Shallow sequencing: a cost-effective and accurate alternative to WGS for taxonomic profiling ?

Benoit Goutorbe, Anne-Laure Abraham, Mahendra Mariadassou, Valentin Loux, Olivier Rué, Anne Plauzolles, Ghislain Bidaut, Philippe Halfon, Sophie Schbath

Shallow sequencing: a cost-effective and accurate alternative to WGS for taxonomic profiling ?

Benoit Goutorbe[1,2,3], Anne-Laure Abraham[1], Mahendra Mariadassou[1], Valentin Loux[1,4], Olivier Rué[1,4], Anne Plauzolles[2], Ghislain Bidaut[3], Philippe Halfon[2] and Sophie Schbath[1]

[1] Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France

[2] Laboratoire Alphabio, 1 rue Melchior Guinot, 13003, Marseille, France

[3] CRCM, Inserm, CNRS, Institut Paoli-Calmettes, Aix-Marseille Université, 13009, Marseille, France

[4] Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

Corresponding author: benoit.goutorbe@inrae.fr

Context

In the rising research area of **microbiota-associated health outcomes**, clinical researchers have to deal with the critical choice of the analytic technique used to characterize patients' microbiota. This choice is usually binary, with **metabarcoding**, a low-cost and an efficient way to identify and quantify organisms present in an ecosystem (taxonomic profiles) which has a limited resolution and suffers from well known biases (amplification biases, variation in copy numbers, etc), and **whole genome sequencing (WGS)**, which offers deep insights about both taxonomic and functional profiles but is much more expensive (about 10 times the cost of metabarcoding) and produces data massively more complex to analyse.

Due to a huge inter-patient variability, **large cohorts are needed** to extract reliable information. Thus, a large majority of studies are carried out using metabarcoding techniques. Shallow WGS (WGS at very low sequencing depth, down to 500 K reads/sample) is one of the techniques that could fit into this **technological gap** : a recently published paper [1] demonstrated the huge potential of this approach but left **many important questions unanswered**.

Materials & Methods

Using several public human gut microbiota datasets, covering several diseases and health conditions [2,3,4] we evaluated the information content of deep and shallow WGS. In order to do so, we subsampled in each dataset the reads at several sequencing depths, down to 10K reads/sample. Using *Unified Human Gut Genome* [5] as reference database, we compared two aligners (*BWA MEM*, *bowtie2*) and a k-mer based method (*kraken2*). We also focused on assessing how redistributing ambiguously mapped reads could enhance the profiles at critical sequencing depths. We assessed how low-depth sequencing degrades (i) the taxonomic profile and (ii) downstream statistical analysis (here biomarker discovery and diagnosis ability).

Results

Our preliminary results indicate that the taxonomic profiles seem to be robust at low sequencing depth. As expected, relatively abundant populations are less impacted than rare ones. Diagnosis-like classifications, often focused on relatively abundant populations, don't seem to be impacted by sequencing depth. Further investigations are needed to identify in which contexts shallow sequencing is enough to characterize an ecosystem and discriminate patients regarding to clinical outcomes.

References

[1] B. Hillmann *et al.* Evaluating the Information Content of Shallow Shotgun Metagenomics. 3(6):12, 2018.

[2] R. Loomba *et al.* Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection of Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease. *Cell Metabolism*, 25(5):1054–1062.e5, May 2017.

[3] V. Matson *et al.* The commensal microbiome is associated with anti–PD-1 efficacy in metastatic melanoma patients. *Science*, 359(6371):104–108, January 2018.

[4] N. Qin *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature*, 513(7516):59–64, September 2014. Number: 7516 Publisher: Nature Publishing Group.

[5] A. Almeida *et al.* A unified sequence catalogue of over 280,000 genomes obtained from the human gut microbiome. preprint, Microbiology, September 2019.