

BioNumerics Tutorial:

Importing non-numerical character data

1 Aims

This tutorial shows how to import non-numerical data in a BioNumerics database and link the data to a character type experiment. It illustrates the use of *character mappings* in BioNumerics. Character mappings in BioNumerics are used to map categorical names (e.g. Present/Absent, Yes/No, Susceptible/Intermediate/Resistant, etc.) to character values (e.g. 0 and 1) or a range of values to according to predefined criteria.

2 Preparing the database

The **DemoBase Connected** will be used in this tutorial 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 **Download example databases** link, located in the lower right corner of 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 <http://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.

3 Example data

The example Excel file from which we will import data in this tutorial contains presence/absence information for four genes and can be downloaded from <http://www.applied-maths.com/download/sample-data>.

1. Open the file `Binary_data.xlsx` in Excel to examine the data that will be imported.

The sheet **Sheet1** contains for 47 isolates following information: a unique identifier (Strain number) and information about the presence (Yes) or absence (No) of four genes (Gene1, Gene2, Gene3 and Gene4) (see Figure 1).

2. Close the Excel file again.

	A	B	C	D	E	F	G	H
1	Strain number	Gene1	Gene2	Gene3	Gene4			
2	52441	Yes	No	No	Yes			
3	52449	Yes	Yes	No	Yes			
4	52416	Yes	No	No	Yes			
5	25680	No	No	Yes	No			
6	42815	No	Yes	No	No			
7	42816	No	Yes	No	No			
8	42853	No	Yes	No	No			
9	52435	Yes	No	No	Yes			
10	52433	Yes	No	No	Yes			
11	52413	Yes	No	No	Yes			
12	52422	Yes	No	No	Yes			

Figure 1: Part of the Excel file.

4 Creating a new character type

1. In the *BioNumerics Startup* window, double-click on the **DemoBase Connected** database to open it.

Since we will import character data, we will first create a character type to hold this data.

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 “Genes” and press **<Next>**.

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

4. Choose **Binary data** since only two possible states are present in our dataset: **Yes** (gene present) and **No** (gene absent). Press **<Next>**.

The wizard asks if the character type has an open (**Yes**) or closed (**No**) character set.

5. Answer **No** and press the **<Finish>** button to complete the setup of the new character type.

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

6. Double-click on the experiment **Genes** to call the *Character type* window.

In order to import the character states present in our Excel file (**Yes** and **No**) into our character type experiment, we need to specify character mappings that will map the two states to a character value. To be able to work with character mappings, the *Mapping* panel should be displayed in the *Character type* window.

7. Click on the *Mapping* panel.
8. To add a criterion, select **Mapping > Add new mapping...** This pops up the *Edit character map* dialog box.
9. In the *Edit character map* dialog box, enter “Yes” as **Name**, “1” as **Range start** and “1” as **Range end** (see Figure 2). Next, press **<OK>** to create the mapping.

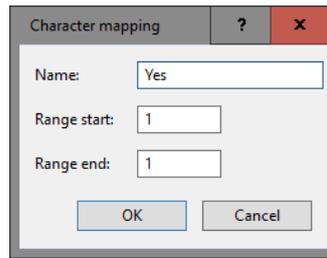


Figure 2: Create a new mapping.

10. In a similar way, create a “No” mapping (between 0 and 0) (see Figure 3).

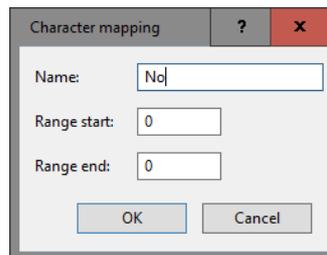


Figure 3: Create a new mapping.

The *Mapping* panel now lists the two character mappings (see Figure 4).

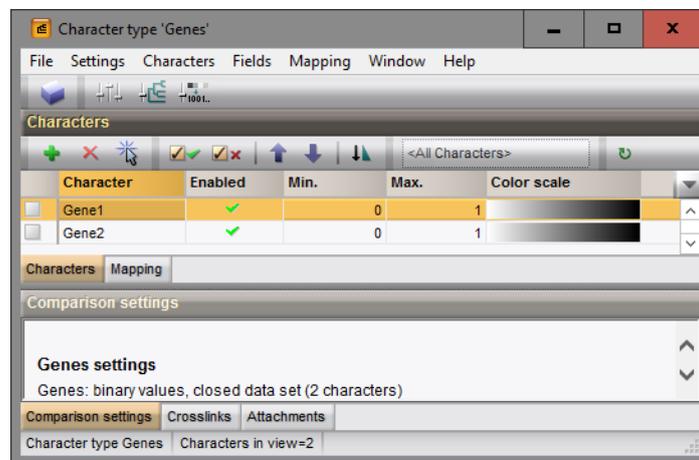


Figure 4: The *Character type* window with mappings defined.

11. Close the *Character type* window with **File** > **Exit**.

5 Import procedure

1. Select **File** > **Import...** (📁, **Ctrl+I**) to open the *Import* dialog box.
2. Choose the option **Import fields and characters (Excel file)** under the *Character type data* item in the tree (see Figure 5) and press <**Import**>.
3. Press <**Browse**> and browse for the downloaded `Binary_data.xlsx` file. Next, press <**Open**>.

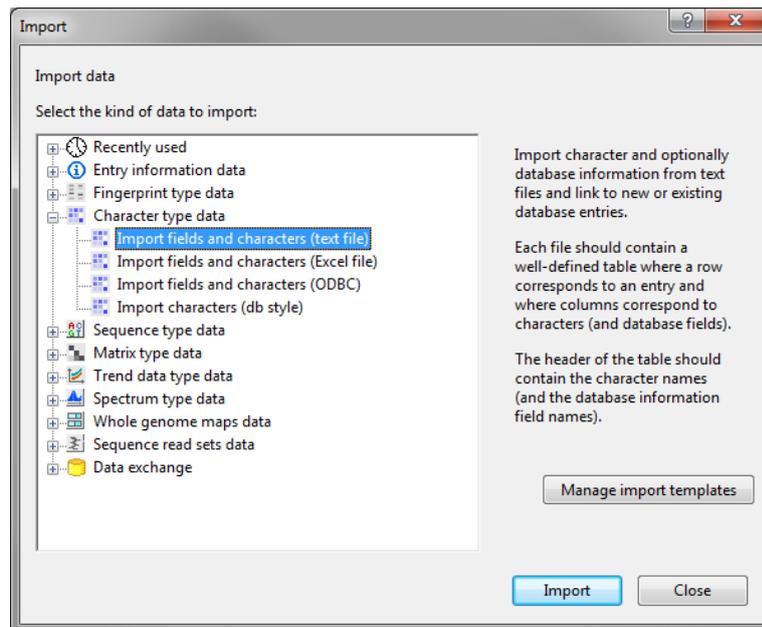


Figure 5: The *Import* dialog box.

4. Make sure the correct sheet is selected (i.e. "Sheet1") and press *<Next>*.

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

5. Select "Strain number" in the list and click *<Edit destination>* or simply double-click on "Strain number".
6. Select "Strain number" as the BioNumerics destination field in the *Edit data destination* dialog box under *Entry info field* (see Figure 6) and press *<OK>*. If this field is not present in the database, choose the *Create new* option under *Entry info field* and confirm the creation of the new field twice.

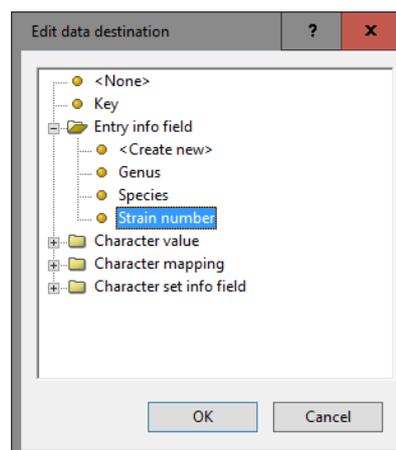


Figure 6: Strain number link.

7. Select "Gene1", "Gene2", "Gene3" and "Gene4" from the list by holding down the **Ctrl**-key. Click on *<Edit destination>*, select **Genes** under the **Character mapping** option (see Figure 7) as destination (NOT under the "Character value" option) and click *<OK>*.
8. Press *<OK>* and then *<Yes>* to confirm the creation of new characters.

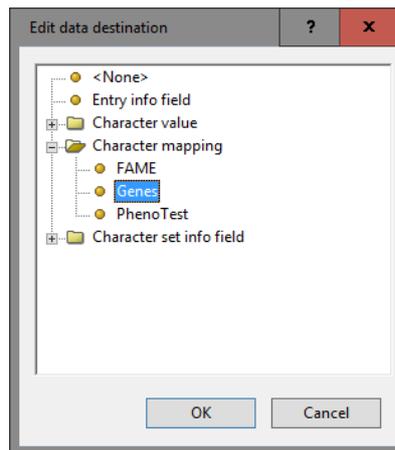


Figure 7: Link to character type experiment.

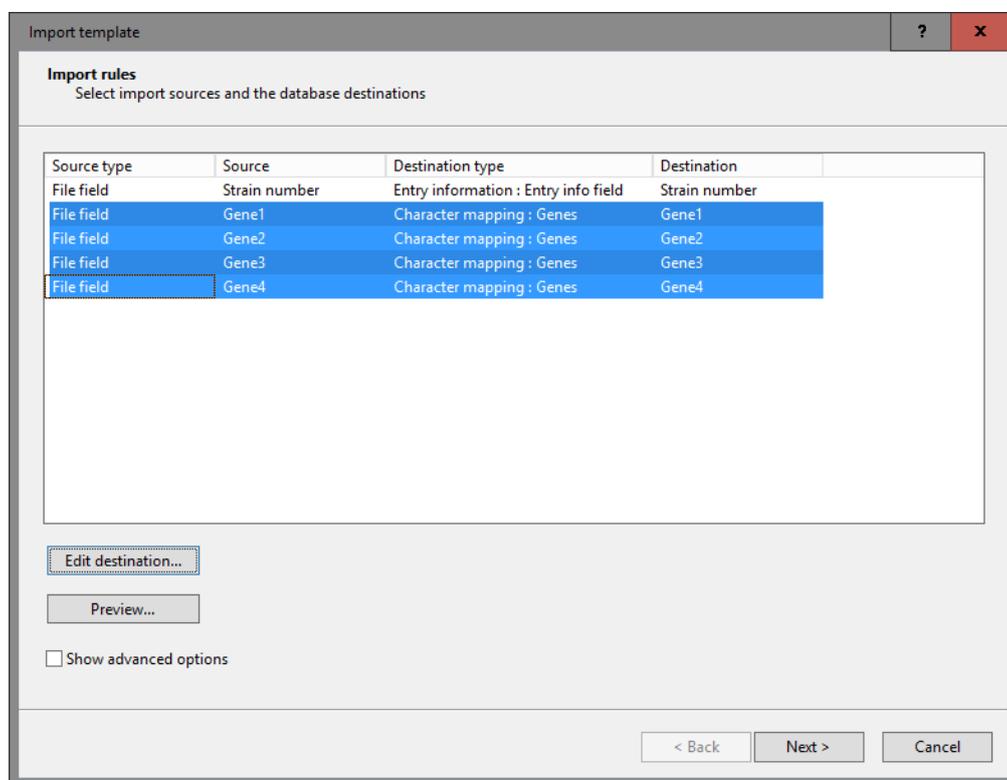


Figure 8: The import rules.

The grid panel is updated (see Figure 8).

9. Press <Preview> to see what you are about to import.
10. Press the <Close> button to close the preview.
11. Press <Next> and check the *Strain number* option.

The character data will be linked to an existing entry in the database if the Strain number in the Excel file and the BioNumerics Strain number correspond.

12. Press <Finish>, specify a template *Name* (e.g. "Import binary data") and press <OK>.

In the *Import template* wizard page, the newly created template is automatically selected.

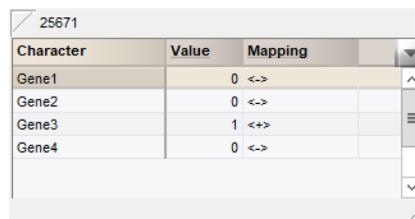
13. Click <*Next*>.

14. Press <*Finish*> to start the actual import. The progress of the import is shown while the data is added to the BioNumerics database.

The character data is stored in the character type **Genes**.

15. Double-click on the experiment **Genes** in the *Experiment types* panel, select *Settings* > *General settings...* (🔧), select the *Experiment card tab* and change the representation to *List*. Close the two windows.

16. Click on a green colored dot in the *Experiment presence* panel to open the **Genes** experiment card for an entry (see Figure 9).



Character	Value	Mapping
Gene1	0	<->
Gene2	0	<->
Gene3	1	<->
Gene4	0	<->

Figure 9: The character experiment card.

The **Mapping** column displays the data imported from the Excel file. The corresponding character values, based on the mappings defined in section 4, are displayed in the **Value** column.

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