



This document presents what is new in the version 3.2 of KoriBlast.

A) Summary of major updates

1. Explore blast hit common ontologies within a new dedicated classification viewer (Taxonomy, GeneOntology, InterPro, Enzyme) in the Analyse module.
2. Analyse large blast result dataset (1.000.000+ hits) in the improved version of the Global Viewer.

B) Summary of minor updates

3. Search module - display size of job folders in the "Job Info" dialogue box.
4. Enable editing and batch editing of sequences.
5. Global Viewer - "PathWay Explorer" gateway - add popup action to select/unselect categories (Enzyme) within selected hit list.
6. Document Viewer: get Genbank/Genpept data files from the NCBI

C) Compliance with the new "KoriBlast server"

"KoriBlast server" is a new product designed to manage and dispatch a large number of Blast queries over a grid computing framework . This product is distinct from "KoriBlast" and is not required for "KoriBlast" to be fully functional. Nonetheless "KoriBlast 3.2" is compliant with "KoriBlast server" and offers two functionalities when a "KoriBlast server" is available :

7. Send remote Blast queries on a grid computing framework in a true generic "client/server" mode interacting with the new "KoriBlast server".
8. Annotate automatically Blast query hits using the "KoriBlast server" annotation retrieval functionality.

A) Major updates

1. Explore blast hit common ontologies within a new dedicated classification viewer (Taxonomy, GeneOntology, InterPro, Enzyme) in the Analyse module.

The viewer is available through the "classification" tab pane. It offers two views on the ontology classification (in the example: GeneOntology) built from all hit features gathered for the query being analysed. Data can be refreshed at any time using the "compute" button (after an update of the reference classification for instance).

The "table" report is a spread sheet of all hit features found for the selected ontology classification. When you select a classification item within the table then all the corresponding hits (those having this feature) are displayed below. A tool bar enables you to export in csv format or print the selected features.

The screenshot displays the KoriBlast software interface. The main window is titled "Classification" and shows a table report for Gene Ontology. The table lists various GO terms and their associated hit definitions. A red arrow points to the "Classification" tab in the top navigation bar. Another red arrow points to the "Table Report" button. A third red arrow points to the "Gene Ontology" dropdown menu. Below the table, a detailed view of a selected hit is shown, including an alignment map and a list of features.

ID	Name	Hits	Score	Hit Definition
GO:0003677	DNA binding	13	390	RecName: Full=Probable transposase for insertion sequence element ISRM3-like;
GO:0004803	transposase activity	13	390	RecName: Full=Probable transposase for insertion sequence element ISRM3-like;
GO:0006313	transposition, DNA-mediated	13	390	RecName: Full=Probable transposase for insertion sequence element ISRM3-like;
GO:0006527	arginine catabolic process	3	34	RecName: Full=Biosynthetic arginine decarboxylase; Short=ADC; EC=4.1.1.19;
GO:0008295	spermidine biosynthetic process	3	34	RecName: Full=Biosynthetic arginine decarboxylase; Short=ADC; EC=4.1.1.19;
GO:0008792	arginine decarboxylase activity	3	34	RecName: Full=Biosynthetic arginine decarboxylase; Short=ADC; EC=4.1.1.19;
GO:0046872	metal ion binding	3	34	RecName: Full=Biosynthetic arginine decarboxylase; Short=ADC; EC=4.1.1.19;

#	Accession	Definition	E-Value	Quality	# HSPs	Identity	Positive	Query Coverage	Hit Coverage
1	Y4PO_R...	RecName: Full=Probable transposase for insertio...	1.E-107	☺	1	49.7%	68.9%	91.8%	92.2%
2	TRA3_R...	RecName: Full=Transposase for insertion sequen...	6.E-107	☺	1	49.2%	68.4%	91.8%	92.2%
3	TRA_CO...	RecName: Full=Probable transposase for insertio...	1.E-81	☺	1	41.5%	62.3%	80.1%	99.7%
4	TRA5_LA...	RecName: Full=Transposase for insertion sequen...	4.E-57	☺	1	32.5%	52.6%	93.5%	96.2%
5	TRA1_LA...	RecName: Full=Transposase for insertion sequen...	7.E-52	☺	1	31.5%	53.4%	86.8%	94.9%
6	TRAS_R...	RecName: Full=Transposase for insertion sequen...	9.E-37	☺	1	26.8%	49%	86.8%	87.9%

Classification viewer: table report.

Important notice: it is required to upgrade the Gene Ontology databank to enable the use of these new features; use the Databank Manager for the upgrade.

The "tree" report is an interactive tree viewer that displays each hit feature with all its hierarchy within the ontology classification. When you select a classification item within the tree then all the corresponding hits (those having this feature) are displayed below. A tool bar enables you to export in csv format or print the selected features.

The screenshot shows the KoriBlast interface with the following components:

- Search Panel:** Contains the query "angolac blastp vs. yersinia GJ162351655[Gb]ABX85603.1| MIOC P...".
- Navigation Sidebar:** Includes Search, Analyse, Global Analyse, Databank Manager, Document Viewer, Preferences, and Help and Tutorial.
- Classification View:**
 - Buttons: List, Conserved regions, Alignment & Tree, Position Specific Matrix, Classification.
 - Tree Report: A hierarchical tree of Gene Ontology terms. Red arrows point to the "Tree Report" button and a specific node "GO:0003676 nucleic acid binding".
 - Table of Hits:

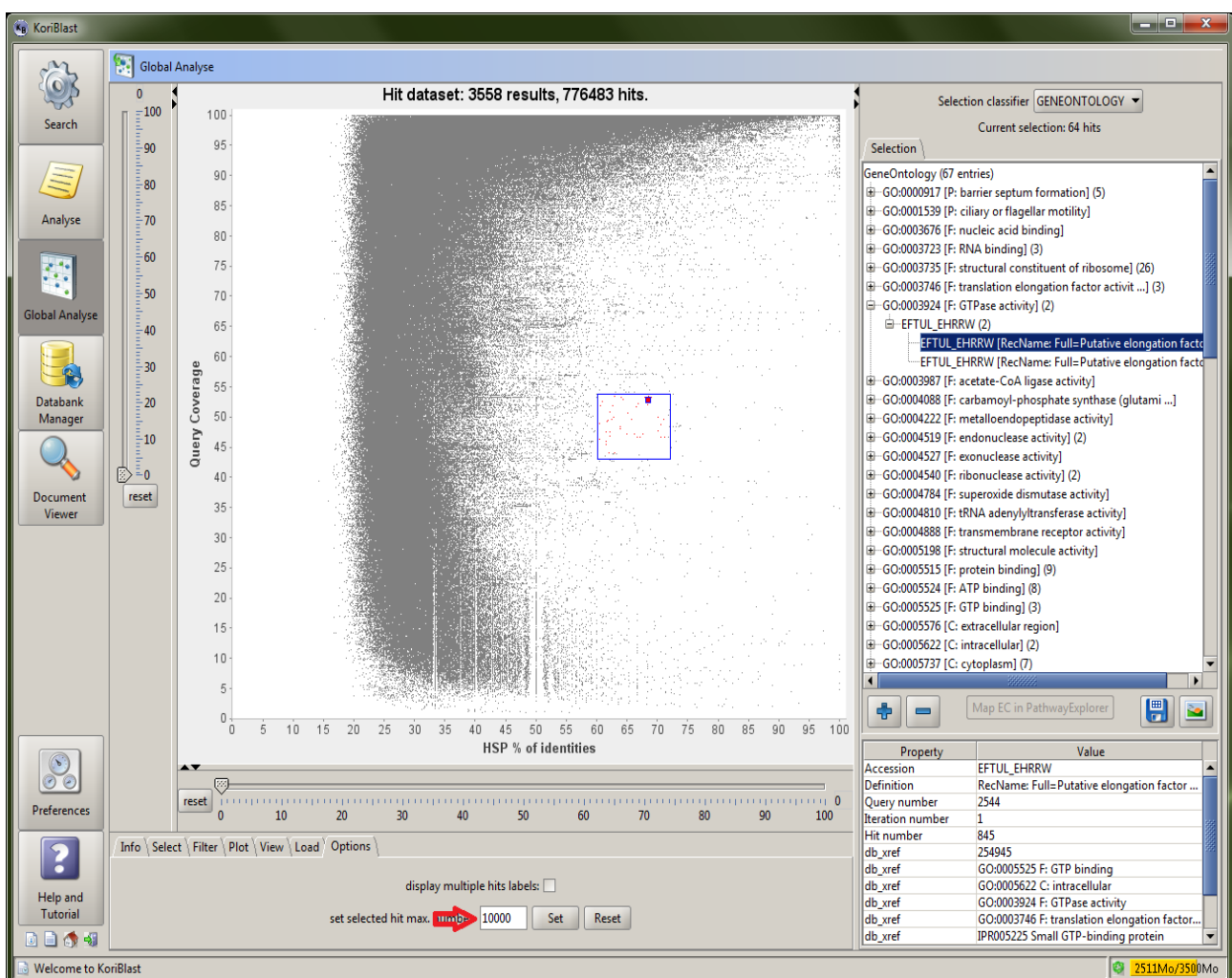
#	Accession	Definition	E-Value	Quality	# HSPs	Identity	Positive	Query Coverage	Hit Coverage
1	Y4PO_R...	RecName: Full=Probable transposase for insertio...	1,E-107	🟢	1	49%	68.9%	91.8%	92.2%
2	TRA3_R...	RecName: Full=Transposase for insertion sequen...	6,E-107	🟢	1	49.2%	68.4%	91.8%	92.2%
3	TRA_CO...	RecName: Full=Probable transposase for insertio...	1,E-81	🟢	1	41.5%	62.3%	80.1%	99.7%
4	TRA1_LA...	RecName: Full=Transposase for insertion sequen...	4,E-57	🟢	1	32.5%	52.6%	93.5%	96.2%
5	TRA1_LA...	RecName: Full=Transposase for insertion sequen...	7,E-52	🟢	1	31.5%	53.4%	86.8%	94.9%
6	TRA5_R...	RecName: Full=Transposase for insertion sequen...	9,E-37	🟢	1	26.8%	49%	86.8%	87.9%
 - Alignment Viewer:** Shows an alignment between the query (402 aa) and the hit Y4PO_RHISN (400 aa) with a pI of 9.07 and Mw of 42101.55 Da. It includes an HSP map and a sequence alignment view.

Classification viewer: tree report.

2. Analyse large blast result dataset (1.000.000+ hits) in the improved version of the Global Viewer.

The Global Viewer enables you to analyse and explore large blast result datasets provided you have enough RAM memory available on your computer.

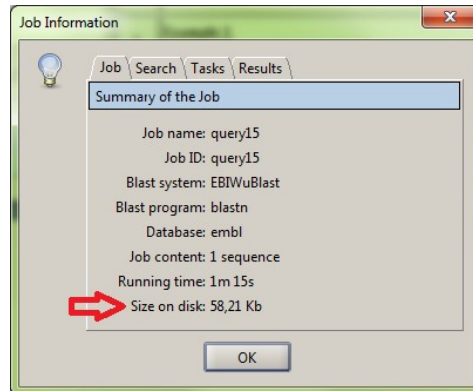
In order to prevent unwanted lacks of RAM memory, when a selection is performed on the plot, the number of hits displayed (in the selection panel) is limited. Default value is set to 10.000 hits maximum. If the selection exceeds the limit then the hits are sorted by their evalule and only the best are kept within the selection panel. You can change this limit in the "option" tab panel below the plot (the field to edit is pointed by the red arrow in the snapshot).



Global Viewer: large result dataset.

B) Minor updates

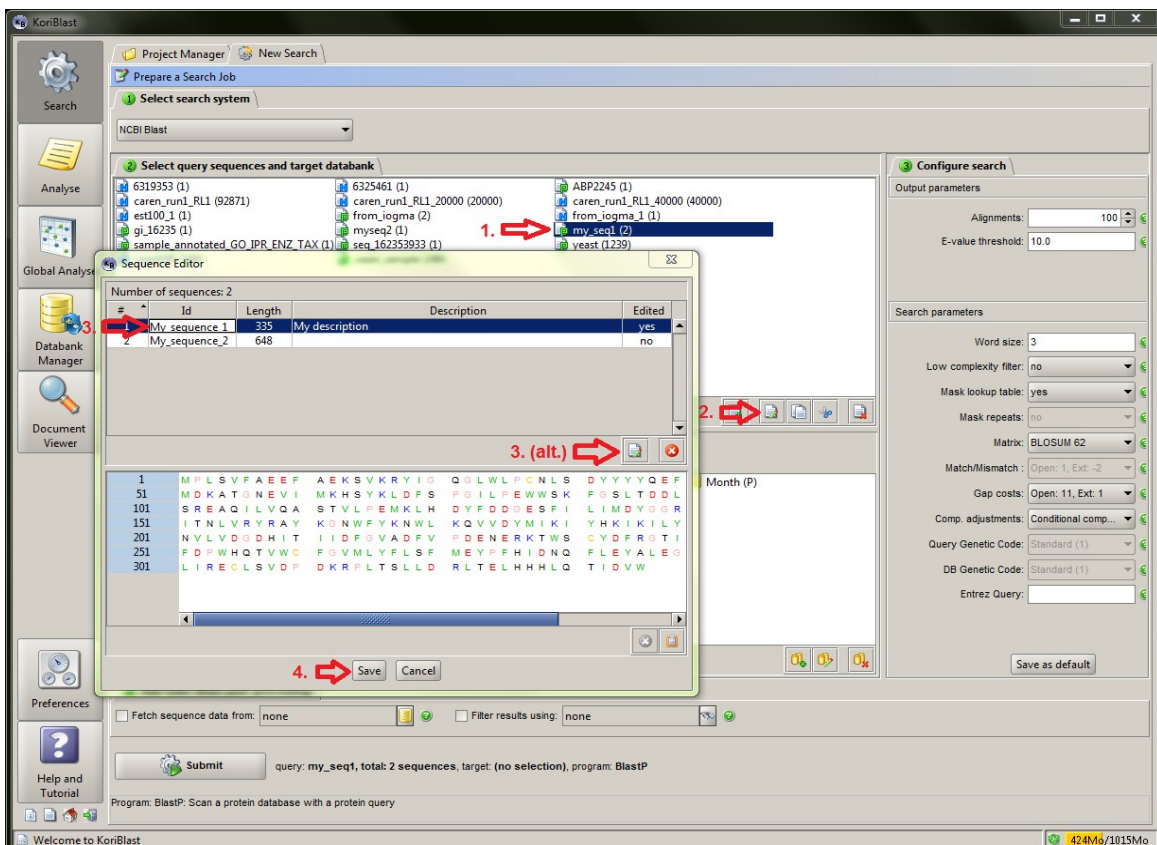
3. Search module - project manager - display size of job folders in the "Job Info" dialogue box.



Project manager: job information window.

4. Search module - new search - Enable editing and batch editing of sequences.

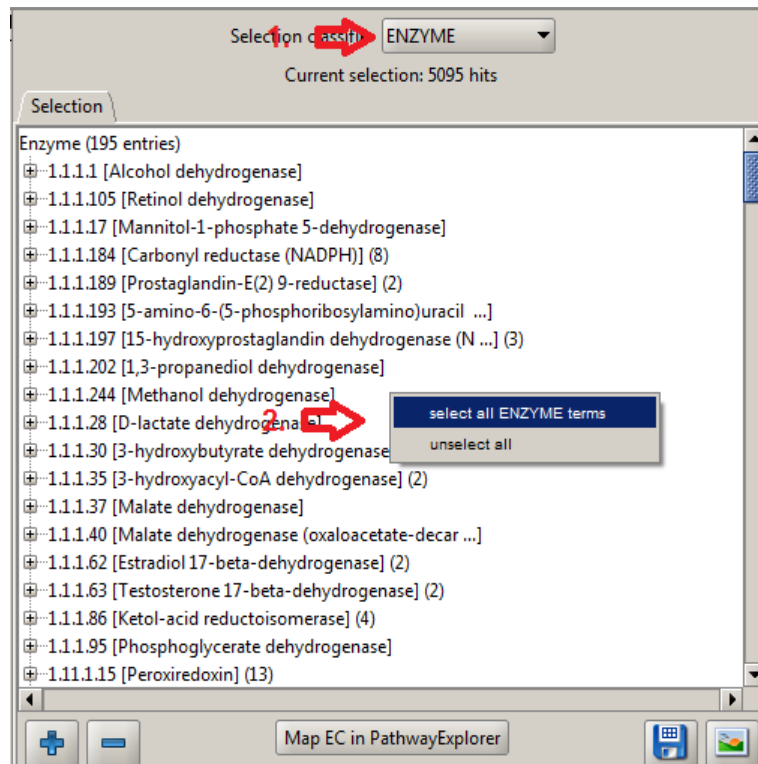
1. Select the sequence file to edit.
2. Click on the "edit" button.
3. Within the sequence identifiers list, double-click in the "Id" or "Description" cell to edit the current sequence identifier or description field.
- 3 (alt.). If you need to edit all sequence identifiers once for all, click on the "Batch rename sequence" button: a specific popup edit window is then displayed.
4. Save modifications.



Search module : single sequence or batch sequences editor.

5. Global Viewer - "PathWay Explorer" gateway - add popup action to select / unselect categories ("Enzyme") within selected hit list.

1. select "ENZYME" category display.
2. right-click within the tree view to display the category selection popup.



Global Viewer - "Pathway Explorer" gateway : selection popup.

6. Document Viewer: from the Databank Explorer module, it is now possible to retrieve Genbank/Genpept formatted data files, in addition to the Fasta format already available.

With that new feature, it is now possible to prepare personal annotated data sets: use the Databank Explorer to get annotated sequences, then the Databank Manager to convert them to data index. In turn, such index can be used to annotate Blast results.

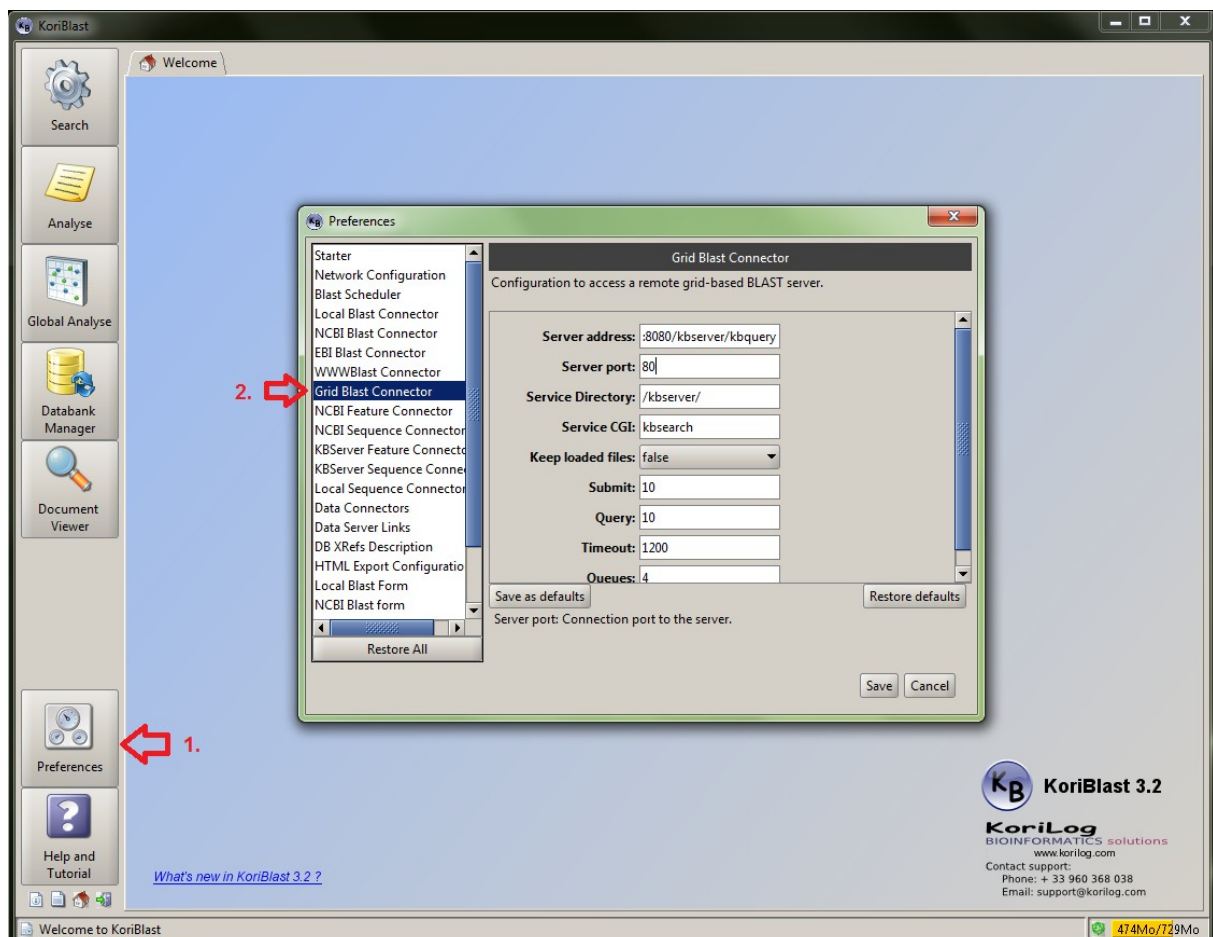
C) Compliance with the new "KoriBlast server"

1. True "client/server" mode blast querying system with "KoriBlast server".

You can send remote blast queries to your grid computing framework in a client/server mode provided the "KoriBlast server" has been deployed on the grid. In this mode, multiple KoriBlast users can concurrently submit blast queries to the grid computing framework. The request queuing system and load balance policy will be fully taken in charge by the "KoriBlast server".

Here are the main steps in order to enable a grid search using "KoriBlast Server" from KoriBlast (please, refer to the specific documentation for the detailed configuration) :

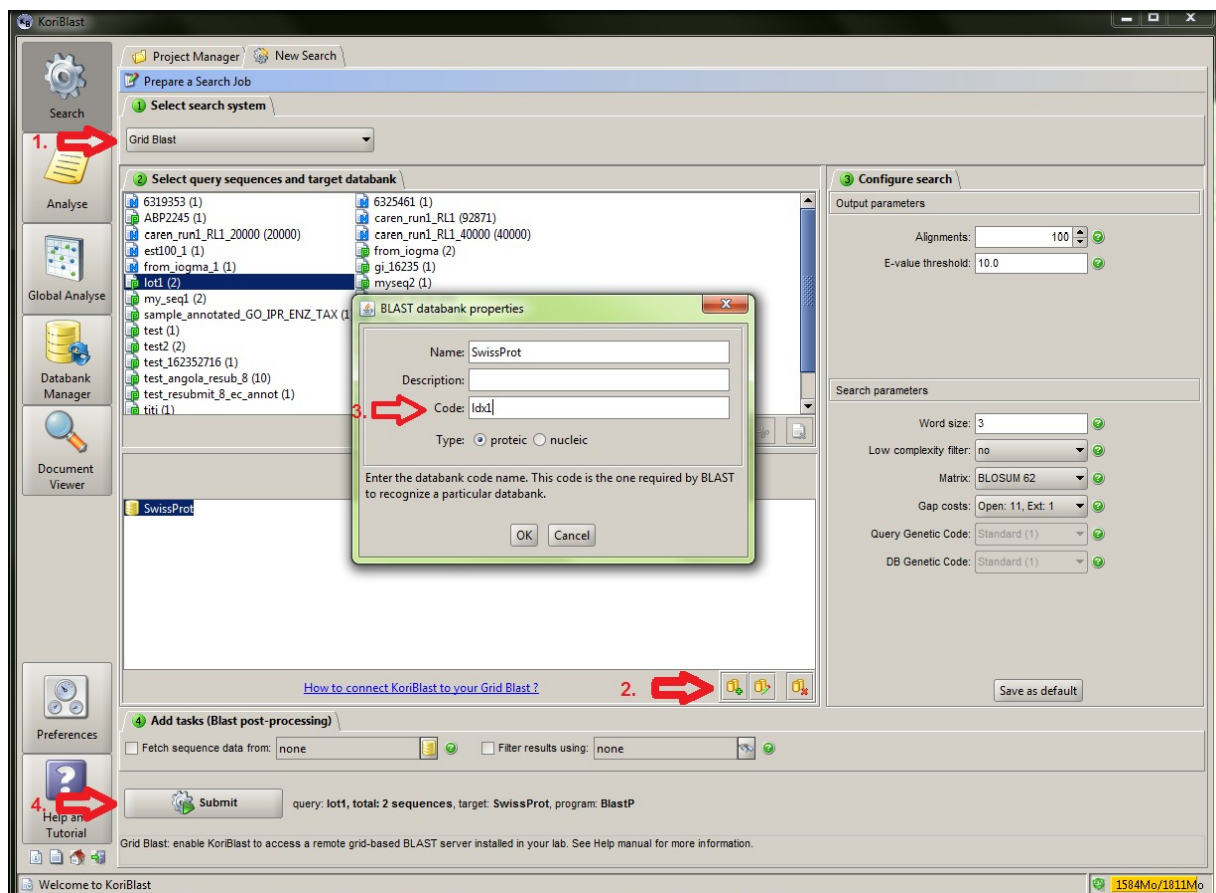
- a. pre-requisite is to have a "KoriBlast server" deployed. You'll need the "KoriBlast server" url when configuring Koriblast.
- b. first step in KoriBlast is to configure a grid Blast connector in order to address queries to the "KoriBlast server".



Configure a grid blast connector for "KoriBlast server".

c. When preparing a new search on the grid system, specify the databank you want the "KoriBlast server" to access in order to process your query.

(please note that you will have to enter the databank code given by the "KoriBlast server" administrator. Please refer to the specific documentation for details).



Specify the databank to be accessed by the "KoriBlast server".

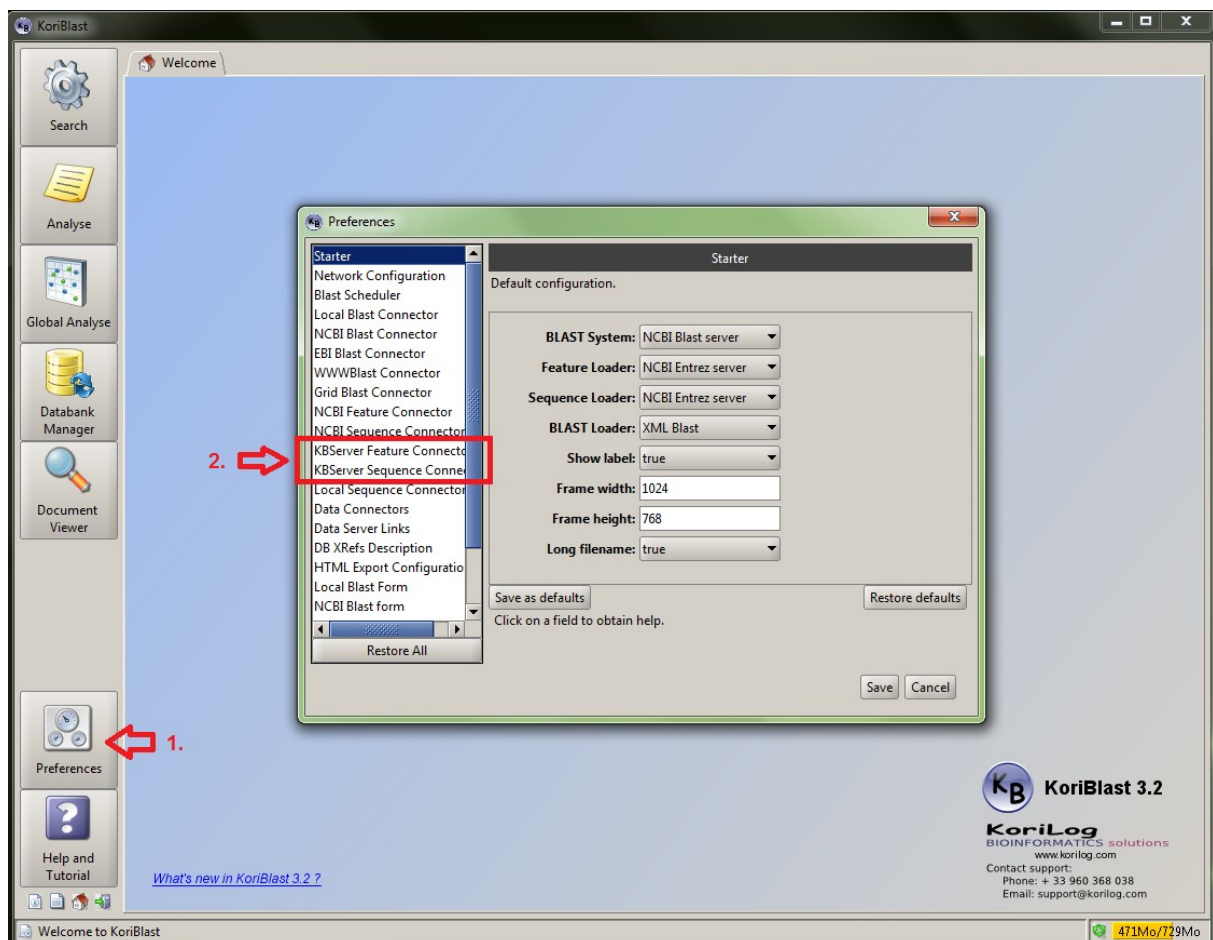
d. Submit the query.

2. Automatic annotation of Blast hits with "KoriBlast server".

The hits returned by any Blast system can be automatically annotated using the "KoriBlast server" data annotation functionality.

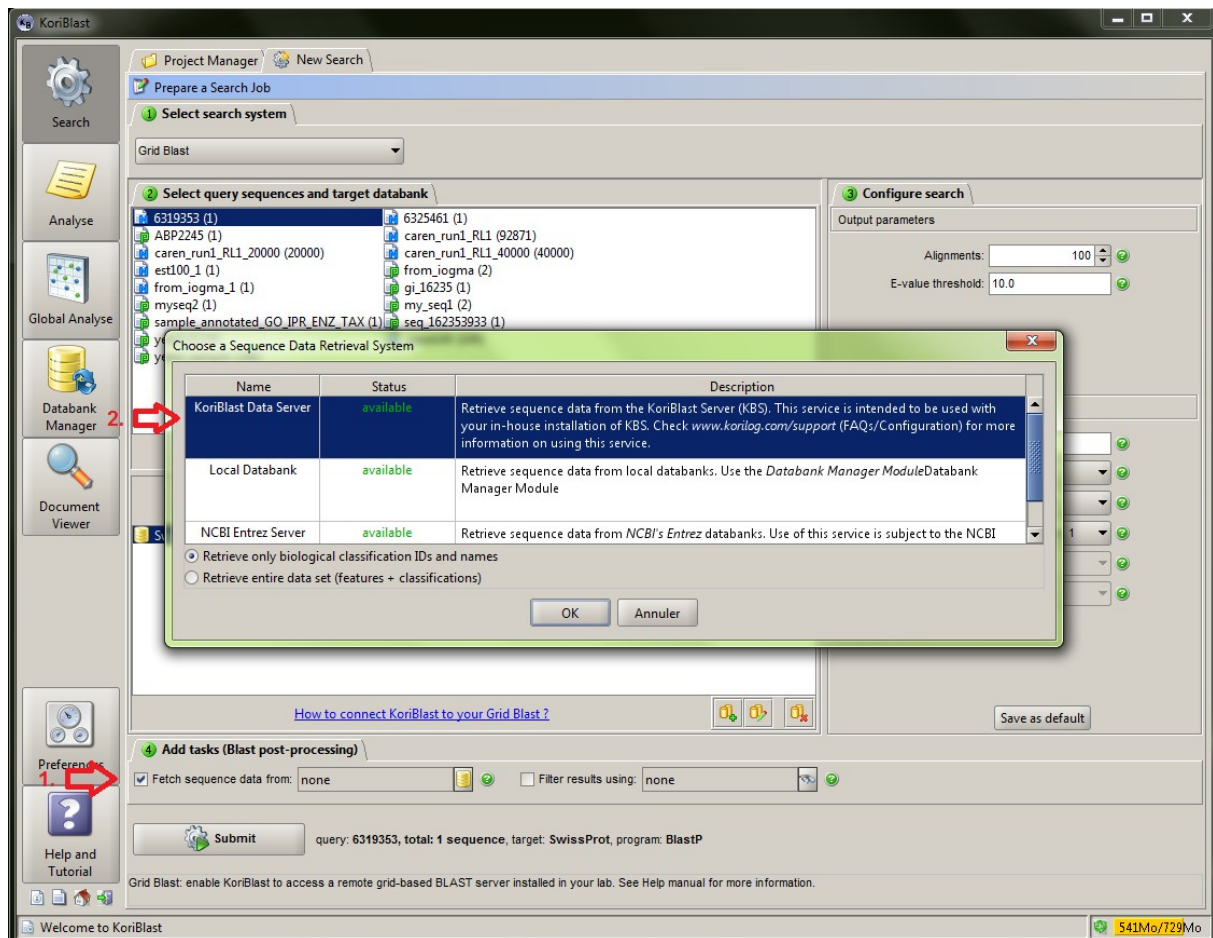
Here are the main steps in order to enable a grid search using "KoriBlast Server" from KoriBlast (please, refer to the specific documentation for the detailed configuration) :

- a. first step is to configure a "KServer feature" connector and a "KServer sequence connector" in order to address annotation queries to the "KoriBlast server".



"KoriBlast server" connectors for data annotation from "KoriBlast server".

- b. When preparing a new search, check the hit data annotation option and select the "KoriBlast Server" data annotation retrieval system.



Select the "KoriBlast server" data annotation retrieval system.

- c. Submit the query.