Call for postdoc candidates in statistics with application to quantitative genetics

We are looking for candidates to apply to `https://www.pepr-agroeconum.fr`.

Location MaIAGE unit (`https://maiage.inrae.fr`) and UMR Génétique Quantitative et Évolution-Le Moulon (`https://moulon.inrae.fr/`).

Duration and salary The duration is 24 months. The salary depends on the candidate's experience. Please contact us for more information.

Contact Gildas Mazo (`gildas.mazo@inrae.fr`), Mélisande Blein-Nicolas (`melisande.blein-nicolas@inrae.fr`) and Renaud Rincent (`renaud.rincent@inrae.fr`).

Topic In spite of the variety and sophistication of the statistical methods that have appeared in the last two decades, statistical modeling and inference of gene regulatory networks is increasingly challenging as biologists have been starting to embrace a more holistic approach and combine data at different scales and of different type to understand the system as a hole.

As an example, as part of an effort to adapt to climate change, a dataset was collected from a panel of 254 maize genotypes [6] to study the response of maize to drought. This dataset includes the abundances of about 2,000 proteins [2] and 20,000 transcripts (Nicolas et al., unpublished), phenotypic measurements for six ecophysiological traits (e.g., leave area) [1], and one million single nucleotide polymorphisms (SNPs) [5]. (A SNP is a change at one specific location in the DNA sequence.)

One goal is to discover variables ("biomarkers") that influence maize response to drought and how they interact with each other. Combining Gaussian graphical models on the protein and trait variables with quantitative genetics models, it was shown that modeling multiscale networks allowed to better explain the variance of the genotype x water deficit interaction and have a finer understanding of the genotype-phenotype relationship [3].

However, to improve our understanding of the genotype-phenotype relationship, the different types of data (SNPs, transcripts, proteins, traits) need to be combined. Since some are categorical (SNPs) and some others are continuous (proteins, traits), Gaussian graphical models are not adapted anymore.

One possible approach to address such questions is to assume that the Gaussian vector is only latent, meaning that the observed variables are only transformations of the hidden Gaussian vector [4]. This considerably relaxes the Gaussian model assumption because one can construct statistical models with *any* given marginals. In particular, both discrete and continuous variables can be mixed together.

The topic of the postdoc will be to (i) develop copula methods and theory for network inference and (ii) investigate their potential on the maize dataset mentioned above.

Candidate We are looking for applications from outstanding candidates with a PhD in statistics, biostatistics or machine learning.

References

[1] Santiago Alvarez Prado, Llorenç Cabrera-Bosquet, Antonin Grau, Aude Coupel-Ledru, Emilie J. Millet, Claude Welcker, and François Tardieu. Phenomics allows identification of genomic regions affecting maize stomatal conductance with conditional effects of water deficit and evaporative demand. *Plant, Cell & Environment*, 41(2):314–326, February 2017.

[2] Mélisande Blein-Nicolas, Sandra Sylvia Negro, Thierry Balliau, Claude Welcker, Llorenç Cabrera-Bosquet, Stéphane Dimitri Nicolas, Alain Charcosset, and Michel Zivy. A systems genetics approach reveals environment-dependent associations between SNPs, protein coexpression, and drought-related traits in maize. *Genome Research*, 30(11):1593–1604, November 2020.

[3] Yacine Djabali. *Harnessing multi-omics data to unravel the genetic and molecular basis of complex traits : A systems genetics study of maize drought response.* Theses, Université Paris-Saclay, December 2023.

[4] Gildas Mazo, Dimitris Karlis, and Andrea Rau. A Randomized Pairwise Likelihood Method for Complex Statistical Inferences. *Journal of the American Statistical Association*, 0(0):1–11, 2023. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2023.2257367.

[5] Sandra S. Negro, Emilie J. Millet, Delphine Madur, Cyril Bauland, Valérie Combes, Claude Welcker, François Tardieu, Alain Charcosset, and Stéphane D. Nicolas. Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. *BMC Plant Biology*, 19(1):318, December 2019.

[6] R. Rincent, S. Nicolas, S. Bouchet, T. Altmann, D. Brunel, P. Revilla, R. A. Malvar, J. Moreno-Gonzalez, L. Campo, A. E. Melchinger, W. Schipprack, E. Bauer, C.-C. Schoen, N. Meyer, M. Ouzunova, P. Dubreuil, C. Giauffret, D. Madur, V. Combes, F. Dumas, C. Bauland, P. Jamin, J. Laborde, P. Flament, L. Moreau, and A. Charcosset. Dent and Flint maize diversity panels reveal important genetic potential for increasing biomass production. *Theoretical and Applied Genetics*, 127(11):2313–2331, November 2014. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 11 Publisher: Springer Berlin Heidelberg.