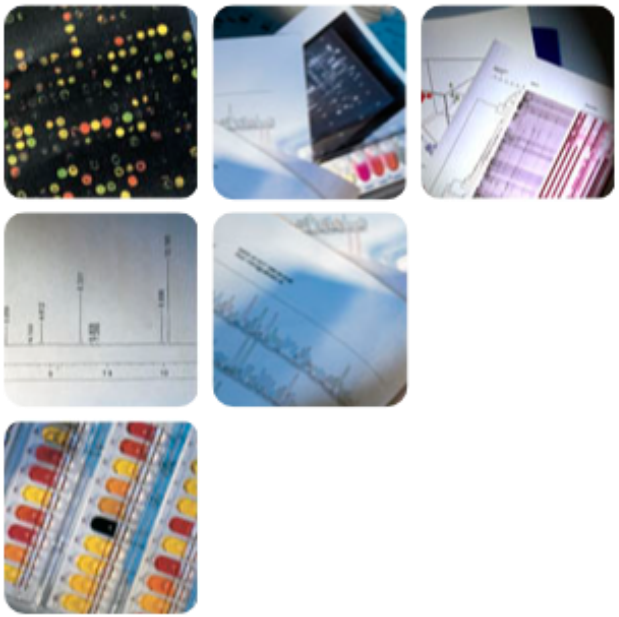
Whole genome sequence analysis using BioNumerics

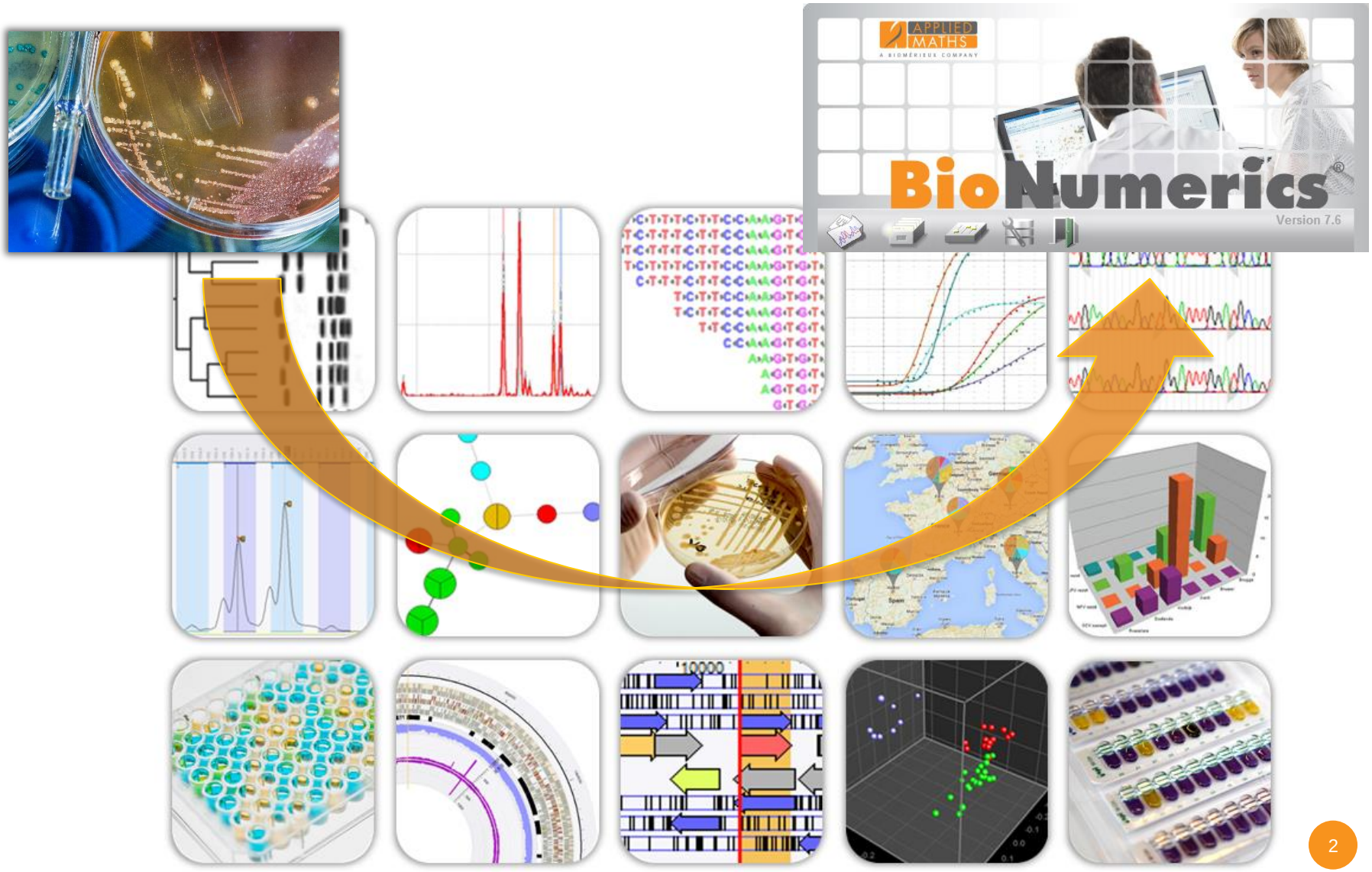


Katrien De Bruyne

APPLIED MATHS

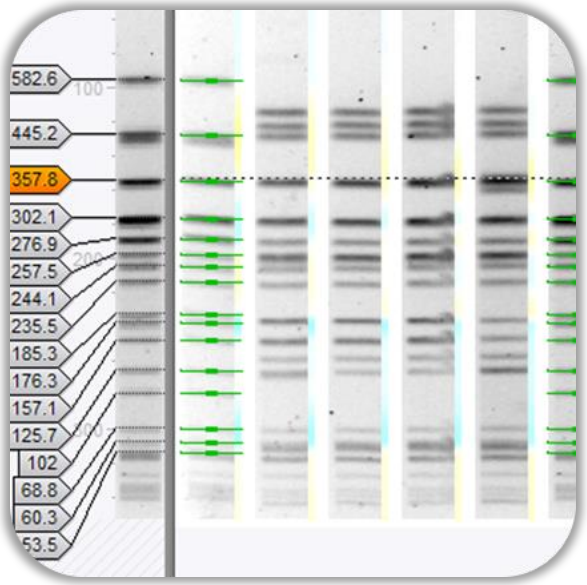
A BIOMÉRIEUX COMPANY

All biological data and metadata integrated in one software platform,

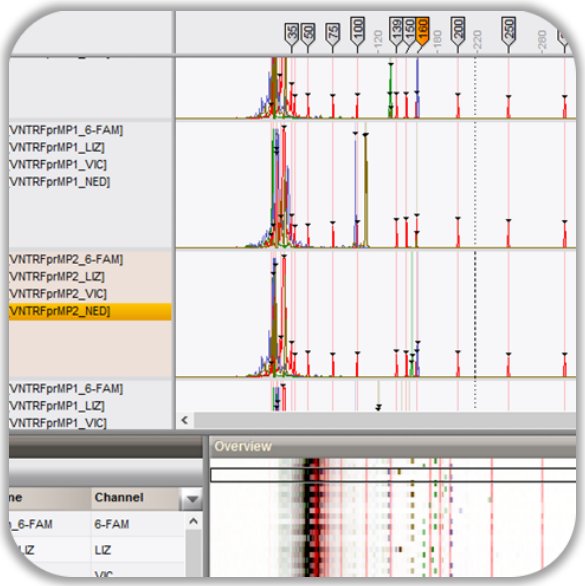


both classical techniques...

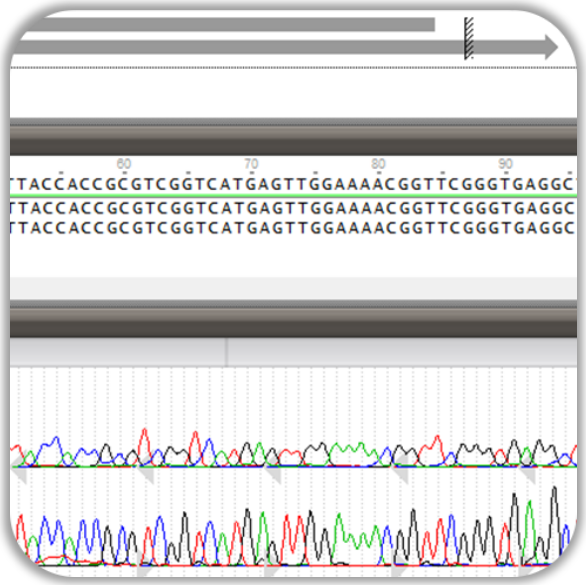
PFGE



MLVA

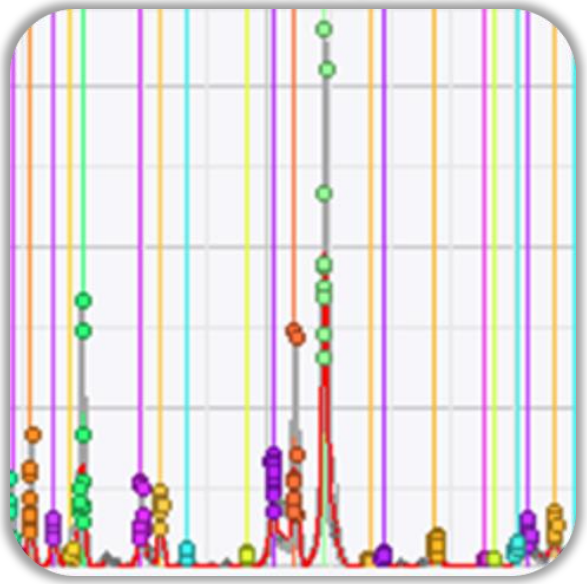


MLST

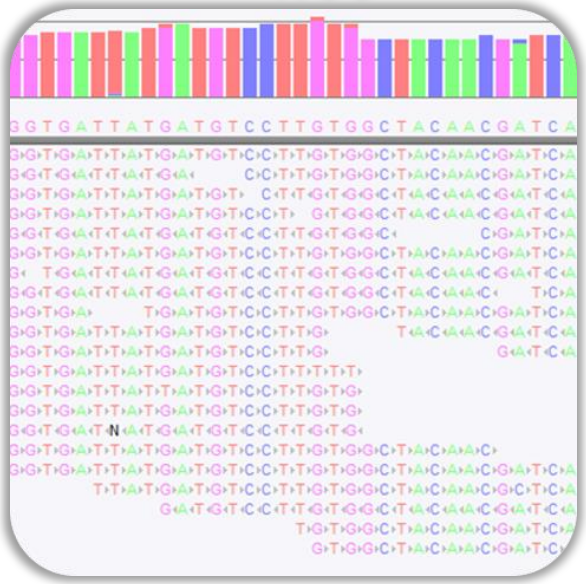


... and new technologies !

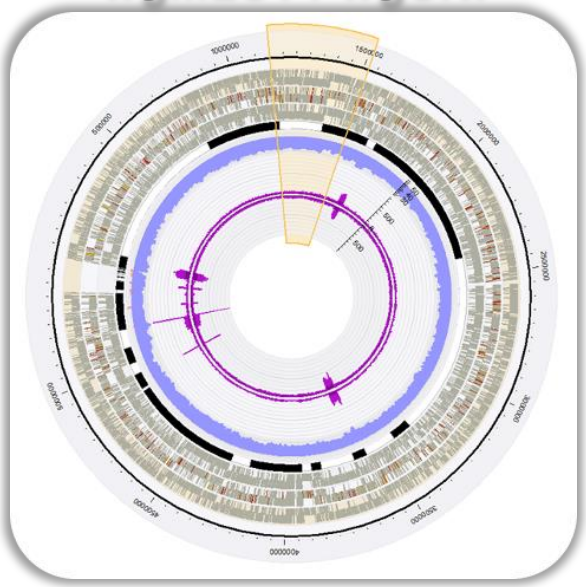
MALDI-TOF MS



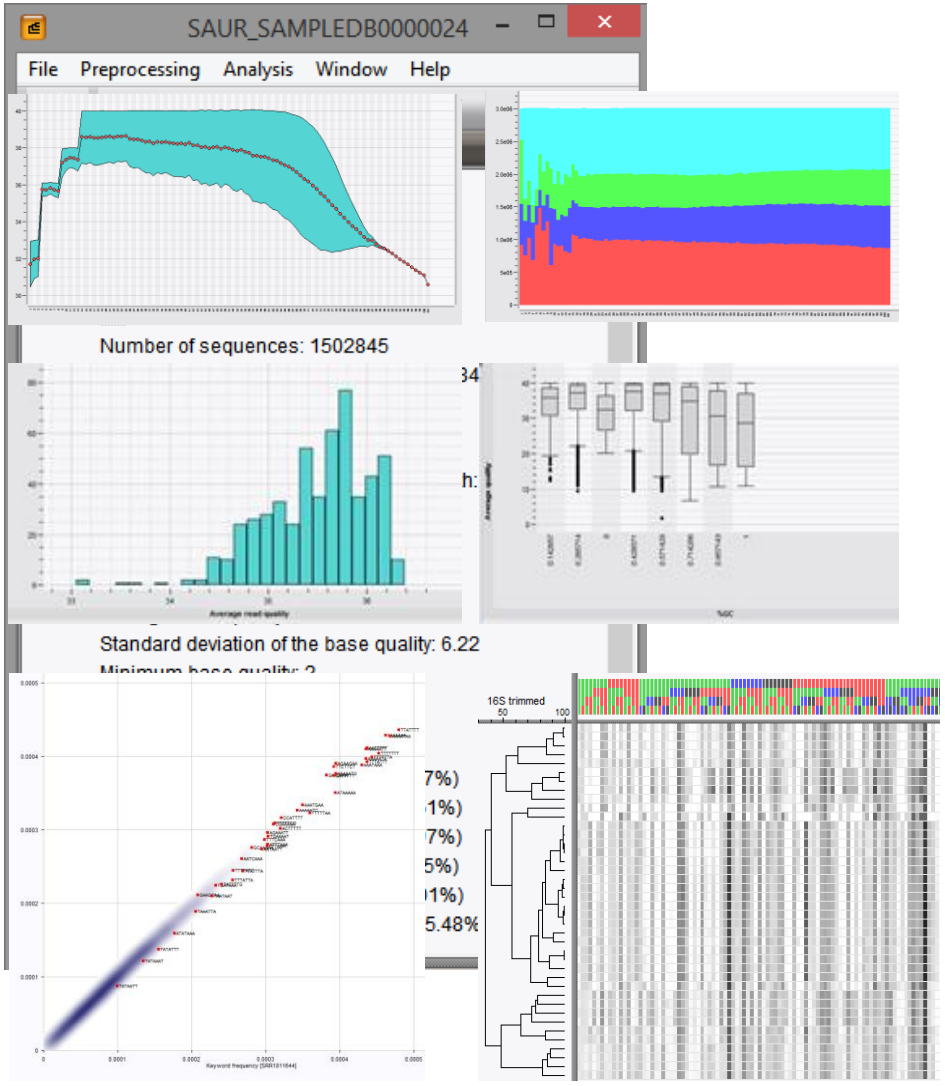
NGS



wgMLST / wgSNP



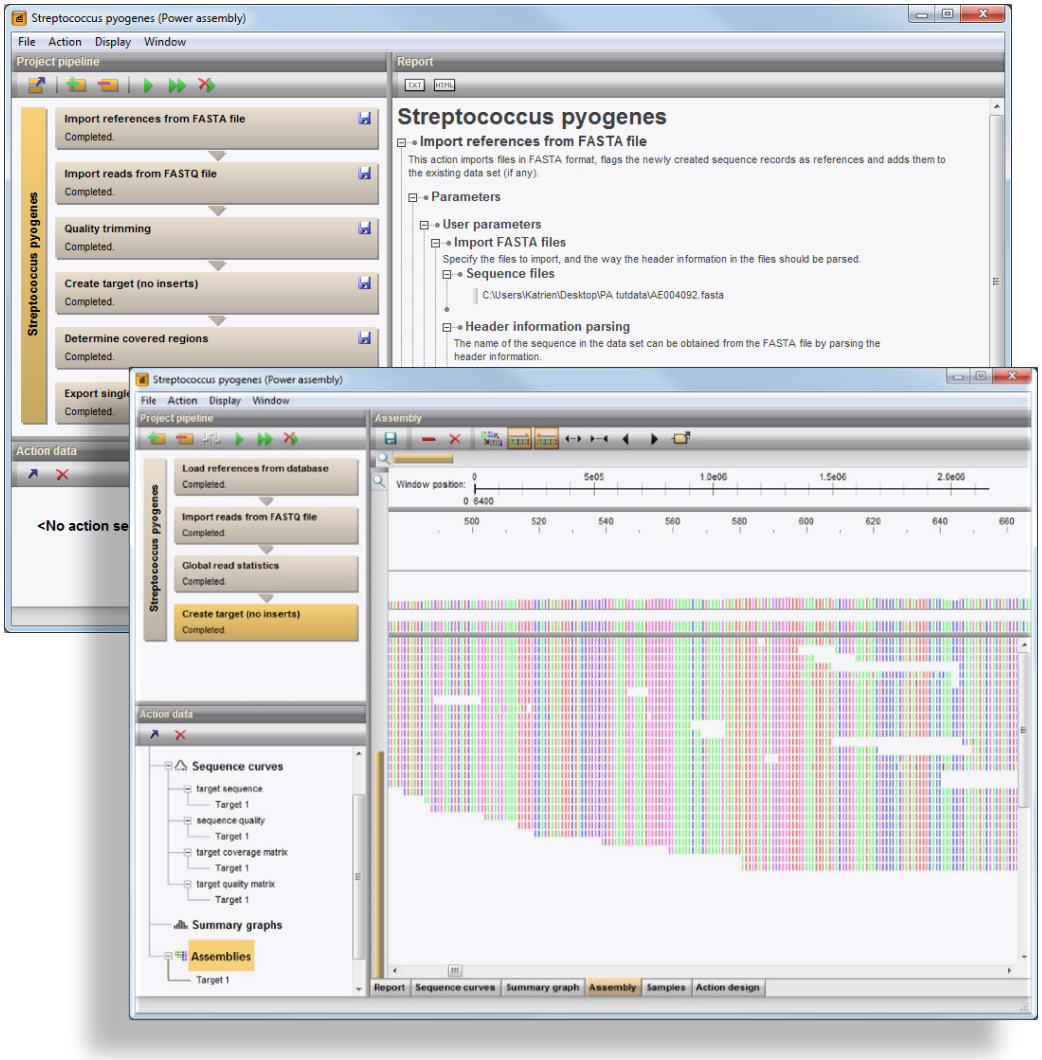
Read quality assessment



- Raw data statistics
- Per-position statistics, per-sequence statistics, per-oligo statistics for data exploration & quality assessment
- Pairwise & multi-sample comparison of keyword profiles:
  - Reproducability
  - Identify contamination
  - Find most suitable reference

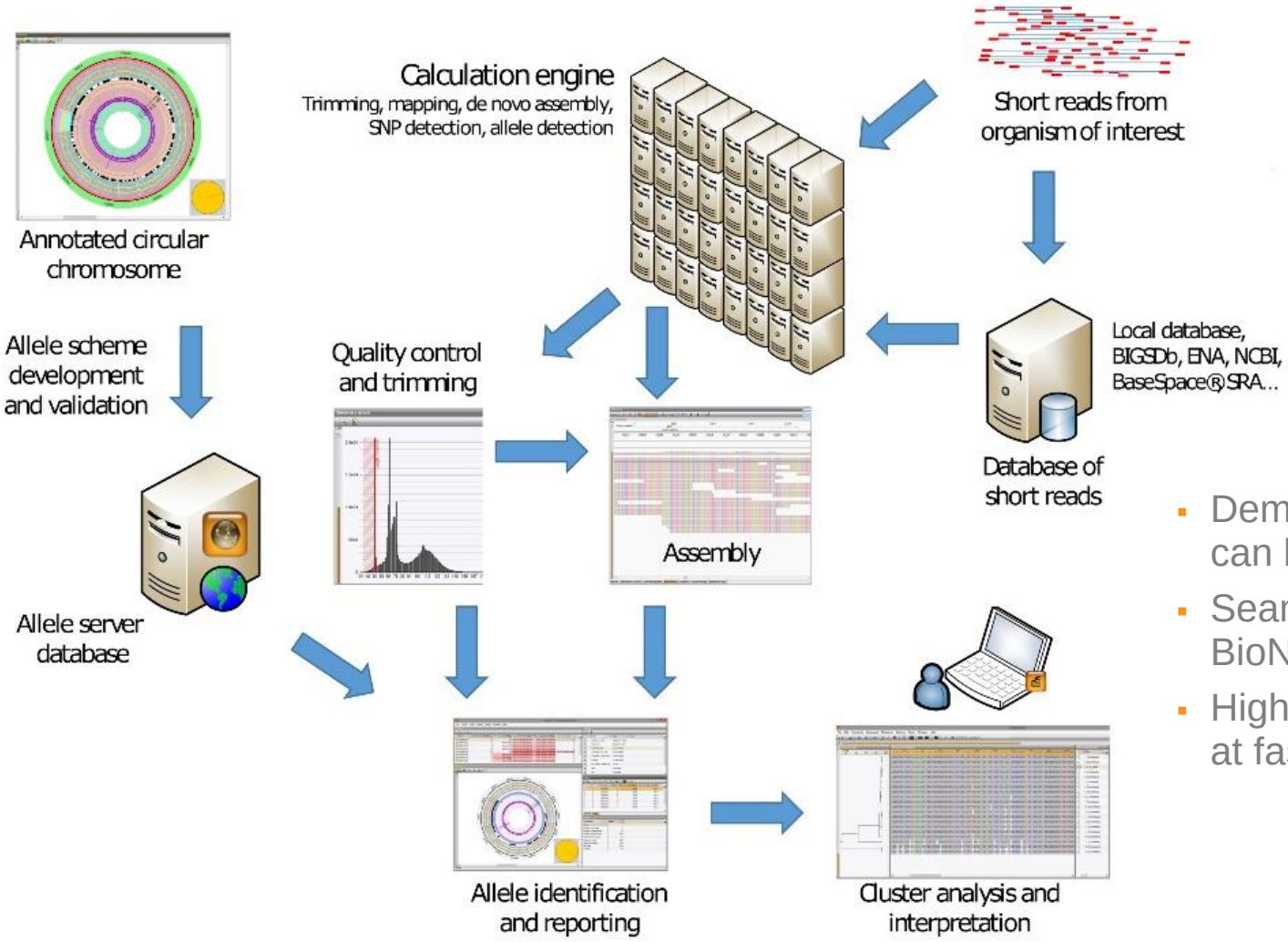
Local WGS tools using the Power Assembler

Point & click application for mapping and assembly of WGS data



- Build new or modify predefined project templates
- Easy access to all actions and parameters
- Save and share analysis templates
- Both mapping and de novo assembly
- Extensive logging and reporting
- Batch analysis from entry selection

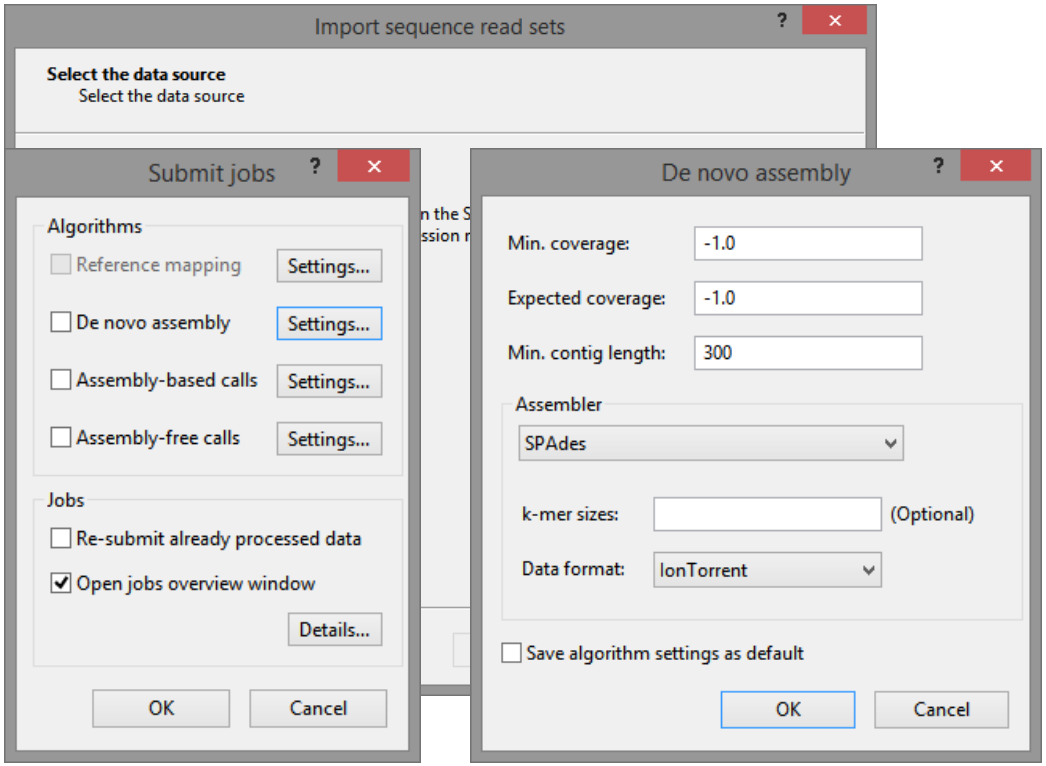
WGS tools integrated on the calculation engine



- Demanding calculations can be performed on CE
- Seamless integration with BioNumerics
- High throughput analyses at fast turnaround times

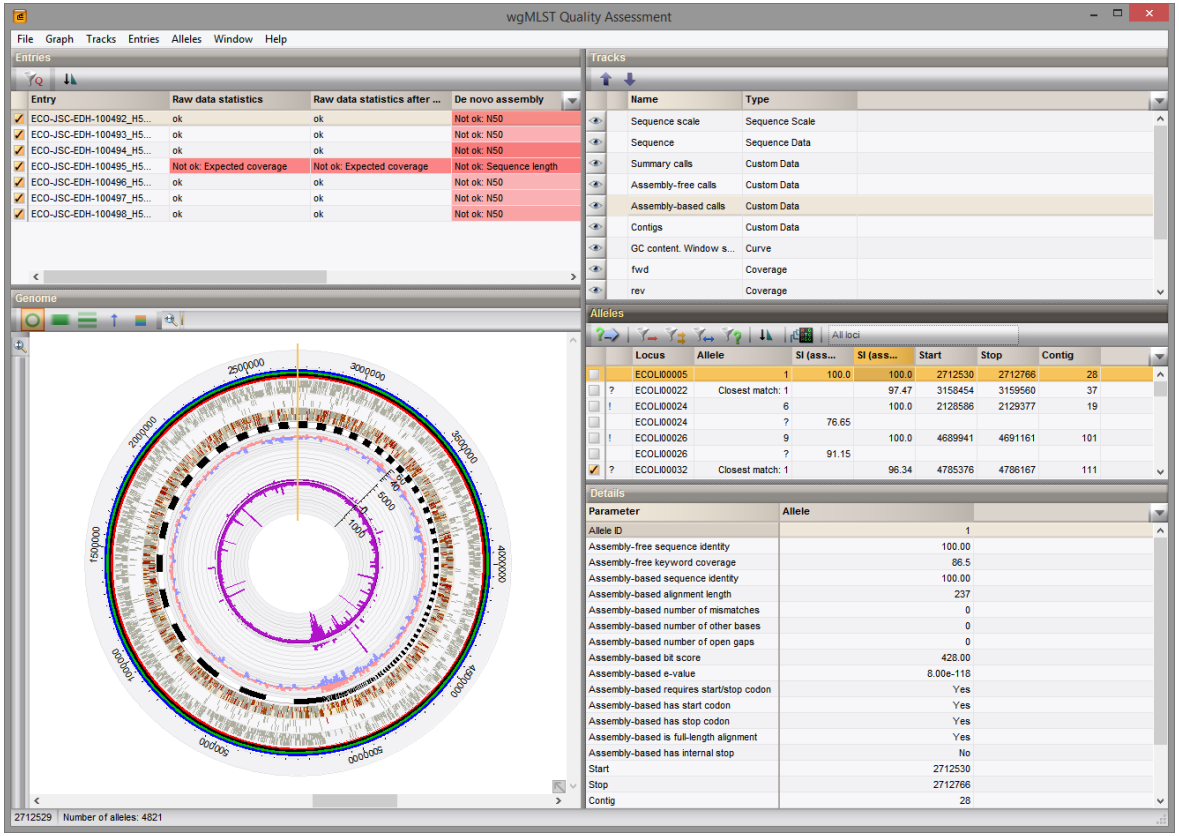
wgMLST for cluster detection

1. Automated import from sequence read sets
2. Batch job processing on the calculation engine
3. Quality assessment of imported results
4. Automated submission of new alleles
5. Automated assignments and sample reporting
6. Calculate population modelling networks



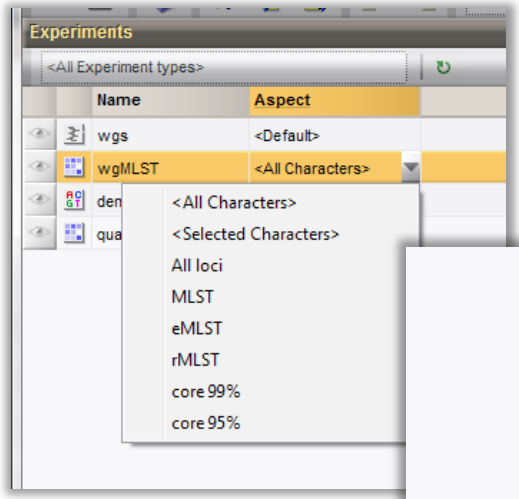
wgMLST for cluster detection

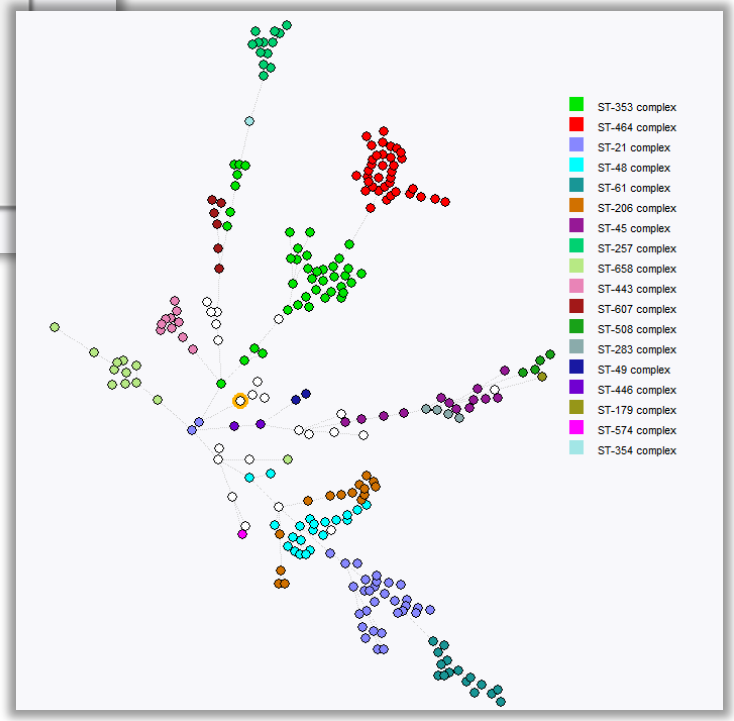
1. Automated import from sequence read sets
2. Batch job processing on the calculation engine
3. Quality assessment of imported results
4. Automated submission of new alleles
5. Automated assignments and sample reporting
6. Calculate population modelling networks



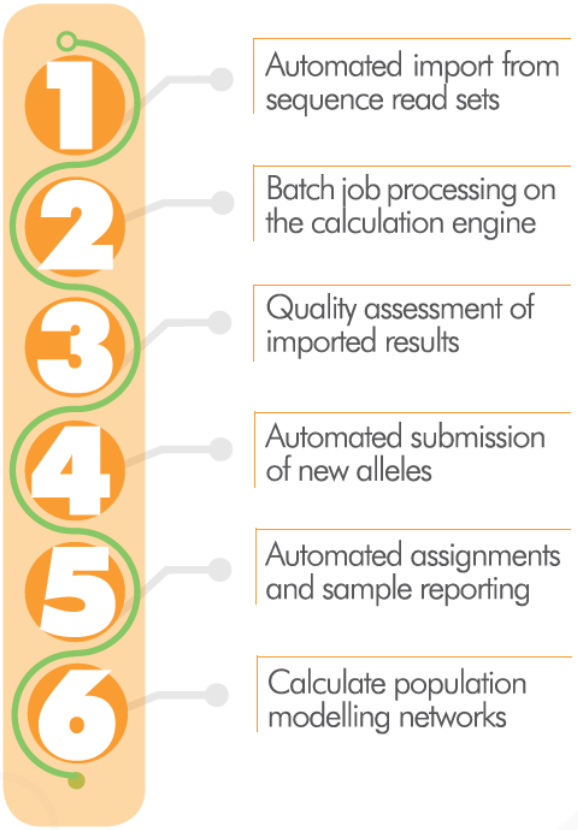
wgMLST for cluster detection

1. Automated import from sequence read sets
2. Batch job processing on the calculation engine
3. Quality assessment of imported results
4. Automated submission of new alleles
5. Automated assignments and sample reporting
6. Calculate population modelling networks





wgMLST for cluster detection



BioNumerics wgMLST schemes

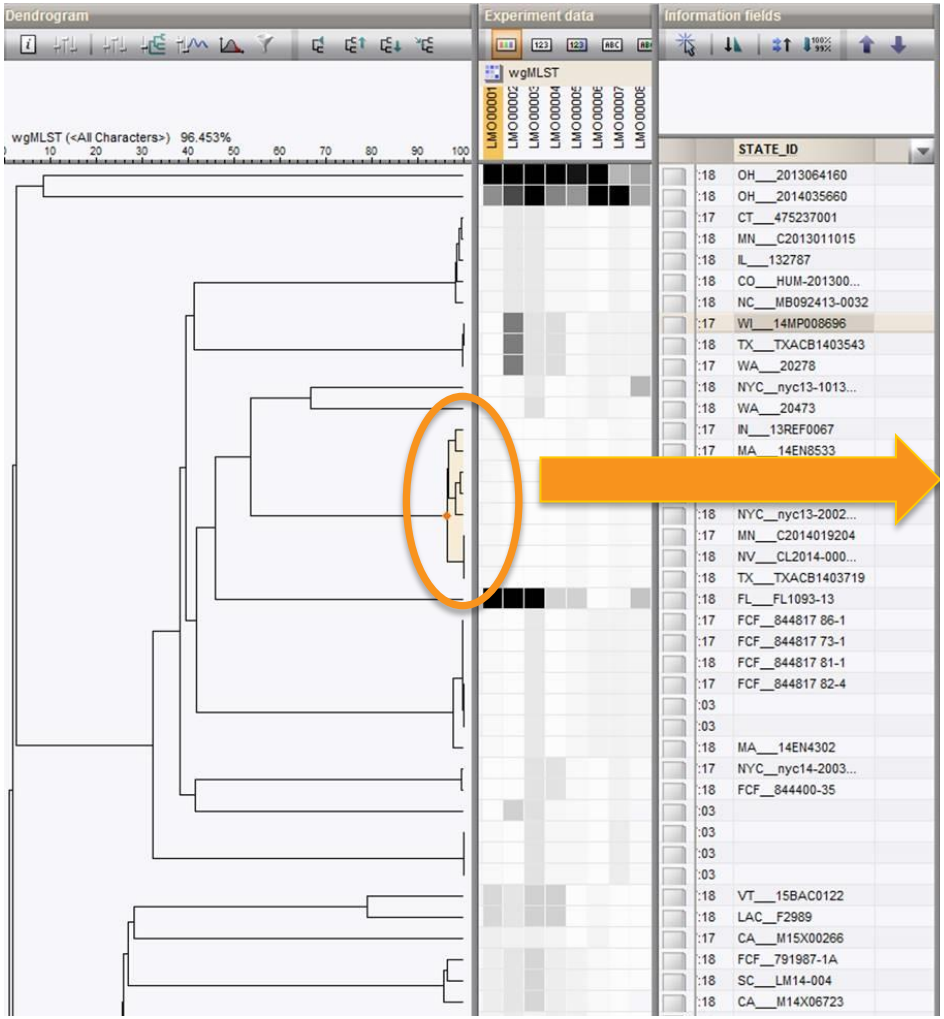
Released

- *Staphylococcus aureus*
- *Listeria monocytogenes*

In preparation:

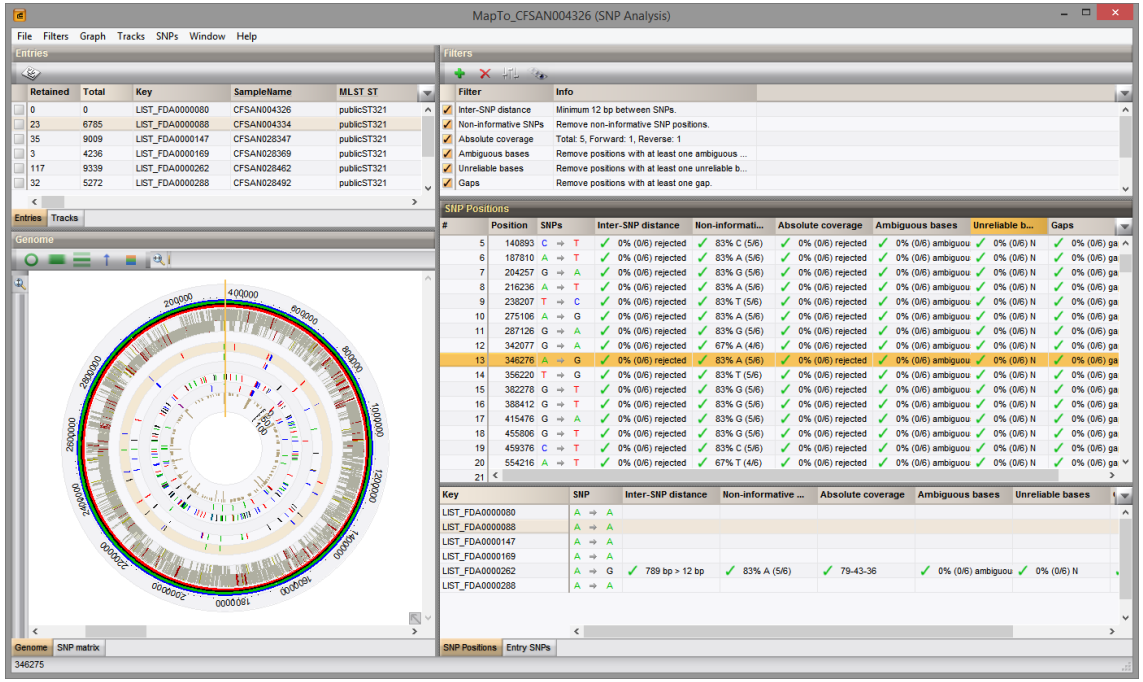
- *Campylobacter*
- *E. coli / Shigella*
- *Salmonella enteritidis*
- *Mycobacterium tuberculosis*
- *Pseudomonas aeruginosa*
- *Legionella pneumophila*
- *Bacillus anthracis/subtilis/cereus*
- *Neisseria gonorrhoeae*
- *Clostridium difficile*
- …

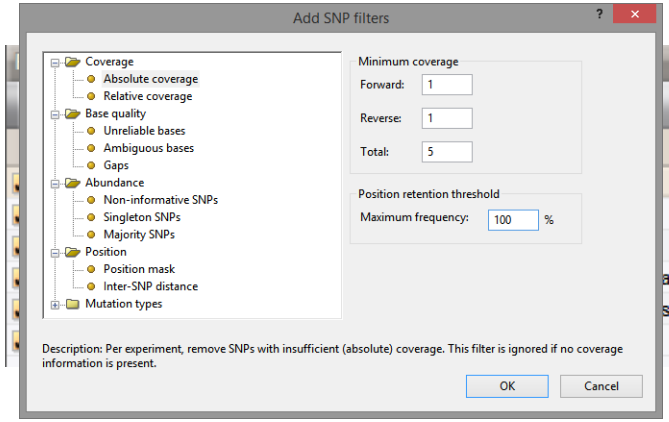
wgMLST for cluster detection: linked to epi data

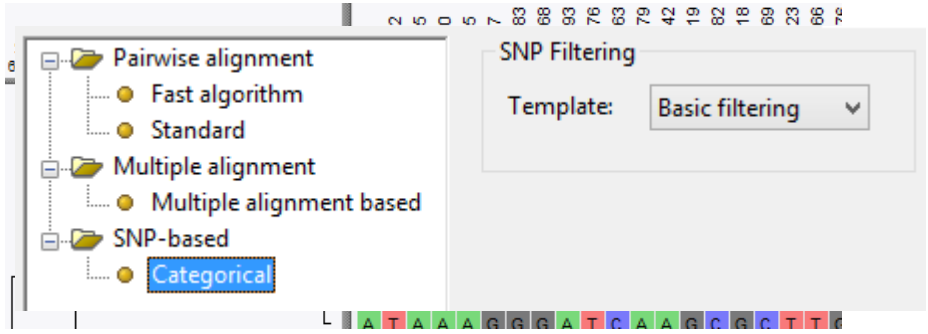


wgSNP for cluster resolution

1. Automated import from sequence read sets
2. Choice of reference sequence
3. Mapping against the reference sequence
4. wgSNP assessment and filtering
5. Calculate population modelling networks

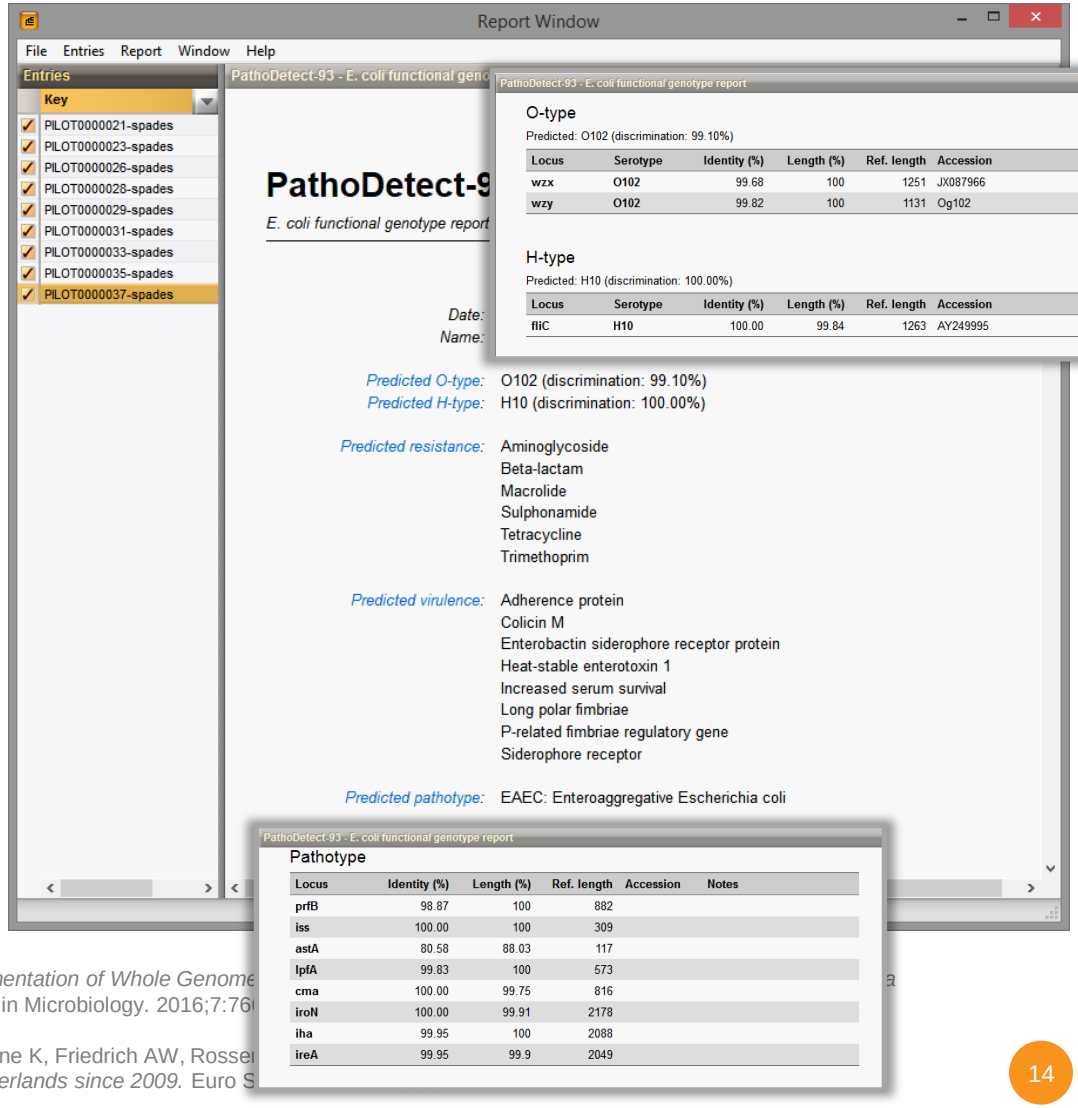






BioNumerics genotyping plugins

- **Epidemiological metadata**
  isolation date, isolation place, linked patient information
- **Typing data**
  MLST, coreMLST, wgMLST typing
- **Pathotype detection**
- **O/H serotyping**
- **Resistance and virulence detection**
- **Plasmids and prophage detection**
- **In-silico PCR**



Lindsey RL, Pouseele H, Chen JC, Strockbine NA, Carleton HA. *Implementation of Whole Genome*
Toxin-Producing Escherichia coli (STEC) in the United States. Frontiers in Microbiology. 2016;7:76

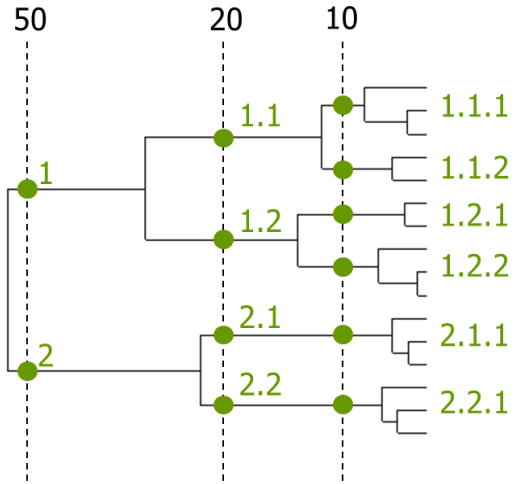
Kluytmans–van den Bergh MF, Huizinga P, Bonten MJ, Bos M, De Bruyne K, Friedrich AW, Rossel
Enterobacteriaceae in retail chicken meat but not in humans in the Netherlands since 2009. Euro S
7917.ES.2016.21.9.30149

Strain nomenclature principles

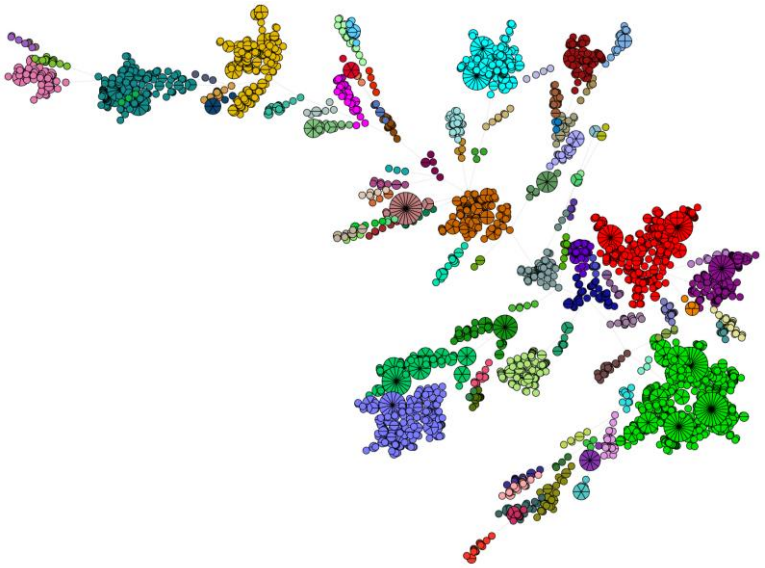
- **The ST idea, i.e. 100% identity rule won't work**

  Almost every strain has some unique mutations, therefore would have a unique ST

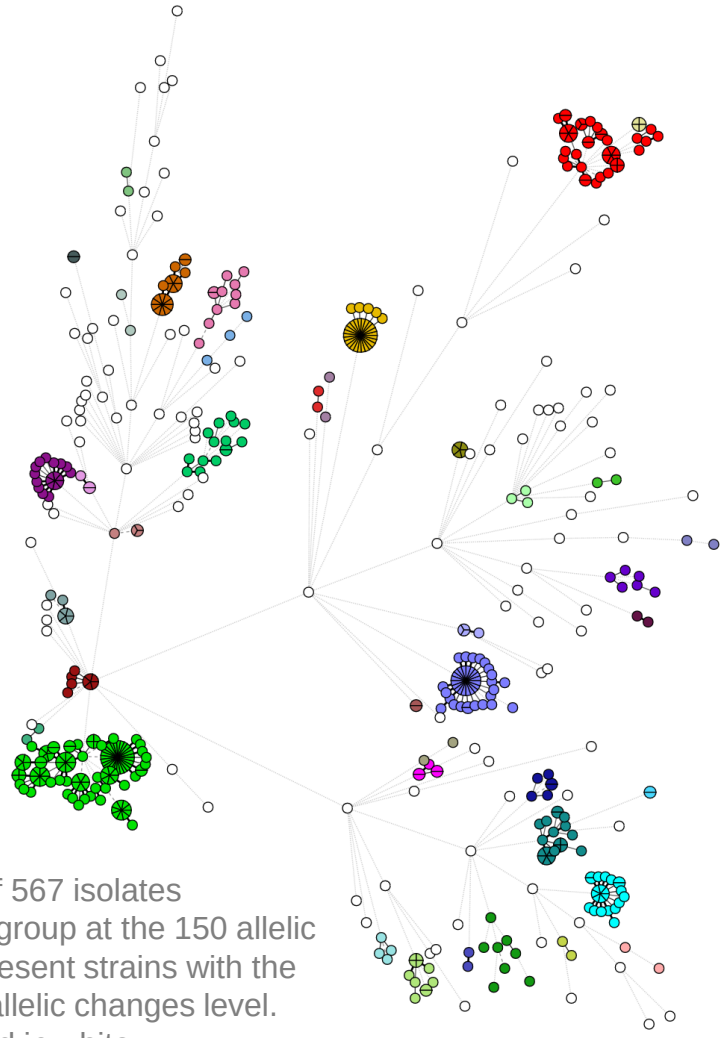
- **Hierarchical strain nomenclature**
  - Reflects the distance between strains
  - Based on single linkage clustering at different distance cutoffs (e.g. using cgMLST)
  - The cutoff for each level has been optimized and validated to minimize the error rate of assignments



Strain nomenclature applied on Listeria



Minimum spanning tree of 3652 isolates
Colors represent strains with the same name at the 150 allelic changes level.
Sporadic groups (<3 isolates) have been removed



Minimum spanning tree of 567 isolates
All isolates in the biggest group at the 150 allelic changes level.Colors represent strains with the same name up to the 11 allelic changes level.
Sporadic isolates indicated in white.

A BIOMÉRIEUX COMPANY

BioNumerics GET A FREE TRIAL

Katrien_DeBruyne@applied-maths.com