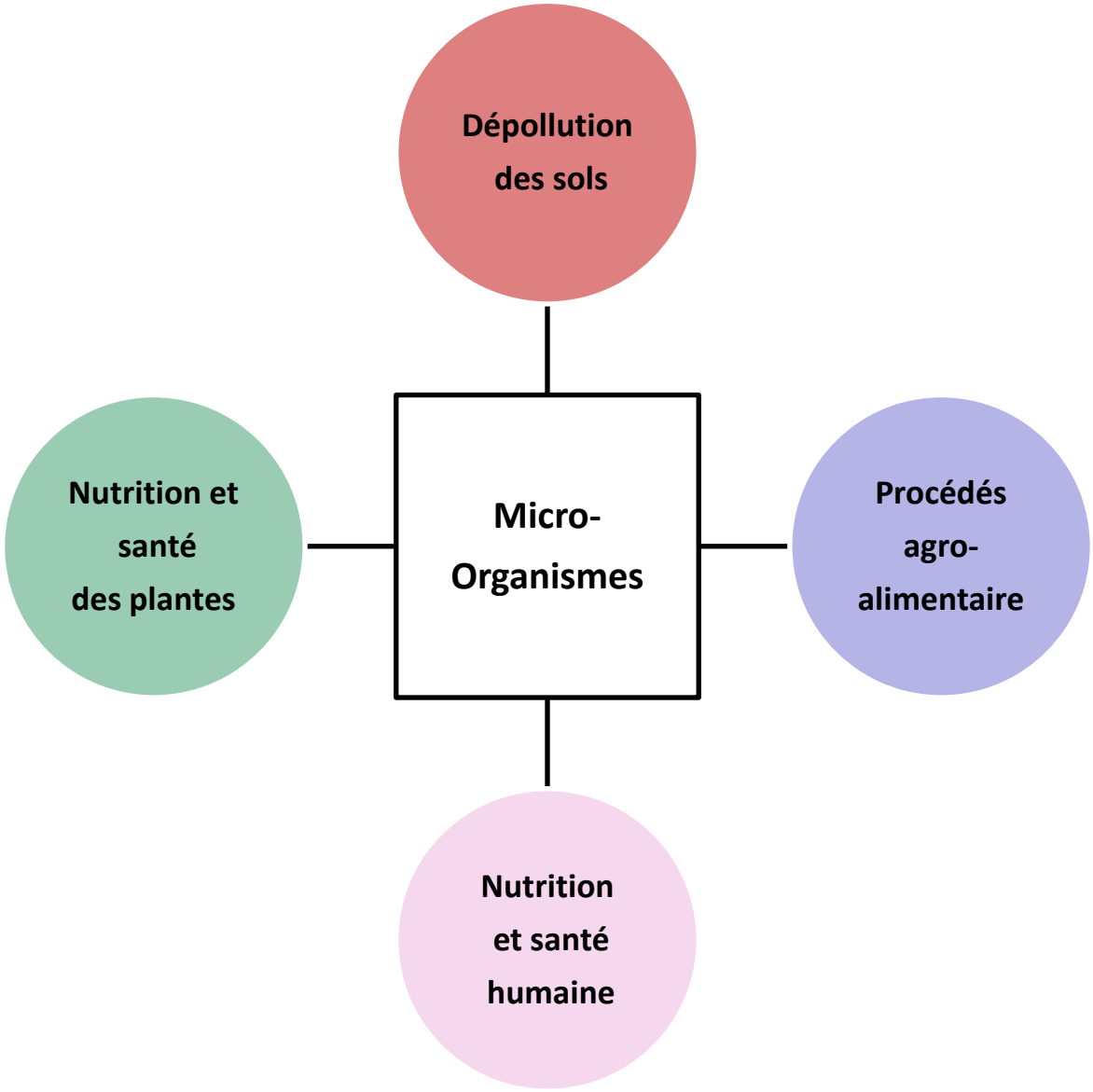
La métagénomique au service de l'étude des communautés microbiennes

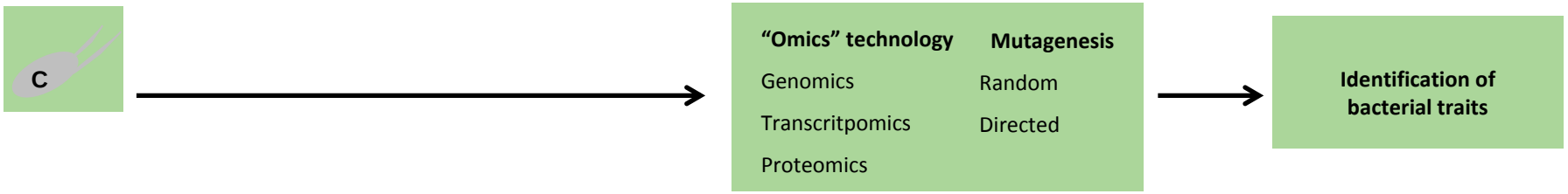
Matthieu Barret

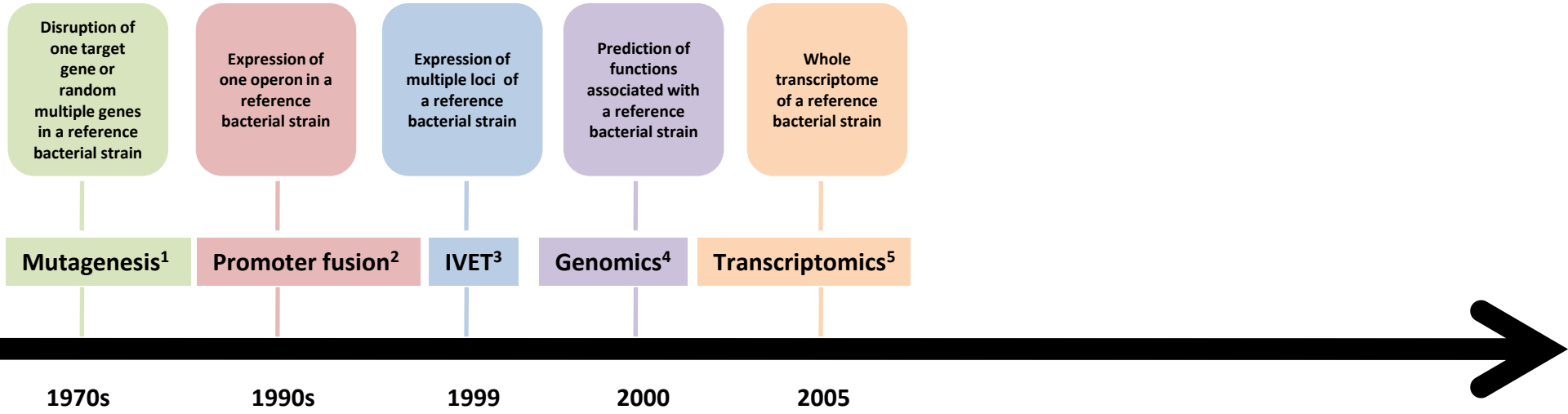
Importance des micro-organismes



Approches culture-dépendantes

- Quelles sont les micro-organismes/fonctions impliqués dans ces processus ?





- **Approches restreintes aux micro-organismes cultivables**

Bactéries non-cultivables

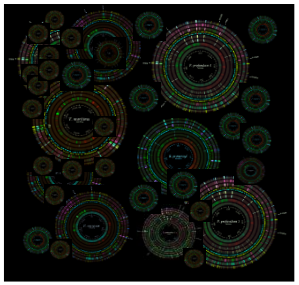
- Aucun milieu de culture ne permet la croissance de la totalité des bactéries
- 1-10% bactéries sont actuellement cultivables

Habitat	Cultivability (%)
Freshwater	0,25
Seawater	0,1 – 3
Sediments	0,25
Soil	0,3
Food	0,5 -10

- Alternatives :
  - Développer de nouvelles méthodes de culture
  - Utiliser des approches culture-indépendantes (ex: métagénomique)

Métagénomique (quelques définitions)

« *procédé méthodologique qui vise à étudier le contenu génétique d'un échantillon issu d'un environnement complexe* »



« *étude de l'ensemble des génomes de la microflore environnementale* »
Handelsman *et al.*, 1998

- *Métagénomique*: étude des métagénomes
- *Métagénome :* génome d'un microbiote
- *Microbiote* : ensemble des micro-organismes issus d'un environnement

Métagénomique (quelques applications)

Permet d'étudier les micro-organismes cultivables et non-cultivables dans leur environnement natif

1- Permet d'estimer la diversité microbienne associée à un environnement

2- Permet d'estimer l'abondance de gènes (fonctions) au sein d'un environnement

3- Permet d'isoler de nouveaux gènes codant de nouvelles enzymes

- composés antimicrobiens
- biotransformation de la biomasse végétale

Challenges in metagenomic




eDNA extraction

Sequencing

Sequences analyses

Challenges in metagenomic




eDNA extraction

cellular compartmentalization of individual genotypes is destroyed

Sequencing

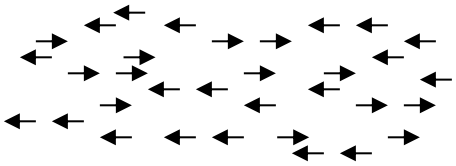
DNA fragments are typically broken by mechanical or enzymatic methods

Sequences analyses

1-10 millions reads (100-300 bases each)

Sequences analyses

Reads

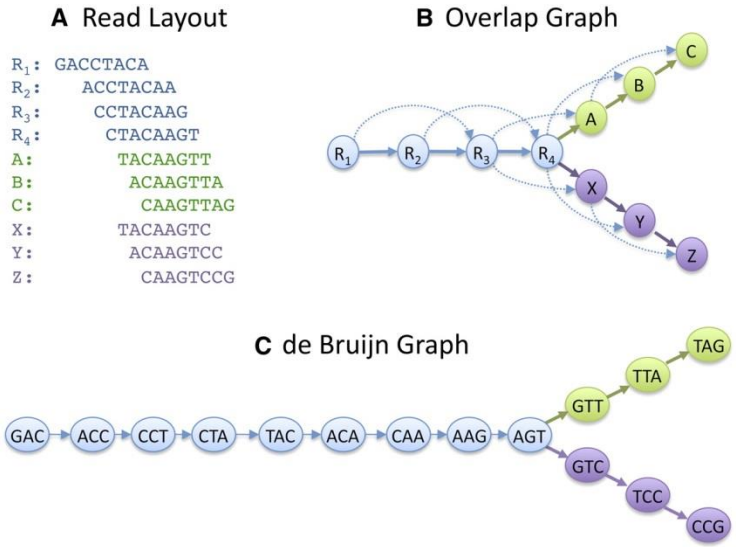


Genomes

- Which populations are represented in the metagenome?
- What is their relative abundance?
- Can I distinguish sequences from different populations?
- Is my population of interest similar/dissimilar to isolate genome(s) in terms of gene content?
- What is the functional content of the genome of my population of interest?

De novo assembly of metagenomes

- Based on de Bruijn graph (MetaVelvet, Ray Meta, IDBA-UD, Omega)



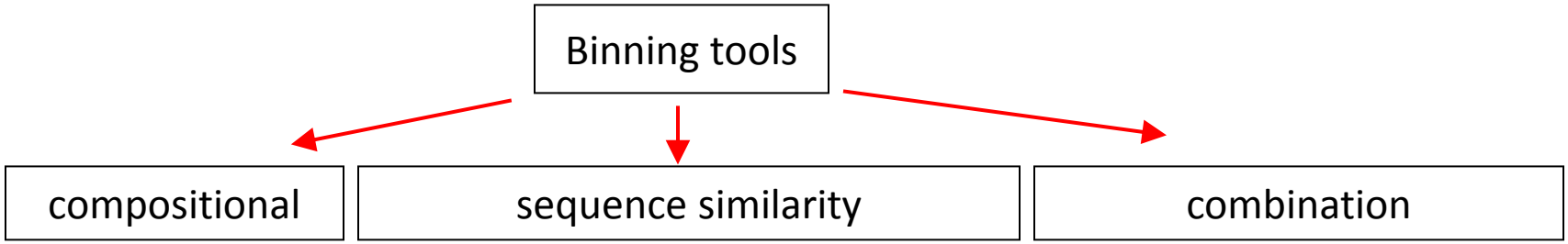
- If k is large = edges are missing (gaps)

- If k is small = branching problem (chimera)

- Main problem in metagenomic assembly = presence of multiple related strains that act like poison to the assembler

Binning

- **unassembled reads** can be **binned** into sets of sequences (**bins**) that likely originated from the same species or broader taxonomic lineages



Binning

- Sequence-similarity : compare reads to database (MG-RAST; IMG-M)

  MG-RAST = 59,86 Tb

  IMG-M = 3,7 Tb

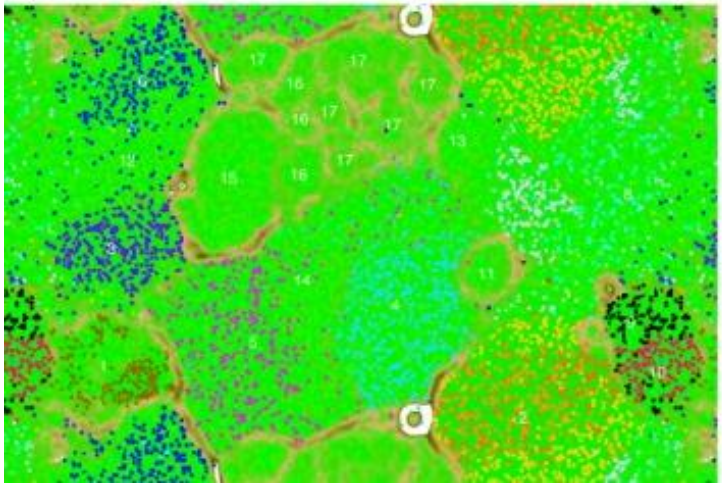
- Group reads based on taxonomic information (species, genus, family…)

- Require great computational resources as every read is aligned to a large volume of sequence

Binning

- Sequence-composition : group reads according to their sequence composition (GC% or tetranucleotide frequency)

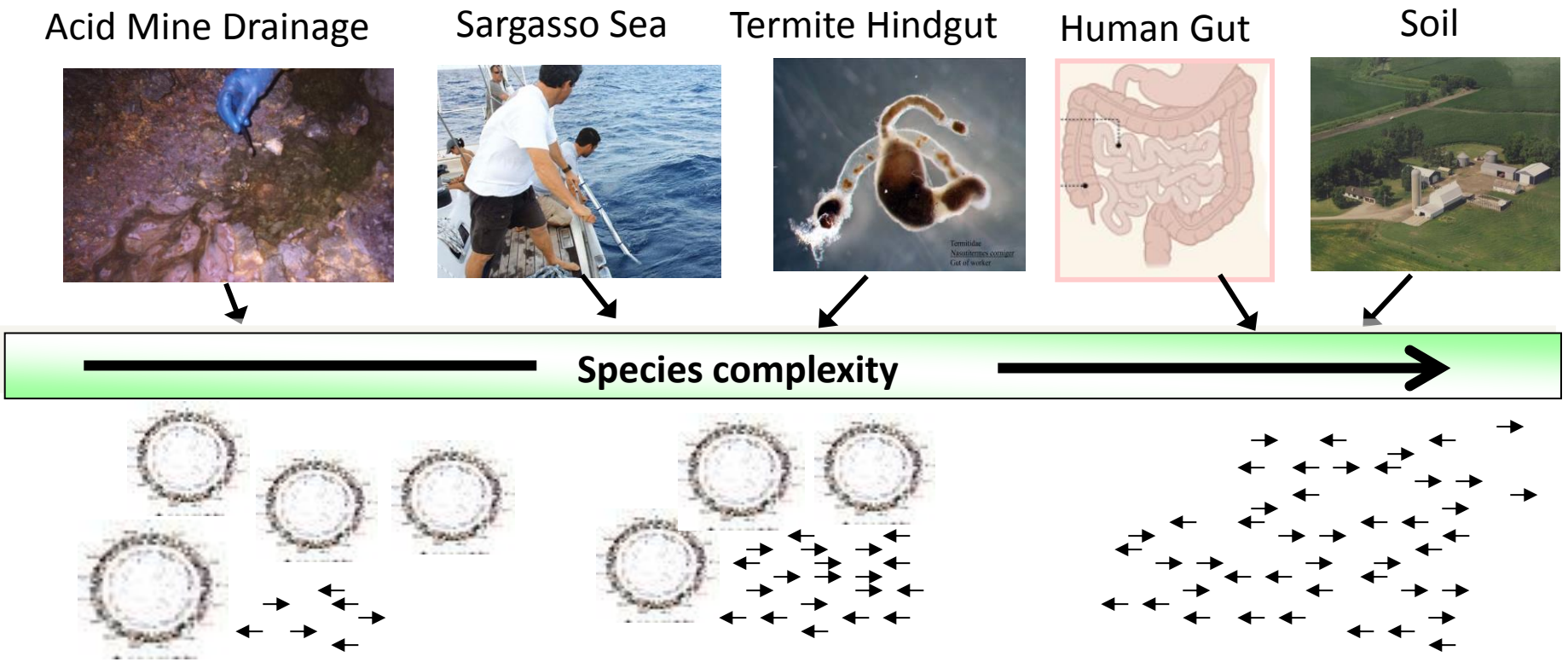
- Emergent Self-Organizing Maps are used to cluster reads (Wrighton *et al*., 2012)



- However many of the obtained groups are not clearly separated

Conclusions

- Analyzing metagenomic data is a major bottleneck
- Assembly tests show that contigs are slightly larger than read sizes
- Binning could help assembly of abundant taxa



Future directions : comparative metagenomics

- Could we exploited similarities/difference between metagenomes ?
- How ? Using partial knowledge (strongly biaised)
- How ? Using compositional approaches (read sequence level)
- Provides reads similar/different bewteen datasets
- How to scale up to large datasets ?

COMMET: comparing and combining multiple metagenomic datasets

Nicolas Maillet*, Guillaume Collet†, Thomas Vannier‡, Dominique Lavenier* and Pierre Peterlongo*
*INRIA / IRISA-UMR CNRS 6074, EPI GenScale, Rennes, France
†INRIA / IRISA-UMR CNRS 6074, EPI Dyliss, Rennes, France
‡Direction des Sciences du Vivant, Commissariat à l'Énergie Atomique (CEA),
Institut de Génomique, Genoscope, Laboratoire d'Analyses BioInformatique des Séquences -
UMR CNRS 8030 - Université d'Évry Val d'Éssonne (UEVE), France

References

tool	category	references
MetaPhyler	marker gene analysis	Liu et al., 2011
MetaPhlAn2	marker gene analysis	Segata et al., 2012
AMPHORA2	marker gene analysis	Wu and Scott, 2012
PhyloSift	marker gene analysis	Darling et al., 2014
IMG/M 4	database	Markowitz et al., 2014
MG-RAST	database	Meyer et al., 2008
metAMOS	assembly	Treangen et al., 2013
MetaVelvet	assembly	Namiki et al., 2012
IDBA-UD	assembly	Peng et al., 2012
Ray Meta	assembly	Boisvert et al., 2012
Omega	assembly	Haider et al., 2014
GroopM	binning	Imelfort et al., 2014
MaxBin	binning	Wu et al., 2014
Commet	comparative metagenomic	Maillet et al., 2014