BIOINFORMATICS ORIGINAL PAPER

Vol. 28 no. 13 2012, pages 1752–1758
doi:10.1093/bioinformatics/bts267

Systems biology

Advance Access publication May 4, 2012

GEMSiRV: a software platform for GEnome-scale metabolic model simulation, reconstruction and visualization

Yu-Chieh Liao[1,*,†], Ming-Hsin Tsai[1,2,†], Feng-Chi Chen[1] and Chao A. Hsiung[1,*]

[1]Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan 350 and [2]Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Genome-scale metabolic network models have become an indispensable part of the increasingly important field of systems biology. Metabolic systems biology studies usually include three major components—network model construction, objective- and experiment-guided model editing and visualization, and simulation studies based mainly on flux balance analyses. Bioinformatics tools are required to facilitate these complicated analyses. Although some of the required functions have been served separately by existing tools, a free software resource that simultaneously serves the needs of the three major components is not yet available.

Results: Here we present a software platform, GEMSiRV (GEnome-scale Metabolic model Simulation, Reconstruction and Visualization), to provide functionalities of easy metabolic network drafting and editing, amenable network visualization for experimental data integration and flux balance analysis tools for simulation studies. GEMSiRV comes with downloadable, ready-to-use public-domain metabolic models, reference metabolite/reaction databases and metabolic network maps, all of which can be input into GEMSiRV as the starting materials for network construction or simulation analyses. Furthermore, all of the GEMSiRV-generated metabolic models and analysis results, including projects in progress, can be easily exchanged in the research community. GEMSiRV is a powerful integrative resource that may facilitate the development of systems biology studies.

Availability: The software is freely available on the web at http://sb.nhri.org.tw/GEMSiRV.

Contact: jade@nhri.org.tw

Received on January 9, 2012; revised on March 14, 2012; accepted on May 1, 2012

1 INTRODUCTION

Genome-scale metabolic networks (GEMs) are important for understanding systems-level cellular behaviors. These networks have been successfully applied to the studies of bacteria evolution, metabolic engineering, biological network analysis and generation and validation of novel biochemical hypotheses (Durot *et al*., 2009; Feist and Palsson, 2008; Mahadevan *et al*., 2011). With the growing availability of complete genomes, the number of GEMs is expected to increase rapidly. However, the genome-scale reconstructions have been progressing at a slow pace. Recently, a web-based resource (Model SEED) was built to speed up the construction of new GEMs by generating functional draft models on the basis of genomic sequences (Henry *et al*., 2010). While GEMs can be generated automatically through the Model SEED platform, an alternative, bottom-up approach (Thiele and Palsson, 2010) has been highly regarded and widely applied over the last decade because of its expert opinion-reinforced accuracy and iterative refinements that integrate both computational and experimental advantages (Becker and Palsson, 2005; Duarte *et al*., 2004; Feist *et al*., 2007; Navid and Almaas, 2009; Nogales *et al*., 2008; Puchalka *et al*., 2008; Suthers *et al*., 2009). However, a user-friendly tool that accommodates the requirements of the bottom-up approach (which includes recursive model refinements that go back and forth between computational and experimental analyses) remains unavailable. For example, the researchers may wish to incorporate gene expression data into the metabolic network by shutting down the reactions of which the responsible genes are not expressed. This function, as far as we understand, has not been provided by any existing tools.

Furthermore, tools that integrate functionalities of both GEM reconstruction and subsequent computational analyses are very rare. An important application of GEMs is flux balance analysis (FBA). FBA simultaneously monitors hundreds to thousands of biological reactions and yields the integrated outputs of a single cell according to the stoichiometric matrix and the biological/physicochemical constraints that are imposed on the GEM of interest (Orth *et al*., 2010). A limited number of software tools and packages have been developed to enable FBA, thus allowing quantitative prediction of cellular metabolisms. Examples include Acorn (Sroka *et al*., 2011), BioMet Toolbox (Cvijovic *et al*., 2010), CellNetAnalyzer (Klamt *et al*., 2007), COBRA (Becker *et al*., 2007), FBA-SimVis (Grafahrend-Belau *et al*., 2009), Model SEED (Henry *et al*., 2010), OptFlux (Rocha *et al*., 2010) and SBRT (Wright and Wagner, 2008) (Table 1). However, none of these tools has the dual function of FBA and GEM reconstruction. Meanwhile, free and open-source reconstruction tools, including METANNOGEN (Gille *et al*., 2007) and MetNetMaker (Forth *et al*., 2010) provide an environment for constructing metabolic networks based on the KEGG database. Yet these tools can neither perform FBA simulations nor import existing models for further curations. Although YANAsquare (Schwarz *et al*., 2007) is an integrated network reconstruction, visualization and analysis tool, it cannot perform FBA. Furthermore, the major data resource of these tools (the KEGG database) may include metabolic

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Table 1. Feature comparison of software tools for metabolic reconstruction and analysis

Software/feature	Reconstruction					Simulation			Visualization			Note
	Reference database	Model importing	Reconstruction refinement	Reconstruction to model conversion	Model exporting	User-friendly interface	Adjustable parameter	FBA	Map creation	Flux visualization	Gene expression visualization	
Acorn												Web-based interface
BioMet												Web-based resource
CellNetAnalyzer												Matlab is required
COBRA												Matlab is required
FBA-SimVis												VANTED (Junker *et al.*, 2006) add-on tool
Model SEED												Automatic tool for constructing functioning model drafts
OptFlux												Open-source tool allows designing metabolic engineering.
SBRT												Open-source tool supports for external software
METANNOGEN												KEGG is the primary information source
MetNetMaker												KEGG is the primary information source
YANAsquare												Performs elementary mode analysis and uses KEGG as resource
GEMSiRV												Open-source tool developed in this study

A dark/gray box indicates a feature has been fully/partly implemented in a software tool.

reactions that are mass- or charge imbalanced, which will seriously compromise the accuracy of FBA. Another drawback of these tools is that they require manual input of metabolic reactions (rather than automatic reconstruction or import of existing models for modifications), which is very time-consuming and error-prone in view of the scale of GEMs. In sum, these publicly available tools (listed in Table 1), though providing important functions in one or a few aspects, do not satisfy the multi-facet needs of metabolic network studies.

In view of the unmet needs of metabolic network studies, here we provide GEMSiRV (GEnome-scale Metabolic model Simulation, Reconstruction and Visualization), an integrated, downloadable resource that can be used locally for semi-automatic GEM reconstruction, manual curation, computational analyses and visualization. GEMSiRV has the following advantages: first, it can be run on the local computers/servers, thus the restrictions due to data size or internet speed can be minimized. Second, it enables easy import and manual curation of existing models. Expert opinions can be easily added into the draft networks generated by other tools (e.g. Model SEED) to increase their accuracy and reliability. Third, it allows researchers to perform FBA by incorporating a freely available linear programming solver (note that the user needs to install the solver before using the FBA function). Fourth, GEMSiRV can accept high-throughput experimental data for model reconstruction or refinement. Finally, GEMSiRV provides an interface for interactive, editable network visualization. The interface is particularly useful for gap filling, scientific discussions/publications and visualizing the changes in network when high-throughput experimental data are incorporated in the model. The GEMSiRV package is freely downloadable at http://sb.nhri.org.tw/GEMSiRV.

2 METHODS

GEMSiRV is implemented by using the Java language, which is becoming increasingly popular in the scientific community for its portability. Note that the GNU Linear Programming Kit (GLPK) is required (the user needs to install this solver) for GEMSiRV to perform FBA. To demonstrate the features of GEMSiRV, we have constructed a web resource that includes online help (http://sb.nhri.org.tw/GEMSiRV). The web resource contains instructions for software download, installation and operation, a repository of metabolic models, reference databases, metabolic maps and some useful toolboxes.

3 RESULTS

GEMSiRV consists of three main modules: (i) metabolic network reconstruction module, (ii) simulation module and (iii) visualization module. It allows importing, editing, simulation analyses and visualization of GEMs (Fig. 1).

3.1 Metabolic network reconstruction

3.1.1 Model importing and editing Existing GEMs can be used as references for the model construction of closely related species, and can greatly decrease the time and efforts required in building a new metabolic model. The capability of taking existing GEMs for model development is thus indispensable for an integrated systems biology tool. The Systems Biology Markup Language (SBML) has been developed as a standard format for data exchange between different systems biology tools. Although GEMs in the SBML format can be easily imported into a variety of tools (Hucka *et al*., 2003), information loss during SBML file import usually occur because it is difficult for existing tools to exhaustively accommodate the wide variety of model-specific data features (Durot *et al*., 2009). To tackle this issue, GEMSiRV provides four user-editable reference databases, all of which include the metabolites and metabolic reactions retrieved from a large number of existing GEMs (the details can be found in the following section). The loss of information can thus be compensated by manually adding or editing metabolic reactions with reference to these databases, and model-specific features can be tailored by the user. Note that such data integration and model editing are not provided by other existing tools.

3.1.2 Construction of reference databases Since a GEM is an assembly of biochemical reactions, a reference database that includes rich information regarding metabolites and reactions will greatly accelerate the process of model reconstruction. Considering that most of the published GEMs adopted the naming conventions of BiGG (Schellenberger *et al*., 2010), Model SEED (Henry *et al*., 2010), MetaCyc (Caspi *et al*., 2012) or KEGG, we collected metabolite and reaction information separately from these resources to build four metabolic reference databases (Table 2). The reference databases are freely available at the GEMSiRV website via the

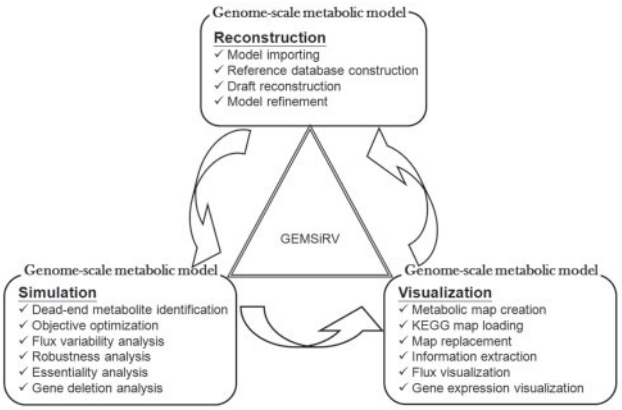

Fig. 1. A schematic overview of GEMSiRV. GEMSiRV includes three major modules (simulation, reconstruction and visualization). Each module contains several different functions that support metabolic network studies

Table 2. A summary of reference databases

Reference database	No. of metabolites	No. of reactions	Note
Ref_BiGG	2833	7107	6913 reactions are charge and formula balance
Ref_KEGG	16 399	8383	7443 reactions are formula balance
Ref_MetaCyc	9270	9946	5785 reactions are formula balance
Ref_Model SEED	15 304	12 723	11 528 reactions are formula balance

'Reference Databases' hyperlink (http://sb.nhri.org.tw/GEMSiRV/en/Reference_Databases). Since the information of metabolite charge is unavailable for KEGG's and Model SEED's compounds, 'not available (NA) for charge balance' will be noted when the reactions that include these metabolites are evaluated. Users can also start their own model reconstruction projects from scratch with the pre-constructed reference database if appropriate reference models are unavailable. For metabolites or reactions not present in the reference database, users can define theses metabolites/reactions and create their own databases using the GEMSiRV interface. In addition, the metabolites and reactions in the imported models can be overwritten or added to the reference databases. The customized reference databases can be saved and exported as spreadsheet files, which are easily portable and exchangeable. Note that unified nomenclature across the reference databases is strongly recommended to prevent redundant data entries and inconsistencies within or between models.

3.1.3 Draft reconstruction An imported model can be used as a reference for drafting a new GEM if the reference species is genetically close to the target species. Given a user-provided list of orthologous genes, GEMSiRV enables automatic extraction from the reference model the 'orthologous reactions' (i.e. reactions that are catalyzed by the protein products of orthologous genes) that conform to the gene–protein reaction (GPR) associations defined in the reference model. This function can not only greatly reduce the efforts required for manual model construction but also help evaluate whether the GPR associations are conserved between the reference and the target species (Liao *et al*., 2011a).

3.1.4 Model refinement In GEMSiRV, users can curate and further refine the draft models generated by Model SEED or GEMSiRV. By using the simulation and visualization modules, users can identify dead-end metabolites and blocked reactions in the models, which represent the model gaps to be corrected by the users. In addition, the reference database built in GEMSiRV can provide candidate metabolic reactions for gap filling.

3.2 Simulation

GEMSiRV implements the major FBA functions that are included in the COBRA toolbox (Fig. 2). With a built-in linear programming solver, the intuitive interface of GEMSiRV allows users to perform various *in silico* analyses for their imported/constructed GEMs and generate biological hypotheses according to the simulation results. The GEMSiRV Simulation module provides the following functions.

3.2.1 Dead-end metabolite identification A metabolic network can be converted to a stoichiometric matrix-based mathematical model. The mathematic model describes how the network reactions are connected to each other and defines the boundaries of the biological system (which usually refers to a cell) and cross-boundary material exchanges. GEMSiRV can examine the connections of all metabolites in a network and identify and mark dead-end metabolites with a cross in the map of a metabolic network (e.g. in Fig. 2A, fru is a dead-end metabolite). Identification of dead-end metabolites is critical for debugging the draft model because such metabolites will lead to errors in subsequent simulation analyses.

3.2.2 Objective optimization With a built-in linear programming solver (e.g. GLPK), GEMSiRV can be used to perform simulations for the imported or newly constructed metabolic network models. For example, the user may wish to simulate the growth of a bacterial species under certain nutritional constraints. GEMSiRV can then search in the constraint-based solution space of reaction fluxes and optimize the fluxes for the biomass objective function. The reaction fluxes can be saved and presented in the map of metabolic network with the GEMSiRV Visualization module. And the most effective and efficient pathway through the network will be highlighted by the composed reactions carrying fluxes with different color codes (Fig. 2B).

3.2.3 Flux variability analysis Flux variability analysis can be used to study the redundancy of reactions in a network. GEMSiRV can determine the minimum and maximum flux for each reaction in the model and thus can also identify the blocked reactions (which always carry zero fluxes). These reactions can be marked with a cross in the map of a metabolic network (in Fig. 2C, for example, FRUPts2 is a blocked reaction).

3.2.4 Robustness analysis Robustness analysis can be used to study how changing a reaction flux can affect other reaction fluxes or certain objective functions (e.g. growth rate). In GEMSiRV, at most two reaction fluxes can be altered simultaneously. The results of the robustness analyses can therefore be presented as 2D (one altered flux) or 3D (two altered fluxes) plots (Fig. 2D).

3.2.5 Gene/reaction essentiality analysis GEMSiRV can simulate deletion of a reaction by constraining its flux to zero. Similarly, simulating deletion of a gene can be done by stopping all of the reaction fluxes corresponding to the gene of interest. If the cell (model) fails to grow after a reaction/gene is deleted *in silico*, this reaction/gene is considered as essential for the *in silico* growth (and possibly, the biological growth) of the cell. GEMSiRV can perform essentiality analysis for genes and reactions separately, and determine the ratio of the objective flux (the flux of reaction/gene-deleted model divided by the flux of the wild type) (Fig. 2E).

3.2.6 Gene deletion analysis In GEMSiRV, gene deletion analysis is carried out by *in silico* knocking out one or more than two genes simultaneously, as described in the above section. Such *in silico* knock-outs can simulate biological gene knockouts or transcriptional regulatory constraints (i.e. regulatory suppression of gene expression). GEMSiRV can generate the constraint-based reaction fluxes and an SBML file for the *in silico* knock-out model. Such knock-out models can be again imported into GEMSiRV for other network analyses or editing.

3.3 Visualization

In addition to network reconstruction and simulation, GEMSiRV provides an interface for editing and visualizing the metabolic networks of interest. KEGG pathway maps or the maps in CellDesigner format can be directly loaded into GEMSiRV, and a user-modifiable map can be easily produced based on the loaded map. For example, a genome-scale map for *i*YL1228 (Liao *et al.*, 2011b) generated by GEMSiRV is available on the web. With the unified nomenclature for metabolites and reactions in the

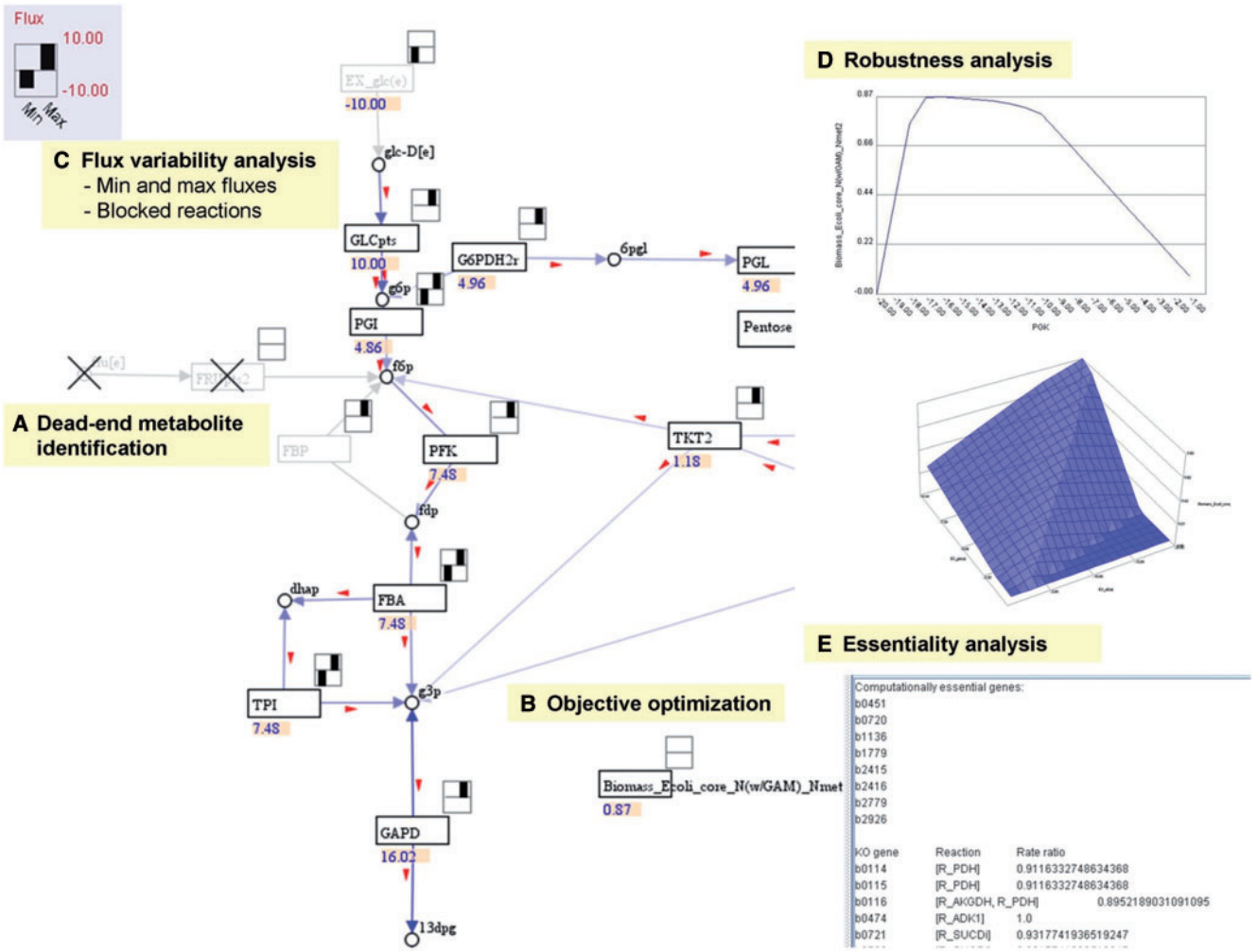

Fig. 2. The simulation functions provided in GEMSiRV: (**A**) dead-end metabolite identification, (**B**) objective optimization, (**C**) flux variability analysis, (**D**) Robustness analysis and (**E**) essentiality analysis. For demonstration, the *E.coli* textbook model has been imported and simulated in the GEMSiRV. More metabolic models can be found in the GEMSiRV website via the 'Metabolic Models' hyperlink (http://sb.nhri.org.tw/GEMSiRV/en/Metabolic_Models)

network model, GEMSiRV allows data exchange between network reconstruction, model simulation and visualization of reaction fluxes and gene expression levels. Fig. 3 demonstrates an example of network map in which simulated reaction fluxes and experimental gene expression levels under aerobic and anaerobic conditions are simultaneously imported into GEMSiRV. Such data integrations enable the user to evaluate whether the FBA solutions are consistent with the gene expression data. To our knowledge, this is the first open-source software to provide such interactive features for genome-scale network reconstructions.

4 DISCUSSION

The recently published systems biology tool Model SEED has greatly reduced the efforts needed to construct GEMs. Applying the models constructed by this type of tool to further studies requires a user-friendly, integrated, interactive interface for model editing, simulation and visualization. GEMSiRV is designed to accelerate network reconstruction and to facilitate generation of new biological hypotheses by supporting simulation analyses of metabolic networks and providing functionalities of data integration and visualization. GEMSiRV also provides ready-to-use downloadable reference databases, metabolic maps and published metabolic models in its website. Therefore, GEMSiRV along with its web resource may greatly expedite GEM reconstructions, and can accelerate the development of biomedical applications of metabolic network reconstructions. One of the advantages of GEMSiRV is that, unlike other metabolic network tools, GEMSiRV allows all of the projects-in-progress to be easily exchanged or shared between researchers. GEMSiRV can therefore facilitate information exchanges and communications (including metabolic network models, simulation results and network maps) in the research community for high-quality metabolic network reconstructions and applications. GEMSiRV is, to the best of our knowledge, the only tool that simultaneously provides functionalities of GEM reconstruction, simulation analyses, data integration, model editing

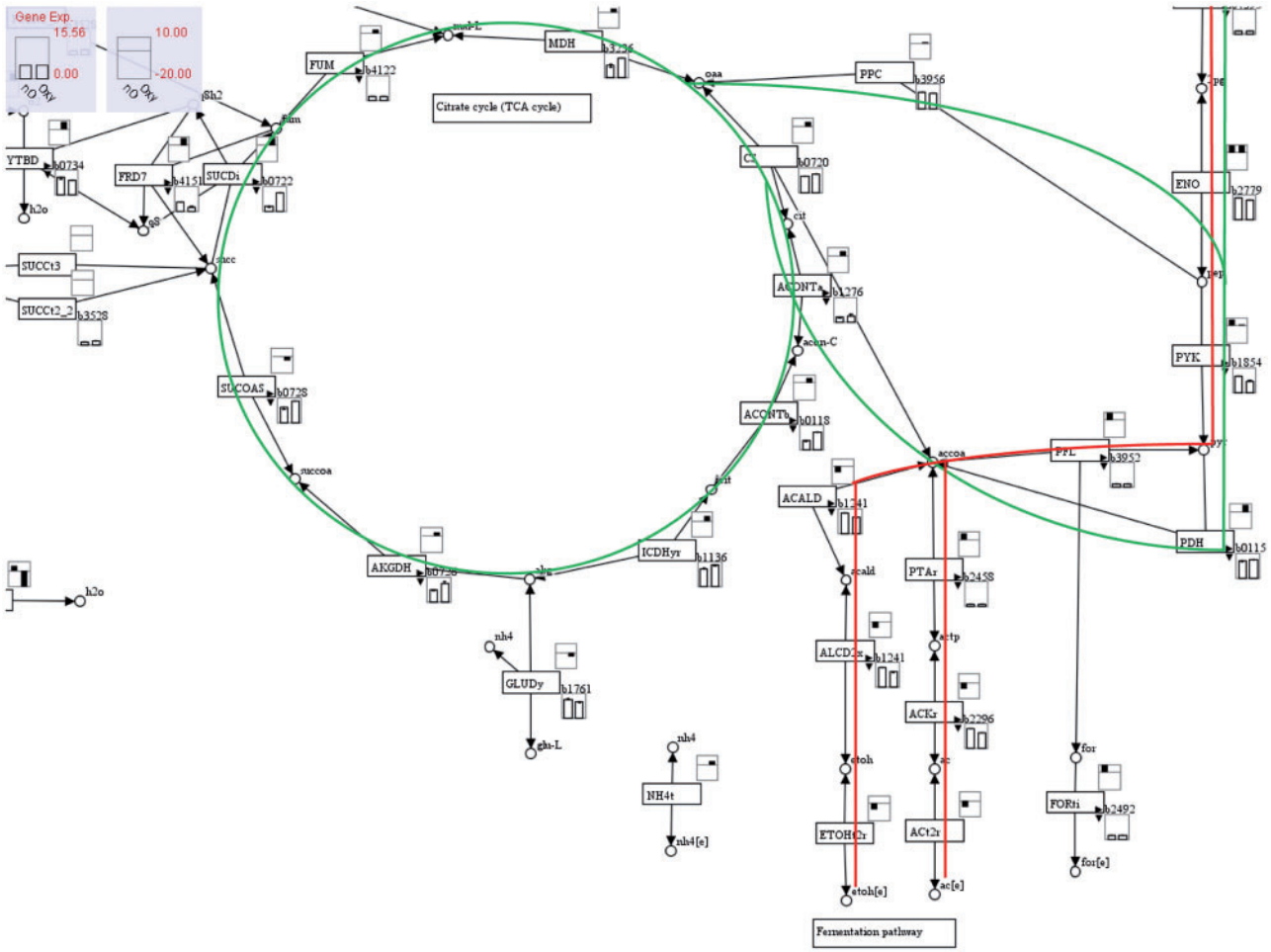

Fig. 3. An example network map generated by GEMSiRV that incorporates multiple layers of large-scale information. In this example, reaction fluxes (the upper right panel of each reaction) and gene expression levels (the lower right panel of each reaction) for aerobic and anaerobic conditions are shown. The aerobic and anaerobic pathways manually highlighted in green and red, respectively, are consistent with the trancriptomic and fluxomic data

and interactive visualization as stand-alone software. Furthermore, GEMSiRV is not only an integrated network reconstruction tool but also a tool for knowledge and learning of systems biology. GEMSiRV is expected to add momentum to metabolic network researches.

ACKNOWLEDGEMENTS

This work was motivated by collaboration with Dr. Bernhard Palsson's laboratory at UCSD on establishing metabolic network reconstructions. We are grateful for many interactions with Dr Palsson's Lab. The authors thank Shin-Hung Lin for constructing the webserver of GEMSiRV and developing toolboxs (GBKParser and TextReplacer), Chieh-Hua Lin and Ying-Hsiang Chen for testing and valuable discussion and Yi-Chiu Lin for editing a genome-scale metabolic map.

Funding: This work was supported by National Health Research Institutes intramural funding (PH-099-SP-10). The computational facilities for this work were partly supported by the National Research Program for Genomic Medicine, National Science Council, Taiwan (NSC99-3112-B-400-012).

Conflict of Interest: none declared.

REFERENCES

Becker,S.A. *et al.* (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protoc.*, **2**, 727–738.

Becker,S.A. and Palsson,B.O. (2005) Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation. *BMC Microbiol.*, **5**, 8.

Caspi,R. *et al.* (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **40**, D742–D753.

Cvijovic,M. *et al.* (2010) BioMet Toolbox: genome-wide analysis of metabolism. *Nucleic Acids Res.*, **38** (Suppl.2), W144–W149.

Duarte,N.C. *et al.* (2004) Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.*, **14**, 1298–1309.

Durot,M. *et al.* (2009) Genome-scale models of bacterial metabolism: reconstruction and applications, *FEMS Microbiol. Rev.*, **33**, 164–190.

Feist,A.M. *et al.* (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.

Feist,A.M. and Palsson,B.O. (2008) The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nat. Biotechnol.*, **26**, 659–667.

Forth,T. *et al.* (2010) MetNetMaker: a free and open-source tool for the creation of novel metabolic networks in SBML format. *Bioinformatics*, **26**, 2352–2353.

Gille,C. *et al.* (2007) METANNOGEN: compiling features of biochemical reactions needed for the reconstruction of metabolic networks. *BMC Syst. Biol.*, **1**, 5.

Grafahrend-Belau,E. *et al.* (2009) FBA-SimVis: interactive visualization of constraint-based metabolic models. *Bioinformatics*, **25**, 2755–2757.

Henry,C.S. *et al.* (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.*, **28**, 977–982.

Hucka,M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Junker,B.H. *et al.* (2006) VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, **7**, 109.

Klamt,S. *et al.* (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol.*, **1**, 2.

Liao,Y.C. *et al.* (2011a) MrBac: a web server for draft metabolic network reconstructions for bacteria. *Bioengineered Bugs*, **2**, 284–287.

Liao,Y.C. *et al.* (2011b) An experimentally validated genome-scale metabolic reconstruction of Klebsiella pneumoniae MGH 78578, iYL1228. *J. Bacteriol.*, **193**, 1710–1717.

Mahadevan,R. *et al.* (2011) In situ to in silico and back: elucidating the physiology and ecology of Geobacter spp. using genome-scale modelling. *Nat. Rev. Microbiol.*, **9**, 39–50.

Navid,A. and Almaas,E. (2009) Genome-scale reconstruction of the metabolic network in Yersinia pestis, strain 91001. *Mol. Biosyst.*, **5**, 368–375.

Nogales,J. *et al.* (2008) A genome-scale metabolic reconstruction of Pseudomonas putida KT2440: iJN746 as a cell factory. *BMC Syst. Biol.*, **2**, 79.

Orth,J.D. *et al.* (2010) What is flux balance analysis?, *Nature Biotechnology*, **28**, 245–248.

Puchalka,J. *et al.* (2008) Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology. *PLoS Comput. Biol.*, **4**, e1000210.

Rocha,I. *et al.* (2010) OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Syst. Biol.*, **4**, 45.

Schellenberger,J. *et al.* (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, **11**, 213.

Schwarz,R. *et al.* (2007) Integrated network reconstruction, visualization and analysis using YANAsquare. *BMC Bioinformatics*, **8**, 313.

Sroka,J. *et al.* (2011) Acorn: a grid computing system for constraint based modeling and visualization of the genome scale metabolic reaction networks via a web interface, *BMC Bioinformatics*, **12**, 196.

Suthers,P.F. *et al.* (2009) A genome-scale metabolic reconstruction of Mycoplasma genitalium, iPS189. *PLoS Comput. Biol.*, **5**, e1000285.

Thiele,I. and Palsson,B.O. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.*, **5**, 93–121.

Wright,J. and Wagner,A. (2008) The Systems Biology Research Tool: evolvable open-source software. *BMC Syst. Biol.*, **2**, 55.