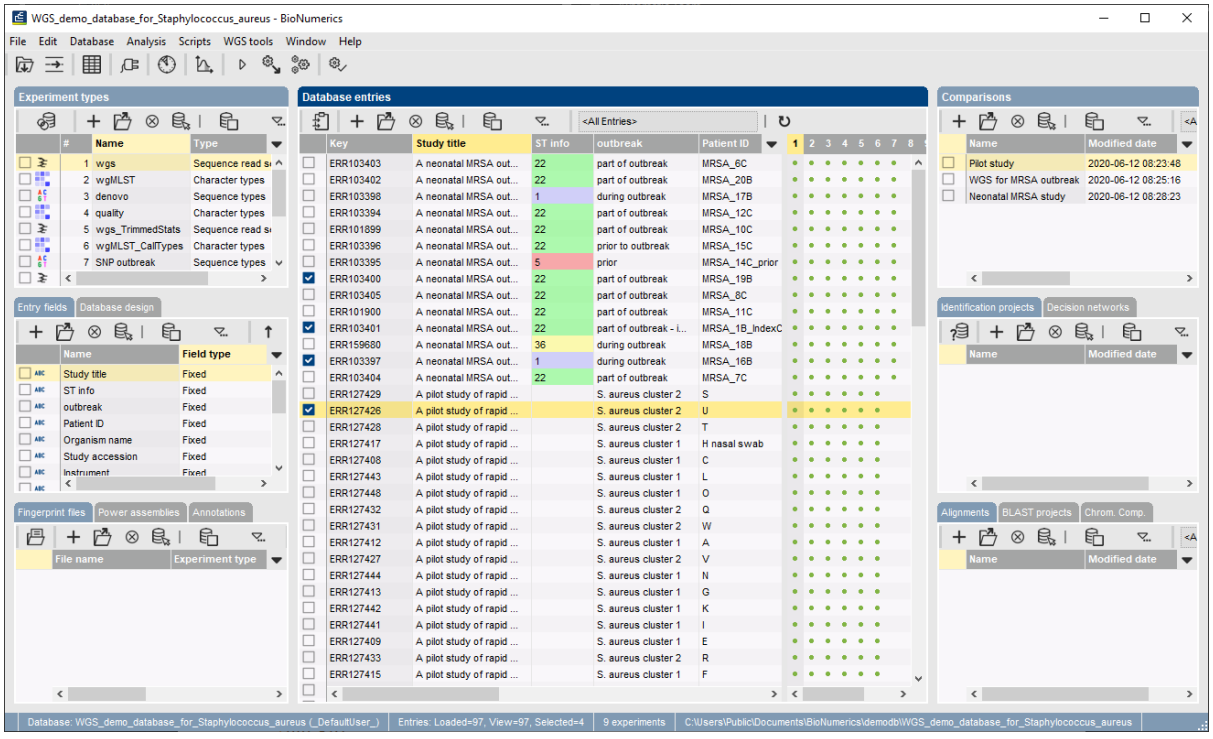
BIONUMERICS 8.0 release notes

Applied Maths is proud to announce BIONUMERICS version 8.0. This major update features a new user interface and focuses on whole genome sequencing, with the introduction of the local calculation engine, three new genotyping plugins, two new maximum likelihood clustering methods and many additional algorithms on the calculation engine for e.g. hybrid assembly, genome annotation and SNP analysis.

The BIONUMERICS offering is simplified and now consists of four different configurations: BIONUMERICS-SUITE, BIONUMERICS-SEQ, BIONUMERICS-MALDI and BIONUMERICS-GEL.

NEW LOOK AND FEEL

The BIONUMERICS user interface received a complete overhaul, featuring a contemporary flat design in the bioMérieux corporate color scheme. New icons are used throughout the software for a more consistent look and feel.

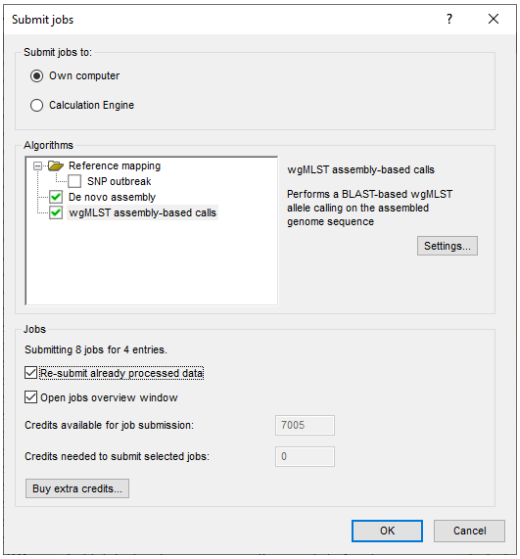


The software offers a different user experience depending on the configuration. BIONUMERICS-SUITE has all available user interface elements and hence the most complex interface. For other configurations, elements that cannot be used are hidden from view. This includes components such as menu items, buttons and panels and further extends to import-, export- and process routines if not relevant for a given configuration.

Each BIONUMERICS configuration has its own default layout of the main window. The goals are (1) to emphasise the Database entries panel as “main” panel (larger and centrally placed), (2) to inspire a workflow (data import, processing & database configuration on the left, analyses on the right of the Database entries panel) and (3) to hide any useless panels from view (see previous paragraph).

LOCAL CALCULATION ENGINE

The local calculation engine allows you to launch jobs on your own computer, much in the same way as launching jobs on the calculation engine. These jobs run as a separate process and hence do not block the use of BIONUMERICS while they are being processed. Since no data are sent to the calculation engine, this provides an answer for labs that need to deal with extremely strict data regulations. Job types offered on the local calculation engine include de novo assembly via SKESA, reference mapping using SNAP, assembly-based wgMLST allele calling and the maximum likelihood clustering algorithms RAxML and FastTree.



The re-designed Submit job dialog box deals with the expanded algorithm offering (see below) and allows to run jobs either on the local calculation engine or on the calculation engine. It clearly indicates the algorithm to be used, hence it is no longer needed to click the settings button to access this information.

Just as on the calculation engine, sequence read set data are downloaded transparently from NCBI's SRA, EMBL's ENA, Amazon S3 buckets, BaseSpace and Aliyun OSS buckets. A download job is automatically launched when needed and the de novo assembly or reference mapping job is started after the download job completes.

A comparison of the calculation engine and local calculation engine in use:

	Calculation engine	**Local calculation engine**
Cost per sample	Yes, credits are used	No, except electricity bill
Performance	Powerful calculation nodes	Depending on your computer's specifications
Time to result	Predictable	Variable, depending on job queue
Job concurrency	Scalable	Max. 2 jobs run in parallel
Algorithms available	All	Limited set
Dependence with BIONUMERICS client	Jobs run if BIONUMERICS is closed, even if computer is shut down	Main window should be open for new jobs in queue to start

To facilitate the use of the local calculation engine, demo projects can be accessed by simply selecting an organism from a drop-down list. Demo projects do not have credits assigned to them, but allow for a connection to the corresponding allele database for downloading search data and organism settings. The relatively bulky search data for assembly-based wgMLST allele calling are downloaded reliably by splitting the files up into segments and verifying the integrity of each file segment through its md5 checksum.

BIONUMERICS 8.0 release notes

IMPROVEMENTS ON THE APPLIED MATHS CLOUD CALCULATION ENGINE

Several additional algorithms were added as jobs on the Applied Maths Cloud Calculation Engine and are accessible in BIONUMERICS version 8.0 via the WGS tools plugin.

HYBRID GENOME ASSEMBLY

Hybrid genome assembly combines output of different sequencing technologies to obtain the best possible result. Hybrid assembly is implemented in BIONUMERICS version 8.0 via Unicycler on the calculation engine. Unicycler creates a short-read assembly graph (typically using Illumina data) and then uses long reads data (PacBio or NanoPore) to build bridges, often resulting in a complete genome assembly.

A new sequence read set experiment type setting ("Define as long reads") was introduced to indicate that this experiment contains long reads, i.e. from PacBio or NanoPore sequencers.

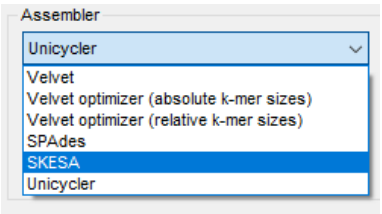
An additional sequence read set experiment type (default "wgsLong") is created by the WGS tools plugin for storage of long reads. A different experiment type can be specified in the WGS tools settings, as long as the experiment type is defined as a long read sequence read set.

The user should explicitly check "Hybrid assembly" in the de novo assembler settings for a hybrid assembly to be performed. If this option is unchecked, the software will default to short read assembly, even if long read data is available.

NEW / UPDATED DE NOVO ASSEMBLY ALGORITHMS

As mentioned above, the open source de novo assembly algorithm Unicycler is now available for hybrid assembly and for regular short read assembly. Final polishing of the assembly is done via Pilon and/or Racon.

NCBI's fast and performant SKESA assembler is added to the list of available de novo assembly algorithms. Final base calling is done via SNAP reference mapping to ensure consistent results between assemblies on the calculation engine and local calculation engine.

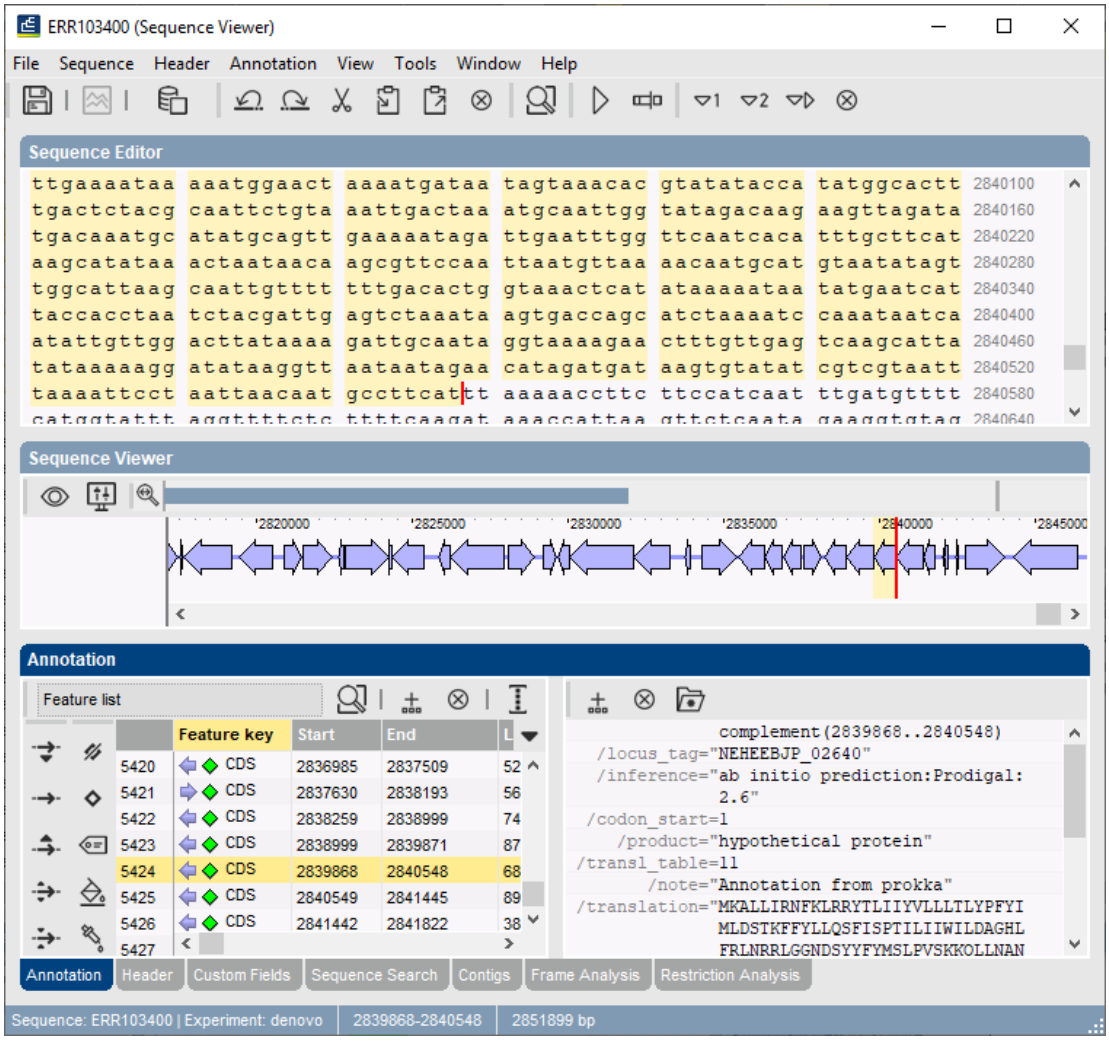


The open source SPAdes de novo assembly algorithm is updated from v3.7.1 to v3.13.1. Final base calling is done via Bowtie2 reference mapping.

PROKKA ANNOTATION

The Prokka annotation pipeline by Torsten Seemann is offered as a job on the calculation engine, complementing the already existing annotation tool in BIONUMERICS. Although primarily intended for

whole genome sequences, Prokka jobs can be launched on all nucleic acid sequence types except reference mapped sequence types.



Annotation features from Prokka all have the same note (“Annotation from prokka”) and are added to the annotation, i.e. any existing annotation features are not overwritten, even if defined on the same position as a Prokka feature. However, a new Prokka annotation overwrites all existing Prokka features.

CFSAN SNP PIPELINE

A CFSAN (US FDA, Center for Food Safety and Applied Nutrition) SNP analysis can be started from a comparison, using sequence read sets as input data. The resulting SNP matrix is stored as an aspect of the sequence read set.

BIONUMERICS 8.0 release notes

MAXIMUM LIKELIHOOD CLUSTERING

Two open source Maximum Likelihood (ML) clustering algorithms (RAxML and FastTree) are offered as jobs on the calculation engine. RAxML is a very popular and widely used ML algorithm in the scientific community. However, FastTree generates trees significantly faster than RAxML with little (and in most cases no) degradation in tree accuracy. RAxML and FastTree jobs can be launched from the comparison window.

Both ML algorithms only use aligned sequences as input data, e.g.:

- Gene sequences in the comparison window on which a multiple alignment was calculated.
- A SNP matrix generated by a wgSNP analysis.
- A SNP matrix generated by the CFSAN SNP pipeline.

REFERENCE MAPPING

The open source SNAP sequence aligner is added as reference mapper algorithm, offering accurate results and short calculation times for mapping sequence read sets against a reference sequence, either as pre-processing for a wgSNP analysis or as final base calling step in a de novo genome assembly.

DOWN SAMPLING FOR OVER-SEQUENCED SAMPLES

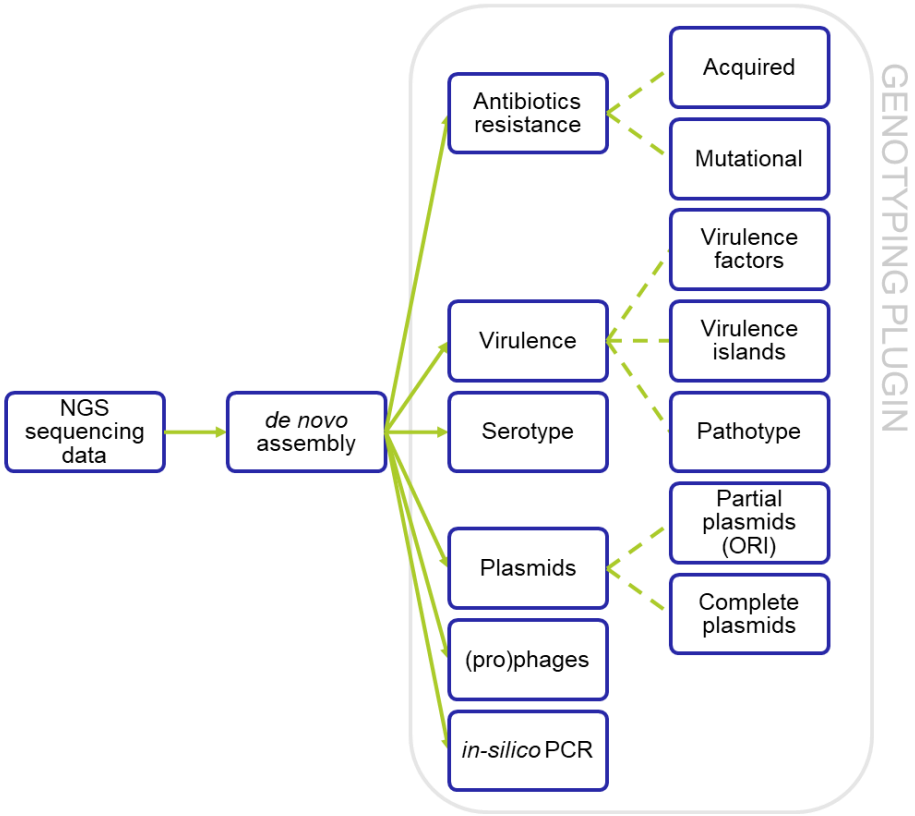
For all de novo assemblies, a down sampling procedure is implemented. This procedure assesses coverage based on the genome length as specified in the curator settings. If the coverage is lower than 200, no down sampling is performed and the de novo assembly proceeds with the original data set. If coverage is higher, the data set is down sampled for assembly. This procedure avoids long-running or failed de novo assembly jobs, while maintaining accurate results.

NEW FUNCTIONAL GENOTYPING PLUGINS

Functional genotyping plugins allow the prediction of phenotypic traits based on whole genome sequences. Additional organism-specific genotyping plugins are available in BIONUMERICS version 8.0 for *Listeria*, *Salmonella* and *Staphylococcus aureus* and the existing plugins for *Escherichia coli* and *Mycobacterium tuberculosis* complex (MTBC) are improved. The genotyping plugins are included in the installation while the knowledge bases are downloadable from the website. As knowledge bases are separated from plugin features, this allows you to specify the KB version to use, for each feature separately. The change logging of the knowledge bases has been improved and reports can be generated per strain/sample. There is a general report that highlights the main conclusions and a detailed report that provides the evidence on which the main conclusions are based. Analysis times are optimized and now range between 30 seconds and 2 minutes per sample.

E. COLI FUNCTIONAL GENOTYPING

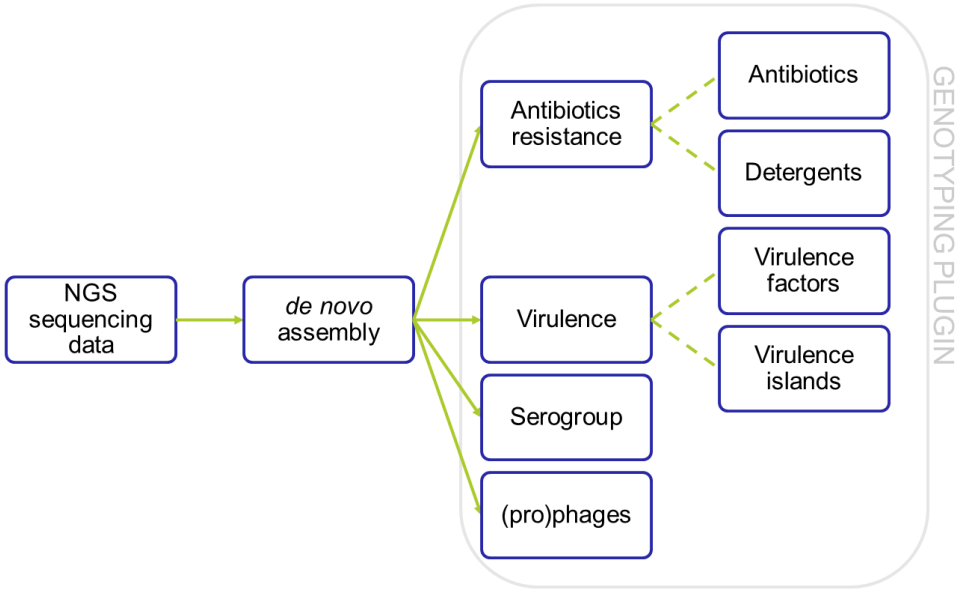
The *Escherichia coli* functional genotyping plugin allows you to screen *E. coli* whole genome sequences to predict serotype, virulence, and antibiotic resistance. It also allows you to detect phages and plasmids and to extract PCR amplicon sequences. The existing plugin was updated to benefit from the separation between plugin code and search data and from faster execution times.



LISTERIA FUNCTIONAL GENOTYPING

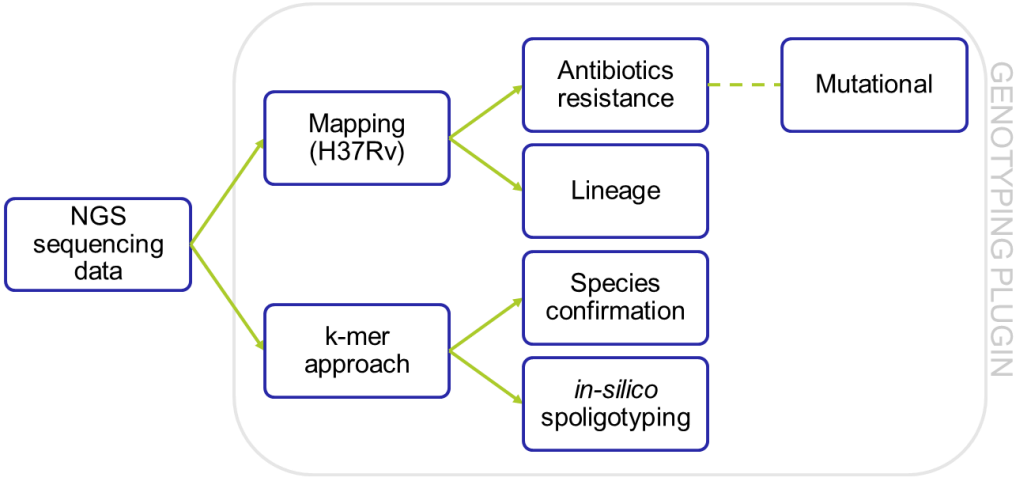
The *Listeria* functional genotyping plugin is new in BIONUMERICS version 8.0 and allows you to screen *Listeria* whole genome sequences to predict serotype, virulence, and resistance against antibiotics and detergents. It also allows you to detect phages.

BIONUMERICS 8.0 release notes



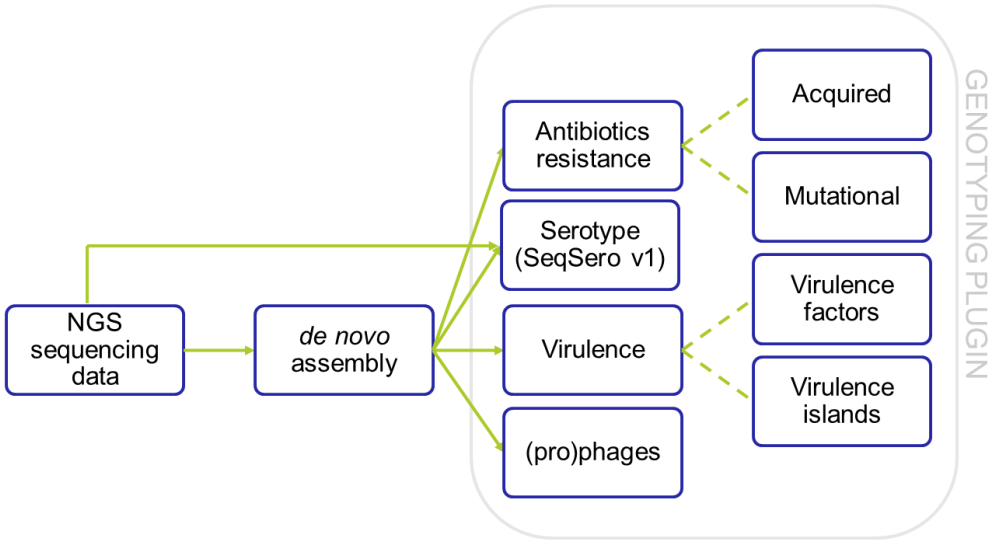
MTBC FUNCTIONAL GENOTYPING

The MTBC functional genotyping plugin allows you to screen *Mycobacterium tuberculosis* complex whole genome sequences to predict spoligo type, species, lineage, and resistance against antibiotics.



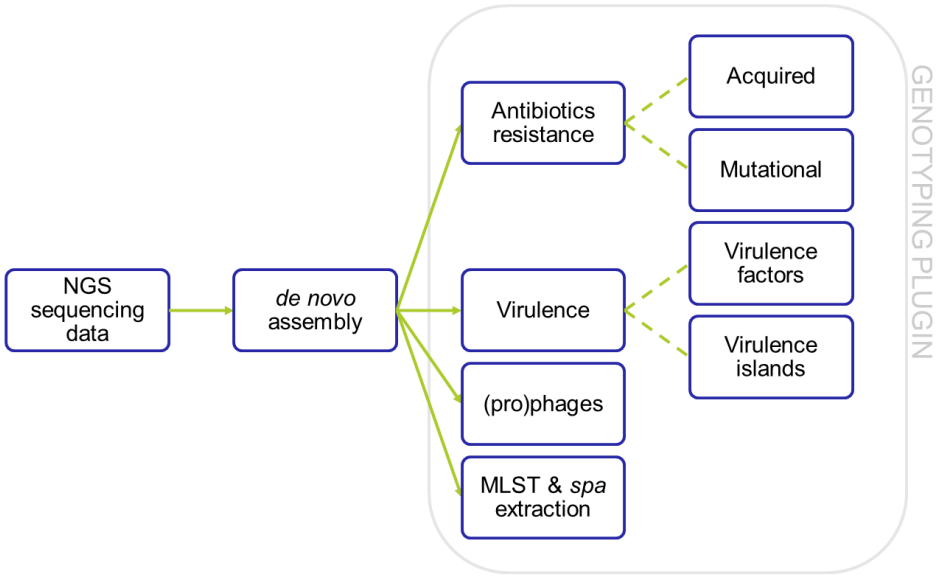
SALMONELLA FUNCTIONAL GENOTYPING

The *Salmonella* functional genotyping plugin is new in BIONUMERICS version 8.0 and allows you to screen *Salmonella* whole genome sequences to predict serotype (via SeqSero on the calculation engine), virulence, and antibiotic resistance. It also allows you to detect phages.



STAPHYLOCOCCUS AUREUS FUNCTIONAL GENOTYPING

The *Staphylococcus aureus* functional genotyping plugin allows you to screen *S. aureus* whole genome sequences to predict virulence and antibiotic resistance. It also allows you to detect phages and to extract MLST and spa repeat sequences.



BIONUMERICS 8.0 release notes

IMPROVED PERFORMANCE

Nucleic acid and amino acid sequences in a BIONUMERICS database are now stored as GZIP archives instead of plain text. This compression, combined with a better indexing of the relevant database tables, decreases the loading time for sequences and therefore improves responsiveness, especially when working with whole genome sequence databases over the network.

Results from wgMLST assembly-based allele calling jobs are compressed on the calculation engine to speed up their download.

OTHER IMPROVEMENTS

BIONUMERICS version 8.0 received numerous smaller improvements:

- New import routines for Vitek MS data and Bruker FTIR spectra.
- Additional options are available for exporting sequences from the comparison window to FASTA files.
- New Chromium embedded web browser to replace the outdated Gecko browser. The embedded web browser is used e.g. for displaying the online help, running online scripts and rendering functional genotyping reports.
- The Database tools, Dendrogram tools, Sequence translation tools and Sequence extraction plugins are now installed by default.
- The number of currently selected entries in the database is shown in the caption of the Comparison window.
- Two additional buttons are available in the Startup screen, to facilitate the download of demonstration databases and for finding the database you are looking for in a potentially long list of databases, respectively.
- The BIONUMERICS user interface can be translated in Chinese. The language is selected in the Startup screen.
- Faster database backup and restore functionality due to the implementation of a new 7-zip version.
- High-DPI compatibility setting (scaling by Windows) is enabled by default on BnStart.exe and Bn.exe to improve the rendering of the software on high-resolution (4K) screens.
- The BIONUMERICS installer does not have a dependency on the .NET 3.5 framework anymore, resulting in an easier installation.