anses
French agency for food, environmental and occupational health & safety

Laboratory for Food Safety
Maisons-Alfort location

EURL Lm

European Union Reference Laboratory for
Listeria monocytogenes
http://eurl-listeria.anses.fr

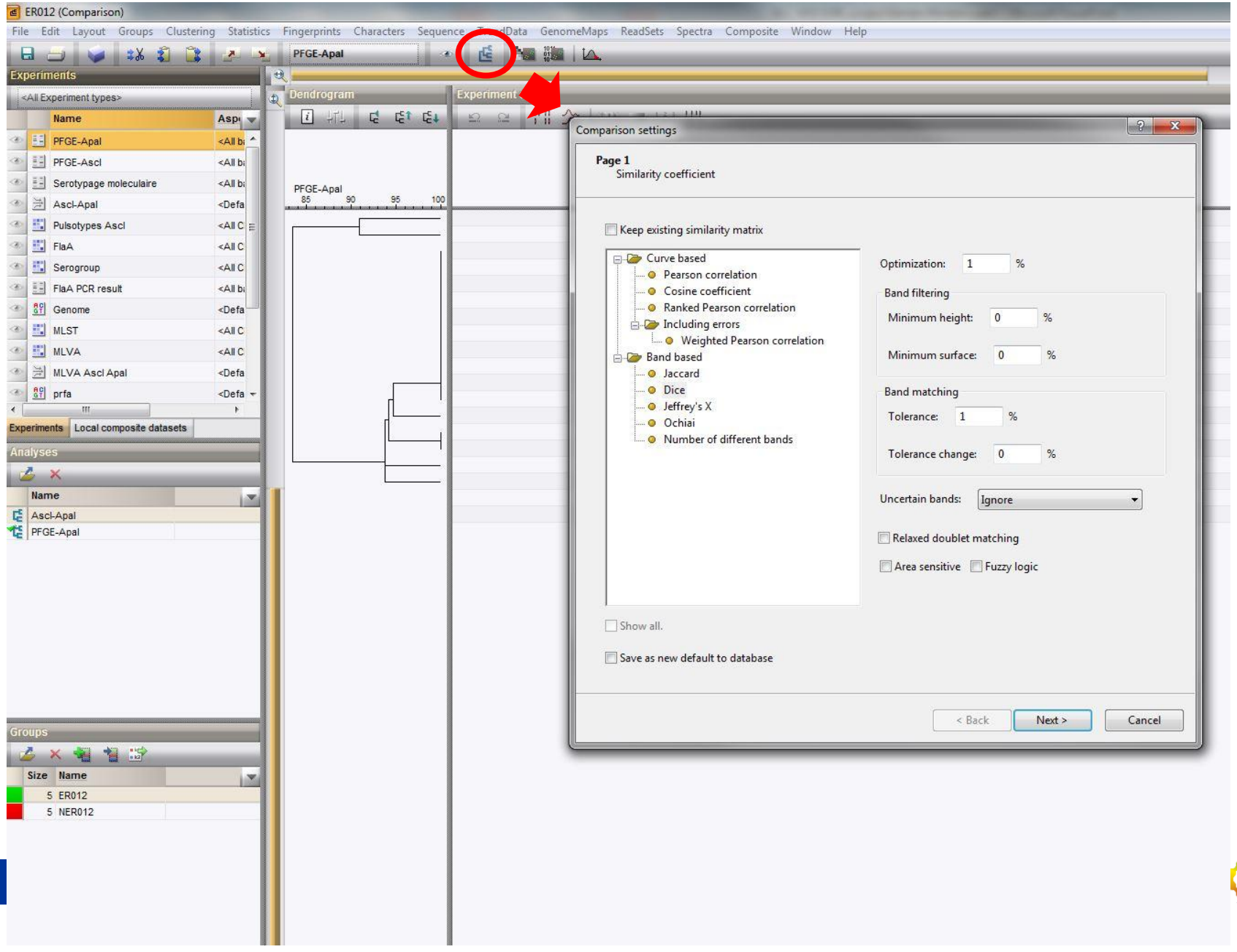
Introduction to cluster analysis purpose and parameters and Global Database Management

Benjamin FELIX

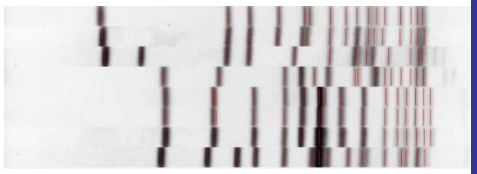
Joint training course on the use of BioNumerics software to analyse PFGE data
Rome 3-4 July 2017

French Agency for Food, Environmental and Occupational Health Safety – Food Safety Laboratory
Maisons-Alfort location and Boulogne-sur-Mer location
14, rue Pierre et Marie Curie - F94701 Maisons-Alfort Cedex - Telephone: + 33 149 771 350 - Fax: + 33 149 771 380
Boulevard du Bassin Napoléon - F62200 Boulogne-sur-Mer - Téléphone : +33 321 992 500 – Télécopie : + 33 321 991 725
www.anses.fr - @Anses_fr

Clustering

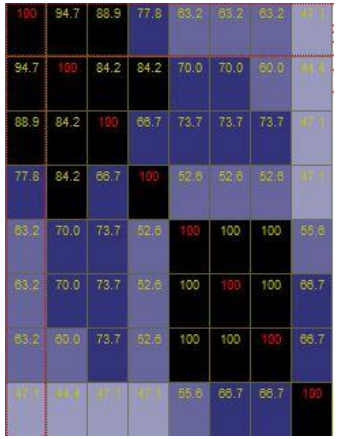

ER012 (Comparison)
File Edit Layout Groups Clustering Statistics Fingerprints Characters Sequence GenomeMaps ReadSets Spectra Composite Window Help
PFGE-ApaI
Experiments
<All Experiment types>
Name
PFGE-ApaI
PFGE-AscI
Serotypage moleculaire
AscI-ApaI
Pulsotypes AscI
FlaA
Serogroup
FlaA PCR result
Genome
MLST
MLVA
MLVA AscI ApaI
prfa
Experiments
Local composite datasets
Analyses
Name
AscI-ApaI
PFGE-ApaI
Groups
Size
Name
5 ER012
5 NER012
Dendrogram
PFGE-ApaI
85
90
95
100
Comparison settings
Page 1
Similarity coefficient
Keep existing similarity matrix
Curve based
Pearson correlation
Cosine coefficient
Ranked Pearson correlation
Including errors
Weighted Pearson correlation
Band based
Jaccard
Dice
Jeffrey's X
Ochiai
Number of different bands
Optimization: 1 %
Band filtering
Minimum height: 0 %
Minimum surface: 0 %
Band matching
Tolerance: 1 %
Tolerance change: 0 %
Uncertain bands: Ignore
Relaxed doublet matching
Area sensitive
Fuzzy logic
Show all.
Save as new default to database
< Back
Next >
Cancel

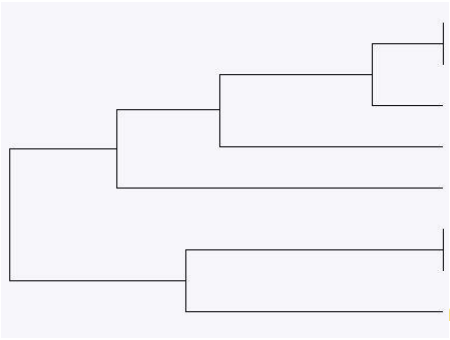
Cluster Analysis



PFGE Profiles

Set of band-based profiles



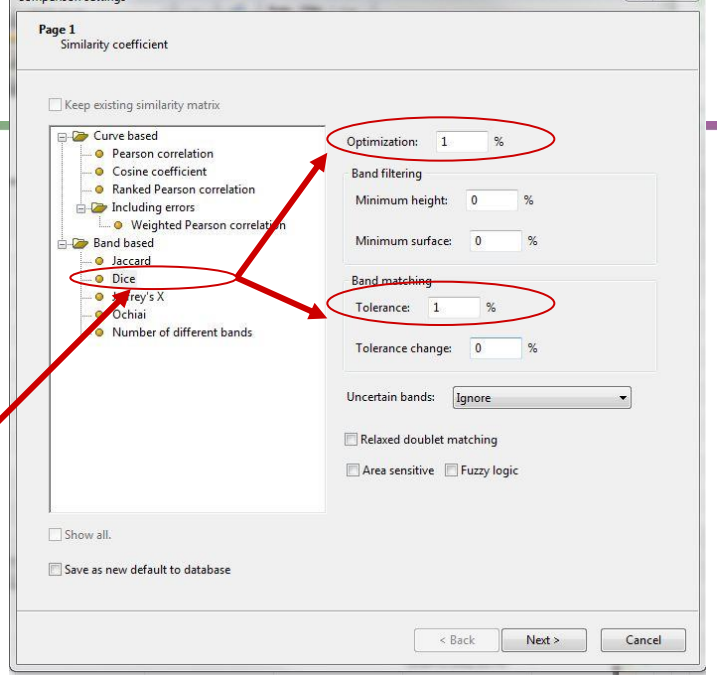


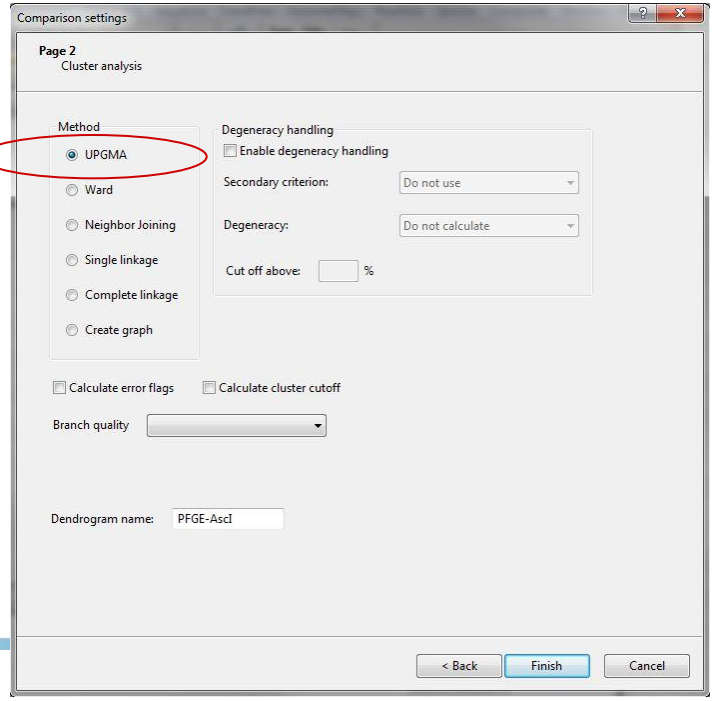
Coefficient calculation:
Algorithm « Dice »
Binary: Band absence or presence

Similarity matrix

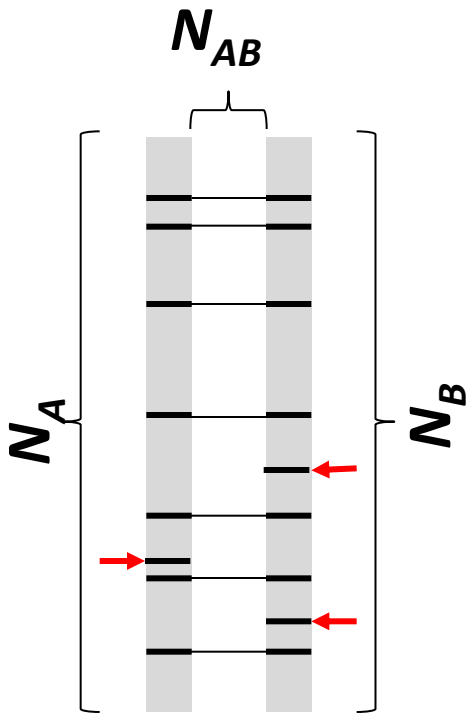
Clustering:
Algorithm « UPGMA »

Dendrogram/clustering





Dice coefficient: Band-based similarities



Dice:

$$S_D = \frac{2N_{AB}}{N_A + N_B} \qquad S_D = \frac{2N_{common}}{N_{Total}}$$

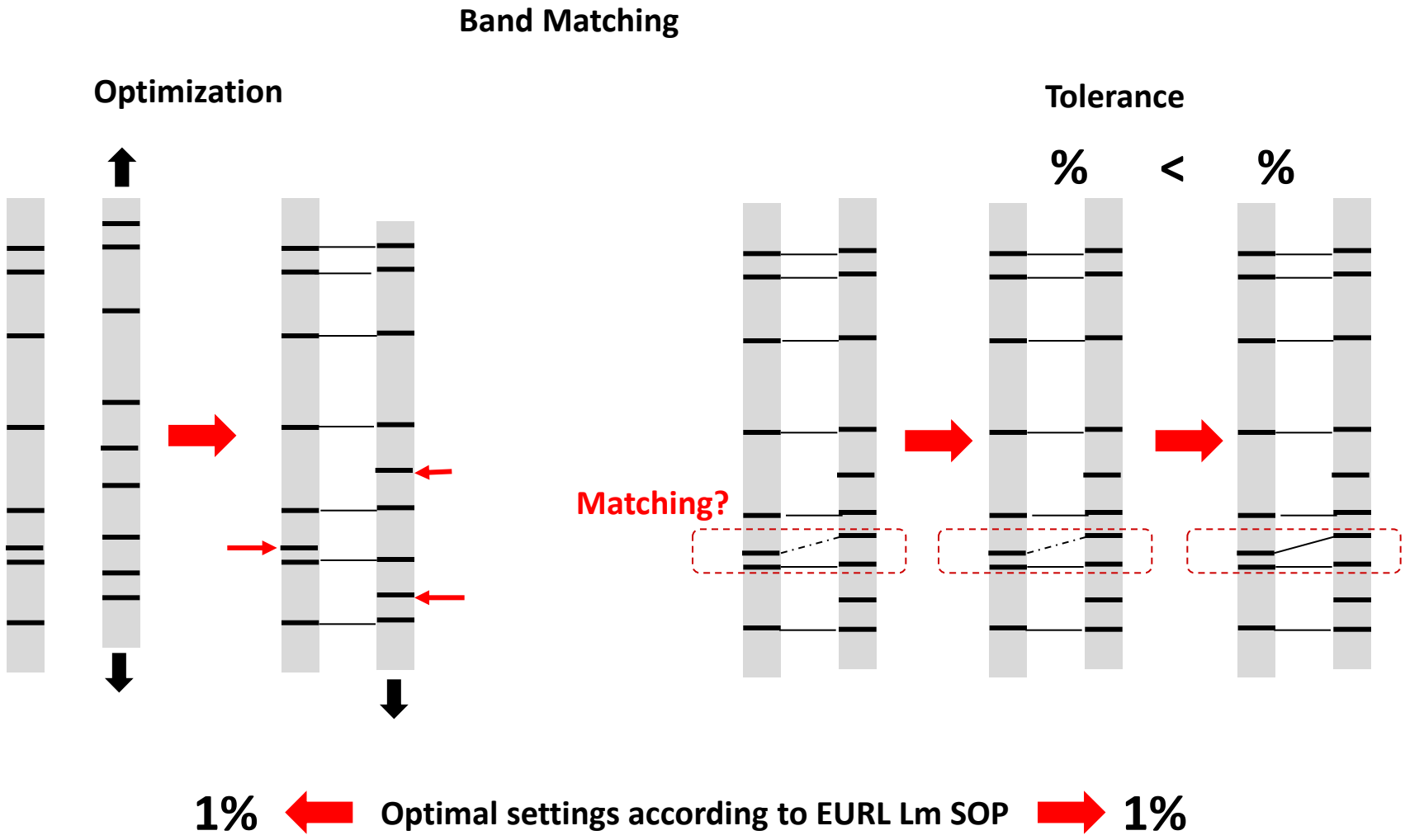
$$S_D = \frac{2x7}{8+9} = 82\%$$

3 bands of difference

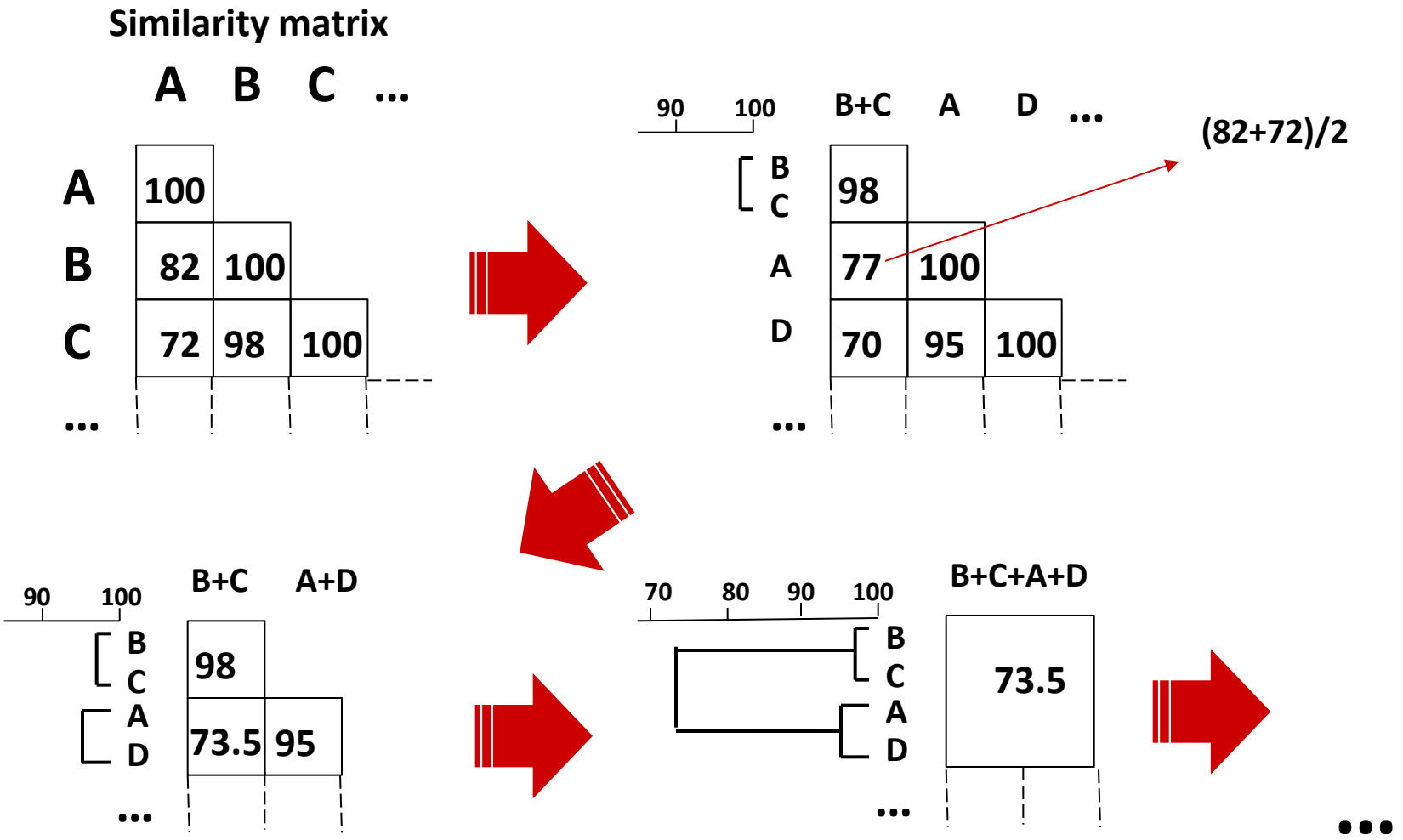
Similarity matrix

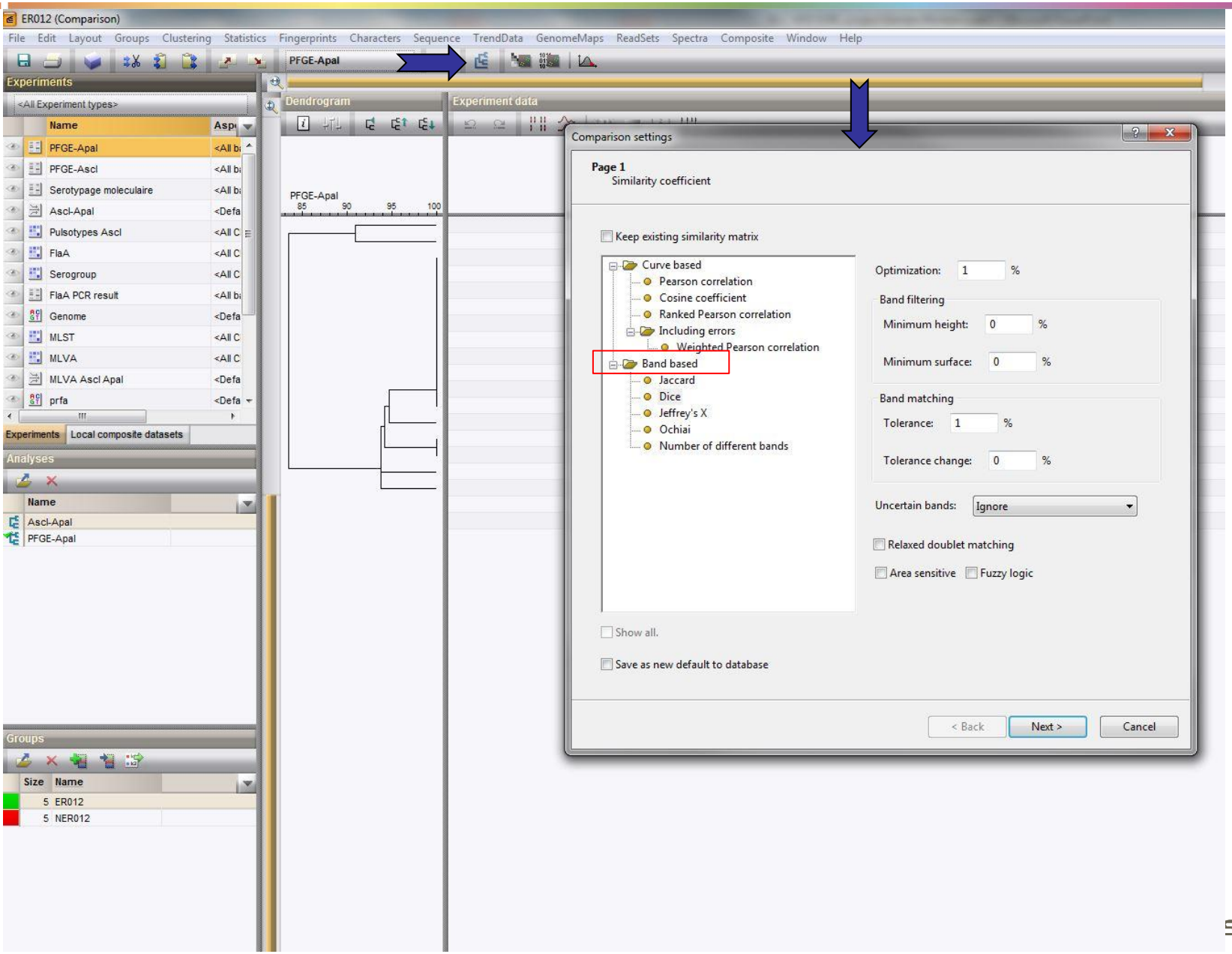
	A	B	C	…
A	100			
B	82	100		
C	72	90	100	
…				

Dice coefficient: Optimization and tolerance



UPGMA: Unweighted Pair Group Method using Arithmetic average

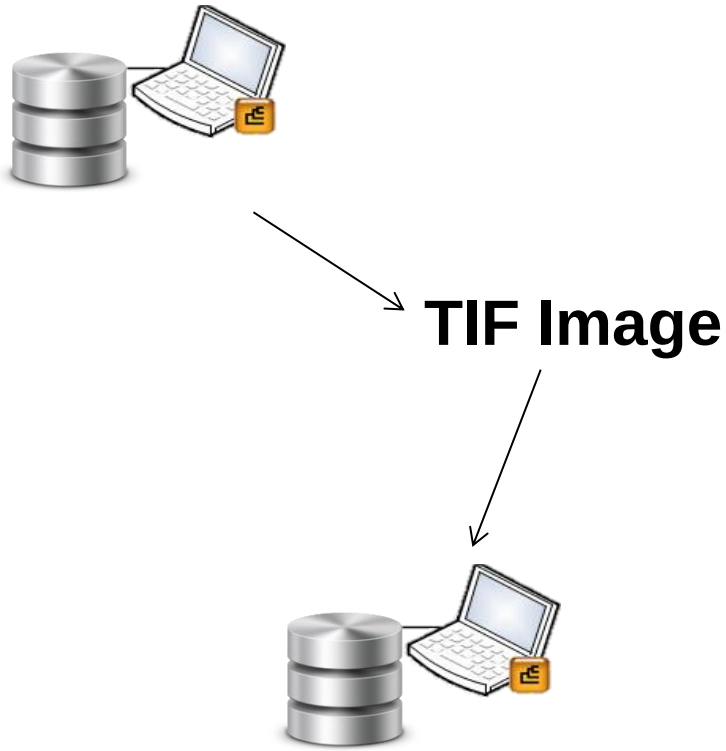



ER012 (Comparison)
File Edit Layout Groups Clustering Statistics Fingerprints Characters Sequence TrendData GenomeMaps ReadSets Spectra Composite Window Help
PFGE-ApaI
Experiments
<All Experiment types>
Name
PFGE-ApaI
PFGE-AscI
Serotypage moleculaire
AscI-ApaI
Pulsotypes AscI
FlaA
Serogroup
FlaA PCR result
Genome
MLST
MLVA
MLVA AscI ApaI
prfa
Experiments
Local composite datasets
Analyses
Name
AscI-ApaI
PFGE-ApaI
Groups
Size
Name
5 ER012
5 NER012
Dendrogram
Experiment data
PFGE-ApaI
85
90
95
100
Comparison settings
Page 1
Similarity coefficient
Keep existing similarity matrix
Curve based
Pearson correlation
Cosine coefficient
Ranked Pearson correlation
Including errors
Weighted Pearson correlation
Band based
Jaccard
Dice
Jeffrey's X
Ochiai
Number of different bands
Optimization: 1 %
Band filtering
Minimum height: 0 %
Minimum surface: 0 %
Band matching
Tolerance: 1 %
Tolerance change: 0 %
Uncertain bands: Ignore
Relaxed doublet matching
Area sensitive
Fuzzy logic
Show all.
Save as new default to database
< Back
Next >
Cancel

Database management

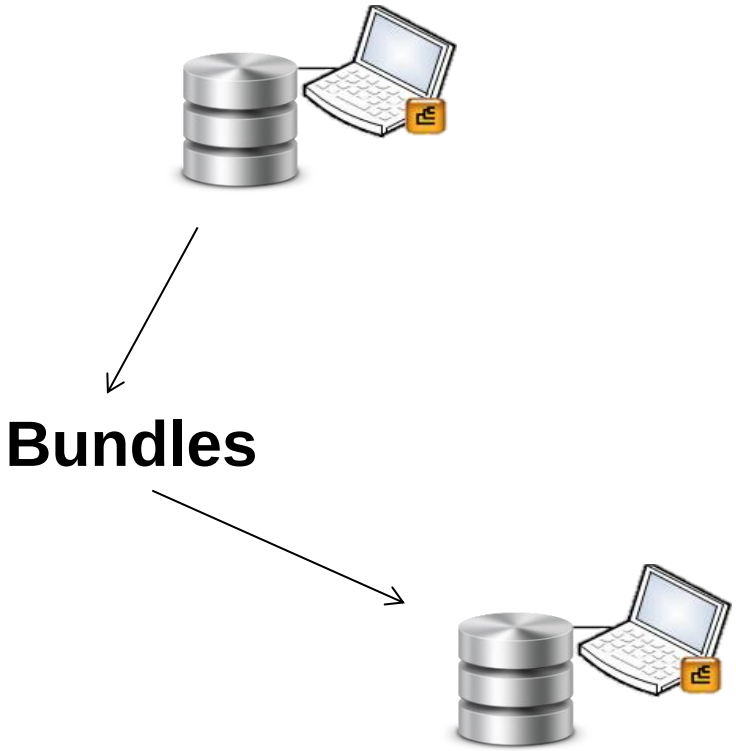
Communication between databases

Database exchange



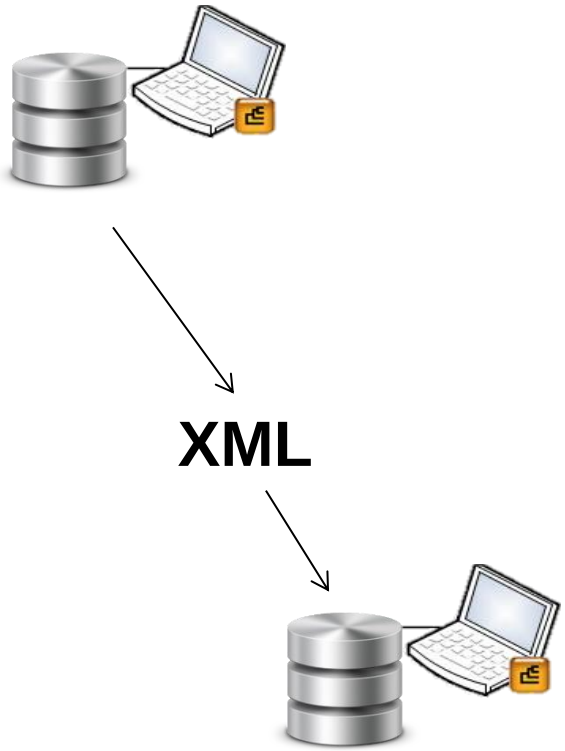
- Permant import of fingerprint, need image processing, normalization and analysis

Database exchange



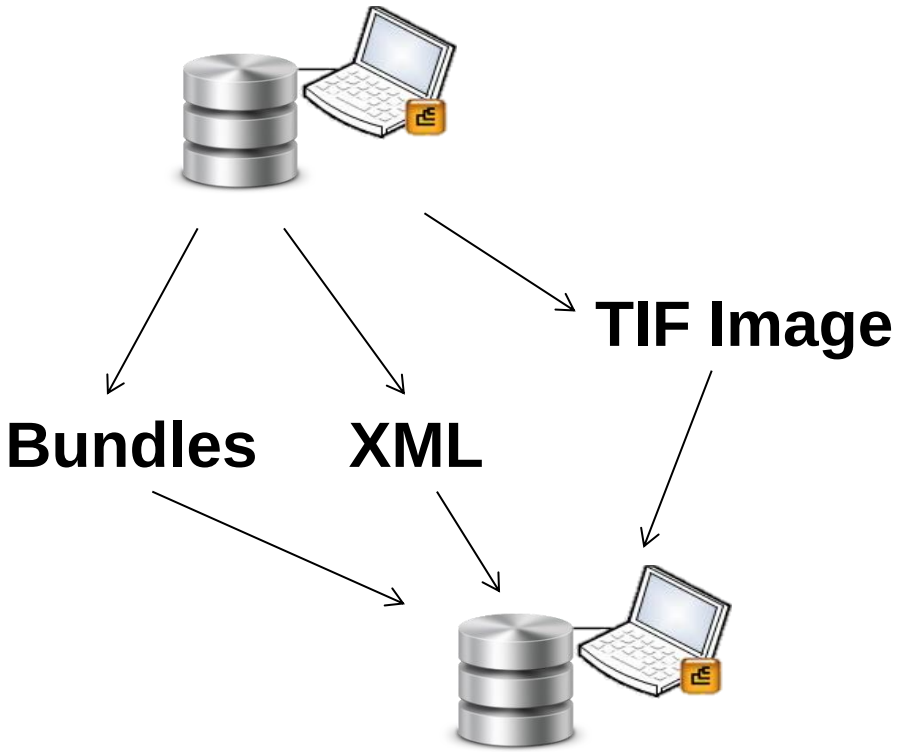
- Profile are temporally available in the database (profiles, sequences and fields) until the software shut down

Database exchange

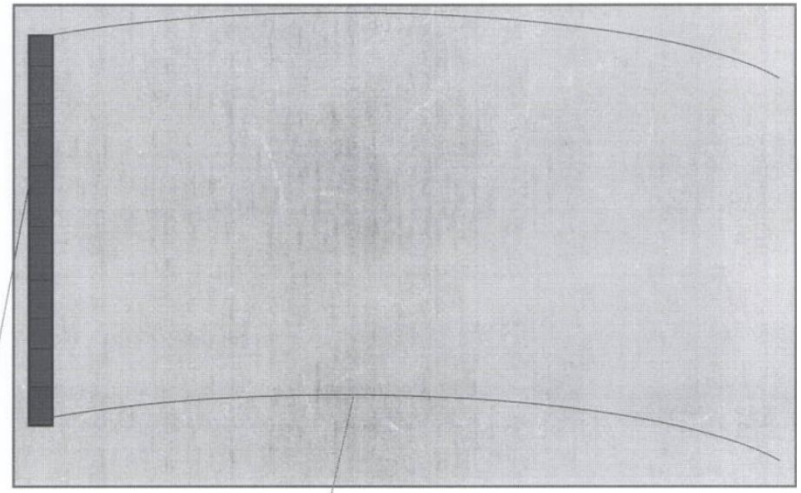


- Permanent import of data (profiles, sequences and fields) without any analysis.

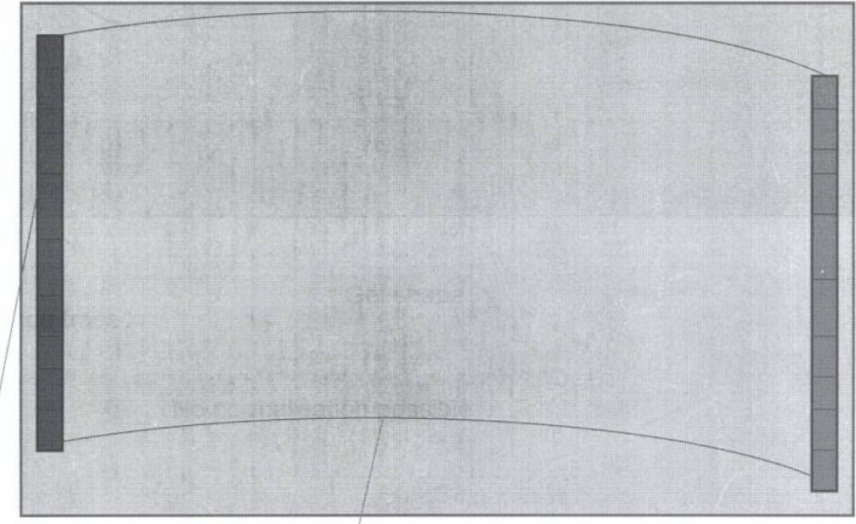
Database exchange



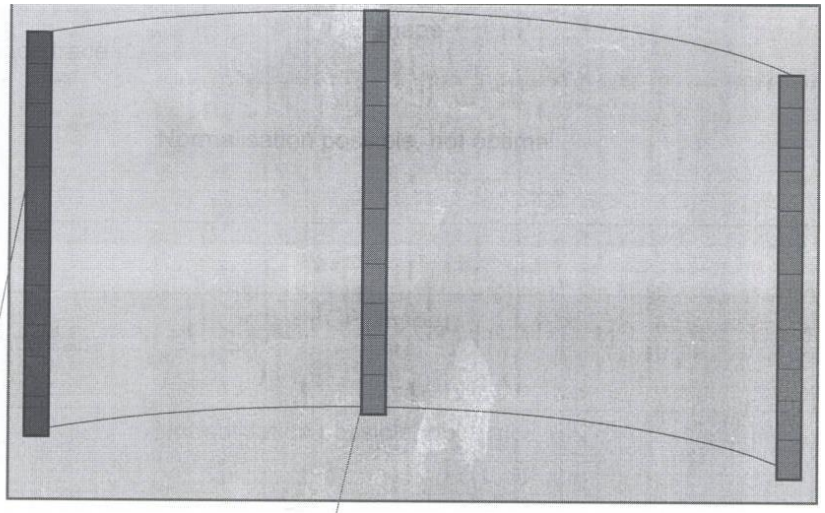
Database compatibility



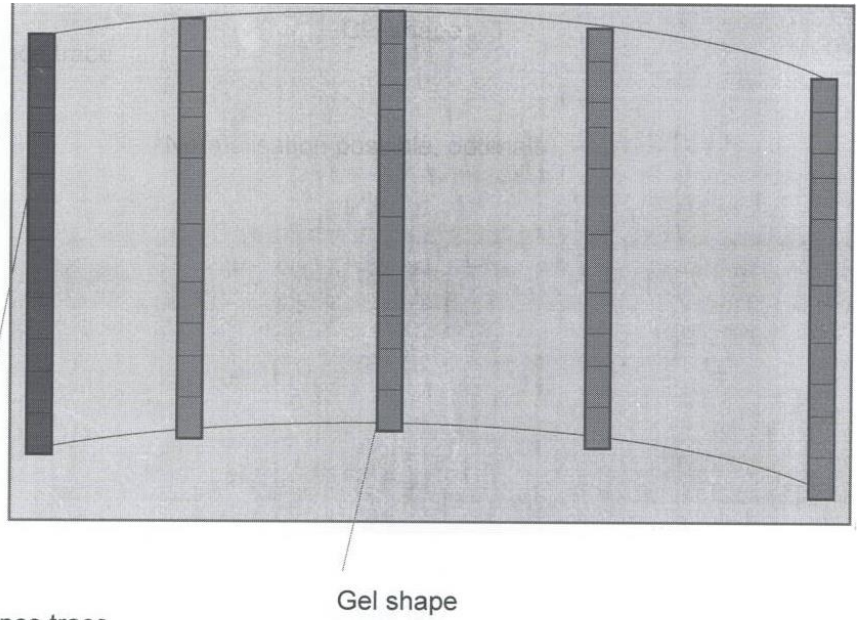
No normalisation possible



No normalisation possible



Normalisation possible, not optimal



Normalisation possible, optimal!

Reference system

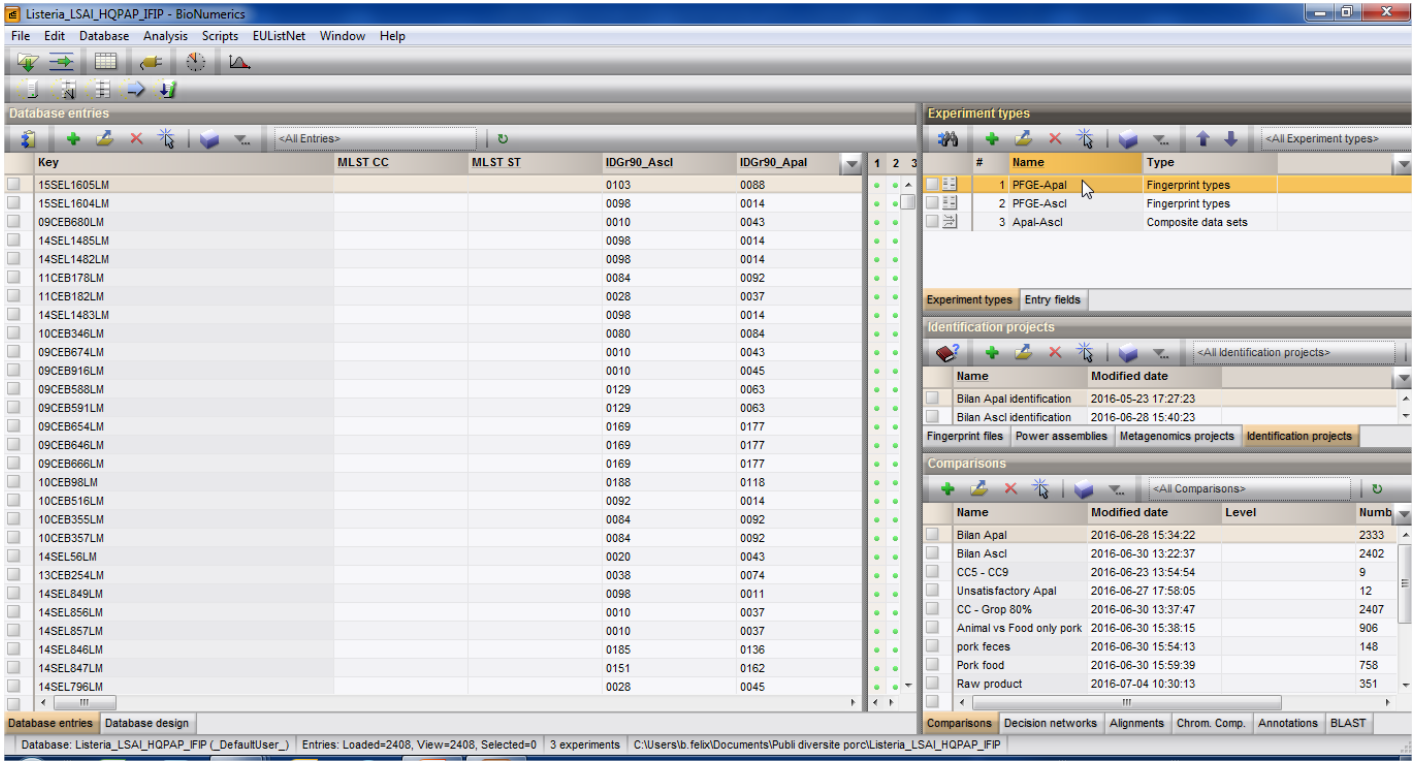
Reference system included in the gel:

- Salmonella Braenderup H9812 *Xba*I restriction profile

Reference system

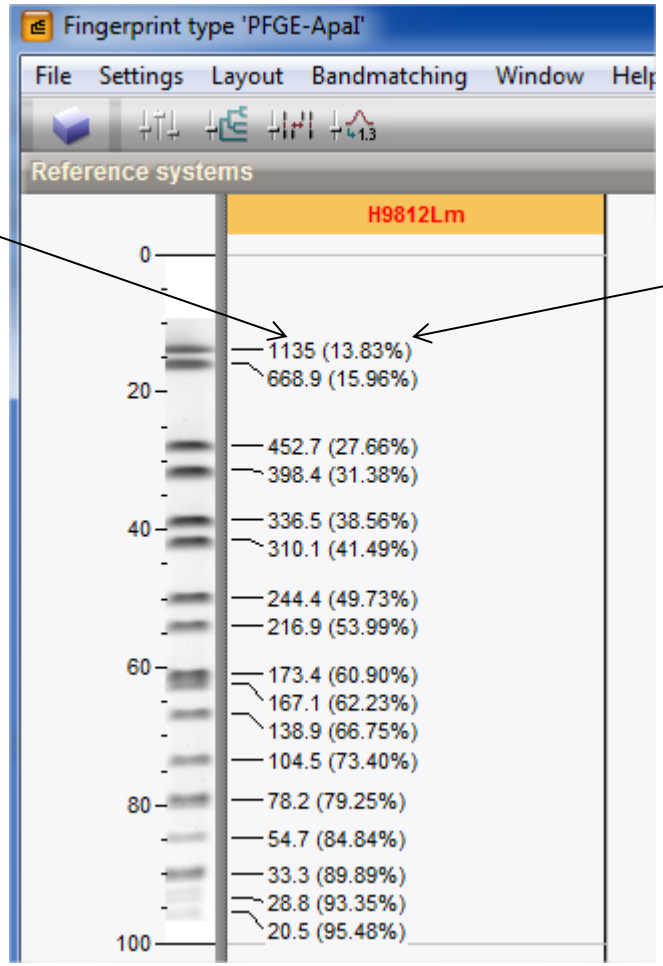
Database main reference system included in the BioNumerics experiment:

- Salmonella Braenderup H9812 *Xba*I restriction profile



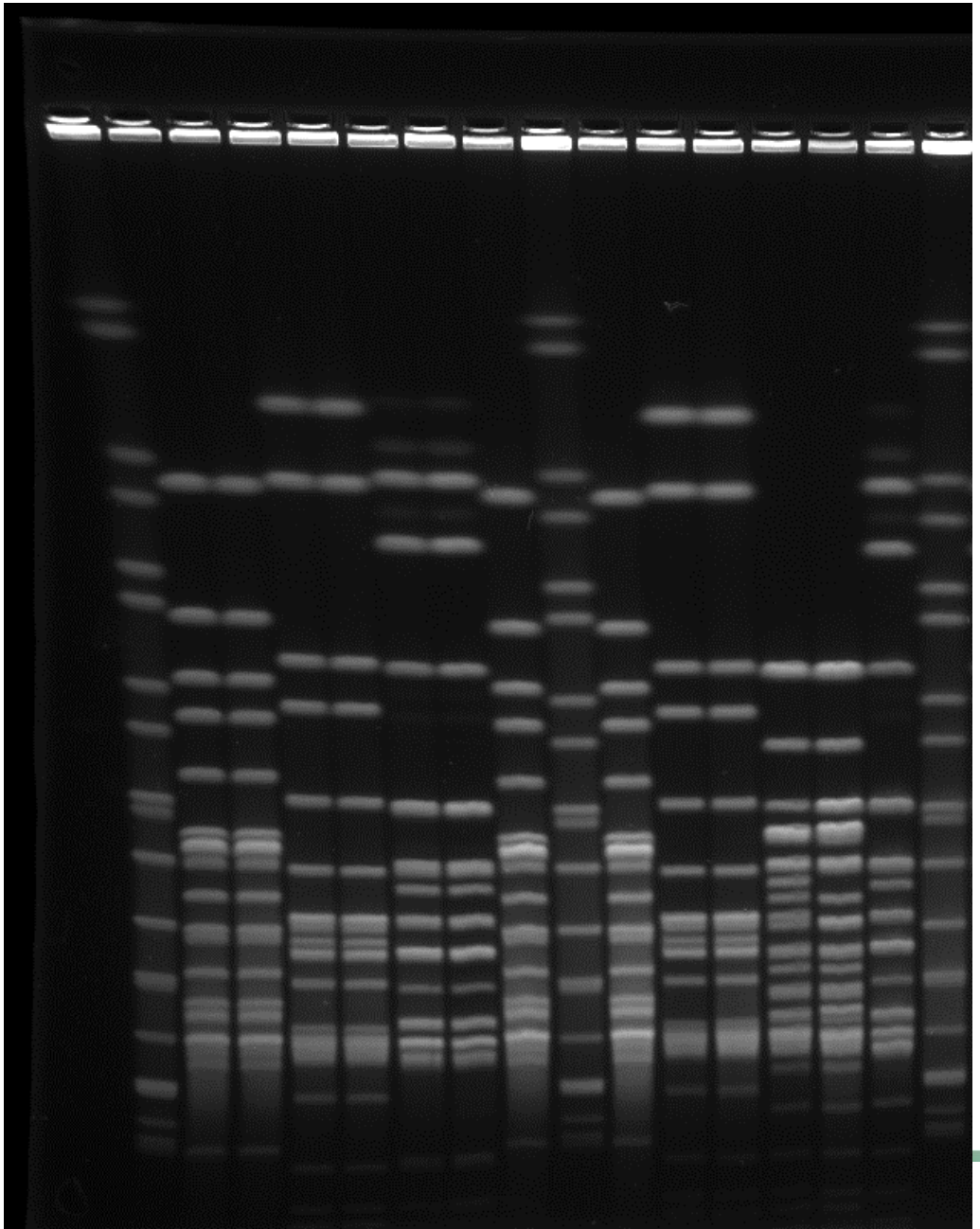
Normalization process

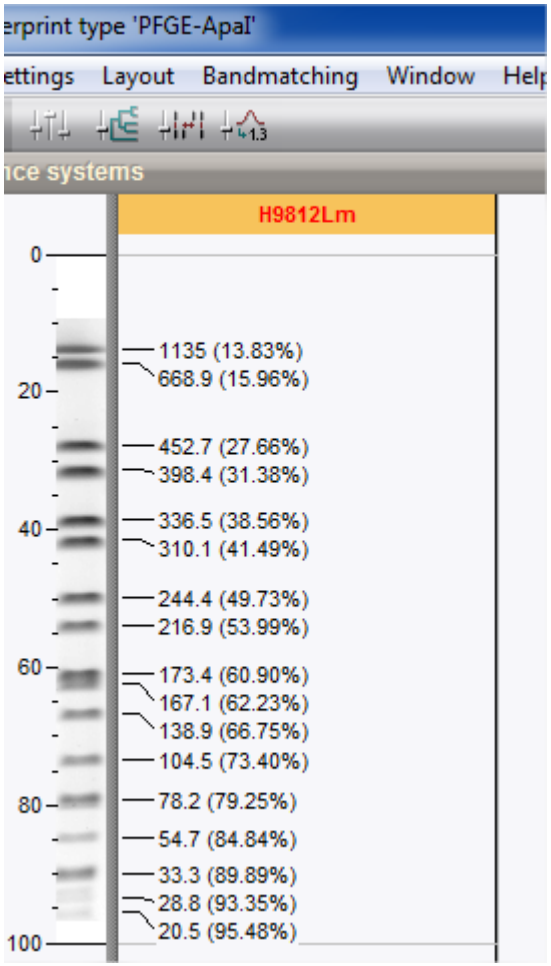
Metric = molecular size kbp



Percentage of migration

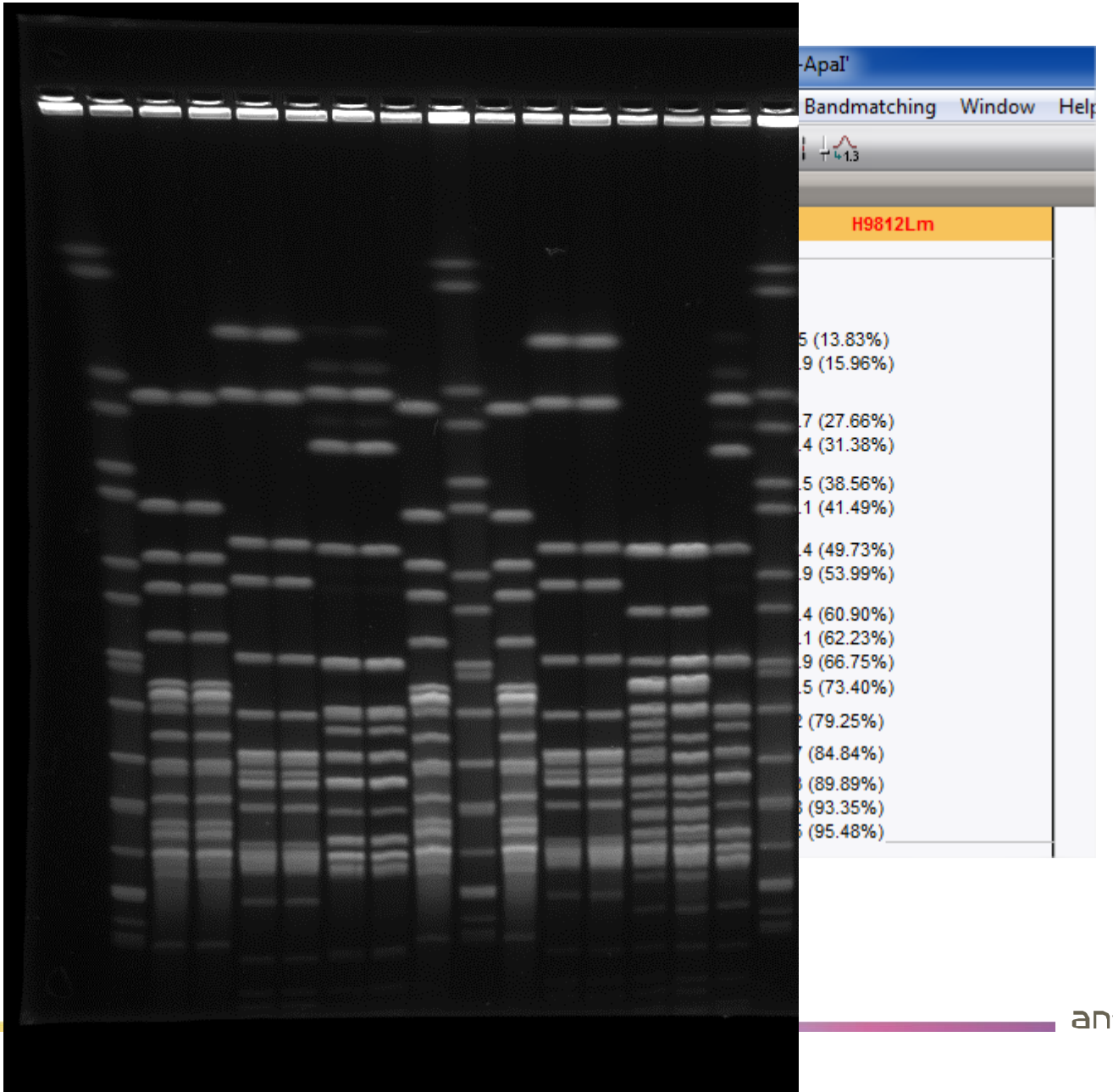
Normalization process





anses

Normalization process



Normalization process

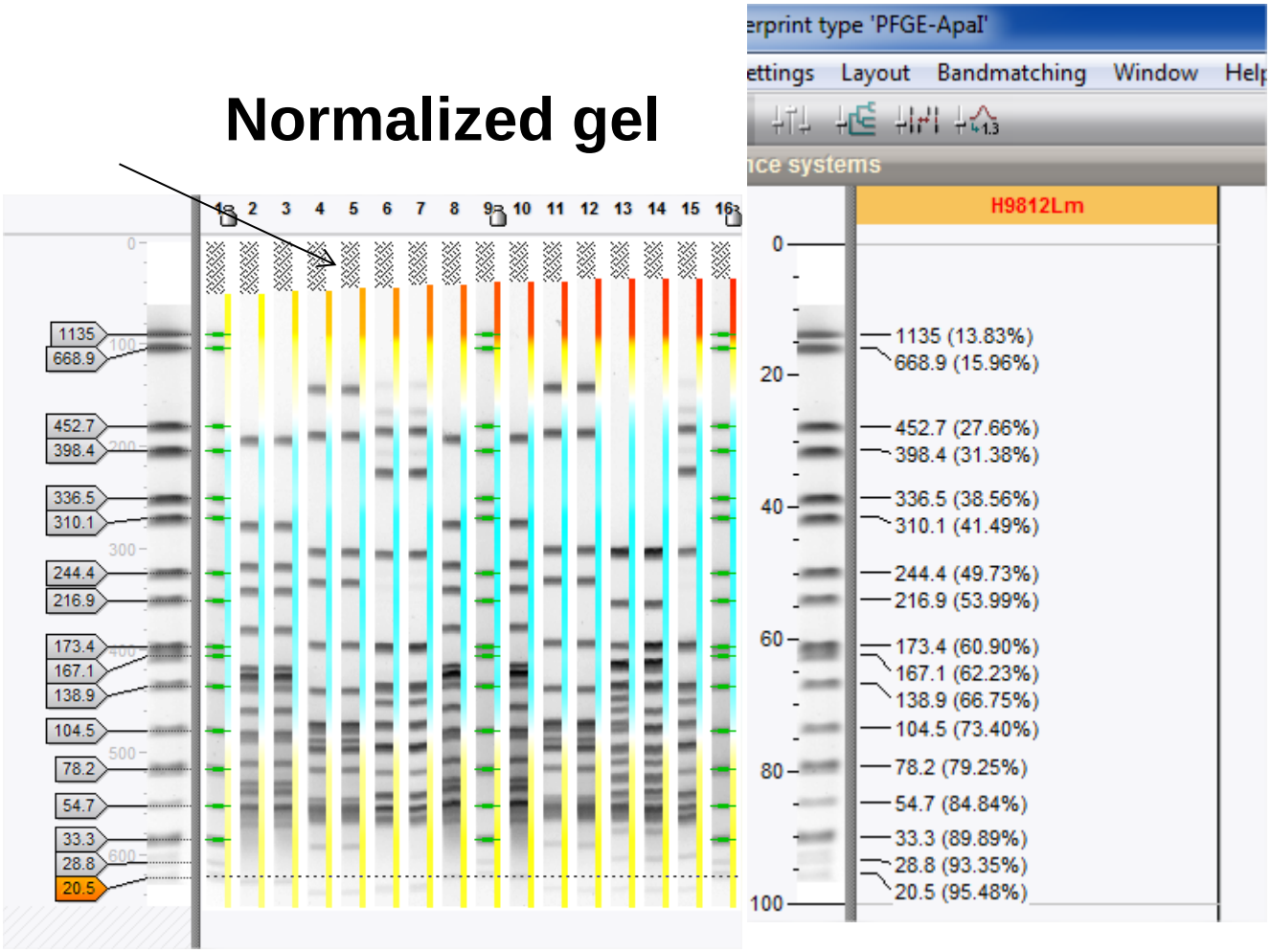
Decale

Normalization process

Distortion

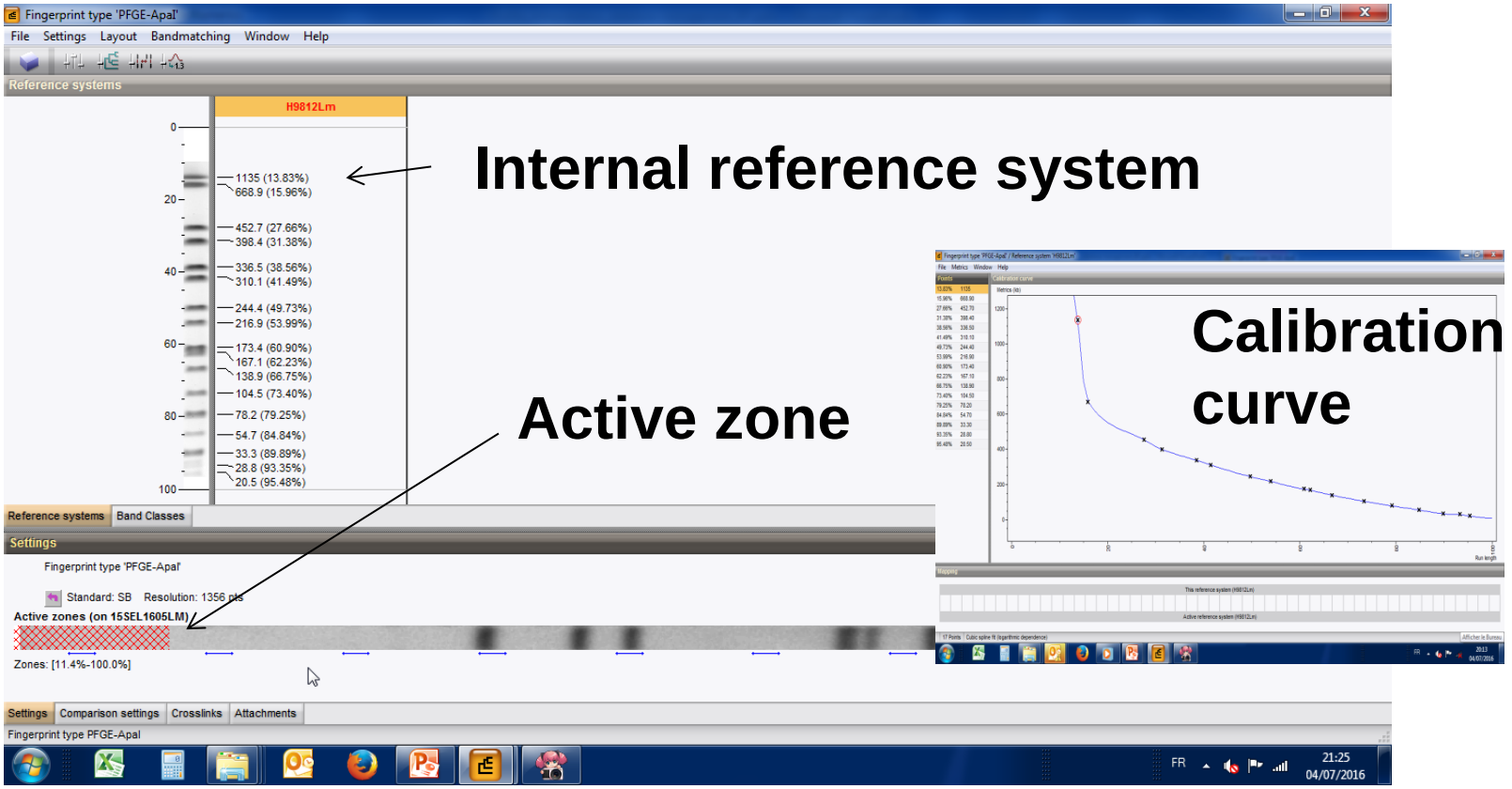
Normalization process

Normalized gel



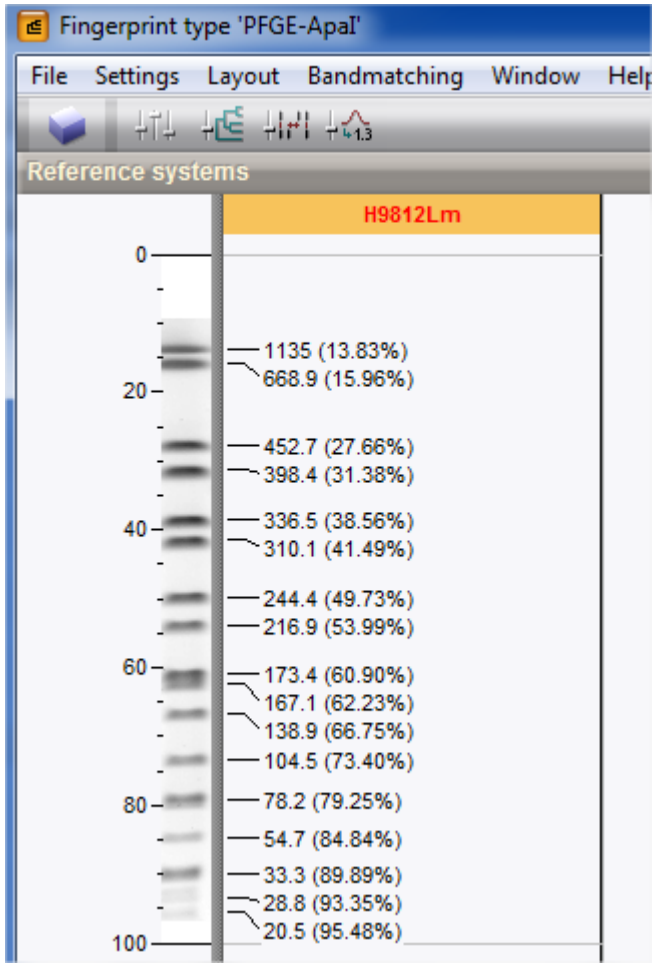
Reference system

Database main reference parameters



How to proceed when the reference systems are different ?

Re-mapping



Several reference systems can work together

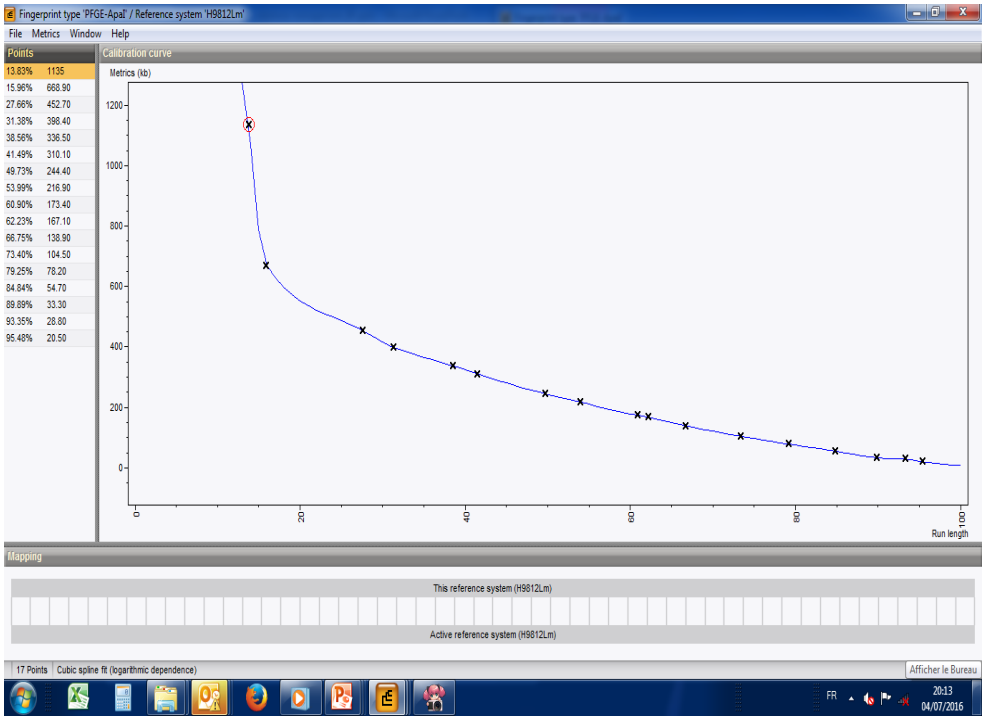
Re-mapping

The re-mapping can be used for bundles and XML import

Bundles need re-mapping to be set beforehand

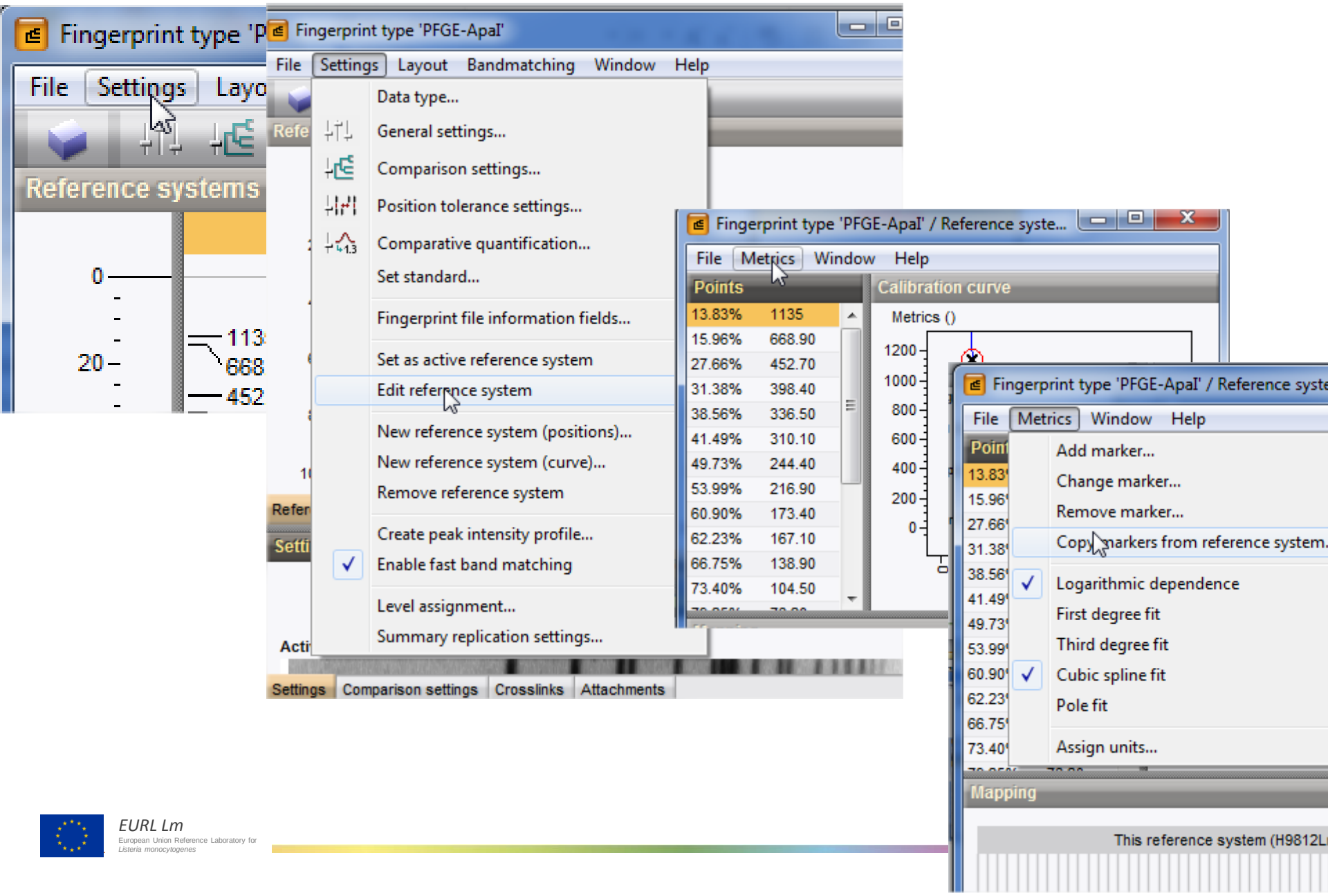
XML import implement automatically parallel reference system

Re-mapping

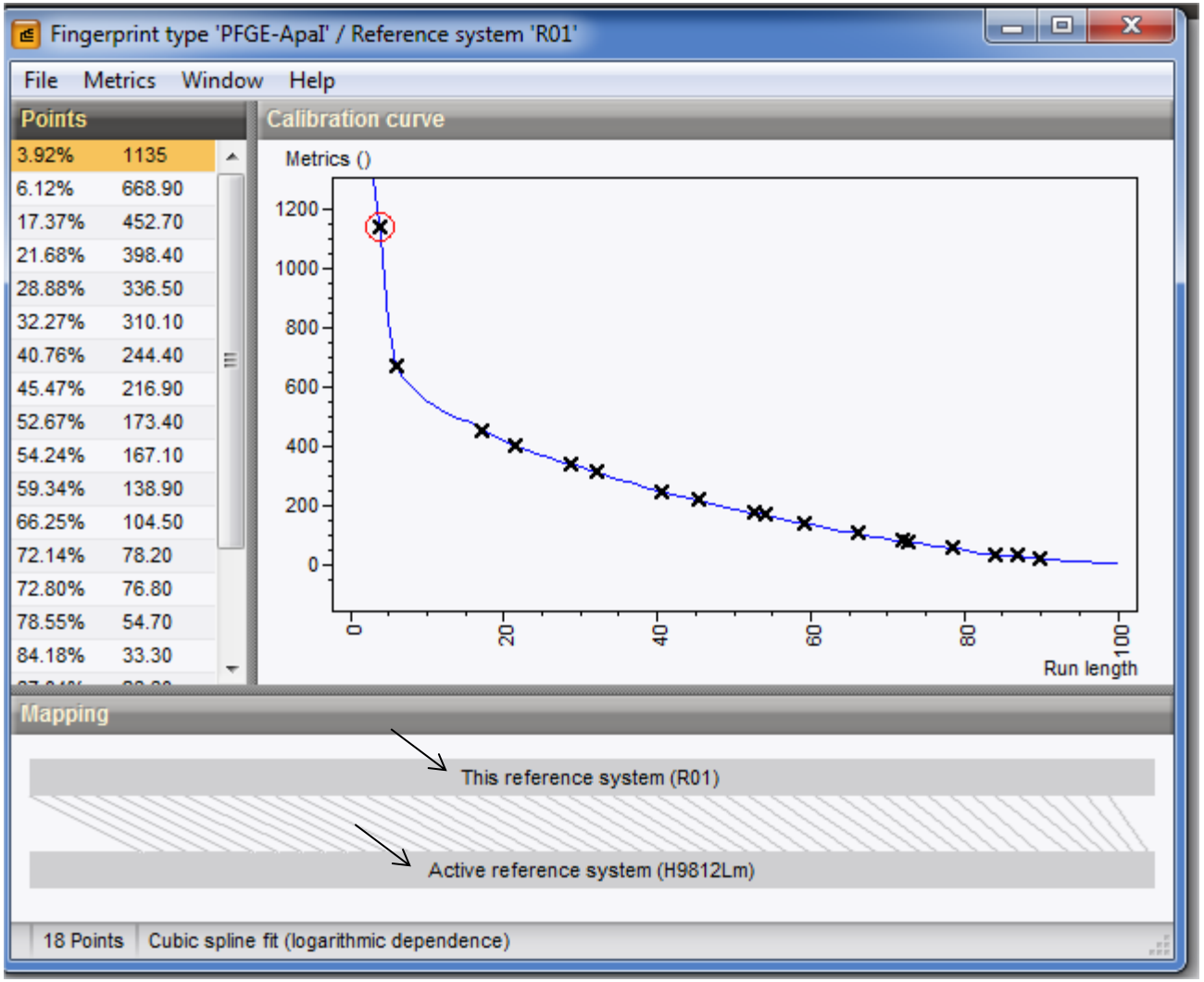


The calibration curve is used to perform the re-mapping process

Re-mapping


Fingerprint type 'PFGE-ApaI'
File
Settings
Layout
Bandmatching
Window
Help
Data type...
General settings...
Comparison settings...
Position tolerance settings...
Comparative quantification...
Set standard...
Fingerprint file information fields...
Set as active reference system
Edit reference system
New reference system (positions)...
New reference system (curve)...
Remove reference system
Create peak intensity profile...
Enable fast band matching
Level assignment...
Summary replication settings...
Settings
Comparison settings
Crosslinks
Attachments
Fingerprint type 'PFGE-ApaI' / Reference syste...
Metrics
Points
Calibration curve
Metrics ()
13.83% 1135
15.96% 668.90
27.66% 452.70
31.38% 398.40
38.56% 336.50
41.49% 310.10
49.73% 244.40
53.99% 216.90
60.90% 173.40
62.23% 167.10
66.75% 138.90
73.40% 104.50
Add marker...
Change marker...
Remove marker...
Logarithmic dependence
First degree fit
Third degree fit
Cubic spline fit
Pole fit
Assign units...
Mapping
Reference systems

Re-mapping is implemented



Re-normalization

- Specific script provided by Applied Maths
- Convert_refsys.BNS
- The script car re-analysed automatically all fingerprint into another internal reference system
- However this script only to convert similar reference system e.g. SB01 into SB02
- The script works by batch of 20 gels
- Beware to backup your system before to proceed
- Beware to verify the output of the re-normalisation