



*19th International Conference on
Computational Methods in Systems Biology*

Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species

Arnaud Belcour^{*2}, **Clémence Frioux**^{*1,2,3}, Méziane Aite²,
Anthony Bretaudeau^{2,3}, Falk Hildebrand⁴ and Anne Siegel²



¹Inria CNRS INRAe, Univ. Bordeaux, France

²Inria IRISA CNRS, Univ. Rennes, France

³INRAe IGEPP – Genouest platform, Rennes, France

⁴Quadram Institute – Earlham Institute, Norwich UK

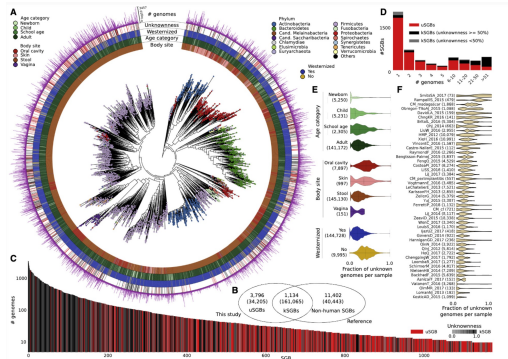
September 24, 2021

clemence.frioux@inria.fr

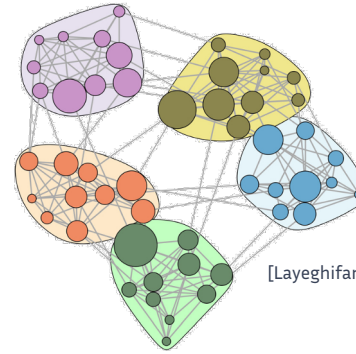
Inria Bordeaux Sud-Ouest – CNRS – INRAe – Univ. Bordeaux



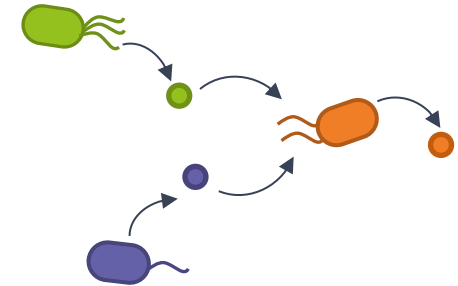
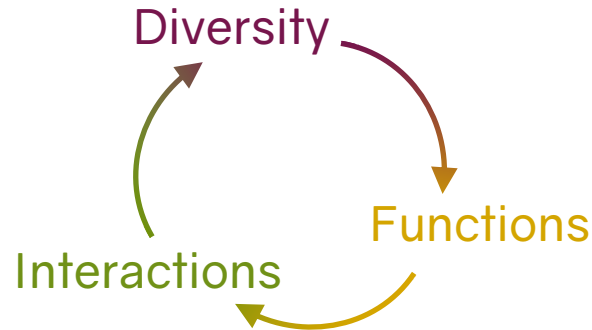
An era of microbiomes



[Pasolli et al 2019]



[Layeghifard et al 2017]

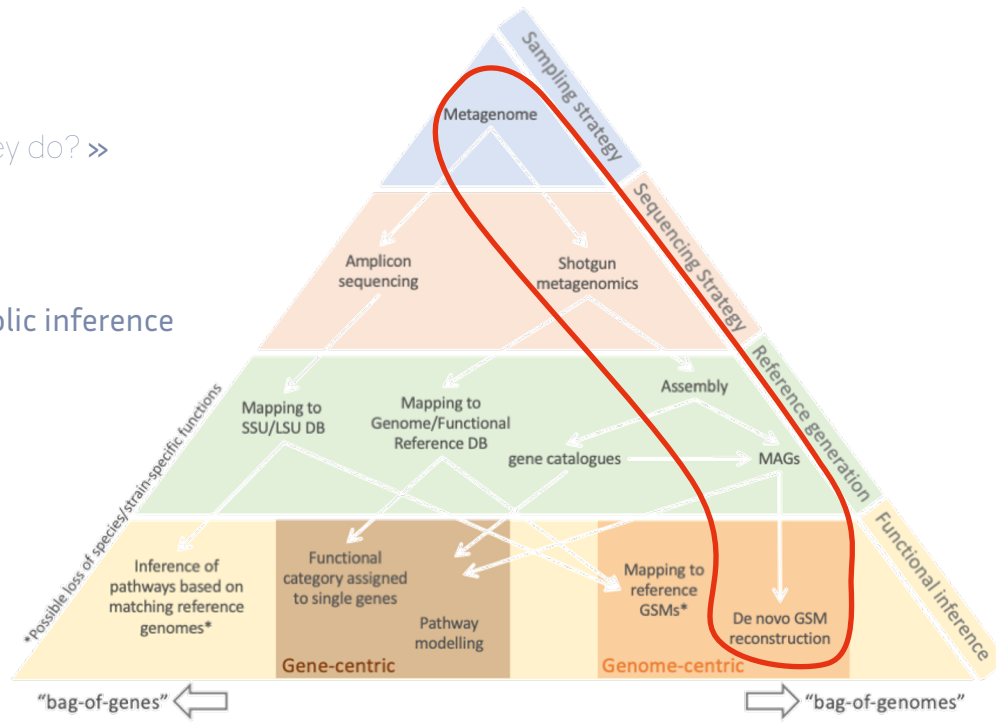


Identification of microbial interactions from metagenomic data is key to better understand complex ecosystems

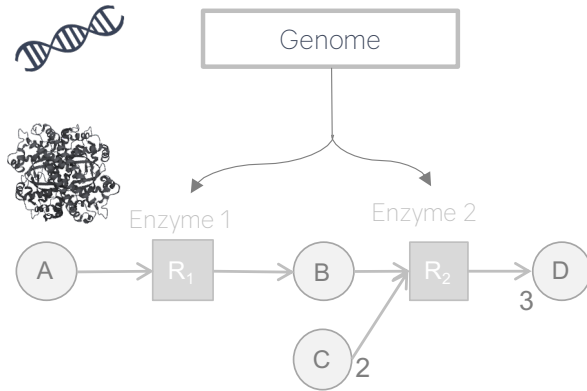
Metagenomics meets metabolism

- From « who is there? » to « what might they do? »
- Inferring functions in metagenomes
- Genome-centric approaches enable metabolic inference

Combine genome reconstruction and metabolic modelling to unravel the functional repertoire and community organisation of uncultivated species



Metabolic networks and communities

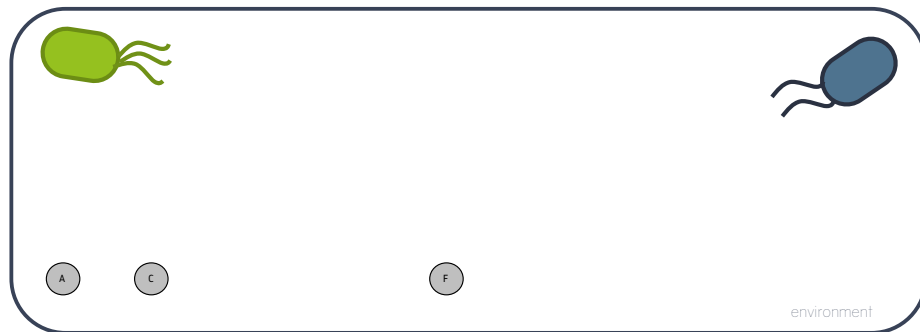
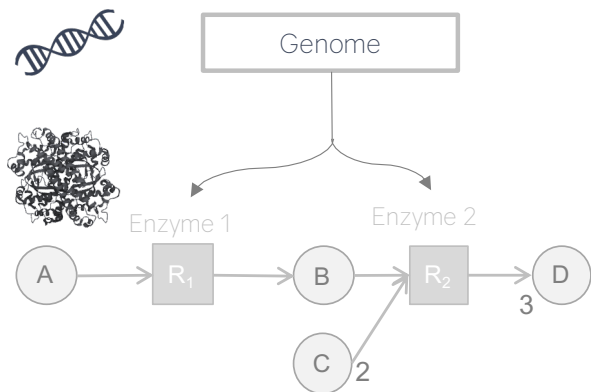


1 genome → 1,000 to 5,000 reactions

Genome-scale metabolic network (GSMN)

Functional repertoire of the cell

Metabolic networks and communities

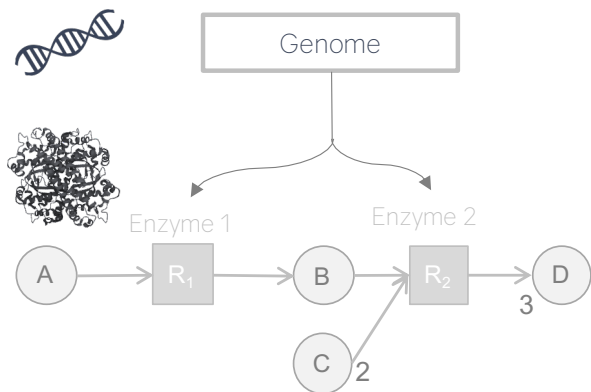


1 genome → 1,000 to 5,000 reactions

Genome-scale metabolic network (GSMN)

Functional repertoire of the cell

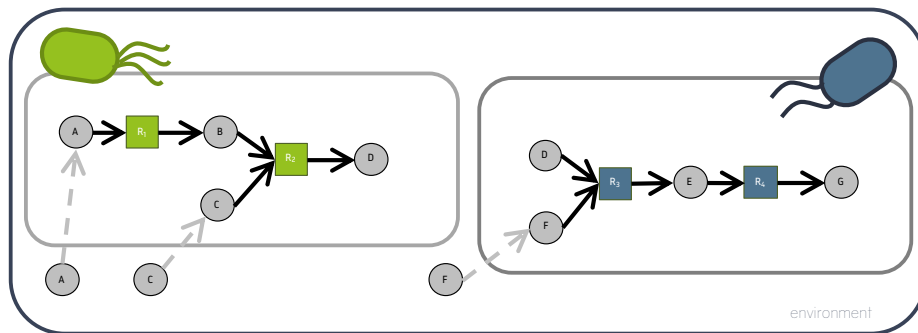
Metabolic networks and communities



1 genome → 1,000 to 5,000 reactions

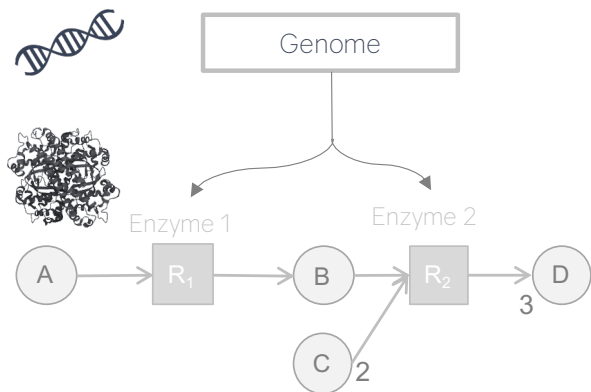
Genome-scale metabolic network (GSMN)

Functional repertoire of the cell



1. Reconstruct GSMNs for all genomes

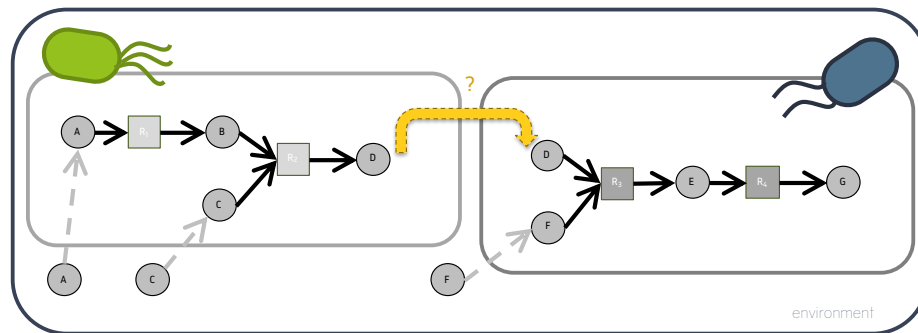
Metabolic networks and communities



1 genome \rightarrow 1,000 to 5,000 reactions

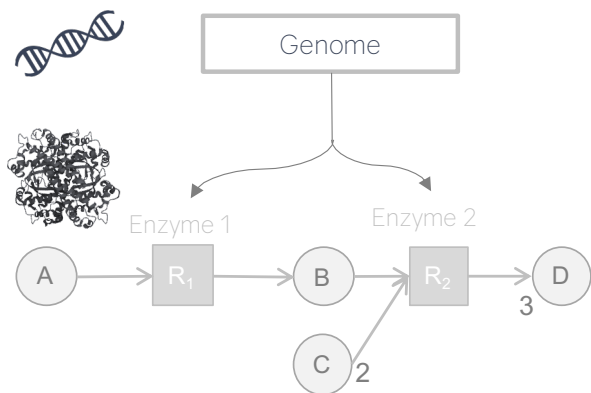
Genome-scale metabolic network (GSMN)

Functional repertoire of the cell



1. Reconstruct GSMNs for all genomes
2. Assess putative complementarity between them

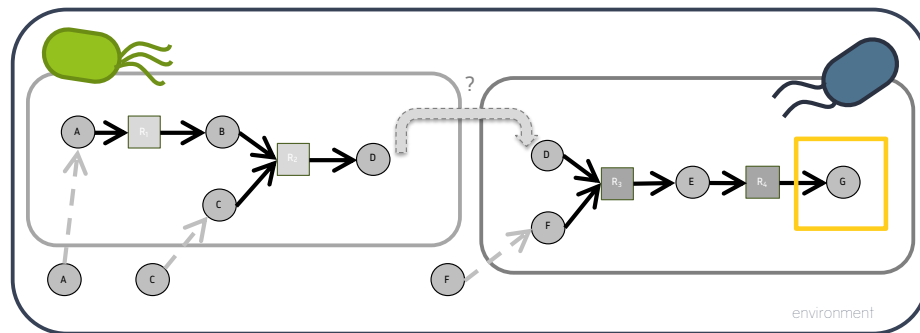
Metabolic networks and communities



1 genome → 1,000 to 5,000 reactions

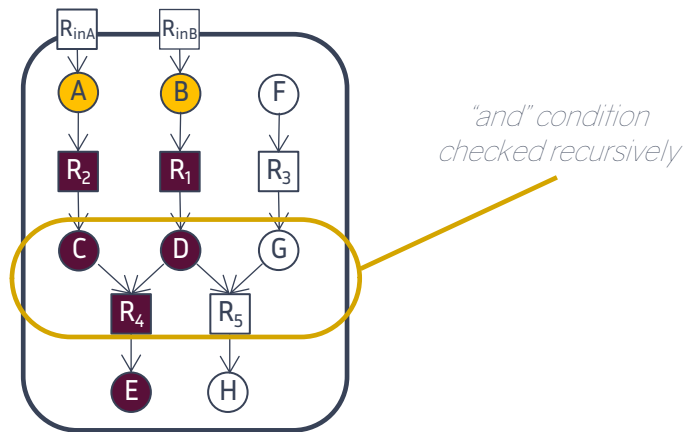
Genome-scale metabolic network (GSMN)

Functional repertoire of the cell



1. Reconstruct GSMNs for all genomes
2. Assess putative complementarity between them
3. Identify key species associated to a function

Metabolic modelling



Network expansion [Ebenhöh et al 2004]

- Scope of seeds
- Qualitative
- Implementation in Answer Set programming

$$\text{scope}(G, S) = \bigcup_i M_i,$$

where $M_0 = S$ and

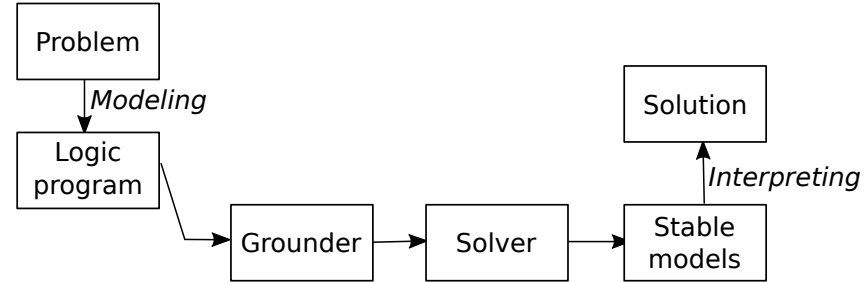
$$M_{i+1} = M_i \cup \text{products}(\{r \in R \mid \text{reactants}(r) \subseteq M_i\})$$

Robustness to incomplete data, similarity to prediction from qualitative simulations \rightarrow suitability to a metagenomic context

Answer Set Programming

[Gebser et al 2012]

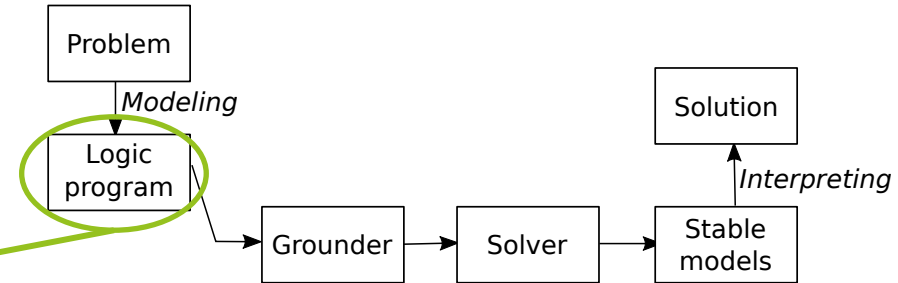
- Logic paradigm
- Knowledge representation & reasoning
- Optimisation



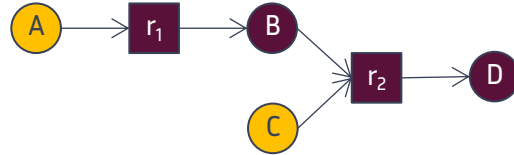
Answer Set Programming

[Gebser et al 2012]

- Logic paradigm
- Knowledge representation & reasoning
- Optimisation



```
reaction("r1").
reactant("A", "r1").
product("B", "r1").
%
reaction("r2").
reactant("B", "r1").
reactant("C", "r1").
product("D", "r1").
%
seed("A").
seed("C").
%
scope(M) :- seed(M).
scope(M) :- product(M,R); reaction(R); scope(N) : reactant(N,R).
```



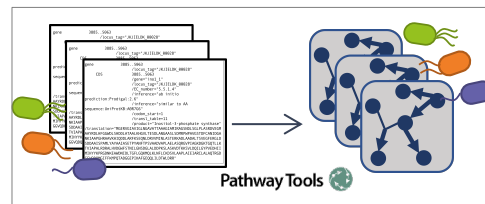
A pipeline for metabolic screening of communities

Microbiota

- MAGs
- Reference genomes



Automatic GSMN reconstruction



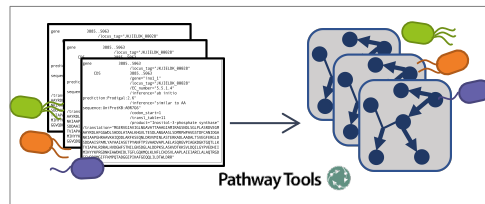
A pipeline for metabolic screening of communities

Microbiota

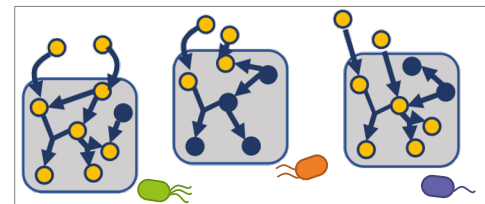
- MAGs
- Reference genomes



Automatic GSMN reconstruction



Individual metabolic capabilities

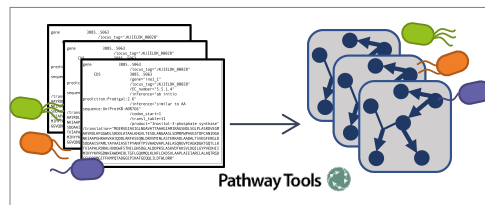


A pipeline for metabolic screening of communities

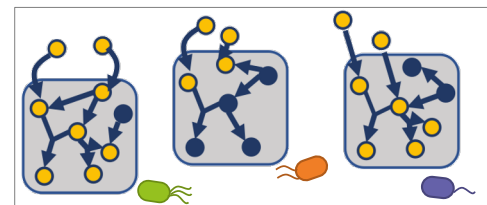
Microbiota

- MAGs
- Reference genomes

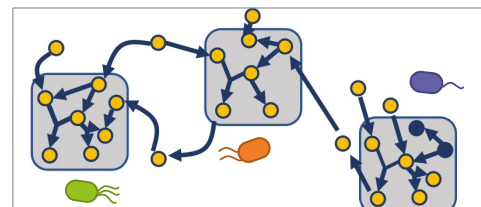
Automatic GSMN reconstruction



Individual metabolic capabilities



Collective metabolic capabilities

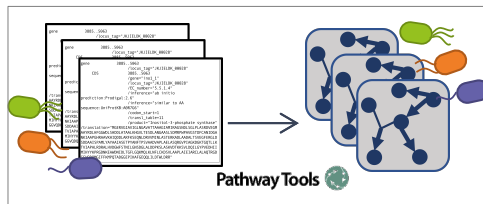


A pipeline for metabolic screening of communities

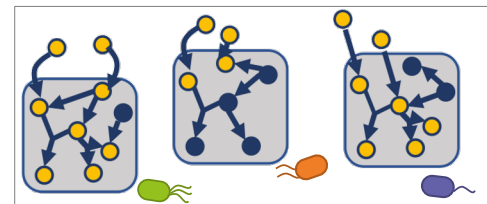
Microbiota

- MAGs
- Reference genomes

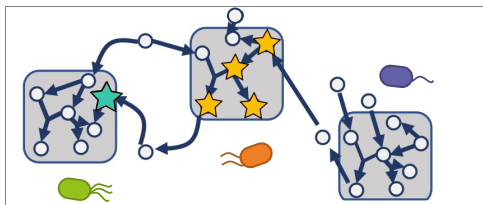
Automatic GSMN reconstruction



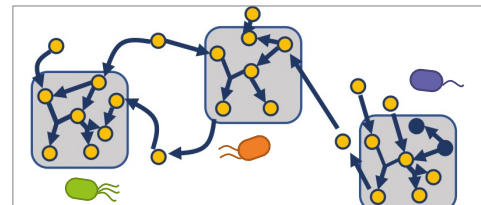
Individual metabolic capabilities



Added value of cross feeding



Collective metabolic capabilities

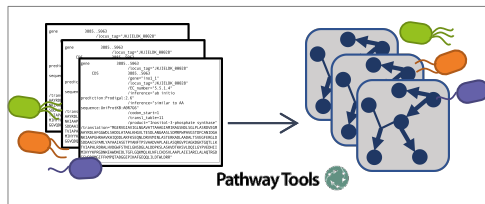


A pipeline for metabolic screening of communities

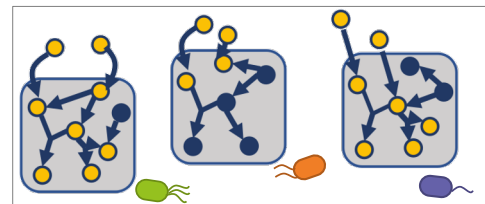
Microbiota

- MAGs
- Reference genomes

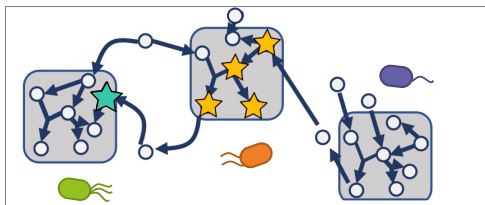
Automatic GSMN reconstruction



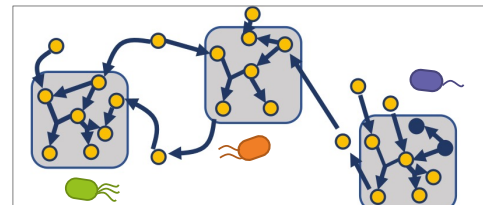
Individual metabolic capabilities



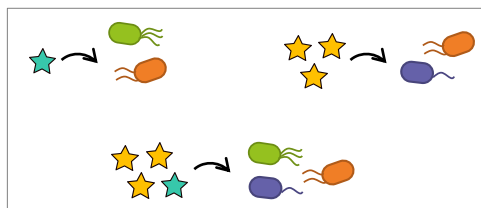
Added value of cross feeding



Collective metabolic capabilities



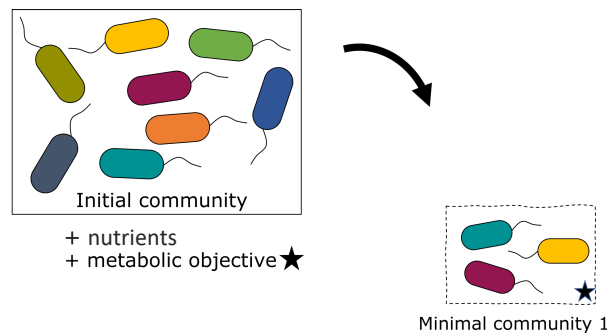
Minimal community selection



Systematic screening of **metabolic potential** and **mutualistic potential** in a microbiota

Key species

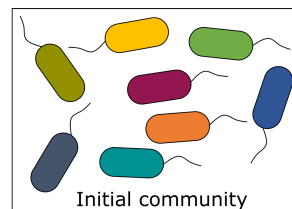
- Minimal community selection:
combinatorial problem solving



The concept of **key species** addresses the **combinatorial** challenge brought by **functional redundancy** in microbiomes

Key species

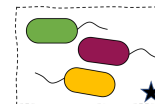
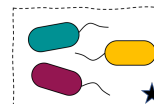
- Minimal community selection:
combinatorial problem solving
- However... 1 solution (= 1 minimal community) among
millions? **Huge combinatorics**



+ nutrients
+ metabolic objective★

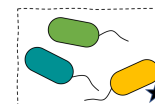
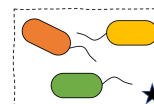


All minimal communities satisfying
the metabolic objective



Minimal community 1

Minimal community 2



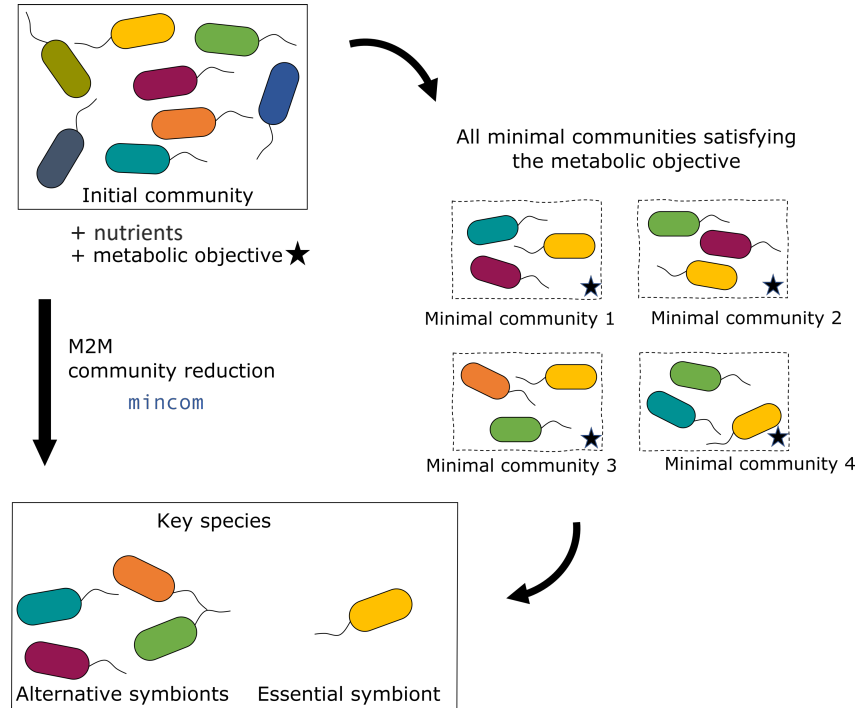
Minimal community 3

Minimal community 4

The concept of **key species** addresses the **combinatorial** challenge brought by **functional redundancy** in microbiomes

Key species

- Minimal community selection:
combinatorial problem solving
- However... 1 solution (= 1 minimal community) among millions? **Huge combinatorics**
- *Key species* = species found in at least one of the predicted minimal communities (MC)
 - > *Essential symbionts* : species predicted in **every** MC
 - > *Alternative symbionts* : species predicted in some but not all MC



The concept of **key species** addresses the **combinatorial** challenge brought by **functional redundancy** in microbiomes

Application to 1,520 reference genomes of cultivable species of the human gut microbiota

	Gut dataset
initial data	draft reference genomes
number of genomes	1,520
<i>GSMN reconstruction</i>	
all reactions	3,932
all metabolites	4,001
avg reactions per GSMN	1,144 (\pm 255)
avg metabolites per GSMN	1,366 (\pm 262)
avg genes per mn	596 (\pm 150)
% reactions associated to genes	74.6 (\pm 2.17)
avg pathways per mn	163 (\pm 49)
<i>metabolic potential</i>	
number of seeds	93
avg scope per mn	286 (\pm 70)
union of individual scopes	828

156 metabolites that cannot be produced by individual species

Classification into categories: lipids (28), carbohydrate derivatives (58)...

Key species computation for categories of metabolites + minimal communities enumeration

Analysis and visualisation

Key species by groups of metabolic end-products

		Firm.	Bact.	Acti.	Prot.	Fuso.	total
aminoacids and derivatives (5 targets)	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
aromatic compounds (11 targets)	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
carboxyacids (14 targets)	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
coA derivatives (10 targets)	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
lipids (28 targets)	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
sugar derivatives (58 targets)	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

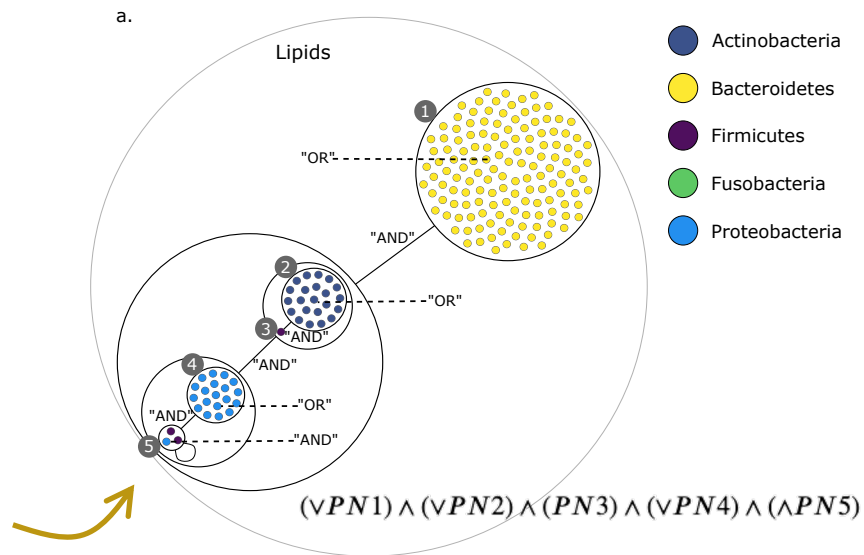
Enumeration of all minimal communities



How do bacteria associate within them?
Any patterns?

Key species by groups of metabolic end-products

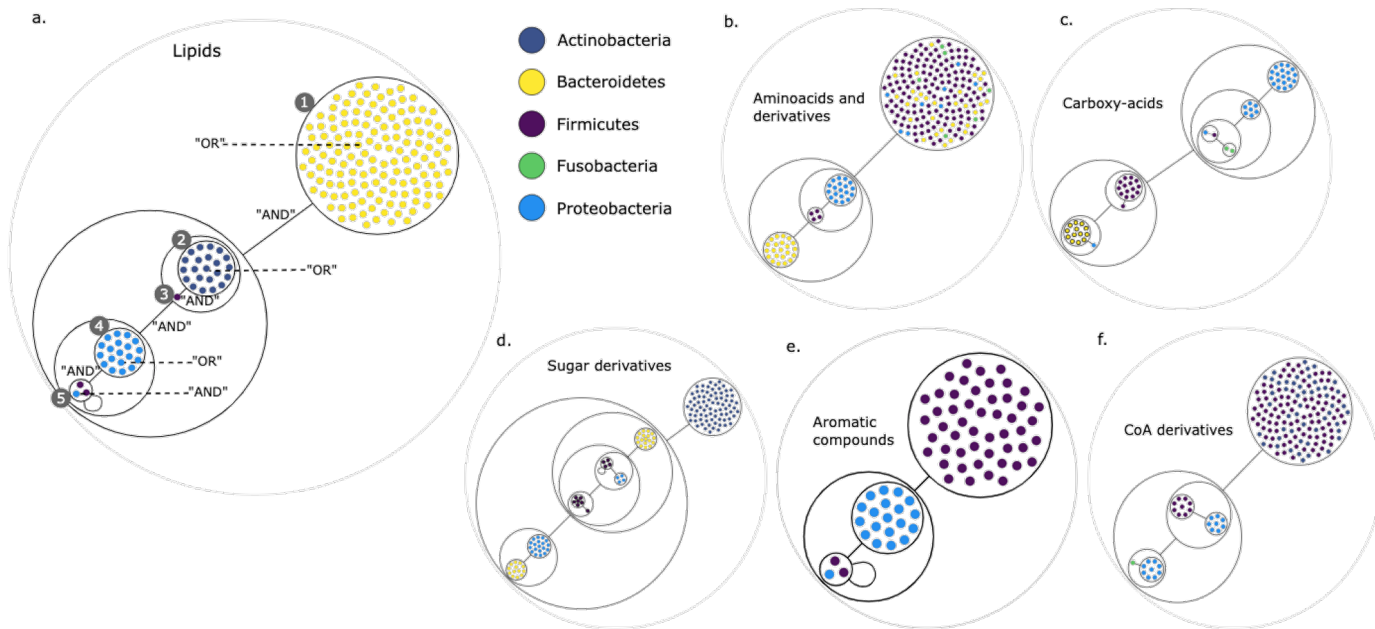
		Firm.	Bact.	Acti.	Prot.	Fuso.	total
aminoacids and derivatives (5 targets)	KS	142	52	0	27	6	227
4 bact. per community	ES	0	0	0	0	0	0
120,329 communities	AS	142	52	0	27	6	227
aromatic compounds (11 targets)	KS	52	0	0	20	0	72
5 bact. per community	ES	2	0	0	1	0	3
950 communities	AS	50	0	0	19	0	69
carboxyacids (14 targets)	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
coA derivatives (10 targets)	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
lipids (28 targets)	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
sugar derivatives (58 targets)	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



58,520 minimal communities associated to the producibility of the 28 lipids can be visualised as a compressed association graph.

→ Species with equivalent roles in the communities

Same “reading pattern” for all groups



Essential symbionts (ES) shared between several groups of targets
 → species with unique functions among the 1,520?
 4/5 ES for carbohydrates are studied as probiotics

MAGs vs reference genomes

Metagenome-assembled genomes (MAGs)

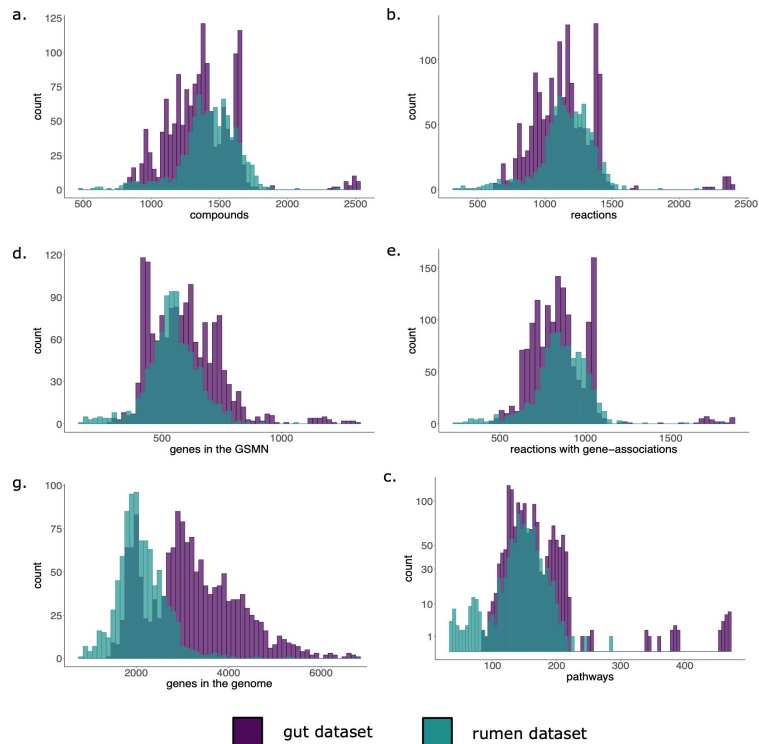
- > Possibly incomplete
- > Possibly contaminated

Experimental setup

- 913 cow rumen MAGs [Stewart et al 2018]
- Random degradation
 - > Removal of 2% genes in all MAGs
 - > Removal of 5% genes in 80% of MAGs
 - > Removal of 5% genes in all MAGs
 - > Removal of 10% genes in 70% of MAGS
- Reconstruct & compare GSMNs, apply M2M screening and community selection

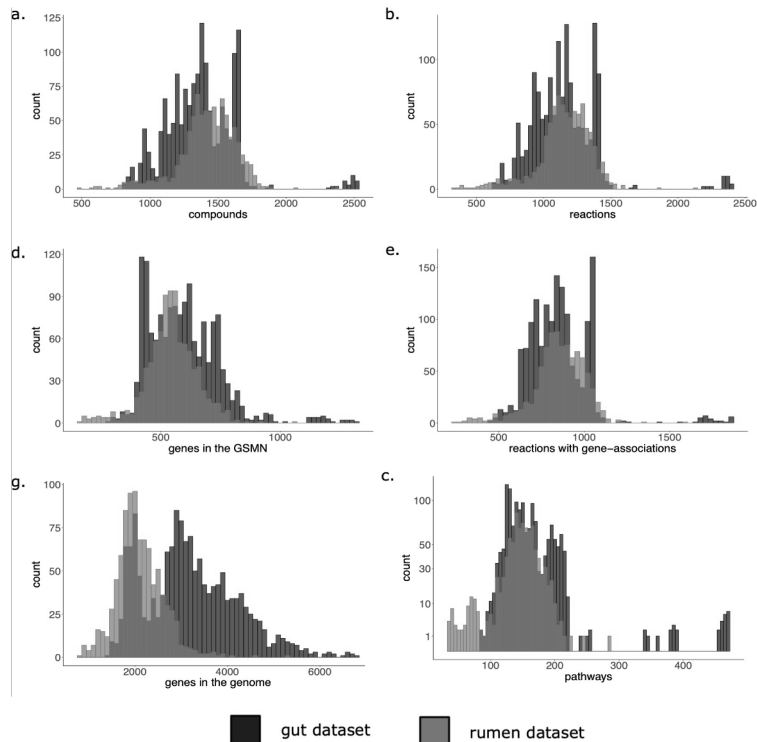
It's ok to use MAGs

GSMN characteristics are similar for reference genomes and MAGs

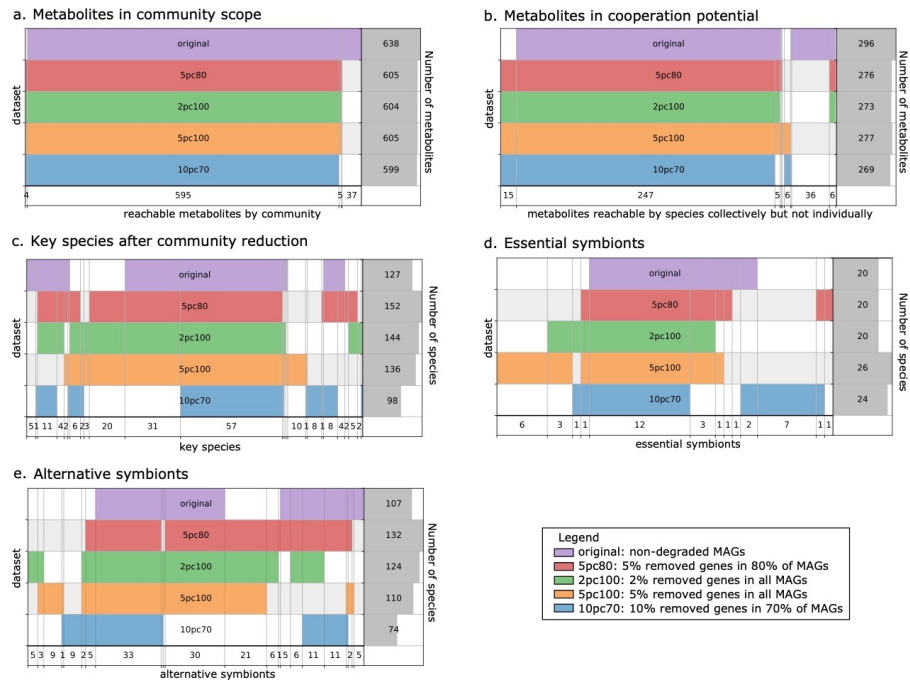


It's ok to use MAGs

GSMN characteristics are similar for reference genomes and MAGs



Modelling predictions are stable despite MAG degradation



Arnaud Belcour^{1†}, Clémence Frioux^{2,1,3,4†}, Méziane Aite¹, Anthony Bretaudeau^{1,5,6}, Falk Hildebrand^{3,4}, Anne Siegel¹

¹Univ Rennes, Inria, CNRS, IRISA, F-35000 Rennes, France; ²Inria Bordeaux Sud-Ouest, France; ³Gut Microbes and Health, Quadram Institute, NR4 7GJ, Norwich, United Kingdom; ⁴Digital Biology, Earlham Institute, NR4 7GJ, Norwich, United Kingdom; ⁵INRAE, UMR IGEP, Bioinformatics Platform for Agroecosystems Arthropods (BIPAA), 35042 Rennes, France; ⁶Inria, IRISA, GenOuest Core Facility, 35042 Rennes, France

A flexible and scalable pipeline

Metage2Metabo

- > Global analysis of metabolism in microbiomes: redundancy - complementarity
- > Key species screening: complexity reduction
- > Robust modelling framework that supports data incompleteness

Use cases:

- > Comparison of metabolic repertoire in genome collections
- > Comparison of cohort samples through their MAGs composition



[aureme/metage2metabo](https://github.com/aureme/metage2metabo)



- Arnaud Belcour
- Méziane Aïte
- Anne Siegel



- Falk Hildebrand
- NBI Computing
Infrastructure for Science



- Pleiade project team
<https://inria.fr/en/pleiade>



- Simon Dittami



- Anthony Bretaudeau



We're hiring!

Postdoc position - 2 years - at Inria Bordeaux
systems biology – metabolism
gut microbiota dynamics – logic modelling
<https://tinyurl.com/postdoc-sysbio-inria>

