JOBIM 2022 – Keynote abstracts

Anne-Laure Boulesteix

A replication crisis in methodological computational research?

Statisticians are often keen to analyze the statistical aspects of the so-called "replication crisis". They condemn fishing expeditions and publication bias across empirical scientific fields applying statistical methods. But what about good practice issues in their own - methodological - research, i.e. research considering statistical (or more generally, computational) methods as research objects? When developing and evaluating new statistical methods and data analysis tools, do statisticians and data scientists adhere to the good practice principles they promote in fields which apply statistics and data science? I argue that methodological researchers should make substantial efforts to address what may be called the replication crisis in the context of methodological research in statistics and data science, in particular by trying to avoid bias in comparison studies based on simulated or real data. I will discuss topics such as publication bias, the design and necessity of neutral comparison studies and the importance of appropriate reporting and research synthesis in the context of methodological computational research by drawing an analogy with clinical research.

Guillaume Bourque

Transposable elements are a source of innovation and variability in our response to infection

Transposable elements (TEs) have expanded the binding repertoire of many transcription factors and, through this process, have been co-opted in different transcriptional networks. In particular, TEs have contributed new regulatory sequences to immune-related genes. However, the extent to which TEs have shaped the evolution of the mammal immunity against pathogens remains poorly understood. In this presentation, I will discuss the results from a new study looking at epigenetic changes following influenza infection in monocyte-derived macrophages extracted from 35 individuals of different ancestry. Specifically, I will show that different TEs families are associated with regions of shared, but also variable, accessibility across individuals upon infection. I will also demonstrate that TEs can predict a significant portion of the heterogeneity that we observe in the response to infection. Finally, I will present new bioinformatics methods that make use of graph genomes to analyse epigenomic data. Such graphs can reveal signal that would have been missed using a standard reference genome.

Hélène Morlon

Analyzing the evolution of the microbiome using metabarcoding data

Phylogenetic analyses are central to study evolutionary dynamics, such as past dynamics of speciation and extinction and co-evolutionary dynamics. While such analyses are widespread in well-characterized groups of species, such as animals and plants, they are much more challenging in groups for which diversity is mostly known through environmental DNA techniques, as is the case of most microbial groups. I will present several approaches to addressing the difficulty of carrying phylogenetic analyses using metabarcoding data. In the case of diversification analyses, I will show that at least some results can be robust to high phylogenetic uncertainty, and that this uncertainty can be reduced by combining metabarcoding data with reference sequences. In the case of the (co)evolution of host associated microbiota, I will present new models for detecting vertical transmission as well as covariations between microbial taxa abundances through time. I will illustrate the talk with examples from the oceanic plankton, arbuscular mycorrhizal fungi, and the gut microbiota.

Cédric Notredame

Nextflow and NF-Core : growing your analysis with the right data sprinkler

Pipelines are everywhere these days, and their rapidly growing interconnection presents us with increasingly complex challenges. Scientific reproducibility was the first issue to be pointed out, but the social relevance of analytic pipelines means they will soon impact our everyday lives at various levels. Reproducibility and readability are much more than mere technical details in a medical setup. In this talk, I will present the genesis of Nextflow and how it grew from an internal lab tool up to becoming one of the work-horses of global genomics analysis. I will provide a simple overview of the underlying algorithm and what distinguishes it from alternative solutions. I will also discuss the impact of production software development in a research lab and how to deal with this. I will conclude with a few examples of massive scale-up going on in the lab and the importance of interoperable genomic pipelines for data integration in collective projects such as BovReg and EuroFAANG that bring together the farmed animals genomics community. I will wrap up with our future plans for pipeline integration, including NF-Benchmark, our ongoing project for dynamic biological benchmarking.

Pierre Peterlongo

Swim in the data tsunami

In a few years we have witnessed an incredible change in the genomic landscape. Sequencing has become a matter of a few hours of experiment for a low price. As a result, our community is flooded with data.
In this context, many analysis methods have become routine. One can think for example of mapping sequencing data to good reference genomes to determine simple variants such as SNPs or short indels. However, these methods become inefficient in many situations such as the analysis of non-model species for which the reference is poor or non-existent, or the analysis of complex metagenomic data, or more globally to decompartmentalize the analyses, usually performed at the scale of one or a few datasets.
In this talk, I will present various methods of analysis of sequencing data, seen through the prism of breaking down the barriers between the projects, for a better exploitation of the wealth of information offered by the democratization of sequencing.

Raphaël Guérois

Exploring interactomes in 3D under the lens of coevolution

Predicting protein-protein interactions and characterizing their structural organization provides essential information to elucidate the molecular mechanisms underlying cross-talk between cellular pathways. Exploiting coevolution information has emerged over the last years as a major strategy for predicting the mode of recognition between proteins. Our recent methodological developments [1, 2] have provided key insights to unravel the regulatory subtleties that might exist between pairs of interacting proteins [3, 4]. The advent of machine learning in the field, culminating in the Alphafold breakthrough, is pushing the boundaries from analyses of binary to multiple interactions. We will discuss about strategies to further integrate predictions with experimental assessment and further unravel the complexity of the molecular logic behind large networks of proteins interacting together.

[1] Quignot C, Granger P, Chacon P, Guerois R, Andreani J. Atomic-level evolutionary information improves protein-protein interface scoring. Bioinformatics. 2021.
[2] Quignot C, Postic G, Bret H, Rey J, Granger P, Murail S, Chacón P, Andreani J, Tufféry P, Guerois R. InterEvDock3: a combined template-based and free docking server with increased performance through explicit modeling of complex homologs and integration of covariation-based contact maps. Nucleic Acids Res. 2021 Jul 2;49(W1):W277-W284.
[3] Pyatnitskaya A, Andreani J, Guerois R, De Muyt A, Borde V. The Zip4 protein directly couples meiotic crossover formation to synaptonemal complex assembly. Genes Dev. 2022 36(1-2):53-69.
[4] Acharya A, Kasaciunaite K, Göse M, Kissling V, Guerois R, Seidel R, Cejka P. Distinct RPA domains promote recruitment and the helicase-nuclease activities of Dna2. Nat Commun. 2021 12(1):6521.