Metagrowth: a new resource for the building of metabolic hypotheses in microbiology

Hiroyuki Ogata* and Jean-Michel Claverie

Information Génomique et Structurale, CNRS UPR 2589, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France

ABSTRACT

Metagrowth is a new type of knowledge base developed to guide the experimental studies of culture conditions of obligate parasitic bacteria. We have gathered biological evidences giving possible clues to the development of the axenic (i.e. 'cell-free') growth of obligate parasites from various sources including published literature, genomic sequence information, metabolic databases and transporter databases. The database entries are composed of those evidences and specific hypotheses derived from them. Currently, 200 entries are available for *Rickettsia prowazekii*, *Rickettsia conorii*, *Tropheryma whipplei*, *Treponema pallidum*, *Mycobacterium tuberculosis* and *Coxiella burnetii*. The web interface of Metagrowth helps users to design new axenic culture media eventually suitable for those bacteria. Metagrowth is accessible at http://igs-server.cnrs-mrs.fr/axenic/.

INTRODUCTION

A number of bacteria resist axenic (i.e. 'cell-free') culture in the laboratory. Those include obligate parasites causing serious human diseases, such as *Rickettsia* (1,2) and *Mycobacterium leprae* (3). They adapt to a limited environment that provides appropriate physical conditions, nutriments and other factors required for their replication and growth. Current culture systems of obligate parasitic bacteria depend on eukaryotic cells (e.g. for *Rickettsia*) or even entire living animals (e.g. for *M.leprae*). The lack of cell-free culture media poses a critical problem in studying these bacteria. Without cell-free culture, it is impossible to use modern experimental approaches (e.g. transcriptomics, proteomics) that depend on non-contaminated RNA or protein extractions. Thus, the establishments of axenic culture media for those pathogens would have a significant impact on the medical and biological research communities working on these diseases. In a more fundamental way, this type of study might help to unraveling various type of signals involved in their host–parasite relationships.

With the recent development of metabolic databases (4,5), genome-based metabolic reconstruction has become an efficient approach to tackle this problem (6,7). By examining the metabolic pathways predicted from genomic sequence analyses, one can generate testable hypotheses for the improvement of bacterial culture conditions. We recently analyzed the complete genome sequence of a human pathogen, *Tropheryma whipplei* strain Twist, and identified significant deficiencies in the biosynthesis of nine amino acids (8). Remarkably, this knowledge effectively guided the development of the first axenic culture medium to grow this fastidious microorganism (7) that had been previously cultured only in association with a fibroblast cell line (HEL) (9). We believe that this type of approach should be generalized and could allow more obligate parasitic bacteria to be grown in a cell-free culture medium.

It is clear that explicit hypotheses (e.g. 'required nutriments') and supporting evidences (e.g. 'deficiencies of the *de novo* synthesis') are determinants for this type of study. However, such a biological knowledge is usually dispersed in literature and various biological databases. Till date, no existing database exhaustively collects and systematically provides biological knowledge about the cultivation of obligate parasites. This prompted us to gather evidences and hypotheses that are relevant to the improvement of the culture conditions of obligate parasites and make them available in a knowledge-base named Metagrowth (http://igs-server.cnrs-mrs.fr/axenic/). In this paper, we describe the source of the data accessible in Metagrowth as well as its web interface guiding the user to design new cell-free culture media of obligate parasites.

DATA IN METAGROWTH

Metagrowth is gathering 'evidences' and derived 'hypotheses' relevant to the improvement of the culture conditions of parasitic bacteria as follows:

Example 1. Evidence: 'The genome does not encode enzymes for the biosynthesis of compound X, but encodes a

---

transporter for *X*'. → Hypothesis: 'Adding *X* may improve the growth'.

Example 2. Evidence: 'The genome encodes enzymes requiring cofactor *Y*, but does not encode genes for the biosynthesis of cofactor *Y*'. → Hypothesis: 'Adding *Y* may improve the growth'.

These kinds of information are collected from the published literature, genomic sequence databases (4), metabolic databases (4,5,10,11) and transporter databases (12). Table 1 shows a tentative Metagrowth entry describing the supplementation of *S*-adenosyl-L-methionine for the improvement of *Rickettsia* culture conditions. The 'Evidence' and 'Hypothesis' records are the two main components of the database entry. The evidence record is a free text describing experimentally validated facts or predicted metabolic features. Associated hypotheses for the improvement of the culture condition are stored in the hypothesis record. In the hypothesis record, we currently describe the supplementation of organic or inorganic compounds in the medium, or appropriate physical conditions such as oxygen concentration. A prefix 'IN=' in the hypothesis record designates a preferential input (a compound or a physical condition) to the culture medium that could be experimentally tested. Hyperlinks to relevant genes, pathways and literature in the source

Table 1. A Metagrowth entry

Entry	E0435
Organism	*R.prowazekii* (TAX:782 Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; Rickettsiaceae; Rickettsieae; Rickettsia; typhus group) *R.conorii* (TAX:781 Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; Rickettsiaceae; Rickettsieae; Rickettsia; spotted fever group)
Evidence	Presence of transporters for *S*-adenosylmethionine (AdoMet) (24), and the lack (pseudogene status) of AdoMet synthetase, MetK, in Rickettsia (25–27). AdoMet is an important substrate for methyltransferase reaction in the cell.
Hypothesis	IN = *S*-Adenosyl-L-methionine cpd:C00019, (ET = DT)
Gene	RP076 RC0106

databases are provided to direct the users to the original data. Currently, we have accumulated 220 Metagrowth entries for 6 species of bacteria (5 genera): *Rickettsia prowazekii* and *Rickettsia conorii* (agents of typhus and spotted fever, respectively; 40 entries), *T.whipplei* (Whipple's disease; 38 entries), *Treponema pallidum* (13) (syphilis; 42 entries), *M.leprae* (leprosy; 63 entries) and *Coxiella burnetii* (14) (Q-fever; 37 entries).

Rating of the reliability of collected scientific evidences is an important issue in constructing a database of biological hypotheses (15). In Metagrowth, the relationships between evidences and hypotheses were classified into several categories. We refer to the relationships as 'evidence types' in Metagrowth. The evidence types could be used for the prioritization of different hypotheses supported by different kinds of evidences. We defined four major classes of evidence types. Class I evidences describe the inability to synthesize a compound, either by metabolic deficiencies or by general incapability of biosynthesis (e.g. inorganic molecules such as metal ions). Class II evidences refer to the importing capability of a compound, either by active transporters or by passive membrane permeability. Class III evidences refer to the requirement or utilization of a compound by the bacteria. Those include cofactors required for known or predicted enzymatic reactions in the cell, and basic building blocks of macromolecules such as the 20 amino acids. Class IV evidences refer to the other type of evidences, mostly experimentally validated facts. Each class was further divided, leading to a total of 10 subclasses. The precise definitions of these evidence classes and subclasses are provided in the Supplementary Material (Table S1). The evidence subclasses are identified in the hypothesis record of Metagrowth entries with a 'ET=' prefix (Table 1). Figure 1 shows the current status of the number of hypotheses supported by different types of evidences.

DESIGN OF A NEW CELL-FREE CULTURE MEDIUM

Browsing Metagrowth entries, users can easily obtain a list of nutriment compounds and the corresponding list of evidences
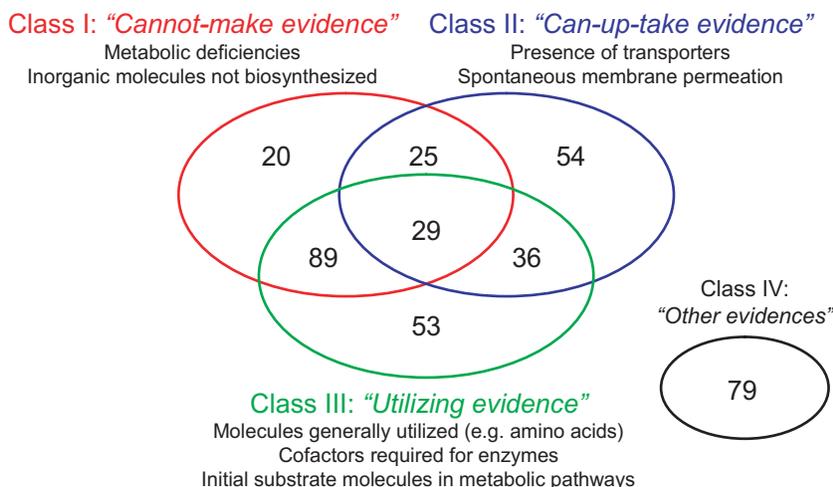

Figure 1. Number of the hypotheses in Metagrowth supported by different classes of evidences.

suggesting the supplementation of the medium by these compounds. An important practical issue is the determination of the concentrations of those supplemented molecules. To help the users with this respect, Metagrowth proposes a range of concentrations for different compounds. The ranges were determined according to the roles of molecules in bacterial cells and the concentrations of the molecules of the same role in several reference culture media. As reference culture media, we selected two complex culture media for fastidious (i.e. 'difficult to grow') bacteria: BSK-H medium designed to support the growth of the Lyme disease spirochete *Borrelia burgdorferi* (16), and TWH medium supporting the growth of *T.whipplei* (7). The concentrations of the components in the two media and the upper and lower limits of the concentrations within each compound category are provided in the Supplementary Material (Figure S1).

The users may specify one or more values within the suggested range of concentration of each molecule. If the users specify more than one concentration in the range, the combination of different concentrations for different molecules could lead to a huge number of experiments even if the number of molecules remains relatively small. The full combinatorial testing of 20 different nutriments, at two concentrations each, corresponds to $2^{20} \approx 10^6$. Such a large number of screening experiments can be avoided by the use of the incomplete factorial design approach. Incomplete factorial design is a mathematical method to effectively reduce the number of experiments required by a full combinatorial-screening of parameters (17). SAmBA is an implementation of the incomplete factorial design, which has been extensively used for the determination of protein crystallization conditions (18), and more recently for optimizing recombinant protein experiments (19). Metagrowth outputs the list of molecules and their concentrations in a format compatible with the SAmBA program (http://igs-server.cnrs-mrs.fr/samba/). In the above example with 20 compounds, 40 representative experimental protocols are proposed using SAmBA. In theory, the incomplete factorial design provides a minimal set of experiments in which the influence of each parameter can be examined rationally by statistical methods such as a multiple linear regression analysis.

FUTURE DIRECTIONS AND CONCLUSIONS

Many evidences in Metagrowth originate in metabolic analyses described in the literature such as whole genome sequencing papers. They are usually based on the visual inspection of predicted metabolic pathways. The use of *in silico* simulation studies with more sophisticated mathematical metabolic models (20–23) than those available in the current metabolic databases (4,5) is clearly the next improvement in the generation of metabolic hypotheses. With these approaches, one may, more precisely, examine Metagrowth evidences and derived hypotheses such as 'a metabolic pathway from *X* to *Y* lacks an enzyme, thus the addition of *Y* in the medium may improve the culture of bacteria'. *In silico* simulation studies may reveal an alternative pathway to the metabolite *Y* bypassing the missing reaction steps.

In the current Metagrowth, we only present predictions for preferential inputs to the culture conditions (designated

by the 'IN=' prefix). We plan to incorporate other kinds of information such as an 'unnecessary' or 'toxic' in association with a compound. Compound nomenclature in Metagrowth is based on LIGAND (10), in which hierarchical relationships between individual and generic compound names are not well treated. Standardization of the compound name in Metagrowth will be required to facilitate data update and to automatically detect data redundancies.

Genome sequence analysis and metabolic reconstruction of *T.whipplei* led to the establishment of the first cell-free culture medium allowing this fastidious bacteria to grow outside its cellular host. Further improvement of the culture condition using Metagrowth may lead to an even faster growth, which would further facilitate the manipulation and study of this microorganism. The development of an axenic culture medium for other bacteria could be more challenging. For *M.leprae* and *T.pallidum*, a large body of research has been carried out to explore the possibility of axenic cultivation as can be seen in Metagrowth. The study of bacterial culture conditions offers new testable and valuable challenges for the whole cell metabolic modeling and simulation studies. It helps in improving genomic annotation by identifying deficient or alternative metabolic pathways. It helps better characterization of host–parasite relationships, eventually giving us clues about new therapeutic targets. We hope to help the scientific community working on those pathogens by providing comprehensive information about their culture conditions through Metagrowth.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online. Table S1 gives the definition of the classes and the subclasses of the relationships between evidences and hypotheses. Figure S1 gives the concentrations of the components in existing complex culture medium.

REFERENCES

1. Andersson,S.G., Zomorodipour,A., Andersson,J.O., Sicheritz-Ponten,T., Alsmark,U.C., Podowski,R.M., Naslund,A.K., Eriksson,A.S., Winkler,H.H. and Kurland,C.G. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, **396**, 133–140.
2. Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.-E., Barbe,V., Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.-M. *et al.* (2001) Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. *Science*, **293**, 2093–2098.
3. Cole,S.T., Eiglmeier,K., Parkhill,J., James,K.D., Thomson,N.R., Wheeler,P.R., Honore,N., Garnier,T., Churcher,C., Harris,D. *et al.* (2001) Massive gene decay in the leprosy bacillus. *Nature*, **409**, 1007–1011.
4. Kanehisa,M., Goto,S., Kawashima,S., Okuno,Y. and Hattori,M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.
5. Krieger,C.J., Zhang,P., Mueller,L.A., Wang,A., Paley,S., Arnaud,M., Pick,J., Rhee,S.Y. and Karp,P.D. (2004) MetaCyc: a multiorganism

database of metabolic pathways and enzymes. *Nucleic Acids Res.*, **32**, D438–D442.

6. Lemos,E.G., Alves,L.M. and Campanharo,J.C. (2003) Genomics-based design of defined growth media for the plant pathogen *Xylella fastidiosa*. *FEMS Microbiol. Lett.*, **219**, 39–45.

7. Renesto,P., Crapoulet,N., Ogata,H., La Scola,B., Vestris,G., Claverie,J.-M. and Raoult,D. (2003) Genome-based design of a cell-free culture medium for *Tropheryma whipplei*. *Lancet*, **362**, 447–449.

8. Raoult,D., Ogata,H., Audic,S., Robert,C., Suhre,K., Drancourt,M. and Claverie,J.-M. (2003) *Tropheryma whipplei* Twist: a human pathogenic actinobacteria with a reduced genome. *Genome Res.*, **13**, 1800–1809.

9. Raoult,D., Birg,M.L., La Scola,B., Fournier,P.E., Enea,M., Lepidi,H., Roux,V., Piette,J.C., Vandenesch,F., Vital-Durand,D. *et al.* (2000) Cultivation of the bacillus of Whipple's disease. *N. Engl. J. Med.*, **342**, 620–625.

10. Goto,S., Okuno,Y., Hattori,M., Nishioka,T. and Kanehisa,M. (2002) LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.*, **30**, 402–404.

11. Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.

12. Paulsen,I.T., Nguyen,L., Sliwinski,M.K., Rabus,R. and Saier,M.H.,Jr (2000) Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *J. Mol. Biol.*, **301**, 75–100.

13. Fraser,C.M., Norris,S.J., Weinstock,G.M., White,O., Sutton,G.G., Dodson,R., Gwinn,M., Hickey,E.K., Clayton,R., Ketchum,K.A. *et al.* (1998) Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science*, **281**, 375–388.

14. Seshadri,R., Paulsen,I.T., Eisen,J.A., Read,T.D., Nelson,K.E., Nelson,W.C., Ward,N.L., Tettelin,H., Davidsen,T.M., Beanan,M.J. *et al.* (2003) Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc. Natl Acad. Sci. USA*, **100**, 5455–5460.

15. Karp,P.D., Paley,S., Krieger,C.J. and Zhang,P. (2004) An evidence ontology for use in pathway/genome databases. *Pac. Symp. Biocomput.*, 190–201.

16. Pollack,R.J., Telford,S.R.,III and Spielman,A. (1993) Standardization of medium for culturing Lyme disease spirochetes. *J. Clin. Microbiol.*, **31**, 1251–1255.

17. Carter,C.W.,Jr and Carter,C.W. (1979) Protein crystallization using incomplete factorial experiments. *J. Biol. Chem.*, **254**, 12219–12223.

18. Audic,S., Lopez,F., Claverie,J.-M., Poirot,O. and Abergel,C. (1997) SAmBA: an interactive software for optimizing the design of biological macromolecules crystallization experiments. *Proteins*, **29**, 252–257.

19. Abergel,C., Coutard,B., Byrne,D., Chenivesse,S., Claude,J.B., Deregnaucourt,C., Fricaux,T., Gianesini-Boutreux,C., Jeudy,S., Lebrun,R. *et al.* (2003) Structural genomics of highly conserved microbial genes of unknown function in search of new antibacterial targets. *J. Struct. Funct. Genomics*, **4**, 141–157.

20. Arita,M. (2003) *In silico* atomic tracing by substrate-product relationships in *Escherichia coli* intermediary metabolism. *Genome Res.*, **13**, 2455–2466.

21. Heiner,M., Koch,I. and Will,J. (2004) Model validation of biological pathways using Petri nets-demonstrated for apoptosis. *Biosystems*, **75**, 15–28.

22. Ibarra,R.U., Edwards,J.S. and Palsson,B.O. (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature*, **420**, 186–189.

23. Price,N.D., Papin,J.A., Schilling,C.H. and Palsson,B.O. (2003) Genome-scale microbial *in silico* models: the constraints-based approach. *Trends Biotechnol.*, **21**, 162–169.

24. Tucker,A.M., Winkler,H.H., Driskell,L.O. and Wood,D.O. (2003) *S*-Adenosylmethionine transport in *Rickettsia prowazekii*. *J. Bacteriol.*, **185**, 3031–3035.

25. Andersson,S.G., Zomorodipour,A., Andersson,J.O., Sicheritz-Ponten,T., Alsmark,U.C., Podowski,R.M., Naslund,A.K., Eriksson,A.S., Winkler,H.H. and Kurland,C.G. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, **396**, 133–140.

26. Ogata,H., Audic,S., Renesto-Audiffren,P., Fournier,P.E., Barbe,V., Samson,D., Roux,V., Cossart,P., Weissenbach,J., Claverie,J.M. and Raoult,D. (2001) Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. *Science*, **293**, 2093–2098.

27. Andersson,J.O. and Andersson,S.G. (1999) Genome degradation is an ongoing process in *Rickettsia*. *Mol. Biol. Evol.*, **16**, 1178–1191.