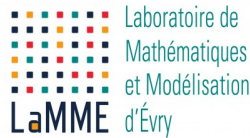
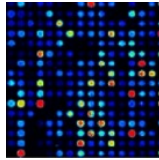


SÉMINAIRE – MATH FOR GENOMICS

SÉANCE DU MERCREDI 29 MAI 2019. 10H30.

EVRY. IBGBI. LAMME.

Biclustering via Latent Block Model for studying plant-microbiote interactions



Christophe Mougel (INRA Rennes)

Soils are a crucial part of the earth system and play fundamental roles in its functioning, linking the atmospheric, subsurface and aquatic compartments. Soils are the product of the activities of plants, which supply organic matter and play a pivotal role in weathering rocks and minerals. Soils supply many of mankind's needs and contribute greatly to the economic well-being of developed and developing nations. Soil is a non renewable resource that performs key environmental, social and cultural functions which are vital to human life and the sustainability of global ecosystems and defined as ecosystem services (MEA, 2005). These services result from the functioning of and interactions between living soil organisms (i.e. microbes, fauna and plant roots). Among them, soil microbes play key roles in ecosystems and influence a large number of important ecosystem processes, including nutrient acquisition (Smith & Read 1997; Sprent 2001), nitrogen cycling (Tiedje 1988; Kowalchuk & Stephen 2001), carbon cycling (Hogberg et al. 2001) and soil formation (Rillig & Mummey 2006). Moreover, soil microbes represent the unseen majority in soil and comprise a large portion of the genetic diversity on Earth (Whitman et al. 1998). For instance, it has been estimated that one gram of soil contains as many as 10^{10} – 10^{11} bacteria (Horner-Devine et al. 2003), 6000–50 000 bacterial species (Curtis et al. 2002), and up to 200 m fungal hyphae (Leake et al. 2004). However, while it is widely recognized that microbes perform crucial roles in biogeochemical cycling, the impact of microbes on plant functioning and productivity is still poorly understood in agrosystem compare to more natural environment like grassland and forest.

The major objective of the project is to better understand the plant – microbial communities' interactions in the rhizosphere thanks to a multidisciplinary approach coupling microbial ecology, ecophysiology and plant genetics. To achieve this goal we planified to (i) analyze the diversity of the total and active microbial communities in the rhizosphere of different *M. truncatula* genotype corresponding to the core collection used in genome-wide association studies, (ii) characterize of the C and N nutrition strategies of the plant during their interaction with the microbial communities in the rhizosphere and (iii) identify potential plant genes involved into plant functioning and/or the selection of particular microbial communities in the rhizosphere.

Significant advances have been achieved over the last decade on (i) our ability to characterize complex microbial diversity in the soil thanks to NGS and adequate bioinformatic pipeline and statistical methods combined with (ii) the possibility to analyze the plant functioning during plant-microbial communities interaction and (iii) with the advances in plant genotyping technology, including rapid increases in the number of genetic markers available for genome-wide association studies. Altogether the genome-wide association analysis has become a viable approach for the dissection of complex traits as plant genetic determinism of plant-microbial communities interactions and the consequence on plant nutrition strategies.

Julie Aubert (INRA Jouy)

Microbial ecology focuses how microorganisms interact with their environment, with each other and with their host. Different studies have shown that microbial communities living in animals (humans included), in or around plants have a significant impact on health and disease of their host and on various services, such as adaptation under stressing environment. Our work is motivated by the analysis of summarized microbial ecological data as provided by high-throughput sequencing, such as metabarcoding. Amplicon-based sequencing provides abundance matrices where rows correspond to microorganisms and columns to environmental samples they come from. Biclustering is one way to study the major question of associations between bacterial communities and biological samples. Indeed, biclustering aims at simultaneously dividing a data matrix into several homogenous subsets of rows and columns. This technique has been applied in various fields such as collaborative filtering or gene expression data. It has proved its usefulness in discovering local patterns in which a subset of genes exhibits a similar pattern over a subset of samples. Two different strategies may be used to this aim: model-based or algorithmic methods. We will consider here the latent block model (LBM) framework introduced by Govaert et Nadif (2010).

The LBM assumes that unobserved labels exist for both microorganisms and samples. Latent variables are not independent conditionally on observed ones and therefore the classical maximum likelihood inference is intractable. We propose a latent block model adapted to amplicon-based sequencing or metagenomic data. We focus on the Poisson-Gamma distribution to model the overdispersion observed in such data and introduce row and column effects to take the sampling effort and mean of microorganism abundance into account. We derive a variational-based inference algorithm and propose a strategy to select the number of biclusters. We implement the proposed method in the `cobiclust` R package and illustrate the flexibility and performance of this approach on different amplicon-based datasets.