

M1 Research Internship (2020-2021)

**Knowledge transfer using multivariate gene expression projections  
onto a large-scale reference database**

Andrea Rau

*BioEcoAgro* Joint Cross-Border Research Unit

Address : 2 Chaussée Brunehaut, CS 50136 Estrées-Mons, 80203 Péronne Cedex

Email : [andrea.rau@inrae.fr](mailto:andrea.rau@inrae.fr)

Web : <https://www.bioecoagro.eu>  
<https://www.bioecoagro.eu/Structure/Cluster-1/Team-1>

**Team name:** *Cluster 1*: Innovative cropping systems for agro-ecological and bioeconomic transition in the context of climate change / *Team 1*: Functioning and adaptation of the plant in interaction with its environment

**Keywords :** Data integration, gene expression, multivariate methods

**Project context:**

Large-scale curated transcriptome datasets are available for several well-characterized stresses in controlled conditions for model species, providing a valuable opportunity to reproduce existing results in the literature and transfer knowledge to new studies. Although these data represent a rich resource for deciphering the pathways involved in various stress responses under controlled or field conditions, they are typically under-utilized and tend to rely on gene-by-gene queries for a subset of candidates in the species of interest (e.g., based on their expression profile or genetic position). This simultaneously represents a serious gap, and an opportunity to better leverage existing knowledge to achieve a more complete characterization of stress-related pathways in other experimental conditions or non-model species.

In this internship, our objective is to develop a **multivariate statistical approach to project gene expression data** from a given experiment onto a **large-scale multi-stress reference 'omics dataset** in a model species, to build a **bridge of interpretation** between related stress conditions:

**science for people, life & earth**

**UMR1313 GABI**  
**Animal Genetics and Integrative  
Biology unit**  
Domaine de Vilvert  
78350 Jouy-en-Josas - France  
+33 1 (0) 1 34 65 22 01

Join us



UMR's website :  
[https://www6.jouy.inrae.fr/gabi\\_eng/](https://www6.jouy.inrae.fr/gabi_eng/)

- (1) The [GEM2NET](#) module of the CATdb database is made up of 387 identically processed microarray experiments corresponding to 18 categories of biotic or abiotic stresses in *Arabidopsis thaliana*, a flowering plant widely used as a model organism. We will extract, pre-process, and format these data for use as a multi-stress reference dataset.
- (2) We will perform a straightforward dimensionality reduction on expression profiles for curated gene sets extracted from the multi-stress reference data. A supervised learning approach will be fit to a subset of components using the known stress subgrouping category labels. Independent samples will be projected into the feature-space, and the trained supervised learning approach will be used to assign them conditional probabilities of membership for each of the stress subgroups.
- (3) We will perform a benchmarking study of the validity and interpretability our approach with hold-out data from GEM2Net. If time permits, we will also validate the approach with independent *Arabidopsis* experiments under related stress conditions extracted from other public databases.

### Required skills:

- Solid knowledge of the R programming language for data analysis (experience using Bioconductor packages would be appreciated but is not required);
- Knowledge of and experience with approaches for dimensionality reduction (e.g., principal components analysis, nonnegative matrix factorization) and supervised learning (e.g., Gaussian mixture models, support vector machines);
- Comfortable reading/writing English (ability to read and understand scientific articles and write software documentation);
- Candidates should have motivation and interest for genomic applications in general, but experience in this area is not required.

The research internship will take place in a research environment that brings together biostatisticians, bioinformaticians, and biologists. The research internship will take place at the [Inrae Hauts-de-France Research Center](#) in Estrées-Mons (80). The work will be supervised by Dr. Andrea Rau.

### Remuneration:

- Standard research internship salary (approximately 580 € / month)

science for people, life & earth

UMR1313 GABI  
Animal Genetics and Integrative  
Biology unit  
Domaine de Vilvert  
78350 Jouy-en-Josas - France  
+33 1 (0) 1 34 65 22 01

Join us  
      
UMR's website ;  
[https://www6.jouy.inrae.fr/gabi\\_eng/](https://www6.jouy.inrae.fr/gabi_eng/)

  
RÉPUBLIQUE  
FRANÇAISE  
Liberté  
Égalité  
Fraternité

INRAE