

Inria

Logic modelling of metabolism: from individual networks to communities

Clémence Frioux

Workshop Modélisation du métabolisme - INRAE

November 19, 2021

From network composition to constraint-based modelling



```
> protein_1
ATGGAATCGCATAGCA...
> protein_2
AATGCCTGATCGAATA...
```

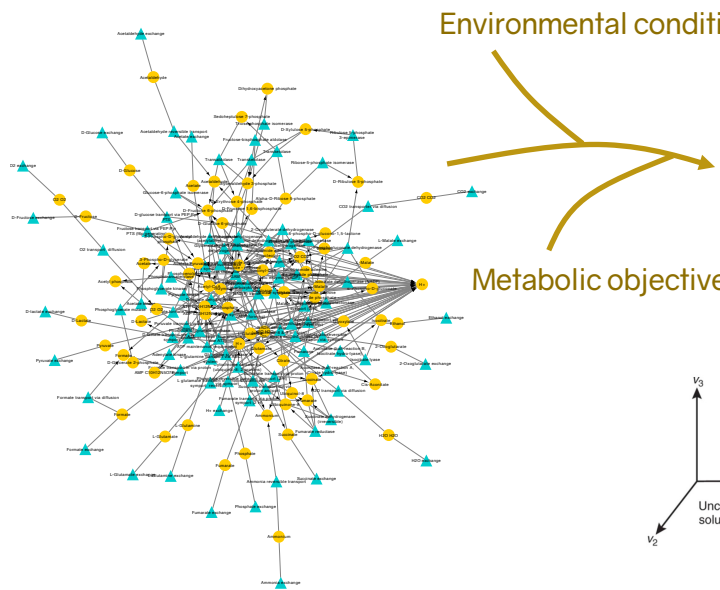
Functional annotation



protein	name	EC-number	GO	KEGG	...
protein_1	pfkA	2.7.1.11	-	K00850	...
protein_2	eno	4.2.1.11	-	K01689	...
...

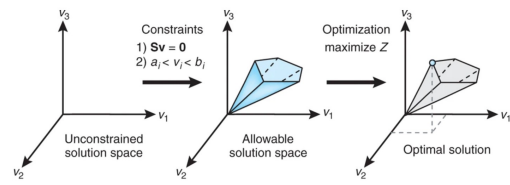
Metabolic network reconstruction & curation

Function	
Carbohydrate degradation	■
Lipid synthesis	■
Vitamin biosynthesis	■
Antibiotic resistance	■
...	...



Flux analysis (FBA, FVA...)
Constraint-based modelling

$$S = \begin{pmatrix} \tau_1 & \dots & \tau_n \\ +1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ -1 & \dots & +1 \end{pmatrix} \begin{matrix} M_1 \\ \dots \\ M_n \end{matrix}$$

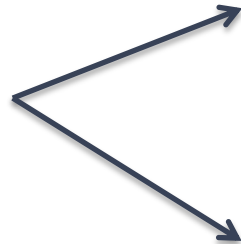


Metabolic constraint-based modelling necessitates curated metabolic networks



A rise in non-optimal use-cases

Cheaper sequencing
Metagenomics



Non-model organisms

- Limited experimentations
- Unknown biomass composition



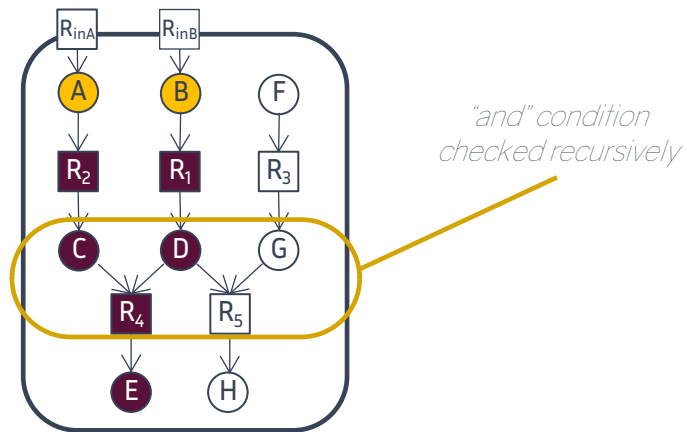
Large-scale community modelling

- Impossible manual curation
- Uncertainty on/missing functions
- Scalability of optimisations-based simulation



Needed alternatives for metabolic screening,
complexity reduction, hypothesis generation

Boolean abstraction of metabolic producibility



Network expansion [Ebenhöh et al 2004]

- Scope of seeds
- Discrete modelling

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

where $M_0 = S$ and

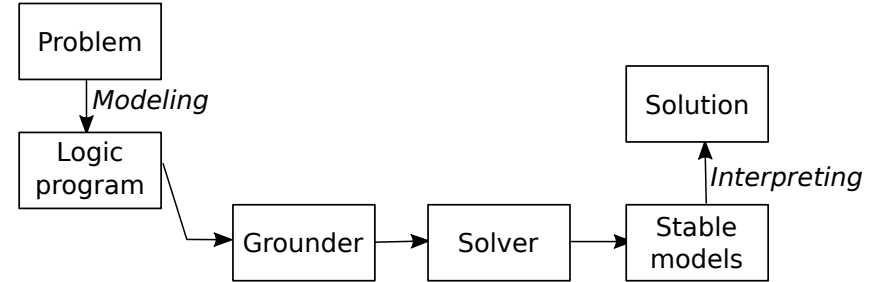
$$M_{i+1} = M_i \cup \text{products}(\{r \in R \mid \text{reactants}(r) \subseteq M_i\})$$

Qualitative simulation of metabolic producibility ignoring stoichiometric coefficients and w/o objective function

A flexible implementation system

[Gebser et al 2012]

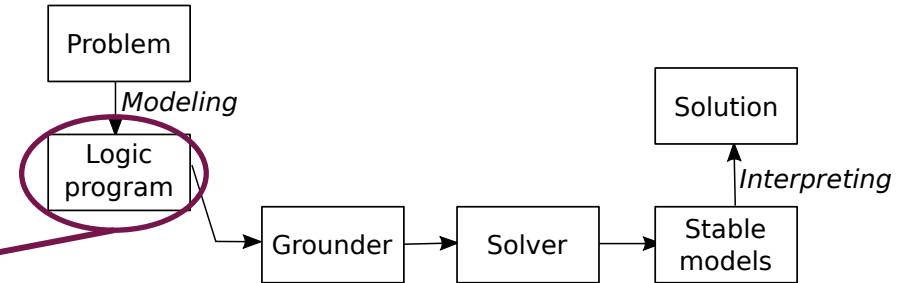
- Answer Set Programming
- Logic paradigm
- Knowledge representation & reasoning
- Optimisation



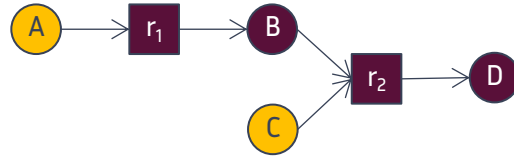
A flexible implementation system

[Gebser et al 2012]

- Answer Set Programming
- Logic paradigm
- Knowledge representation & reasoning
- Optimisation



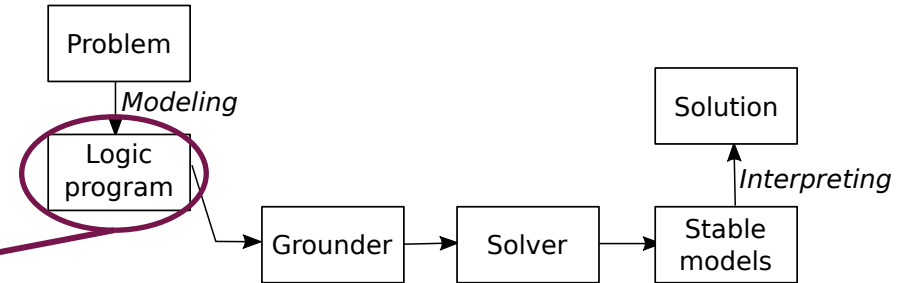
```
reaction("r1").
reactant("A", "r1").
product("B", "r1").
%
reaction("r2").
reactant("B", "r1").
reactant("C", "r1").
product("D", "r1").
%
seed("A").
seed("C").
%
scope(M) :- seed(M).
scope(M) :- product(M,R); reaction(R); scope(N) : reactant(N,R).
```



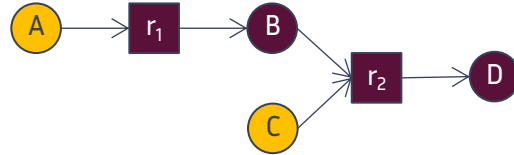
A flexible implementation system

[Gebser et al 2012]

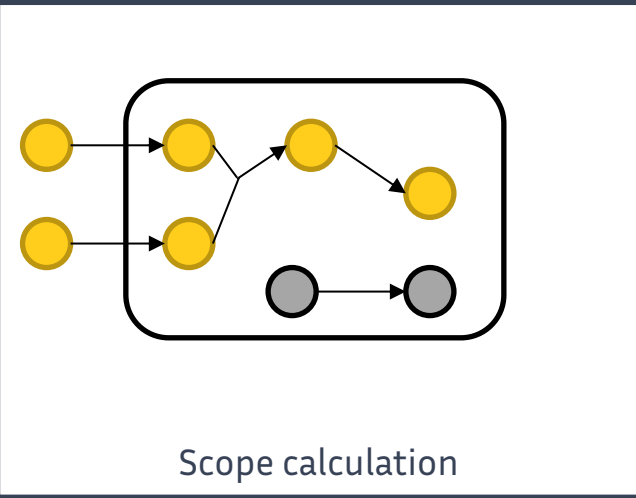
- Answer Set Programming
- Logic paradigm
- Knowledge representation & reasoning
- Optimisation

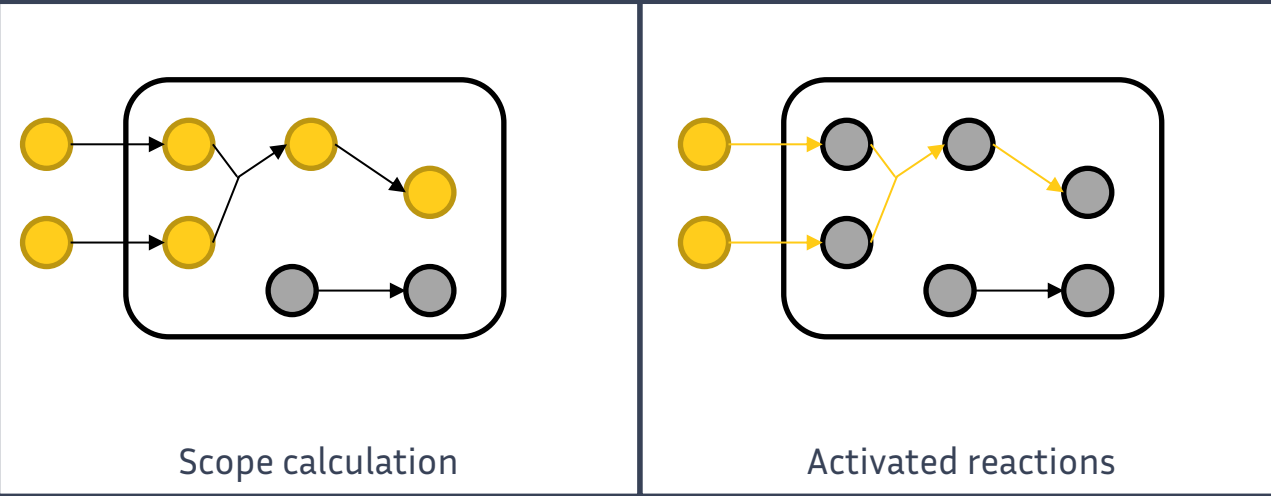


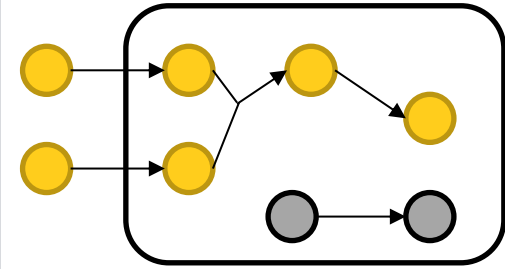
```
reaction("r1").
reactant("A", "r1").
product("B", "r1").
%
reaction("r2").
reactant("B", "r1").
reactant("C", "r1").
product("D", "r1").
%
seed("A").
seed("C").
%
scope(M) :- seed(M).
scope(M) :- product(M,R); reaction(R); scope(N) : reactant(N,R).
```



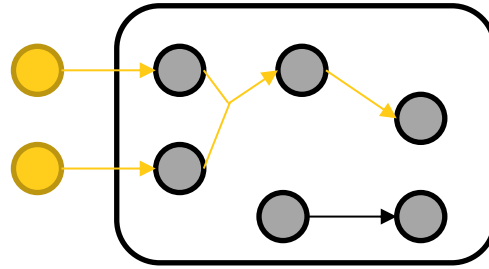
- Existing variation for pseudo-state modelling [Thuillier et al 2021]
- Building block for many applications



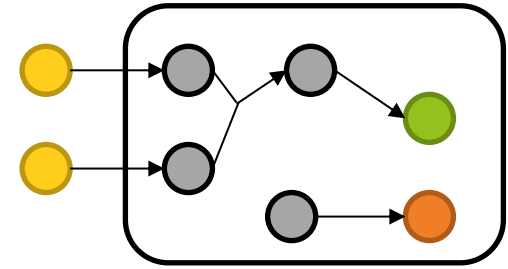




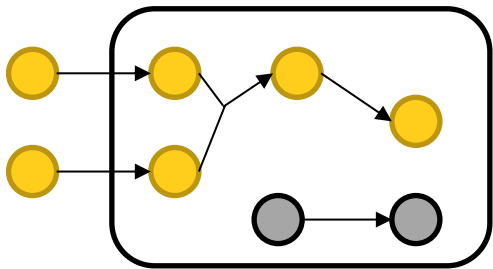
Scope calculation



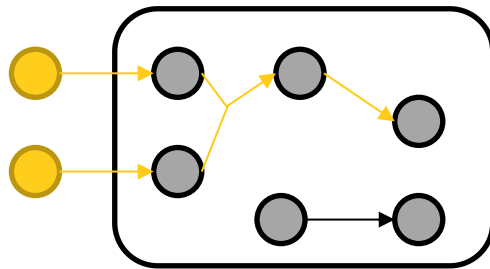
Activated reactions



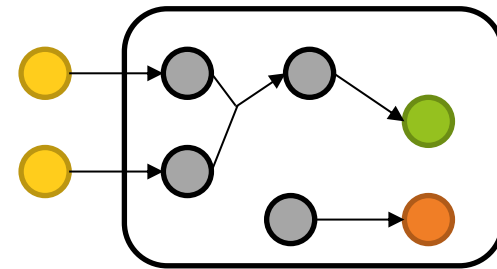
Producibility checking



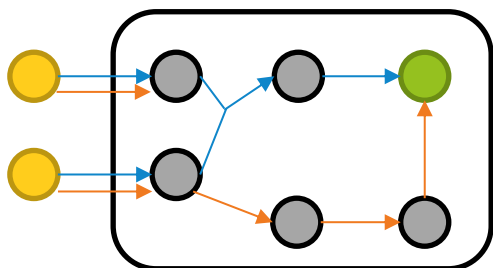
Scope calculation



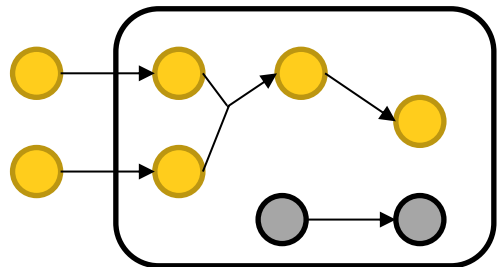
Activated reactions



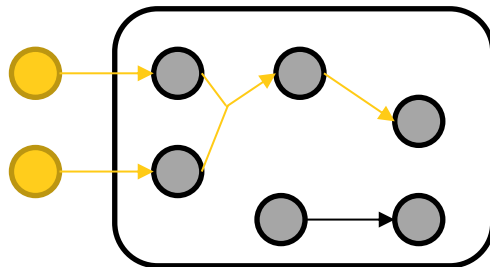
Producibility checking



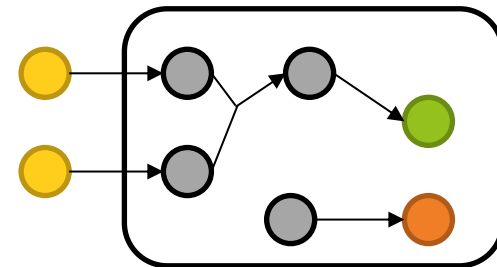
Pathway detection



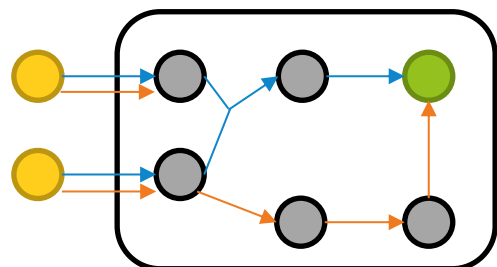
Scope calculation



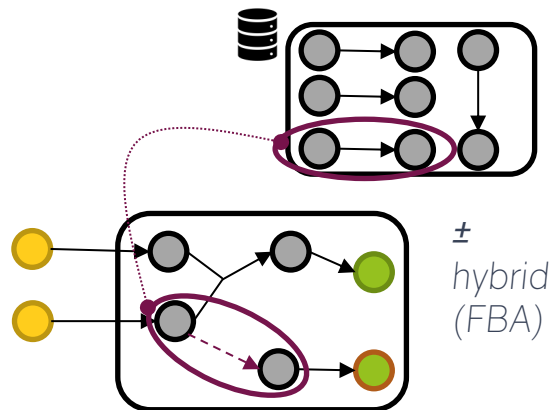
Activated reactions



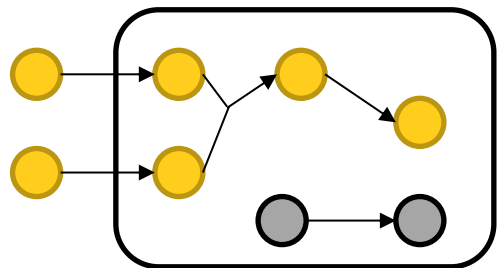
Producibility checking



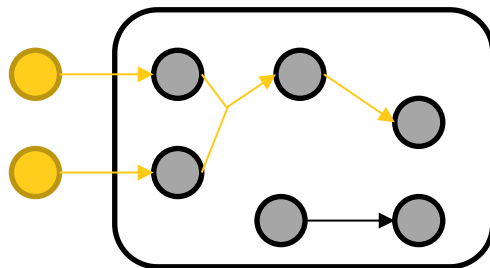
Pathway detection



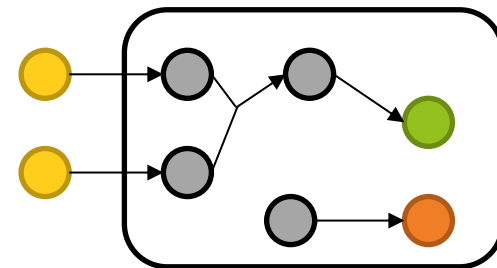
Gap-filling



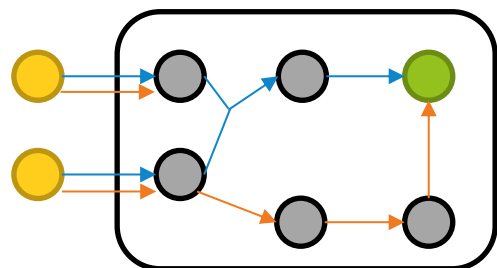
Scope calculation



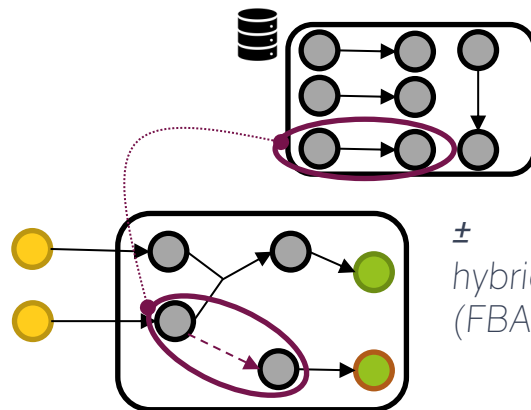
Activated reactions



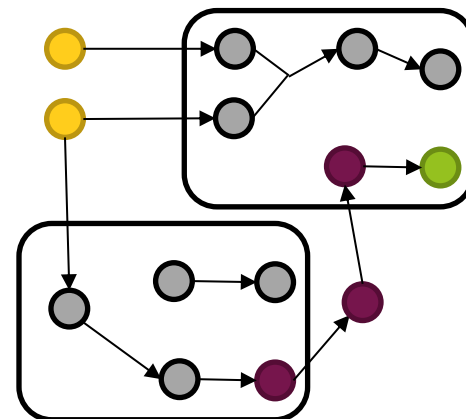
Producibility checking



Pathway detection

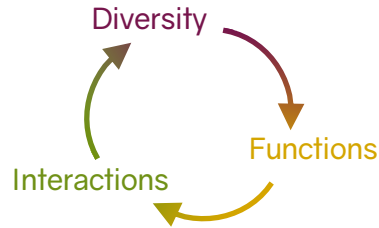
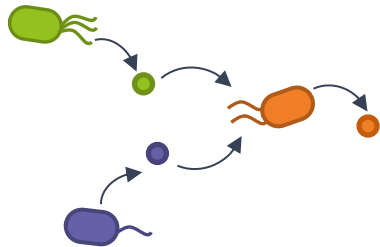


Gap-filling

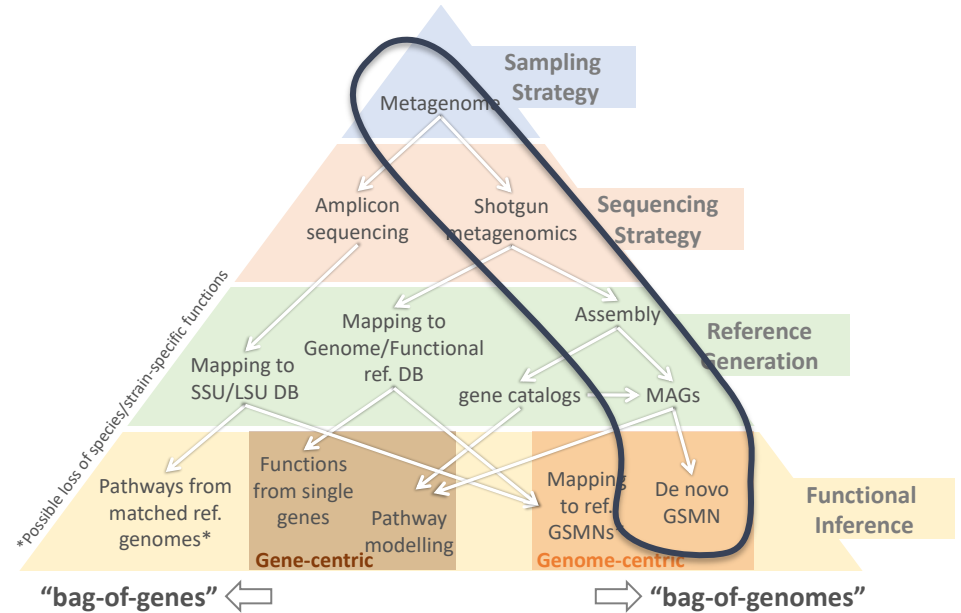


Community analysis

Metagenomics meets metabolism



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



[Frioux et al. CSBJ 2020]

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

Metabolic networks and communities

Shotgun metagenomic sequencing



Metagenome-Assembled Genomes
(MAGs)

Metabolic networks and communities

Shotgun metagenomic sequencing



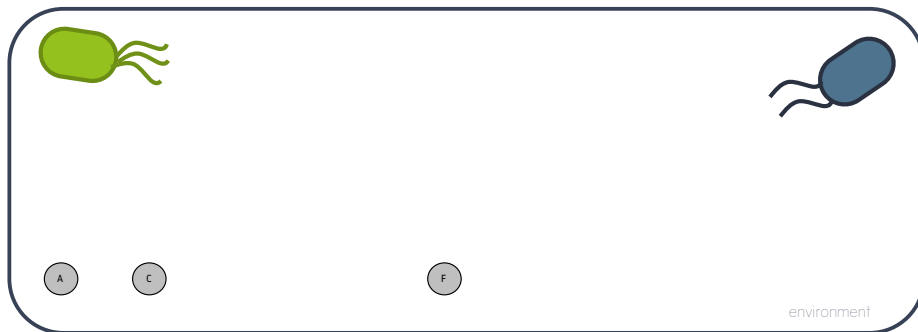
Metagenome-Assembled Genomes
(MAGs)



One sample → One set of MAGs



What can we learn about the community?



Metabolic networks and communities

Shotgun metagenomic sequencing



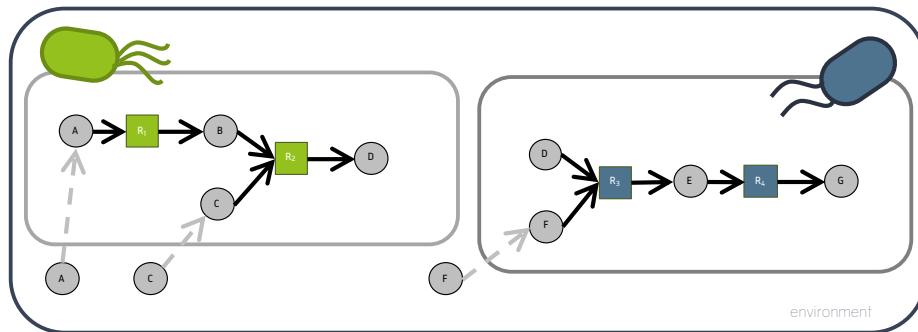
Metagenome-Assembled Genomes (MAGs)



One sample → One set of MAGs
→ One set of metabolic networks



What can we learn about the community?



1. Reconstruct GSMNs for all genomes

Metabolic networks and communities

Shotgun metagenomic sequencing



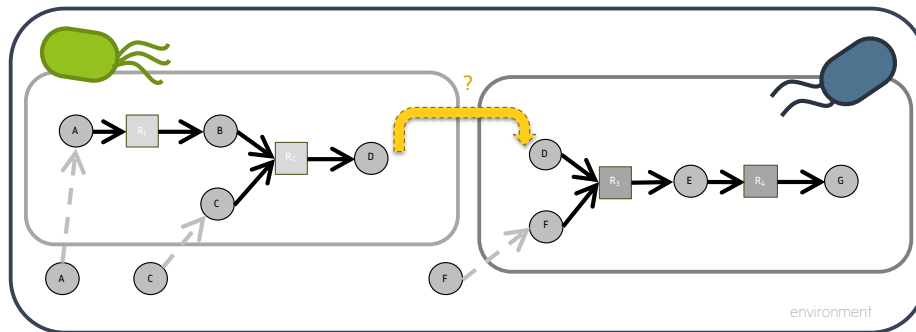
Metagenome-Assembled Genomes (MAGs)



One sample → One set of MAGs
→ One set of metabolic networks



What can we learn about the community?



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

Metabolic networks and communities

Shotgun metagenomic sequencing



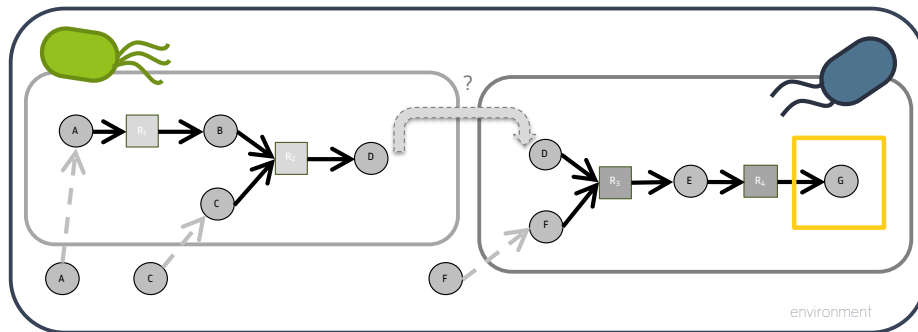
Metagenome-Assembled Genomes (MAGs)



One sample → One set of MAGs
→ One set of metabolic networks



What can we learn about the community?



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

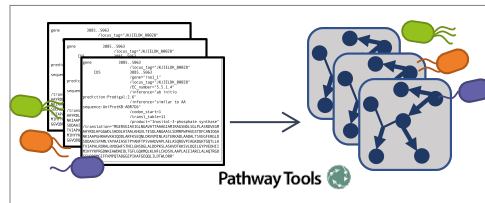
A pipeline for metabolic screening of communities

Microbiota

- MAGs
- Reference genomes



Automatic GSMN reconstruction



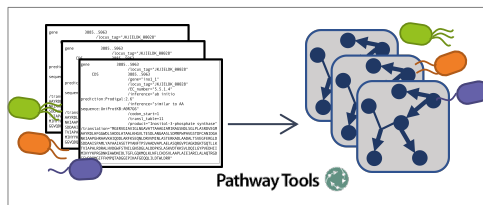
A pipeline for metabolic screening of communities

Microbiota

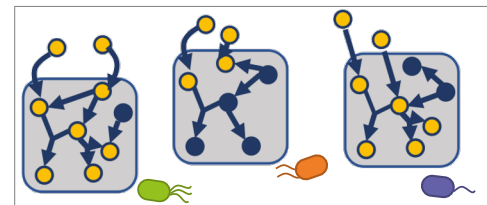
- MAGs
- Reference genomes



Automatic GSMN reconstruction



Individual metabolic capabilities



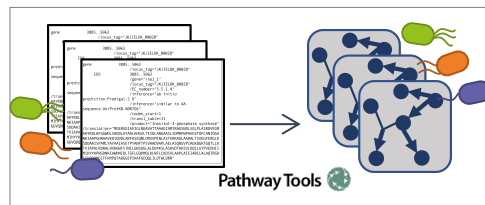
A pipeline for metabolic screening of communities

Microbiota

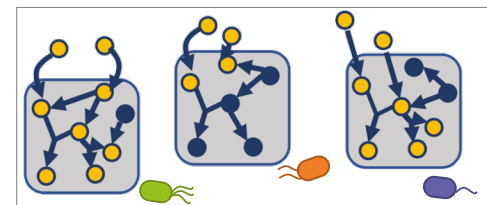
- MAGs
- Reference genomes



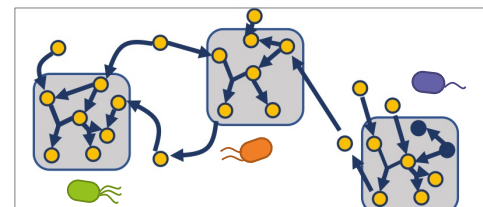
Automatic GSMN reconstruction



Individual metabolic capabilities



Collective metabolic capabilities

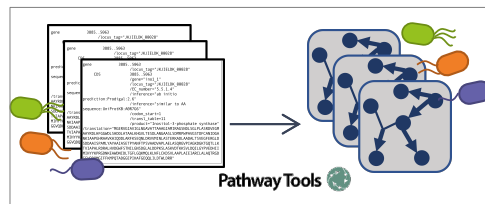


A pipeline for metabolic screening of communities

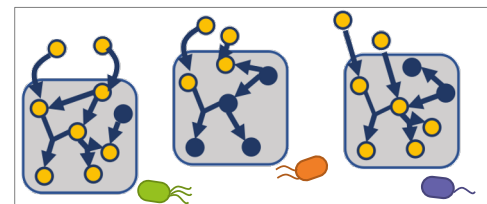
Microbiota

- MAGs
- Reference genomes

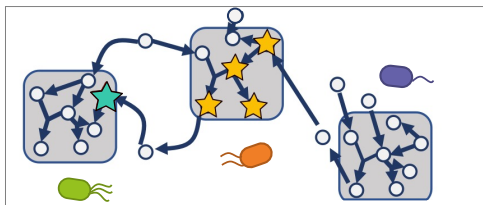
Automatic GSMN reconstruction



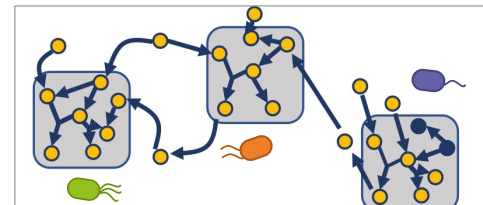
Individual metabolic capabilities



Added value of cross feeding



Collective metabolic capabilities

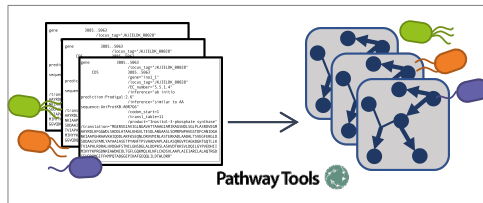


A pipeline for metabolic screening of communities

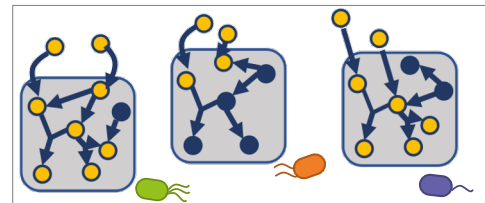
Microbiota

- MAGs
- Reference genomes

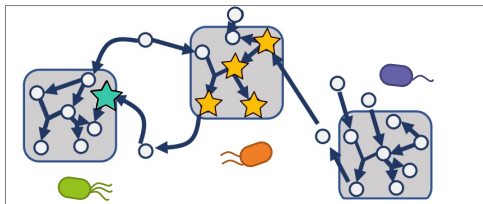
Automatic GSMN reconstruction



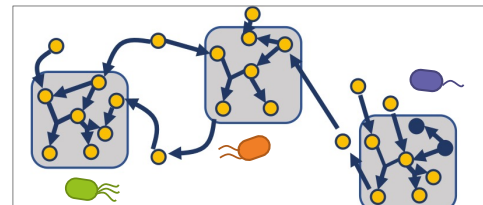
Individual metabolic capabilities



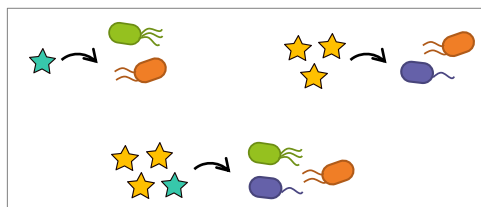
Added value of cross feeding



Collective metabolic capabilities



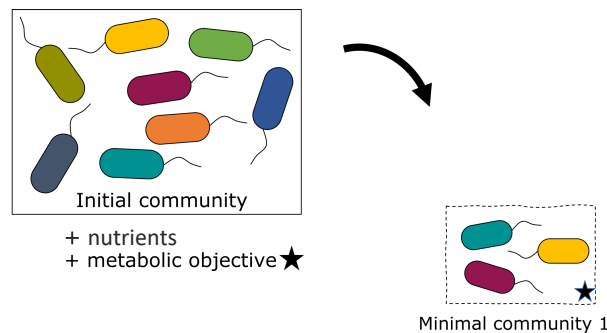
Minimal community selection



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

Key species

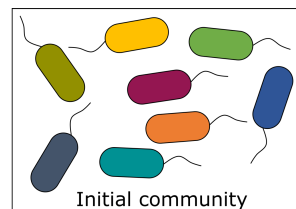
- Minimal community selection:
combinatorial problem solving



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

Key species

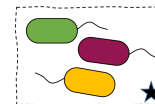
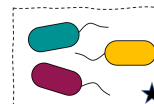
- Minimal community selection:
combinatorial problem solving
- However... 1 solution (= 1 minimal community) among
millions? **Huge combinatorics**



+ nutrients
+ metabolic objective★

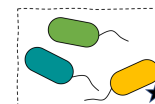
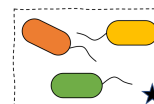


All minimal communities satisfying
the metabolic objective



Minimal community 1

Minimal community 2



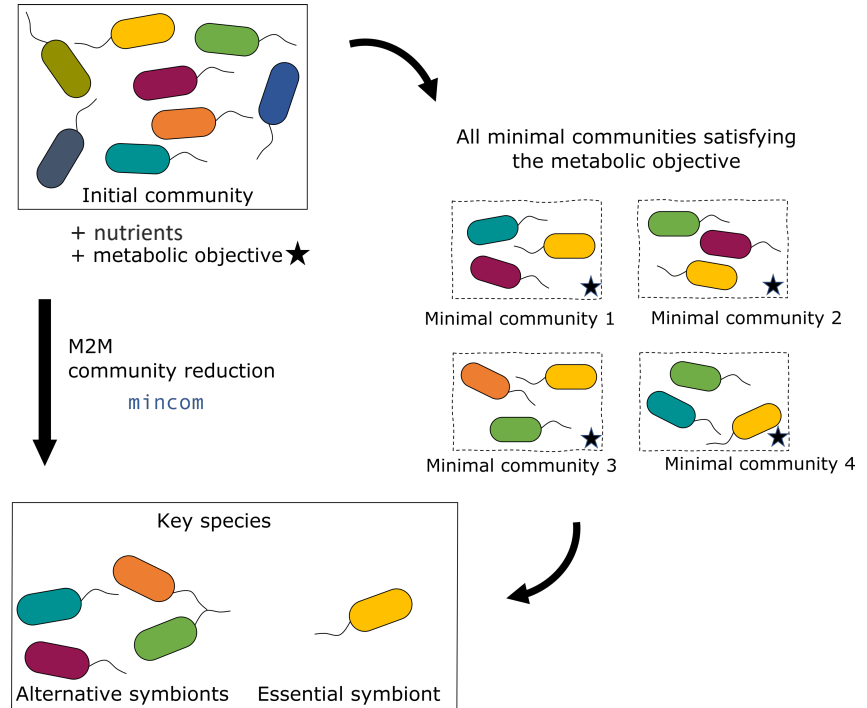
Minimal community 3

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

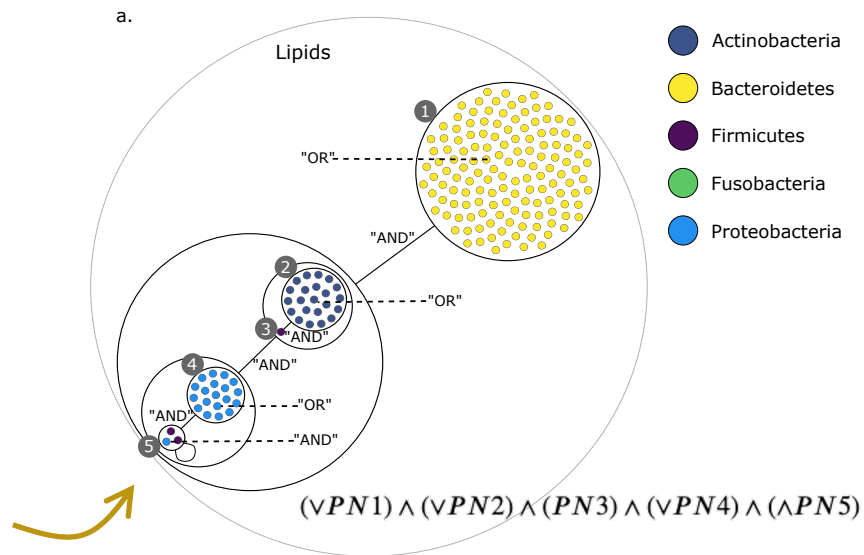
Enumeration of all minimal communities



How do bacteria associate within them?
Any patterns?

Key species by groups of metabolic end-products

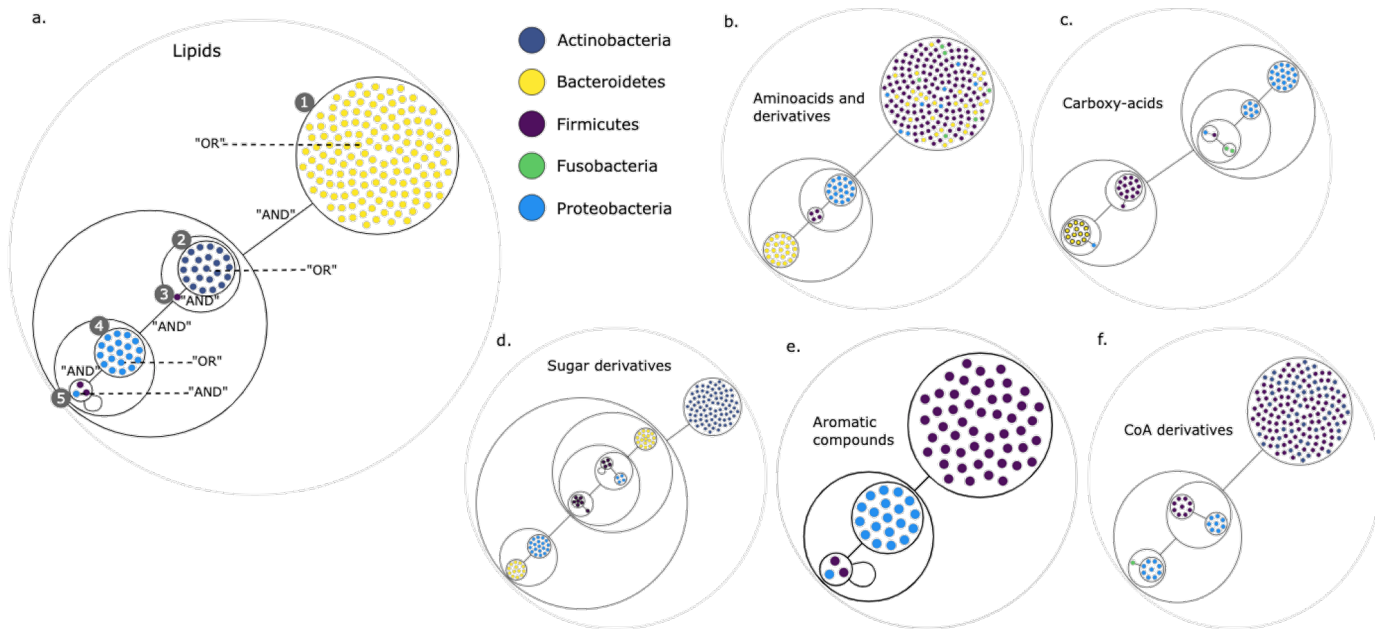
		Firm.	Bact.	Acti.	Prot.	Fuso.	total
aminoacids and derivatives (5 targets)	KS	142	52	0	27	6	227
4 bact. per community	ES	0	0	0	0	0	0
120,329 communities	AS	142	52	0	27	6	227
aromatic compounds (11 targets)	KS	52	0	0	20	0	72
5 bact. per community	ES	2	0	0	1	0	3
950 communities	AS	50	0	0	19	0	69
carboxyacids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

Take home messages

Boolean abstraction of metabolic producibility

- > Modelling framework complementary to constraint-based modelling

An integrative workflow - Metage2Metabo

- > Global analysis of metabolism: redundancy - complementarity
- > Key species screening: complexity reduction
- > Suitability to MAGs incompleteness

Use cases for microbiome functional analyses:

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



aureme/metage2metabo
cfrioux/MeneTools
bioasp/meneco



- 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!

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

