**PhD project of Inf-HOLOBIONT ANR (PRT-S)**

Pr Laurence Delhaes – CHU de Bordeaux et INSERM U1045

Co-encadrement avec Pr Olivier Barraud – CHU de Limoges et INSERM U1092

**ANR grant for a PhD October 2023-2026; contact: *laurence.delhaes@u-bordeaux.fr***

***Context:*** Pneumonia, especially ventilator-associated pneumonia (VAP), remains a leading cause of infection in critically ill patients1 with an increased risk of death in ICU patients leading to a high attributable mortality that we need to improve. In fact, infections are frequent among ICU patients, with an estimated prevalence of suspected or proven infections of 54% in a cross-sectional worldwide study in 2017, nearly 50% of infected patients receiving mechanical ventilation (MV)1. The site of infection is mainly the respiratory tract (60%), followed by the abdomen (18%), and bloodstream (15%). After microbiological cultures, Gram-negative microorganisms are identified in 67% of cases, Gram-positive microorganisms in 37%, and fungal microorganisms in 16% mainly in patients receiving MV, and they are responsible for VAP1. In addition, VAP treatment is known to be responsible for more than half of the antibiotic consumption in ICU. The delay in microorganism identification of respiratory samples by standard methods (direct examination (DE) and microbial cultures) is the main factor explaining those antibiotics overconsumption, associated to increased microbial resistance1.

In this context, we aim at developing a rapid molecular diagnostic tool: **Inf-HOLOBIONT allowing to evaluate the 3 components (infection/pathogen occurrence, microbial dysbiosis, and immune host response) simultaneously from the same respiratory sample** (endotracheal aspirate (ETA) or bronchoalveolar lavage (BAL)), by following a holistic deep-sequencing (NGS) approach by Oxford Nanopore Technologies (ONT), associated with a unique bioinformatics analysis.

Since it relies on unbiased, non-targeted deep-sequencing, this approach may improve the management of severe ICU patients, for whom the delay to diagnose infection is known to worsen the prognosis. A similar metagenomics approach has been recently proposed with an optimal turnaround time of 5-6h2,3, while it may take 24 to 48h using conventional microbiology tools. Nevertheless, these studies only focused on infection diagnosis and did not assess microbial dysbiosis nor immune host response which could provide crucial information about the patients’ condition4. In addition to optimize the sample treatment, we aim at setting up the most efficient bioinformatic work-flow that will allow us well conduce the Inf-HOLOBIONT project.

***PhD objective:*** Build an efficient work-flow to perform the bioinformatics analysis associated with Inf-HOLOBIONT project (i.e. analysis of the raw data obtained from metagenomics ONT, using modified read-until approach). In parallel, several microbiology experimentations will be conducted to handle and test/validate this work-flow and approach.

***Methods:***The metagenomics analysis will be based on recent developments of the ONT community; namely:

(i) the "UNCALLED" approach (for Utility for Nanopore Current ALignment to Large Expanses of DNA), (ii) the use of the WIMP workflow allowing real-time assignment of any microbial cDNA sequence (bacterial, viral, fungal or archaeal) combined with (iii) the "Read Until" function which allows selective sequencing on ONT (by bioinformatics enrichment or depletion)5,6.

We propose to adapt the raw signal mapping algorithm published by Kovaka et al.5, which allows to adjust the sampling of the fully sequenced dsDNA. We will favor a strategy of depletion of the dsDNA strands corresponding to the human genome by creating a "human response" database, in order to separate the non-human sequences (microbiological ones) that will be analyzed via the WIMP tool to document any dysbiosis, and via antibiotic resistance gene databases to identify the corresponding resistance profile (ARG databases for review see7). These data will be compared to results obtained by conventional approaches to validate the Inf-HOLOBIONT project (i.e. a rapid diagnostic tool particularly useful to face VAP, but also easily adaptable to any other infectious clinical situation if needed).

Recently, Bordeaux Transcriptome Genome Platform - PGTB’s team (scientific coordination: O. Lepais & L. Delhaes) has set up experiments to locally validate the Read Until program by ONT, and confirmed that less than 2% of microbial DNA in an initial mixture are enriched up to 98% of the reads sequenced by ONT using the Read-Until program.

The major risk is to establish a bioinformatics pipeline able to produce results which will be difficult to interpret in terms of presence or absence of pathogens (with the corresponding antimicrobial susceptibility profile), of existence of any dysbiosis and of the host immune response quality. To overcome this pitfall, we started to explore the key point to make a decision on the software to use, and to benchmark existing pipelines (Table 1). After a collective discussion with bioinformatic colleagues, we compared the 3 methods available to perform dry-lab depletion on ONT analysis. Their advantages and disadvantages are summarized in table 1. These methods will be compared, in order to select the most efficient protocol in the context of Inf-HOLOBIONT.

In parallel, several microbial experiments will be conducted during the PhD: First, artificial samples composed of human bronchial epithelial cells (BEAS2B) inoculated with either bacteria, fungi and/or viruses will be used to mimic real respiratory sample from ICU patients and to validate each step of our process.

Secondly, a set of documented samples retrospectively selected will be used to confirm the efficiency of Inf-HOLOBIONT tool in a real clinical context.

**Table 1:** Comparison of the 3 methods available to perform dry-lab depletion on ONT analysis that could be used

|  |  |  |
| --- | --- | --- |
| **Softwares** | **Advantages** | **Disadvantages** |
| Adaptative sampling (Read-Until) | * Integrated software
* Reference file not too big = fast decision making so a lot more read are going through the pores so the enrichment is better (1)
 | * Using CPU instead of GPU: the reference file is less big
* Usage of minimap2 for read classification is not optimal (5)
* Marquet et al. observed 25% of human reads could not accurately be rejected when they tried to deplete all human host reads from vaginal samples with ONT’s adaptive sampling option (5)
 |
| Read Fish | * GPU basecalling is faster (3)
* Select specific human chromosomes, illustrating gigabase references file is not a constraint (3)
* These methods can be used to efficiently screen any target panel of genes without specialised sample preparation (3)
* Easy to test our configuration and out selcetive sequencing thanks to the Github (4)
* This approach works on any device capable of real time base calling
* Open source with an installable via PyPI
* used an updated version of Read Until API required for Python3 compatibility
 | * Added software (1)
* Needs a specific and powerfull GPU to work (3)
* Requires Guppy version 3.4.5. and MinKNOW version core 3.6 and LINUX OS (Old version compatible with MinKNOW 4.0 and Guppy 4) (4)
* Needs an additional base-calling server for real-time basecalling (5)
* Usage of minimap2 for read classification is not optimal bc target reads were correctly classified at only 83% (5)
 |
| Read Bouncer | * Facilitates both GPU base-calling with ONTs Guppy as well as CPU base-calling with DeepNano-blitz
* Improved read classification
* Smallest reference sequence index size and peak memory usage
* Make faster and more reliable rejection decisions than Readfish and MinKNOW.
* For Windows and Linux with an open source
 | * Added software(1)
 |

Table references are as follows: (1) Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02582-x>; (2) The Reliability of Metagenome-Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample: <https://journals.asm.org/doi/10.1128/AEM.02593-20>; (3) Readfish enables targeted nanopore sequencing of gigabase-sized genomes (2020 Nov 30): <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7610616/>; (4) <https://github.com/looselab/readfish> (2020, latest release README 9 months ago); (5) ReadBouncer: Precise and Scalable Adaptive Sampling for Nanopore Sequencing (February 02, 2022.): https://www.biorxiv.org/content/10.1101/2022.02.01.478636v1

***Working environment:*** *The project is headed by Pr Laurence Delhaes, under U1045 Inserm group and the CHU Parasitology-Mycology department; with collaboration with the PGTB (Plateform Genome Transcriptome de Bordeaux (C. Boury, B Penaud, E Guichoux et O Lepais)*, and the bioinformatician group of CHU (L Gaston*).* Pr Olivier Barraud will co-direct this PhD and a Bioinformatic engineer will be recruited during this ANR in relation with the bioinformatician group of CHU (L Gaston*).*

Several preliminary analyses have been done during the last 2 years and will be shared with the PhD student.

****

**Figure 1:** Main microbial approaches from conventional to NGS approaches and proposed workflow for ONT long-read-based sequencing of Inf-HOLOBIONT tool (green boxes). The technical points identified as key to the success of Inf-HOLOBIONT project are shown in yellow boxes; cited references are listed in the bibliography section.

***Bibliography***

**1.** Vincent JL, Sakr Y, Singer M, et al. Prevalence and Outcomes of Infection Among Patients in Intensive Care Units in 2017. JAMA. 2020;323(15):1478–1487. https//jamanetwork.com/article.aspx? doi=10.1001/jama.2020.2717.

**2.** Charalampous T., Kay G.L., Richardson H., et al. Livermore & Justin O’Grady. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol. 2019;37(7):783-792. <https://doi.org/10.1038/s41587-019-0156-5>.

**3.** Wu N, Ranjan P, Tao C, et al. Rapid identification of pathogens associated with ventilator-associated pneumonia by Nanopore sequencing. Respir Res. 2021 Dec 10;22(1):310. <https://doi.org/10.1186/s12931-021-01909-3>

**4.** Dickson RP, Schultz MJ, van der Poll T, et al. Lung Microbiota Predict Clinical Outcomes in Critically Ill Patients. Am J Respir Crit Care Med. 2020;201(5):555-563. <https://doi.org/10.1164/rccm.201907-1487oc>

**5.** Kovaka S, Fan Y, Ni B, Timp W, Schatz MC.Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. <https://doi.org/10.1038/s41587-020-0731-9>

**6.** Payne A, Holmes N, Clarke T, et al. Nanopore adaptive sequencing for mixed samples, whole exome capture and targeted panels. bioRxiv 2020.02.03.926956; Published in Nature Biotechnology <http://dx.doi.org/10.1038/s41587-020-00746-x>

**7.** d'Humières C, Salmona M, Dellière S, et al. The Potential Role of Clinical Metagenomics in Infectious Diseases: Therapeutic Perspectives. Drugs. 2021; 81:1453-66. <https://doi.org/10.1007/s40265-021-01572-4>

**8.** ReadBouncer: Precise and Scalable Adaptive Sampling for Nanopore Sequencing (February 02, 2022.) [https://www.biorxiv.org/content/10.1101/2022.02.01.478636v1](https://www.biorxiv.org/content/10.1101/2022.02.01.478636v1%20%20www.biorxiv.org/content/10.1101/2022.02.01.478636v1)