



**HAL**  
open science

## Unlocking the Soil Microbiome: Unraveling Soil Microbial Complexity Using Long-Read Metagenomics

Carole Belliardo, Nicolas Maurice, Clémence Frioux, Claire Lemaitre, Riccardo Vicedomini, Samuel Mondy, Arthur Péré, Marc Bailly-Bechet, Etienne Danchin

### ► To cite this version:

Carole Belliardo, Nicolas Maurice, Clémence Frioux, Claire Lemaitre, Riccardo Vicedomini, et al.. Unlocking the Soil Microbiome: Unraveling Soil Microbial Complexity Using Long-Read Metagenomics. EAGS 2024 - The International Environmental and Agronomical Genomics symposium, Feb 2024, Toulouse, France. , pp.1-1, 2024. hal-04509213

**HAL Id: hal-04509213**

**<https://hal.science/hal-04509213>**

Submitted on 19 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Unlocking the Soil Microbiome : Unraveling Soil Microbial Complexity using Long-Read Metagenomics



Carole Belliaro<sup>1</sup> Nicolas Maurice<sup>2,3</sup>, Clémence Frioux<sup>3</sup>, Claire Lemaitre<sup>2</sup>, Riccardo Vicedomini<sup>2</sup>, Samuel Mondy<sup>4</sup>, Arthur Pere<sup>1</sup>, Marc Bailly-Bechet<sup>1</sup>, Etienne Danchin<sup>1</sup>  
<sup>1</sup> INRAE, Institut Sophia Agrobiotech, Sophia Antipolis, France; <sup>2</sup> Inria/IRISA Genscale, Campus de Beaulieu, 35042 Rennes, France  
<sup>3</sup> Inria/INRAE Pleiade, Campus de Talence, 33405 Bordeaux, France; <sup>4</sup> INRAE, Institut AGRO Dijon, Université de Bourgogne, 21065 DIJON  
 carole.belliardo@inrae.fr ; nicolas.maurice@inria.fr



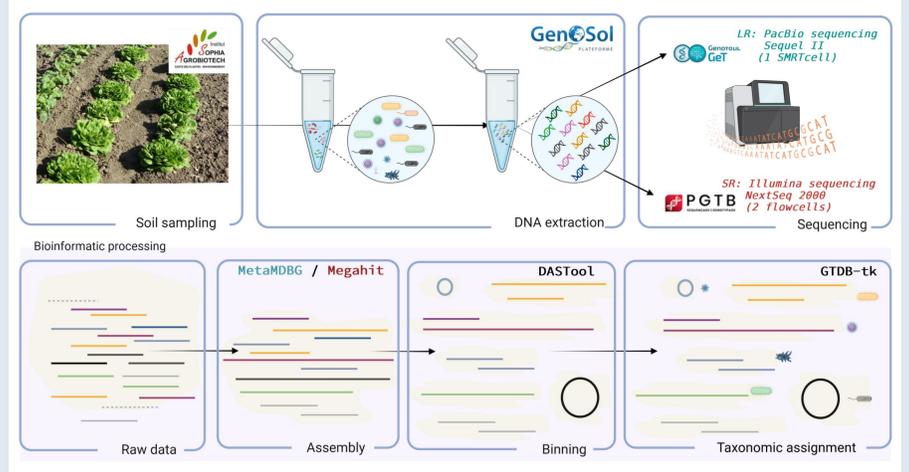
## Context

Plants can recruit micro-organisms from the millions ( $10^3$ - $10^8$ /g) found in soil, to regulate key functions such as [1] :

- Nutrition and growth,
- Resistance to biotic and abiotic stress.

Still, soil microbiomes are poorly characterized [2], as culturomics, transcriptomics and short-read metagenomics cannot recover enough complete genomes needed to infer their functions and model those communities.

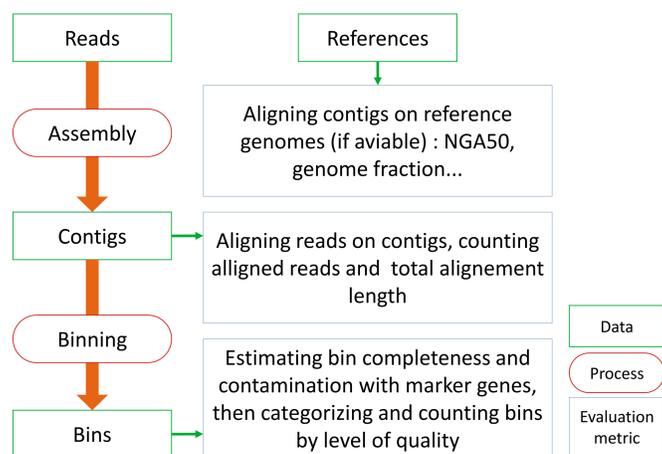
We produced both short-reads (SR) and HiFi long-reads (LR) metagenomes from tunnel culture soils, with the aim to reconstructing metagenome-assembled genomes (MAGs). We hypothesize that HiFi long-read sequencing technologies could enhance metagenomics by improving genome reconstruction. We compared several methods to reconstruct metagenomes from HiFi long reads and compared the results to deep short-reads sequencing.



## Evaluation of existing highly accurate long-read metagenome assemblers

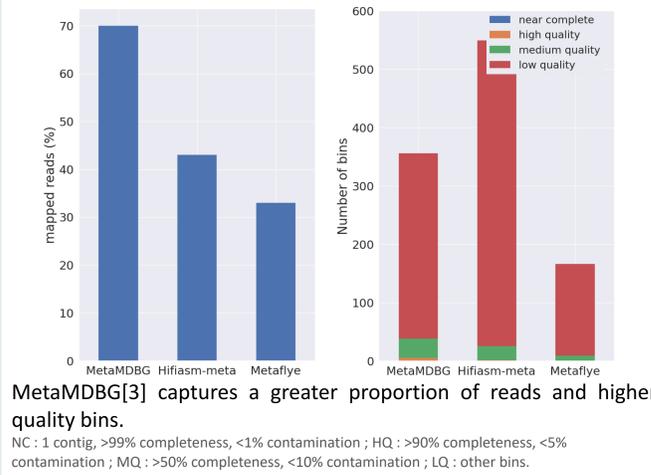
### Mapler

A pipeline built for evaluating Hi-Fi long read metagenome assemblers.



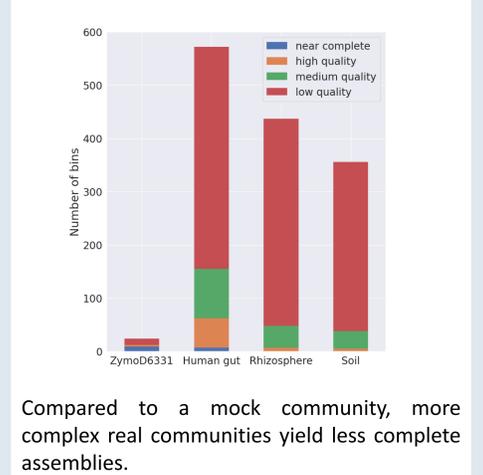
### Assembler comparison

On the soil of a salad culture in PACA region (1SMRT Cell)



### Sample comparison

Using metaMDBG



## Long-reads versus short-reads technologies comparison

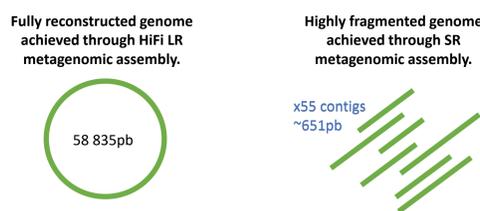
### Soil metagenomes summary

		Short-reads	Long-reads
Read nb. (N50)		2 x 991M (151)	2,9M (7,6 kb)
Contigs	Contig nb. (N50)	31M (620)	493 030 (7,6 kb)
	Predicted proteins (% complete)	41 72 9705 (39%)	1 379 489 (97%)
	Average depth	6X[0:763 551]	3X [0:37]
Bins	Bin nb.	228	55
	MAGs nb.	139	13
	Average depth	48X	3X

Unassembled HiFi LR surpass the contig length of SR contigs. Although SR predicted 30 times more proteins, LR predictions are less redundant and almost complete, with more than 60% of prediction from SR being fragmented.

### Example of a Metagenome-Assembled genome

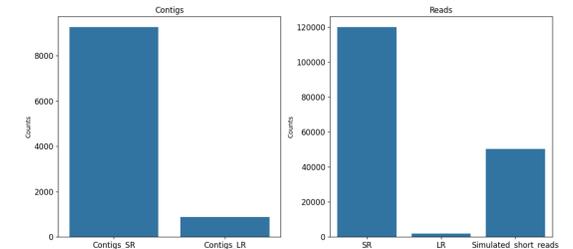
*Actinoplanes phage phiAsp2*, a tiny genome of 58.6kb



LR provides higher contiguity of reconstructed genomes, yielding more suitable data for their structural and functional analysis.

### More reliable taxonomic assignment

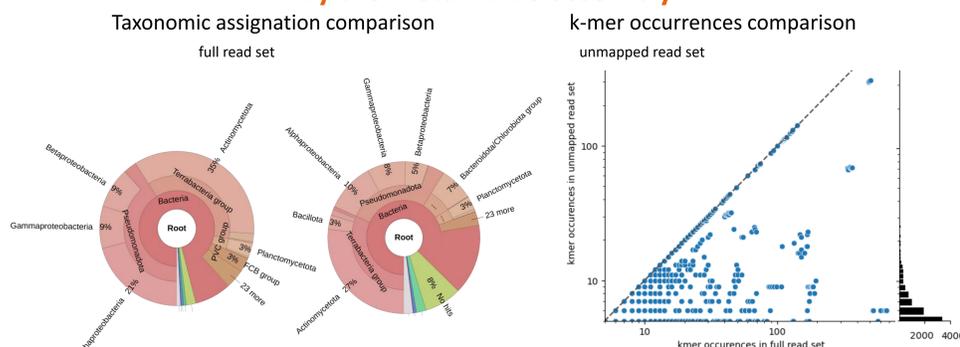
Dozens of MAGs were obtained from LR after assembly, binning and dereplication, encompassing bacterial, archaeal, and viral genomes from terrestrial environments taxonomically annotated with high accuracy.



Shorter sequences result in a significant amount of mis-annotation.

## Lack of captured diversity

### By the metaMDBG assembly

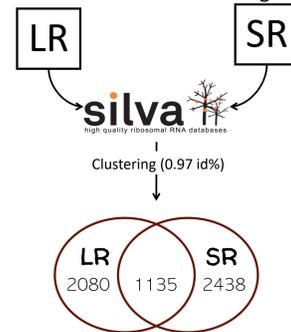


Assembly fail at capturing some or the soil's biodiversity, as reads that failed to map to the assembly are evidence that they aren't properly assembled :

- Rare taxa are more represented in unmapped reads,
- Some k-mers, especially rare ones, belong to unmapped reads.

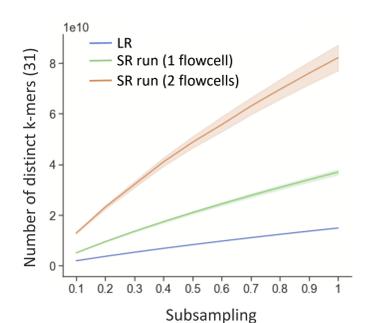
### By sequencing

Assessment of the genetic diversity captured by sequencing technologies via Virtual metabarcoding



No sequencing technology seems to have successfully captured all the soil's biodiversity yet

Genetic diversity captured by sequencing technologies assessed by k-mer occurrence analysis



## Conclusion

LR metagenomics promises a more comprehensive insights into the genomic structure, functional potential of soil microbial genomes and their taxonomic assignment. Since none of the existing technologies appear to offer an optimal representation of the terrestrial microbiome, the integration of both technology into a hybrid assembly and/or adapting assembly algorithms for low coverage HiFi LR could enhance metagenome recovery.



[1] Hoysted G. et al., 2018, Curr. Op. in Pl. Bio  
 [2] Fierer, N., 2017, Nat Rev Microbiol  
 [3] G. Benoit et al., 2024, Nature