

Welcome and Introduction

Peter D. Karp, Ph.D.
Bioinformatics Research Group
SRI International

Pathway Tools Workshop

- **Workshop format**

- First three days formal presentations
- Last two days presentations and hackathon / working groups
- Let's accumulate suggestions for the working sessions

- **Hold workshops every 18 months**

Workshop Goals

- Hear great science
- Present new developments in Pathway Tools
- Learn what new directions PTools should take
- See how other groups are making use of PTools
- Learn of PTools extensions being developed by other groups
- Work together on new directions
- Learn how to program PTools

Workshop Will be Available Live via Internet

- **Via WebEx plus Conference Call**
 - WebEx shares the computer's desktop
- **Make your presentation from**
 - SRI laptop
 - Your WebEx-enabled laptop
- **If you are willing, email your slides to taltman@ai.sri.com for posting on workshop Web site**

SRI Campus is Secure

Pathway Tools Capabilities

- **Create and maintain an organism database integrating genome, pathway, regulatory information**
 - Computational inference tools
 - Interactive editing tools
- **Query and visualize that database**
- **Use the database to interpret omics data**
- **Comparative analysis tools**
- **Systems biology analyses**
- **Model repository**

Model Organism Databases / Organism Specific Databases

- **DBs that describe the genome and other information about an organism**
- **Every sequenced organism with an active experimental community requires a MOD**
 - Integrate genome data with information about the biochemical and genetic network of the organism
 - Integrate literature-based information with computational predictions
- **Curated by experts for that organism**
 - No one group can curate all the world's genomes
 - Distribute workload across a community of experts to create a community resource

Rationale for MODs

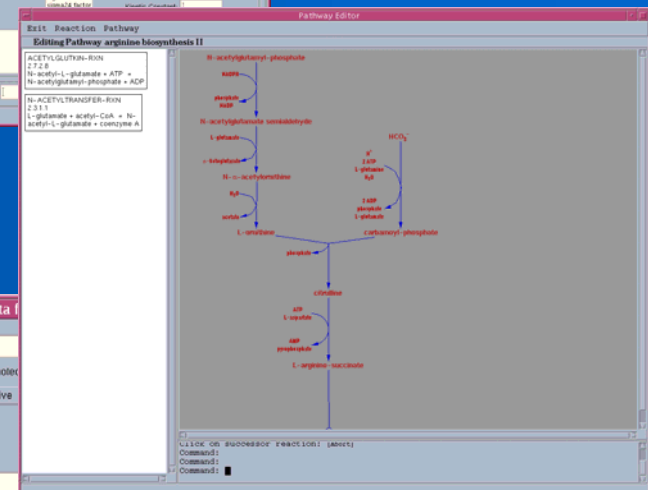
- **Each “complete” genome is incomplete in several respects:**
 - 40%-60% of genes have no assigned function
 - Roughly 7% of those assigned functions are incorrect
 - Many assigned functions are non-specific
- **Need continuous updating of annotations with respect to new experimental data and computational predictions**
 - Gene positions, sequence, gene functions, regulatory sites, pathways
- **MODs are platforms for global analyses of an organism**
 - Interpret omics data in a pathway context
 - *In silico* prediction of essential genes
 - Characterize systems properties of metabolic and genetic networks

Pathway Tools Software: PathoLogic

- **Computational creation of new Pathway/Genome Databases**
- **Transforms genome into Pathway Tools schema and layers inferred information above the genome**
- **Predicts operons**
- **Predicts metabolic network**
- **Predicts which genes code for missing enzymes in metabolic pathways**
- **Infers transport reactions from transporter names**

Pathway Tools Software: Pathway/Genome Editors

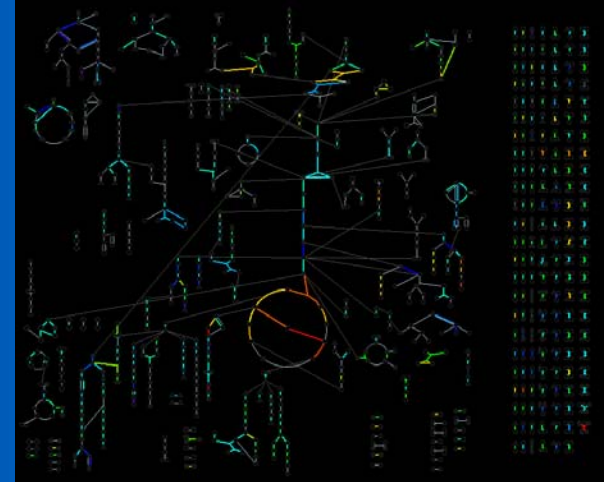
- Interactively update PGDBs with graphical editors
- Support geographically distributed teams of curators with object database system
- Gene and protein editor
- Reaction editor
- Compound editor
- Pathway editor
- Operon editor
- Publication editor



What is Curation?

- Ongoing updating and refinement of a PGDB
- Correcting false-positive and false-negative predictions
- Incorporating information from experimental literature
- Authoring of comments and citations
- Updating database fields
- Gene positions, names, synonyms
- Protein functions, activators, inhibitors
- Addition of new pathways, modification of existing pathways
- Defining TF binding sites, promoters, regulation of transcription initiation and other processes

Pathway Tools Software: Pathway/Genome Navigator



- Querying and visualization of:

- Pathways
- Reactions
- Metabolites
- Genes/Proteins/RNA
- Regulatory interactions
- Chromosomes



- Two modes of operation:

- Web mode
- Desktop mode
- Most functionality shared, but each has unique functionality

Obtaining a PGDB for Organism of Interest

- **Find existing PGDB**
 - In BioCyc
 - From third party
- **Create your own**

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>

Acknowledgements

●SRI

- Suzanne Paley, Ron Caspi, Ingrid Keseler, Carol Fulcher, Markus Krummenacker, Alex Shearer, Tomer Altman, Joe Dale, Fred Gilham, Pallavi Kaipa

●EcoCyc Collaborators

- Julio Collado-Vides, Robert Gunsalus, Ian Paulsen

●MetaCyc Collaborators

- Sue Rhee, Peifen Zhang, Kate Dreher
- Lukas Mueller, Anuradha Pujar

●Funding sources:

- NIH National Institute of General Medical Sciences
- NIH National Center for Research Resources

BioCyc.org

Learn more from BioCyc webinars: biocyc.org/webinar.shtml