

New Standard Resources for Systems Biology: BiGG Models Database and Visual Pathway Editing with Escher

Andreas Dräger

Introduction

- Genome-scale metabolic network reconstructions
 1. Simulation and analysis of complex biological networks
 2. Insights into how thousands of genes together influence cell phenotypes
- Requirements for accuracy in systems biology
 - Standards for model construction
 - Specific software tools
 - Access to high-quality metabolic networks



BIGG MODELS

BiGG Models

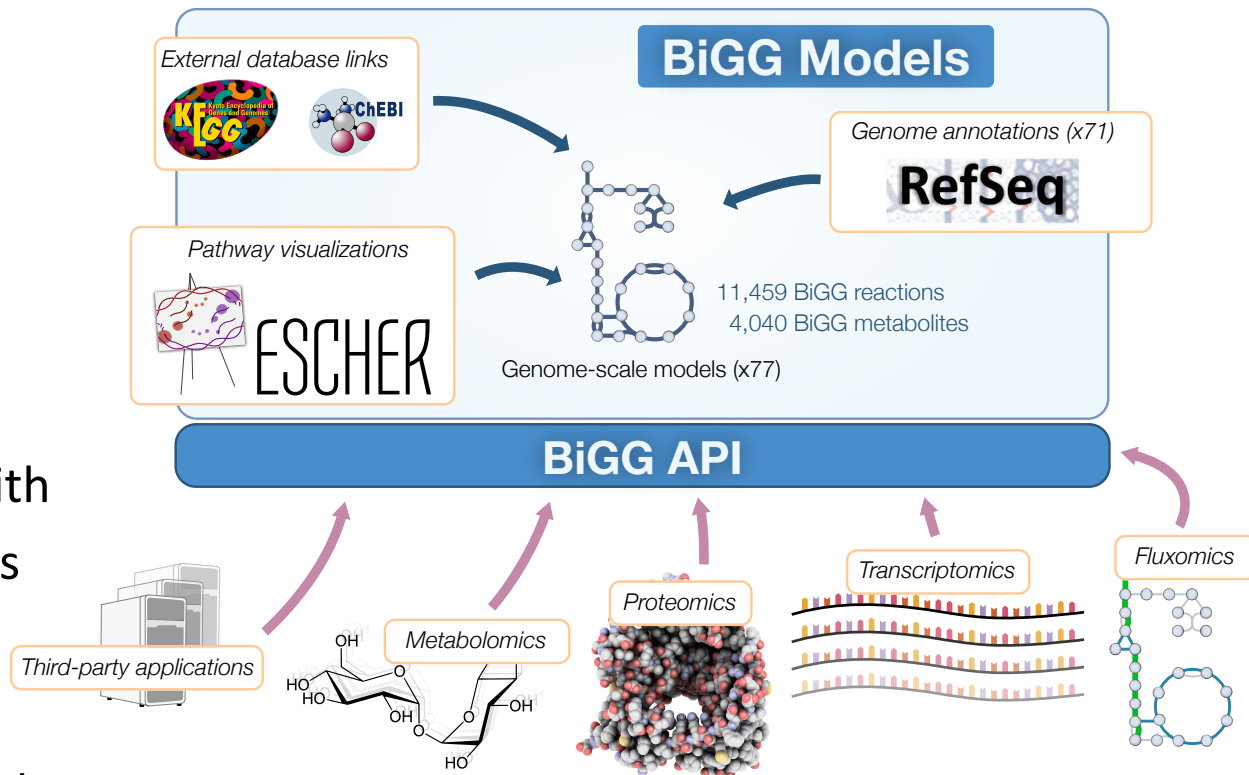
- Collection of software solutions for model
 - building
 - curation
 - visualization
 - simulation
- Currently > 75 genome-scale metabolic reconstructions
 - high-quality
 - manually-curated
- Search and browse functions
- Interactive pathway map visualization with the web-based Escher tool
- MIRIAM-compliant (minimal information required in the annotation of models)
 - BiGG Models usable as annotation resource
 - Database source code, API, and content freely available
 - New SBO terms specifically defined for more precise description of BiGG models
 - High-value bottom-up model building resource



<http://bigg.ucsd.edu>

BiGG Models Content

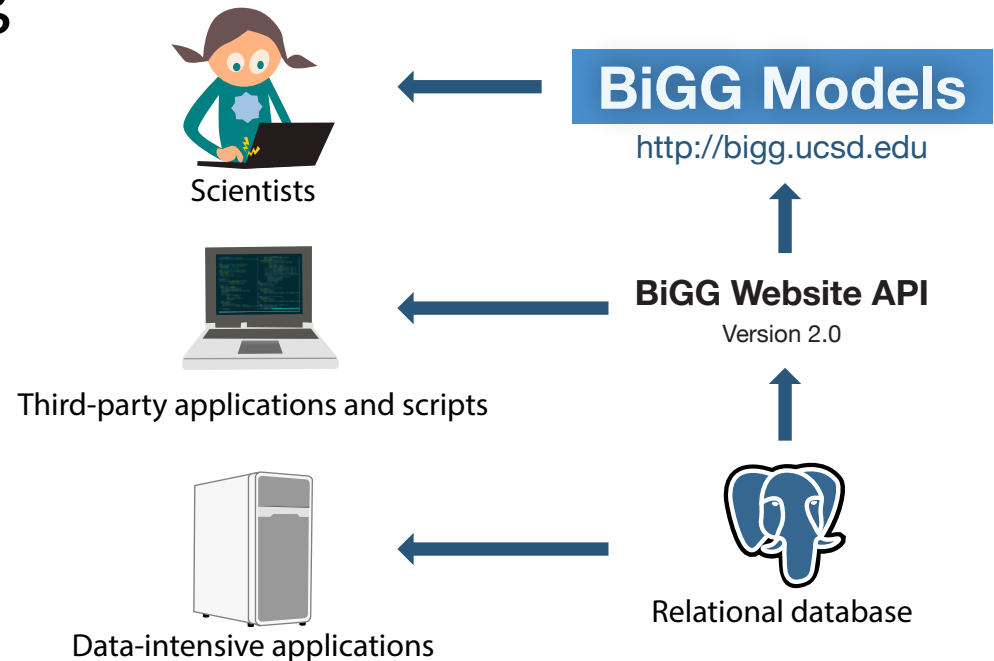
- Collection of currently 77 GEMs (Genome-Scale Model)
- Integrated into single database with shared reaction and metabolite identifiers
- Core database enriched with
 - external database links
 - Escher pathway maps
 - Genome annotation
- → resource for analysis and contextualization of many omics data types



<http://bigg.ucsd.edu>

Accessing BiGG

- User-friendly website for browsing and searching the knowledgebase
- Programmatical access through web API
- Local version of BiGG database for computationally intensive operations



<http://bigg.ucsd.edu>

Standardizing GEMs

- Standardization of metabolite and reaction IDs by removing extraneous characters and using a single format for referring to compartments
- Reaction ID clash → new identifier assigned to one of the reactions
- Manual correction of invalid gene-reaction rules
- Recording of all genes without corresponding genome annotation → future updates to both the GEMs and the genome annotations

- A** BiGG IDs follow a specification
- No special characters (e.g. [] () -)
 - Metabolite compartments defined in the database

Published GEM: `_g3p_c` `g3p(c)` `g3p[c]`

BiGG 2: metabolite *g3p*, compartment *c*

- C** Require valid Boolean logic for *gene reaction rules*
- Fixed 60 gene reaction rules

Published GEM: `(sll1102 and sll1103and sll1104)`

BiGG 2: `(sll1102 and sll1103and sll1104)`

- B** Reactions must be unique
- Fixed 251 conflicting reaction BiGG IDs

Published GEM: **ACPS1** `coa_c + apoACP_c → ACP_c + pap_c + h_c`
ACPS1 `coa_c + apoACP_c → ACP_c + pap_c`

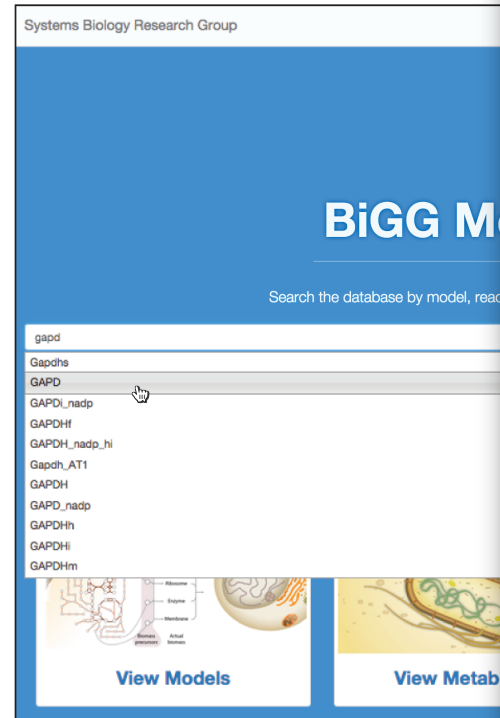
BiGG 2: **ACPS1** `coa_c + apoACP_c → ACP_c + pap_c + h_c`
ACPS1_1 `coa_c + apoACP_c → ACP_c + pap_c`

- D** Identified 1211 genes that do not map to genome annotations

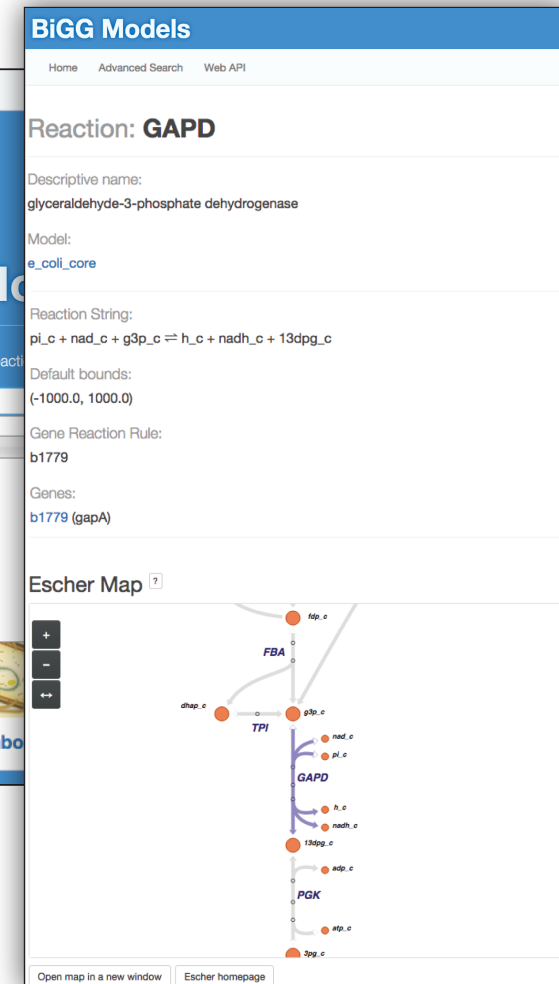
E.g. UNKNOWN from model iMM904

Search assistance

- Central search box for pages in BiGG Models, such as models and their reactions, metabolites, and genes
- Convenient links to the most popular pages about models, metabolites, and reactions below the search box
- “About” at the top of the page: general information about BiGG Models can be found by clicking



<http://bigg.ucsd.edu>



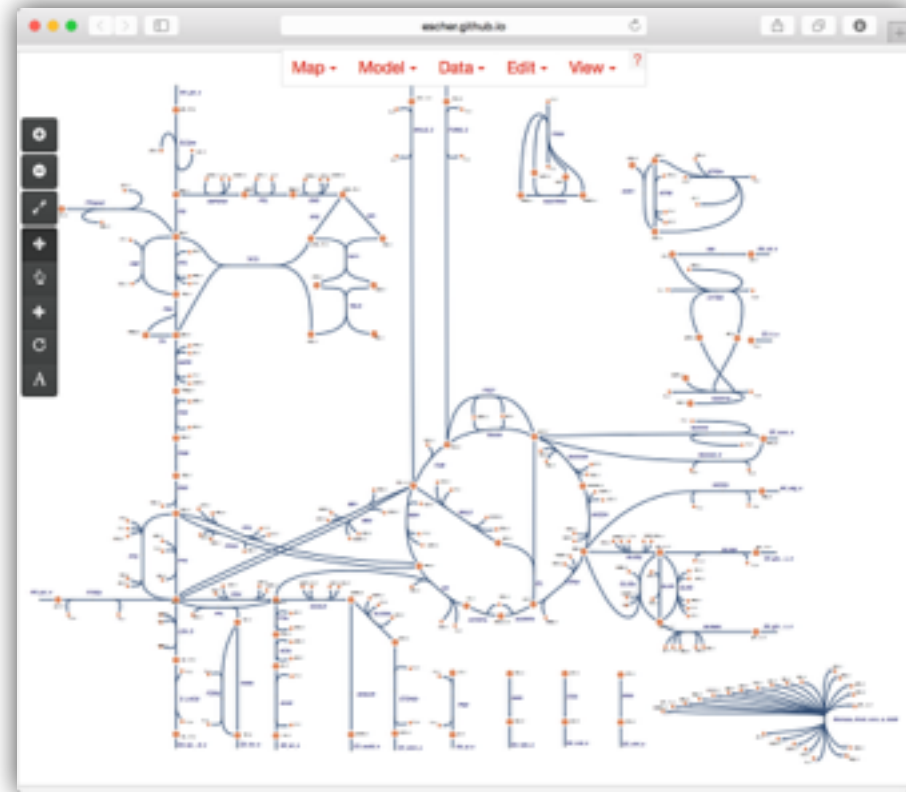


VISUAL PATHWAY EDITOR ESCHER

Escher: A Web application for visualizing data on biological pathways

Key features

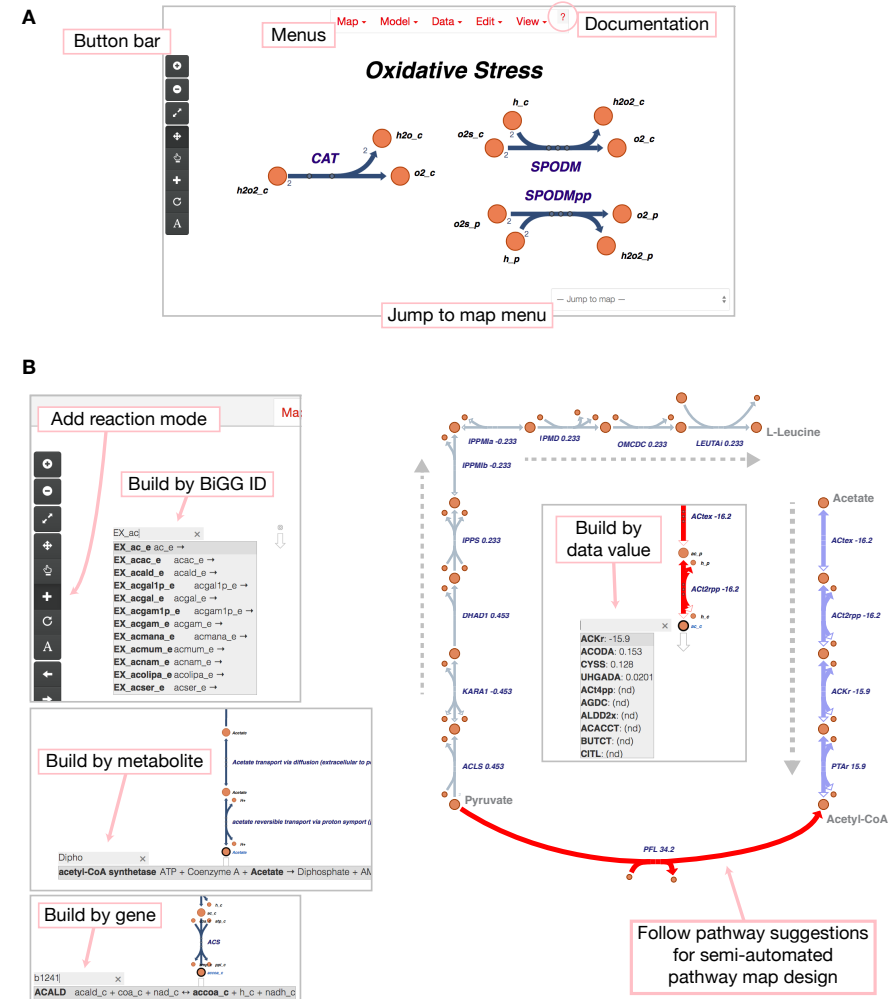
1. Semi-automated design of new pathway maps
 - pathway suggestions based on user data and genome-scale models
2. Visualization of data related to genes or proteins on associated reactions and pathways
 - gene-protein-reaction rules → identification of trends in data types (e.g., RNA-Seq, proteomics, ChIP)
 - metabolite- and reaction-oriented data types (e.g., metabolomics, fluxomics)
3. Rapid adaptation, extension, sharing, and embedding because of web technologies (SVG, D3, developer tools)



<http://escher.github.io>

Editing functions for pathway maps

- A. Set of menus with a link to the documentation, a button bar for accessing common features, and a menu for jumping to maps that were built with the same model
- B. Building pathway maps
 - Enter the Add reaction mode using the Edit menu or the button bar.
 - Click on the canvas or an existing metabolite to see a search menu.
 - Reactions can be searched by reaction ID, by metabolite, and by gene.
 - When a gene dataset or reaction dataset is loaded, suggestions appear for the reactions with the largest values in the dataset.

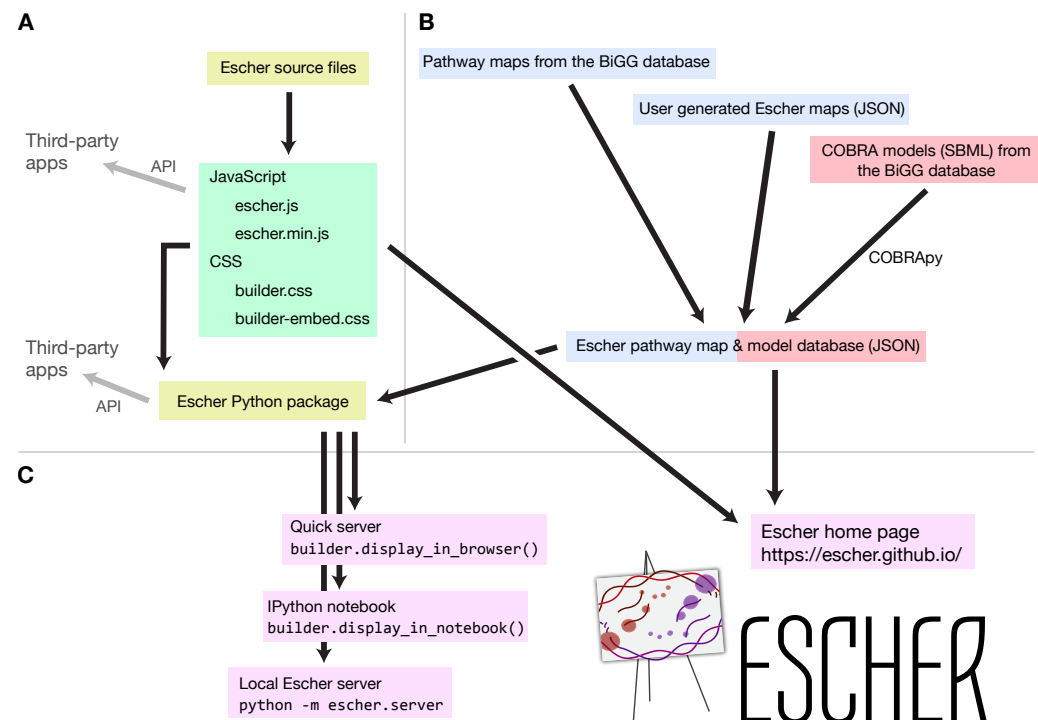


- A. The results of an *in silico* flux simulation visualized on the reactions
- B. Metabolomics data for *E. coli* aerobic growth visualized on the metabolites
- C. RNA-Seq data showing the shift from aerobic to anaerobic conditions in *E. coli*
 - Green: reactions downregulated in anaerobic growth
 - Red: gene upregulated in anaerobic growth, based on the log fold change.



Organization of the Escher project

- A. Escher source code can be compiled to a single JavaScript file (either minified or not minified) and two style sheets. The Python package is used to serve the Escher web application in various ways. APIs exist for both JavaScript and Python.
- B. Escher maps are generated from the BiGG database or built by users. COBRA models are generated using COBRApy.
- C. The Escher web application can be viewed on the Escher website, or, for local access, using various methods in the Python package

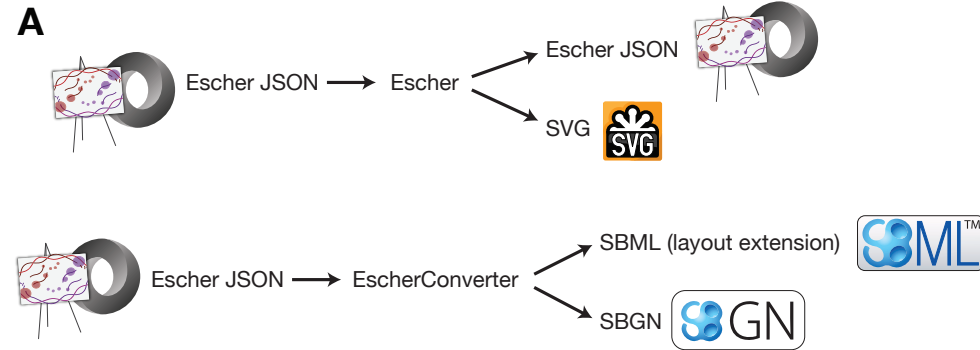




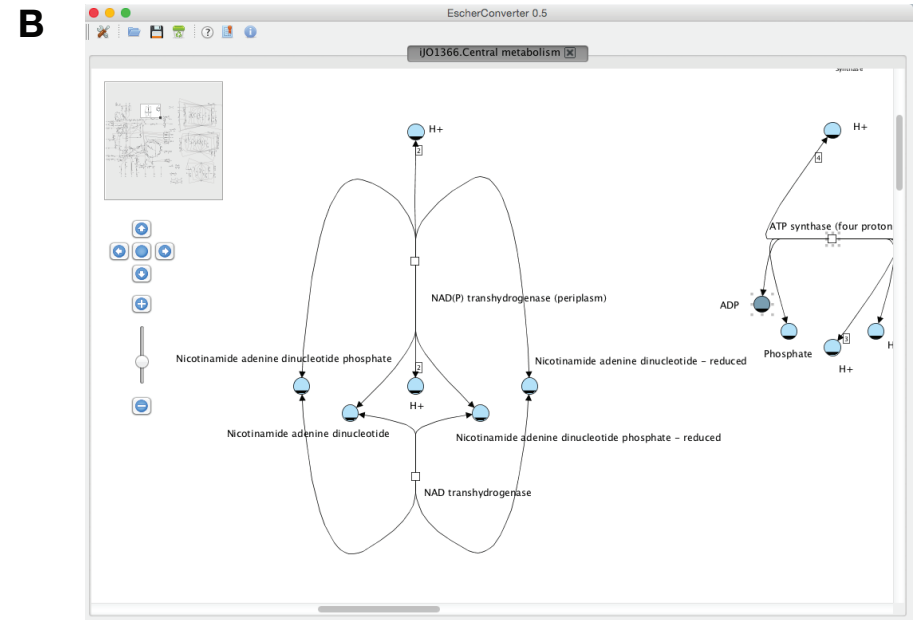
CONVERSION OF ESCHER MAPS TO SBGN-ML AND SBML LAYOUT

Import and export file types in Escher and the EscherConverter

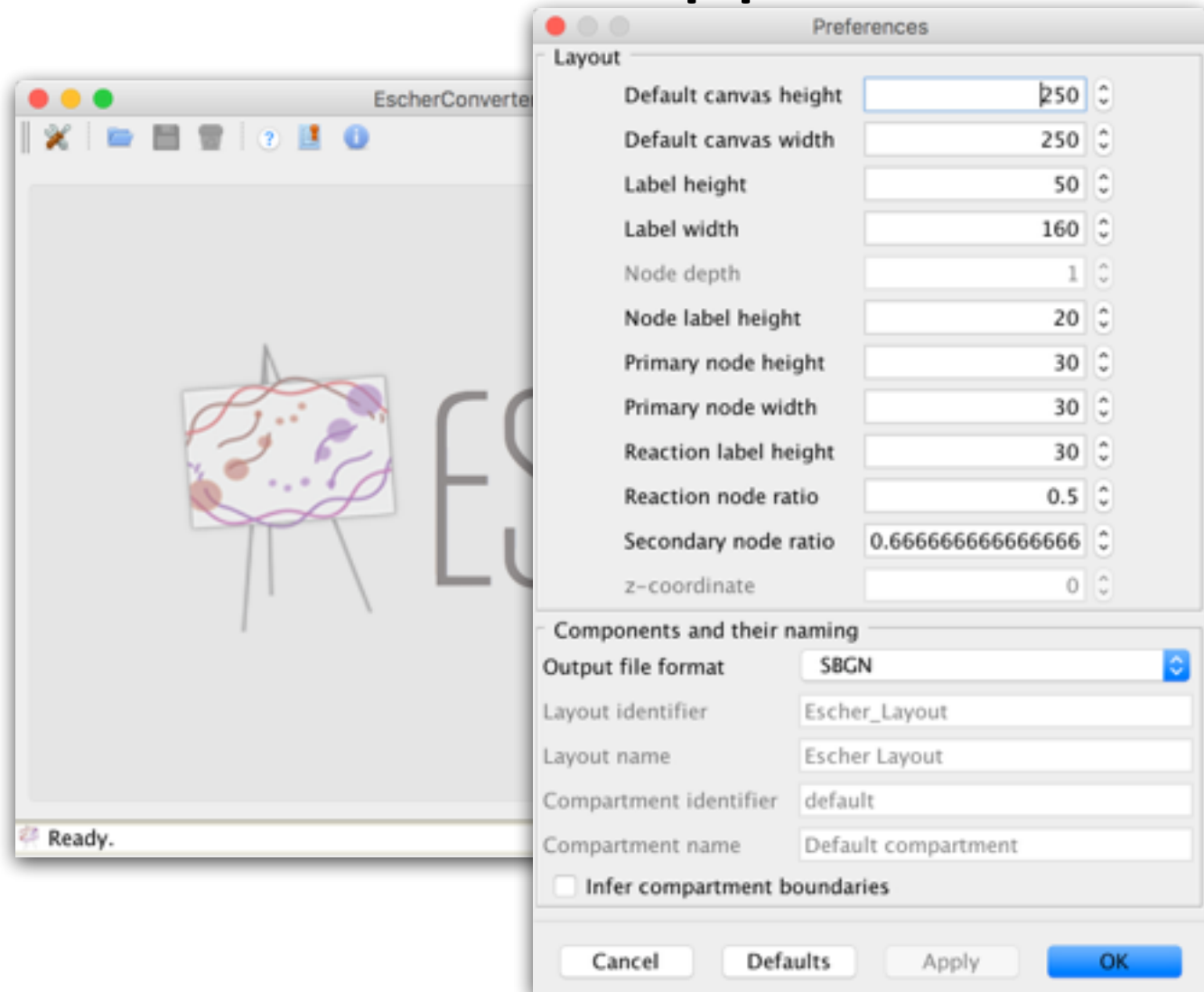
A. Escher can save to the Escher JSON file format or export to a SVG image. EscherConverter can be used to generate files in the SBML and SBGN-ML formats.



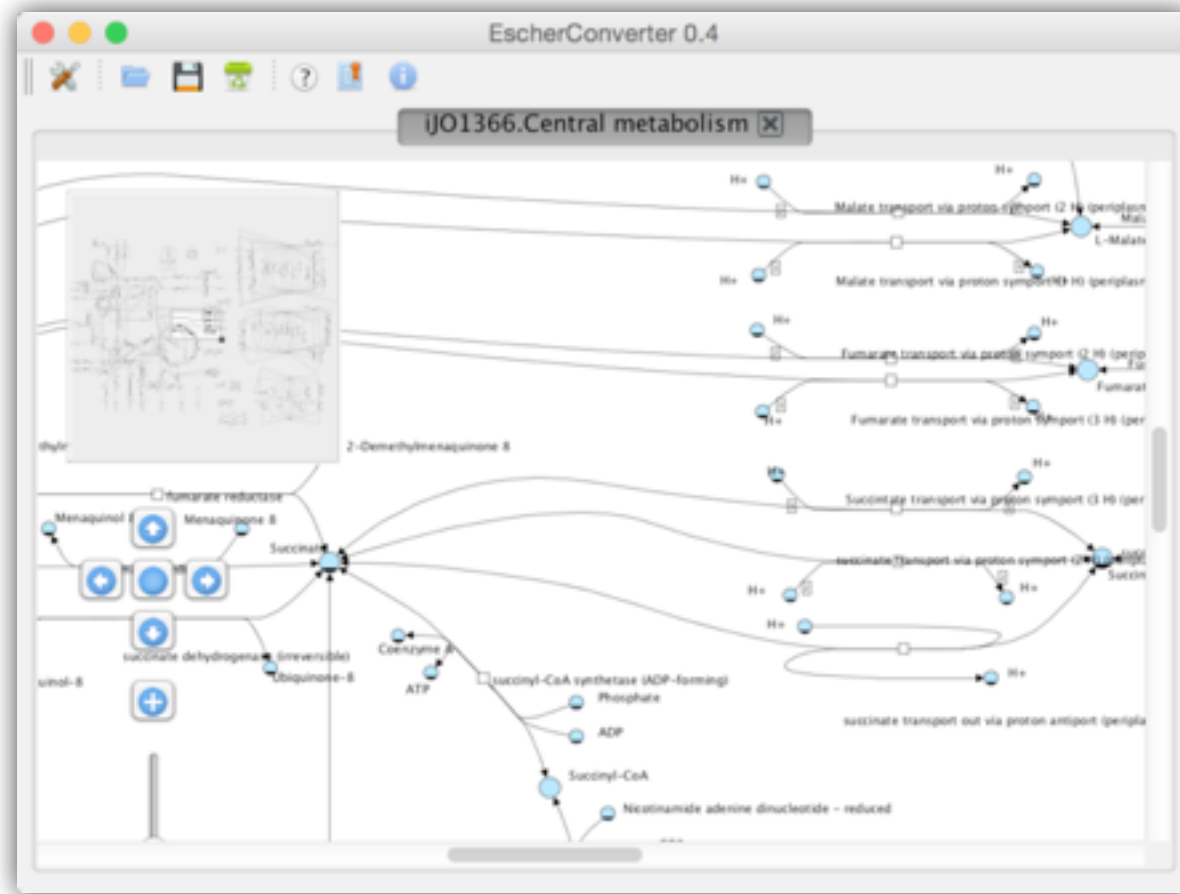
B. The EscherConverter graphical user interface.



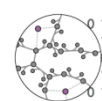
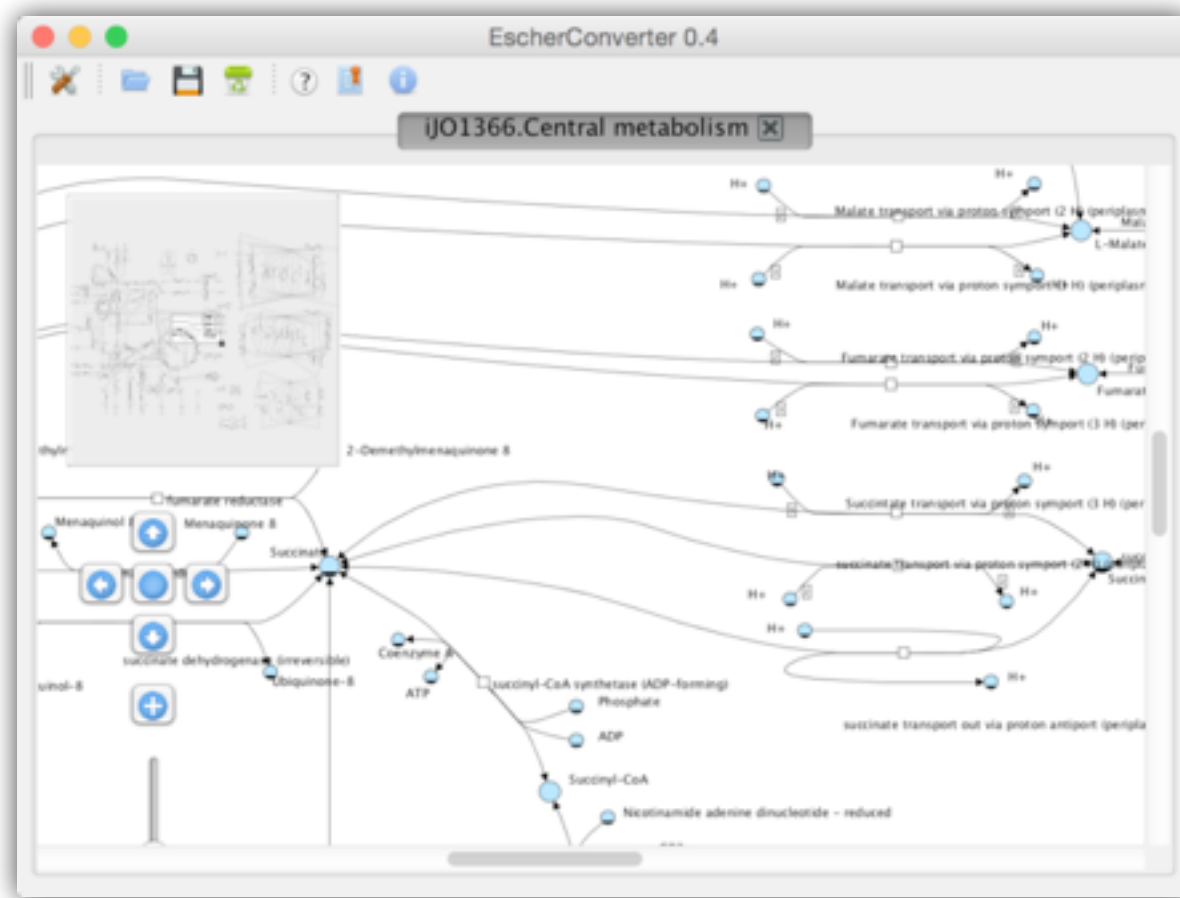
The EscherConverter App: Customization



Preview display



Preview display





MODEL POLISHING

COBRA Models

Problem:

- genome-scale models often difficult to comprehend for external researchers

Solution:

- Making use of wealth of information in BiGG
- Typos in external references (discovered with id pattern matching)
- Resolve errors in BiGG ids and unified identifiers accross models
- Automatic determination of missing information, e.g., compartments for species
- Add external references, e.g., to NCBI taxonomy
- Add specific SBO terms to model components
- Links to MIRIAM resources through identifiers.org

Result:

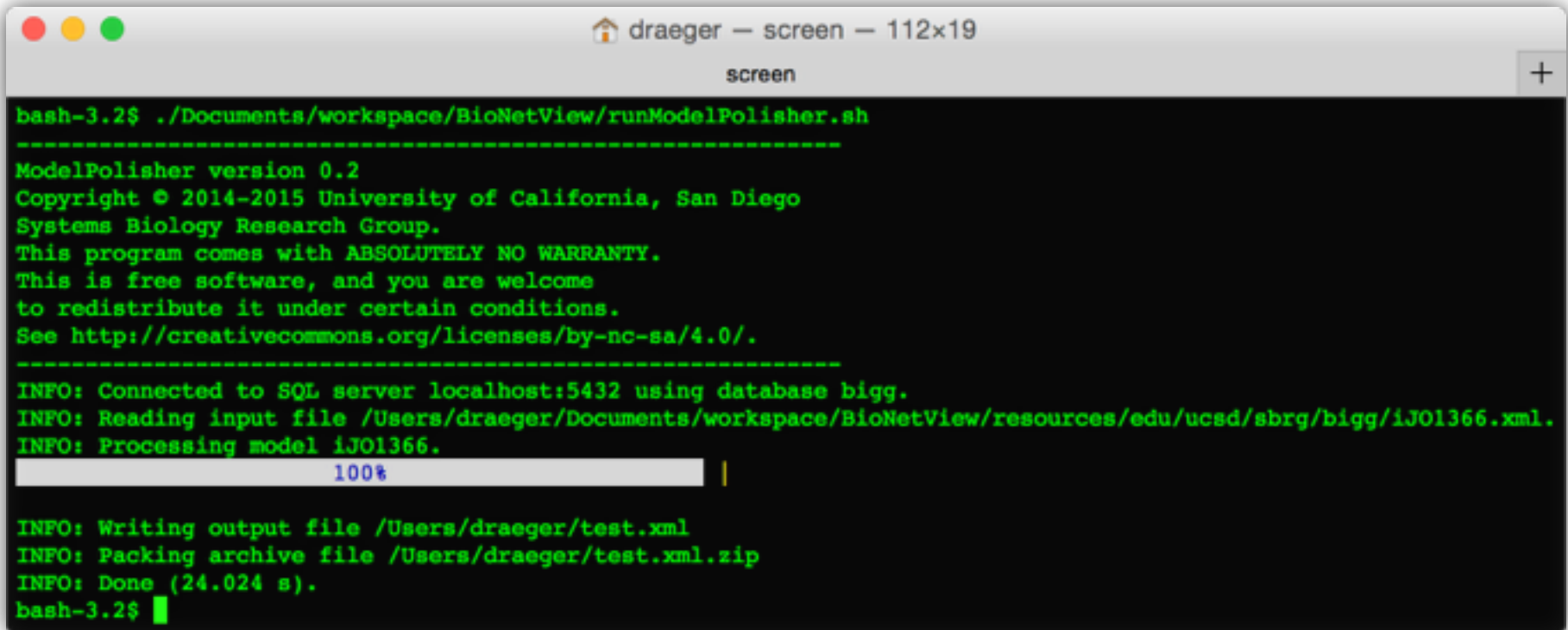
- More informative models
- Easier to interpret for external researchers, hence less ambiguous

Compartment codes

BiGG id	Name
c	cytosol
e	extracellular space
f	flagellum
g	golgi apparatus
h	chloroplast
l	lysosome
m	mitochondria
n	nucleus
p	periplasm
r	endoplasmatic reticulum
s	eyespot
u	thylakoid
v	vacuole
x	peroxisome/glyoxysome

Model Polisher

Input	SBML model(s) generated with COBRApy
Output	highly curated, well annotated SBML model(s) with all information from BiGG



```
draeger — screen — 112x19
screen
bash-3.2$ ./Documents/workspace/BioNetView/runModelPolisher.sh
-----
ModelPolisher version 0.2
Copyright © 2014–2015 University of California, San Diego
Systems Biology Research Group.
This program comes with ABSOLUTELY NO WARRANTY.
This is free software, and you are welcome
to redistribute it under certain conditions.
See http://creativecommons.org/licenses/by-nc-sa/4.0/.
-----
INFO: Connected to SQL server localhost:5432 using database bigg.
INFO: Reading input file /Users/draeger/Documents/workspace/BioNetView/resources/edu/ucsd/sbrg/bigg/iJO1366.xml.
INFO: Processing model iJO1366.
100%
INFO: Writing output file /Users/draeger/test.xml
INFO: Packing archive file /Users/draeger/test.xml.zip
INFO: Done (24.024 s).
bash-3.2$
```

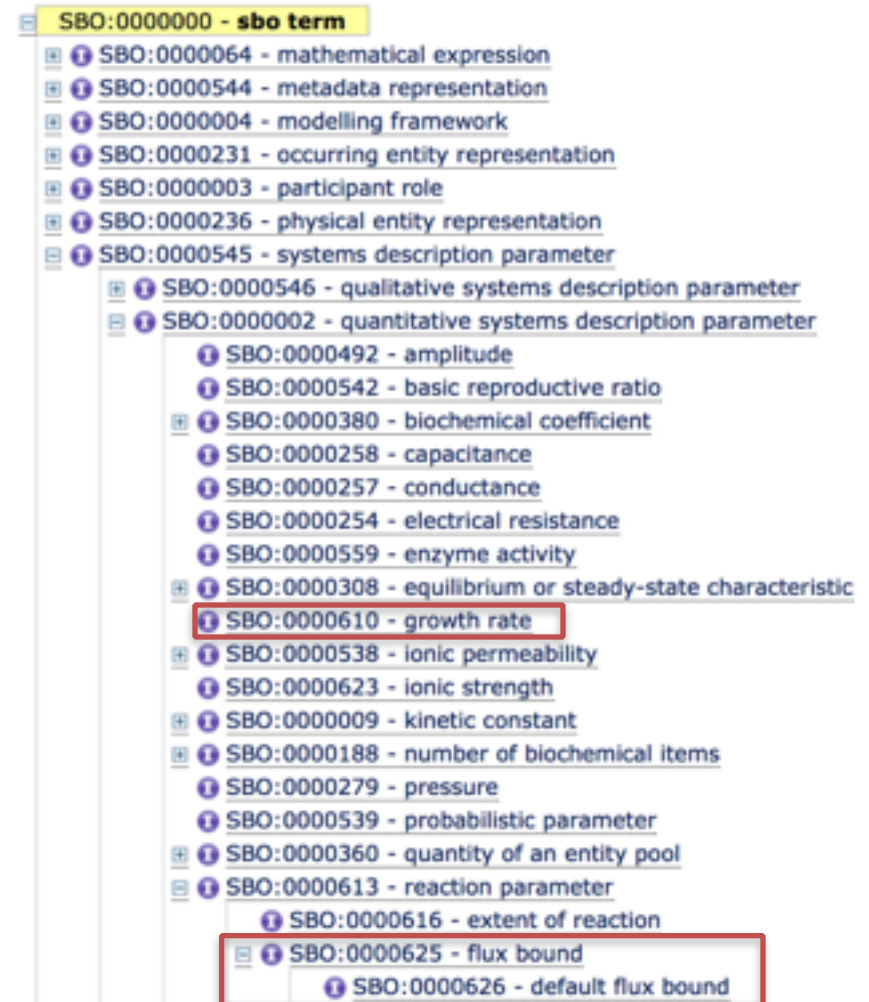
Adding semantics to models: new SBO terms for BiGG and COBRA

Aim:

- Better definition and easier interpretation of model components
- Discrimination flux bound parameters from, e.g., kinetic constants
- Reducing problems of semantic identifier overloading

Solution:

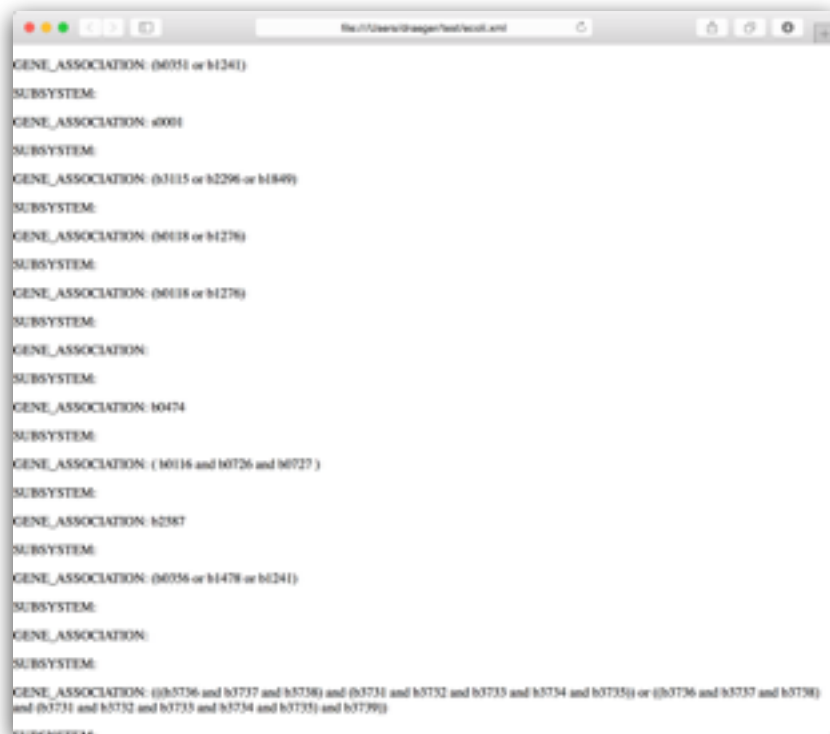
- New specific SBO terms for COBRA and BiGG models
- For the annotation of flux bounds:
 - (Default) flux bound (SBO:0000625 and SBO:0000626)
- For the annotation of reaction types:
 - exchange (SBO:0000627)
 - demand (SBO:0000628)
 - biomass production (SBO:0000629)
 - ATP maintenance (SBO:0000630)
- For the annotation of ME models.
 - growth rate (SBO:0000610)
 - effective catalytic rate (SBO:0000611)



Minimal Information Required in the Annotation of Models (MIRIAM)

- BiGG Database now part of MIRIAM registry:
 - <http://identifiers.org/bigg.model/>
 - <http://identifiers.org/bigg.compartment/>
 - <http://identifiers.org/bigg.metabolite/>
 - <http://identifiers.org/bigg.reaction/>
- All new BiGG models contain links to BiGG and unambiguously identify each component through its BiGG id
- New resources added to MIRIAM for use in BiGG models:
 - <http://identifiers.org/seed/> for subsystems
 - <http://identifiers.org/seed.compound/> for compounds
 - <http://identifiers.org/unipathway.compound/> for compounds

Display of models in web browser



```
GENE_ASSOCIATION: (b0351 or b1241)
SUBSYSTEM:
GENE_ASSOCIATION: s0004
SUBSYSTEM:
GENE_ASSOCIATION: (b1115 or b1296 or b1849)
SUBSYSTEM:
GENE_ASSOCIATION: (b0138 or b1278)
SUBSYSTEM:
GENE_ASSOCIATION: (b0138 or b1278)
SUBSYSTEM:
GENE_ASSOCIATION:
SUBSYSTEM:
GENE_ASSOCIATION: b0474
SUBSYSTEM:
GENE_ASSOCIATION: ( b0116 and b0726 and b0727 )
SUBSYSTEM:
GENE_ASSOCIATION: b2387
SUBSYSTEM:
GENE_ASSOCIATION: (b0356 or b1478 or b1241)
SUBSYSTEM:
GENE_ASSOCIATION:
SUBSYSTEM:
GENE_ASSOCIATION: (((b5736 and b0737 and b5738) and (b5731 and b5732 and b5733 and b5734 and b5735)) or ((b5736 and b5737 and b5738) and (b5731 and b5732 and b5733 and b5734 and b5735) and b5739))
SUBSYSTEM:
```

A COBRA model



ISB619 - Staphylococcus aureus subsp. aureus N315

Terms of use

Copyright © 2015 The Regents of the University of California.

Permission to use, copy, modify and distribute any part of this model from BiGG Database for educational, research, and non-profit purposes, without fee, and without a written agreement is hereby granted, provided that the above copyright notice, this paragraph, and the following three paragraphs appear in all copies.

Those desiring to incorporate this BiGG database or parts of its content into commercial products or use for commercial purposes should contact the

Technology Transfer & Intellectual Property Services
University of California, San Diego
8500 Gilman Drive # 0910
La Jolla, CA 92093-0910
United States of America
Phone: +1-858-534-5815
Fax: +1-858-534-7345
E-mail: ttvent@ucsd.edu

IN NO EVENT SHALL THE UNIVERSITY OF CALIFORNIA BE LIABLE TO ANY PARTY FOR DIRECT, INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES, INCLUDING LOST PROFITS, ARISING OUT OF THE USE OF THIS BiGG DATABASE, EVEN IF THE UNIVERSITY OF CALIFORNIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

THE BiGG DATABASE PROVIDED HEREIN IS ON AN "AS IS" BASIS, AND THE UNIVERSITY OF CALIFORNIA HAS NO OBLIGATION TO PROVIDE MAINTENANCE, SUPPORT, UPDATES, ENHANCEMENTS, OR MODIFICATIONS. THE UNIVERSITY OF CALIFORNIA MAKES NO REPRESENTATIONS AND EXTENDS NO WARRANTIES OF ANY KIND, EITHER IMPLIED OR EXPRESS, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, OR THAT THE USE OF THE BiGG DATABASE WILL NOT INFRINGE ANY PATENT, TRADEMARK OR OTHER RIGHTS.

Description

This is a metabolism model of Staphylococcus aureus subsp. aureus N315 in [SBML](#) format.

ISB619

The content of this model has been carefully created in a manual research effort. This file has been exported from the software [COBRApy](#) and further processed with a [JGEM](#)-based application.

This file has been produced by the [Systems Biology Research Group](#) and is currently hosted on [BiGG knowledgebase](#) and identified by: [ISB619](#).

References

When using content from BiGG database in your research works, please cite

Schellenberger, J., Park, J. C., Conrad, T. C., and Palsson, B. O. (2012). [BiGG, a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions](#), BMC Bioinformatics, 11:213.

B Polished BiGG model

Conclusions

Escher

- Easily usable graphical editing function for biochemical networks
- Multitude of data mapping functions
- Access to flux balance analysis (FBA) for knock-out analyses
- Bringing the gap between traditional display of networks and community standards
- API and open source

BiGG Models

- valuable database
- Pathway visualization
- Structured for easy access
- Improvement of quality, standardization, accessibility of all genome-scale models
- Boost of development of community standards for constraint-based modeling
- API and open source

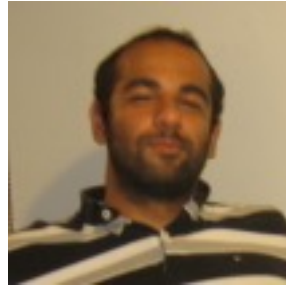
Acknowledgments



Zachary A.
King



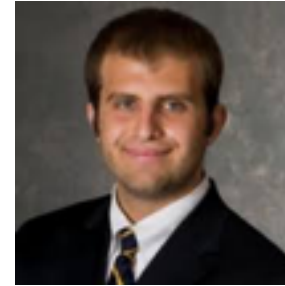
Justin S.
Lu



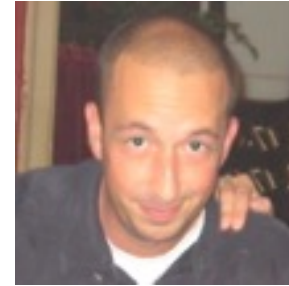
Ali
Ebrahim



Steve
Federowicz



Joshua A.
Lerman



Nikolaus
Sonnenschein



Philip C.
Miller



Bernhard O.
Palsson



Nathan E.
Lewis

Funding

- National Science Foundation Graduate Research Fellowship
- European Commission: Marie Curie International Outgoing Fellowship within the EU 7th Framework Program for Research and Technological Development (EU project AMBiCon)
- Novo Nordisk Foundation provided to the Center for Biosustainability at the Technical University of Denmark
- The Department of Energy
- National Institutes of Health