

## BioNumerics Tutorial:

# Peak matching and follow up analysis of spectra

## 1 Aim

This tutorial aims to familiarize the user with the process of peak matching and also to give the user some examples of possible follow up analyses available with peak matching.

There are three important terms for the peak matching: peak, peak class and peak class view. A peak is defined on the level of the spectrum during preprocessing, performing a peak matching does not make any changes to the defined peaks. A peak class is defined on a group of spectra and is similar to the band classes for fingerprint types. During peak matching, peaks will be assigned to a peak class. A set of peak classes can be stored as a peak class view. Several peaks class views containing different peak classes can be defined and stored in your database.

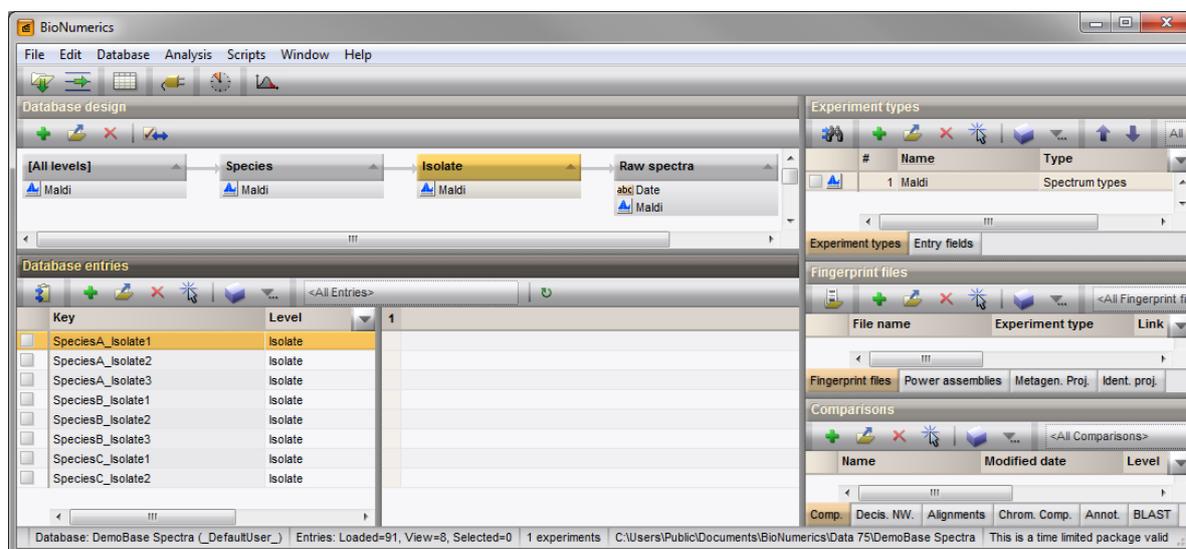
## 2 Preparing a sample database

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.

In this tutorial, we will be working on the data in the lowest level, raw spectra, though the techniques may also be applied to summary spectra. As most analysis will be performed to distinguish between the species, it is easy to have this information in a field at the level 'Raw spectra'.

3. Click on the level 'Isolate' in the *Database design* panel and press **Edit > Create new object...** (👤) in the *Entry fields* panel. In the *Create new entry information field* dialog box, fill in the name 'Species' and check the second option, **Calculate field content from other fields**. Press <Edit> to open the *Calculated database field settings* dialog box and here, press <Add information field>. In the *Add information field* dialog box, select '<Parent Key>' and then press <OK> three times to create the information field.
4. Right-click on this new information field in the *Database entries* panel and select **Field properties** to open the *Database field properties* dialog box and press <Level assignment>.
5. In the *Level assignment* dialog box, use the drop down arrow in the **Assignment** column at the level 'Raw spectra' to assign the field as 'Replicated' (see Figure 2). Press <OK> twice.

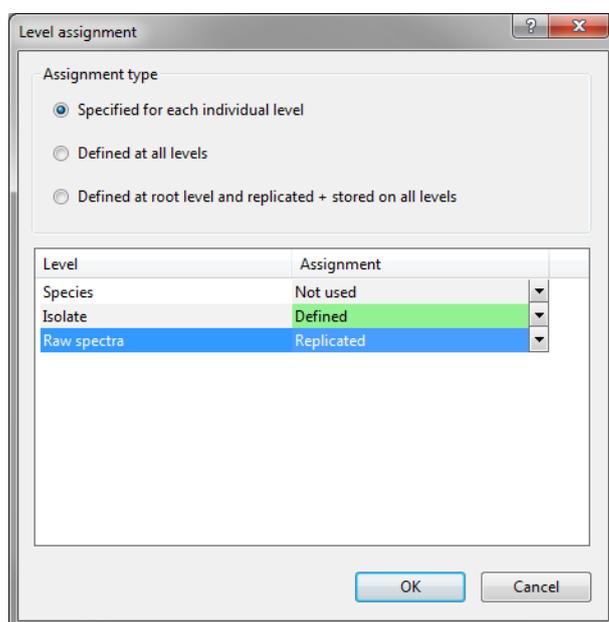


Figure 2: Level assignment of information field 'Species'

### 3 Peak matching

1. Select all 80 entries at the lowest level 'Raw spectra' and press **Edit > Create new object...** (👤) in the *Comparisons* panel to create a comparison containing all the imported spectra.
2. Click on the spectrum type **Maldi** in the *Experiments* panel and select **Layout > Show image** or press the eye button (👁) next to the experiment name in the *Experiments* panel.
3. Select **Spectra > Do peak matching** (👤).

This pops up the *Peak class matching* wizard. The only option currently available is **Recreate peak classes**. This will create new peak classes and add these to the default peak class type.

4. Pressing <Next> will open the second page of the *Peak class matching* wizard (see Figure 3). Fill in a constant tolerance of "1.9", a linear tolerance of "550" and a peak detection rate of "10%" and press <Finish>.

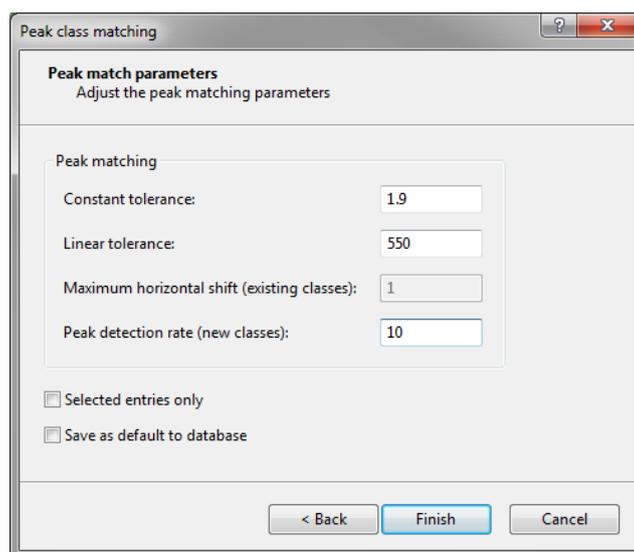


Figure 3: Second page of *Peak class matching* wizard

With a peak matching present, the user can proceed to a follow up analysis. In order to visualize the results better, groups can be created based on the Species information field.

5. Right-click on the header of the information field 'Species' in the *Comparison* window and select **Groups** > **Create groups from database field**. Leave all settings at default and press <OK> to create three groups based on the three species.

## 4 Exporting a peak matching table

Exporting a peak matching table cannot be performed directly on the spectral experiment. The peak matching table is character data derived from the spectral type and can be accessed using composite datasets.

1. Save the comparison with **File** > **Save as...** and close the *Comparison* window with **File** > **Exit**.
2. In the *Experiment types* panel, select **Edit** > **Create new object...** (+) to create a new experiment type and select **Composite data set** from the list. Name the new composite dataset "MALDI" and press <OK>.
3. Open the properties of the new experiment type by double clicking on it in the *Experiment types* panel.
4. In the *Composite data type* window, select the spectral type **Maldi** and select **Experiment** > **Include experiments** (✓) to base the composite dataset on our spectral type. Close the *Composite data type* window.
5. Open the comparison again in which the peak matching was performed in the previous step.
6. Click on the composite dataset **MALDI** in the *Experiments* panel and select **Layout** > **Show image** (👁) or press the eye button (👁) next to the experiment name in the *Experiments* panel.

The icons displayed at the top of the *Experiments* panel will determine how the data is visualized and exported. The first icon (👁) will result in a binary representation, with only absence and presence of the peak classes shown. The second (🌈) and third icon (123) result in a representation of the intensity values as color or as value respectively.

7. The information displayed can be exported to a tab delimited file by selecting **Composite** > **Export character table...**

Depending on the visualization, the exported file will contain either a binary peak matching table (presence/absence of peaks) or a peak matching table containing the intensity values.

## 5 Principal component analysis of peak classes

Principal component analysis (PCA) is a powerful technique that can be used in this context to reduce the complexity of the data and make it easier to identify groups and visualize the data in two or three dimensions. PCA works on character sets only, so to apply this technique to spectral types, it is necessary to perform the peak matching first.

1. To perform a PCA analysis on our spectra, make sure the spectral experiment 'Maldi' is highlighted in the *Experiments* panel and select *Statistics > Principal Components Analysis...* (🔍). Leave all settings at default (see Figure 4) and press <OK>.

This will start the calculation of the PCA and result in Figure 5.

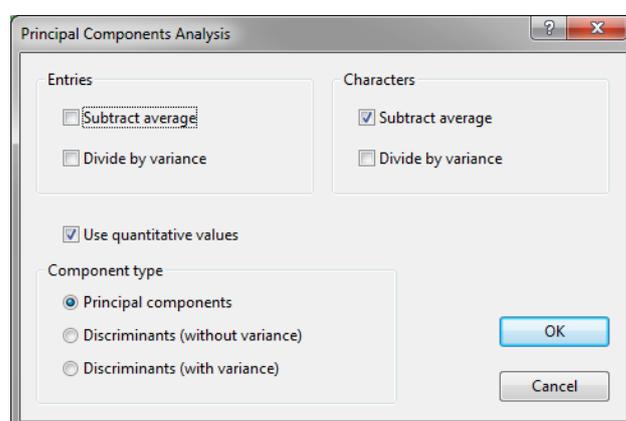


Figure 4: Settings to perform a PCA

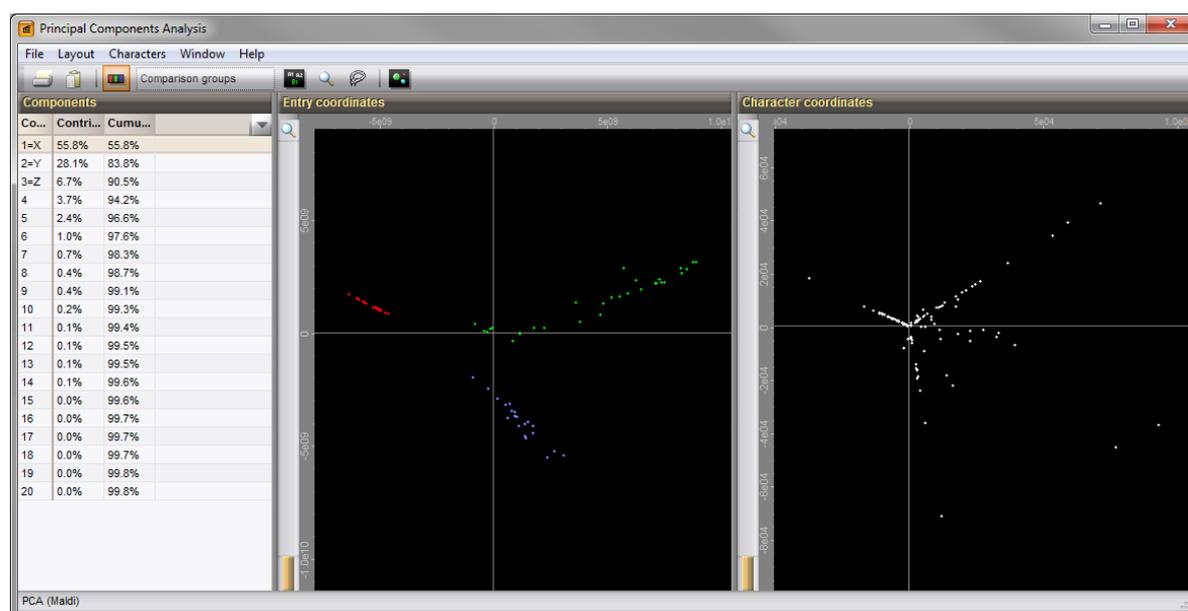


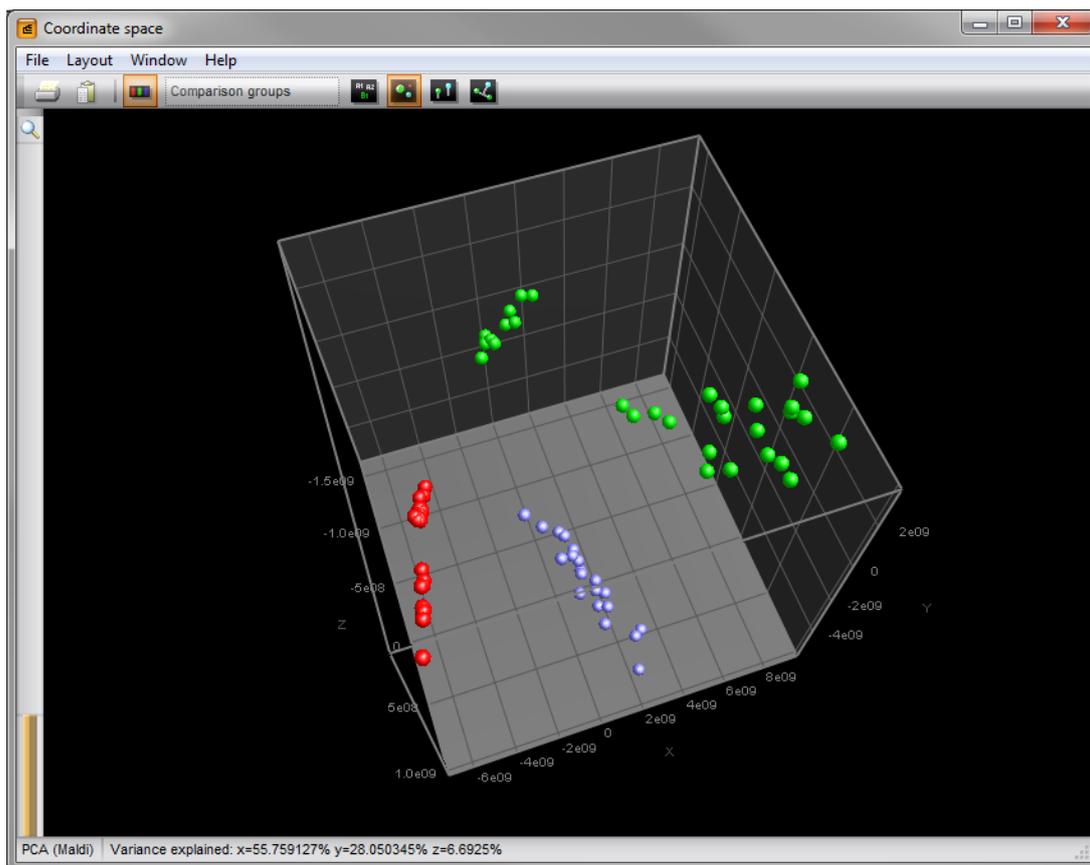
Figure 5: 2D results of PCA

In the 2 dimensional images, the spectra from different species are clearly separated. For species A and B, there are peaks that are specifically linked to these species. This can be seen by looking at the coordinates

of both the entries and the characters. All entries for species A can be found in the second quadrant. There is also a group of peaks located in this second quadrant, they are likely linked specifically to species A. The same can be seen for species B, but not for species C. This species will likely be defined by the combination of several peaks, that can also be present individually in other species.

2. To obtain a three dimensional view of the PCA analysis of the entries, select **Layout > Show 3D plot** (  ).

In the 3D view (see Figure 6) the same can be seen, the three species form distinct groups. Species B shows the lowest variance with the entries grouped closest together, species A shows the highest variance. For both species, distinct subgroups can be seen, possibly correlated to individual isolates.



**Figure 6:** 3D results of PCA

A PCA allows for a good visualization of the data, resulting in a quick visual interpretation of the data.

## 6 Analysis of peak classes using the matrix mining

The functionalities available in the *Matrix Mining* window are very useful for spectra. It can be used to make sub-selections of peaks with certain characteristics and perform statistical analysis on the peak classes. Similar to the PCA, the *Matrix Mining* window works with character data, meaning a peak matching is required to use these functionalities. It is not the aim of this tutorial to offer an exhaustive description of all functionalities available in the *Matrix Mining* window, but to offer some examples that can provide useful conclusions for analysis of spectral data.

1. Make sure the spectral experiment **Maldi** is highlighted in the *Experiments* panel and select **Statistics > Matrix mining...** in the *Comparison* window. This will open the *Matrix Mining* window (see Figure 7).

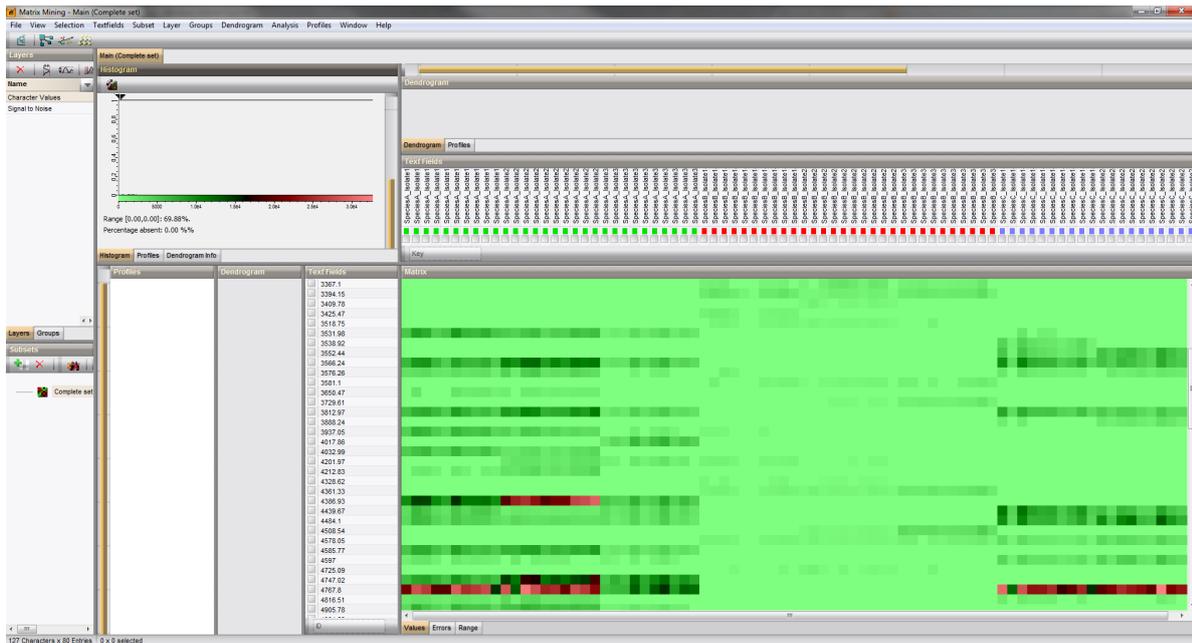


Figure 7: The *Matrix Mining* window

In the **Matrix panel**, the intensity of the peaks matched to the peak classes is represented by colors, green meaning low intensity, red high intensity. A two-way clustering can be performed to produce a heat map:

2. Select **Analysis > Cluster analysis...** (🔍), make sure **Characters** is selected, leave the remaining settings at default and press **<Next>**.
3. In the **Calculate dendrogram** dialog box, select the similarity coefficient **Spearman Ranks** and press **<Next>**.
4. In the last page, leave the settings at default and press **<Finish>**.
5. Repeat these steps choosing **Entries** in the first step to obtain the two-way clustering.

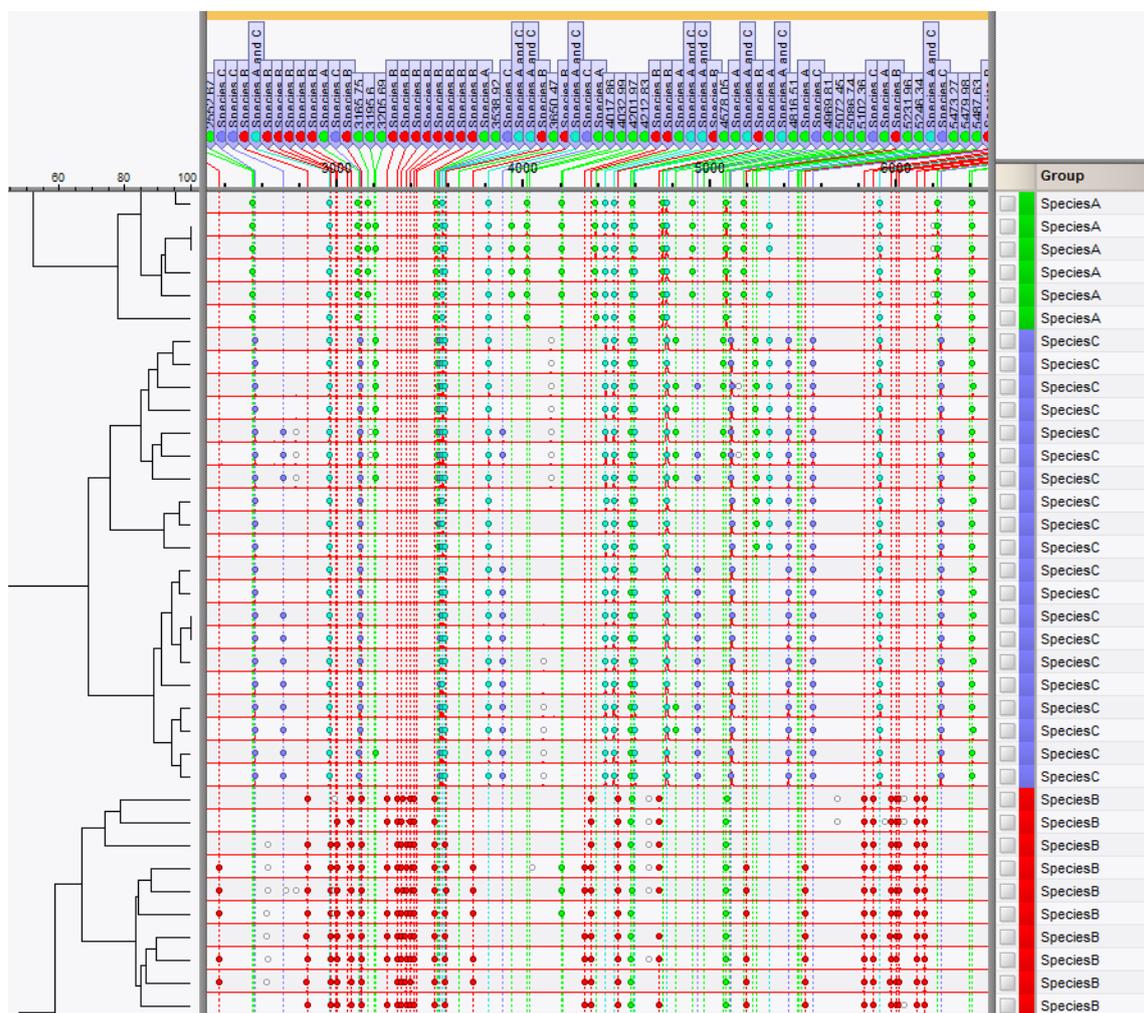
From the two-way clustering, it can very quickly be deduced that species B has very few peak classes in common with species A and C. Species A and C do have a set of peak classes in common, but each species also has a set of species specific peak classes. Some sets only belonging to a certain isolate can also be seen. As the selection of peak classes is synchronized between the *Matrix Mining* window and the *Comparison* window, we can use the *Matrix Mining* window to select peaks sharing a certain characteristic. Selecting peak classes in the *Matrix Mining* window is very similar to selecting entries in the *Comparison* window.

6. For instance, select all peak classes belonging to species B by holding down the **Ctrl**-key and clicking on the node in the dendrogram of the peak classes that contains all peak classes specific for species B.

Go back to the *Comparison* window (leave the *Matrix Mining* window open), the peak classes selected in the *Matrix Mining* window are also selected here. We can modify the properties of these peak classes to define that these peak classes are specific for species B.

7. Select **Spectra > Manage peak class types** (🔍), go to the tab **Custom Fields** and press **<Create New>**. Enter the name 'Species' and press **<OK>** and **<Close>**.
8. Next, select **Spectra > Display settings** (🔍), choose the field 'Species' as display label and press **<OK>**.
9. Now we will edit this new field for the selected peak classes by pressing **Spectra > Edit peak class properties** (🔍). Fill in 'Species B' as property for Species and optionally change the color. Press **<OK>**.

All selected peak classes will have the label species B and will be colored, including the peaks matched to these classes. The same procedure can be followed for the peak classes common to both species A and C, specific for Species A and specific for species C, resulting in an image similar to Figure 8.



**Figure 8:** The *Comparison* window after editing the peak class properties based on the two-way clustering

Default all peak classes found after peak matching are displayed in the *Comparison* window: the view in the *Aspect* column in the *Experiments* panel is set to *<All peak classes>*. It is possible to store a subset of peak classes in a new peak class *view*. Consider the case that these three species are hard to distinguish phenotypically, but species C is a pathogen and species A and B are harmless commensals. In this case, we will only be interested in the peak classes that reliably distinguish species C from species A and B.

10. Go back to the *Matrix Mining* window and select *Profiles > Statistics wizard...* (🔧).

This will open the first step of the *Statistics Wizard* dialog box.

11. Under **Orientation**, select the first option, to *Calculate a statistic for each Character* and press *<Next>*.
12. In the second step, select *Mann-Witney test* under **Independent tests (two groups)** and press *<Next>*.
13. In the third step, choose **Species C** as group 1 and **All other groups** as group 2 and press *<Next>* and *<Finish>*.

In the profiles panel, there is now a profile present with the p-value from our Mann-Witney test. All peak

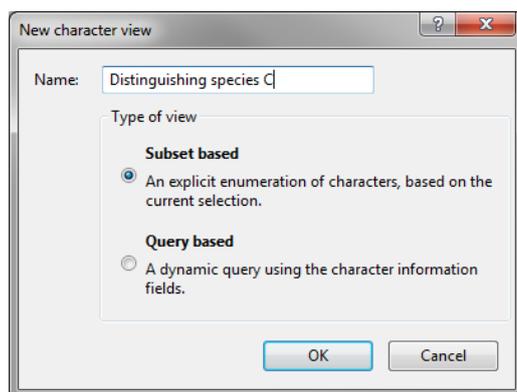
classes with a p-value lower than 0.05 are significantly different between species C and the other two species.

14. To select peak classes significantly different between species C and the other species, make sure the profile with the p-values is highlighted in the profiles panel and go to **Profiles > Selected profile > To query...** (🔍). In the *Profile To Query* dialog box, select '<=' in the first box and fill in 0.05 in the second box and press <OK>.

15. Go back to the comparison.

All peak classes now selected are significantly different and can be used to distinguish between species C and the other two species. This set can be stored as a new peak class *view*.

16. Select **Spectra > Manage views...** (🔍) and press <Add>. Name the new peak class view 'Distinguishing species C' (see Figure 9) and press <OK> and <Exit>.



**Figure 9:** New subset based view.

The new *view* is automatically selected in the *Aspect* column in the *Experiments* panel. Defining peak class views containing a subset of peaks can be important for certain analyses, such as clustering with a composite dataset based on the peak matching. In identification projects based on spectra, peak class views can be used to base the classification on.