

Research project for Master student (Cirad / University of Yaoundé 1) - 2016

General information:

- Title: **Genomic selection in rubber tree**
- Keywords: plant breeding, quantitative genetics and genomics, R programming
- 6 months, at the **University of Yaoundé 1**
- contact: **david.cros@cirad.fr**

Details:

Genomic selection (GS) is a method of marker assisted selection (MAS) efficient to improve quantitative traits. It relies on dense marker coverage of the genome and on statistical methods able to take advantage of the information of all the markers jointly, usually the BLUP methodology and Bayesian approaches. The model of GS is calibrated on a pool of individuals called “training set” using their data records (usually phenotypes) and their genotypes. It is then applied to selection candidates with genotypes for the same markers but no data records, in order to estimate their additive genetic value or their total genetic value (ie the sum of additive and non additive genetic values). All the markers are used for selection, which differs over the previous MAS approaches, where only a subset of significant markers was kept. GS is particularly promising when traditional breeding requires extensive phenotyping, as in this case it becomes possible to reduce the generation interval and to increase selection intensity. The potential of GS is therefore very high for perennial crops.

This research project aims at evaluating the potential of GS for rubber tree using empirical data:

- Question 1: Do GS models give better estimates of the total genetic value of phenotyped individuals than conventional models?
- Question 2: Do GS models allow predicting the total genetic value of non-phenotyped sibs of the training individuals?

Phenotypic and molecular data are available on two families of around 190 full-sibs. The phenotypes were collected for 4 traits (latex production, sucrose, circumference, branching), on several clones of each full-sib individual. The molecular data are the genotypes for 309 SSR markers.

Planning:

Literature review (GS / rubber tree)	3 weeks
Preparing molecular and phenotypic data (including imputation of missing SSR data)	1 week
Analyzing the whole dataset with conventional approach (without SSR data) and with GS models (cf Question 1). This will be done with R-asreml (“G-BLUP” with various genomic matrices and models), on the 4 traits and varying the number of markers.	1.5 month
Estimating the prediction accuracy of GS for non-phenotyped individuals (cross-validation approach) (cf Question 2). This will be done with the BGLR package for R (“G-BLUP”, Bayesian methods and a semi-parametric method ; and several models), on the 4 traits and varying the number of markers.	2 months
Report writing	1.5 months

There is **no field work, no lab work**. All analyzes will be done with the R program. The student will mostly have to adapt pre-existing R scripts, but also to write a few new specific scripts. **Knowledge of R is required. Knowledge of quantitative genetics and / or statistics is required.** No prior knowledge about GS, MAS or rubber tree is required.

The results are expected to give a scientific article.