

# *Recent and Planned Enhancements to Pathway Tools and BioCyc*

**Peter Karp**

**Bioinformatics Research Group**

**SRI International**

**[pkarp@ai.sri.com](mailto:pkarp@ai.sri.com)**

# Pathway/Genome Databases on the Web

- <http://biocyc.org/otherpgdbs.shtml>

## Other Pathway/Genome Databases on the Internet

<a href="#">AraCyc</a>	<i>Arabidopsis thaliana</i>	S. Rhee, <a href="#">Department of Plant Biology, Carnegie Institution</a> , USA
<a href="#">CvioCyc</a>	<i>Chromobacterium violaceum</i>	Artiva Maria Goudel, Federal University of Santa Catarina, Brazil
<a href="#">CryptoCyc</a>	<i>Cryptosporidium parvum</i> Iowa <i>Cryptosporidium hominis</i> TU502	<a href="#">Cryptosporidium Genome Resources</a>
<a href="#">DictyCyc</a>	<i>Dictyostelium discoideum</i>	dictyBase, Northwestern U., USA
<a href="#">LacPlantCyc</a>	<i>Lactobacillus plantarum</i> WCFS1	F. H. J. van Enckevort, <a href="#">CMBI</a> , The Netherlands
<a href="#">LeishCyc</a>	<i>Leishmania major</i> Friedlin	Bio21 Institute, <a href="#">University of Melbourne</a> , Australia
<a href="#">MedicCyc</a>	<i>Medicago truncatula</i>	Samuel Roberts Noble Foundation, USA
<a href="#">MicroScope</a>	PGDBs for 60 Genomes	C. Medigue, <a href="#">Genoscope</a> , France
<a href="#">MouseCyc</a>	<i>Mus musculus</i>	C. Bult, <a href="#">Jackson Laboratory</a> , USA
<a href="#">PATRIC</a>	<i>Brucella suis</i> 1330 <i>Coxiella burnetii</i> RSA 493 <i>Rickettsia typhi</i> str. Wilmington	PathoSystems Resource Integration Center, <a href="#">Virginia Bioinformatics Institute</a> , USA
<a href="#">PseudoCyc</a>	<i>Pseudomonas aeruginosa</i>	F. Brinkman, <a href="#">Pseudomonas Genome Project</a> , Simon Fraser U., Canada
<a href="#">RetliDB</a>	<i>Rhizobium etli</i>	<a href="#">Center for Genomic Sciences</a> , Mexico
<a href="#">RiceCyc</a>	<i>Oryza sativa</i>	<a href="#">Gramene</a> curators, Cornell U. and CSHL
<a href="#">ScoCyc</a>	<i>Streptomyces coelicolor</i> A3(2)	V. Armendarez, G. Chandra, M. Bibb, <a href="#">John Innes Centre</a> , UK
<a href="#">SolCyc</a>	<i>Solanum lycopersicum</i> <i>Solanum tuberosum</i>	<a href="#">Sol Genomics Network</a> , USA
<a href="#">TBestDB</a>	Taxonomically Broad EST Database	TBestDB Group, Canada
<a href="#">Yeast Biochemical Pathways</a>	<i>Saccharomyces cerevisiae</i>	<a href="#">SGD</a> curators, Stanford U., USA
<a href="#">Contact us</a> if you'd like your PGDB added to this list.		

# *Pathway/Genome Databases on the Web*

# *Pathway Tools Software: PGDBs Created Outside SRI*

- 2,580+ licensees: 205 groups applying software to 1,750 organisms
- **Saccharomyces cerevisiae**, SGD project, Stanford University
  - 135 pathways / 565 publications
- FungiCyc, Broad Institute -- 23 fungi
- **Candida albicans**, CGD project, Stanford University
- dictyBase, Northwestern University
  
- **Mouse**, MGD, Jackson Laboratory
- **Drosophila**, FlyBase, Harvard University
- Under development:
  - *C. elegans*, WormBase
  
- **Arabidopsis thaliana**, TAIR, Carnegie Institution of Washington
  - 288 pathways / 2282 publications
- ChlamyCyc, GoFORSYS
- PlantCyc, Carnegie Institution of Washington
- Six *Solanaceae* species, Cornell University
- GrameneDB, Cold Spring Harbor Laboratory
- **Medicago truncatula**, Samuel Roberts Noble Foundation

# *Pathway Tools Software: PGDBs Created Outside SRI*

- E. Uberbacher, ORNL 33 Bioenergy-related organisms
- G. Serres, MBL and ORNL: 18 *Shewanella* genomes
- M. Bibb, John Innes Centre, *Streptomyces coelicolor*
- TBDB Project, *Mycobacterium tuberculosis*
- F. Brinkman, Simon Fraser Univ, *Pseudomonas aeruginosa*
- Genoscope, *Acinetobacter*
- R.J.S. Baerends, University of Groningen, *Lactococcus lactis* IL1403, *Lactococcus lactis* MG1363, *Streptococcus pneumoniae* TIGR4, *Bacillus subtilis* 168, *Bacillus cereus* ATCC14579
- Matthew Berriman, Sanger Centre, *Trypanosoma brucei*, *Leishmania major*
- Sergio Encarnacion, UNAM, *Sinorhizobium meliloti*
- Mark van der Giezen, University of London, *Entamoeba histolytica*, *Giardia intestinalis*

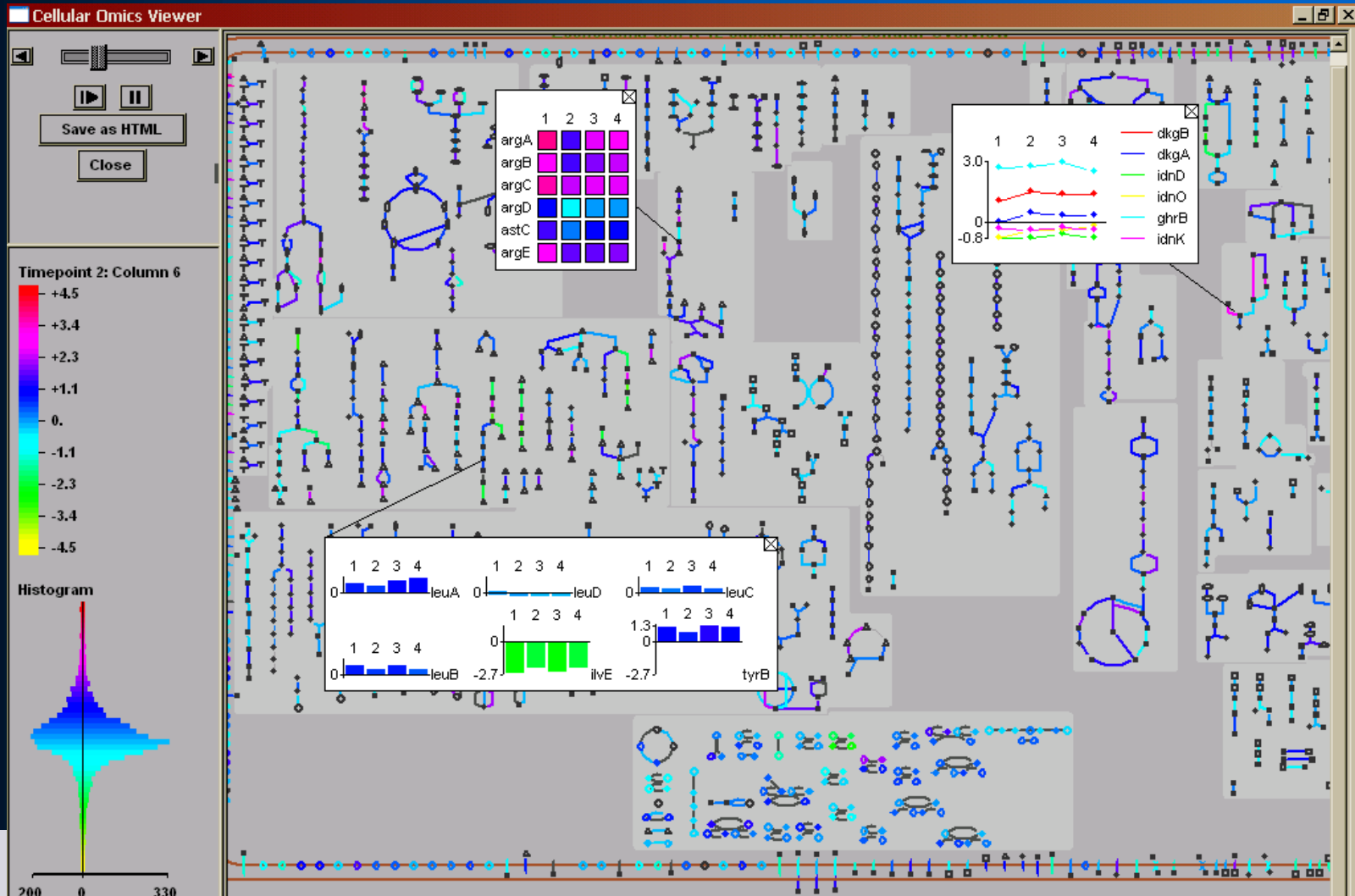
# *Pathway Tools Software: PGDBs Created Outside SRI*

- **Large scale users:**
  - *C. Medigue, Genoscope, 200+ PGDBs*
  - *G. Sutton, J. Craig Venter Institute, 80+ PGDBs*
  - *G. Burger, U Montreal, 60+ PGDBs*
  - *Bart Weimer, UC Davis, Lactococcus lactis, Brevibacterium linens, Lactobacillus acidophilus, Lactobacillus plantarum, Lactobacillus johnsonii, Listeria monocytogenes*
- **Partial listing of outside PGDBs at <http://biocyc.org/otherpgdbs.shtml>**

Version 13.5

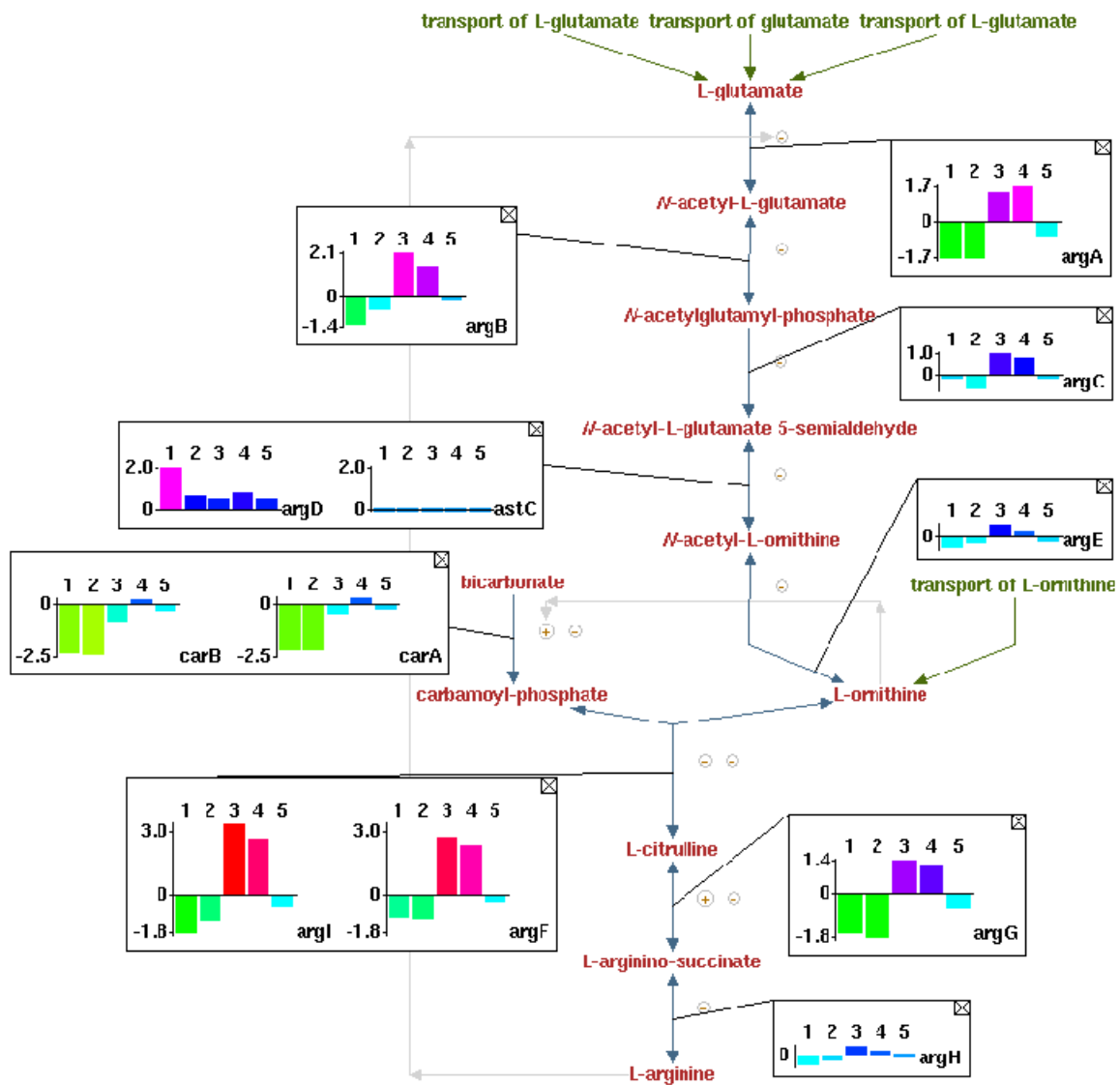
Sept 2009

- Omics graphing capabilities



Escherichia coli K-12 substr. MG1655 Pathway: arginine biosynthesis I

More Detail Less Detail Species Comparison





*Version 13.5*

*Sept 2009*

- **Object groups**
  - Define groups of pathways, genes, metabolites, etc
  - Perform transformations on groups
- **Enrichment analysis**
  - Genes -> Pathways, GO terms, regulators
  - Metabolites -> Pathways

*Version 13.5*

*Sept 2009*

- **Protein editor upgrades**
- **Enhancements to ChIP-chip display tracks**
- **Cross reference MetaCyc to Pubchem and KEGG**
- **Import of protein features from SwissProt into a PGDB**

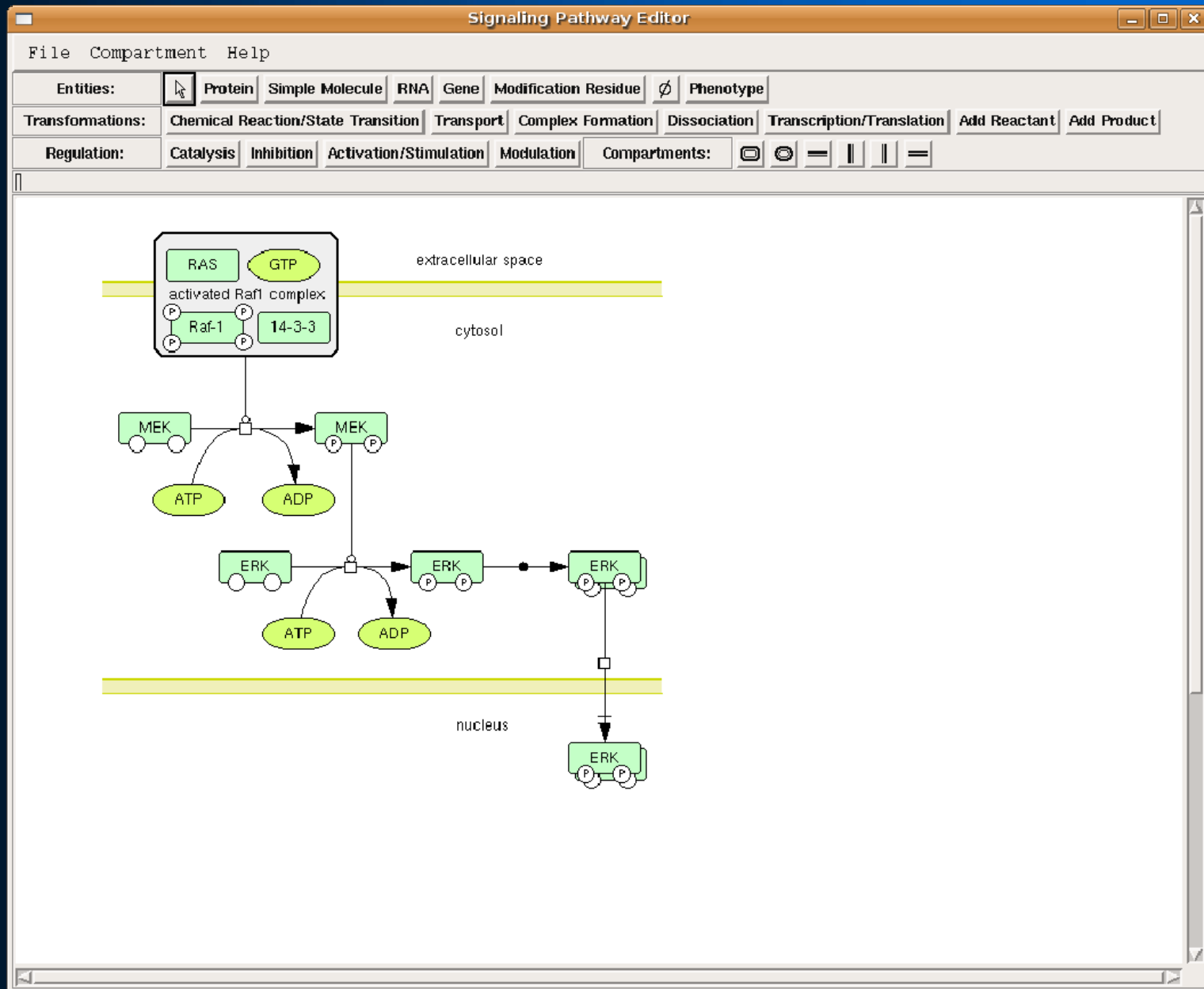
*Version 14.0*

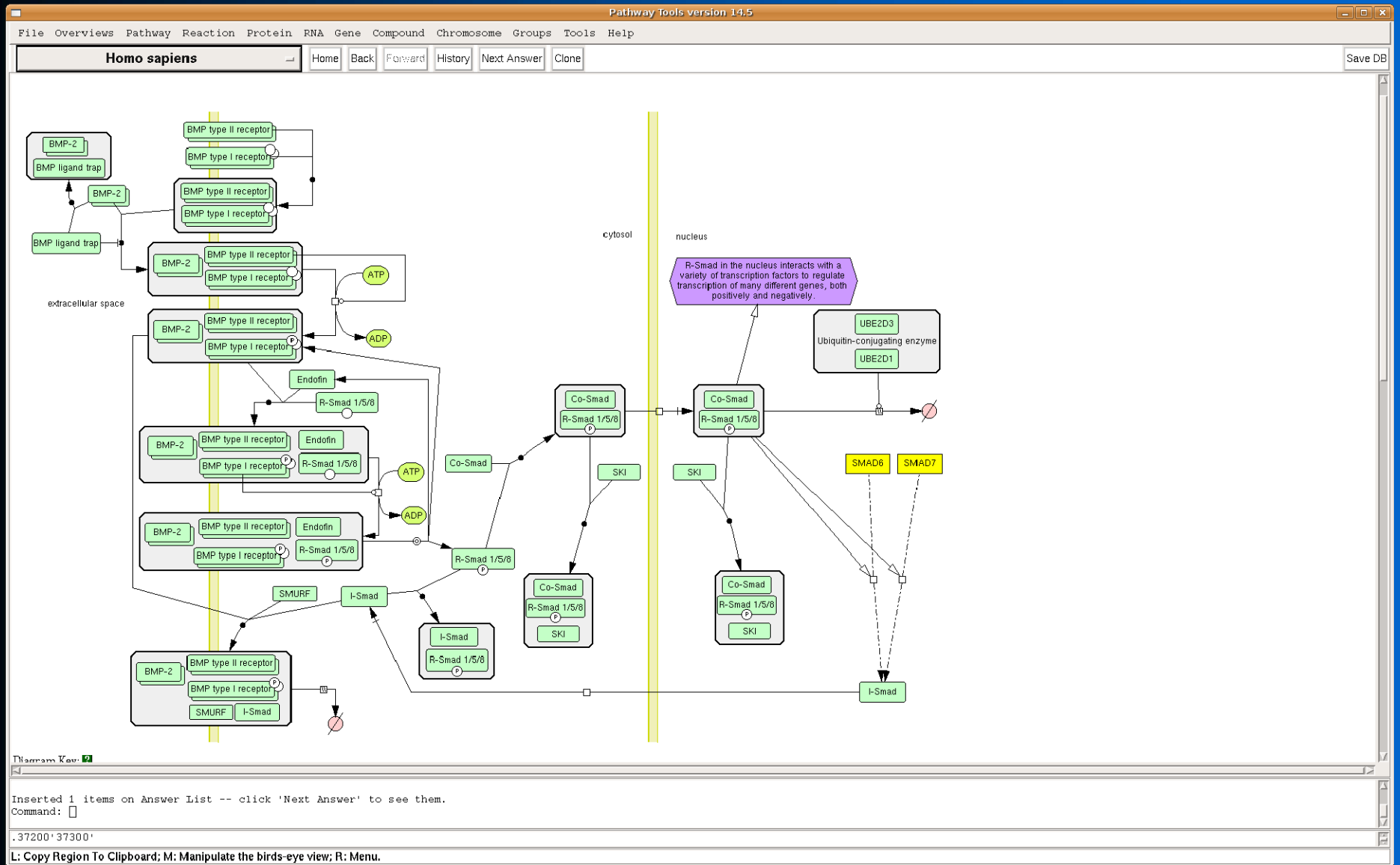
*March 2010*

- **Increased scalability based on Allegro Cache**

# *Pathway Tools Supports Signaling Pathways*

- **Schema updated to accommodate signaling pathways and reactions**
- **Signaling pathway editor**
  - Pathway -> New Signaling Pathway
- **Signaling pathway display page**
- **Closely modeled on CellDesigner software and SBGN standard**
- **User positions elements of signaling pathways interactively**





*Version 14.0*

*March 2010*

- **Cellular Overview re-implemented for the Web**
  - Javascript and AJAX based
  - Semantic zooming, searchable
  - New omics viewer with zooming
  
- **Tools -> Cellular Overview**

*Version 14.5*

*September 2010*

- **Regulation summary diagram**
- **Web services**
- **Monoisotopic mass search**
- **Omics viewer for regulatory overview**
- **Molfile support for V3000 format**
- **Multi-strain gene/protein editor**
- **Performance improvements**
- **Folders available for groups**



*Version 14.5*

*September 2010*

- **BioCyc contains 1,004 genomes**
- **BioCyc is switching genome data sources**
  - CMR → RefSeq
- **BioCyc pipeline**
  - RefSeq → BioWarehouse → PathoLogic-format files
  - PathoLogic
    - ◆ Predict pathways
    - ◆ Predict pathway hole fillers
    - ◆ Predict operons
    - ◆ Transport inference parser
  - Generate Cellular Overview
  - Compute orthologs

## *MetaCyc Data -- Version 14.5*

<b>Pathways</b>	<b>1,583</b>
<b>Reactions</b>	<b>8,837</b>
<b>Enzymes</b>	<b>6,758</b>
<b>Small Molecules</b>	<b>8,763</b>
<b>Organisms</b>	<b>2,007</b>
<b>Citations</b>	<b>26,009</b>

## *Coming in Later Versions*

- **Version 15.0:**
- **Groups available through BioCyc Web site**
- **Enhanced pathway layout algorithms**
- **Automatic generation of flux balance models from PGDBs**
- **Multi-strain curation tools**
- **Annotation normalization tool**
- **New pathway prediction algorithm**
- **High throughput computation of orthologs**

# *Suggestions*

- **Submit pathways to MetaCyc (author credit system)**
- **Submit PGDBs to registry**
  - Facilitate comparisons
  - Enable local installation of your PGDBs