




What's new in the 2.6 release of KoriBlast

Summary

1. PSI-Blast searches are available (Search Module)
2. Prepare personal Blast databanks from native *GenBank*, *Genpept*, *EMBL*, *SwissProt* and *Uniprot* files, using either plain text or gzipped compressed files (Search Module)
3. Install pre-formatted NCBI Blast databanks directly from their native 'tar.gz' files (Search Module).
4. Design search filter enabling pattern matching on hit sequences (Search & Filter Modules)
5. PSI-Blast adapted hits viewer (Analyse Module)
6. Compare search results to find new, common or deleted hits (Search & Analyse Modules)
7. Export KoriBlast data files and filters to share with colleagues having KoriBlast or BlastViewer (Analyse & Filter Modules).
8. New Sequence Viewer to display and manipulate sequences (extract, translate, get compositions) downloaded from public repositories (Document Viewer Module).
9. Better handling of large projects


The following sections give more details about all the novelties of KoriBlast 2.6

A. Search Module

- **add PSI-Blast search capabilities.** This new search engine is available for both online (EBI server¹) and local searches². The Analyse Module has been adapted to display results produced by PSI-Blast (see below).
- **update local Blast databanks creation.** It is now possible to process directly sequence data files in the following formats: *GenBank*, *Genpept*, *EMBL*, *SwissProt* and *Uniprot* (in addition to the Fasta format available in previous release). These files can be provided as plain text or gzipped compressed files. For example, this feature enables the direct processing of files retrieved from the *Genomes Division* on the NCBI ftp site. Please refer to the Korilog online support for more information: <http://www.korilog.com/support>, section 'Blast resources'.
- **install native NCBI Blast databanks** from archived files. When installing locally a Blast databank (i.e. using the Local Blast search system), it is now possible to process NCBI Blast files in their native format 'tar.gz'. More generally, when using the dialogue box 'Available Blast Databanks', the new command  provides an easy-to-use tool to unarchive the following files: '.tar', '.tar.gz', '.gz', '.tgz' and '.zip'. Please refer to the Korilog online support for more information: <http://www.korilog.com/support>, section 'Blast resources'.
- **new layout for the display of percent values.** This concerns the following columns in the Job Result table: identity, similarity, query coverage and hit coverage. The display of % values is now associated to a colour-coded bar allowing a better evaluation of these values.

1 The NCBI PSI-Blast server is not accessible since there is no support from the NCBI staff for using this service programmatically.

2 both services are available in the 'Search System' drop down list in the Search Module. EBI PSI-Blast is available as a dedicated service, whereas local PSI-Blast is available as part of the Local Blast service (check the list of Blast programs).

- **new selection commands.** These are available in the Job Result table's popup menu (accessible through a right-click on this table). It is now quite easy to select all or no hits, as well as inverting a previous selection.
- **new commands to compare results.** When comparing search results (command  located at the bottom of the Job History table), in addition to finding new hits, it is now possible to find also common or deleted hits between a selection of results.

B. Analyse Module


Comparison of results

The Analyse Module enables the direct **selection and comparison of result entries** from the list of opened results (left side of the Analyse Module) to find new, deleted or common hits. This command is accessible by using the toolbar located on the top of the list of opened results.

Export result entries

It is now possible to **export any results from the Analyse Module**. This feature is compatible with the BlastViewer, i.e. a search result exported from KoriBlast can be opened in the BlastViewer. This export command is accessible by using the toolbar located on the top of the list of opened results.

Hit List View Enhancements

- **new layout for the display of percent values.** This concerns the following columns: identity, similarity, gaps, query coverage and hit coverage. The display of % values is now associated to a colour-coded bar allowing a better evaluation of these values.
- **new columns** are displayed, namely: division, taxonomy, creation date and update date. These are the data available for every hit when features have been downloaded.
- **new selection commands** are available in the Hit List's popup menu (accessible through a right-click on this table). It is now quite easy to select all or no hits, as well as inverting a previous selection.
- **PSI-Blast adapted display.** In the context of the analysis of a PSI-Blast result, the top of the Hit List table displays the list of iterations produced by PSI-Blast. This feature is only available for data files produced with the local PSI-Blast, since the EBI PSI-Blast always report the results from the latest search iteration.
- **Document Viewer Link:** a new command  has been added to allow the direct download and display of hit sequences in the new Sequence Viewer (see Document Viewer, below).

Feature Viewer

- The Feature Viewer, located on top of the pairwise sequence alignment, enables the **display by feature types**.

C. Filter Manager Module

- **Import/export filters.** These commands enable the sharing of filters with other users using the two new [Import] and [Export] buttons located at the bottom of the filters table.
- **Pattern mathing on query and hit sequences.** It is possible to design a filter that uses pattern matching on the query and hit sequences. Such a filtering system enables a generalized PHI-Blast like search system for all search systems available with KoriBlast.

D. Document Viewer Module

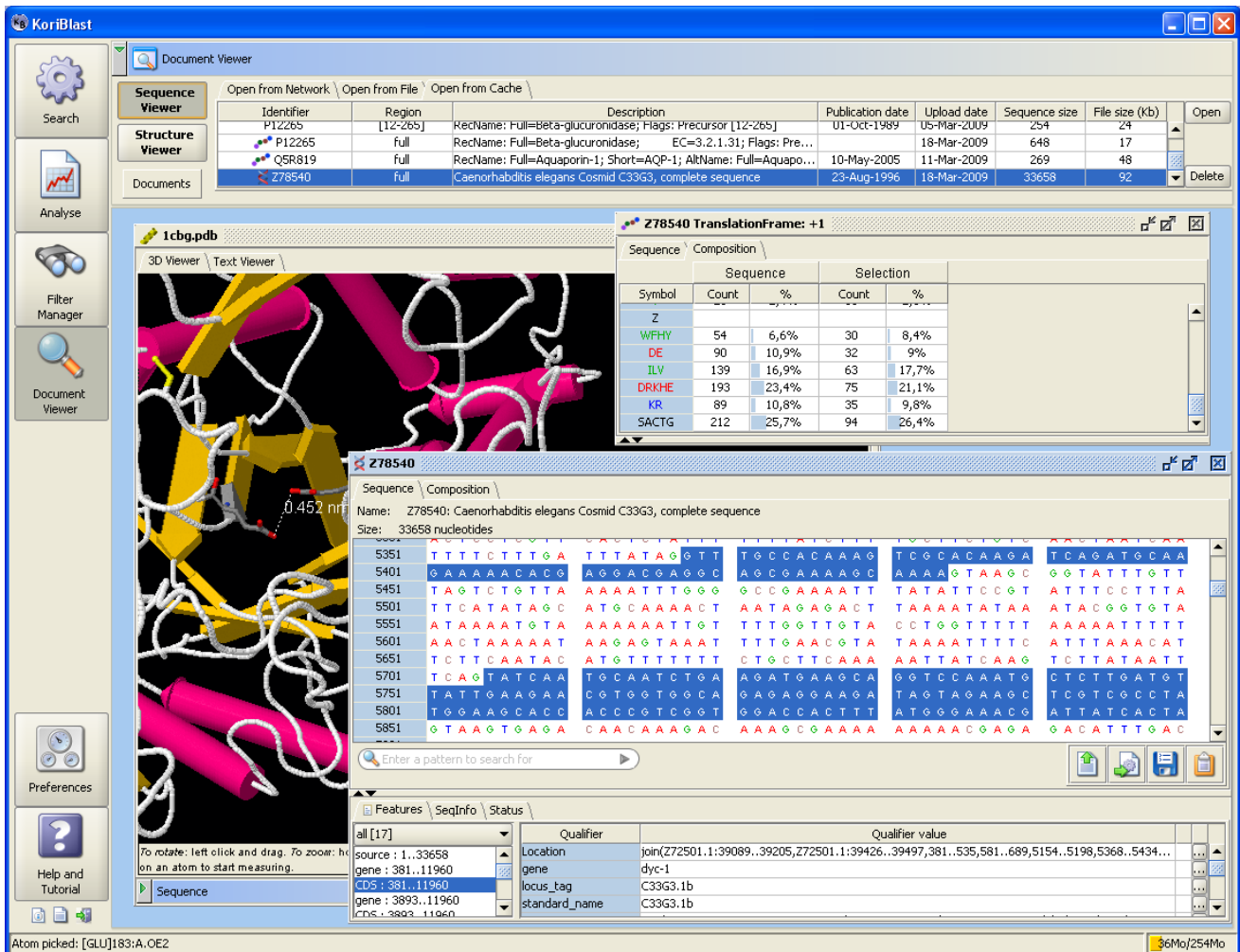
'Structure' Viewer

- **Open local file.** In addition to the download of 3D structures from the RCSB server (www.rcsb.org), it is now possible to open PDB data files available locally on your disk drive.
- **Better cache management.** The viewer manages a cache where downloaded 3D structures are stored so that they remain available over time, as well as when you work offline.

E. A brand new viewer available in the Document Viewer Module

'Sequence' Viewer

- The new Sequence Viewer enables the **display of full and partial sequence entries retrieved from the NCBI (Genbank, Genpept, RefSeq, Genomes) and Uniprot (SwissProt and TrEmbl)** online servers. It is also possible to open a sequence file using one of the following formats: Fasta, GenBank, Genpept, EMBL, SwissProt and Uniprot.
- All **downloaded sequences are stored in a cache** on your local disk drive so that they remain available over time, as well as when you work offline.
- The viewer is capable of **displaying a colour-coded sequence, its composition** (full sequence and selected region) and known features, as shown on this picture:



The screenshot shows the KoriBlast software interface. On the left is a sidebar with icons for Search, Analyse, Filter Manager, Document Viewer, Preferences, and Help and Tutorial. The main window is divided into several panes:

- Document Viewer:** A table listing documents with columns for Identifier, Region, Description, Publication date, Upload date, Sequence size, and File size (kb). The selected document is Z78540.
- 3D Viewer:** Displays a 3D ribbon model of a protein structure in yellow and pink. A distance of 0.452 nm is indicated between two atoms.
- 278540 TranslationFrame: +1:** A composition table showing the distribution of amino acids in the selected sequence.

| Symbol | Count | % | Count | % |
|--------|-------|-------|-------|-------|
| Z | | | | |
| WFHY | 54 | 6,6% | 30 | 8,4% |
| DE | 90 | 10,9% | 32 | 9% |
| ILV | 139 | 16,9% | 63 | 17,7% |
| DRKHE | 193 | 23,4% | 75 | 21,1% |
| KR | 89 | 10,8% | 35 | 9,8% |
| SACTG | 212 | 25,7% | 94 | 26,4% |
- 278540:** A sequence viewer showing the full sequence of 33658 nucleotides. The sequence is displayed in a grid format with a search bar below it.
- Features:** A table showing sequence features with columns for Qualifier and Qualifier value. The selected feature is 'cds'.

| Qualifier | Qualifier value |
|---------------|---------------------------------------------------------------------------------------------|
| all [17] | |
| Location | join(272501.1:39089..39205,272501.1:39426..39497,381..535,581..689,5154..5198,5368..5434... |
| gene | dyc-1 |
| locus_tag | C33G3.1b |
| standard_name | C33G3.1b |

- The viewer is highly interactive: selection of sub-sequences in various ways, and selected features are immediately highlighted on the sequence.
- The viewer provides various functions to extract and transform (sub-)sequences. Transformations concern: complement, reverse, reverse/complement and translation. Extracted and/or transformed sequences can be displayed in a separate sequence view, submitted to a Blast search system, saved in a Fasta files or copied to the Clipboard.
- The viewer provides a pattern matching engine to locate sequence patterns.
- More details on this viewer and its use is available in the KoriBlast User Manual.

F. Other improvements:

- Fetching data from the NCBI servers uses the new recommendations: no more than 3 queries every second. (Your current configuration is not updated during the 2.6 installation. If you want to modify this timer, please go to the Preferences Dialogue Box, select the section 'NCBI Feature Connector', then update the field called 'Submit').
- Software configuration: it is possible to enable a proxy configuration using an authentication.
- To enable a better management of large search projects, KoriBlast is about 10 times faster at startup. (The first time you start KoriBlast 2.6 (i.e. just after its installation), it will convert your projects and it will start 'as slower as' the 2.5 release. However, next starts of KoriBlast 2.6 will be faster).