BIONUMERICS Tutorial:

Importing data in a database with levels

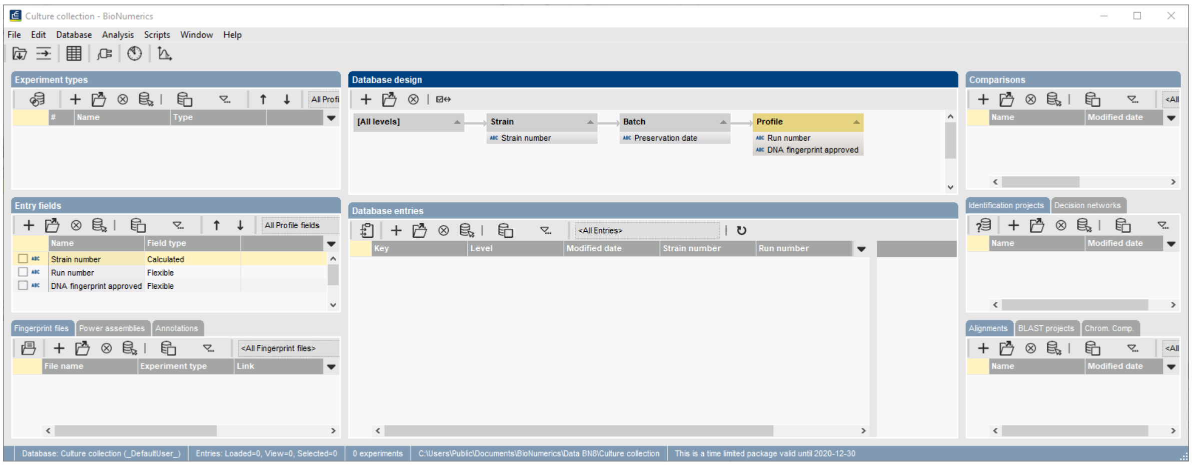
1 Aim

In this tutorial you will learn how to import data in a BIONUMERICS database with levels and how to replicate and summarize level-specific information and experimental data to other levels.

2 Preparing a sample database

1. Create a new database and define the levels as described in the tutorial: ”Setting up a BIONUMERICS database with levels”.

Three hierarchical levels should now be present in your example database (see Figure 1):



Name	Field type
Strain number	Calculated
Run number	Flexible
DNA fingerprint approved	Flexible

Figure 1: The *Main* window with levels defined.

1. **Strain:** This level will contain all strains. The ***Strain number*** is defined and is replicated to the **Batch** and **Profile** levels.
2. **Batch:** This level will information about the batches. A batch is always prepared from a single strain. From any given strain in the database, multiple batches can be prepared. The ***Preservation date*** field has been defined at this level.
3. **Profile:** The ***Run number*** and ***DNA fingerprint approved*** fields has been defined at this level and two approved states have been specified: ***Yes*** and ***No***. The ***Run number*** is replicated to the **Batch** level.

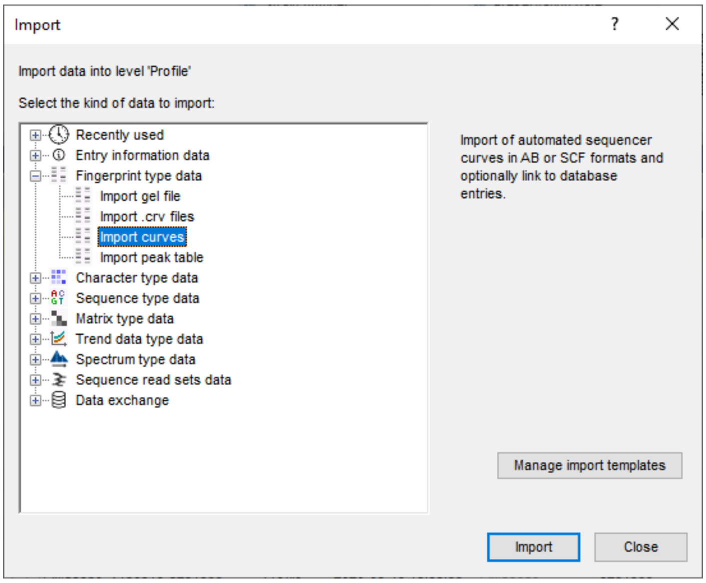
As an exercise, we will import a set of capillary electrophoresis profiles (= electropherograms) produced by a Beckman automated sequencer (.SCF files) in this database.

2. The curves can be downloaded from the Applied Maths website: go to https://www.applied-maths.com/download/sample-data and click on ”Data set leveled database”. When the download is complete, unzip the file.

3 Importing data

When importing data (descriptive information and/or experimental data) in a database with levels, it is important to highlight the corresponding level first. Typically, information will most often be imported at the deepest child level (i.e. **Profile** in our example database).

1. Make sure the **Profile** level is selected in the *Database design* panel and select ***File*** > ***Import...*** (📂, **Ctrl+I**).
2. Select ***Fingerprint type data*** > ***Import curves*** and press <***Import***> (see Figure 2).



Import data into level 'Profile'

Select the kind of data to import:

Import of automated sequencer curves in AB or SCF formats and optionally link to database entries.

Figure 2: Import tree.

3. Browse to the downloaded and unzipped example data folder containing the DNA fingerprints. Select all 56 .SCF files in this folder and press <***Open***>.

The files are displayed in the *Input* wizard page and the default suggested ***Fingerprint file name*** is the folder name.

4. Press <***Next***>.

The *Import rules* dialog box lists the information present in the selected files as **Source**, their linked **Source type** and the **Destination** component they are associated with (currently all set to <None>).

5. Select ***Curve dye*** from the list, select <***Edit destination***> and select ***Fingerprint dye*** as corresponding field. Press <***OK***>.

We will now link the file names to the ***Key*** field of the **Profile** level:

6. Double-click on the ***File*** row available in the grid or select the row and press <***Edit Destination***>.

7. Select ***Profile key*** in the *Edit data destination* dialog box (see Figure 3) and press <***OK***>.

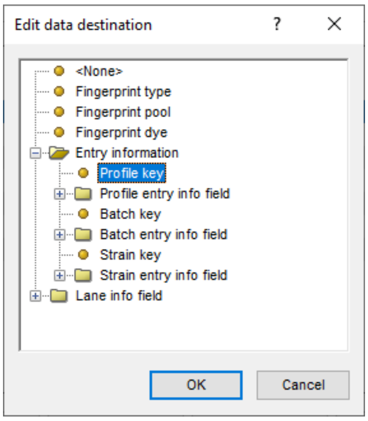


Figure 3: Link to ***Profile key***.

We will now link part of the file names to the ***Key*** fields of the **Batch** and **Strain** levels:

8. Visualize the advanced options for the *Import template* dialog box by clicking on the check box next to ***Show advanced options***.

9. Press <***Add rule***> to open the *Add data conversion rule* wizard.

10. In the first page of the *Add data conversion rule* wizard select ***File*** > ***Name*** and press <***Next***> (see Figure 4).

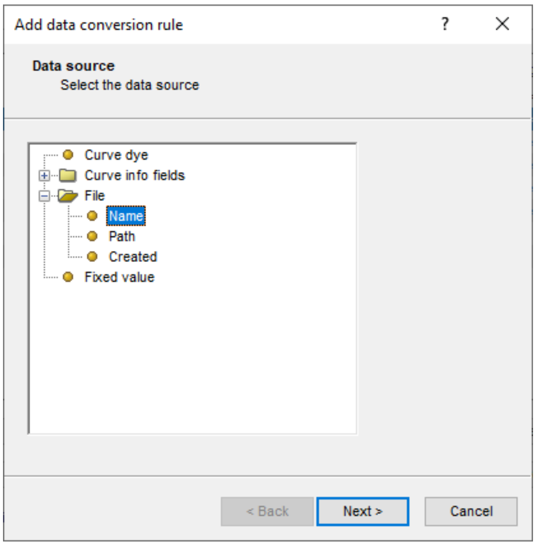


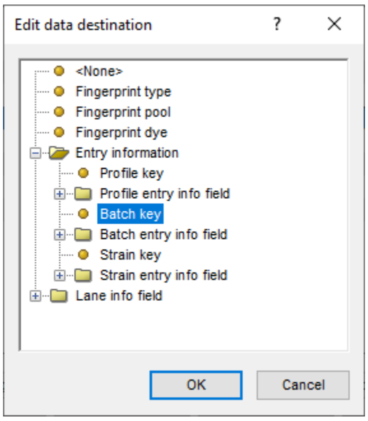
Figure 4: Link ***File name***.

11. In the second page of the *Add data conversion rule* wizard, select ***Batch key*** and press <***Next***> (see Figure 5).

12. In the *Data parsing* dialog box, fill in following data parsing string: “[DATA]-*”. Press the <***Preview***> button (see Figure 6).

This parsing string will only retain the text before the ”-” and will store the text in the ***Batch Key*** field.

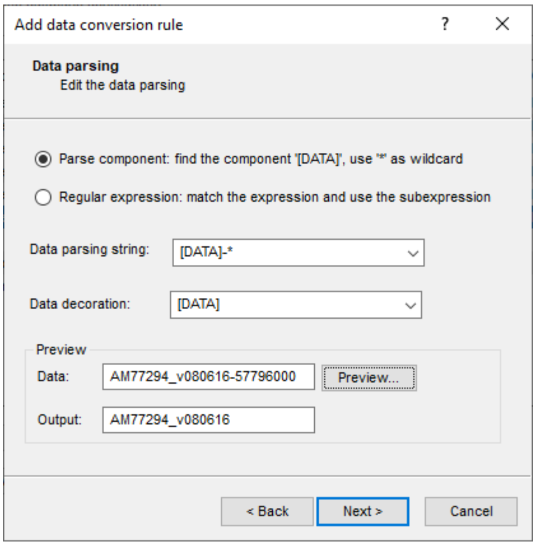
13. When the information is parsed correctly press <***Next***> and <***Finish***>.



Edit data destination

- <None>
- Fingerprint type
- Fingerprint pool
- Fingerprint dye
- Entry information
  - Profile key
  - Profile entry info field
  - Batch key
  - Batch entry info field
  - Strain key
  - Strain entry info field
- Lane info field

Figure 5: Link to ***Batch key***.



Add data conversion rule

Data parsing
Edit the data parsing

● Parse component: find the component '[DATA]', use '*' as wildcard

○ Regular expression: match the expression and use the subexpression

Data parsing string: [DATA]-*

Data decoration: [DATA]

Preview

Data: AM77294_v080616-57796000

Output: AM77294_v080616

Figure 6: Parsing string.

14. Press <***Add rule***> again to open the *Add data conversion rule* wizard.

15. In the first page of the *Add data conversion rule* wizard select ***File*** > ***Name*** and press <***Next***>.

16. In the second page of the *Add data conversion rule* wizard, select ***Strain key*** and press <***Next***> (see Figure 7).

17. In the *Data parsing* dialog box, fill in following data parsing string: "[DATA]_*". Press the <***Preview***> button (see Figure 8).

This parsing string will only retain the text before the "_" and will store the text in the ***Strain key*** field.

18. When the information is parsed correctly press <***Next***> and <***Finish***>.

The grid is updated and should now look like Figure 9.

19. In the *Import template* dialog box, press <***Preview***> and verify the preview of the import.

Extra information, such as the profile ***Run number*** and the batch ***Preservation date*** can also be

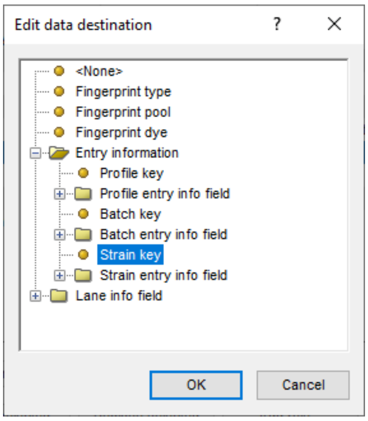
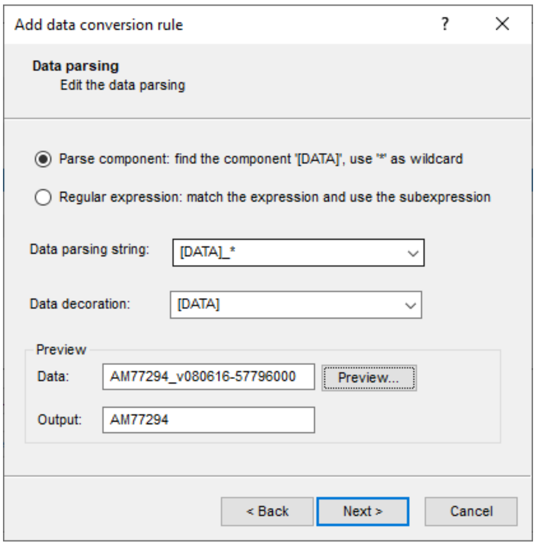


Figure 7: Link to ***Strain key***.



● Parse component: find the component '[DATA]', use '*' as wildcard

○ Regular expression: match the expression and use the subexpression

Data parsing string: [DATA]_*

Data decoration: [DATA]

Preview

Data: AM77294_v080616-57796000

Output: AM77294

Figure 8: Parsing string.

parsed from the file names.

20. Press <***Add rule***> again to open the *Add data conversion rule* wizard. Select ***File*** > ***Name*** and press <***Next***>. Select ***Run number*** under ***Profile entry info field*** and press <***Next***>. Fill in following data parsing string: "*-[DATA]" (see Figure 10). Press the <***Preview***> button. Press <***Next***> and <***Finish***>.

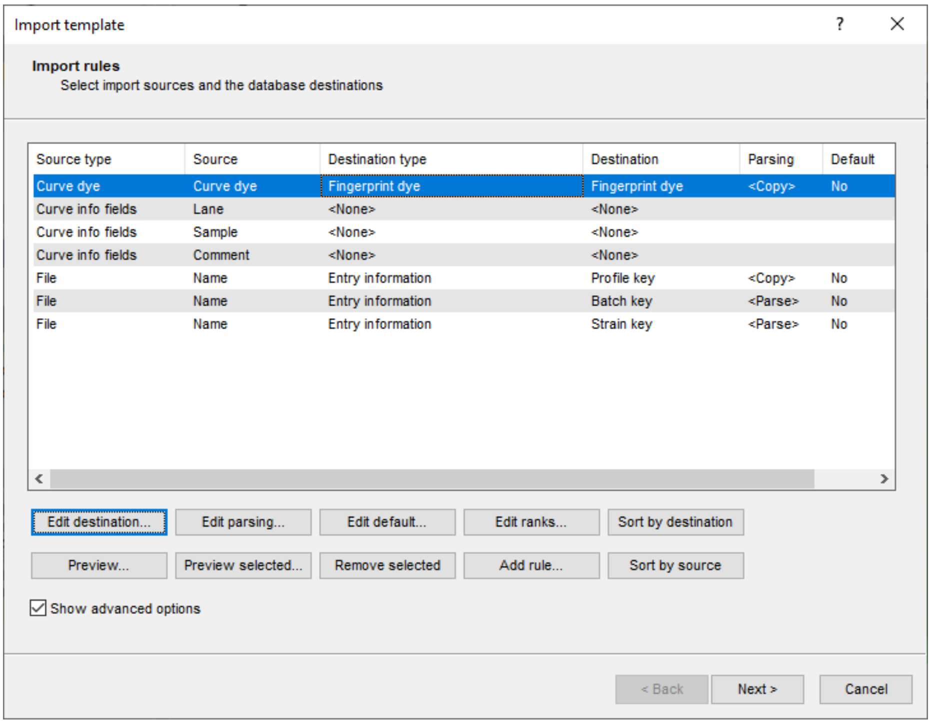
21. Press <***Add rule***> again to open the *Add data conversion rule* wizard. Select ***File*** > ***Name*** and press <***Next***>. Select ***Preservation date*** under ***Batch entry info field*** and press <***Next***>. Fill in following data parsing string: "*v[DATA]-" (see Figure 11). Press the <***Preview***> button. Press <***Next***> and <***Finish***>.

The grid is updated and should now look like Figure 12.

22. In the *Import template* dialog box, press <***Preview***> and verify the preview of the import (see Figure 13).

23. Press <***Next***> to go to the next step.

In the example data set, channel 1 contains the size standard and channel 4 represents the actual



Import template

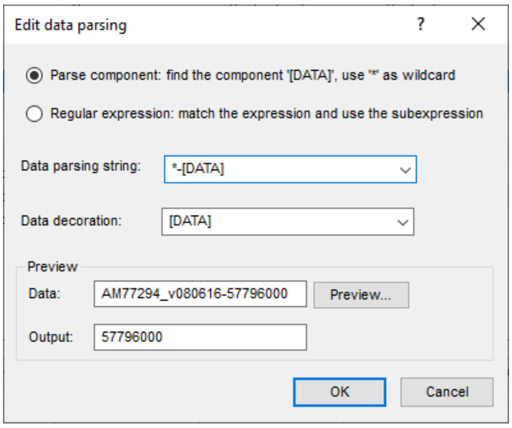
Import rules

Select import sources and the database destinations

Source type	Source	Destination type	Destination	Parsing	Default
Curve dye	Curve dye	Fingerprint dye	Fingerprint dye	<Copy>	No
Curve info fields	Lane	<None>	<None>		
Curve info fields	Sample	<None>	<None>		
Curve info fields	Comment	<None>	<None>		
File	Name	Entry information	Profile key	<Copy>	No
File	Name	Entry information	Batch key	<Parse>	No
File	Name	Entry information	Strain key	<Parse>	No

☑ Show advanced options

Figure 9: Import and parsing rules.



Edit data parsing

● Parse component: find the component '[DATA]', use '*' as wildcard

○ Regular expression: match the expression and use the subexpression

Data parsing string: *-[DATA]

Data decoration: [DATA]

Preview

Data: AM77294_v080616-57796000

Output: 57796000

Figure 10: Parsing string for the ***Run number***.

samples.

24. Make sure ***1*** is selected as reference dye and uncheck the dyes ***2*** and ***3*** to prevent the import of these channels (see Figure 14).

25. Press <***Next***> and <***Finish***>.

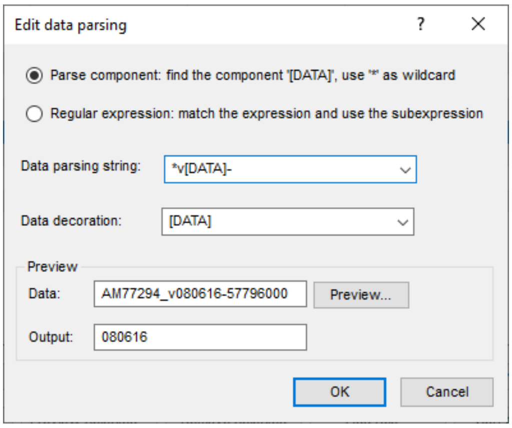
26. Specify a template name, e.g. **Import AB curve files** and press <***OK***>.

27. Make sure the newly created template is selected and press <***Next***> (see Figure 15).

28. Specify an experiment name e.g. **AFLP** (see Figure 16), press <***OK***> and confirm the action.

29. Specify an OD range of ***65536*** gray values (= 16-bit) and press <***OK***>.

A fingerprint type needs to be present in the database for each dye. The names of these fingerprint



Edit data parsing

● Parse component: find the component '[DATA]', use '*' as wildcard

○ Regular expression: match the expression and use the subexpression

Data parsing string: *v[DATA]-

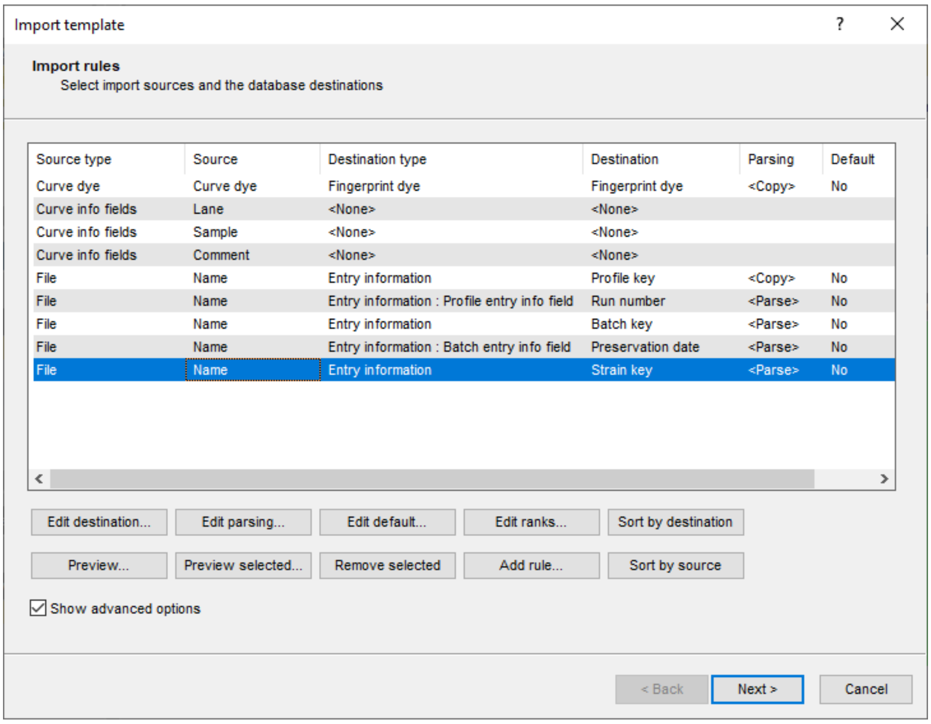
Data decoration: [DATA]

Preview

Data: AM77294_v080616-57796000

Output: 080616

Figure 11: Parsing string for the ***Preservation date***.



Import template

Import rules

Select import sources and the database destinations

Source type	Source	Destination type	Destination	Parsing	Default
Curve dye	Curve dye	Fingerprint dye	Fingerprint dye	<Copy>	No
Curve info fields	Lane	<None>	<None>		
Curve info fields	Sample	<None>	<None>		
Curve info fields	Comment	<None>	<None>		
File	Name	Entry information	Profile key	<Copy>	No
File	Name	Entry information : Profile entry info field	Run number	<Parse>	No
File	Name	Entry information	Batch key	<Parse>	No
File	Name	Entry information : Batch entry info field	Preservation date	<Parse>	No
File	Name	Entry information	Strain key	<Parse>	No

☑ Show advanced options

Figure 12: Import and parsing rules.

types are composed of the base fingerprint type name, followed by the dye name. A new dialog box pops up, listing all missing fingerprint types.

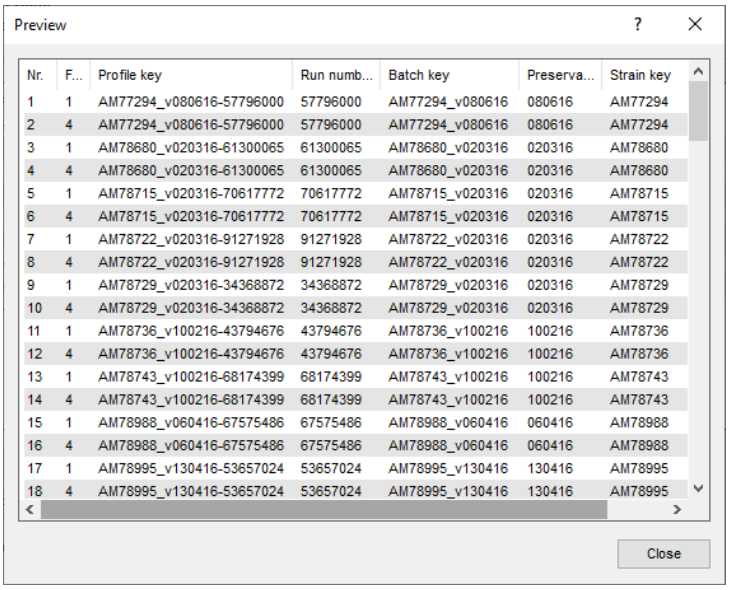
30. Confirm the creation of the two missing experiments in the database.

31. Press <***Next***> to confirm the creation of the new entries in each level (see Figure 18).

32. Make sure ***Open curve preprocessing window*** is checked in the last step and press <***Finish***>.

For each dye checked in the *Dyes panel* of the *Import data* dialog box, a new fingerprint file is created, composed of the file name specified and the name of the dye (e.g. DNA fingerprints_1). These files are displayed in the *Fingerprint files* panel. The reference file is shown in the **Link** column. Double-clicking on a fingerprint file opens the *Fingerprint* window.

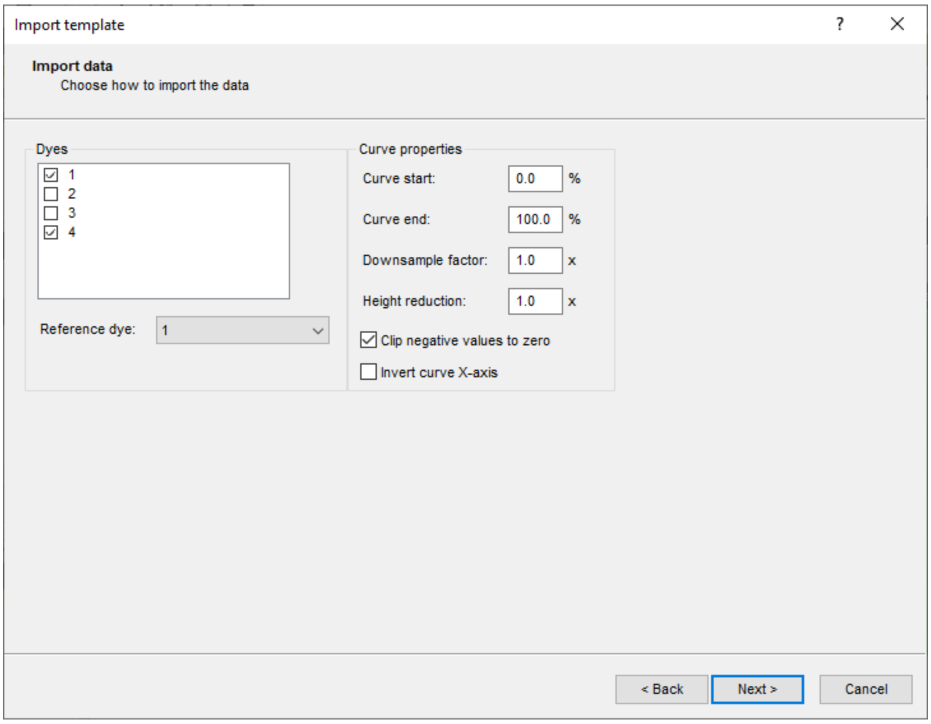
In the ***Profile*** level, 56 entries are added to the *Database entries* panel (see Figure 19). The ***Run***



Preview

Nr.	F...	Profile key	Run numb...	Batch key	Preserva...	Strain key
1	1	AM77294_v080616-57796000	57796000	AM77294_v080616	080616	AM77294
2	4	AM77294_v080616-57796000	57796000	AM77294_v080616	080616	AM77294
3	1	AM78680_v020316-61300065	61300065	AM78680_v020316	020316	AM78680
4	4	AM78680_v020316-61300065	61300065	AM78680_v020316	020316	AM78680
5	1	AM78715_v020316-70617772	70617772	AM78715_v020316	020316	AM78715
6	4	AM78715_v020316-70617772	70617772	AM78715_v020316	020316	AM78715
7	1	AM78722_v020316-91271928	91271928	AM78722_v020316	020316	AM78722
8	4	AM78722_v020316-91271928	91271928	AM78722_v020316	020316	AM78722
9	1	AM78729_v020316-34368872	34368872	AM78729_v020316	020316	AM78729
10	4	AM78729_v020316-34368872	34368872	AM78729_v020316	020316	AM78729
11	1	AM78736_v100216-43794676	43794676	AM78736_v100216	100216	AM78736
12	4	AM78736_v100216-43794676	43794676	AM78736_v100216	100216	AM78736
13	1	AM78743_v100216-68174399	68174399	AM78743_v100216	100216	AM78743
14	4	AM78743_v100216-68174399	68174399	AM78743_v100216	100216	AM78743
15	1	AM78988_v060416-67575486	67575486	AM78988_v060416	060416	AM78988
16	4	AM78988_v060416-67575486	67575486	AM78988_v060416	060416	AM78988
17	1	AM78995_v130416-53657024	53657024	AM78995_v130416	130416	AM78995
18	4	AM78995_v130416-53657024	53657024	AM78995_v130416	130416	AM78995

Figure 13: Preview.



Import template

Import data
Choose how to import the data

Dyes
☑ 1
☐ 2
☐ 3
☑ 4

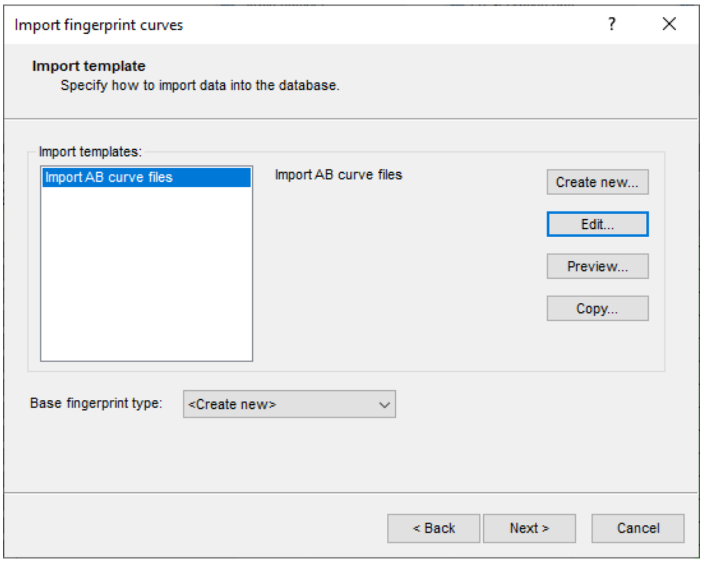
Reference dye: 1

Curve properties
Curve start: 0.0 %
Curve end: 100.0 %
Downsample factor: 1.0 x
Height reduction: 1.0 x
☑ Clip negative values to zero
☐ Invert curve X-axis

Figure 14: Import data.

number information is parsed from the file names and the ***DNA fingerprint approved*** state is default set to ***No***. The ***Strain number*** is replicated from the ***Strain*** level. The imported fingerprint lanes are linked to new entries in the database and to the corresponding fingerprint "dye" type (***AFLP1*** and ***AFLP4***). The fingerprint type experiments are displayed in the *Experiment types* panel.

In the ***Batch*** level, 50 entries are present in the *Database entries* panel (see Figure 20). The ***Preservation date*** information is parsed from the file names and the ***Run numbers*** are summarized from the ***Profile*** level. The ***Strain number*** is replicated from the ***Strain*** level.


Import fingerprint curves

Import template

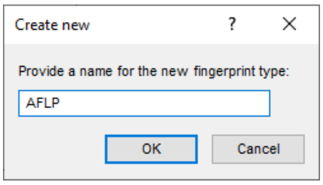
Specify how to import data into the database.

Import templates: Import AB curve files

Import AB curve files

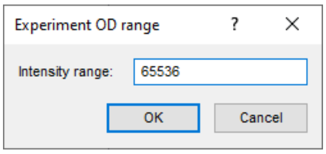
Base fingerprint type: <Create new>

Figure 15: Import template and experiment.


Create new

Provide a name for the new fingerprint type: AFLP

Figure 16: Base fingerprint type experiment.


Experiment OD range

Intensity range: 65536

Figure 17: OD range.

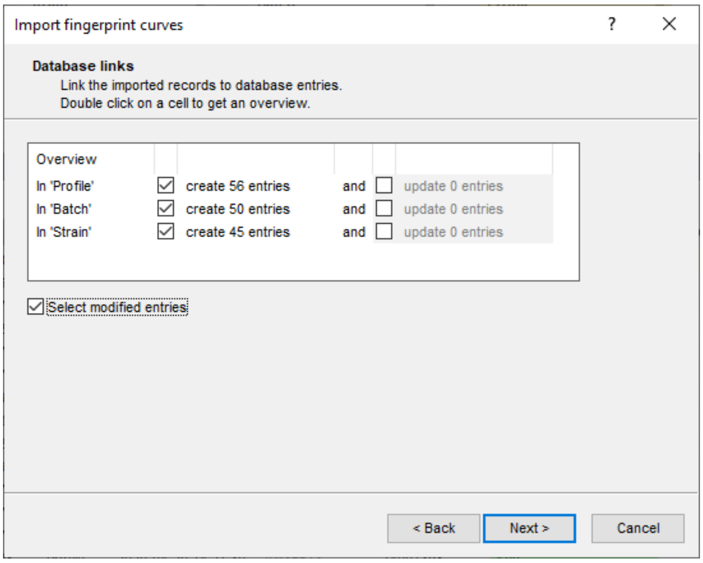
In the ***Strain*** level, 45 entries are added to the *Database entries* panel.

4 Processing data

When the option ***Open curve preprocessing window*** was checked in the last step of the import routine, the *Fingerprint curve processing* window opens when pressing the <***Finish***> button. The two channels from the run are automatically loaded and displayed in the *Fingerprint curve processing* window.

The *Fingerprint curve processing* window can also be called from the *Main* window by highlighting one of the channels in the *Fingerprint files* panel and select ***Open fingerprint data...*** (🖹). Alternatively, you can first open the *Fingerprint* window with ***Edit*** > ***Open highlighted object...*** (🗁, **Enter**) and then select ***File*** > ***Edit fingerprint data...*** (▦).

1. Click on the 👁 icon left of the fourth channel in the *Channels* panel.



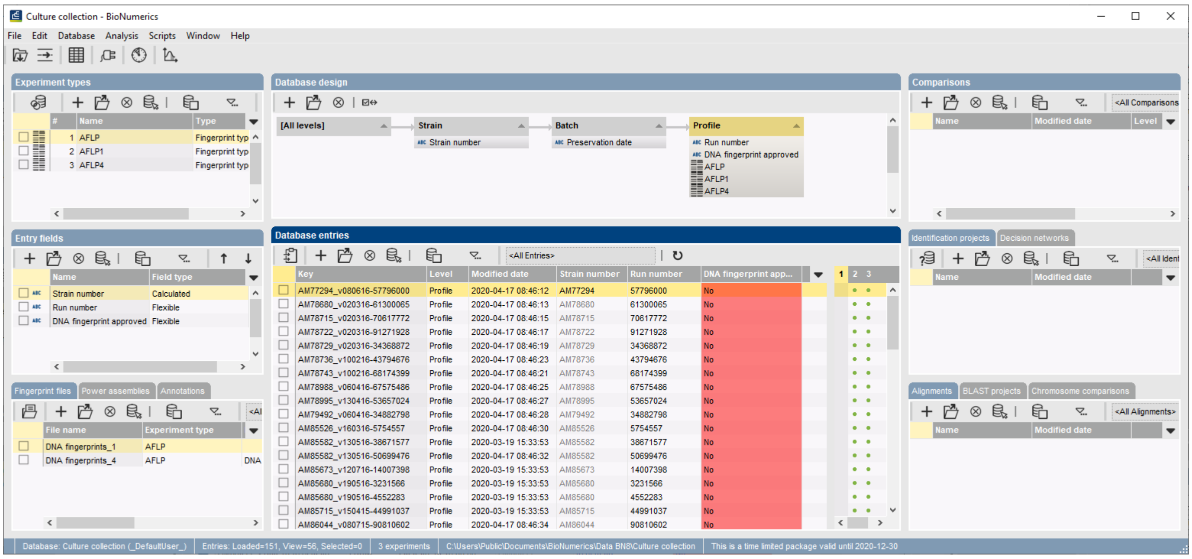
Import fingerprint curves

Database links
Link the imported records to database entries.
Double click on a cell to get an overview.

Overview				
In 'Profile'	☑ create 56 entries	and	□ update 0 entries
In 'Batch'	☑ create 50 entries	and	□ update 0 entries
In 'Strain'	☑ create 45 entries	and	□ update 0 entries

☑ Select modified entries

Figure 18: Database entries per level.



Culture collection - BioNumerics

#	Name	Type
1	AFLP	Fingerprint typ
2	AFLP1	Fingerprint typ
3	AFLP4	Fingerprint typ

Name	Field type
Strain number	Calculated
Run number	Flexible
DNA fingerprint approved	Flexible

File name	Experiment type
DNA fingerprints_1	AFLP
DNA fingerprints_4	AFLP

Key	Level	Modified date	Strain number	Run number	DNA fingerprint app...
AM77294_v080616-57796000	Profile	2020-04-17 08:46:12	AM77294	57796000	No
AM78680_v020316-61300065	Profile	2020-04-17 08:46:13	AM78680	61300065	No
AM78715_v020316-70617772	Profile	2020-04-17 08:46:15	AM78715	70617772	No
AM78722_v020316-91271928	Profile	2020-04-17 08:46:17	AM78722	91271928	No
AM78729_v020316-34368872	Profile	2020-04-17 08:46:19	AM78729	34368872	No
AM78736_v100216-43794676	Profile	2020-04-17 08:46:23	AM78736	43794676	No
AM78743_v100216-68174399	Profile	2020-04-17 08:46:21	AM78743	68174399	No
AM78988_v060416-67575486	Profile	2020-04-17 08:46:25	AM78988	67575486	No
AM78995_v130416-53657024	Profile	2020-04-17 08:46:27	AM78995	53657024	No
AM79492_v060416-34882798	Profile	2020-04-17 08:46:28	AM79492	34882798	No
AM85526_v160316-5754557	Profile	2020-04-17 08:46:30	AM85526	5754557	No
AM85582_v130516-38671577	Profile	2020-03-19 15:33:53	AM85582	38671577	No
AM85582_v130516-50699476	Profile	2020-04-17 08:46:32	AM85582	50699476	No
AM85673_v120716-14007398	Profile	2020-03-19 15:33:53	AM85673	14007398	No
AM85680_v190516-3231566	Profile	2020-03-19 15:33:53	AM85680	3231566	No
AM85680_v190516-4552283	Profile	2020-03-19 15:33:53	AM85680	4552283	No
AM85715_v150415-44991037	Profile	2020-03-19 15:33:53	AM85715	44991037	No
AM86044_v080715-90810602	Profile	2020-04-17 08:46:34	AM86044	90810602	No

Database: Culture collection (_DefaultUser_) | Entries: Loaded=151, View=56, Selected=0 | 3 experiments | C:\Users\Public\Documents\BioNumerics\Data BN8\Culture collection | This is a time limited package valid until 2020-12-30

Figure 19: The *Main* window after import of the curves: ***Profile*** level.

The data channel is now hidden from the view and its icon is displayed as 👁.

2. Use the zoom sliders on the left and on top to optimize the display of the fingerprint curves.

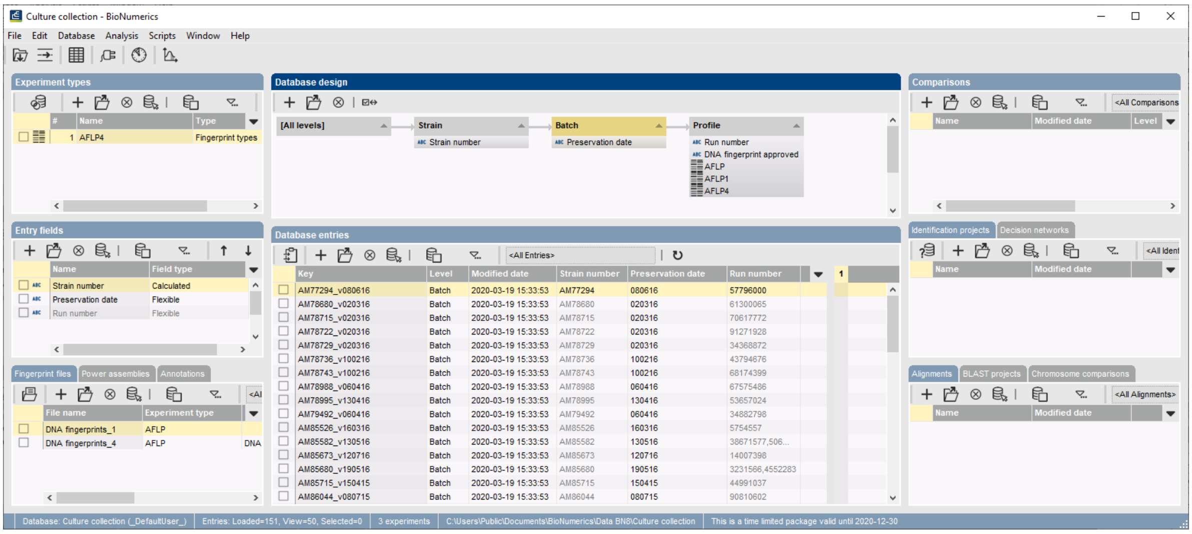
Since the raw chromatogram files have not undergone any preprocessing, normalization will have to be performed. This requires a *reference system* to be defined, based upon the marker peaks available in the reference dye.

3. Make sure the reference dye is the only dye visible in the upper panel (see Figure 21).

4. Select ***Bands*** > ***Search reference bands...*** (, **Ctrl+F**) to call the *Search reference bands* dialog box.

5. Specify a peak detection ***OD range*** of **2** (in %) and a peak detection ***Curve range*** of **5** (in %). Press <***OK***>.

The bands that fall within the specified criteria are marked with a solid line at the band's position



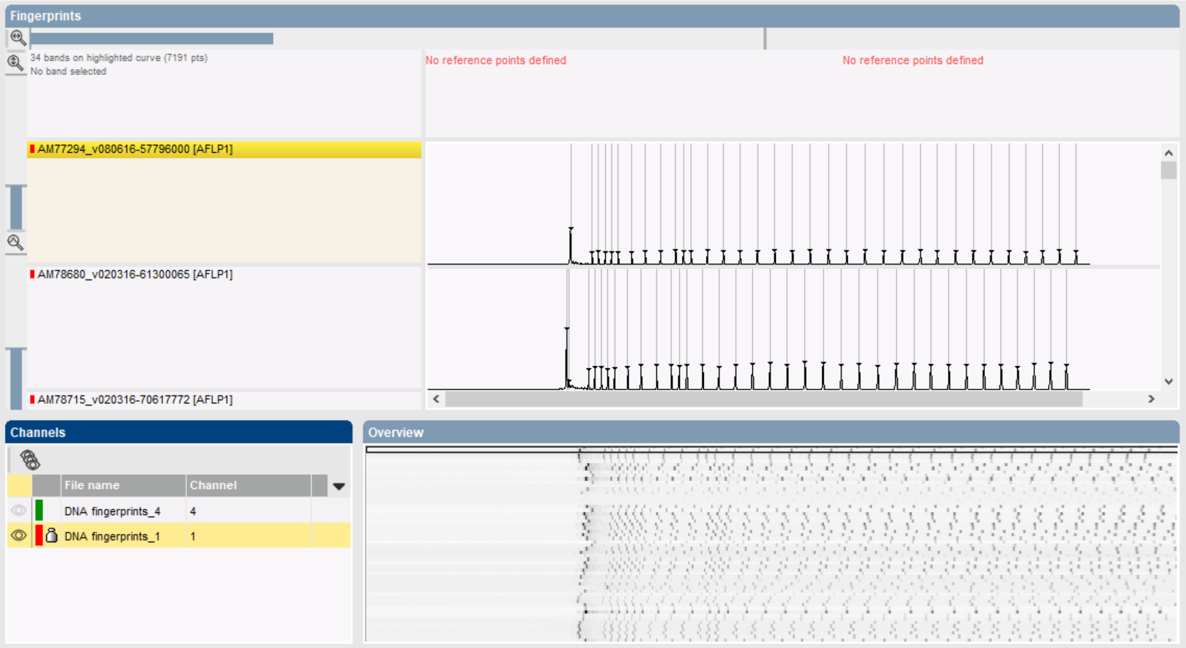
Database entries

Key	Level	Modified date	Strain number	Preservation date	Run number
AM77294_v080616	Batch	2020-03-19 15:33:53	AM77294	080616	57796000
AM78680_v020316	Batch	2020-03-19 15:33:53	AM78680	020316	61300065
AM78715_v020316	Batch	2020-03-19 15:33:53	AM78715	020316	70617772
AM78722_v020316	Batch	2020-03-19 15:33:53	AM78722	020316	91271928
AM78729_v020316	Batch	2020-03-19 15:33:53	AM78729	020316	34368872
AM78736_v100216	Batch	2020-03-19 15:33:53	AM78736	100216	43794676
AM78743_v100216	Batch	2020-03-19 15:33:53	AM78743	100216	68174399
AM78988_v060416	Batch	2020-03-19 15:33:53	AM78988	060416	67575486
AM78995_v130416	Batch	2020-03-19 15:33:53	AM78995	130416	53657024
AM79492_v060416	Batch	2020-03-19 15:33:53	AM79492	060416	34882798
AM85526_v160316	Batch	2020-03-19 15:33:53	AM85526	160316	5754557
AM85582_v130516	Batch	2020-03-19 15:33:53	AM85582	130516	38671577,506...
AM85673_v120716	Batch	2020-03-19 15:33:53	AM85673	120716	14007398
AM85680_v190516	Batch	2020-03-19 15:33:53	AM85680	190516	3231566,4552283
AM85715_v150415	Batch	2020-03-19 15:33:53	AM85715	150415	44991037
AM86044_v080715	Batch	2020-03-19 15:33:53	AM86044	080715	90810602

Database: Culture collection (_DefaultUser_) | Entries: Loaded=151, View=50, Selected=0 | 3 experiments | C:\Users\Public\Documents\BioNumerics\Data BN8\Culture collection | This is a time limited package valid until 2020-12-30

Figure 20: The *Main* window after import of the curves: ***Batch*** level.

(see Figure 21).



34 bands on highlighted curve (7191 pts)
No band selected

No reference points defined

No reference points defined

AM77294_v080616-57796000 [AFLP1]

AM78680_v020316-61300065 [AFLP1]

AM78715_v020316-70617772 [AFLP1]

File name	Channel
DNA fingerprints_4	4
DNA fingerprints_1	1

Figure 21: The *Fingerprint curve processing* window only displaying the reference dye.

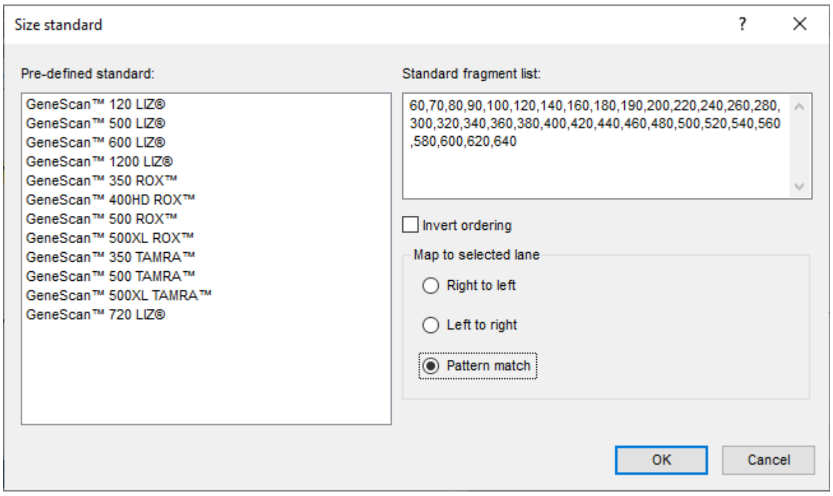
6. Select a suitable lane and then select ***References*** > ***Define size standard...***.

This will display the *Size standard* dialog box, from which a size marker can be selected. In the example curve files, the size standard is not listed. The molecular weights of the reference can be copied from the text file `SizeMarker.txt`.

7. Paste the copied weights to the *List* panel and check ***Pattern match*** (see Figure 22). Press <***OK***> twice.

8. Save the data to the database with ***File*** > ***Save*** (🖫, **Ctrl+S**) and confirm.

The software will automatically create the reference system and calibration curve for each of the



Size standard

Pre-defined standard:

GeneScan™ 120 LIZ®
GeneScan™ 500 LIZ®
GeneScan™ 600 LIZ®
GeneScan™ 1200 LIZ®
GeneScan™ 350 ROX™
GeneScan™ 400HD ROX™
GeneScan™ 500 ROX™
GeneScan™ 500XL ROX™
GeneScan™ 350 TAMRA™
GeneScan™ 500 TAMRA™
GeneScan™ 500XL TAMRA™
GeneScan™ 720 LIZ®

Standard fragment list:

60,70,80,90,100,120,140,160,180,190,200,220,240,260,280,300,320,340,360,380,400,420,440,460,480,500,520,540,560,580,600,620,640

□ Invert ordering

Map to selected lane

○ Right to left

○ Left to right

● Pattern match

Figure 22: Define size standard.

fingerprint types. Since this allows the calculation of metrics information, a metrics scale now becomes available in the upper part of the *Fingerprints* panel.

Normalization is achieved by assigning bands in the reference channel to external reference positions.

9. To normalize a complete run at once, select ***Normalization*** > ***Auto assign reference positions (all lanes)...*** (, **Ctrl+A**), leave all settings unaltered and press <***OK***>.

10. When the assignment of the marker bands to reference positions is made, the data can be shown in normalized mode with ***Normalization*** > ***Show normalized view*** (, **Shift+N**).

11. Click on the icon left of the **AFLP4** and **AFLP1** channels in the *Channels* panel.

The data channel is now shown and the reference channel is hidden from the view.

12. Select ***Bands*** > ***Search data bands...*** (, **Ctrl+Shift+F**) to call the *Search data bands* dialog box.

13. Specify a peak detection ***OD range*** of **2** (in %) and a peak detection ***Curve range*** of **1** (in %). Check ***Filter by fragment length***, specify a ***Minimum fragment length*** of **35** and press <***OK***>.

14. Optionally the *Search data bands* dialog box can be called again to optimize the band search settings.

If you want to calculate the similarity between the AFLP4 data curves in the *Comparison* window using a curve based coefficient (e.g. ***Pearson correlation***), the search for bands can be skipped in the data channel.

15. Save the changes and close the *Fingerprint curve processing* window.

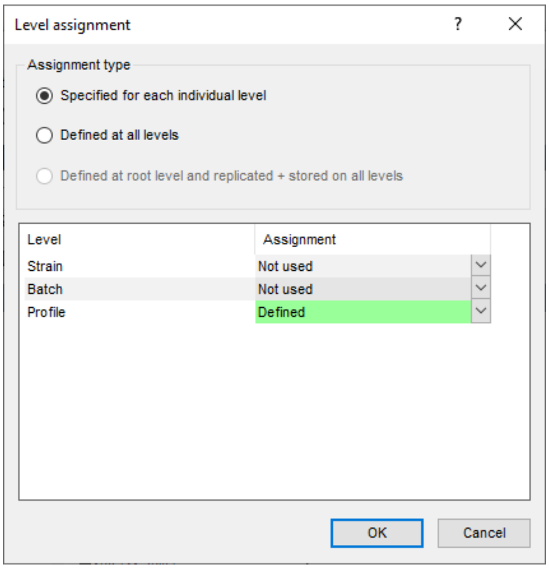
5 Summarizing data

Similar to information fields (see tutorial ”Setting up a BIONUMERICS database with levels”) experiment types can also be defined at a specific database levels and replicated over other levels if desired.

In this example, the profiles are linked to the fingerprint experiment type **AFLP4** in the **Profile** level. Replicated measurements are present and since we want to pick a representative pattern and use it at the **Batch** and **Strain** levels, we need to replicate the experiment type at these levels:

1. Double-click on **AFLP4** in the *Experiment types* panel to open the *Fingerprint type* window for this experiment type.

2. Select ***Settings*** > ***Level assignment...*** to show the *Level assignment* dialog box (see Figure 23).



Level assignment

Assignment type

● Specified for each individual level

○ Defined at all levels

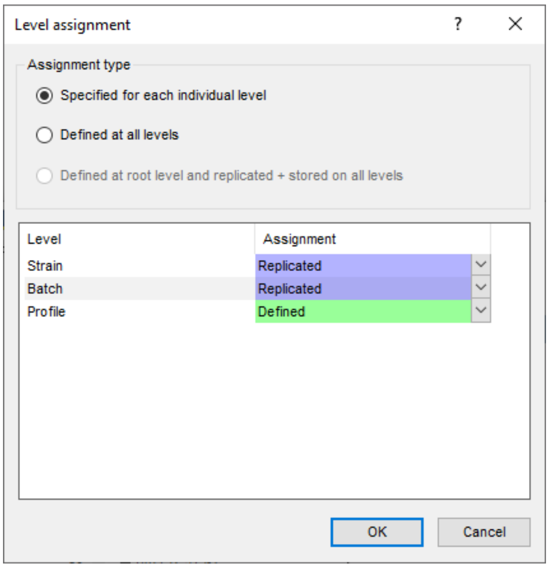
○ Defined at root level and replicated + stored on all levels

Level	Assignment
Strain	Not used
Batch	Not used
Profile	Defined

Figure 23: The initial level assignment settings for ***AFLP4***.

3. Use the corresponding drop-down list next to **Strain** to change the ***Assignment*** from “Not used” to “Replicated”.

4. Repeat the previous step for the **Batch** level and press <***OK***> (see Figure 24).



Level assignment

Assignment type

● Specified for each individual level

○ Defined at all levels

○ Defined at root level and replicated + stored on all levels

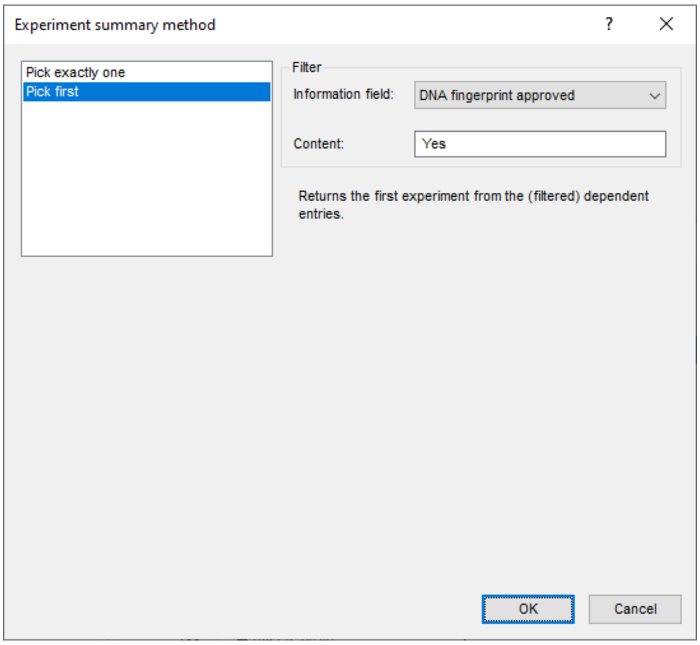
Level	Assignment
Strain	Replicated
Batch	Replicated
Profile	Defined

Figure 24: Updated level assignments for ***AFLP4***.

> An experiment type that was created in a database without levels or an experiment type that was created on-the-fly (e.g. during import of experiment data) will be defined at all levels. In this case, you need to check ***Specified for each individual level*** and manually specify at which level the experiment type should be defined and replicated.

The method used to replicate the experimental information is specified in the summary replication settings.

5. In the *Fingerprint type* window, select ***Settings*** > ***Summary replication settings...*** to open the *Experiment summary method* dialog box (see Figure 25).



Experiment summary method

Pick exactly one
Pick first

Filter

Information field: DNA fingerprint approved

Content: Yes

Returns the first experiment from the (filtered) dependent entries.

Figure 25: Summary replication settings for **AFLP4**.

We will specify following filter: if the field ***DNA fingerprint approved*** contains the text “Yes”, this means that a curator has checked the profile and the profile can be considered for replication to the parent levels.

With the option ***Pick exactly one***, no experiment will be created at the parent level(s) if there is more than one fingerprint approved. With the option ***Pick first*** the first encountered pattern will be replicated.

6. Select ***Pick first***, pick ***DNA fingerprint approved*** from the drop-down list as ***Information field***, enter “Yes” as ***Content*** (see Figure 25) and press <***OK***>. Close the *Fingerprint type* window.

Since no fingerprints are yet approved (***DNA fingerprint approved*** all set to ”No”) no profiles are summarized in the ***Batch*** and ***Strain*** levels.

The imported profiles can be checked and compared with the original DNA fingerprints from the strains. The original DNA fingerprints can be imported and stored in the database and flagged with a specific information field (e.g. an information field ***Reference*** with the states ***Yes*** and ***No***).

In this example database, the reference profiles are not present so they cannot be included in the clustering. For completeness, the clustering steps are given below:

7. Make a selection of profiles in the **Profile** level. Press <***Ctrl+A***> to select all profiles at once and press ***Edit*** > ***Create new object...*** (+) in the *Comparisons* panel to create a comparison.

8. Click on the 👁 next to the experiment name **AFLP4** in the *Experiments* panel to display the **AFLP4** patterns in the *Experiment data* panel.

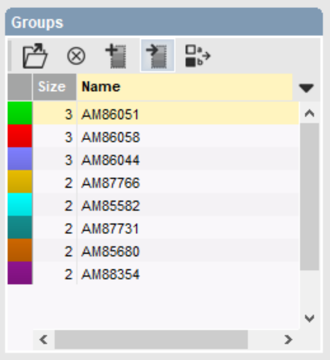
Comparison groups can be defined from clusters, from database fields, or just from any selection you want. Here we will let BIONUMERICS create groups based on the **Strain number**.

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

10. Keep the first option selected, specify a ***Maximum count*** of ”8” since only 8 strains have replicates in this comparison and confirm.

A message will appear listing the number of groups found in the selected field and the number of groups created based on the defined settings.

The groups are listed in the *Groups* panel (see Figure 26). The group color is displayed next to each entry in the *Information fields* panel.



	Size	Name
🟩	3	AM86051
🟥	3	AM86058
🟦	3	AM86044
🟨	2	AM87766
🟦	2	AM85582
🟩	2	AM87731
🟧	2	AM85680
🟪	2	AM88354

Figure 26: Comparison groups.

11. Make sure **AFLP4** is selected in the *Experiments* panel and select ***Clustering*** > ***Calculate*** > ***Cluster analysis (similarity matrix)...***.

12. Select ***Pearson correlation*** from the list.

13. Enter an ***Optimization*** of 0.1%, and leave all the other settings at their defaults.

The ***Optimization*** setting limits the amount of movement for each fingerprint as a whole.

14. Press <***Next***>.

15. Select ***UPGMA*** and press <***Finish***> to start the cluster analysis.

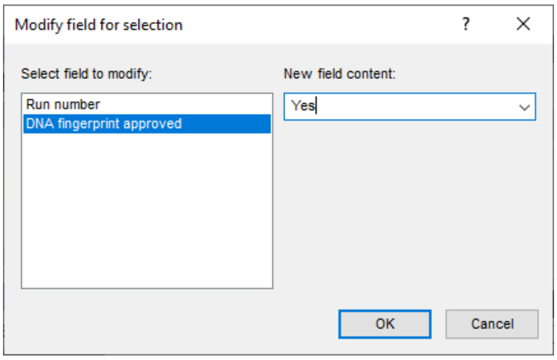
When finished, the dendrogram and the similarity matrix are displayed in their corresponding panels. The cluster analysis is listed in the *Analyses* panel of the *Comparison* window.

If the reference profiles were present for each strain in the *Comparison* window, one could very easily see which profiles have a perfect 100% match with their reference. For these accepted profiles the state of the information field ***DNA fingerprint approved*** can be changed from ”No” to ”Yes”:

16. Select all profiles that have a 100% match with their reference profile.

17. Close the *Comparison* window and make sure the *Database entries* panel is the active panel in the *Main* window. Select ***Edit*** > ***Information fields*** > ***Edit field in selection...*** (**Ctrl+M**), select ***DNA fingerprint approved*** as field, and set the new field content to ***Yes*** (see Figure 27). Press <***Yes***> twice to confirm the action.

The field content is updated (see Figure 28) and the accepted profiles are replicated to the parent levels.



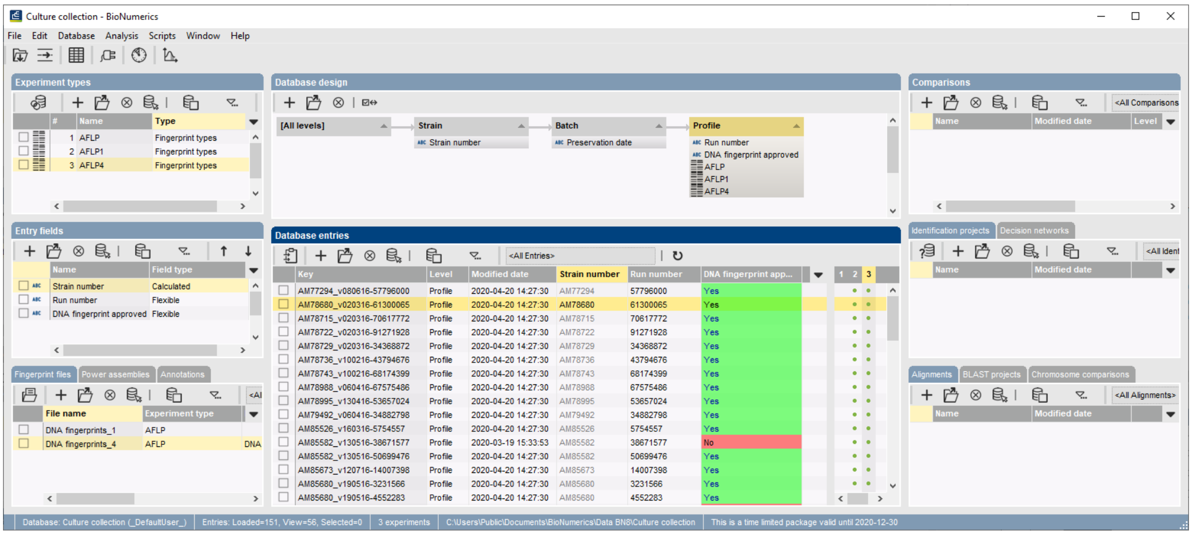
Modify field for selection

Select field to modify:

- Run number
- DNA fingerprint approved

New field content: Yes

Figure 27: Modify field for selection.



#	Name	Type
1	AFLP	Fingerprint types
2	AFLP1	Fingerprint types
3	AFLP4	Fingerprint types

Name	Field type
Strain number	Calculated
Run number	Flexible
DNA fingerprint approved	Flexible

File name	Experiment type
DNA fingerprints_1	AFLP
DNA fingerprints_4	AFLP

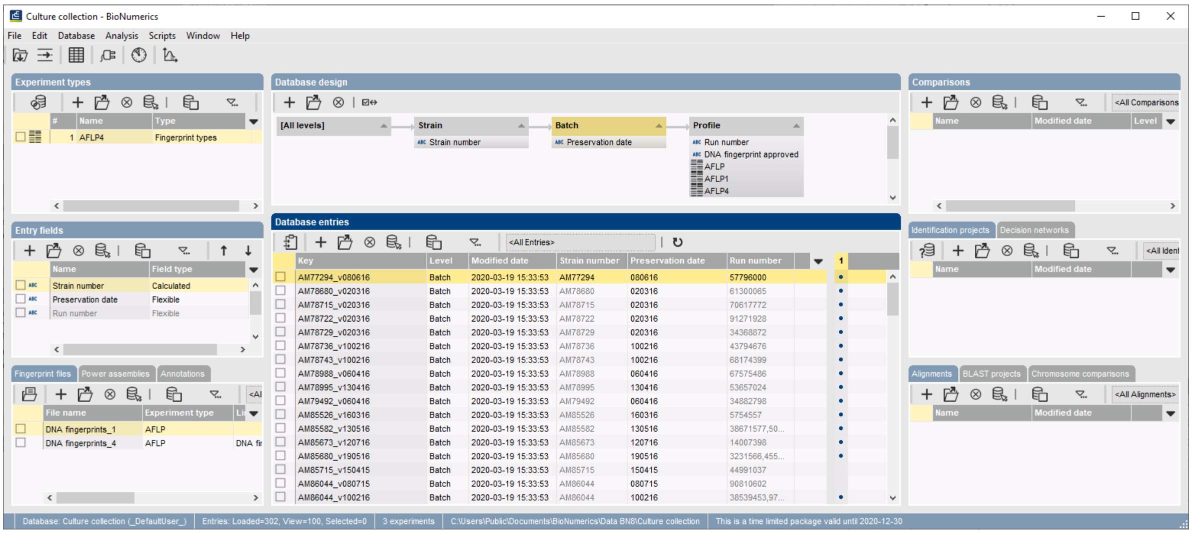
Database entries

Key	Level	Modified date	Strain number	Run number	DNA fingerprint app...
AM77294_v080616-57796000	Profile	2020-04-20 14:27:30	AM77294	57796000	Yes
AM78680_v020316-61300065	Profile	2020-04-20 14:27:30	AM78680	61300065	Yes
AM78715_v020316-70617772	Profile	2020-04-20 14:27:30	AM78715	70617772	Yes
AM78722_v020316-91271928	Profile	2020-04-20 14:27:30	AM78722	91271928	Yes
AM78729_v020316-34368872	Profile	2020-04-20 14:27:30	AM78729	34368872	Yes
AM78736_v100216-43794676	Profile	2020-04-20 14:27:30	AM78736	43794676	Yes
AM78743_v100216-68174399	Profile	2020-04-20 14:27:30	AM78743	68174399	Yes
AM78988_v060416-67575486	Profile	2020-04-20 14:27:30	AM78988	67575486	Yes
AM78995_v130416-53657024	Profile	2020-04-20 14:27:30	AM78995	53657024	Yes
AM79492_v060416-34882798	Profile	2020-04-20 14:27:30	AM79492	34882798	Yes
AM85526_v160316-5754557	Profile	2020-04-20 14:27:30	AM85526	5754557	Yes
AM85582_v130516-38671577	Profile	2020-03-19 15:33:53	AM85582	38671577	No
AM85582_v130516-50699476	Profile	2020-04-20 14:27:30	AM85582	50699476	Yes
AM85673_v120716-14007398	Profile	2020-04-20 14:27:30	AM85673	14007398	Yes
AM85680_v190516-3231566	Profile	2020-04-20 14:27:30	AM85680	3231566	Yes
AM85680_v190516-4552283	Profile	2020-04-20 14:27:30	AM85680	4552283	Yes

Database: Culture collection (_DefaultUser_) | Entries: Loaded=151, View=56, Selected=0 | 3 experiments | C:\Users\Public\Documents\BioNumerics\Data BN8\Culture collection | This is a time limited package valid until 2020-12-30

Figure 28: Modified fields.

If you click on the ***Batch*** or ***Strain*** level in the *Database design* panel the replicated profiles are indicated with a blue dot in the *Experiment presence* panel (see Figure 29).



#	Name	Type
1	AFLP4	Fingerprint types

Name	Field type
Strain number	Calculated
Preservation date	Flexible
Run number	Flexible

File name	Experiment type
DNA fingerprints_1	AFLP
DNA fingerprints_4	AFLP

Database entries

Key	Level	Modified date	Strain number	Preservation date	Run number	1
AM77294_v080616	Batch	2020-03-19 15:33:53	AM77294	080616	57796000	●
AM78680_v020316	Batch	2020-03-19 15:33:53	AM78680	020316	61300065	●
AM78715_v020316	Batch	2020-03-19 15:33:53	AM78715	020316	70617772	●
AM78722_v020316	Batch	2020-03-19 15:33:53	AM78722	020316	91271928	●
AM78729_v020316	Batch	2020-03-19 15:33:53	AM78729	020316	34368872	●
AM78736_v100216	Batch	2020-03-19 15:33:53	AM78736	100216	43794676	●
AM78743_v100216	Batch	2020-03-19 15:33:53	AM78743	100216	68174399	●
AM78988_v060416	Batch	2020-03-19 15:33:53	AM78988	060416	67575486	●
AM78995_v130416	Batch	2020-03-19 15:33:53	AM78995	130416	53657024	●
AM79492_v060416	Batch	2020-03-19 15:33:53	AM79492	060416	34882798	●
AM85526_v160316	Batch	2020-03-19 15:33:53	AM85526	160316	5754557	●
AM85582_v130516	Batch	2020-03-19 15:33:53	AM85582	130516	38671577,50...	●
AM85673_v120716	Batch	2020-03-19 15:33:53	AM85673	120716	14007398	●
AM85680_v190516	Batch	2020-03-19 15:33:53	AM85680	190516	3231566,455...	●
AM85715_v150415	Batch	2020-03-19 15:33:53	AM85715	150415	44991037	
AM86044_v080715	Batch	2020-03-19 15:33:53	AM86044	080715	90810602	
AM86044_v100216	Batch	2020-03-19 15:33:53	AM86044	100216	38539453,97...	●

Database: Culture collection (_DefaultUser_) | Entries: Loaded=302, View=100, Selected=0 | 3 experiments | C:\Users\Public\Documents\BioNumerics\Data BN8\Culture collection | This is a time limited package valid until 2020-12-30

Figure 29: Replicated data displayed in parent levels.