Programme du mini-symposium Jobim :

« **Bioinformatique des voies métaboliques, des séquences aux molécules.** »

7 juillet 2022, 15h-18h

15:00-15:15 **Introduction générale, présentation de la thématique**

par les membres du comité d'organisation et de programme

15:15-17:15 **Conférences invitées**

15:15-15:45 Maude Pupin (Université de Lille): bioinformatic tools to analyze nonribosomal peptides and other metabolites

15:45-16:15 Thomas Dussarrat (Pontificia Universidad Católica de Chile & Université de Bordeaux, INRAE): Multi-species predictive metabolomics via GLM approach unveils a generic metabolic toolbox for plant response to an extreme abiotic gradient in the Atacama Desert

16:15-16:45 Pause café

16:45-17:15 Yves Moreau (University of Leuven): Bayesian matrix factorization and deep learning for drug-target activity prediction

17:15-18:00 Table ronde

La priorité sera donnée aux étudiantes et étudiants pour les questions.

Bioinformatic tools to analyze nonribosomal peptides and other metabolites

Maude Pupin (Université de Lille)

Non-ribosomal peptides are natural peptides synthesized by specialized enzymes, the synthetases, encoded by large gene clusters. This synthetic pathway produces a great diversity of compounds because it selects and incorporates not only several hundred different amino acids, but also fatty acids or monosaccharides. These basic building blocks are called monomers. Moreover, bonds of different kinds are made between monomers, leading to various structures that may contain cyles and/or branches.

I will present the Norine software platform dedicated to these molecules. It proposes a specific notation to describe the monomer structure: a graph in which the nodes are the monomers and the edges are the chemical bonds between them. This notation facilitates the association between a metabolite and its biosynthetic gene cluster. It can be used for any metabolite synthesized by the assembly of monomers.

Norine offers tools to analyze and visualize secondary metabolites based on this notation. s2m and rBAN deduce a monomeric structure from a chemical structure. Kendrick Formula Predictor and NRPro help analyze mass spectrometry results. And Norine is a database dedicated to non-ribosomal peptides.

https://bioinfo.cristal.univ-lille.fr/NRP/

Multi-species predictive metabolomics via GLM approach unveils a generic metabolic toolbox for plant response to an extreme abiotic gradient in the Atacama Desert

Thomas Dussarrat[1,2], *Sylvain Prigent*[2], *Claudio Latorre*[4], *Stéphane Bernillon*[2], *Amélie Flandin*[2], *Francisca Diaz*[1], *Cédric Cassan*[2], *Pierre Van Delft*[3], *Daniel Jacob*[2], *Jérôme Joubes*[3], *Yves Gibon*[2], *Dominique Rolin*[2], *Rodrigo A. Gutiérrez*[1], *Pierre Pétriacq*[2]

[1] *Departamento de Genética Molecular y Microbiología, Pontificia Universidad Católica de Chile, FONDAP Center for Genome Regulation and Millenium Institute for Integrative Biology (iBio), Av Libertador Bernado O'Higgins 340, Santiago, Chile.* [2] *Univ. Bordeaux, INRAE, UMR 1332 BFP, 71 av E. Bourlaux, 33140 Villenave d'Ornon, France.* [3] *Univ. Bordeaux, CNRS, UMR 5200 Laboratoire de Biogenèse Membranaire, 71 av E. Bourlaux, 33140 Villenave d'Ornon, France.* [4] *Departamento de Ecología, Pontificia Universidad Católica de Chile, Av Libertador Bernado O'Higgins 340, Santiago, Chile.*

Performances of the best ideotypes are threatened by the increased aridity worldwide. Developing new strategies to improve our understanding of abiotic plant tolerance is urgent for more sustainable agriculture. Wild plant species that inhabit extreme lands such as deserts and high mountains represent a unique resource of adaptive molecular mechanisms. Yet, plant adaptation to hostile biomes is mainly considered species-specific[1]. However, the contribution of generic mechanisms remains unexplored. Here, we propose a comprehensive approach to investigate the role of metabolic processes in the adaptation of multiple selected species from the Atacama Desert, the driest non-polar environment on earth[2]. First, we compared gene expansion and expression patterns between 32 selected Atacama species thriving on an elevation gradient from 2500 to 4500m and 32 closest sequenced species. Reaction enrichment analyses via Pathway Tools identified genetic evolutions governing convergent biochemical pathways. Subsequently, the biochemical diversity of Atacama species was accessed using multi-platform metabolomics. Metabolome and plant environment were linked through generalised linear models to study the role of shared metabolic mechanisms in plant resilience. This predictive metabolomics approach unveiled a generic metabolic toolbox predicting plant environment, independently of plant species and year[3]. The predictive metabolites were detected in agronomic and ornamental species. Overall, while this study enhanced our understanding of the place of generic mechanisms in adaptation to extreme climates, our multi-species approach combined with machine learning offered promising perspectives in agronomy and ecology.

Keywords: Plant metabolism, adaptation, predictive metabolomics, GLM, multiple species.

References:

1. Scossa, F. & Fernie, A. R. The evolution of metabolism: How to test evolutionary hypotheses at the genomic level. *Computational and Structural Biotechnology Journal* **18**, 482–500 (2020).
2. Eshel, G. *et al.* Plant ecological genomics at the limits of life in the Atacama Desert. *Proc Natl Acad Sci USA* **118**, e2101177118 (2021).
3. Dussarrat, T. *et al.* Predictive metabolomics of multiple Atacama plant species unveils a core set of generic metabolites for extreme climate resilience. *New Phytologist* nph.18095 (2022) doi:10.1111/nph.18095.

Bayesian matrix factorization and deep learning for drug-target activity prediction

Yves Moreau (University of Leuven)

Learning latent representations via matrix completion or deep learning provides an attractive framework to handle sparsely observed data, also called "scarce" data. A typical setting for scarce data is the prediction of biological activity of chemical compounds against drug targets, where only 0.1% to 1% of all compound-target pairs are measured. Matrix factorization searches for latent representations of compounds and targets that allow an optimal reconstruction of the observed measurements. These methods can be further combined with linear regression models to create multitask prediction models. In our case, fingerprints of chemical compounds are used as "side information" to predict target activity. Such representations can also be leveraged to link drug targets and pathway activation. Similar results can be achieved via deep learning. Next to classical chemical fingerprints, high-content imaging can also be used to predict drug-target activity with improved performance for scaffold hopping. Our methods are available as two open source Python/C++ libraries - Macau & SparseChem (https://github.com/jaak-s/macau/tree/master/python/macau; https://github.com/melloddy/SparseChem).