiment presence* panel to open the experiment card for an entry.

The imported values are now displayed in a list (see Figure 22).

32. Close the experiment card by clicking in the left upper corner of the card.

Information about the follow-up analysis of binary data sets in BIONUMERICS can be found in the tutorial ”Clustering a binary data set”, available on the Applied Maths website.

MT036

Character	Value	Mapping
1-1	1	<+>
2-1	0	<->
3-1	1	<+>
4-1	1	<+>
5-1	1	<+>
6-1	1	<+>
7-1	1	<+>
8-1	1	<+>
9-1	1	<+>
10-1	1	<+>
11-1	1	<+>
12-1	1	<+>

Figure 22: The character experiment card.