University of Groningen

New surveyor tools for charting microbial metabolic maps

Breitling, Rainer; Vitkup, Dennis; Barrett, Michael P.

Published in:
Nature Reviews Microbiology

DOI:
10.1038/nrmicro1797

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2008

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Breitling, R., Vitkup, D., & Barrett, M. P. (2008). New surveyor tools for charting microbial metabolic maps. *Nature Reviews Microbiology*, *6*(2), 156-161. https://doi.org/10.1038/nrmicro1797

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 25-08-2022

INNOVATION

New surveyor tools for charting microbial metabolic maps

Rainer Breitling, Dennis Vitkup and Michael P. Barrett

Abstract | The computational reconstruction and analysis of cellular models of microbial metabolism is one of the great success stories of systems biology. The extent and quality of metabolic network reconstructions is, however, limited by the current state of biochemical knowledge. Can experimental high-throughput data be used to improve and expand network reconstructions to include unexplored areas of metabolism? Recent advances in experimental technology and analytical methods bring this aim an important step closer to realization. Data integration will play a particularly important part in exploiting the new experimental opportunities.

More than 300 complete microbial genomes have been sequenced since the publication of the *Haemophilus influenzae* genome 12 years ago[1] and thousands of genomes are currently in the sequencing pipeline. The availability of these genomic 'parts lists' has spurred intense interest in how the interaction of these parts determines cellular physiology. This systems-biology perspective seeks to describe the emergent properties (see Glossary) of a complex system — that is, functions of an organism that cannot be understood (or even defined) from isolated components, but emerge if all parts are integrated. Global stoichiometric models of microbial metabolism have been successful in predicting such emergent properties, in particular for well-studied organisms[2] (BOX 1).

The main ingredient of a stoichiometric model is a comprehensive description of the biochemical connectivity between cellular metabolites (a stoichiometric matrix), which is similar to classical biochemical-pathway maps (BOX 1). Currently, the information that is required for such models is derived primarily from automated genome annotations and computational predictions, which, for well-studied organisms, are refined by careful manual curation based on published literature and comparative genomics[3,4]. Here we discuss new analytical strategies that enable the experimental completion and expansion of these cellular metabolic models (FIG. 1). Such expansion will be crucial as we move towards those areas of metabolism that distinguish organisms rather than the conserved core that unites them. We highlight a number of technological advances that could allow the *de novo* reconstruction of large unexplored parts of the metabolic map directly from experimental observations. In particular, we discuss how ultra-high-resolution mass spectrometry allows the identification of metabolites and their chemical relationships, how correlation analysis and genetical genomics can reveal links within and between metabolic pathways and, finally, how flux and mutant-phenotype measurements can be used to verify and correct the reconstructed biochemical maps.

Gaps in current metabolic-pathway maps

Metabolite networks have been reconstructed automatically for a large number of microorganisms[5–7]. The network reconstructions are primarily accomplished using computational functional transfer from experimental results that have been obtained from a small number of model organisms. The presence of a particular reaction is inferred if an orthologue of the corresponding enzyme is encoded in the genome. In cases in which genome divergence or the presence of non-orthologous enzymes interferes with genome annotation, the resulting networks will be incomplete. For example, in recently reconstructed models of the *Geobacter sulphurreducens* and *Methanosarcina barkeri* metabolomes[8,9], many hypothetical reactions (comprising approximately 20% of the total number of reactions) had to be added to the initial metabolic models to enable fulfilment of all the basic biochemical requirements. In *G. sulphurreducens*, established pathways for the synthesis of lysine, serine, alanine and threonine were all missing in the genome annotation, even though the bacterium can synthesize these amino acids[9]. Various bioinformatic approaches have been proposed to fill in the gaps in those cases in which it is known that a particular reaction or pathway does exist in a given metabolic system[10–12]. These gap-filling methods take into account genomic correlations (such as mRNA co-expression, chromosomal clustering across genomes and protein fusions) between known and missing parts of metabolic networks. Metabolic gap closing is aggravated by the problem of globally orphan metabolic activities. Indeed, for 30–40% of the known metabolic activities that are classified by the Enzyme Commission (see Further information) there are no known sequences in any organism — these activities remain globally orphan[12–14].

Importantly, currently unknown metabolic reactions cannot be discovered using automatic reconstruction methods. However, the complete structure of biochemical pathways is well-established only for the central metabolism of model organisms — indeed, untargeted large-scale metabolite screens yield an abundance of unknown compounds[15]. Even in *Escherichia coli*, the quintessential model organism, a novel pathway of glucose metabolism was discovered in 2003 (REF. 16) and a previously undescribed pathway for pyrimidine catabolism in 2006 (REF. 17). Those parts of the metabolic map that are not required to fulfil immediate growth demands tend to fall outside the scope of available tools for network reconstruction and analysis. They generally contribute to performance only under specific environmental conditions, such as physical or antibiotic stress, and their

© 2008 Nature Publishing Group

absence rarely influences viability under standard laboratory conditions. Nonetheless, the knowledge of these condition- and organism-specific activities is essential for a complete understanding of the metabolic capacities of microbial species.

Mass spectrometric metabolome mapping

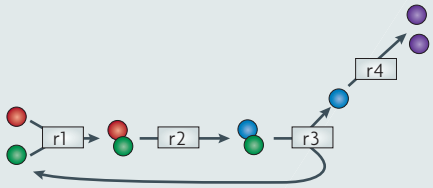
How then can we discover new metabolites and biochemical pathways? Of fundamental importance for large-scale metabolic-network reconstruction is a complete inventory of cellular metabolites. Although complete genomes, transcriptomes and, to a lesser extent, proteomes have been described for numerous cell types, measurements of cellular metabolism have lagged behind owing to the variability and chemical complexity of low-molecular-weight compounds.

The ability to collect biochemically relevant information about metabolites using high-throughput techniques has only recently become available[18]. A number of reviews that summarize the various metabolomic platforms have been published[19–21] and a full discussion is not warranted here. Mass spectrometry has long been the favourite method for comprehensive metabolome screening; it has the largest scope for metabolite analysis, in terms of the universality, sensitivity and specificity of the measurements[19,22]. Several mass-spectrometric platforms can measure large numbers of metabolites simultaneously in any given sample[18,23–26]. Ultra-high-accuracy mass analysers[27] achieve molecular-mass resolutions of up to 1,000,000 and mass accuracies that are better than 1 part per million. The first study to use Fourier transform ion cyclotron resonance mass spectrometry in metabolomics showed that it is possible to separate more than 5,000 unique masses in ripening strawberry extracts and assign single empirical formulae to approximately half, based solely on molecular mass[27]. A comparable performance is now available using the magnet-free (and therefore more economical) Orbitrap mass analyser, which is equally suitable for metabolomics experiments[23,24]. Further advances in sample preparation that maximize metabolite yield, and minimize losses that are due to ion suppression and related phenomena, will further enhance the usefulness of this exceptionally promising approach.

In a metabolomic study of the protozoan pathogen *Trypanosoma brucei*, a high-accuracy mass-spectrometry approach was used to obtain not only accurate mass measurements of metabolites, but also infer accurate mass differences between related metabolites[28]. As specific mass differences imply corresponding biochemical transformations, the accurate difference information was used to predict the potential connectivity between all measured mass peaks (FIG. 2). In this way, an entire hypothetical metabolic map was reconstructed *de novo*. The follow-up statistical analysis showed that the structure of the network was far from random and contained more meaningful biochemical links than would have been expected by chance[28].

Ultra-high-accuracy mass spectrometers also allow the prediction of hypothetical structural formulae for unknown compounds by combining their exact molecular masses with knowledge about their possible relationship to other masses in the sample. For example, although several formulae could all have highly similar molecular masses, the chemical relationships to other compounds in the mixture will indicate the most likely true formula. The fact that most metabolites contain simple building

Box 1 | **The power of stoichiometric models of metabolism**



$$\begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{cccc} r1 & r2 & r3 & r4 \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 2 \end{pmatrix} \end{array} = S$$

A stoichiometric matrix, such as a metabolic-pathway map, is a method of providing a complete description of all the possible cellular biochemistries. In a metabolic map, metabolites are connected to the reactions (enzymes) that interconvert them (see figure for a theoretical metabolic map). The stoichiometric matrix (S) contains the same information in a mathematical form: for example, in the figure, the entries show that in reaction (r) 1, one molecule each of red and green combine (are consumed) to form one molecule of red–green product. The network can be reconstructed from S , but the opposite is not necessarily the case. For example, in r4, two molecules of the purple compound are produced for every blue molecule that is consumed. Such information is not always included in the biochemical-pathway map. In steady state, the condition Sv=0 needs to be satisfied, in which v is the vector that describes the fluxes through each reaction of the system. This limits the possible combinations of metabolic fluxes that can maintain a steady state. Additional constraints can specify that some reactions are irreversible or have certain maximal rates, which further restricts the range of allowed metabolic behaviours.

For a genome-scale metabolic network, the pathway map can contain thousands of metabolites and reactions, so leading to a huge matrix that describes a diverse number of possible metabolic phenotypes, and can be used to discover global emergent properties that determine cellular physiology. A computational analysis of the stoichiometric matrix using flux-balance analysis can reveal which genes are essential in different environmental conditions[2]. This is an important emergent property, as the essentiality of a gene depends on its network context, such as the availability of alternative pathways or other back-up mechanisms. The accuracy of lethality predictions using flux-balance analysis can be as high as 70–80% for well-studied organisms[2].

A flux-balance analysis of the stoichiometric matrix can also be used to predict evolutionary responses to gene deletion[62,63] and to analyse evolutionary trajectories, if augmented with a few basic assumptions, such as 'microbial cells are optimized to achieve maximal growth rate'. For example, by interpreting experimentally observed fluxes based on stoichiometric constraints, Fong and colleagues[64] observed that rapid evolutionary adaptation in *Escherichia coli* is initially achieved by activating latent alternative pathways and, subsequently, by increasing pathway capacity, but rarely by evolving new functionality.

Growth yield and flux through specific pathways are other emergent properties that are predicted successfully by analysis of the metabolic network, without requiring quantitative information on kinetics or regulation[2]. Flux predictions are also surprisingly accurate for mutant cells[62,65], and this forms the basis for new strategies in bioengineering. For example, the Optknock approach[66] identifies sets of mutations (gene knock outs) that change the metabolic system in such a way that high growth rates (the objective of the cell) are biochemically coupled to high fluxes through particular reactions, such as secretion of a specific metabolite (the objective of the engineer). This approach has been used to engineer bacterial strains that overproduce lactic acid for industrial use[67]. Although the results of the Optknock strategy agree with manually derived designs, the advantage of the approach will be realized in the overproduction of new targets.

© 2008 Nature Publishing Group

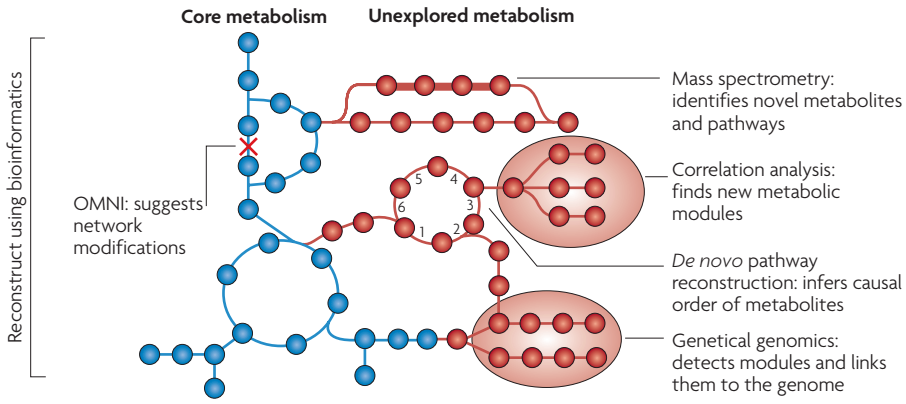


Figure 1 | **Methods for correcting and expanding the metabolic map.** Many methods have been developed recently for expanding our knowledge of microbial metabolism on a large scale. Bioinformatics can be used to create a reconstructed core model using comparative genomics information. Correlation analysis and its modern derivative, which is based on genetical genomics, can be used to identify novel metabolic modules and link them to the responsible enzymes. It can also order metabolites along reaction pathways. Ultra-high-accuracy mass spectrometry can be used to provide the metabolomics data that is required by each of these methods and, additionally, to infer novel pathways directly from the observed exact metabolite-mass differences. In combination with flux measurements, optimal metabolic network identification (OMNI) can then be used to identify incorrect or inactive reactions, both in the bioinformatically predicted core and among the newly discovered areas of metabolism.

blocks facilitates this analysis; Nobeli and colleagues[29] have shown that as few as 57 common structural groups (ranging from porphyrin and nicotinamide units to phenyl and phosphate groups) account for as much as 90% of all atoms in the known metabolome of *E. coli*.

Pathway reconstruction from correlation

Although ultra-high-accuracy mass spectrometry can predict potential connectivity between masses, this method on its own is inadequate to prove that metabolites are biochemically connected. Metabolic correlation analysis is probably the most direct technique that can infer biochemical connectivity from high-throughput measurements of metabolite concentrations. It is expected that compounds in which concentrations are correlated in multiple samples are linked by an enzymatic reaction. The feasibility of *de novo* inference of entire reaction pathways based on metabolite correlations was first demonstrated experimentally by Arkin and colleagues[30]. In this pioneering study, the aim was to reconstruct a segment of glycolysis (consisting of 8 purified enzymes and 14 metabolites) from actual time-series measurements obtained using a continuous-flow stirred-tank reactor. Correlated responses to perturbations linked the metabolites in a manner that recreated much of the chemical and regulatory connectivity that is already known for glycolysis. Refinements of this approach, using more extensive system perturbations, have so far been restricted to small systems under optimal experimental conditions[31,32].

The application of correlation analysis to entire metabolite networks presents far greater challenges. This approach was demonstrated using *Arabidopsis thaliana* in one of the first large screens of a cellular metabolome[33]. In this analysis, strong correlations were indeed observed between some metabolites in which biochemical connections were already well established (such as glucose 6-phosphate and fructose 6-phosphate). However, numerous non-adjacent metabolites also showed strong correlations, whereas the majority of metabolic neighbours did not correlate at all.

The basis for these observations has been examined by several groups[34–37]. Metabolic networks are not merely causal networks; they are also biotransformation networks in which compounds are physically transformed into each other. As a consequence, reactions that are not directly connected can still have a profound influence on one another. Indirect effects and non-observed concentrations, which are common in metabolic networks, as well as various kinds of time delays that are imposed by intermediary reactions and compartmentalization, also influence the observed correlations[38]. In terms of classical metabolic-control analysis, the correlation patterns of metabolite concentrations reflect the enzyme control coefficients that are involved in their turnover[34]. In many cases, these control coefficients do not reflect direct biochemical connectivity in an intuitive way.

In spite of the previously discussed caveats, these initial studies demonstrate the feasibility of using correlation patterns of a set of metabolites to deduce the connectivity between them. Clearly, although the correlations can be used as fingerprints for network structures, accurate network reconstruction — particularly at the global cellular level — will require integration with additional lines of evidence.

Genetical genomics of metabolism

One recently introduced methodology that can enhance correlation-based network reconstruction is based on the genetical

Glossary

De novo pathway reconstruction
The inference of metabolic pathways directly from experimental measurements, without any prior information.

Emergent property
A property that emerges only in the context of an integrated system, not in its components; also called systems property.

Flux-balance analysis
A computational method that is used to obtain feasible flux distributions in metabolic networks. Linear constraints on nutrient uptake, reaction irreversibility and steady-state conservation of metabolite concentrations are applied using a stoichiometric model. The fluxes that are optimal for a given objective function (for example, biomass production or ATP synthesis) are then obtained using linear optimization.

Genetical genomics
The combination of high-throughput measurements of gene expression, protein levels or metabolite concentrations with classical genetic strategies.

Metabolomics
The analysis of the concentration and dynamics of small cellular molecules (the metabolome).

Optimal metabolic network identification
(OMNI). A computational method for correcting stoichiometric models based on a small number of pathway flux measurements.

Stable-isotope flux analysis
An analysis that traces the metabolic fate of non-radioactive atoms from labelled precursors to biomass components. The steady-state labelling pattern can be used to infer the activity of metabolic pathways (fluxes) with the help of stoichiometric models.

Stoichiometric model
A detailed description of metabolism without information on the kinetic or thermodynamic parameters. The model specifies how many molecules of each substrate are used and how many product molecules are generated (the reaction stoichiometry) for every reaction.

© 2008 Nature Publishing Group

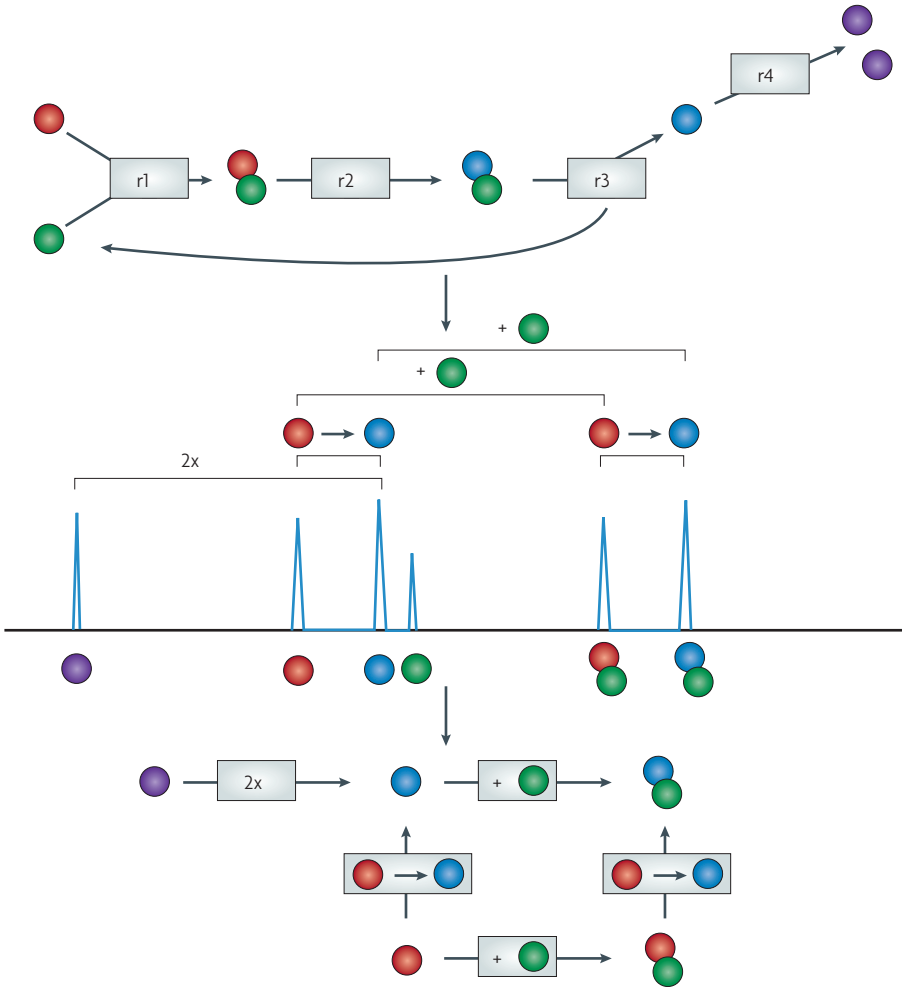


Figure 2 | **Network reconstruction using ultra-high-accuracy mass spectrometry.** If metabolite masses in a cellular sample are measured comprehensively and with high accuracy, the difference between masses can be used to identify compounds that are related by common metabolic transformations. This information can be used to infer a putative metabolic network. The reconstruction will contain spurious edges, for example, the red and blue metabolites are not directly connected by an enzymatic transformation in the real network (top). Such reactions can be pruned from the reconstruction using measured C^{13} stable-isotope-flux measurements in combination with optimal metabolic network identification analysis, or by correlation measurements. Genetical genomics experiments can lead to further refinement and help to link the inferred reactions to the catalysing enzymes. r, reaction.

genomics approach[39–41]. In genetical genomics, segregating populations (for example, recombinant inbred lines) are created by crossing two divergent parental lines (FIG. 3). As each offspring line is a 50:50 ratio mosaic of the parental genotypes, genetical genomics is a multiple perturbation approach. In several eukaryotic species, including yeast, the approach has been used successfully in transcriptomic studies[42]. For prokaryotic microorganisms, genetical genomics experiments can also be performed by exploiting the natural variation that is present in a bacterial population rather than using controlled genetic crosses. For human gene expression, such a population-based genetical-genomics approach has already been tested[43].

The approach can also be used in metabolomics. In a pilot study, Keurentjes and colleagues[40] crossed two *A. thaliana* accessions that had different metabolite profiles and then screened and compared the metabolomes of the resulting offspring lines. Large groups of metabolites were linked to shared genetic loci. In several cases in which the identity of the metabolites was determined, these shared loci were also shown to harbour key enzymes from the corresponding pathway. In a further step, the authors recovered a fraction of the biochemical connectivity within the glucosinolate biosynthesis pathway based purely on the population-wide correlation patterns.

The study by Keurentjes and colleagues[40] demonstrates that genetical genomics can be used to assign observed metabolites to local functional modules. Using larger population sample sizes will permit more in-depth analyses of the connectivity between metabolites based on genetical genomics and allow robust separation of correlated genetic variation from intrinsic fluctuations within the system.

Flux measurements and phenotypes

Although metabolomics and genetical genomics provide an inventory of metabolites and suggest possible chemical relationships between them, dynamic information from flux analysis can be used to further refine the stoichiometric reaction models. Metabolic fluxes are measured by adding stable-isotopic tracers to microbial growth medium and observing their flow through the metabolic network[44,45]. Technologies that enable stable-isotope flux analysis at high throughput have recently become available[44].

Recent work by Herrgard and colleagues[46] suggests an experimental method for both validating and correcting metabolic models based on experimentally measured fluxes. Their approach (optimal metabolic network identification) uses a two-level optimization procedure to find modifications in existing models that both minimize the discrepancy between predicted and observed fluxes and maximize metabolic performance (for example, biomass production), which is assumed to be optimized by evolution in most microbial species.

In addition to validation by experimentally measured fluxes, cellular metabolic models are now routinely checked against available high-throughput functional-genomics data, such as growth phenotypes in various environmental conditions[47]. The flux-balance metabolic models are used to predict mutant-growth phenotypes and wrong predictions guide the iterative model improvement[48–51]. Initially, adjustments to metabolic models that were based on inconsistencies between experimental and predicted mutant phenotypes were made manually. Recently, however, several approaches have been developed to correct the models in an automatic or semi-automatic fashion. For example, Reed and colleagues[52] used an optimization-based approach to predict missing reactions and improve agreement between experiments and model predictions. Several of their predictions were verified experimentally. Kumar and colleagues[53] subsequently suggested an algorithm to identify metabolites in reconstructed networks that cannot be produced under any conditions. Using a database of

© 2008 Nature Publishing Group

known metabolic activities, these workers then identified a minimal number of additional (missing) reactions that were required to restore the connectivity of all metabolites in the network.

Several recent studies extended the constraint-based approach to metabolomics data. Notably, Kümmel and colleagues[54] introduced network-embedded thermodynamic (NET) analysis to allow the network-level interpretation of metabolomics data. This approach uses known reaction directionalities and metabolite-formation energies to calculate possible concentration ranges of metabolites. NET analysis was successfully used to resolve the concentrations of several pooled metabolites and check the internal consistency of available metabolomics datasets for *E. coli*[54].

The future of metabolic reconstruction

In this Innovation, we have outlined a number of new technologies that enable the identification of novel metabolites and their connections within microbial metabolic networks. These approaches will be particularly useful for an intensified exploration of the peripheral and less-studied areas of metabolism. A number of studies highlight the importance of peripheral metabolism for microbial diversification and physiology. Peripheral metabolites are the most evolutionarily volatile, being rapidly lost and gained during evolution[55]. A large-scale study of horizontal gene transfer in bacteria revealed that recently transferred genes are predominantly active at the periphery of the metabolic map and confer specific advantages in changing environments[56]. The so-called 'bow tie' architecture of metabolism[57] emphasizes the functional role of peripheral metabolism: a large number of substrates converge towards a restricted small-core metabolism, from which they fan out again into a multitude of synthesized cellular products. Perhaps of even greater importance is the fact that most attempts to manipulate microbial metabolism in order to produce compounds of biotechnological value are focused on these peripheral pathways[58,59].

Integrated approaches are required to fully exploit the potential of the new experimental technologies that are described above. There is no single experimental technology that will be able to reliably discover metabolic networks on a high-throughput scale. In combination, however, the described approaches offer unprecedented power to reconstruct metabolic networks. For example, a computational sequence analysis of a target genome can chart an initial metabolic map of an organism. Experiments using ultra-high-accuracy mass spectrometry can then provide precise molecular identification and *de novo* pathway reconstruction. Metabolites that correlate across multiple experimental conditions and are linked by a predicted chemical transformation can be connected into pathways with a high degree of confidence. If the predicted network is also supported by metabolic flux measurements and predicted mutant phenotypes, under multiple environmental conditions, the reliability of the metabolic map becomes almost certain. The drive towards data synthesis is not unique to metabolic network reconstructions; integrative methods are now widely used in the context of protein–protein interaction and regulatory networks[60,61]. Methods developed in that context facilitate the weighted integration of various experimental- and genomic-data sources to predict molecular interactions.

In contrast to the labour-intensive and time-consuming experimentation on individual metabolites and enzymes that has been used previously, the experimental tools and analytical methods that are presented here will enable automatic or semi-automatic reconstructions of novel metabolic pathways with unprecedented speed and accuracy. Although much of the work in this young and promising area of research has been limited to successful proof-of-principle studies, larger studies are already underway in numerous laboratories. Such experiments will help us to elucidate new areas of metabolism for all sequenced and yet-to-be-sequenced organisms.

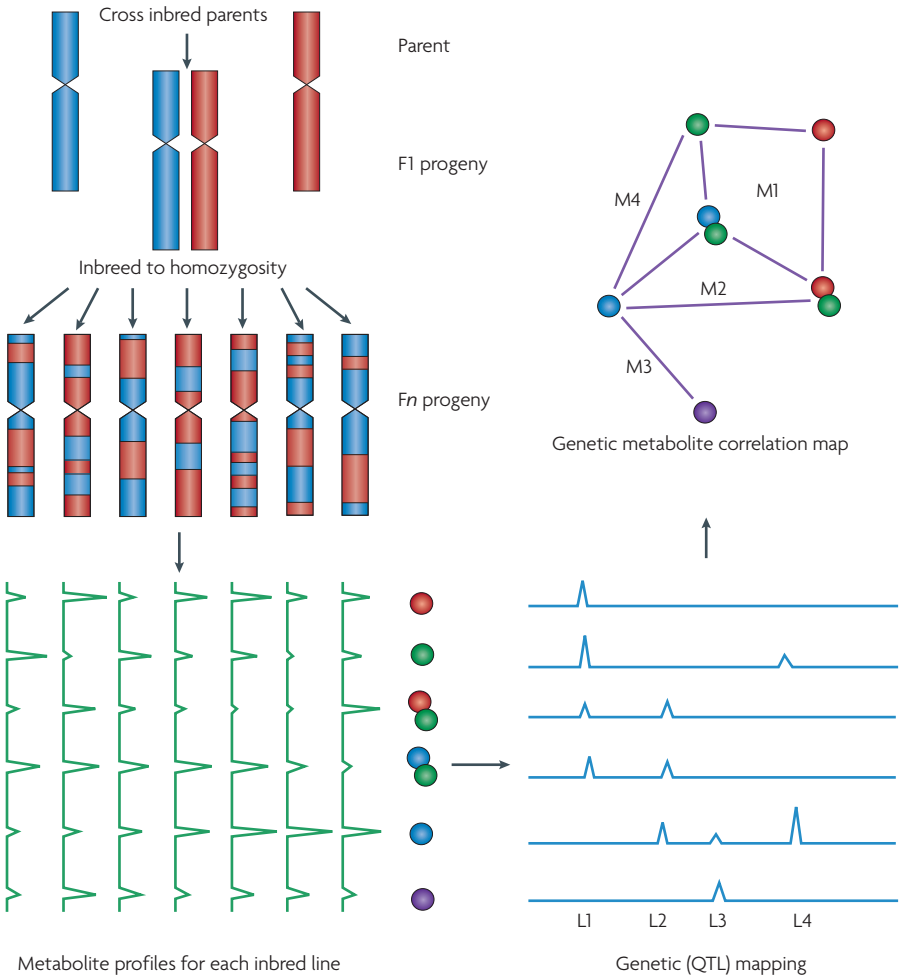


Figure 3 | **Metabolite module identification using genetical genomics.** Two inbred strains with different metabolic capacities are crossed. The heterozygous F1 generation is propagated by selfing or brother–sister mating to generate recombinant inbred lines that are homozygous mosaics of the parental strains. The metabolite profile for each strain is measured, for example, by chromatography and mass spectrometry. The genetic loci that control the metabolite concentrations are mapped using techniques that were developed for quantitative trait locus (QTL) mapping. The correlation between these genetic maps can be used to create a metabolite network, in which genes that share common control loci are connected. Each control locus (L1–4) corresponds to a putative metabolic module (M1–4).

© 2008 Nature Publishing Group

Rainer Breitling is at the Groningen Bioinformatics Centre, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9751 NN Haren, The Netherlands.

Dennis Vitkup is at the Center for Computational Biology and Bioinformatics, Department of Biomedical Informatics, Columbia University, New York, New York 10032, United States.

Michael P. Barrett is at the Division of Infection and Immunity, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom.

Correspondence to R.B. and D.V.
e-mails: r.breitling@rug.nl; vitkup@dbmi.columbia.edu

doi:10.1038/nrmicro1797
Published online 19 November 2007

1. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
2. Price, N. D., Reed, J. L. & Palsson, B. Ø. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Rev. Microbiol.* **2**, 886–897 (2004).
3. Notebaart, R. A., van Enckevort, F. H., Francke, C., Siezen, R. J. & Teusink, B. Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics* **7**, 296 (2006).
4. Reed, J. L., Famili, I., Thiele, I. & Palsson, B. O. Towards multidimensional genome annotation. *Nature Rev. Genet.* **7**, 130–141 (2006).
5. Ma, H. & Zeng, A. P. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* **19**, 270–277 (2003).
6. Kanehisa, M. *et al.* From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* **34**, D354–D357 (2006).
7. Caspi, R. *et al.* MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* **34**, D511–D516 (2006).
8. Feist, A. M., Scholten, J. C., Palsson, B. O., Brockman, F. J. & Ideker, T. Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Mol. Syst. Biol.* [online], http://www.nature.com/msb/journal/v2/n1/pdf/msb4100046.pdf (2006).
9. Mahadevan, R. *et al.* Characterization of metabolism in the Fe(III)-reducing organism *Geobacter sulfurreducens* by constraint-based modeling. *Appl. Environ. Microbiol.* **72**, 1558–1568 (2006).
10. Kharchenko, P., Chen, L., Freund, Y., Vitkup, D. & Church, G. M. Identifying metabolic enzymes with multiple types of association evidence. *BMC Bioinformatics* **7**, 177 (2006).
11. Kharchenko, P., Vitkup, D. & Church, G. M. Filling gaps in a metabolic network using expression information. *Bioinformatics* **20**, (Suppl. 1), I178–I185 (2004).
12. Green, M. L. & Karp, P. D. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC Bioinformatics* **5**, 76 (2004).
13. Chen, L. & Vitkup, D. Distribution of orphan metabolic activities. *Trends Biotechnol.* **25**, 343–348 (2007).
14. Lespinet, O. & Labedan, B. Orphan enzymes? *Science* **307**, 42 (2005).
15. Fiehn, O. & Weckwerth, W. Deciphering metabolic networks. *Eur. J. Biochem.* **270**, 579–588 (2003).
16. Fischer, E. & Sauer, U. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *J. Biol. Chem.* **278**, 46446–46451 (2003).
17. Loh, K. D. *et al.* A previously undescribed pathway for pyrimidine catabolism. *Proc. Natl Acad. Sci. USA* **103**, 5114–5119 (2006).
18. Breitling, R., Pitt, A. R. & Barrett, M. P. Precision mapping of the metabolome. *Trends Biotechnol.* **24**, 543–548 (2006).
19. Dettmer, K., Aronov, P. A. & Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **26**, 51–78 (2007).
20. Dunn, W. B., Bailey, N. J. & Johnson, H. E. Measuring the metabolome: current analytical technologies. *Analyst* **130**, 606–625 (2005).
21. Hollywood, K., Brison, D. R. & Goodacre, R. Metabolomics: current technologies and future trends. *Proteomics* **6**, 4716–4723 (2006).
22. Want, E. J., Nordstrom, A., Morita, H. & Siuzdak, G. From exogenous to endogenous: the inevitable imprint of mass spectrometry in metabolomics. *J. Proteome Res.* **6**, 459–468 (2007).
23. Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **40**, 430–443 (2005).
24. Makarov, A., Denisov, E., Lange, O. & Horning, S. Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J. Am. Soc. Mass Spectrom.* **17**, 977–982 (2006).
25. Brown, S. C., Kruppa, G. & Dasseux, J. L. Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrom. Rev.* **24**, 223–231 (2005).
26. Koulman, A. *et al.* High-throughput direct-infusion ion trap mass spectrometry: a new method for metabolomics. *Rapid Commun. Mass Spectrom.* **21**, 421–428 (2007).
27. Aharoni, A. *et al.* Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. *OMICS* **6**, 217–234 (2002).
28. Breitling, R., Ritchie, S., Goodenowe, D., Stewart, M. L. & Barrett, M. P. *Ab initio* prediction of metabolic networks using Fourier transform mass spectrometry data. *Metabolomics* **2**, 155–164 (2006).
29. Nobeli, I., Ponstingl, H., Krissinel, E. B. & Thornton, J. M. A structure-based anatomy of the *E.coli* metabolome. *J. Mol. Biol.* **334**, 697–719 (2003).
30. Arkin, A., Shen, P. & Ross, J. A test case of correlation metric construction of a reaction pathway from measurements. *Science* **277**, 1275–1279 (1997).
31. Vance, W., Arkin, A. & Ross, J. Determination of causal connectivities of species in reaction networks. *Proc. Natl Acad. Sci. USA* **99**, 5816–5821 (2002).
32. Torralba, A. S., Yu, K., Shen, P., Oefner, P. J. & Ross, J. Experimental test of a method for determining causal connectivities of species in reactions. *Proc. Natl Acad. Sci. USA* **100**, 1494–1498 (2003).
33. Fiehn, O. *et al.* Metabolite profiling for plant functional genomics. *Nature Biotechnol.* **18**, 1157–1161 (2000).
34. Camacho, D., de la Fuente, A. & Mendes, P. The origin of correlations in metabolomics data. *Metabolomics* **1**, 53–63 (2005).
35. Steuer, R. On the analysis and interpretation of correlations in metabolomic data. *Brief. Bioinform.* **7**, 151–158 (2006).
36. Steuer, R., Kurths, J., Fiehn, O. & Weckwerth, W. Interpreting correlations in metabolomic networks. *Biochem. Soc. Trans.* **31**, 1476–1478 (2003).
37. Steuer, R., Kurths, J., Fiehn, O. & Weckwerth, W. Observing and interpreting correlations in metabolomic networks. *Bioinformatics* **19**, 1019–1026 (2003).
38. Voit, E. O., Marino, S. & Lall, R. Challenges for the identification of biological systems from *in vivo* time series data. *In Silico Biol.* **5**, 83–92 (2005).
39. Jansen, R. C. & Nap, J. P. Genetical genomics: the added value from segregation. *Trends Genet.* **17**, 388–391 (2001).
40. Keurentjes, J. J. *et al.* The genetics of plant metabolism. *Nature Genet.* **38**, 842–849 (2006).
41. Fu, J., Swertz, M., Keurentjes, J. & Jansen, R. MetaNetwork: a computational protocol for the genetic study of metabolic networks. *Nature Protoc.* **2**, 685–694 (2007).
42. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
43. Cheung, V. G. *et al.* Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**, 1365–1369 (2005).
44. Sauer, U. High-throughput phenomics: experimental methods for mapping fluxomes. *Curr. Opin. Biotechnol.* **15**, 58–63 (2004).
45. Sauer, U. Metabolic networks in motion: 13C-based flux analysis. *Mol. Syst. Biol.* [online], http://www.nature.com/msb/journal/v2/n1/pdf/msb4100109.pdf (2006).
46. Herrgard, M. J., Fong, S. S. & Palsson, B. O. Identification of genome-scale metabolic network models using experimentally measured flux profiles. *PLoS Comput. Biol.* **2**, e72 (2006).
47. Dudley, A. M., Janse, D. M., Tanay, A., Shamir, R. & Church, G. M. A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol. Syst. Biol.* [online], http://www.cs.tau.ac.il/~rshamir/papers/aimee.pdf (2005).
48. Forster, J., Famili, I., Fu, P., Palsson, B. O. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244–253 (2003).
49. Reed, J. L., Vo, T. D., Schilling, C. H. & Palsson, B. O. An expanded genome-scale model of *Escherichia coli* K-12 (*i*JR904 GSM/GPR). *Genome Biol.* **4**, R54 (2003).
50. Kuepfer, L., Sauer, U. & Blank, L. M. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* **15**, 1421–1430 (2005).
51. Oh, Y. K., Palsson, B. O., Park, S. M., Schilling, C. H. & Mahadevan, R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.* **282**, 28791–28799 (2007).
52. Reed, J. L. *et al.* Systems approach to refining genome annotation. *Proc. Natl Acad. Sci. USA* **103**, 17480–17484 (2006).
53. Kumar, V. S., Dasika, M. S. & Maranas, C. D. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* **8**, 212 (2007).
54. Kümmel, A., Panke, S. & Heinemann, M. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol. Syst. Biol.* [online], http://www.nature.com/msb/journal/v2/n1/pdf/msb4100074.pdf (2006).
55. Guimera, R. & Nunes Amaral, L. A. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
56. Pal, C., Papp, B. & Lercher, M. J. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nature Genet.* **37**, 1372–1375 (2005).
57. Csete, M. & Doyle, J. Bow ties, metabolism and disease. *Trends Biotechnol.* **22**, 446–450 (2004).
58. Clardy, J., Fischbach, M. A. & Walsh, C. T. New antibiotics from bacterial natural products. *Nature Biotechnol.* **24**, 1541–1550 (2006).
59. Fischbach, M. A. & Walsh, C. T. Biochemistry: directing biosynthesis. *Science* **314**, 603–605 (2006).
60. Jansen, R. *et al.* A Bayesian networks approach for predicting protein–protein interactions from genomic data. *Science* **302**, 449–453 (2003).
61. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. & Eisenberg, D. A combined algorithm for genome-wide prediction of protein function. *Nature* **402**, 83–86 (1999).
62. Fong, S. S. & Palsson, B. O. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nature Genet.* **36**, 1056–1058 (2004).
63. Ibarra, R. U., Edwards, J. S. & Palsson, B. O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420**, 186–189 (2002).
64. Fong, S. S., Nanchen, A., Palsson, B. O. & Sauer, U. Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *J. Biol. Chem.* **281**, 8024–8033 (2006).
65. Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99**, 15112–15117 (2002).
66. Burgard, A. P., Pharkya, P. & Maranas, C. D. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84**, 647–657 (2003).
67. Fong, S. S. *et al.* *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol. Bioeng.* **91**, 643–648 (2005).

Acknowledgements
The authors thank M. Spitzer, J. Wildenhain, M. Swertz, R. Jansen, J. Fu, Y. Li, E. Takano, D. Höller and R. Steuer for their constructive comments on the manuscript. We apologize to those authors whose relevant work we could not cite owing to space constraints.

DATABASES
Entrez Genome Project: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj
Arabidopsis thaliana | *Escherichia coli* | *Geobacter sulphurreducens* | *Haemophilus influenzae* | *Methanosarcina barkeri* | *Trypanosoma brucei*

FURTHER INFORMATION
Rainer Breitling's homepage: http://gbic.biol.rug.nl/~rbreitling/
Enzyme Commission: http://www.chem.qmul.ac.uk/iubmb/enzyme/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2008 Nature Publishing Group