

REVIEW

Metabolic pathway databases and model repositories

Abraham A. Labena^{1,2}, Yi-Zhou Gao¹, Chuan Dong³, Hong-li Hua³ and Feng-Biao Guo^{3,*}

¹ School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China

² College of Computational and Natural Sciences, Dilla University, Dilla P.O.box. 419, Ethiopia

³ School of Life Science and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China

* Correspondence: fbguo@uestc.edu.cn

Received March 15, 2017; Revised April 17, 2017; Accepted April 17, 2017

Background: The number of biological Knowledge bases/databases storing metabolic pathway information and models has been growing rapidly. These resources are diverse in the type of information/data, the analytical tools, and objectives. Here we present a review of the most popular metabolic pathway databases and model repositories, focusing on their scope, content including reactions, enzymes, compounds, and genes, and applicability. The review aims to help researchers choose a suitable database or model repository according to the information and data required, by providing an insight look of each pathway resource.

Results: Four pathways databases and three model repositories were selected on the basis of popularity and diversity. Our review showed that the pathway resources vary in many aspects, such as their scope, content, access to data and the tools. In addition, inconsistencies have been observed in nomenclature and representation of database entities. The three model repositories reviewed do not offer a brief description of the models' characteristics such as simulation conditions.

Conclusions: The inconsistencies among the databases in representing their contents may hamper the maximal use of the knowledge accumulated in these databases in particular and the area of systems biology at large. Therefore, it is strongly recommended that the database creators and the metabolic network models developers should follow international standards for the nomenclature of reactions and metabolites. Besides, computationally generated models that could be obtained from model repositories should be utilized with manual curations as they lack some important components that are necessary for full functionality of the models.

Keywords: metabolic pathway; database; model repository

Author summary: Four metabolic pathway databases and three model repositories were reviewed with regard to their scope, content, and applicability. Despite their innumerable use in the fields of systems biology and metabolic engineering, these pathway databases and model repositories are not in harmony with each other due to the inconsistencies in the way they represent their contents. Besides, the automatically generated metabolic models that can be obtained from the model repositories are not accurate enough for further scientific usage without additional manual curation. Therefore, international standards such as IUBMB principles should be strictly obeyed in creating such metabolic pathway resources.

INTRODUCTION

Metabolic pathways are among the most important components of the cellular systems. A metabolic pathway by definition is a linked series of enzyme-catalyzed chemical reactions that occur in a cell [1]. In these series

of reactions, metabolites (reactants and intermediates) are modified into products that can be used in various cellular activities including cellular growth and maintenance. Recently, the number of internet-accessible biological knowledge bases/databases storing pathway information has been growing rapidly [2]. These pathway resources

aim to systematically bring together metabolic pathway and network information and expedite research and developments in systems biology, as well as aid accessibility of this information to the scientific community [3]. The databases can be categorized based on the type of information/data they store, the analytical tools

they contain and their objectives [4]. Table 1 presents a list of available metabolic pathway resources with their descriptions and website links. A comprehensive list of the pathway resources from different categories of life sciences is available at <http://www.expasy.org/links.html> and <http://www.pathguide.org>.

Table 1 List of available metabolic pathway resources

Name	Description	Website
<i>Arabidopsis</i> Reactome	Curated knowledge base of plant biological pathways	http://www.arabidopsisreactome.org
AtIPD	<i>Arabidopsis thaliana</i> isoprenoid pathway database	http://www.atipd.ethz.ch
BiGG	Knowledge base of genome-scale metabolic network models	http://bigg.ucsd.edu
BioCyc	A collection of pathway/genome databases (PGDBs) and software tools for understanding their data	http://biocyc.org
BioModels	Online reference repository for quantitative, dynamic models of biological network models	https://www.ebi.ac.uk/biomodels-main
BioPath	Database on biochemical pathways	https://www.mn-am.com/databases/biopath
BioSilico	A web-based database system that facilitates the search and analysis of metabolic pathways	http://biosilico.kaist.ac.kr
BRENDA	Comprehensive enzyme information system	http://www.brenda-enzymes.info
BsubCyc	Database of the bacterium <i>Bacillus subtilis</i> and is based on the updated <i>B. subtilis</i> 168 genome sequence and annotation	https://bsubcyc.org
CATHACyc	Metabolic pathway database of <i>Catharanthus roseus</i>	http://www.cathacyc.org
ECMDB	<i>Escherichia coli</i> metabolome database	http://ecmdb.ca
EcoCyc	Encyclopedia of <i>E. coli</i> genes and metabolism	https://ecocyc.org
EcoCyc	Scientific database for the bacterium <i>E. coli</i> K-12 MG1655	https://ecocyc.org
ENZYME	A repository of information relative to the nomenclature of enzymes	http://enzyme.expasy.org
ExpASY	Biochemical pathway maps	http://web.expasy.org/pathways
FlyReactome	A curated knowledgebase of <i>Drosophila melanogaster</i> pathways	http://fly.reactome.org
HMDB	The human metabolome database	http://www.hmdb.ca
HPD	An integrated human pathway database	http://discern.uits.iu.edu:8340/HPD
HUMANCyc	An encyclopedic reference on human metabolic pathways	https://humancyc.org
KaPPA-View4	Kazusa Plant Pathway Viewer	http://kpv2.kazusa.or.jp/kpv4
KEGG	Kyoto Encyclopedia of Genes and Genomes	http://www.kegg.jp
LAMP	Library of Apicomplexan metabolic pathways	http://www.llamp.net
LIPID MAPS	LIPID metabolites and pathways strategy	http://www.lipidmaps.org/resources/resources.html
MaizeGDB	Metabolic pathways in maize	http://maizecyc.maizegdb.org
Malaria	Malaria parasite metabolic pathways	http://mpmp.huji.ac.il
MedicCyc	A biochemical pathway database for <i>Medicago truncatula</i>	http://mediccyc.noble.org
MetaCyc	Knowledge of experimentally validated metabolic pathways	https://metacyc.org
MetaNetX/MetanetX.org	Repository and webserver for genome-scale metabolic network Models	http://www.metanetx.org/
MetNetDB	Contains information on networks of metabolic and regulatory and interactions in <i>Arabidopsis</i>	http://metnetweb.gdcb.iastate.edu/MetNet_db.htm
MMMDB	Mouse multiple tissue metabolome dataBase	http://mmdb.iab.keio.ac.jp
Model SEED	Web-based resource for high-throughput generation, optimization and analysis of genome-scale metabolic pathway models	http://modelseed.org
MouseCyc	Manually curated database of both known and predicted metabolic pathways for the laboratory mouse	http://mousecyc.jax.org
PathCase	Pathways database system	http://nashua.cwru.edu/PathwaysWeb

(Continued)

Name	Description	Website
PC2	Pathway Commons 2 (integrates a number of pathway and molecular interaction databases supporting BioPAX and PSI-MI formats into one large BioPAX model)	http://www.pathwaycommons.org/pc2
PMN	Plant metabolic network	http://www.plantcyc.org
Reactome	A free, open-source, curated and peer-reviewed pathway database	http://www.reactome.org
RGB	Rat resource center	http://rgd.mcw.edu/wg/home/pathway2/molecular-pathways2
SABIO-RK	Biochemical reaction kinetics database	http://sabio.villa-bosch.de
SSER	Species specific essential reactions database	http://cefg.cn/sser/index.html
UniPathWay	Metabolic pathways database	http://www.grenoble.prabi.fr/obiwarehouse/unipathway
YEASTNET	A consensus reconstruction of yeast metabolism	http://www.comp-sys-bio.org/yeastnet
YMDB	Yeast metabolome database	http://www.ymdb.ca

Information from pathway resources has been extensively applied in *in-silico* reconstructions and analysis of biological models representing vital cellular systems such as signaling, regulatory and metabolic networks [5]. Accordingly, hundreds of metabolic pathway models have been reconstructed and represented in the form of predictive mathematical models for those organisms with their genome sequenced [6–8]. These reconstructions aim to integrate metabolic entities (genes, enzymes, reactions, and compounds) for a better understanding of the cellular structure and functions [9]. Well curated and experimentally tested reconstructions have been applied in wide variety of areas such as identifying essential gene/reactions and drug targets [5], studying protein-protein interactions, producing and increasing the yields of important industrial chemicals [5,10,11]. Therefore, it is of great value to assess and provide an insight view of the existing databases and pathway resources so that to help the researchers in the field choose the most suitable one according to the information and data pertinent to a particular research.

Unlike the previous reviews such as [12,13] presenting an in-depth review of few pathway databases, here we will discuss relatively larger number and different types of pathway resources and model repositories that have been commonly used in metabolic network reconstructions and analysis in two perspectives. Firstly, our review compares the pathway resources with respect to their content, scope and data representation, and presents the most up-to-date information and statistics on pathways, reactions, compounds and enzymes for each database and model repository. Secondly, the review presents a brief comparison of four of the databases based on additional functions. For that, we summarized the available tools and their functions. The databases reviewed in this article include Reactome [14], MetaCyc [15], Kyoto Encyclopedia of Genes and Genomes (KEGG) [16], and plant

metabolic network (PMN) [17], and the web servers and model repositories and web servers include Biochemical Genomic and Genetic knowledgebase (BiGG) [18], BioModels [19] and MetaNetX [20]. (Table 1).

METABOLIC PATHWAY DATABASES

Although they are diverse in their scope, metabolic databases should be able to provide reliable information on four of the basic constituents of metabolic networks. These include biochemical reactions, enzymes catalyzing the reactions, pathways, and metabolites. Metabolites should be represented in their appropriate charged or neutral states in the databases. Similarly, the databases should incorporate reaction directionality information, elemental and charge balance and compartment to which the reaction and compounds belong to. Some comprehensive databases include genomic information, software/tools for analysing and visualizing the pathways and reactions. Availability and accuracy of the tools can also determine the extent to which the databases would be applied. The databases also vary in data source and export formats. Some databases are organism or species specific such as PMN and Reactome whereas the others describe multiple organisms or reference databases such as MetaCyc and KEGG. The following section describes some database-specific features for each database. Table 2 represents a summary of contents of the databases assessed accordingly in this review.

Reactome

Reactome is publically accessible, open-source, manually curated and peer-reviewed database of human pathways [21]. The primary goal of Reactome is to provide molecular details of signal transduction, transport, metabolism, DNA replication and other cellular processes

Table 2 Contents of the databases

Database/content	Reactions	Compounds	Pathways	Proteins	Version
KEGG	10307	17787	474838	6836	80.2
MetaCyc	15162	13585	2884	17090	20.5
Reactome	78927	88597	22072	120935	59
PMN	74013	2509	13117	279301	11.0

as an ordered network of molecular transformations of *Homo sapiens* [14,22]. The latest version includes organisms from other domains of life. Users can exploit three important features of Reactome. They can browse, visualize and analyze reactions, metabolites, and enzymes in each pathway and download the information in various formats.

However, there are some drawbacks that need attention. For instance, comparison function for the pathways in various organisms returns the number of pathways, reactions or compounds with no description about the type of the database entities shared across the organisms compared. Furthermore, there is no way to download part of reactions or compounds at the pathway browsing interface or further description page other than programmatically accessing the whole data in the database. Most database visitors seek for information related to a particular project such as searching for gap filling reaction in metabolic network reconstruction. In such cases users pay more attention to fragments of information than the whole, thus offering such option is invaluable. Besides, Gene-Protein-Reaction (GPR) association information is one important quality that Reactome lacks, hence one has to refer other linked databases for such information, and this may cost additional time of the user.

In general, the Reactome knowledgebase has been applied in developing other species specific and general reactomes. Fly Reactome, Gallus Reactome, microme and plant Reactome [23] can be mentioned in this regard.

MetaCyc

MetaCyc is a highly curated database of experimentally validated metabolic pathways from all domains of life [15,24–26]. It contains pathways derived from extensively large number of primary literature [15]. MetaCyc aims to serve as a general reference database on metabolism. An important feature of MetaCyc is its pathway tool that can be used to computationally predict metabolic network models of any organism from a sequenced genome.

However, the data retrieval and download system require expertise in programming skills that most users may not have. This may hinder accessibility of useful information in the database. Most reactions in MetaCyc are elementally balanced and accompanied by thermodynamic information (Standard Gibbs free energy values,

$\Delta_r G^\circ$) although the substrates and products are represented in full names rather than in simplified molecular formulae or abbreviations. This makes the browsing process cumbersome to some extent.

Regardless of these minor set-backs, MetaCyc has been applied to create more than 5,700 PGDBs for large number of organisms, including *Saccharomyces cerevisiae* [27], *Arabidopsis thaliana* [28], *Oryza sativa* [29], *Mus musculus* [30], *Bos taurus* [31], *Medicago truncatula* [32], *Populus trichocarpa* [33], *Dictyostelium discoideum* [34], *Leishmania major* [35], *Chlamydomonas reinhardtii* [36], several *Solanaceae* species [37], bioenergy-related organisms (BeoCyc) and many pathogenic organisms [38].

PMN

Plant metabolic network is a metabolic pathways database that hosts one reference database (PlantCyc) and 22 species-specific databases. All of them especially focus on plant metabolic pathways. At the center of PMN is PlantCyc, which is a metabolic pathway reference database containing more than 900 pathways, their catalytic enzymes and genes. Furthermore, PlantCyc contains compounds from over 350 plant species. The data source in PlantCyc covers pathways from experimentally validated literature and curated by PMN and its collaborators [17].

PlantCyc can be considered as a derivative of MetaCyc. Accordingly, many features of PlantCyc resemble MetaCyc. The databases schema and classification of the pathways, reactions, and compounds are all in a similar fashion as in MetaCyc. Hence, the drawbacks in MetaCyc are also reflected in PlantCyc. In addition, PlantCyc contains some computationally predicted and hypothetical (which are not part of plant metabolism but utilized in the metabolic pathways of other organisms) pathways but there are no special markers or other forms of identifiers used for these classes of database entities on the pathway diagram.

The datasets and the pathway from PMN have been applied for in-depth investigation of plant metabolism including pathway predictions, metabolic network reconstructions and developing bioinformatics tools [33,39–44], etc. For instance, in [43] PMN was used to investigate genomic signatures of specialized metabolism in plants. The MORPH algorithm [44], ranking candidate

genes for membership in *Arabidopsis* and tomato pathways, was also developed based on the pathway information from PMN.

KEGG

KEGG is a database resource developed to understand high-level functions and utilities of the biological system, from molecular-level information. The objective of KEGG was to create a reference knowledge base of metabolism and other cellular processes [45]. The most recent version of the KEGG contains 16 databases which are grouped into four categories, namely systems, genomic, chemical and health information databases [16].

KEGG is well known for a detailed description, representation, and visualization of its contents. However, there are some discrepancies arise in reaction representations. For instance, R00472 is represented as a sum of two reactions, R00473 + R10612, but there is no description how these reactions combined and resulted in R00472. KEGG can be accessed by users of varying levels of programming skill, but the search function seems complicated for those with a little or no programming skill at both KEGG and DBGET searching interfaces. In fact, DBGET offers keyword searching option, still, it requires KEGG prefixes as search criteria. Providing examples, one for each method in the briefings at the help page may maximize the usage in this regard. In addition, browsing some pathways such as Glycolysis (Embden-Meyerhof pathway) D-glucose to Pyruvate ends up with long non-zooming maps. This forces the user to scroll up and down while browsing parts of that particular pathway.

The datasets and the tools of KEGG have been largely utilized in a remarkable number of scientific studies, including genome annotations, pathway predictions, metabolic network reconstructions, and developing bioinformatics tools [46–52]. For instance, the KEGG omics data visualization tool was integrated with a desktop application, KegArray and many more web-based tools to visualize and interpret large amount of data obtained from high-throughput measurement techniques such as microarray, metagenome, and metabolome analyses [49]. Pathway Inspector [50], a pathway-based web application for RNAseq analysis of model and non-model organisms is another tool created based on KEGG database. The KEGG GO terms have been used in developing CGDB [51] a database of circadian genes in eukaryotes and many more.

Comparison of the pathway databases

The pathways databases can also be compared in terms of the software or tools that can facilitate analysis and

interpretation of their contents. Table 3 summarizes unique tools specific to each database.

MODEL REPOSITORIES

Since the models come from diverse research groups or individuals, one of the current challenges related to metabolic network models is the inconsistencies in representing their components such as reaction, or compound names, and formulae, and symbols used to designate cellular compartments. Therefore, the online model repositories should go beyond simply serving as a repository to platforms for systematic analysis and standardization of these models, the efforts of BiGG and MetaNetX.org can be mentioned in this regard. Furthermore, the model repositories are expected to provide concise descriptions of the models such as simulation conditions, this may include *in-silico* growth media condition, energy requirements, objective function, and file compatibility information. Bearing this in mind, the following section will assess some of the most common metabolic network model databases.

BiGG

Biochemical, Genetic and Genomic (BiGG) is a knowledge base of genome-scale metabolic network (GEM) models. The latest version of BiGG comprises 80 high-quality, manually curated genome-scale metabolic models (including *E. coli* core model), 15,288 reactions and 5,175 metabolites (as of Nov. 28th 2016). Additional features include pathway visualization with Escher map (a web-based tool for building, viewing, and sharing visualizations of biological pathways), various model export formats (SBML Level 3, MAT and JSON) and model validation function.

Despite its numerous good qualities, BiGG has some minor drawbacks that should be addressed. One such point to be considered is that there should be information about each model describing whether a given model is experimentally validated or not. In addition, the model validation tool is limited to SBML level 3 v1 and flux balance constraints (FBC) v2 files, since significant numbers of metabolic models still exist in SBML level 2 and incorporating an option for this category of file format is invaluable.

The GEMs from BiGG have been applied to develop bioinformatics tools such as Fast-SNP (Fast matrix pre-processing algorithm for efficient loopless flux optimization of metabolic models) [53], the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) [54] including integrative view of protein, gene and 3D structural information, etc.

Table 3 Comparison of the databases based on additional functions

Name	Tools	Function
Reactome	Pathway browser	<ul style="list-style-type: none"> • A tool for visualizing and interacting with Reactome biological pathways
	Analyze data	<ul style="list-style-type: none"> • Merges pathway identifier mapping, over-representation, and expression analysis tools into a single tabbed data analysis portal with integrated visualization and summary features, which can accept a gene, protein or small molecule list, or an expression dataset
	Species comparison	<ul style="list-style-type: none"> • Allows users to compare pathways between human and any of the other species inferred from Reactome by orthology
	Reactome FI network	<ul style="list-style-type: none"> • Cytoscape plugin designed to find network patterns related to cancer and other types of diseases
MetaCyc	Pathway tools	<ul style="list-style-type: none"> • Development of organism-specific databases
		<ul style="list-style-type: none"> • Metabolic reconstruction and modeling
		<ul style="list-style-type: none"> • Scientific visualization, web publishing
		<ul style="list-style-type: none"> • Visual analysis of gene expression and metabolomics datasets
		<ul style="list-style-type: none"> • Computational inferences
		<ul style="list-style-type: none"> • Comparative genome and pathway analyses • Analysis of biological networks
PMN	E2P2	<ul style="list-style-type: none"> • Enzyme function annotation software. Predicts metabolic enzymes in a sequenced genome
	SAVI	<ul style="list-style-type: none"> • Pathway validation software. Processes predicted metabolic pathways using pathway metadata such as taxonomic distribution and key reactions and makes decisions about which pathways to keep, remove, or subject to manual validation
KEGG	PlantClusterFinder	<ul style="list-style-type: none"> • A pipeline to predict metabolic gene clusters from plant genomes
	KegHier	<ul style="list-style-type: none"> • Java application for browsing BRITE hierarchy files
	KegArray	<ul style="list-style-type: none"> • Java application for microarray data analysis
	KegDraw	<ul style="list-style-type: none"> • Java application for drawing compound and glycan structures

BioModels

BioModels is an online reference repository that hosts peer-reviewed quantitative, dynamic models of biological networks of numerous organisms [19]. Currently, BioModels contains over 1,200 models reported in the literature and 140,000 models automatically generated from pathway resources using a Path2Models project, a tool developed by Büchel *et al.* [55].

However, the automatically generated models still need some degree of manual interventions as the models show some incapability in accurately predicting the defined biomass precursors. For example, the automatically generated *Homo sapiens* metabolic network model was incapable of synthesizing nine of the amino acids namely cysteine, histidine, isoleucine, leucine, lysine, methionine, threonine, tryptophan, and valine in a defined minimal medium [55]. In another example, FBA of BMID000000142305 model resulted in the growth rate of zero for wild-type (WT) strain on a simulation carried out at MetaNetX.org and local machine using COBRA toolbox version 2.0 when the model is used as obtained from BioModels database. Therefore, it is critical to carry out initial curation of the models prior to any research or investigation that will depend on these models.

Models from BioModels have been utilized in large

numbers of scientific studies involving metabolic network analysis. The manually curated models have been widely used in gene and reaction essentiality analysis, protein-protein interaction, identifying drug targets; developing bioinformatics tools [56–58] etc.

MetaNetX/MetaNetX.org

MetaNetX serves as a repository for genome-scale metabolic network models and MetaNetX.org [20] is an online platform for accessing, analyzing and manipulating genome-scale metabolic networks (GSM). The website and the tools in MetaNetX.org provide multiple functions: i) free access to MNXref reconciliation data and collection of published GSMNs, ii) it also allows users to upload, manipulate, analyze or modify their own genome-scale metabolic network models and export them in SBML or MXN tab-limited format, iii) tools for metabolic network analyses based on network structure, FBA or nested pattern methods [59].

MetaNetX.org has tremendous advantages users can exploit. However, there are issues still that need to be considered. For instance, there is no option for programmatic access to the wealth of metabolic knowledge in this repository. In addition, some models that can be picked up from the repository for analysis are automatically

Table 4 Comparison of tools in model repositories

Name	Tools	Function
BiGG	Escher map and model validation tool	It is a web-based tool for building, viewing and sharing visualizations of biological pathways
BioModels	Path2Models (automatic generation of GSM)	It automatically generates metabolic models from biochemical pathway maps
MetaNetX	Flux balance analysis (FBA), flux variability analysis(FVA), group of coupled reactions (GCR), reaction knock-out (RKO), peptide/gene knock-out (PKO), gap-filling (GAP), predict direction (DIR)	A web-based resource for accessing, analysing and manipulating genome-scale metabolic networks. It also provides interactive comparison of two or more models and integrates data from various public resources

generated by computational methods. Therefore, providing evidence or references for experimental validation of this group of models is critical.

The resources and tools in MetaNetX.org have been generally applied to the metabolic network reconstruction and analysis. FBA and other interactive tools were utilized to develop new reconstructions, compare existing reconstructions, and extract components of reconstructions including compounds, reactions, gene-protein associations, media components and the biomass equation components with its coefficients [60–62].

Comparison of the model repositories

The model repositories are also compared based on the additional function they provide to the users. Table 4 represents some database specific tools available.

CONCLUSION AND PERSPECTIVE

The present review discusses pathway knowledge bases/databases and model repositories that have been commonly used in metabolic network reconstructions and analysis in two perspectives. Firstly, although each database and knowledge base has its own aim and scope, the review presents a brief comparison of pathway resources based on the scope, contents, and applicability. Accordingly, some inconsistencies have been observed in nomenclature and representation of database entities. For example, KEGG uses an R number as a reaction identifier such as R00259 for acetylation of L-glutamate, while BiGG and MetaCyc use quite different reaction identifiers such as ACGS and EC number 2.3.1.1 or enzyme name respectively for the same reaction. On the other hand, MetaNetX uses MNX prefix followed by a single letter indicating the type of object: R for reaction; M for metabolite; C for cellular compartment and followed by an integer (e.g., MNXR1234). These inconsistencies hamper maximal use of the knowledge accumulated in these databases and in the area of systems biology at large. Hence, it is strongly recommended that the database creators and the metabolic network models developers should follow international standards for nomenclature of

reactions and metabolites such as IUBMB (International Union for Biochemistry and Molecular Biology) rule thereby to facilitate the integration of the databases and related pathway resources.

Secondly, we observed that three of the databases reviewed, PMN, KEGG, and Reactome, have well-organized summary and statistics about their content on the website. In fact, other pathway resources also contain such information but not in a well-structured fashion. Hence, it would be of great value if the pathway resources offer such information in clear and precise way so that users can easily catch up a clear image about the resources in database or repository, and this enables them to choose an appropriate database or model repository according to a particular research interest.

Finally, all of the reviewed model repositories do not provide a brief description of the models' characteristics such as simulation conditions. Providing such information including *in-silico* growth media information, objective function, energy requirements, evidence for experimental validation of each model, and so on, is a critical point to be considered by existing model repositories and those that will be built in future.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (No. 31470068), Sichuan Youth Science and Technology Foundation of China (No. 2014JQ0051) and the Fundamental Research Funds for the Central Universities of China (Nos. ZYGX2015Z006 and ZYGX2015J144). The funders had no role in study design, data collection, analysis, decision to publish, and preparation of the manuscript.

COMPLIANCE WITH ETHICAL GUIDELINES

The authors Abraham A. Labena, Yi-Zhou Gao, Chuan Dong, Hong-li Hua and Feng-Biao Guo declare that they have no conflict of interest.

This article is a review article and does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Berg, J. M., Tymoczko, J. L. and Stryer, L. (2002) *Metabolism* Is

- Composed of Many Coupled, Interconnecting Reactions. In *Biochemistry*. 5th ed. New York: Lippincott Williams & Wilkins
- Wren, J. D. and Bateman, A. (2008) Databases, data tombs and dust in the wind. *Bioinformatics*, 24, 2127–2128
 - Likić, V. A. (2006) Databases of metabolic pathways. *Biochem. Mol. Biol. Educ.*, 34, 408–412
 - Stobbe, M. D., Houten, S. M., Jansen, G. A., van Kampen, A. H. and Moerland, P. D. (2011) Critical assessment of human metabolic pathway databases: a stepping stone for future integration. *BMC Syst. Biol.*, 5, 165
 - Stobbe, M. D. (2015) Metabolic Pathway Databases: A Word of Caution. In *Computational Systems Toxicology*. Hoeng, J., Peitsch, M.C., eds. pp. 27–63. New York: Springer
 - Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M. and Palsson, B. Ø. (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.*, 7, 535
 - Thiele, I., Vo, T. D., Price, N. D. and Palsson, B. O. (2005) Expanded metabolic reconstruction of *Helicobacter pylori* (iT341 GSM/GPR): an *in silico* genome-scale characterization of single- and double-deletion mutants. *J. Bacteriol.*, 187, 5818–5830
 - Henry, C. S., Zinner, J. F., Cohoon, M. P. and Stevens, R. L. (2009) iBsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome Biol.*, 10, R69
 - Radrach, K., Tsuruoka, Y., Dobson, P., Gevorgyan, A., Swainston, N., Baart, G. and Schwartz, J.-M. (2010) Integration of metabolic databases for the reconstruction of genome-scale metabolic networks. *BMC Syst. Biol.*, 4, 114
 - Zhang, C. and Hua, Q. (2015) Applications of genome—scale metabolic models in biotechnology and systems medicine. *Front. Physiol.*, 6, 413
 - Liu, L., Agren, R., Bordel, S. and Nielsen, J. (2010) Use of genome-scale metabolic models for understanding microbial physiology. *FEBS Lett.*, 584, 2556–2564
 - Ooi, H. S., Schneider, G., Lim, T. T., Chan, Y. L., Eisenhaber, B. and Eisenhaber, F. (2010) Biomolecular pathway databases. *Methods Mol. Biol.*, 609, 129–144
 - Wittig, U. and De Beuckelaer, A. (2001) Analysis and comparison of metabolic pathway databases. *Brief. Bioinform.*, 2, 126–142
 - Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M. R., *et al.* (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res.*, 42, D472–D477
 - Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., *et al.* (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, 44, D471–D480
 - Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, 44, D457–D462
 - Dreher, K. (2014) Putting the Plant Metabolic Network Pathway Databases to Work: Going Offline to Gain New Capabilities. In *Plant Metabolism: Methods and Protocols*. Sriram G. Totowa, ed., pp. 151–171. NJ: Humana Press
 - King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O. and Lewis, N. E. (2016) BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.*, 44, D515–D522
 - Chelliah, V., Laibe, C. and Le Novère, N. (2013) BioModels Database: a repository of mathematical models of biological processes. *Methods Mol. Biol.*, 1021, 189–199
 - Ganter, M., Bernard, T., Moretti, S., Stelling, J. and Pagni, M. (2013) MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics*, 29, 815–816
 - Croft, D., O’Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, 39, D691–D697
 - Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., *et al.* (2016) The Reactome pathway Knowledgebase. *Nucleic Acids Res.*, 44, D481–D487
 - Naithani, S., Preece, J., D’Eustachio, P., Gupta, P., Amarasinghe, V., Dharmawardhana, P. D., Wu, G., Fabregat, A., Elser, J. L., Weiser, J., *et al.* (2017) Plant Reactome: a resource for plant pathways and comparative analysis. *Nucleic Acids Res.*, 45, D1029–D1039
 - Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., *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
 - Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., Holland, T. A., Keseler, I. M., Kothari, A., Kubo, A., *et al.* (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.*, 42, D459–D471
 - Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. A., Gilham, F., Kaipa, P., Karthikeyan, A. S., Kothari, A., Krummenacker, M., *et al.* (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, 38, D473–D479
 - Christie, K. R., Weng, S., Balakrishnan, R., Costanzo, M. C., Dolinski, K., Dwight, S. S., Engel, S. R., Feierbach, B., Fisk, D. G., Hirschman, J. E., *et al.* (2004) *Saccharomyces* Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res.*, 32, D311–D314
 - Mueller, L. A., Zhang, P. and Rhee, S. Y. (2003) AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.*, 132, 453–460
 - Liang, C., Jaiswal, P., Hebbard, C., Avraham, S., Buckler, E. S., Casstevens, T., Hurwitz, B., McCouch, S., Ni, J., Pujar, A., *et al.* (2008) Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.*, 36, D947–D953
 - Evsikov, A. V., Dolan, M. E., Genrich, M. P., Patek, E. and Bult, C.

- J. (2009) MouseCyc: a curated biochemical pathways database for the laboratory mouse. *Genome Biol.*, 10, R84
31. Seo, S. and Lewin, H. A. (2009) Reconstruction of metabolic pathways for the cattle genome. *BMC Syst. Biol.*, 3, 33
 32. Urbanczyk-Wochniak, E. and Sumner, L. W. (2007) MedicCyc: a biochemical pathway database for *Medicago truncatula*. *Bioinformatics*, 23, 1418–1423
 33. Zhang, P., Dreher, K., Karthikeyan, A., Chi, A., Pujar, A., Caspi, R., Karp, P., Kirkup, V., Latendresse, M., Lee, C., *et al.* (2010) Creation of a genome-wide metabolic pathway database for *Populus trichocarpa* using a new approach for reconstruction and curation of metabolic pathways for plants. *Plant Physiol.*, 153, 1479–1491
 34. Fey, P., Gaudet, P., Curk, T., Zupan, B., Just, E. M., Basu, S., Merchant, S. N., Bushmanova, Y. A., Shaulsky, G., Kibbe, W. A., *et al.* (2009) dictyBase—a *Dictyostelium* bioinformatics resource update. *Nucleic Acids Res.*, 37, D515–D519
 35. Doyle, M. A., MacRae, J. I., De Souza, D. P., Saunders, E. C., McConville, M. J. and Likić, V. A. (2009) LeishCyc: a biochemical pathways database for *Leishmania major*. *BMC Syst. Biol.*, 3, 57
 36. May, P., Christian, J. O., Kempa, S. and Walther, D. (2009) ChlamyCyc: an integrative systems biology database and web-portal for *Chlamydomonas reinhardtii*. *BMC Genomics*, 10, 209
 37. Bombarely, A., Menda, N., Teclé, I. Y., Buels, R. M., Strickler, S., Fischer-York, T., Pujar, A., Leto, J., Gosselin, J. and Mueller, L. A. (2011) The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. *Nucleic Acids Res.*, 39, D1149–D1155
 38. Snyder, E. E., Kampanya, N., Lu, J., Nordberg, E. K., Karur, H. R., Shukla, M., Soneja, J., Tian, Y., Xue, T., Yoo, H., *et al.* (2007) PATRIC: the VBI PathoSystems Resource Integration Center. *Nucleic Acids Res.*, 35, D401–D406
 39. Vincent, J., Dai, Z., Ravel, C., Choulet, F., Mouzeyar, S., Bouzidi, M F., Agier, M., Martre, P. (2013) dbWFA: a web-based database for functional annotation of *Triticum aestivum* transcripts. Database (Oxford). 2013, bat014.
 40. Obertello, M., Shrivastava, S., Katari, M. S. and Coruzzi, G. M. (2015) Cross-species network analysis uncovers conserved nitrogen-regulated network modules in rice. *Plant Physiol.*, 168, 1830–1843
 41. Shiratake, K. and Suzuki, M. (2016) Omics studies of citrus, grape and rosaceae fruit trees. *Breed. Sci.*, 66, 122–138
 42. Cho, K., Cho, K. S., Sohn, H. B., Ha, I. J., Hong, S. Y., Lee, H., Kim, Y. M. and Nam, M. H. (2016) Network analysis of the metabolome and transcriptome reveals novel regulation of potato pigmentation. *J. Exp. Bot.*, 67, 1519–1533
 43. Chae, L., Kim, T., Nilo-Poyanco, R. and Rhee, S. Y. (2014) Genomic signatures of specialized metabolism in plants. *Science*, 344, 510–513
 44. Tzfadia, O., Amar, D., Bradbury, L. M. T., Wurtzel, E. T. and Shamir, R. (2012) The MORPH algorithm: ranking candidate genes for membership in *Arabidopsis* and tomato pathways. *Plant Cell*, 24, 4389–4406
 45. Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, 42, D199–D205
 46. Schmeier, S., Alam, T., Essack, M. and Bajic, V. B. (2017) TcoF-DB v2: update of the database of human and mouse transcription co-factors and transcription factor interactions. *Nucleic Acids Res.*, 45, D145–D150
 47. Li, P., Tompkins, R. G. and Xiao, W. (2017) KERIS: kaleidoscope of gene responses to inflammation between species. *Nucleic Acids Res.*, 45, D908–D914.
 48. Wang, Y., Xu, L., Thilmony, R., You, F. M., Gu, Y. Q. and Coleman-Derr, D. (2016) PIECE 2.0: an update for the plant gene structure comparison and evolution database. *Nucleic Acids Res.*, 45, 1015–1020
 49. Kotera, M., Hirakawa, M., Tokimatsu, T., Goto, S. and Kanehisa, M. (2012) The KEGG databases and tools facilitating omics analysis: latest developments involving human diseases and pharmaceuticals. *Methods Mol. Biol.*, 802, 19–39
 50. Bianco, L., Riccadonna, S., Lavezzo, E., Falda, M., Formentin, E., Cavalieri, D., Toppo, S. and Fontana, P. (2016) Pathway Inspector: a pathway based web application for RNAseq analysis of model and non-model organisms. *Bioinformatics*, btw636
 51. Li, S., Shui, K., Zhang, Y., Lv, Y., Deng, W., Ullah, S., Zhang, L. and Xue, Y. (2016) CGDB: a database of circadian genes in eukaryotes. *Nucleic Acids Res.*, 45, D397–D403
 52. Chen, I. A., Markowitz, V. M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Andersen, E., Huntemann, M., *et al.* (2017) IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.*, 45, D507–D516
 53. Saa, P. A. and Nielsen, L. K. (2016) Fast-SNP: a fast matrix pre-processing algorithm for efficient loopless flux optimization of metabolic models. *Bioinformatics*, 32, 3807–3814
 54. Rose, P. W., Prlic, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., Costanzo, L. D., Duarte, J. M., Dutta, S., Feng, Z., *et al.* (2017) The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.*, 45, D271–D281.
 55. Büchel, F., Rodriguez, N., Swainston, N., Wrzodek, C., Czauderna, T., Keller, R., Mittag, F., Schubert, M., Glont, M., Golebiewski, M., *et al.* (2013) Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst. Biol.*, 7, 116
 56. Naldi, A., Monteiro, P. T., Müssel, C., the Consortium for Logical Models and Tools, Kestler, H. A., Thieffry, D., Xenarios, I., Saez-Rodriguez, J., Helikar, T. and Chaouiya, C. (2015) Cooperative development of logical modelling standards and tools with CoLoMoTo. *Bioinformatics*, 31, 1154–1159
 57. Le Novère, N. (2015) Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.*, 16, 146–158
 58. Zhang, P., Tao, L., Zeng, X., Qin, C., Chen, S., Zhu, F., Li, Z., Jiang, Y., Chen, W. and Chen, Y. Z. (2016) A protein network descriptor server and its use in studying protein, disease, metabolic and drug targeted networks. *Brief. Bioinform.*, bbw071
 59. Moretti, S., Martin, O., Van Du Tran, T., Bridge, A., Morgat, A. and Pagni, M. (2016) MetaNetX/MNXref – reconciliation of

- metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Res.*, 44, D523–D526
60. Ye, Y. N., Ma, B. G., Dong, C., Zhang, H., Chen, L. L. and Guo, F. B. (2016) A novel proposal of a simplified bacterial gene set and the neo-construction of a general minimized metabolic network. *Sci. Rep.*, 6, 35082
61. Thompson, R. A., Dahal, S., Garcia, S., Nookaew, I. and Trinh, C. T. (2016) Exploring complex cellular phenotypes and model-guided strain design with a novel genome-scale metabolic model of *Clostridium thermocellum* DSM 1313 implementing an adjustable cellulosome. *Biotechnol. Biofuels*, 9, 194
62. van Heck, R. G., Ganter, M., Martins Dos Santos, V. A. and Stelling, J. (2016) Efficient reconstruction of predictive consensus metabolic network models. *PLoS Comput. Biol.*, 12, e1005085