



Title : Prediction of antimicrobial resistance in metagenomics

Research project within an international company ; no PhD is foreseen.

Advisor: Magali Jaillard Dancette, magali.dancette@biomerieux.com , +33 4 78 87 2000

Laboratory : bioMérieux company, Big Data Analytics team, trainee located at Marcy l'Etoile (West suburbs of Lyon)

Context: The global acceleration and spread of antimicrobial resistance has been identified as one of the most serious threats to global health by the World Health Organization (WHO). In this context, predicting the antibiotic resistance status from a patient's flora sample (*microbiome*), without requiring to first isolate the pathogen strain through a culture step, is a critical clinical challenge: it would both save valuable time to adapt the patient's antibiotic therapy and contribute to limiting the spread of antibiotic resistance.

Objective : The internship aims at setting up methods and tools for predicting the antibiotic resistance status of pathogens present in a microbiome from (meta)genomic information obtained by next generation sequencing, using supervised machine learning approaches. Prediction models learned from isolated pathogens are already available in the literature (e.g, [1]) and will serve as the basis for the work. The first step will be to assess if these models can easily be applied to metagenomes. The next step will be the development of an alternative strategy aiming to learn novel prediction models from metagenomic samples.

You will work with classification algorithms, as well as with bioinformatics tools dedicated to metagenome analysis (e.g. Kraken, MetaCherchant [2]) and to the simulation of next generation sequencing data. You will have access to the bioMérieux computing cluster and join a team composed of data scientists, statisticians and bioinformaticians. Your missions will be to:

- design and implement machine learning strategies based on genomic and metagenomic data;
- create learning and validation metagenome datasets, primarily using simulated data;
- evaluate the performance of the models built and propose improvements based on these assessments.

Skills required : We are looking for a Master student in statistics, machine learning or data science with a great interest for biological applications, or in bioinformatics with good skills in statistics, machine learning or data science. The trainee will read and produce documents in English and will use python and R statistical language and scripting tools for data processing.

Bibliography :

[1] Nguyen, M. *et al.* (2018). **Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae***. Scientific reports, 8(1), page 421.

[2] Olekhnovich, E.I., *et al.* (2017). **MetaCherchant: analyzing genomic context of antibiotic resistance genes in gut microbiota**. Bioinformatics, 34(3), 434--444.

Contact : magali.dancette@biomerieux.com. Application on bioMérieux recruitment platform at <http://biomerieux-recruitment.com>, using job reference : 53849.