Metagenome-scale metabolic network reconstruction

F. Guktnecht [(a)], G. Collet [(b)], J. Cambefort [(b)], J. Andres [(a)], S. Prigent [(b)], D. Eveillard [(c)], A. Siegel [(b)], F. Plewniak [(a)]

[(a)] Ecophysiologie Moléculaire des Microorganismes. G.M.G.M - UdS/CNRS UMR7156. 28 rue Goethe. 67083 Strasbourg Cedex
[(b)] Projet Dyliss IRISA CNRS/Univ Rennes 1 & Inria Rennes. Campus de Beaulieu. 35042 Rennes cedex.
[(c)] LINA. Univ-Nantes LINA UMR 6241 Université de Nantes. 2, rue de la Houssinière. BP 92208. 44322 Nantes Cedex 03

Genome-scale metabolic networks have been successfully used to understand a wide range of organism's physiology from *Escherichia coli* to *Homo sapiens* [4]. However, despite their interest, individual genomes remain difficult to extract from their natural environment, which seriously limits our ability to model the metabolism and metabolic interactions within natural ecosystems. The capacity to use alternative omic sources, such as metagenomic data, is therefore highly desirable for metabolic network reconstructions, in particular at the scale of a whole microbial community if one is interested to better understand bacterial consortia behaviors.

Supporting this hypothesis, we present herein the reconstruction of a single metabolic network for the microbial community of Carnoulès Acid Mine Drainage (AMD) that exhibits high concentrations of metals and arsenic. This community was shown to be intricate and dominated by seven bacterial strains whose genomes were reconstructed from metagenome sequencing [1]. Five of these strains represent yet uncultivated bacteria, including two that belong to a novel bacterial genus (*Candidatus* Fodinabacter communificans). *Euglena mutabilis*, a photosynthetic protist frequently found in AMDs, was identified as the dominant eukaryotic species in the Carnoulès community where it probably plays the role of a primary producer.

As inputs, metabolic network reconstruction techniques consider genomic knowledge. Since the genome of *Euglena mutabilis* is not publicly available, its metabolic network was reconstructed from *de novo* transcriptomics and metabolomic data obtained from previous studies [2, 3] and complementary metabolomic experiments. The transcript sequences obtained by pyrosequencing of total cDNA libraries from *E. mutabilis* cultures in minimal medium with and without arsenic were translated in all six frames and searched using BLASTP for similarity with MetaCyc 16.0 enzyme sequences. A MetaCyc reaction was considered to be present if a corresponding enzyme had a significant hit with a putative transcript. A first draft of the *E. mutabilis* metabolic network was thus obtained containing 630 reactions involving 937 different metabolites.

In parallel, metabolomic data from GC/MS analyses of Carnoulès water and *E. mutabilis* culture supernatant allowed the identification of 72 metabolites produced by *E. mutabilis*. These metabolites were used as targets within the Meneco metabolic network completion software [5] (http://bioasp.github.io/meneco/) to suggest missing reactions in the draft network. More precisely, this tool makes use of a logic paradigm (Answer Set Programming) to identify all minimal sets of reactions from the MetaCyc repository which allows restoring the capability of the draft metabolic network to produce target metabolites. This functional completion was iteratively performed in a supervised manner to ensure that the proposed completion for every pathway was in accordance with biological knowledge on the system and not influenced by any completion requirements in other pathways (cofactor biosynthesis,...). The appropriate seed metabolites were thus added at the appropriate steps to avoid unwanted penalties. Manual addition of reactions was also performed based on literature and *E. mutabilis* physiology. The ambiguities and gaps due to reactions involving generic compounds were solved by replacing generic compounds by the appropriate metabolite instances. It is worth noticing that these additions of reactions were validated by relaxed similarity search of translated transcripts by HMMsearch and BLASTP with reference enzyme sequences. Significant hits were confirmed by

reciprocal blast hits in the NCBI non-redundant protein database. Overall, after addition of 337 reactions to the initial draft, the *E. mutabilis* metabolic network is able to produce the 72 target metabolites from 27 seed metabolites.

Complementary, the metabolic network for each of the seven dominant bacterial strains was reconstructed from the automated annotations of their genome sequence on the MicroScope platform [6]. We thus obtained seven networks containing each from 392 to 1111 reactions depending on the degree of completion of the corresponding genome. All the remaining metagenomic sequences that could not be assembled into genomes and which represented 70851 fragments of length up to 122 kb with an average of 1 kb, were annotated using the same method as *E. mutabilis* transcripts. This allowed us to obtain a draft metabolic network containing 914 reactions involving 1182 metabolites representing the non-dominant strains in the community, considered in the model as a single compartment.

The whole set of nine metabolic networks that compose the AMD environment were subsequently combined as a single large network, where each member of the community is considered as a single compartment. This combination allows investigating all potential interactions between the nine partners by the sake of introduction of hypothetical exchange reactions to link the non-produced reactants of every partner to the corresponding metabolite produced in others. The large proportion of such links between *E. mutabilis* and *Candidatus* Fodinabacter communificans was consistent with the suspected importance of the interactions between those organisms [1]. Interestingly, a majority of non-produced reactants in the dominant bacterial strains can be produced by the non-dominant ones, suggesting that the latter may have a major role in this ecosystem and should not be overlooked in the final model.

Although all metabolic networks have not been completed yet, and more elaborate analyses still remain to be done in order to identify in detail the actual exchanges at stake, these preliminary observations suggest that metabolic networks reconstructed at the scale of a whole microbial community from metagenomic data provide insights into the potential interactions between its members.

References

[1] P. N. Bertin, A. Heinrich-Salmeron, E. Pelletier, F. Goulhen-Chollet, F. Arsène-Ploetze, S. Gallien, B. Lauga, C. Casiot, A. Calteau, D. Vallenet, V. Bonnefoy, O. Bruneel, B. Chane-Woon-Ming, J. Cleiss-Arnold, R. Duran, F. Elbaz-Poulichet, N. Fonknechten, L. Giloteaux, D. Halter, S. Koechler, M. Marchal, D. Mornico, C. Schaeffer, A. A. T. Smith, A. Van Dorsselaer, J. Weissenbach, C. Médigue, and D. Le Paslier. Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J*, 5(11):1735–1747, Nov 2011.

[2] D. Halter, J. Andres, F. Plewniak, J. Poulain, C. Da Silva, F. Arsène-Ploetze, and P. N. Bertin. Arsenic hypertolerance in the protist e. mutabilis is mediated by specific transporters and functional integrity maintenance mechanisms. *Environ Microbiol*, Apr 2014.

[3] D. Halter, F. Goulhen-Chollet, S. Gallien, C. Casiot, J. Hamelin, F. Gilard, D. Heintz, C. Schaeffer, C. Carapito, A. Van Dorsselaer, G. Tcherkez, F. Arsène-Ploetze, and P. N. Bertin. In situ proteo-metabolomics reveals metabolite secretion by the acid mine drainage bio-indicator, euglena mutabilis. *ISME J*, 6(7):1391–1402, Jul 2012.

[4] J. Monk, J. Nogales, and B. Ø. Palsson. Optimizing genome-scale network reconstructions. *Nature Biotechnology*, 32(5):447–452, May 2014.

[5] T. Schaub and S. Thiele. Metabolic network expansion with answer set programming. In P. Hill and D. Warren, editors, *Logic Programming*, volume 5649 of *Lecture Notes in Computer Science*, pages 312–326. Springer Berlin Heidelberg, 2009.

[6] D. Vallenet, E. Belda, A. Calteau, S. Cruveiller, S. Engelen, A. Lajus, F. Le Fèvre, C. Longin, D. Mornico, D. Roche, Z. Rouy, G. Salvignol, C. Scarpelli, A. A. Thil Smith, M. Weiman, and C. Médigue. Microscope–an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res*, 41(Database issue):D636–D647, Jan 2013.