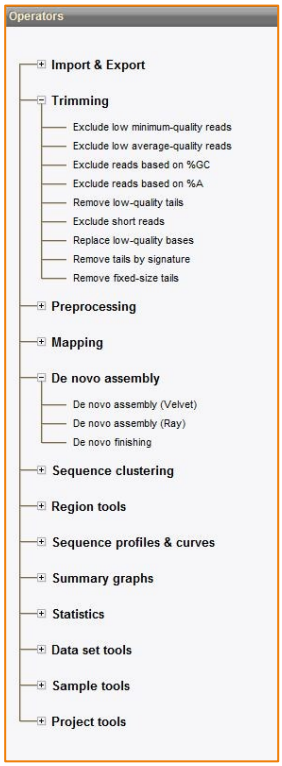
RELEASE NOTE: BIONUMERICS VERSION 6.6

While in the coming months all activity at Applied Maths will be centered on the next major BioNumerics upgrade version 7.0, we are pleased to announce a second interim upgrade of version 6.x. This upgrade, version 6.6, contains important enhancements to the Power Assembler and a new *de novo* assembly tool for next generation sequence (NGS) data. Furthermore, a number of improvements have been made to the Import plugin, and a number of bugs reported in the BioNumerics main program have been resolved.

POWER ASSEMBLER



De novo assembly

De novo assembly is based on two open source third party tools: *Velvet*[1] & *Ray*[2]. Both applications are GPL-licensed, and have been revised to compile under Windows in 32 bit as well as 64 bit modes. Both assembly tools are complemented by an in-house developed greedy algorithm for finishing.

To assess the resulting contigs or scaffolds, the operator *Contig statistics* has been added. This operator calculates summary statistics on a set of sequences, including the N10, N25, N50 and N95 count and length.

Reduced RAM requirements

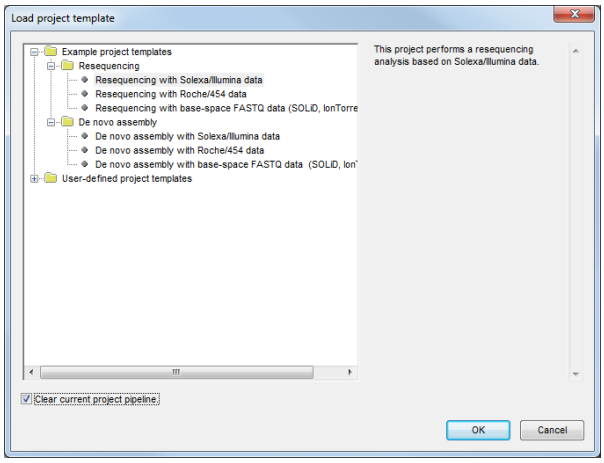
Data is kept as much as possible on the hard disk rather than in memory, in order to relax the RAM memory requirements. This means that the number of reads is no longer a restrictive factor in resequencing projects, as long as there is sufficient hard disk space. Example: The memory footprint of the Power Assembler in a resequencing experiment of an *E. coli* genome (4.6Mb) with 5.4 million reads (about 670Mb on disk) is about 475Mb (of which 326Mb is occupied by the index structures).

The disk-based approach has been optimized to work even faster than the former memory-based approach. All data is stored in the *working directory* (by default, the OS temporary directory). The calculation time for an assembly can further be reduced significantly by having the working directory on a Solid State Disk.

[1] Zerbino D. R. and Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. 2008. Genome Research 18: 821-829.
[2] Boisvert S., Laviolette F. and Corbeil J. Ray: Simultaneous Assembly of Reads from a Mix of High-Throughput Sequencing Technologies. 2010. Journal of Computational Biology 11: 1519-1533.

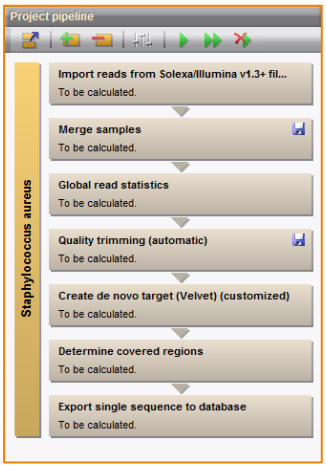
Project templates

The pipeline of a project (that is, the sequence of actions in the project) can now be saved as an analysis template in the database. Functionality has been added to load templates from the database into a project, and to remove existing templates. The pipeline of a project can also be exported to a file (in XML format), or can be imported from a previously created file. A number of new, pre-defined pipelines are available.



Log file visualization

During the processing, the activity log file can be visualised in real-time in the user interface. This makes it easier to keep track of the execution of the project actions. Within a session, for every action in the project the last log file is kept.



Resolving imperfect multiplex identifiers

Since multiplex identifier tags (barcodes) are subject to sequencing errors, the tags attached to a read sequence can be slightly different from the expected tags. Therefore, a procedure has been developed that resolves imperfect barcodes in order to assign as many read sequences as possible to their respective samples.

BIONUMERICS PROGRAM

NEW FEATURES

- Improved usage statistics reporting in the NetKey+ license server.
- ParseUsage.exe to parse NetKey+ log files.
- Possible to "fill range" with frequency bar graphs.

BUG FIXES

- Resolved crash when blue nodes are dragged on linked gel.
- Quoting Oracle sequences with double quotes (to avoid problems with Oracle schemas).
- Some corrections on translation tools.
- Restored feature to remember location in *Open file* dialog box in scripts.
- SNP detection tool in the chromosome comparison does not report unresolved bases ('N') as SNPs anymore.
- Possible to use HTTPS downloads from Python scripts.
- Fixed a bug in the Python script function that loads entry attachments.
- Sequence viewer can now open NGS projects after modifying a sequence in the editor.

IMPORT PLUGIN

New version of the import plugin: v2.14

NEW FEATURES

- ODBC import: treat “None” fields as empty strings (same behaviour as BNS script language and also what most users expect). No need anymore to go to the advanced options and select a default value.
- ODBC import: ignore 'empty' records (all field content is empty).
- Import text files: if headers are duplicated, report which ones.
- Character import: when creating new characters for binary type, automatically set maximum range to ‘1’.
- Import template preview: show all rows instead of first 100.

BUG FIXES

- Sequence import: fixed the error that occurred when importing sequences without import template.
- Sequence import: last selected import template is now selected in a subsequent import.
- Sequence import: possibility to select “[None]” to avoid using a template for import (previously, this was not possible if at least one template was defined).
- Import template: binding source 'Fixed value' can be used.
- Entry link field: an error message appears if the content is non-unique, but non-existing in the database (previously, separate entries were created).
- Large file import: display meaningful error "File is too large" (previously, no error message was displayed).
- Templates are compatible with local databases.

OTHER PLUGINS

- Batch Sequence Assembler plugin v1.34
  In rare cases, the assembly would 'hang'. This has been fixed.
  The former shortcut key to “Set all warnings/errors to solved, save and close”, Shift+S, induced ‘S’ in the sequence before closing. This short key has been changed to Ctrl+Shift+S.

- SNP calling plugin v1.13
  Export opens the report directly in Notepad.