

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

### General information:

- Title: **Investigating genomic selection in oil palm with computer simulations**
- Keywords: R programming, linear mixed model, quantitative genetics
- 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. 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 a pool of individuals (selection candidates) with genotypes for the same markers but no data records, in order to predict their genetic value. 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.

Oil palm is among the most advanced perennial species regarding GS. Empirical comparisons of breeding strategies are very difficult to implement in oil palm, as they would require a lot of time and money. Computer simulations are therefore of major interest for this species. A simulation was written in R to compare the current phenotypic breeding scheme and several strategies of GS over four generations for two correlated traits (Cros et al 2015). It led to the identification of two GS strategies that were relevant alternatives to the current breeding strategy, as they led to much larger annual responses to selection. It also showed that RRGS required inbreeding management because of a higher annual increase in inbreeding than RRS.

The goal of this research project is to extend this previous work studying the effects on selection response of new factors:

- univariate versus bivariate models,
- combining two generations of data in the training set,
- inbreeding management

### Planning:

Literature review (GS / oil palm breeding / HaploSim R package), Understand the existing R script	1 month
Implement univariate analyzes and compare uni- and bivariate approaches So far only bivariate analyzes are made. Univariate analyzes should be added in the modified script and it should be possible for the user to choose between uni- and bivariate analysis.	1 month
Add possibility of using two generations to train the GS models So far the GS models are calibrated with the data of a single generation. The modified script should have a scenario where the data of 2 generations (ie n°1 and 3) are used to train the GS model applied in the generation n°4.	3 weeks
Literature review (inbreeding management)	1 week
Implement breeding strategies with inbreeding management	2 months
Finalizing report	1 month

The student will have to adapt pre-existing R scripts and to write new specific scripts. **Good knowledge of R is required. Knowledge of linear mixed model, statistics or quantitative genetics is required.** Knowledge of quantitative genetics would be appreciated.

The results are expected to give a scientific article.