

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

General information:

- Title: **Comparison of genetic parameters estimated from SSR markers and genotyping-by-sequencing in oil palm breeding populations**
- Keywords: plant breeding, population genetics and genomics, R programming
- 6 months, at the **University of Yaoundé 1**
- contact: **david.cros@cirad.fr**

Details:

Several genetic parameters are of major importance for breeding. They help in predicting the response to (marker assisted) selection, heterosis, etc. They are also necessary to calibrate computer simulations that allow studying breeding strategies that could not be implemented in real life.

In oil palm, linkage disequilibrium (LD), population structure (existence of subpopulations, F_{ST}) and effective size for instance have only been studied with SSR markers, at a low density, while SNP markers are becoming the current reference marker type for genetic studies. Genotyping-by-sequencing (GBS) is an appealing strategy to obtain high density SNP coverage of the whole genome at a reasonable cost. However, the different characteristics of SSR and GBS-SNP (percentage of missing data, percentage of genotyping errors, etc) make the comparison of genetic parameters computed from these 2 marker types necessary. The goal of this project is to implement such a comparison using oil palm empirical data.

The available dataset is made of 140 group A oil palm individuals (Asiatic Deli population mostly, Angola) and 131 group B individuals (a mixture of African populations, mostly Côte d'Ivoire and Democratic Republic of Congo), genotyped with 313 SSR markers and 3000+ SNP from GBS (group A and group B are the two parental heterotic groups used to produce oil palm commercial hybrids).

Planning:

Literature review (population genetics / R packages for population genetics / oil palm breeding populations)	1 month
Studying population structure by principal component analysis with SSR and GBS-SNP For SNP the SNPrelate R package will be used.	0.5 months
Computing F_{ST} between breeding populations with SSR and GBS-SNP. For SNP the Weir and Cockerham estimate of Fst will be computed using the SNPrelate or Hierfstat R packages. For SSR the diveRsity R package will be used.	0.5 months
Computing LD with SSR and GBS-SNP For SNP this will use the statistics implemented in the package LDcorSV, that can correct for population structure and relatedness between individuals (Mangin et al 2012). We will also use the approach of de Roos et al (2008) in cattle to estimate number of generations since divergence between the major breeding populations (Deli, La Mé, Yangambi) and past evolution of effective size.	2.5 months
Report writing	1.5 months

Other genetic parameters can be taken into consideration (% heterozygosity, inbreeding).

There is **no field work, no lab work**. All analyzes will be done with the R program, using various packages (the above mentioned, adegenet, etc). **Knowledge of R is required. Knowledge of genetics and / or statistics is required.** No prior knowledge oil palm is required.

The results are expected to give a scientific article.