



**HAL**  
open science

## Metage2Metabo: metabolic complementarity applied to genomes of large-scale microbiotas for the identification of keystone species

Arnaud Belcour, Clémence Frioux, Méziane Aite, Anthony Bretaudeau, Falk Hildebrand, Anne Siegel

► **To cite this version:**

Arnaud Belcour, Clémence Frioux, Méziane Aite, Anthony Bretaudeau, Falk Hildebrand, et al.. Metage2Metabo: metabolic complementarity applied to genomes of large-scale microbiotas for the identification of keystone species. JOBIM 2020 - Journées Ouvertes de Biologie, Informatique et Mathématique, Jun 2020, Montpellier / Virtual, France. 10.1101/803056 . hal-03151934

**HAL Id: hal-03151934**

**<https://inria.hal.science/hal-03151934>**

Submitted on 25 Feb 2021

**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.

# Metage2Metabo

metabolic complementarity applied to genomes of  
large-scale microbiotas for the identification of  
keystone species

Arnaud Belcour\*, **Clémence Frioux\***, Méziane Aite, Anthony Bretaudeau, Falk Hildebrand, Anne Siegel

Source code: <https://github.com/AuReMe/metage2metabo>

Documentation: <https://metage2metabo.readthedocs.io/>

Preprint: <https://doi.org/10.1101/803056>



UMR

IRISA

INRAE



Quadram  
Institute  
Science • Health • Food • Innovation



Earlham  
Institute

# Context

## Metagenomics flooding:

- Newly-discovered (uncultured) microbes/taxa in microbiota
- Assemblies of numerous new reference genomes
- Incoming golden era of Metagenome-Assembled Genomes (MAGs) and metagenomic species

→ **Collections of thousands of genomes/MAGs**

Rely on relevant modelling approaches to tackle:

- **incompleteness** of poorly-studied genomes
- **impossibility to manually curate** thousands of metabolic networks (GSMNs)
- The **high combinatorics** brought by the analysis of all possible communities in the initial collection

→ **Network expansion algorithm** to model producibility



[Almeida et al 2019](#),  
[Nayfach et al 2019](#),  
[Pasolli et al 2019](#),  
[Parks et al 2017](#),  
[Almeida et al 2019](#)

**Going beyond « who is there? »** in metagenomics datasets and large collections of genomes

- What can microbes do?
- What does putative cooperation between microbes brings to the community?
- Which microbes matter most for specific functions?



[Edgar et al 2017](#),  
[De Oliveira Martins et al 2020](#),  
[Sun et al 2020](#)



[Handorf et al 2005](#),  
[Van der Ark et al 2017](#),  
[Bernstein et al 2019](#)

# Workflow

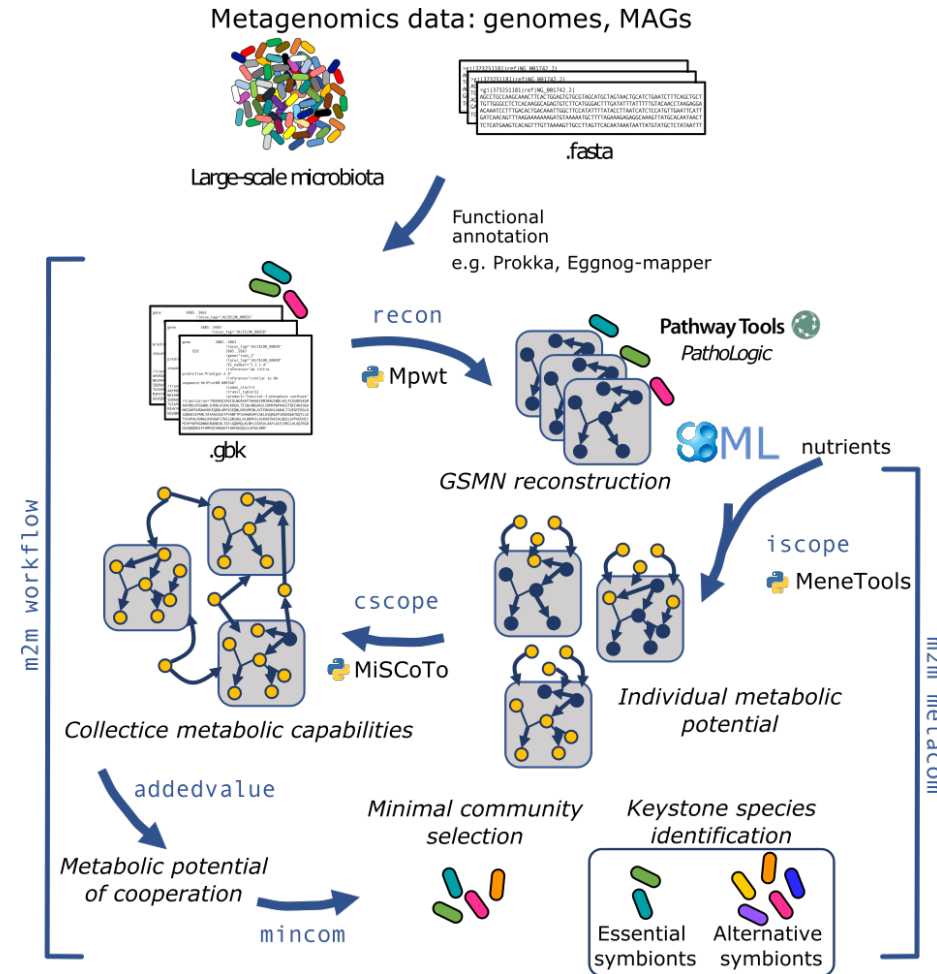
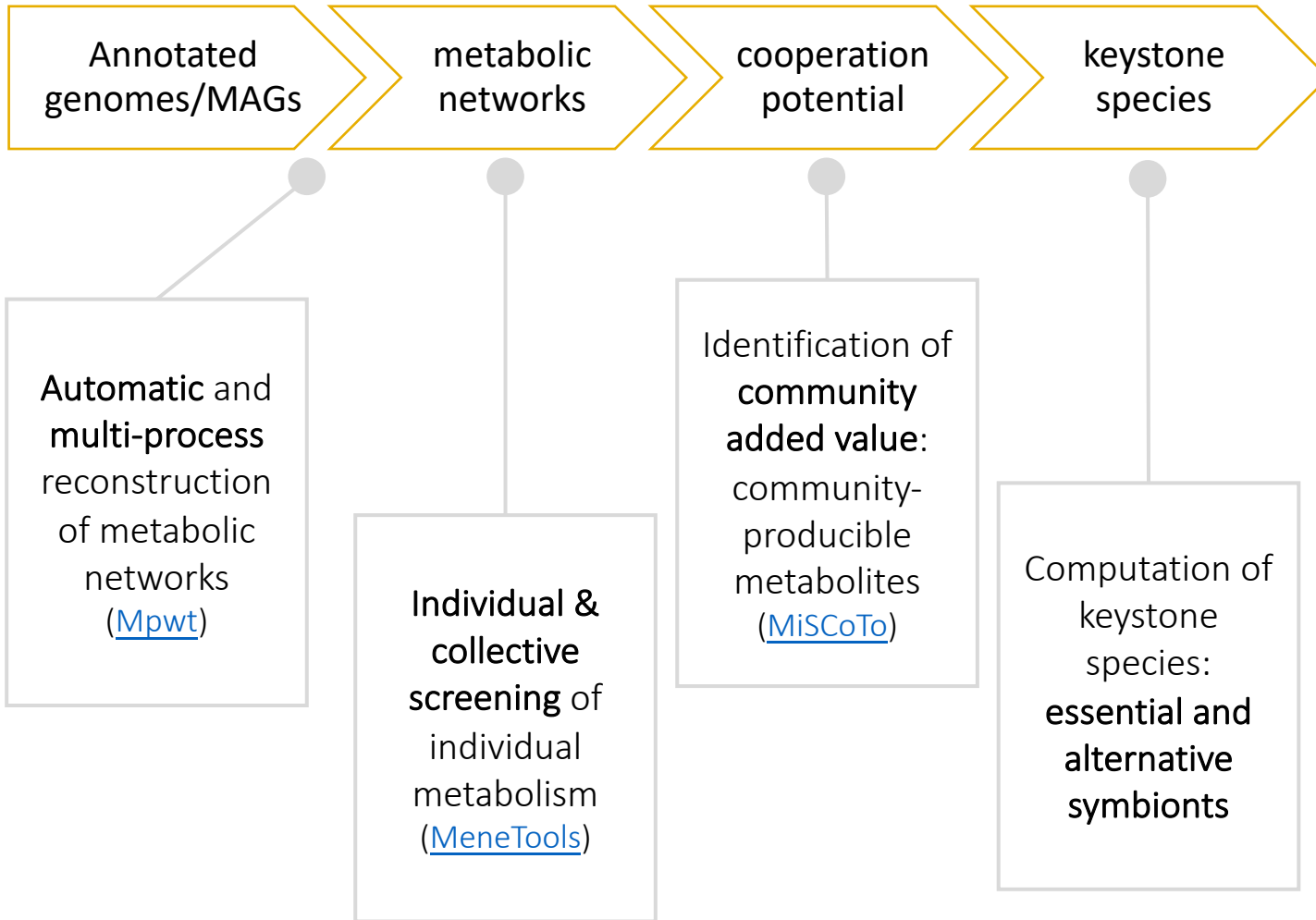


Fig. 1: flexibility of m2m's global workflow. The metacom command can run the workflow using existing metabolic networks (SBML)

**M2M reconstructs metabolic networks for genomes/MAGs and screens the complementarity of metabolisms in a large community.**

# Application to datasets

Keystone species (KS) = *microbes occurring in at least one minimal community ensuring the producibility of a set of target metabolites from given nutrients*. Efficient solving: no need to enumerate all communities to retrieve them.

KS = essential symbionts (ES) + alternative symbionts (AS)  
*ES are found in every minimal community*  
*AS are found in some minimal communities*

913 cow rumen MAGs ([Stewart et al 2018](#))  
 1520 draft bacterial reference genomes ([Zou et al 2019](#))

↓  
 m2m workflow

## Insights:

- Metabolic networks from reference genomes and MAGS display similar characteristics (Fig. 2)
- M2M metabolic screening and KS identification are robust to low genome degradations (Fig. 3)

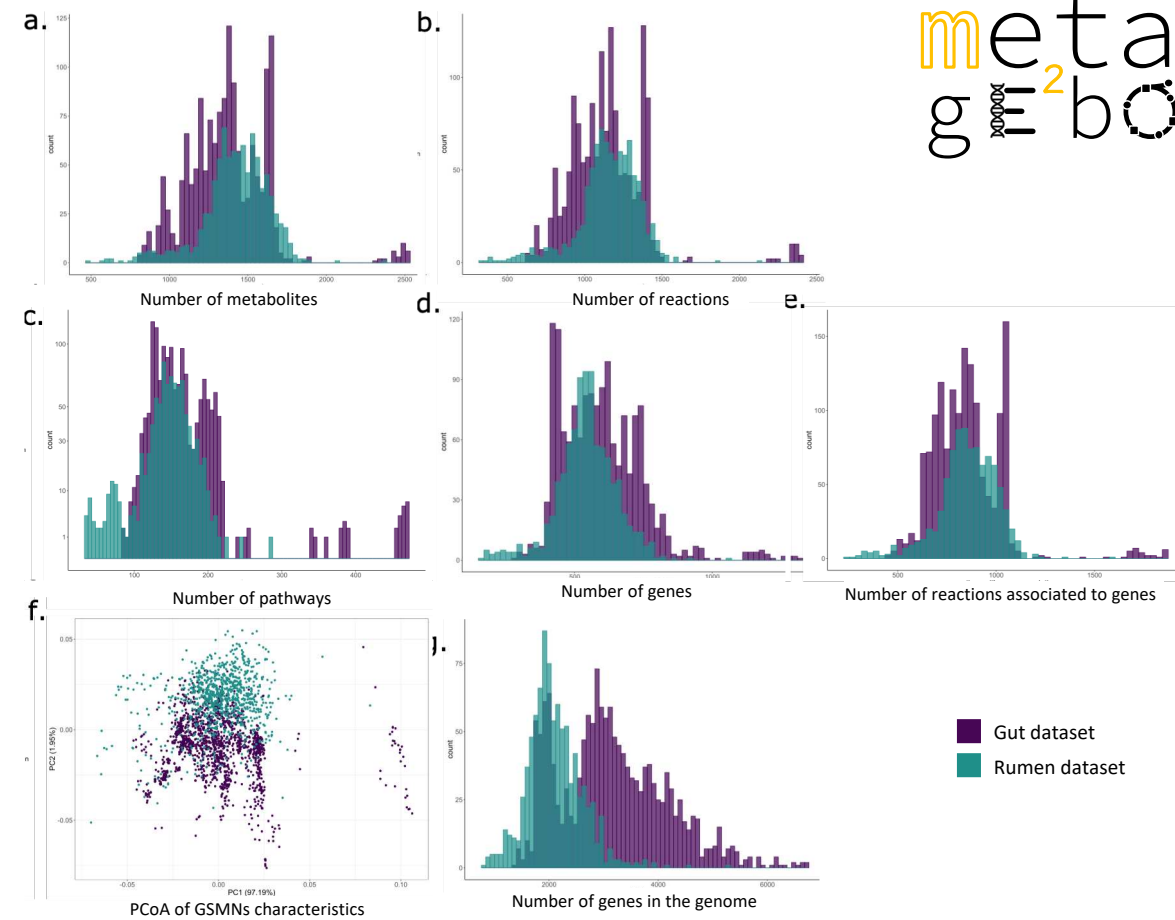


Fig. 2: Comparison of GSMNs from both datasets

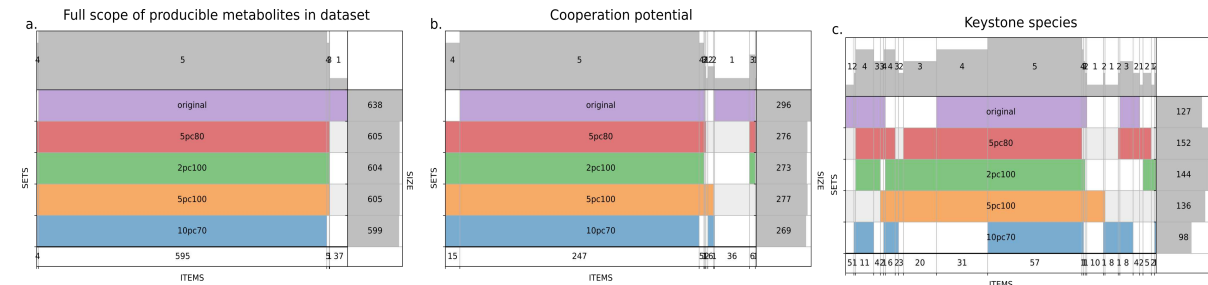


Fig. 3: Robustness analysis following MAGs degradation for the rumen dataset (upper row: original genomes, up to down: increasing genome degradation)

# Analysis of gut communities

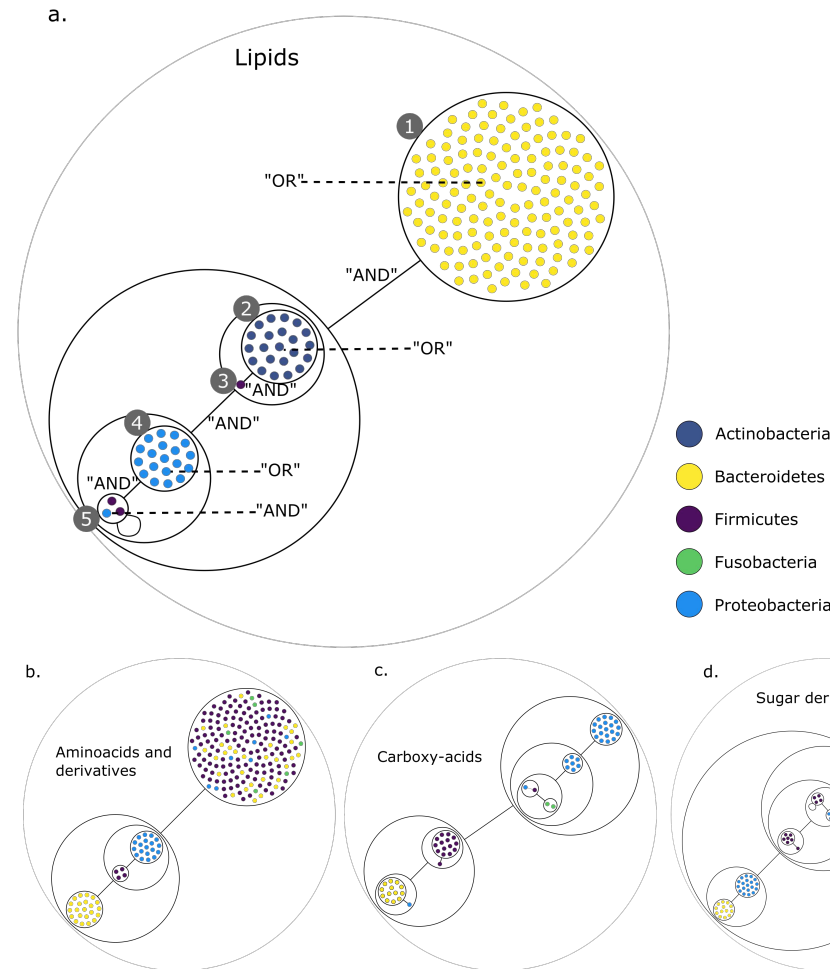
Collection of genomes = 1520 draft bacterial reference genomes ([Zou et al 2019](#))

Nutrients = typical EU diet (qualitative)

Targets = metabolites producible only through cooperation processes, grouped by categories

→ Enumeration of all minimal communities with m2m-analysis to study putative associations of microbial metabolisms

|   |    | Firm. | Bact. | Acti. | Prot. | Fuso. | total |
|---|----|-------|-------|-------|-------|-------|-------|
| <b>all (156 targets)</b>                      | KS | 41    | 47    | 103   | 12    | 2     | 205   |
| 25 bact. per community                        | ES | 8     | 0     | 1     | 2     | 0     | 11    |
| not enumerated                                | AS | 33    | 47    | 102   | 10    | 2     | 194   |
| <b>aminoacids and derivatives (5 targets)</b> | KS | 142   | 52    | 0     | 27    | 6     | 227   |
| 4 bact. per community                         | ES | 0     | 0     | 0     | 0     | 0     | 0     |
| 120,329 communities                           | AS | 142   | 52    | 0     | 27    | 6     | 227   |
| <b>aromatic compounds (11 targets)</b>        | KS | 52    | 0     | 0     | 20    | 0     | 72    |
| 5 bact. per community                         | ES | 2     | 0     | 0     | 1     | 0     | 3     |
| 950 communities                               | AS | 50    | 0     | 0     | 19    | 0     | 69    |
| <b>carboxyacids (14 targets)</b>              | KS | 16    | 13    | 0     | 28    | 2     | 59    |
| 9 bact. per community                         | ES | 2     | 0     | 0     | 2     | 0     | 4     |
| 48,412 communities                            | AS | 14    | 13    | 0     | 26    | 2     | 55    |
| <b>coA derivatives (10 targets)</b>           | KS | 106   | 0     | 50    | 17    | 1     | 174   |
| 5 bact. per community                         | ES | 0     | 0     | 0     | 0     | 1     | 1     |
| 95,256 communities                            | AS | 106   | 0     | 50    | 17    | 0     | 173   |
| <b>lipids (28 targets)</b>                    | KS | 3     | 140   | 22    | 20    | 0     | 185   |
| 7 bact. per community                         | ES | 3     | 0     | 0     | 1     | 0     | 4     |
| 58,520 communities                            | AS | 0     | 140   | 22    | 19    | 0     | 181   |
| <b>sugar derivatives (58 targets)</b>         | KS | 11    | 30    | 78    | 23    | 0     | 142   |
| 11 bact. per community                        | ES | 5     | 0     | 0     | 0     | 0     | 5     |
| 7,860,528 communities                         | AS | 6     | 30    | 78    | 23    | 0     | 137   |



- Graph compression enables representing up to millions of minimal communities for metabolic groups of targets.  
→ dealing with high combinatorics
- Identification of microbial groups with similar roles in communities, that follow phyla decomposition  
→ more informative than providing a single community solution

Table 1: Enumeration of minimal communities and their phyla composition for groups of target metabolites. KS = keystone species, ES = essential symbionts, AS = alternative symbionts

Fig. 4: Power graph analysis (graph compression) of the association of genomes in the predicted communities, coloured by phylum. In the non-compressed graphs, two nodes (microbes) were linked if they were found together in at least one minimal community.