

## BioNumerics Tutorial:

# Combined analysis of character data

## 1 Aim

---

In BioNumerics it is possible to obtain a combined dendrogram from multiple experiments. In this tutorial you will learn how to create a dendrogram based on character data coming from the two phenotypic tests defined in the **DemoBase Connected** database.

## 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 Creating a composite data set of character data

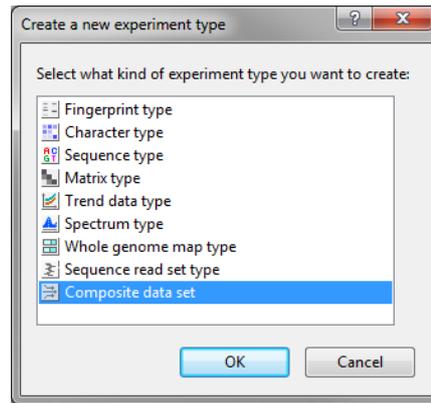
---

A composite data set in BioNumerics is a "container" of experiment types, i.e. it holds data coming from one or several experiments. In a first step we need to define a *composite data set*, containing the character type experiments as input.

1. In the *BioNumerics Startup* window, double-click on the **DemoBase Connected** database to open it.
2. In the *Main* window, click on  in the toolbar of the *Experiment types* panel and select **Composite data set** from the list.
3. Press **<OK>**, enter a name, for example "All-pheno" and press **<OK>**.

The *Experiment types* panel now lists the composite data type **All-pheno**.

4. Double-click on **All-pheno** in the *Experiment types* panel.



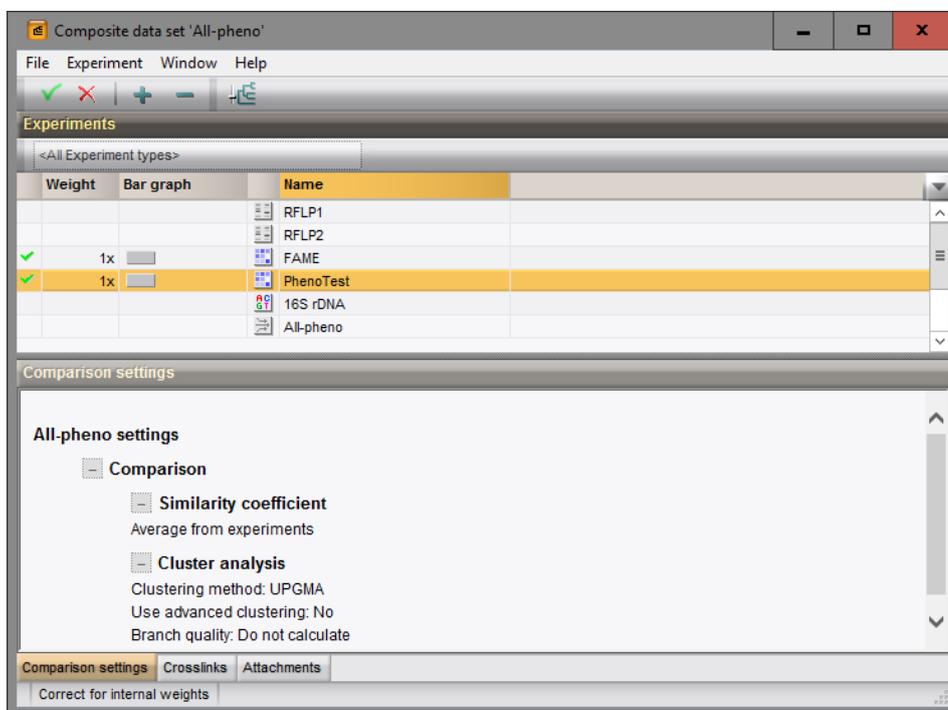
**Figure 1:** Create a new composite data set.

The *Composite data type* window is shown for **All-pheno**.

5. Click on **FAME** and select *Experiment > Include experiments* (✓).
6. Click on **PhenoTest** and select *Experiment > Include experiments* (✓).

Both FAME and PhenoTest are selected in the composite data set, and are marked with a green V-sign (see Figure 2).

The bar that appears in the 'Bar graph' column displays the weights assigned to each experiment type. Weights can be increased or decreased with *Experiment > Increase weight* (+) and *Experiment > Decrease weight* (-) respectively.



**Figure 2:** The Composite data set.

In order to treat individual characters on an equal basis while averaging matrices, BioNumerics can automatically use weights proportional to the number of tests each experiment contains. This correction is achieved

as follows:

7. Select *Experiment* > *Correct for internal weights*. The caption now shows "Correct for internal weights".
8. Close the *Composite data type* window with *File* > *Exit*.
9. Double-click on the character type **FAME** in the *Experiment types* panel and select *Settings* > *Comparison settings...* (⊕) to access the settings. Make sure *Euclidian distance* is selected and press <OK>. Close the *Character type* window.
10. Double-click on the character type **PhenoTest** in the *Experiment types* panel and select *Settings* > *Comparison settings...* (⊕) to access the settings. Make sure *Pearson correlation* is selected and press <OK>. Close the *Character type* window.
11. In the *Database entries* panel of the *Main* window, select all entries except STANDARD: select **Ctrl+A** to select all entries and use the **Ctrl-** key to unselect the entries defined as STANDARD. Alternatively unselect the entries by unchecking the check boxes next to the entries.
12. Highlight the *Comparisons* panel in the *Main* window and select *Edit* > *Create new object...* (+) to create a new comparison for the selected entries.

## 4 Combined analysis of character data

---

1. Click on the (◀) next to the experiment **All-pheno** in the *Experiments* panel to display the characters in the *Experiment data* panel.
2. Select *Composite* > *Show quantification (values)* (123) to display the character values.
3. Select *Composite* > *Show presence/absence* (☐) to display the presence and absence states again.
4. Make sure the experiment **All-pheno** is selected in the *Experiments* panel and choose *Clustering* > *Calculate* > *Cluster analysis (similarity matrix)...*

When selecting the option *Average from experiments* (see Figure 3), the similarity matrices of the individual experiments (in this example **PhenoTest** and **FAME**) are calculated first using the comparison settings defined for each individual experiment (double-click on the character type in the *Experiment types* panel and select *Settings* > *Comparison settings...* (⊕) to access the settings). From these matrices, a combined matrix is calculated by averaging the values. The averaging can happen in two ways: each value can be considered equally important, or BioNumerics can assign a weight proportional to the number of tests in an experiment ("correct for internal weights"), or the user can define an extra weight for each experiment manually.

If the combined experiments are comparable in terms of biological meaning, reaction type and numerical range, a coefficient under *Character data* can be selected. All characters from the different experiment types are merged to one character table and based on this character table a similarity matrix is calculated.

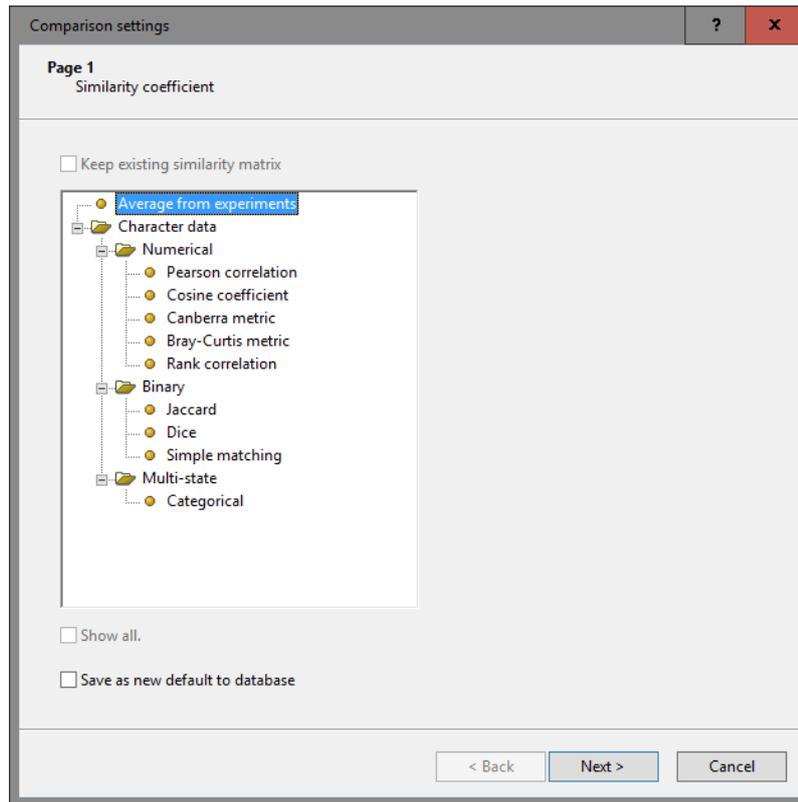
5. The character sets of **PhenoTest** and **FAME** have different ranges, so select *Average from experiments* and press <Next> and <Finish>.

The dendrogram is displayed in the *Dendrogram* panel (see Figure 4).

6. To view the settings used for the calculation of the individual matrices, select *Clustering* > *Show information* (i).

Comparison groups can be defined from clusters, from database fields, or just from any selection you want. As an example, we will let BioNumerics create groups based on the **Genus** names.

7. In the *Comparison* window, right-click on the field name **Genus** in the *Information fields* panel, and select *Create groups from database field*.



**Figure 3:** Similarity coefficients.

8. Keep the first option selected and confirm.

In our example three groups are created. The groups are listed in the *Groups* panel. The group color is displayed next to each entry in the *Information fields* panel (see Figure 4).

9. Select **Clustering > Dendrogram display settings...** (🔧) to call the *Dendrogram display settings* dialog box.
10. Enable **Show group colors** and press <OK>.

The dendrogram branches are now colored according to the group colors.

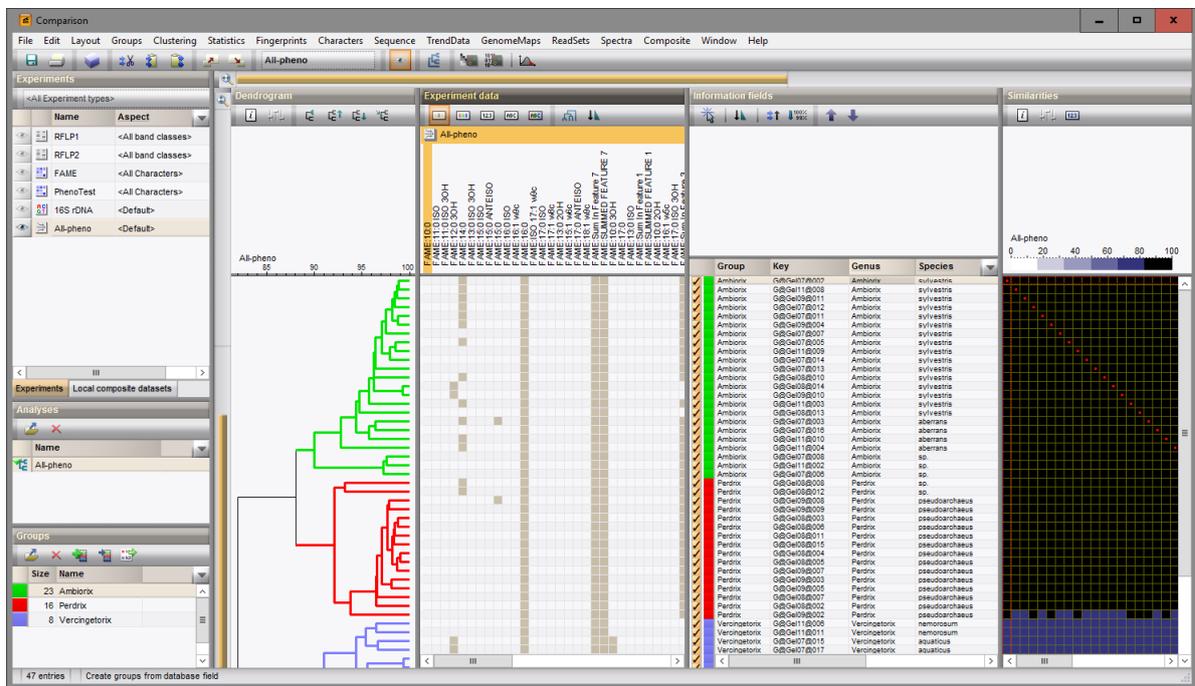


Figure 4: The Comparison window.