BioNumerics Tutorial:

Creating a minimum spanning tree based on MLVA repeat numbers

1 Aim

In this tutorial we will create a minimum spanning tree based on MLVA repeat numbers. We will also see how we can alter the layout of the minimum spanning tree and how to export the picture to use it in a publication, presentation, etc.

2 Preparing the database

1. Create a new database (see tutorial "Creating a new database") or open an existing database.

2. Import the MLVA repeat numbers from the example text file `MLVA repeat numbers.txt` as described in the tutorial: "Importing MLVA repeat numbers from a text file". This sample file contains repeat numbers for about 500 strains.

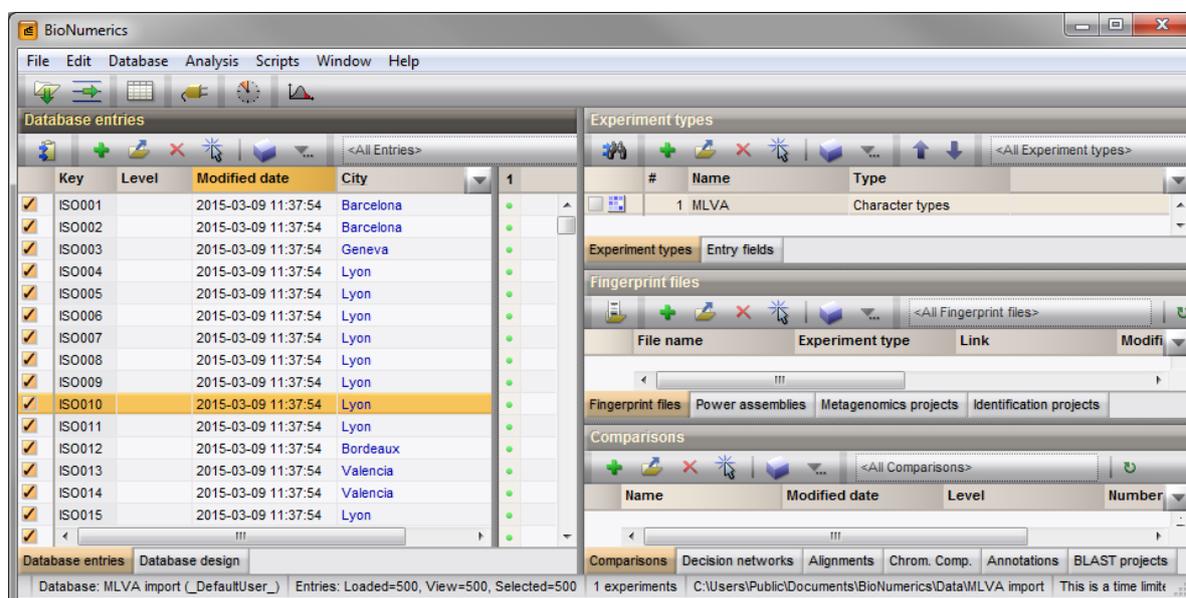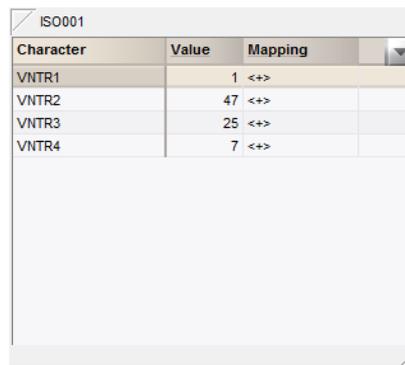After import the *Main* window should look like Figure 1.

Figure 1: The *Main* window after import of the data.

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

3. To view the values in a list, double-click on the experiment **MLVA** in the *Experiment types* panel, select **Settings > General settings...** (⊞), select the *Experiment card tab* and change the representation to *List*. Close the two windows.

4. Click on a green colored dot in the *Experiment presence* panel to open the experiment card for an entry.

The imported repeat numbers are displayed in the experiment card next to the corresponding locus name (see Figure 2).

Figure 2: The experiment card.

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

6. Right-click on the **City** information field in the *Main* window and choose **Field properties** from the floating menu (see Figure 3).
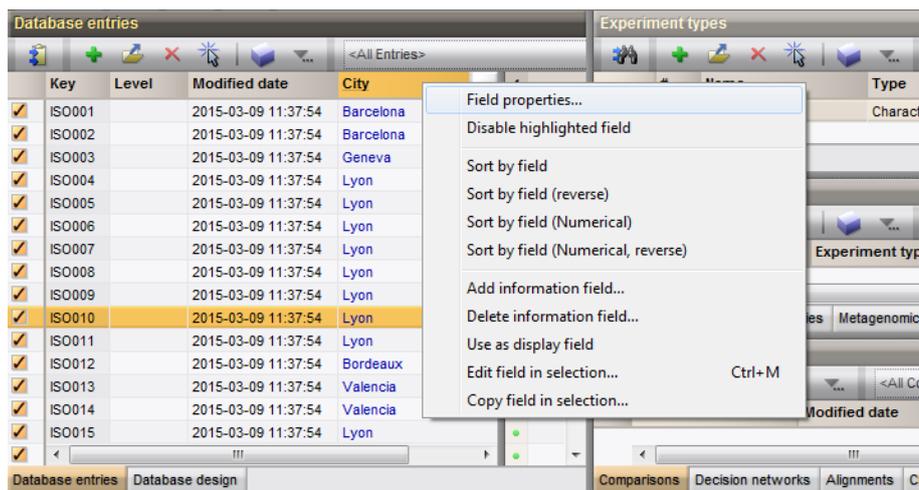

Figure 3: Field properties.

7. Press <**Add all**> to create all existing states for the **City** field. Confirm the action.

8. Check **Use colors** to display a specific color code for each field state (see Figure 4).

9. Press <**OK**> to accept the new settings.

The *Database entries* panel is updated (see Figure 5).

It is also possible to create groups based on the field content in the *Comparison* window with **Groups > Create groups from database field**.

3 Comparison window

1. In the *Database entries* panel of the *Main* window, select all entries using **Edit > Select all** (**Ctrl+A**).
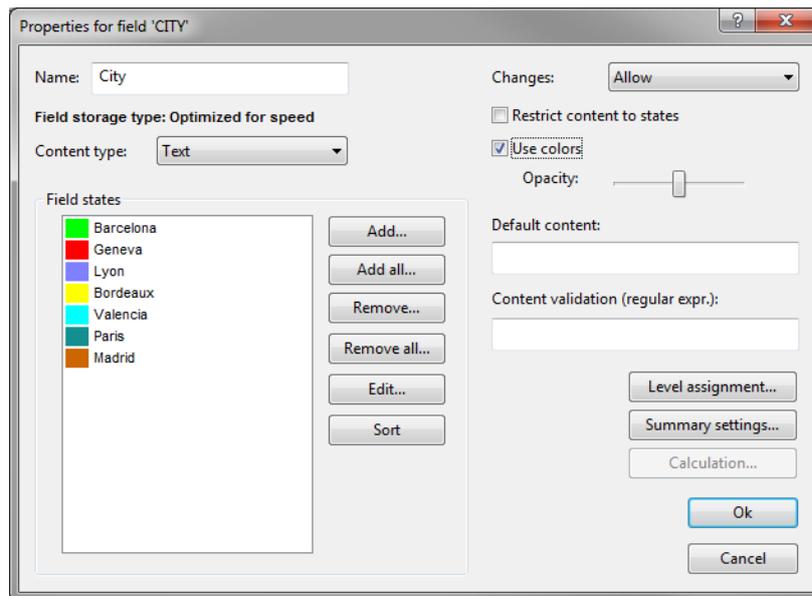
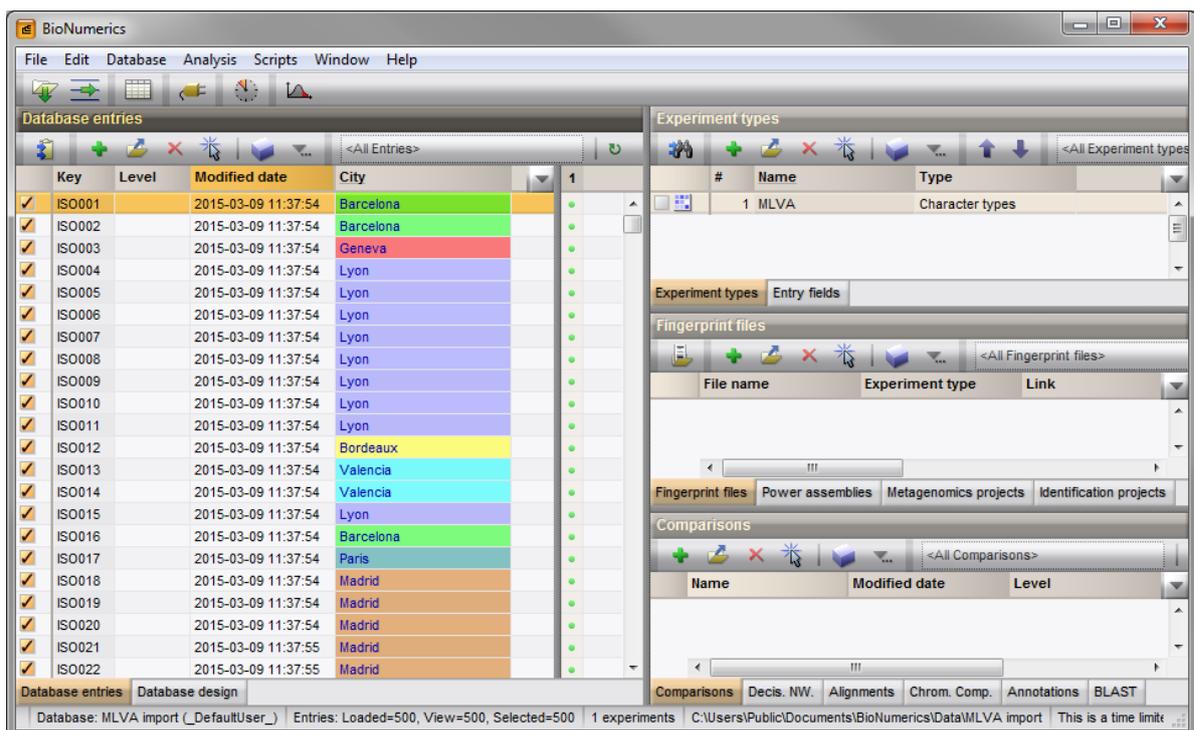**Figure 4:** The *Database field properties* dialog box.

Figure 5: The *Main* window.

2. Highlight the *Comparisons* panel in the *Main* window and select *Edit > Create new object...* ( ) to create a new comparison for the selected entries.

3. Click on the next to the experiment name **MLVA** in the *Experiments* panel and select *Characters > Show values* ( ) to display the repeat numbers in the *Experiment data* panel (see Figure 6).
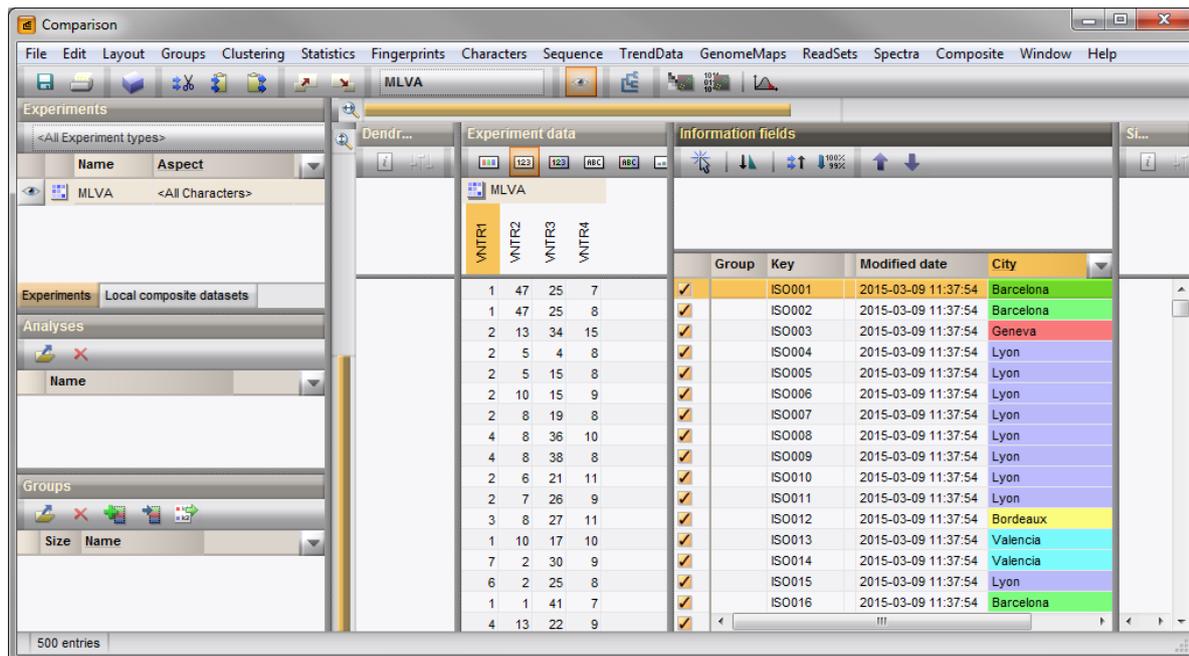
Figure 6: The *Comparison* window.

4 Advanced clustering window

A minimum spanning tree in BioNumerics is calculated in the *Advanced cluster analysis* window. This window can be launched from the *Comparison* window.

1. Select ***Clustering*** > ***Calculate*** > ***Advanced cluster analysis...*** or press the ⟨icon⟩ button and select ***Advanced cluster analysis*** to launch the *Create network* wizard.

Due to the arbirariness of the repeat numbers, the similarity coefficient for clustering MLVA repeat numbers is the categorical coefficient. The categorical coefficient compares the repeat numbers to see if they are the same or different but does not quantify the difference. The predefined template ***MST for categorical data*** uses the categorical coefficient for the calculation of the similarity matrix, and will calculate a standard minimum spanning tree with single and double locus variance priority rules.

2. Specify an analysis name (for example **MLVA1**), make sure ***MLVA*** is selected, select ***MST for categorical data***, and press <***Next***>.

⟨icon⟩ To view and modify the settings of a selected template check the option ***Modify template settings for new analysis***.

The *Advanced cluster analysis* window pops up. The *Network panel* displays the minimum spanning tree, the upper right panel (*Entry list*) displays the entries that are present in the tree. The *Cluster analysis method panel* displays the settings used, in this example the priority rules that result in the displayed network.

The colors of the comparison groups are automatically shown as node colors, but since no groups were defined in our comparison, no nodes are colored in our example. The coloring can very easily be changed to a field state grouping defined in the *Main* window:

3. Press ⟨icon⟩ or choose ***Display*** > ***Display settings*** to open the *Display settings* dialog box.

4. In the *Node colors tab* select the **City** from the list and press <***OK***>.

The node colors are updated according to the city names. A node or branch can be selected by clicking on them. To select several nodes/branches hold the **Shift**-key, or click and hold down the left mouse button and

drag the mouse pointer over the nodes to be selected.

    5. Hold the **Shift**-key and select a few nodes.

The *Selection entry list panel* displays the entries currently selected on the network. The *Entry data panel* displays the character data for the selected entries.

    6. The zoom slider on the left always further zooming in or out on the network. The zoom slider on top adjusts the size of the nodes.

    7. Select **Display** > **Zoom to fit** or press 🔍 to optimize the view of the tree.

    8. Press 🎚 or choose **Display** > **Display settings** to open the *Display settings* dialog box again.

    9. Uncheck the option **Separate entries** in the *Node colors tab*.

    10. Press <**OK**> to apply the new settings.

The *Advanced cluster analysis* window should now look like Figure 7.
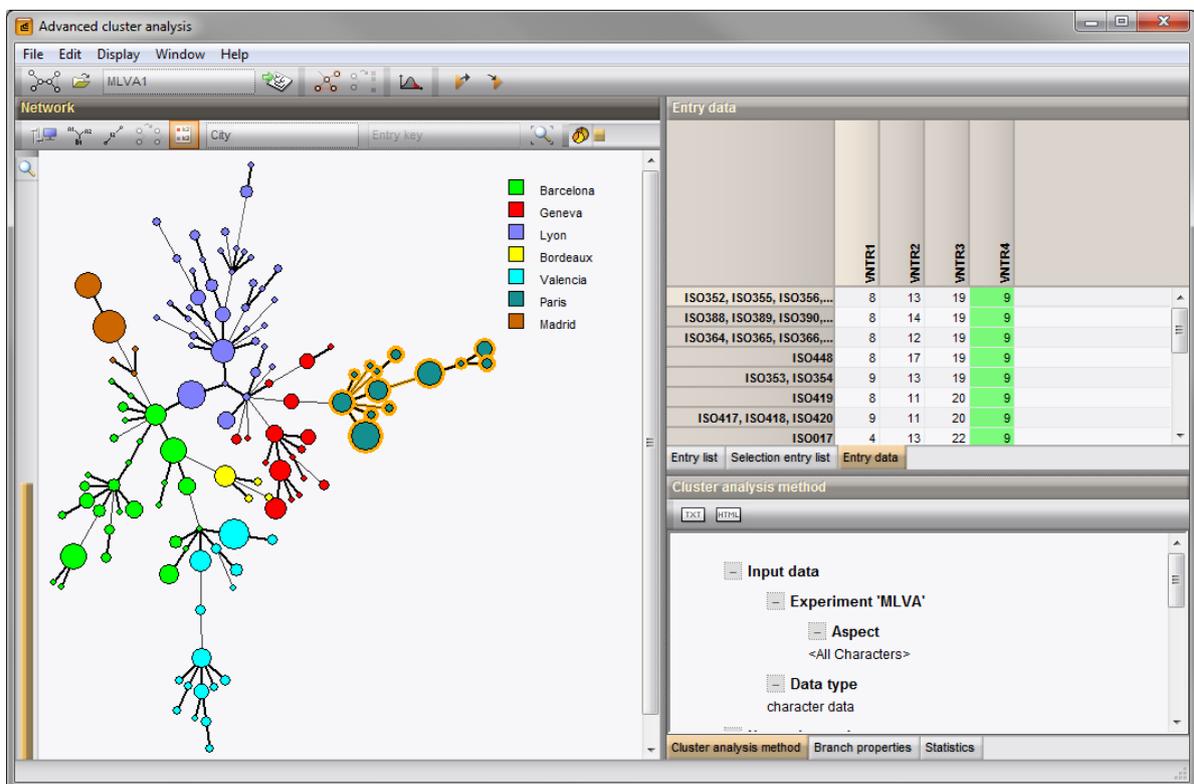

Figure 7: The *Advanced cluster analysis* window.

    11. The image can be exported with **File** > **Export image**.

    12. Close the *Advanced cluster analysis* window and *Comparison* window with **File** > **Exit**.