

# A Comparison of the BioCyc™ and KEGG® Pathway Databases and Web Portals

March 2023

Peter D. Karp  
SRI International®  
pkarp@ai.sri.com



# Introduction

- This document compares the KEGG and [BioCyc](#) pathway databases and websites
- Main criteria for comparison are:
  - Database content
  - Software tools

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

# Number of Genomes and Curation Levels

	<b>BioCyc version 26.5</b> December 2022	<b>KEGG v104.0+</b> December 2022
<b>Genome Databases*</b>	<b>20,028</b>	<b>8,611</b>
<b>Curated Genome Databases</b>	<b>69**</b>	<b>***</b>
<b>Curated Publications****</b>	<b>138,452</b>	<b>74,938</b>

\* Includes non-viral genomes only

\*\* For list of BioCyc curated organism databases [click here](#)

\*\*\*Unknown


\*\*\*\*We assume that all publications cited by KEGG and BioCyc have been curated. This is true for BioCyc; it may or may not be true of KEGG.

BioCyc genome-specific databases undergo substantial manual literature-based curation to correct computational inferences and add additional information from the scientific literature, such as experimentally determined gene functions

# MetaCyc Curated Metabolic DB Compared to KEGG

- MetaCyc is a reference metabolic pathway database containing pathways from all domains of life
- MetaCyc, like KEGG Reference, is the source of pathways predicted in individual organisms – more reference pathways means more pathways can be computationally predicted in genome databases
- Mini-reviews are multi-paragraph summaries of pathways and gene functions authored by curators. The numbers given are the sum of the length in characters of all mini-reviews in each database.

	<b>MetaCyc version 26.5</b> December 2022	<b>KEGG v104.0+</b> December 2022
<b>Pathways</b>	<b>3,085</b>	<b>425 metabolic modules</b>
<b>Reactions</b>	<b>18,391</b>	<b>11,860</b>
<b>Metabolites</b>	<b>18,785</b>	<b>19,019</b>
<b>Mini-reviews (textbook pages)</b>	<b>10,392</b>	<b>1,557</b>



# Summary of BioCyc/KEGG Comparison

December 2022

- BioCyc has more data (2.3x genomes, 7.3x pathways, 1.5x reactions)
- BioCyc has more accurate data (curated from 138,000 publications)
- BioCyc has broader types of data
- BioCyc has many more informatics tools

# Comparison of Data Content – March 2023

- Each database contains a number of additional types of data – some is curated and some is computationally predicted

	BioCyc	KEGG
Genes, proteins	✓	✓
Reactions	✓	✓
Metabolic pathways	✓	✓
Signaling pathways		✓
Metabolites	✓	✓
Enzyme activators, inhibitors, cofactors, kinetic constants, localization	✓	
Protein features	✓	
Protein subunit composition	✓	
Protein 3-D structures		✓
Gene Ontology terms	✓	
Evidence codes	✓	
Reaction atom mappings	✓	
Gene essentiality data	✓	
Phenotype Microarray data	✓	
<u>Transcriptional regulatory networks</u>	11 organisms	0 organisms
Diseases, drugs		✓

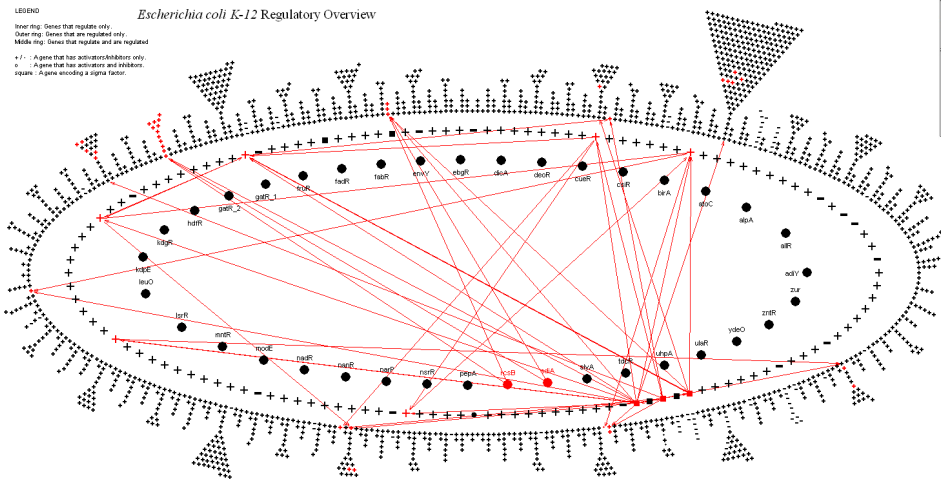
# Comparison of Informatics Tools – March 2023

- This comparison does not consider KEGG software tools that are not present in BioCyc, such as KEGG genome annotation
- Clicking on hyperlinks in this table will show an example of the tool

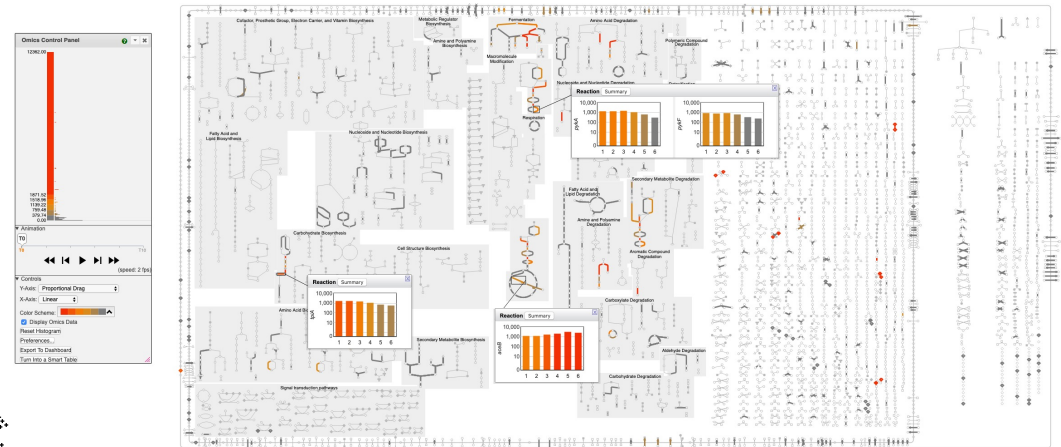
	BioCyc	KEGG
Genome browser	✓	✓
Genome browser depicts sequence, regulatory elements, customizable tracks	✓	
<a href="#">Comparative genome browser</a>	✓	
<a href="#">Regulatory network browser</a>	✓	
SmartTables	✓	
<a href="#">Advanced search tools</a>	✓	
BLAST search, sequence pattern search	✓	✓
Multiple sequence alignments	✓	
<a href="#">Depicts substrate-level and genetic regulation</a>	✓	
<a href="#">Metabolic Network Explorer</a>	✓	
<a href="#">Metabolic Route Search</a>	✓	✓
<a href="#">Comparative analysis tool suite</a>	✓	

# Example BioCyc Visualization Tools

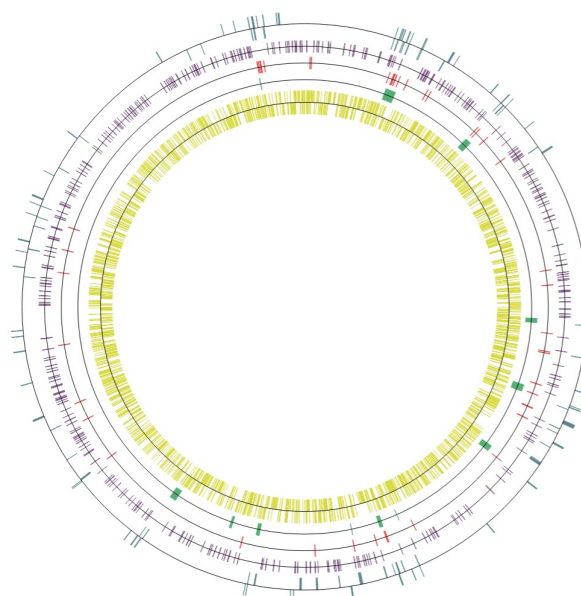
## Regulatory Network Viewer



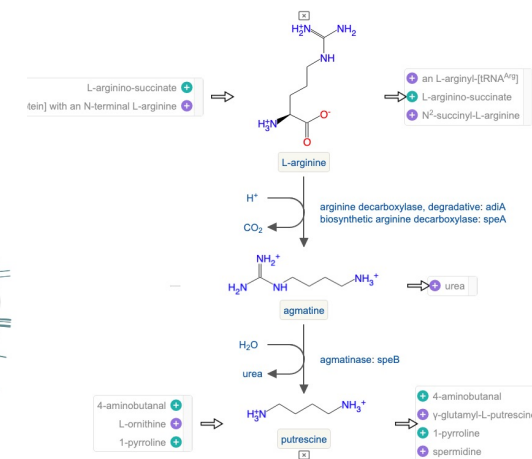
## Metabolic Network Viewer with Omics Data



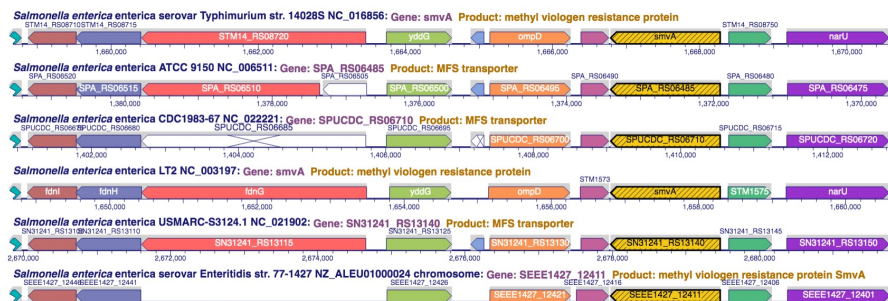
## Circular Genome Viewer



## Metabolic Network Explorer



## Comparative Genome Browser





# Gene Expression Data Analysis Tools – March 2023

	BioCyc	KEGG
Paint Gene Expression Data onto Pathway Diagram	✓	✓
Paint Gene Expression Data onto Pathway Collage*	✓	
<a href="#">Paint Gene Expression Data onto Zoomable Metabolic Network Diagram</a>	✓	✓
<a href="#">Omics Dashboard</a>	✓	
Enrichment analysis for metabolic pathways and GO terms	✓	

- A pathway collage is a multi-pathway diagram where the user chooses what pathways to include

# Metabolomics Data Analysis Tools – March 2023

	BioCyc	KEGG
Metabolic Pathway Enrichment Analysis	✓	
Paint Metabolomics Data onto Pathway Diagram	✓	✓
<a href="#">Paint Metabolomics Data onto Zoomable Metabolic Network Diagram</a>	✓	✓
Omics Dashboard	✓	
<a href="#">Metabolite Translation Service</a>	✓	
<a href="#">Pathway Covering Sets</a>	✓	
<a href="#">Search by Monoisotopic Mass</a>	✓	
Search by Molecular Weight	✓	
Search by Chemical Formula	✓	
Search by Chemical Substructure	✓	✓
Search by InChI or InChI Key	✓	

P.D. Karp et al, "Computational Metabolomics Operations at BioCyc.org,"

[Metabolite](#) 5:291-310 2015

# BioCyc: Gene Expression Data on Single Pathway



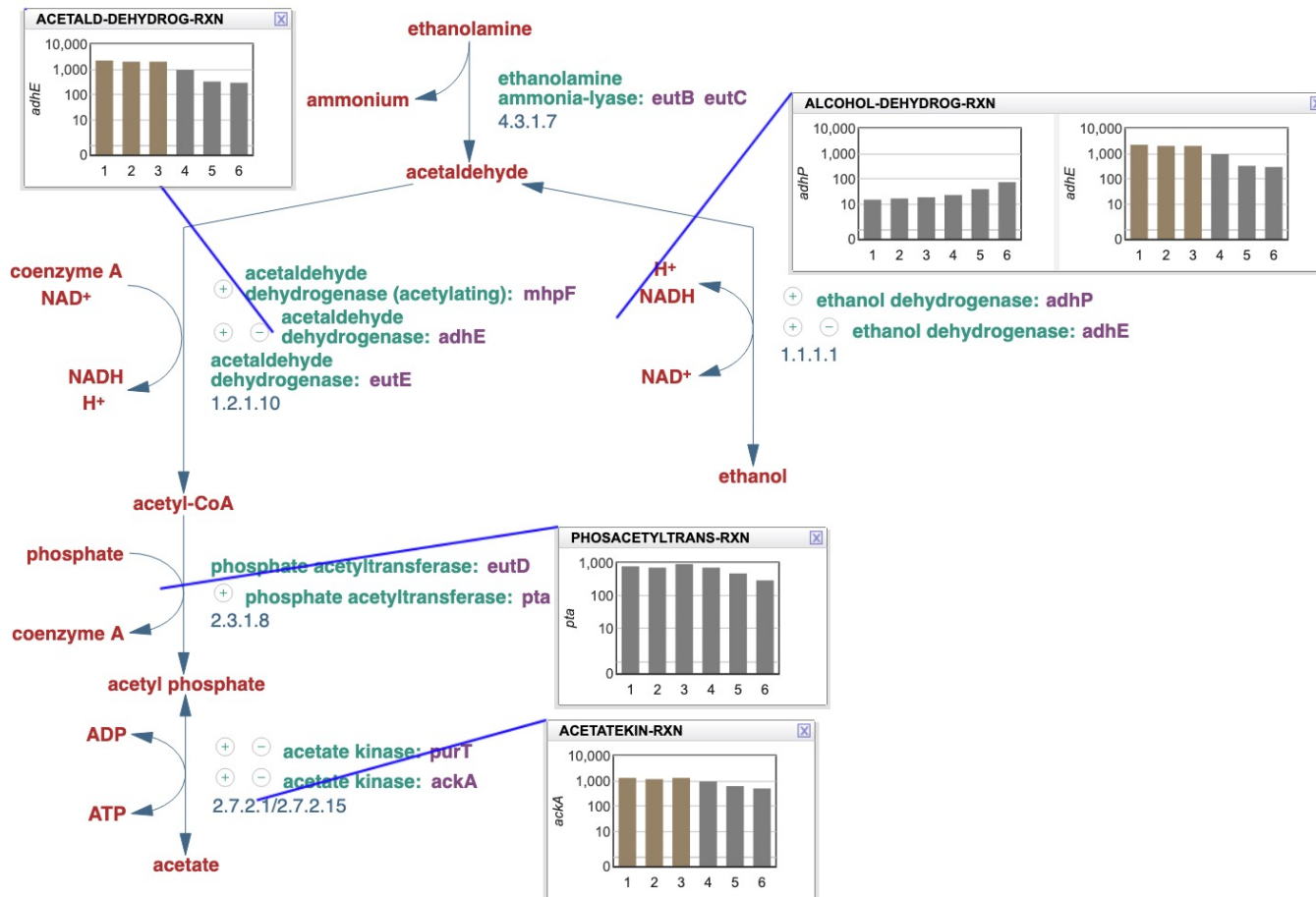
Sites Search Genome Metabolism Analysis SmartTables Help

Provide Feedback

Add to SmartTable

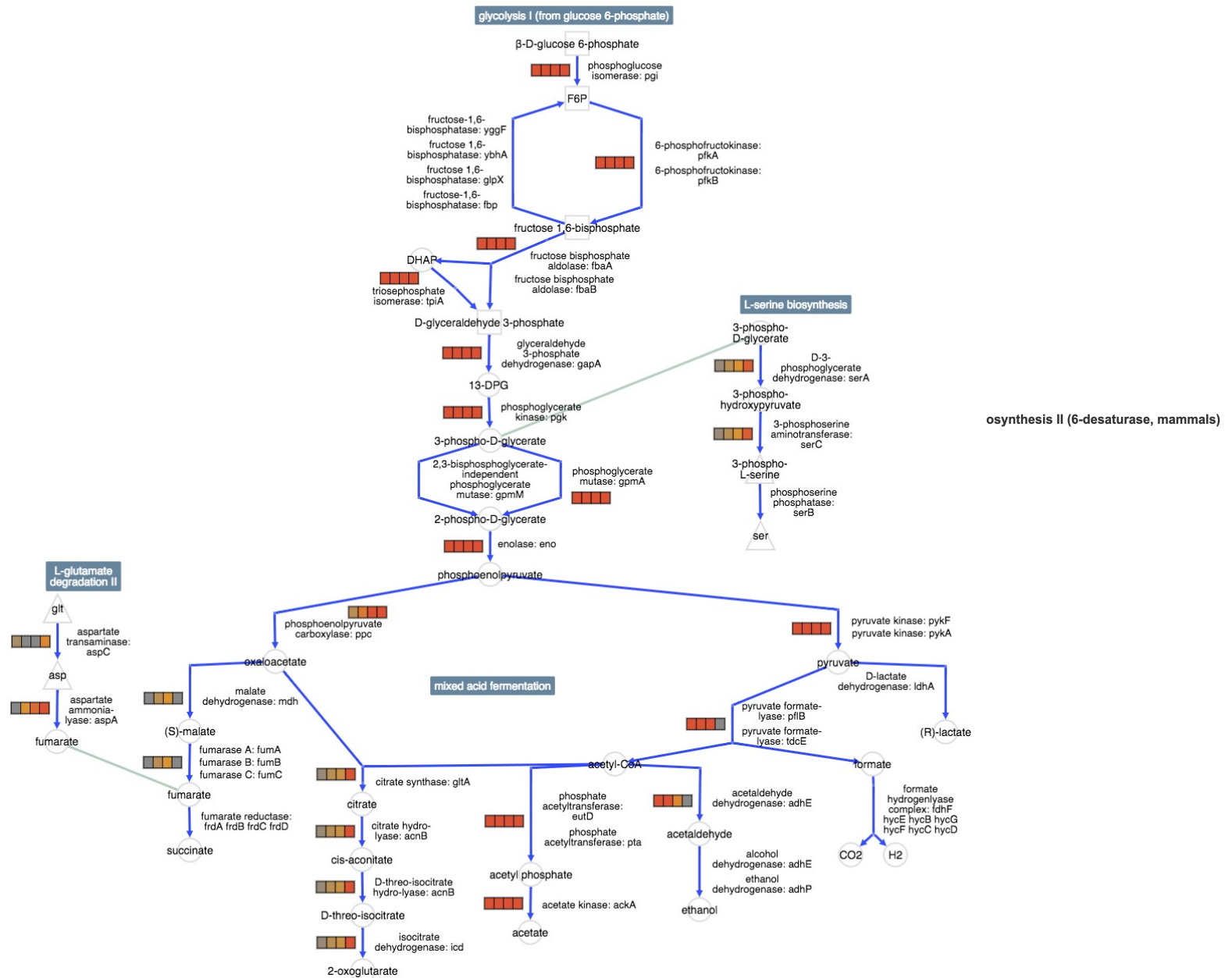
*Escherichia coli* K-12 substr. MG1655 Pathway: ethanolamine utilization

Detail Level: All compounds, enzymes Show Regulation Details



(base, mammals)

# BioCyc: Pathway Collage with Gene Expression Data

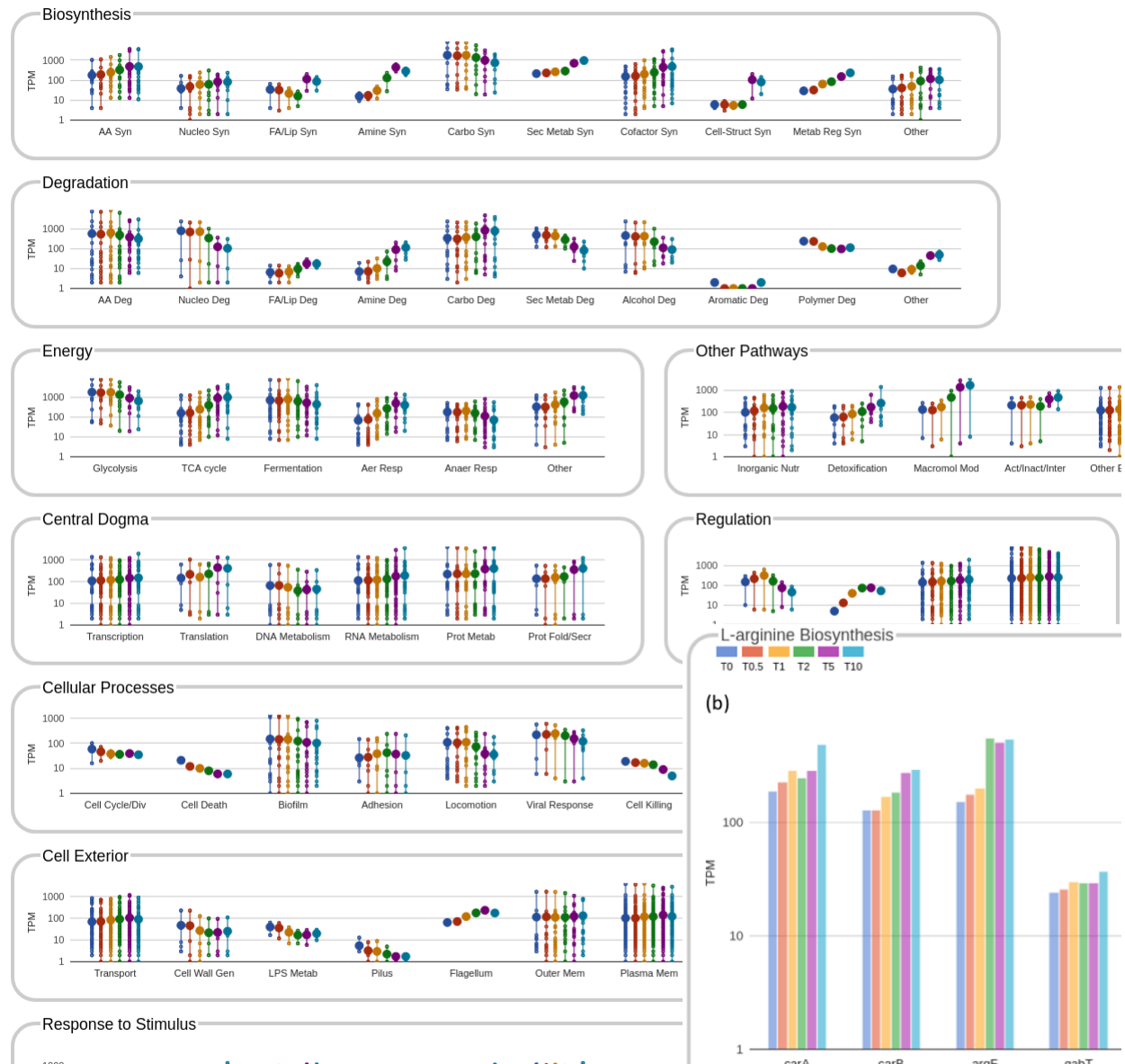


# BioCyc Omics Dashboard with Gene Expression Data

- A series of **panels** summarize omics data for different cellular systems
- Each panel contains a set of **plots** (subsystems)
- Large dots **average** measurements
- Drill down to individual genes (far right)

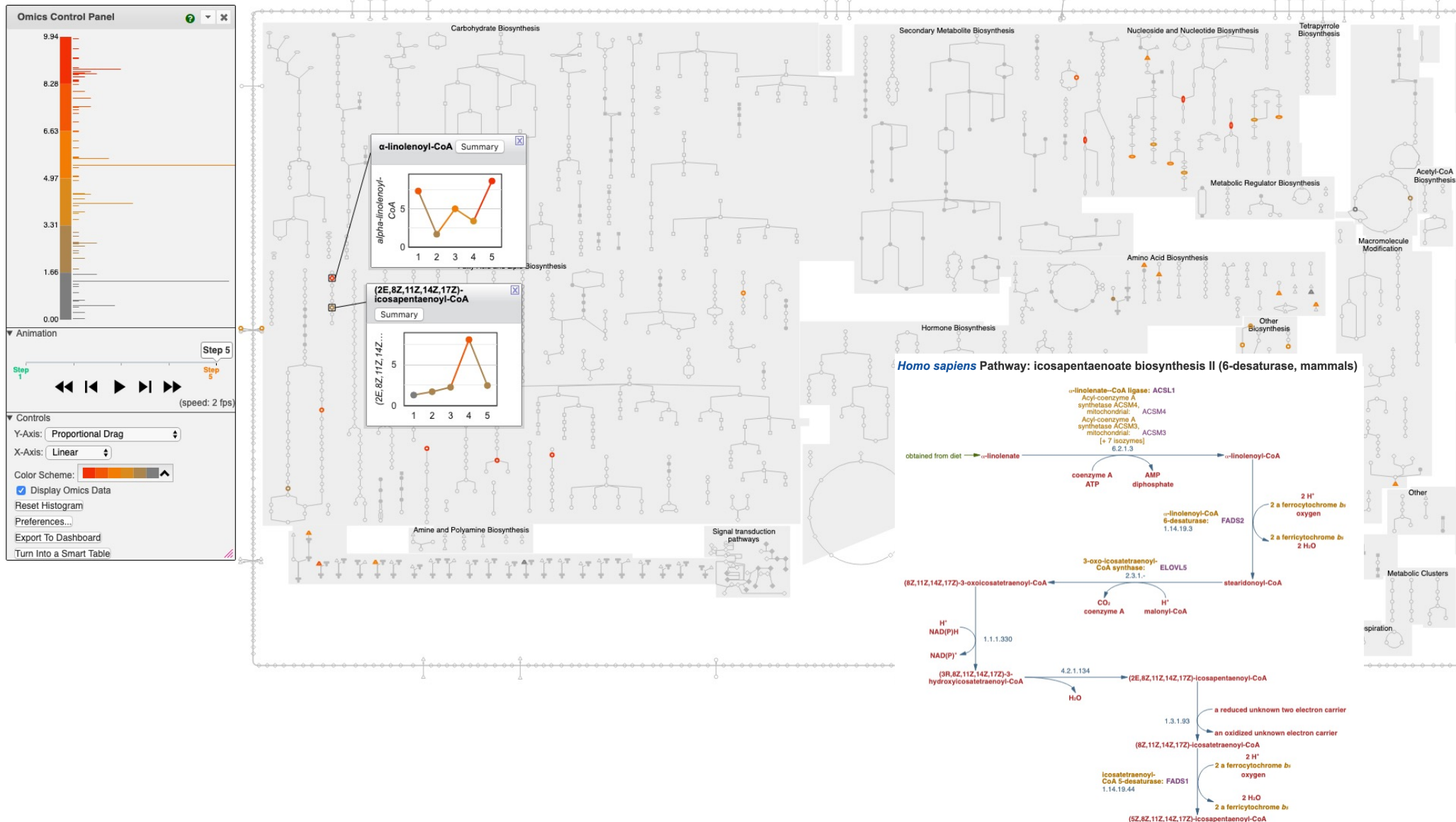
Pathway Tools Omics Dashboard for *Escherichia coli* K-12 substr. MG1655  
GSE71562 Anaerobic-Aerobic transition, significant genes only.

T0 T0.5 T1 T2 T5 T10



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

*Homo sapiens*





# BioCyc SmartTables for Metabolomics

- Collect and save lists of database objects
  - Metabolites, pathways, genes, sequence regions, ...
- Import from files, explore interactively
- Filter and combine (union, intersection, subtraction)
- Transform them into related objects (eg: metabolite list → pathway list)
- Share with public or specific collaborators, publish
- Pathway enrichment analysis for metabolites

# Using SmartTables: Browsing Database Attributes

Sites ▾ Search ▾ Genome ▾ Metabolism ▾ Analysis ▾ SmartTables ▾ Help ▾

[SmartTables directory](#) [SmartTables Help](#)

## SmartTable: Human Metabolomics Synthetic

[Click to add description](#)

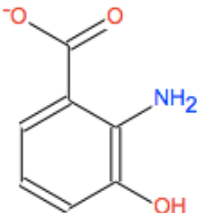
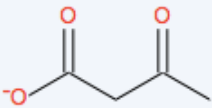
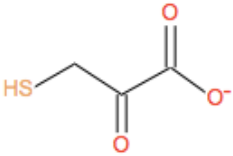
66 rows of compounds from *H. sapiens*

Owner: [Peter Karp](#), Created: 25-Mar-2014 10:14:02

ADD TRANSFORM COLUMN		ADD PROPERTY COLUMN		ENRICHMENTS	
choose a transform... ?		Database Links ?		choose an enrichment... ?	
Show paged Show all					
<input type="checkbox"/>	column 1	ChEBI	HMDB	KEGG	
<input type="checkbox"/> 1	<a href="#">3-hydroxyanthranilate</a>	36559		C00632	
<input type="checkbox"/> 2	<a href="#">acetoacetate</a>	13705		C00164	
<input type="checkbox"/> 3	<a href="#">3-mercaptopyruvate</a>	57678		C00957	
<input type="checkbox"/> 4	<a href="#">3-oxo-cholyl-CoA</a>				
<input type="checkbox"/> 5	<a href="#">3-phospho-hydroxypyruvate</a>	18110		C03232	
<input type="checkbox"/> 6	<a href="#">3-pyridylacetate</a>				
<input type="checkbox"/> 7	<a href="#">5-amino-1-(5-phospho-B-D-ribose)imidazole</a>	2655		C03373	
	<a href="#">(5Z)-1,5C)-11-<math>\alpha</math>-hydroxy-9,15-dioxohepta-1,2-enate</a>	57400		C04707	



# Using SmartTables: Browsing Database Attributes

	column 1	SMILES	Structure of compound
1	3-hydroxyanthranilate	<chem>C1(C=C(C(N)=C(O)C=1)C([O-])=O)</chem>	
2	acetoacetate	<chem>CC(=O)CC([O-])=O</chem>	
3	3-mercaptopyruvate	<chem>C(C(C=O)[O-])=O)S</chem>	

# More Information on BioCyc

- BioCyc subscriptions
  - <https://biocyc.org/Product-summary.shtml>
  - [biocyc-sales@biocyc.org](mailto:biocyc-sales@biocyc.org)
- BioCyc publication
  - [“The BioCyc collection of microbial genomes and metabolic pathways”](#), *BMC Bioinformatics*