# Benefits of PacBio HiFi long reads for metagenomic whole genome analysis

Adrien Castinel, Jean Mainguy, Olivier Bouchez, Sylvie Combes, Carole Iampietro, Christine Gaspin, Denis Milan, Cecile Donnadieu, Géraldine Pascal, Claire Hoede

# Benefits of PacBio HiFi long reads for metagenomic whole genome analysis

Adrien CASTINEL[1], Jean MAINGUY[2], Sylvie COMBES[3], Christine GASPIN[2], Denis MILAN[1 3], Cécile DONNADIEU[1], Carole IAMPIETRO[1], Géraldine PASCAL[3], Olivier Bouchez[1] and Claire HOEDE[2]

1 GeT-PlaGe, US 1426, Genotoul, INRAE, 31320 Castanet-Tolosan, France
2 MIAT, PF Bioinfo GenoToul, Université de Toulouse, INRAE, Chemin de Borde Rouge, 31320 Castanet-Tolosan, France
3 GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, Chemin de Borde Rouge, 31320 Castanet Tolosan, France

Contact :
olivier.bouchez@inrae.fr
claire.hoede@inrae.fr

## Introduction

Whole metagenomics analysis aim to provide an overview of the biodiversity and the functions of bacterial communities. Currently, those analysis are performed on Illumina short read sequencers that provide the advantage of generating less than 1% sequencing error. The arrival of the **HiFi technology (PacBio Sequel II)** allow us to sequence long read fragments of several Kb with a low error rate.

One of the aim of the **SeqOccIn project** (Sequencing Occitanie Innovation, https://get.genotoul.fr/seqoccin/) is to acquire expertise in metagenomics to better characterize microbial communities inside complex matrices. In the framework of this project, we developed the **metagenomics shotgun protocol** on the PacBio Sequel II sequencer and carry out a comparative **study of the assemblies** obtained with PacBio long reads and Illumina short reads data.

## DNA extraction for long read sequencing

For metagenomic study, an extraction method able to produce **DNA fragment up to 10Kb** while maintaining a good **species representativeness** is necessary. That's why we tested 5 different protocols (InnuPREP Stool DNA, Mag-Bind Stool DNA, QIAamp PowerFecal Pro, DNA miniprepKit, Quick DNA fecal/soil) on a **microbial mock community** (zymoBIOMICS D6300). Three methods out of five produced fragments of sufficient size and quality. We selected them to assess the representativeness of the bacterial communities.

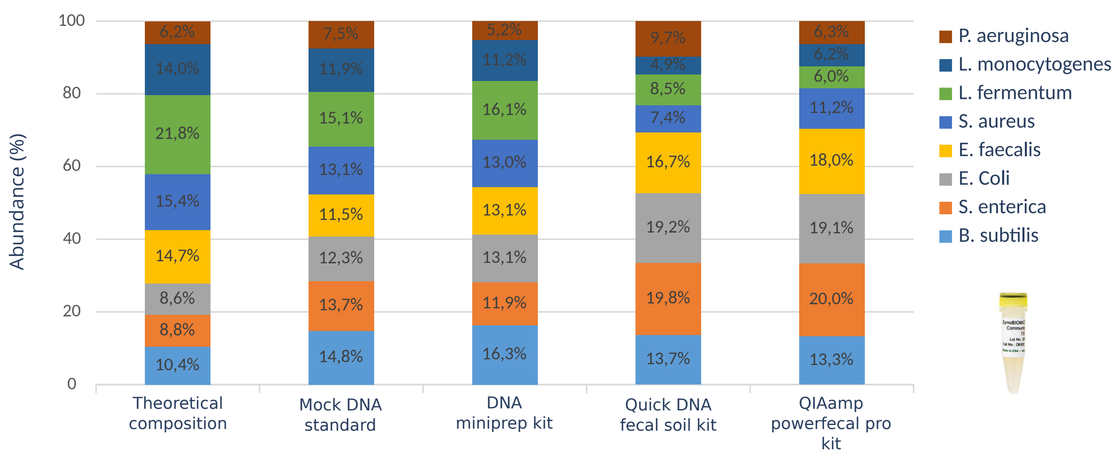



**Fig. 1: The DNA Miniprep kit showed less biases compared to the two other kits.** The 16S rRNA gene sequenced on Illumina MiSeq allowed us to evaluate the percentage of each species found in the microbial mock community and to compare it with the expected theoretical percentage. Among the three kits tested, the DNA Miniprep kit showed less biases compared to the two others.

## Metagenomic assembly results

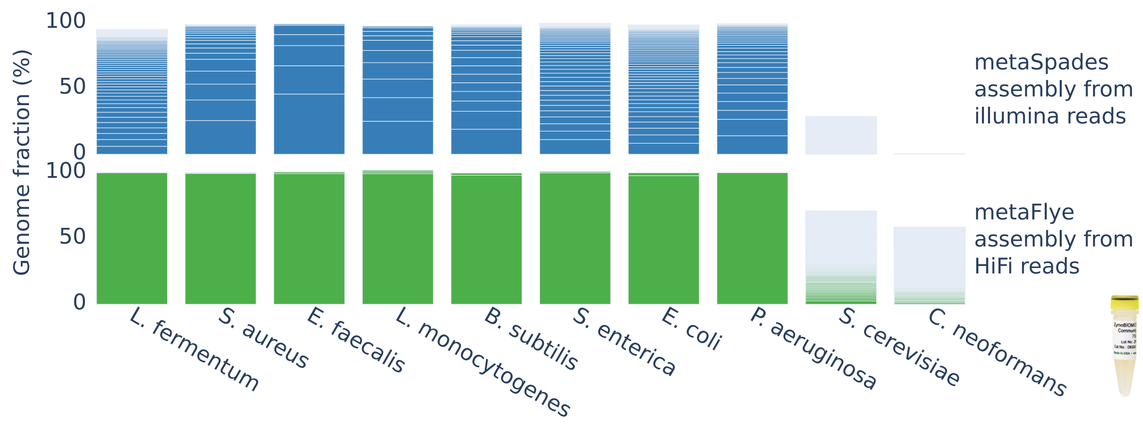



**Fig. 2: Bacterial genomes of mock Zymobiomics are assembled in one contig by HiFi reads.** The genome fraction is the fraction of the reference genomes covers by the assembly. Each block in a bar represents a contig. Bacterial genomes are well covered by all assemblies but Illumina assembly is much more fragmented

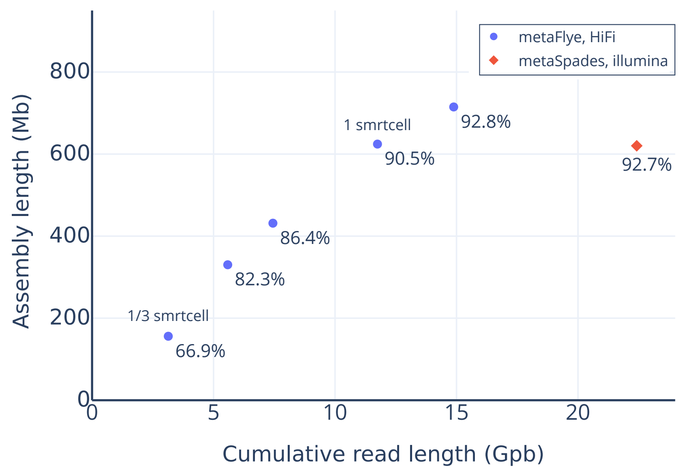



**Fig. 3: Complexe samples of pig feces required high sequencing depth.** The percentage of mapped reads on the assembly is for each read set

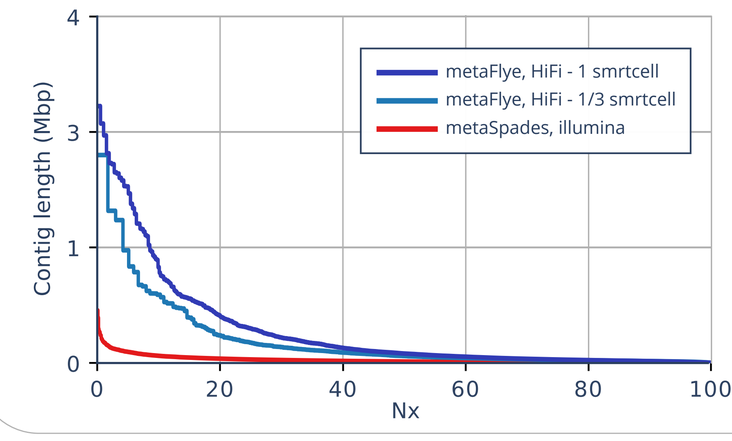
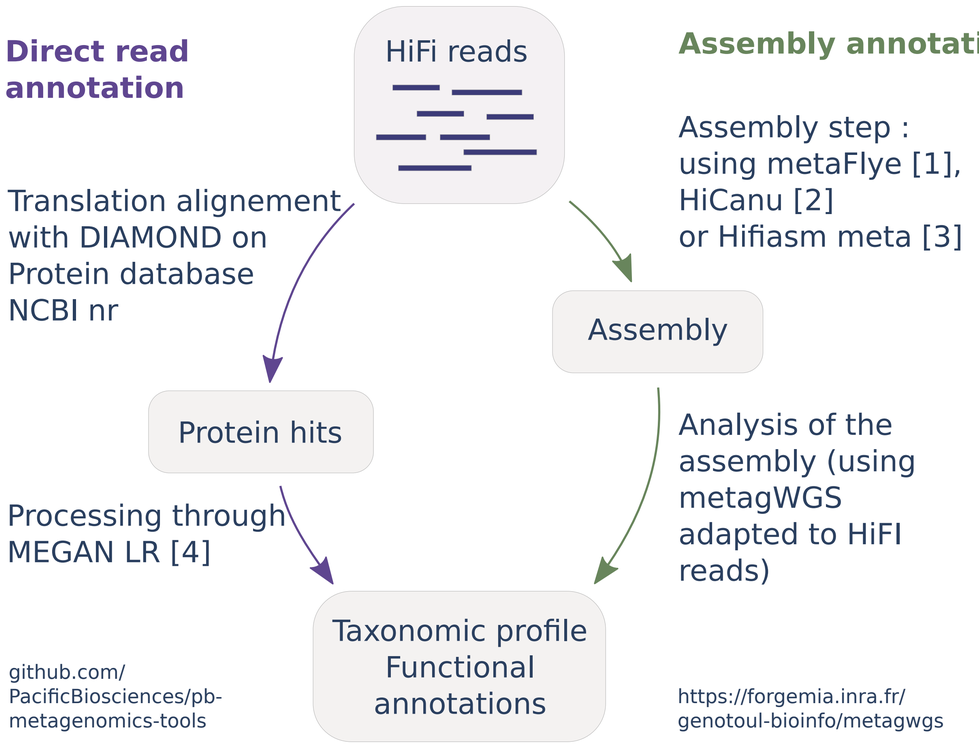



**Fig. 4: The assemblies are far less fragmented with HiFi reads than with Illumina reads.** Nx is a measure of the assembly contiguity.

## How to analyse metagenomic HiFi reads?




References:
1. Kolmogorov, Mikhail, et al. "metaFlye: scalable long-read metagenome assembly using repeat graphs." Nature Methods 17.11 (2020).
2. Nurk, Sergey, et al. "HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads." Genome research 30.9 (2020).
3. Cheng, Haoyu, et al. "Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm." Nature Methods 18.2 (2021).
4. Huson, Daniel H., et al. "MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs." Biology direct 13.1 (2018).

## Taxonomic profiles

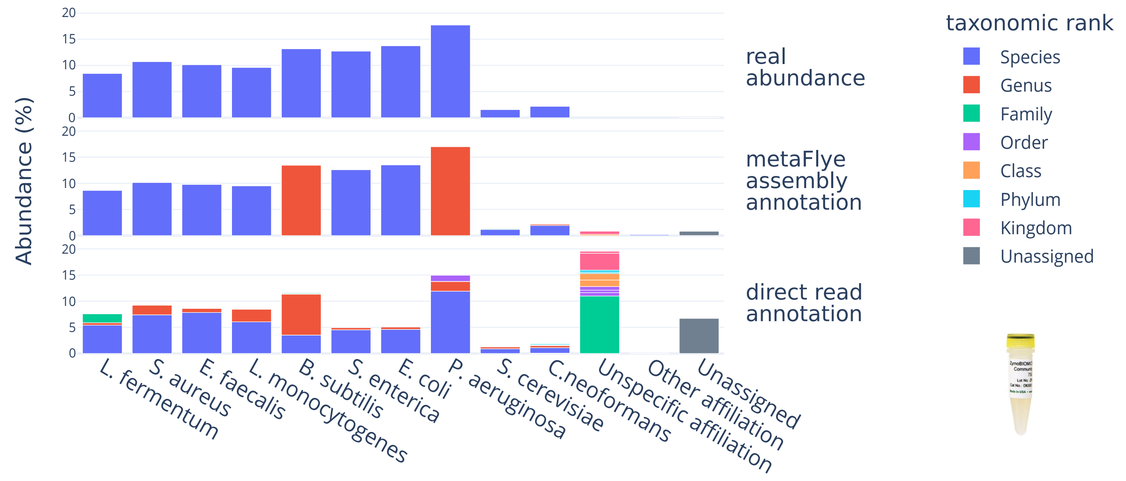



**Fig. 5: Direct read annotation produces less accurate taxonomic profile than assembly annotation on mock Zymobiomics.**

## Conclusion

The **DNA miniprep** kit has been selected to extract DNA for long read sequencing.
With enough sequencing depth, HiFi reads produce **high quality assemblies** but also imply a **higher cost**. That's why **direct read annotation** seems promising as it could required **less sequencing depth**. Howewer our first analysis on the mock shows that this approach is less accurate than assembly annotation results. We are currently working on the method to improve the results.