

12 Months Engineer Position in Statistics at INRA-MIA

Distance-based clustering of ecological community data

Community ecologists are often interested in clustering samples collected in different sites or mediums. In many cases, the genomic data are summarized as a matrix of pairwise dissimilarities or distances between samples. A huge literature exists about on the one hand the various dissimilarities that can be considered and on the other hand clustering algorithms that can then be applied to a given dissimilarity or distance matrix. Parallel to this, meta-data or covariates (depth, salinity, pH, etc) are often collected to describe the sites, mediums or conditions under which the samples have been observed.

The HYDROGEN ANR project is interested in such problems in order to analyse samples collected as part of the TARA ocean project. The project focuses on *k*-mer based comparisons between samples for which only dissimilarity matrices between samples can be stored. In this context we are typically interested in

1. Assessing the robustness of clustering methods to the dissimilarity/distance choice;
2. Understanding which covariates drive those dissimilarities.

Depending on the advances of the project, the recruited engineer will be involved in either of these tasks.

Context:

This contract is funded by the French ANR project HydroGen (Comparative Metagenomic for Measuring Biodiversity. Application to Ocean Life Studies):

[http://www.agence-nationale-recherche.fr/en/anr-funded-project/?tx_lwmsuivibilan_pi2\[CODE\]=ANR-14-CE23-0001](http://www.agence-nationale-recherche.fr/en/anr-funded-project/?tx_lwmsuivibilan_pi2[CODE]=ANR-14-CE23-0001)

This project is a collaboration between 3 research teams with complementary competences in algorithmics, statistics and genomics: INRIA-GenScale, INRA (MaIAGE+AgroParisTech) and CEA-CNS-LABIS.

Skills:

A strong background in statistics and applied mathematics is required, in addition to a good experience in computer programming. An interest for applications in biology and ecology is also more than welcome.

Location:

The position can be held either at INRA MaIAGE (Jouy-en-Josas) or at AgroParisTech (Paris). Visits to the 2 other partners in Rennes and Evry will be planned.

Salary: 2000€ a month (BT).

Starting: The position is available now.

Contact:

- Stéphane Robin (UMR MIA-Paris, AgroParisTech/INRA) <robin@agroparistech.fr>
- Mahendra Mariadassou (UR MaIAGE INRA, Jouy-en-Josas) <mahendra.mariadassou@inra.fr>