Practical aspects of genomic selection in oil palm (*Elaeis guineensis*)

D. Cros, M. Denis, L. Sánchez, B. Cochard, A. Flori, T. Durand-Gasselin, B. Nouy,
A. Omoré, V. Pomiès, V. Riou, E. Suryana, J.M. Bouvet

D. Cros (✉), M. Denis, B. Cochard, A. Flori, V. Pomiès, V. Riou, J.M. Bouvet
CIRAD, UMR AGAP (Genetic Improvement and Adaptation of Mediterranean and Tropical Plants Research Unit), 34398 Montpellier, France. e-mail: david.cros@cirad.fr

L. Sánchez, INRA, UR0588, UAGPF (Forest Tree Improvement, Genetics and Physiology Research Unit), 45075 Orléans, France

E. Suryana, P.T. SOCFINDO Medan, Medan 20001, Indonesia

A. Omoré, INRAB, CRAPP, Pobè, Benin

B. Nouy, T. Durand-Gasselin, PalmElit SAS, 34980 Montferrier sur Lez, France

Genomic selection (GS) is a form of marker assisted selection that improves breeding schemes in plants and animals. It relies on dense genome wide marker coverage to produce genomic estimated breeding values (GEBV) from a joint analysis of all markers. The GS model is calibrated using individuals with known phenotypes and genotypes (training set) and predicts the GEBV of selection candidates. GS can increase the genetic gain in plants. In perennial crops, this can be achieved via shortened breeding cycles and increased selection intensity. In oil palm (*Elaeis guineensis*) the main challenge is to obtain sufficient accuracy to calibrate the GS models, despite small populations.

We obtained empirical estimates of GS accuracy in oil palm for eight yield traits using real data from an on-going breeding program. We used three parental populations (Deli, Group B, La Mé) involved in conventional reciprocal recurrent selection (RRS) with 93 to 131 individuals, genotyped with 265 SSR. We estimated the within population GS accuracy when predicting masked estimated breeding values. We used three methods to sample training sets and the GBLUP statistical method to obtain GEBV. The results showed that in Group B and La Mé, GS could achieve higher accuracy than the pedigree-based model, indicating that GS could account for family effects and Mendelian sampling terms. The GS accuracy ranged from -0.41 to 0.94 and was correlated with the relationship between training and test sets (a_{max}). Training sets optimized with CDmean gave the highest a_{max} and accuracies, ranging from 0.49 to 0.94. Finally, Group B and La Mé individuals could be preselected for progeny tests by applying GS to key yield traits.

However, many aspects related to the potential of GS are difficult to study with real data. Wong and Bernardo (2008) carried out a simulation study to evaluate over three generations the potential of GS to increase yield in oil palm. They concluded that GS gave the highest annual response to selection. In order to extend their promising results, we simulated two realistic oil palm breeding populations and compared over four generations the current RRS with reciprocal recurrent genomic selection (RRGS). The aim was to select the best individuals in the two parental populations to increase hybrid performance on bunch production. For RRGS, we applied the GBLUP method with the phenotype of hybrids as data

records to obtain the parental genomic estimated breeding value. We studied the effects of four parameters on the selection response in hybrids: (1) the molecular data used to calibrate the GS model: in RRGS_PAR, we only used parental genotypes and in RRGS_HYB we also used genotypes of hybrid individuals; (2) frequency of calibration of GS model; (3) number of candidates; (4) number of genotyped hybrids in RRGS_HYB. We concluded that both RRGS_PAR and RRGS_HYB could lead to a higher selection response than RRS because they allowed decreasing the generation interval and increasing selection intensity.