



Tour d'horizon des méthodes d'exploration du microbiote en 2023

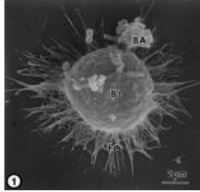
Mathieu Almeida
INRAE MetaGenoPolis, Jouy-en-Josas

INRAE



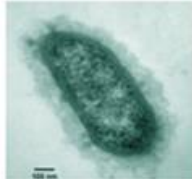
L'Humain est microbien, écosystème, symbiose

[Zaman *et al.*, 1999]



Protozoaire
(*Blastocystis hominis*)

[Samuel *et al.*, 2007]

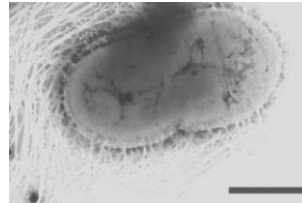


Archée
(*Methanobrevibacter smithii*)



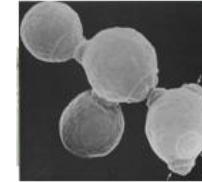
Cellules de l'hôte

[Derrien *et al.*, 2004]



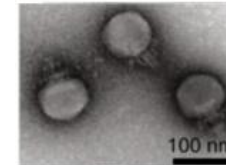
Bactérie
(*Akkermansia muciniphila*)

[Nollin and Borgers, 1975]



Champignon
(*Candida albicans*)

[Sausset *et al.*, 2020]

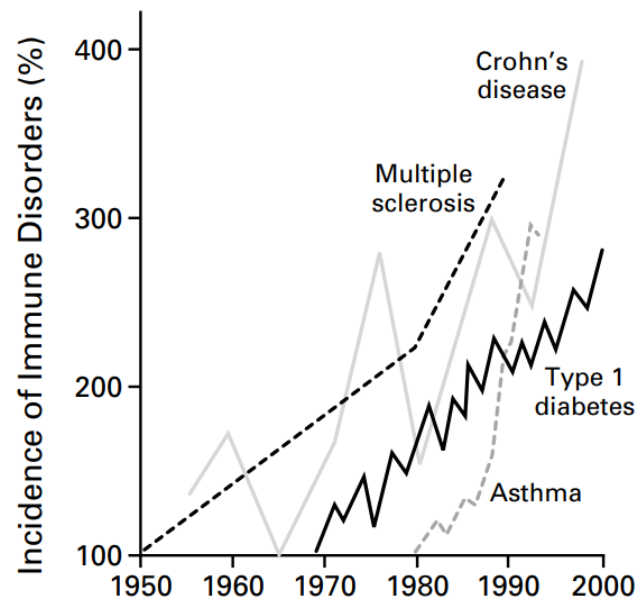
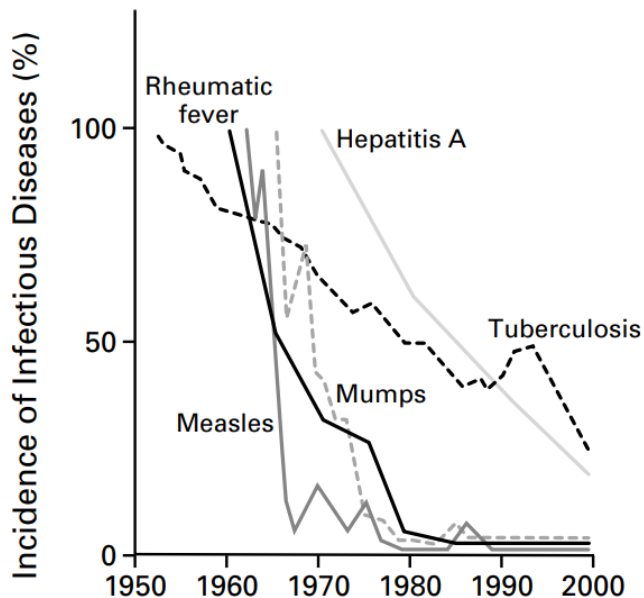


Virus
(*crAssphage*)

**De nombreux micro-organismes qui
peuvent jouer un rôle dans la santé de
l'hôte.**

Vers une augmentation des maladies chroniques et immunitaires

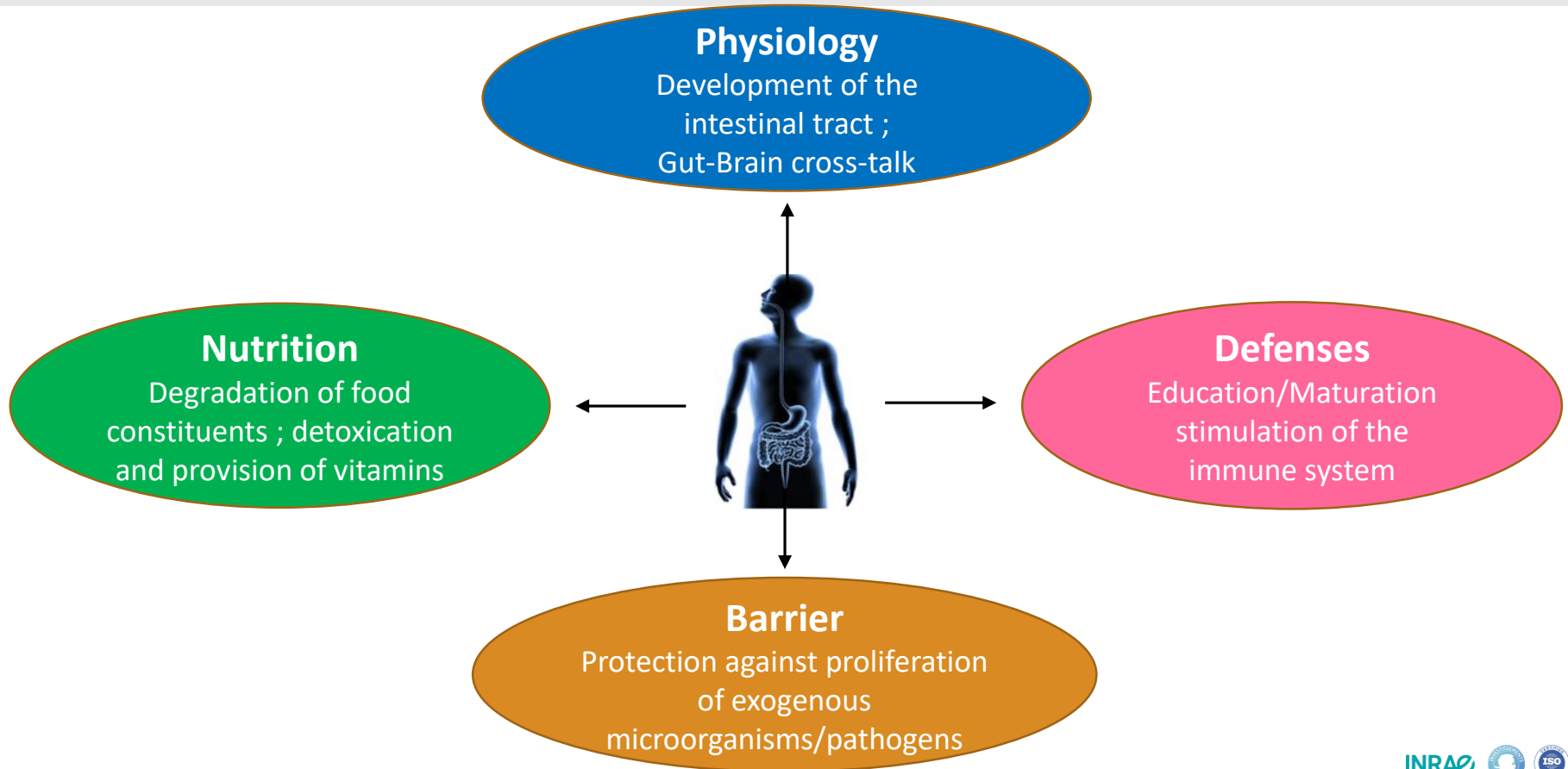
[Bach JF, N Eng J Med, 2002]

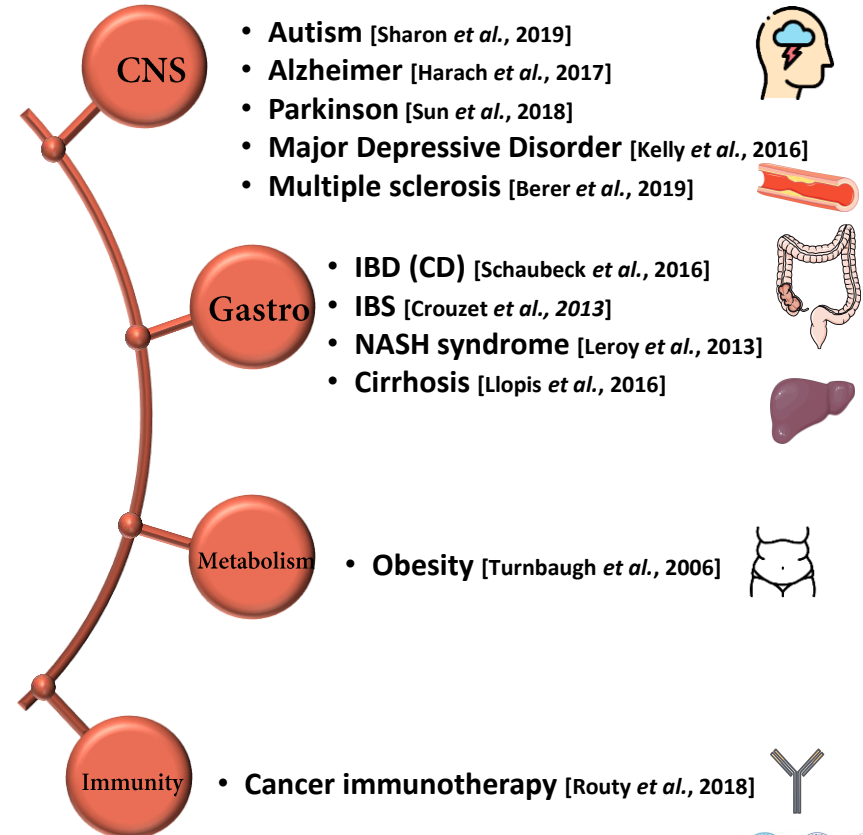
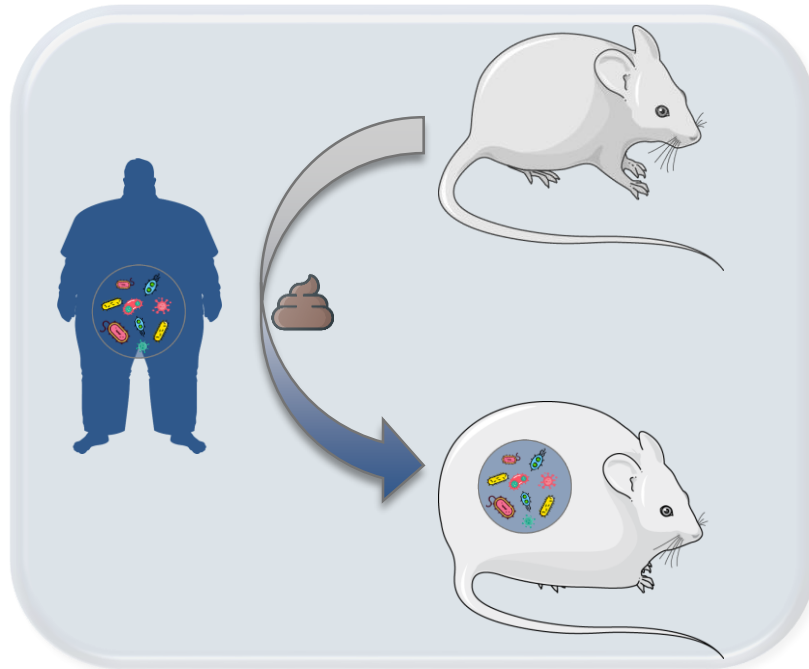


Diminution des maladies infectieuses mais **augmentation des maladies chroniques/immunitaire** ces 60 dernières années

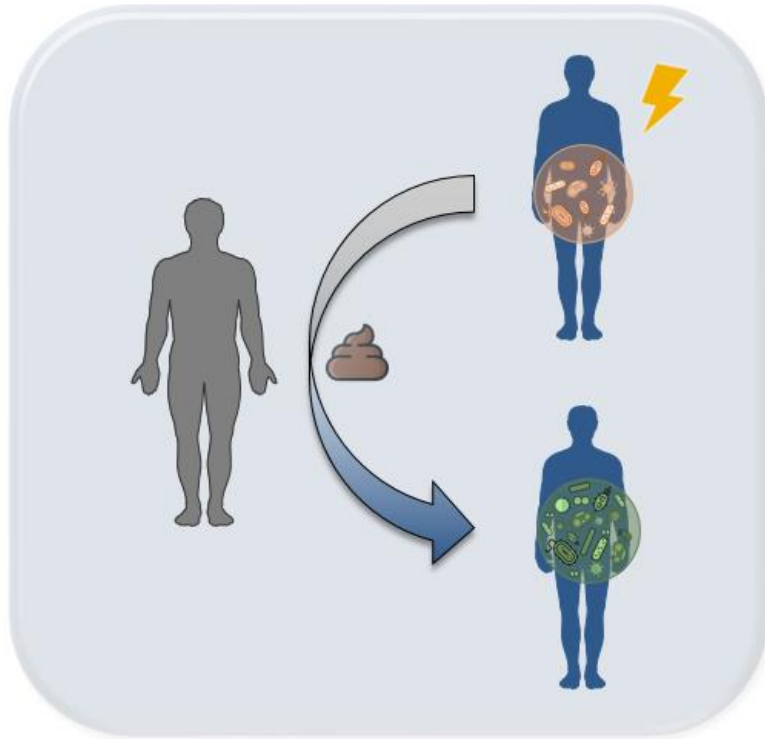
25% des humains peuvent être concernés d'ici 2025

Le rôle du microbiote dans la santé humaine

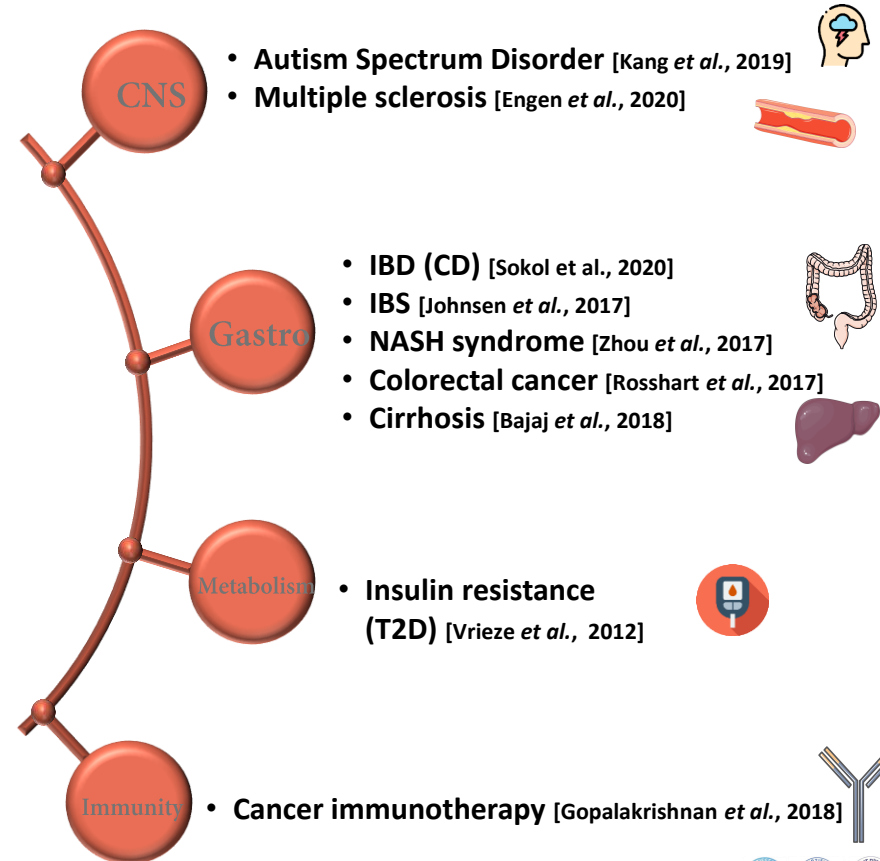




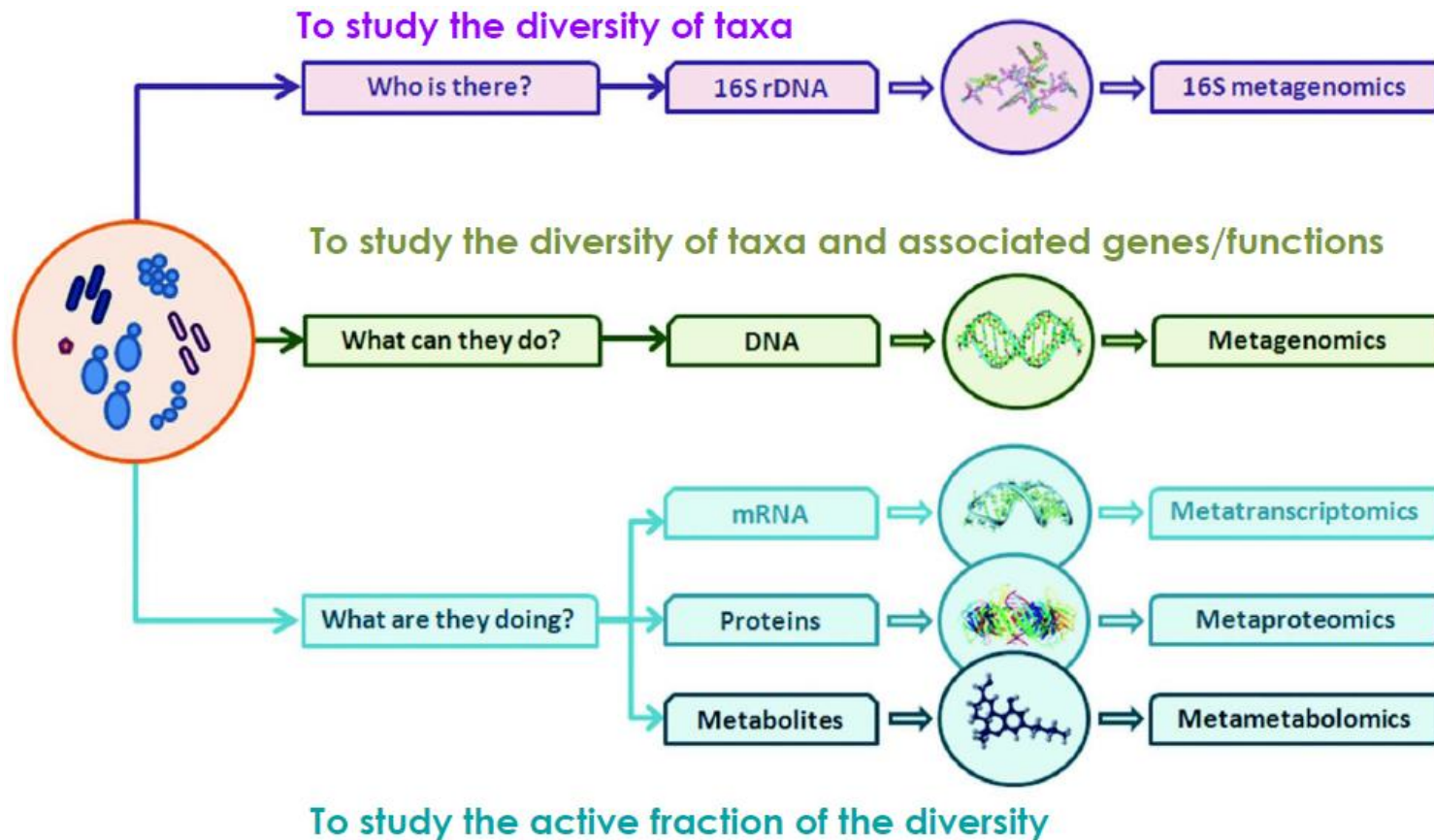
Preuve de causalité par transfert fécal chez l'humain



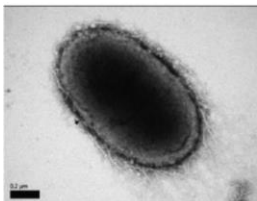
Comment explorer la composition microbienne et sa fonction dans la maladie et la santé ?



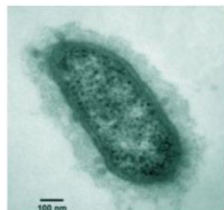
Comment explorer le microbiote?



16S rRNA



Bacteria

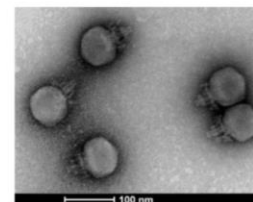


Archaea



Ev Host cells

Marker-gene sequencing strategies
Can't fish-out the whole diversity at once



Centrifugation
Size filtering
Enrichment...

No Marker !

Viruses

18S rRNA / ITS

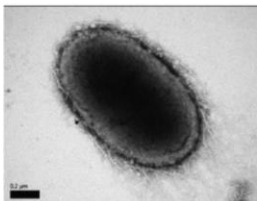


Protozoa



Fungi

Comment explorer le microbiote? 16S vs metagenomique

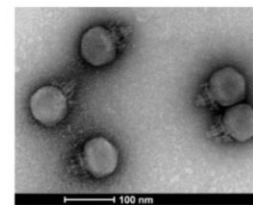


Bacteria

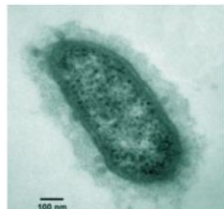


Ev Host cells

Shotgun metagenomics =
Random complete sequencing of DNA



Viruses



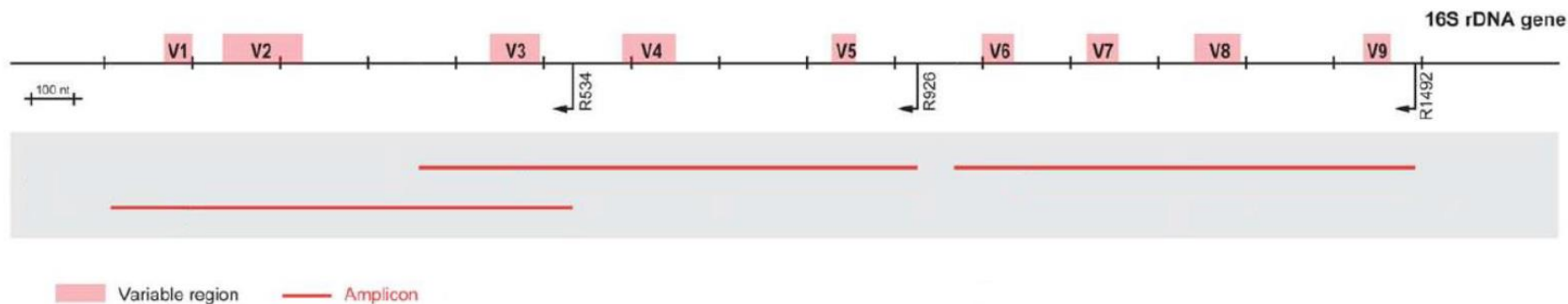
Archaea



Protozoa



Fungi

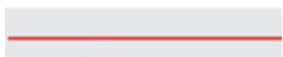


- Prokaryota (Bacteria, Archaea): contain 16S rRNA genes
- Conserved regions between variable regions used to target and amplify 16S rRNA genes.
- Variable regions used to annotate taxa.

Comment explorer le microbiote? 16S vs metagenomique



Streptococcus thermophilus



Streptococcus salivarius

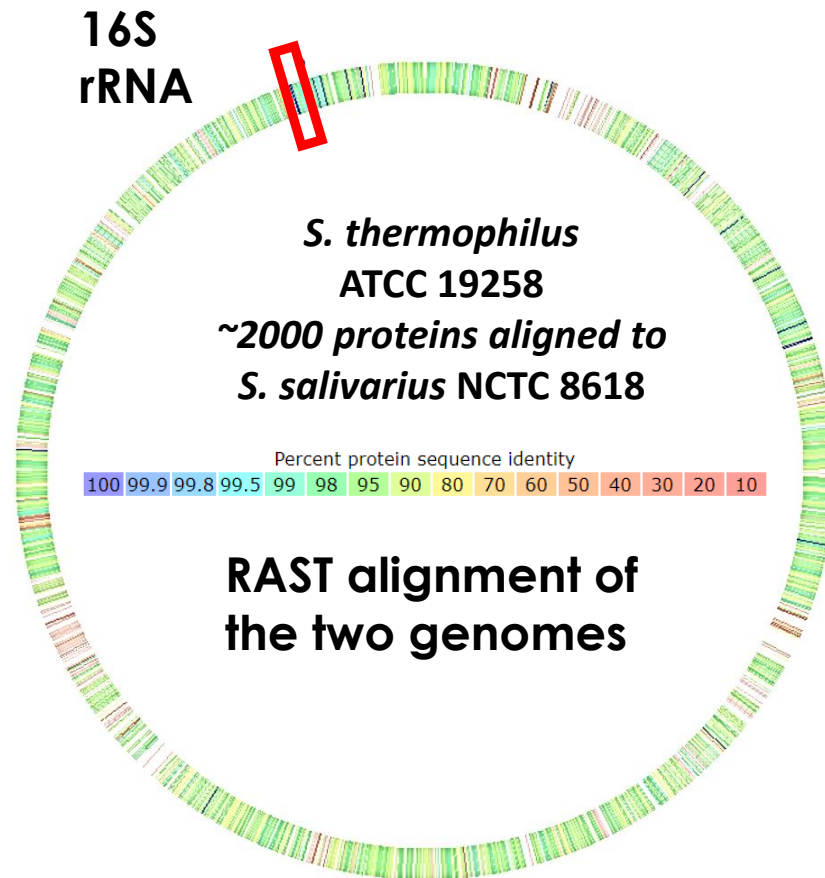
	Max Score	Total Score	Query Cover	E value	Per. Ident
Streptococcus thermophilus strain ATCC 19258	789	3948	100%	0.0	100.00%
Streptococcus salivarius strain NCTC8618	784	4704	100%	0.0	99.77%

Exact match between the two species

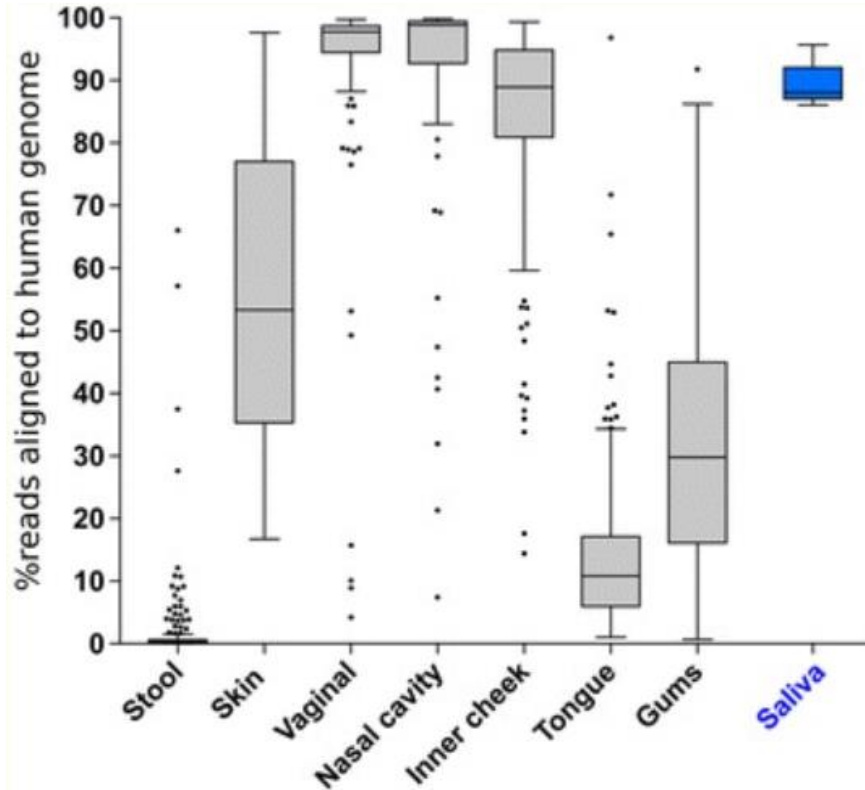
Streptococcus salivarius strain NCTC8618 genome assembly, chromosome: 1

Score	Expect	Identities	Gaps	Strand
784 bits(424)	0.0	426/427(99%)	0/427(0%)	Plus/Plus
Query 1	TAGGGAATCTTCGGCAATGGGGCAACCTGACCGAGCAACGCCGCGTGAGTGAAGAAGG	60		
Sbjct 15213	TAGGGAATCTTCGGCAATGGGGCAACCTGACCGAGCAACGCCGCGTGAGTGAAGAAGG	15272		
Query 61	TTTTCGGATCGTAAAGCTCTGTTGTAAGTCAAGAACGGGTGTGAGAGTGGAAAGTTCACA	120		
Sbjct 15273	TTTTCGGATCGTAAAGCTCTGTTGTAAGTCAAGAACGAGTGTGAGAGTGGAAAGTTCACA	15332		
Query 121	CTGTGACGGTAGCTTACCAGAAAGGGACGGCTAACTACGTGCCAGCAGCCGGTAAATAC	180		
Sbjct 15333	CTGTGACGGTAGCTTACCAGAAAGGGACGGCTAACTACGTGCCAGCAGCCGGTAAATAC	15392		
Query 181	GTAGGTCCTCGAGCGTTGTCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTTTGATAA	240		
Sbjct 15393	GTAGGTCCTCGAGCGTTGTCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTTTGATAA	15452		
Query 241	GTCTGAAGTTAAAGGCTGTGGCTCAACCATAGTTCGCTTTGAAACTGTCAAAC TTGAGT	300		
Sbjct 15453	GTCTGAAGTTAAAGGCTGTGGCTCAACCATAGTTCGCTTTGAAACTGTCAAAC TTGAGT	15512		
Query 301	GCAGAAGGGGAGAGTGGAAATTCATGTGTAGCGGTGAAATGCGTAGATATATGGAGGAAC	360		
Sbjct 15513	GCAGAAGGGGAGAGTGGAAATTCATGTGTAGCGGTGAAATGCGTAGATATATGGAGGAAC	15572		
Query 361	ACCGTGGCGAAAGCGGCTCTGGTCTGTAAC TGACGCTGAGGCTCGAAAGCGTGGGGA	420		
Sbjct 15573	ACCGTGGCGAAAGCGGCTCTGGTCTGTAAC TGACGCTGAGGCTCGAAAGCGTGGGGA	15632		
Query 421	GCGAACA 427			
Sbjct 15633	GCGAACA 15639			

When exploring the entire genome, specific genes detected for each species



Cependant, le microbiote peut être difficile à explorer sans amplification



- If analyzing nasal cavity samples : almost 100% of the reads in the trash !
- So for some ecosystem : 16S is still a better option
- Unless we can deplete host DNA prior to DNA sequencing !

[Marot et al, Microbiome, 2018]

Le séquençage métagénomique: Un outil performant pour explorer le lien entre microbiote et sa fonction vis-à-vis d'une pathologie

Collecter & préserver

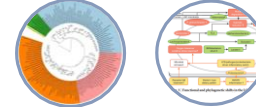
Transférer & Identifier

Aliquoter & Stocker

Extraction ADN & Séquençage

Exploration Taxa & Fonction

Modélisation & Analyse statistique

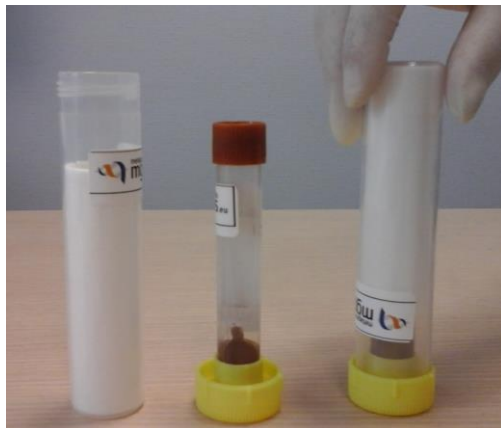


Chaque étape est cruciale dans l'exploration du lien microbiote/santé

Kit de collecte pour stocker et préserver les échantillons de selle.



Kit de collecte pour stocker et préserver les échantillons de selle.

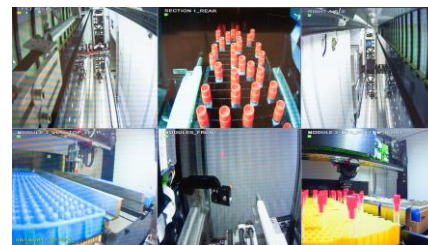
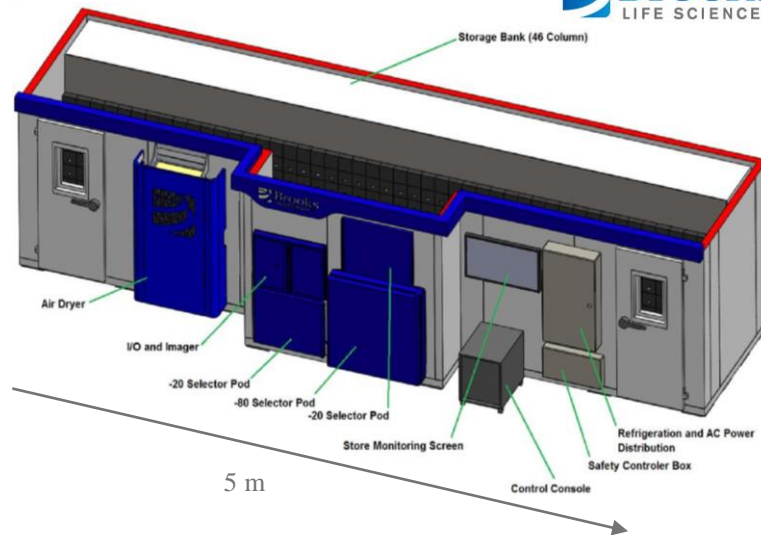


- Les échantillons sont conservés par la solution de RNA-later ou DNA/RNA Shield présente dans le tube.
- Préserve les échantillons même à température ambiante sur au moins 20 jours.

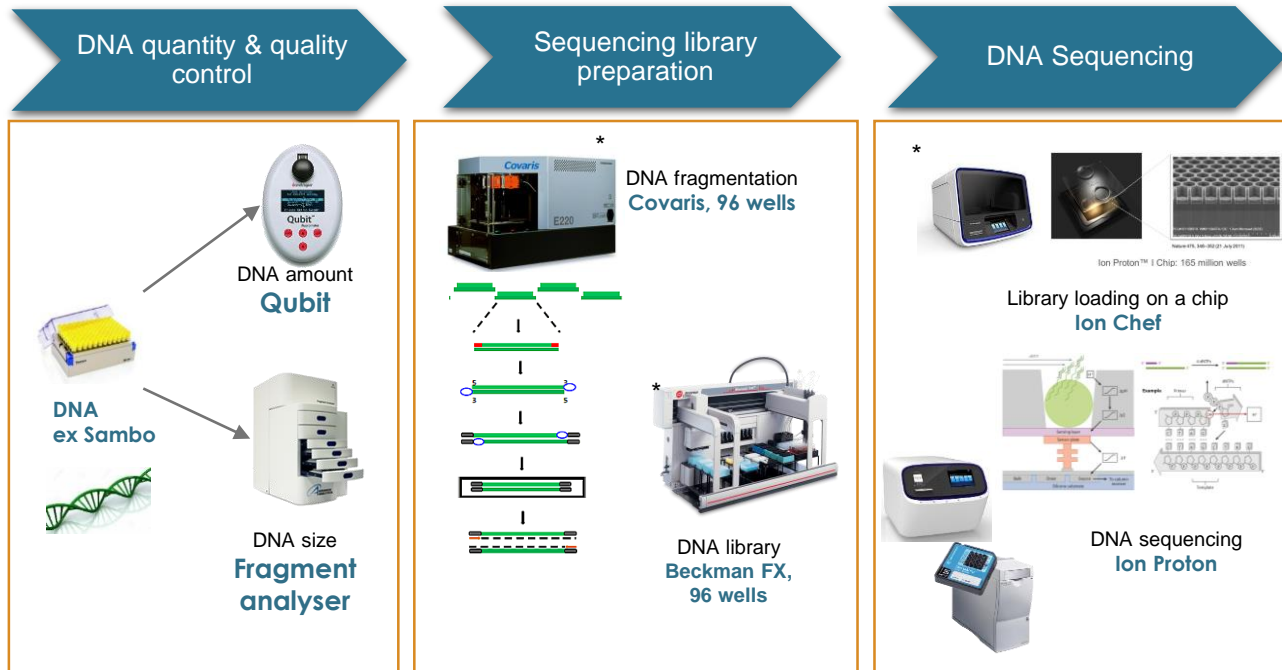
Kit de collecte pour stocker et préserver les échantillons de selle.



- Capacité de stockage à MGP : 600 000 échantillons
- Tube barcodés :
 - 3 aliquots par échantillon :
 - 2 conservés dans la biobanque
 - 1 envoyé à la plateforme de séquençage



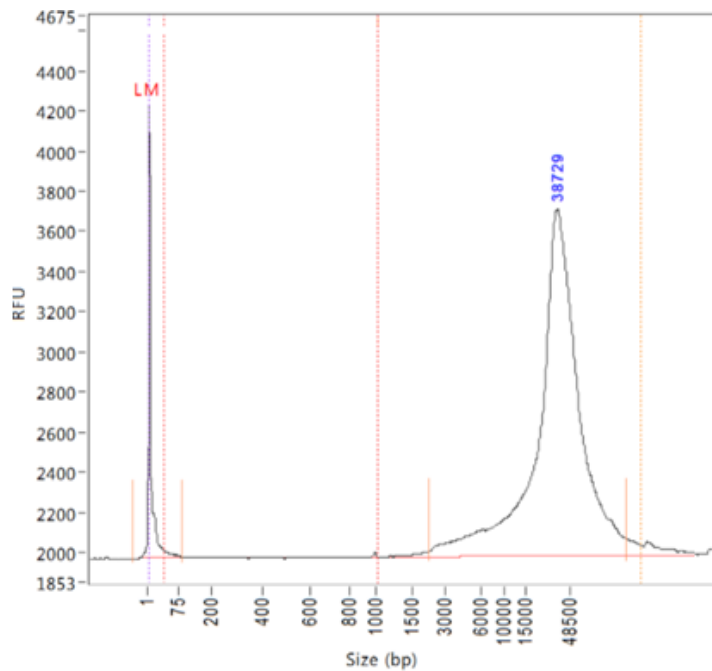
Kit de collecte pour stocker et préserver les échantillons de selle.



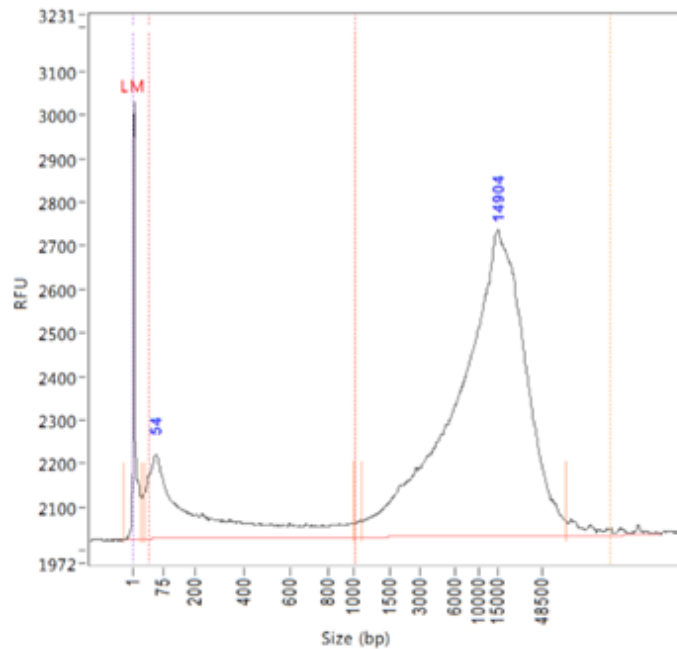
* automatisé

- 2019: Semi automatisation et la standardisation des process permet d'effectuer 2 500 métagénomés / an.
- 2023: Automatisation complète permet de traiter 10 000 métagénomés / an

Contrôle qualité de l'échantillon : Evaluation de la fragmentation de l'ADN



Qualité haute



Qualité moyenne

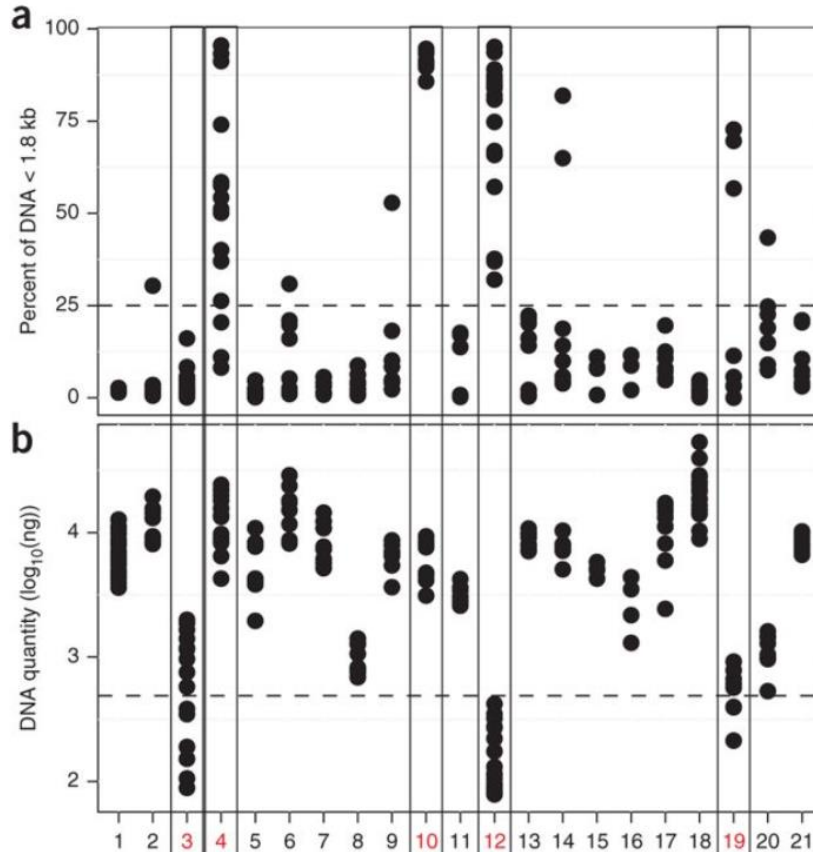


Contrôle qualité de l'échantillon : Evaluation de la fragmentation de l'ADN



- 21 instituts de recherche de 13 pays différent sur 3 continents reçoivent les mêmes échantillons (4 aliquots * 2 individus différents)

IHMS : Evaluer les protocoles pour maximiser la concentration et la taille de l'ADN.



- Les protocoles en rouges n'arrivent pas à optimiser la concentration en ADN et/ou la longueur des fragments extraits.

Le séquençage métagénomique: Un outil performant pour explorer le lien entre microbiote et sa fonction vis-à-vis d'une pathologie

[Burz *et al.*, Scientific Reports, 2019]

[Pedersen *et al.*, Nature Protocole, 2018]

[Qin *et al.*, Nature, 2010]

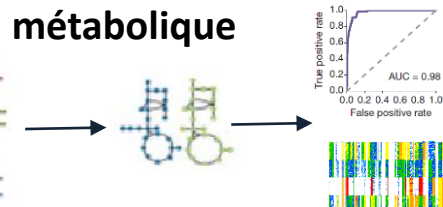
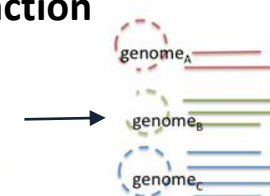
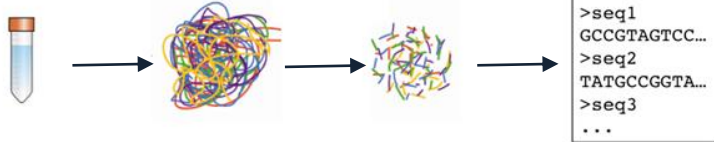
[Bidkhorji *et al.*, bioRxiv, 2021]

Collecte et stockage échantillon

Séquençage métagénomique

Annotation Gene & Fonction

Modélisation métabolique



Extraction & Fragmentation ADN

Filtre QC & Assemblage Fragment ADN

Pan-genome reconstruction & Quantification

Analyses Statistiques

[Costea *et al.*, Nature Biotechnology, 2017]

[Nielsen, Almeida *et al.*, Nature Biotechnology, 2014]

[Plaza Oñate *et al.*, Bioinformatics, 2019]

[Le Chatelier *et al.*, Nature, 2013]

Quelle technologie de séquençage utiliser ?



**Illumina
HiSeq**



Ion S5



**DNBSeq-
G400**

**2nd generation
sequencing
(short reads)**



**MinION
R10**



**PacBio
Sequel II**

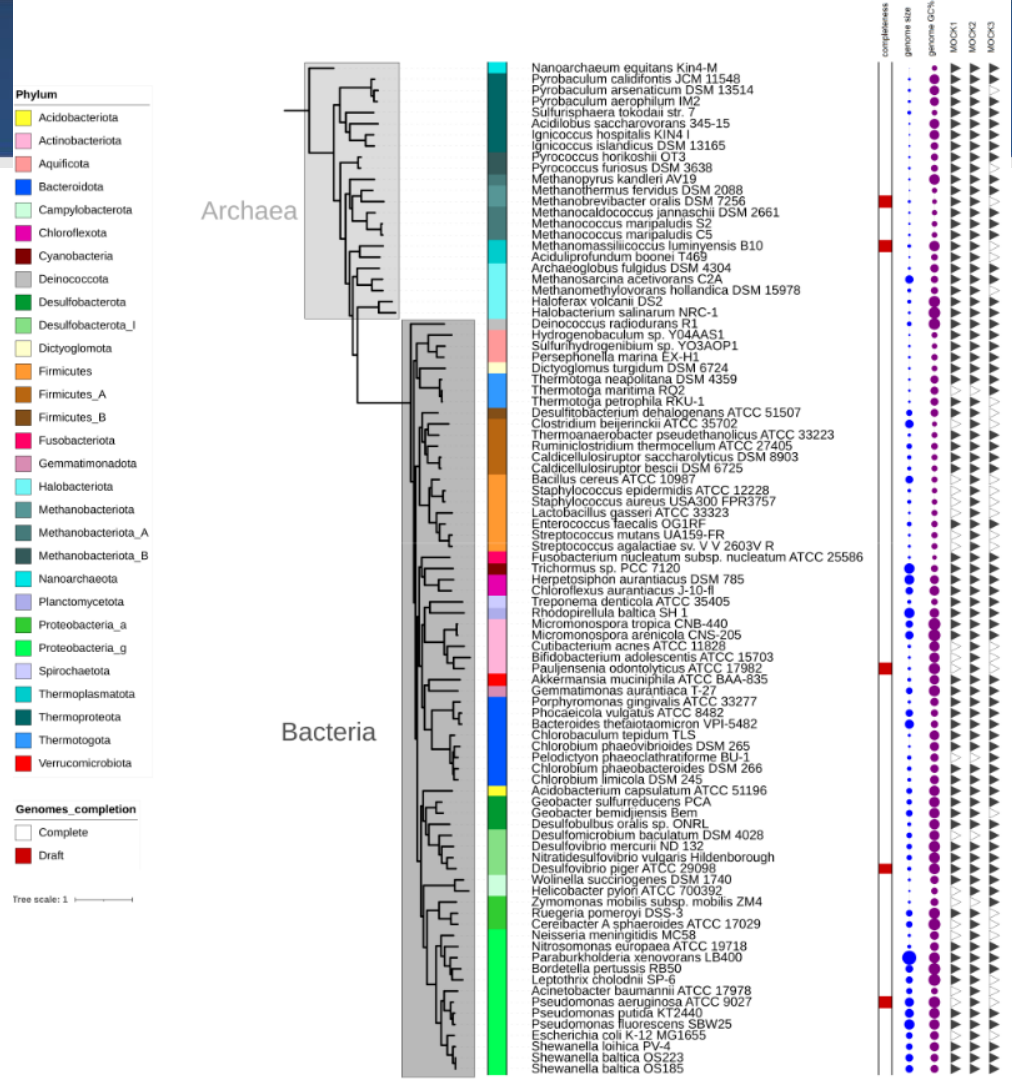
**3rd generation
sequencing
(long reads)**

Créer des communautés synthétiques pour évaluer la performance des séquenceurs (mock)



Dr. Mircea Podar

- Synthetic mock with known species and genome associated
- From 30% GC enriched to 70% GC enriched organisms.



Créer des communautés synthétiques pour évaluer la performance des séquenceurs (mock)



Dr. Mircea Podar

- Synthetic mock with known species and genome associated
- From 30% GC enriched to 70% GC enriched organisms.

[Meslier *et al.*, Scientific Data, 2022]

	<i>Illumina HiSeq 3000</i>	<i>Ion Proton P1</i>	<i>Ion S5</i>	<i>DNBSEQ- G400</i>	<i>DNBSEQ- T7</i>	<i>ONT MinION R9</i>	<i>PacBio Sequel II</i>
<i>Average reads length ± stdv after trimming (bp)</i>	149 ± 4.24	144.041 ± 28.43	145.76 ± 28.12	99.91 ± 0.96	99.52 ± 2.58	4408.41 ± 2831.95	10289.7 ± 4036.27
<i>Max read length after trimming (bp)</i>	150	373	347	100	100	60869	40278
<i>SE / PE</i>	PE	SE	SE	PE	PE	SE	SE
<i>Average insert size ± stdv (bp)</i>	433.47 ± 92.37	-	-	245.13 ± 51.04	235.56 ± 54.80	-	-
<i>Total Run Time</i>	4d	4h	4h	3d	3d	48h*	30h

Créer des communautés synthétiques pour évaluer la performance des séquenceurs (mock)

<i>Sample ID</i>	<i>Sequencing Technology</i>	<i>N. reads after trimming (M)</i>	<i>%Mapped end-to-end</i>	<i>%Avg end-to-end best mapped identity</i>	<i>%Avg end-to-end best mapped SNPs</i>	<i>%Avg end-to-end best mapped indels</i>
MOCK (<i>N=71 species</i>)	Illumina HiSeq 3000	20.59*2	99.62	99.45	0.46	0.09
	Ion Proton P1	20.00	99.29	99.42	0.12	0.46
	Ion S5	28.51	99.35	99.61	0.08	0.31
	ONT Minion R9	0.696	99.75	89.08	3.37	7.55
	PacBio Sequel II	0.524	99.65	99.72	0.06	0.22
	DNBSEQ-G400	35.42*2	99.22	99.70	0.30	0.003
	DNBSEQ-T7	375.12*2	98.92	99.42	0.58	0.003

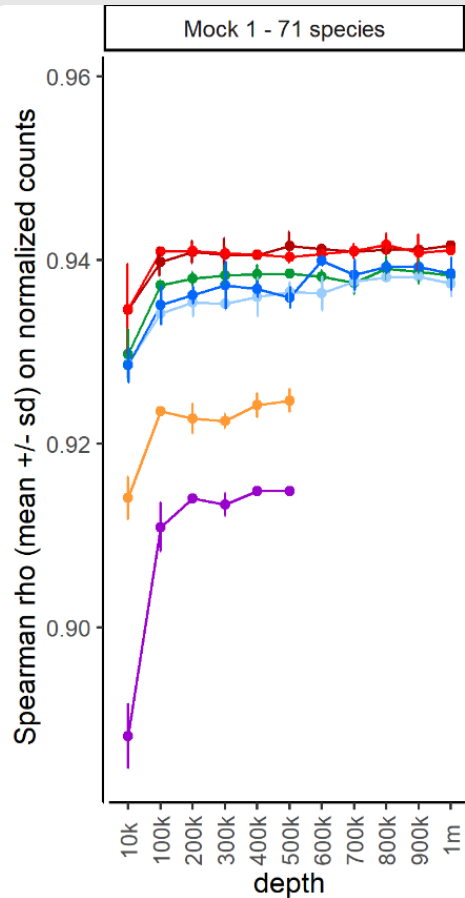
Technology choice depends on the biological question !

[Meslier *et al.*, Scientific Data, 2022]

Are the functions included in mobile elements ? -> long reads.

What is the microbial richness and function? -> short reads.

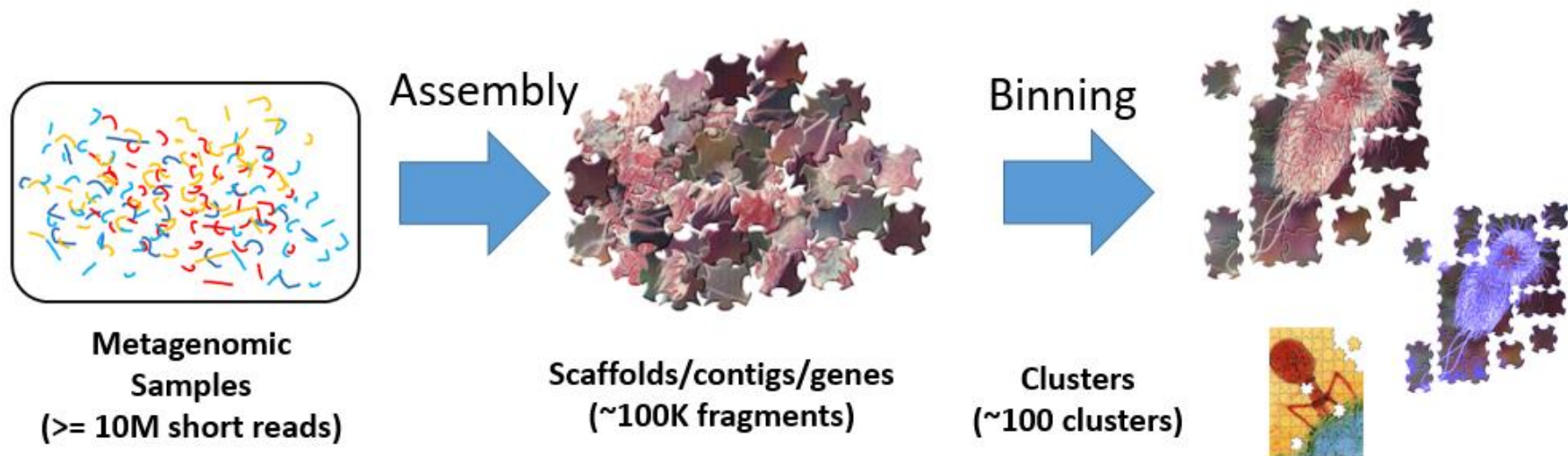
Créer des communautés synthétiques pour évaluer la performance des séquenceurs (mock)



- Some technologies introduce biases in the microbial composition observation (DNA filtering, high error rate...)
- Depending on the scientific question, need to adapt the methodology for DNA preparation and chose the sequencer accordingly

[Meslier *et al.*, Scientific Data, 2022]

La métagénomique: des milliers de puzzle mélangés !



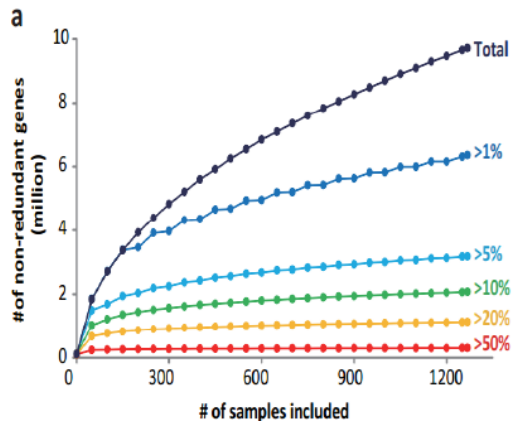
Il est nécessaire de reconstruire les puzzles pour explorer le microbiote

Exploration guidée par utilisation de génomes de référence



741 species

Qin et al Nature, 2010
Nielsen*, Almeida* et al.,
Nature Biotech, 2014



1661 metagenomic species

Li et al Nature, 2014
Plaza Oñate, Bioinformatics, 2018

MGNify
Submit, analyse, discover and compare microbiome data

Overview | Submit data | Text search | Sequence search | Browse data | Genomes

Getting started

Search by

Name, biome, or keyword Sequence similarity

Or by data type

173631 amplicon	3526 studies
16247 assemblies	178459 samples
2050 metacoding	235044 analyses
27479 metagenomes	
2075 metatranscriptomes	

Or by selected biomes

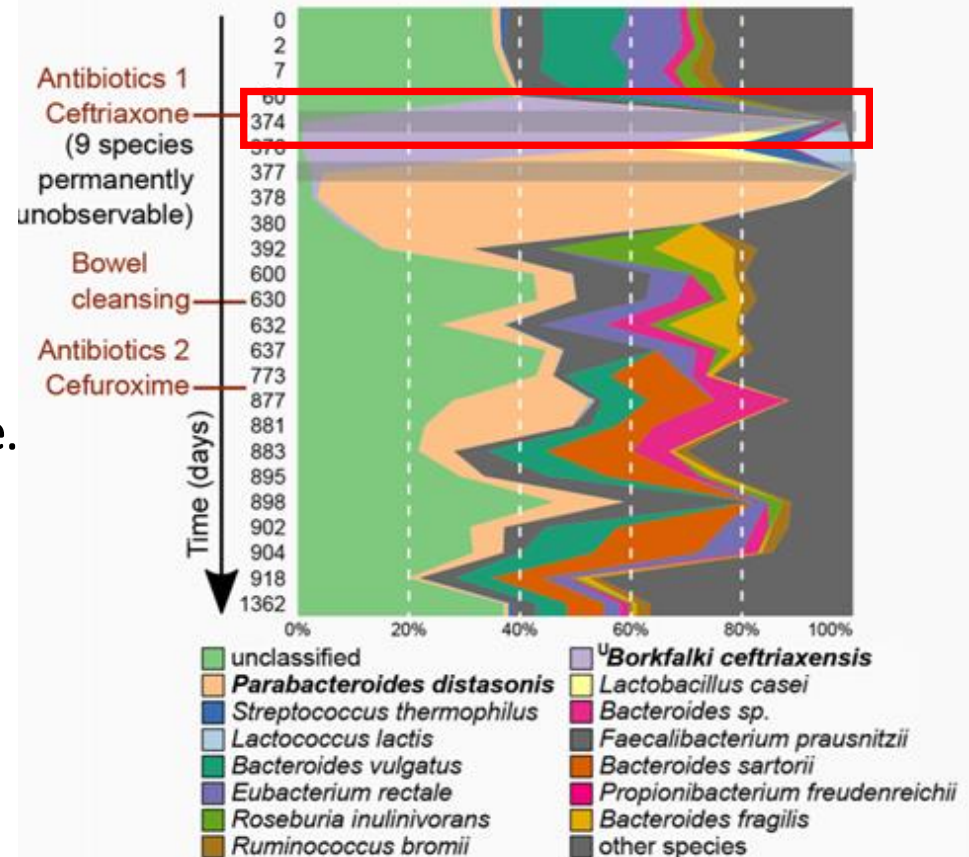
Human (63360)	Digestive system	Aquatic (40229)	Marine (30268)	Soil (13518)
---------------	------------------	-----------------	----------------	--------------

>4000 species

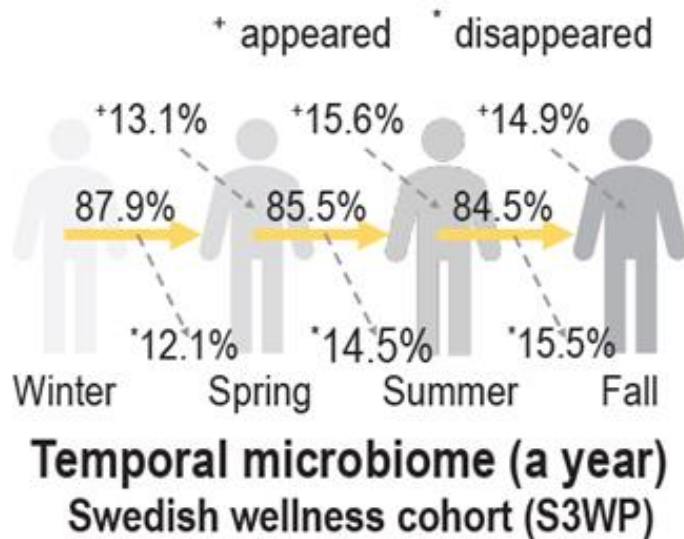
Mitchell et al, Nucleic Acids Research, 2017,
Alexandre Almeida et al, Nature, 2019

Meilleure façon de reconstruire un nouveau génome: découvrir un échantillon avec une forte abondance d'un nouveau génome!

- After ceftriaxone treatment, new organism increased up to 90% of the total microbiota in one sample.
- *Borkfalki ceftriaxensis* prevalent in 1/3 of public gut samples in low abundance.



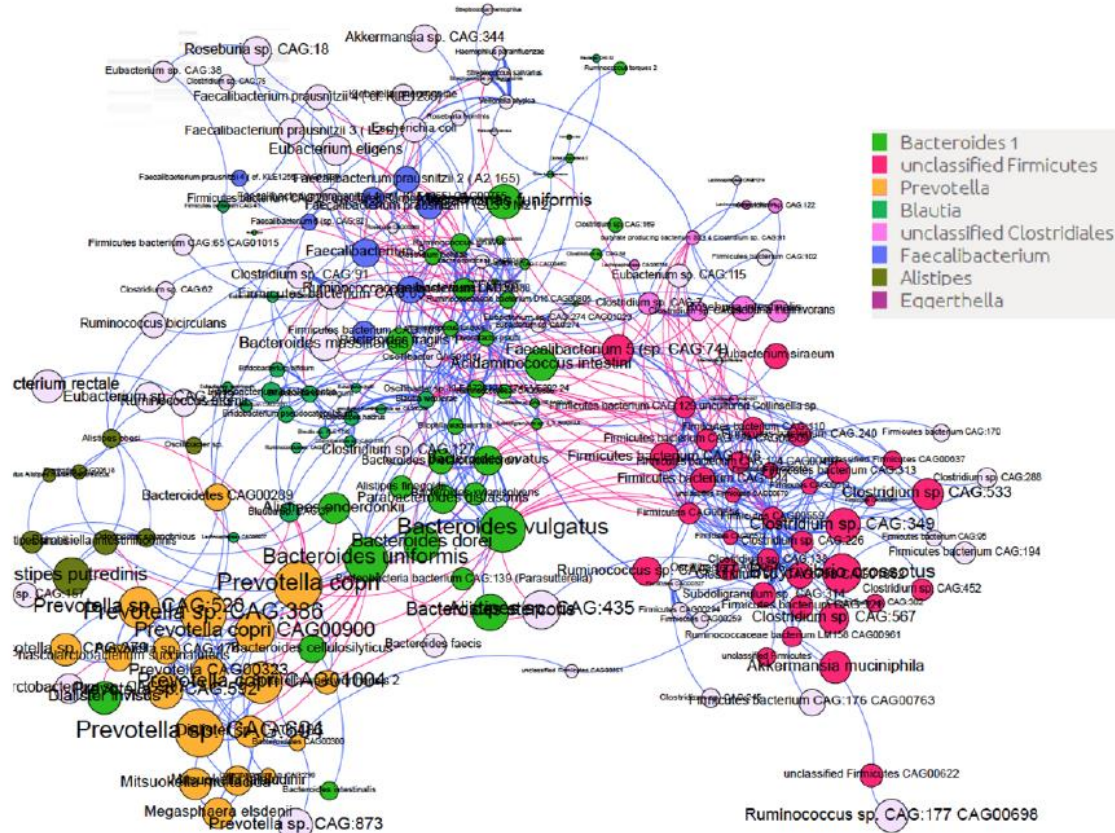
[Hildebrand et al, Gut, 2019]



[Shoaie *et al.*, <https://www.researchsquare.com/article/rs-339282/v1>]

- Wellness cohort: 86 healthy Swedes followed during a year, with faecal samples (344 samples in total) and clinical data collected every 3 months.
- During the study, >80% of the gut microbiome detected independently of the season for these healthy individuals.
- Some species tend to persist when detected, or tend to vanish.

Des espèces aux communautés microbiennes



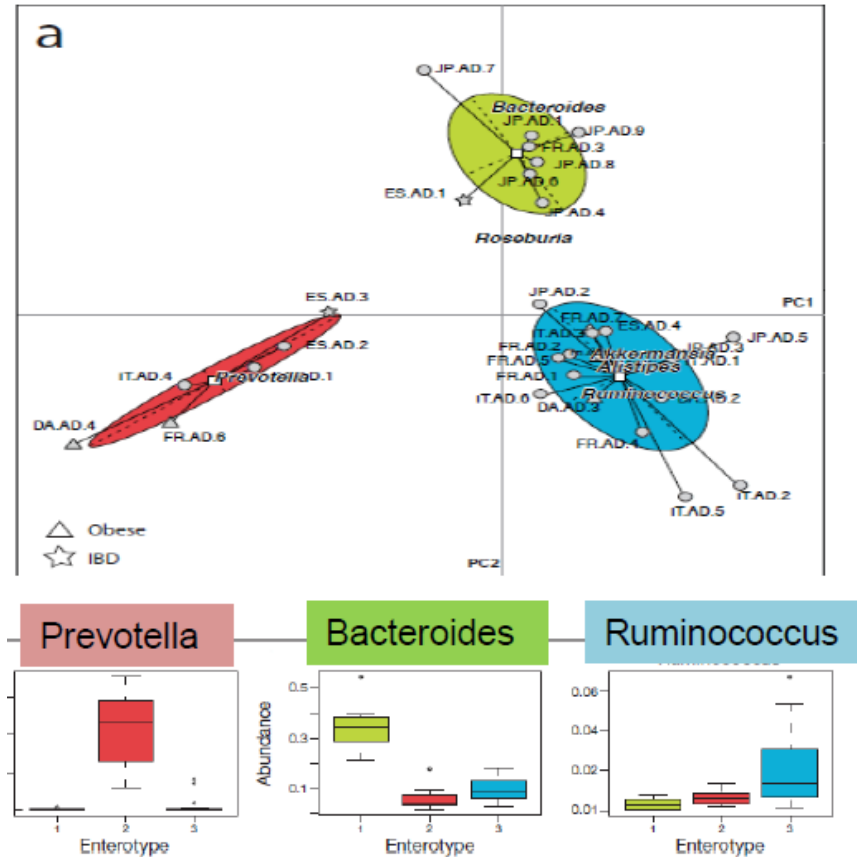
Europeans, Americans,
Asians. n=33

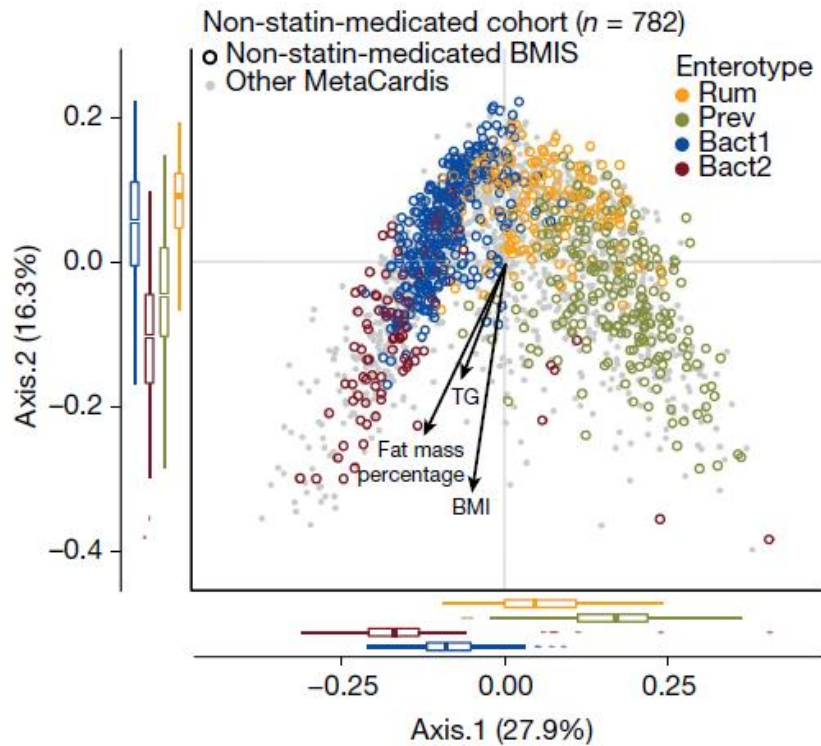
Identification of 3 clusters
(enterotypes)

that are not nation or
continent specific

debates about number,
robustness, resilience...

Arumugam, Raes et al, Nature 2011



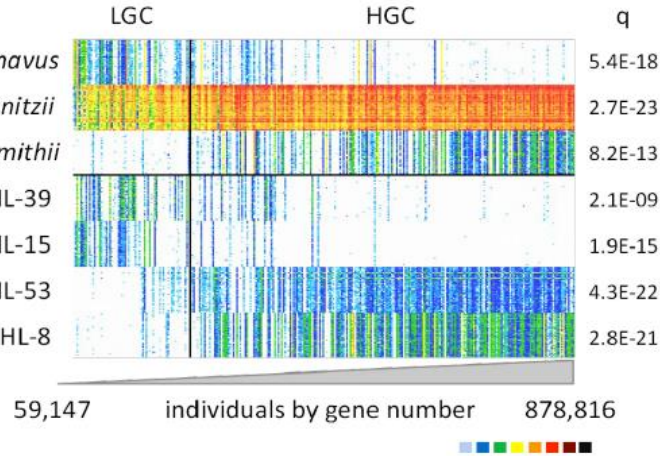
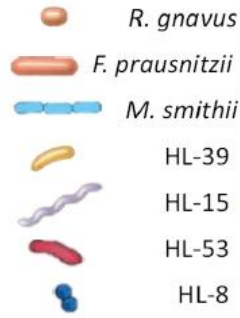
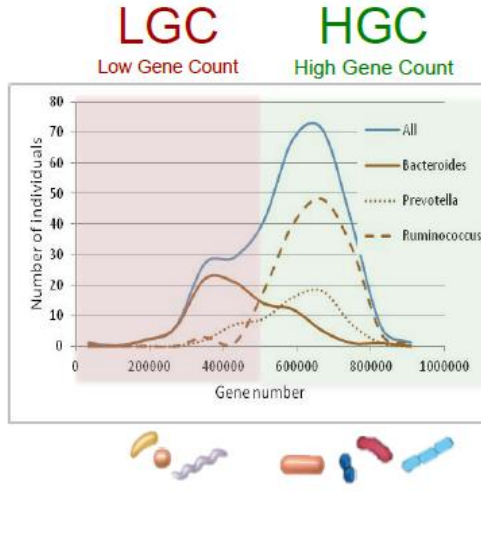


[Vieira silva *et al.*, Nature, 2020]

- n = 888 Européens (Français, Allemand, Danois)
- Objectif: explorer le microbiote intestinal chez des individus obèses
- Détection d'un nouvel enterotype appelé B2, dominé par le genre Bacteroides, faible détection Faecalibacterium.
- B2 corrélé aux maladies chroniques (Crohn) et 78% des patients présentant une maladie inflammatoire.
- Détectés dans ~4% des sujets sains sans surpoids : individus à risque de développer des maladies ?

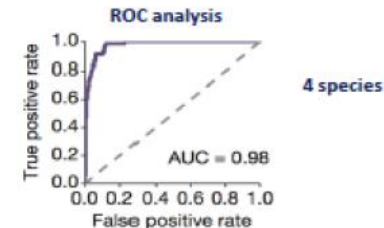
Quel est le lien entre la composition microbienne intestinale, la santé et le mode de vie ?

Stratifier les individus par leur richesse microbienne pour prédire le risque de développer des maladies

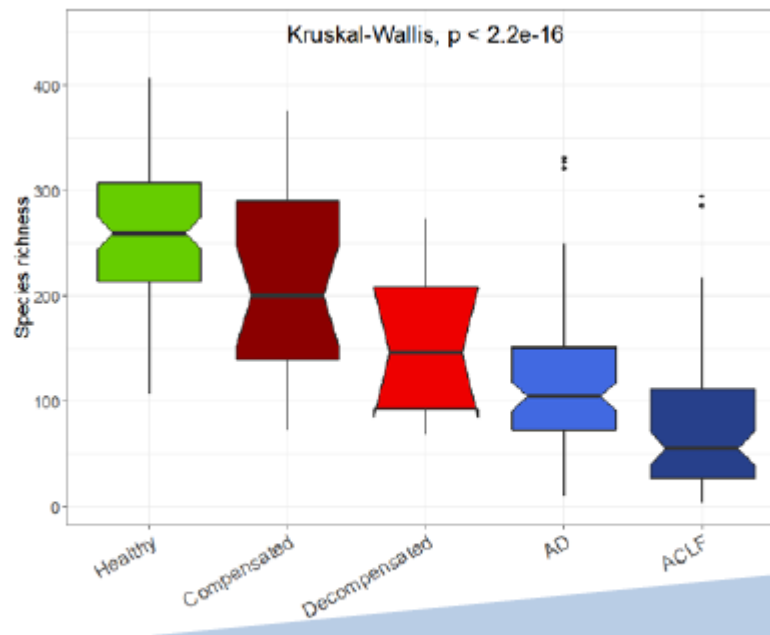
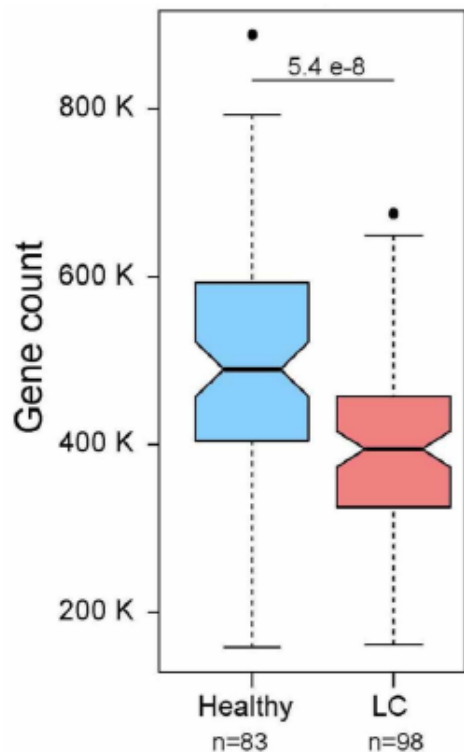


Diversity (Le Chatelier*, Nielsen*, Qin*, Prifti* et al, Nature 2013)

- diversity = number of present genes in a given subject
- identification of **two robust groups of patients** : low diversity and high diversity.
- linked with a number of phenotypic parameters (**inflammation, insulin resistance**). “High as Healthy”



Utiliser le microbiote intestinal pour le diagnostic de la cirrhose du foie

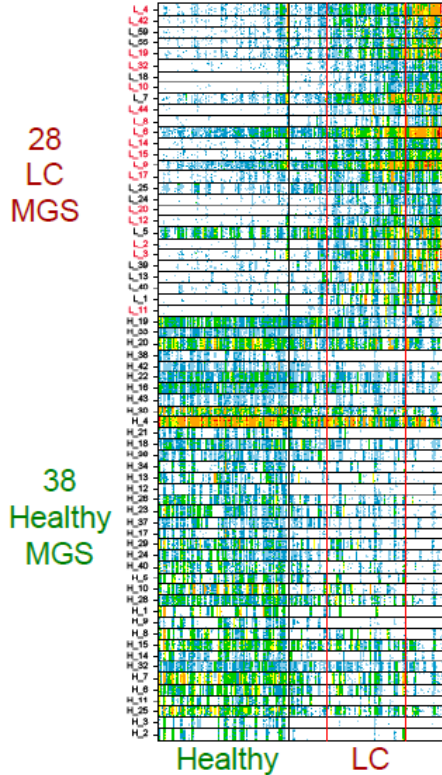


liver cirrhosis gravity scores

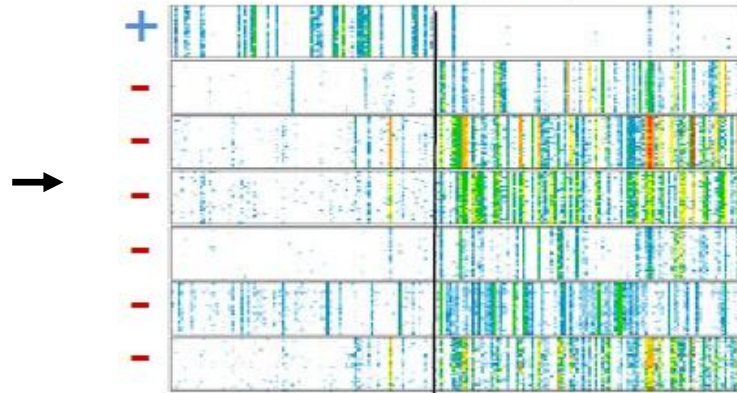
Qin N. et al. Nature 2014

Utiliser le microbiote intestinal pour le diagnostic de la cirrhose du foie

66 species

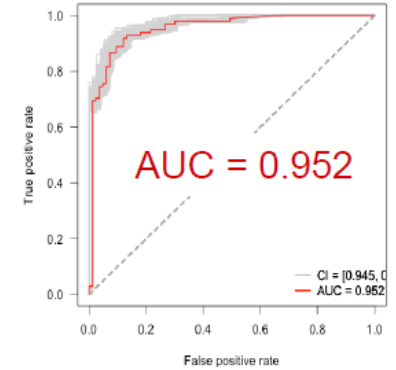


7 model species

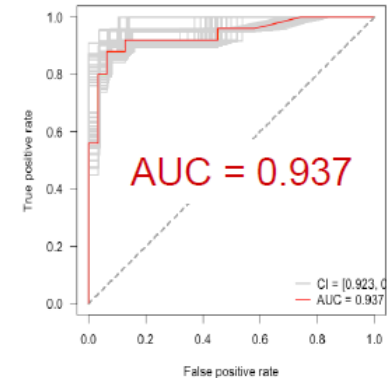


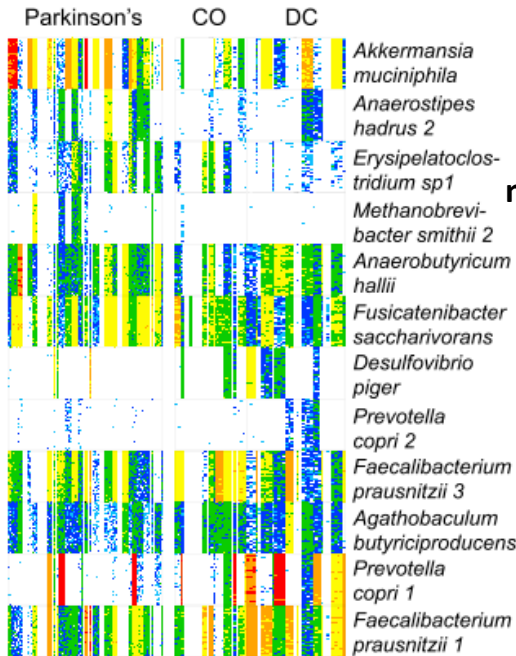
Accurate diagnostics irrespective of disease origin: alcohol, hepatitis viruses
disease status: compensated & de-compensated (ascites, encephalopathy...),
medication: antivirals, beta blockers, proton pump inhibitors or none

discovery (83+98)



validation (31+25)





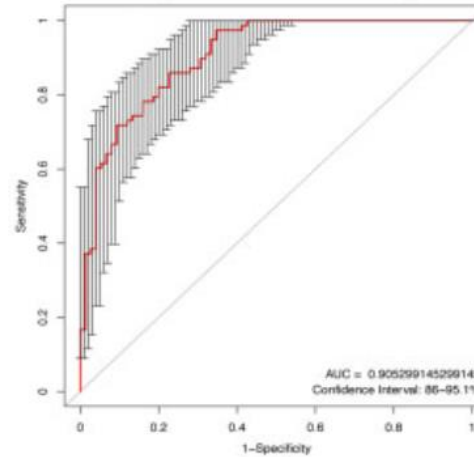
[Rosario *et al*, Cell Reports, 2021]
31 PD vs 28 CTR
matchés sur l'âge.

Akkermansia muciniphila
Anaerostipes hadrus 2
Erysipelatoclostridium sp1
Methanobrevibacter smithii 2
Anaerobutyricum hallii
Fusicatenibacter saccharivorans
Desulfovibrio piger
Prevotella copri 2
Faecalibacterium prausnitzii 3
Agathobaculum butyriciproducens
Prevotella copri 1
Faecalibacterium prausnitzii 1

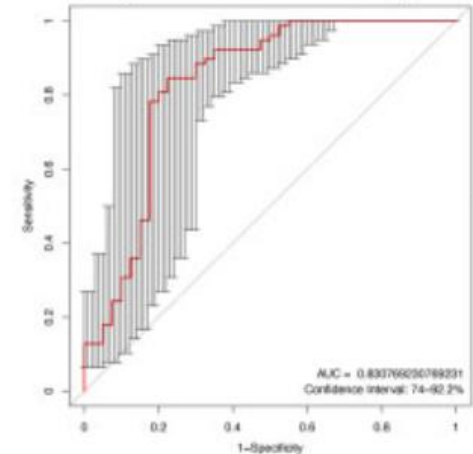


[Qian *et al*, Brain, 2020]

78 PD vs 75 Healthy
(Real-time PCR)



78 PD vs 40 MSA
(Real-time PCR)



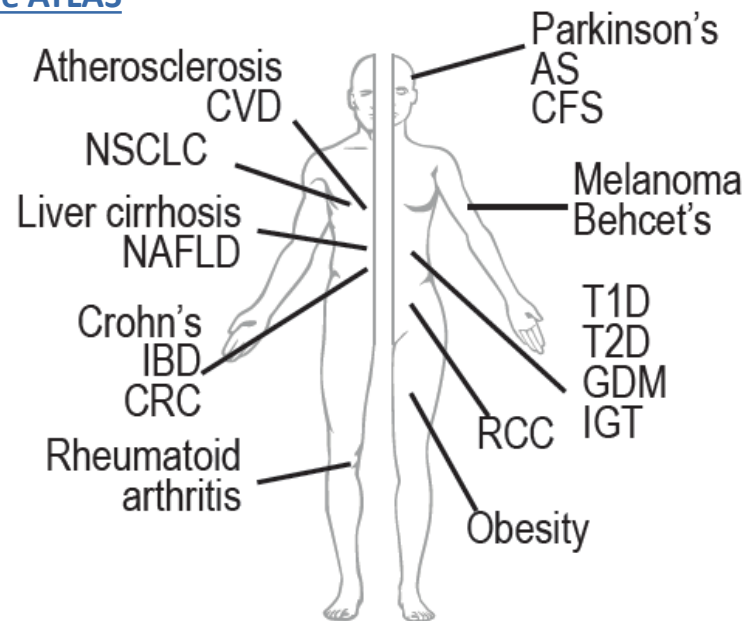
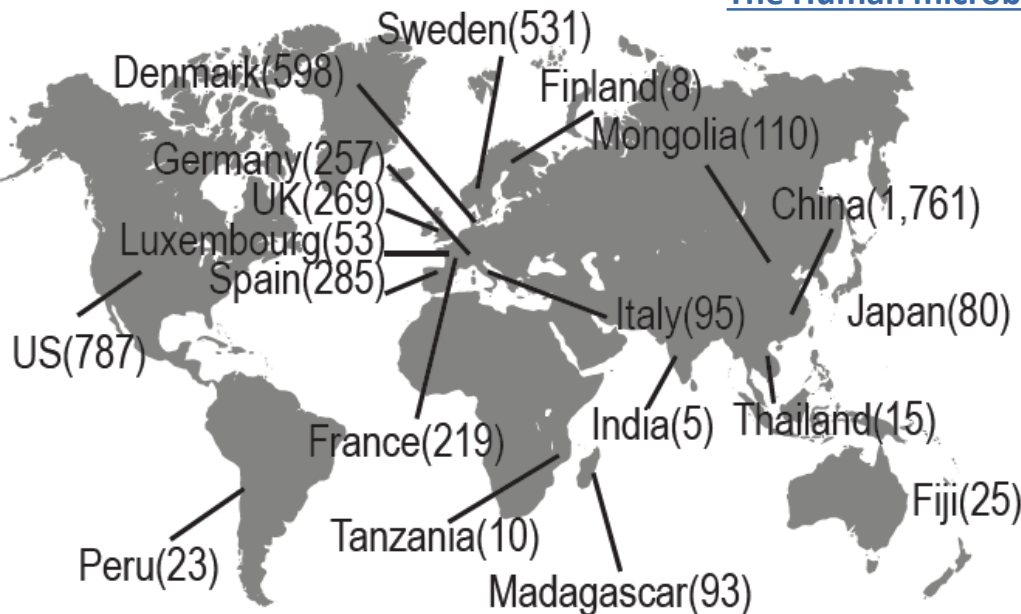
Prédiction PD vs Healthy ou MSA par utilisation de biomarqueur en PCR sur cohorte indépendante

Parkinson: individus avec une forte richesse microbienne (constipation?)

Mais des signatures microbiennes fortes détectées (A. muciniphila fort, Faecalibacterium faible)

Dans le monde: de nombreuses études métagénomiques pour explorer le lien entre microbiote fécal, sa fonction et la santé

The Human microbiome ATLAS



De nombreux projets mais un faible nombre d'individus analysés...

[Shoaie, Lee, Almeida *et al.*,
<https://www.researchsquare.com/article/rs-339282/v1>, in preparation]

Pour répondre aux problèmes de faible effectif: le projet French Gut



Le microbiote français
Le French Gut

Contribution citoyenne nationale visant à analyser 100 000 échantillons fécaux de français majeurs ainsi que des données cliniques et nutritionnelles jusqu'à 2025 pour:

- Définir l'hétérogénéité du microbiote intestinal français
- Identifier les facteurs environnementaux et le mode de vie qui influencent la composition microbienne intestinale, dans la santé comme dans la maladie.

French Gut Consortium

ASSISTANCE
PUBLIQUE  HÔPITAUX
DE PARIS

INRAE



Inserm

La science pour la santé
From science to health

**INSTITUT
PASTEUR**



**Institutions
privées**

Données et métadonnées

Le French Gut

✓ DONNÉES DE SÉQUENÇAGE

Séquences brutes du microbiote intestinal, obtenues par séquençage métagénomique shotgun des échantillons fécaux.
> 20 M de sequences générées par échantillon.

✓ DONNÉES DE SANTÉ ET NUTRITIONNELLE

- 1 Set minimal de métadonnées “MMHP” (ID, date, site d'échantillonnage, âge, sexe, IMC, pays, santé ou maladie).
- 1 Set de ~60 métadonnées “Le French gut” tirées de questionnaires, permettant une analyse croisée des métadonnées et des données de la métagénomique

Le questionnaire French Gut



UN REMPLISSAGE RAPIDE ET SECURISÉ

~15 minutes sur un système en ligne certifié à l'hébergement des données personnelles.



VOLET ELIGIBILITÉ ET INSCRIPTION

Age, prise d'antibiotique, coloscopie, intervention chirurgicale, ...



VOLET DÉMOGRAPHIQUE

Sexe, lieu de résidence, tabagisme, sédentarité, activité physique...



VOLET SANTÉ

IMC, affections de longue durée (ALD 30), symptômes gastro-intestinaux, bien-être...



VOLET ALIMENTATION

Score fréquentiel alimentaire DASH, portions en légumes, fruits, aliments et boissons fermentés...



Recruitment via public media, French cohorts and patient networks

Online
registration

1

Through the secured
French Gut website

Eligibility &
Questionnaire

2

Including clinical and
nutritional forms

3

Sampling kit
at home

Using MetaGenoPolis
custom design collect kit

5

Collective
feedback

At each recruitment
milestones

4

Sequencing
and Analysis

Following IHMS high
quality standards



Le microbiote français :
Le French Gut

Pilot phase in 2022



<https://lefrenchgut.fr>

Le French Gut inclus dans le consortium Million Microbiome of Humans Project (MMHP)



- But: collecter et analyser le microbiote (fécal, oral, vaginal, ...) d'1 million d'individus à l'horizon 2026.
- MGP est un des 5 membres fondateur du MMHP (via S. Dusko Ehrlich)
- MGP va contribuer au MMHP via le French Gut.

<https://db.cngb.org/mmhp/>

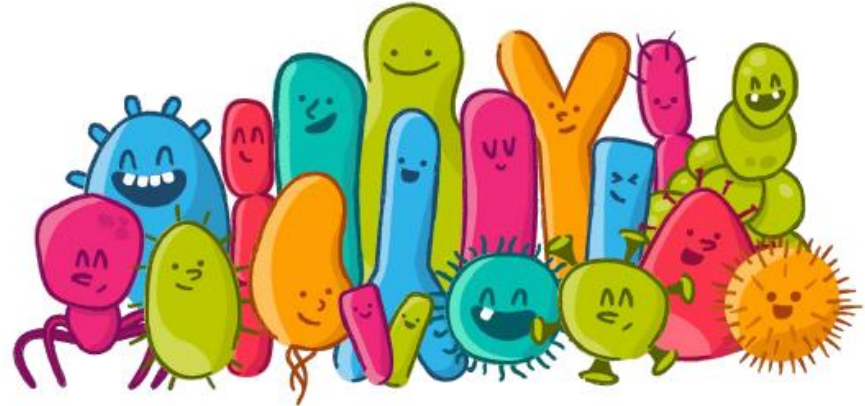


Avec les partenaires (ouvert pour collaboration) :
Allemagne, Italie, Pays-Bas, Espagne...

Diversity is healthy !

Take care of your microbiota

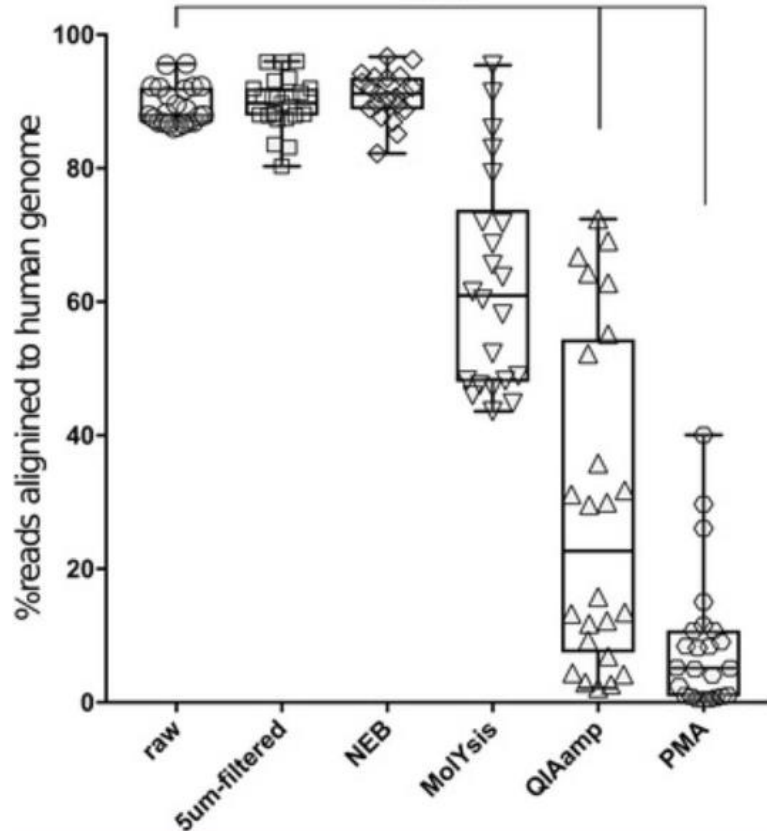
Thanks for your attention



Pour en savoir plus contacter :
mathieu.almeida@inrae.fr

Et visiter notre site web:
<http://mgps.eu/>

Cependant, le microbiote peut être difficile à explorer sans amplification



- Marotz et al new PMA technic sample preparation:
- 1/ Osmotic lysis (mammalian cell disruption)
- 2/ Free DNA degradation using propidium monoazide PMA (DNA intercalator)
- 3/ Normal DNA extraction procedure
- Allow to reduce the amount of human DNA reads in saliva sample

[Marot et al, Microbiome, 2018]