



# *Recent Developments in the MetaCyc Database*

Ron Caspi

# *Introduction*

# *MetaCyc Scope*

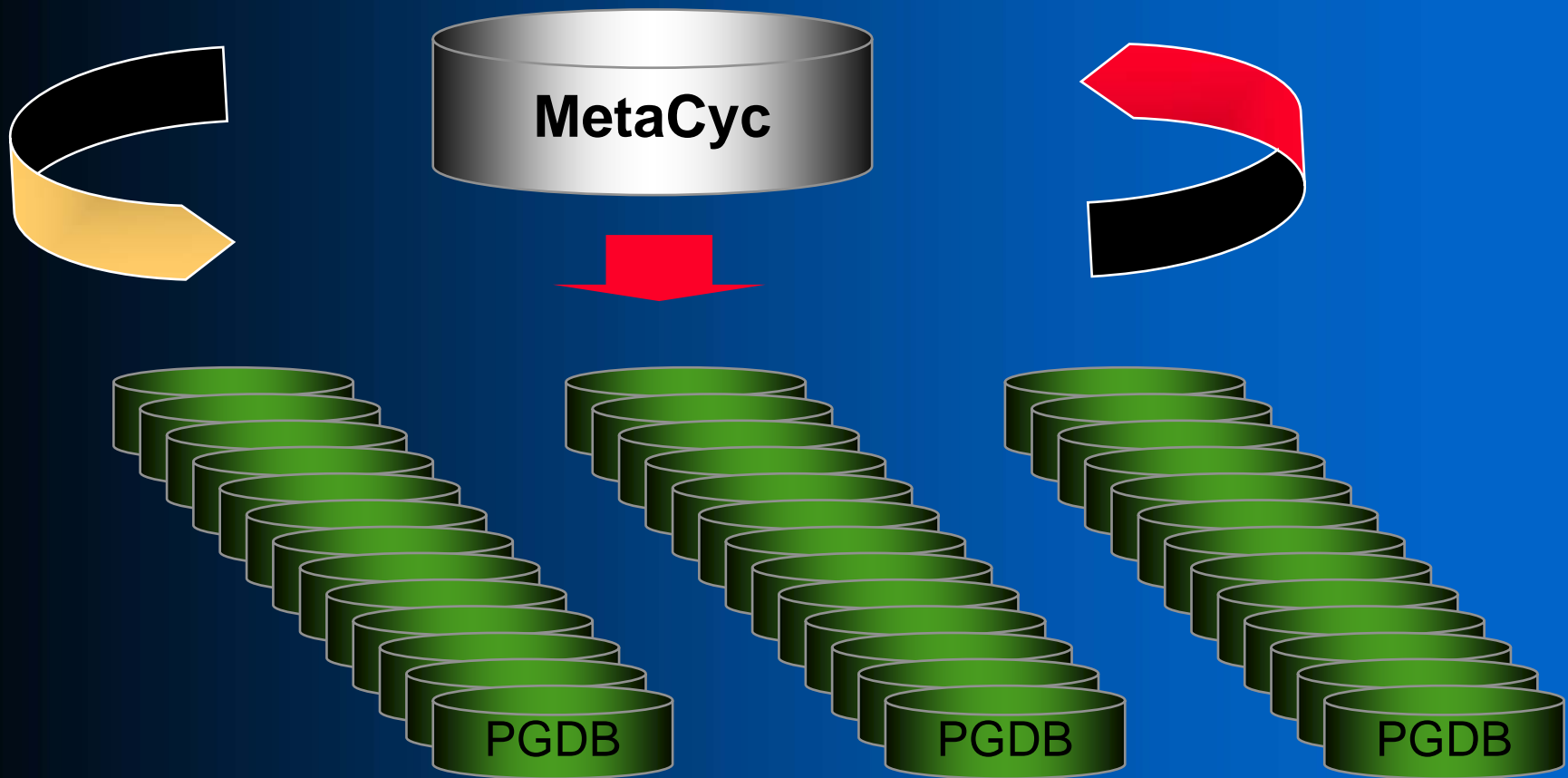
- MetaCyc is a collection of metabolic pathways and enzymes from a wide variety of organisms, primarily microorganisms and plants
- The goal of MetaCyc is to contain a representative sample of every experimentally elucidated pathway, thereby cataloging the universe of known pathways and enzymes.

# *MetaCyc Curation*

- Continuous curation by Ph.D level curators
  - Information gathered from biological and biomedical literature
- An update is released every three months

<http://biocyc.org/metacyc/release-notes.shtml>

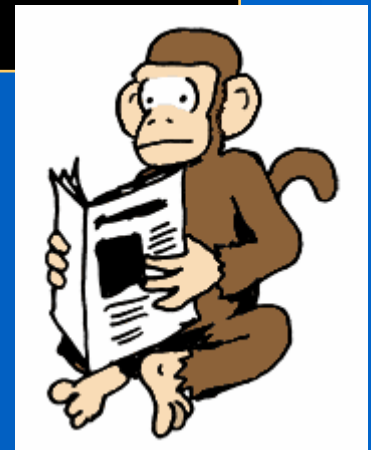
# *MetaCyc Is Used to Create Computationally-Derived Organism-Specific Databases*



# *Curation Conventions*

- Curate pathway enzymes for species with most experimental data
- Add enzymes from additional species if compelling reason exists (e.g., different subunit structure)
- Create pathway variants if reactions of variant differ from existing pathway
- Species list for pathway is often not exhaustive

A detailed description of the curation conventions is found in the [Curator's Guide](#)



# *Criteria for Defining Pathway End Points*

- End points should be stable compounds.
- Biosynthetic pathways should begin with one of the 13 intermediates of central metabolism from which all biosyntheses begin.
- A pathway (or a transport reaction) link should be created to indicate the pathway that produces the precursor metabolite.
- Degradative pathways that produce an intermediate of central metabolism should stop at that point.
- If appropriate, a pathway link should be created to indicate the pathway that processes the resulting metabolite.
- Very large or complex pathways should usually be defined as superpathways that combine several smaller base pathways divided at breakpoints.

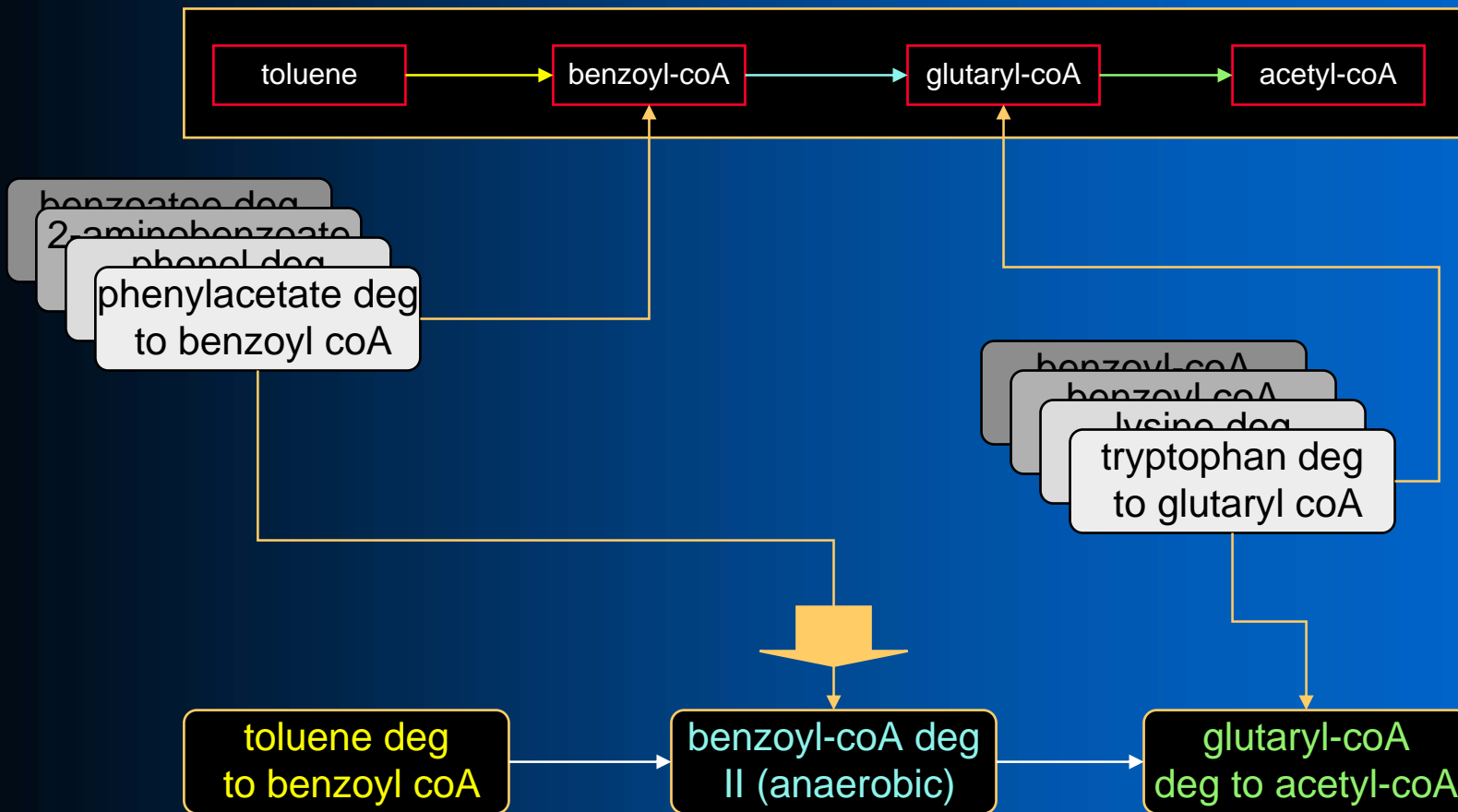
# *The 13 intermediates of central metabolism from which all biosyntheses begin*

- pyruvate
- acetyl CoA
- oxaloacetate
- triose phosphate
- 3-phosphoglycerate
- phosphoenolpyruvate
- ribose-5-phosphate
- $\alpha$ -ketoglutarate
- succinyl CoA
- erythrose-phosphate
- fructose-6-phosphate
- glucose-6-phosphate
- sedoheptulose-7-phosphate



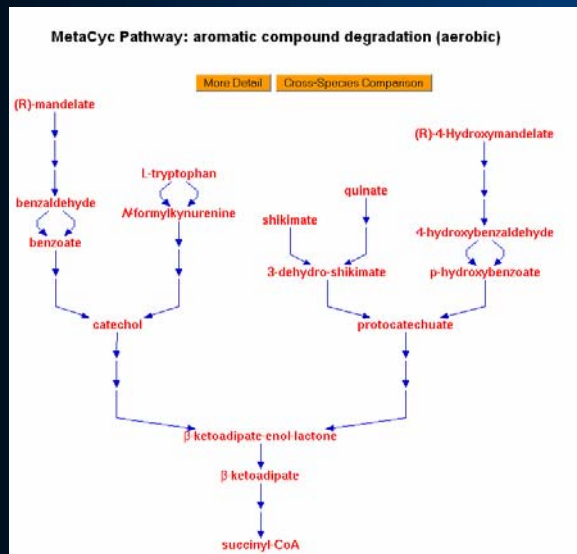
# Modular Pathway Design

toluene degradation VI (anaerobic)



# Different Kinds of Superpathways

- Groups of pathways linked by their substrates
- Super-pathways are defined by their component pathways
- Multiple levels of super-pathways can be defined
- Pathway prediction algorithms accommodate super-pathways



We currently have over 65 superpathways in MetaCyc

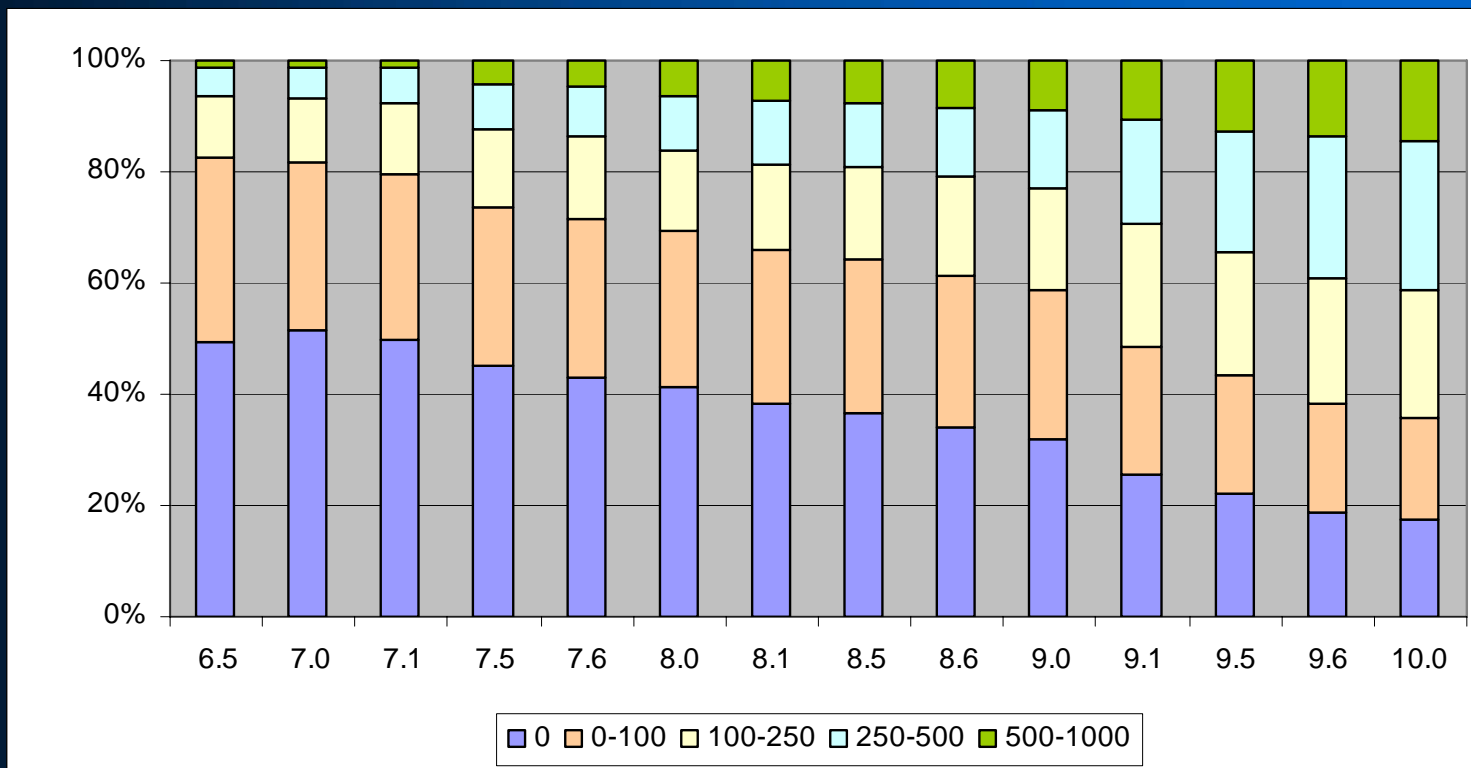
# *Data*

# Curation Progress

	Feb 2002	June 2006	Difference
Pathways	<b>449</b>	<b>785</b>	+ 336 (75%)
Pathways with comments	<b>161</b>	<b>618</b>	+ 457 (283%)
Reactions	<b>4214</b>	<b>5869</b>	+ 1655 (39%)
Enzymes	<b>1147</b>	<b>3446</b>	+ 2299 (200%)
Enzymes with comments	*	<b>3146</b>	+ 3146
Chemical compounds	<b>2339</b>	<b>5177</b>	+ 2838 (121%)
Chemical structures	<b>1858</b>	<b>4658</b>	+ 2800 (150%)
Organisms	<b>158</b>	<b>711</b>	+ 553 (350%)
Citations	<b>2597</b>	<b>10039</b>	+7442 (286%)

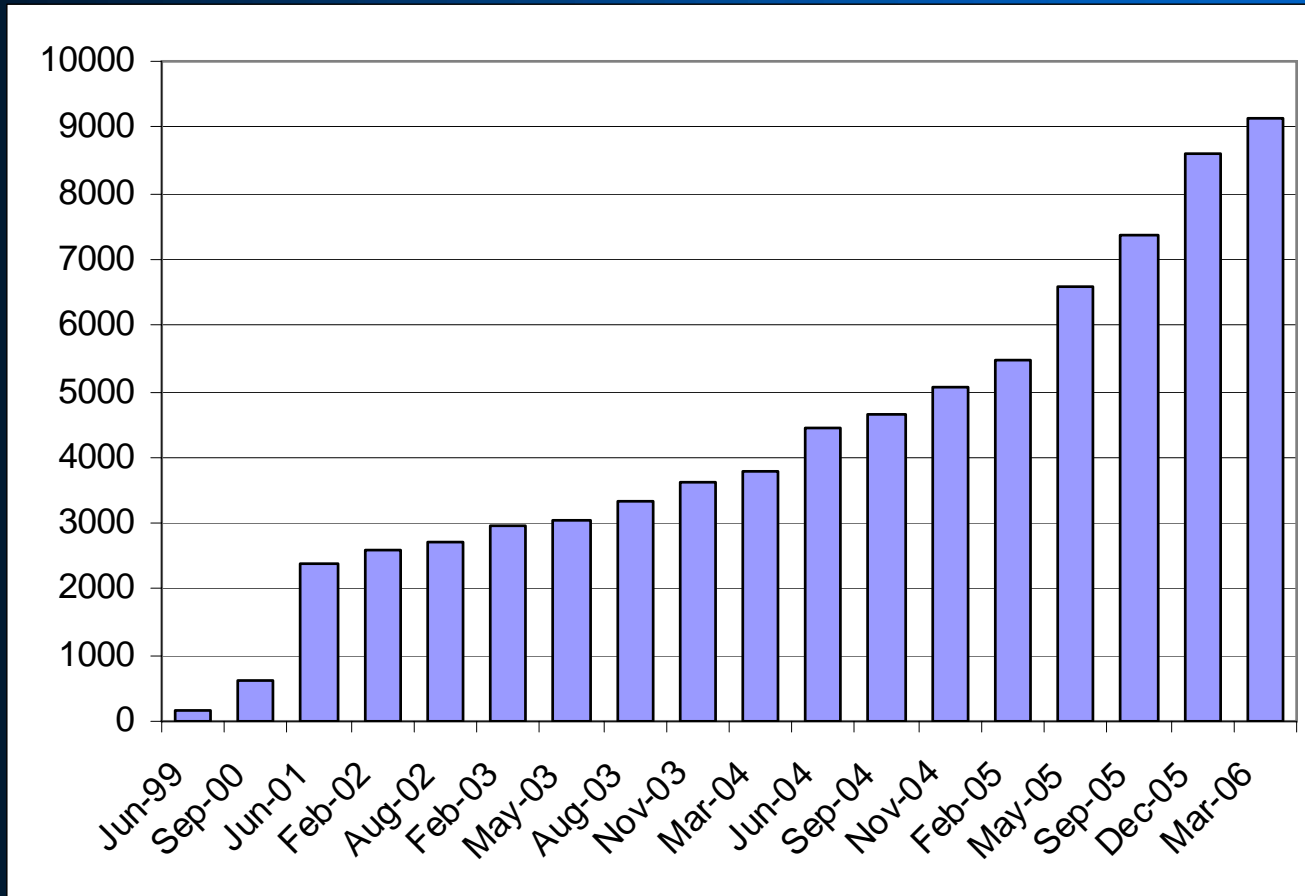
\* data not available

# *Distribution of Comment lengths*



Pathways were divided to groups based on comments length, measured by the number of characters.

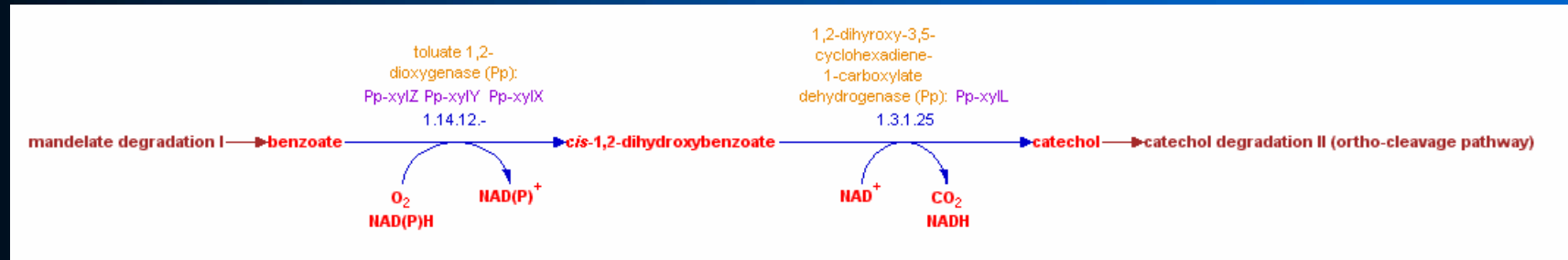
# *Increase in the Number of Citations*



# *Features*

# Hyperlinks in Pathways and Comments

Use links to other pathways whenever possible



Use hyperlinks to other database objects within comments

Comment:

Most organisms degrade [L-leucine](#) to [acetoacetate](#) and [acetyl-CoA](#) (see [leucine degradation I](#)). The anaerobic bacterium *Clostridium sporogenes* degrades [L-leucine](#) to different products, which include [isobutyrate](#) and [acetate](#) [ [Poston76](#) ]. This alternative pathway starts with isomerization of [L-leucine](#) to [β-leucine](#), performed by the enzyme [leucine 2,3-aminomutase](#). [β-leucine](#) is then converted, presumably by a transaminase, to [β-ketoisocaproate](#), which is then broken into [isobutyrate](#) and [acetate](#) [ [Poston86](#) ].

The only enzyme that has been characterized from this pathway is [leucine 2,3-aminomutase](#) [ [Poston76](#) ]. The enzyme was reported to also be present in mammalian tissues [ [Poston80](#) ] and green plants [ [Freer81](#) ]. However, recent work disputes this claim, having found no [β-leucine](#) in human blood and no [leucine 2,3-aminomutase](#) activity in rat liver [ [Stabler88](#) ].

Variants: [leucine degradation III](#), [leucine degradation I](#)



# Pathway Variant Classes

Pathway Info Editor for PWY-5254

Class: Biosynthesis->Cofactors, Prosthetic Groups, Electron Carriers

Common Name:

Create Variant Class for methanofuran biosynthesis

Parent class: Cofactor-Biosynthesis

Variant class ID:

Variant class common name (if different):

Documentation string:

Other members of variant class:

- rhizobactin 1021 biosynthesis
- methylerythritol phosphate pathway
- coenzyme B biosynthesis
- coenzyme M biosynthesis
- formylTHF biosynthesis II
- polyisoprenoid biosynthesis
- S-adenosylmethionine biosynthesis
- pyridoxamine anabolism
- lipoate salvage and modification

Variants: ***β-alanine biosynthesis IV***, ***β-alanine biosynthesis II***, ***β-alanine biosynthesis III***

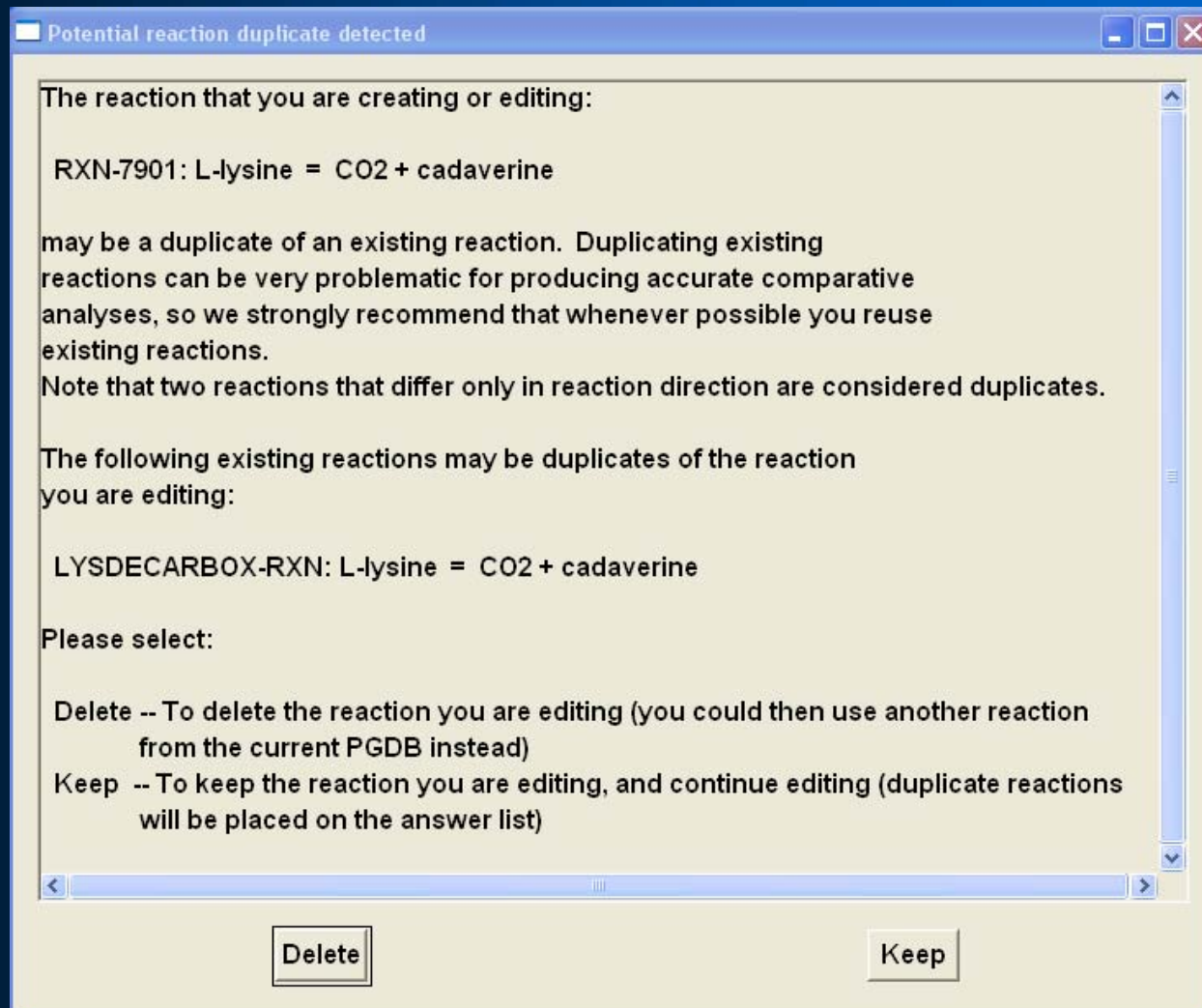
Created by: christ on 30-Jun-2005

# MetaCyc2GO Mapping

- The MetaCyc2GO mapping file contains the correspondence between MetaCyc pathways and GO process terms, and between MetaCyc reactions and GO function terms.
- It allows to propagate annotations from one database to another.
- The MetaCyc2GO file was originally maintained by GO.
- It is now continuously updated in an ongoing effort by Carnegie and SRI and the GO consortium.
  - The existing GO reactions and function terms have been checked against the latest EC nomenclature.
  - Discrepancies between GO and MetaCyc have been reviewed and corrected manually.



# Eliminating Unwanted Duplicate Frames



# Compound Pages Link to Regulated Enzymes

Compound pages list the enzymes that are regulated by the compounds

## example: L-glutamate

Activator (Allosteric) of: [glutaminase B](#)

Inhibitor (Competitive) of: [glutaminase A](#), [glutaminase B](#), [L-glutamine:D-fructose-6-phosphate aminotransferase](#), [aspartate 1-decarboxylase](#), [tyrosine/phenylalanine aminotransferase](#), [glutamate synthase \(NADH-dependent\)](#)

Inhibitor (Noncompetitive) of: [glutamate dehydrogenase \(NAD-dependent\)](#), [formiminoglutamate formiminohydrolase](#), [glutamate dehydrogenase \(NADP-dependent\)](#)

Inhibitor (Allosteric) of: [pyruvate kinase](#)

Inhibitor (Mechanism unknown) of: [UDP-N-acetylmuramoylalanine-D-glutamate ligase](#), [L-cysteine desulfhydrase](#), [pyridoxamine-oxaloacetate transaminase](#)

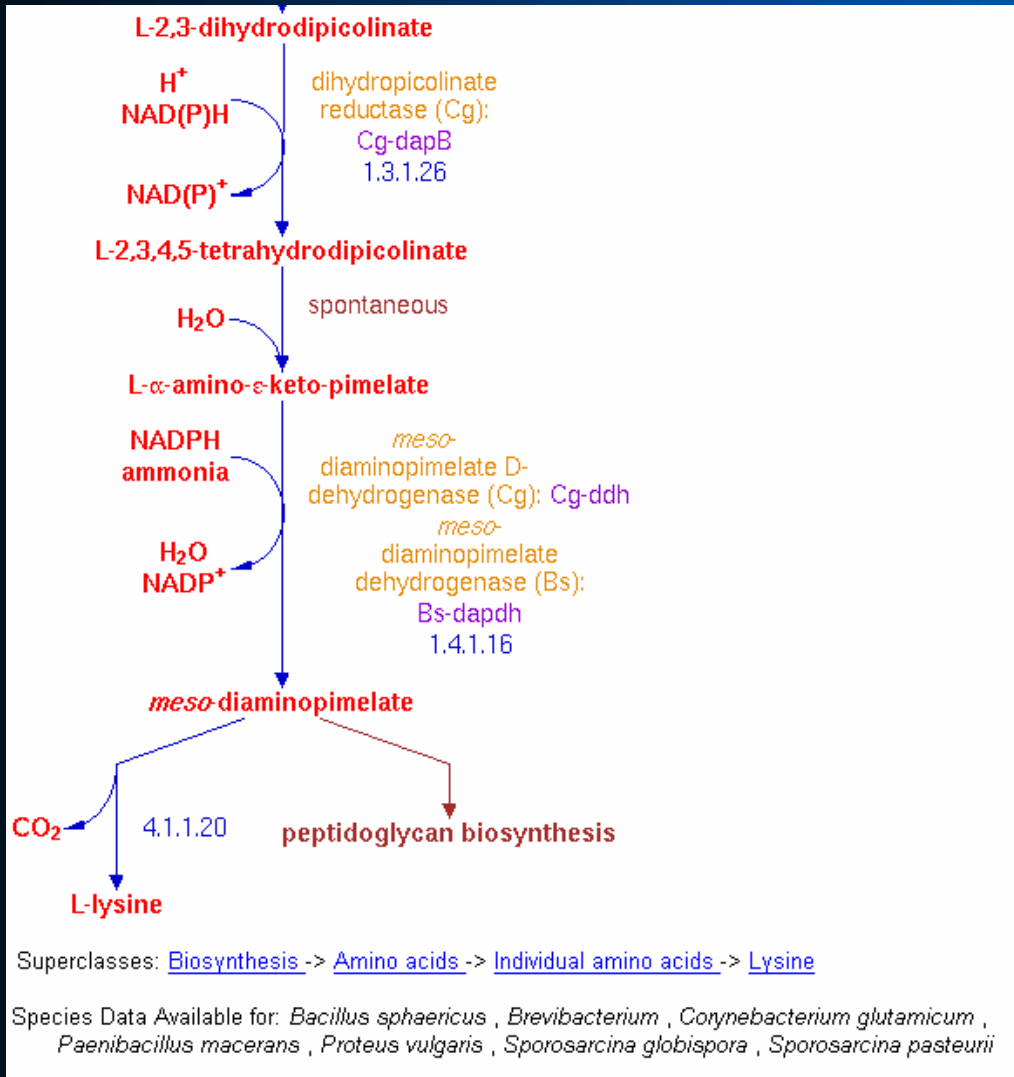
[Query Page](#)

[Advanced Query Page](#)

[BioCyc Home](#)

[Report Errors or Provide Feedback](#)

# Species Abbreviations in Pathway Display



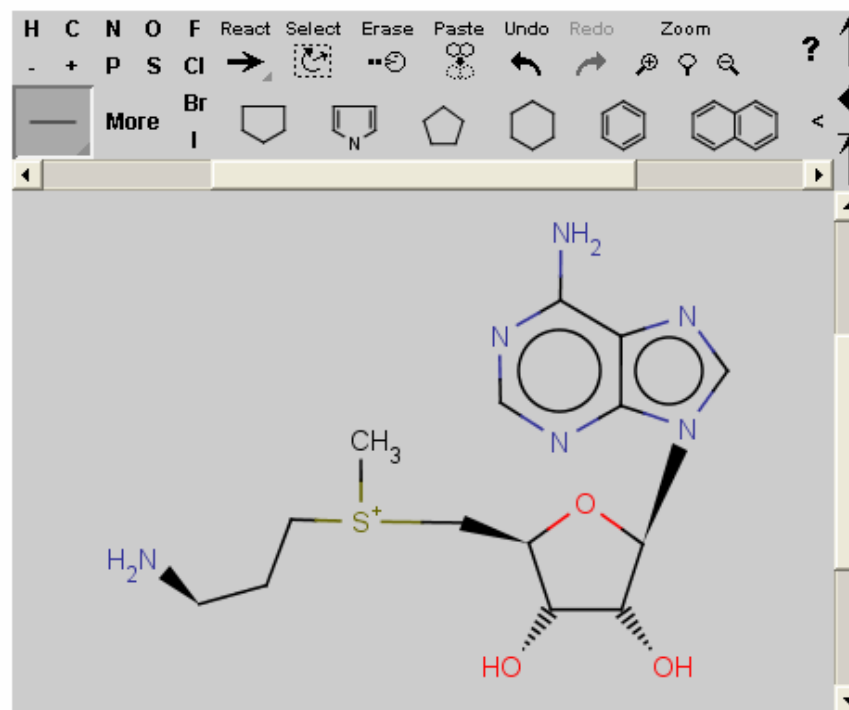
Genes and proteins are labeled in graphical diagrams with the initials of the organism's genus and species names

# Support for More Chemical Structure Editors

S-ADENOSYLMETHIONINAMINE : S-adenosyl-L-methioninamine

Choose operation for S-ADENOSYLMETHIONINAMINE

- Compound Editor
- Frame Editor
- Synonym Editor
- Relationships Editor
- Ontology Editor
- Marvin Compound Structure Editor**
- JME Compound Structure Editor
- Import Compound Structure from Molfile...
- Export Compound Structure to Molfile...
- Merge Compounds...
- Export Compound to KB...
- Copy Frame Name to Clipboard
- Duplicate Frame and Edit
- Create Frame
- Delete Frame



[Marvin MSketch Editor](#)

To use the editor, please enable Java and JavaScript in your WWW browser.

Submit Will write back the structure into frame S-ADENOSYLMETHIOI

# Expected Taxonomic Range

Superclasses: **Biosynthesis** -> **Cofactors, Prosthetic Groups, Electron Carriers**

Species Data Available for: *Archaeoglobus fulgidus*, *Methanobacteria*, *Methanobrevibacter*  
*Methanothermobacter thermautotrophicus*, *Methanothermus*

Expected Taxonomic Range: *Archaeoglobi*, *Methanobacteria*, *Methanococci*, *Methanomicrobia*

Created by: caspi on 1-Jun-2006

Credits: Created 01-Jun-2006 by Caspi R, SRI

Choose a class

Organisms

- Archaea
  - Archaeoglobi
  - Halobacteria
  - Methanobacteria
  - Methanococci
  - Methanomicrobia
  - Methanopyr
  - Thermococci
  - Thermoplasmata
  - Thermoprotei
- Bacteria
  - Actinobacteria
  - Aquificae
  - Bacteroidetes/Chlorobi group
  - Chlamydiae/Verrucomicrobia group
  - Chloroflexi (Green non-sulfur bacteria)
  - Cyanobacteria
  - Deinococcus-Thermus
  - Firmicutes
  - Planctomycetes
  - Proteobacteria
  - Spirochaetes
  - Thermotogae
- Eukaryota
  - Fungi
  - Magnoliophyta
- + Metazoa

# *Other MetaCyc-related Pathway Tools Improvements*

- **Pathways:**
  - Made pathway displays more compact
  - Search for pathways and reactions by substrates
- **Chemical structures:**
  - Import/export of mol files
  - Search for compounds by combining name, MW, formula, and substructure
- **Auto-checkpointing**



# *Future Directions*

# *Future Curation*

- **Environmental** pathways (degradation of explosives, pesticides etc)
- **Medical pathways** (antibiotics biosynthesis, microbial secondary metabolites etc)
- **Pathways of commercial importance** (food industry, vitamins, chemicals etc)
- **General metabolism**

# *Submitting Pathways to MetaCyc*

- The pathway must be experimentally proven and described in one or more published journal articles
- Pathways and enzymes should have comments, literature citations and evidence codes
- Genes and enzymes should have links to sequence databases (e.g. Entrez and UniProt) and citations, if available
- Enzymes should have as much information as possible, such as optimal pH and temperature, Km values, inhibitor, activator and cofactor information, etc.
- Once in MetaCyc, your pathway will be used routinely by PathoLogic during the creation and updating of PGDBs

More about the pathway submission can be found at <http://www.metacyc.org/MetaCycPosting.shtml>

# Author Crediting System

Added the ability to credit authors and institutions for creating, reviewing, or revising a pathway

MetaCyc Author: **Eurie Hong**

Email: eurie@genome.stanford.edu

Affiliations: **Saccharomyces Genome Database**

Comment:

Head, Scientific Curation, SGD

Note: Listed below are contributions the author has made to MetaCyc

Created: **allantoin degradation II on 27-Apr-2005,**  
**superpathway of NAD biosynthesis in eukaryotes on 27-Apr-2005,**  
**glutamate biosynthesis IV on 27-Apr-2005,**  
**NAD salvage pathway I on 27-Apr-2005,**  
**ergosterol biosynthesis on 21-Dec-2004,**  
**lysine biosynthesis IV on 21-Dec-2004,**  
**sphingolipid metabolism on 21-Dec-2004**

Editor:

Credits:	Date	Authors	Organizations
-----	none yet	<input type="button" value="Select/Change"/> <input type="button" value="Create"/>	<input type="button" value="Select/Change"/> <input type="button" value="Create"/>
-----		Current selection(s): Caspi R	Current selection(s): SRI
Created			
Reviewed			
Revised			
<input type="button" value="OK"/> <input type="button" value="Cancel"/>			

Display:

Credits: Created 27-Apr-2005 by **Balakrishnan R, Hong E, Saccharomyces Genome Database**  
Reviewed 03-Apr-2006 by **Caspi R, SRI**

*End*