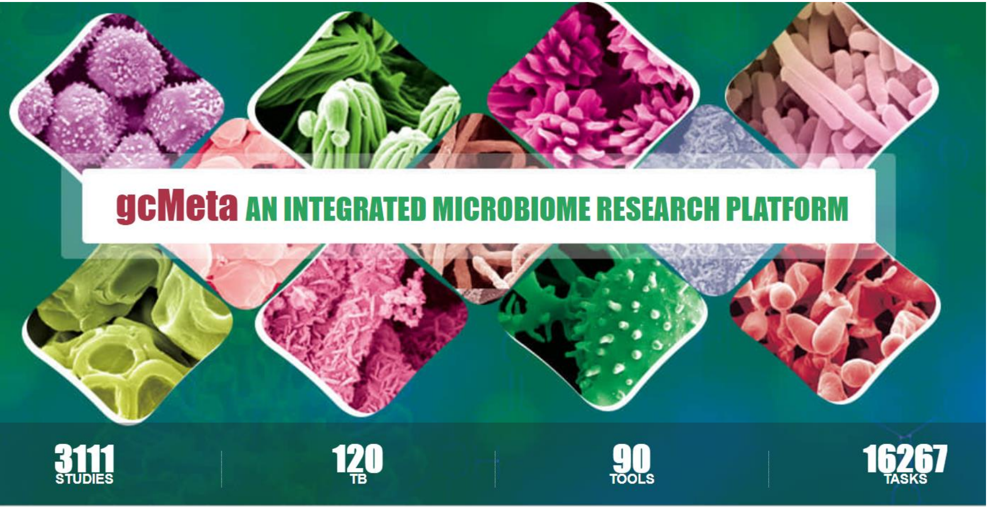
metagenopolis
mgps.eu

Une caractérisation standardisée à haute résolution du microbiote intestinal pour une meilleure compréhension du lien avec notre santé.

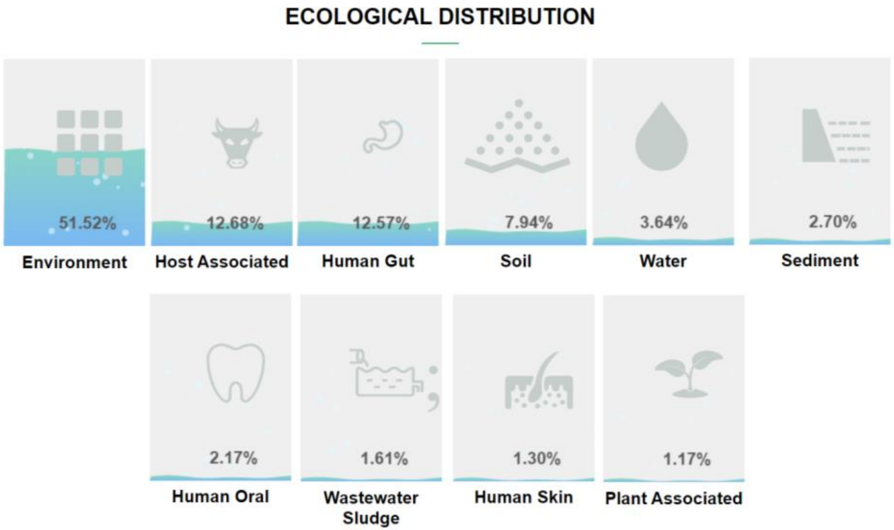
Mathieu Almeida
Chargé de Recherche bioinformaticien

INRAE

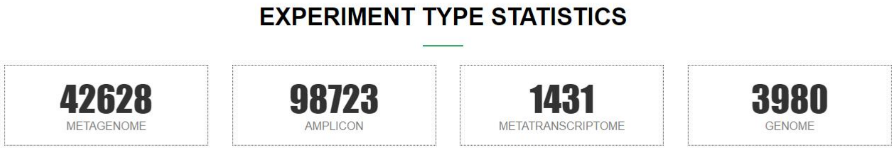
The microbiota exploration enter a new era of Big Data


gcMeta AN INTEGRATED MICROBIOME RESEARCH PLATFORM

3111	120	90	16267
STUDIES	TB	TOOLS	TASKS


ECOLOGICAL DISTRIBUTION

Category	Share
Environment	51.52%
Host Associated	12.68%
Human Gut	12.57%
Soil	7.94%
Water	3.64%
Sediment	2.70%
Human Oral	2.17%
Wastewater Sludge	1.61%
Human Skin	1.30%
Plant Associated	1.17%

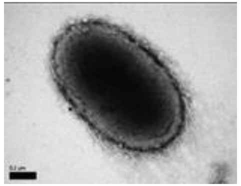

EXPERIMENT TYPE STATISTICS

42628	98723	1431	3980
METAGENOME	AMPLICON	METATRANSCRIPTOME	GENOME

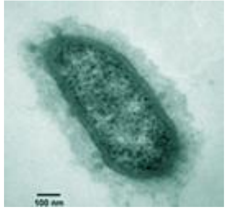
[Shi *et al.*, NAR, 2018]

- Many cohorts and samples coming from various environment.
- Different sequencing strategies, bioinformatic analysis tools…

From targeted exploration to global metagenomic exploration of the human microbiota



Bacteria



Archaea

16S rRNA

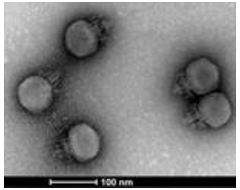
Prokaryota

Protozoa

Fungi

18S rRNA / ITS

Eukaryota

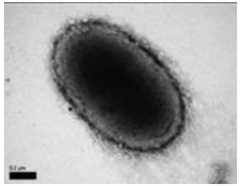


Virus

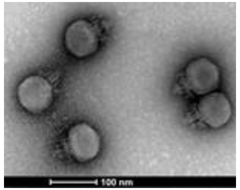
Filtration, centrifugation…

Virus-like particles

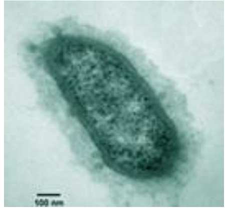
From targeted exploration to global metagenomic exploration of the human microbiota



Bacteria



Virus



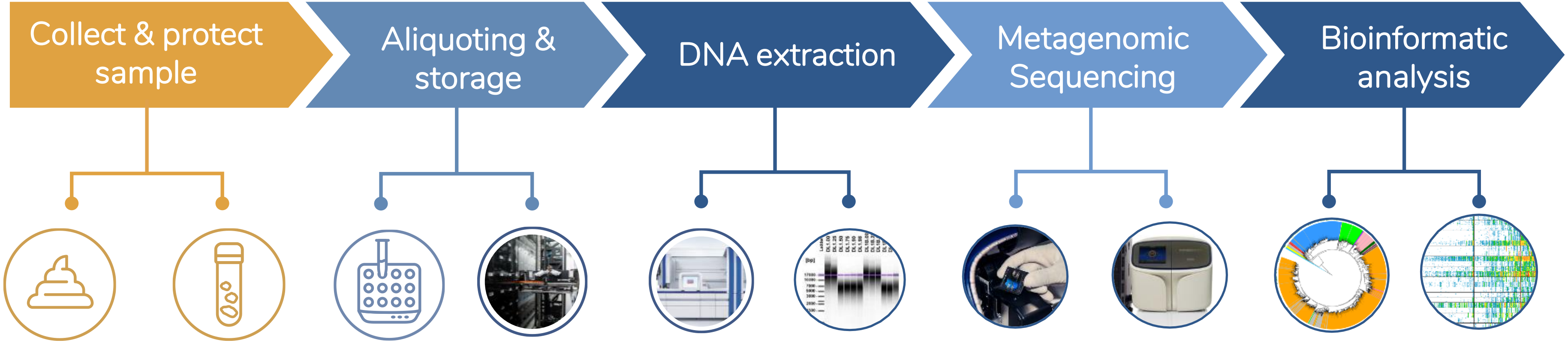
Archaea

Protozoa

Fungi

Shotgun metagenomics: Random sequencing of all accessible DNA

Need to standardize key metagenomic analysis steps for optimal microbial exploration



Collect & protect sample

Aliquoting & storage

DNA extraction

Metagenomic Sequencing

Bioinformatic analysis

[Costea *et al.*, Nat Biotech, 2014]

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

[Santiago *et al.*, BMC Microbiology, 2014]

[Meslier *et al.*, Gut, 2020]

[Nielsen *et al.*, Nat Biotech, 2014]

MetaGenoPolis leading the International Human Microbiome Standards consortium for benchmarked and best practices in human microbiota exploration

Which sequencing technology for microbial exploration ?

Illumina

Ion Torrent

DNBSeq

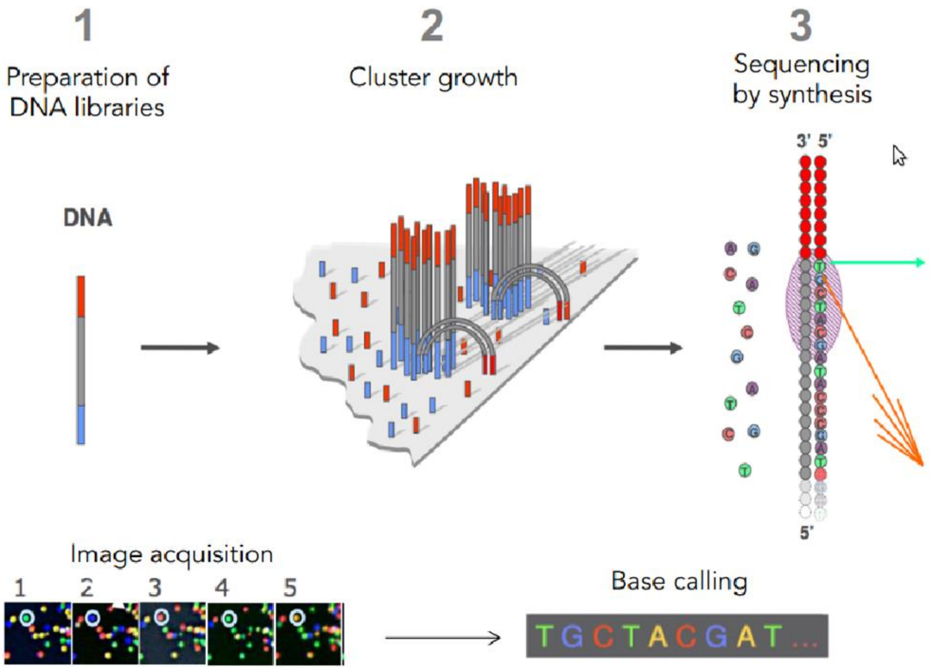
2nd generation sequencing (short reads)

MinION

PacBio

3rd generation sequencing (long reads)

Illumina HiSeq : the most popular

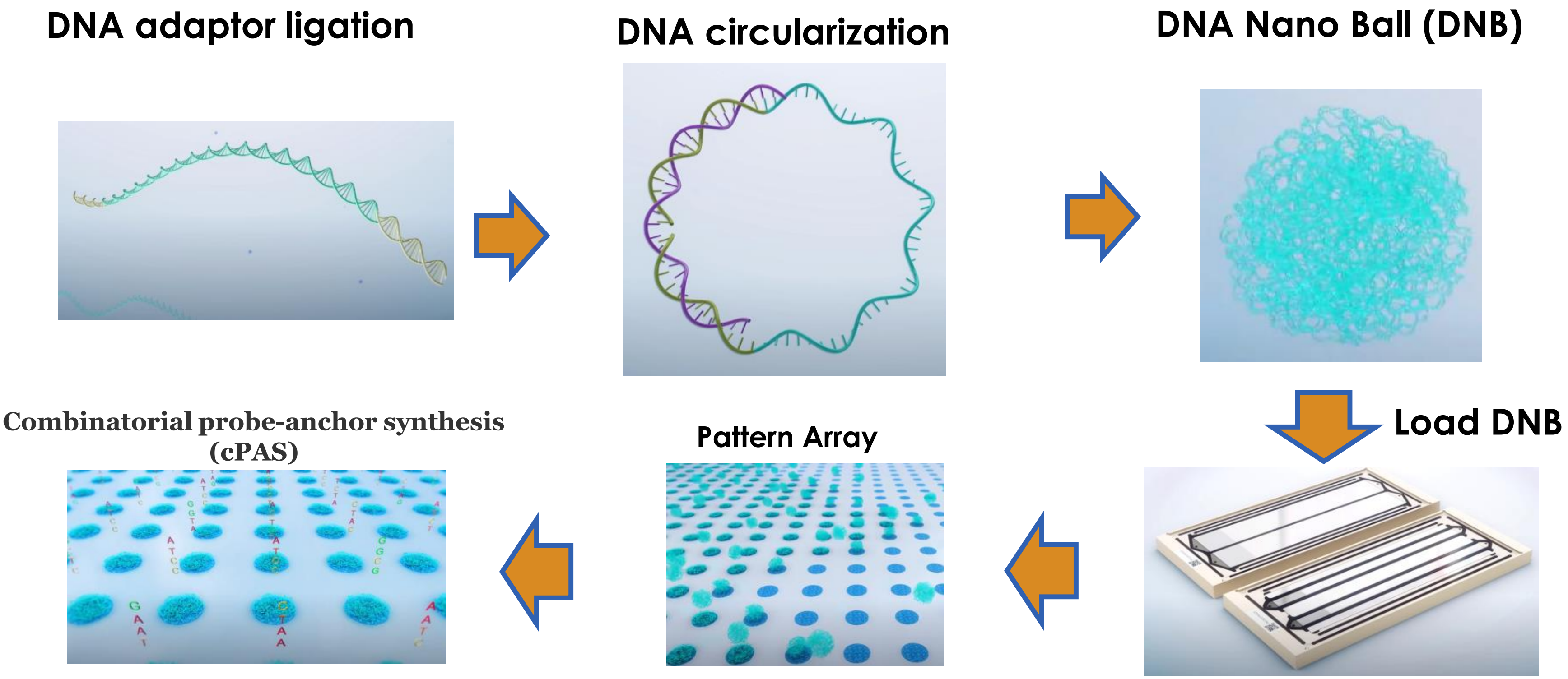


PE reads of 150 bp, up to 6 Tb of data per run (~20 billions reads). DNA amplification by clustering on lane

Pro: very accurate, cheap (price per base), very high throughput

Con: slow (~3 days for a run), short reads.

DNBSeq: the DNA Nano ball avalanche



DNA adaptor ligation

DNA circularization

DNA Nano Ball (DNB)

Load DNB

Pattern Array

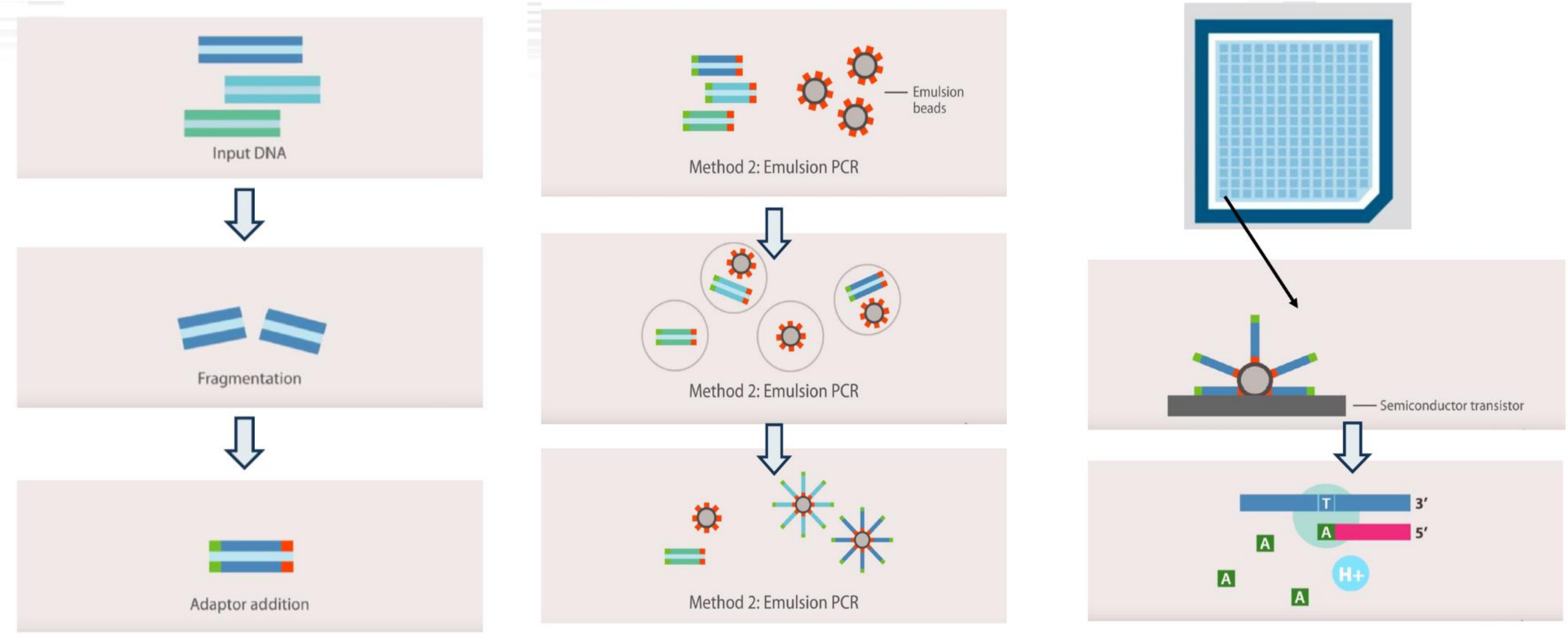
Combinatorial probe-anchor synthesis (cPAS)

PE Reads of ~150 bp but up to 6 Tb of data per run (~10 billions reads).

Pro: accurate, cheap (price per base), very high throughput.

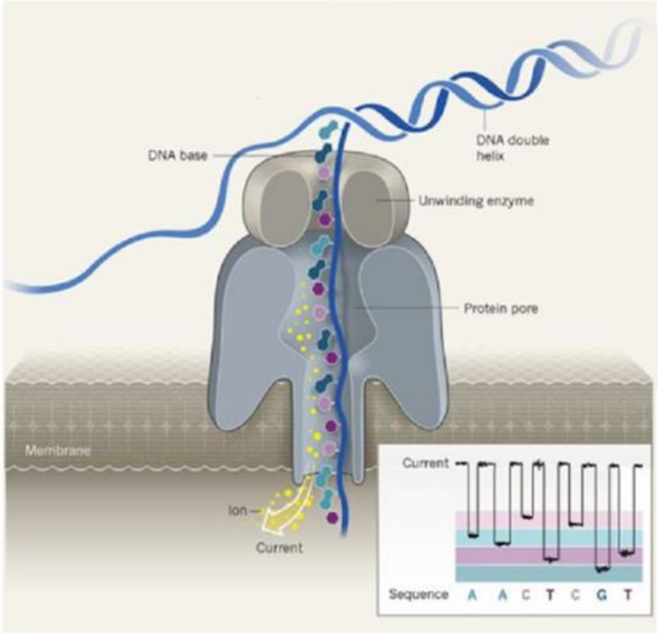
Con: short reads.

Ion Torrent S5: cheap and fast



Pro: cheap, short reads of ~200 bp, fast ~3 hours per run
Con: in/del errors, average throughput level ~ 50 Gb per run (~ 160 Millions reads)

minION : long read sequencing on your bench

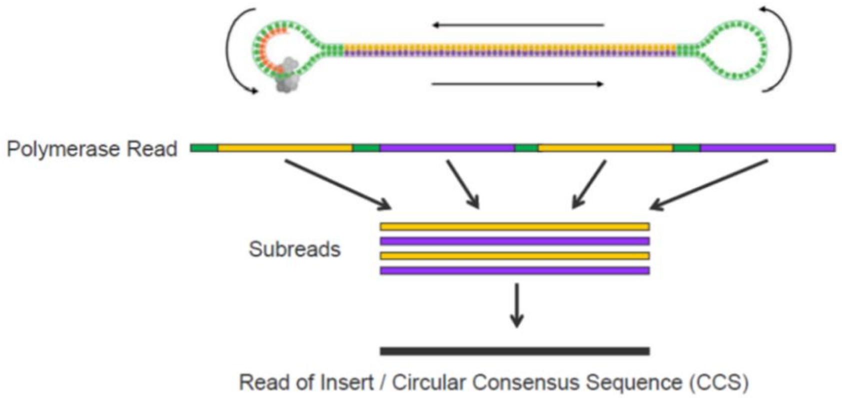


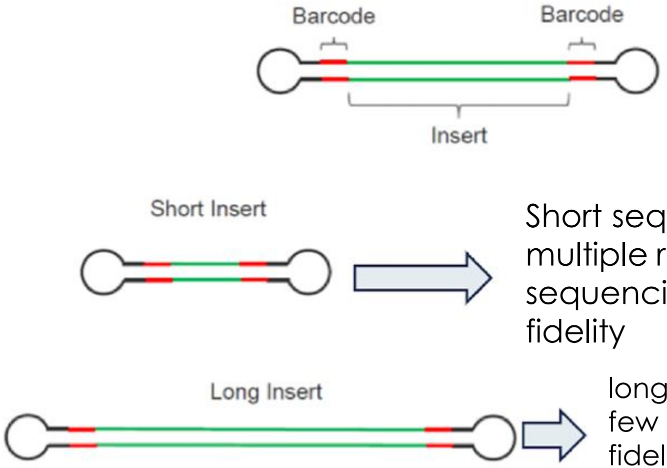
Oxford NANOPORE Technologies

Pro : No amplification bias.
cheap and very fast (results in few minutes).
long reads (~5 kb).

Con : High error rate ~3%. Low throughput : ~20 Gb (~5 Millions reads)

PacBIO Sequel: the war tank





Short sequence = multiple rounds of sequencing = High fidelity

long sequence = few rounds = lower fidelity.

Pro: **long reads (> several kb), very accurate.**

Con: **heavy equipment. Low throughput 1.5 Gb per run (~1 Million reads)**

Metagenomic sequencing comparative study using synthetic microbial mix

OAK RIDGE National Laboratory

Dr. Mircea Podar

Sample ID	*Sequencing Technology*	*N. reads after trimming (M)*	*Total run time*	*Average read size*	*Max read size*	*%Avg identity*
MOCK (N=71 species)	**Illumina HiSeq 3000**	20.59*2	4d	149	150	99.45
	Ion Torrent S5	28.51	36h	145.76	347	99.61
	DNBSEQ-G400	35.42*2	3d	99.91	100	99.70
	DNBSEQ-T7	375.12*2	3d	99.52	100	99.42
	ONT Minion R9	0.696	48h	4408.41	60869	89.08
	PacBio Sequel II	0.524	30h	10289.7	40278	99.72

- **2nd generation high throughput and high accuracy but short reads**
- **3rd generation low throughput and variable accuracy but long reads**

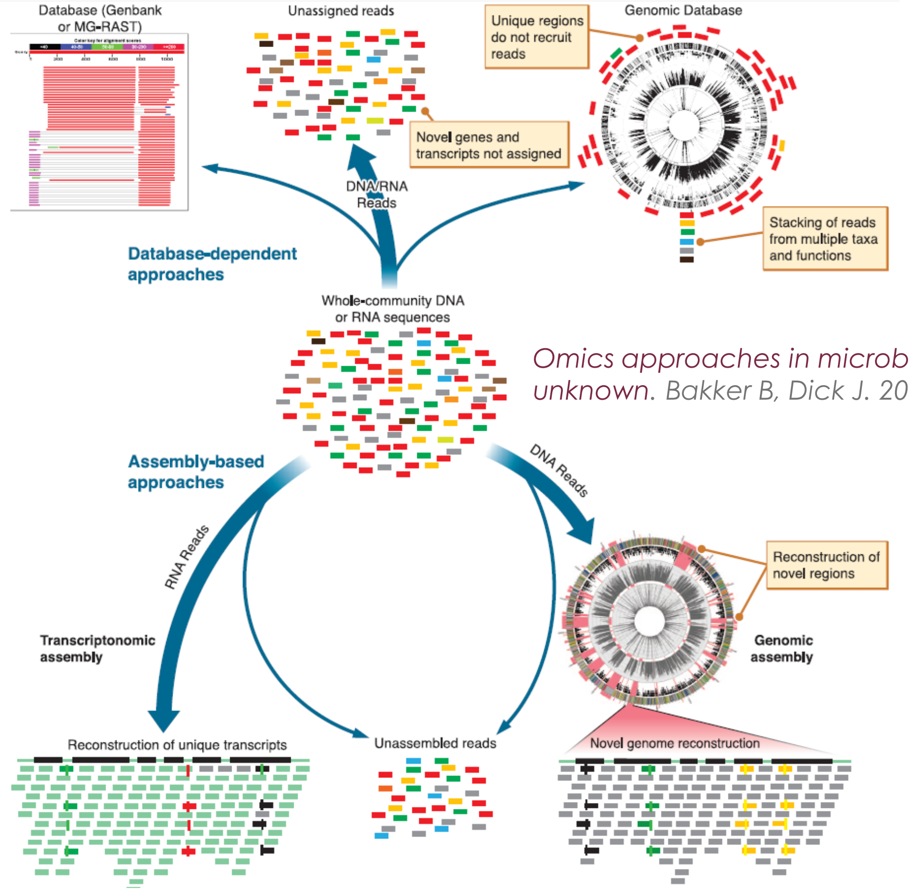
Millions of fragments to explore the microbiota

Mapping to references

Low computational cost

Genome *de novo* reconstruction

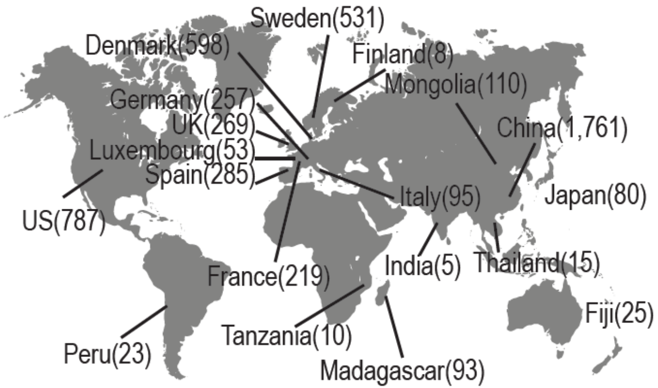
High computational cost

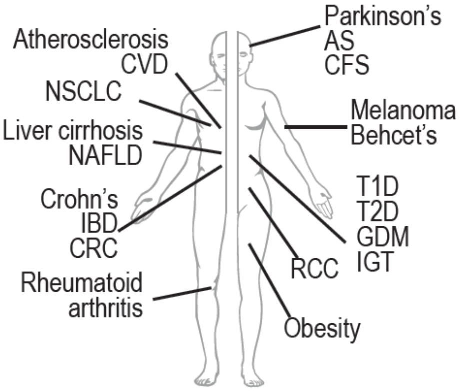


Omics approaches in microbial ecology: charting the unknown. Bakker B, Dick J. 2013. Microbes 8:353-360

Exploring thousands of gut microbial samples to identify microbial association with health and disease

The Human Microbiome ATLAS Project





>5000 gut metagenomic samples coming from 19 countries and exploring 20 diseases.

Deep functional and taxonomical exploration of microbiome in health and disease

All data available on ATLAS website (https://www.microbiomeatlas.org/)

[Shoaie *et al.*, in review]

metagenopolis mgps.eu

KING'S College LONDON

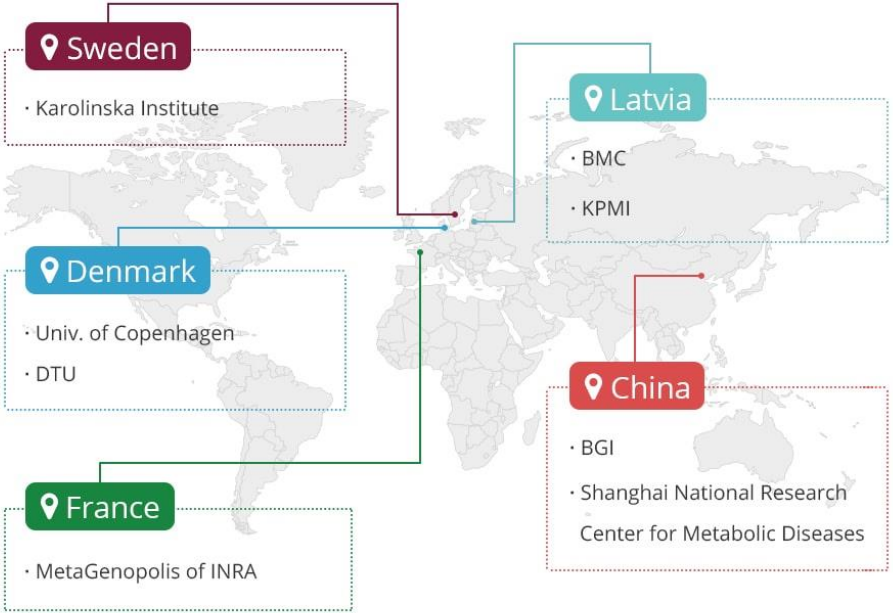
SciLifeLab

INRAE

Exploring millions of microbiome with the Million Microbiome of Humans Project (MMHP)

metagenopolis mgps.eu

Founding members of the MMHP



Sweden
- Karolinska Institute

Latvia
- BMC
- KPMI

Denmark
- Univ. of Copenhagen
- DTU

China
- BGI
- Shanghai National Research Center for Metabolic Diseases

France
- MetaGenopolis of INRA

MMHP

Collect and analyse 1 million human microbiota (gut, oral, skin, …) around the world by 2025.

Started 2019, 5 founding members, including MetaGenoPolis.

Metadata and samples accessible each year

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

BGI 华大 · MGI 华大智造 · Karolinska Institutet · INRAE · Danmarks Tekniske Universitet DTU · SciLifeLab · CSIC · BMC Latvian Biomedical Research and Study Centre · Maastricht UMC+ · Københavns Universitet · University of Latvia Institute of Clinical and Preventive Medicine · ikmb

Le microbiote français

Le French Gut

National citizen contribution aiming at collecting 100,000 faecal samples & associated nutritional and clinical data by 2025 to:

- Define the heterogeneity of the French gut microbiomes
- Identify the environmental and lifestyle factors driving microbial composition and their deviations in diseases

French Gut Consortium

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

INRAE

Inserm
La science pour la santé
From science to health

Private Institutions

Le French Gut: Overview

Recruitment via public media, French cohorts and patient networks

Le microbiote français :

Le French Gut

1 Online registration

Through the secured French Gut website

2 Eligibility & Questionnaire

Including clinical and nutritional forms

3 Sampling kit at home

Using MetaGenoPolis custom design collect kit

4 Sequencing and Analysis

Following IHMS high quality standards

5 Collective feedback

At each recruitment milestones

Pilot phase in 2021/2022

Some other MetaGenoPolis publications

Cancer

2018: *Routy et al.* **Science**, Microbiome and epithelial cancer immunotherapy
2018: *Gopalakrishnan et al.* **Science**, Microbiome and melanoma cancer immunotherapy

2020: Derosa et al. **European Urology**, Microbiome and Resistance to Cancer Immunotherapy in Renal Cell Carcinoma Patients

Metabolic disorders

2012: *Qin et al.* **Nature**, Type II Diabetes
2013: *Le Chatelier et al.* **Nature**, Richness of gut microbes and metabolic markers
2014: *Qin et al.* **Nature**, Human gut microbiome alterations in liver cirrhosis
2015: *Qin et al.* **Nature**, Accurate liver cirrhosis diagnostic
2015: *Forslund et al.* **Nature**, Drug confounders in microbiome analysis
2016: *Pedersen et al.* **Nature**, Microbiome & insuline resistance

2020: Vieira-Silva et al. **Nature**, Microbiome and Statin therapy
2021: Malard et al. **Nature Com**, *Leukemia and autologous fecal tranfer*

Diet

2011: *Arumugam et al.* **Nature**, Enterotypes
2013: *Cotillard et al.* **Nature**, Impact of diet on gut microbiome
2013: *Le Chatelier et al.* **Nature**, Richness of gut microbes and metabolic markers
2019: Cox et al. **Gastroenterology**, Low FODMAP Diet in Inflammatory Bowel Disease patients
2020: Meslier et al. **Gut**, Beneficial effects of Mediterranean diet

Antibiotic resistance

2019: *Ruppé et al.* **Nature Microbiology**, Prediction of the intestinal resistome

Technologies

2010: *Qin et al.* **Nature**, The human gut reference catalogue
2013: *Sunagawa et al.* **Nature Methods**, Universal phylogenetic markers
2014: *Nielsen et al.* **Nature Biotech**, Method for identifying metagenomic species
2014: *Li et al.* **Nature Biotech**, 10 millions genes reference catalog
2015: *Xiao et al.* **Nature Biotech**, A mouse gut gene catalogue
2016: *Xiao et al.* **Nature Microbiology**, A pig gut gene catalog
2017: *Costea et al.* **Nature Biotech**, Standards for microbiome studies
2018: *Plaza Onate et al.*, ***Bioinformatics***, Reconstitution of metagenomic pangenome species
2020: *Herold et al.* **Nature Com**, *Time series meta-omics in dysbiosis*

Thank you!

Business contact: karine.valeille@inrae.fr
Scientific contact: joel.dore@inrae.fr ; mathieu.almeida@inrae.fr