Shallow Shotgun Metagenomics as a cost-effective and accurate alternative to WGS for taxonomic profiling and clinical diagnosis

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

Shallow Shotgun Metagenomics as a cost-effective and accurate alternative to WGS for taxonomic profiling and clinical diagnosis

Benoit Goutorbe[1,2,3], Anne-Laure Abraham[1], Mahendra Mariadassou[1], 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] Centre de Recherche en Cancérologie de Marseille, Cibi plateform, Aix Marseille Université U105, Inserm U1068, CNRS UMR7258, Institut Paoli Calmettes, Marseille, France

Corresponding author: benoit.goutorbe@inrae.fr

Abstract *Shallow shotgun metagenomics has been recently suggested as a promising strategy to study human microbiota, providing nearly identical taxonomic profiles than deep shotgun metagenomics with a sequencing cost similar to metabarcoding. With shallow sequencing approach (typically <1M reads/samples), taxonomic profiles are directly built by mapping reads on a catalog of reference genomes, without assembly step.*

In the present study, we first used simulated data set to design a dedicated workflow in order to obtain reliable taxonomic profiles from shallow sequencing reads. We propose a novel data-driven filtering method based on machine learning techniques that largely outperformed basic filtering methods. We then used this approach on 3 real data sets, covering patients from several continents and clinical conditions. Even if one looses some information like rare taxa, our results clearly show that shallow shotgun metagenomics is able to correctly retrieve structures like differences between groups of patients and diagnosis-like classification.

Keywords Shallow shotgun metagenomics, Gut microbiota, Sequencing depth, Clinical research

1 Introduction

Allowing culture-free analysis of microbial ecosystems, high throughput sequencing revolutionized our comprehension of the role that plays human associated microbiota in health and disease. It is nowadays a very active clinical research field, with thousands of studies carried out each year, covering many diseases [1] [2] [3]. Two sequencing strategies emerged to study microbiota, and the choice depends on the sequencing cost, the size of the cohort, the expected level of taxonomic resolution and, when possible, functional annotation. On the one hand, metabarcoding, which consists in targeted sequencing of a phylogenetic marker (often rRNA gene 16S for bacteria, ITS for fungi), is a very cost-efficient way to characterize diversity within and between samples and to obtain an approximate taxonomic identification of microorganisms (often down to the genus level). On the other hand, shotgun sequencing consists in sequencing all DNA material present in an environment, which allows deeper taxonomic resolution (species, or even strain level), functional profiling (identification and quantification of genes, metabolic pathways), and *de novo* assembly of uncultured organism genomes as Metagenome Assembled Genomes (MAGs) [4].

Due to huge inter-patients variability and new insights into microbiome's plasticity [5], clinical studies need to include many patients [6], and have a longitudinal approach when possible, to extract reliable information. Despite the continuous drop in sequencing costs, metabarcoding is thus often preferred to shotgun sequencing to carry out clinical studies, providing limited information and hindering our comprehension of microbiota.

Shallow shotgun metagenomics has been recently suggested as an alternative [7], cost-competitive to metabarcoding (allowing analysis of large cohorts) and providing nearly the same information as *deep* shotgun sequencing. 20M reads/sample were typically used to characterize a human gut microbiota samples with shotgun sequencing, while so called *shallow* shotgun metagenomics typically deals with fewer than 1M reads/sample, drastically reducing sequencing costs. This is made possible

by assembly-free processing of reads, thus requires an exhaustive genome reference catalog for the studied environment to process mapping [8] [9].

In the present study, we aim to provide new insights towards usage of shallow shotgun metagenomics in clinical research, assessing the reliability of information that can be recovered from shallow shotgun sequencing in comparison with deep sequencing. We first used simulations to design and calibrate filters that efficiently identify organisms genuinely present in the mapping data providing reliable taxonomic profiles at each sequencing depth. We then applied this approach to real data sets, and assessed information recovery at a sample level and a study level. Our results show that some information is lost if we want to obtain reliable profiles (rare taxa are filtered out to avoid having massive identification of spurious taxa), but that structures like differences between groups of patients and diagnosis-like classification are very well conserved using shallow shotgun metagenomics.

2 Material and Methods

2.1 Data

Simulated data sets. We retrieved taxonomic profiles of 19 human gut microbiomes from Qin 2014 [10] through *curatedMetagenomicData* [11], with a complexity of 98 $\pm$ 15 species per sample, and species' relative abundance ranging from 5.10^{-1} down to 10^{-6} (average 10^{-3}). These profiles were given using the NCBI's taxonomy and were translated into UHGG's taxonomy [12] by choosing the UHGG species with the closest taxonomic assignation to the NCBI species (if several species tied, one was chosen randomly), resulting in profiles with the exact same complexity, approximately the same phylogenetic composition and some uncultured organisms (MAGs). UHGG genomes are clustered into "species clusters" (thereafter referred as species), that share at least 95% of identity on 30% of the genomes, and one representative genome is chosen in each cluster for inclusion in the mapping catalog. We generated profiles using either the species representative genome or using a randomly selected genome belonging to the same species. The second scenario introduces noise in the mapping data because of the intra-species diversity and corresponds to the more realistic case where the genome is not necessarily in the database used for mapping. It is the one shown in the results. For each sample, we used *Grinder* [13] to simulate 10M paired end reads (length of 2*125bp, insert size normally distributed with an average of 500bp and standard deviation of 50 bp without sequencing errors) and subsampled at 5 M, 1 M, 500 K, 100 K, 50 K and 10 K reads/sample.

Real data set. We used data from 3 clinical studies, for a total of N=439 samples covering patients from several continents and clinical conditions (healthy patients, hepatic diseases at different stages, cancer patients). Loomba *et al.* (2017) [14] compares the gut microbiota of patients suffering from hepatic diseases at different stages (fibrosis vs NAFLD). Matson *et al.* (2018) [15] compares, among patients having metastatic melanoma, those who responded to anti-PD-1 immunotherapy and those who didn't. Qin *et al.* (2014) [10] compares patients having liver cirrhosis and a group of healthy controls, with a discovery and a validation cohort for both groups. We analyzed these data sets at full depth and subsampled them to mimic shallow sequencing in the remainder.

2.2 Bioinformatics pipeline

Reads were pre-processed using *trimmomatic* [16], removing low quality reads and reads shorter than 80 nucleotides. For real data sets, reads were also mapped to human genome to filter out host contamination. Remaining reads were then mapped to UHGG catalog [12] using *bwa mem* [17] local aligner. We also used *bwa aln*, and *bowtie2* [18] in its *end-to-end* and local settings for comparison in the simulated data, but we only present results for *bwa mem* as it resulted in a better trade-off between overall mapping rate, false positive and multi-mapping rate.

Multi-mapping (*ie* reads that map to several genomes) occurs frequently when mapping shotgun metagenomics reads to a catalog of reference genomes (26% and 42% of the mapped reads in simulated and real data sets respectively), due to highly conserved genes and mobile elements notably. Thus, we split mapped reads into unambiguous reads that mapped to one genome only, and other reads. For each genome identified, we retrieved the reads count (RC) and the fraction of the genome covered

(FC) by at least one read, using either all reads or unambiguous reads only (uRC and uFC), as well as a specificity ratio (SR) defined by the number of unambiguous reads divided by the total number of reads mapped to this genome.

In order to estimate species' relative abundances, we first compute the representative genomes' average coverage $C_s = \frac{1}{\ell_s} \sum_i r_{i,s}$, with ℓ_s being the length of the representative genome of species s and $r_{i,s}$ the length of read i that is unambiguously mapped to s, and then we obtain the relative abundance by normalizing across species to sum to 1 : $A_s = \frac{C_s}{\sum_j C_j}$. We refine this estimation by reallocating the ambiguous reads by randomly assigning them to one of their hits, with a probability proportional to the previously computed relative abundances.

2.3 Simulations analysis

Direct mapping of short reads on reference genomes produces false positives (genomes covered by reads but not present) that need to be filtered out. We used simulated profiles, with known composition, to determine the most efficient way to classify the genomes into true positives (TP) and false positives (FP). In order to assess methods and compare them to each other, we computed the area under the receiver operating characteristic (ROC) curve (AUC) for this classification task, using *evabic* R package. We also implemented an automated threshold search, that allows for a false discovery rate ($FDR = \frac{FP}{TP+FP}$) of at most 10%, and compared false negative (FN) rates at this threshold across methods and sequencing depths.

We first evaluated how genomes features (RC, uRC, FC, uFC and SR) can be used independently to classify the genomes, and then combined them to train classifiers. We used logistic regression, linear discriminant analysis (LDA) and random forests (RF), to perform classification, with uRC, uFC, SR and total sequencing depth as input features. Finally, we used a 4-fold cross validation process to evaluate the performance of these methods and determine suitable thresholds for each method and sequencing depths.

2.4 Real data sets analysis

We analyzed real data set using (1) RF-based filters fitted on the simulations data and thresholds that control FDR at each sequencing depth, and (2) a basic filtering that discards all species with a relative abundance beyond 10^{-4}, FC beyond 10^{-2} or uFC beyond 10^{-4}. This filtering is inspired by what was done in [8] and corresponds to currently used methods with a quite permissive threshold due to low sequencing depths on which it will be applied.

We evaluated α-diversity using species richness and Shannon diversity, and β-diversity using Jaccard distance and Bray-Curtis dissimilarity index using *phyloseq* [19]. In order to assess the impact of sequencing depth on taxonomic profiles, we evaluated the correlation between subsampled and deep α-diversity measures using Spearman and Pearson correlation as well as the correlation between species relative abundance at full depth and shallower depths. We also measured the distance between low depth samples and their full depth counterpart. Finally, for each data set, we evaluated the differences between groups of interest, according to the sequencing depth, at different levels:

— differences in α-diversity between groups, through a Wilcoxon test using different metrics described earlier,
— structure in the β-diversity matrix, through a PERMANOVA analysis using different metrics described earlier,
— biomarker discovery using a Wilcoxon test with Benjamini-Hochberg correction that shows differentially abundant species,
— patients' classification in their group of interest, using random forests trained on taxonomic profiles, with a feature selection step as performed in [14].

In order to perform unbiased comparison of p-values and AUCs across sequencing depths, we used only samples having at least 10M high quality reads per sample for Loomba-2017 ($N = 77$), Matson-2018 ($N = 39$) and 5M reads for Qin-2014 ($N = 235$, we reduced maximum sequencing depth to include more patients).

3 Results

3.1 Filters design and performance on simulated data

Our raw mapping data from simulations showed that a small number of reads (8% of unambiguously mapped reads) are mapped to an unexpected genome, resulting in a great number of false positives genomes (FDR = 91% prior to any filter). The basic (threshold-based) filtering technique yielded in overall FNR = 0.39 and FDR = 0.45. We sought optimal thresholds on read counts (RC), but it appeared to be sub-optimal (AUC = 0.75). Using genomes' fraction covered (FC) to determine such a threshold was better (AUC = 0.85), and retrieving only unambiguous reads to compute this statistics enhanced the classification (AUC = 0.856 using uRC, and AUC = 0.896 for uFC, see table 1) thus was used for the following. As seen on figure 1A, a threshold based on uFC and/or uRC values, which would correspond to horizontal and/or vertical line to discriminate TPs and FPs, is suboptimal as it would miss the long tail of genomes with low uRC but comparatively high uFC values. These results motivated our attempt to train classifiers able to take benefit from this pattern. To train such classifiers, we used uRC, uFC, SR, as well as sequencing depth to predict genomes' status (present or absent).

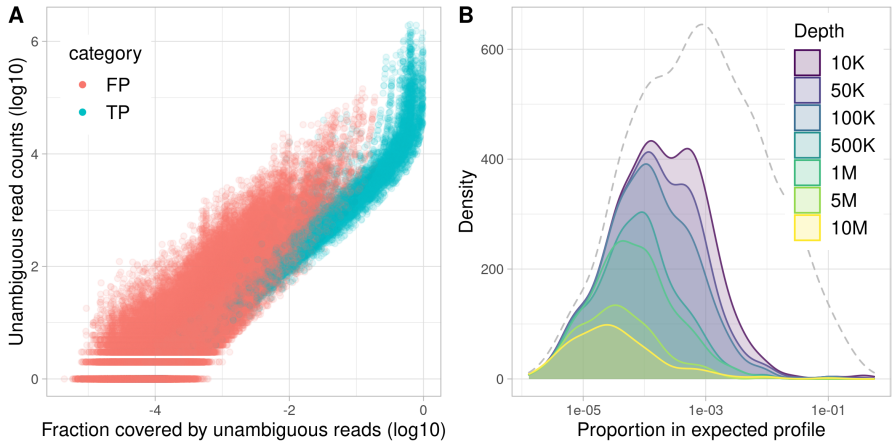

Fig. 1. Simulations results: (A) Unambigous fraction covered (uFC) and unambiguous read counts (uRC) of genomes present in the expected profiles (TPs, blue points) or absent (FPs, red points). (B) Distribution of FN species according to their relative abundances in the expected profiles, using RF-based filters with a 10% FDR on the testing set of cross validation. The dot line represents the distribution of all expected species.

method	AUC		FN rate at threshold	
	training	testing	training	testing
uRC	0.856		0.883	
uFC	0.896		0.580	
LDA	0.944 ± 0.002	0.944 ± 0.006	0.391 ± 0.010	0.391 ± 0.025
Logistic regression	0.955 ± 0.002	0.955 ± 0.006	0.386 ± 0.012	0.387 ± 0.032
Random forest	0.999 ± 0.0001	0.969 ± 0.007	0.017 ± 0.002	0.291 ± 0.040

Tab. 1. Area under ROC curves and false negative rates when threshold is set to tolerate 10% FDR. For machine learning based methods, these measures are split into training and testing sets, using a 4-fold cross validation.

We can see on table 1 that sophisticated classifiers largely outperform basic filtering. LDA and logistic regression perform similarly, and yield nearly identical results in training and testing in the cross validation process, highlighting very good generalization capabilities. RF appeared to be the best method, yielding in a nearly perfect classification in training set, and still better than others in the testing sets; RF will thus be used in the remainder. When setting a threshold that control the FDR at 10%, we can see the interest of refining this classification, to drastically lower the FN rate, although it remains quite high in the test samples.

We further characterized the information loss in the context of shallow sequencing metagenomics by plotting the distribution of expected relative abundances of species that were absent in profiles with respect to the sequencing depth, as seen on figure 1B, using RF-based filtering. We can see clearly the inflation of FN while lowering sequencing depth, but we can also notice that, as expected, the populations that are lost are relatively rare. For instance, at 500K reads/sample, all populations with relative abundance greater than 10^{-2} were detected.

While focusing on TPs, we noticed that Pearson correlation between expected and estimated species relative abundances went up from $\rho = 0.54$ to $\rho = 0.60$ by reallocating ambiguous reads.

3.2 Performance on real data sets for taxonomic profiling

Applying the RF-based filters on real data sets resulted in high quality profiles, with an average diversity of 128 ± 66 species per sample at full depth, which gradually decreased with sequencing depth, down to 45 ± 21 at 500K reads/sample for example (fig 2A). In comparison, basic filtering were more permissive, producing profiles with increased diversity and more resilient towards reduced sequencing depths (fig 2D). The Shannon diversity index (fig 2B,E) was much less impacted by sequencing depth, indicating that the species lost at low sequencing depth were mostly rare ones. Down to 500K reads per sample, the correlation between full depth and subsampled Shannon indices was nearly perfect using basic filtering and remained very high with RF-based filters. Distances between subsamples and their reference, defined as the corresponding sample at full depth, gradually augmented when decreasing the sequencing depth, and these distances were much more important using RF-based filters than basic filters (fig 2C,F, both graphs share the same scale), and showed high replicability across data sets.

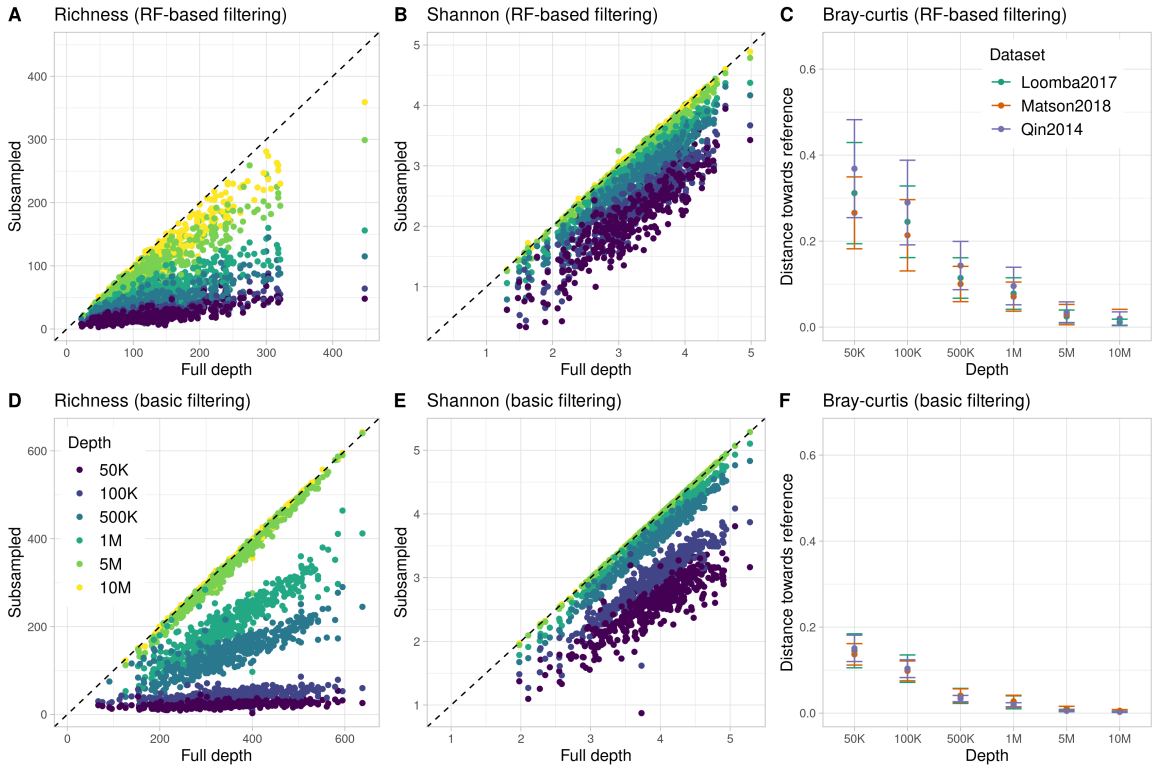

Fig. 2. Comparison between full depth and subsampled profiles for samples from the 3 data sets considered : richness observed, Shannon diversity and Bray-Curtis distance between subsampled data and reference (full depth data) using RF-based filtering (A, B and C respectively) and basic filtering (D, E and F respectively)

Comparison between filtering strategies highlighted that filtering plays a key role while dealing with shallow metagenomics data. If taxonomic profiles were less impacted by sequencing depth using basic filters than RF-based filters, we know according to our simulations that lots of false positives are present

in the profiles, which introduces an important noise and jeopardizes the biological interpretation of results.

3.3 Performance on real data sets for patients stratification

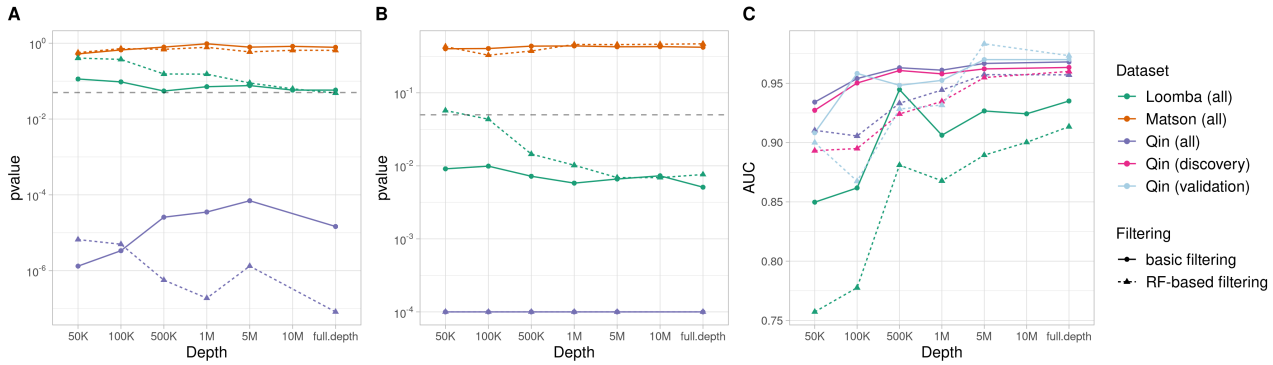

Fig. 3. Differences between patients groups in different studies : significance of inter-group difference regarding Shannon diversity index (A), PERMANOVA analysis (B). AUC corresponding to random forest classification (C) was performed in Loomba-2017 and Qin-2014, with a split between discovery and validation cohorts in Qin-2014 as performed on the original paper of this study.

Here, we assess the robustness of biological signal found in the different data sets towards sequencing depth. As expected according to previous results, differences in α-diversity between groups were maintained using shallow sequencing : p-values were concordant across sequencing depths (see Fig. 3A), with a strong difference in Qin-2014 data set, a slight difference between groups that is not significant in Loomba-2017 and no differences between groups in Matson-2018. PERMANOVA analysis led to similar results (Fig. 3B), showing that the structure of the matrix distances between samples is very marginally impacted by sequencing depths. The p-value regarding Loomba-2017 on RF-based filtered data increased, but stayed significant, even under 1M reads/sample, traducing the absence of key populations at such sequencing depths. Out of the 7 differentially abundant taxa in Loomba-2017 with RF-based filtering found at full depth ($FDR < 0.1$), 5 taxa were still identified at 1M reads/sample and 4 at 500K reads/sample. Basic filtering, allowing more taxa in the profiles, identified more differentially abundant taxa (19 at full depth, 12 at 500K reads/sample) but the reliability of these taxa is questionable. As previously discussed, signal was much more important in Qin-2014 data set: 25 differentially abundant taxa were identified at full depth with RF-based filtering ($FDR < 0.05$ in both discovery and validation cohorts), 13 taxa at 1M reads/sample and 9 at 500K reads/sample. Again, basic filtering allowed to identify more taxa (124 at full depth, 72 at 500K reads/sample). Finally, classification of patients using RF was performed in Loomba-2017 and Qin-2014 (see Fig. 3C). In Loomba-2017, we could perform a better classification using basic filtering, with an AUC similar to full depth AUC down to 500K reads/sample, while it gradually decreased as sequencing depth decrease using RF-based filters, due to some key taxa for the classification being lost. On Qin-2014 data set, we could perform a very good classification on both discovery and validation cohorts even at low sequencing depth, with performance very stable using basic filtering down to 100K reads/sample using basic filtering and that gradually decreased with RF-based filters under 5M reads/sample.

4 Discussion

Direct mapping of reads on catalogs of reference genomes was previously suggested as the most suitable way to build taxonomic profiles from shallow sequencing metagenomics data [7] [9], as it produced nearly identical taxonomic profiles across sequencing depths. Our simulations highlighted the need to refine filters on genomes identified by such mapping, and to perform depth dependent thresholds to obtain reliable profiles at each sequencing depth. This step is crucial to prevent misleading interpretations and to provide trustful biological knowledge. Controlling the FDR in the taxonomic profiles had the direct consequence of decreasing the number of identified species, especially at low sequencing depth. The benefit of RF-based filters, and to a lesser extent other machine learning based

models tested, over simple filtering based on species features independently (RC, uRC, FC and uFC) was remarkable, allowing to identify more species and rarer ones for equivalent FDR. The application of such techniques, as they rely on a learning step, is by definition limited to the training conditions. In our case, usage of our RF-based model to filter genomes should be limited to ecosystems with similar complexity, sequenced with short reads at depth included in the range used for the training and mapped to a catalog similar to representative genomes of UHGG in terms of completeness, intra and inter species diversity.

On the three real data sets considered, our analysis showed that differences between groups of patients observed at full depth where still recovered at low sequencing depth. Permissive and depth-independent filtering, as performed in previously published papers on shallow shotgun metagenomics, allowed a little improvement in structure recovery than our stringent RF-based filters: these structures were less sensible to the noise introduced by FPs in the profiles using basic filtering, than to the removal of key species induced by our stringent RF-based filter.

Overall, our results show that (1) one needs to perform stringent and depth-dependent filters to obtain reliable profiles in shallow sequencing data, (2) resulting taxonomic profiles are limited to most abundant taxa in shallow sequencing context, and (3) shallow shotgun metagenomics can be a suitable approach to perform diagnosis-like classification of patients even if further investigations should be led to assess generalization capability and interpretability of signatures obtained with shallow sequencing.

Shallow shotgun metagenomics requires an exhaustive reference database regarding the studied ecosystem to build taxonomic profiles. Although it produced reduced complexity profiles if we want to ensure reliability of results, it appeared to be a very good alternative for clinical studies, and sufficient to classify patients, when discrimination between groups is expected to be important and to rely on relatively dominant taxa. Therefore, it can be profitable in such cases to favour the number of patients included or to introduce a longitudinal aspect, rather than per sample sequencing depth. For other body sites (vaginal, oral or skin microbiota) host contamination should be taken into account when determining sequencing depth, as host reads will be discarded. Shallow shotgun metagenomics could also be used to perform functional analysis, for example for coarse grain identification of family of genes (like KOs), as the sequencing depth could not allow the identification of very specific genes and SNPs that require assembly, such as antibiotic resistance.

Acknowledgements

We are grateful to the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi: 10.15454/1.5572390655343293E12) for providing help, computing and storage resources, as well as CRCM's DISC platform. This work was financially supported by ANRT and Laboratoire Alphabio thanks to a CIFRE scholarship.

References

[1] Yong Fan and Oluf Pedersen. Gut microbiota in human metabolic health and disease. *Nature Reviews Microbiology*, 19(1):55–71, January 2021.

[2] Sunny H. Wong and Jun Yu. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nature Reviews Gastroenterology & Hepatology*, 16(11):690–704, November 2019.

[3] Jose C Clemente, Julia Manasson, and Jose U Scher. The role of the gut microbiome in systemic inflammatory disease. *BMJ*, page j5145, January 2018.

[4] Donovan H. Parks, Christian Rinke, Maria Chuvochina, Pierre-Alain Chaumeil, Ben J. Woodcroft, Paul N. Evans, Philip Hugenholtz, and Gene W. Tyson. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, 2(11):1533–1542, November 2017.

[5] J. Grembi *et al.* Gut microbiota plasticity is correlated with sustained weight loss on a low-carb or low-fat dietary intervention. *Sci Rep*, 10(1):1405, December 2020.

[6] Christopher Quince, Alan W Walker, Jared T Simpson, Nicholas J Loman, and Nicola Segata. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9):833–844, September 2017.

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

[8] Tasha M Santiago-Rodriguez, Aaron Garoutte, Emmase Adams, Waleed Nasser, Matthew C Ross, Alex La Reau, Zachariah Henseler, Tonya Ward, Dan Knights, Joseph F Petrosino, and Emily B Hollister. Metage-

nomic Information Recovery from Human Stool Samples Is Influenced by Sequencing Depth and Profiling Method. page 17, 2020.

[9] Federica Cattonaro, Alessandro Spadotto, Slobodanka Radovic, and Fabio Marroni. Do you cov me? Effect of coverage reduction on metagenome shotgun sequencing studies. *F1000 Research*, 7:1767, 2020.

[10] 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.

[11] Edoardo Pasolli, Lucas Schiffer, Paolo Manghi, Audrey Renson, Valerie Obenchain, Duy Tin Truong, Francesco Beghini, Faizan Malik, Marcel Ramos, Jennifer B Dowd, Curtis Huttenhower, Martin Morgan, Nicola Segata, and Levi Waldron. Accessible, curated metagenomic data through ExperimentHub. page 4, 2018.

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

[13] Florent E. Angly, Dana Willner, Forest Rohwer, Philip Hugenholtz, and Gene W. Tyson. Grinder: a versatile amplicon and shotgun sequence simulator. *Nucleic Acids Research*, 40(12):e94–e94, July 2012.

[14] 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.

[15] 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.

[16] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, August 2014.

[17] H. Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio]*, May 2013. arXiv: 1303.3997.

[18] B. Langmead *et al.* Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4):357–359, April 2012. Number: 4 Publisher: Nature Publishing Group.

[19] Paul J. McMurdie and Susan Holmes. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, 8(4):e61217, April 2013.