



**HAL**  
open science

## Generation of metabolomic-informed models of metabolism in complex microbial communities

Coralie Muller, Sylvain Prigent, Clémence Frioux

### ► To cite this version:

Coralie Muller, Sylvain Prigent, Clémence Frioux. Generation of metabolomic-informed models of metabolism in complex microbial communities. Journées scientifiques Agroécologie et Numérique 2025, Jan 2025, Dijon, France. . hal-04919231

**HAL Id: hal-04919231**

**<https://inria.hal.science/hal-04919231v1>**

Submitted on 29 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Generation of metabolomic-informed models of metabolism in complex microbial communities

coralie.muller@inria.fr

Coralie Muller<sup>1,2</sup>, Sylvain Prigent<sup>\*2</sup>, Clémence Frioux<sup>\*1</sup>

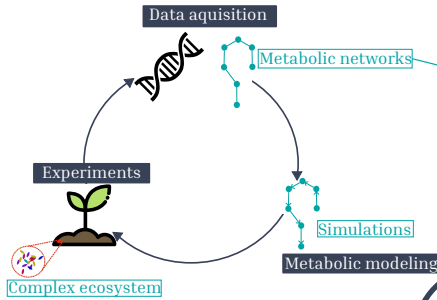
\* These authors contributed equally to the work

1. Inria, Univ. Bordeaux, INRAE, F-33400 Talence, France  
2. Univ. Bordeaux, INRAE, UMR1332 BFP, 33140 Villenave d'Ornon, France

Keywords: Systems biology -- Ecology -- Metabolic modelling -- Heterogeneous data

## A routine analysis...

One purpose of metabolic modelling [1] is to decipher the functions and interactions occurring in microbial communities, especially in complex environments.



## but incomplete!

Genome Scale Metabolic network (GSMN) a solid tool ...but with loopholes[2]

- Experimental studies are costly.
- Adaptability (custom parameters).
- Facility of reproducibility.
- Serve as a proxy for experimental design.
- Generalist: built with generic database (Kegg, Metacyc ...) main pathways.
- Biased towards model organisms.
- Misses important functions related to the secondary metabolism.[2]

**Primary metabolism**  
(Growth, cell division, respiration, photosynthesis, reproduction)

**Secondary metabolism**  
(Specialised pathways: antibiotic, color, odour, taste, protection)

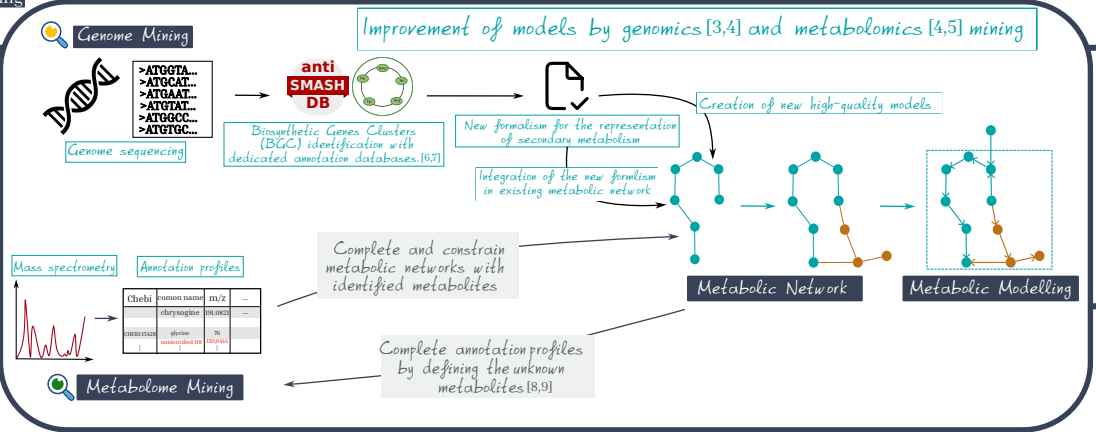
## Objectives

To better understand how plants recruit microbes, it is essential to develop a comprehensive description of secondary metabolism, which is currently lacking.  
Build high quality models of microbial metabolism could be a key point in development of methods for simulating and describing secondary metabolic pathways.

- 🔍 Define a formalism for the representation of secondary metabolism obtained from the genome mining.
- 🔍 Combine metabolic modelling and metabolomic profile annotation in a virtuous cycle.
- 🔍 Provide a methodological approach that fosters using metabolomic to capture the set of functions that are really active.

## Fitting step

To identify new metabolites potentially linked to secondary metabolism, we carry out a fitting step. This compares the metabolites in the metabolomics data with those in the metabolic network. Metabolites that are present in the metabolomics data and absent from the networks are kept.



The main challenge in this step is the heterogeneity of unique identifiers (IDs) for each metabolite across the various databases used to construct the metabolic networks. To address this, we will build a datatable that integrates information from multiple databases, enabling rapid and automated ID conversion.

## A real complexity

There is a real disparity between known and lesser known species in the identification of metabolites due to the fact that they are less studied and their data is poorly referenced in databases. Furthermore, finding metabolomic data for certain species can be a real challenge, in addition to the fact that there are many unidentified metabolites, as is the case here with *Penicillium polonicum*.

### 1. Build a datatable conversion file

Construction of a datatable enabling rapid and automatic conversion of IDs between the main databases (Metacyc [10], ModelSeed [11], BiGG [12], ...) for the reconstruction of metabolic networks.



UNIQUE-ID	ADD-COMPLEMENT	BiGG
PANTOTHENATE	panothethenic acid	fol
CPD-12826	lysine b	lys
PROHEMIE	lysine b	lys

manually curated complementary file

UNIQUE-ID	CHEBI	ADD-COMPLEMENT	SEED	BiGG
PANTOTHENATE	29032	panothethenic acid	cpd00644	pano_R
CPD-12826	62501	lysine b	cpd00393	lys
PROHEMIE	57305	lysine b	cpd00033	gly

### 2. Fit annotation profiles vs metabolic networks

Identify shared and unique metabolites between metabolomics data and a metabolic network for a given species.

CHEBI	common name	m/z	Annotation profiles
CHRYSOGENE	chrysoгене	91.0829	...
FOLATES	folates	139.037	...
GLYCINE	glycine	75	...
UNIDENTIFIED 118	unidentified 118	104.044	...

Match CHEBI	Match COMMON-NAME	Unmatch
CPD-12826 - 62501	GLY lysine	chrysoгене unidentified 118

**Note**  
The absence of a metabolite in a metabolic network while it is present in the metabolomics data is due to define and explore secondary metabolism.

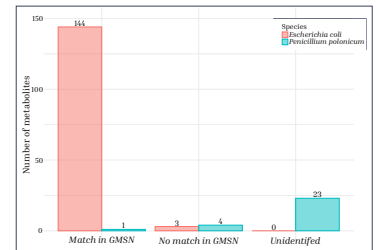


Fig: Comparison number of fit, unfit and unidentified between two species: one well-known *Escherichia coli* [13] and one lesser-known *Penicillium polonicum* [5]

## Take-home message

- Secondary metabolism represents specific pathways which are of great interest in many fields, but which are difficult to explore in models.
- Improving the quality of metabolic networks is essential if we want to be able to simulate secondary metabolism pathways.
- Real complexity in identifying secondary metabolites due to their absence in generalist databases and in lesser-known organisms.

## Perspectives

### Genome Mining

- Use specific databases and methods to identify BGCs of interest in genomes.
- Create a formalism to standardise BGC information and integrate it into an existing metabolic network.

### Metabolome Mining

- Developing a strategy to identify unknown metabolites in metabolomics data by gaining insights from metabolic networks.
- Further identification of secondary metabolites present in the metabolomic data and their pathways to integrate them to the metabolic network.

## References

[1] Cerk, K. et al. Microb. Biotechnol. P. Nedjic, C.G., Lecomte, M., Muller, C., Sherman, D.J., Hildebrand, F., Ishihara, S., and Frioux, C. (2020). Community-scale models of microbiomes: Articulating metabolic modelling and metagenome sequencing. *Microb. Biotechnol.*

[2] Qin, S., et al. (2021). Flux balance analysis based metabolic modeling of microbial secondary metabolism: Current status and outlook. *PLoS Comput. Biol.*

[3] Prigent, Sylvain, et al. « Reconstruction of 24 *Penicillium* Genome-scale Metabolic Models Shows Diversity Based on Their Secondary Metabolism ». *Biotechnology and Bioengineering* (2018)

[4] Han, JR., Ji, S., Li, WJ. et al. Mining microbial and metabolic dark matter in extreme environments: a roadmap for harnessing the power of multi-omics data. *Adv. Biotechnol.* 2, 26 (2014).

[5] Manshadi, Ayush, et al. « MNNO-An Open Source Software for Refining Metabolic Networks and Investigating Complex Network Activity Using Empirical Metabolomics Data ». *Analytical Chemistry*, (2020).

[6] Blin, K., Shaw, S., Kautsar, S.A., Medema, M.H., and Weber, T. (2020). The antiSMASH database version 7: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Res* 49.

[7] Floist, A., Ricart, E., Campari, C., Chevalier, M., Dufresse, V., Michalik, J., Jacques, P., Flaissat, C., Lissacq, P., Leclercq, V., et al. (2020). Norine: update of the nonribosomal peptide resource. *Nucleic Acids Res* 48.

[8] Paulino, N., Caulet, C., Damont, A., Peyriga, L., Durand, S., Deboerde, C., Alves, S., Berillon, S., Berton, T., Bir, R., et al. (2022). NoaForest: a multi platform digital infrastructure for interoperable metabolomic spectral data and metadata management. *Metabolomics*

[9] Cao, L., Guler, M., Tsigirlakou, A., Lee, Y.Y., Gutierrez, A., and Mohlman, H. (2021). Multitasking: learning mass spectrometry fragmentation of small molecules. *Nat. Commun.*

[10] [Metacyc] Caspi, R., Altman, T., Billington, R., Dreier, K., Foerster, H., Fulcher, C.A., Holland, T.A., Kosler, T.M., Kothari, A., Nishino, K., Okamoto, M., Latendresse, M., Mueller, L.A., Ogas, O., Paley, S., Subhraveti, P., Weaver, B.S., Weerasinghe, R., Zhang, P. and Karp, P.D. (2010)

[11] Henry, C.S., DeLough, M., Best, A.R., Frybarger, P.M., Linay, B., and R.I. Stevens. High throughput Generation and Optimization of Genome-scale Metabolic Models. *Nature Biotechnology*, (2008).

[12] King ZA, Liu JS, DeGler A, Miller PC, Federovics S, Terman JA, Ehrlich A, Palsson BO and Lewis NE. BiGG Models: A platform for integrating, standardizing, and sharing genome-scale models (2016) *Nucleic Acids Research*

[13] Lepp, M., Farke, N., Kuntz, M., et al. Systematic identification of metabolites controlling gene expression in *E. coli*. *Nat Commun* 10, 4463 (2019).