Hybrid approach for explainable metabolic modelling of microbial ecosystems

Approche hybride de modélisation explicable du métabolisme des écosystèmes microbiens

PhD defense of Maxime LECOMTE

December 11, 2023

Jury members

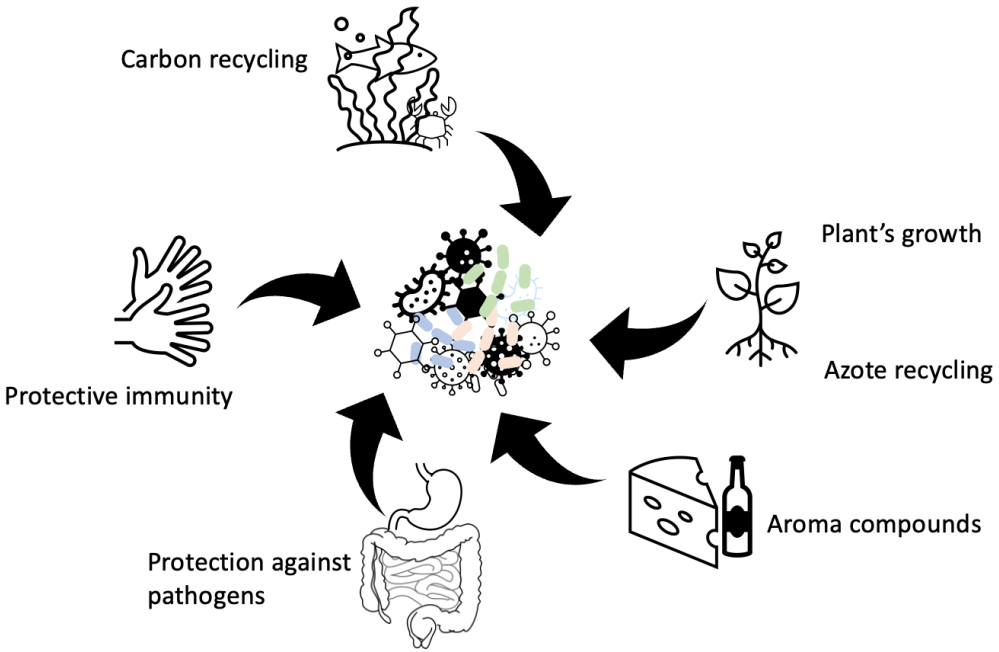
Reviewers: BAROUKH Caroline
COCAIGN-BOUSQUET Muriel

Examinators: COTTRET Ludovic
LAROCHE Béatrice
MARKOV Gabriel
SIMON Laurent

Supervised by : David SHERMAN, Hélène FALENTIN and Clémence FRIOUX

École doctorale Mathématiques et informatique

Microbial ecosystems are complex



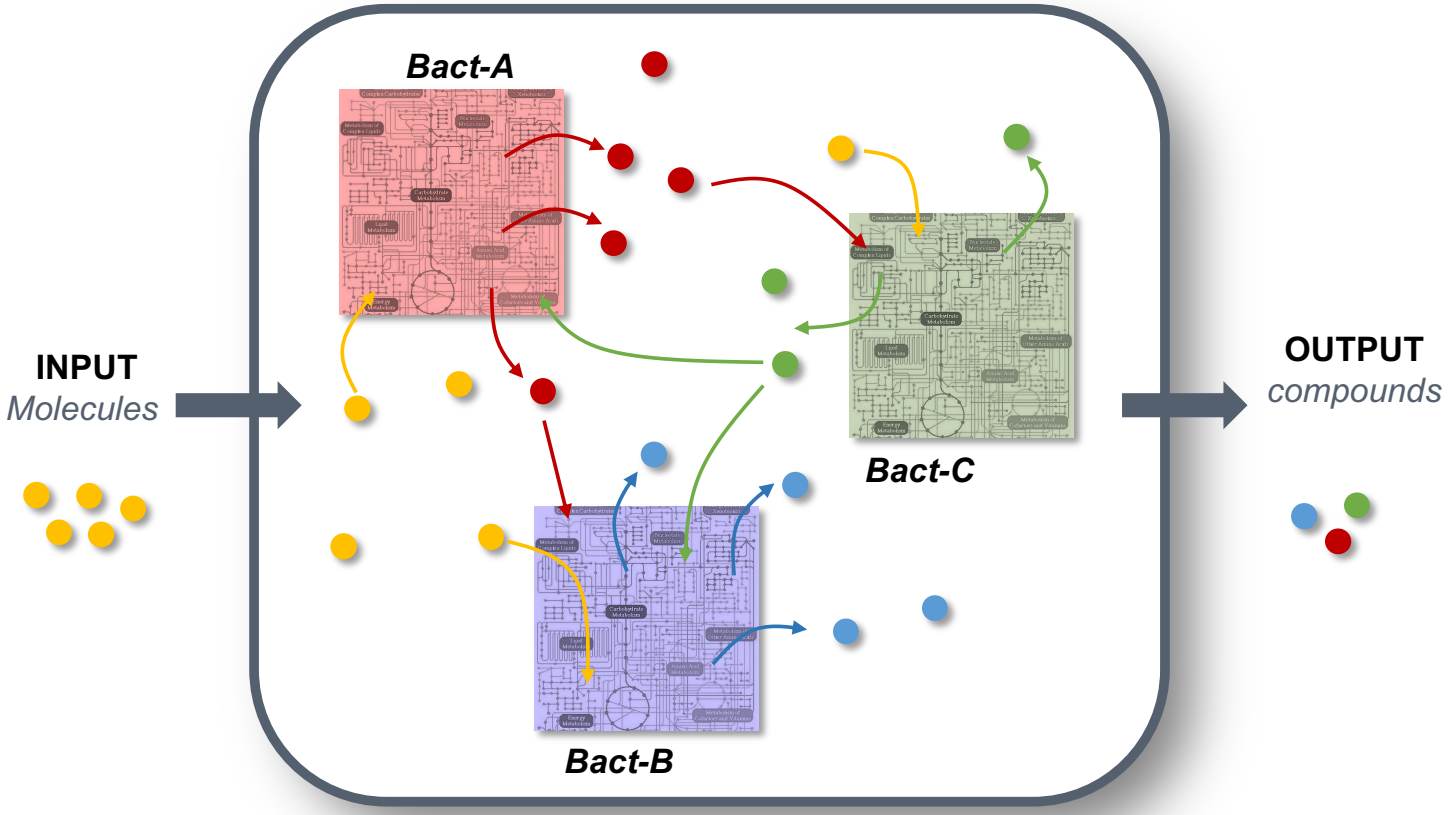
In the environment

- High taxonomic diversity
- High functional diversity
- Complex bacterial interactions

Computational methods are required to tackle the complexity of microbial ecosystems

Royet and Plailly, 2004; Belkaid and Hand, 2014; Zhang et al., 2015; Hoorman, 2011; McSweeney and Sousa, 2000.

Bacterial interaction process



- The direction of the exchange, species involved, ecosystem output

Mechanistic models of bacterial interactions

Modeling as a computational method to study microbial communities

Modeling

Mathematical object used to simulate or predict a function

	Numerical (quantitative)	Discrete (qualitative)
Mathematical formalism	ODE or constraints	Knowledge base and logical rules
Solver	Numeric	Logic
Data	Multi-omics data (genomic, transcriptomics, metabolomics, growth measures...)	Genomes annotations
Scope	Controlled ecosystems	Screening bacterial communities
Computational object	Metabolic networks	Metabolic networks

- Mathematical or computational formalism choice is determined by the issue
- Metabolic modeling for understanding complex ecosystems

Objective and contributions

Objective

Contribute to analyzing metabolic interactions of bacterial communities with a hybrid approach

1. Numerical method applied to controlled community involved in cheese production → tool designed (*in revision in Metabolic Engineering*)
   - Multi-omics data integration
   - Dynamic model → **iterative process**
2. Discrete model for screening metabolic bacterial potential in natural ecosystems → CoCoMiCo software (*in prep.*)
   - Inference of **interaction properties** from KB
3. Improvement of discrete models
   - Inference of **temporal information** from KB
   - Selection of communities based on **biological constraints**

Metabolic modeling for understanding complex ecosystems

https://forgemia.inra.fr/tango/tango_models

https://gitlab.inria.fr/ccmc/CoCoMiCo

Metabolism explains observable phenotype

Metabolism

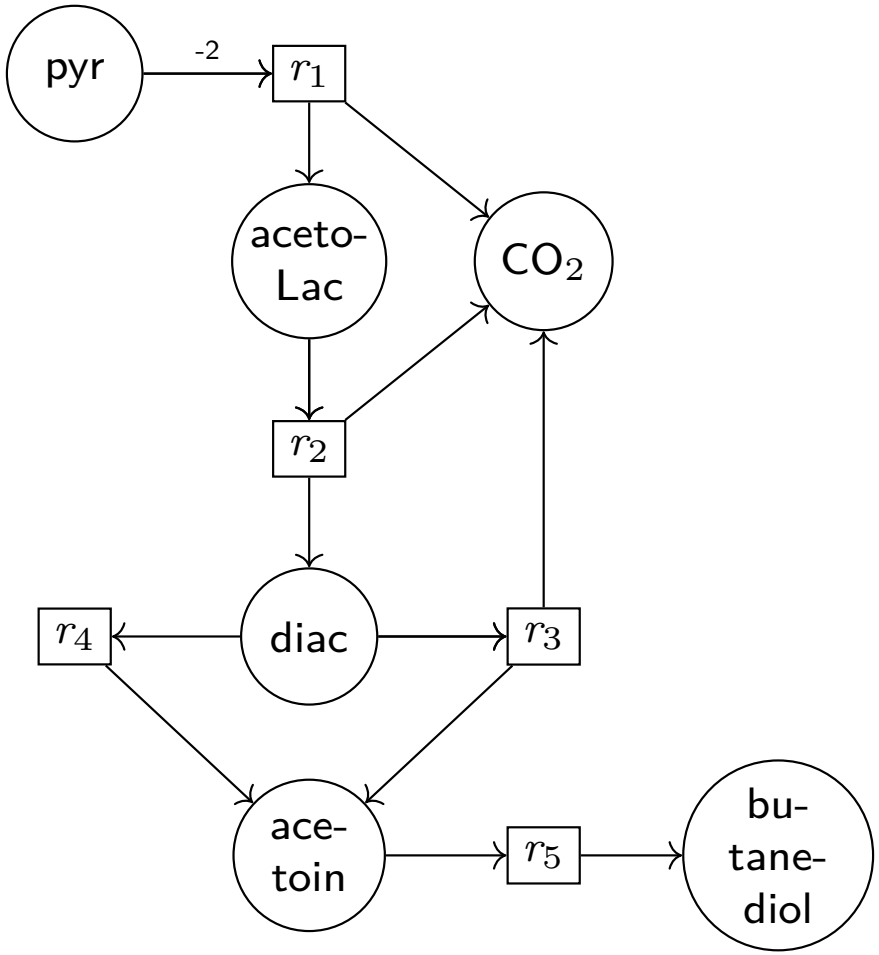
Set of all biochemical reactions occurring in the cell of an organism that permit the production of energy and metabolic goods.

r_1 :2 pyr $\rightarrow$ 1 aceto-Lac + 1 CO_2
r_2 :1 aceto-Lac $\rightarrow$ 1 diac + 1 CO_2
r_3 :1 aceto-Lac $\rightarrow$ 1 acetoin + 1 CO_2
r_4 :1 diac $\rightarrow$ 1 acetoin
r_5 :1 acetoin $\rightarrow$1 butanediol

Stoichiometry matrix

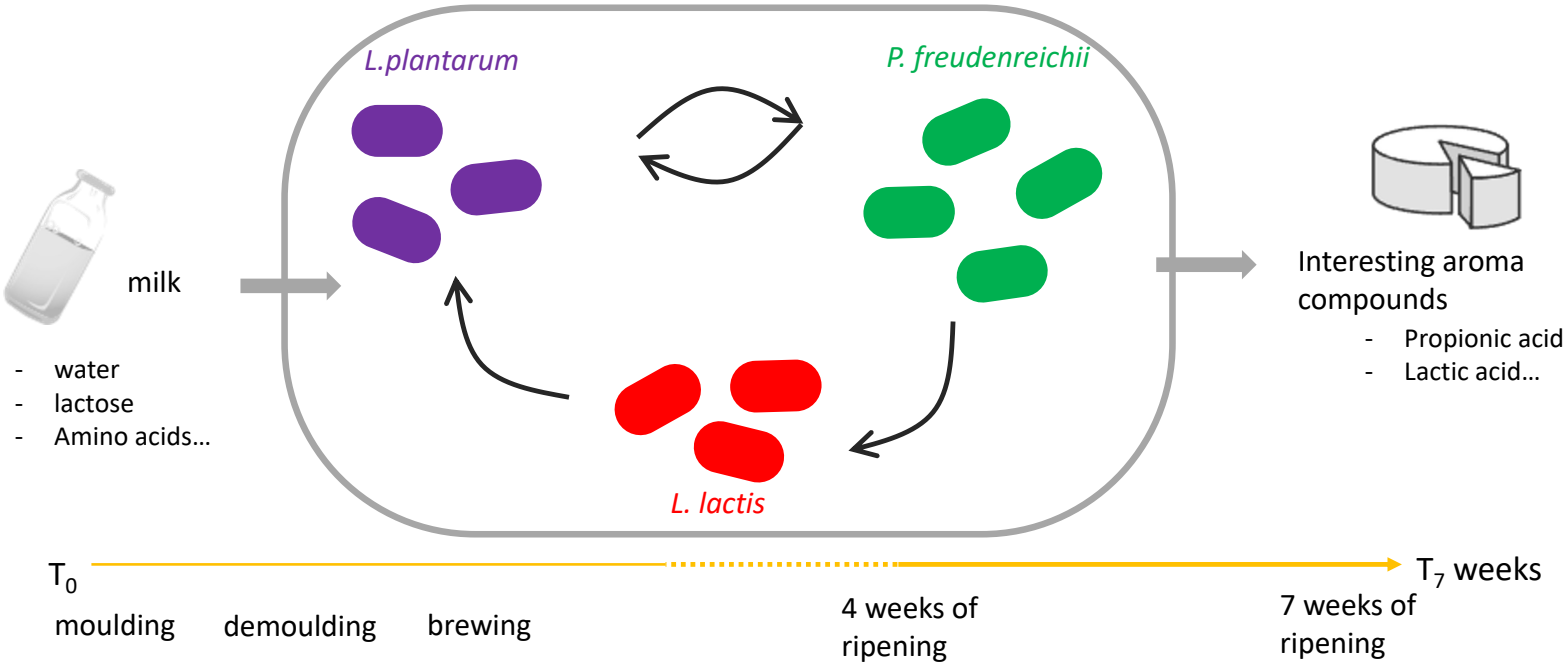
$$\begin{array}{c} \\ \text{pyr} \\ \text{aceto-Lac} \\ \text{diac} \\ CO_2 \\ \text{acetoin} \\ \text{butanediol} \end{array} \begin{array}{c} \begin{array}{ccccc} r_1 & r_2 & r_3 & r_4 & r_5 \end{array} \\ \begin{pmatrix} -2 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Bipartite graph relationship



Sánchez López de Nava A, 2023.

Bacterial fermentation: a complex dynamic process



Challenge

- Build coherent metabolic networks (iterative refinement)
- Nutrient concentration over time
- The dynamics of bacterial density
- Resource sharing between populations

Solution

- Create a dynamic and a numeric model of the metabolism (FBA, dFBA)
- TANGO implementation for characterizing bacterial interaction

Orth, Thiele, and Palsson, 2010; Mahadevan, Edwards, and Doyle, 2002.

Numerical model of the metabolism – FBA

definition

Metabolic model

From a GEM, a model metabolic has the capacity to simulate and to predict on the metabolic activity

Orth, Thiele, and Palsson, 2010.

Numerical model of the metabolism – FBA

Flux Balance Analysis

Metabolic model

From a GEM, a model metabolic has the capacity to simulate and to predict on the metabolic activity

Constraint-based approaches

$$\text{maximiser/minimiser } f_{obj}$$
$$\text{tel que } (S.v)_{int} = 0$$
$$\text{et } v_{i_{min}} \leq v_i \leq v_{i_{max}}$$

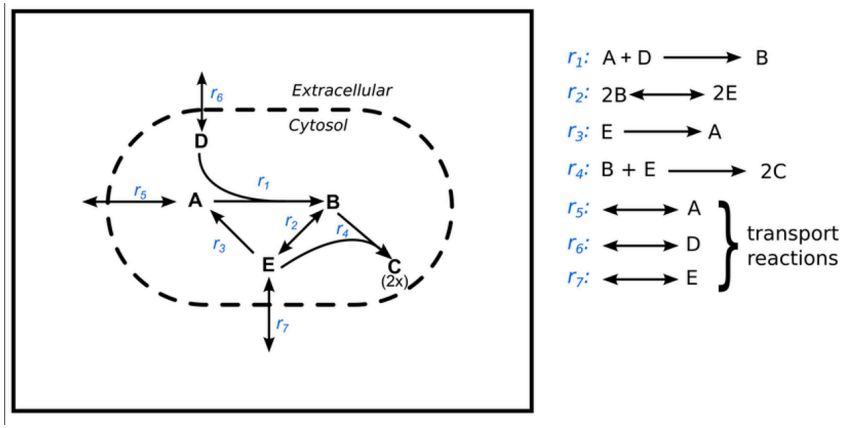


Figure 1: Example of a metabolic network

$$S = \begin{array}{c|ccccccc} & r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 \\ \hline A & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ B & 1 & -2 & 0 & -1 & 0 & 0 & 0 \\ C & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ D & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ E & 0 & 2 & -1 & -1 & 0 & 0 & 1 \end{array} \qquad \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix}$$

(Stoichiometric values) (Metabolic flux values)

Figure 2: A. Stoichiometry matrix representation and the flux vector v

Orth, Thiele, and Palsson, 2010.

Numerical model of the metabolism – FBA

Flux Balance Analysis

Metabolic model

From a GEM, a model metabolic has the capacity to simulate and to predict on the metabolic content

Constraint-based approaches

$$\text{maximiser/minimiser } f_{obj}$$
$$\text{tel que } (S.v)_{int} = 0$$
$$\text{et } v_{i_{min}} \leq v_i \leq v_{i_{max}}$$

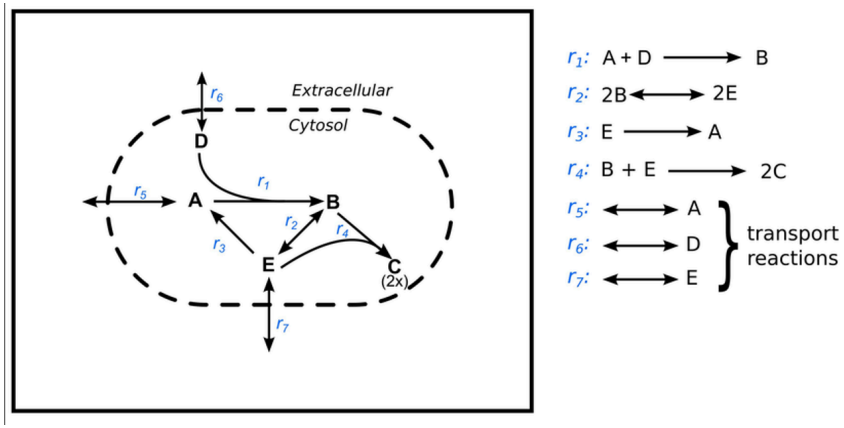


Figure 1: Example of a metabolic network

Subject to

$$S\vec{v} = \vec{0} = \begin{cases} \frac{dA}{dt} = -v_1 + v_3 + v_5 \\ \frac{dB}{dt} = v_1 - 2v_2 - v_4 \\ \frac{dC}{dt} = 2v_4 \\ \frac{dD}{dt} = -v_1 + v_6 \\ \frac{dE}{dt} = 2v_2 - v_3 - v_4 + v_7 \end{cases} \qquad \begin{matrix} 0 \leq v_1 < \infty \\ -\infty < v_2 < \infty \\ 0 \leq v_3 < \infty \\ 0 \leq v_4 < \infty \\ 0 \leq v_5 \leq \infty \\ -\infty < v_6 < \infty \\ 0 \leq v_7 \leq \infty \end{matrix}$$

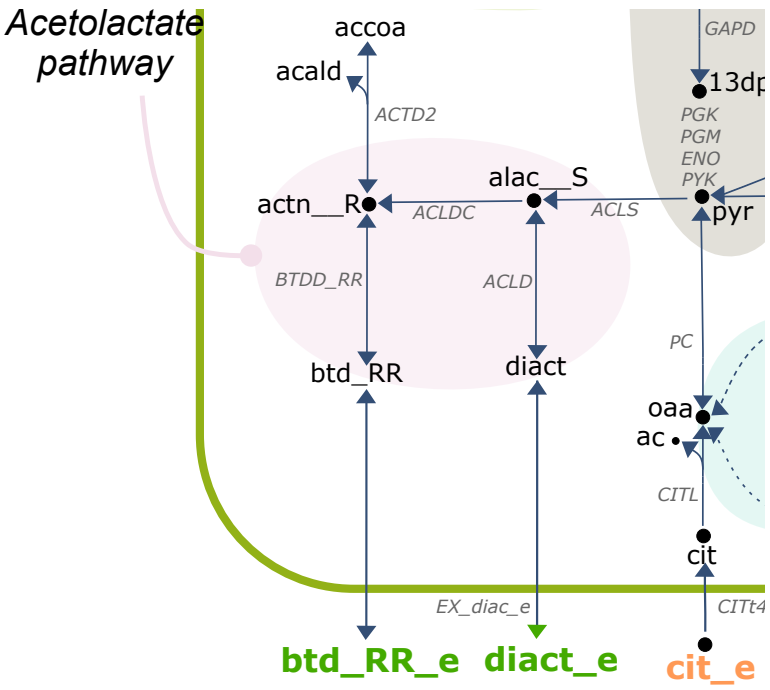
(Steady state system) (Reaction bounds)

Figure 2: B. Linear programming problem.

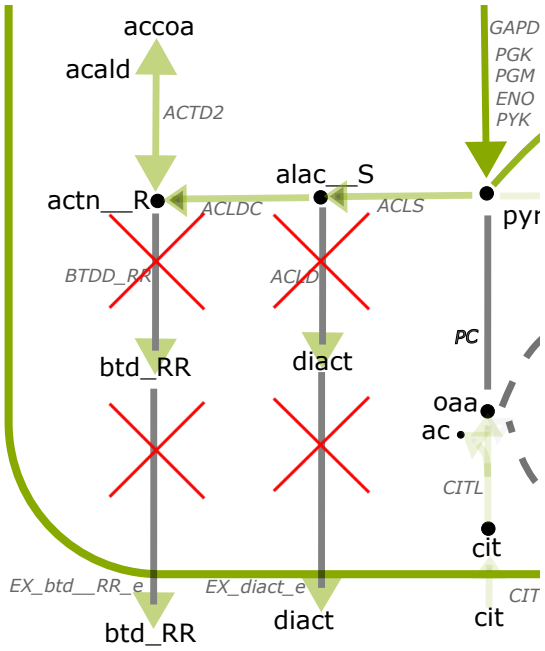
Orth, Thiele, and Palsson, 2010.

FBA genome-scale metabolic model refinement

- Modify reactions (flux, add, remove)
- Verify known pathway and metabolic activities



Initial metabolic map

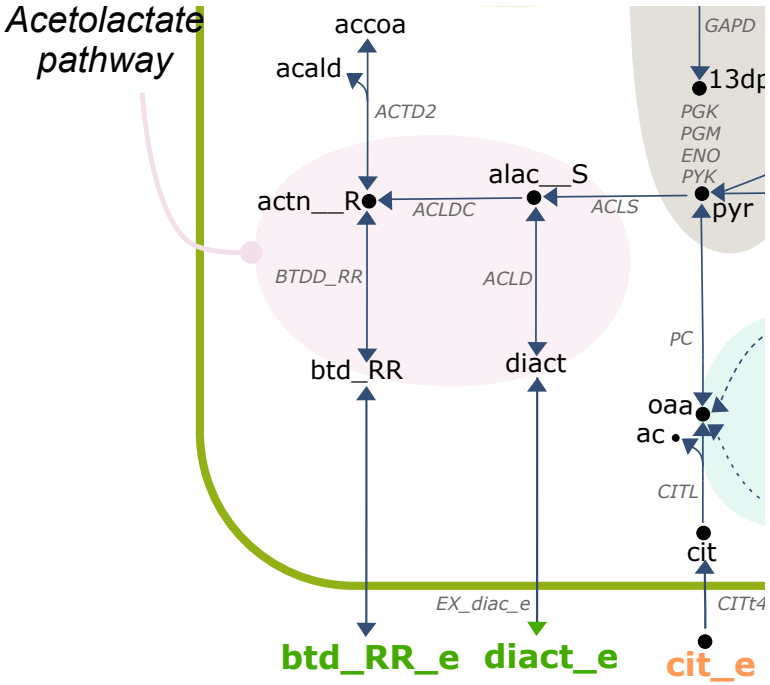


FBA show missing expected activities

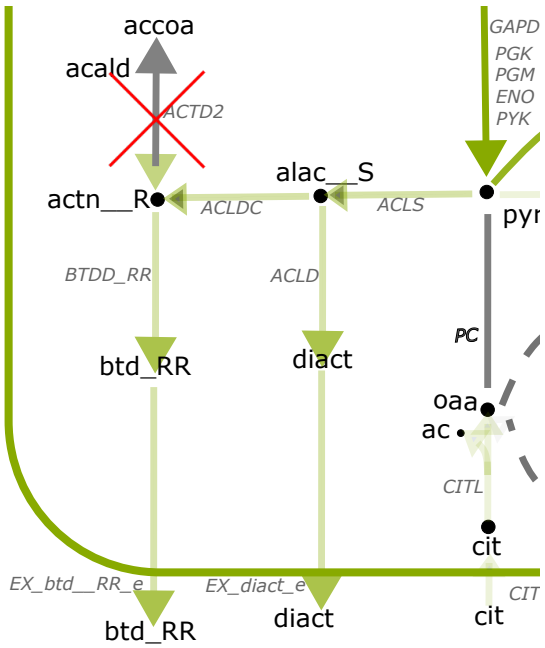
Carroll et al., 1999; Swindell et al., 1996; Makhlouf, 2006.

FBA genome-scale metabolic model refinement

- Modify reactions (flux, add, remove)
- Verify known pathway and metabolic activities



Initial metabolic map



Constraining ACTD2 reaction

Improvement

Modification reaction bounds in the model and in the software

Carroll et al., 1999; Swindell et al., 1996; Makhlouf, 2006.

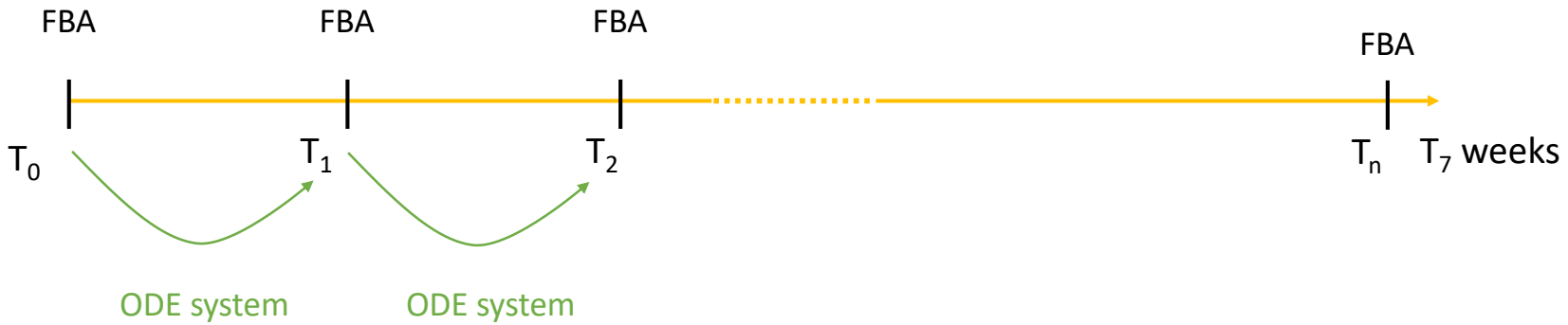
Dynamic model of the metabolism

dFBA

- FBA at each time point
- Previous time point output used as input for the next time point

FBA

$$F_j = \sum_{i \in \mathcal{B}} \mu_{i,j} \left((c^{ex}_{min,i}, c^{ex}_{max,i})(b^n, m^n) \right) b_i$$



ODE System

$$b_i^{n+1} = b_i^n + \Delta t * F_{b_i}$$

$$m_j^{n+1} = \begin{cases} m_i^n + \Delta t * F_j & \text{si } F_j > 0 \\ m_j^n / (1 - \Delta t * F_j / m_j^n) & \text{sinon} \end{cases}$$

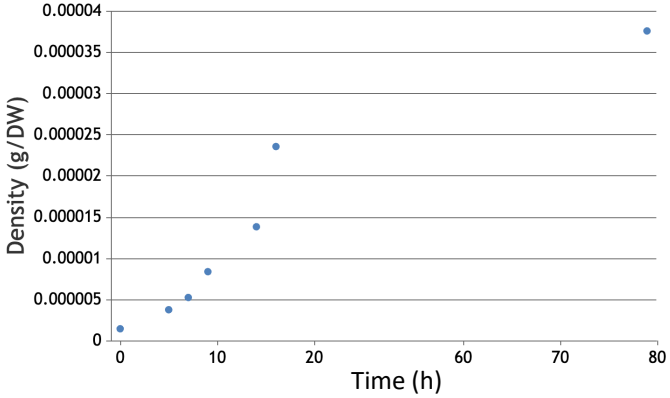
Mahadevan, Edwards, and Doyle, 2002.

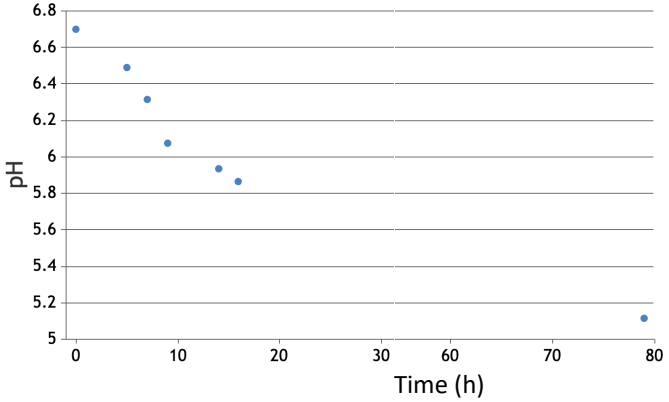
Parameter inference

- Make individual GSMN accurate for inferring mechanistic bacterial behavior
- Finding optimal parameters for explaining **species level** biological observations
- Quantitative check of metabolic goods and biomass density

Lactic Acid Bacteria (LAB)

$$J(b_i, pH|\theta_i, b_{i,exp}, pH_{exp}) = \left\| \frac{\log_{10}(b_i) - \log_{10}(b_{i,exp})}{\sigma_{log,i,exp}} \right\|^2 + \alpha \left\| \frac{pH - pH_{exp}}{\sigma_{pH,exp}} \right\|^2 \quad (1)$$





Environmental condition

- Milk

Parameter inference

- Make individual GSMN accurate for inferring mechanistic bacterial behavior
- Finding optimal parameters for explaining **species level** biological observations
- Quantitative check of metabolic goods and biomass density

P. freudenreichii

$$J(b, m|\theta_i, b_{exp}, m_{exp}) = \left\| \frac{\log_{10}(b) - \log_{10}(b_{exp})}{\sigma_{log,b,exp}} \right\|^2 + \alpha \left\| \frac{m - m_{exp}}{\sigma_{m,exp}} \right\|^2 \quad (1)$$

Acids	Concentration T_0	Concentration $T_f = 89$ h	Standard error
Lactate	16.5	7.88	0.08
Acetate	0	3.07	0.02
Succinate	0	0.371	0.063
Propionate	0	8.31	0.02

Table 1: Acids concentrations data in g/L

Environmental condition

- Milk + lactate + peptone

Dynamic control of the metabolism

Defining consumption limitation in the model

$$c^{ex}_{min,i,j} = max(-\frac{m_j}{\Delta t * \sum_{i \in \mathcal{M}_|} b_i}, v^{int}_{i,j}) \quad (2)$$

- $\mathcal{M}_|$ Bacteria subset can metabolize $_j$
- $v^{int}_{i,j}$ close to the FBA value
- Balanced resource sharing

Lactose consumption case for LAB

1- Calculation of pH from lactate concentration

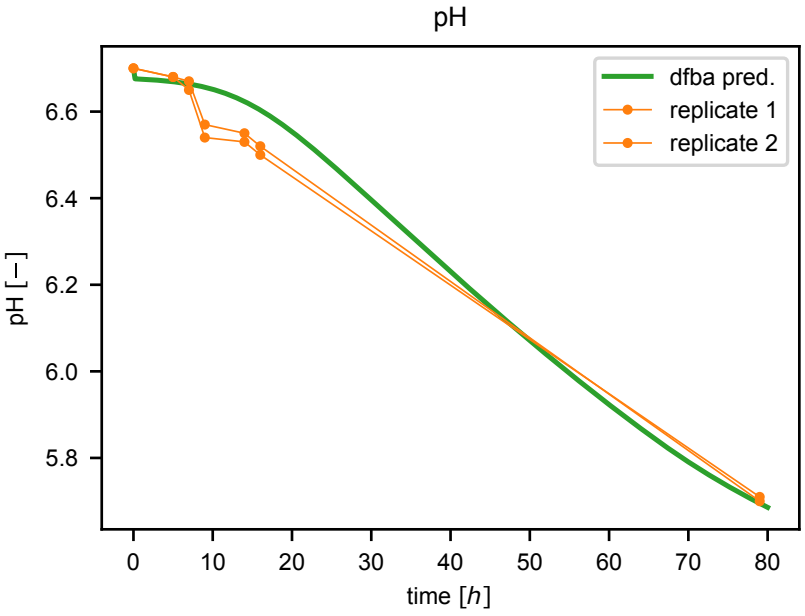
$$pH = pKa_{lac} c_1 * (m_{lac__L_e} + m_{lac__D_e}) + c_2$$

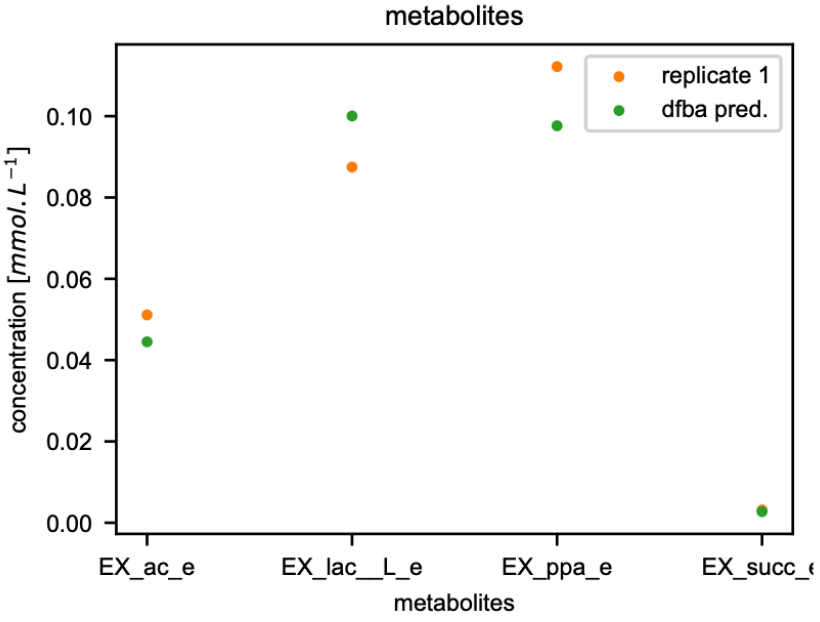
2- Regulation of lactose consumption

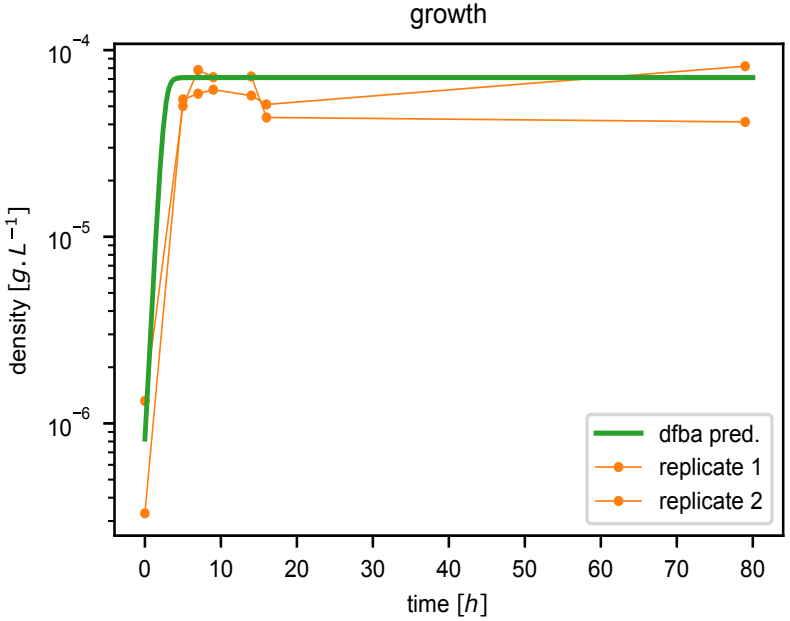
$$c^{ex}_{min,i,j} = max(-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, -\mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} - \mu_{min,lcts}) \quad (3)$$

Ozcan et al., 2020.

Individual model validation



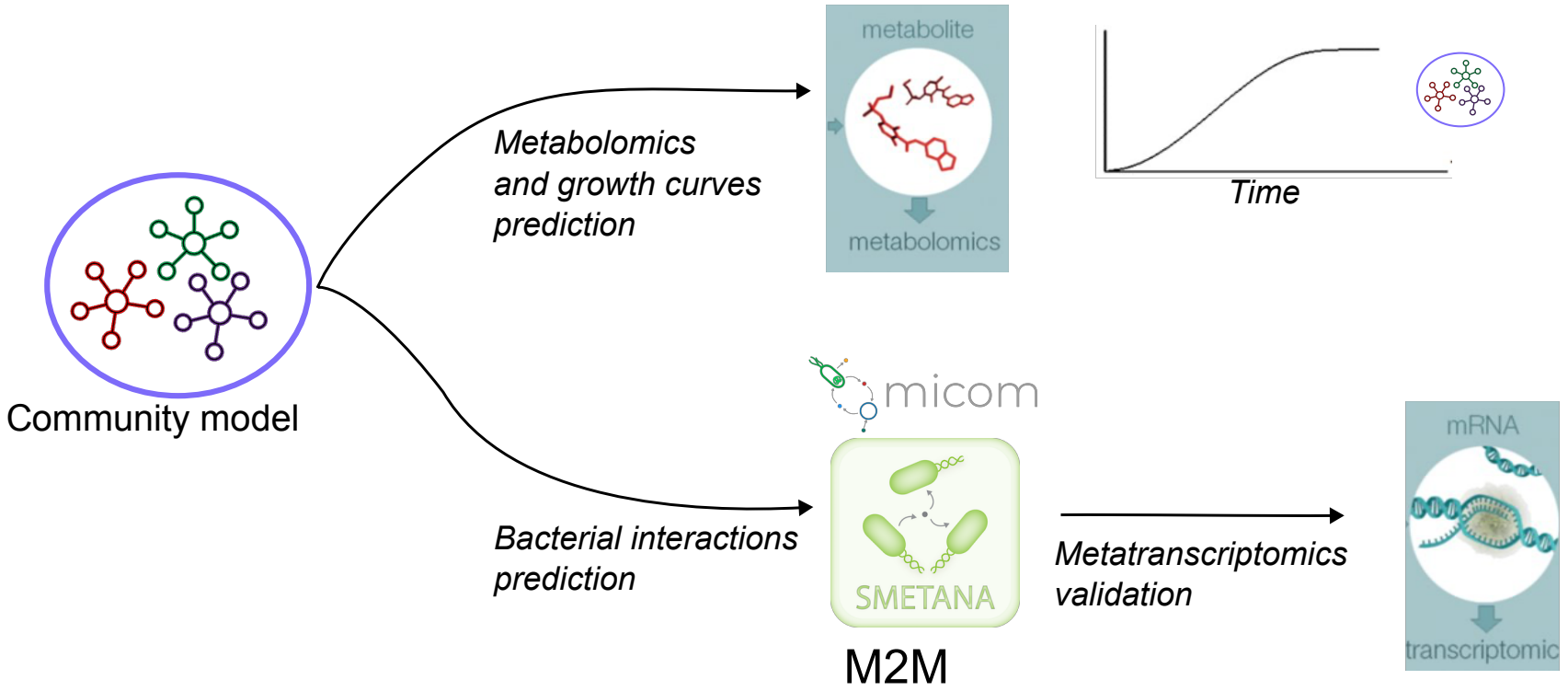




Results after parameters inference

Individual metabolic models are well calibrated → reproduce experimental data

Community validation

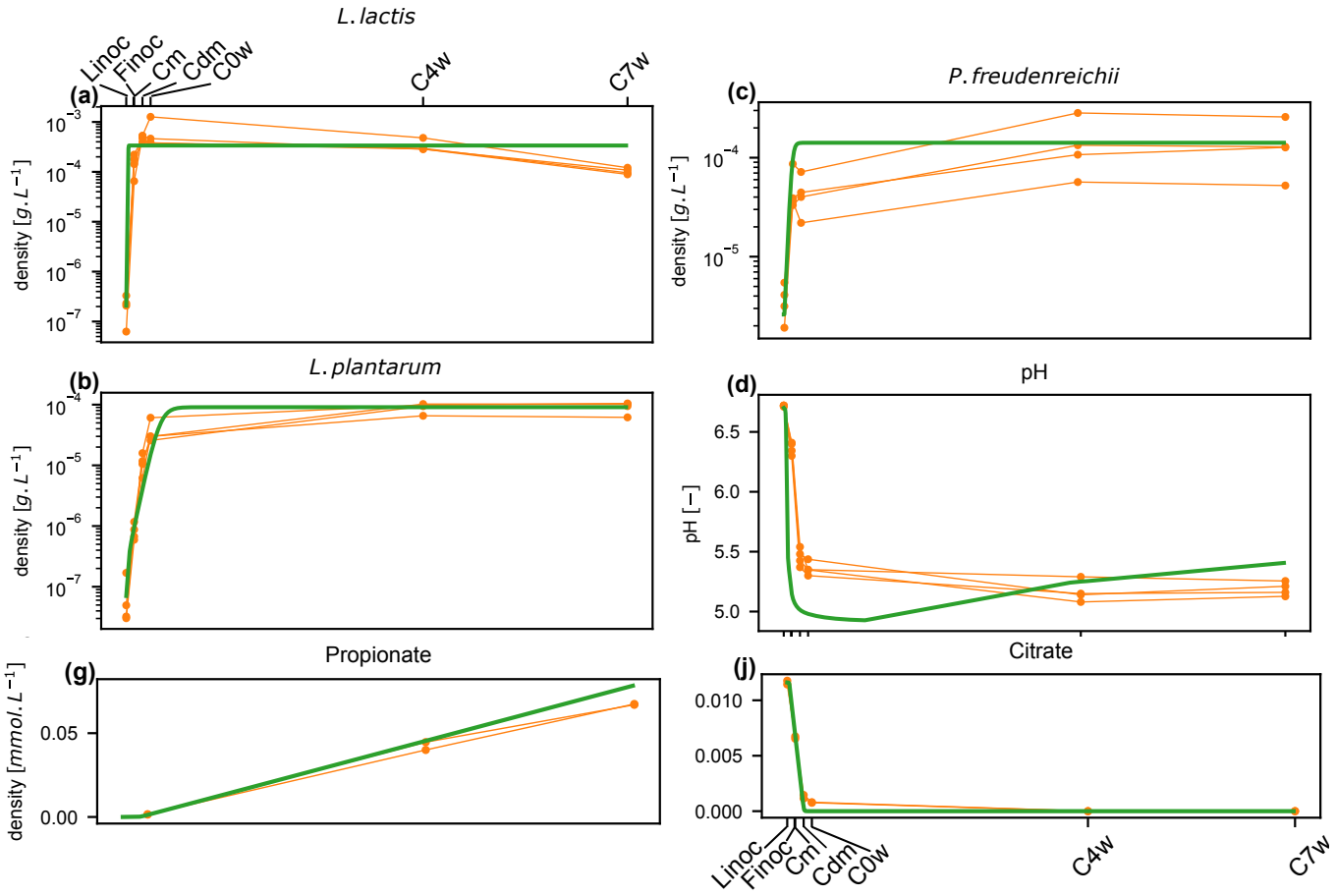


- Bacterial interaction **prediction**
- Metabolic explanation of biological observations
- **No community calibration**

Cao et al., 2021.

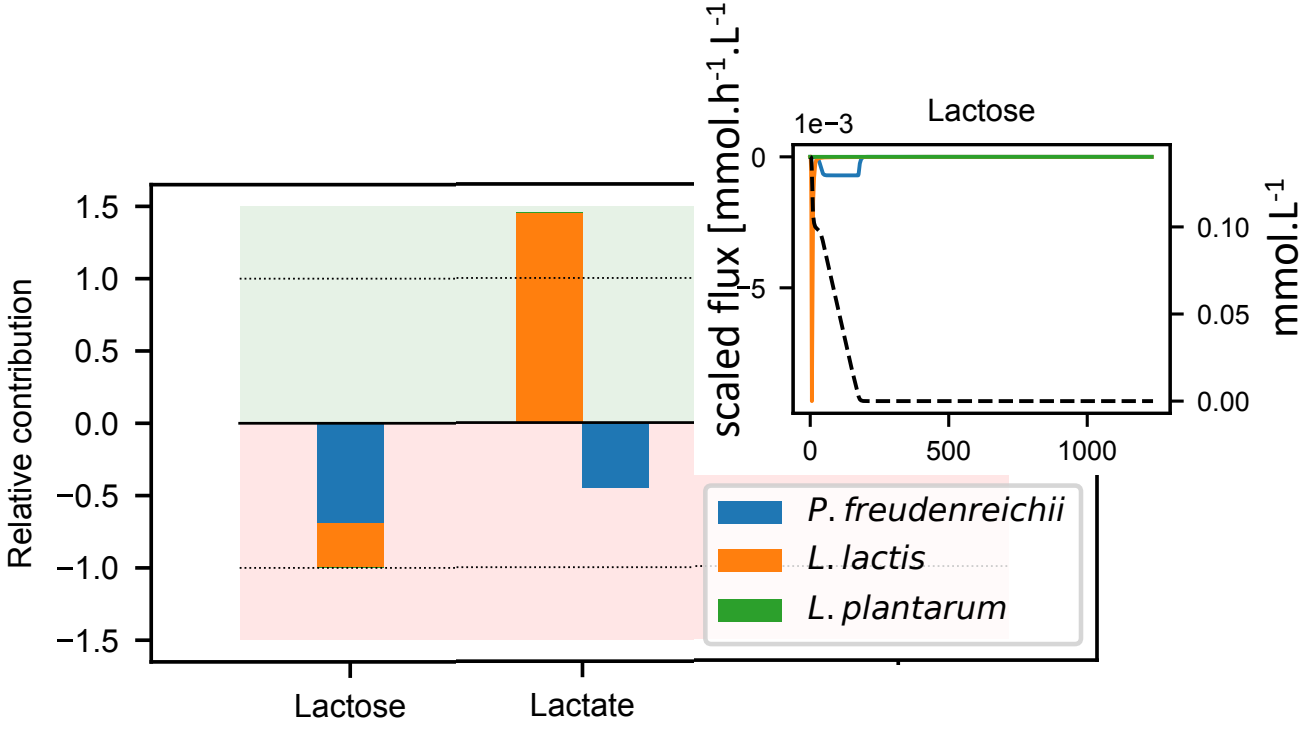
Community prediction

Growth and pH



- Growth well predicted for all bacteria
- Lactate proxy production can explain the observed pH
- Metabolomic data is well predicted

Highlighted bacterial community behavior



- Temporal segregation for lactose
- *L. lactis* main lactate producer
- Commensalism between *L. lactis* and *P. freudenreichii*

- 11 shared metabolites predicted with SMETANA, MiCOM (phenylalanine, succinate, xanthine..)

Zelezniak et al., 2015; Diener, Gibbons, and Resendis-Antonio, 2020.

Numerical model conclusion

Numerical modelling of a controlled community

Strength

- Faithful numerical model on small communities
- Iterative process (refinement and calibration)

Weaknesses

- Greedy process
- *A priori* knowledge

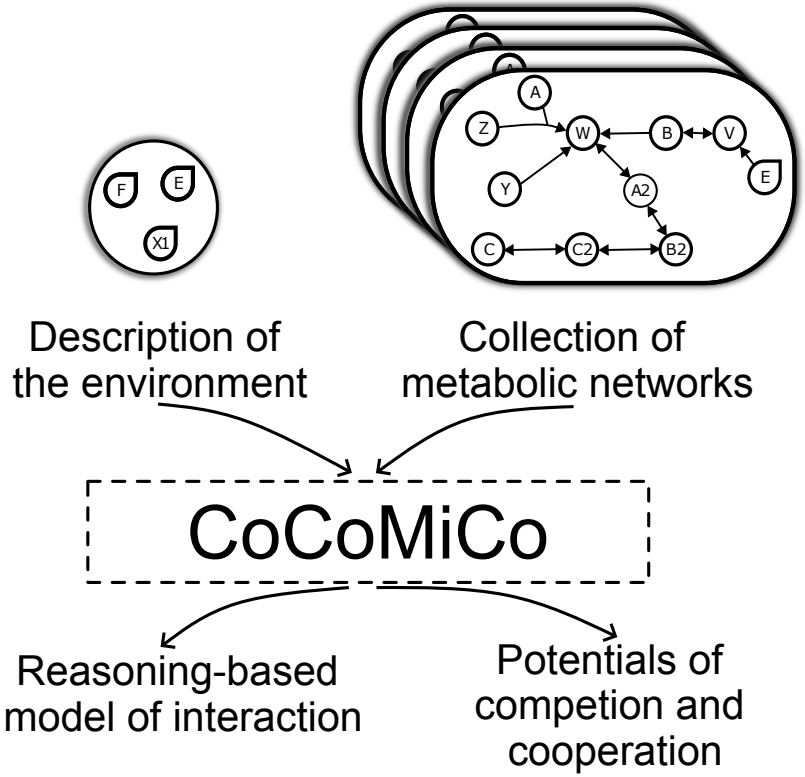
Opportunities

In silico model usable in industrial contexts

Threats

Scalability

Discrete approach for characterizing large-scale bacterial interaction



Challenge

- Scalable to natural environment
- Whole community
- Define cooperation and competition
- Characterize metabolic interaction properties

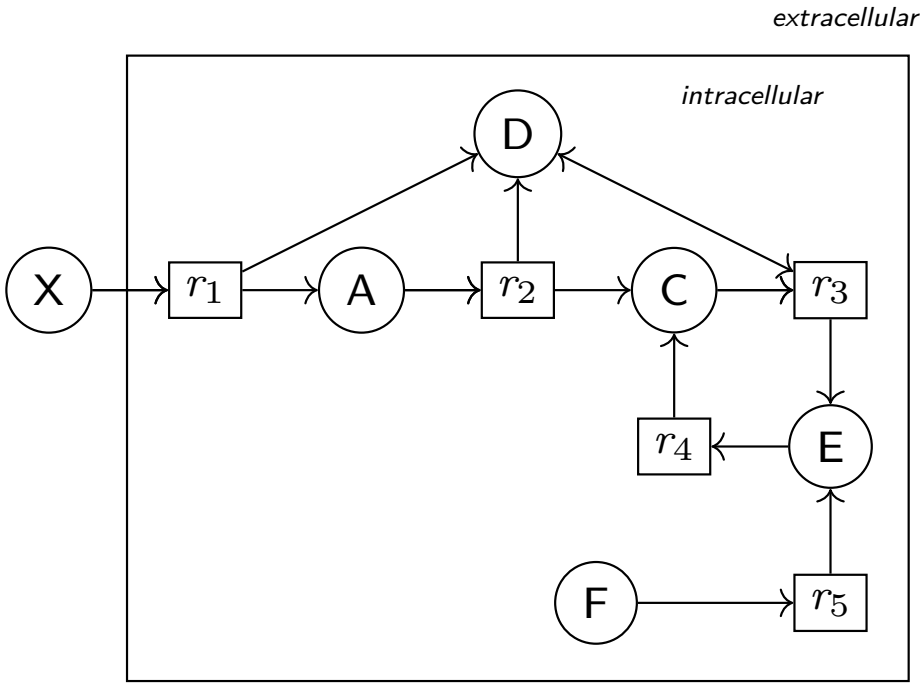
Solution

- Create a discrete approach (reasoning-based)
- Create a consistent KB and rules
- CoCoMiCo implementation for characterizing bacterial interactions

KB and logical rules constructions

- Report relations between metabolites and reactions

Source of the KB



Source of the rules

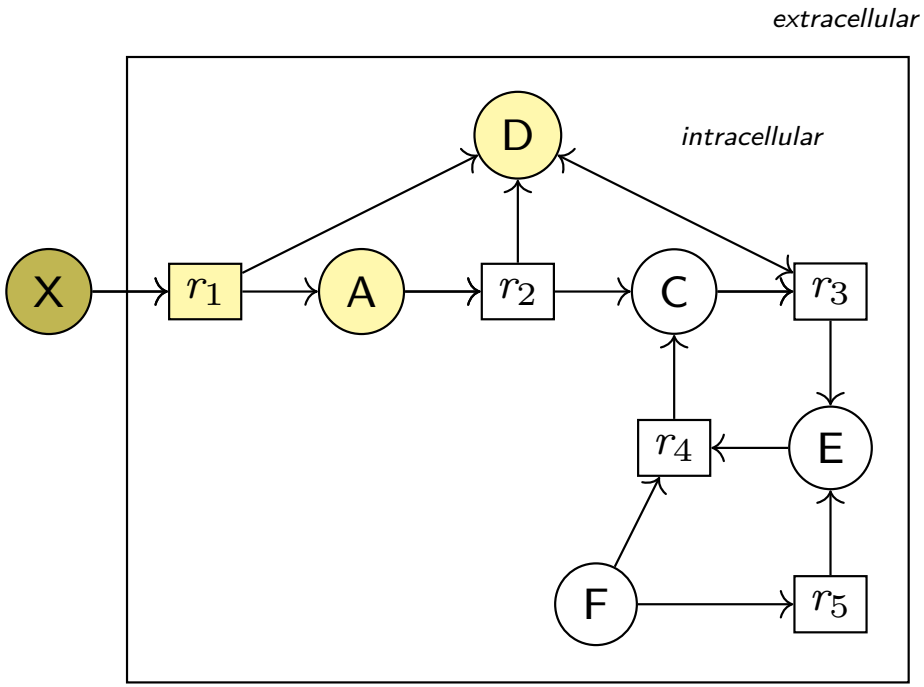
Modeler and biological expert

- Producibility is initiated by the presence of nutrients,
- The products of a reactions are producible if all reactants of this reaction are themselves producible

The scope, *i.e.* the metabolic capacity of a network reasons on the biochemical reactions

Ebenhöh, Handorf, and Heinrich, 2004.

Illustration of the scope

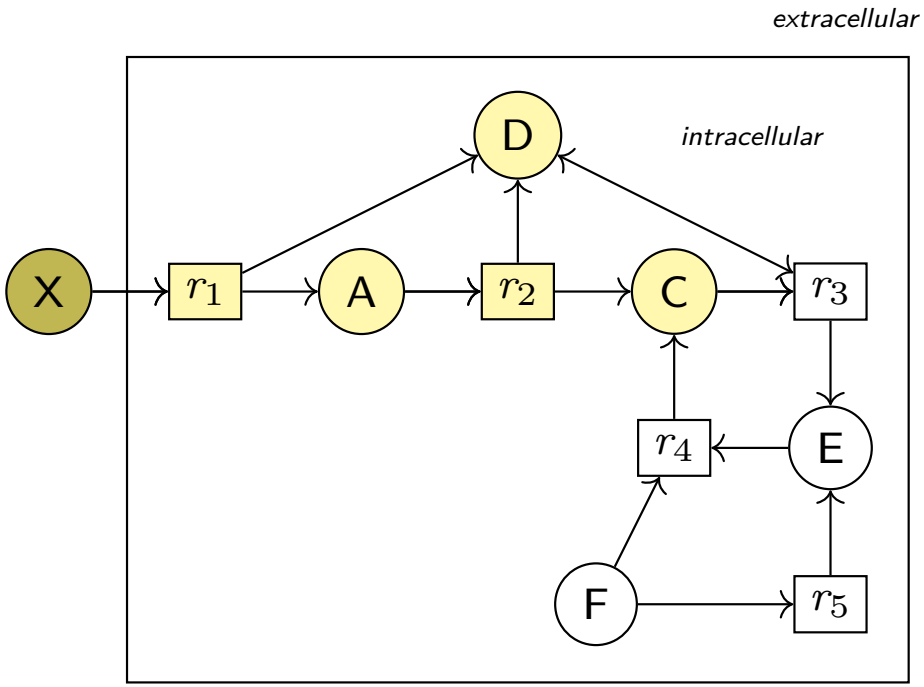


Modeler and biological expert

- Producibility is initiated by the presence of nutrients,
- The products of a reactions are producible if all reactants of this reaction are themselves producible

X is a seed
scope : X,A,D

Illustration of the scope

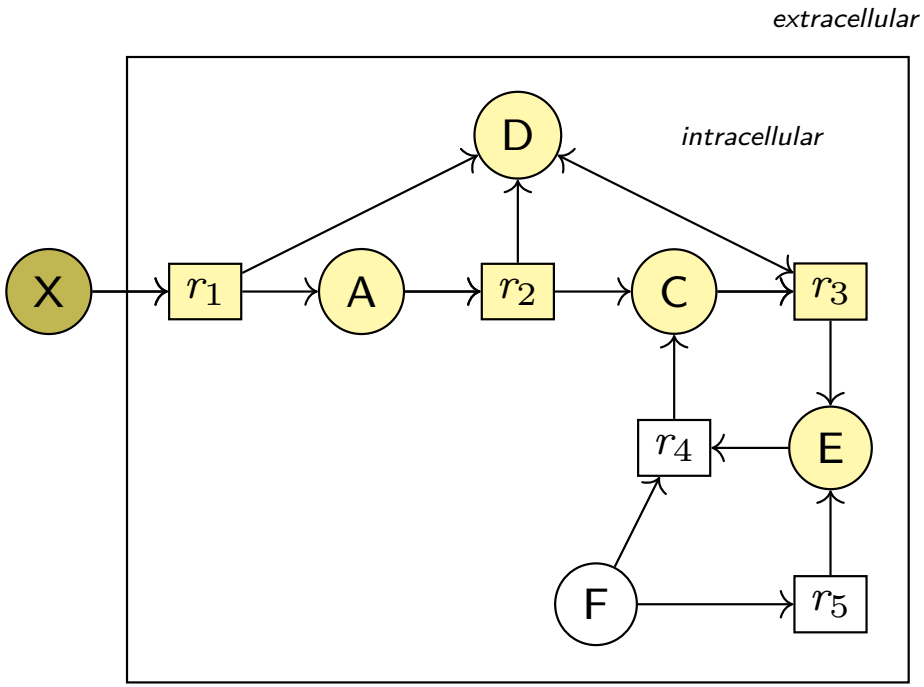


Modeler and biological expert

- Producibility is initiated by the presence of nutrients,
- The products of a reactions are producible if all reactants of this reaction are themselves producible

X is a seed
scope : X,A,D,C

Illustration of the scope



Modeler and biological expert

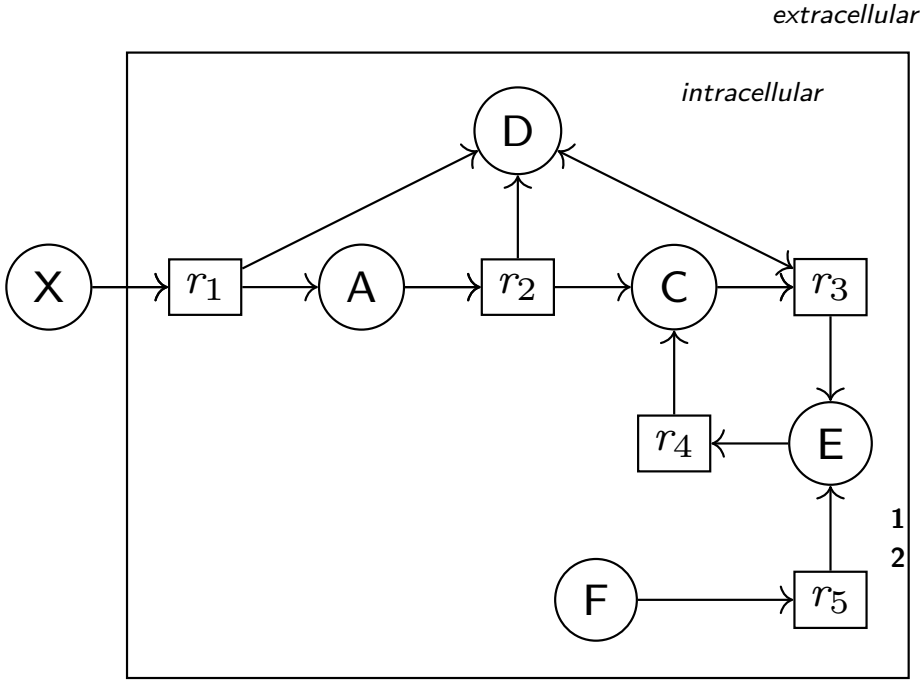
- Producibility is initiated by the presence of nutrients,
- The products of a reactions are producible if all reactants of this reaction are themselves producible

X is a seed
scope : X,A,D,C,E

KB and logical rules constructions

- Report relations between metabolites and reactions

Source of the KB



Source of the rules

Modeler and biological expert

- Producibility is initiated by the presence of nutrients,
- The products of a reactions are producible if all reactants of this reaction are themselves producible

```
scope(M) :- seed(M).
scope(M) :- bacteria(B), product(M,R,B),
      reaction(R,B), scope(M2) : reactant(M2,R,
      B).
```

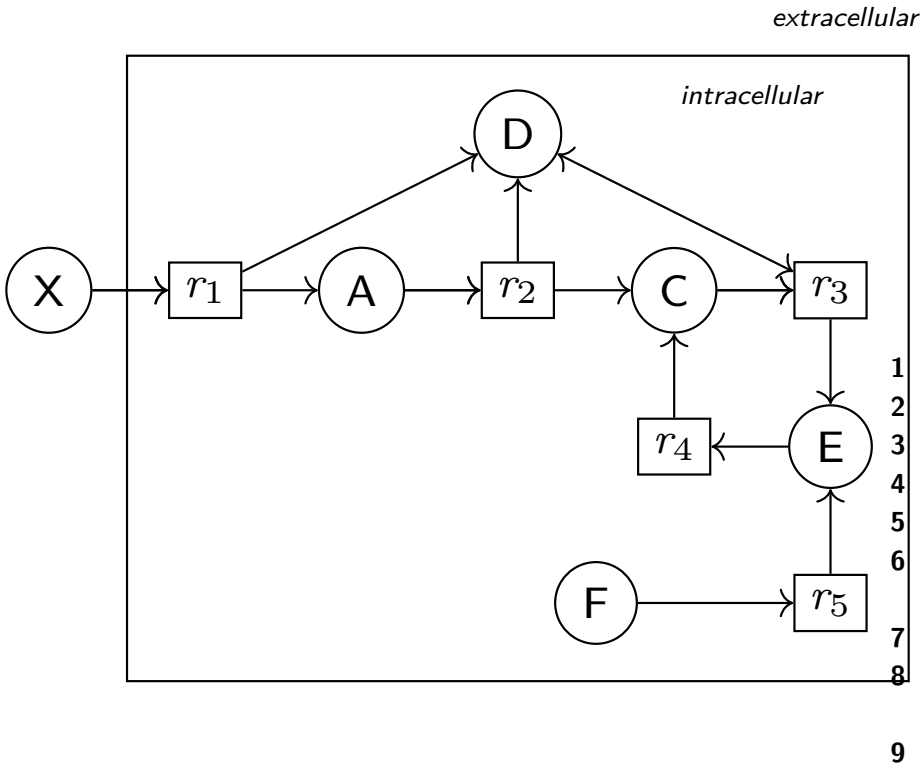
The scope of a network is formalized in 2 logical rules

Ebenhöh, Handorf, and Heinrich, 2004.

KB and logical rules constructions

- Report relations between metabolites and reactions

Source of the KB



Source of the rules

Modeler and biological expert

- Producibility is initiated by the presence of nutrients,
- The products of a reactions are producible if all reactants of this reaction are themselves producible

```
scope(metabolite(M,B), Community) :-
                taxon(B),
                in_community(Community, B),
                reaction(R,B),
                product(M,R,B),
                available(M', Community) :
                     reactant(M',R,B).

available(M, Community) :- seed(M), community(
     Community).
available(M, Community) :- scope(metabolite(M,
     B), Community).
```

Ebenhöh, Handorf, and Heinrich, 2004.

Answer Set Programming formalism

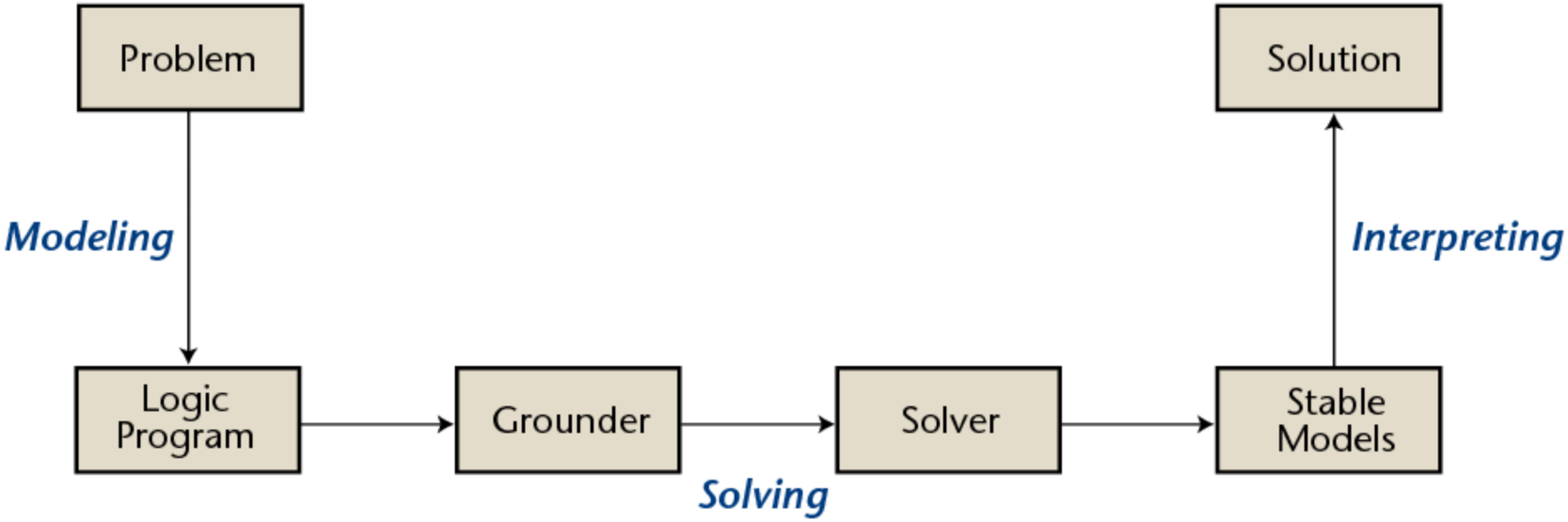


Figure 3: The workflow of Answer set programming (ASP)

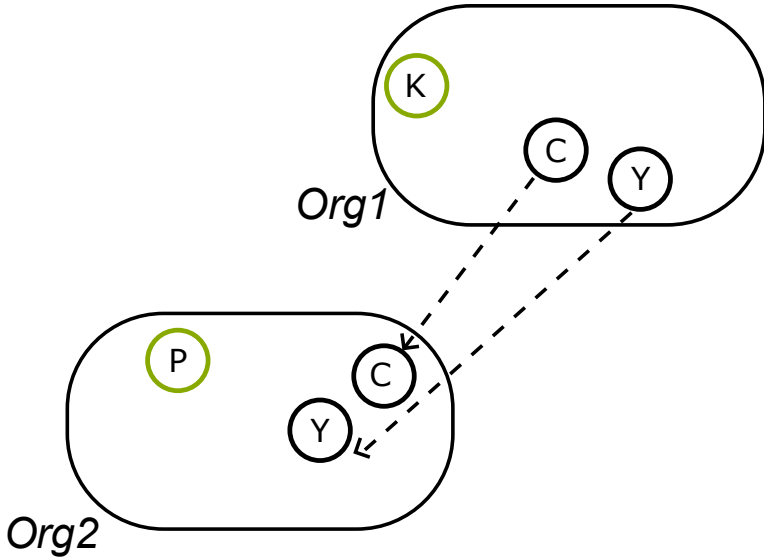
- Closed world assumption → unknown information is false
- Mechanistic model
- Solve combinatorial problems

Kaufmann et al., 2016.

Challenge: Characterize cooperation properties

Cooperation properties

- Exchanged metabolites

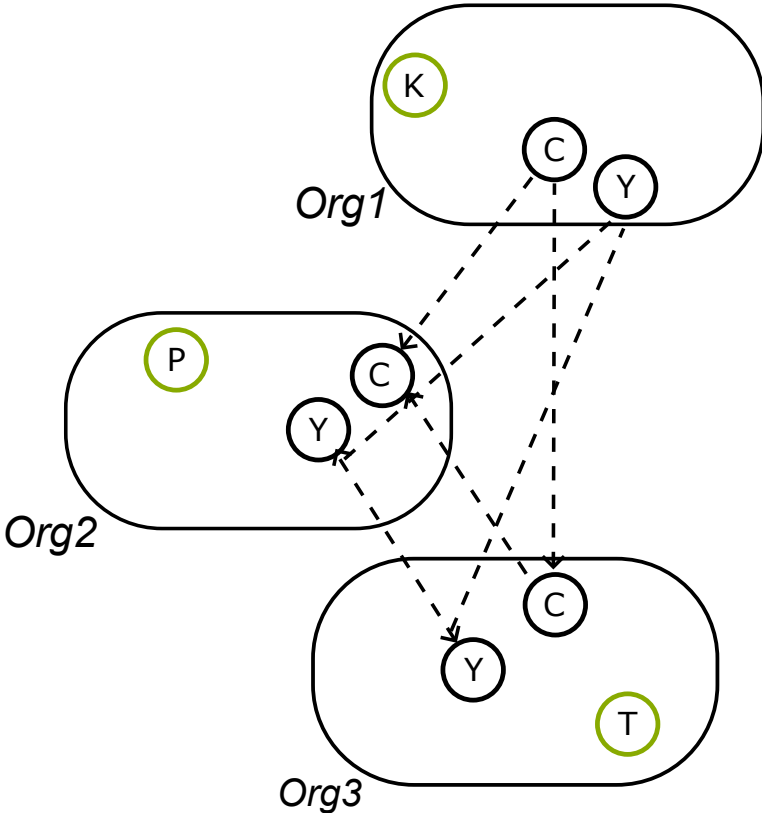


- Redesigned microbial ecology definition of cooperation → no growth effect
- Effect on donors are not taken into account (0, +, -)

Challenge: Characterize cooperation properties

Cooperation properties

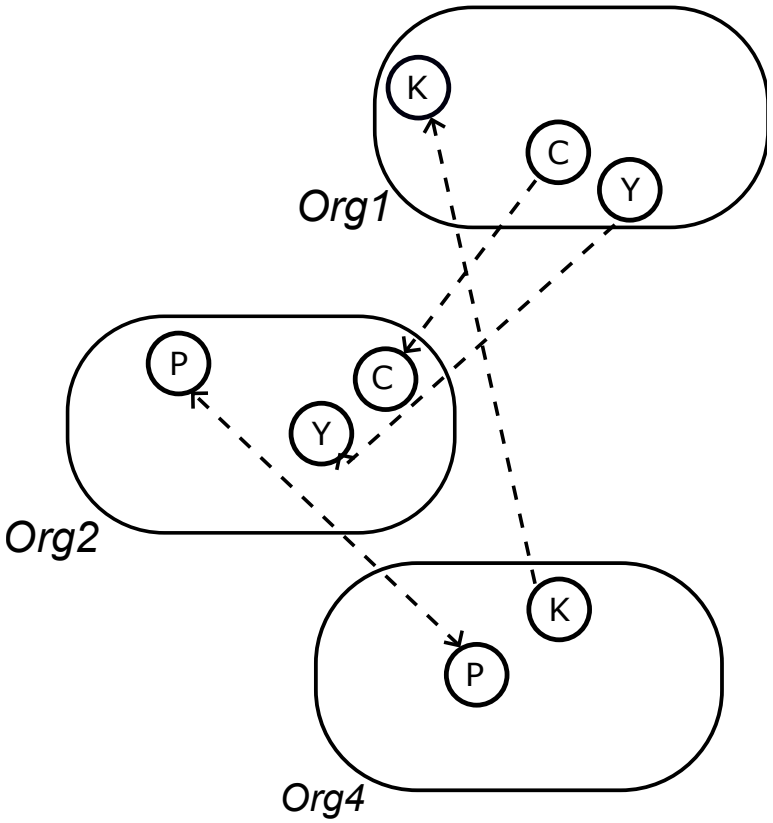
- Exchanged metabolites
- Can a species give to everyone in the system ?



Challenge: Characterize cooperation properties

Cooperation properties

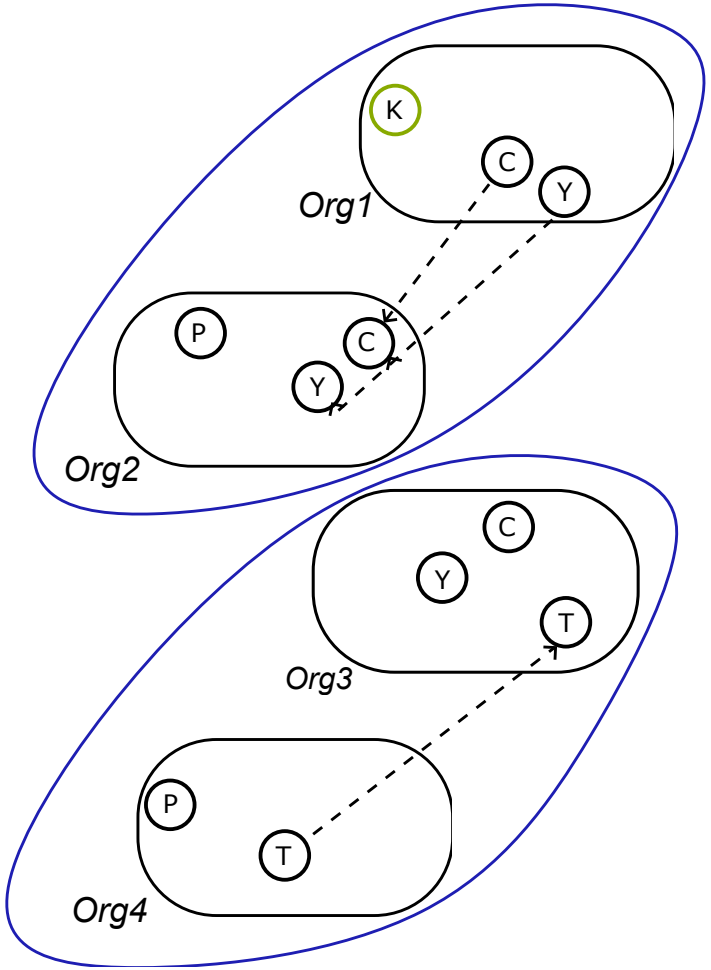
- Exchanged metabolites
- Can a species give to everyone in the system ?
- Added-value of adding a species



Challenge: Characterize cooperation properties

Cooperation properties

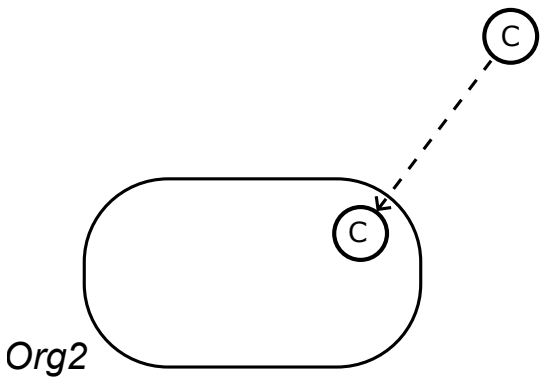
- Exchanged metabolites
- Can a species give to everyone in the systeme ?
- Added-value of adding a species
- Compare cooperation indexes



Characterize competition interaction, economic point of view

- Monopsony

Monopsony comes from two Greek words: "monos" meaning "single" and "opsonia" meaning "purchase."

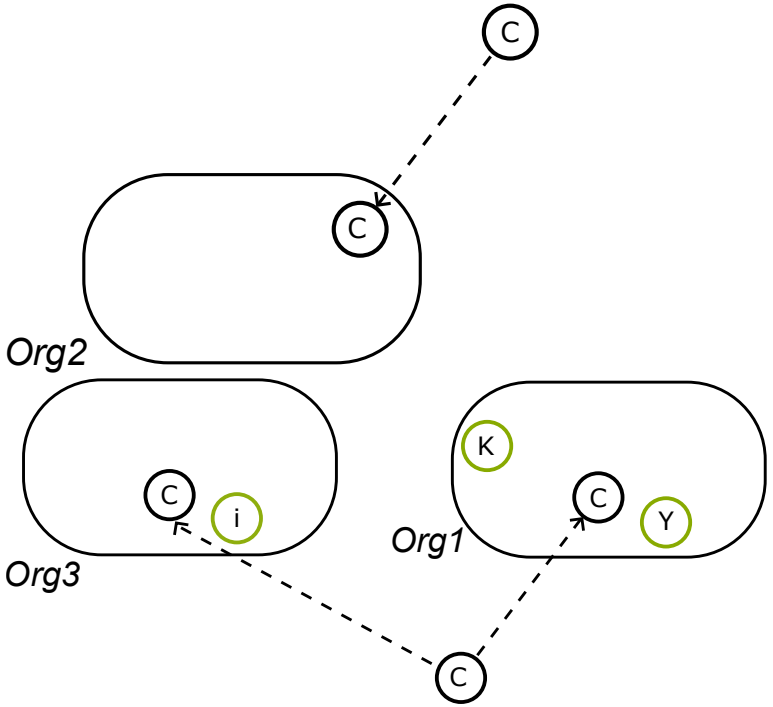


Characterize competition interaction, economic point of view

- Monopsony

Monopsony comes from two Greek words: "monos" meaning "single" and "opsonia" meaning "purchase."

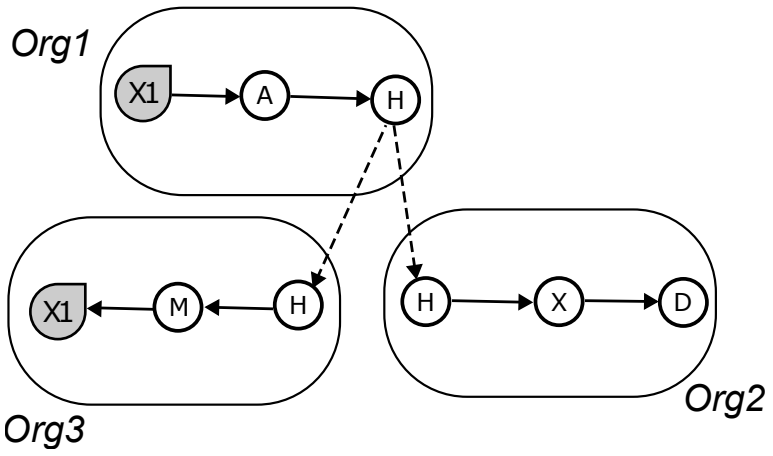
- **Polyopsony**



- How cooperation and competition (polyopsony) rules are defined ?

Potential interactions rules

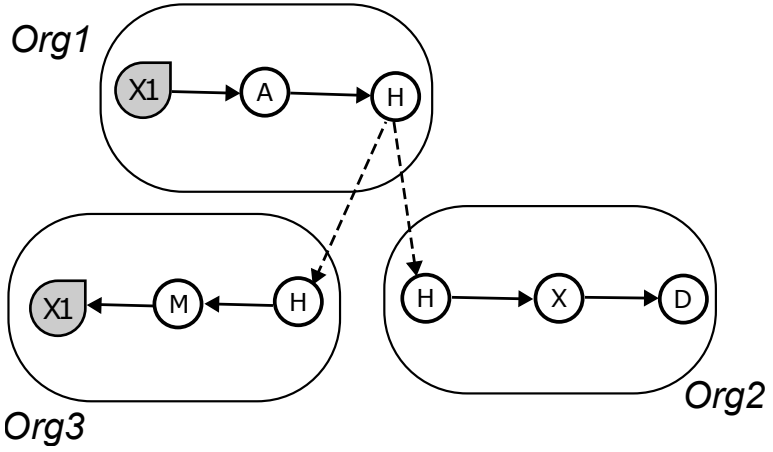
```
exchange(M,P,C) :- taxon(P),
taxon(C),
P != C,
reactant(M,_,C),
product(M,_,P),
scope(metabolite(M,P),
 all),
not scope(metabolite(M,C),
self(C)).
```



Potential interactions rules

```
exchange(M,P,C) :- taxon(P),
taxon(C),
P != C,
reactant(M,_,C),
product(M,_,P),
scope(metabolite(M,P),
 all),
not scope(metabolite(M,C),
self(C)).
```

```
polyopsony(M,N) :-
N=#count{C,M:exchange(M,_,C)},
exchange(M,_,_), N > 1.

polyopsony(S,N) :-
N=#count{B:seed_consumed_by_
    taxon(S,B)},
N > 1, seed(S).
```



Scope of CoCoMiCo

Use of logical formalism for characterizing cooperation and competition properties

- *Exchanged compounds*
- Polyopsony, monopsony
- Bacteria involved in an exchange
- ρ, δ metrics

Calculate indexes of cooperation and competition

- Compare communities
- Use as optimization criteria

Cooperation (exchanged compounds) and competition (polyopsony) properties used for calculating cooperation and competition indexes

Frioux et al., 2018.

Cooperation indexes

Cooperation weight

Can a species give to everyone in the systeme ?
How to qualitatively model the contributions of exchange species?

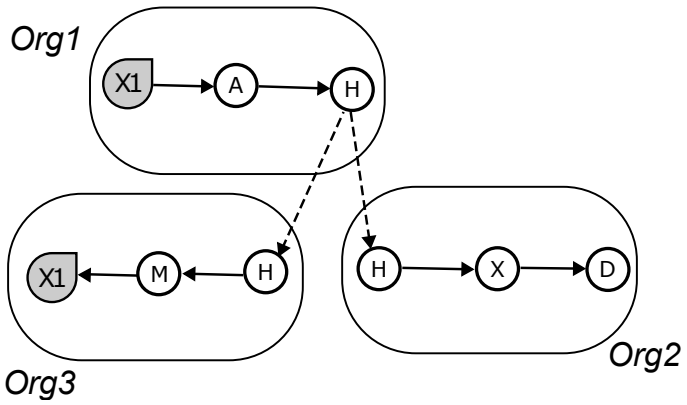
Exchangeable metabolites

```
H:{Org1} % producer
H:{Org2,Org3} % consumers
```

$$w(k) = 2 - 0.5^{k-1}$$

```
H:{1} % producer
H:{1.5} % consumers
```

$$\text{CooP} = \sum_{m \in M} w(|P_m|) + \sum_{m \in M} w(|C_m|)$$
$$= 2.5$$



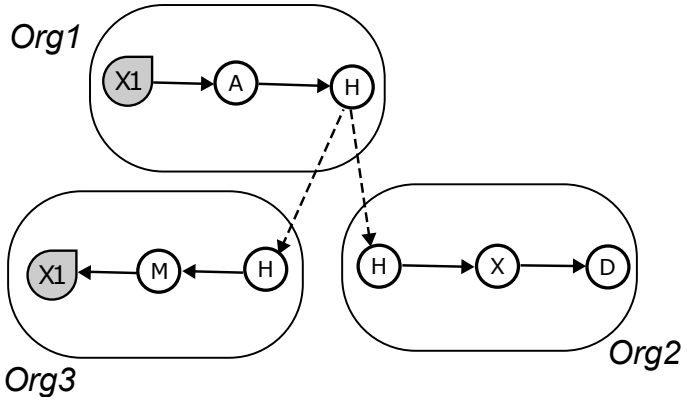
Competition indexes

Polyopsonistic

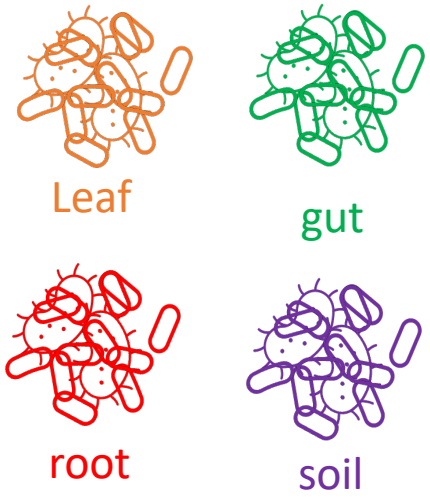
Number of consumers involved in exchangeable metabolites and seed > 1

```
H:{Org2,Org3} % exchanged metabolite
X1:{Org1} % the nutrient
```

```
Comp = sum(polyopsonist.values()) / len(community.taxa) = 2/3
```



Workflow for method assessment



Reference genomes of cultivable species from 4 ecosystems

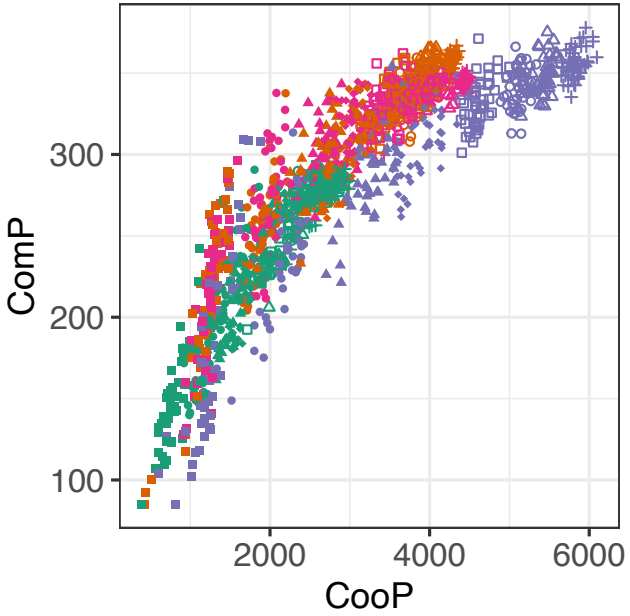
→ Automatic GEM reconstruction → 50 random communities of size X of each ecosystem

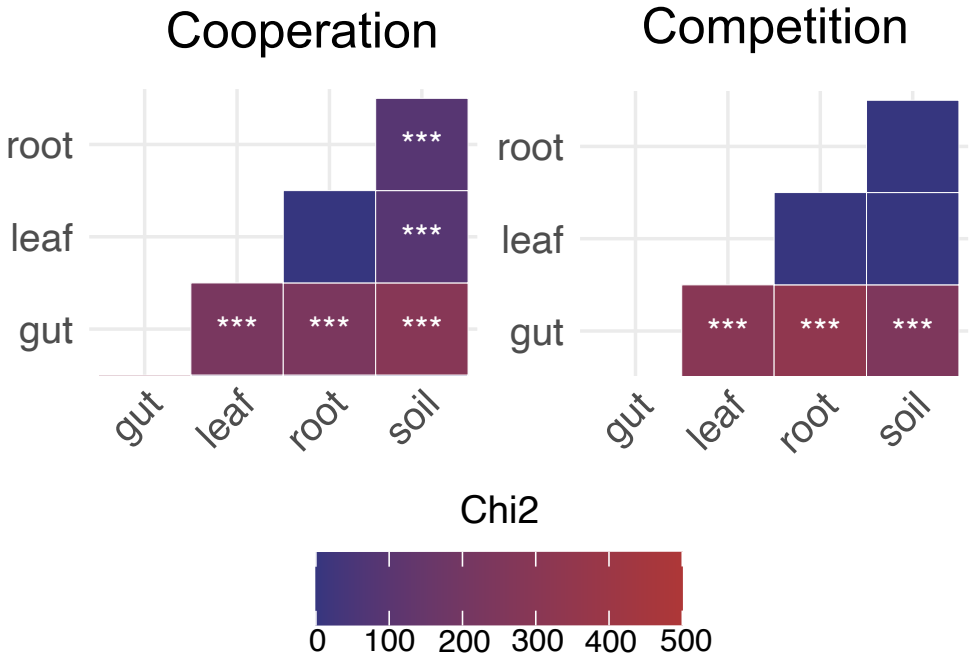
Goal

- Ecosystem consistency
- Added-value of removing/adding species

Are scores ecosystem consistent?

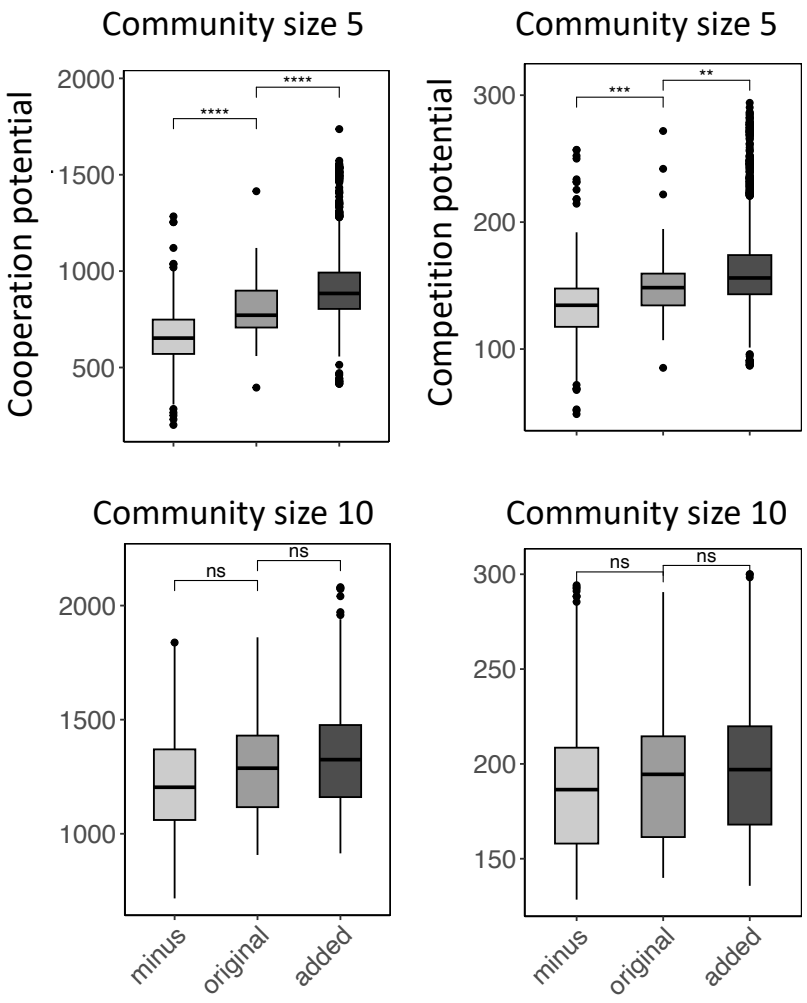
Specific seeds





- Leaf and root ecosystems are not distinguishable from cooperation and competition (polyopsony) scores

Added-value of removing/adding species on the gut microbiota



- Cooperation and competition (polyopsony) indexes are impacted with a taxonomic modification only for small size communities

Comparison with quantitative state-of-the-art tools for deciphering cooperation and competition in microbial communities

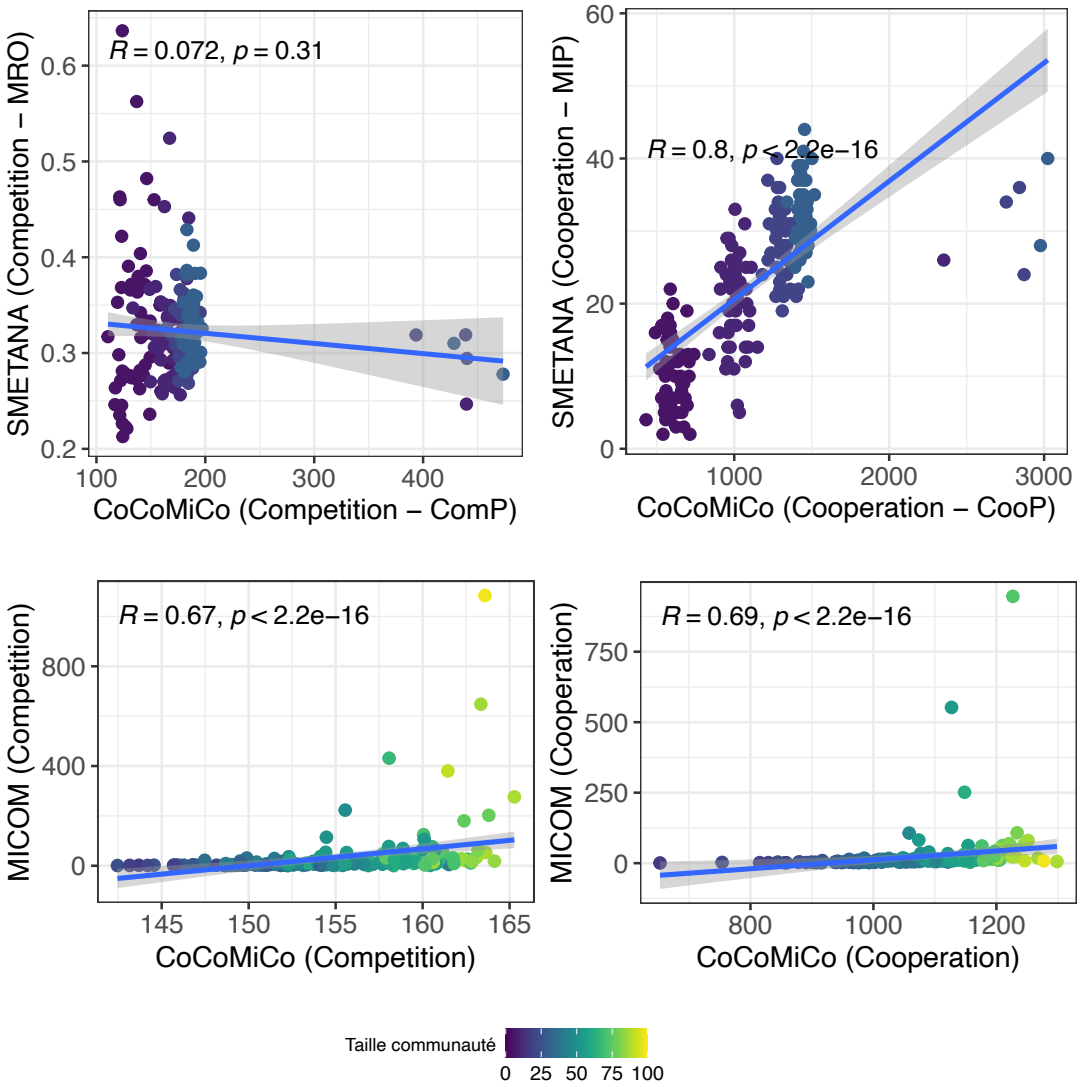
SMETANA	micom	
Up-to community size of 30	Up-to community size of 50	Up-to community size of 800
Cooperation and competition scores	No scores --> deduced from all pair combinations of KO process	Cooperation and competition scores
Up to 2 days	From 1h to a days	From 1s to hours

Deduced scores from MICOM accepted by Diener.

- CoCoMiCo tool is faster than SMETANA and MiCOM for characterizing cooperation and competition (polyopsony) in large-scale communities

Zelezniak et al., 2015; Diener, Gibbons, and Resendis-Antonio, 2020.

Cooperation and competition potentials



- Competition (polyopsony) and cooperation coherent with MiCOM

Discrete model conclusion

Reasoning based model of interactions

Strength

- Scalable mechanistic model on natural communities
- Faithful of logical rules inferred from expert

Weakness

Dependent of the topology of the network

Opportunities

In silico model usable for screening cooperation and competition at coarse grain unknown communities

Threats

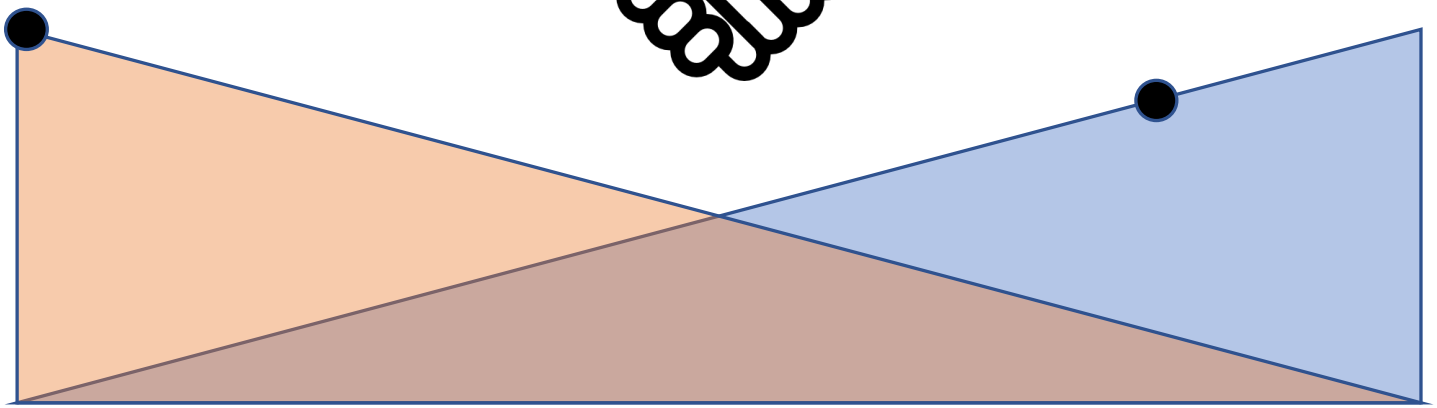
Risk of false-positive

Can we go further with discrete models?

Hybrid approach of the metabolism

Numerical accuracy
Small communities
Deep understanding

Scalability
Large communities
Screening process



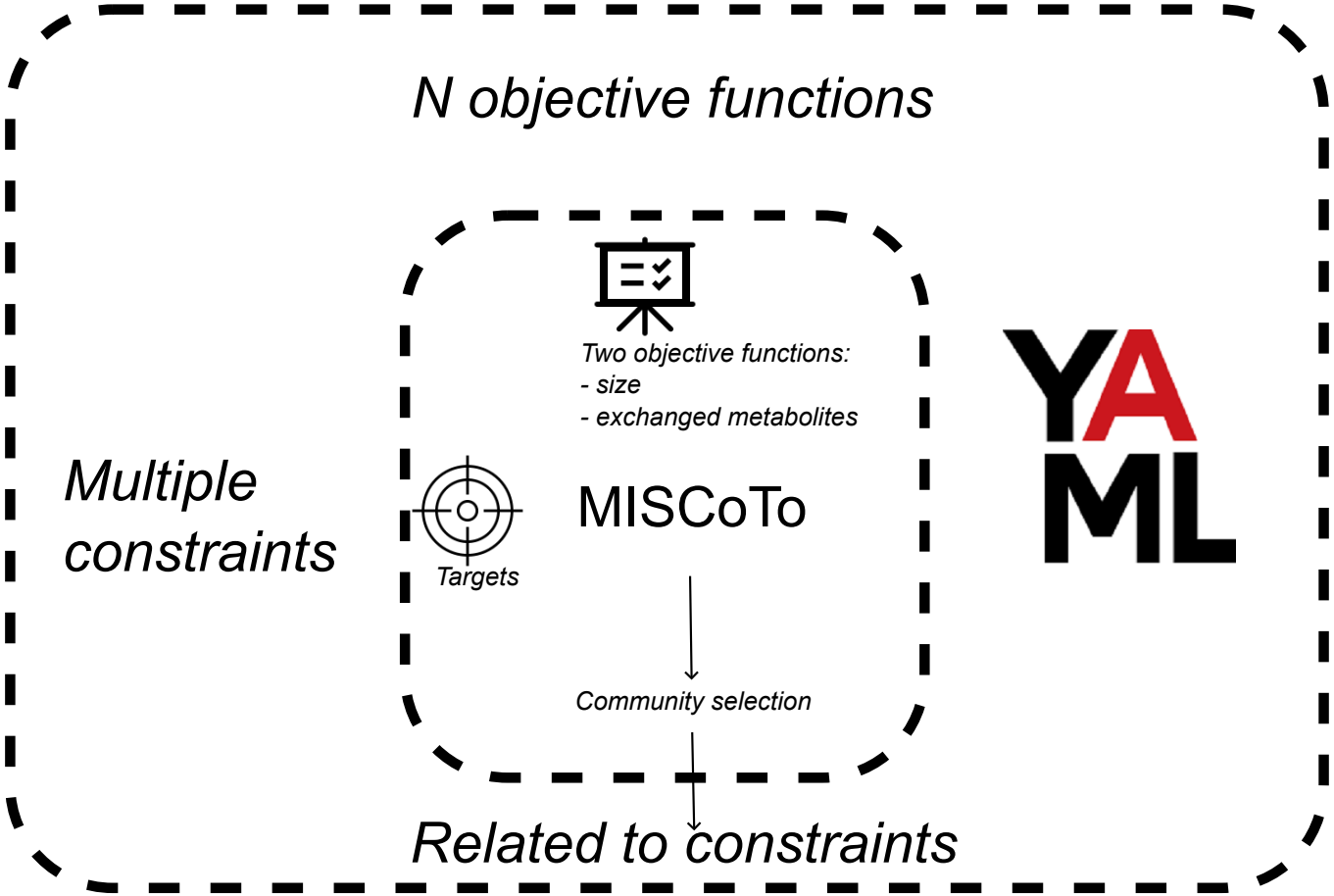
Numerical models

Discrete models

- Not always need to create hybrid model for inferring community behaviors
- Can we improve accuracy to discrete model ?

Adding *a priori* constraints

- Reduce CoCoMiCo output



Frioux et al., 2018.

Example of adding *a priori* constraints

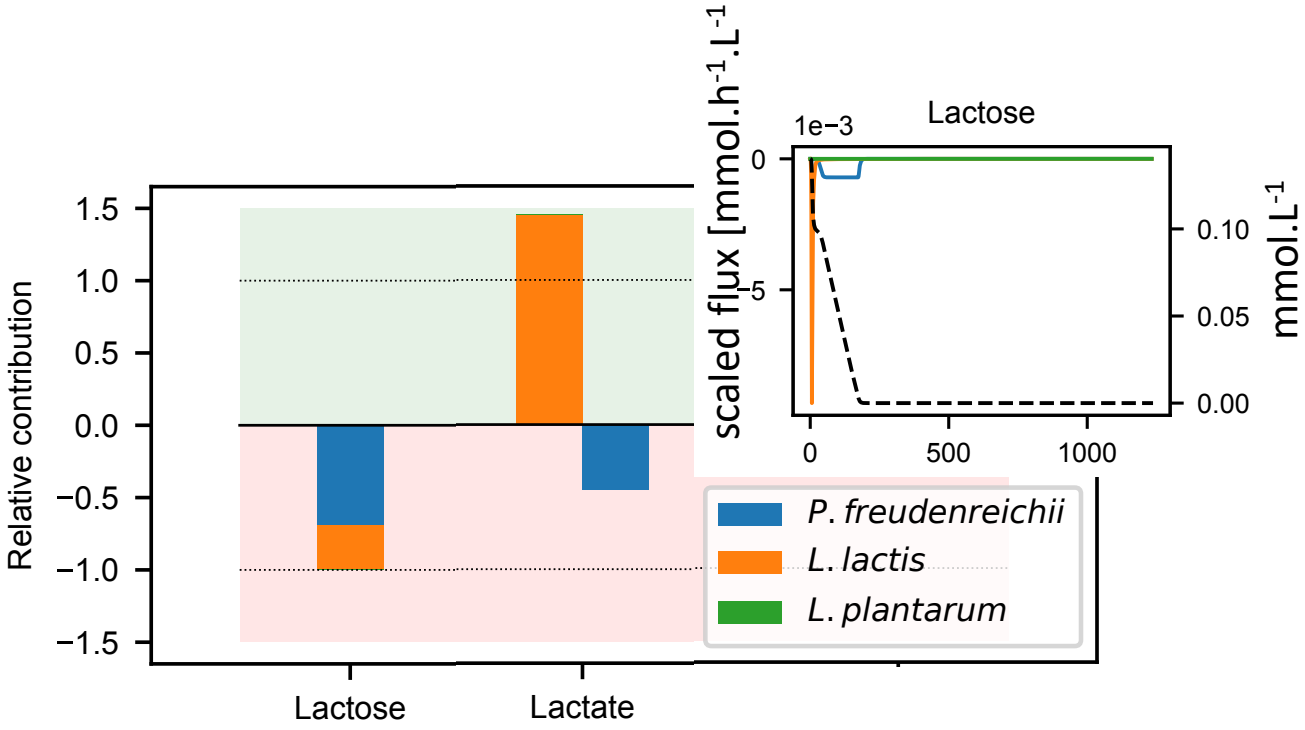
For a community composed of 6 taxa, which community guaranty the largest metabolic scope with the lowest number of metabolic exchange ?

```yaml
# Size constraints
SIZE:
  size: 6

  # Optimisation
OPTIMISATION:
  - max
  - min

# Objectif
OBJ:
  - scope
  - exchanged
```

Cooperation and competition properties use as optimization criteria

Adding temporal information: context

- Temporal segregation for lactose consumption

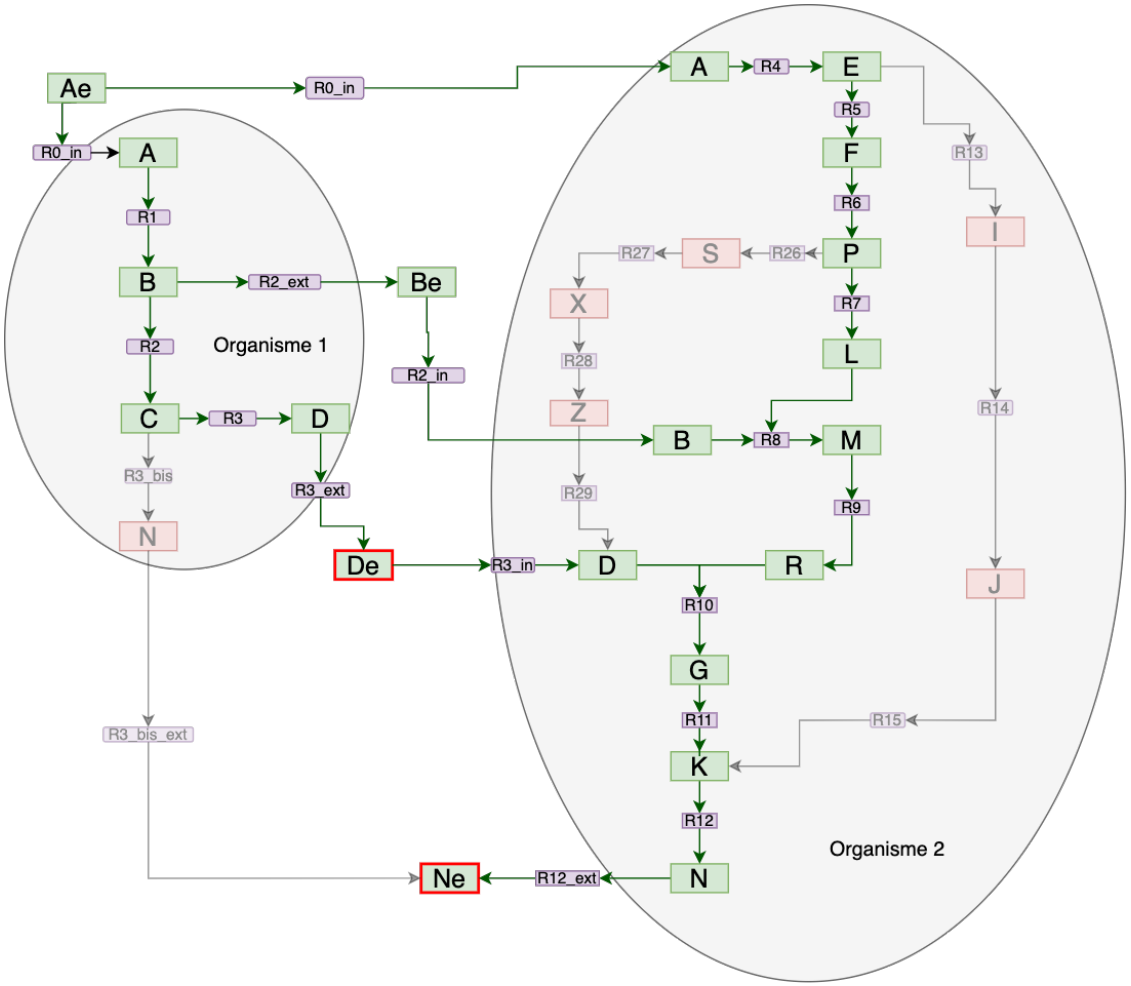


Goal

Based on the same formalism of the discrete model, we use temporal operator to characterize bacterial communities

Example of utilization: Explication

Reasoning about actions (RAC) and changes extends our formalism to temporal formalism using linear temporal logic.



time 0 **R0_in** activated

time 1 **R1_in**, **R4** activated, E,B producible

time 2 ...

Highlight the need of an potential exchange metabolite, **Be**, in order to produce **Ne** given that **De** is observed.

Conclusion and perspectives

Hybrid approach for studying bacterial interactions

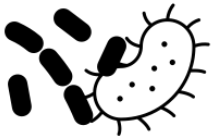
Natural community

CoCoMiCo software

Reasoning-based modeling

Interaction properties
Potenial interactions

Controlled community



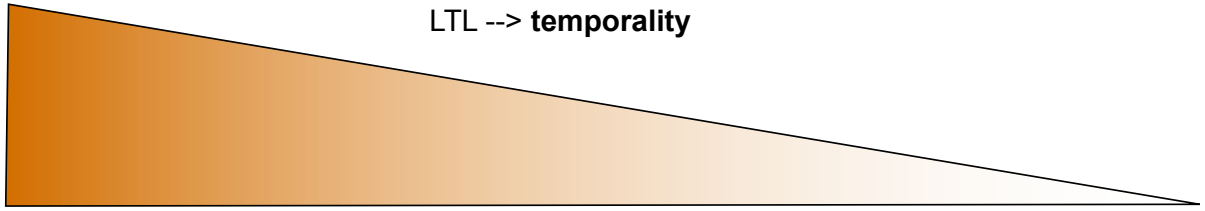
Tango **Implementation**

Numerical-based modeling

Temporal segregation
Potential interactions

SelCOM --> interaction properties as optimization constraints

LTL --> **temporality**



+ Scalability -

Conclusion and perspectives

Hybrid approach for studying bacterial interactions

Natural community

CoCoMiCo software

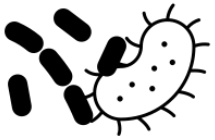
Reasoning-based modeling

Interaction properties
Potenial interactions

Validate bacterial interactions

Metatranscriptomics data

Controlled community



Tango **Implementation**

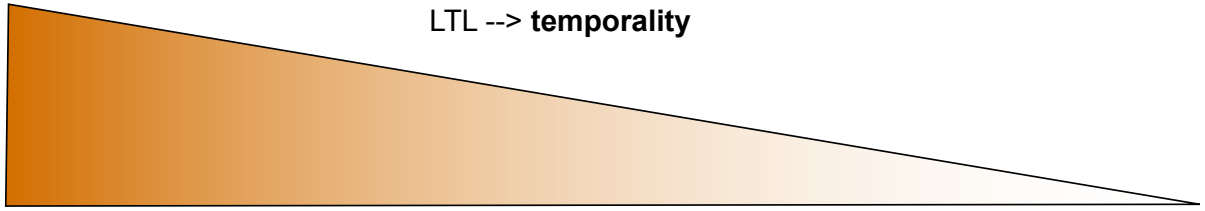
Numerical-based modeling

Temporal segregation
Potential interactions

SelCOM --> interaction properties as optimization constraints

Software development

LTL --> **temporality**



+ Scalability -

STLO

- David Sherman
- Clémence Frioux
- Simon Labarthe
- Coralie Muller
- Pleiade team

- Hélène Falentin

Thanks for your attention