







PhD Position: Statistical approaches to investigate and predict genomic vulnerability of populations and species, with applications to teosinte and maize.

Context and objectives

Climate change has already stared altering the productivity of agricultural systems through increased temperatures, changing precipitation patterns, and more frequent extreme weather events. In France, for example, crop yields have recently been impacted by drought and heat stress and by extremely wet conditions. More broadly, the effects of climate change may reduce global crop production by at least 10% in 2050. As global temperatures rise and precipitation patterns shift, the ability of crops and livestock to adapt is critical for maintaining food security, and predicting which populations are at risk of maladaptation under future climate scenarios has become a major issue. In this context, Genome-Environment Association (GEA) and Genomic Offset (GO) approaches, that explore the relationship between the genetic content of a population and the environmental conditions it experiences have received a growing interest (Fitzpatrick et al, 2015, Gautier, 2015). These methodologies can help to anticipate the potential maladaptation of species, thereby guiding strategies for conservation, breeding, and land management to mitigate risks and enhance resilience.

The goal of the PhD project is to contribute to the development of innovative GEA and GO procedures that will build on recent advances in optimization and statistical inference. More specifically the recruited PhD student will consider variational inference approaches (Zhang et al, 2018) for GEA and stochastic optimization to speed up the inference, with the objective of scaling up to modern genomic datasets that may involve hundreds of populations. He/she will also develop probabilistic GO models inspired from the Redundancy Analysis approach (Tipping & Bishop, 1999, Capblancq & Forester, 2021) and extend it by introducing Neural Networks in order to handle non-linear relationships between covariates and response variables. To this aim, the PhD student will join a consortium of researchers issued from different disciplines with a long experience in interdisciplinary projects. The developed methodology will be applied to public datasets for benchmarking purposes, as well as, to an innovative set of two datasets corresponding to a domestic plant (maize) and its wild "ancestor" (teosinte). This will enable us to investigate how adaptation of crop wild relatives to dry environments could be informative about the response of maize to a drier climate.

Work environment

The PhD will be hosted by the UMR MIA Paris Saclay, expert in statistical modeling and machine learning methodology with applications in genomics and genetics, and supervised by J. Aubert and T. Mary-Huard. Applications will be realized in

collaboration with M. Tenaillon and L. Moreau, experts in quantitative and population











genetics and maize breeding at UMR Génétique Quantitative et Évolution-Le Moulon. The PhD student will also benefit from the co-supervision of Mathieu Gautier (CBGP, Univ Montpellier) for the development of GEA approaches.

The PhD is co-funded by the <u>CLand project</u> and the <u>PEPR Maths-Vives</u>.

Required profile

The PhD student will be affiliated to the Graduate School "Ecole Doctorale de Mathématiques Hadamard" of Université Paris-Saclay.

The requirements are the following:

- A Master's degree in Applied Mathematics / Statistics / Machine Learning,
- Advanced programming skills in Python and/or R.
- Prior experience in biostatistics and genomic data analysis is desirable but not essential.

A strong interest in applications in biology and ability to work in a multidisciplinary team.

How to candidate ?

Send your CV, a cover letter and the contact details of two references to Tristan Mary-Huard (<u>tristan.mary-huard@agroparistech.fr</u>) Julie Aubert (julie.aubert@agroparistech.fr)

The contract will be for 3 years, starting in September or October 2025.

Bibliography

Capblancq & Forester (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. Methods in Ecology and Evolution, 12(12), 2298-2309.

Fitzpatrick et al (2015). Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. Ecology letters, 18(1), 1-16.

Gautier (2015). Genome-Wide Scan for Adaptive Differentiation and Association Analysis with population-specific covariables. Genetics, 201(4):1555-1579.

Tipping & Bishop (1999). Probabilistic principal component analysis. Journal of the Royal Statistical Society Series B: Statistical Methodology, 61(3), 611-622.

Zhang et al (2018). "Advances in variational inference." IEEE transactions on pattern analysis and machine intelligence 41.8 : 2008-2026.