BioNumerics Tutorial:

Summarizing spectra

1 Aim

This tutorial illustrates how to combine several spectra to create summary spectra, according to the levels and dependencies present in the database. How to import the spectra and obtain the correct levels and dependencies is explained in the tutorial: "Importing and preprocessing raw spectrum data".

2 Preparing a sample database

1. Create a new database and import the example raw spectra files as described in the tutorial: "Importing and preprocessing raw spectrum data".

When working with levels, it is easier to switch between levels if the *Database design* panel is docked above the *Database entries* panel:

2. Click on the *Database design tab* and - while keeping the mouse button pressed - drag it upwards in the *Database entries* panel. Drop the floating panel on the top part of the docking guide that appears.

The *Database design* panel is now shown above the *Database entries* panel (see Figure 1). The entries belonging to the selected level are displayed in the *Database entries* panel.
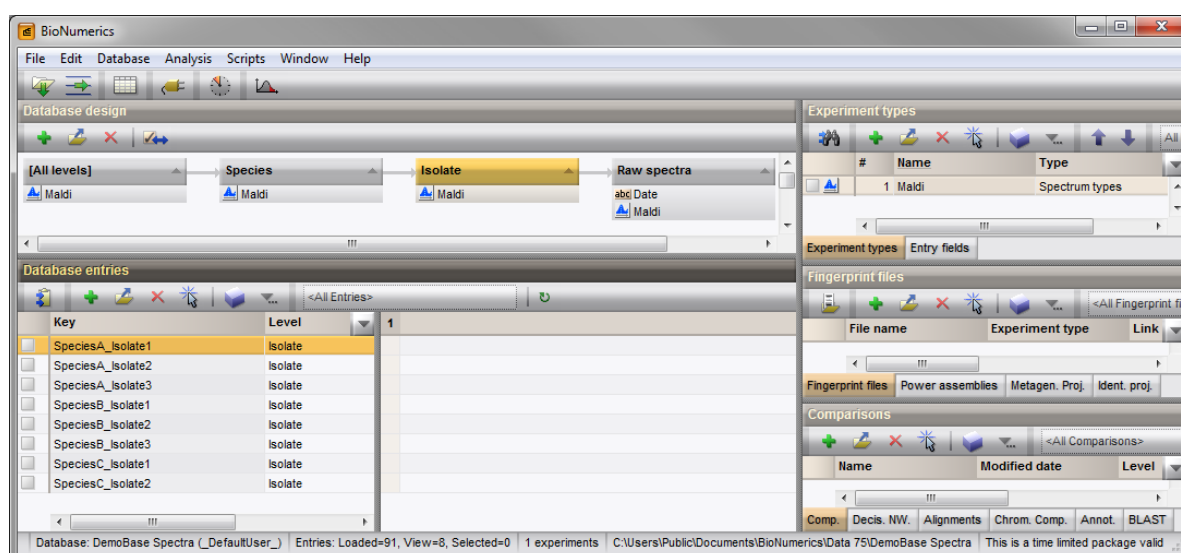

Figure 1: *Database design* panel shown above the *Database entries* panel: the entries belonging to the level 'Isolate' are shown in the *Database entries* panel.

3. Verify the creation of the correct dependencies in each level, by double-clicking on an entry to open the *Entry* window. Examples of how the dependencies should look like can be found in Figure 2.
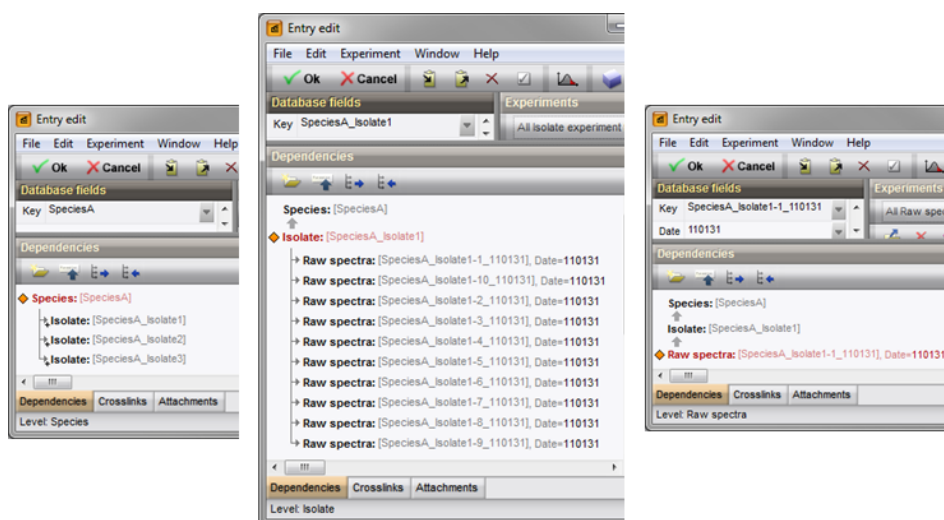
Figure 2: Example of dependencies at different levels.

3   Creating summary spectra from raw spectra

1. At the lowest level, "Raw spectra", select all entries (***Edit*** > ***Select all*** (**Ctrl+A**)) and select ***Analysis*** > ***Spectrum types*** > ***Summarize...*** to open the *Create summary spectra* wizard.

2. In the first page of the *Create summary spectra* wizard, select the option *Higher level* and press *<Next>*.

3. In the second page, select the level "Isolate" as target level and press *<Next>* twice.

In the fourth step of the wizard (see Figure 3), the user can select an appropriate workflow for the summary of the spectra.
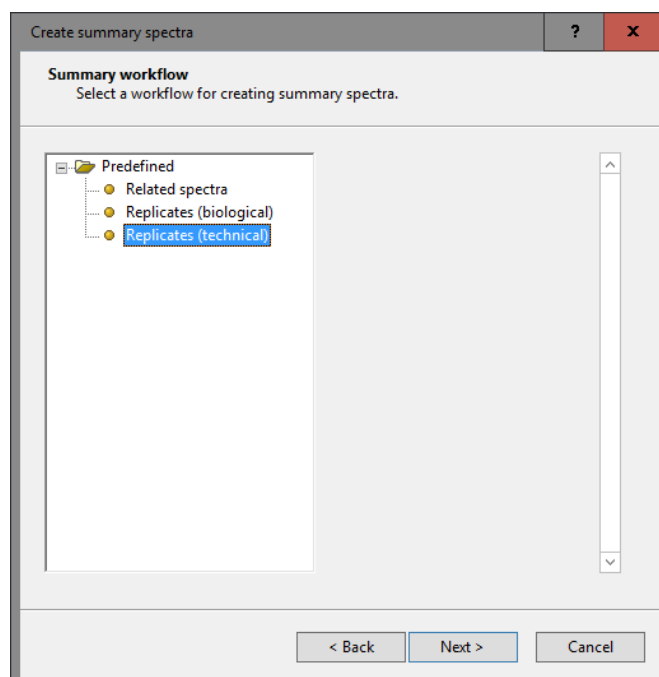

Figure 3: Summary templates.

Three predefined summary templates are present and a short description of the selected template can be

found in the panel on the right: one for technical replicates (***Replicates (technical)***), one for biological replicates (***Replicates (biological)***)) and one general template for processing for instance several strains of the same subgroup or the same species (***Related spectra***). The settings of each template can be changed in the *Summary Spectrum* window and saved to the database (see further).

4. Select the predefined summary workflow ***Replicates (technical)*** and press <***Next***> and <***Finish***>.

The *Summary Spectrum* window will now open with the raw spectra organized into 8 summary spectra (see Figure 5).

5. Select a summary spectrum from the list in the *Summary Spectra* panel.

The member spectra are updated in the *Members* panel and in the *Peak List* panel. The average similarity of all member spectra is calculated and the similarity of each individual spectrum with this average similarity is indicated in the ***Simil*** column. Member spectra that are grayed out have been inactivated because their similarity (***Simil***) is below the similarity threshold that is specified. The similarity threshold depends on the selected summary template:

- ***Replicates (technical)***: similarity threshold placed at 95%

- ***Replicates (biological)***: similarity threshold placed at 80%

- ***Related spectra***: placed at 40%

6. To check (and optionally update) the similarity threshold, double-click on the ***Import & Similarity filter*** operator in the *Workflow* panel to call the *Workflow* window. Select the ***Similarity filter*** operator from the flow chart, and choose ***File*** > ***Operator settings...*** (![icon]). The similarity threshold can be updated by the user and the updated template can be saved to the database with ***File*** > ***Save workflow as template***. Close the *Workflow* window.
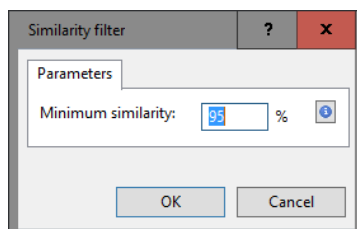

Figure 4: The similarity threshold for the technical replicates template.

7. Press <***Peak Matching***> to calculate the summary spectra and to identify summary peaks.

After execution of the workflow, the *Summary Spectrum* window will look like Figure 6. The summary spectra will appear as a thick red line. Each peak from the summary spectrum appears as thin vertical line, adjacent peaks are marked with a different color.

8. Save the summary spectra to the database with ***File*** > ***Save spectrum data*** (![icon]).

Each entry at the level **Isolate** will now have an experiment present for the spectrum type **Maldi**.

4 Comparison of member spectra

To review the quality of the raw spectra and the subsequent summary, it is possible to create a comparison with all the member spectra with ***Summary*** > ***Create comparisons*** (![icon]).

1. In the *Create summary comparisons* dialog box, select ***Summary*** > ***Create comparisons*** (![icon]) and select the option to create a comparison from the member spectra and press <***OK***>.
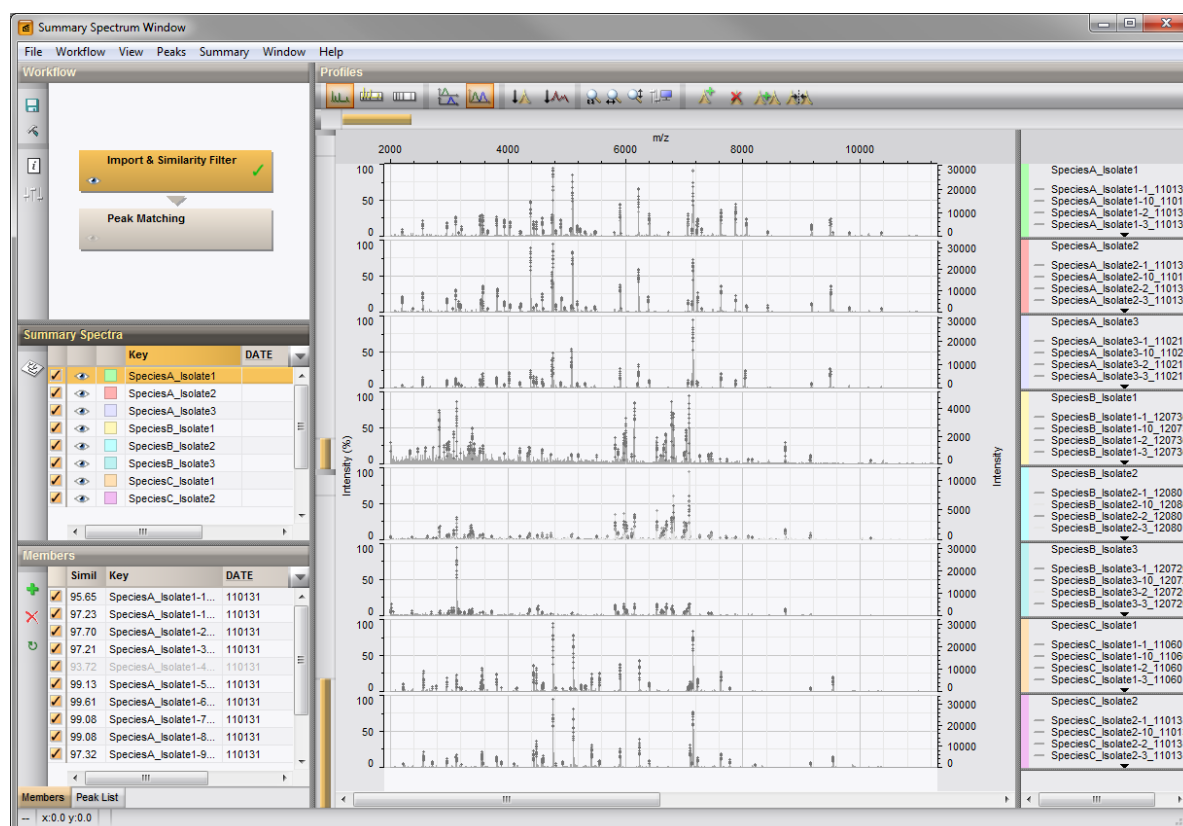
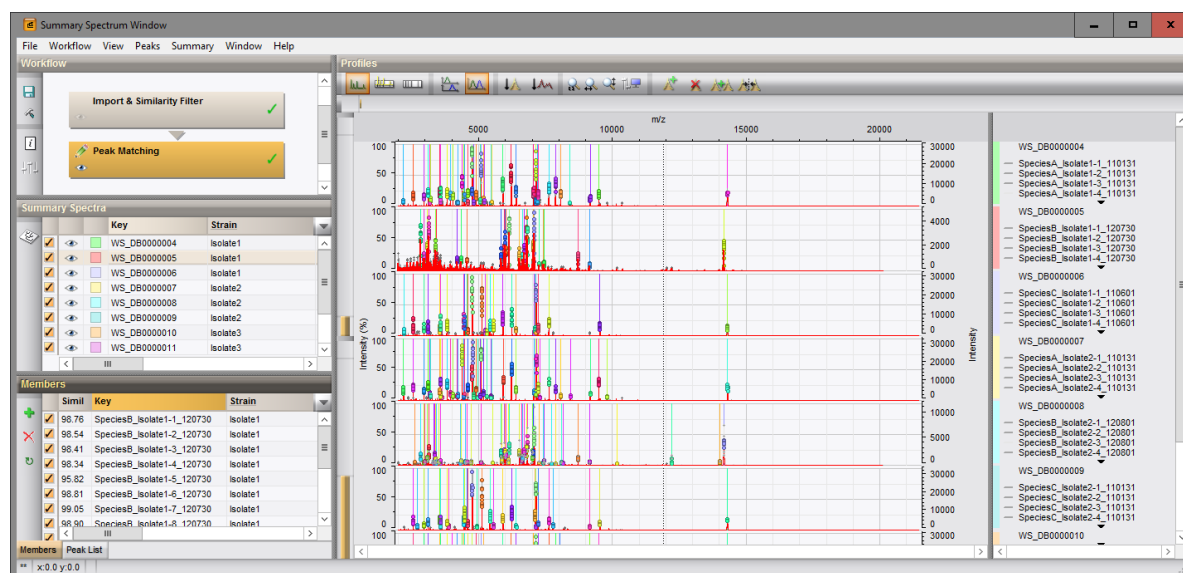**Figure 5:** *Summary Spectrum* window before running the workflow.

Figure 6: *Summary Spectrum* window after running the workflow.

This action will open a comparison window including all member spectra grouped according to their summary spectra.

2. Select *Clustering* > *Calculate* > *Cluster analysis (similarity matrix)...* to calculate a *UPGMA* tree based on the curve-based *Pearson correlation* coefficient.

In the resulting dendrogram the member spectra are clustered together with the other members. If the member spectra of several summary spectra are clustered together, either the preprocessing or summary

workflow need to be reviewed. In case of closely related isolates, it might also be that they cannot be distinguished from one another based on their spectra. In this case, the two isolates from Species C are very similar and it is likely that the distinction between them will be less reliable than the distinction between the isolates from Species A (see Figure 7).
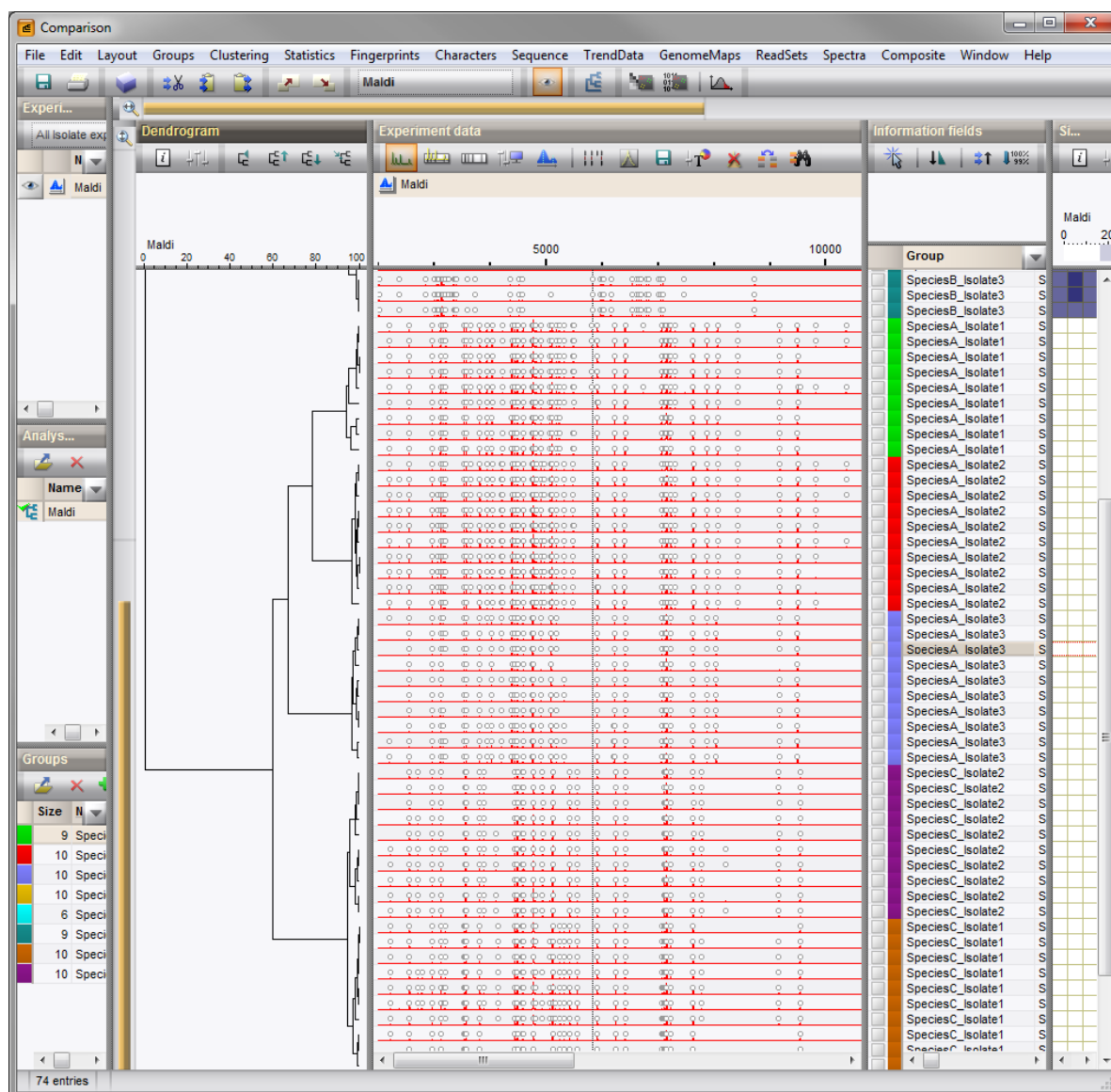

Figure 7: *Comparison* window of member spectra generated from the *Summary Spectrum* window.

5   Creating summary spectra from summary spectra

1. Close the *Comparison* window (save if desired) and the *Summary Spectrum* window.

2. Unselect all entries in the *Main* window by pressing the **F4** key.

The summary spectra of the isolates can themselves by summarized per species:

3. Go to the level "Isolate", select all entries and select ***Analysis > Spectrum types > Summarize...***. Select the option *Higher level* and press *<Next>*. Choose "Species" as target level and choose the workflow "Related spectra". Leave all other settings at default.

4. Press <*Peak Matching*> to execute the summary workflow. For these spectra, the tolerance settings are a bit to strict, this is seen best in the summary spectrum of Species A (see Figure 8). To change the tolerance settings, double-click on <*Peak Matching*> to open the *Workflow* window. Highlight peak matching and select *File > Operator settings...* (⊞). This will open the dialog box with the parameters of the Peak matching operator. Increase the constant tolerance from 1 to 2 and increase the linear tolerance from 300 ppm to 400.

5. Close the *Workflow* window, rerun the workflow by pressing <*Peak Matching*> and save the results with *File > Save spectrum data* (🖫).

The entries at the level *Species* will now also have an experiment present for *Maldi*. All summary spectra can be used for follow-up analysis such as peak matching, determining common or discriminative peaks, identifying unknown spectra against a database, etc..
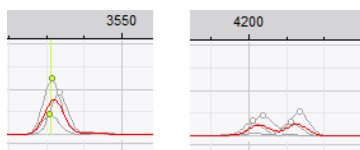

Figure 8: Summary spectra of Species A at default tolerance settings.