Published online 29 November 2016 *Nucleic Acids Research, 2017, Vol. 45, Database issue* ***D543–D550***
doi: 10.1093/nar/gkw1003

The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12

Ingrid M. Keseler[1], Amanda Mackie[2], Alberto Santos-Zavaleta[3], Richard Billington[1], César Bonavides-Martínez[3], Ron Caspi[1], Carol Fulcher[1], Socorro Gama-Castro[3], Anamika Kothari[1], Markus Krummenacker[1], Mario Latendresse[1], Luis Muñiz-Rascado[3], Quang Ong[1], Suzanne Paley[1], Martin Peralta-Gil[3], Pallavi Subhraveti[1], David A. Velázquez-Ramírez[3], Daniel Weaver[1], Julio Collado-Vides[3], Ian Paulsen[2] and Peter D. Karp[1,*]

[1]SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025, USA, [2]Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW, Australia and [3]Programa de Genómica Computacional, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, A.P. 565-A, Cuernavaca, Morelos 62100, Mexico

Received September 21, 2016; Editorial Decision October 09, 2016; Accepted November 07, 2016

ABSTRACT

EcoCyc (EcoCyc.org) is a freely accessible, comprehensive database that collects and summarizes experimental data for *Escherichia coli* K-12, the best-studied bacterial model organism. New experimental discoveries about gene products, their function and regulation, new metabolic pathways, enzymes and cofactors are regularly added to EcoCyc. New SmartTable tools allow users to browse collections of related EcoCyc content. SmartTables can also serve as repositories for user- or curator-generated lists. EcoCyc now supports running and modifying *E. coli* metabolic models directly on the EcoCyc website.

INTRODUCTION

Over the course of many decades, thousands of researchers have contributed to the experimental study of *Escherichia coli*. Considering its moderate-size genome that contains roughly 4500 genes, it is easy to assume that we already know most of what there is to know about *E. coli*—its genes, gene products and their functions, as well as its cellular metabolism and interactions with the environment. In fact, however, many facets of *E. coli* biology are still unknown; for example, there are still a number of essential genes of unknown function (e.g. (1)). Basic *E. coli* research continues to generate surprising insights, and new experimental methods enable researchers to fill holes in our knowledge and solve long-standing mysteries. Because *E. coli* serves as a model bacterium for many less well-studied organisms, new discoveries should have broad impact. One of the most significant uses of a model organism database such as EcoCyc is to collect and disseminate both longstanding knowledge as well as recent research advances in an easily accessible format. EcoCyc continues to serve in this role for the *E. coli* research community and, through its integration within the BioCyc collection of thousands of organism-specific databases, provides a simple way to find and compare orthologous genes and metabolic pathways across a wide spectrum of organisms.

Since our last publication on EcoCyc in the 2013 NAR Database Issue (2), many improvements to EcoCyc, the Pathway Tools software, the BioCyc family of websites and the display and analysis tools available there have been described elsewhere (3–6). Here, we focus on recent updates to the EcoCyc web site and on additions to the database that reflect new knowledge about *E. coli* biology.

UPDATES TO ECOCYC

Manual curation within EcoCyc continues to adopt a two-pronged approach. Priority is given to adding new functions for gene products when they are reported in the literature. Considerable effort also goes into updating older entries in the database; typically this is undertaken by reviewing all the proteins belonging to a specific family, metabolic pathway or regulon. Occasionally, large datasets containing, for example, gene essentiality or protein localization information are added computationally. An overview of the current data content of EcoCyc is shown in Table 1.

Update of the sequence version in EcoCyc from GenBank U00096.2 to U00096.3

As of May 2016 (release 20.0), EcoCyc uses the U00096.3 version of the *E. coli* K-12 MG1655 genome sequence. The

*To whom correspondence should be addressed. Tel: +1 650 859 4358; Fax: +1 650 859 3735; Email: pkarp@ai.sri.com

Table 1. EcoCyc content and *E. coli* gene product functions

Data type	Number (Release 20.1)
Genes	4505
Gene products covered by a mini-review	3884
Gene products with GO terms with EXP evidence	3350
Enzymes	1567
Metabolic reactions	1913
Compounds	2699
Transporters	282
Transport reactions	485
Transported substrates	338
Transcription factors	204
Regulatory interactions	6399
Transcription initiation	3457
Transcription attenuation	23
Regulation of translation	212
Enzyme modulation	2675
Other	32
Literature citations	31 999

GenBank sequence record had been updated from version U00096.2 to U00096.3 in 2013, because the U00096.2 sequence did not precisely correspond to a specific isolate of the MG1655 strain (7). Importantly for next-generation sequencing studies, U00096.2 differed from the genome of the strain deposited as the sequenced MG1655 strain at the American Type Culture Collection (ATCC 700926) and the Yale Coli Genetic Stock Center (CGSC7740). Sequence differences between common *E. coli* strains have been described by Freddolino *et al.* (8). Most significantly, the final sequenced strain differs from other MG1655 strains by carrying mutations that inactivate the transcriptional regulators encoded by *crl* and *glpR* and the galactitol transporter component encoded by *gatC*. Due to an IS1 element insertion and other indels, the nucleotide coordinates of genes and other features differ between U00096.2 and U00096.3. To ease the transition for researchers with datasets that use the prior genome coordinates, the EcoCyc web site is offering a coordinate mapping service that translates data files containing the old genome coordinates to new coordinates. The 'Map Sequence Coordinates' function can be found under the Genome menu on the EcoCyc homepage. However, researchers should keep in mind that the genome of any given laboratory stock of MG1655 will also differ from the published genome sequence, and that some differences will be physiologically significant.

Updates to the data content in EcoCyc

The transporter systems in EcoCyc. Since our last publication in the 2013 NAR database issue (2), 14 new transport proteins or complexes have been characterized and curated accordingly (Table 2). We have also reviewed and updated the curation of 48 proteins, both membrane and cystosolic, which belong to the functional superfamily of the phosphoenolpyruvate (PEP)-dependent, sugar transporting phosphotransferase systems (PTS^{sugar}). This work extends our representation of the range of substrates, both physiological and non-physiological, that *E. coli* can transport across the membrane. For example, methyl α-D-glucoside was added as a non-physiological substrate of the glucose PTS, and D-sorbitol was added as a physiologically relevant substrate of the galactitol PTS. Our coverage of the literature was improved through the addition of a further 144 citations.

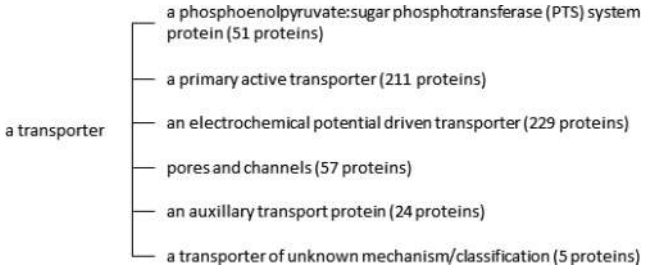

Figure 1. Transporter classes in EcoCyc.

Update of transporter classification. We have reviewed and updated transporter classes within the Pathway Tools ontology. The class 'a transporter' within EcoCyc now contains 6 child classes (Figure 1) plus numerous sub-classes, named according to IUBMB recommendations as based on the Transporter Classification Database (9). All 577 transport proteins in EcoCyc are classified within this ontology, making it straightforward for a user to accurately determine the number and identity of proteins within a particular class. For example, searching EcoCyc with the term 'a primary active transporter' will return the transporter class of the same name containing the two child classes of 'P-type ATPases' (four proteins) and 'an ATP binding cassette transporter' (52 substrate binding proteins, 64 ATP binding proteins and 86 integral membrane subunits).

Update of electron transport pathways and respiratory enzymes. We have completed a long-term project to update the curation of electron transport (ET) pathways and respiratory enzymes in EcoCyc. An initiative to represent ET pathways in EcoCyc was first described in 2009 (10) along with the subsequent addition of 11 ET pathways. We have now added a further 15 pathways (Table 3), bringing the total number to 26. All pathways contain a fully referenced text summary, including (when known) information on energetics, isoenzyme involvement and the identity of membrane quinone(s). The curation of all 23 respiratory enzymes involved in ET pathways has also been updated. Particular attention has been given to ensuring that the correct cofactors of each respiratory enzyme are identified (when known). In addition, a new, recently-described cofactor, the 4Fe-3S iron–sulfur cluster of hydrogenase I (11), was added to the database as part of this project. Supplementary Table S1 summarizes the respiratory enzymes and their associated cofactors as currently represented in EcoCyc. Just over 200 references were added to the database, the majority of these dating from the last quarter of the 20th century, a period of intense research activity in *E. coli* bioenergetics.

New transcription factors and updated information related to transcriptional regulation in EcoCyc. A total of 14 new transcription factors (TFs) regulating a variety of different biological processes have been identified in the experimental literature and have been added to the database since fall

Table 2. New membrane transporters characterized in *E. coli* K-12 and curated in EcoCyc

Gene name (old gene name)	Protein Function	Reference
dauA (ychM)	C_4 dicarboxylate transporter	(23)
ghxP (yjcD)	Guanine/hypoxanthine transporter	(24)
ghxQ (ygfQ)	Guanine/hypoxanthine transporter	(24)
adeQ (yicO)	Adenine transporter	(24)
satP (yaaH)	Succinate/acetate:H+ symporter	(25)
yihO	Putative sulfoquinovose transporter	(17)
cysZ	Sulfate:H^+ symporter	(26)
yijE	Cystine exporter	(27)
tcyJ tcyL tcyN (fliY yecS yecC)	Cystine/cysteine ABC transporter	(28,29)
nimT (yeaN)	Drug efflux transporter	(30)
lysO (ybjE)	L-Lysine exporter	(31)
yjeH	L-Methionine/branched chain amino acid exporter	(32)
osmF yehY yehX yehW	Glycine betaine ABC transporter	(33)

Table 3. Electron transport pathways added to EcoCyc

Pathway name	Enzymes involved	Quinone species used
NADH to cytochrome *bo* oxidase electron transfer II	NADH:ubiquinone oxidoreductase II; cytochrome *bo* oxidase	Ubiquinone
NADH to cytochrome *bd* oxidase electron transfer II	NADH:ubiquinone oxidoreductase II; cytochrome *bd*-1 oxidase	Ubiquinone
Nitrate reduction VIIIb (dissimilatory)	NADH:ubiquinone oxidoreductase II; nitrate reductase A	Ubiquinone
Pyruvate to cytochrome *bo* oxidase electron transfer	Pyruvate oxidase; cytochrome *bo* oxidase	Ubiquinone
Pyruvate to cytochrome *bd* oxidase electron transfer	Pyruvate oxidase; cytochrome *bd*-I oxidase	Ubiquinone
Glycerol-3-phosphate to cytochrome *bo* oxidase electron transfer	Aerobic glycerol-3-phosphate dehydrogenase; cytochrome *bo* oxidase	Ubiquinone
Glycerol-3-phosphate to fumarate electron transfer	Anaerobic glycerol-3-phosphate dehydrogenase; fumarate reductase	Menaquinone
Nitrate reduction IX (dissimilatory)	Anaerobic glycerol-3-phosphate dehydrogenase; nitrate reductase A; nitrate reductase Z	Menaquinone
Nitrate reduction X (dissimilatory, periplasmic)	Aerobic glycerol-3-phosphate dehydrogenase; periplasmic nitrate reductase	Ubiquinone
D-Lactate to cytochrome *bo* oxidase electron transfer	D-Lactate dehydrogenase; cytochrome *bo* oxidase	Ubiquinone
Proline to cytochrome *bo* oxidase electron transfer	Proline dehydrogenase; cytochrome *bo* oxidase	Ubiquinone
Formate to nitrite electron transfer	Formate dehydrogenase; formate dehydrogenase N; formate-dependent nitrite reductase	Menaquinone
Hydrogen to dimethyl sulfoxide electron transfer	Hydrogenase I; hydrogenase II; dimethyl sulfoxide reductase	Menaquinone
Hydrogen to fumarate electron transfer	Hydrogenase II; fumarate reductase	Menaquinone
Hydrogen to trimethylamine *N*-oxide electron transfer	Hydrogenase I; hydrogenase II; trimethylamine *N*-oxide reductase	Menaquinone

2012. The functions of these TFs are summarized in Table 4. In addition to the new TFs, there has been an increase in other database objects like transcription units, regulatory interactions and transcription factor binding sites (TFBSs), promoters and terminators (Table 5).

Table 6 summarizes updates to existing TFs within EcoCyc. For several TFs, active or inactive protein conformations have been identified. For example, it was shown that only the homotetrameric conformation of the quorum-sensing regulator LsrR is active, while the autoinducer-bound conformation (LsrR-AI-2) is inactive. Binding of AI-2 to LsrR may disrupt the tetrameric assembly of LsrR, resulting in its dissociation from DNA (12).

The newly discovered iron-sulfur cluster-bound conformation of IscR (IscR-[2Fe-2S]) was shown to regulate the expression of genes involved the iron-sulfur cluster assembly pathway through negative feedback that depends on the cellular Fe-S cluster demand (13). IscR-[2Fe-2S] negatively regulates the expression of *iscRSUA,* genes of the Isc Fe–S cluster biosynthesis pathway, and activates expression of the *sufABCDSE* operon, which encodes the Suf Fe-S cluster biosynthesis pathway. Coordinated regulation of these two pathways maintains differential control of Fe–S cluster biogenesis and ensures viability under a variety of growth conditions (14). Both apo-IscR and the IscR-[2Fe–2S] holoprotein conformations were able to activate the Suf pathway (14).

Genomic SELEX screening usually results in the discovery of many target sites for transcriptional regulators. Surprisingly, only one target was found for each of the two newly discovered regulators, CecR and DecR, These regulators were found to be associated with novel roles in the control of sensitivity to cefoperazone and chloramphenicol (CecR) (15) and cysteine detoxification systems (DecR) (16).

In EcoCyc, evidence codes are attached to many types of data and generally contain a supporting literature citation. Filling gaps in our representation of transcriptional regulation, we have added missing references to the published experimental evidence to a set of 225 promoter objects.

Table 4. New transcription factors characterized in *E. coli* K-12 and curated in EcoCyc

Gene name (old gene name)	Processes regulated by the transcription factor	Reference
nimR (*yeaM*)	Resistance to 2-nitroimidazole	(30)
sutR (*ydcN*)	Sulfur metabolism	(34)
ydcI	Stress response	(35)
mraZ (*yabB*)	Cell division, cell wall	(36)
rclR (*ykgD*)	Survival under reactive chlorine stress	(37)
pgrR (*ycjZ*)	Peptidoglycan degradation	(38)
rcdA (*ybjK*)	Stress response, biofilm formation	(39)
ydfH	*rspAB* operon expression	(40)
yjjQ	Flagellar synthesis, capsule formation	(41)
yebK	Adaptation to growth in cellobiose minimal medium	(42)
ypdB	Carbon control network, nutrient scavenging	(43)
yedW	H_2O_2 sensing	(44)
cecR (*ybiH*)	Sensitivity to cefoperazone and chloramphenicol	(15)
decR (*ybaO*)	Cysteine detoxification	(16)

Table 5. Data related to transcriptional regulation

Data type	Total	New
Transcription Unit	3553	95
Promoter	3841	73
Terminator	283	31
Transcription Factor	205	14
Transcription Factor Binding Site	2836	199
Regulatory Interaction	3374	183

Predicted transporters and transcription factors. To encourage further research on *E. coli* gene products that have no known function to date, the EcoCyc project has generated a set of public SmartTables that contain (i) inner membrane proteins with minimal or no characterization, (ii) predicted, but uncharacterized inner membrane transporters and (iii) predicted transcription factors. These tables will be updated on an ongoing basis and are available under the 'SmartTables > Public SmartTables' menu or by using the following links:

http://ecocyc.org/group?id=biocyc13-4655-3682893892
(uncharacterized inner membrane proteins)
http://ecocyc.org/group?id=biocyc17-4655-3682299327
(predicted inner membrane transporters)
http://ecocyc.org/group?id=biocyc17-1553-3682788185
(predicted transcriptional regulators)

Static versions of these tables are available as Supplementary Tables S2–S4.

E. coli *metabolism.* Although the basic metabolic capabilities of *E. coli* have been well studied, surprising new discoveries continue to be made. A recent example is *E. coli*'s ability to utilize the six-carbon sulfur sugar sulfoquinovose as the sole source of carbon and energy for growth (17). Sulfoquinovose is a major component of organo-sulfur compounds in nature (18). It is synthesized by higher plants, mosses, ferns, algae and most photosynthetic bacteria and serves as the polar headgroup of the sulfolipid in photosynthetic membranes (19,20). Sulfoquinovose is structurally similar to glucose, and degradation of this sugar follows a pathway that is highly similar to glycolysis (Figure 2). The sulfur-containing three-carbon degradation product sulfopropanediol is excreted and can be utilized as both a carbon and sulfur source by other organisms (21). However, open questions remain. Although proteins with suggestive predicted functions or mutant phenotypes are encoded in the genomic vicinity of the sulfoquinovose-degrading enzymes, neither the importer for sulfoquinovose nor the exporter for sulfopropanediol has been firmly established, and no regulatory mechanisms are yet known.

New discoveries in *E. coli* K-12 also further our understanding of the physiology of microbial cell envelopes as exemplified by the recent characterization of a periplasmic methionine sulfoxide reducing system (22). This system, comprised of a periplasmic methionine sulfoxide reductase (MsrP) and an inner membrane, heme-binding, quinol dehydrogenase (MsrQ), functions to protect periplasmic proteins from oxidative damage and is conserved throughout Gram-negative bacteria (22). The use of membrane quinols as a source of reducing power in the cell envelope is a novel finding and represents a notable advance in our understanding of how bacteria repair damaged proteins.

Updates to the EcoCyc website and Pathway Tools software

Updates to the genome browser. We have added a new zoom level to the BioCyc genome browser. It is now possible to zoom to the sequence level, which will show details such as transcription start sites, transcription factor binding sites, other regulatory sites such as attenuators and interaction sites for small regulatory RNAs, as well as gene and protein sequences. This zoom level thus enables inspection of the relative location of sites within the sequence. Figure 3 shows the new zoom level, comparing it to two previously available zoom levels.

Updates to SmartTables. The SmartTables tools for manipulating sets of genes, chemical compounds, and other objects within EcoCyc and other BioCyc databases have been expanded in several respects. A new set of special SmartTables (see menu SmartTables→Special SmartTables) allows the user to explore and manipulate various sets of entities within EcoCyc, such as all chemical compounds, all metabolic pathways, all promoters, all terminators, all riboswitches, all non-coding RNAs, and all proteins or protein subtypes such as all transporters and all transcription factors.

Table 6.

Type of update	TFs
Updated summaries	AcrR, AraC, ArcA, BaeR, BasR (PmrA), BluR, BolA, CadC, ChbR, CpxR, CRP, CspA, CueR, DecR (YbaO), DksA, ExuR, FadR, FeaR, Fur, H-NS, HipB, HU, HypT, IHF, IscR, LacI, LeuO, LsrR, MalT, MarA, MarR, MarRAB, MazE, McbR, MlrA, MqsR, MraZ, NarL, NemR, NikR, NorR, NrdR, OmpR, PhoB, PspF, RbsR, RcnR, RcsA, RcsB, RcsB- BglJ, Rob, RpoD, RpoE, RpoH, RpoS, RstA, RutR, SdiA, SoxR, SoxS, TyrR, UxuR, YehT, YpdB, Zur
New conformations	LsrR (homotetramer), LsrR-AI-2, MetJ-MTA, MetJ-adenine, IscR-2Fe-2S
Relocalization of TFBSs	PuuR

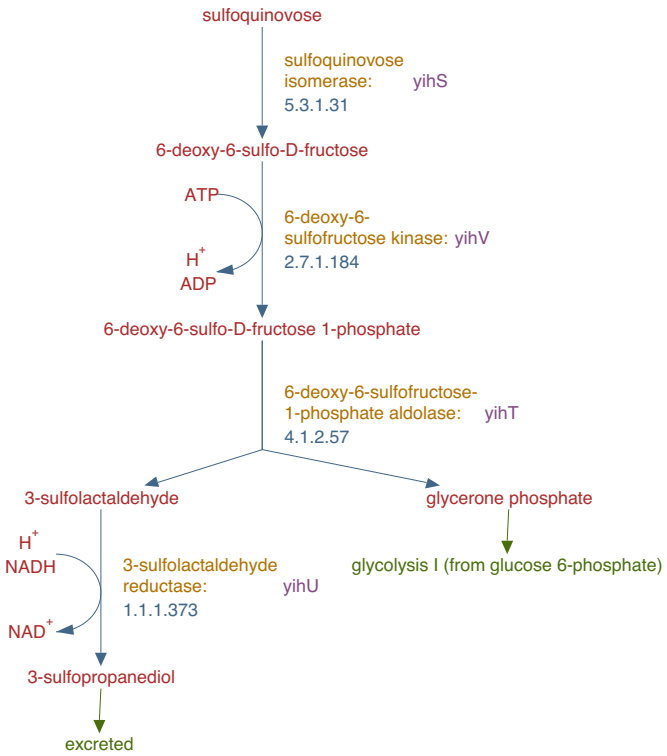

Figure 2. Sulfoquinovose degradation pathway.

Execution of the E. coli *metabolic model via the EcoCyc Website.* Because the metabolic model associated with EcoCyc is derived directly from EcoCyc using the MetaFlux module of Pathway Tools, its data content has continued to evolve as we update the set of metabolic reactions and transporters within EcoCyc. Previously, it was only possible to execute the EcoCyc metabolic model using the downloadable Pathway Tools software. To make it more accessible to users, the model can now be executed directly on the EcoCyc web site. Web-based metabolic models are also available for two other gut microbiome organisms in the BioCyc database collection, *Bacteroides thetaiotaomicron* and *Eubacterium rectale*.

In basic terms, a metabolic model consists of a set of active reactions plus the conditions of growth of the organism; the models stored within the EcoCyc website contain both. The active reactions correspond to those reactions that are active at a given time based on cellular regulation, and can be either the full set or a subset of the reactions stored within EcoCyc. For each modeled growth situation, the conditions of growth consist of the **nutrients** available to the growing *E. coli* cell, the set of **biomass metabolites** (the end products of the metabolic network) produced by the cell, and the metabolic end-products that are secreted by the cell (the **secretions**). The process of running a metabolic model through the EcoCyc website consists of the following steps.

First, choose EcoCyc as your current organism by clicking Select Organism Database in the upper-right corner of the screen.

Second, click Run Metabolic Model under the Metabolism menu.

Third, log in to your EcoCyc (BioCyc) account if you are not already logged in.

At this point you can either select an existing model to run, or you can create a new model if none of the existing models cover exactly the situation you wish to model. Usually, it is easiest to create a new model by copying and editing an existing model. You can select an existing model to run from either the list of models that other people have made public, or from a list of models that you may have saved in the past. For example, select the public Glucose Fermentation model to select a model that anaerobically ferments glucose. Once you select a model you can inspect the nutrients, reactions, biomass metabolites, and secretions that it contains by selecting a tab of the corresponding name toward the bottom of the model-summary page (see Figure 4). To actually run the model, click Execute within the Results tab. The result of running a model is a list of steady-state reaction flux values for those metabolic reactions that carry non-zero flux, which are presented in a table. For example, Figure 4 shows that the two highest fluxes in the entire metabolic network during glucose fermentation are through two reactions in glycolysis. Those fluxes can be painted on the EcoCyc metabolic map diagram by clicking the button **Show Fluxes on Cellular Overview.** Additional information is available from the solution file and the log file (which can be accessed via buttons in the Reactions tab), such as the uptake fluxes of each nutrient. Imagine you want to run a model that ferments galactose instead of glucose. To do so, click the Copy button at the top of the model-selection page, enter the Nutrients tab, and then replace β-D-glucopyranose with β-D-galactose. The Nutrients tab allows you to place upper and lower bounds on the uptake rates of different nutrients. Since models typically attempt to optimize the cellular growth rate, an upper bound must be provided for some nutri-

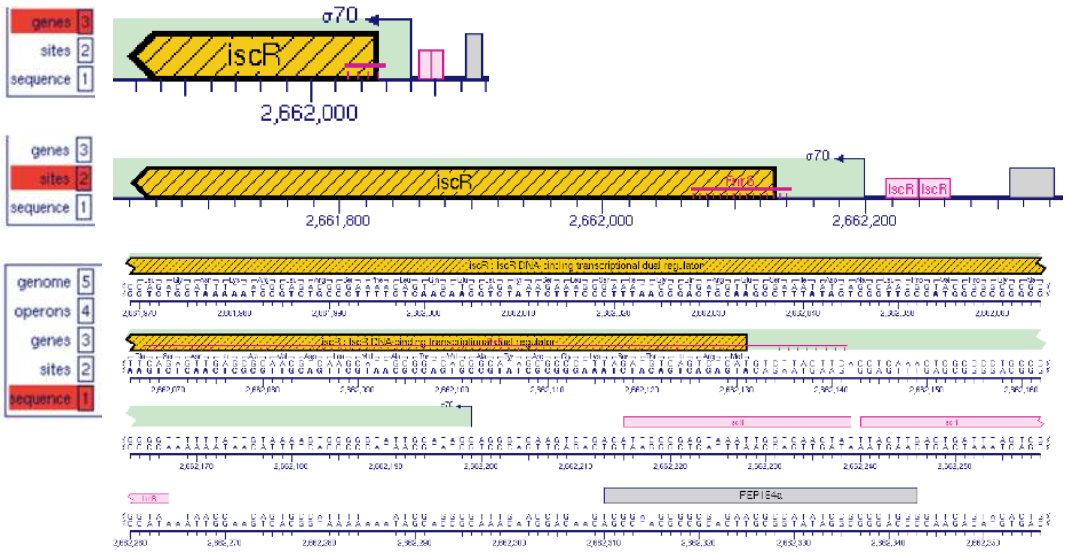

Figure 3. Zoom levels of the genome browser, with the previously available 'genes' and 'sites' levels compared to the new 'sequence' level.

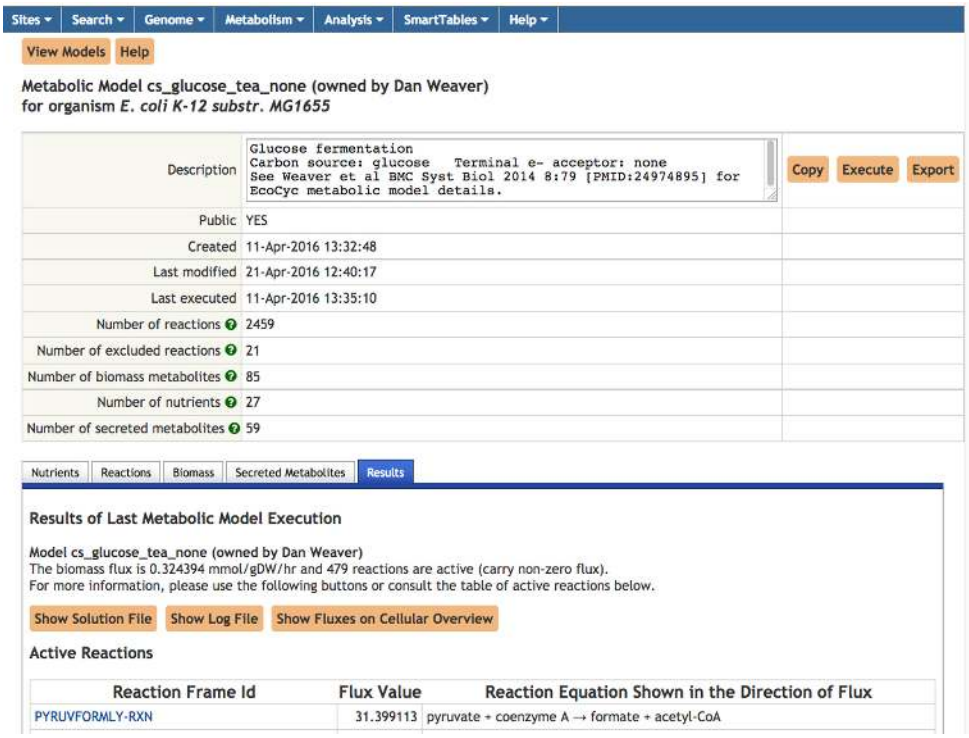

Figure 4. Running metabolic models on the EcoCyc web site.

ent, otherwise the model would attempt to produce infinite growth, which would stymie the mathematical solver software. Gene knock-outs can be simulated by specifying reactions to remove from the model from within the Reactions tab. Detailed instructions for running a metabolic model through the EcoCyc website are available by selecting the 'Getting Started Guide' link (http://ecocyc.org/PToolsWebsiteHowto.shtml#metabolicmodels).

ECOCYC AVAILABILITY

EcoCyc is freely and openly available to all. See http://ecocyc.org/download.shtml for download information. New versions of the downloadable EcoCyc data files and of the EcoCyc website are released three times per year. Access to the website is free; users are required to register for a free account after viewing more than 30 pages in a given month.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Institute of General Medical Sciences of the National Institutes of Health under Award Numbers U24GM077678 to P.D.K. and R01GM110597 to J.C.-V. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Funding for open access charge: National Institute of General Medical Sciences of the National Institutes of Health.
Conflict of interest statement. None declared.

REFERENCES

1. McClung,D.J., Calixto,A., Mosera,M.N., Kumar,R., Neidle,E.L. and Elliott,K.T. (2016) Novel heterologous bacterial system reveals enhanced susceptibility to DNA damage mediated by yqgF, a nearly ubiquitous and often essential gene. *Microbiology*, **162**, 1808-1821.
2. Keseler,I.M., Mackie,A., Peralta-Gil,M., Santos-Zavaleta,A., Gama-Castro,S., Bonavides-Martinez,C., Fulcher,C., Huerta,A.M., Kothari,A., Krummenacker,M. *et al.* (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.*, **41**, D605–D612.
3. Karp,P.D., Weaver,D., Paley,S., Fulcher,C., Kubo,A., Kothari,A., Krummenacker,M., Subhraveti,P., Weerasinghe,D., Gama-Castro,S. *et al.* (2014) The EcoCyc database. *EcoSal Plus*, **2014**.
4. Weaver,D.S., Keseler,I.M., Mackie,A., Paulsen,I.T. and Karp,P.D. (2014) A genome-scale metabolic flux model of Escherichia coli K-12 derived from the EcoCyc database. *BMC Syst. Biol.*, **8**, 79.
5. Karp,P.D., Latendresse,M., Paley,S.M., Krummenacker,M., Ong,Q.D., Billington,R., Kothari,A., Weaver,D., Lee,T., Subhraveti,P. *et al.* (2015) Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Brief Bioinform.*,**17**, 877-890.
6. 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**, D471–480.
7. Plunkett,G. 3rd, Richardson,A.J. and Rudd,K.E. (2013) Unpublished.
8. Freddolino,P.L., Amini,S. and Tavazoie,S. (2012) Newly identified genetic variations in common Escherichia coli MG1655 stock cultures. *J. Bacteriol.*, **194**, 303–306.
9. Saier,M.H. Jr., Reddy,V.S., Tsu,B.V., Ahmed,M.S., Li,C. and Moreno-Hagelsieb,G. (2016) The transporter classification database (TCDB): recent advances. *Nucleic Acids Res.*, **44**, D372–D379.
10. Keseler,I.M., Bonavides-Martinez,C., Collado-Vides,J., Gama-Castro,S., Gunsalus,R.P., Johnson,D.A., Krummenacker,M., Nolan,L.M., Paley,S., Paulsen,I.T. *et al.* (2009) EcoCyc: a comprehensive view of Escherichia coli biology. *Nucleic Acids Res.*, **37**, D464–470.
11. Pandelia,M.E., Bykov,D., Izsak,R., Infossi,P., Giudici-Orticoni,M.T., Bill,E., Neese,F. and Lubitz,W. (2013) Electronic structure of the unique [4Fe-3S] cluster in O2-tolerant hydrogenases characterized by 57Fe Mossbauer and EPR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 483–488.
12. Wu,M., Tao,Y., Liu,X. and Zang,J. (2013) Structural basis for phosphorylated autoinducer-2 modulation of the oligomerization state of the global transcription regulator LsrR from Escherichia coli. *J. Biol. Chem.*, **288**, 15878–15887.
13. Giel,J.L., Nesbit,A.D., Mettert,E.L., Fleischhacker,A.S., Wanta,B.T. and Kiley,P.J. (2013) Regulation of iron-sulphur cluster homeostasis through transcriptional control of the Isc pathway by [2Fe-2S]-IscR in Escherichia coli. *Mol. Microbiol.*, **87**, 478–492.
14. Mettert,E.L. and Kiley,P.J. (2014) Coordinate regulation of the Suf and Isc Fe-S cluster biogenesis pathways by IscR is essential for viability of Escherichia coli. *J. Bacteriol.*, **196**, 4315–4323.
15. Yamanaka,Y., Shimada,T., Yamamoto,K. and Ishihama,A. (2016) Transcription factor CecR (YbiH) regulates a set of genes affecting the sensitivity of Escherichia coli against cefoperazone and chloramphenicol. *Microbiology*, **162**, 1253–1264.
16. Shimada,T., Tanaka,K. and Ishihama,A. (2016) Transcription factor DecR (YbaO) controls detoxification of L-cysteine in Escherichia coli. *Microbiology*, **162**, 1698–1707.
17. Denger,K., Weiss,M., Felux,A.K., Schneider,A., Mayer,C., Spiteller,D., Huhn,T., Cook,A.M. and Schleheck,D. (2014) Sulphoglycolysis in Escherichia coli K-12 closes a gap in the biogeochemical sulphur cycle. *Nature*, **507**, 114–117.
18. Harwood,J.L. and Nicholls,R.G. (1979) The plant sulpholipid– a major component of the sulphur cycle. *Biochem. Soc. Trans.*, **7**, 440–447.
19. Benning,C. (2007) Questions remaining in sulfolipid biosynthesis: a historical perspective. *Photosynth Res.*, **92**, 199–203.
20. Benning,C. (1998) Biosynthesis and Function of the Sulfolipid Sulfoquinovosyl Diacylglycerol. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **49**, 53–75.
21. Denger,K., Huhn,T., Hollemeyer,K., Schleheck,D. and Cook,A.M. (2012) Sulfoquinovose degraded by pure cultures of bacteria with release of C3-organosulfonates: complete degradation in two-member communities. *FEMS Microbiol. Lett.*, **328**, 39–45.
22. Gennaris,A., Ezraty,B., Henry,C., Agrebi,R., Vergnes,A., Oheix,E., Bos,J., Leverrier,P., Espinosa,L., Szewczyk,J. *et al.* (2015) Repairing oxidized proteins in the bacterial envelope using respiratory chain electrons. *Nature*, **528**, 409–412.
23. Karinou,E., Compton,E.L., Morel,M. and Javelle,A. (2013) The Escherichia coli SLC26 homologue YchM (DauA) is a C(4)-dicarboxylic acid transporter. *Mol. Microbiol.*, **87**, 623–640.
24. Papakostas,K., Botou,M. and Frillingos,S. (2013) Functional identification of the hypoxanthine/guanine transporters YjcD and YgfQ and the adenine transporters PurP and YicO of Escherichia coli K-12. *J. Biol. Chem.*, **288**, 36827–36840.
25. Sa-Pessoa,J., Paiva,S., Ribas,D., Silva,I.J., Viegas,S.C., Arraiano,C.M. and Casal,M. (2013) SATP (YaaH), a succinate-acetate transporter protein in Escherichia coli. *Biochem. J.*, **454**, 585–595.
26. Zhang,L., Jiang,W., Nan,J., Almqvist,J. and Huang,Y. (2014) The Escherichia coli CysZ is a pH dependent sulfate transporter that can be inhibited by sulfite. *Biochim. Biophys. Acta*, **1838**, 1809–1816.
27. Yamamoto,K., Nonaka,G., Ozawa,T., Takumi,K. and Ishihama,A. (2015) Induction of the Escherichia coli yijE gene expression by cystine. *Biosci. Biotechnol. Biochem.*, **79**, 218–222.
28. Deutch,C.E., Spahija,I. and Wagner,C.E. (2014) Susceptibility of Escherichia coli to the toxic L-proline analogue L-selenaproline is dependent on two L-cystine transport systems. *J. Appl. Microbiol.*, **117**, 1487–1499.
29. Chonoles Imlay,K.R., Korshunov,S. and Imlay,J.A. (2015) Physiological roles and adverse effects of the two cystine importers of Escherichia coli. *J. Bacteriol.*, **197**, 3629–3644.
30. Ogasawara,H., Ohe,S. and Ishihama,A. (2015) Role of transcription factor NimR (YeaM) in sensitivity control of Escherichia coli to 2-nitroimidazole. *FEMS Microbiol. Lett.*, **362**, 1–8.
31. Pathania,A. and Sardesai,A.A. (2015) Distinct paths for basic amino acid export in Escherichia coli: YbjE (LysO) mediates export of L-lysine. *J. Bacteriol.*, **197**, 2036–2047.
32. Liu,Q., Liang,Y., Zhang,Y., Shang,X., Liu,S., Wen,J. and Wen,T. (2015) YjeH is a novel exporter of l-methionine and branched-chain amino acids in Escherichia coli. *Appl. Environ. Microbiol.*, **81**, 7753–7766.
33. Lang,S., Cressatti,M., Mendoza,K.E., Coumoundouros,C.N., Plater,S.M., Culham,D.E., Kimber,M.S. and Wood,J.M. (2015) YehZYXW of Escherichia coli Is a Low-Affinity, Non-Osmoregulatory Betaine-Specific ABC Transporter. *Biochemistry*, **54**, 5735–5747.
34. Yamamoto,K., Nakano,M. and Ishihama,A. (2015) Regulatory role of transcription factor SutR (YdcN) in sulfur utilization in Escherichia coli. *Microbiology*, **161**, 99–111.
35. Solomon,L., Shah,A., Hannagan,S. and Wilson,J.W. (2014) Bacterial genus-specific tolerance for YdcI expression. *Curr. Microbiol.*, **69**, 640–648.
36. Eraso,J.M., Markillie,L.M., Mitchell,H.D., Taylor,R.C., Orr,G. and Margolin,W. (2014) The highly conserved MraZ protein is a

transcriptional regulator in Escherichia coli. *J. Bacteriol.*, **196**, 2053–2066.
37. Parker,B.W., Schwessinger,E.A., Jakob,U. and Gray,M.J. (2013) The RclR protein is a reactive chlorine-specific transcription factor in Escherichia coli. *J. Biol. Chem.*, **288**, 32574–32584.
38. Shimada,T., Yamazaki,K. and Ishihama,A. (2013) Novel regulator PgrR for switch control of peptidoglycan recycling in Escherichia coli. *Genes Cells*, **18**, 123–134.
39. Shimada,T., Katayama,Y., Kawakita,S., Ogasawara,H., Nakano,M., Yamamoto,K. and Ishihama,A. (2012) A novel regulator RcdA of the csgD gene encoding the master regulator of biofilm formation in Escherichia coli. *Microbiologyopen*, **1**, 381–394.
40. Sakihama,Y., Mizoguchi,H., Oshima,T. and Ogasawara,N. (2012) YdfH identified as a repressor of rspA by the use of reduced genome Escherichia coli MGF-01. *Biosci. Biotechnol. Biochem.*, **76**, 1688–1693.
41. Wiebe,H., Gurlebeck,D., Gross,J., Dreck,K., Pannen,D., Ewers,C., Wieler,L.H. and Schnetz,K. (2015) YjjQ Represses Transcription of flhDC and Additional Loci in Escherichia coli. *J. Bacteriol.*, **197**, 2713–2720.
42. Parisutham,V. and Lee,S.K. (2015) Novel functions and regulation of cryptic cellobiose operons in Escherichia coli. *PLoS One*, **10**, e0131928.
43. Fried,L., Behr,S. and Jung,K. (2013) Identification of a target gene and activating stimulus for the YpdA/YpdB histidine kinase/response regulator system in Escherichia coli. *J. Bacteriol.*, **195**, 807–815.
44. Urano,H., Umezawa,Y., Yamamoto,K., Ishihama,A. and Ogasawara,H. (2015) Cooperative regulation of the common target genes between H(2)O(2)-sensing YedVW and Cu(2)(+)-sensing CusSR in Escherichia coli. *Microbiology*, **161**, 729–738.