BIONUMERICS Tutorial:

Partition mapping

1 Introduction

Classification is very important in (micro)biology. By an organism's membership of a class or group, we can deduce some of its characteristics. Organisms can be classified according to different criteria, such as exposure of surface antigens (serotyping), susceptibility to bacteriophages (phage typing), banding patterns obtained with various fingerprint techniques (PFGE, RFLP,...), etc..

When a set of organisms is classified by more than one system, the question arises as to how the classification systems relate to each other. This is what the ***partition mapping tool*** is designed for in BIONUMERICS: to compare the outcome of two classification systems. In this tutorial some of the features of this ***partition mapping tool*** will be illustrated using a sample data set (see 2).

2 Sample data

The partition mapping tool requires the class information of both partitions to be stored as an information field or as comparison groups. To illustrate the partition mapping tools, classification data is available on download page of the website (https://www.applied-maths.com/download/sample-data, click on "Partition sample field data").

The sample data set contains classification results obtained via PFGE and serotyping. For each method, a letter code (type) is assigned to each typical PFGE profile or serological behavior, respectively. The data are available as a tab-delimited text file, designated `Partition_data.txt`.

3 Preparing the database

The **DemoBase Connected** will be used to illustrate the partition mapping tool available in BIONUMERICS and can be downloaded directly from the *BIONUMERICS Startup* window or restored from the back-up file available on our website:

1. To download the database directly from the *BIONUMERICS Startup* window, click the ⬇ button, located in the toolbar in the *BIONUMERICS Startup* window. Select **DemoBase Connected** from the list and select ***Database*** > ***Download*** (🗄). Confirm the download action.

2. To restore the database from the back-up file, first download the file `DemoBase_Connected.bnbk` from https://www.applied-maths.com/download/sample-data, under 'DemoBase Connected'. In the *BIONUMERICS Startup* window, press the button, select ***Restore database***, browse for the downloaded file and select ***Create copy***. Specify a name and click <***OK***>.

In contrast to other browsers, some versions of Internet Explorer rename the `DemoBase_Connected.bnbk` database backup file into `DemoBase_Connected.zip`. If this happens, you should manually remove the `.zip` file extension and replace with `.bnbk`. A warning will appear ("If you change a file name extension, the file might become unusable."), but you can safely confirm this action. Keep in mind that Windows might not display the `.zip` file extension if the option "Hide extensions for known file types" is checked in your Windows folder options.

1. Open the database **DemoBase Connected**.

2. In the *Main* window, select ***File*** > ***Import...*** (, **Ctrl+I**).

3. In the *Import* dialog box, expand ***Entry information data***, highlight ***Import fields (text file)*** and press <***Import***>.

4. Browse for the `Partition_data.txt` file and press <***Next***>.

5. Highlight the row that corresponds to "Key" and press <***Edit destination***>.

6. In the *Edit data destination* dialog box, highlight "Key" and press <***OK***>.

7. Highlight the two remaining file fields and press <***Edit destination***> again.

8. Highlight "Entry info field" and press <***OK***>.

9. Leave the default names unaltered and press <***OK***>. Confirm the action.

The *Import rules* dialog box should now look like in Figure 1.

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

11. Specify a template name and press <***OK***>.

12. Press <***Next***> and <***Finish***> to import the information in the database.

The data is now imported as information fields in the **DemoBase Connected** demonstration database. Optionally, we can set information field properties and a color coding for "PFGE type".

13. Right-click the information field header of "PFGE type" and select ***Field properties*** from the floating menu (see Figure 2).

14. In the dialog box that appears, press <***Add all***> to add all current field states and confirm. Check ***Use colors*** and close the dialog.

Each PFGE type (A, B, C or D) is now displayed in its own color.

Furthermore, we will create a new information field to store mapping information in.

15. Right-click the information fields header of the *Database entries* panel and select ***Add information field*** from the floating menu (see Figure 3).

16. Enter a name for the new information field (e.g. "Mapping"), leave the other settings unaltered and press <***OK***>.
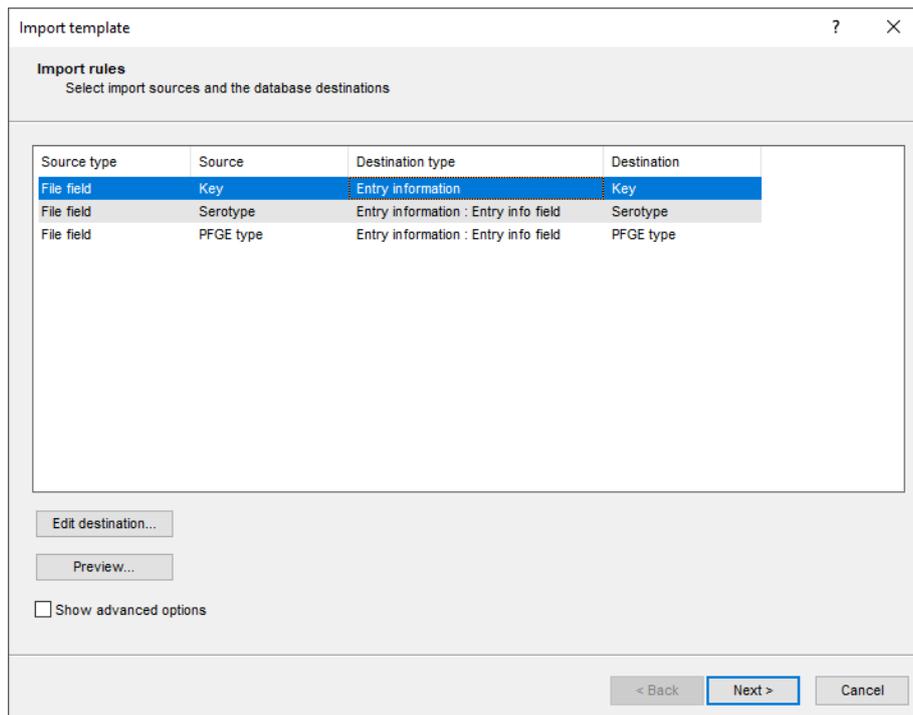
Figure 1: The *Import rules* dialog box after setting up the template for import of the example data.
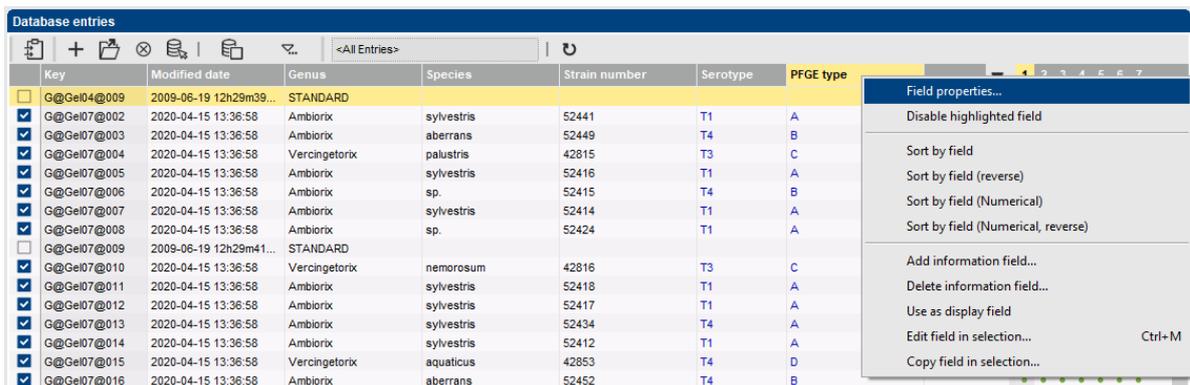
**Figure 2:** Set the field properties for the "PFGE type" field.
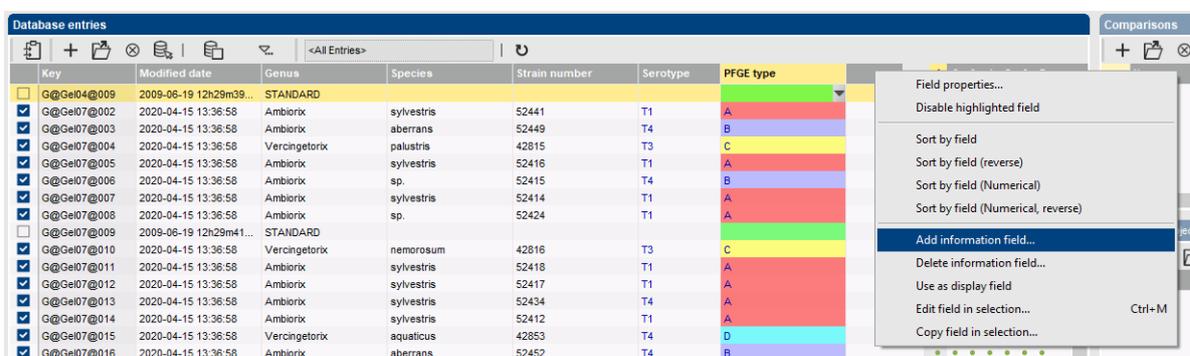
Figure 3: Add a new information field.

4 Performing a partition mapping

1. In the *Database entries* panel of the *Main* window, select all entries using **Edit** > **Select all** (**Ctrl+A**) and unselect the STANDARDs.

2. Click on the + button in the *Comparisons* panel to create a new comparison for the selected entries.

3. In the *Comparison* window, select **Statistics** > **Partition mapping...**.

The *Partition mapping* dialog box pops up.

4. For the example data, select "PFGE type" as **First partition** from the corresponding drop-down list and "Serotype" as **Second partition**.

5. Under **Mapping method**, highlight **Maximum-likelihood partition mapping** and leave the **Precision** versus **Recall** slider at its default position.

6. Press <**OK**> to calculate the partition mapping.
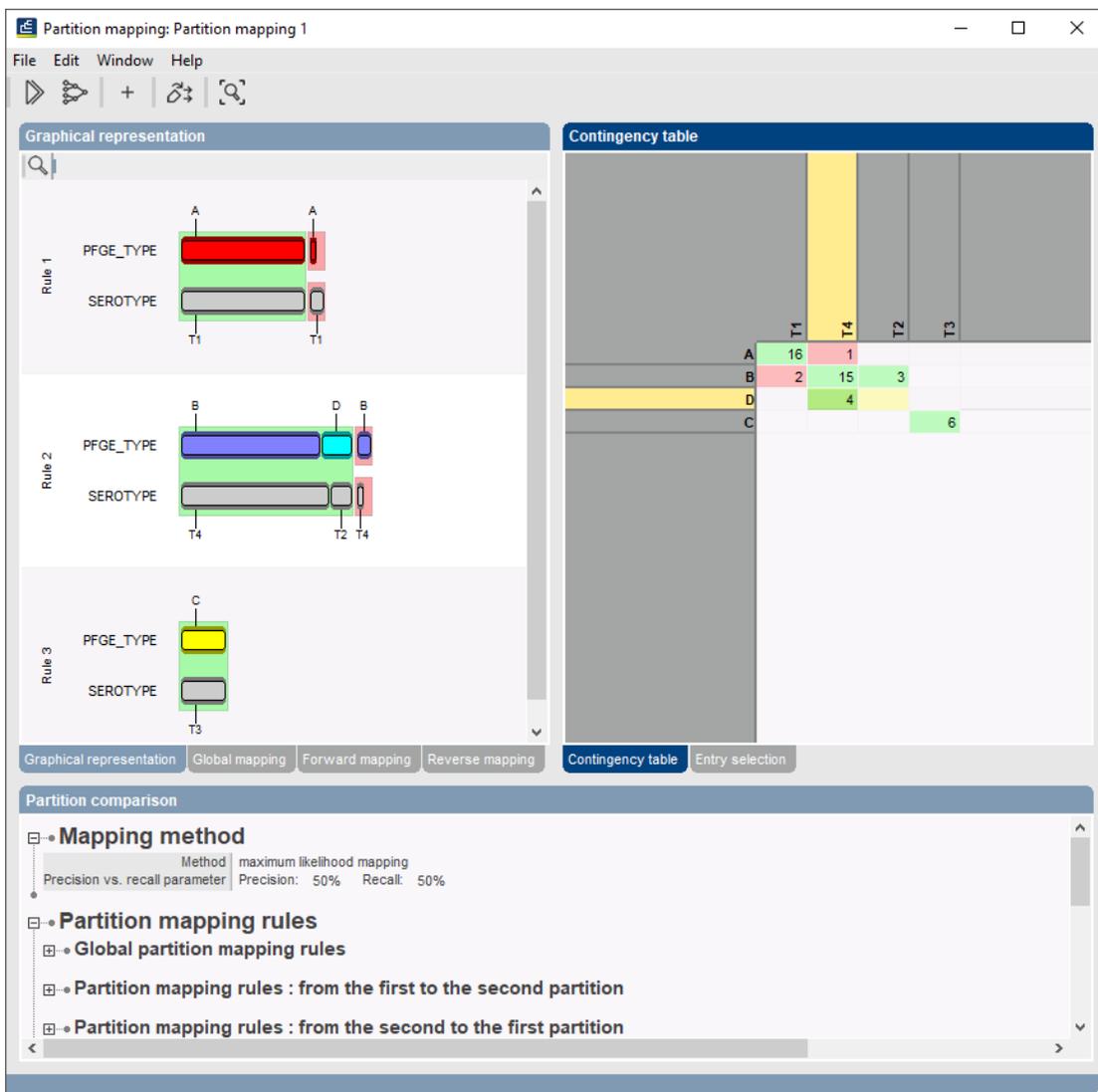
Figure 4: The *Partition mapping* window.

The *Partition mapping* window consists of seven panels (see Figure 4): *Graphical representation*,

Global mapping, *Forward mapping*, *Reverse mapping*, *Contingency table*, *Entry selection* and *Partition comparison panel*.

In the *Contingency table panel*, the contingency table of the first partition (rows) and second partition (columns) is displayed. All mapping rules are calculated based on this table.

The cells in the contingency table are colored according to their compliance with the mapping rules:

- green is a confirmation of the mapping rules

- red means a violation

- yellow means missing entries (not observed in the dataset)

The *Graphical representation panel* visualizes the mapping rules in a graphical fashion. Classes from the first and second partition are represented by bars, which are sized proportionally with the number of entries they contain. For each rule, the classes from the first partition are shown above the classes from the second partition. Classes are colored according to the colors defined in the information field properties; if no colors are defined, the bars appear in gray. Classes from the first and second partition that confirm the mapping rule are enclosed in a green rectangle. Violations of the mapping rule are shown in two, vertically separated, red rectangles.

The *Global mapping panel* is a grid view panel that lists all mapping rules that were calculated based on the contingency table and the selected mapping method.

The *Forward mapping panel* lists how each class from the first partition is mapped, according to the mapping rules, on the second partition.

The *Reverse mapping panel* lists how each class of the second partition is mapped on the first partition.

7. Click on the *Graphical representation tab* to make this panel the active panel again.

8. Zoom in or out on the graphical representation using the zoom slider.

9. Scale the graphical representation so it fits the *Graphical representation panel* exactly with **Edit** > **Zoom to fit** or press the button.

10. In the *Graphical representation panel* click on the first mapping rule while holding the **Ctrl**-key.

11. Select **Edit** > **Transfer selection** or press the button.

12. Leave **Selected partition mapping rules** checked, choose **Select all violations for first partition** from the drop-down list and press <**OK**>.

In the *Entry selection panel*, a single entry is selected: entry G@Gel07@013, the only entry with PFGE type A that does not map on Serotype T1.

The *Partition comparison panel* provides a complete report of the partition mapping.

13. To copy the complete report, displayed in the *Partition comparison panel* to the clipboard as tab-delimited text choose **Edit** > **Copy report**.

5 Applying a partition mapping

If no information is available for a number of entries for one of the partitions, the classes can be predicted by means of the mapping rules. This information can either be stored in an information

field or a comparison group.

1. Select **File** > **Apply mapping** or press the ▷ button in the *Partition mapping* window.

2. For example, check **Forward mapping** as mapping to apply and apply the mapping to **all entries in the database**. Select the previously defined "Mapping" information field (see 3) as **Output field** and use the label from the **Second partition** (see Figure 5).
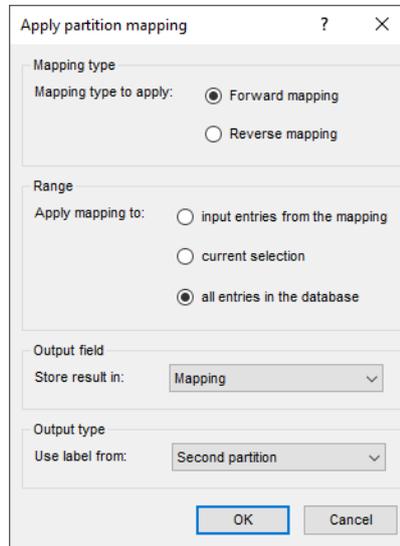

Figure 5: The *Apply partition mapping* dialog box.

3. Press <**OK**> and confirm the action.

The "Mapping" information field now contains the mapping information from the second partition (see Figure 6). For entries where the information in the "PFGE type" field cannot be mapped (e.g. for the STANDARD entries in the **DemoBase Connected**), the "Mapping" field will be left unaltered.
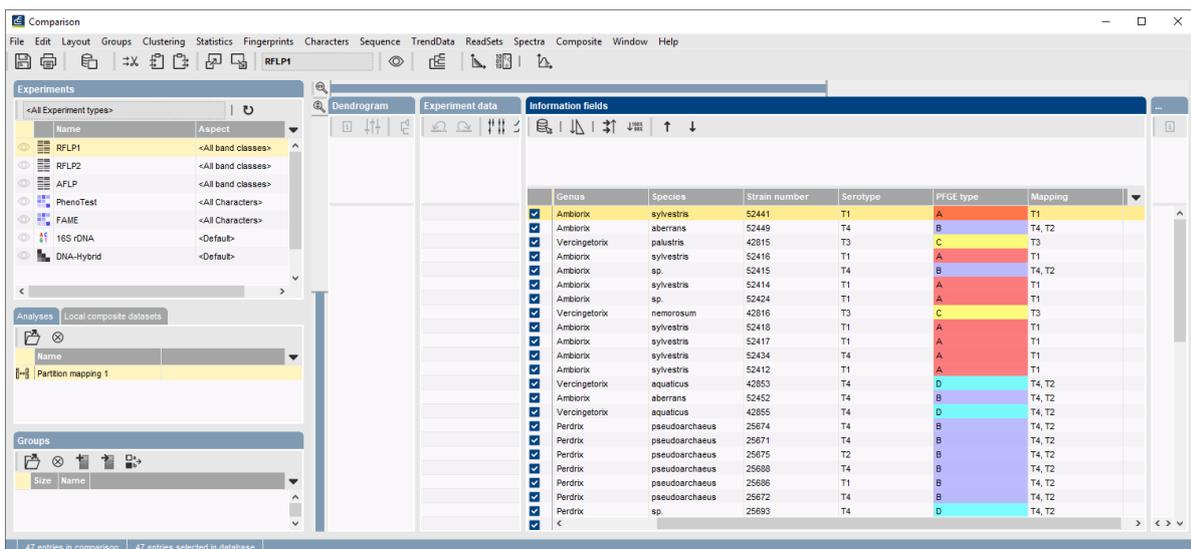

Figure 6: The *Comparison* window.

Instead of directly writing mapping information into an information field, a decision network can automatically be generated based on the partition mapping. Decision networks generated this

way can either be used to write mapping information to an information field or to select database entries.

4. Select ***File*** > ***Create decision network*** in the *Partition mapping* window or press the 🔀 button in the toolbar.

5. Enter **Forward Mapping** as decision network name and leave ***network for applying mapping rules*** selected.

6. Select ***Forward mapping*** as ***Mapping type***, "Mapping" as ***Output field*** and use the label from the ***Second partition***.

7. Press <***OK***> to create the decision network.

8. Click the *Decision networks tab* in the *Main* window and double-click **Forward Mapping** to open the decision network (see Figure 7).
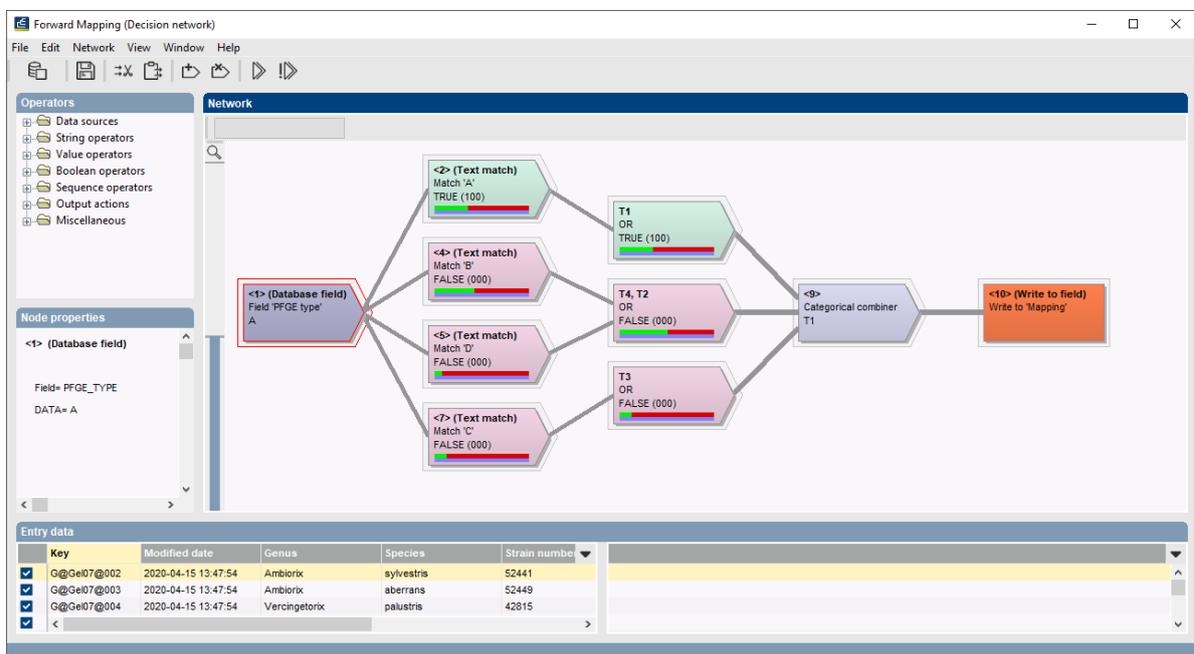
Figure 7: Decision network to apply the mapping rules of the example data.

9. For safety reasons, the output actions are not executed automatically when the network is calculated using the ▷ button. To calculate the network and execute the output action(s), press the ❗▷ button.

This decision network performs the same action as we did earlier via the *Apply partition mapping* dialog box, but has the advantage that it can be applied in a flexible way to any selection of entries.