BIONUMERICS 8.1 release notes

Applied Maths is proud to announce BIONUMERICS version 8.1. This update features an easy export of trees and metadata to iTOL, minhashing-based genome comparison, a comeback of the Geo plugin, the new and versatile Custom genotyping plugin, the new Acquired resistance plugin, genotyping enhancements, an improved fast character matching and much more!

EASY DENDROGRAM ANNOTATION WITH ITOL

The Interactive Tree Of Life (iTOL) hosted by EMBL is a popular online tool for tree visualization, annotation and export to publication grade figures.

With the new iTOL integration plugin, you can upload a dendrogram and associated meta data directly from the BIONUMERICS *Comparison window* to a project within your iTOL account with a single click. The iTOL website opens in the default web browser, where the exported tree can be visualized and annotated just like any other tree in iTOL.

Any data you choose to display in the *Comparison window* will be exported to iTOL. This includes the active dendrogram, entry information, character aspects of any experiment data, comparison groups and similarity values.

MINHASING-BASED SEQUENCE CLUSTERING

Genome sequences (e.g. de novo assemblies) can now quickly and easily be compared without the need for sequence alignment or any additional typing method. This is achieved through an additional Minhashing based clustering option in the *Comparison window*.

Our tests indicate that the dendrograms obtained through minhashing show a high correlation with wgMLST and wgSNP dendrograms from the same sequences.

RETURN OF THE GEO PLUGIN

Back by popular demand, the Geographical mapping plugin is available again and works with the embedded Chromium web browser, introduced in BIONUMERICS version 8.0. Through a simple wizard, this plugin lets you to plot locations, bar graphs, pie charts and dendrograms on an interactive geographical map. Locations can be addresses or coordinates (i.e., longitude and latitude), stored as entry information in the BIONUMERICS database. The extra information visualized in bar graphs and pie charts is collected from either entry information fields or character values.
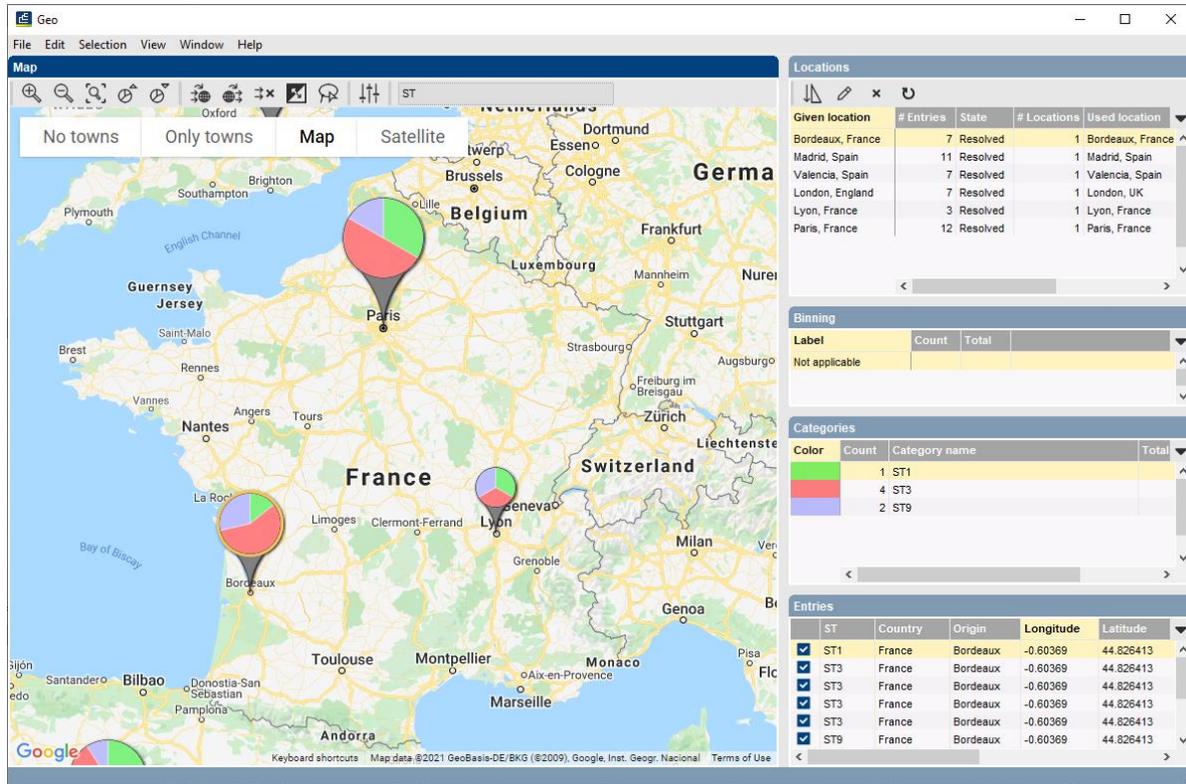
CUSTOM GENOTYPING PLUGIN

The Custom genotyping plugin is a new powerful and extremely versatile tool to detect and extract sequences from whole genome sequences using BLAST or in silico PCR approaches.

The plugin allows you to define your own custom genotyping models, consisting of one or more features with their settings and custom knowledge base. Since knowledge bases for this plugin are user defined, a genotyping model can be tweaked to detect and/or extract virtually any trait or sequence present in a genome. Common applications include the detection of acquired traits such as antibiotics resistance and virulence factors, extracting genes of interest from genomes for further analysis and extracting sets of genes or gene fragments for typing purposes (e.g. MLST).
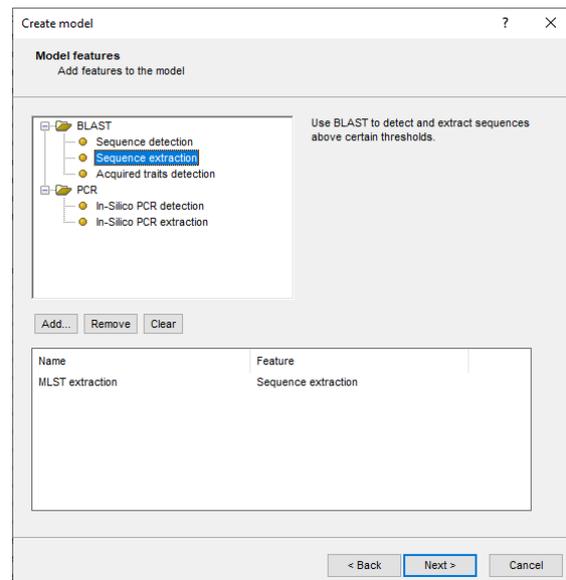
Knowledge bases for this plugin consist of a small set of structured text files, adhering to existing formats (JSON, FASTA and TSV). The plugin offers a convenient knowledge base validation tool and can export example knowledge bases, including an extensive description of the expected format. To

facilitate MLST allele extraction, the MLST online plugin offers an export of allele sequences in the knowledge base format required by the Custom genotyping plugin.

As with other genotyping plugins, the obtained results are presented in clear reports and stored in user defined experiment types.

ACQUIRED RESISTANCE PLUGIN

The new Acquired resistance plugin is a generic (i.e., not specific to any organism) plugin for the detection of acquired antibiotic resistance. This plugin provides a solution for customers aiming to screen Gram-negative organisms for which no specific genotyping plugin is available.

The online knowledge base used by the plugin contains a broad selection of resistance markers and can easily be updated.

IMPROVED ORGANISM-SPECIFIC GENOTYPING PLUGINS

The *Escherichia coli*, *Listeria*, *Salmonella* and *Staphylococcus aureus* functional genotyping plugins all benefit from updates in the genotyping framework:

- The improved knowledge base management makes selecting and updating online knowledge bases more intuitive and prevents the user from accidentally selecting an incompatible knowledge base.
- The redesigned genotyping reports feature hyperlinks to sequence accession numbers and PubMed identifiers (which open in the web browser) and sequence positions (which open in the BIONUMERICS Sequence viewer).
- Sequence annotation of the original sequence experiment based on genotyping results.
- The NCBI toolkit that comes with BIONUMERICS was updated to version 2.11.0 to avoid any issues with spaces in database names while running BLAST.

E. COLI FUNCTIONAL GENOTYPING PLUGIN

Extended spectrum beta-lactamase (ESBL) and carbapenemase-producing *Enterobacteriaceae* (CPE) resistance typing was added to the already rich set of features of the *Escherichia coli* functional genotyping plugin.

The detection of full plasmids is now achieved through minhashing. Not only is this method fast and specific, it also allows us to use a much smaller yet more comprehensive online knowledge base.

BIONUMERICS 8.1 release notes

## *SALMONELLA* FUNCTIONAL GENOTYPING PLUGIN

Serotype prediction in the *Salmonella* functional genotyping plugin is now accomplished with the open source tool SeqSero2 (previously SeqSero was used). In addition to the improved accuracy that the new tool offers, the new implementation does not require a job to be submitted to the calculation engine, since the analysis is done locally on the de novo sequence.

IMPROVED FAST CHARACTER MATCHING

The fast character matching in BIONUMERICS is a method to screen one or more entries against the complete database, based on distances calculated on a selected character experiment.

Different from previous BIONUMERICS versions, in which always the full character set was used, a character aspect can now be selected. This allows e.g. to screen entries based upon their cgMLST profiles (if a core set is defined) instead of using the complete wgMLST pattern.

A minimum overlap between the character sets can be specified, which is especially useful for character sets with many missing values to avoid false matches.
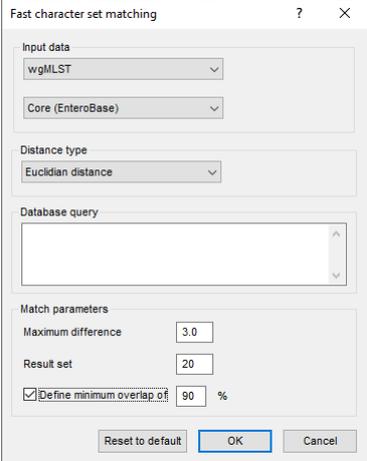
OTHER IMPROVEMENTS

BIONUMERICS version 8.1 received numerous other improvements, including:

- An additional import option was added to import sequence read sets with metadata directly from an NCBI SRA or EMBL ENA search result.
- The embedded Python interpreter was updated from version 2.7 to 3.8. Consequentially, all plugins and scripts were converted to the Python 3 syntax.
- A new Windows installer with fewer dependencies facilitates installation of the software.
- Convenience tools such as the ability to select a database in the start-up screen using the keyboard and the option to copy the content from the active cell in any grid panel speed up common tasks.
- On the Calculation Engine, the annotation tool Prokka was updated to version 1.14.5. Bowtie2 was replaced by SNAP as reference mapping algorithm after a SPAdes de novo assembly for more consistent results.