Comparison of BioCyc, UniProt, and KEGG SRI International March 2025

This document compares the BioCyc, UniProt, and KEGG websites according to both their data content and their software tools.

Most of the comparisons simply assess which sites contain a given type of data, or which sites include a given type of software tool. In some cases a quantitative comparison of the amount of data present is provided.

Versions being compared: BioCyc version 29.0 March 2025 KEGG version 113.0+ March 2025 UniProt version March 2025

KEGG[®] is a U.S. registered trademark owned by Minoru Kanehisa and is not affiliated, endorsed, or sponsored in any way by SRI International. BioCyc TM and MetaCyc TM are trademarks owned by SRI International.

Software Tool Comparison

Tables 1 and 2 compare the software tools provided by the websites.

	BioCyc	UniProt	KEGG
Genome Informatics Tools:			
Linear genome browser	\checkmark		\checkmark
Circular genome browser	\checkmark		
Comparative genome browser	\checkmark		
Regulatory network browser	\checkmark		
Gene/protein/RNA advanced search	\checkmark	\checkmark	
Gene/protein/RNA information pages	\checkmark	\checkmark	\checkmark
BLAST search	\checkmark	protein only	\checkmark
Sequence pattern search	\checkmark	protein only	
Multiple sequence alignment	\checkmark	\checkmark	
Retrieve amino-acid sequence	\checkmark	\checkmark	\checkmark
Retrieve nucleotide sequence	\checkmark		\checkmark
Pathway Informatics Tools:			
Pathway/reaction/metabolite advanced search	\checkmark		
Pathway/reaction/metabolite information pages	\checkmark		\checkmark
Pathway diagrams	\checkmark		\checkmark
Zoomable metabolic network diagrams	\checkmark		\checkmark
Metabolic network explorer	\checkmark		
Search for optimal metabolic route from starting to	\checkmark		\checkmark
ending metabolite			
Additional Tools:			
Visualize promoters, operons, TF binding sites, RBSs	\checkmark		
Visualize regulon of transcription factor	\checkmark		
SmartTables manipulate gene and metabolite sets	✓		
Comparative analysis tool suite	\checkmark		

Table 1

	BioCyc	UniProt	KEGG
Transcriptomics Data Analysis Tools			
Enrichment analysis for metabolic pathways and GO terms	\checkmark		
Paint transcriptomics data onto pathway diagrams	\checkmark		\checkmark
Paint transcriptomics data onto zoomable metabolic network	\checkmark		\checkmark
Animate display of transcriptomics data on metabolic network	\checkmark		
Omics Dashboard	\checkmark		
Metabolomics Data Analysis Tools			
Metabolic pathway enrichment analysis	\checkmark		
Paint metabolomics data onto pathway diagrams	\checkmark		\checkmark
Paint metabolomics data onto zoomable metabolic network	\checkmark		\checkmark
Omics Dashboard	\checkmark		
Metabolite translation service	\checkmark		
Pathway covering sets	\checkmark		
Search by monoisotopic mass, chemical formula, InChI	\checkmark		
Search by chemical substructure	\checkmark		\checkmark

Table 2

Data Content Comparison

Table 3 lists the different types of data content present in the databases, in some cases noting the amount of data content.

	BioCyc	UniProt	KEGG
Genomes	20,050		10,713
Genes	\checkmark		\checkmark
RNAs	\checkmark		
Protein monomers	\checkmark	\checkmark	\checkmark
Protein complexes	\checkmark		
Reactions	19,692	\checkmark	12,242
Pathways	3,231		453 modules
Metabolites	20,072		19,460
Transcriptional regulatory networks	11*		
Gene essentiality data	\checkmark		
Phenotype Microarray data	\checkmark		
Enzyme activators, inhibitors, cofactors	\checkmark	\checkmark	
Protein localization	\checkmark	\checkmark	
Protein features (e.g., active sites)	\checkmark	\checkmark	
Kinetic parameters, pHopt, Tempopt	\checkmark	\checkmark	
Protein 3D structures			\checkmark
Gene Ontology annotations	\checkmark	\checkmark	
Evidence codes, citations justify annotations	\checkmark	\checkmark	
Curator-authored mini-reviews	\checkmark	\checkmark	unknown
Literature citations / publications curated	153,000	unknown	unknown

Table 3: Genomes: Includes non-viral genomes only. Click <u>here</u> for a list of BioCyc curated organism databases.

* BioCyc contains 11 organisms with significant transcriptional regulatory network data.

BioCyc contains 1.9 times as many genomes as does KEGG and 1.6 times as many reactions as does KEGG.

The large difference in the number of metabolic pathways between BioCyc and KEGG is particularly important. Both groups predict metabolic pathways from sequenced genomes by recognizing the presence of known pathways from their respective reference pathway databases (for BioCyc, that database is called MetaCyc). Because MetaCyc has 7.1 times as many pathways as does the KEGG reference database of modules, BioCyc is capable of recognizing 7.1 times as many pathways as is KEGG.

Using KEGG for Enrichment Analysis

Enrichment is a common technique for analyzing sets of differentially expressed genes from transcriptomics data. The method tries to find cases where the differentially expressed genes contain more genes from a given pathway than expected by chance, thus identifying differentially expressed pathways. If the following KEGG pathway were found to be highly enriched for a given set of genes, which of these 21 pathways would we say has been activated? The answer is that we don't know – the point being that mixing so many pathways into one diagram renders the diagram essentially unusable for enrichment analysis.



KEGG Map map00260 Contains 21 Pathways!

Fig. 1 KEGG map00260. The biological processes present in this map are listed along the right. Each process name is color coded, and those colors identify that process within the map. Each colored square is positioned next to a reaction within that process. KEGG diagram from [21], downloaded

© 2022 SRI International

Example BioCyc Visualization Tools



BioCyc: Gene Expression Data on Single Pathway

ECOCYC



BioCyc: Metabolic Network Diagram and Pathway Diagram Painted with Metabolomics Data



BioCyc Omics Dashboard with Gene Expression Data

- A series of panels summarize omics data for different cellular systems
- Each panel contains a set of **plots** (subsystems)
- Large dots average measurements

© 2022 SRI International

 Drill down to individual genes (far right)

