

BioNumerics Tutorial:

Importing and processing Agilent 2100 BioAnalyzer curve files

1 Aim

The 2100 BioAnalyzer is a capillary electrophoresis instrument from the company Agilent (<https://www.agilent.com/>). The generated electropherograms can be exported from the Agilent 2100 BioAnalyzer software in csv format and imported in BioNumerics after installation of the *Import fingerprint tools plugin*. In this tutorial the installation of the *Import fingerprint tools plugin* and the import and processing of Agilent 2100 BioAnalyzer curve files is covered.

2 Sample data

Two batches of Agilent 2100 BioAnalyzer curve files can be downloaded from the Applied Maths website (<http://www.applied-maths.com/download/sample-data>, click on "Agilent 2100 BioAnalyzer curve files"). These example csv files will be used to illustrate the import and processing steps in this tutorial. Each csv file contains the full densitometric curve for a single profile, with some additional meta data.



When exporting data from the Agilent 2100 BioAnalyzer software make sure to check **Sample data** ("Creates csv files containing the sample data, one file per sample") in the "Electrophoresis Export Options" dialog. It does not matter whether **aligned sample data** or **unaligned sample data** are exported, since the data are normalized in BioNumerics anyway. **Result tables** need not to be checked.

3 Preparing the database

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

If a database is opened for the first time, the *Plugins* dialog box will appear by default (see Figure 1).

If the database has already been opened previously, the *Plugins* dialog box can be called from the *Main* window by selecting **File > Install / remove plugins...** (.

Once a plugin is installed, it is marked with a green V-sign. It can be removed again with the **<Deactivate>** button.

If the selected plugin is documented, pressing **<Show Manual>** will open its manual in the *Help* window.

The *Import fingerprint tools plugin* is provided as an *online plugin*. Online plugins are available from the Applied Maths website, from which they can be downloaded and installed in the database in just a few mouse clicks. Since administrator rights are not required for installation of a plugin in the database, online plugins can be easily updated to take advantage of the latest improvements in program code and search data.

Proceed as follows to install the *Import fingerprint tools plugin*, starting from the *Plugins* dialog box:

2. Select the *Database Functionality tab* in the *Plugins* dialog box and press the **<Add / Update...>** button.

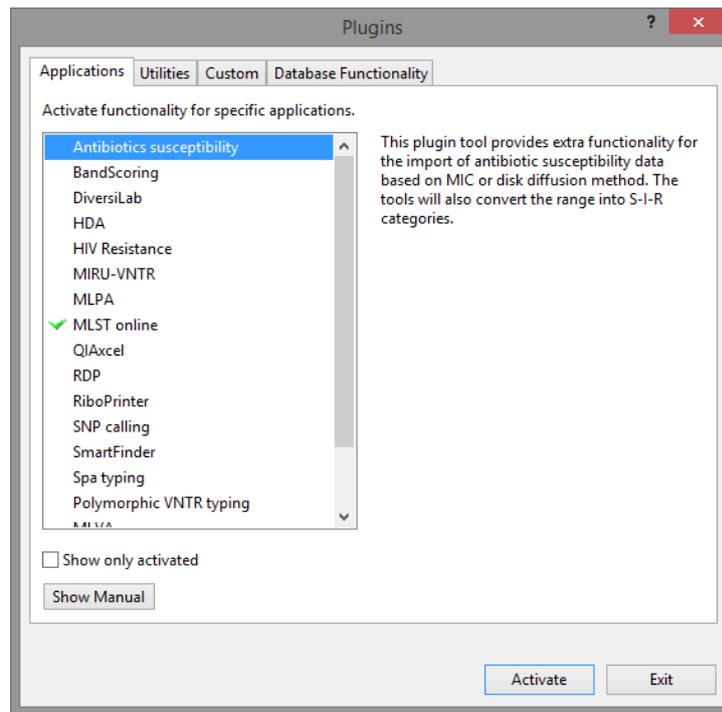


Figure 1: The *Plugins* dialog box.

3. Check the box that corresponds to the *Import fingerprint tools plugin* (see Figure 2) and press <OK>.

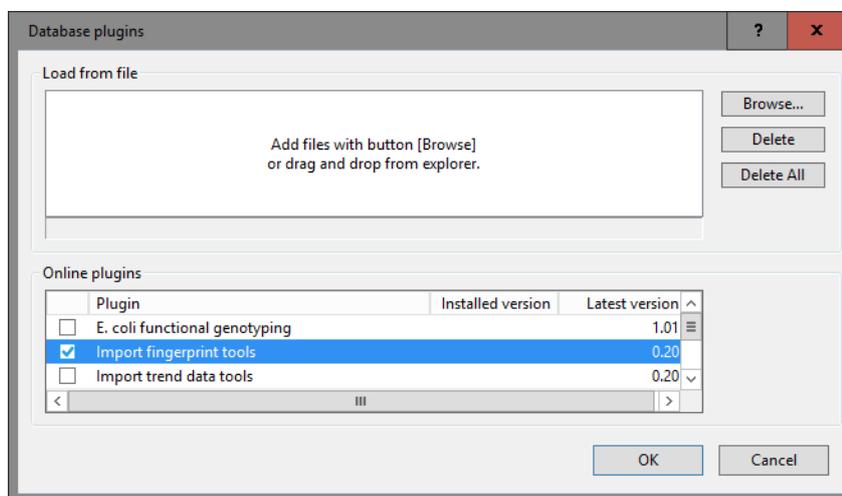


Figure 2: Adding the *Import fingerprint tools plugin*.

The actual download of the plugin file depends on the speed of your internet connection.

4. Confirm the installation of the plugin.

Once the plugin is successfully downloaded and installed, it is marked with a green V-sign in the *Database Functionality tab* in the *Plugins* dialog box (see Figure 3).

5. Close the *Plugins* dialog box.

Initially, a new fingerprint type experiment needs to be created which will hold the imported Agilent 2100

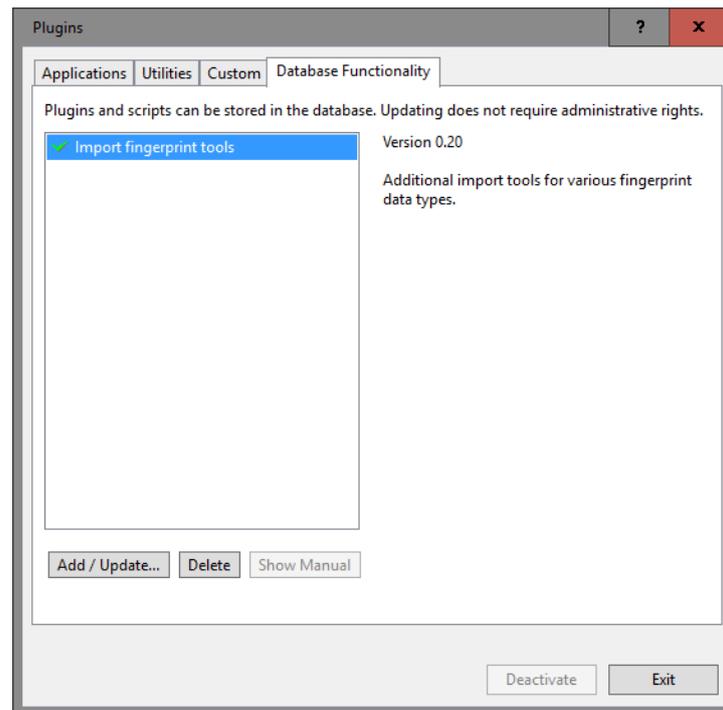


Figure 3: Installed plugin.

BioAnalyzer curve data.

6. In the *Main* window, click on  in the toolbar of the *Experiment types* panel and select **Fingerprint type** from the list (see Figure 4).

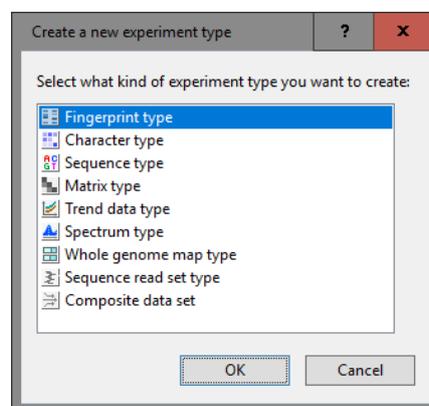


Figure 4: Create new fingerprint type experiment.

7. Press **<OK>**, enter a name, for example “Agilent 2100 Bioanalyzer” and press **<Next>** (see Figure 5).
8. In the page of the *New fingerprint type* wizard, make sure that **Two-dimensional TIFF files** is selected and select the dynamic range of the data: i.e. **8-bit** for the example files (see Figure 6). Press **<Next>**.
9. Press **<Next>** twice without altering the settings and press **<Finish>** to complete the creation of the new fingerprint type.

The *Experiment types* panel now lists the fingerprint type **Agilent 2100 Bioanalyzer**. The two reference positions, corresponding to the upper and lower marker will be defined during the import of the first fingerprint file (see 4.1).

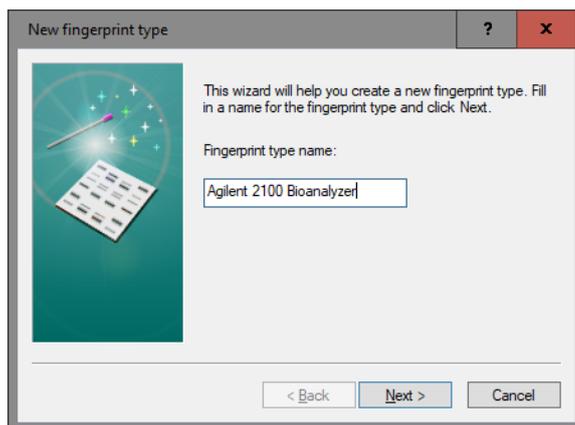


Figure 5: Specify an experiment name.

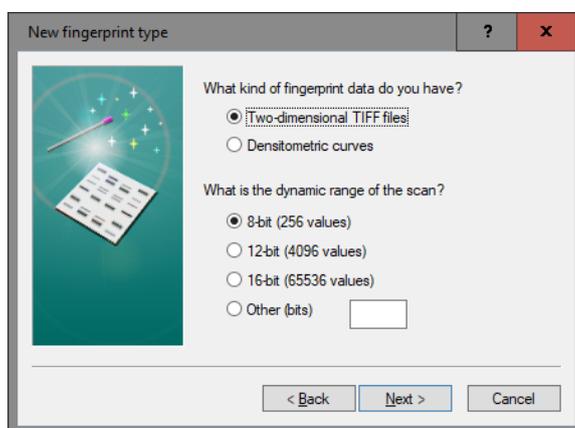


Figure 6: Fingerprint type settings.

4 Importing BioAnalyzer curve files

4.1 First time import

Since it is the first time that we will import Agilent 2100 BioAnalyzer csv files in our BioNumerics database, an *import template* should be created. After the first import, each subsequent import action will be much faster, since the saved import template can simply be re-used (see 4.2).

1. Select **File** > **Import...** (📁, **Ctrl+I**) to call the *Import* dialog box (see Figure 7).
2. Select **Import BioAnalyzer curves** under **Fingerprint type data** and press **<Import>** to start with the import of the data.

The *Import Agilent 2100 Bioanalyzer data* wizard page appears. Since a separate fingerprint file will be created for each import batch, it makes sense to select all samples from a single chip.

3. Browse to the downloaded and unzipped example data folder **Agilent** and select all Agilent 2100 BioAnalyzer csv files in the **Chip01** folder (see Figure 8). Press **<Next>**.



If you are importing your own data sets, make sure not to select any ***_Results.csv** files, as these have a very different format from the curve data files (***_Ladder.csv** and ***_Sample.csv** files).

The *Import rules* dialog box is displayed. Agilent 2100 BioAnalyzer csv files provide meta data as they

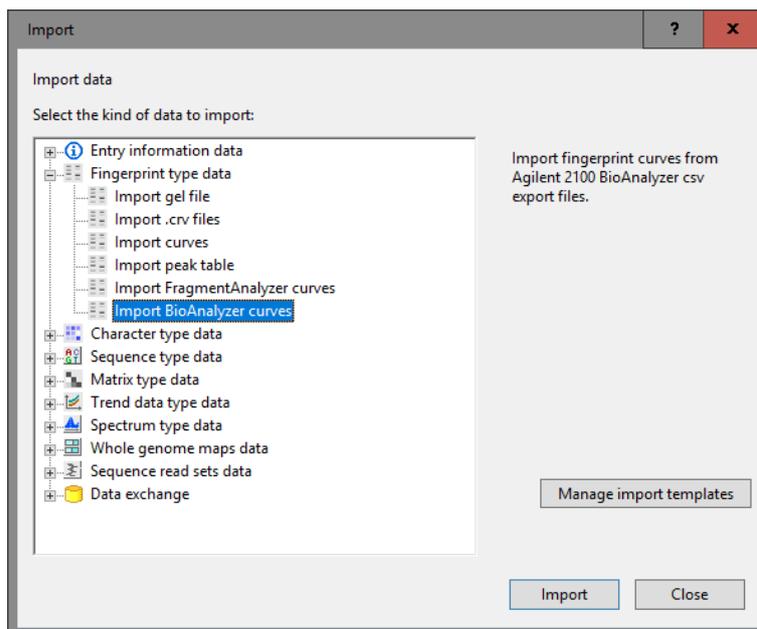


Figure 7: The *Import* dialog box.

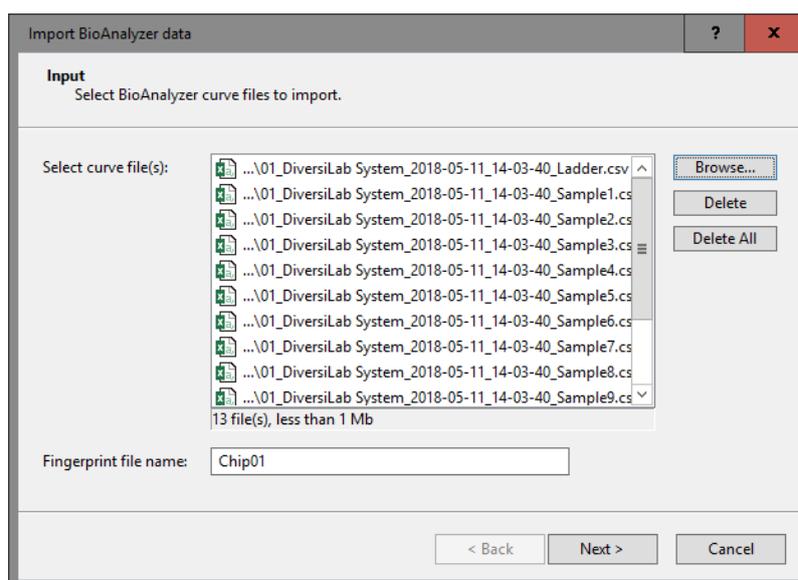


Figure 8: Select curve files from Chip01.

were entered in the Agilent software. Each piece of information can be linked to a destination: either an entry information field or a fingerprint information field in BioNumerics.

In this example, we will store all meta data from the csv files into BioNumerics entry information fields:

4. Click on the first file field in the list ("Data File Name") to highlight it. While holding the **Shift**-key, click on the last file field ("Number of Events") to highlight all file fields.
5. Press the **<Edit destination>** button.
6. In the *Edit data destination* dialog box, click on **Entry info field** (see Figure 9) and press **<OK>**.
7. Press **<OK>** in the *Create new* dialog box and confirm the creation of the entry info fields.

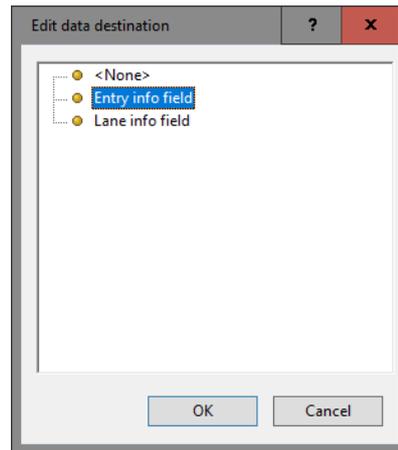


Figure 9: Link to an "Entry info field".

The rows in the grid are updated (see Figure 10).

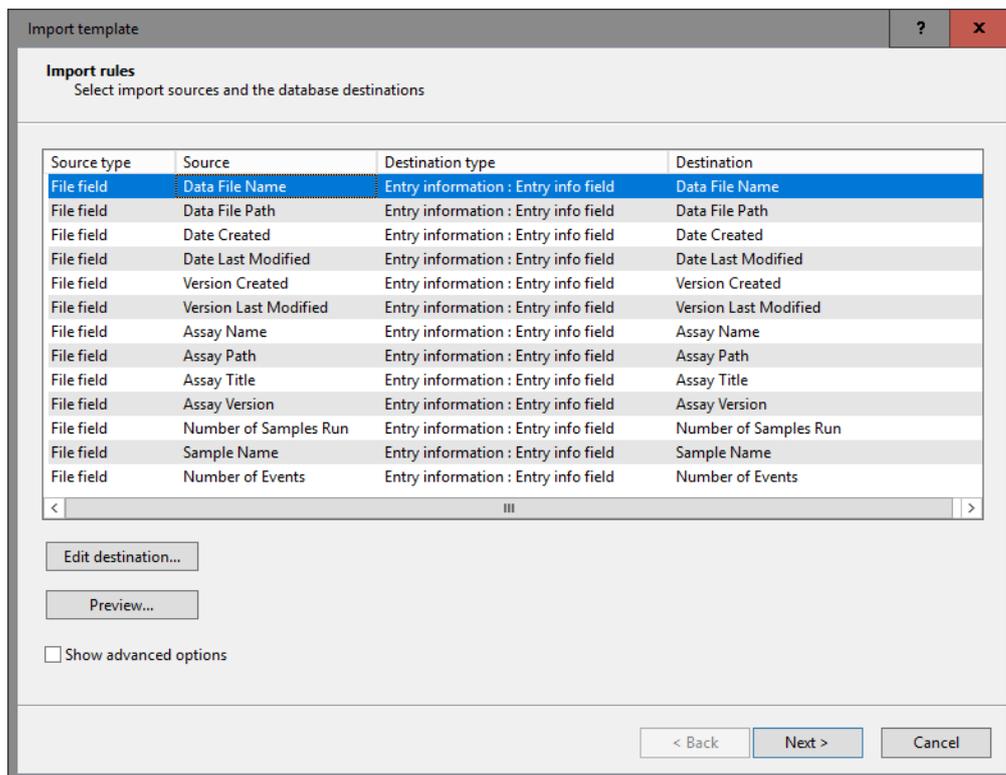


Figure 10: The *Import rules* dialog box.

8. Press *<Next>*.

This action displays the *Import curves* wizard page. In most import scenarios, these settings can be left to their default values.

9. Press *<Next>* to accept the default settings.

The *Import links* dialog box appears, listing the fields from the Agilent 2100 BioAnalyzer csv files that are

imported into entry information fields.

10. Since none of the fields contain a unique identifier for the imported samples, no fields should be linked to let BioNumerics automatically create unique entry keys. Leave all fields unchecked and press <**Finish**>.
11. Enter a **Name** (e.g. “Agilent CSV”) for the newly created import template and press <**OK**> to save the import template in the database.
12. Make sure the **Agilent CSV** import template is highlighted and the **Agilent 2100 Bioanalyzer** experiment is selected in the *Import template Bioanalyzer data* wizard page (see Figure 11). Press <**Next**>.

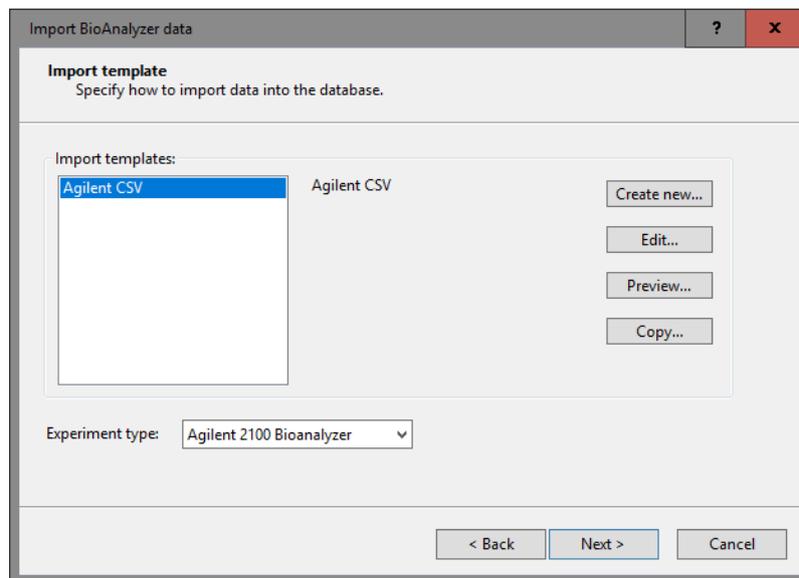


Figure 11: The *Import template Bioanalyzer data* wizard page.

The last page in the wizard indicates that 13 entries will be created with this import action.

13. Press <**Finish**> to import the csv files into the database.

Since this is the first import, the warning message “No reference system found. Please assign reference positions manually.” appears (see Figure 12).

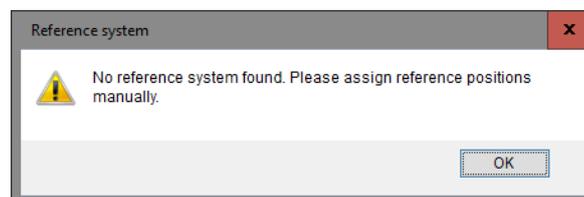


Figure 12: Warning message.

14. Press <**OK**> to close the warning message.
15. Make sure the *Normalization* tab is selected in the *Fingerprint processing* window.
16. Show the gel image in normalized view with *Normalization* > *Show normalized view* (🔍, Shift+N).
17. Click on the band corresponding to the upper marker in one of the lanes, e.g. lane 6, and select *References* > *Add external reference position*.
18. In the dialog that appears, enter “7000” as band name and press <**OK**>.

19. Scroll down in the gel image and click on the band corresponding to the lower marker in lane 6.
20. Select **References** > **Add external reference position**, enter “150” as band name and press <OK>.

This creates the necessary reference system. Since automatic normalization was not possible during the import of this first chip, we need to normalize the gel manually.

21. Make sure the **150** label is the active reference position (the **150** label should be highlighted in orange).
22. With the reference position **150** highlighted, **Ctrl+click** on the lower marker band in each profile to assign the band to this reference position.
23. Click on **7000** label to make it the active reference position. The **7000** label should now be highlighted in orange.
24. **Ctrl+click** on the upper marker in each profile to assign the band to this reference position.
25. Update the normalization with **Normalization** > **Update normalization** (, **Ctrl+U**).

Next, we will assign bands on the fingerprint profiles. This step is not required when only curve-based similarity coefficients will be used to compare profiles.

26. Select **File** > **Next step** () to proceed to the next step: the *Bands* tab is now selected.
27. Select **Bands** > **Auto search bands** () , leave all settings at defaults and press <Search on all lanes>.
28. Check the band assignments.

If too many band assignments are wrong, repeat the automatic band search with different settings. Minor tweaks to the band assignment can be done manually.

29. Save your work with **File** > **Save** (, **Ctrl+S**).

A warning message will appear about the normalized gel resolution. It is safe to update the normalized track resolution to the actual resolution of the gel.

30. Press <Yes> in the confirmation message.
31. Close the *Fingerprint processing* window with **File** > **Exit**.

The *Main* window should now look like Figure 13.

The 13 csv files are imported as 13 individual lanes of the *Chip01* fingerprint file. The *Chip01* fingerprint file is displayed in the *Fingerprint files* panel. Each lane is linked to a database entry: the 13 entries are displayed in the *Database entries* panel, and the linked lanes are displayed with green dots in the *Experiment presence* panel.

32. Open the lane for one of the entries in the database by clicking on a colored dot in the *Experiment presence* panel (see Figure 14).
33. Close the card by clicking in the small triangle-shaped button in the upper left corner.

4.2 Routine import

In our first import round, we have created our import template and defined the reference system for our fingerprint type experiment. Importing Agilent 2100 BioAnalyzer files from other chips is now a very quick and easy process:

34. Select **File** > **Import...** (, **Ctrl+I**) to call the *Import* dialog box again.

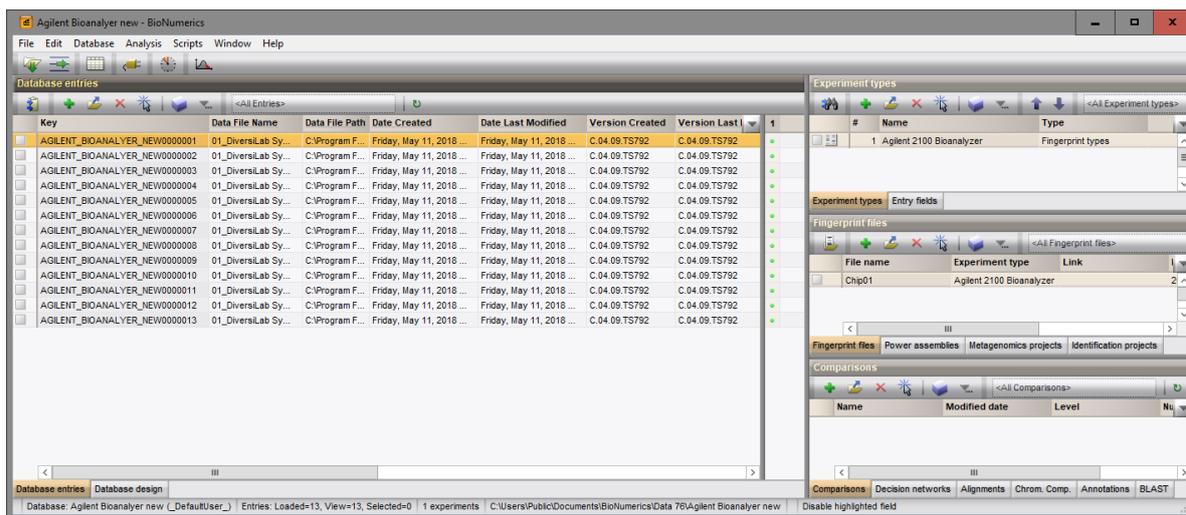


Figure 13: The Main window after import.

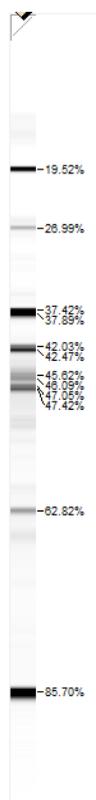


Figure 14: Gel card.

35. Select **Import BioAnalyzer curves** under **Fingerprint type data** (or under **Recently used**) and press **<Import>**.
36. Press the **<Browse>** button and browse for the csv files of the Chip02 folder (see Figure 15).
37. Press **<Next>**.
38. Make sure the **Agilent CSV** import template is highlighted and the **Agilent 2100 Bioanalyzer** experiment is selected in the **Import template Bioanalyzer data** wizard page (see Figure 11). Press **<Next>**.
39. Press **<Finish>** to start the import.

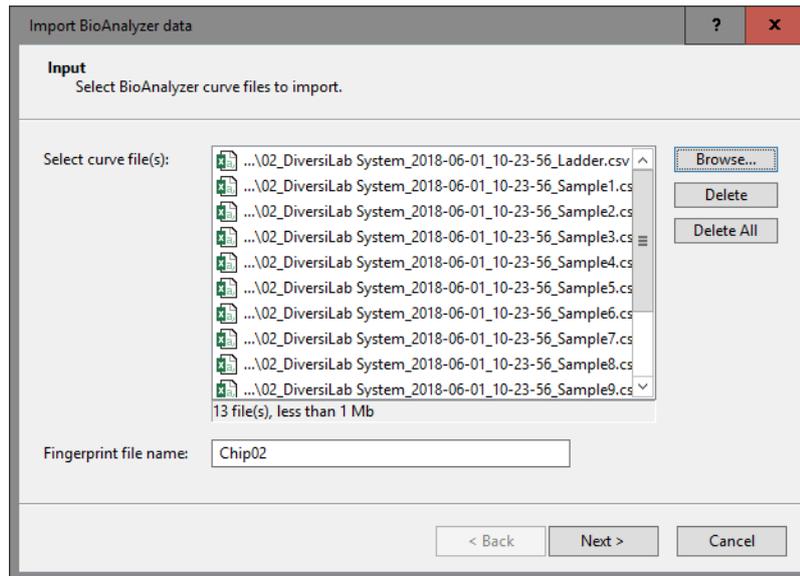


Figure 15: Chip02 curve files.

All curves of **Chip02** are grouped into a single fingerprint file and the fingerprint file is automatically normalized based on the upper and lower markers in each profile and a band search is performed.

In the curves files of **Chip02**, the plugin is unable to find the upper and lower markers in lane 8 and 10 (see Figure 16).

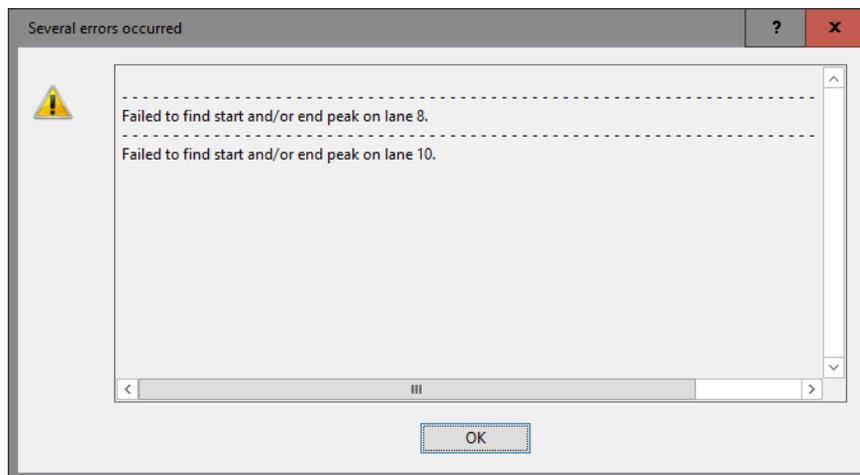


Figure 16: Error messages for Chip02.

40. Press **<OK>** to close the error message.

We will first normalize lane 8 and 10 to avoid errors during comparison.

41. In the *Fingerprint files* panel of the *Main* window, double-click on **Chip02** to call the *Fingerprint* window.

42. Select **File > Edit fingerprint data...** () to open the fingerprint file in the *Fingerprint processing* window.

43. In the bottom of the window, click on the *Normalization* tab or select **File > Next step** () twice to show the normalization step.

In lane 8 and lane 10, the upper marker was not identified correctly (see Figure 17).

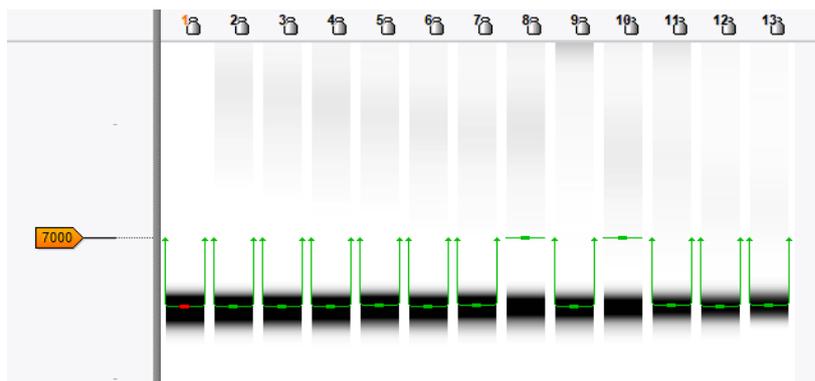


Figure 17: Example of a normalization issue: the upper marker was not found on lanes 8 and 10.

44. Make sure the **7000** label is highlighted in orange (= active reference position).
45. In lane 8, **Ctrl+Click** on the correct position of the upper marker to link that peak to the reference position (see Figure 18).

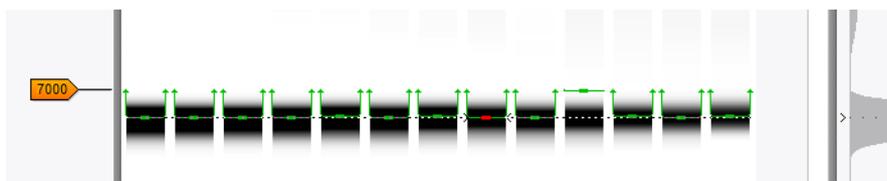


Figure 18: Correction lane 8.

46. In lane 10, **Ctrl+Click** on the correct position of the upper marker to link that peak to the reference position (see Figure 19).

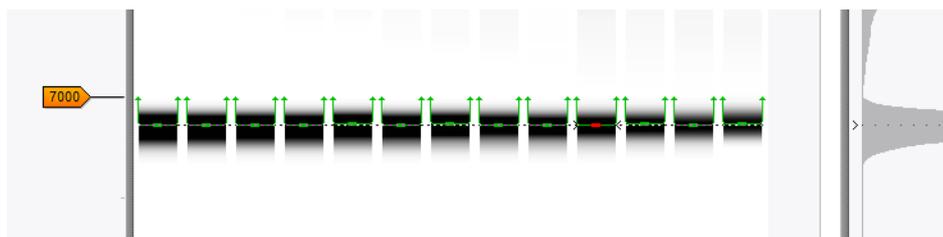


Figure 19: Correction lane 10.

Next, we will check for any issues with the lower marker:

47. Scroll down the gel image and click on reference position **150**, corresponding to the lower marker. The **150** label should now be highlighted in orange (= active reference position).

In lane 8 and lane 10, the lower marker was not identified correctly.

48. In lane 8, **Ctrl+Click** on the correct position of the low marker to link that peak to the reference position (see Figure 20).
49. In lane 10, **Ctrl+Click** on the correct position of the low marker to link that peak to the reference position (see Figure 20).
50. Save the edits with **File > Save** (📁), **Ctrl+S** and close the *Fingerprint processing* window.

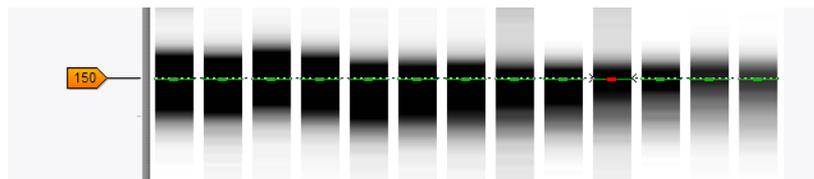


Figure 20: Correction of lower markers.

5 Analyzing BioAnalyzer curve files

1. In the *Database entries* panel of the *Main* window, select all entries with **Ctrl+A**.
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 (left arrow) next to the experiment name **Agilent 2100 Bioanalyzer** in the *Experiments* panel to display the patterns in the *Experiment data* panel (see Figure 21).

In order to correct for the normalization problems associated with having only two markers, the *Import fingerprint tools plugin* can perform an on-the-fly alignment of the densitometric curves present within a comparison. More information about this correction can be found in the *Import fingerprint tools plugin* manual. In this tutorial no details about the correction tool will be discussed.

The correction is calculated on the fly on the densitometric curves loaded in the *Comparison* window, so we first need to show the densitometric curves in the *Experiment data* panel before performing the correction:

4. Select *Fingerprints > Show densitometric curves* (mountain icon) to display the densitometric curves (see Figure 21).

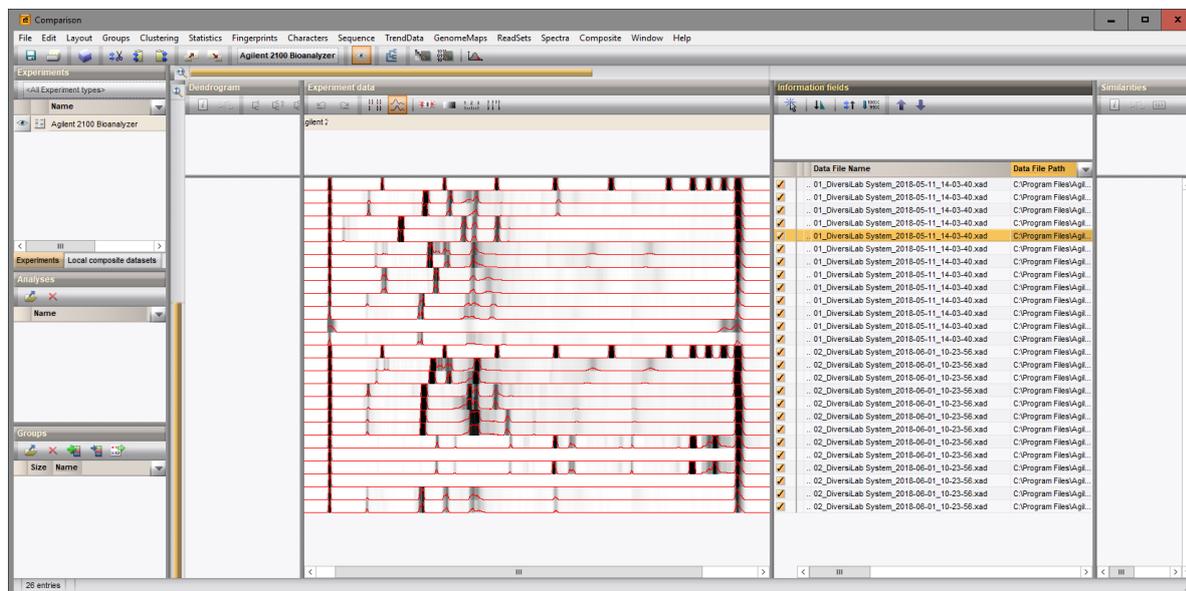


Figure 21: Show densitometric curves.

Next, a similarity matrix needs to be calculated for the fingerprints. In order to obtain the best results with the alignment algorithm, it is recommended to apply a large optimization value, e.g. 3%.

5. Make sure **Agilent 2100 Bioanalyzer** is selected in the *Experiments* panel and select *Clustering > Calculate > Cluster analysis (similarity matrix)*...

6. Make sure a curve-based coefficient is selected, e.g. *Pearson correlation*, specify an optimization of 3%, press <OK> and <Finish>.

The similarity matrix (and dendrogram) is calculated for the fingerprints present in the *Comparison* window.

7. Select *Fingerprints > Auto-correct curves* to perform the auto-correction.
8. As a first example, check *Keep edges of fingerprints fixed*. Use 5.0% as *Maximum shift* and enter "5" as *Number of iterations*. Press <OK>.

The gel strips display the original traces, whereas the curves are corrected by the auto-correct tool. This viewing mode can be useful to monitor the settings and the results of the alignment algorithm.

9. Select *Fingerprints > Auto-correct curves* to perform the auto-correction again with different settings.
10. Switch off *Keep edges of fingerprints fixed* and specify 1.0% as *Maximum stretch/compression* and press <OK>.



The auto-correct function should be used carefully and the result should always be compared with the original traces to verify that the algorithm did not perform excessive shift, stretching or compression on (groups of) patterns. It is recommended to use the function at first with very low values for shift and stretch/compression, and gradually increase these until the results are satisfactory without observing "over-corrected" traces.

11. A clustering can now be calculated on the aligned curves (*Clustering > Calculate > Cluster analysis (similarity matrix)...*).

If the auto-correct algorithm performed additional correction to the traces that could not be obtained by the *Optimization* function, the clusters will become more homogeneous.



It is important to realize that the alignment is based on the chosen set of entries within the current comparison. By leaving out or adding entries to the comparison, other alignments might be obtained. The alignment is therefore not saved in the database.