

Optimizing genome-scale network reconstructions

Jonathan Monk^{1,3}, Juan Nogales¹⁻³ & Bernhard O Palsson¹

Metabolic reconstructions remain limited in their scope and content, and improvements in biochemical knowledge and collaborative research are required.

A genome-scale network reconstruction (GENRE) is built systematically using genome annotation, 'omics' data sets and legacy knowledge¹. Thus, GENREs should provide the best representation of the metabolic capabilities of a target organism on the basis of the information available at the time of reconstruction. They allow researchers to test and share new hypotheses about metabolic functions in a target organism. As a result, interest in network reconstructions and the scope of their applications has grown rapidly²⁻⁶.

The first GENRE was built for *Haemophilus influenza* in 1999 (ref. 7) just a few years after the first whole genome sequence was published in 1995 (ref. 8). This initial reconstruction represented a conceptual basis for building GENREs and demonstrated that the genotype-to-phenotype relationship of metabolic pathways could be discerned mechanistically at genome scale. Subsequently, guidance for generating metabolic reconstructions was developed and adopted on the basis of experiences with well-characterized model organisms. For example, in updates to the GENRE for *Escherichia coli* we suggested standards for modeling the relationships between genes, proteins and reactions involved in a particular biochemical transformation through the gene-protein-reaction association⁹. Next, mass and charge balancing for each reaction in the network and the addition of thermodynamic information¹⁰ were included. Updates

to the GENRE for *Saccharomyces cerevisiae* suggested a standard way to describe cellular compartmentalization¹¹. These guidelines have enabled reconstructions of human metabolism^{12,13}, photosynthesis¹⁴ and light-driven metabolism¹⁵, and have been used in various applications²⁻⁶. Furthermore, automated reconstruction approaches are now available to create draft reconstructions, reducing the time and effort required to make a metabolic reconstruction¹⁶⁻¹⁸.

During the past five years, the number of GENREs has grown rapidly (Fig. 1a) and expanded the 'metabolic space' that can be analyzed computationally³. Furthermore, GENREs have become accepted as valuable tools to teach and analyze biological processes at the systems level¹⁹. Therefore, more than a decade after the publication of the first GENRE, it is timely to analyze the metabolic knowledge represented in published network reconstructions to assess the overall progress and status of this field.

Coverage of metabolic reactomes

Although the metabolic network reconstruction field might appear to be mature, many challenges remain. Our analysis of the number of new metabolic reactions that have been incorporated into new GENREs in recent years shows that only a few reconstructions have added a substantial number of new reactions (Fig. 1) (see Supplementary Table 1 for a list of the 117 GENREs analyzed). Therefore, the metabolic coverage of GENREs has not progressed in line with the rising number of publications. Comparing enzymatic activities found in current GENREs to the BRENDA²⁰ enzyme database shows that only 33% of the enzymatic activities in BRENDA assigned to metabolism are included in the group of GENREs that we analyzed (Fig. 1b). Although this result could be

biased owing to incomplete mapping or redundant enzyme commission (EC) nomenclature, it indicates that current GENREs give incomplete coverage of known metabolic reactions.

Many new GENREs are based on existing reconstructions. Therefore, analogously to mistakes made with automated genome annotation²¹, the inclusion of an incorrect gene-protein-reaction association or an incorrect reaction in a GENRE can be disseminated to a new reconstruction. If the metabolic knowledge included in a GENRE reflects the metabolic capabilities of the target organism, we would expect clustering of GENRE content to reflect evolutionary relationships among organisms. However, our similarity analysis of GENRE reaction content shows that this is not the case. Multiple correspondence analysis (MCA) of the content of 53 curated GENREs out of the 117 published by February 2013 (Supplementary Tables 1 and 2 and Supplementary Data) (see also the University of California, San Diego, Systems Biology Research Group (SBRG) website, <http://sbrg.ucsd.edu/optimizing-genres/>), shows a high degree of similarity among many existing GENREs, regardless of their location in the phylogenetic tree (Fig. 2a). Many GENREs cluster close to the center of the diagram, showing that reconstructed organisms as metabolically diverse as *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Clostridium beijerinckii* and *Synechocystis sp.* PCC 6803 have similar reaction content (Fig. 2b). This clustering suggests that the metabolic space of currently published GENREs is largely limited to well-conserved metabolic pathways, rather than offering a comprehensive representation of the biochemical capabilities of these organisms, and there is an over-representation of primary metabolic pathways in GENREs relative to secondary metabolic pathways.

¹Department of Bioengineering, University of California, San Diego, La Jolla, California, USA. ²Present address: Department of Environmental Biology, Centro de Investigaciones Biológicas, CSIC, Madrid, Spain. ³These authors contributed equally to this work. Correspondence should be addressed to B.O.P. (palsson@ucsd.edu).

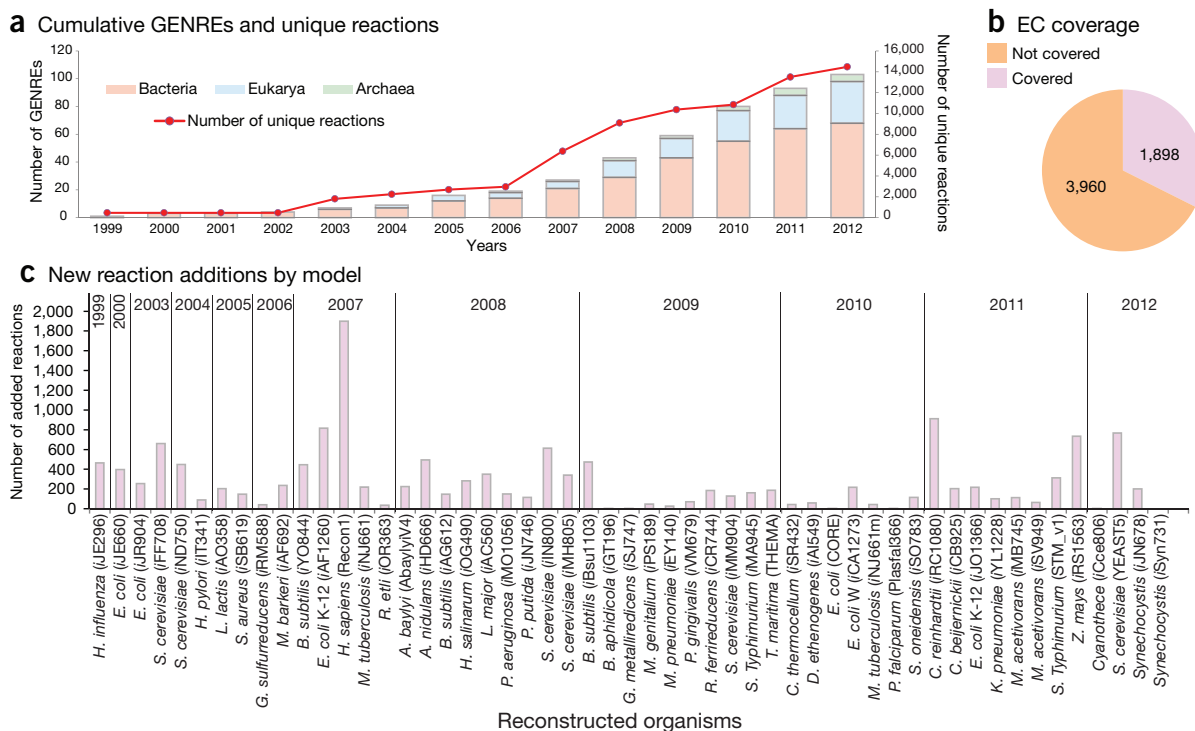


Figure 1 Expansion of metabolic networks and global reactome coverage over time. **(a)** By year, the cumulative number of GENREs published (vertical bars) and unique reactions included in all GENREs (red dots and line). **(b)** The proportion of Enzyme Commission (EC) numbers included in published GENREs. **(c)** Contribution to the coverage of metabolic space of each GENRE publication, as determined by the number of unique reactions added by each GENRE at the time of publication. The GENREs are ordered by publication date from *H. influenza* (iJE296) published in 1999, to *Synechocystis* (iSyn731), published in 2012.

By contrast, three groups of GENREs—dominated by enterobacteria, yeasts and photosynthetic eukaryotes, respectively—have distinct reactomes (Fig. 2a). The first GENRE of *E. coli*, iJE660 (ref. 22), is in the cluster at the center of the diagram, but later updates—iJR904 (ref. 9), iAF1260 (ref. 10) and iJO1366 (ref. 23)—are farther from the center, reflecting increasing organism-specific metabolic content (Fig. 2c). A similar pattern is observed for *S. cerevisiae*, from the initial GENRE (iFF708)¹¹ through to the latest version (YEAST5)²⁴. Given these two examples, it seems reasonable to expect that coverage of unique metabolic characteristics for other target organisms can be expanded through extensive manual curation of legacy metabolic knowledge, where it exists.

GENREs represent biochemically and genetically structured knowledge bases; therefore, well-curated, species-specific reactomes for many organisms might be a way to assess biological diversity. However, the range of organisms for which GENREs exist is limited, raising questions about the breadth of coverage of metabolism across the biosphere. We used the National Center for Biotechnology Information (NCBI) taxonomy database to examine the phylogenetic distribution of published GENREs (Supplementary Table 1). This examination reveals that some phyla have multiple GENREs, but many others lack GENREs (Fig. 3). For

example, >40% (32 of 78) of the species for which there are metabolic reconstructions are members of the proteobacteria phylum. By contrast, there are 15 phyla containing species with sequenced genomes but without any GENREs. All three domains—Bacteria, Archaea and Eukarya—include such unrepresented phyla. Thus, the phylogenetic coverage of current GENREs provides an incomplete representation of the metabolic capabilities found on Earth, and further reconstructions of diverse organisms are needed to achieve broad phylogenetic coverage.

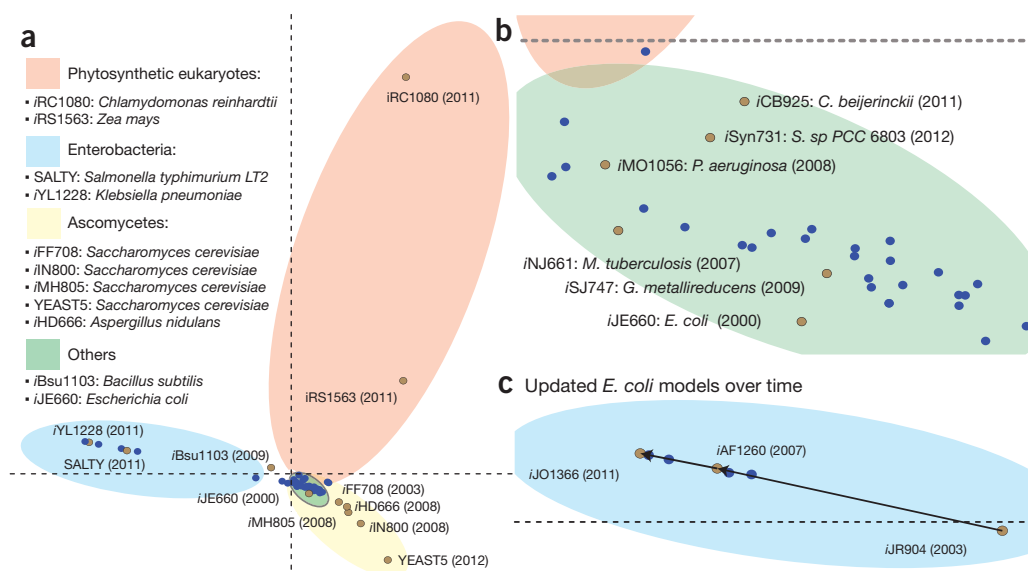
In summary, more organisms across the tree of life need to be reconstructed, and many current metabolic network reconstructions are lacking a substantial portion of their target organism's reactome. The development stage of such GENREs can be likened to the first reconstruction of *E. coli*, iJE660, published 14 years ago. An additional consideration is that the notions of 'quality' and 'completeness' of a metabolic network reconstruction are not well defined (Box 1 and Supplementary Table 3).

Limitations on GENRE development

Limited biological knowledge of the target organism is the main reason for the limited metabolic coverage in current GENREs (Fig. 4). Even for a microorganism as well studied as *E. coli*, only half (~54%) of the protein-encoding gene products have direct experimental evidence for their

function²⁵, and up to one-third of the proteome remains functionally un-annotated²⁶. This limitation is more substantial for less-studied organisms with a low species knowledge index (SKI)²⁷. The SKI is defined as the ratio of publications divided by the number of protein-encoding genes for a given organism. Although GENREs for organisms with low SKIs often have limited evidence to support modeled reactions, there is no clear relationship between the amount of biochemical information available for a given organism and the final metabolic coverage of its GENRE; species with high SKI values can still have GENREs with limited metabolic coverage. For example, the *P. aeruginosa* (SKI = 8.7) reconstruction has a similar reactome to the *C. beijerinckii* (SKI = 0.05) model (Fig. 2a). This indicates that some reconstructions might not fully utilize existing biological knowledge. The main reason for such under-utilization is a lack of substantial manual curation, probably owing to the extensive time and resources required. Genome annotation is often used as the main source of content in new GENREs, but genome annotation cannot be considered a complete and accurate source of biochemical information²¹. In addition, many organisms have a large number of un-annotated protein-encoding genes, and a substantial portion of the metabolic space known on Earth is excluded because 30–40% of known enzymatic activities are 'orphans' (that is, no genes encod-

Figure 2 Similarity analysis of GENREs. **(a)** The GENREs cluster in four groups. Most are grouped close to the center of coordinates (green ellipse), reflecting minimal differences between them. However, enterobacteria (yellow), yeasts (blue) and photosynthetic eukaryotes (pink) are more distant, indicating that these reconstructions differ from each other and have content that covers different sections of the metabolic space. **(b)** Detail of the main group of GENREs at the center of the coordinates in **a**. The brown circles show examples of metabolically distinct organisms with GENREs that clustered together. **(c)** Detail of the multiple iterations of GENREs for the model organism *E. coli*. Of 117 GENREs published and curated (as of February 2013), up to 53



could be consistently represented and were subject to multiple correspondence analysis. The other reconstructions were unavailable or had inconsistent nomenclature for reactions and metabolites⁵⁴. The human GENRE (Recon 1) was removed from this analysis because it is substantially different from the other GENREs analyzed (see also SBRG website (<http://sbrg.ucsd.edu/optimizing-genres/>) and **Supplementary Table 2**).

ing these activities have been discovered)^{28–30}. Thus, the combination of a paucity of biological knowledge, extensive reliance on genome annotation and incomplete use of legacy data from biochemical and/or physiological studies is likely to contribute to substantial underrepresentation of the metabolic potential of the target organism.

The limited use of legacy data in reconstructions is particularly problematic for secondary metabolism. Secondary metabolism tends to be organism specific and is therefore difficult to reconstruct based on genome annotation alone. Secondary metabolic pathways often include the biosynthesis of cofactors, vitamins, lipids and the cell envelope. Therefore, underrepresentation of these pathways in GENREs affects the biomass objective function (BOF), which we have proposed to define cellular growth requirements³¹, resulting in BOFs that do not reflect the physiology of the organism. The incomplete modeling of secondary metabolism biochemical pathways also reduces the computable metabolic space by creating ‘blocked’ reactions³² (that is, reactions that cannot be used owing to missing connections in the network). This prevents the use of GENREs for fundamental systems biology studies, such as gene essentiality prediction and systematic interpretation of high-throughput data. Lack of biological knowledge and inadequate curation of legacy data can be partially mitigated through the use of high-throughput phenotyping. More thorough use of such technologies would be desirable; however, the application of high-throughput data sets for modeling is still too costly to be readily accessible to individual reconstruction efforts, and these technologies,

although useful, often remain untapped in metabolic reconstructions.

A lack of rigor in applying established reconstruction protocols is another factor contributing to the limited metabolic coverage of current GENREs. Despite the existence of standards for reconstruction, GENREs that include unbalanced mass and charge reactions, incomplete pathways and noncompartmentalized networks are still being published. For example, it is striking that from 54 reconstructions of Gram-negative bacteria, just 11 (~20%) include periplasm as a cellular compartment (**Supplementary Table 2** and SBRG website (<http://sbrg.ucsd.edu/optimizing-genres/>)). Another weakness in the reconstruction process is a lack of a standardized representation for common metabolites and reactions. This shortcoming can make metabolic reconstructions difficult to interpret and impede automated high-throughput data mapping. It also prevents comparative analysis of GENREs; indeed, lack of a standardized representation is the main reason we compared only 53 reconstructions in the analysis presented here. Clearly, the field needs to recognize these limitations and continue to develop and adhere to best practices.

Toward more comprehensive GENREs and broader deployment

In light of the points we have raised, optimizing the content of GENREs is a key goal. In this section we highlight three issues that we believe are crucial to achieving this end.

Targeted application of high-throughput technologies. Biological knowledge can be increased by carefully applying high-throughput technolo-

gies. For example, several recent studies have used targeted high-throughput metabolomic, transcriptomic and mutant screen data sets as well as computational structure and metabolite-docking predictions to discover new functions^{33–35}.

There is also a need to develop high-throughput technologies to characterize reactomes, but large-scale biochemistry is, unfortunately, a major challenge. Efforts aimed at determining metabolite-protein interactions have shown promise; for example, a method for systematic large-scale investigation of *in vivo* protein-metabolite interactions in yeast has been developed³³, leading to the discovery of several new metabolite-protein regulatory interactions. This technology could also be applied to discover metabolic interactions (such as cofactor-enzyme interactions) and the energy sources used in novel biochemical reactions. Another approach, untargeted metabolomics, has been used to assign functions to new genes in a high-throughput manner³⁴.

The reconstruction process itself presents an exciting opportunity to increase the biological knowledge of a target organism. That is, because GENREs represent biochemically and genetically structured knowledge bases, they can be interrogated using constraint-based reconstruction and analysis methods such as BNICE³⁶, GapFill³⁷ and GrowMatch³⁸, and the iterative process of reconstruction can drive the generation of new hypotheses and subsequently new biological knowledge³² (Fig. 4). An example of using a GENRE to interpret experimental results and discover novel biochemistry is a study that combined GENREs with systematic multiple-

gene knockout strains to discover new reactions carried out by phosphofructokinase and aldolase, two enzymes extensively studied in *E. coli* glycolysis³⁵. Experimental results were compared with computational predictions, and disparities suggested missing reactions in the reconstruction. The putative reactions were then confirmed using metabolome analyses and *in vitro* enzymatic assays. This study shows that there is still more to learn, even for extensively studied areas of metabolic biochemistry.

Building high-quality reconstructions with community participation and buy-in. Accurate

and complete GENRE development is a multidisciplinary activity; it requires the participation of experts from diverse disciplines. An ideal team would include researchers who have strong biological knowledge of the target organism and access to legacy data. As an initial attempt, ‘jamboree’ efforts—gatherings of multidisciplinary researchers that seek to create a high-quality metabolic reconstruction for a target organism through intensive collaboration³⁹—have led to reconstructions for three target organisms (namely, *S. cerevisiae*⁴⁰, *Salmonella typhimurium* LT2 (ref. 41) and *Homo sapiens*⁴²). However, more structured efforts to

engage expert researchers with the metabolic network reconstruction field are needed to curate existing content and expand the scope of GENREs. One possible mechanism is crowdsourcing, in which many individuals can contribute to a reconstruction so that it contains as much legacy data as possible.

To form crowdsourced teams, the reconstruction community must reach out to domain experts, many of whom are currently unfamiliar with the metabolic reconstruction process. Recently, a multidisciplinary team of researchers (including experts in pharmaceutical chemistry, genomic biology, biochemistry, bioengineer-

Metabolic reconstructions for 78 species across the tree of life

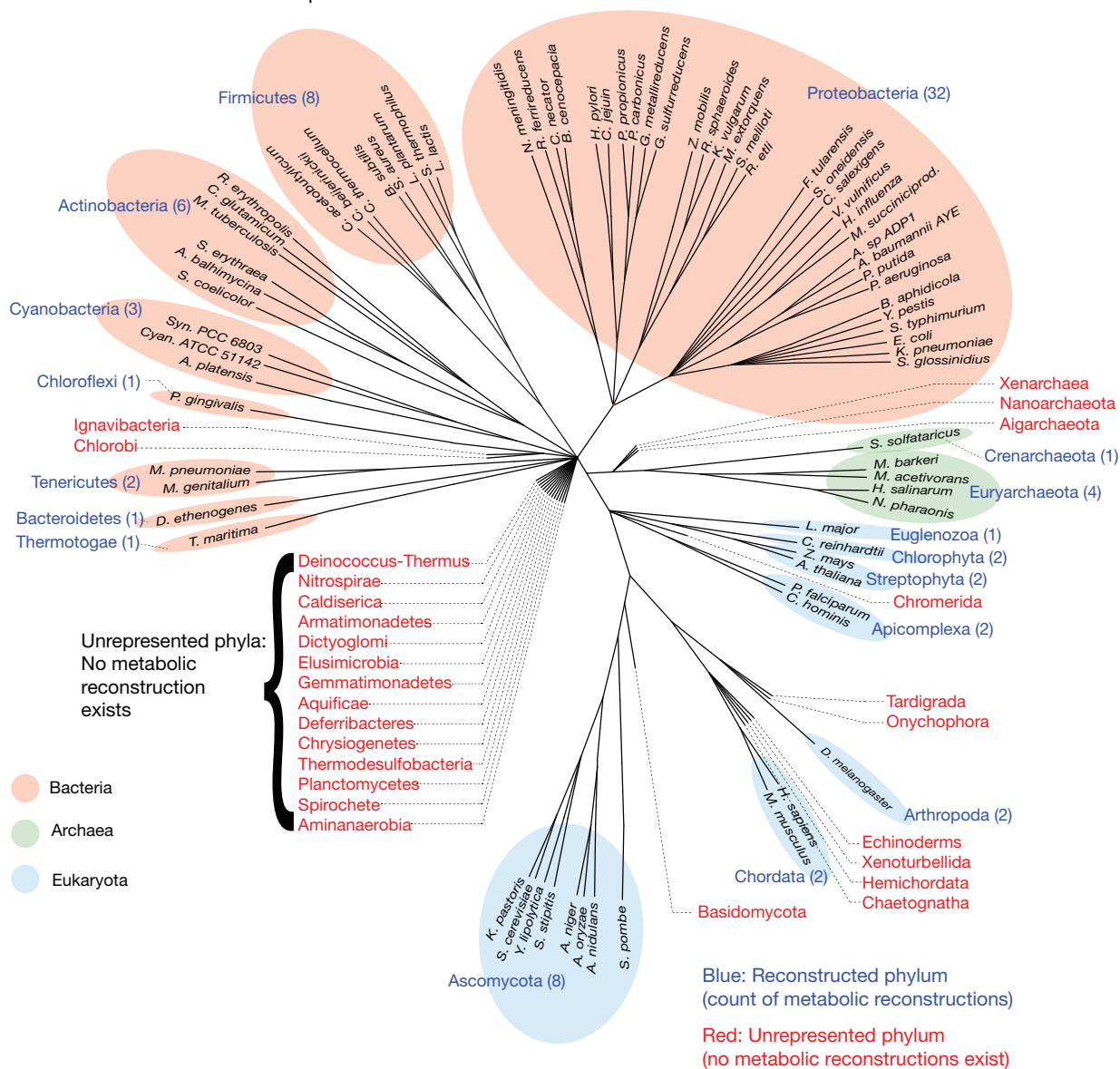


Figure 3 Phylogenetic coverage of GENREs. Distribution of GENREs across the phylogenetic tree of life for 78 species with existing GENREs (as of February 2013). The Bacteria domain has the most organisms with reconstructed GENREs. Within Bacteria the Proteobacteria phylum has the most organisms (32) with reconstructed GENREs. There are many phyla for which no GENREs have been reconstructed (red). See the SBRG website (<http://sbrg.ucsd.edu/optimizing-genres/>) for an up-to-date representation of reconstructed species and their location in the tree of life.

© 2014 Nature America, Inc. All rights reserved. nsg

Box 1 Assessing GENRE quality and completeness

Currently, there are no agreed-on metrics for the quality and completeness of GENREs. It is likely that this absence of well-defined and broadly accepted quality metrics encourages 'quantity over quality' in metabolic reconstructions. Therefore, an open discussion to define the minimal quality criteria for any genome-scale reconstruction is required. These criteria should be based on the scope, level of accuracy and application of the target GENRE. Because a GENRE is a highly curated, organism-specific knowledge base, we propose to define its quality and grade of completeness on the basis of three aspects: metabolic coverage, extent of manual curation and use of established standard operating procedures (SOPs) and standards. Individual parameters can be defined and quantified within these three general quality criteria (**Supplementary Table 3**).

Metabolic coverage. An intuitive way to estimate the completeness of a GENRE is to compute its overall metabolic space. This assessment can be done directly, by measuring the proportion of the genome covered (in terms of the proportion of protein-encoding genes included), or indirectly, by computing its phenotype coverage (for instance, in terms of the number of nutrient sources supporting growth). Both parameters show the degree of completeness of a reconstruction. For example, successive models of *E. coli* have substantially increased the number of genes included and the number of growth-supporting nutrient sources (**Supplementary Table 3**). However, these parameters have limitations, because another important (and often overlooked) factor is the portion of the represented metabolic space that is completely computable. This can be estimated by computing the number of blocked reactions (those reactions that cannot be used owing to missing connections in the network)³². The number of blocked reactions varies among reconstructions—for example, from 12% in the latest *E. coli* model (iJO1366)²³ to > 45% in the model of *C. beijerinckii* (iCM925)⁵³. Some blocked reactions probably reflect gaps in biological knowledge. In other cases, they might indicate inclusion of reactions from a different organism owing to incorrect genome annotation and/or lack of manual curation. Because these nonspecific reactions are likely to act on nonnative metabolites and/or have different cofactor usage, they tend to be unconnected to the network.

Extent of manual curation. The degree and quality of manual curation for any reconstruction is difficult to quantify objectively. However, parameters such as the number of publications contributing to the reconstruction and the confidence score¹ per reaction are useful. The confidence score is a quantitative parameter that indicates the degree of biological knowledge supporting the inclusion of a given gene-protein-reaction association in the model:

a value of 1 indicates *in silico* modeling evidence only; 2 indicates physiological and/or genome annotation evidence; 3 indicates direct or indirect genetic evidence such as knockout characterization; 4 indicates detailed knowledge of gene-product function and biochemical characterization. Therefore, the number of publications links the reconstruction with legacy data, and the confidence score is an objective parameter of the degree of support for the inclusion of each reaction.

The degree of manual curation and completeness can also be inferred from the level of detail of the biomass objective function (BOF). The BOF can be formulated at three levels³¹, which strongly depend on the detail of biosynthetic pathway modeling: the basic level uses precursors and/or simple metabolites contributing to the major macromolecular content of the cell that do not need detailed modeling of biosynthetic pathways; the intermediate level includes the biosynthetic energy needed to synthesize the macromolecules from their building blocks; the advanced level takes into account macromolecular structures in addition to DNA, RNA, fatty acids and proteins, such as strain-specific cell-envelope structures, cofactors, vitamins and so forth. Thus, BOF and biosynthetic modeling depend on each other, and an advanced BOF needs thorough biosynthetic modeling to provide the biomass components. This inevitably requires exhaustive manual curation. **Supplementary Table 3** shows the increasing level of detail in biosynthetic pathways and BOFs in *E. coli* models. However, most of the 53 GENREs we analyzed have an intermediate level of detail for both BOF and biosynthetic pathways. This level of detail may be indicative of a lack of manual curation and/or biological knowledge, and it may partly explain the incomplete metabolic coverage found in most of the reconstructions analyzed.

The ratio between reactions and genes included in any reconstruction is indicative of the level of metabolic coverage. GENREs with a high level of modeling detail have ratios >1, and GENREs with low level of detail have ratios <1 (**Supplementary Table 3**). This difference arises because GENREs with basic or intermediate levels of detail often include many reactions in which several gene products and their enzymatic transformations are 'lumped' into a single reaction.

Use of well-established SOPs and standards. The use of standards is not affected by a lack of biological knowledge or other limitations; thus, it should be mandatory. In addition, the inclusion of a standard representation of metabolites and reactions (such as Kyoto Encyclopedia of Genes and Genomes code (KEGG ID), International Union of Pure and Applied Chemistry (IUPAC) International Chemical Identifier (InChI) and so forth) would allow automatic mapping among different GENREs.

ing, chemistry and microbiology) used protein structure and genome context to functionally annotate enzymes in *Pelagibaca bermudensis*⁴³. And we were involved in a study⁴⁴ that integrated protein structure information into GENREs, thus allowing GENREs to be used in similar analyses in the future. The success of the work on *P. bermudensis* should encourage similar efforts for metabolic reconstruction. Multidisciplinary teams might be motivated by the growing appreciation of the power of

GENREs and the likelihood that a more comprehensive reconstruction will bring greater prestige.

Increasing coverage of the phylogenetic tree. If we wish to understand and study the full diversity of metabolic capabilities on Earth, reconstruction efforts must be undertaken for diverse organisms spread throughout the tree of life, in a manner analogous to the Genomic Encyclopedia for Bacteria and Archaea (GEBA)

project to sequence genomes of diverse organisms⁴⁵. The organisms on phylogenetic branches for which there are biochemical and legacy data but not reconstructions should be targeted first. Once high-quality reconstructions are completed for such target organisms, the content can be mapped to closely related species, as in the strategy used to generate reconstructions for *Klebsiella*⁴⁶, *Yersinia*⁴⁷ and *Salmonella*⁴¹ on the basis of the *E. coli* reconstruction¹⁰.

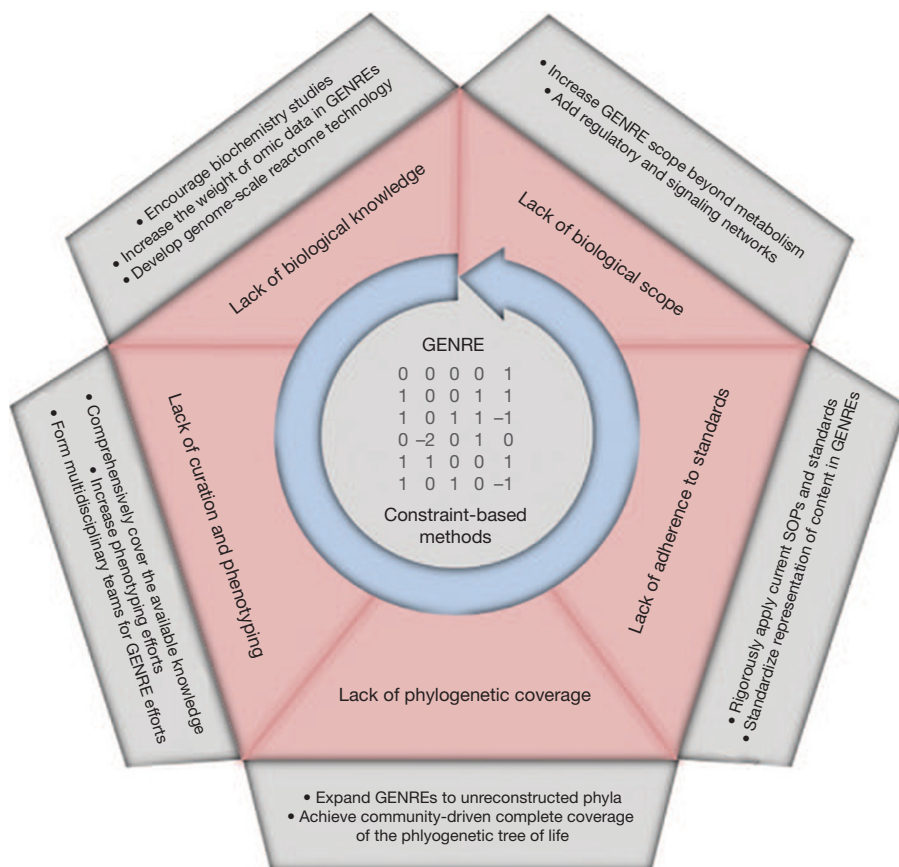


Figure 4 Factors limiting GENRE completeness and strategies for improvement. Several limitations (red) are now hindering the further development of the reconstruction field. Each shortcoming can potentially be mitigated by specific actions (gray), but a broad and integrated strategy is needed. The reconstruction and analysis process also provides an opportunity to work toward a complete GENRE of a given organism in an iterative manner (circular arrow). Several constraint-based methods⁴ can be applied to optimize GENRE (grey stoichiometric matrix) refinement.

Outlook

GENREs have already enabled a wide range of basic and applied biological studies^{2–6}, but the field remains immature. Most current metabolic reconstructions cannot be considered ‘genome scale’; instead, they are models of primary metabolism that may be unsuitable for deeper systems-biological studies of the target organism. We need to undertake a concerted effort to improve metabolic coverage of well-studied organisms and to capture known metabolic capabilities in various branches of the phylogenetic tree. Particular effort is also required to improve biochemical knowledge. Furthermore, as we suggested more than 10 years ago⁴⁸, GENREs could be expanded to include other cellular processes such as transcription and translation^{49,50}, transcriptional regulation⁵¹ and metabolic maintenance⁵². More comprehensive inclusion of such processes and their integration with metabolism would allow a quantitative assessment of the relationships among these processes. However, high-quality metabolic reconstructions must be in place before such extensions are considered.

GENREs are fundamental for discerning quantitative genotype-phenotype relationships; thus, the more comprehensive a GENRE is, the more phenotypic functions can be computed. Increased scope and quality of computed phenotypic functions and their experimental validation would in turn increase our understanding of target organisms and the complexities of the genotype-phenotype relationships.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nbt.2870).

ACKNOWLEDGMENTS

We would like to thank B. Heavner, A. Bordbar, A. Feist and J. Lerman for helpful discussions and critical review of the manuscript. J.M. is funded by US National Institutes of Health grant R01 GM057089. J.N. was funded by the Spanish Ministry of Education through the National Plan for Scientific Research, Development and Technological Innovation (2008–2011).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

- Thiele, I. & Palsson, B.O. *Nat. Protoc.* **5**, 93–121 (2010).

- Österlund, T., Nookaew, I. & Nielsen, J. *Biotechnol. Adv.* **30**, 979–988 (2012).
- Kim, T.Y., Sohn, S.B., Kim, Y.B., Kim, W.J. & Lee, S.Y. *Curr. Opin. Biotechnol.* **23**, 617–623 (2012).
- Lewis, N.E., Nagarajan, H. & Palsson, B.O. *Nat. Rev. Microbiol.* **10**, 291–305 (2012).
- McCloskey, D., Palsson, B.O. & Feist, A.M. *Mol. Syst. Biol.* **9**, 661 (2013).
- Bordbar, A., Monk, J.M., King, Z.A. & Palsson, B.O. *Nat. Rev. Genet.* **15**, 107–120 (2014).
- Edwards, J.S. & Palsson, B.O. *J. Biol. Chem.* **274**, 17410–17416 (1999).
- Fleischmann, R.D. *et al. Science* **269**, 496–512 (1995).
- Reed, J.L., Vo, T.D., Schilling, C.H. & Palsson, B.O. *Genome Biol.* **4**, R54 (2003).
- Feist, A.M. *et al. Mol. Syst. Biol.* **3**, 121 (2007).
- Förster, J., Famili, I., Fu, P., Palsson, B.O. & Nielsen, J. *Genome Res.* **13**, 244–253 (2003).
- Duarte, N.C. *et al. Proc. Natl. Acad. Sci. USA* **104**, 1777–1782 (2007).
- Thiele, I. *et al. Nat. Biotechnol.* **31**, 419–425 (2013).
- Nogales, J., Gudmundsson, S., Knight, E.M., Palsson, B.O. & Thiele, I. *Proc. Natl. Acad. Sci. USA* **109**, 2678–2683 (2012).
- Chang, R.L. *et al. Mol. Syst. Biol.* **7**, 518 (2011).
- Henry, C.S. *et al. Nat. Biotechnol.* **28**, 977–982 (2010).
- Vitkin, E. & Shlomi, T. *Genome Biol.* **13**, R111 (2012).
- Agren, R. *et al. PLOS Comput. Biol.* **9**, e1002980 (2013).
- Rabinowitz, J.D. & Vastag, L. *Nat. Chem. Biol.* **8**, 497–501 (2012).
- Schomburg, I. *et al. Nucleic Acids Res.* **41**, D764–D772 (2013).
- Schnoes, A.M., Brown, S.D., Dodevski, I. & Babbitt, P.C. *PLOS Comput. Biol.* **5**, e1000605 (2009).
- Edwards, J.S. & Palsson, B.O. *Proc. Natl. Acad. Sci. USA* **97**, 5528–5533 (2000).
- Orth, J.D. *et al. Mol. Syst. Biol.* **7**, 535 (2011).
- Heavner, B.D., Smallbone, K., Barker, B., Mendes, P. & Walker, L.P. *BMC Syst. Biol.* **6**, 55 (2012).
- Riley, M. *et al. Nucleic Acids Res.* **34**, 1–9 (2006).
- Hu, P. *et al. PLoS Biol.* **7**, e1000096 (2009).
- Janssen, P., Goldovsky, L., Kunin, V., Darzentas, N. & Ouzounis, C.A. *EMBO Rep.* **6**, 397–399 (2005).
- Lespinet, O. & Labeledan, B. *Science* **307**, 42 (2005).
- Chen, L. & Vitkup, D. *Trends Biotechnol.* **25**, 343–348 (2007).
- Pouliot, Y. & Karp, P. *BMC Bioinformatics* **8**, 244 (2007).
- Feist, A.M. & Palsson, B.O. *Curr. Opin. Microbiol.* **13**, 344–349 (2010).
- Orth, J.D. & Palsson, B. *BMC Syst. Biol.* **6**, 30 (2012).
- Li, X., Gianoulis, T.A., Yip, K.Y., Gerstein, M. & Snyder, M. *Cell* **143**, 639–650 (2010).
- Baran, R. *et al. ACS Chem. Biol.* **8**, 189–199 (2013).
- Nakahigashi, K. *et al. Mol. Syst. Biol.* **5**, 306 (2009).
- Hatzimanikatis, V. *et al. Bioinformatics* **21**, 1603–1609 (2005).
- Satish Kumar, V., Dasika, M.S. & Maranas, C.D. *BMC Bioinformatics* **8**, 212 (2007).
- Kumar, V.S. & Maranas, C.D. *PLOS Comput. Biol.* **5**, e1000308 (2009).
- Thiele, I. & Palsson, B.O. *Mol. Syst. Biol.* **6**, 361 (2010).
- Herrgård, M.J. *et al. Nat. Biotechnol.* **26**, 1155–1160 (2008).
- Thiele, I. *et al. BMC Syst. Biol.* **5**, 8 (2011).
- Thiele, I. *et al. Nat. Biotechnol.* **31**, 419–425 (2013).
- Zhao, S. *et al. Nature* **502**, 698–702 (2013).
- Chang, R.L. *et al. Science* **340**, 1220–1223 (2013).
- Wu, D. *et al. Nature* **462**, 1056–1060 (2009).
- Liao, Y.-C. *et al. J. Bacteriol.* **193**, 1710–1717 (2011).
- Charusanti, P. *et al. BMC Syst. Biol.* **5**, 163 (2011).
- Reed, J.L. & Palsson, B.O. *J. Bacteriol.* **185**, 2692–2699 (2003).
- Thiele, I., Jamshidi, N., Fleming, R. & Palsson, B. *PLOS Comput. Biol.* **5**, e1000312 (2009).
- Lerman, J.A. *et al. Nat. Commun.* **3**, 929 (2012).
- Covert, M.W., Knight, E.M., Reed, J.L., Herrgård, M.J. & Palsson, B.O. *Nature* **429**, 92–96 (2004).
- Linster, C.L., Van Schaftingen, E. & Hanson, A.D. *Nat. Chem. Biol.* **9**, 72–80 (2013).
- Milne, C.B. *et al. BMC Syst. Biol.* **5**, 130 (2011).
- Ganter, M., Bernard, T., Moretti, S., Stelling, J. & Pagni, M. *Bioinformatics* **29**, 815–816 (2013).