Title of the thesis : Machine learning and graph-based techniques to predict long-term bacterial community structure

Location: Lorraine Research Laboratory in Computer Science and its Applications (LORIA), Nancy, France

Project-team: CAPSID and ORPAILLEUR, LORIA

We propose a fully funded 3-years PhD position in computer science with application to biomolecule analysis. The proposed position is funded by the Lorraine Université d'Excellence (LUE) through a multidisciplinary project that involves 2 researchers in computer science and 4 researchers in microbiology. A PhD thesis in microbiology will be conducted in parallel.

Context and motivations

Bacteriocins are antimicrobial peptides of bacterial origin with a very high economic potential in the agri-food sector. They are used in biopreservation/biocontrol applications to fight against undesirable microorganisms in the agronomy and food industry. LIBio has recently developed a technology based on the selection of two strains of the lactic acid bacterium *Carnobacterium maltaromaticum* producing anti-*Listeria monocytogenes* bacteriocins **[9]**. These strains inhibit the growth of this pathogen in cheeses when added to the manufacturing milk to produce the antimicrobial agents in the cheese matrix. These remarkable properties have led to a patent **[10]** that has very recently been licensed to a ferment producer. However, like the vast majority of biopreservation technologies, the effect is at best bacteriostatic: there is little or no decay of the pathogenic bacteria, which can then be maintained at low concentrations in the food. The biopreservation technologies described in the literature are based on engineering approaches that do not take advantage of the properties of the microbial communities forming the microbiomes of food products. Yet microbiome engineering is among the 12 promising technologies that could transform food systems over the next decade **[11]**. Indeed, in the case of biopreservation, assemblies of microorganisms could allow obtaining communities producing multiple antimicrobial agents and moreover being able to occupy the ecological niche of the undesirable microorganism to exclude it more efficiently. However, knowledge in the field of microbial community engineering is insufficient to fully exploit their potential. Indeed, due to the complexity of microbial communities, there is no available method to predict microbial community structure based on the knowledge of the ecological properties of microorganisms. Moreover, assembling microorganisms whose properties is to produce antimicrobial agents is a major difficulty because these agents can lead to the mutual exclusion of the microorganisms producing them.

Positioning

In a microbial ecosystem in which members produce antimicrobial substances like bacteriocins, three actors can be considered: the bacteriocin-producing microorganism (P) and the microorganisms sensitive (S) and resistant (R) to this bacteriocin. It was experimentally shown that in simple ecosystems mixing three such actors, all three actors are able to maintain equilibrium **[12]**. In these systems, S is more competitive than R because it does not pay the cost of resistance, and R is more competitive than P because it does not pay the cost of bacteriocin production. This cyclical relationship between P, S, and R is similar to that of the popular game "rock paper scissors" where no one player has an

advantage over the other two: each player can overtake one player and each can be defeated by another. These simple experimental systems suggest that it is possible to implement engineering tools to predict the structure of complex communities based on the interactive properties of microorganisms. Thanks to the emergence of high throughput investigation methods, it is now possible to produce interaction data between large sets of microorganisms and thus reconstitute models of microorganism interaction networks. [6] [7] Lately, Ramia et al. [4] [5] built the interaction network corresponding to 73 *Carnobacterium maltaromaticum* strains. Like previously, the graph is sender-determined and also shows a highly nested structure [7], which means that it is different from a randomly built network with the same number of nodes and edges. The results also show that the competitive interaction network is very dense making *C. maltaromaticum* a very interesting model to develop community engineering approaches producing high performance antimicrobial substances cocktails for the fight against undesirable microorganisms. This project will use the data published in Ramia et al. [5] and will try to provide a rather computer science approach to the study of those **interaction graph properties**.

The **originality of this project** is that it will make it possible to integrate experimental variables describing the properties of interaction between microorganisms for the prediction of community structure which is not possible by existing methods.

Objectives of the thesis

The main goal of the thesis is to use advanced machine learning and graph-based approaches in order to predict the long-term community structure in microbiological ecosystems [3] [1]. Particularly, it aims at providing approaches to deduce diversity directly from the static, inner properties of the interaction graph the entities are involved in. The practical objectives of this **interdisciplinary PhD project**, which will be carried out in collaboration with researchers from the Laboratoire d'Ingénierie des Biomolécules (**LIBio**), are as follows:

1. to study existing research works on the analysis of interaction networks and long-term diversity prediction in bacteria.
2. to propose machine learning and graph-based approaches in order to learn models that are able to predict diversity based on the interaction graphs. In this context, regression methods could be used to learn the relation between the interaction graph properties and the diversity.
3. to study how graph embedding could help in predicting the level of development for each strain. In this context, we aim to study the impact of graph embedding methods on the prediction results. A specific embedding method could be proposed in the context of this project.

Skills and profile:

Required qualification: Candidates must have a master degree in computer science. Good programming skills in a procedural language are essential. Experience of machine learning and graph mining is also desirable but not essential. A strong interest in bioinformatics would also be highly desirable.

Application deadline: June 10, 2023

Additional information:

- **Supervision and contact:**
    - Yannick Toussaint, LORIA, yannick.toussaint@loria.fr
    - Sabeur Aridhi, LORIA, sabeur.aridhi@loria.fr

- Cécile Mangavel, cecile.mangavel@univ-lorraine.fr

- **Duration:** 3 years
- **Starting date:** between Sept. 1st 2023 and Dec. 1st 2023

<u>The required documents for applying are the following :</u>

- CV;

- a motivation letter;

- your degree certificates and transcripts for Bachelor and Master (or the last 5 years if not applicable).

- Master thesis (or equivalent) if it is already completed, or a description of the work in progress, otherwise;

- all your publications, if any (it is not expected that you have any).

- At least one recommendation letter from the person who supervises(d) your Master thesis (or research project or internship); you can also send at most two other recommendation letters. The recommendation letter(s) should be sent directly by their author to the prospective PhD advisor.

All the documents should be sent in at most 2 pdf files; one file should contain the publications, if any, the other file should contain all the other documents. These two files should be sent to:

- Yannick Toussaint, LORIA, yannick.toussaint@loria.fr
- Sabeur Aridhi, LORIA, sabeur.aridhi@loria.fr
- Cécile Mangavel, cecile.mangavel@univ-lorraine.fr

References

[1] Sabeur Aridhi and Engelbert Mephu Nguifo. Big graph mining: Frameworks and techniques. *BigData Res*., 6:1–10, 2016.
[2] Tristan Durey, Sabeur Aridhi, Frédéric Borges, and Yannick Toussaint. Competition by interference: network analysis and simulations. *preprint*, *under submission*, 2023.
[3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org.
[4] Nancy Ramia. Compétition par interférence et diversité génétique à l'échelle intraspécifique chez la bactérie lactique Carnobacterium maltaromaticum. PhD thesis, Université de Lorraine and Libanese University, 2018.
[5] Nancy Ramia, Cécile Mangavel, Claire Gaiani, Aurélie Muller-Gueudin, Samir Taha, Anne-Marie Revol-Junelles, and Frédéric Borges. Nested structure of intraspecific competition network incarnobacterium maltaromaticum. *Nature*, 2020.
[6] Kalin Vetsigian, Rishi Jajoo, and Roy Kishony. Structure and evolution of streptomyces interaction networks in soil and in silico. *PLOS Biology*, 2011.
[7] Eneas Aguirre-von-Wobeser and Gloria Soberón-Chávez and Luis E. Eguiarte and Gabriel Yaxal Ponce-Soto and Mirna Vázquez-Rosas-Landa and Valeria Souza. Two-role model of an interaction network offree-livingy-proteobacteria from an oligotrophic environment. *Environmental Microbiology*, 2013.

[8] Hadia Jalil, Sabeur Aridhi, Frédéric Borges, and Yannick Toussaint. Predicting long-term diversity in bacteria populations by graph-based machine learning approaches. *preprint*, *under submission*, 2023.

[9] El Kheir Sara M., Cherrat Lamia, Awussi Ahoefa A., Ramia Nancy E., Taha Samir, Rahman Abdur, Passerini Delphine, Leroi Françoise, Petit Jeremy, Mangavel Cécile, Revol-Junelles Anne-Marie, Borges Frédéric. High-Throughput Identification of Candidate Strains for Biopreservation by Using Bioluminescent Listeria monocytogenes. Frontiers in Microbiology, 9, p.1883, 2018.

[10] Borges Frédéric, Revol-Junelles Anne-Marie, Novel Strains of Carnobacterium Maltaromaticum and uses thereof. WO2021078612 (A1). 2021-04-29.

[11] D'Hondt, K., Kostic, T., McDowell, R. *et al.* Microbiome innovations for a sustainable future. *Nat Microbiol* **6,** 138–142 (2021).

[12] Abrudan MI, Brown S, Rozen DE. Killing as means of promoting biodiversity. Biochem Soc Trans. 2012 Dec 1;40(6):1512-6.