

KEGG

KEGG is a collection of databases (Figure 1) and connected software tools that integrates genes into pathways linking them functions, diseases, drugs and the like. The KEGG identifiers that are used for the various databases are listed in Figure 2).



Figure 1: Overview of the databases hosted by KEGG. The table is provided via the [overview page](#) from KEGG



Figure 2: Overview of the identifiers used in the various KEGG databases. The table is provided via the [overview page](#) from KEGG

KEGG integrates this information via KEGG orthologs (KO) groups across a comprehensive yet obviously limited set of species. To allow the custom propagation of the information stored in KEGG to sequences from species that are not represented in the KEGG database, they provide the annotation tool [BlastKOala](#) (Kanehisa et al. 2016). In a nutshell, BlastKOala makes heavy use of phylogenetic relationships among sequences in order to propagate functional annotation of so called KEGG ortholog groups (KO) to further sequences. It is a Blast based tool that additionally compares domain architectures between the sequences in a KO group and a new candidate.



Figure 3: Workflow of the two tools for assigning KO numbers to an unannotated protein sequence. While KAAS is relying on orthology assignments, BlastKOala recruits unidirectional Blast searches.

From:
<https://applbio.biologie.uni-frankfurt.de/teaching/wiki/> - Teaching

Permanent link:
<https://applbio.biologie.uni-frankfurt.de/teaching/wiki/doku.php?id=general:bioseqanalysis:functionassignment:kegg:blastkoala>

Last update: 2022/03/11 08:08

