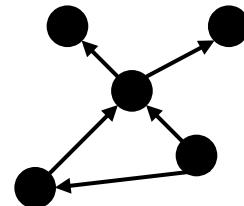
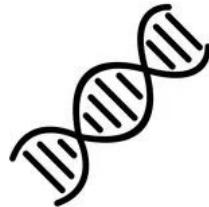


Functional role of the gut microbiota

Through the modelling of metabolic networks via omics analysis, and applications on metabolic diseases

The slides and results are partially provided by A. Weber (PhD in Bioinformatics,
Co-supervised by H. Soula)

Motivation



**Environme
nt**

From genomes/
metagenomes

Build metabolic
networks and
study properties

Link the networks to
various phenotypes/
environmental
variables

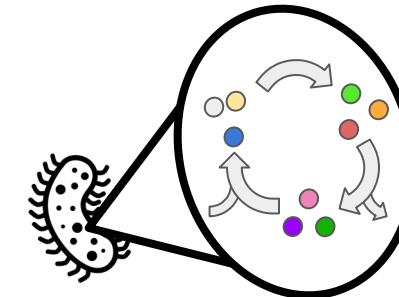
What is studied?

- **prokaryotic species vs environmental variables**
- **human gut microbiota vs clinical variables**



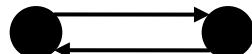


Metabolic networks (1)



Metabolism : the set of all chemical reactions

Networks : Objects (**nodes**) linked by links (**edges**).



Directed



Undirected

Eg: Social networks

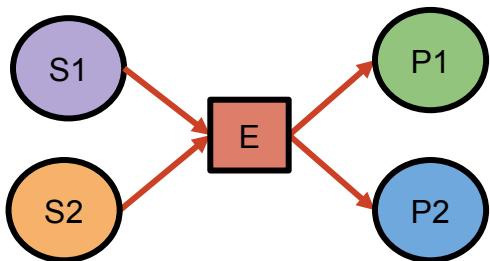
- nodes: people
- edges: social interactions



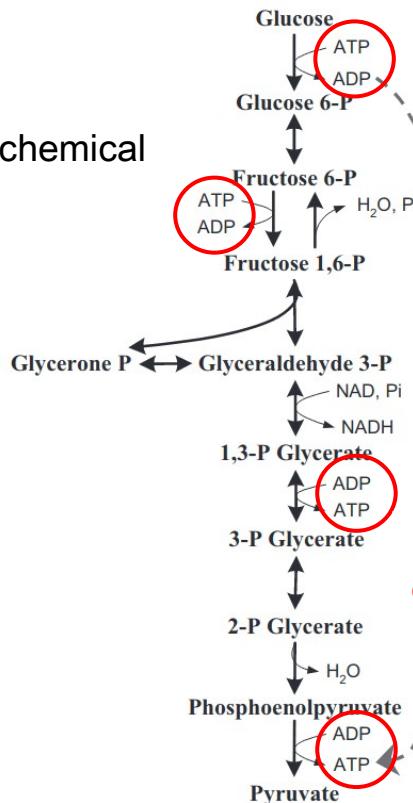
Metabolic networks (2)

Metabolic networks : Representation of all chemical reactions

Eg: Chemical reaction:

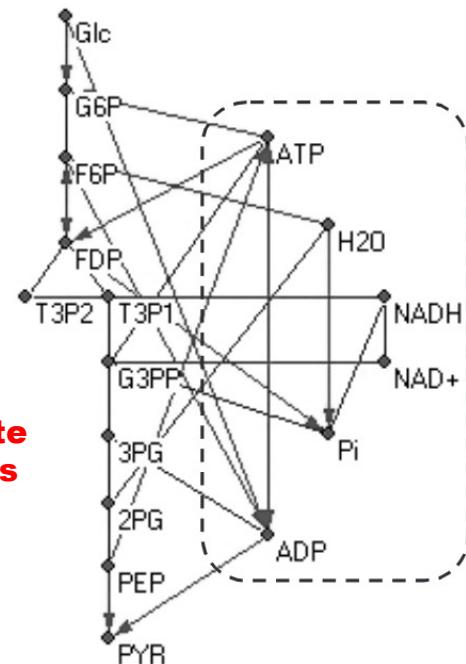


Metabolites and enzymes (**nodes**),
Substrate/products (**edges**)



Classic biology representation of
a metabolic pathway (Glycolysis)

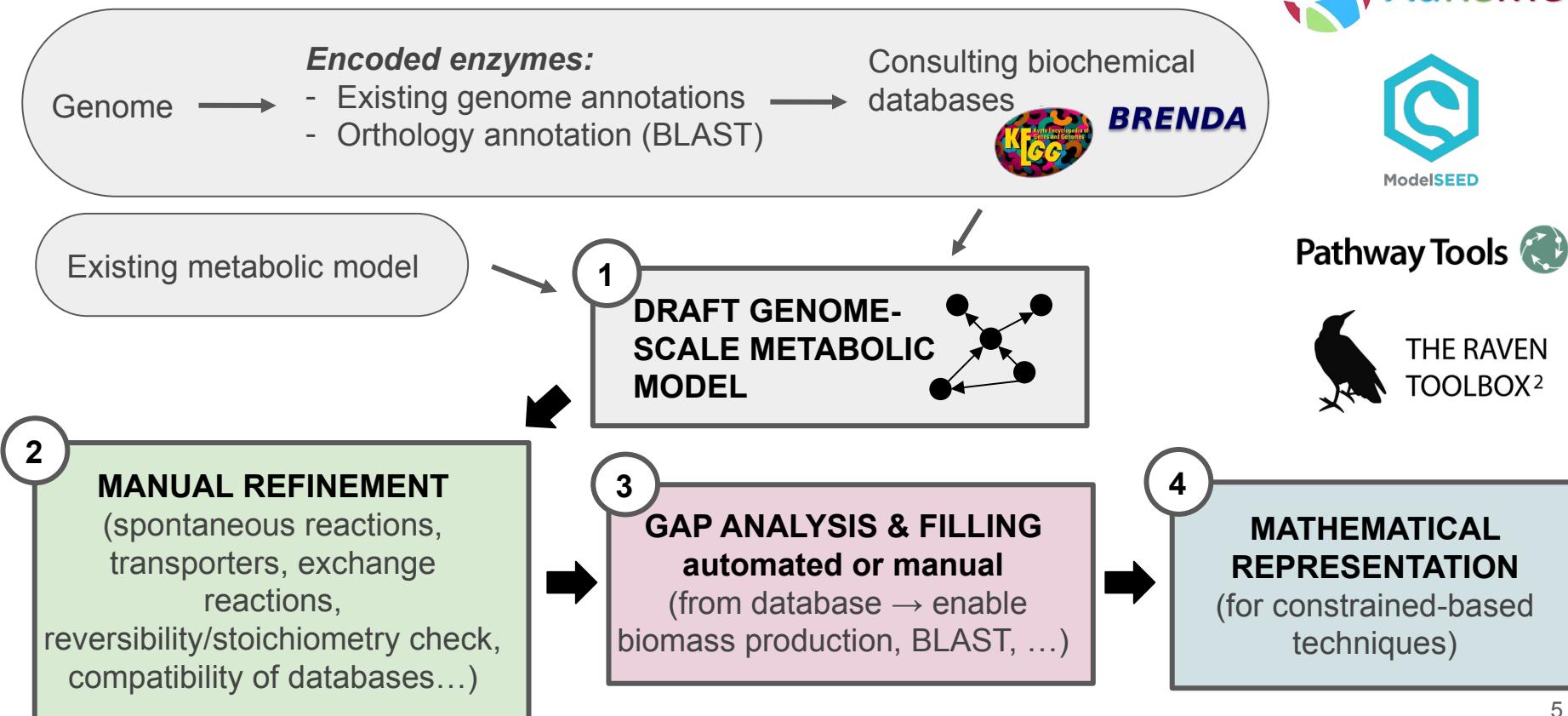
Ma and Zeng, 2003



Repeate
d nodes

Network representation

Metabolic network reconstruction approaches

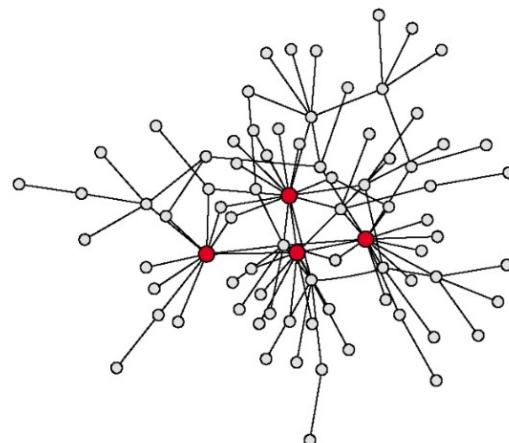


Main metabolic network analysis strategies

Topological approach

Purely based on the **structure (topology)** of the network

Impact of structure on cell

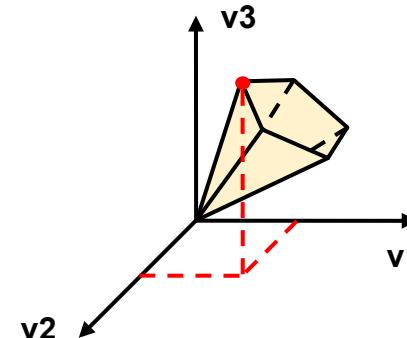


Constraint-based approach

Topology + stoichiometry (quantities of substrates/products)

Predicting **metabolic fluxes**

Needs **strong constraints** to find the optimal solution in the solution space



Constraint-based approaches

Flux Balance Analysis (FBA)

Steady-state of input and output of biomass:

**Concentratio
ns**

$$\frac{dx}{dt} = S \cdot v = 0$$

**Stoichiometry
coefficients matrix**

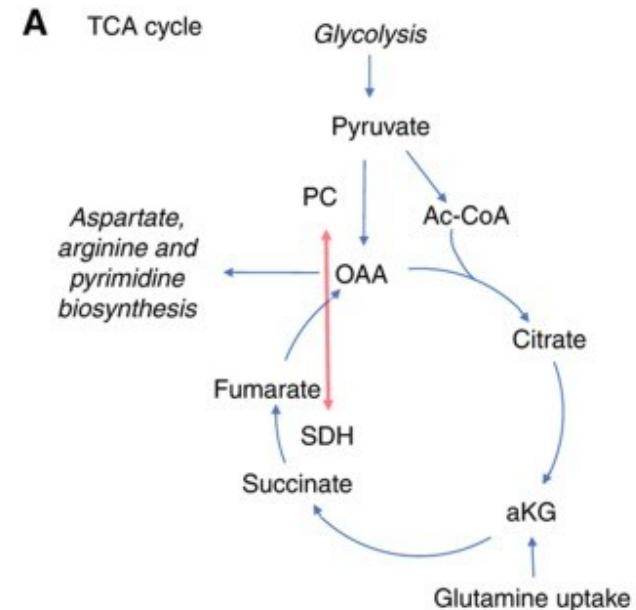
**Fluxes of
reactions**

Other constraints: $\alpha_i \leq v_i \leq \beta_i$

Objective function **Z** to optimise (biological objective) in the solution space:

$$Z = c^T \cdot v$$

**Vector of contribution to objective
function**



Example: Identification of essential genes in cancer cells and their combinations

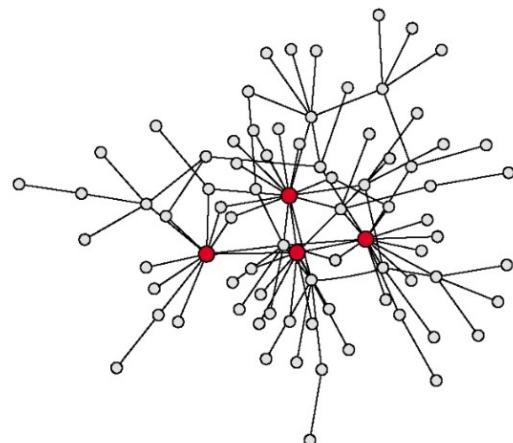


Main metabolic network analysis strategies

Topological approach

Purely based on the **structure (topology)** of the network

Impact of structure on cell



Constraint-based approach

Topology + stoichiometry (quantities of substrates/products)

Predicting **metabolic fluxes**

Needs **strong constraints** to find the optimal solution in the solution space

v3

- Metabolite production
⇒ precise networks
- Biomass objective function ?
- Harder to apply on a wide scale
- Strong constraints

Metabolic network reconstruction

List of a species' genes

List of microbiota genes
(KEGG* Orthology KO numbers)

ENZYME: 2.7.2.4	
Entry	EC 2.7.2.4
Name	aspartate kinase; aspartokinase;
Class	Transferases; Transferring phosphorus-containing groups; Phosphotransferases with a carboxy group as acceptor <small>BRITE hierarchy</small>
Sysname	R00480
Reaction (IUBMB)	ATP + L-aspartate = ADP + 4-phospho-L-aspartate [RN:R00480]
Reaction (KEGG)	[R00480]

Enzyme entries in KEGG* database
(associated to ECs)

REACTION: R00480	
Entry	R00480
Name	ATP:L-aspartate 4-phosphotransferase
Definition	ATP + L-Aspartate <-> ADP + 4-Phospho-L-aspartate
Equation	C00002 + C00049 <-> C00008 + C03082

Substrates Products

CC(=O)N[C@@H](C)COP(=O)([O-])[O-] <-> CC(=O)N[C@@H](C)COP(=O)([O-])O

Reaction entries in KEGG* database
(associated to R numbers)

1113
2.7.2.4
...
Enzyme Codes (EC)

List of enzyme codes (EC)
encoded by the genes from
KEGG*

* Kyoto Encyclopedia of Genes and Genomes



L-aspartate

2.7.2.4

4-phospho-L-aspartate

Construction of edges for each reaction

Ubiquitous metabolites removed : H_2O , ATP, ADP, NAD⁺, NADH, NADPH, NADP⁺, CO_2 , NH_3 , SO_4^{2-} , thioredoxin, PO_4^{3-} , PP_i , H^+

Network topological analysis

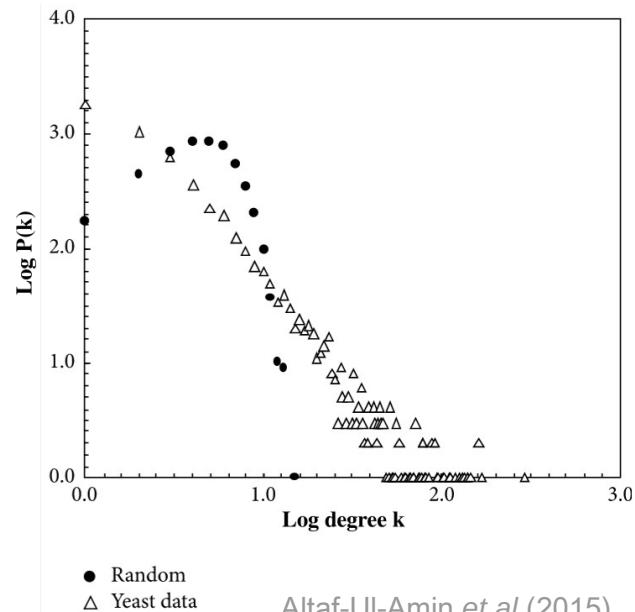
$$\text{Edge density} = \frac{\# \text{ edges}}{\# \text{ nodes}}$$

Mean number of reactions per metabolite

Degree of a node: # edges connected to the node
 Number of reactions of a metabolite

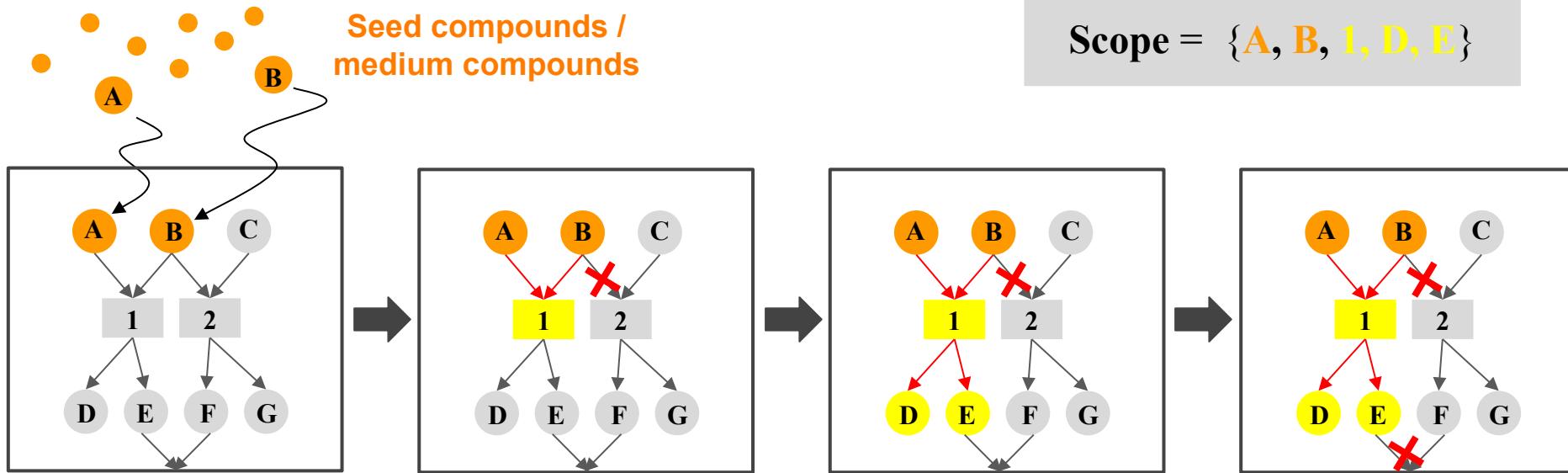
Degree distribution: probability distribution of the degrees
 Probability distribution of the number of reactions for a metabolite

For networks where nodes are metabolites and edges reactions



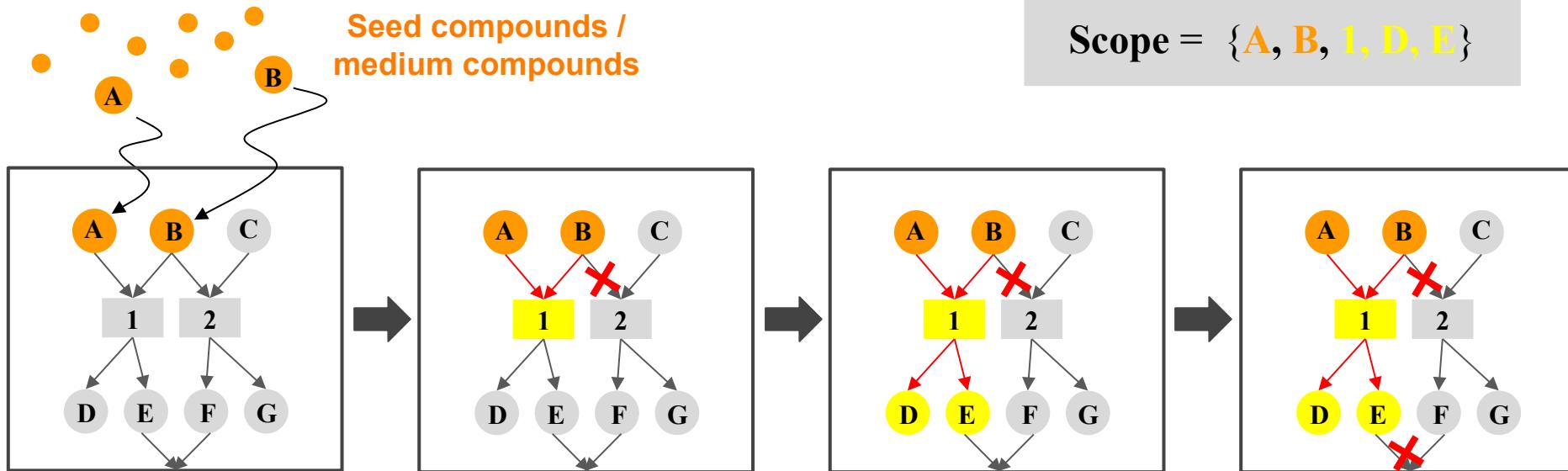
➡ Not very interpretable in terms of metabolic function

Network topological analysis: Scope



Scope of a network: Nodes (metabolites and enzymes) reachable in the network from a set of input nodes from the environment.

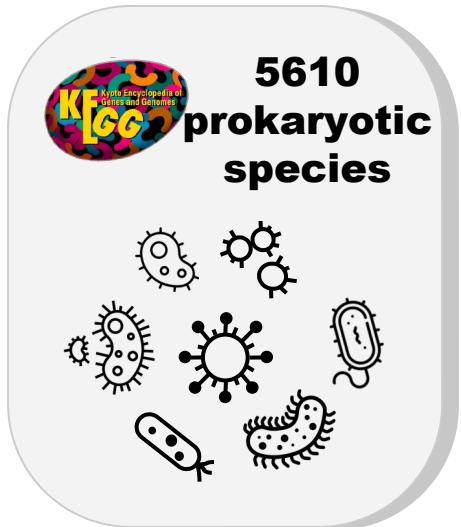
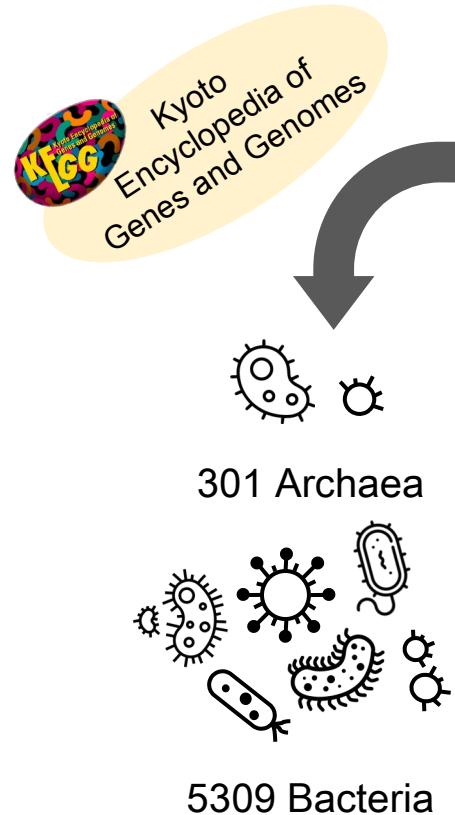
Network topological analysis: Scope



- Gives potential biosynthetic capability of a network
- Very medium dependent

Functional prediction of environmental variables using metabolic networks

Dataset (1)



**Growth temperature and classes
for 3392/5610 species**
(190 Archaea, 3202 Bacteria)



76
Hyperthermophiles
(> 80°C)



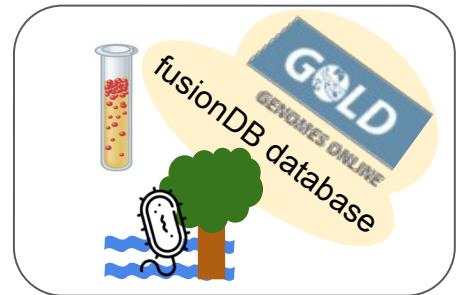
188
Thermophiles
(45-80°C)



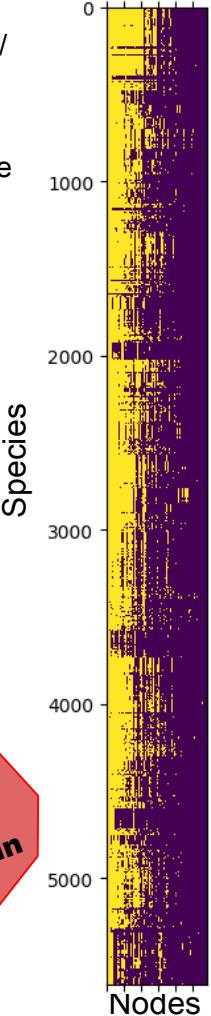
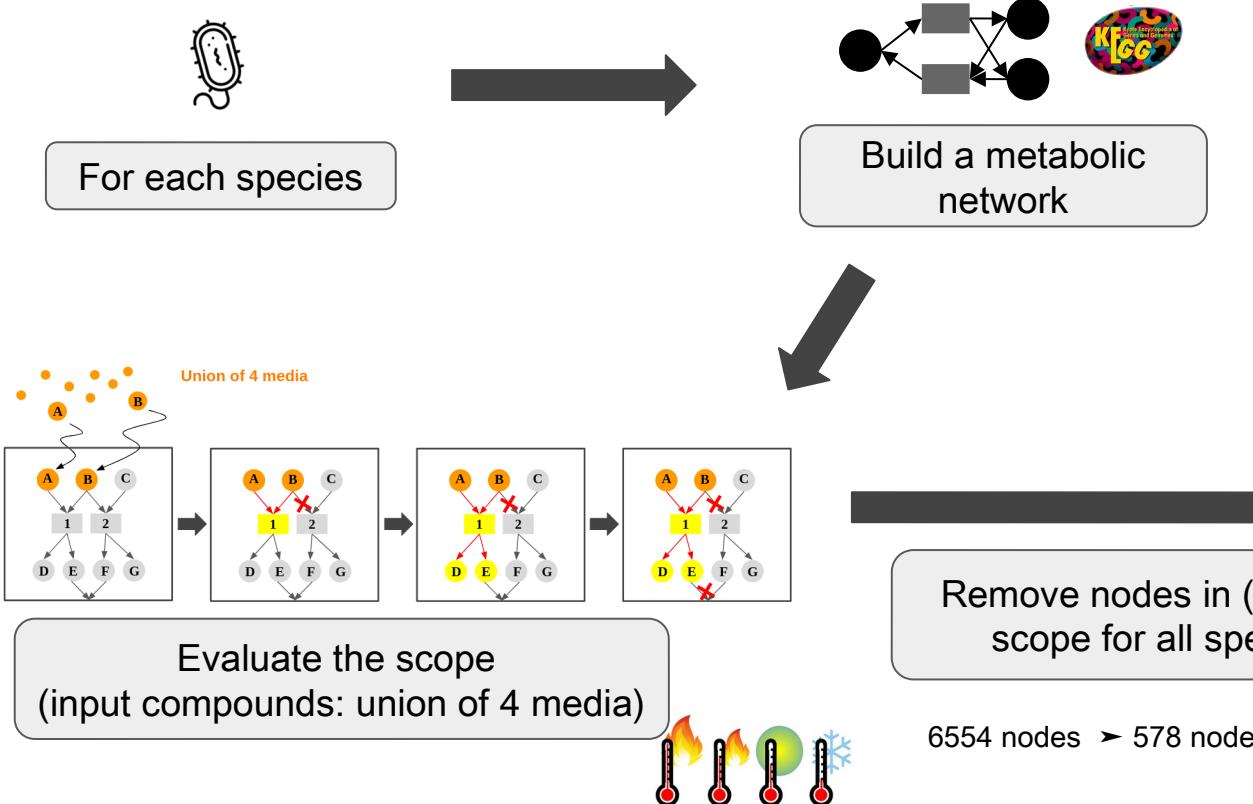
2910
Mesophiles
(25-45°C)



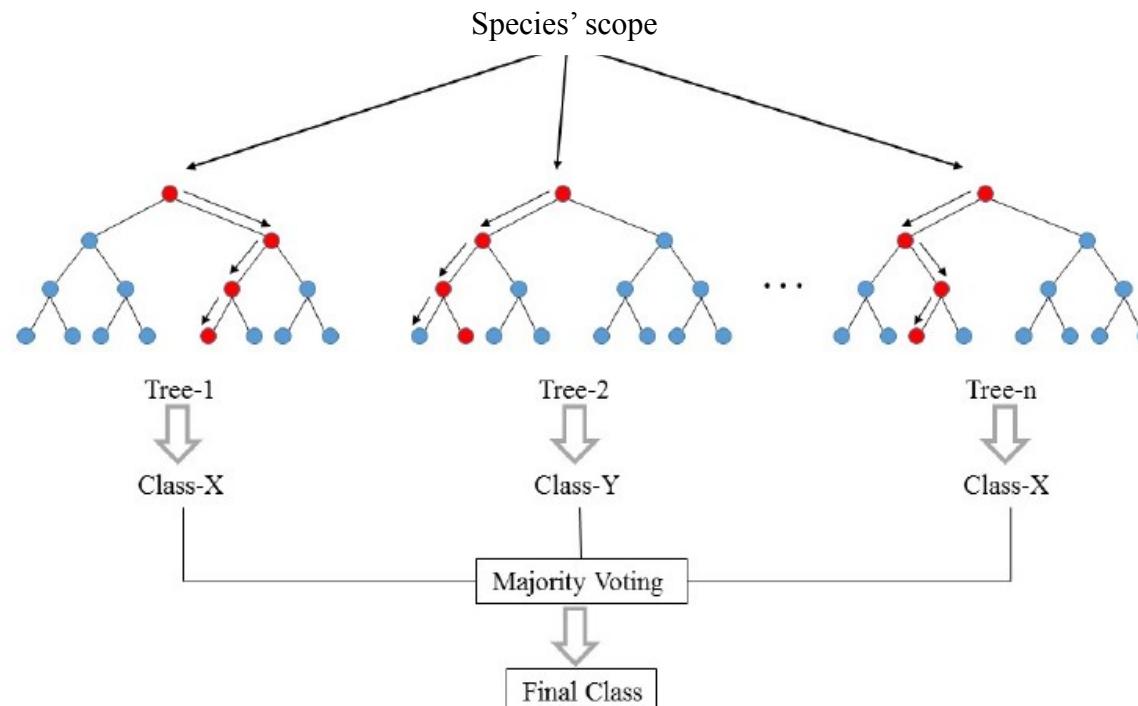
218
Psychrophiles



Scope of the networks (1)



Random forests to predict environmental classes

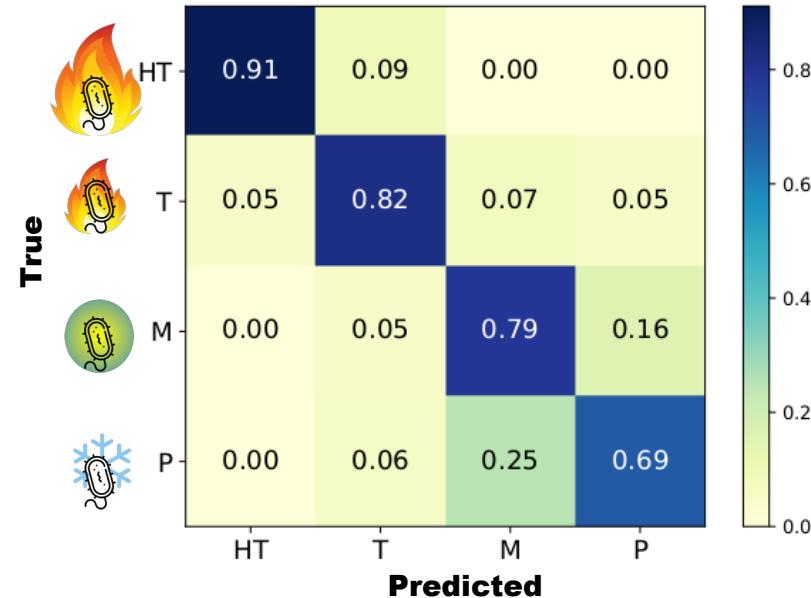


- Train the forest to **predict the environmental variables**
- Division of dataset into **training** and **testing** sets, and a small **validation** set
 - 1000 trees, 300 times (**cross-validation**)
- **Gini impurity**: measures the quality of split in the tree
- Classes are weighted

Random forests: growth temperature class (1)



- **3392 species** with temperature information
- **Balancing temperature classes:** 300 randomly selected mesophile species for each cross-validation fold (300)

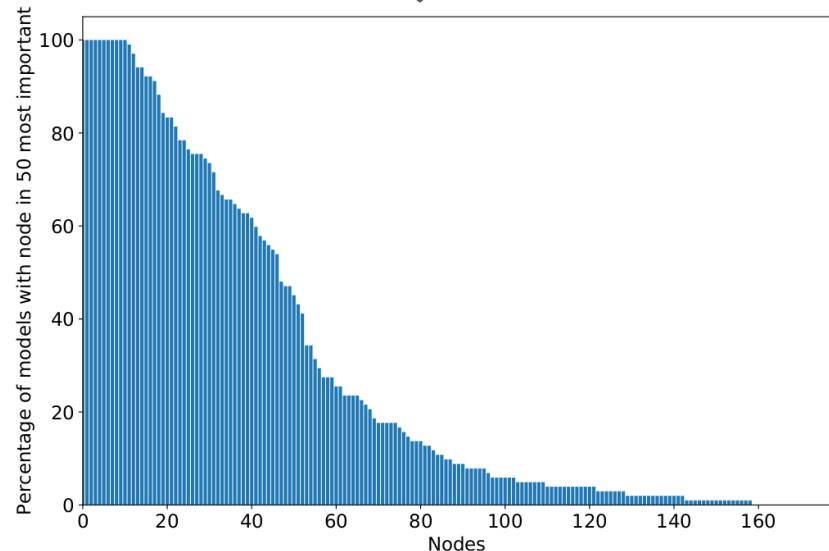
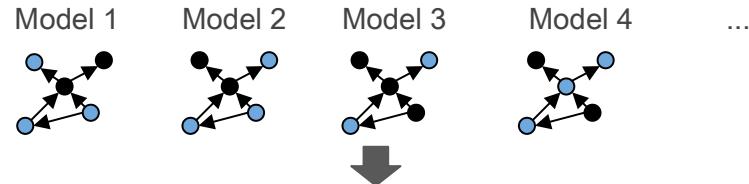


→ **Good accuracy per class**

- Weaker for Psychrophiles, often confused with Mesophiles

Random forests: growth temperature class (2)

Selection of the **50 most predictive nodes in scope** for each cross-validation model

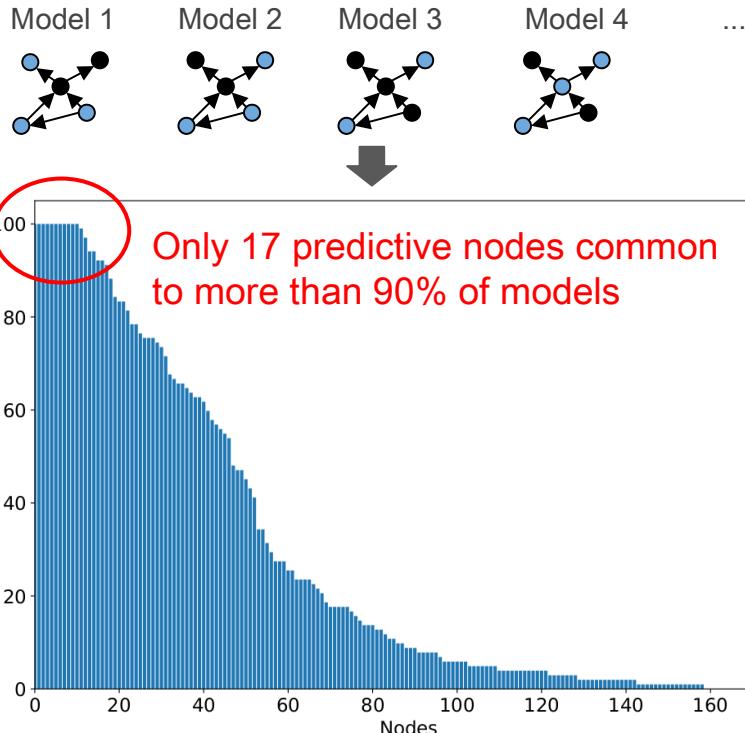


What are these nodes?



Random forests: growth temperature class (2)

Selection of the **50 most predictive nodes in scope** for each cross-validation model



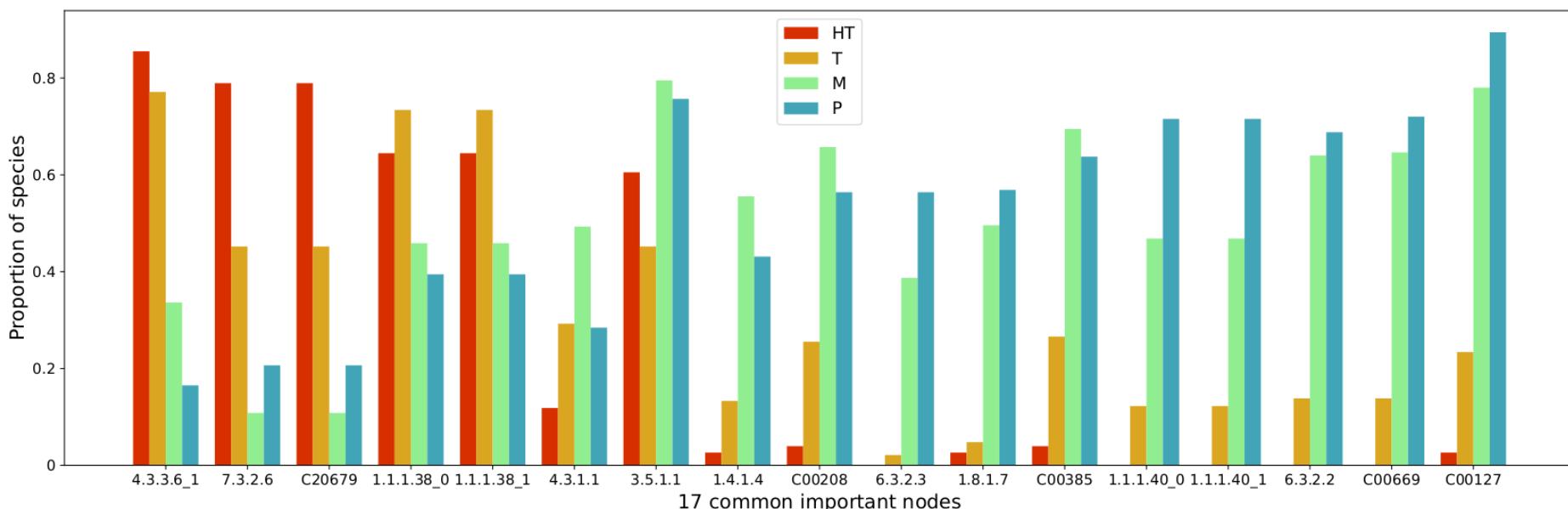
KEGG pathway	Mean number of nodes
Alanine, aspartate and glutamate metabolism	8.70 ± 1.92
Pyruvate metabolism	8.14 ± 1.48
Cysteine and methionine metabolism	7.12 ± 2.01
Glutathione metabolism	6.33 ± 0.84
Carbon fixation in photosynthetic organisms	5.34 ± 1.21
Glycine, serine and threonine metabolism	5.21 ± 1.91
Arginine biosynthesis	4.69 ± 1.14
Purine metabolism	3.59 ± 1.14
ABC transporters	3.12 ± 1.11

→ They belong to basic metabolic pathways

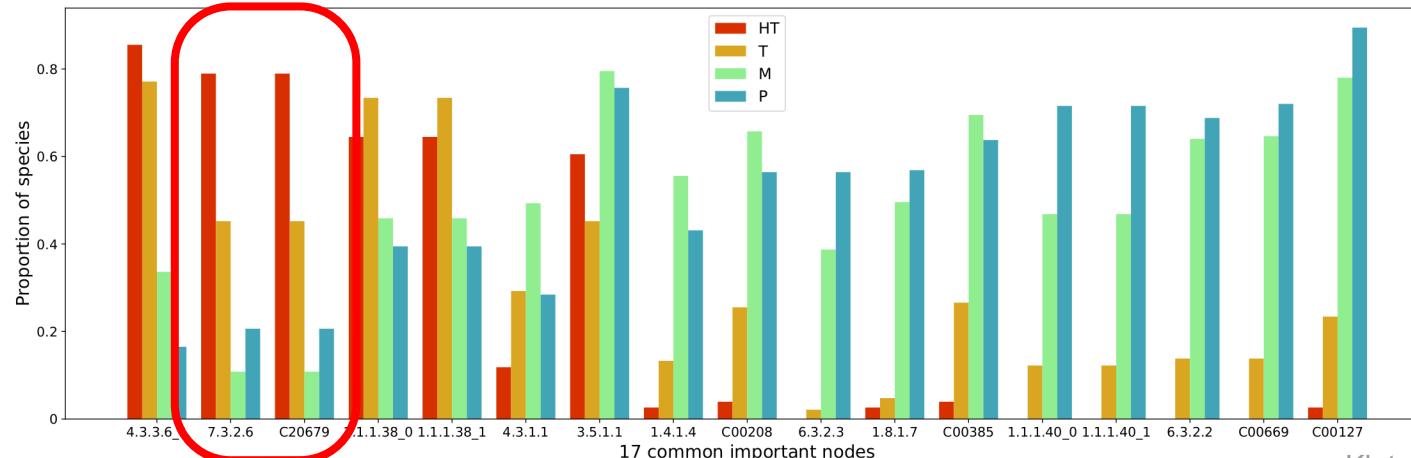


Random forests: growth temperature class (3)

17 predictive nodes common for models

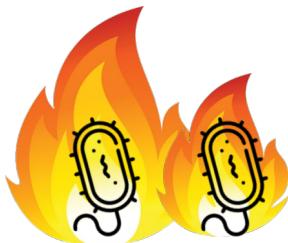


Random forests: growth temperature class (4)



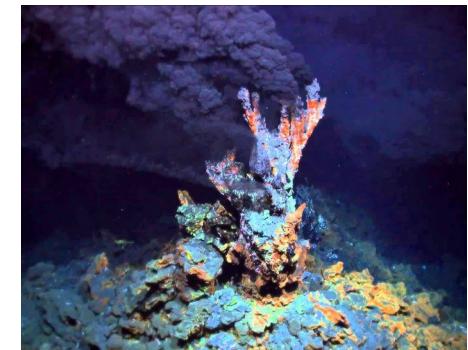
7.3.2.6: ABC-type tungstate transporter
C20679: Tungstate

Kletzin, Adams, 1996



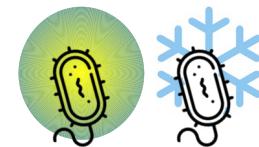
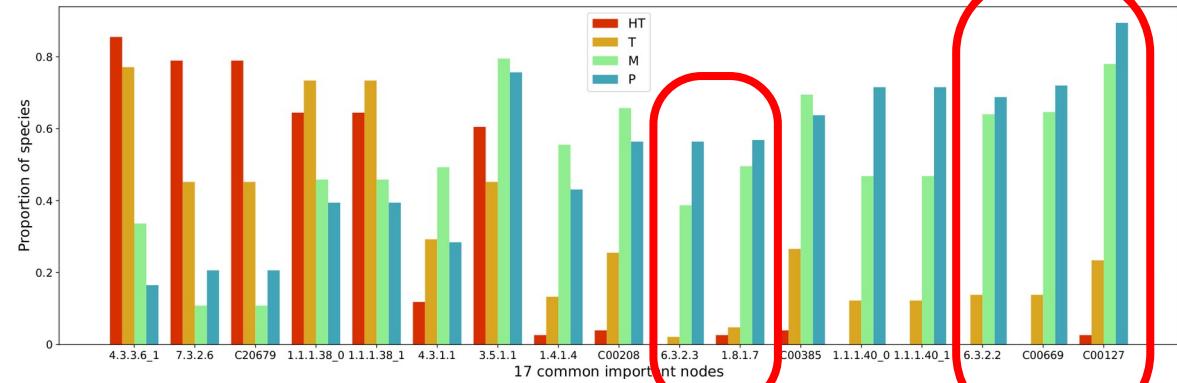
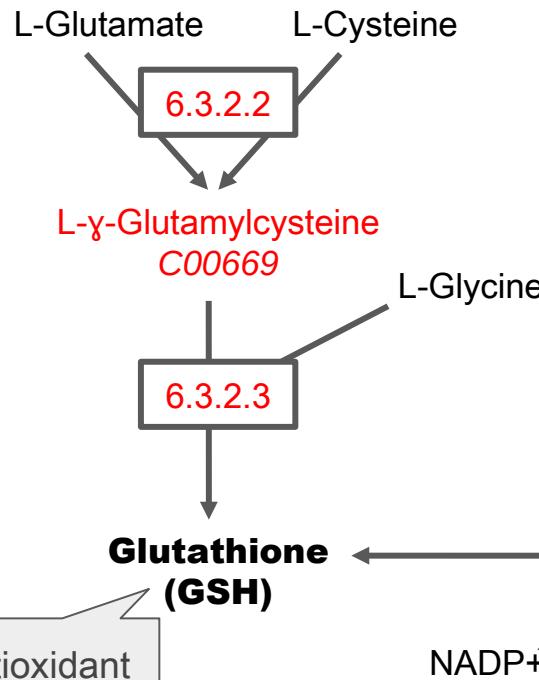
→ Tungstate metabolism ⇌ high temperature

→ Tungsten in **high concentration** in **thermophilic environments**
(hydrothermal vents, hot springs)

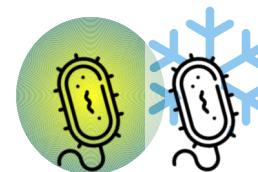
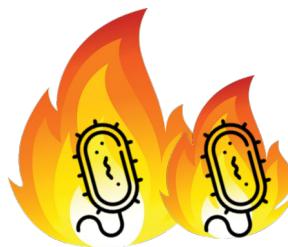


Random forests: growth temperature class (5)

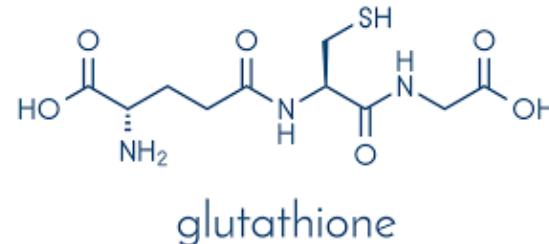
Glutathione



Random forests: growth temperature class (6)



- No glutathione
- Other antioxidant thiols
(coenzyme A)
 more stable at high temperatures



Artificial neural networks: growth temperature prediction

- **Growth temperature** normalisation:

$$\frac{T - T_{\min}}{T_{\max} - T_{\min}}$$

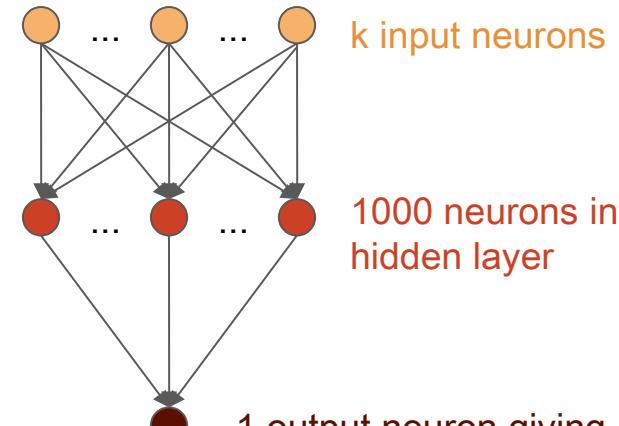
- Chose 300 random mesophiles out of 2910, to balance the classes
- Divided into train(66%)/test datasets
- Parameters:

- Learning rate = **1e-4** for 1000 epochs, then **1e-5** for 2000 epochs
- weighted MSELoss
- Adam optimizer
- 3000 epochs
- 1000 batch size

For k nodes in scope

*Linear function
Dropout ($p=0.2$)*

*Linear function
Sigmoid*

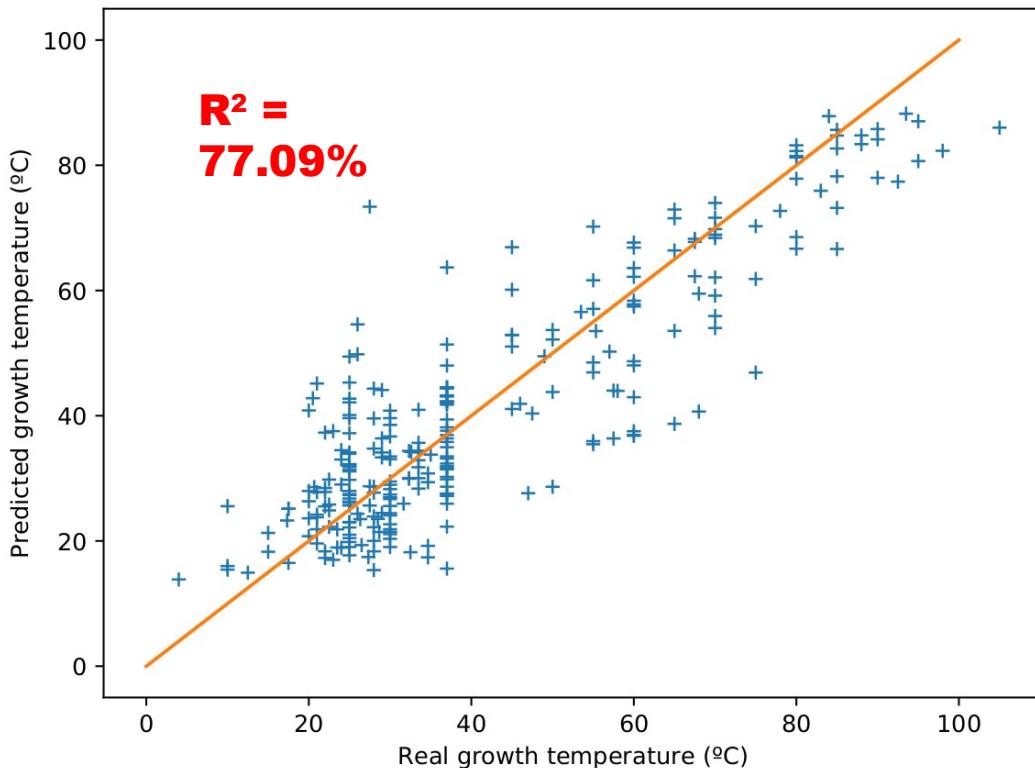


1 output neuron giving
the predicted
temperature

k input neurons

1000 neurons in
hidden layer

Artificial neural network: growth temperature prediction



- **Input:** scope matrix
- **Artificial neural network:** predicting growth temperature values
- **Balancing temperature classes:** 300 randomly selected mesophile species, weighted objective function

→ Can predict growth temperature from the scope

In summary

The scope is:

- An **efficient embedding**: reduces dimension and keeps sufficient information to:
 - ◆ Characterise growth temperature classes, habitat and oxygen tolerance (Random Forests)
 - ◆ Estimate growth temperature values (Artificial Neural Networks)
- **Metabolically interpretable**: pinpoints important metabolic molecules and pathways for the classification
 - ◆ Temperature class prediction:
 - Tungsten metabolism > warmer species
 - Glutathione metabolism > colder species

Gut microbiota metabolic reconstruction to assess severity of type 2 diabetes in obesity

Obesity and type 2 diabetes (T2D)



Obesity: accumulation of excess body fat that may impair health

$$\text{BMI} = \frac{\text{weight (kg)}}{\text{height}^2 (\text{m}^2)}$$

Approximated by the **Body Mass Index > 30 kg/m²** (BMI kg/m²) and **waist circumference**

(cm)

- ➔ Multifactorial (diet, gut microbiota...)
- ➔ Closely related



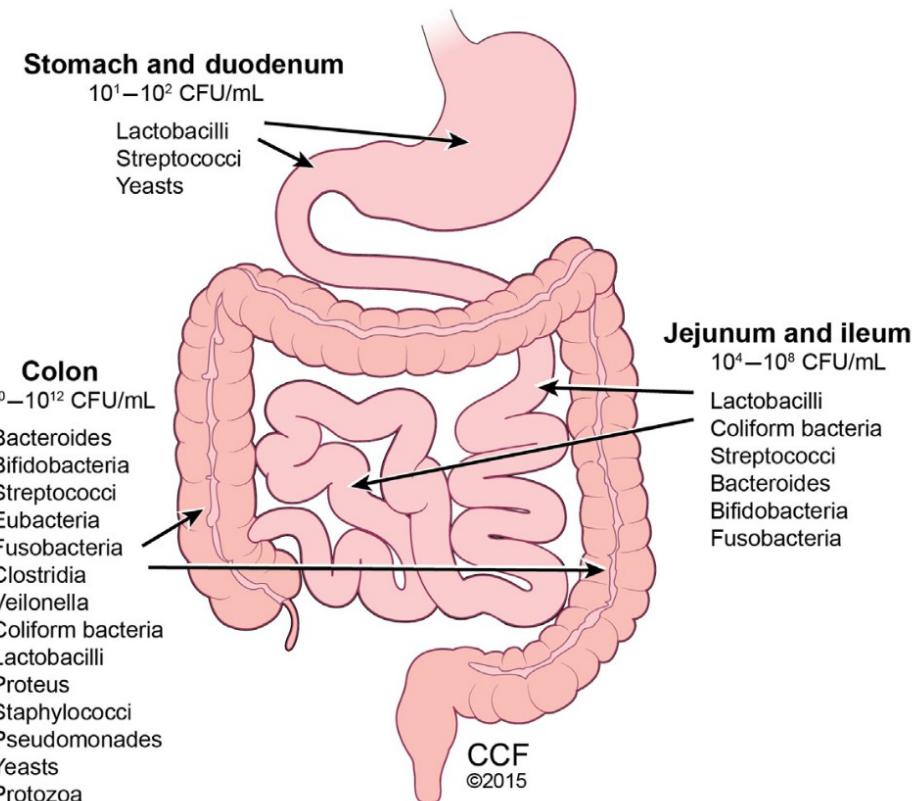
Type 2 diabetes: chronic fasting hyperglycemia due to defects in insulin sensitivity and/or insulin secretion

Diagnostic test	Diabetes cut-offs
Fasting plasma glucose	≥ 7.0 mmol/L
2-hour post-load plasma glucose after a 75g oral glucose tolerance test (OGTT)	≥ 11.1 mmol/L
HbA1c levels	≥ 6.5%
Random blood glucose, with diabetic signs and symptoms	≥ 11.1 mmol/L

T2D diagnosis

The human gut microbiota

- **Community of microorganisms**
(Bacteria, Eukaryotes, Archaea, viruses) **inhabiting the gut**
- $\sim 10^{13}$ bacteria in the gut Sender, Fuchs et al, 2016
- 200-1000 species per person Qin, Li et al, 2010
Yang, Pu et al, 2020
- **Roles:** help immunity, nutrient metabolism, vitamin biosynthesis, etc Jandhyala, Talukdar et al, 2015
- Modulated by diet, drugs, other environmental factors Yatsunenko, Rey et al, 2012
- Composition and abundance varies across the intestinal tract

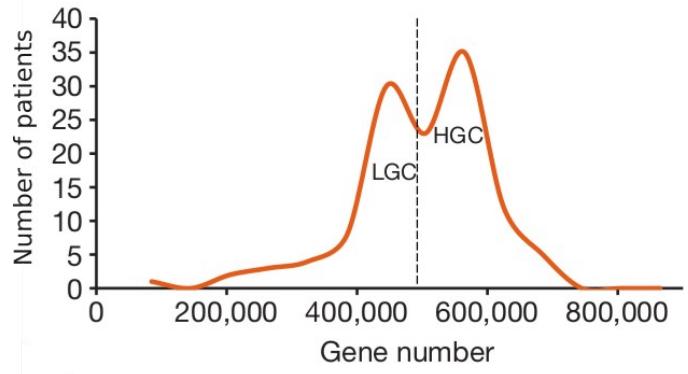




The human gut microbiota in obesity and T2D

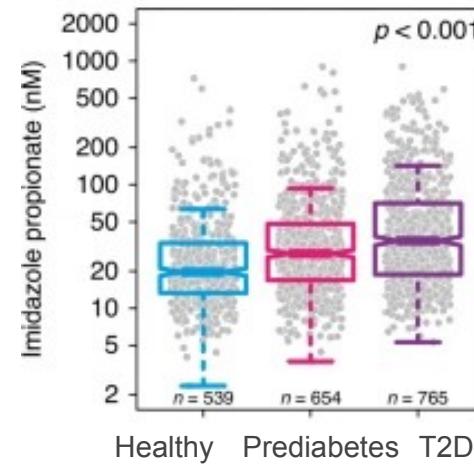


Alterations to the gut microbiota is associated to diseases



Low microbiota gene count (LGC) associated to more severe obesity-related phenotypes than high gene count (HGC) patients

Cotillard, Kennedy et al, 2013



Serum levels of imidazole propionate (metabolite produced by microbiota) increased in prediabetes and T2D

Molinaro, Bel Lassen et al, 2020

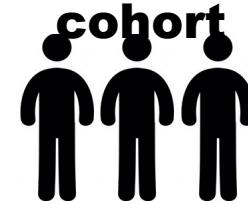
Dataset: clinical cohorts

Bariatric cohort



89 severe and morbidly obese women candidate to bariatric surgery

MetaCardis cohort



314 French severe and morbidly obese (BMI > 35kg/m²)

Fecal metagenomics on gut microbiota, and
morphological and clinical variables

Kim, Marchand, Henegar et al, 2011

Genser, Poitou-Bernert et al, 2015

Aron-Wisnewsky, Prifti, Belda et al, 2018



Karine Clément

Judith Aron-Wisnewsky

Christine Poitou-Bernert

Eugeni Belda



Molinaro, Bel Lassen et al, 2020

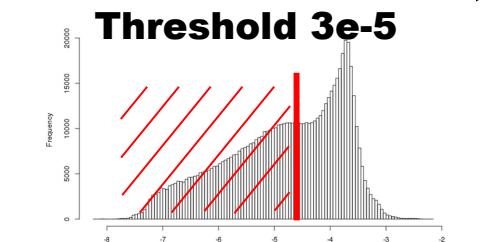
Le Roy, Debedat et al, 2019

Gut microbiota metabolic networks: reconstruction

	KO1	KO2	KO3	KO4
Patients	0,0	...		
	...			

KO abundance matrix

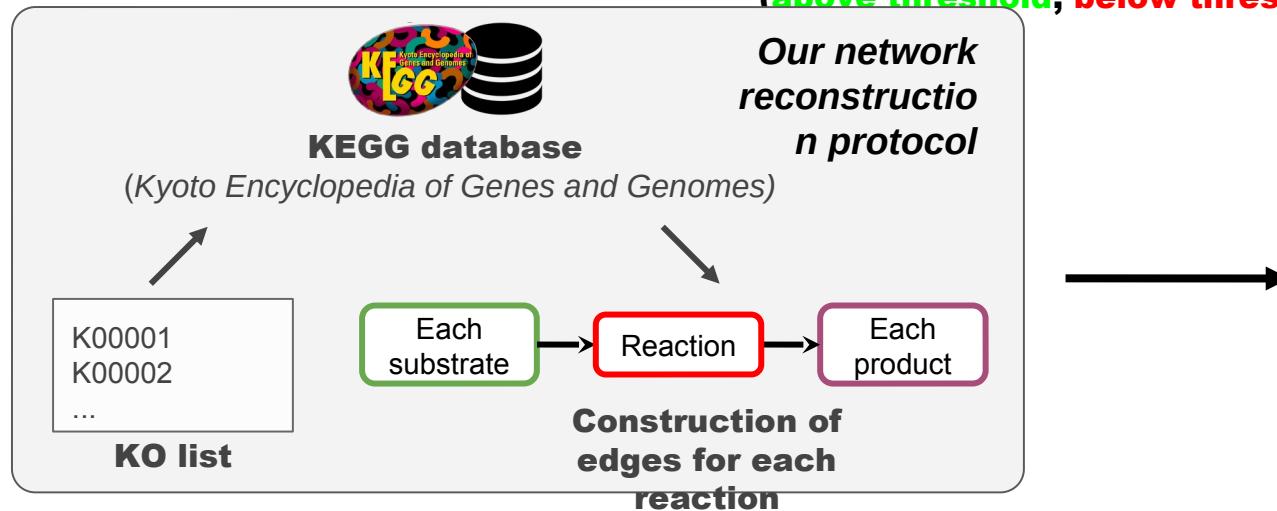
KO = KEGG Orthology
= orthology genes



	KO genes			
Patients	Red	Green	Green	Green
	Green	Green	Red	Green

Matrix of KO presence/absence in patients
(above threshold, below threshold)

Eugenio Belda

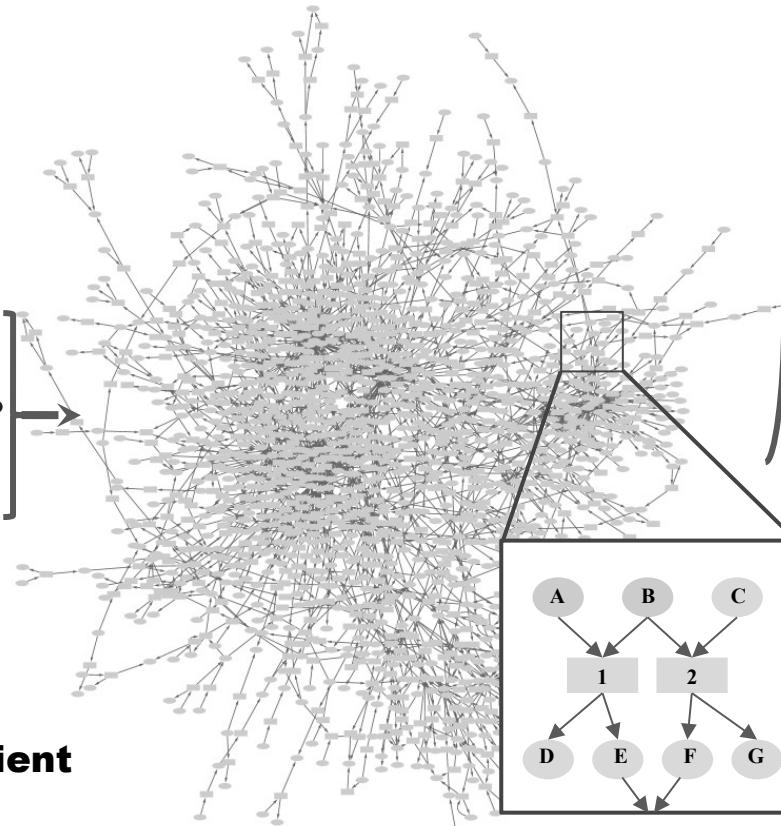




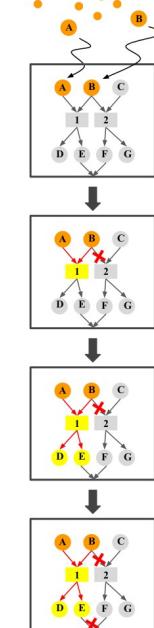
Gut microbiota metabolic networks: scope



**Microbiota
(community)
network per patient**

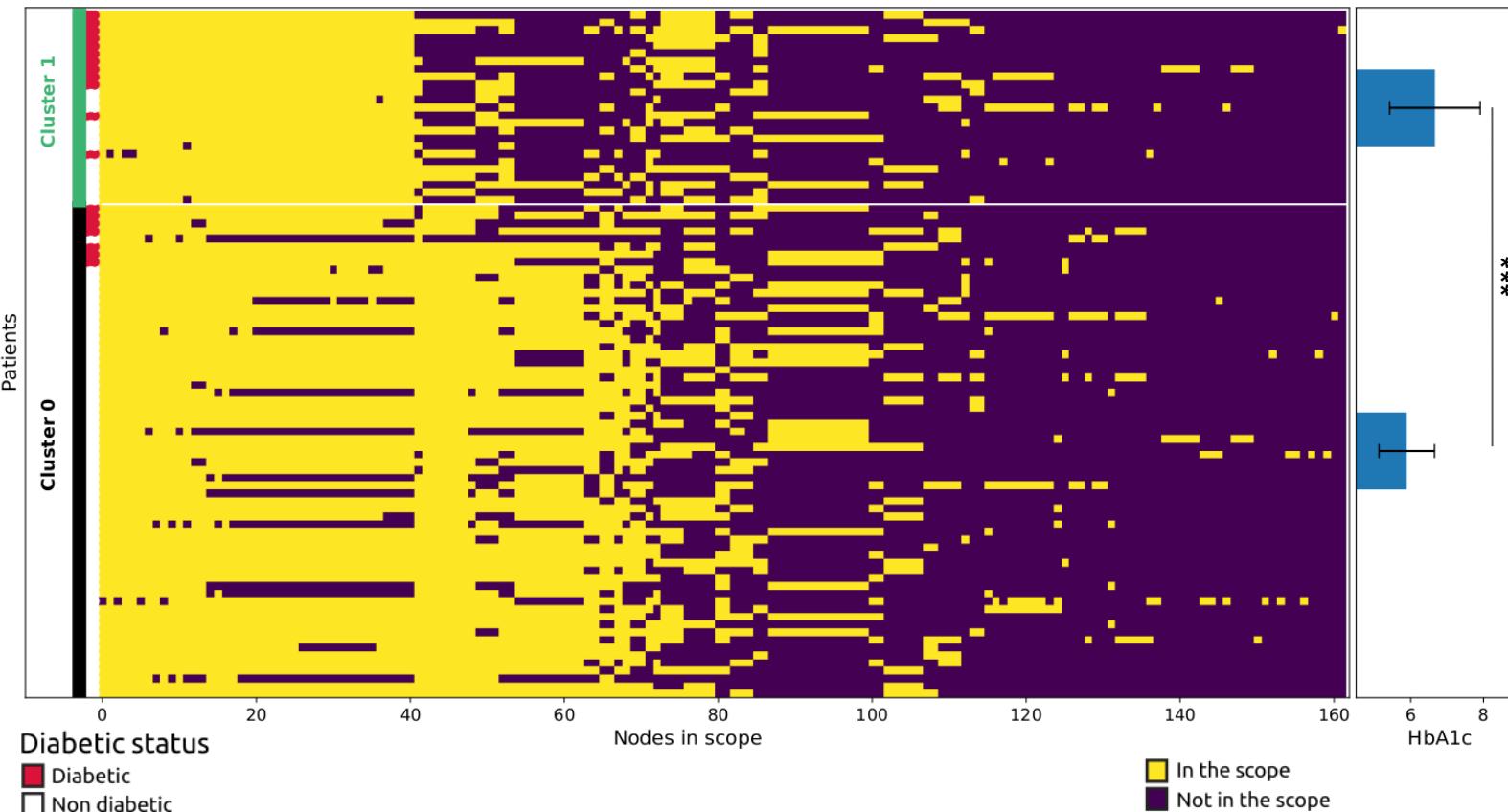


Dulbecco's
Modified Eagle's
medium (DMEM)
molecules



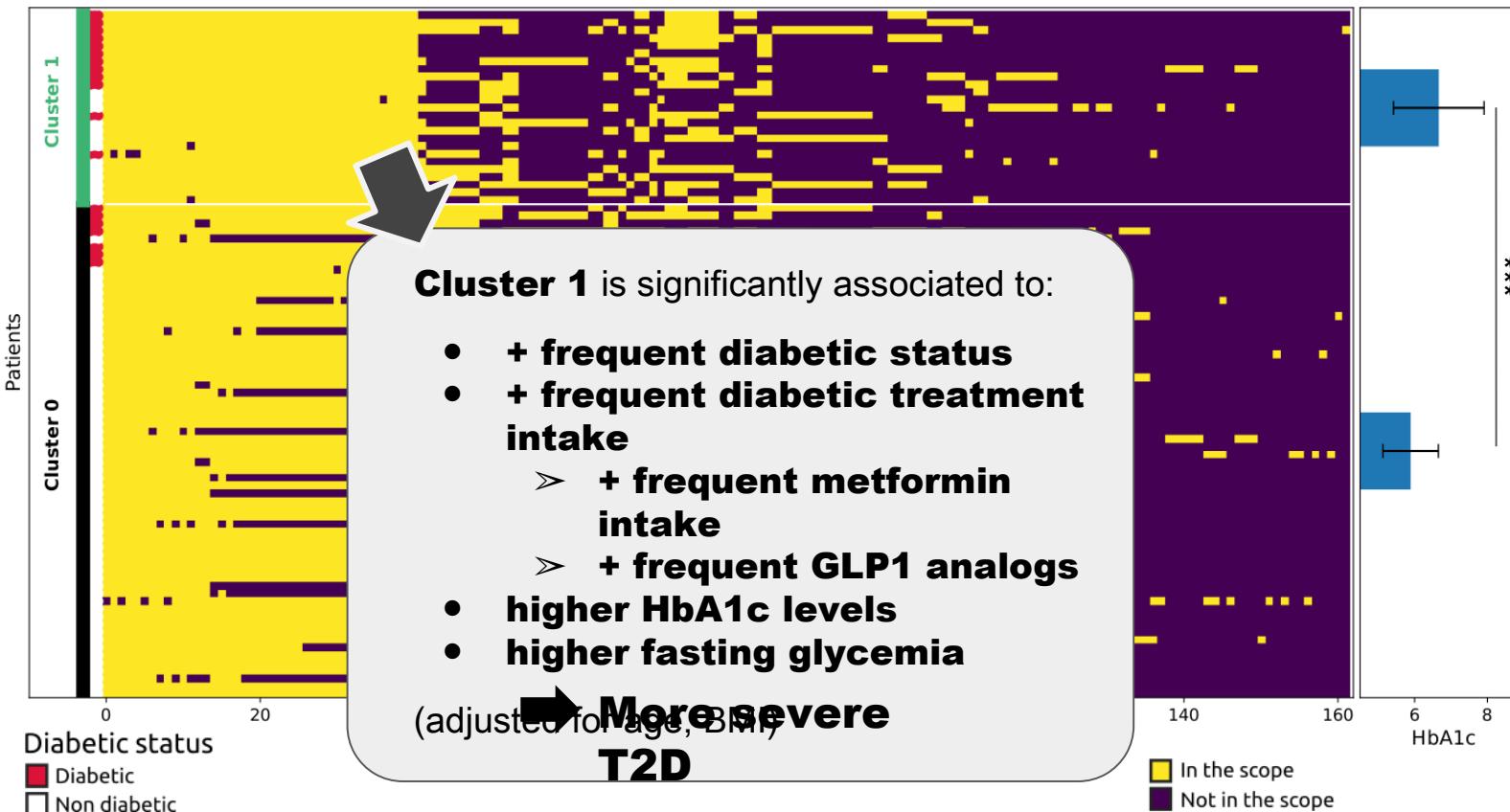
Spectral clustering of scope matrix (2 clusters)

Bariatric cohort



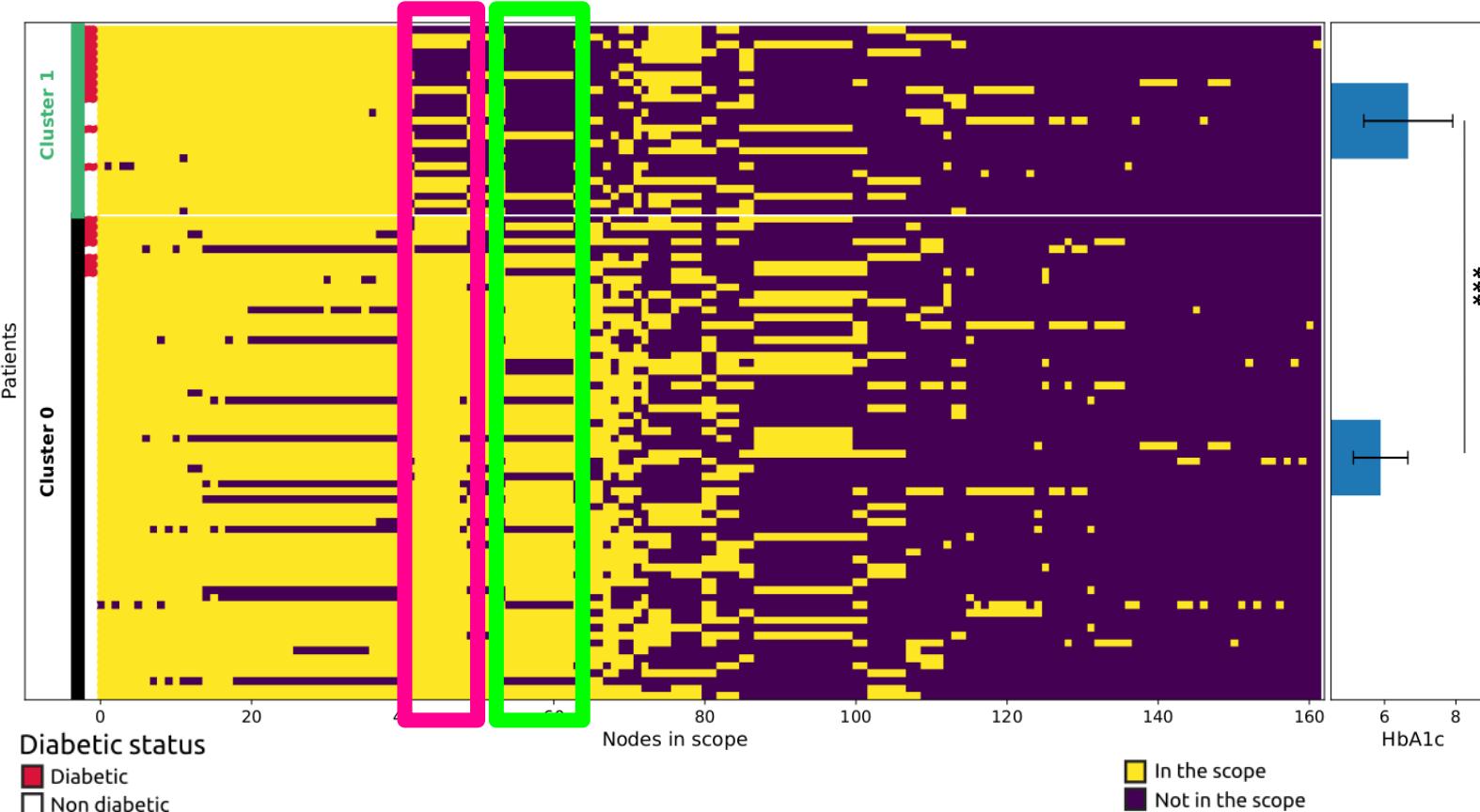
Spectral clustering of scope matrix (2 clusters)

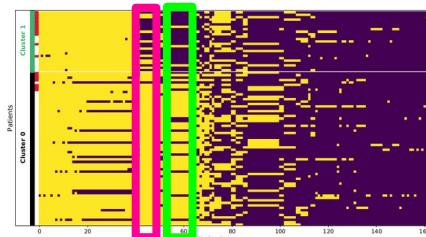
Bariatric cohort



Spectral clustering of scope matrix (2 clusters)

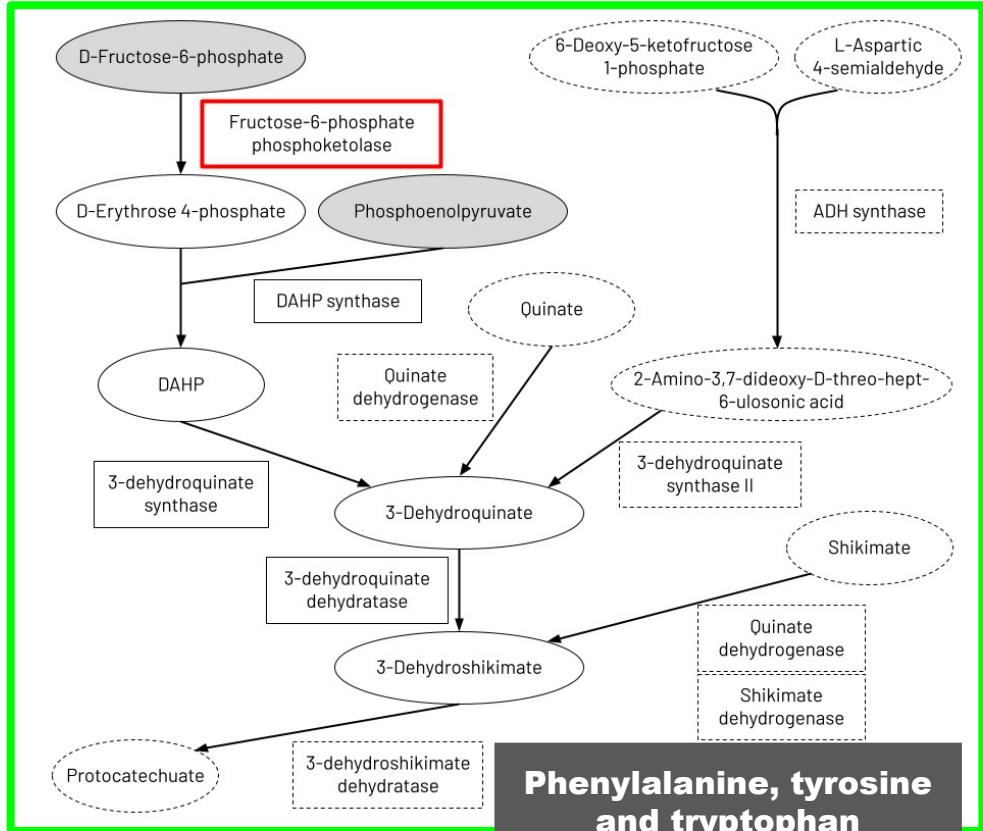
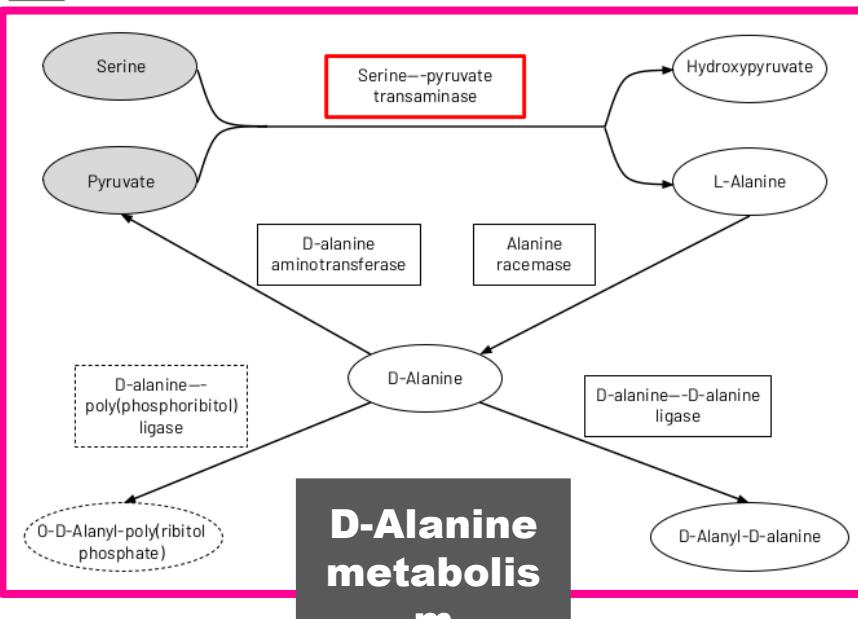
Bariatric cohort

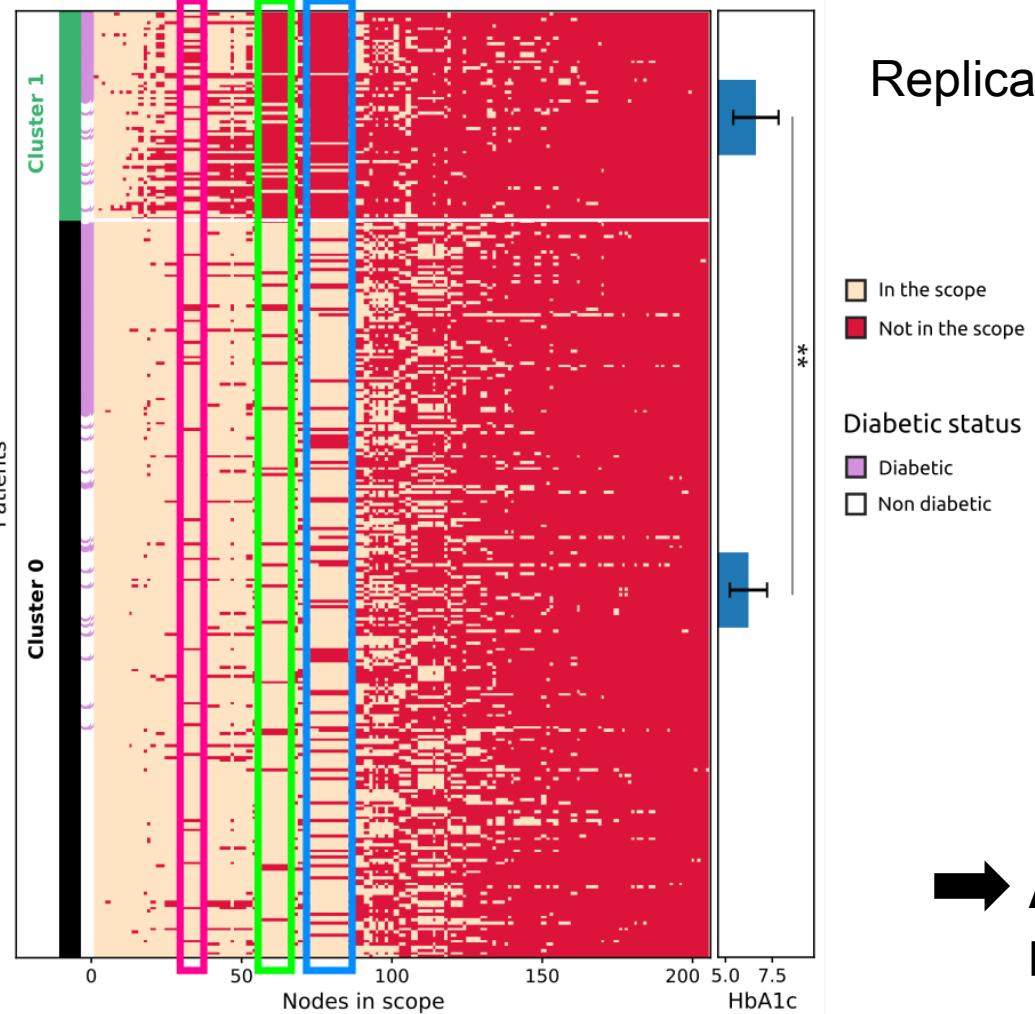




Impaired pathways in cluster 1

Bariatric cohort





Replication in an independent cohort (1)



MetaCardis cohort (314 patients)

Cluster 1 is significantly associated to:

- + freq **diabetic status**
- + freq **diabetic treatment intake**
 - + freq metformin intake
- **higher HbA1c levels**
- **higher fasting glycemia**
- **less prediabetic patients**

(adjusted for age, BMI, gender)

→ **Also a more severe T2D profile**

Literature on the enzymes and pathways of interest (1)

- **Phenylalanine, tyrosine and tryptophan biosynthesis** (Fructose-6-phosphate phosphoketolase)
 - **Enzyme specific to *Bifidobacterium*'s bifid shunt.**
 - Less *Bifidobacterium* in T2D and in obese gut microbiotas when compared to healthy individuals.

→ Points in the **same direction** as the identified **T2D severity of cluster 1**

Sedighi, Razavi et al, 2017
Wu, Ma et al, 2010
Gao, Zhu et al, 2018

Literature on the enzymes and pathways of interest (2)

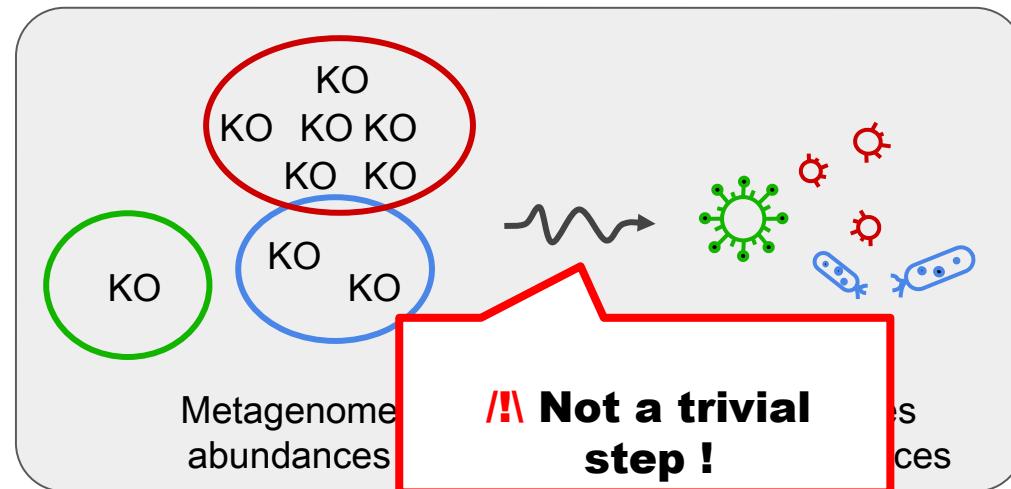
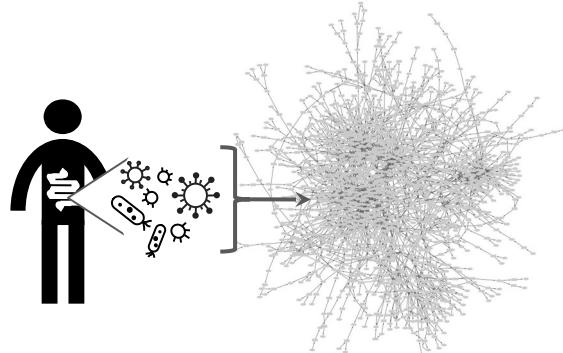
Less direct, would need experimental confirmations:

- **D-Alanine metabolism** (Serine-pyruvate transaminase)
 - No info on enzyme role in the gut microbiota.
 - In **rat livers**:
 - **Two pathways of gluconeogenesis from serine, with one using Ser-Pyr transaminase.**
 - **The other pathway** has been shown to be **favoured under starvation and diabetes.**
- **Thiamine metabolism (vitamin B1)** (Thiaminase II)
 - Bacterial enzyme involved in a **thiamine salvage pathway**.
 - Thiamine **cannot be synthesized by the host** (but gut microbiota can)
 - Thiamine is **decreased in T2D patients**, who also have a **higher thiamine demand**.
→ **New targets for experimental validation**

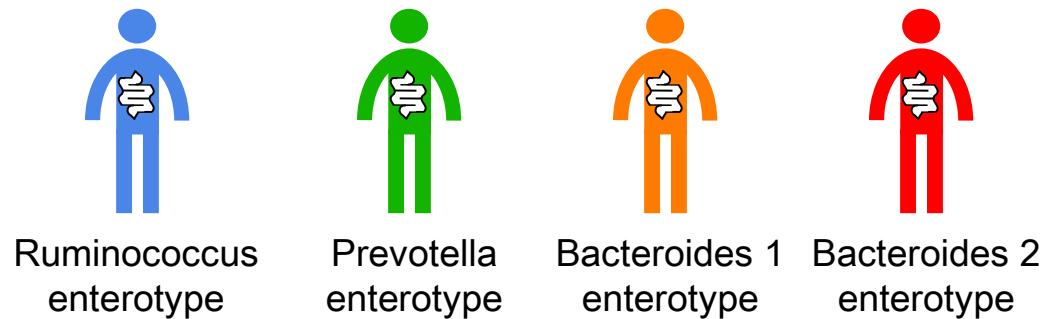
Hagopian, Ramsey et al, 2005
Snell, 1984

Thornalley et al, 2007
Pácal, Kuricová et al, 2014

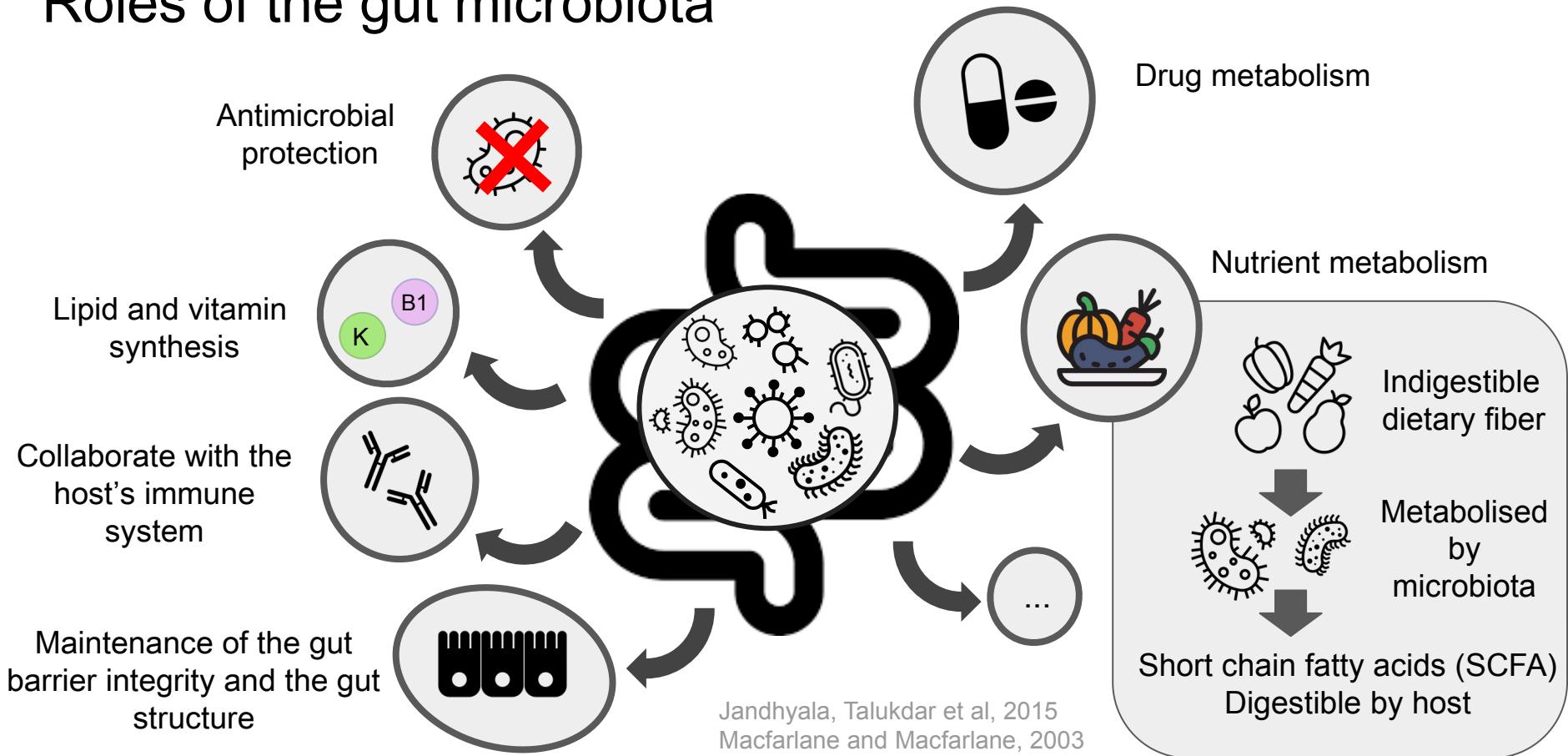
Species characterisation of the gut microbiota (1)



→ Microbiota profiling
(enterotype s)



Roles of the gut microbiota



Statistics

BARIATRIC Cluster 1

- + **T2D** (p-value (p-val) 1.4×10^{-4} , likelihood ratio Chi square value (Chisq) 14.5008, 1 DF)
- + **diabetic treatment intake** (p-val 1.5×10^{-4} , Chisq 14.3341, DF 1)
- + **metformin intake** (p-val 1.5×10^{-4} , Chisq 14.3341, DF 1)
- + **GLP-1 analog intake** (p-val 7.1×10^{-3} , Chisq 7.2457, DF 1).
- + **HbA1c** (p-val 3.6×10^{-4} , F value (Fval) 13.8376, DF 1, DF residuals 85)
- + **Fasting glycemia** (p-val 3.2×10^{-3} , Fval 9.2096, DF 1, DF residuals 85).

METACARDIS Cluster 1

- + **T2D** (p-val 2.6×10^{-3} , Chisq 9.0695, DF 1)
- + **diabetic treatment intake** (p-val 4.9×10^{-4} , Chisq 12.1417, DF 1)
- + **metformin intake** (p-val 8.3×10^{-6} , Chisq 19.8785, DF 1)
- + **HbA1c** (p-val 7.9×10^{-3} , Fval 7.189, DF 1, DF residuals 309)
- + **fasting glycemia** (p-val 1.3×10^{-2} , Fval 6.2456, DF 1, DF residuals 306)
- **prediabetic patients** (p-val 4.2×10^{-3} , Chisq 8.2155, DF 1)