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 and species confirmation, a comeback of the Geo plugin, the new and versatile Custom genotyping plugin, the new MLST for WGS plugin as ultimate tool for MLST analysis starting from genome sequences, 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.

MINHASHING-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*.
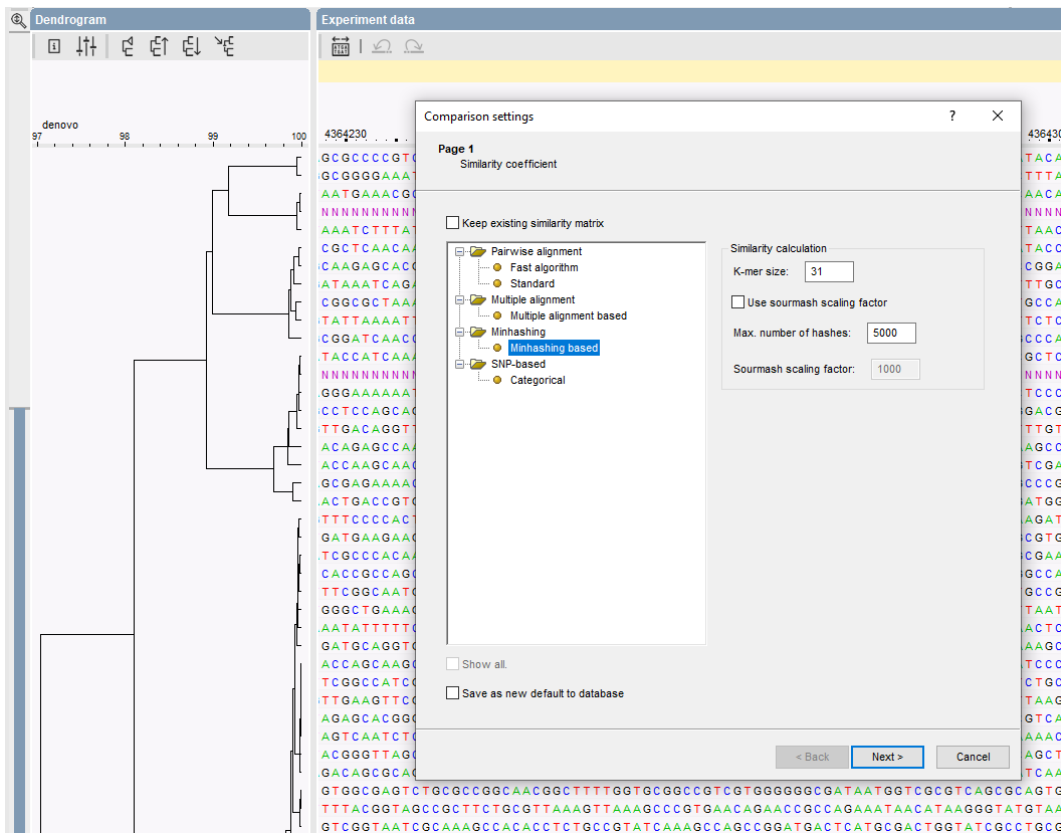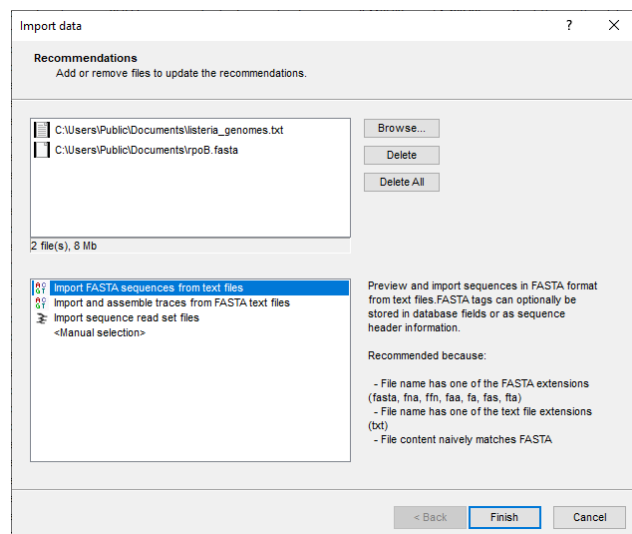
BIONUMERICS 8.1 release notes

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

FILE-FIRST APPROACH FOR IMPORT

The import wizard now includes an additional dialog that allows you to browse for your data files and suggests suitable import routines, based on file type and content.

This new feature greatly facilitates the import process and makes it more intuitive by eliminating the need to select the required import routine in the tree overview with all possible imports.

LONG READ DE NOVO ASSEMBLY

In addition to hybrid assembly (i.e., combining short and long reads in a single assembly) that was introduced earlier, it is now possible to perform long read only assemblies on e.g. PacBio and NanoPore sequence reads. This feature is implemented on the Calculation Engine via the open source Unicycler de novo assembler.

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 allows 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 that can be configured to your own needs. Starting from genome sequences and using BLAST, in silico PCR and minhashing approaches, it can be set up for following analyses:

- Detecting (sub)sequences within genomes
- Extracting sequences into their own experiment
- Detecting acquired traits such as antimicrobial resistance or virulence
- Detecting pre-defined point mutations
- Detecting traits caused by point mutations
- Confirming species identity

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.

Knowledge bases for this plugin consist of a small set of structured text files, adhering to existing formats (JSON, FASTA and TSV). The plugin contains tools for exporting knowledge base files in several formats and offers a convenient knowledge base validation tool. Additionally, example knowledge bases can be exported, including an extensive description of the expected format.

Custom genotyping models can be exported to a file and exported in another database, ensuring that exactly the same analysis can be replicated across databases.

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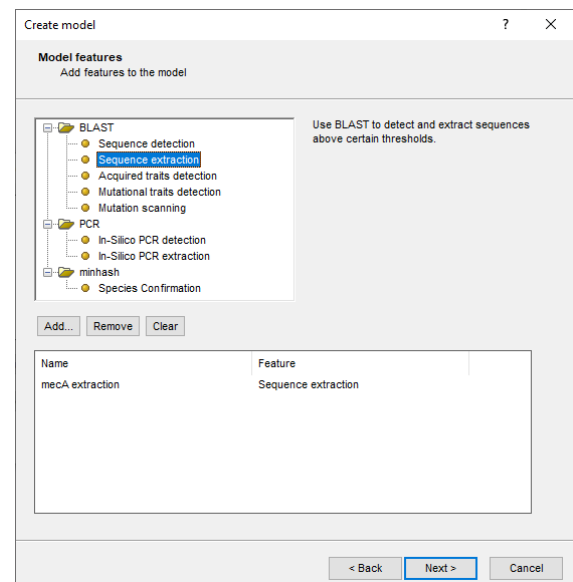
ACQUIRED RESISTANCE PLUGIN

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.
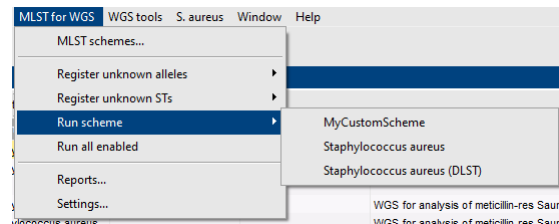
MLST FOR WGS PLUGIN

BIONUMERICS already provided several workflows for extracting MLST alleles from whole genome sequences, but most of these approaches required the MLST online plugin (now referred to as the "MLST for Sanger sequencing plugin") to be installed for the actual allele and profile assignments.

The new MLST for WGS plugin is a standalone tool, specifically designed for performing MLST analysis with genome sequences as input. Based on the genotyping framework, it performs a fast, accurate and complete allele calling and allows running multiple MLST schemes in a single database.

Similar as for the MLST for Sanger sequencing plugin, the MLST for WGS plugin allows you to select MLST schemes from an online list and to create and maintain custom MLST schemes.
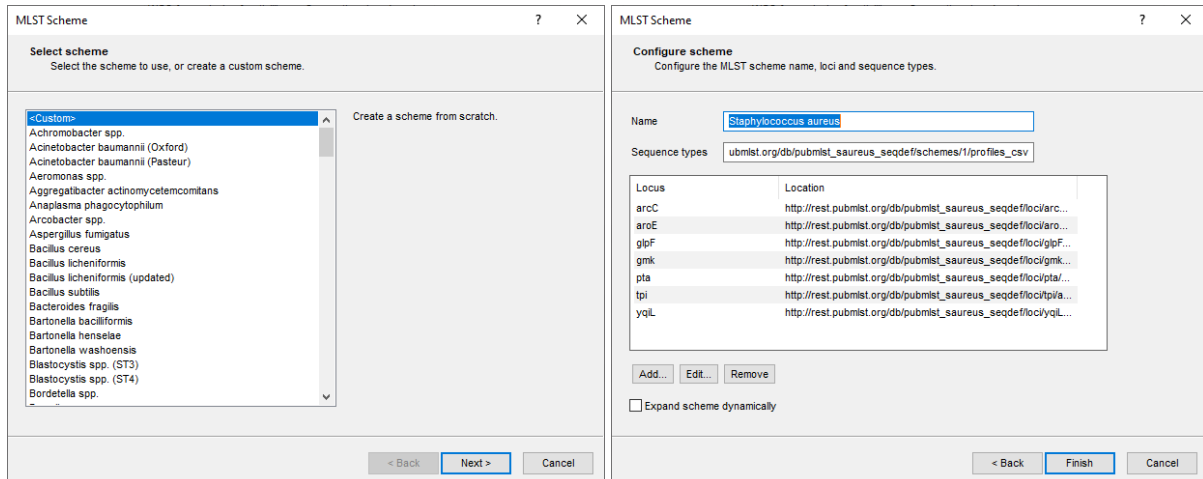
IMPROVED ORGANISM-SPECIFIC GENOTYPING PLUGINS

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.
- Each organism-specific genotyping plugin was extended with a species confirmation feature, which allows you to check if an isolate is indeed a target organism for the plugin.
- 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.

BIONUMERICS 8.1 release notes

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.

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.

With a single command, all matches with the screened entries can be selected for further analysis in the *Comparison window*.

BIONUMERICS version 8.1: New features   Copyright 1998 – 2022 Applied Maths NV. Not for diagnostic use.

BIONUMERICS 8.1 release notes

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.
- The MTBC genotyping plugin was upgraded to the latest genotyping framework.
- To extend the serviceable life of this final BIONUMERICS version, the concept of online plugins was reintroduced and plugins stored in the database take precedence over plugins in the installation directory.
- An improved and faster download of sequence read data from NCBI SRA.
- The default SQLite database can be compacted without the need for third-party software.
- Job logs can be removed in bulk to reduce database sizes.
- The open source tool fastp is now used for trimming sequence read sets on the Calculation Engine and local Calculation Engine.
- Sequences can be imported from zipped files, such as TAR archives downloaded from NCBI Assemblies.
- An additional filter was made available in the *Comparison window* to filter characters based on their value.
- 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, the SPAdes assembler was updated to 3.15.3 and Unicycler to version 0.5.0. SNAP was updated to version 2.0.0 and now replaces Bowtie2 as reference mapping algorithm after a SPAdes de novo assembly for more consistent results.