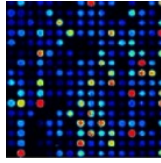# Séminaire – Math For Genomics
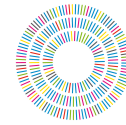## Séance du jeudi 6 mai 2021. 14h.
Evry. IBGBI.

Poisson lognormal models help digging into the mechanisms of species assemblages

---

Virginie Ravigné (chercheuse CIRAD, UMR PVBMT, Saint-Pierre, Réunion)

Joint species distributions reveal the combined effects of host plants, abiotic factors and species competition as drivers of species abundances in fruit flies

The relative importance of ecological factors and species interactions for phytophagous insect species distributions has long been a controversial issue. Using field abundances of eight sympatric Tephritid fruit flies on 21 host plants, we inferred flies' realized niches using joint species distribution modeling and network inference, on the community as a whole and separately on three groups of host plants. These inferences were then confronted to flies' fundamental niches estimated through laboratory-measured fitnesses on host plants. Species abundances were mainly determined by host plants followed by climatic factors, with a minor role for competition between species sharing host plants. The relative importance of these factors mildly changed when we focused on particular host plant groups. Despite overlapping fundamental niches, specialists and generalists had almost distinct realized niches, with possible competitive exclusion of generalists by specialists on Cucurbitaceae and different assembly rules: specialists were mainly influenced by their adaptation to host plants while generalist abundances varied regardless of their fundamental host use.

Julien Chiquet (Senior Researcher in Statistics, INRAE)

Variational inference for Poisson lognormal models: application to multivariate analysis of count data

Many application domains such as ecology or genomics have to deal with multivariate count data. A typical example is the joint observation of the respective abundances of a set of species in a series of sites, aiming to understand the co-variations between these species. The Gaussian setting provides a canonical way to model such dependencies, but does not apply in general. We adopt a framework based on the multivariate Poisson lognormal (PLN) model, which is attractive since it allows one to describe multivariate count data with a Poisson distribution as the emission law, while all the dependencies are kept in a hidden friendly multivariate Gaussian layer. While usual maximum-likelihood-based inference raises some issues in PLN, we show how to circumvent this issue by means of a variational algorithm for which gradient descent applies. We briefly study the asymptotic of those variational estimators. We then derive several variants of our algorithm to apply PLN to dimension reduction, discriminant analysis, clustering with Mixture model and

sparse covariance inference on multivariate count data. We illustrate our method on a microbial ecology dataset studying oaks powdery mildew. Code and slideshow are available for reproducibility at https://pln-team.github.io/PLNmodels/.