

The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases

Ron Caspi¹, Tomer Altman¹, Joseph M. Dale¹, Kate Dreher², Carol A. Fulcher¹, Fred Gilham¹, Pallavi Kaipa¹, Athikkattuvalasu S. Karthikeyan², Anamika Kothari¹, Markus Krummenacker¹, Mario Latendresse¹, Lukas A. Mueller³, Suzanne Paley¹, Liviu Popescu¹, Anuradha Pujar³, Alexander G. Shearer¹, Peifen Zhang² and Peter D. Karp^{1,*}

¹SRI International, 333 Ravenswood, Menlo Park, CA 94025, ²Carnegie Institution, Department of Plant Biology, 260 Panama Street, Stanford, CA 94305 and ³Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, NY 14853, USA

Received September 30, 2009; Accepted October 1, 2009

ABSTRACT

The MetaCyc database (MetaCyc.org) is a comprehensive and freely accessible resource for metabolic pathways and enzymes from all domains of life. The pathways in MetaCyc are experimentally determined, small-molecule metabolic pathways and are curated from the primary scientific literature. With more than 1400 pathways, MetaCyc is the largest collection of metabolic pathways currently available. Pathways reactions are linked to one or more well-characterized enzymes, and both pathways and enzymes are annotated with reviews, evidence codes, and literature citations. BioCyc (BioCyc.org) is a collection of more than 500 organism-specific Pathway/Genome Databases (PGDBs). Each BioCyc PGDB contains the full genome and predicted metabolic network of one organism. The network, which is predicted by the Pathway Tools software using MetaCyc as a reference, consists of metabolites, enzymes, reactions and metabolic pathways. BioCyc PGDBs also contain additional features, such as predicted operons, transport systems, and pathway hole-fillers. The BioCyc Web site offers several tools for the analysis of the PGDBs, including Omics Viewers that enable visualization of omics datasets on two different genome-scale diagrams and tools for comparative analysis. The BioCyc PGDBs generated by SRI are offered for adoption by any party interested

in curation of metabolic, regulatory, and genome-related information about an organism.

INTRODUCTION

MetaCyc (MetaCyc.org) is a highly curated, non-redundant reference database of small-molecule metabolism. It contains metabolic pathway and enzyme data that have been experimentally demonstrated in the scientific literature (1) (Figure 1). Because MetaCyc contains only experimentally determined pathways and enzymes, and due to its tight integration of data and references, MetaCyc is a uniquely valuable resource in fields including genome analysis, metabolism and metabolic engineering. The metabolic pathways and enzymes in MetaCyc are derived from organisms representing all domains of life (Tables 1 and 2). In the past, microbial and plant metabolism were emphasized, but current curation also focuses on vertebrate metabolism.

In conjunction with its role as a general reference on metabolism, MetaCyc is used as a reference database for the PathoLogic component of the Pathway Tools software (2) to computationally predict the metabolic network of any organism having a sequenced and annotated genome (3). In this automated process, a predicted metabolic network is created in the form of a Pathway/Genome Database (PGDB). BioCyc (BioCyc.org) is a collection of more than 500 organism-specific PGDBs that were generated in this way both at SRI and by other groups. The editing capability of Pathway Tools enables computationally predicted PGDBs to be improved and updated by manual curation. Interested scientists may

*To whom correspondence should be addressed. Tel: +1 650 859 4358; Fax: +1 650 859 3735; Email: pkarp@ai.sri.com

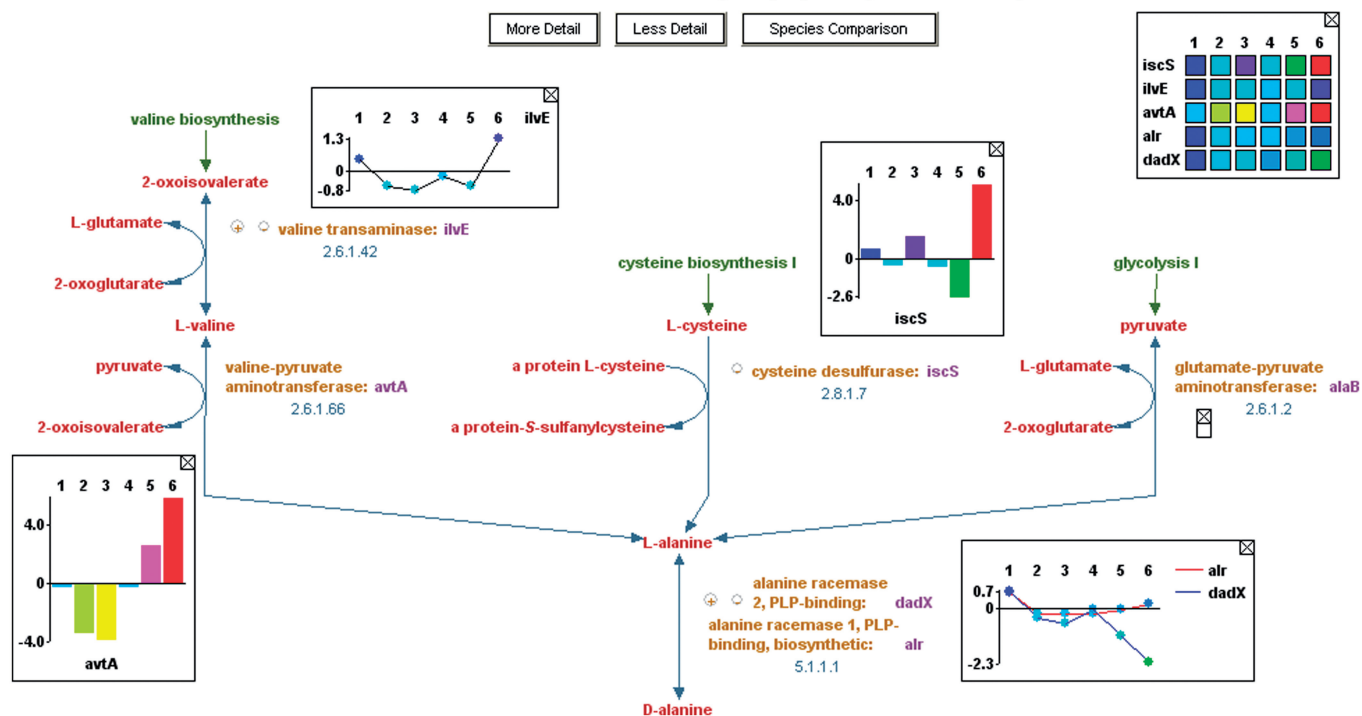
***Escherichia coli K-12 substr. MG1655 Pathway: superpathway of alanine biosynthesis**

Figure 1. An example of a pathway showing omics data pop-ups. Pathways can be displayed at varying levels of detail, and this pathway's display depicts an intermediate level of detail including enzymes, EC numbers and genes, but no chemical structures. Notice the green arrows at the top of the pathway, which provide hyperlinks to related upstream pathways. After an omics dataset has been uploaded, the user can easily add data pop-ups to the pathway diagram displaying the omics data for genes or metabolites. The user can select from heatmap, bar graph or plot representation. Multiple time points are supported.

adopt and curate existing PGDBs through the BioCyc Web site (biocyc.org/intro.shtml#adoption), or create new PGDBs using MetaCyc and Pathway Tools (biocyc.org/download.shtml). More than 80 groups have used Pathway Tools and MetaCyc to create PGDBs for their organisms of interest, including important model organisms such as *Saccharomyces cerevisiae* (4), *Arabidopsis thaliana* (5), *Oryza sativa* (6), *Mus musculus* (7), *Bos taurus* (8), *Medicago truncatula* (9), *Dictyostelium discoideum* (10), *Leishmania major* (11), *Chlamydomonas reinhardtii* (12), several *Solanaceae* species (13) and many pathogenic bacteria (14) (see <http://biocyc.org/otherpgdbs.shtml> for a more complete list).

A web server included in Pathway Tools enables the publishing of PGDBs through either the internet or an internal network. The Navigator component of Pathway Tools allows the browsing and analysis of PGDBs either locally or over the Internet. A detailed description of Pathway Tools can be found at <http://bioinformatics.ai.sri.com/ptools/> and in (15).

PGDBs generated by Pathway Tools and MetaCyc are an excellent platform for the integration of genome information with many other types of data regarding metabolism, regulation, and genetics. They provide powerful tools for analyzing omics datasets from experiments related to gene transcription, metabolomics, proteomics, ChIP-chip analysis, etc. (Figure 2). The PGDBs accelerate research in many fields including biochemistry, molecular biology, biotechnology, bioinformatics, metabolic

engineering and systems biology (16–19). Both MetaCyc and organism-specific PGDBs can also be used as educational tools.

During the past 2 years, we again significantly expanded the data content of MetaCyc and BioCyc. We also added supporting enhancements to the Pathway Tools software. The expanded and enhanced databases and software are described in the following sections.

METACYC ENHANCEMENTS

Expansion of MetaCyc

All pathways in MetaCyc are curated from the experimental literature. Since the last *Nucleic Acids Research* publication (2 years ago) (20), we added 507 new base pathways (pathways comprised of reactions only, where no portion of the pathway is designated as a subpathway) and 129 superpathways (pathways composed of at least one base pathway plus additional reactions or pathways), and updated 104 existing pathways, for a total of 740 new and revised pathways. The total number of base pathways grew by 43%, from 977 (version 11.5) to 1399 (version 13.5) (the total increase is less than 507 pathways because some existing pathways were deleted from the database during this period); while the total number of superpathways grew by 120%, from 106 (version 11.5) to 235 (version 13.5).

degradation/utilization/assimilation, generation of precursor metabolites and energy and detoxification.

In version 13.5, the largest top-level class is Biosynthesis, with 902 base pathways. Its main subclasses are secondary metabolites biosynthesis (351); cofactors, prosthetic groups, and electron carriers biosynthesis (160); amino acids biosynthesis (105); and fatty acids and lipids biosynthesis (101).

The second-largest top-level class is degradation/utilization/assimilation, with 639 base pathways. Within this group, the largest subclasses are aromatic compounds degradation (152), amino acids degradation (113), inorganic nutrients metabolism (72), secondary metabolites degradation (58), and carbohydrates degradation (52).

The third-largest top-level class, generation of precursor metabolites and energy, contains 124 base pathways. Its largest subclasses are fermentation (34), respiration (25), chemoautotrophic energy metabolism (14) and methanogenesis (12).

The final top-level class, detoxification, is much smaller, with only 16 base pathways.

During the previous 2 years, the number of metazoan pathways in MetaCyc increased by 67%, from 104 to 174 pathways.

The list of pathways added to MetaCyc since the last *NAR* publication is too long to give here. For a complete report, please see the MetaCyc release notes history at <http://metacyc.org/release-notes.shtml>.

Electron transfer pathways

Following the introduction of support for electron transfer reactions into the database schema, we added a total of 11 electron transfer pathways to the database. This type of pathways utilizes a different display algorithm that conveys features such as the direction of the electron flow, the cell-compartment locations where the substrates are transformed, and the optional translocation of protons across membranes. For an example of such pathways, see the pathway ‘succinate to cytochrome *bd* oxidase electron transfer’.

Interactions with other databases

IUBMB. MetaCyc is regularly updated with data from the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB), which includes new and modified EC numbers. The last supplement incorporated is supplement 14.

NCBI Taxonomy. Starting with version 12.0, the full NCBI Taxonomy database (21) is integrated into Pathway Tools, enabling specification of the taxa in which MetaCyc pathways occur using NCBI Taxonomy, and allowing taxonomic querying of MetaCyc pathways and enzymes.

Gene Ontology. We continue to update the mapping between MetaCyc and Gene Ontology (GO) process and function terms (22). In May 2009, we submitted to GO updated mappings between MetaCyc pathways and reactions and GO biological process and molecular

function terms. An updated file is found at <http://www.geneontology.org/external2go/metacyc2go>.

PubChem. To improve the mapping between MetaCyc compounds and other compound databases, all MetaCyc compounds were incorporated into the PubChem database (23), and are linked to their PubChem entries.

KEGG. MetaCyc was updated to better link objects to corresponding entries in KEGG. 3269 MetaCyc reactions were mapped to KEGG reactions, and an additional 920 MetaCyc compounds now have links to KEGG compounds. MetaCyc pathways contain more than twice as many reactions (4950) than does KEGG (2463), although KEGG contains more total reactions (9194 in version 50) than MetaCyc (8387). MetaCyc contains 1399 pathways compared to the 155 pathways in KEGG.

BioCyc ENHANCEMENTS

Expansion of BioCyc

The BioCyc databases are organized into three tiers.

- Tier 1 PGDBs were created using intensive manual curation and receive continuous updating.
- Tier 2 PGDBs have received moderate amounts of review, but may not be updated on an ongoing basis.
- Tier 3 PGDBs were created computationally, and received no subsequent manual review or updating.

During the past two years, the number of BioCyc PGDBs increased from 371 (version 11.5) to 508 (version 13.1), out of which two are in Tier 1 (EcoCyc and MetaCyc), 24 are in Tier 2, and the rest belong to Tier 3. Some Tier 2 PGDBs were provided by groups outside SRI [examples include MouseCyc (7) CattleCyc (8) and YeastCyc (4)]. Database authors are identified on the database summary page (Tools → Reports → Summary Statistics).

Innovations in database sharing

The extended family of Pathway Tools-based Pathway/Genome Databases includes both the SRI-created BioCyc collection and many PGDBs created outside SRI by other Pathway Tools users. This DB family exhibits a number of innovations in scientific database sharing. We believe that no one group can curate all the world's genomes; therefore, we strongly emphasize the notion of widely distributing the workload of curating genome databases. All the Tier 3 PGDBs and some of the Tier 2 PGDBs are offered for adoption to interested parties under an open license agreement. Some groups adopt existing PGDBs within BioCyc, assuming responsibility for their ongoing curation. Other groups create their own PGDBs using Pathway Tools. We offer free technical support to facilitate this task for our academic users.

All Pathway Tools-based PGDBs share the same schema, thus facilitating comparative analyses and data exchange. An encouraged form of data exchange is the submission of experimentally determined metabolic pathways curated by curators of other PGDBs for

Table 1. List of species that have more than 15 experimentally elucidated pathways represented in MetaCyc

Bacteria		Eukarya		Archaea	
<i>Escherichia coli</i>	248	<i>Arabidopsis thaliana</i>	239	<i>Methanosarcina barkeri</i>	18
<i>Bacillus subtilis</i>	53	<i>Homo sapiens</i>	126	<i>Methanocaldococcus jannaschii</i>	17
<i>Pseudomonas aeruginosa</i>	47	<i>Saccharomyces cerevisiae</i>	95	<i>Methanosarcina thermophila</i>	17
<i>Pseudomonas putida</i>	41	<i>Glycine max</i>	62	<i>Methanothermobacter thermautotrophicus</i>	16
<i>Salmonella typhimurium</i>	33	<i>Rattus norvegicus</i>	60	<i>Sulfolobus solfataricus</i>	16
<i>Pseudomonas fluorescens</i>	26	<i>Pisum sativum</i>	42		
<i>Mycobacterium tuberculosis</i>	20	<i>Mus musculus</i>	38		
<i>Klebsiella pneumonia</i>	18	<i>Zea mays</i>	38		
<i>Agrobacterium tumefaciens</i>	17	<i>Solanum lycopersicum</i>	34		
<i>Cupriavidus</i> sp. PS12	15	<i>Solanum tuberosum</i>	34		
<i>Mycobacterium smegmatis</i>	15	<i>Oryza sativa</i>	32		

The species are grouped by taxonomic domain and are ordered within each domain based on the number of pathways (number following species name) to which the given species was assigned. Some pathways may be labeled with a higher-level taxon, such as genus, if all the species within that genus are thought to have the given pathway. However, such higher-level taxa are not included in this table.

inclusion in MetaCyc, broadening the pathways available for pathway prediction, and easing the bottleneck of data entry into MetaCyc. Conversely, Pathway Tools now includes the ability to propagate updates made to MetaCyc to other PGDBs derived from earlier versions of MetaCyc. For example, corrections made to MetaCyc chemical structures or reaction equations can be propagated to other PGDBs. Further, Pathway Tools can perform incremental pathway prediction, thus propagating newly curated pathways present in the latest version of MetaCyc to older organism-specific PGDBs.

Another means for facilitating data exchange is the PGDB registry, operated by SRI. Groups that curate PGDBs can register their databases in our PGDB registry (<http://biocyc.org/registry.shtml>) in a process that includes deposition of the PGDB in a downloadable format on the author's FTP or HTTP site. With a few mouse clicks, any Pathway Tools user can download a PGDB listed in the Registry and install it into their working copy of Pathway Tools, making it available for comparative analysis, omics data analysis, etc. Thus, users can share PGDBs as easily as they exchange music files on the Internet.

Compound protonation and reaction balancing

Starting with version 13.0, all MetaCyc compounds have been adjusted to a consistent protonation state for a reference pH of 7.3, common in the cellular cytosol. This adjustment was performed using the Marvin computational chemistry software (ChemAxon Kft, Budapest Hungary). In addition, all reactions that had a mass-imbalance due only to hydrogen atoms were computationally balanced by adding or removing protons from the appropriate side of the reaction. These updated compounds and reactions eventually will be propagated into all BioCyc PGDBs, and to other PGDBs created using Pathway Tools, making it easier to apply flux-balance analysis techniques to these databases.

This change resulted in certain differences between some MetaCyc reactions and the comparable reactions in other databases, such as the ENZYME Database (24). However, we believe that the representation of reactions

Table 2. Distribution of pathways in MetaCyc based on the taxonomic classification of associated species. Taxonomic groups (phyla for Bacteria and Archaea, kingdoms for Eukarya) are grouped by domain and are ordered within each domain based on the number of pathways (number following taxon name) associated with the taxon

Bacteria		Eukarya		Archaea	
Proteobacteria	750	Viridiplantae	643	Euryarchaeota	95
Firmicutes	190	Fungi	204	Crenarchaeota	31
Actinobacteria	171	Metazoa	174		
Bacteroidetes/Chlorobi	54	Euglenozoa	16		
Cyanobacteria	42				
Deinococcus-Thermus	22				
Thermotogae	12				
Spirochaetes	9				
Aquificae	9				
Planctomycetes	5				
Fusobacteria	4				
Chloroflexi	4				
Chlamydiae	2				
-Verrucomicrobia					
Thermodesulfobacteria	2				
Nitrospirae	2				
Chrysiogenetes	1				

Euglenozoa are listed separately as this group does not belong to any of the other eukaryotic kingdoms. A pathway may be associated with multiple organisms.

in MetaCyc is more consistent and, within the limits of the cytosolic pH of 7.3, more accurate.

SOFTWARE IMPROVEMENTS

The following paragraphs list a number of the most salient improvements to Pathway Tools during the past 2 years.

Web site redesign

The BioCyc Web site has undergone a significant overhaul that includes a new toolbar, a new organism selector widget, and new search commands. The new object-specific search commands provide an intermediate level of search complexity that lies between the very easy-to-use Quick Search box and the sophisticated Advanced

Table 3. The distribution of pathways in MetaCyc based on pathway ontology

Biosynthesis (902)	Secondary Metabolites Biosynthesis (351) Cofactors, Prosthetic Groups, Electron Carriers Biosynthesis (160) Amino acids Biosynthesis (105) Fatty Acids and Lipids Biosynthesis (101) Carbohydrates Biosynthesis (70) Hormones Biosynthesis (40) Amines and Polyamines Biosynthesis (32) Cell structures Biosynthesis (31) Nucleosides and Nucleotides Biosynthesis (20) Other Biosynthesis (16) Aromatic Compounds Biosynthesis (13) Siderophore Biosynthesis (7) Metabolic Regulators Biosynthesis (4) Aminoacyl-tRNA Charging (4)
Degradation/Utilization/Assimilation (639)	Aromatic Compounds Degradation (152) Amino Acids Degradation (113) Inorganic Nutrients Metabolism (72) Secondary Metabolites Degradation (58) Carbohydrates Degradation (52) Amines and Polyamines Degradation (40) Chlorinated Compounds Degradation (39) Carboxylates Degradation (30) Degradation/Utilization/Assimilation -Other (26) C1 Compounds Utilization and Assimilation (24) Fatty Acid and Lipids Degradation (18) Alcohols Degradation (14) Hormones Degradation (15) Aldehyde Degradation (12) Nucleosides and Nucleotides Degradation and Recycling (9) Cofactors, Prosthetic Groups, Electron Carriers Degradation (2)
Generation of precursor metabolites and energy (124)	Fermentation (34) Respiration (25) Chemoautotrophic Energy Metabolism (14) Methanogenesis (12) Electron Transfer (11) TCA cycle (9) Glycolysis (6) Photosynthesis (6) Pentose Phosphate Pathways (4) Other (3)
Detoxification (16)	Methylglyoxal Detoxification (8) Arsenate Detoxification (3) Acid Resistance (2) Mercury Detoxification (1)

Pathways classified directly under the top-level classes are not shown. The number of pathways within each class is presented in parentheses. Some pathways may be classified under more than one class.

Query Page. Search pages customized for finding genes/proteins/RNAs, chemical compounds, and pathways are relatively easy to use, yet enable the user to define multi-criteria searches (e.g. find proteins that satisfy specified constraints on their pI, molecular weight, cellular location, small-molecule ligand and chromosomal location). As part of the site redesign, the Regulatory Overview tool that depicts the complete regulatory network stored within a PGDB is now available through the Web site (currently only the desktop version of Pathway Tools supports painting of omics data on the Regulatory Overview).

Web accounts

Users of the BioCyc Web site can now create accounts in which they can store Web site display preferences, specify a default organism for queries, and define and save organism lists for comparative genomics operations.

New omics display and analysis functions

The desktop version of Pathway Tools enables users to graph omics data for selected genes or metabolites, and also provides over-representation analysis for determining whether certain ontology classes (including GO MetaCyc Pathway Ontology, etc.) are over-represented in gene lists and metabolite lists. A new X–Y plot style of tracks for the Pathway Tools genome browser allows the user to visualize ChIP-chip datasets against the genome. ChIP-chip intensity measurements can be visually correlated with promoters, gene positions and operon boundaries.

Customize pathway diagrams for PowerPoint or publications

The appearance of pathway pages can now be customized in many respects (Pathway → Customize Diagram). Options include setting the font size, determining which elements are included in the drawing (such as enzyme names and gene names), and deciding whether chemical structures are displayed. The pathway diagrams can be downloaded as screen-resolution GIF images (for import into PowerPoint presentations), or as high-resolution PostScript or PDF files for import into documents.

Apple port

Pathway Tools now runs on Apple computers.

HOW TO LEARN MORE ABOUT MetaCyc AND BioCyc

The BioCyc.org and MetaCyc.org Web sites provide several informational resources, including an online BioCyc guided tour (25), a MetaCyc user guide (26) and many Webinar videos that combine narration with online demonstration of different topics (27). We routinely host workshops and tutorials (on site and at conferences) that provide training and in-depth discussion of our software for beginning and advanced users. To stay informed about recent changes and enhancements to our software, join the BioCyc mailing list at <http://biocyc.org/subscribe.shtml>. A list of our publications is available online (28).

DATABASE AVAILABILITY

The MetaCyc and BioCyc databases are freely and openly available to all. See <http://biocyc.org/download.shtml> for download information. New versions of the downloadable data files and of the BioCyc and MetaCyc Web sites are released four times per year.

ACKNOWLEDGEMENTS

We thank Dr Carol Bult and Dr Alexei Evsikov from the Jackson Laboratory for their contribution of pathways from the MouseCyc database. We thank Dr Lindsay Eltis and Dr Hao-Ping Chen from the University of British Columbia for their contribution of pathways from the *Rhodococcus jostii* RHA1 database. We also thank Dr Malabika Sarker from SRI International for her contribution of the *Mycobacterium tuberculosis* mycolate biosynthesis pathway. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the preceding agencies.

FUNDING

National Institutes of Health, National Institute of General Medical Sciences (GM080746, GM077678, and GM75742); National Science Foundation, Division of Biological Infrastructure (grant number 0640769 to P.Z., A.K., K.D.). Funding for open access charge: National Institutes of Health, National Institute of General Medical Sciences.

Conflict of interest statement. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the preceding agencies.

REFERENCES

- Krieger, C.J., Zhang, P., Mueller, L.A., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S.Y. and Karp, P.D. (2004) MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.*, **32**, D438–D442.
- Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L. *et al.* (2010) Pathway Tools version 13.0: Integrated Software for Pathway/Genome Informatics and Systems Biology. *Brief. Bioinformatics*, In Press.
- Romero, P., Wagg, J., Green, M.L., Kaiser, D., Krummenacker, M. and Karp, P.D. (2004) Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.*, **6**, R2.1–R2.17.
- Christie, K.R., Weng, S., Balakrishnan, R., Costanzo, M.C., Dolinski, K., Dwight, S.S., Engel, S.R., Feierbach, B., Fisk, D.G., Hirschman, J.E. *et al.* (2004) Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res.*, **32**, D311–D314.
- Mueller, L.A., Zhang, P. and Rhee, S.Y. (2003) AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.*, **132**, 453–460.
- Liang, C., Jaiswal, P., Hebbard, C., Avraham, S., Buckler, E.S., Casstevens, T., Hurwitz, B., McCouch, S., Ni, J., Pujar, A. *et al.* (2008) Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.*, **36**, D947–D953.
- Evsikov, A.V. *et al.* (2009) MouseCyc: a curated biochemical pathways database for the laboratory mouse. *Genome Biol.*, **10**, R84.
- Seo, S. and Lewin, H.A. (2009) Reconstruction of metabolic pathways for the cattle genome. *BMC Syst. Biol.*, **3**, 33.
- Urbanczyk-Wochniak, E. and Sumner, L.W. (2007) MetaCyc: a biochemical pathway database for *Medicago truncatula*. *Bioinformatics*, **23**, 1418–1423.
- Fey, P., Gaudet, P., Curk, T., Zupan, B., Just, E.M., Basu, S., Merchant, S.N., Bushmanova, Y.A., Shaulsky, G., Kibbe, W.A. *et al.* (2009) dictyBase—a *Dictyostelium* bioinformatics resource update. *Nucleic Acids Res.*, **37**, D515–D519.
- Doyle, M.A., MacRae, J.I., De Souza, D.P., Saunders, E.C., McConville, M.J. and Likic, V.A. (2009) LeishCyc: a biochemical pathways database for *Leishmania major*. *BMC Syst. Biol.*, **3**, 57.
- May, P., Christian, J.O., Kempa, S. and Walther, D. (2009) ChlamyCyc: an integrative systems biology database and web-portal for *Chlamydomonas reinhardtii*. *BMC Genomics*, **10**, 209.
- Mazourek, M., Pujar, A., Borovsky, Y., Paran, I., Mueller, L. and Jahn, M.M. (2009) A dynamic interface for capsaicinoid systems biology. *Plant Physiol.*, **150**, 1806–1821.
- Snyder, E.E., Kampanya, N., Lu, J., Nordberg, E.K., Karur, H.R., Shukla, M., Soneja, J., Tian, Y., Xue, T., Yoo, H. *et al.* (2007) PATRIC: the VBI PathoSystems Resource Integration Center. *Nucleic Acids Res.*, **35**, D401–D406.
- Karp, P.D., Paley, S. and Romero, P. (2002) The Pathway Tools software. *Bioinformatics*, **18**(Suppl 1), S225–S232.
- Valdes, J., Veloso, F., Jedlicki, E. and Holmes, D. (2003) Metabolic reconstruction of sulfur assimilation in the extremophile *Acidithiobacillus ferrooxidans* based on genome analysis. *BMC Genomics*, **4**, 51.
- Kim, T.Y., Kim, H.U., Park, J.M., Song, H., Kim, J.S. and Lee, S.Y. (2007) Genome-scale analysis of *Mannheimia succiniciproducens* metabolism. *Biotechnol. Bioeng.*, **97**, 657–671.
- Aanensen, D.M., Mavroidi, A., Bentley, S.D., Reeves, P.R. and Spratt, B.G. (2007) Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J. Bacteriol.*, **189**, 7856–7876.
- Bernal, V., Carinhas, N., Yokomizo, A.Y., Carrondo, M.J. and Alves, P.M. (2009) Cell density effect in the baculovirus-insect cells system: a quantitative analysis of energetic metabolism. *Biotechnol. Bioeng.*, **104**, 162–180.
- Caspi, R., Foerster, H., Fulcher, C.A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S.Y., Shearer, A.G., Tissier, C. *et al.* (2008) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.*, **36**, D623–D631.
- Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetverin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S. *et al.* (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **37**, D5–D15.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat. Genet.*, **25**, 25–29.
- Wang, Y. *et al.* (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.*, **37**, W623–W633.
- Bairoch, A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.
- SRI International, BioCyc online guided tour, <http://biocyc.org/samples.shtml> (1 October 2009, date last accessed).
- SRI International, MetaCyc user guide, <http://www.metacyc.org/MetaCycUserGuide.shtml> (1 October 2009, date last accessed).
- SRI International, BioCyc webinars, <http://biocyc.org/webinar.shtml> (1 October 2009, date last accessed).
- SRI International, BioCyc publication list, <http://biocyc.org/publications.shtml> (1 October 2009, date last accessed).