3iA Côte d'Azur
Interdisciplinary Institute for Artificial Intelligence

Postdoctoral research topic

- Title of the proposed topic: **Sequencing wasted waters to explore the natural history of SARS-CoV2**
- Research axis of the 3IA: Theme 3 "AI for Computational Biology and Bio-Inspired AI"
- Supervisor: Pascal Barbry, barbry@ipmc.cnrs.fr
- Potential co-supervisor (name, affiliation): INRIA
- The laboratory and/or research group: IPMC, "Physiological Genomics Group", Dr. Pascal Barbry.

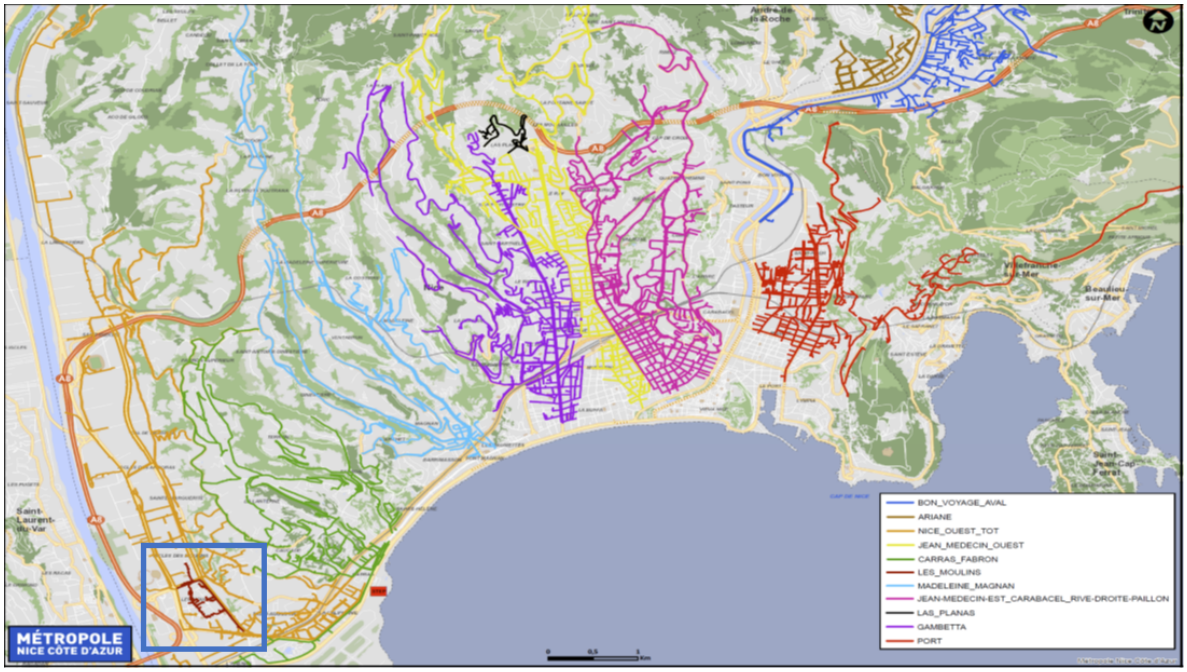
Apply by sending an email directly to the supervisor.
The application will include:

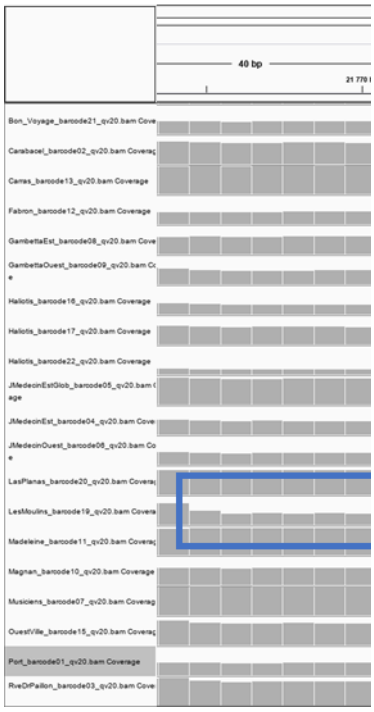
- Letter of recommendation of the supervisor indicated above
- Curriculum vitæ including the list of the scientific publications
- Motivation letter
- Letter of recommendation of the thesis supervisor

- **Description of the topic:**

We are currently looking for a postdoctoral computational biologist to develop innovative data mining approaches to explore and model SARS-CoV2 after RNA sequencing of waste waters, with the aim to set up robust metagenomic analyses at different geographic scales. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the virus that causes coronavirus disease 2019 (COVID-19). The World Health Organization declared the outbreak a Public Health Emergency of International Concern on 30 January 2020, and a pandemic on 11 March 2020.

This project has started in my laboratory in September 2020 after discussions with Nice Metropole. After setting up sampling and RNA extraction, we developed a nanopore sequencing technology that provides reliable results in less than three working days. The approach allows to decipher the whole virus genome and identify its key mutations. We were able in January 2021 to perform the first identification of the SARS-CoV-2 lineage B.1.1.7 in French waste waters. This lineage is now rapidly spreading in the European population. The emergence of new viral strains illustrates well the urgent need for developing innovative genomic and epidemiological surveillance with tools that treat sequence information. Our genomic analysis of SARS-CoV-2 was performed on raw wastewater samples that are collected the 21st of January in 20 different areas of the city of Nice. While the B.1.1.7 lineage was not detected in last December, we detected its expression in the area of "Les Moulins" (Figure 1). Our laboratory is now following since 4 months the Nice area, in close association with Nice Metropole, and we are also processing additional samples coming from representative French treatment plants, in collaboration with VEOLIA. The two series of data provide very complementary information that will be used to develop multiscale epidemiological models of virus spreading. Development of systematic waste water testing can represent an important part of future national pandemic response plans. Testing, both for viral load and mutational status can be performed cost-effectively. Implementation of such a survey instrument could facilitate "re-opening" of public spaces and safe return of workforces in regions where SARS-CoV-2 is already widespread and to track and contain emergence/re-emergence where the prevalence of the virus is lower.





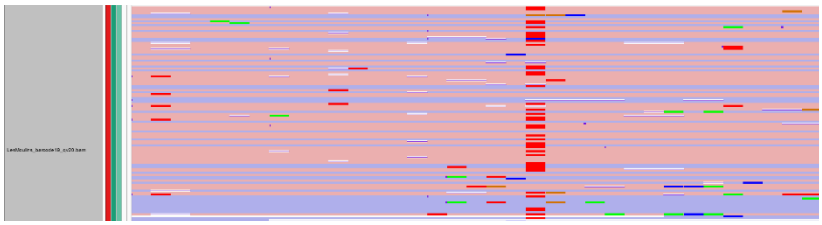


Figure 1. (Top) *Map of the 20 areas that are currently analyzed for the Nice Metropole. Les Moulins, the area in which the B.1.1.7 strain was first identified, is indicated by a blue box.* ***(Right)*** *Illustration of the SARS-CoV2 sequence around nucleotides 217665-21770, which display a deletion in B.1.1.7 (blue box). About 50% of the corresponding reads contains the deletion, suggesting a high prevalence of the strain in this specific area. The mutant has also been detected in the Nice Ouest area and in Haliotis (water plant), as expected from the organization of the network.* ***(Left bottom)*** *The mutation was independently confirmed by identifications of additional mutations, such as N501Y of the spike protein in the same samples, at similar frequency.*

The proposed approach is complementary to individual clinical tests, because it provides access to a whole population that lives in a geographic area, independent of the number of testing that are performed in this zone. It comes also with some specificities. For instance, the template contains many copies of viruses that are derived from different individuals. Thus, a mixture of genotypes is measured in parallel, for which specific metagenomic analyses need to be developed. The postdoc will develop a quantitative to integrate metagenomic information with metadata of interest. This part will be performed in collaboration with the Public Health Department of the Nice University Hospital. At the end of this project, we anticipate the construction of maps recapitulating the circulation of the major SARS-CoV-2 mutations across a large territory, which could contribute to the deployment of innovative

surveillance for the different metropoles. Applied at different scales, from a neighborhood to a metropole or a region, the approach will enable communities to reinforce their public health arsenal besides testing of individuals. This could strengthen local screening strategies in particular areas where certain variants are identified.

My laboratory has already worked on SARS-CoV2, contributing in 2020 to the identification of its receptor cells in the human airways (Sungnak, Ziegler). This project was performed in the context of the HCA Lung Biological Network.

The ideal candidate would have a background in computational/bioinformatics approaches in metagenomics and/or microbiology. Capacity to interact directly with biologists is important to develop AI-based tools that can be closely developed with them and be transferred to a larger audeince. Datasets already exists but additional experiments will be generated in collaboration with the biological laboratory during the course of this postdoctoral fellowship.

Background:

1. W Sungnak, N Huang, C Bécavin, M Berg, R Queen, M Litvinukova, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes 2020. Nature medicine 26 (5), 681-687
2. CGK Ziegler, SJ Allon, SK Nyquist, IM Mbano, VN Miao, CN Tzouanas, et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues 2020. Cell 181 (5), 1016-1035. e19
3. Fassy J, Lacoux C, Leroy S, Noussair L, Hubac S, Degoutte A, Vassaux G, Leclercq V, Rouquié D, Marquette CH, Rottman M, Touron P, Lemoine Corbel A, Herrmann JL, Barbry P, Nahon JL, Zaragosi LE, Mari B. Versatile and flexible microfluidic qPCR test for high-throughput SARS-CoV-2 and cellular response detection in nasopharyngeal swab samples. MedRxiv. 2020. https://doi.org/10.1101/2020.11.09.20228437
4. Pettit SD, Jerome KR, Rouquié D, Mari B, Barbry P, Kanda Y, Matsumoto M, Hester S, Wehmas L, Botten JW, Bruce EA. 'All In': a pragmatic framework for COVID-19 testing and action on a global scale. EMBO Mol Med. 2020 12:e12634. https://doi.org/10.15252/emmm.202012634.