BioNumerics for PulseNet
Today and tomorrow
Hannes Pouseele
Applied Maths
Introduction

Goal
BioNumerics functionality for PulseNet use and beyond

Disclaimer

- Why you should have BioNumerics 6.6
  - Band editing in the comparison
  - Charts and statistics
  - Next-generation sequencing

- Why you will love BioNumerics 7.x
  - Faster and better
  - MLVA, MALDI, Whole Genome Maps, Genome Sequence Scanning
  - Classifiers
  - More next-generation sequencing
Introduction

Disclaimer

For the last 7 years, I was affiliated with the following organization:

- **Organization:**
  Applied Maths NV, Keistraat 120, B-9830 Sint-Martens-Latem, Belgium

- **Relationship:**
  employee
Band editing in the comparison

- Bands can be edited in the comparison
  - Click on a gel position to select it
  - Delete/enter will remove/add the band

- Comparison serves as a playground to assess the changes
- Changes can be stored when storing the comparison (or via the dedicated menu item)
BioNumerics 6.6
Charts and statistics

- Many possibilities to better understand your data
- Unfortunately, rather complicated
  - Custom PulseNet functionality for epi charts
Next-generation sequencing

Power assembler: a user-friendly tool for the assembly of whole genome sequencing data
- Reference mapping (BWA-AM) with small memory footprint
- De novo assembly (Velvet)

Inspection and quality assessment of the assembled results
- Detailed view of the alignment of each read
- Overview tools for global picture and fast navigation
Follow-up tools

- Reference-based SNP detection
- SNP-based sample clustering
- Alignment of de novo assembled genomes
- Chromosome comparison and SNP detection
BioNumerics 7.x

Faster and better

- More powerful database interaction
  - Flexible choice in loading only part of the database
  - Tools to manage large numbers of comparisons, experiment types, ...

- Many limitations have been removed
  Number of loaded entries, fingerprint files, reference bands, band classes, number of comparison groups

- However, this comes at a cost: local databases no longer supported
  - Very robust conversion tool available
  - Working on a general strategy to accommodate for this within PulseNet
  - If you need help, contact us!
Sequencer fingerprints (MLVA)

- Easy import of raw automated sequencer curves.
- Fast and flexible peak searching
- Noise removal, bleed through detection, stutter bands elimination, etc.
- Normalization algorithm with built-in predefined reference patterns for reliable and fast normalization.
BioNumerics 7.x

MALDI

- Easy import of spectrum data from various formats
- Customizable workflow templates for easy preprocessing
- Baseline subtraction, noise elimination and curve smoothing.
- Automatic peak calling with manual editing options.
- Creation of summary spectra based on peak matching and/or member averaging.
  - Filter out spectra of low quality.
  - Similarity values allow easy inspection of the coherence.
Whole genome maps (OpGen)

- High resolution, ordered whole genome restriction maps
- Analysis focused on strain typing and characterization.

- Easy XML import of whole genome map data
- Pairwise alignment indicates concordance and large-scale rearrangements between samples

- Map-based clustering and global alignment allows to distinguish highly related strains using new and fast tolerance- and pattern-based algorithms.
GSS (Pathogenetix)

- High resolution restriction fragments (length and fluorescence pattern)
- Analysis focused on strain typing and characterization.

- Middle ground between PFGE and WGS:
  - GSS groups correlate well with PFGE groups and very well with WGS groups
  - GSS subgroups correspond well with WGS subgroups
Classifiers are used to identify an unknown sample
- Example: rank by similarity, naive Bayesian classifier, support vector machines, ...
- Extremely useful but often fragile procedures
- Validation framework available

Usage:
- MALDI-based genus/species identification
- rMLST-based genus/species identification
- WGS-based serotyping (cfr Luminex SNP arrays)

Example: classifying *Lmo* lineages
More next-generation sequencing

- NGS data is now an experiment type of its own
  - Possibility to link to externally stored data (storage volume, NCBI, ...)

- Quality assessment of the raw data

- Faster alignment algorithms

- State of the art SNP analysis

- Whole genome MLST analysis

- Targeted metagenomics (16S, ...)
BioNumerics 7.x

WGS data processing model

Calculation engine
Trimming, mapping, de novo assembly, SNP detection, allele detection

BioNumerics client

Isolate database

External storage
NCBI, ENA, BaseSpace
**wgMLST pipeline**

BioNumerics 7.x

**BIGSDb**
- Public nomenclature

**Nomenclature server**
- Allele databases

**Isolate database**

**BioNumerics client**

**External storage**
- NCBI, ENA, BaseSpace

**Calculation engine**
- Trimming, mapping, de novo assembly, SNP detection, allele detection

**SQL databases**