

# Nutriomics challenge

## Introduction to the project

Optimization of dietary nutrient  
supplementation for rational optimization of  
human gut microbiome

**Andrei Osterman**

*Laboratory of Metabolic Genomics*

**Sanford-Burnham-Prebys Medical Discovery Institute, La Jolla, CA, USA**

# MICROBIAL RESILIENCE AND VULNERABILITY

*Metabolic interactions and adaptive evolution of microbes and microbial communities*



- “**Phenobiome**”: Predictive phenotype profiling of human gut microbiome from –omics data

- Nutrient requirements, metabolic production potential
- Fundamental mechanisms, diagnostics, therapy

- “**Resistome**”: Experimental evolution, genomics and mechanisms of antibiotic resistance

- Clinical antibiotics, experimental drugs in major bacterial pathogens
- Resistance profiles, treatment regimens, lead optimization

# Phenobiome Project

## Objective

Predict metabolic capabilities and interactions in microbial communities from genomic data

## Approach

[100+ metabolic phenotypes] x [1,000s gut microbes]

- Utilization of major nutrients (carbohydrates)
- Salvage/Synthesis/Exchange of intermediary metabolites (amino acids, B vitamins)
- Catabolic products (SCFA, tryptamine, etc)

## Methodology

*In silico* metabolic reconstruction

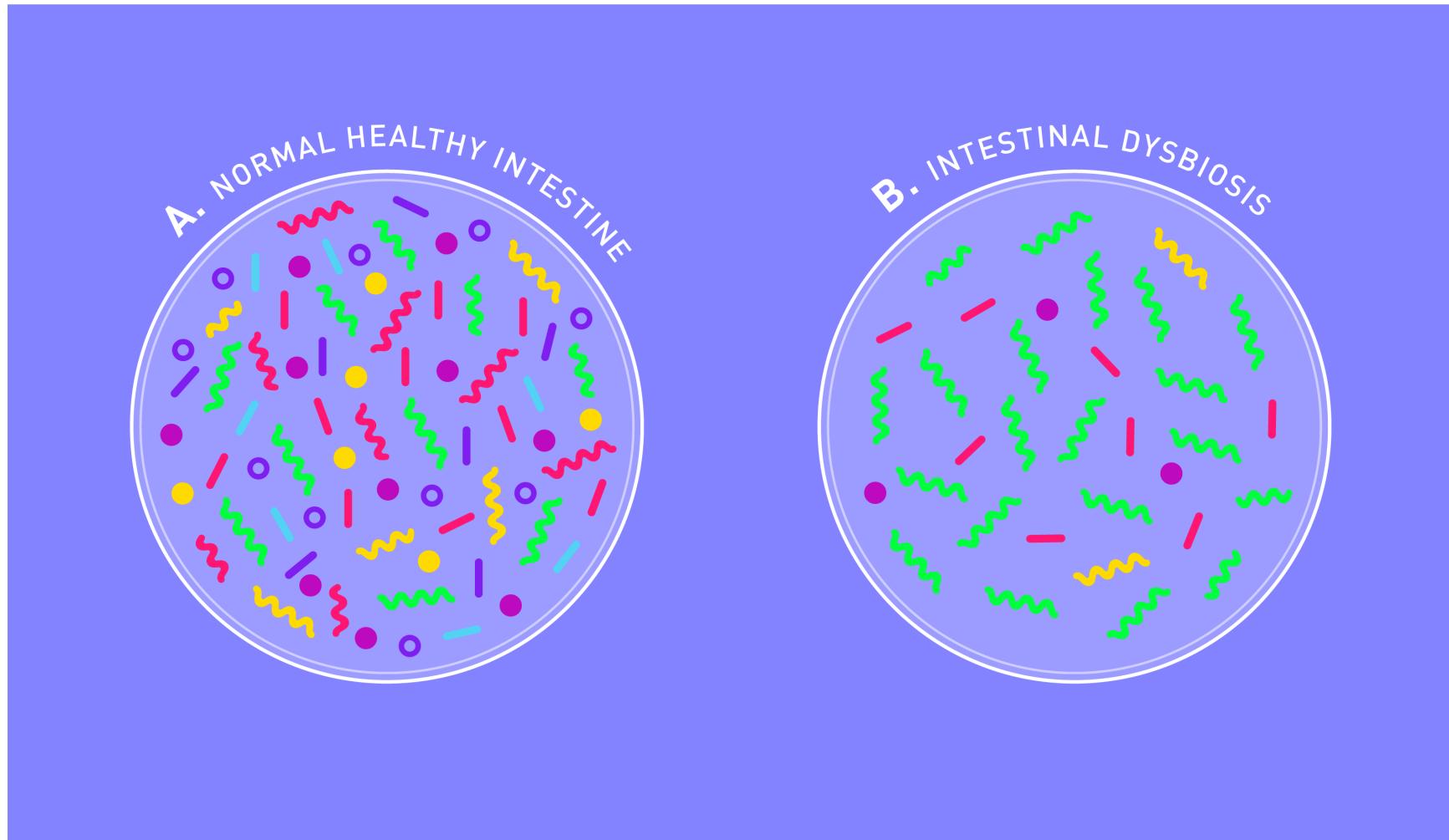
## Applications

Diagnostics, prevention and correction of dysbiosis via rational nutrient supplementation

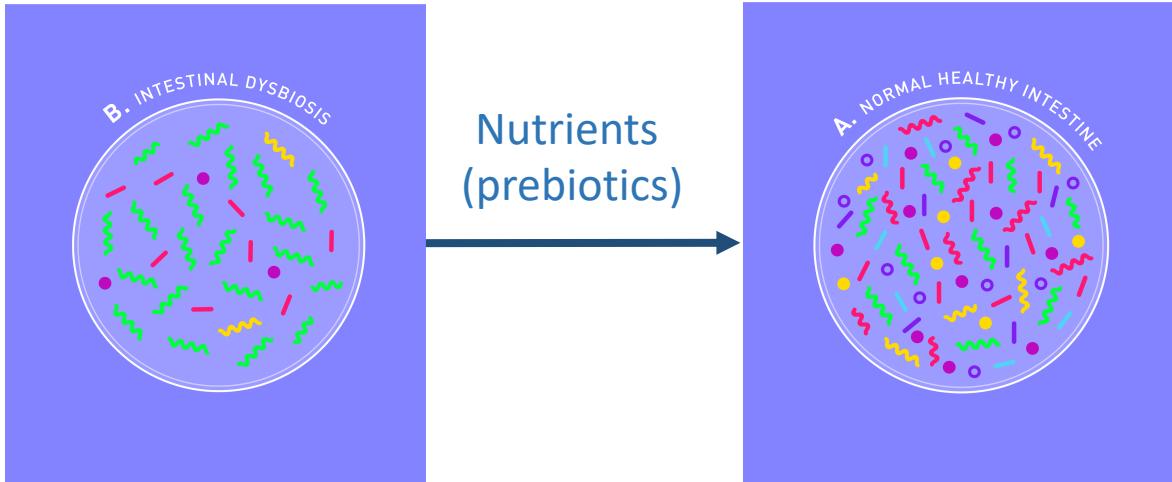
Species ID	BINARY PHENOTYPES MATRIX (PV)							
	Nutrient 1	Nutrient 1	Nutrient 2	Nutrient 3	Nutrient 4	Nutrient 5	Nutrient 6	Nutrient 7
Microbe 1	1	0	0	0	0	1	0	0
Microbe 2	0	1	1	1	0	0	1	0
Microbe 3	1	1	0	0	0	1	0	0
Microbe 4	0	0	1	0	1	0	0	0
Microbe 5	1	1	0	0	0	0	0	1



# Human Gut Microbiome (HGM): norm vs dysbiosis



# Nutriomics challenge

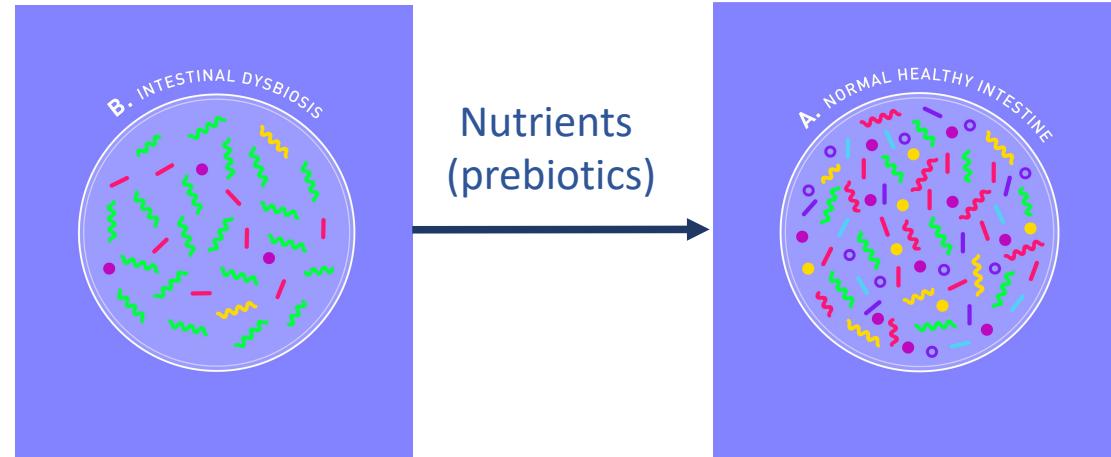


## Biomedical objective:

- Rebalance human gut microbiome (HGM) to enhance the diversity of its bacterial composition (referred to as taxonomic diversity) via supplementary nutrients rationally selected to:
  - promote the preferential growth of “underrepresented” species
  - while providing minimal additional support to “overrepresented” species in the patient’s sample.

# Why do we care?

- Dysbiosis-related syndromes
  - Inflammatory Bowel Disease (IBD)
    - Crohn disease, ulcerative colitis, etc
  - Obesity, diabetes, colon cancer
  - Stunned growth and impaired immunity in infants
  - Metabolism-Associated Liver Disease (MAFLD)
  - Gut-brain axis (CNS syndromes)
- Smart **personalized** prebiotics (food supplements)
  - Microbiota-targeted nutritional supplement
  - Direct treatment of dysbiosys (prebiotics)
  - Supportive treatment (synbiotics, drug response)
  - Prevention, wellness



**BIG CHALLENGE:** How do we progress from knowledge of microbiome taxonomic and functional profiles to rational dietary intervention?

# Nutriomics Strategy

We'll get back to it later, after a brief detour

## Input data for a microbiome sample

Taxonomic profile

TMS1	RA(%)
Species ID	Dysbiotic Microbiome
<i>Microbe 1</i>	Low
<i>Microbe 2</i>	Low
<i>Microbe 3</i>	High
<i>Microbe 4</i>	Norm
<i>Microbe 5</i>	High



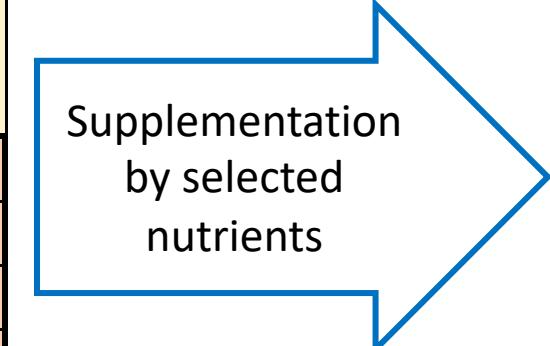
Nutrient Impact Matrix for individual species

Species ID	BINARY PHENOTYPES							
	Nutrient 1	Nutrient 2	Nutrient 3	Nutrient 4	Nutrient 5	Nutrient 6	Nutrient 7	Nutrient 8
<i>Microbe 1</i>	1	0	0	0	0	1	0	0
<i>Microbe 2</i>	0	1	1	1	0	0	1	0
<i>Microbe 3</i>	1	1	0	0	0	1	0	0
<i>Microbe 4</i>	0	0	1	0	1	0	0	0
<i>Microbe 5</i>	1	1	0	0	0	0	0	1

Reference Sample Collection  
(taxonomic profiles for 1,000s  
“normal” microbiome samples)

## Dietary intervention

Optimized Taxonomic profile



TMS1	RA(%)	Change
Species ID	Dysbiotic Microbiome	
<i>Microbe 1</i>	Low	↑
<i>Microbe 2</i>	Low	↑
<i>Microbe 3</i>	High	↓
<i>Microbe 4</i>	Norm	
<i>Microbe 5</i>	High	↓

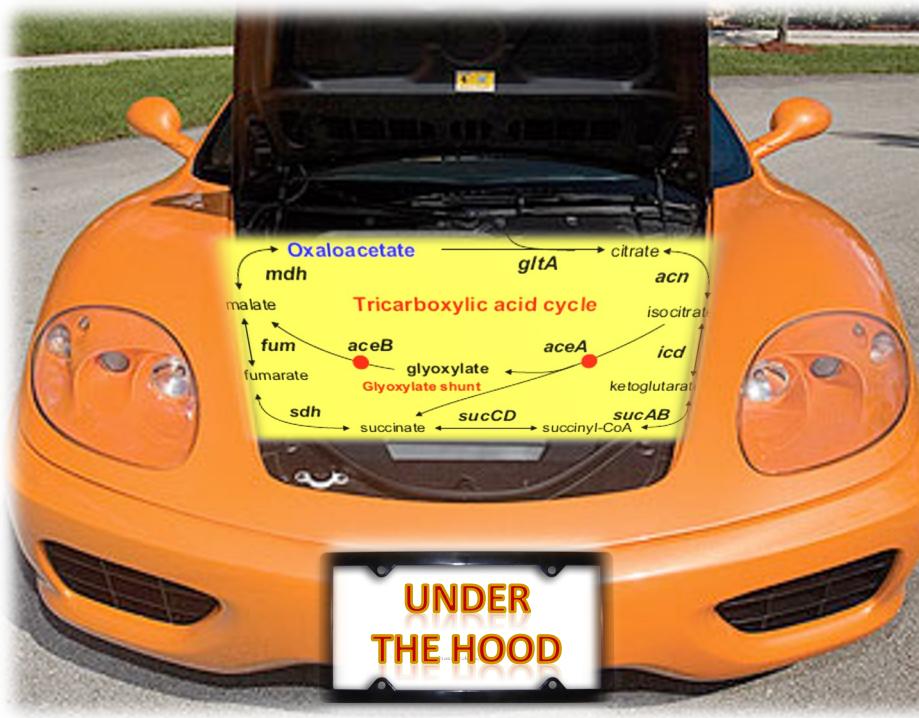
## Bioinformatics objective

Given: (i) Test Microbiome Sample (TMS) with a taxonomic profile;  
 (ii) Reference Sample Collection (RSC) for 1,000s HGM samples;  
 (iii) Nutrient Impact Matrix (NIM) derived from predicted binary phenotypes

Select an optimal sub-set of  $n$  nutrients that collectively would:

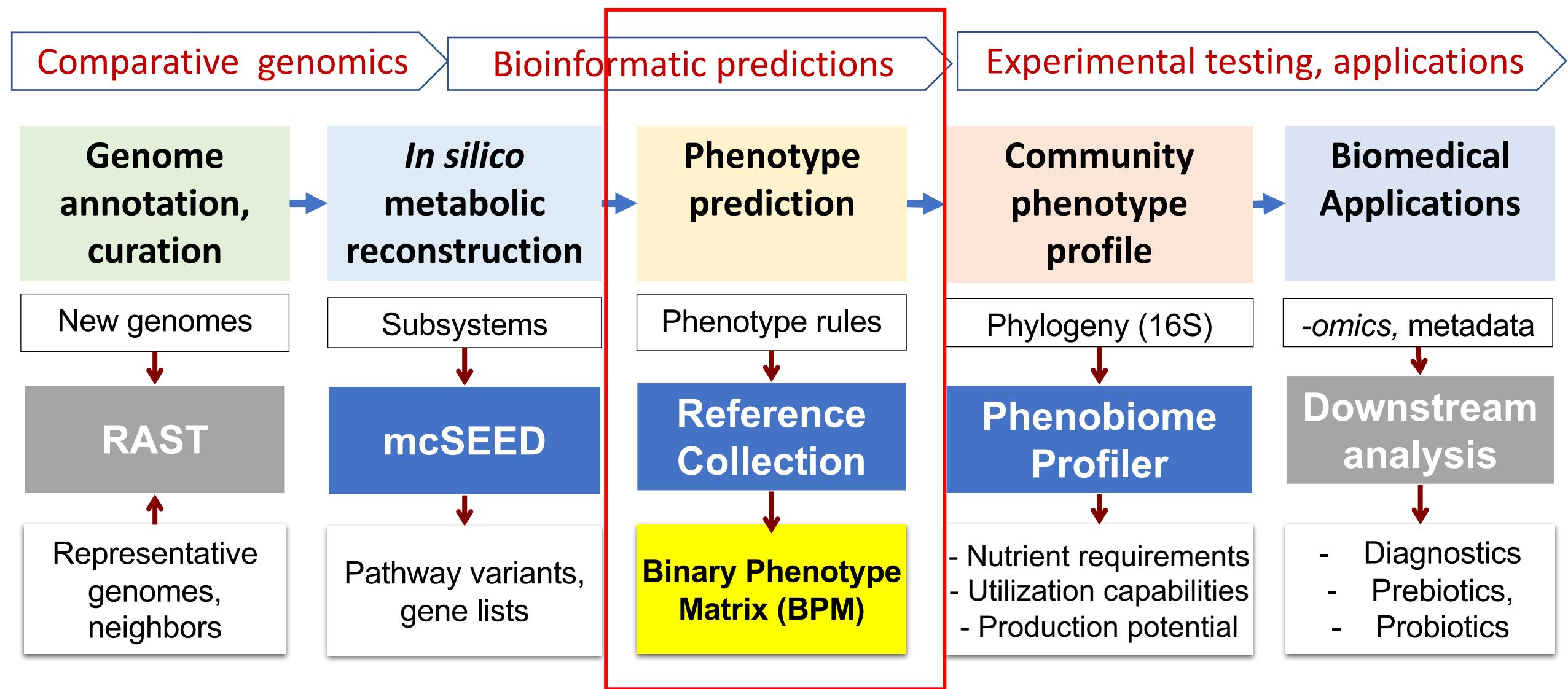
- promote the growth of the maximal number of species deemed underrepresented (low) across multiple samples in RSC,
- while promoting the minimal number of in-range (norm) and overrepresented (highs) species as compared to RSC

# How do we know?



In silico metabolic reconstruction → phenotype prediction

# Phenobiome Approach: Workflow



# Metabolic reconstruction in ~3,000s HGM reference genomes

Comparative genomics: genes → pathways → phenotypes

Amino acid prototrophy		
His	Cys	Trp
1	1	0

B Vitamin prototrophy		
B1	B2	B3
0	0	1

Sugar utilization		
Fuc	Rha	Xyl
1	0	1

SCFA production	
C3	C4
0	1

Binary phenotypes

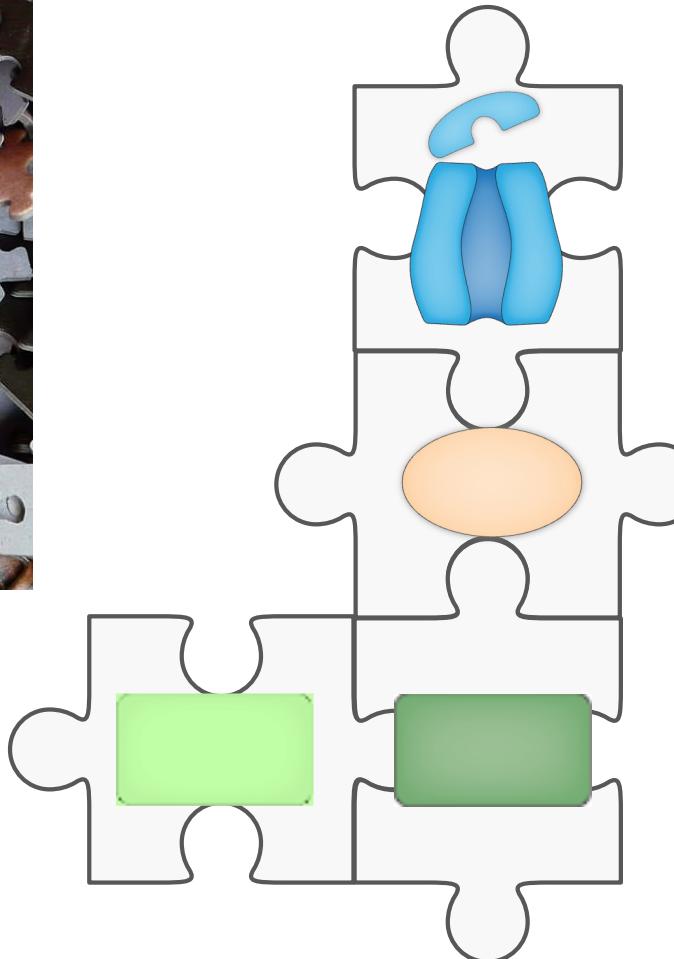
# Reconstruction of carbohydrate utilization pathways, a jigsaw puzzle



Image source: pixcove.com

Transcriptional regulators

Puzzle pieces: pathway elements



Glycan transporters

Glycoside hydrolases  
(GH)

Catabolic enzymes

Searching for genes encoding pathway elements in every genome of the dataset

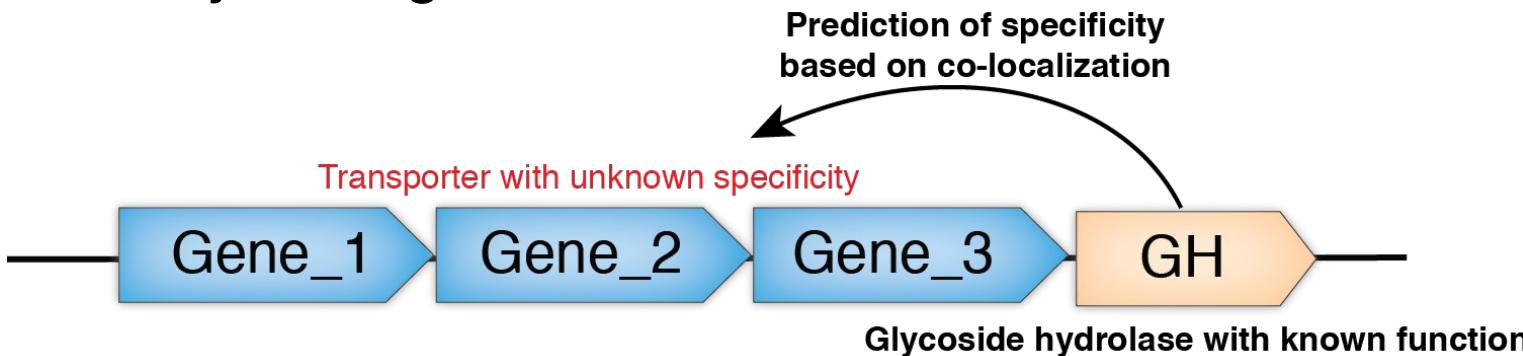
# Comparative-genomic techniques for metabolic reconstruction

## 1. Analysis of previously published data/databases

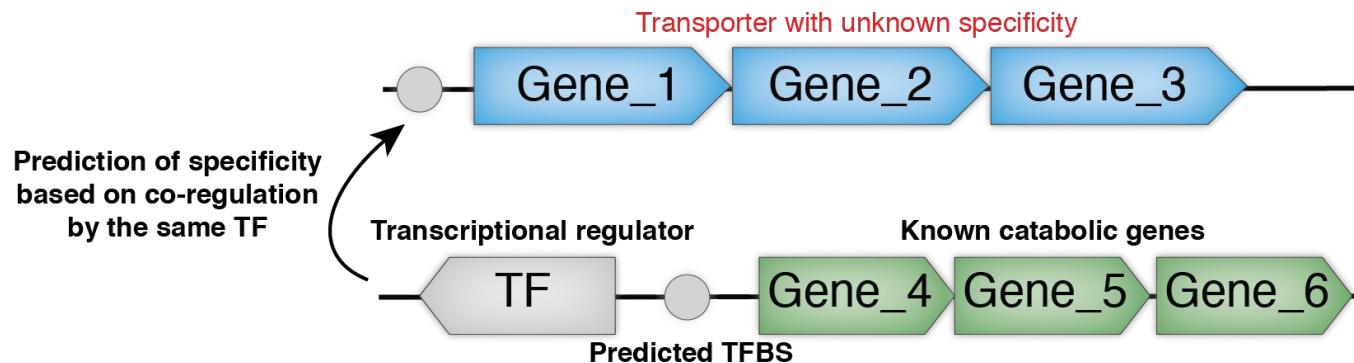
PaperBLAST – Find papers about a protein or its homologs



## 2. Analysis of genomic context



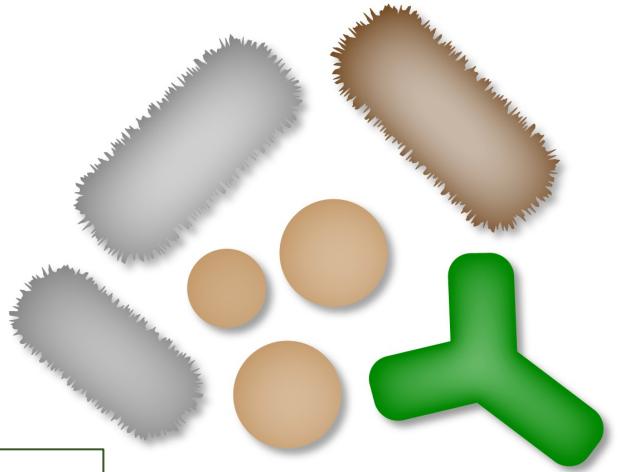
## 3. Reconstruction of transcriptional networks (regulons)



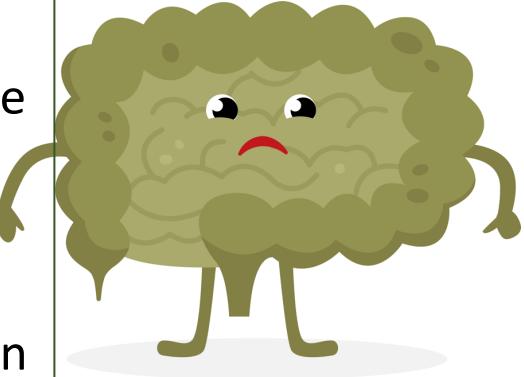
RegPrecise

# Example 1: Bifidobacteria in infants' gut microbiome. Genomics-driven development of probiotics and prebiotics

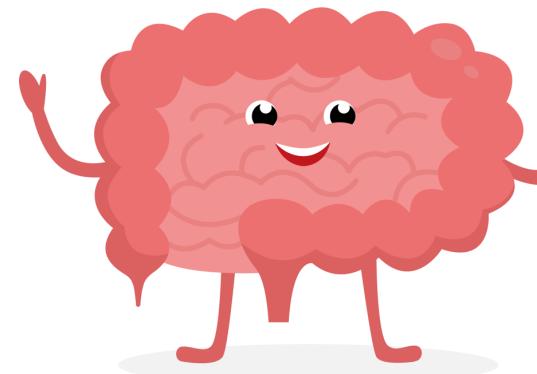
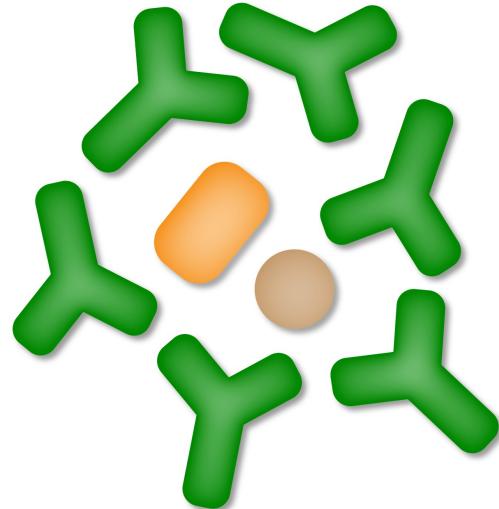
Malnourished children:  
**low *Bifidobacterium* abundance**



- Dysbiosis
- Immature microbiome
- Impaired growth
- Pathogen colonization



Healthy children:  
**high *Bifidobacterium* abundance**



Probiotics  
Prebiotics



Alex  
Arzamasov

- Immunity development
- Pathogens displacement;
- Cross-feeding
- Community building

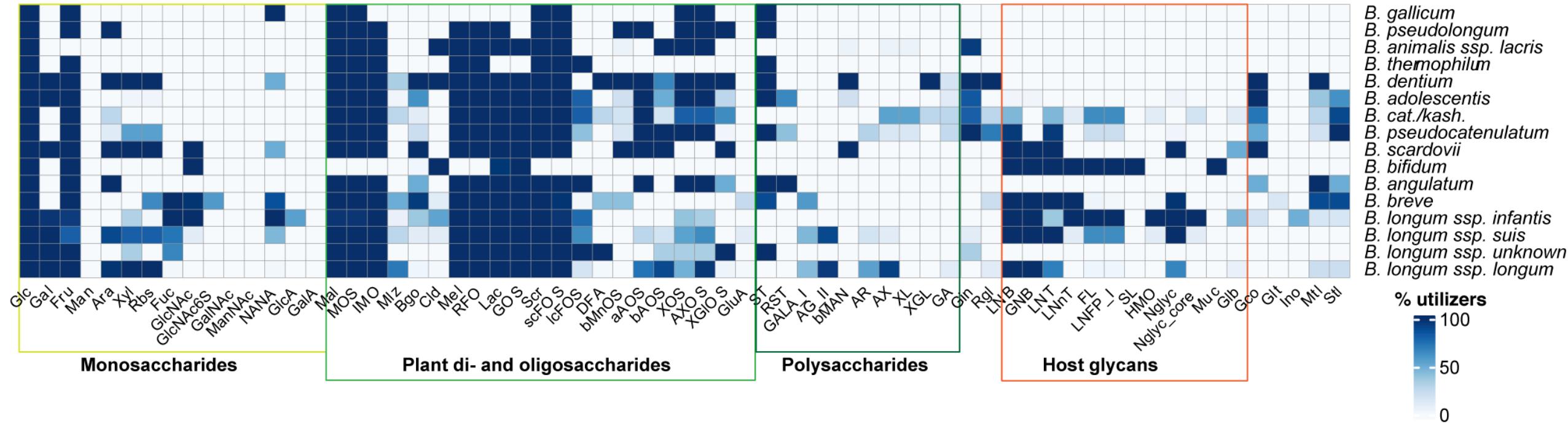
**Probiotics** – Bifidobacteria; **Prebiotics** are **glycans** that are not metabolized by the human body and selectively stimulate growth of bifidobacteria

## **Summary and primary output: Binary Phenotype Matrix (BPM)**

1. 387 genomes, 65 predicted carbohydrate utilization phenotypes:
    - Monosaccharides
    - Oligosaccharides
    - Polysaccharides
    - Other (sugar alcohols)
  2. >100 predicted novel pathway elements

# Variability of predicted carbohydrate utilization phenotypes at species and strain levels

- Dark blue: all strains are predicted **utilizers**
- White: all strains are predicted **non-utilizers**



- **Species-level:** some species are more equipped for plant glycan utilization, others prefer host glycans
- **Strain-level:** substantial variability

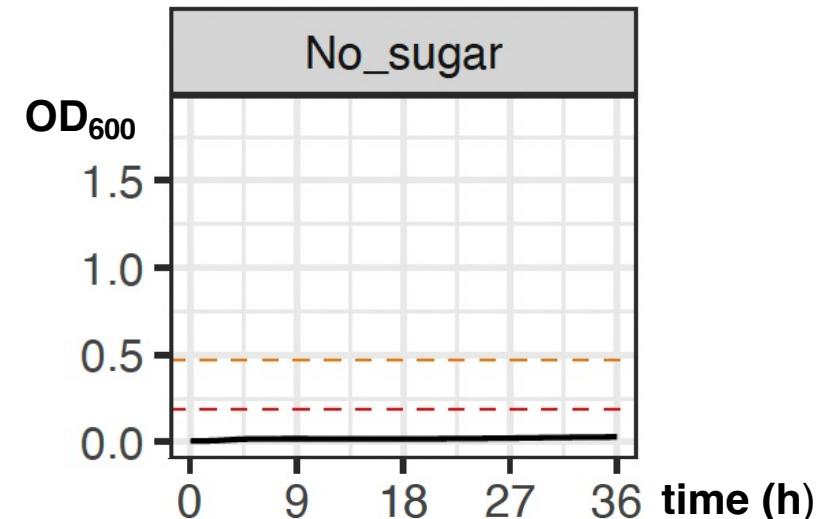
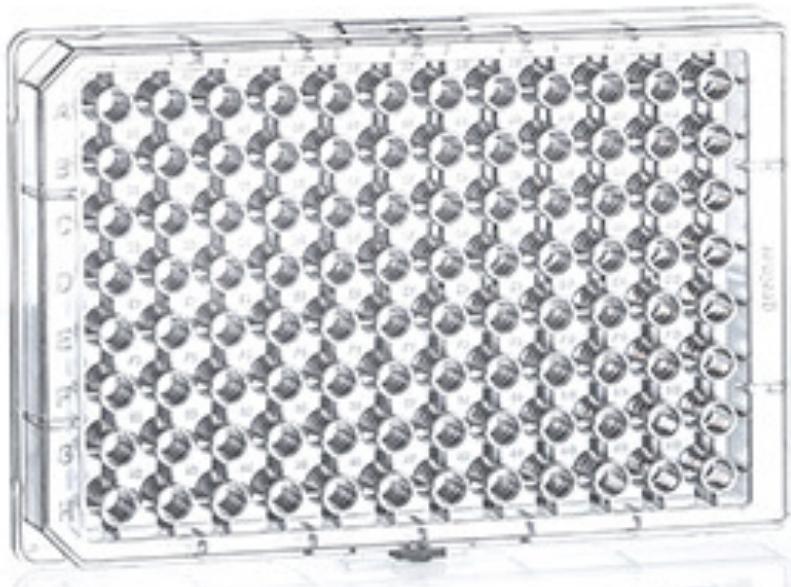
# Experimental testing of predicted phenotypes

- 16 *Bifidobacterium* strains
- 35 carbohydrates

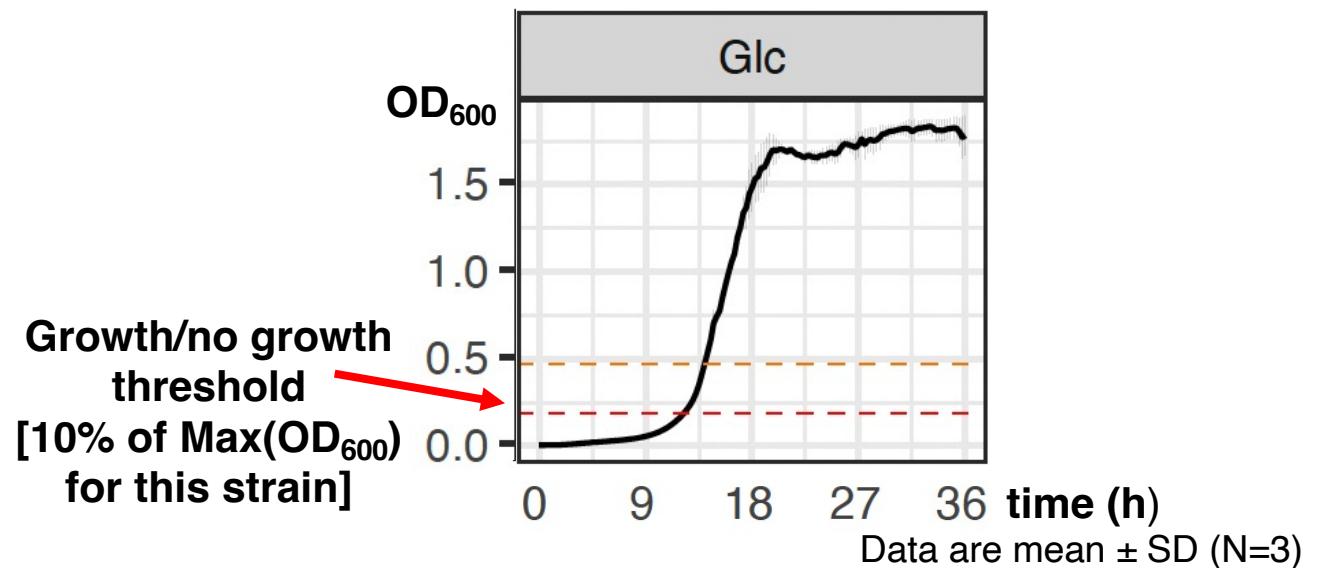
The basal medium does not support the growth unless a carbon source is added



*In vitro* growth experiments in anoxic conditions using a 96-well plate format



Growth/no growth threshold  
[10% of Max(OD<sub>600</sub>) for this strain]



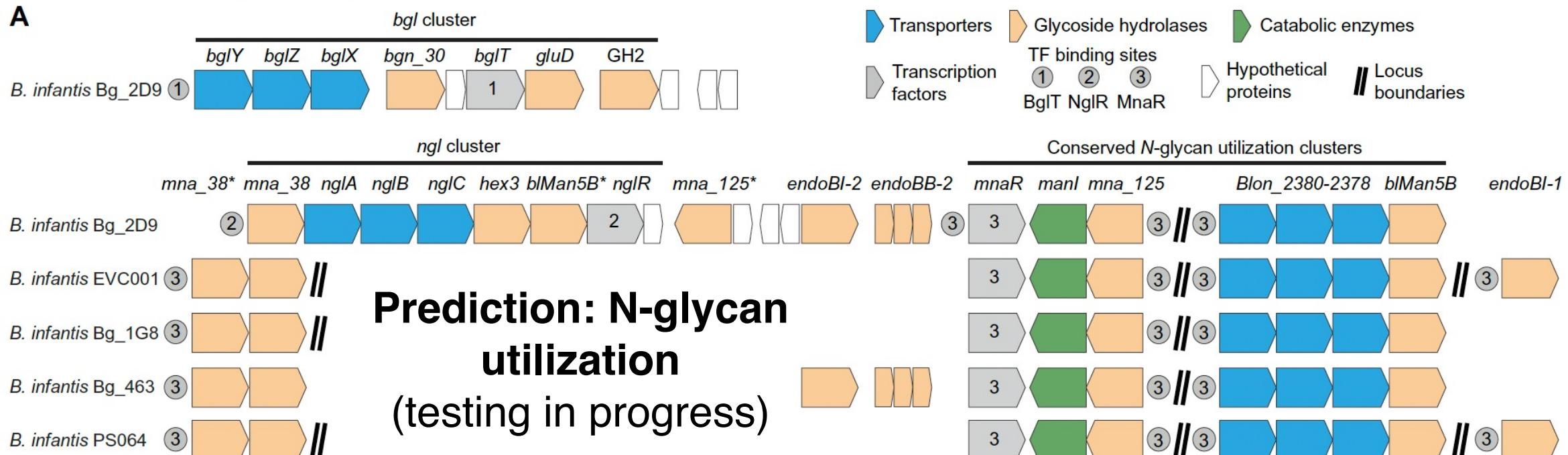
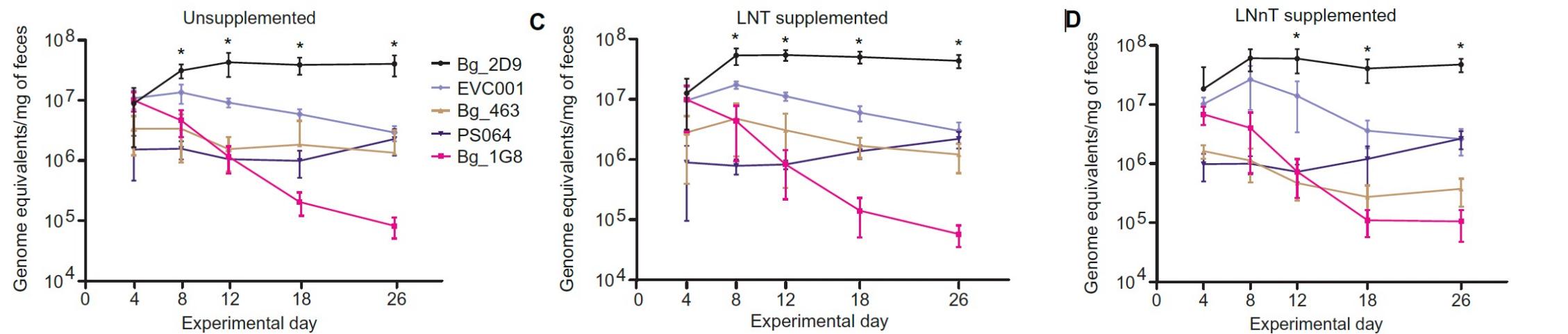
# Experimental testing: summary of results

560 phenotypes in total: 16 strains x 35 carbohydrates

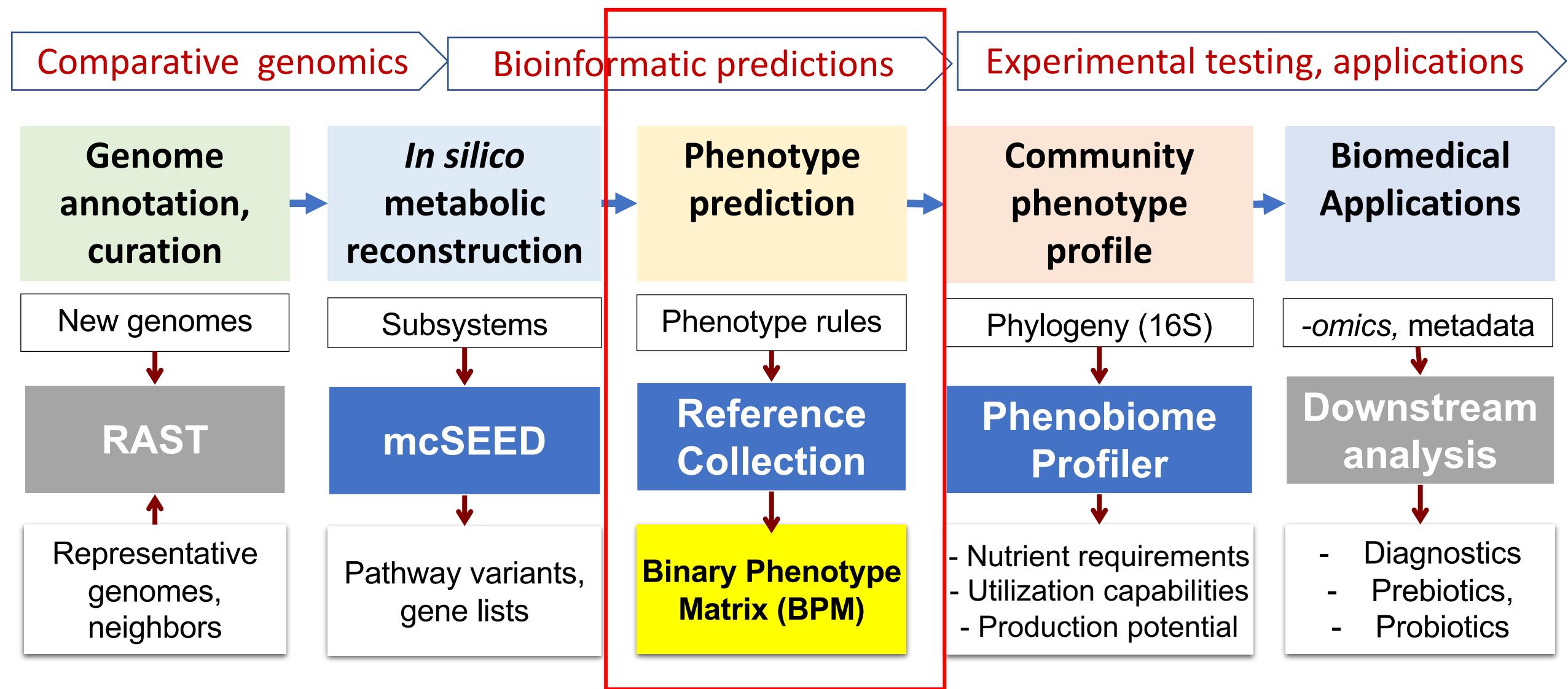
	Mono-saccharides	Di/oligo-saccharides	Poly-saccharides	Other glycans	Total
Tested phenotypes	208	240	48	64	560
Matched predictions	175	233	46	63	517
False positives	8	3	2	1	14
False negatives	25	4	0	0	29
Percent of matched predictions	84%	97%	96%	98%	92%

Overall, predicted phenotypes show excellent agreement with growth data

# *B. infantis* Bg\_2D9 has two unique uncharacterized gene clusters likely driving its elevated fitness in the competition experiment



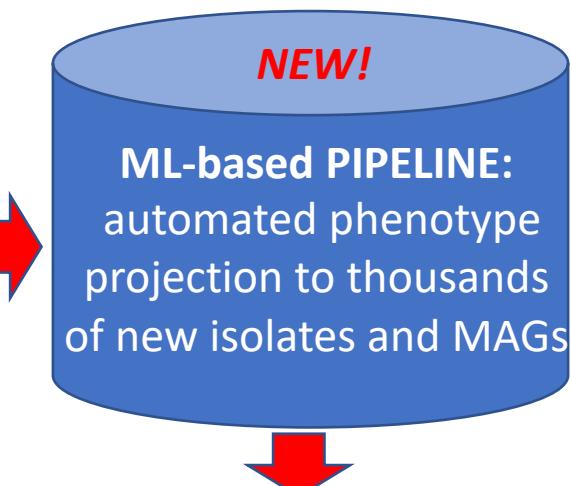
# Phenobiome Approach: Workflow



# HGM Binary Phenotype Matrix (BPM-3,000)

Species/Strains	B-vitamins synthesis			Amino acids synthesis			Carbohydrate utilization										Amino acids degrad			Fermentation									
							Hexoses		Pentoses		Disaccharides		Polyols		Oligosacch.														
	B2	B3	B5	His	Trp	Cys	Tag	GlcA	GlcNAC	NANA	Xyl	Ara	Fuc	Rha	Tre	Chb	Lnb	Lac	Xtl	Srl	FOS	Raf	a(Ara)n	(Fuc)n	His	Met	Trp	Butyr	Propio
<i>Akkermansia muciniphila</i> ATCC	1	1	1	1	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1
<i>Akkermansia muciniphila</i> LFYP55	1	1	1	1	1	1	0	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1
<i>Bacteroides caccae</i> LFYP20	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	0	1
<i>Bacteroides dorei</i> DSM	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	0	0	0	1	0	1	0	1	0	0	0	1	1
<i>Bacteroides uniformis</i> LFYP32	1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	0	0	0	0	1	0	1	1	1	1	0	1	0	1
<i>Bacteroides vulgatus</i> LFYP11	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	0	1	0	1	0	1	0	1	0	1	1
<i>Parabacteroides distasonis</i> LFYP31	1	1	1	1	1	1	0	1	1	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1
<i>Parabacteroides johnsonii</i> DSM	1	1	1	1	1	1	0	1	1	1	1	0	0	1	0	1	0	0	0	1	0	1	0	1	0	0	0	1	0
<i>Citrobacter amalonaticus</i> LFYP1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	1	0	0	1	0	1	0	1	0	1	1
<i>Escherichia fergusonii</i> SS_Bg39	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	1	0	1	0	1
<i>Klebsiella oxytoca</i> LFYP65	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	0	1	0	1	0	1	0
<i>Bifidobacterium adolescentis</i> LFYP80	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0	0
<i>Bifidobacterium bifidum</i> ATCC	0	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0
<i>Bifidobacterium breve</i> LFYP81	0	1	0	1	1	0	0	0	1	1	0	0	1	0	0	1	1	1	1	1	0	0	0	0	0	1	0	0	0
<i>Bifidobacterium catenulatum</i> JG_Bg468	0	1	0	1	1	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	0	1	1	1	0	1	0	1	0
<i>Bifidobacterium longum</i> subsp. <i>infantis</i>	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0
<i>Bifidobacterium longum</i> subsp. <i>longum</i>	0	1	0	1	1	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0
<i>Bifidobacterium pseudocatenulatum</i> LFYP29	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0	1	0
<i>Clostridium bolteae</i> LFYP116	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Clostridium hathewayi</i> DSM	0	0	0	1	1	1	0	0	1	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
<i>Clostridium symbiosum</i> LFYP84	1	1	0	0	1	1	0	0	1	1	1	1	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0
<i>Collinsella aerofaciens</i> LFYP39	0	1	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
<i>Collinsella intestinalis</i> LFYP54	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Collinsella stercoris</i> DSM	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Faecalibacterium prausnitzii</i> A2-165	0	0	0	1	0	1	0	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	1
<i>Faecalibacterium prausnitzii</i> L2-6	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	1	1

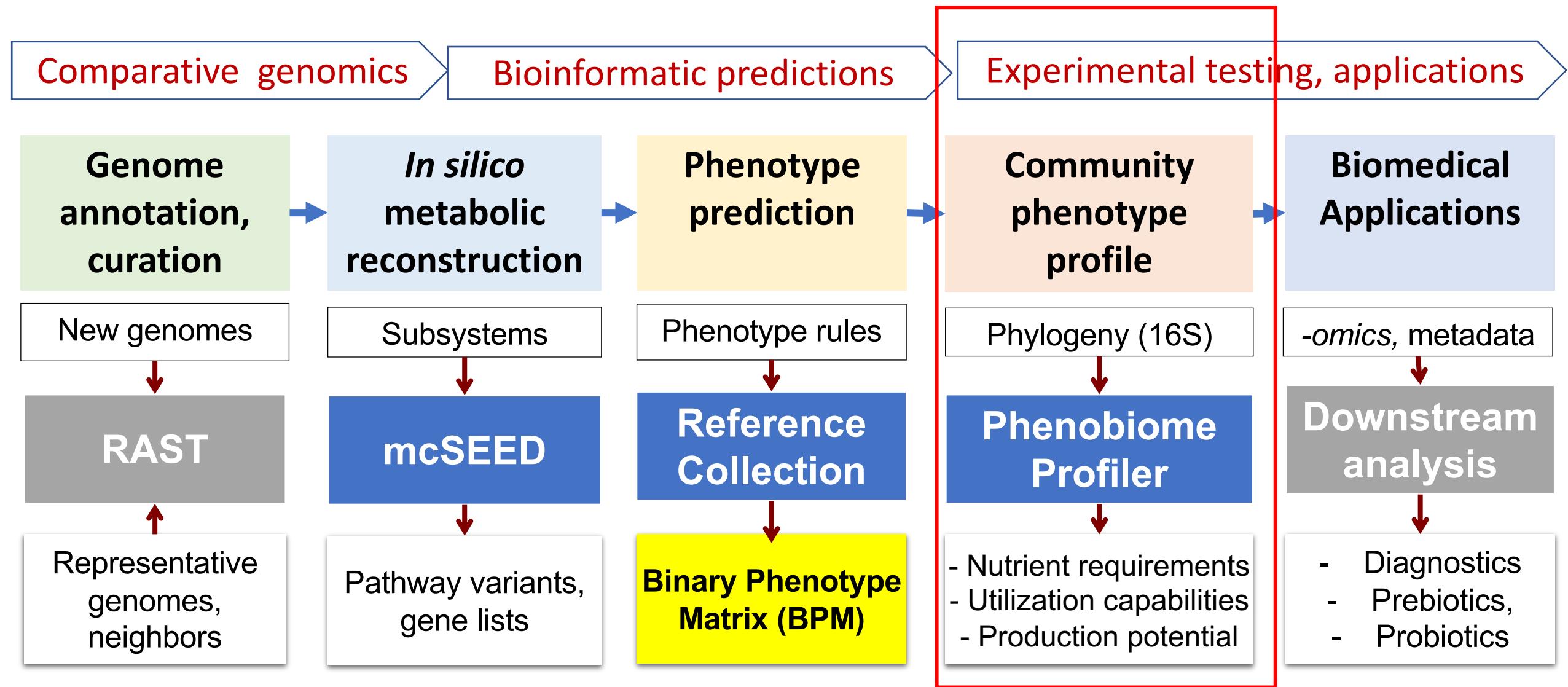
(100+ phenotypes)



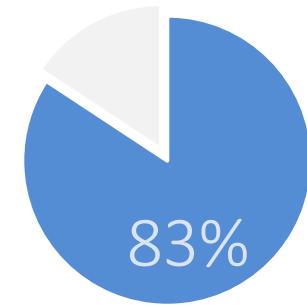
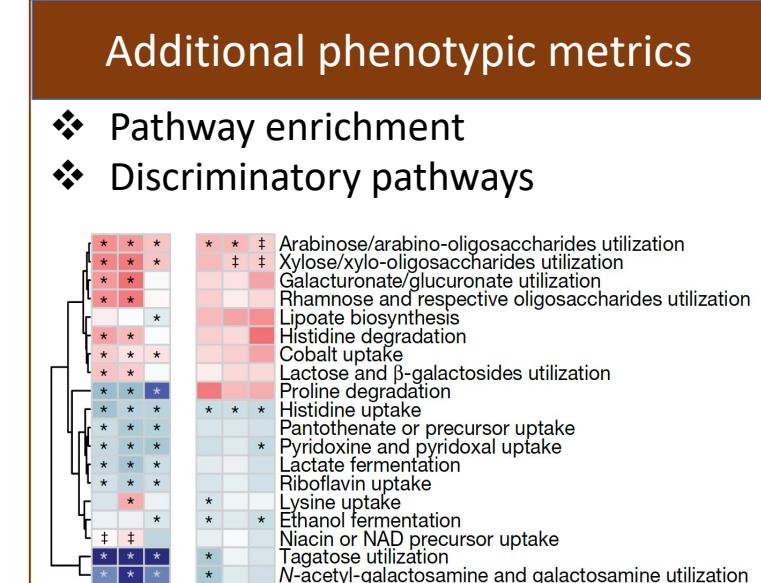
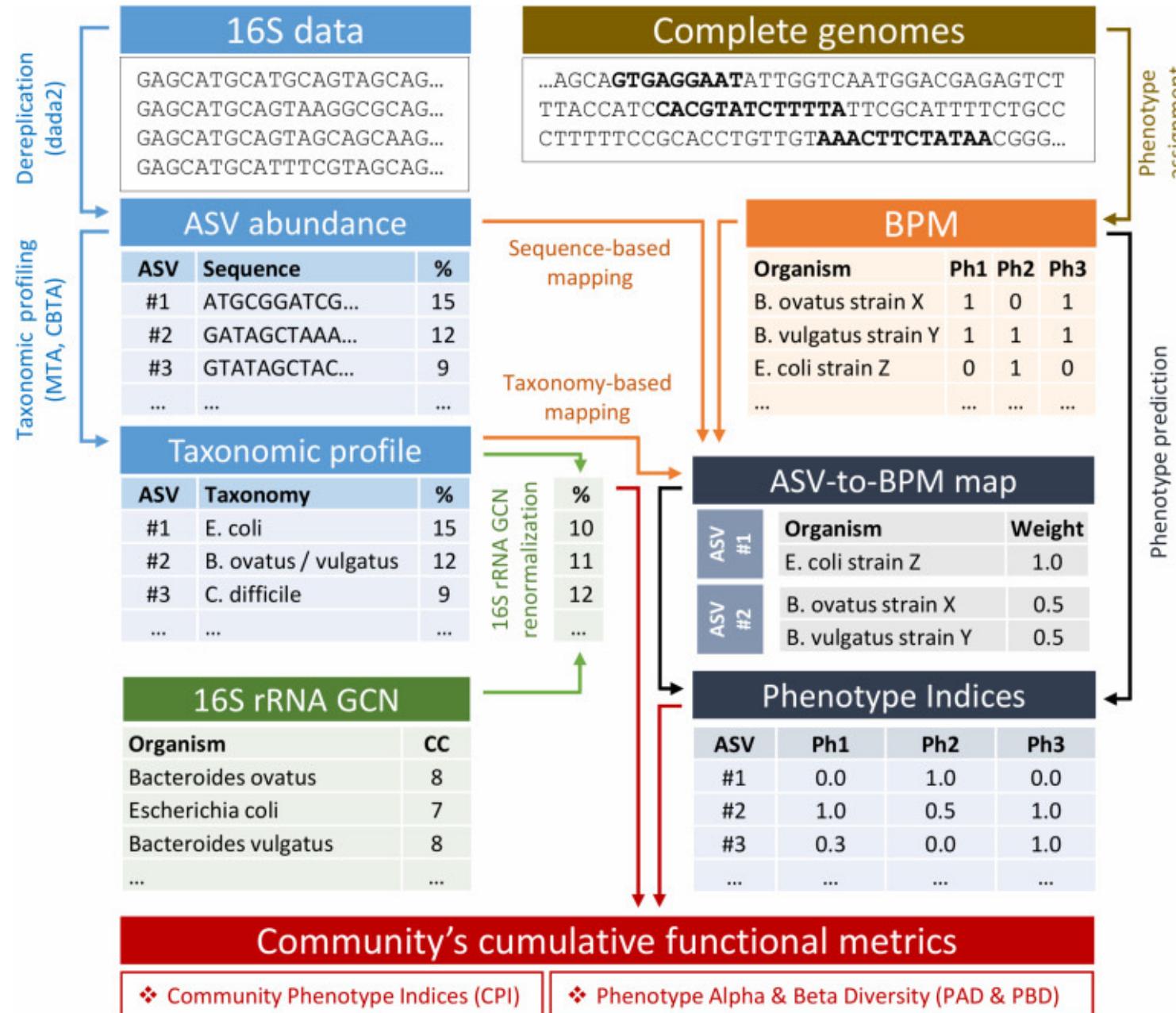
(3,000 HGM genomes)

MAG_species	Coun	NVP	OPVS	B1	B2	B3	B5	B6	B7	B9	B12
<i>Bifidobacterium longum</i>	463	16	2.23	1	0.002	1	0.002	1	0	1	0
<i>Escherichia coli</i>	455	29	2.96	1	1	1	1	1	1	1	1
<i>Bifidobacterium adolescentis</i>	267	15	2.75	0.03	0	0	0	1	0	1	0
<i>Bifidobacterium pseudocatenula</i>	238	16	2.54	1	0	0.01	0	1	0	1	0
<i>Bacteroides vulgatus</i>	223	10	0.26	1	1	1	1	1	1	1	1
<i>Turicibacter sanguinis</i>	223	14	0.21	0	0	0	0.004	1	0	1	0
<i>Bacteroides ovatus</i>	208	13	0.65	1	0.995	1	1	1	1	1	0.02
<i>Bacteroides uniformis</i>	175	10	1.78	1	1	1	1	1	1	1	1
<i>Collinsella aerofaciens</i>	105	16	1.60	0	0	1	0	1	0	1	0.83
<i>Parabacteroides distasonis</i>	71	10	0.23	1	1	1	1	1	1	1	1
[ <i>Eubacterium</i> ] rectale	53	18	3.77	1	1	1	1	1	1	0	0
[ <i>Ruminococcus</i> ] torques	53	42	10.30	0.72	0.283	1	0.283	0.72	0.62	1	1
<i>Blautia wexlerae</i>	53	18	3.77	1	1	1	0	0	0	1	1
<i>Blautia obeum</i>	41	15	1.88	1	0.78	1	0	0	0	1	1
<i>Faecalibacterium prausnitzii</i>	38	29	4.47	0	0.895	0	0	0	0	0	1

# Phenobiome Approach: Workflow



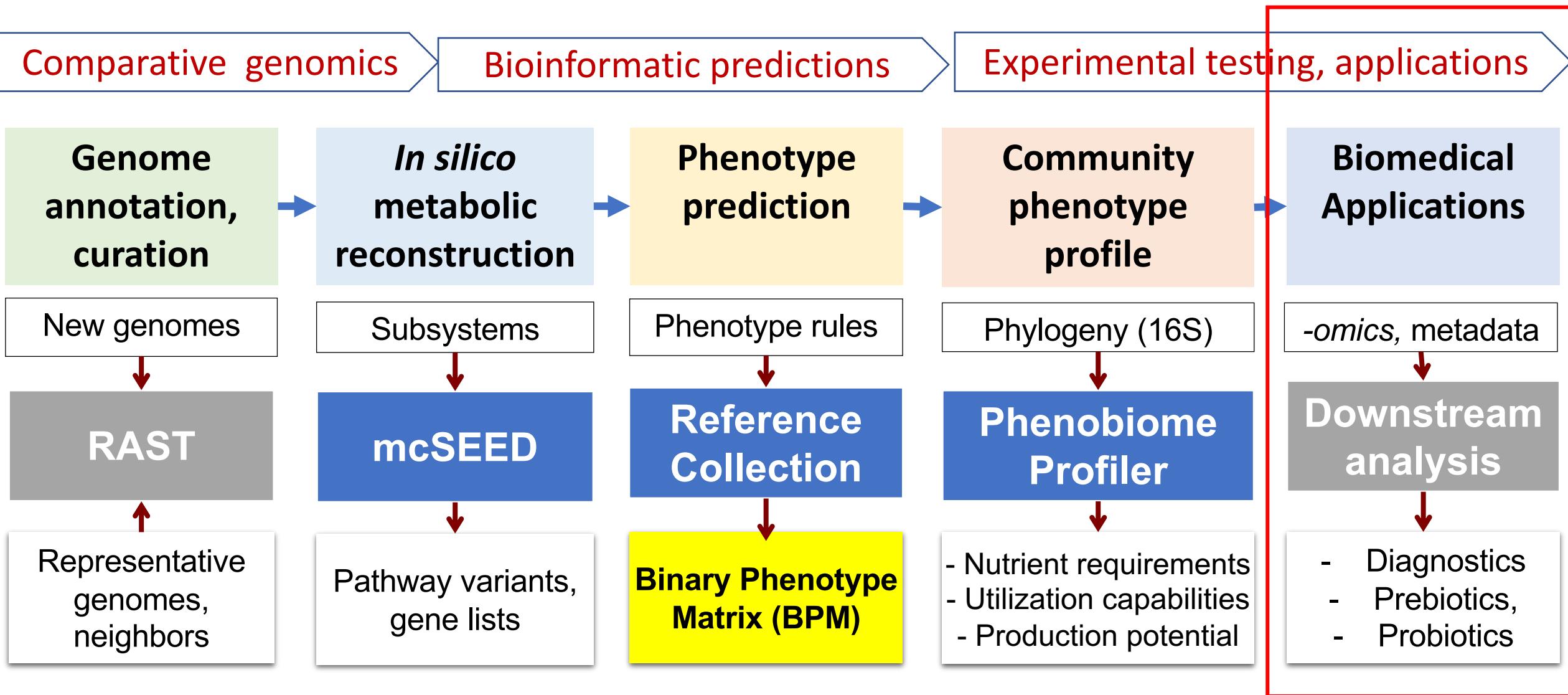
# Phenotype Profiler Workflow



*CPI: fraction of microbiome capable to utilize/produce nutrients and metabolites*

PMID: 34113324

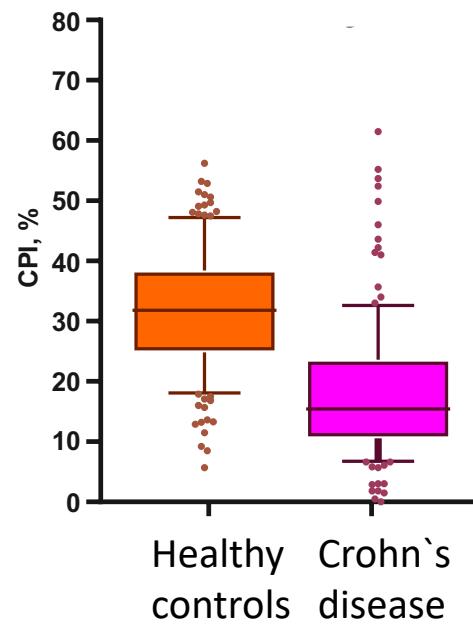
# Phenobiome Approach: Workflow



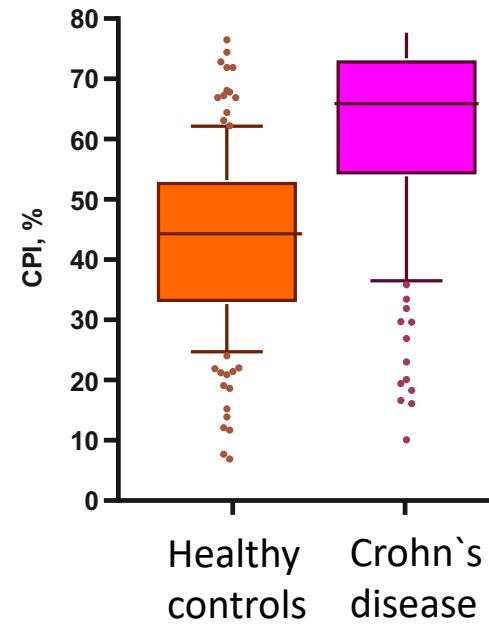
# Example 2: applications in diagnostics, sample classification

Calculated **Community Phenotype Indexes (CPI)** across samples from IBD-Spain study  
Crohn's disease (N=137), Healthy Controls (N=145)

Butyrate production



Propionate production



Increased propionate  
and decreased butyrate  
synthesis potential  
in Crohn's disease samples

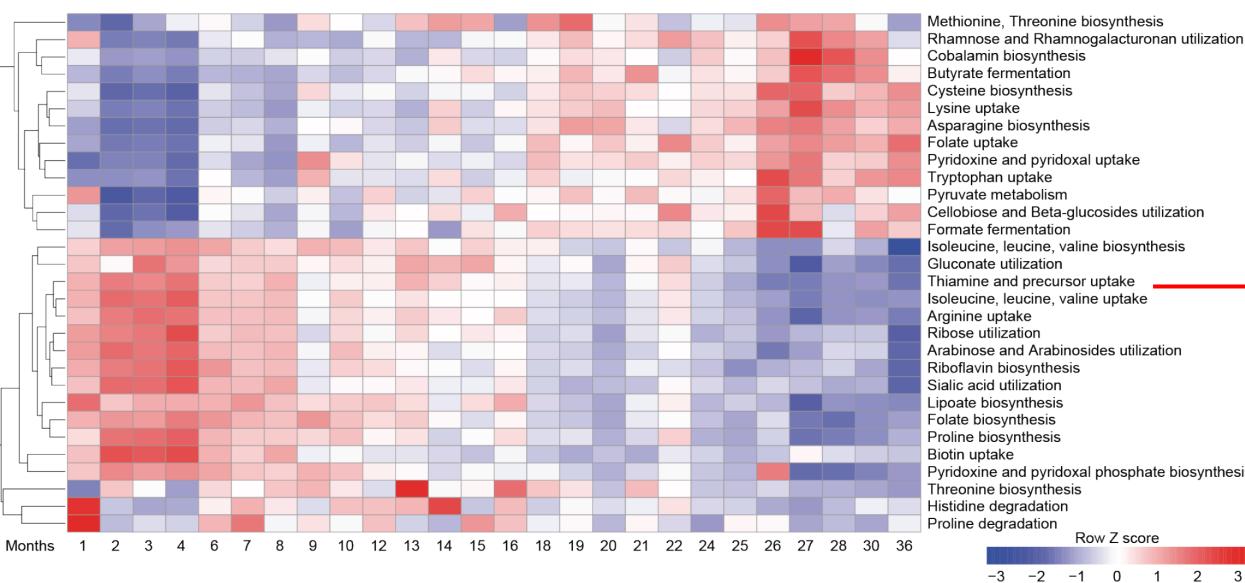
# Example 3: Towards development of therapeutic foods

Randomized Controlled Trial

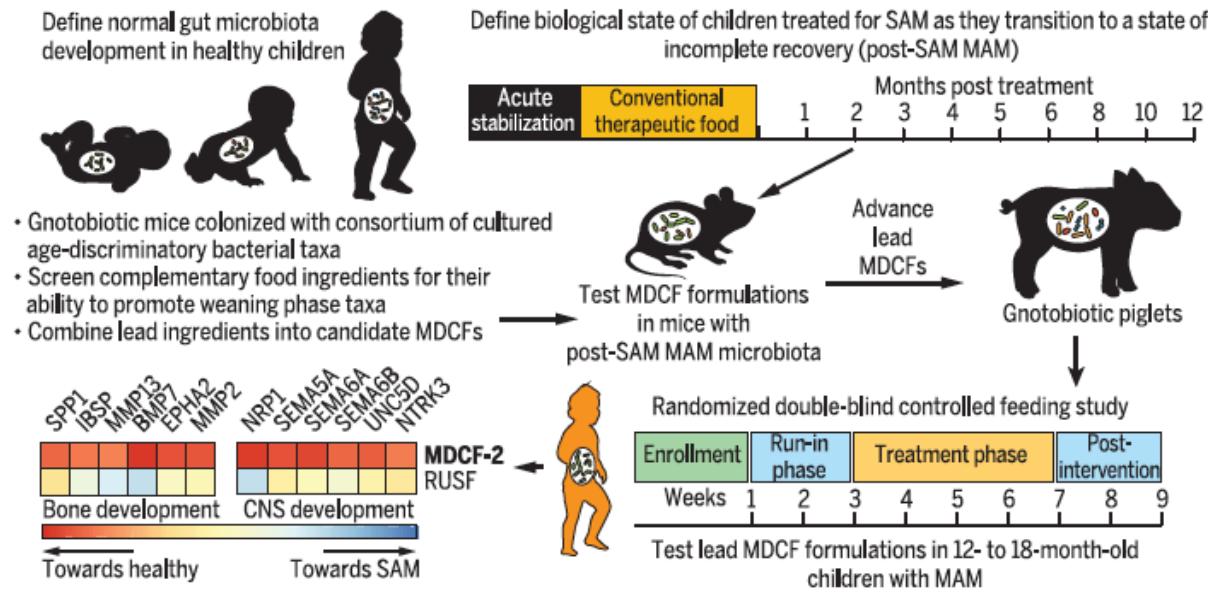
> Science. 2019 Jul 12;365(6449):eaau4732.

## Effects of microbiota-directed foods in gnotobiotic animals and undernourished children

### Top 30 phenotypes in healthy gut microbiota development



### Development/testing of therapeutic foods in children with dysbiosis



Ongoing collaboration with Dr. J. Gordon's Lab at WUSTL

## Example 4:

# Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans

Nature | Vol 595 | 1 July 2021

### Longitudinal microbiome samples from:

- humanized GB mice
- and obese people
- fed diet with fiber snacks

### Monitored by:

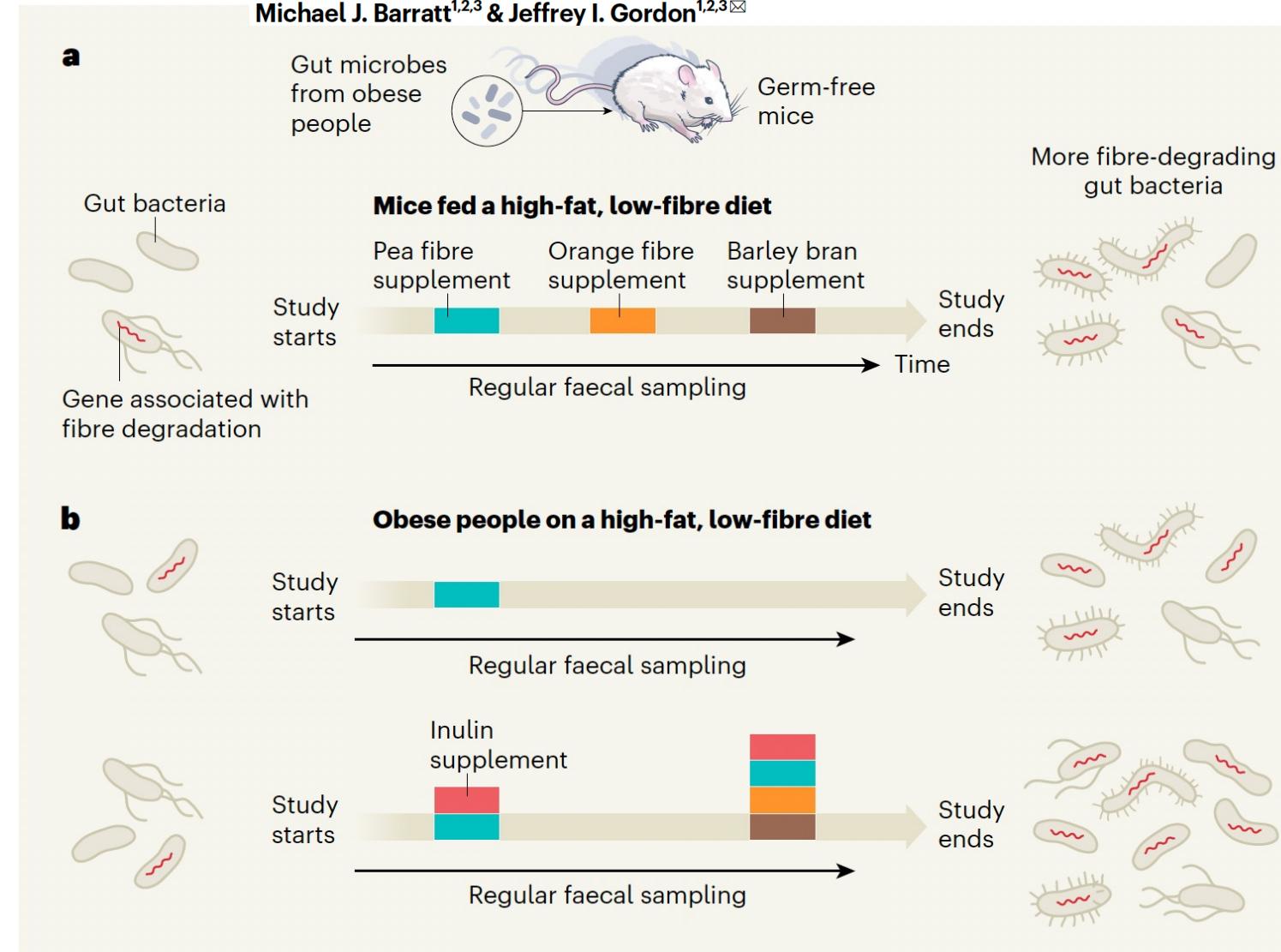
- shotgun metagenomics
- 16S rDNA sequencing

### Followed by Phenobiome analysis:

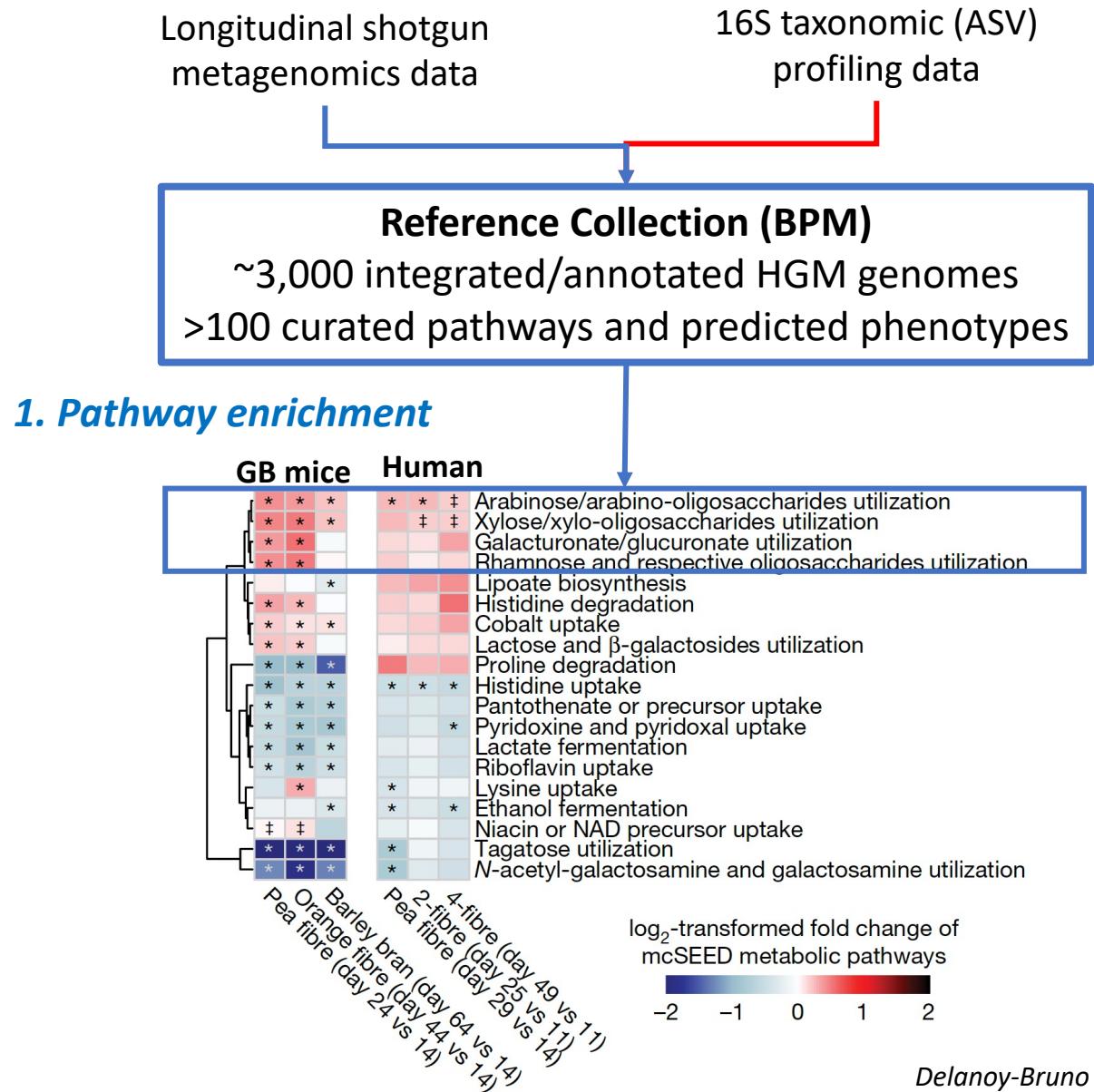
- pathway enrichment;
- CPI profiling

→ Driver phenotypes/nutrients

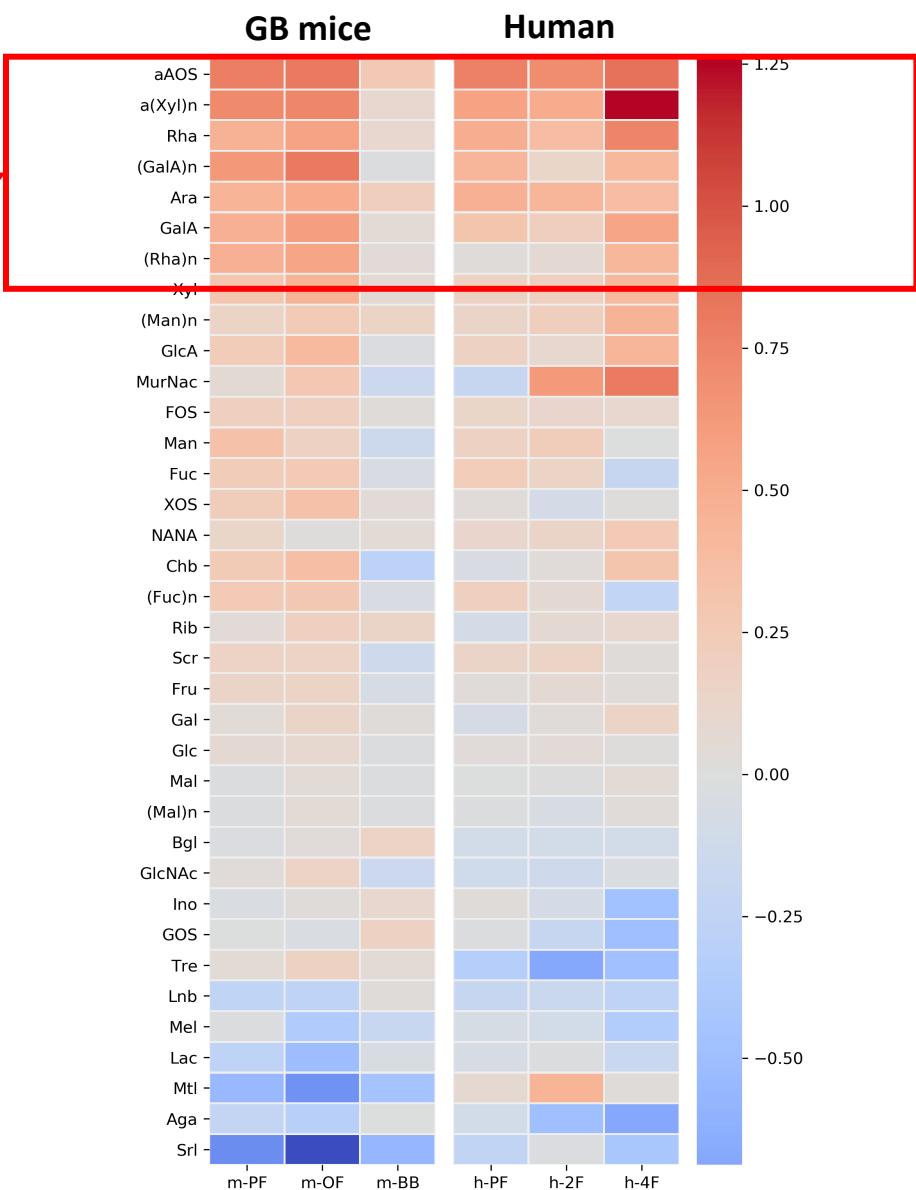
Omar Delannoy-Bruno<sup>1,2</sup>, Chandani Desai<sup>1,2</sup>, Arjun S. Raman<sup>1,2,3</sup>, Robert Y. Chen<sup>1,2</sup>, Matthew C. Hibberd<sup>1,2,3</sup>, Jiye Cheng<sup>1,2,3</sup>, Nathan Han<sup>1,2</sup>, Juan J. Castillo<sup>4</sup>, Garret Couture<sup>4</sup>, Carlito B. Lebrilla<sup>4</sup>, Ruteja A. Barve<sup>5</sup>, Vincent Lombard<sup>6</sup>, Bernard Henrissat<sup>6,7</sup>, Semen A. Leyn<sup>8</sup>, Dmitry A. Rodionov<sup>8,9</sup>, Andrei L. Osterman<sup>8</sup>, David K. Hayashi<sup>10</sup>, Alexandra Meynier<sup>10</sup>, Sophie Vinoy<sup>10</sup>, Kyleigh Kirbach<sup>11</sup>, Tara Wilmot<sup>11</sup>, Andrew C. Heath<sup>12</sup>, Samuel Klein<sup>11</sup>, Michael J. Barratt<sup>1,2,3</sup> & Jeffrey I. Gordon<sup>1,2,3</sup>✉



# Example 4: Driver phenotypes/nutrients from 16S and MGS



## 2. Community Phenotype Index (CPI)



# Acknowledgments

**SBP, Laboratory of Metabolic Genomics.  
Microbiome Bioinformatics Team**



Dmitry  
Rodionov



Simon  
Leyn



Alex  
Arzamasov

**WUSTL, Dr. Jeffrey Gordon's Lab**



**Joint funding: NIDDK, Bill and Melinda Gates Foundation**

**Pavel Pevzner (UCSD)**

**Stan Yablokov (in transit from Russia)**

# Nutriomics Strategy

*Going back to the assignment*

## Input data for a microbiome sample

Taxonomic profile

	RA(%)
Species ID	Dysbiotic Microbiome
<i>Microbe 1</i>	Low
<i>Microbe 2</i>	Low
<i>Microbe 3</i>	High
<i>Microbe 4</i>	Norm
<i>Microbe 5</i>	High



Nutrient Impact Matrix for individual species

Species ID	BINARY PHENOTYPES							
	Nutrient 1	Nutrient 2	Nutrient 3	Nutrient 4	Nutrient 5	Nutrient 6	Nutrient 7	Nutrient 8
<i>Microbe 1</i>	1	0	0	0	0	1	0	0
<i>Microbe 2</i>	0	1	1	1	0	0	1	0
<i>Microbe 3</i>	1	1	0	0	0	1	0	0
<i>Microbe 4</i>	0	0	1	0	1	0	0	0
<i>Microbe 5</i>	1	1	0	0	0	0	0	1

Reference Sample Collection  
(taxonomic profiles for 1,000s  
“normal” microbiome samples)

## Dietary intervention

Optimized Taxonomic profile

TMS1	RA(%)	Change
Species ID	Dysbiotic Microbiome	
<i>Microbe 1</i>	Low	↑
<i>Microbe 2</i>	Low	↑
<i>Microbe 3</i>	High	↓
<i>Microbe 4</i>	Norm	
<i>Microbe 5</i>	High	↓

Supplementation  
by selected  
nutrients

## Bioinformatics objective

Given: (i) Test Microbiome Sample (TMS) with a taxonomic profile;  
 (ii) Reference Sample Collection (RSC) for 1,000s HGM samples;  
 (iii) Nutrient Impact Matrix (NIM) derived from predicted binary phenotypes

Select an optimal sub-set of  $n$  nutrients that collectively would:

- promote the growth of the maximal number of species deemed underrepresented (low) across multiple samples in RSC,
- while promoting the minimal number of in-range (norm) and overrepresented (highs) species as compared to RSC

# Data types and formats

1. A set of “deviant” (potentially pathologic) **TMSs** (up to 100) described by two-column tables ([ASV\_taxonomic ID], [RA%]).
2. **RSC** of S (~ 1,000) microbiome samples (described by the same type of RA tables as above) that are presumed to be “mostly normal”- to be used as a definition of a normal core (with outliers).
3. **NIM** providing NI values for N nutrients for all relevant ASVs. Columns: [ASV\_taxonomic ID], [**NI<sub>n</sub>** values], where “n” is from 1 to “N” (in our case ~100 distinct nutrients/phenotypes).

# Input data: Reference Sample Collection (RSC)

Samples ~ 1,000

ASV (total ~ 2,000; at ave RA $\geq$ 0.02% ~ 300)

	R1382663	ERR1385091	ERR1385532	ERR1383644	ERR1383024	ERR1384548	ERR1384220	ERR1383078	ERR1385168	ERR1382380	Average	Min	Max
<i>Bacteroides uniformis</i>	1.27%	8.49%	26.19%	0.79%	2.13%	3.94%	11.30%	16.42%	0.15%	2.69%	7.03%	0.00%	37.69%
<i>Phocaeicola vulgatus</i>	0.11%	2.23%	0.81%	2.60%	1.75%	6.59%	11.24%	0.00%	0.07%	2.49%	5.16%	0.00%	46.33%
<i>Alistipes onderdonkii</i>	1.15%	4.34%	1.75%	23.46%	0.35%	5.70%	43.45%	0.00%	0.00%	0.00%	4.25%	0.00%	43.45%
<i>Alistipes putredinis</i>	9.10%	7.98%	4.84%	6.42%	2.08%	8.85%	0.00%	0.00%	0.00%	0.00%	3.31%	0.00%	20.57%
<i>Phocaeicola dorei</i>	1.39%	3.06%	1.90%	0.96%	3.54%	1.12%	0.00%	27.18%	0.00%	9.33%	2.90%	0.00%	28.12%
<i>Faecalibacterium prausnitzii</i>	3.62%	2.18%	1.30%	2.32%	2.01%	1.21%	4.03%	3.00%	0.00%	9.06%	2.71%	0.00%	13.08%
<i>Ruminococcus bromii</i>	0.00%	1.29%	0.01%	0.00%	1.92%	0.97%	0.01%	0.00%	0.00%	0.00%	2.26%	0.00%	23.25%
<i>Blautia wexlerae</i>	0.80%	0.21%	1.49%	0.16%	0.07%	0.46%	0.93%	1.14%	6.28%	3.37%	2.17%	0.00%	16.91%
<i>Akkermansia muciniphila</i>	0.11%	0.01%	0.03%	5.00%	0.00%	0.39%	0.46%	0.02%	0.00%	0.00%	2.15%	0.00%	28.14%
<i>[Eubacterium] rectale</i>	1.39%	0.74%	2.82%	1.08%	0.27%	8.85%	0.00%	0.60%	0.00%	0.97%	1.81%	0.00%	30.88%
<i>Prevotella copri</i>	0.00%	0.00%	0.03%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.05%	1.79%	0.00%	64.96%
<i>Bacteroides faecichinchillae/faecis/thetaiotaomicron</i>	0.46%	0.12%	2.93%	0.40%	0.47%	0.70%	0.98%	2.72%	0.00%	0.52%	1.78%	0.00%	61.47%
<i>Dialister invisus</i>	0.00%	0.02%	0.00%	0.50%	0.00%	0.00%	0.00%	0.00%	0.00%	3.50%	1.63%	0.00%	27.72%
<i>Bacteroides stercoris</i>	0.39%	0.72%	1.41%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.98%	1.53%	0.00%	27.75%
<i>Escherichia coli/Escherichia fergusonii/Escherichia marmotae/Shigella</i>	4.23%	0.61%	0.08%	0.94%	0.00%	0.00%	0.00%	0.00%	0.31%	0.03%	1.51%	0.00%	25.81%
<i>Gemmiger formicilis/Faecalibacterium prausnitzii</i>	1.40%	1.81%	0.01%	2.40%	0.34%	0.00%	0.12%	0.00%	0.01%	2.13%	1.34%	0.00%	13.06%
<i>Alistipes finegoldii</i>	1.13%	0.96%	0.75%	1.85%	0.67%	0.68%	1.64%	2.21%	15.27%	4.34%	1.29%	0.00%	15.27%
<i>Bacteroides koreensis/kribbi</i>	0.70%	0.84%	0.00%	0.12%	0.00%	0.00%	0.00%	0.93%	0.00%	6.27%	1.29%	0.00%	13.89%
<i>Alistipes shahii</i>	1.24%	4.73%	1.22%	3.59%	2.74%	0.00%	0.00%	0.00%	0.00%	0.00%	1.27%	0.00%	36.57%
<i>Parabacteroides merdae</i>	0.18%	1.50%	0.54%	0.00%	0.49%	0.48%	0.00%	0.00%	44.62%	1.92%	1.26%	0.00%	44.62%
• • •													
<i>Mediterranea massiliensis</i>	0.00%	0.00%	0.00%	0.00%	2.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	2.20%
<i>Pseudobutyryvibrio ruminis</i>	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.06%	0.00%	0.00%	0.02%	0.00%	0.66%
<i>Spiroplasma monobiae/Falcatimonas natans/Pseudobutyryvibrio ruminis</i>	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	1.41%
<i>Kineothrix alysoides/Robinsoniella peoriensis</i>	0.03%	0.02%	0.00%	0.00%	0.13%	0.00%	0.08%	0.00%	0.00%	0.00%	0.02%	0.00%	0.79%
<i>Eisenbergiella massiliensis</i>	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	1.44%
<i>Spiroplasma monobiae/Holdemania filiformis/Amedibacillus dolichus</i>	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	2.12%
<i>Limosilactobacillus mucosae</i>	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	1.95%
<i>Merdimonas faecis</i>	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.36%	0.00%	0.02%	0.00%	0.36%
<i>Anaeromassilibacillus senegalensis/Ruminococcus flavefaciens/Ruminococcus</i>	0.00%	0.10%	0.00%	0.07%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	0.61%
<i>Lachnoclostridium pacaense/[Clostridium] methoxybenzovorans/Lachnoclostridium</i>	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.07%	0.02%	0.00%	0.46%
<i>Flintibacter butyricus/Intestinimonas butyriciproducens/Ruminiclostridium</i>	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	1.07%
<i>Prevotella baroniae/bivia/bryantii/buccalis/enoeca/marshii/oralis/polyphaga</i>	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	2.02%
<i>Blautia hansenii/hominis/marasmi</i>	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.15%	0.00%	0.02%	0.00%	0.92%
• • •													
	93%	96%	99%	93%	86%	97%	97%	99%	96%	97%			

# Input data: Test Microbiome Sample (TMS)

ASV taxonomic description	ERR1384977
<i>Bacteroides uniformis</i>	0.98%
<i>Phocaeicola vulgatus</i>	4.24%
<i>Blautia wexlerae</i>	0.05%
<i>Prevotella copri</i>	31.75%
<i>Dialister invisus</i>	12.10%
<i>Escherichia coli/Escherichia fergusonii/Escherichia marmotae/Shigella flexneri</i>	0.06%
<i>Bifidobacterium adolescentis/faecale</i>	0.88%
<i>Bifidobacterium breve/longum</i>	0.20%
[ <i>Ruminococcus</i> ] <i>gnavus</i>	0.02%
<i>Flavonifractor plautii</i>	0.39%
<i>Veillonella dispar/infantium</i>	0.22%
<i>Lactobacillus amylovorus/ultunensis</i>	22.58%
<i>Dorea formicigenerans</i>	0.13%
<i>Neglecta timonensis/Caproiciproducens galactitolivorans</i>	0.03%
<i>Fusobacterium nucleatum</i>	11.28%
<i>Sutterella wadsworthensis</i>	1.64%
<i>Coprobacter fastidiosus</i>	0.43%
<i>Bifidobacterium bifidum</i>	0.05%
<i>Megasphaera micronuciformis</i>	5.52%
<i>Mediterraneibacter glycyrrhizinilyticus</i>	0.02%
<i>Sellimonas intestinalis/Drancourtella massiliensis</i>	0.06%
<i>Megasphaera cerevisiae</i>	2.62%
<i>Victivallis vadensis</i>	0.05%
<i>Limosilactobacillus mucosae</i>	1.95%

## From RSC

Average	delta
7.03%	-6.05%
5.16%	-0.91%
2.17%	-2.12%
1.79%	29.96%
1.63%	10.48%
1.51%	-1.45%
0.97%	-0.09%
0.58%	-0.38%
0.51%	-0.49%
0.26%	0.13%
0.24%	-0.02%
0.23%	22.36%
0.14%	-0.02%
0.12%	-0.09%
0.12%	11.17%
0.09%	1.55%
0.06%	0.37%
0.06%	-0.01%
0.06%	5.47%
0.04%	-0.02%
0.04%	0.02%
0.03%	2.60%
0.02%	0.02%
0.02%	1.93%

# Input data: Nutrient Impact Matrix (NIM)

ASV (total ~ 2,000))

Nutrients; N ~ 100; carbohydrates, B-vitamins, amino acids

ASV_taxonomic description	Glc	Gal	Fru	Man	Tag	GlcA	GalA	GalNA	GalN	SicNA	IanNA	NANA	DdGlcA	Xyl	Ara	Fuc	Rha	Rib	Gnt	Mal	Scr	Tre	Chb	Lnb	Lac	Mtl	Atl	Xtl	Srl	Gtl
<i>Bacteroides uniformis</i>	1.00	1.00	1.00	0.49	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	1.00	0.99	0.99	1.00	0.00	1.00	1.00	0.00	1.00	0.00	0.49	0.00	0.00	0.00	0.00	
<i>Bacteroides faecichinchillae/faecis/thetaiotaomi</i>	1.00	1.00	0.90	0.90	0.00	0.70	0.70	0.00	0.00	1.00	0.00	0.20	0.00	0.70	0.70	0.70	0.70	1.00	0.90	0.00	0.80	0.90	0.00	1.00	0.00	0.60	0.20	0.00	0.00	0.00
<i>Alistipes putredinis</i>	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Bacteroides cellulosilyticus/timonensis</i>	1.00	1.00	1.00	0.40	0.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00	0.60	0.00	0.00	0.00	0.00	
<i>Alistipes onderdonkii</i>	1.00	1.00	0.00	0.80	0.00	0.20	0.20	0.00	0.00	0.20	0.00	0.80	0.00	0.20	0.00	0.00	0.80	1.00	0.00	0.80	0.00	0.00	0.20	0.00	0.00	0.80	0.00	0.00	0.00	0.00
<i>Barnesiella intestinihominis</i>	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Phocaeicola dorei</i>	1.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.29	0.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Alistipes communis</i>	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	
<i>Alistipes shahii</i>	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Akkermansia muciniphila</i>	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Parabacteroides distasonis</i>	1.00	1.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Bacteroides ovatus</i>	1.00	1.00	0.57	1.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	1.00	1.00	1.00	0.86	0.00	1.00	1.00	0.00	1.00	0.00	1.00	0.43	0.00	0.00	0.00	0.00
<i>Faecalibacterium prausnitzii</i>	0.80	1.00	1.00	0.20	0.00	0.80	0.80	0.22	0.42	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.18	1.00	0.20	0.29	0.00	0.80	1.00	0.00	0.00	0.20	0.00		
<i>Phascolarctobacterium faecium</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Roseburia inulinivorans</i>	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	
<i>Parabacteroides johnsonii</i>	1.00	1.00	0.33	0.33	0.00	0.67	0.67	0.00	0.33	1.00	0.33	1.00	0.00	0.67	0.00	0.00	0.67	0.67	0.00	0.33	0.33	0.33	0.67	0.00	0.33	0.00	0.00	0.00	0.00	
<i>Ruminococcoides bili</i>	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Coprobacter fastidiosus</i>	1.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Alistipes finegoldii</i>	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	
<i>Blautia wexlerae</i>	1.00	0.33	1.00	0.67	0.00	0.00	0.00	0.00	0.33	0.00	1.00	0.00	0.67	1.00	0.00	0.33	0.00	0.00	0.67	1.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.67	0.00	
<i>Alistipes ihumii</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Bacteroides xyloisolvens</i>	1.00	1.00	0.40	1.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	1.00	1.00	0.60	0.00	1.00	1.00	0.00	1.00	0.00	0.20	0.20	0.00	0.00	0.00	0.00	
<i>Ruminococcus chamanellensis</i>	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Fusicatenibacter saccharivorans</i>	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	
<i>Eubacterium ventriosum</i>	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Oscillibacter valericigenes/Pseudoflavonifactor</i>	0.67	0.67	0.67	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.67	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Anaerobutyricum hallii</i>	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00	0.00
<i>Flintibacter butyricus/Oscillibacter valericigenes</i>	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.50	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Intestinimonas butyriproducens</i>	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	
<i>Bacteroides koreensis/kribbi</i>	1.00	1.00	0.67	1.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	0.67	0.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00	0.33	0.00	0.00	0.00	0.00	
<i>Gemmiger formicilis/Faecalibacterium prausnitz</i>	0.88	1.00	1.00	0.13	0.00	0.88	0.88	0.38	0.50	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.25	1.00	0.13	0.25	0.00	0.88	1.00	0.00	0.00	0.00	0.25	0.00	
<i>Roseburia hominis</i>	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00

...

# Specific tasks

1. Select metrics for microbiome balancing/optimization.
  - You may use a version of a **comparative species distribution** approach outlined above. (*Note: for simplicity, we consider RAs of all ASVs as independent variables, ignoring possible dependencies and interactions.*)
  - Alternatively, if you prefer, you may use standard **metrics of diversity** (eg Shannon entropy) that are popular in the microbiome field and set the goal to maximize diversity of your TMS to make it closer to what is realistically observed in RSC.
2. Based on a selected metric(s), suggest and implement a computational approach that would allow you to analyze the RSC as a foundation for further categorization of ASVs in TMSs.
3. Based on the above, design, implement and test an algorithm for personalized (sample-specific) selection of up to n (=10) nutrients aimed to balance/optimize microbiome composition observed in “deviant” TMSs.

# Suggested workflow

1. Read the provided document for Problem Description
2. Ask for clarification as needed for Problem Formulation
3. Data files will be provided
4. Consider a proposed approach (not algorithm) , think of your own and let's discuss
5. Work on implementation
6. Discuss outcomes

*Please note that this is indeed an open-ended research task. It means that we do not have any preconceived notion of a right or wrong approach to define metrics of success that would enable optimization (selecting between various combinations of nutrients). In fact, the main challenge of this task is to propose, rationalize and test such an approach.*

# More optional reading (applications)

- Delannoy-Bruno O, Desai C, Raman AS, Chen RY, Hibberd MC, Cheng J, Han N, Castillo JJ, Couture G, Lebrilla CB, Barve RA, Lombard V, Henrissat B, Leyn SA, Rodionov DA, Osterman AL, Hayashi DK, Meynier A, Vinoy S, Kirbach K, Wilmot T, Heath AC, Klein S, Barratt MJ, Gordon JI. [Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans](#). *Nature*. 2021 Jul;595(7865):91-95. doi: 10.1038/s41586-021-03671-4. Epub 2021 Jun 23. PubMed PMID: 34163075;
- Iablokov SN, Novichkov PS, Osterman AL, Rodionov DA. [Binary Metabolic Phenotypes and Phenotype Diversity Metrics for the Functional Characterization of Microbial Communities](#). *Front Microbiol*. 2021;12:653314. doi: 10.3389/fmicb.2021.653314. eCollection 2021. PubMed PMID: 34113324;
- Chen RY, Kung VL, Das S, Hossain MS, Hibberd MC, Guruge J, Mahfuz M, Begum SMKN, Rahman MM, Fahim SM, Gazi MA, Haque R, Sarker SA, Mazumder RN, Di Luccia B, Ahsan K, Kennedy E, Santiago-Borges J, Rodionov DA, Leyn SA, Osterman AL, Barratt MJ, Ahmed T, Gordon JI. [Duodenal Microbiota in Stunted Undernourished Children with Enteropathy](#). *N Engl J Med*. 2020 Jul 23;383(4):321-333. doi: 10.1056/NEJMoa1916004. PubMed PMID: 32706533
- Feng L, Raman AS, Hibberd MC, Cheng J, Griffin NW, Peng Y, Leyn SA, Rodionov DA, Osterman AL, Gordon JI. [Identifying determinants of bacterial fitness in a model of human gut microbial succession](#). *Proc Natl Acad Sci U S A*. 2020 Feb 4;117(5):2622-2633. doi: 10.1073/pnas.1918951117. Epub 2020 Jan 22. PubMed PMID: 31969452;
- Raman AS, Gehrig JL, Venkatesh S, Chang HW, Hibberd MC, Subramanian S, Kang G, Bessong PO, Lima AAM, Kosek MN, Petri WA Jr, Rodionov DA, Arzamasov AA, Leyn SA, Osterman AL, Huq S, Mostafa I, Islam M, Mahfuz M, Haque R, Ahmed T, Barratt MJ, Gordon JI. [A sparse covarying unit that describes healthy and impaired human gut microbiota development](#). *Science*. 2019 Jul 12;365(6449). doi: 10.1126/science.aau4735. PubMed PMID: 31296739
- Gehrig JL, Venkatesh S, Chang HW, Hibberd MC, Kung VL, Cheng J, Chen RY, Subramanian S, Cowardin CA, Meier MF, O'Donnell D, Talcott M, Spears LD, Semenkovich CF, Henrissat B, Giannone RJ, Hettich RL, Ilkayeva O, Muehlbauer M, Newgard CB, Sawyer C, Head RD, Rodionov DA, Arzamasov AA, Leyn SA, Osterman AL, Hossain MI, Islam M, Choudhury N, Sarker SA, Huq S, Mahmud I, Mostafa I, Mahfuz M, Barratt MJ, Ahmed T, Gordon JI. [Effects of microbiota-directed foods in gnotobiotic animals and undernourished children](#). *Science*. 2019 Jul 12;365(6449). doi: 10.1126/science.aau4732. PubMed PMID: 31296738;