

UC San Diego

UC San Diego Previously Published Works

Title

BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions

Permalink

<https://escholarship.org/uc/item/5rf0w8df>

Journal

BMC Bioinformatics, 11(1)

ISSN

1471-2105

Authors

Schellenberger, Jan
Park, Junyoung O
Conrad, Tom M
[et al.](#)

Publication Date

2010-12-01

DOI

10.1186/1471-2105-11-213

Peer reviewed

BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions

Jan Schellenberger¹, Junyoung O Park², Tom M Conrad³ and Bernhard Ø Palsson*²

Abstract

Background: Genome-scale metabolic reconstructions under the Constraint Based Reconstruction and Analysis (COBRA) framework are valuable tools for analyzing the metabolic capabilities of organisms and interpreting experimental data. As the number of such reconstructions and analysis methods increases, there is a greater need for data uniformity and ease of distribution and use.

Description: We describe BiGG, a knowledgebase of Biochemically, Genetically and Genomically structured genome-scale metabolic network reconstructions. BiGG integrates several published genome-scale metabolic networks into one resource with standard nomenclature which allows components to be compared across different organisms. BiGG can be used to browse model content, visualize metabolic pathway maps, and export SBML files of the models for further analysis by external software packages. Users may follow links from BiGG to several external databases to obtain additional information on genes, proteins, reactions, metabolites and citations of interest.

Conclusions: BiGG addresses a need in the systems biology community to have access to high quality curated metabolic models and reconstructions. It is freely available for academic use at <http://bigg.ucsd.edu>.

Background

Metabolism is the structure and behavior of chemical reaction networks that occur in living organisms in order to maintain life. It is intrinsically linked to many other cellular functions and metabolic abnormalities are implicated as the cause of various diseases. Over the last 100 years, the list of reactions comprising an organism's metabolism has largely been catalogued. This reductionist process has focused on characterizing individual reactions in great detail. However, as the body of metabolic knowledge grew, so did the desire to integrate it into comprehensive models to simulate, predict and ultimately understand its behavior on a systems level. Kinetic models utilizing a system of differential equations are an established method of modeling biochemical pathways [1]. This field is an active area of research with an extensive number of models [2-5] as well as computational tools [6,7] available. Kinetic modeling suffers from the

difficulty of requiring comprehensive knowledge of kinetic parameters to sufficiently define the system. The parameters have proven difficult to measure in a consistent fashion and are often unknown [8,9]. A consequence is that the scope of kinetic models tends to be limited.

In contrast, constraint based modeling based on genome-scale metabolic reconstructions aim to include every known reaction for an organism, through the integration of genome annotation and biochemical knowledge. Reactions are defined simply by their reaction stoichiometry, and the networks are easily converted to mathematical models on which constraint-based analysis can be applied. In this paradigm, model predictions depend on constraints through reaction fluxes and an inferred metabolic objective, rather than on precisely defined kinetic parameters. Metabolic reconstructions have proven broadly useful for a number of applications. Case studies have been reviewed in [10].

In recent years, the publication of hundreds of genomes, with various databases such as KEGG [11], Biocyc [12] and Reactome [13] describing their annotation, has simplified the task of creating drafts of genome-scale

* Correspondence: palsson@ucsd.edu

² Bioengineering Department, University of California San Diego, 9500 Gilman Dr, La Jolla, California, 92093-0412, USA

Full list of author information is available at the end of the article

metabolic reconstructions [14,15]. This has spurred the development of an ever increasing number of reconstructions [16-22]. It is important to note that reconstructions derived directly from genome annotation may contain several gaps or incorrect annotations, leading to errors in model predictions. In order to be useful for prediction, models must undergo multiple rounds of manual curation and testing [23]. A number of widely-used manually-curated, component-by-component (bottom-up) reconstructions of genomic and bibliomic data have been published, creating the need for a systematized biochemically, genetically and genomically structured (BiGG) knowledgebase of metabolic reconstructions.

Model Reconstruction Process

A general bottom-up metabolic reconstruction process has been formulated and detailed in [16,17]. Initially, a parts list is assembled from existing databases (most notably KEGG [11], EntrezGene [24]) giving a crude reconstruction scaffold. This reconstruction is refined through an extensive review of primary literature, review articles, textbooks, and other specialized databases. A mathematical representation (**S** matrix) of the reconstruction is created and used to validate network structure by testing functionality, such as growth under some condition or the ability to produce a specific metabolite. Furthermore, gap analysis identifies possible missing reactions by finding so called 'dead end' metabolites which can be produced by the network but not consumed. Failure of network validation tests and the existence of gaps suggest targeted literature searches or experiments, which can be used to improve the model. Each reaction is verified individually and a confidence score can be assigned by the curator. A model may undergo several iterative rounds of validation and changes before it reaches a satisfactory state and is published, a process which can take up to a year of time. Because of the great effort involved, there have been attempts to partially automate the process [25-31] and split work through collaboration [32].

Gene-Protein-Reaction associations

Most biological reactions require enzymatic catalysis to occur. Thus the 'on' or 'off' state of each reaction in the network is controlled by the genotype and expression level of associated genes. In the simplest case, a reaction is catalyzed by only a single enzyme which is coded for by a single gene. The expression and translation of that gene implies the feasibility of the reaction, and vice versa. More complex cases involve multiple genes and proteins whose relationship is described using Boolean logic. A single protein may be composed of subunits coded by two (or more) genes. If all of these subunits are required for the catalytic activity of the protein, the activity is modeled

as an 'and' logic ('gene A and gene B'). Alternatively, the model allows for equivalent proteins (isozymes) to catalyze the same reaction. In this case, the presence of either protein is sufficient to establish the activity of the reaction and an 'or' logic is used ('protein A or protein B'). Other phenomena, which are representable in the Boolean framework, are alternative splicing ('or' logic) and obligate protein complexes ('and' logic). Collectively, these Boolean logic statements relating genes, proteins, and reactions are named GPRs. If a GPR statement of a reaction evaluates to 'true,' then its corresponding reaction is said to be feasible. Thus, GPRs may be used to evaluate the effects of gene knockouts and gene regulation on the metabolic reconstructions, ruling out reactions whose genes are not available. GPRs may also be displayed graphically. Figure 1 shows two of the possible GPR associations found in BiGG.

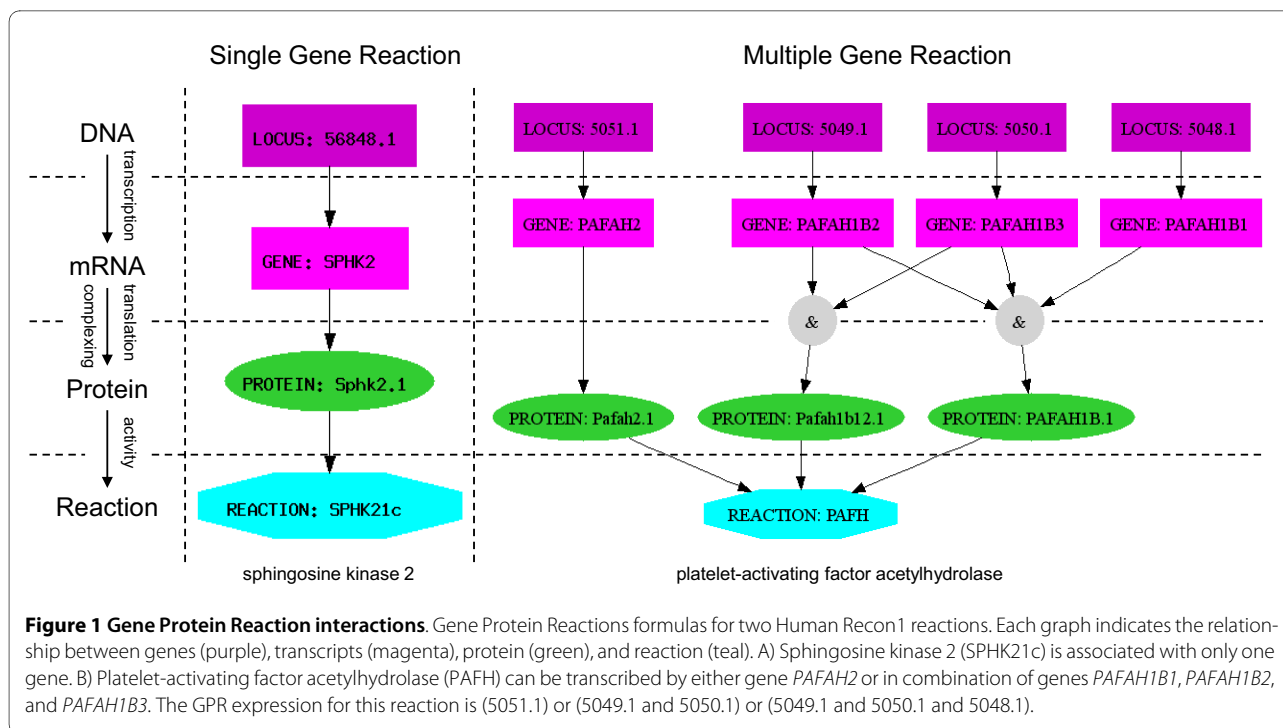
Construction and content Reconstructions

BiGG is currently capable of browsing and exporting the contents of seven different genome-scale reconstructions of six organisms (see Additional File 1): *Homo sapiens* Recon 1 [33], *Escherichia coli* iJR 904 [34] and iAF1260 [35], *Saccharomyces cerevisiae* iND750 [36], *Staphylococcus aureus* iSB619 [37], *Methanosarcina barkeri* iAF692 [38] and *Helicobacter pylori* iT341 [39]. These reconstructions span all three major branches of the tree of life and include two model organisms.

A global reconstruction of the human metabolic network, *H. sapiens* Recon 1, was recently completed [33]. The initial human reconstruction was based on gene information from the KEGG, EntrezGene, and H-Invitational [40] databases and was curated by evaluation of primary literature, reviews, and textbooks. Recon 1 represents a valuable tool as a scaffold for analysis of "omics" data sets.

A variety of microorganisms have also been reconstructed. The *E. coli* reconstructions, iJR904 and more recently iAF1260, are the most complete and most used of these reconstructions. iJR904 has been used for the prediction of adaptive evolution endpoints [41] and the engineering of lactate producing *E. coli* strains [42]. *H. pylori*, another Gram-negative enterobacteria that lives in the human stomach and has been shown to cause ulcers and gastritis, has a reconstruction, iT341. iAF692 is a reconstruction for the methanogenic archaeobacteria *M. barkeri*. iSB619 is a reconstruction of the infectious Gram-positive bacteria *S. aureus* of interest due to high rates of infection and increasing resistance to antibiotics. As more reconstructions are published, they will be added to BiGG.

All reconstructions in BiGG were developed on the Genomatica Simpheny™ platform. This system includes



quality control features to track genes, proteins and reactions, as well as simulation tools to computationally validate models. The models are built from a shared universal database of compounds and reactions. It is therefore not possible to incorporate reconstructions developed with other tools. The reconstructions are stored on a Genomatica (San Diego, CA) supplied server running an Oracle™ database. Access to this database is provided by a read-only client with several tables and views for accessing information on Reactions, Metabolites, Genes, Proteins, Maps and Citations (Figure 2). All queries are performed by a Linux/Apache/Perl Server using the CGI and DBI modules.

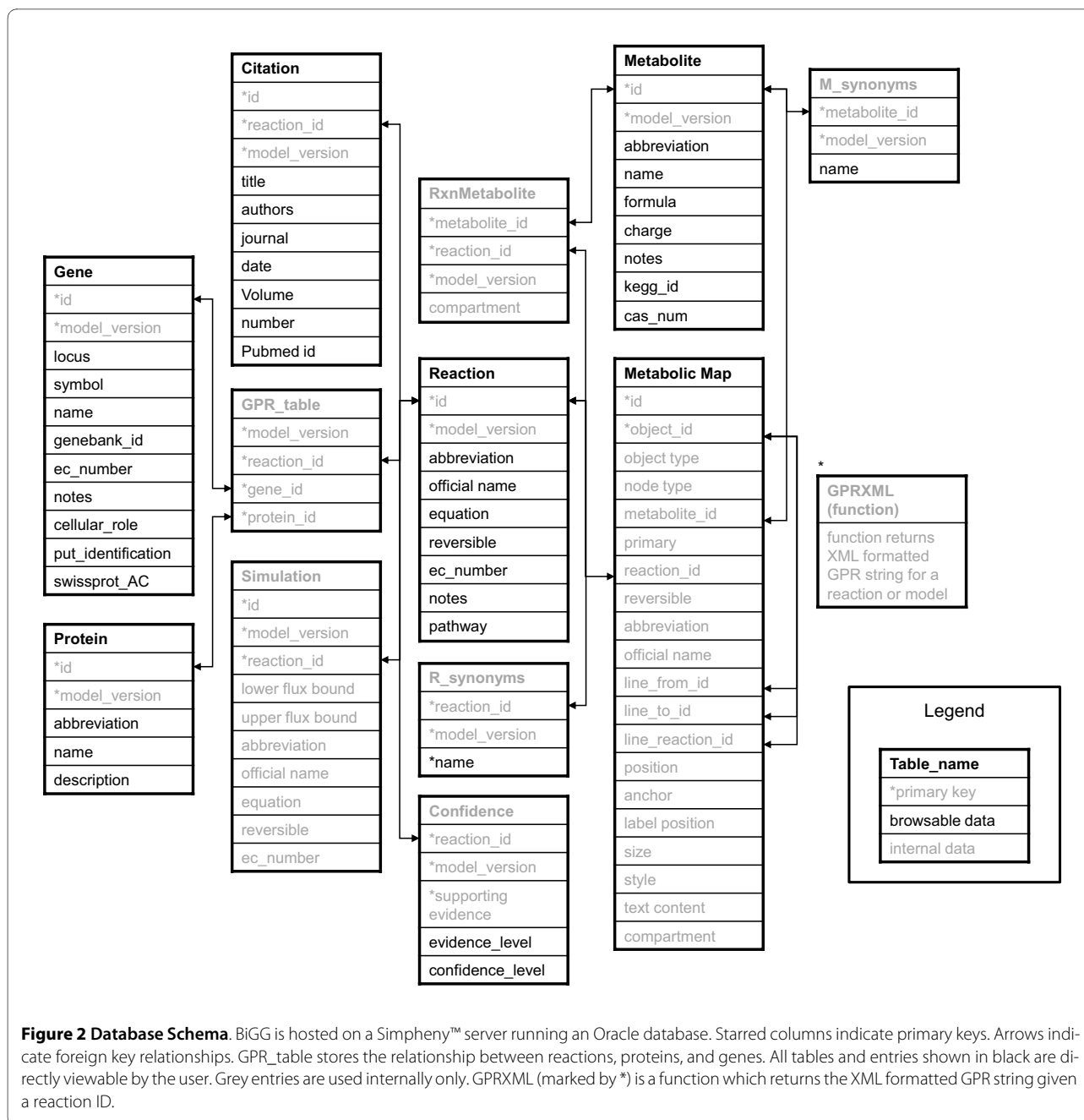
Browsing

The two main functions of BiGG are browsing content and exporting whole reconstructions. The browser is designed for querying the content and comparing different reconstructions whereas the exporter is primarily designed to enable further computational analysis by other software packages.

The BiGG browser contains entries for metabolic reactions, metabolites, genes, proteins, and literature citations (Figure 2). Reaction entries contain information such as the balanced equation, compartment localization, EC number [43], reversibility, author comments, and links to references. Metabolite entries contain information such as chemical formula and charge under physiological conditions. The GPR relationships are displayed as text or graphs using the graphviz package [\[viz.org\]\(http://viz.org\). Hyperlinks to other databases are included whenever provided by the authors of the reconstructions. These include NCBI Entrez gene database \[44\], Uniprot/Swissprot \[45\] for genes, and KEGG and CAS <http://www.cas.org> identifications for metabolites.](http://graph-</p>
</div>
<div data-bbox=)

Reactions and metabolites can be searched through the Search Reactions and Search Metabolites pages. Reactions may be searched for by name, EC number, or associated gene. Alternatively, all reactions in a model may be listed by using the model name as the only search parameter. It is also possible to specify compartment, pathway, or metabolite participation. Results may be limited by only including reactions with known gene associations, high or low confidence, or by excluding transport reactions. In addition, reactions may be searched across reconstructions allowing for model comparison. Lists of reactions matching a set of criteria may be exported as a tab delimited flat file. The exported files can contain information for multiple models, simplifying model comparison.

Metabolites may be searched for by name, KEGG ID, CAS ID, or charge. Just as for reactions, limiting searches by compartment, pathway, and organism is possible. In addition to basic metabolite information such as formula and charge, lists of reactions in which the metabolite participates are listed and categorized by the metabolite's role as a reactant or a product. Lists of metabolites matching a set of search criteria may be exported as tab delimited flat file, and contain information such as



metabolite name, abbreviation, formula, KEGG ID, and CAS ID. The browser interface is shown in Figure 3a.

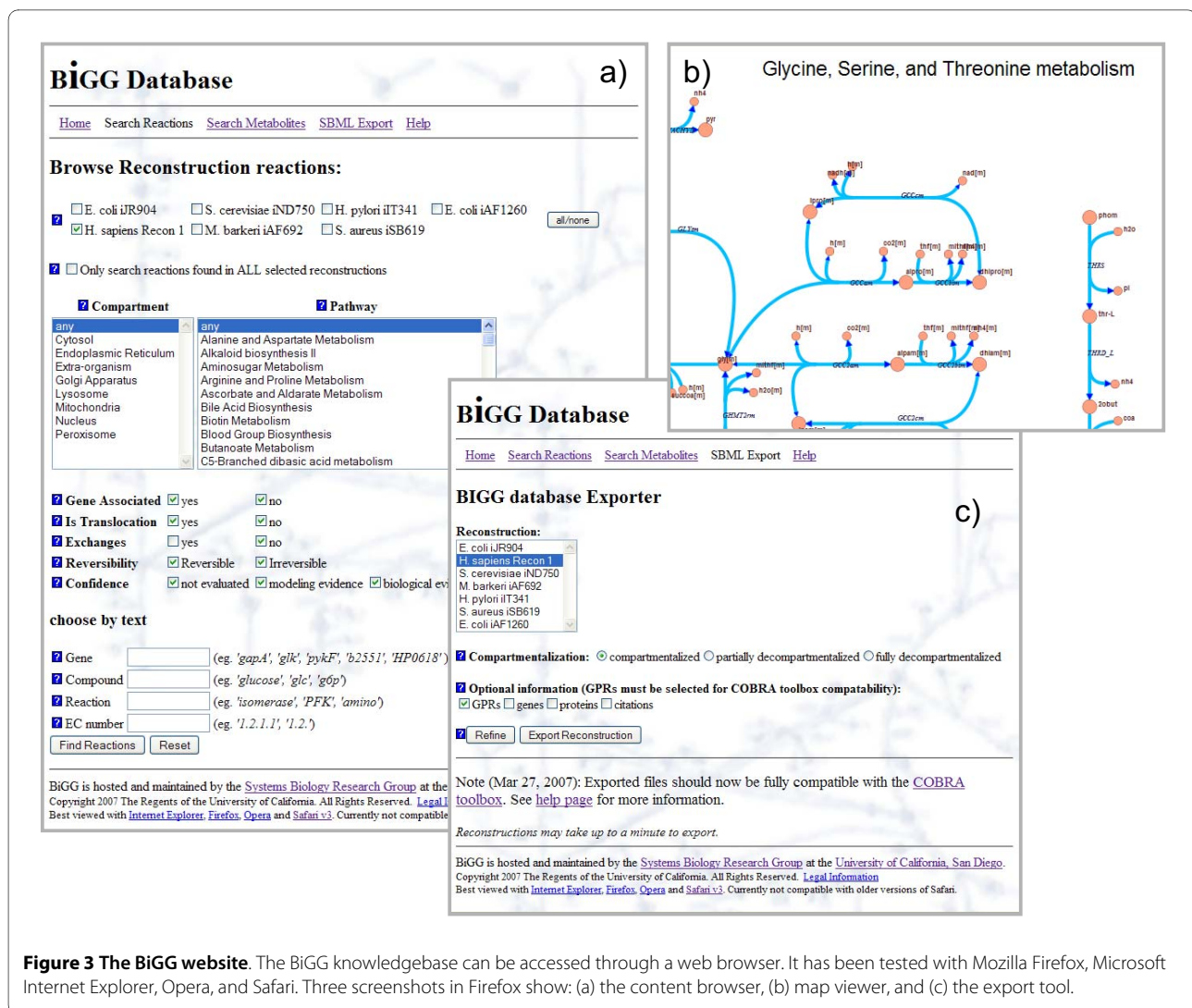
Maps

For visualization, curated metabolic maps are provided for many organisms in BiGG. These maps show metabolites, reactions, and text markup. Some large reconstructions (in particular human Recon 1) have several maps for different metabolic subsystems. Maps are rendered with Scalable Vector Graphics (SVG) for smooth scaling and

are hyperlinked back to the entries in the database. A small portion of a metabolic map from the human reconstruction is shown in Figure 3b.

Exporting

The second function of the BiGG knowledgebase is exporting reconstructions as Systems Biology Markup Language (SBML) files [46], which are specifically designed to work with the Matlab COBRA toolbox [47] and Systems Biology Research Tool [48] for performing



flux balance analysis and other computations. This XML format is widely used for distributing systems biology models [46]. By default, only entries for compartments, metabolites (the <species> tag), and reactions are included. The user has several options available to customize export as detailed below (see Figure 3c for interface).

Decompartmentalization

A reconstruction may be exported as a decompartmentalized model. A compartment in a metabolic reconstruction is a distinct pool of metabolites and their reactions. Metabolites are exchanged among compartments by transporter reactions. All reconstructions included in BiGG have at least two compartments: Cytosol and Extra-organism. Additionally, reconstructions of eukaryotic organisms have internal compartmentalization modeling subcellular organelles. A *partially decompartmentalized* reconstruction removes these internal compartments and relocates their reactions and

metabolites to the Cytosol. A *full decompartmentalization* removes internal compartments as well as the boundary between the Extraorganism and Cytosol compartments, creating a single-compartment system. In either case, internal transporters are removed.

It should be noted that the utility of decompartmentalization lies in model comparison rather than a basis for simulations. For instance, reactions such as ATP synthase are driven by an electrochemical gradient across compartment boundaries. In the decompartmentalized model there can be no gradients thus the ATP Synthase reaction becomes thermodynamically incorrect, creating unpredictable outcomes with some optimizations. As a rule, decompartmentalization should only be used for comparative purposes. Computational studies should only be performed on the full models.

Associated Genes, Proteins, and Citations

The exported SBML file may include information on genes, proteins and citations. Because the SBML specifi-

cation does not include fields for this kind of data, this information is stored in the 'notes' field of the reaction entries.

GPR statements

The notes field of the Reaction entries in the exported SBML file can include Boolean strings corresponding to the GPR statements. The GPR field is read and interpreted by the COBRA toolbox but should be ignored by other programs.

From Reconstructions to Models

Reconstructions are the basis for computational models. The process of converting a reconstruction into a model is performed by the curator and is reviewed in [16,49]. Upper and lower bounds are placed on reaction rates, bounding the space of flux distributions. An objective function is added, often corresponding to the biomass

production. The reconstruction, bounds and objective function together comprise the model exported by BiGG.

Most of the simulations are run by default under parameters simulating aerobic growth condition in glucose minimal medium. This is modeled by the constraints on fluxes of the model's exchange reactions. For instance, modeling of an aerobic environment with glucose minimal media must allow for glucose, oxygen, ammonium ion, salts, and other ions to be up-taken but other carbon sources only excreted. These bounds are included in the SBML file along with the objective coefficients of each reaction and flux distribution. For simulating other conditions there is a web based interface for changing the bounds on any reaction by pressing the "refine" button. In this way, SBML files corresponding to different media compositions can be created.

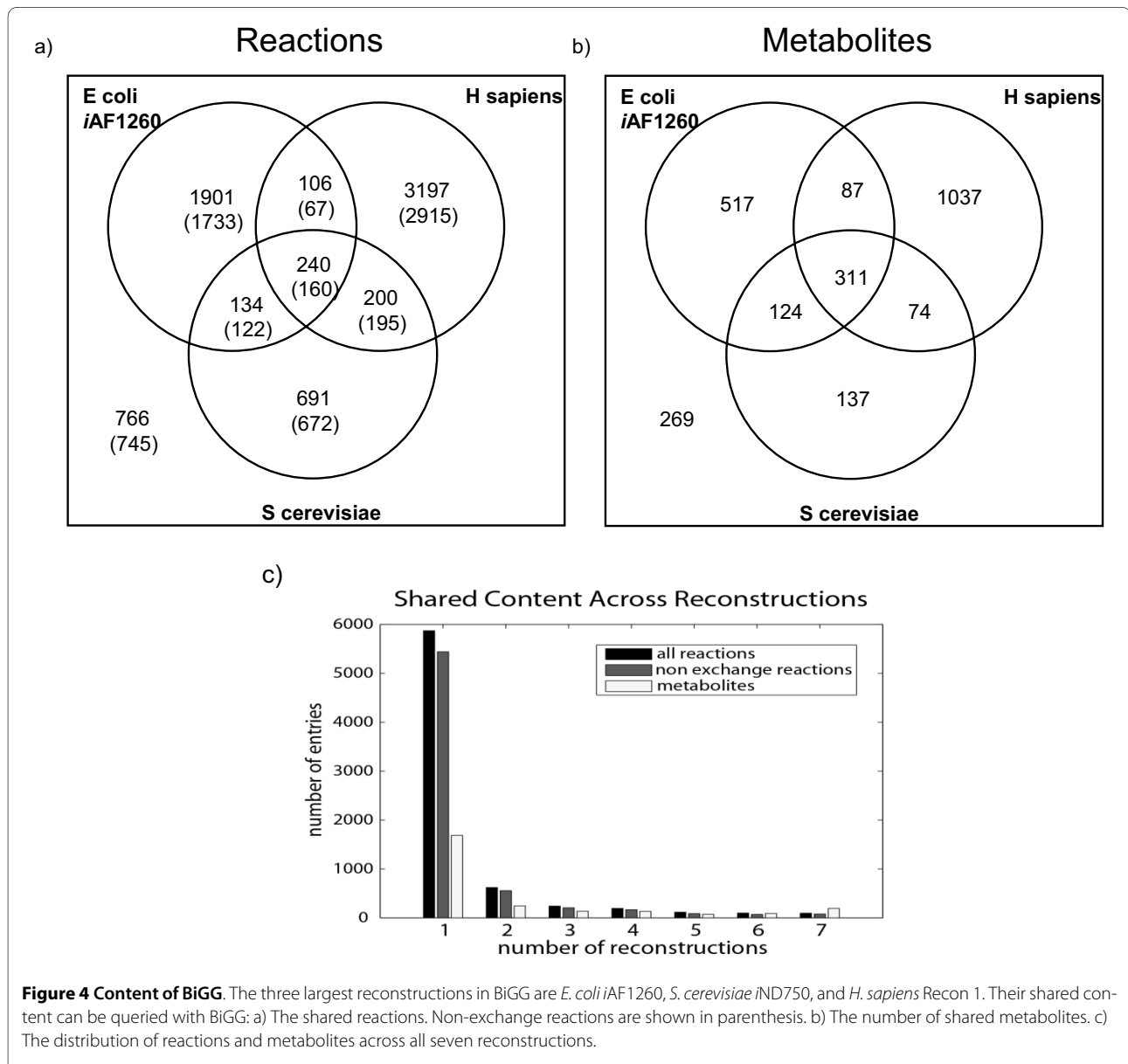


Figure 4 Content of BiGG. The three largest reconstructions in BiGG are *E. coli* iAF1260, *S. cerevisiae* iND750, and *H. sapiens* Recon 1. Their shared content can be queried with BiGG: a) The shared reactions. Non-exchange reactions are shown in parenthesis. b) The number of shared metabolites. c) The distribution of reactions and metabolites across all seven reconstructions.

Compatibility

SBML files conform to the level 2 version 1 specification and are compatible with the COBRA toolbox [47] which contains many computational procedures. Using the COBRA toolbox, the SBML file exported from BiGG may be imported as a network data structure into Matlab. This structure includes the stoichiometric matrix, gene and reaction information, and GPR associations. The toolbox then allows the user to interrogate the model's solution space using a variety of tools, including flux balance and flux variability analysis, random sampling, and robustness and gene deletion analysis. Matlab scripting can be used to combine methods or develop new methods not provided in the toolbox. The JAVA based Systems Biology Research tool [48] is another software package tested to work with the SBML files exported from BiGG.

Utility

Querying General Statistics of Reconstructions

The capability of BiGG to browse and compare multiple reconstructions was used to provide a comparison of the available reconstructions. The seven metabolic reconstructions available via BiGG vary in size from 551 reactions in *H. pylori*, to 3743 reactions in the human reconstruction. The total number of metabolites in BiGG is 2556, of which more than half (1509) are found in the Human Reconstruction (Additional File 1). A set of 96 core reactions is shared by all reconstructions, while most reactions were found in only one reconstruction (Figure 4c). Ubiquitous reactions include those involved in central metabolism, nucleotide and amino acid metabolism, and several exchange reactions (results not shown). Translocation reactions tend to be unique to particular reconstructions. The three largest reconstructions (*H. sapiens*, *E. coli*, *S. cerevisiae*) share a total of 240 reactions, 80 of which are exchange reactions (Figure 4a).

The content distribution usage in BiGG is shown in Figure 4C. Most reactions are only found in one reconstruction although 1167 are shared between at least two. Metabolites are shared more frequently. A smaller fraction of metabolites is unique to just one reconstruction (Figure 4b, c).

Case Study - Orphan reactions

All reconstructions have knowledge gaps where information on components is not available. One example is orphan reactions which are reactions for which the catalyzing enzyme is unknown. The BiGG knowledgebase can be used to study and help fill in these knowledge gaps by 1) listing all orphan reactions and 2) displaying any other reconstruction that use these reactions. *E. coli* metabolism has been studied extensively and most of the predicted open reading frames have at least putative functional assignments. The *E. coli* metabolic network

reconstruction has gone through several iterations and has become more complete [35]. The iJR904 reconstruction contains 58 orphan reactions (Additional File 2). Six are labeled spontaneous, meaning they can proceed without the aid of an enzyme and thus do not require an associated gene. A further reaction is the 'ATP maintenance requirement' which is a virtual reaction representing the turnover of ATP to ADP to maintain cellular functions. A total of seven reactions were removed in iAF1260, including two 'lumped' reactions (reactions which are stoichiometric representations of more complex processes) which are also not gene associated. These two have been replaced with elementary reactions. This leaves 44 reactions with missing gene associations in iJR904. Fourteen now have genes in iAF1260 while the remaining 30 do not. Twelve of these 30 are found in at least one other reconstruction, forming the basis for further searching. Analyses like these provide an overview of the state of reconstructions and can pinpoint areas of future focus. Performing this analysis without the BiGG knowledgebase would be possible although cumbersome.

Future Development

Additional reconstructions

Two notable reconstructions in development are *Bacillus subtilis* and *Haemophilus influenzae*. As they become available, they will be added to BiGG as well. Currently, only *E. coli* has more than one reconstruction version, but in the future, we plan on hosting different (older) versions of other reconstructions as well. Currently, it is only possible to host reconstructions created within the Simpheny software and at the moment there is no way to import other groups' reconstructions. This may change in the future.

Downloadable Maps

The BiGG knowledgebase is designed to work with the COBRA toolbox. Version 2.0 of this toolbox will be released soon and will include a visualization component. The BiGG maps will be downloadable in a custom text format containing coordinates of all metabolites and reaction control points. This is imported into COBRA and displayed in a customizable fashion. Colors and sizes can be changed on a per-reaction basis to visualize biological results.

Pre-made constraints/Media

At present, exported models contain one set of lower and upper flux bounds. Lower bounds of irreversible reactions are automatically set to 0 and upper bounds are either set to arbitrarily large values (eg. 999999) or physiologically determined rates. However, to run meaningful simulations, the bounds of the exchange fluxes must be specified to match the environment. For instance, modeling of an aerobic environment with glucose minimal media must allow for glucose, oxygen, ammonium ion,

and salts to be taken up, but not other extracellular species. Currently, this must be done manually via the export "refine" button, but in the future, libraries of bounds (constraint) vectors will be added to the SBML files to allow the user to specify media conditions.

Discussion and Conclusions

The reconstructions and models in BiGG have several specific features necessary to compute within the COBRA framework: 1) Each reconstruction in BiGG is manually curated. Exotic transformations unique to an organism may be absent from databases and must be pulled from primary literature. 2) BiGG uses both genetics and literature based data to assess whether a reaction is present. If the genetic basis for a reaction is unknown but the reaction is described in the literature, it will be included without associated genes (an "orphan reaction"). 3) The curators of BiGG reconstructions have the option of providing confidence levels for reactions which can be used when evaluating resultant models. These levels, along with reaction notes, provide an assessment of the confidence that a reaction is correctly included in the model. 4) Boolean relationships between genes, proteins and reactions (GPRs) are described in BiGG. This information is necessary for the proper modeling of mutations or gene knockouts. 5) All reactions in BiGG are mass and charge balanced. In some metabolic databases, simple species such as H^+ and H_2O are simply ignored [50]. Failure to balance reactions can result in unrealistic metabolic predictions. 6) Compartmentalization in BiGG gives an accurate description of reactions involving membrane transporters. This is required for simulation of gradient driven pumps [35]. 7) BiGG bridges the gap between a reconstruction and a model. The exported SBML files have all been validated and can be used to make predictions about growth rate, predicting the effect of gene deletions (MOMA [51]), and other COBRA framework methods. Taken together, these 7 features allow BiGG to represent metabolic reconstructions and the underlying chemistry in an accurate way. While individually these features are not unique to BiGG, no other resource including all of these features. The content of other genome-scale metabolic databases cannot be used directly for modeling in the COBRA framework [18]. The advent of genome sequencing has led to an explosion of systems biology methods which attempt to study properties of whole networks rather than individual parts. The results (often referred to as 'emergent properties') cannot be explained by studying the individual parts separately. Due to the scale of the models used, they are quite time consuming to develop and it is beneficial to share them with other researchers. The BiGG knowledgebase provides the first collection of curated, high quality meta-

bolic reconstructions suitable for study with COBRA methods. We expect it to continue to be a useful resource in the future as new and updated models are added to the database.

Availability and Requirements

The BiGG knowledgebase is available online at <http://bigg.ucsd.edu/>. A JavaScript enabled browser is required for browsing and exporting. The map viewer requires SVG support. BiGG data and results require a password which is made freely available for academic use.

Additional material

Additional file 1 Table 1 - Contents of BiGG. BiGG currently contains 7 reconstructions including two versions of *E. coli*. There are a total of 7234 unique reactions and exchanges in the database. Exchange reactions carry metabolites from the extracellular 'compartment' across the system boundary and are not technically part of the metabolic reconstruction. Translocation reactions carry a metabolite between compartments (possibly performing other transformations). Reactions can be gene associated or not. Every reconstruction contains the 'cytosol' and 'extracellular' compartment. Human and yeast contain 'endoplasmic reticulum', 'mitochondria', 'peroxisome', 'nucleus'. The 'periplasm' in iAF1260, vacuole in yeast, and lysosome in human are unique to these reconstructions.

Additional file 2 Table 2 - Orphan reactions. 58 *E. coli* iJR904 orphan (non-gene associated) reactions are categorized by current status. 'Found in 1260' shows whether the reactions are found in reconstruction iAF1260. 'Y*' indicates that the reaction is found in iAF1260 in modified form (usually because of the addition of the periplasm compartment). Notes: spont. - spontaneous, lump - lumped reaction, virtual - virtual reaction, P - *H. pylori*, S - *S. aureus*, Y - *S. cerevisiae*, H - *H. sapiens*,

Abbreviations

BiGG: Biochemically, Genetically and Genomically structured; CAS: Chemical Abstracts Service; COBRA: Constraint Based Reconstruction and Analysis; GPR: Gene, Protein, Reactions; KEGG: Kyoto Encyclopedia of Genes and Genomes; SBML: Systems Biology Markup Language.

Authors' contributions

JS built the database interface and content browser. TC and JP wrote the export tool and GPR visualization. JP wrote the map visualization. JS and TC drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Evelyn Travnik, Raj Thakar and Tom Fahland at Genomatica for helping with the database specifications and interface. Furthermore the authors thank I. Thiele, N. Duarte, S. Becker, N. Jamshidi and A. Feist for helpful suggestions on the user interface. JS is supported by a Ruth L. Kirschstein National Research Service Award - NIH Bioinformatics Training Grant No. GM00806-06. JP was supported by a Calit2 summer research scholarship.

Author Details

¹Bioinformatics Program, University of California San Diego, 9500 Gilman Dr, La Jolla, California, 92093-0419, USA, ²Bioengineering Department, University of California San Diego, 9500 Gilman Dr, La Jolla, California, 92093-0412, USA and ³Department of Chemistry and Biochemistry, University of California San Diego, 9500 Gilman Dr, La Jolla, California, 92093-0332, USA

Received: 4 March 2009 Accepted: 29 April 2010

Published: 29 April 2010

References

1. Segel IH: **Enzyme kinetics: behavior and analysis of rapid equilibrium and steady-state enzyme systems.** New York: Wiley; 1975.

2. Heinrich R, Rapoport TA: **Linear theory of enzymatic chains; its application for the analysis of the crossover theorem and of the glycolysis of human erythrocytes.** *Acta Biol Med Ger* 1973, **31**(4):479-494.
3. Wright BE, Gustafson GL: **Expansion of the kinetic model of differentiation in Dictyostelium discoideum.** *J Biol Chem* 1972, **247**(24):7875-7884.
4. Werner A, Heinrich R: **A kinetic model for the interaction of energy metabolism and osmotic states of human erythrocytes. Analysis of the stationary "in vivo" state and of time dependent variations under blood preservation conditions.** *Biomed Biochim Acta* 1985, **44**(2):185-212.
5. Le Novere N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B, *et al.*: **BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems.** *Nucleic Acids Res* 2006:D689-691.
6. Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC, *et al.*: **E-CELL: software environment for whole-cell simulation.** *Bioinformatics* 1999, **15**(1):72-84.
7. Ander M, Beltrao P, Di Ventura B, Ferkinghoff-Borg J, Foglierini M, Kaplan A, Lemerle C, Tomas-Oliveira I, Serrano L: **SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks.** *Syst Biol (Stevenage)* 2004, **1**(1):129-138.
8. Xia Y, Yu H, Jansen R, Seringhaus M, Baxter S, Greenbaum D, Zhao H, Gerstein M: **Analyzing cellular biochemistry in terms of molecular networks.** *Annu Rev Biochem* 2004, **73**:1051-1087.
9. Famili I, Mahadevan R, Palsson BO: **k-Cone Analysis: Determining All Candidate Values for Kinetic Parameters on a Network Scale.** *Biophys J* 2005, **88**(3):1616-1625.
10. Feist AM, Palsson BO: **The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli.** *Nat Biotechnol* 2008, **26**(6):659-667.
11. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Res* 1999, **27**(1):29-34.
12. Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahren D, Tsoka S, Darzentas N, Kunin V, Lopez-Bigas N: **Expansion of the BioCyc collection of pathway/genome databases to 160 genomes.** *Nucleic Acids Res* 2005, **33**(19):6083-6089.
13. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, *et al.*: **Reactome: a knowledgebase of biological pathways.** *Nucleic Acids Res* 2005:D428-432.
14. Edwards JS, Palsson BO: **Systems properties of the Haemophilus influenzae Rd metabolic genotype.** *J Biol Chem* 1999, **274**(25):17410-17416.
15. Hodgman C, Goryanin I, Juty N: **Reconstructing whole-cell models.** *Drug Discovery Today* 2001, **6**(15):S109-S112.
16. Reed JL, Famili I, Thiele I, Palsson BO: **Towards multidimensional genome annotation.** *Nat Rev Genet* 2006, **7**(2):130-141.
17. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO: **Reconstruction of biochemical networks in microorganisms.** *Nat Rev Microbiol* 2009, **7**(2):129-143.
18. Durot M, Bourguignon PY, Schachter V: **Genome-scale models of bacterial metabolism: reconstruction and applications.** *FEMS Microbiol Rev* 2009, **33**(1):164-190.
19. Oliveira AP, Nielsen J, Forster J: **Modeling Lactococcus lactis using a genome-scale flux model.** *BMC Microbiol* 2005, **5**:39.
20. Oberhardt MA, Puchalka J, Fryer KE, Martins dos Santos VA, Papin JA: **Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1.** *J Bacteriol* 2008, **190**(8):2790-2803.
21. Gonzalez O, Gronau S, Falb M, Pfeiffer F, Mendoza E, Zimmer R, Oesterheld D: **Reconstruction, modeling & analysis of Halobacterium salinarum R-1 metabolism.** *Mol Biosyst* 2008, **4**(2):148-159.
22. David H, Ozcelik IS, Hofmann G, Nielsen J: **Analysis of Aspergillus nidulans metabolism at the genome-scale.** *BMC Genomics* 2008, **9**:163.
23. Thiele I, Palsson BO: **A protocol for generating a high-quality genome-scale metabolic reconstruction.** *Nature protocols* 2010, **5**(1):93-121.
24. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** *Nucleic Acids Res* 2005:D54-58.
25. Karp PD, Paley S, Romero P: **The Pathway Tools software.** *Bioinformatics* 2002, **18**(Suppl 1):S225-232.
26. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, *et al.*: **The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.** *Nucleic Acids Res* 2005, **33**(17):5691-5702.
27. DeJongh M, Formisma K, Boillot P, Gould J, Rycenga M, Best A: **Toward the automated generation of genome-scale metabolic networks in the SEED.** *BMC Bioinformatics* 2007, **8**:139.
28. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formisma K, Gerdes S, Glass EM, Kubal M, *et al.*: **The RAST Server: Rapid Annotations using Subsystems Technology.** *BMC Genomics* 2008, **9**(1):75.
29. Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, Lajus A, Pascal G, Scarpelli C, Medigue C: **MaGe: a microbial genome annotation system supported by synteny results.** *Nucleic Acids Res* 2006, **34**(1):53-65.
30. Breitling R, Vitkup D, Barrett MP: **New surveyor tools for charting microbial metabolic maps.** *Nat Rev Microbiol* 2008, **6**(2):156-161.
31. Notebaart RA, van Enckevort FH, Francke C, Siezen RJ, Teusink B: **Accelerating the reconstruction of genome-scale metabolic networks.** *BMC Bioinformatics* 2006, **7**(1):296.
32. Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Bluthgen N, Borger S, Costenoble R, Heinemann M, *et al.*: **A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology.** *Nat Biotechnol* 2008, **26**(10):1155-1160.
33. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO: **Global reconstruction of the human metabolic network based on genomic and bibliomic data.** *PNAS* 2007, **104**(6):1777-1782.
34. Reed JL, Vo TD, Schilling CH, Palsson BO: **An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR).** *Genome Biology* 2003, **4**(9):R54.1-R54.12.
35. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO: **A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.** *Mol Syst Biol* 2007, **3**(121):.
36. Duarte NC, Herrgard MJ, Palsson B: **Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model.** *Genome Res* 2004, **14**(7):1298-1309.
37. Becker SA, Palsson BO: **Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation.** *BMC Microbiol* 2005, **5**(1):8.
38. Feist AM, Scholten JCM, Palsson BO, Brockman FJ, Ideker T: **Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri.** *Mol Syst Biol* 2006, **2**(2006.0004):1-14.
39. Thiele I, Vo TD, Price ND, Palsson B: **An Expanded Metabolic Reconstruction of Helicobacter pylori (iIT341 GSM/GPR): An in silico genome-scale characterization of single and double deletion mutants.** *J Bacteriol* 2005, **187**(16):5818-5830.
40. Fujii Y, Imanishi T, Gojobori T: **[H-Invitational Database: integrated database of human genes].** *Tanpakushitsu Kakusan Koso* 2004, **49**(11 Suppl):1937-1943.
41. Joyce AR, Fong SS, Palsson BO: **Adaptive Evolution of E. coli on Either Lactate or Glycerol Leads to Convergent, Generalist Phenotypes.** *International E Coli Alliance Second Annual Meeting: 2004; Banff, Alberta* 2004.
42. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO: **In silico design and adaptive evolution of Escherichia coli for production of lactic acid.** *Biotechnol Bioeng* 2005, **91**(5):643-648.
43. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D: **BRENDA, the enzyme database: updates and major new developments.** *Nucleic Acids Res* 2004:D431-433.
44. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** *Nucleic Acids Res* 2007:D26-31.
45. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, *et al.*: **The Universal Protein Resource (UniProt).** *Nucleic Acids Res* 2005:D154-159.
46. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, *et al.*: **The systems biology**

- markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003, **19**(4):524-531.
47. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ: **Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox.** *Nat Protocols* 2007, **2**(3):727-738.
 48. Wright J, Wagner A: **The Systems Biology Research Tool: evolvable open-source software.** *BMC systems biology* 2008, **2**:55.
 49. Price ND, Reed JL, Palsson BO: **Genome-scale models of microbial cells: evaluating the consequences of constraints.** *Nat Rev Microbiol* 2004, **2**(11):886-897.
 50. Poolman MG, Bonde BK, Gevorgyan A, Patel HH, Fell DA: **Challenges to be faced in the reconstruction of metabolic networks from public databases.** *Syst Biol (Stevenage)* 2006, **153**(5):379-384.
 51. Segre D, Vitkup D, Church GM: **Analysis of optimality in natural and perturbed metabolic networks.** *Proc Natl Acad Sci USA* 2002, **99**(23):15112-15117.

doi: 10.1186/1471-2105-11-213

Cite this article as: Schellenberger *et al.*, BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions *BMC Bioinformatics* 2010, **11**:213

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

