

## GENOMIC SELECTION IN OIL PALM (*ELAEIS GUINEENSIS* JACQ.)

David Cros, Marie Denis, Leopoldo Sánchez, Benoit Cochard, Albert Flori, Tristan Durand-Gasselín, Bruno Nouy, Alphonse Omoré, Virginie Pomiès, Virginie Riou, Edyana Suryana, Jean-Marc 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-Gasselín*, PalmElit SAS, 34980 Montferrier sur Lez, France

### 1 ABSTRACT

Genomic selection (GS) can increase the genetic gain in plants. In perennial crops, this can be achieved via shortened breeding cycles and increased selection intensity. Our objective was to obtain the first empirical estimate of GS accuracy in oil palm (*Elaeis guineensis*), where the main challenge is to obtain sufficient accuracy to train GS models, despite small populations. We used two parental populations involved in conventional reciprocal recurrent selection (Deli and Group B) with 131 individuals each, genotyped with 265 SSR. We estimated the within population GS accuracy when predicting masked estimated breeding values for eight yield traits. We used three methods to sample training sets and five statistical methods to estimate genomic breeding values. The results showed that in Group B, 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. The statistical methods did not affect the GS accuracy. Finally, Group B individuals could be preselected for progeny tests by applying GS to key yield traits.

*Keywords: Genomic selection, oil palm, yield, relatedness, GBLUP, Mendelian sampling term*

### 2 INTRODUCTION

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. GEBV are obtained by summing up estimates of marker effects or through a realized additive relationship matrix markers. The model is calibrated using individuals with known phenotypes and genotypes (training set), and subsequently used to produce GEBV on a different set of selection candidates that were only genotyped (test set) (Meuwissen et al. 2001). Depending on the breeding system, genetic gains per year should increase because of the higher accuracy of GS

as compared to conventional selection, shorter generation intervals with the early testing of selection candidates (especially when conventional selection involves progeny testing) and/or higher selection intensity (especially when phenotyping is a limiting factor). Statistical methods to estimate GEBV use two types of information: additive genetic relationships between training and test sets and LD between markers and QTL (Habier et al. 2007; Habier et al. 2010). The GEBV thus implicitly take the two parts of the breeding value of an individual into account, i.e. the average value of its parents (family effects) and the Mendelian sampling term (within-family effects) (Daetwyler et al. 2013). The accuracy of GS, which is the correlation between GEBV and true breeding values, is affected by linkage disequilibrium (LD) between markers and quantitative trait loci (QTL), the relationship between training and test sets, the number of individuals in the training set, the statistical method to estimate GEBV, the trait heritability and the distribution of underlying QTL effects (Lorenz et al. 2011; Grattapaglia 2014).

Currently, few empirical studies have assessed the GS potential in species with long breeding cycles (>10 years) (see Grattapaglia 2014 for a review). Oil palm (*Elaeis guineensis*) is a perennial crop with a highly dynamic breeding sector because of its position as the major world oil crop. Its production is currently over 55 Mt (USDA 2013) and is expected to further increase substantially as demand for palm oil could be between 120 and 156 Mt in 2050 (Corley 2009). Oil palm is a diploid, monoecious and allogamous species with a very high GS potential due to its breeding system. Currently, oil palm genetic improvement is generally based on reciprocal recurrent selection (RRS) (Gascon and de Berchoux 1964). It aims at increasing oil yield, which is a function of bunch number, bunch weight and fruit to bunch, pulp to fruit and oil to pulp ratios. The RRS relies on two populations: the Deli (of Asian origin) and the Group B (a mixture of African populations). These populations show complementary characteristics with respect to total bunch production. Candidate palms sampled from full-sib families in each of the two populations are progeny tested in Deli x Group B crosses and evaluated in long and extensive field trials, in order to get highly reliable estimated breeding values (EBV, with accuracy around 0.90 for all yield components). The best individuals are selected to produce the following generation and commercial hybrid material. Therefore, conventional breeding in oil palm is cost intensive and time consuming, leading to a long breeding cycle (around 20 years, while sexual maturity is reached at around 3 years of age) with a limited number of tested individuals. The main GS challenge for this species is currently to achieve high accuracy of GEBV despite the small training sets that are available (<250 progeny tested individuals).

The only study in which the GS potential was investigated in oil palm is a simulation by Wong and Bernardo (2008). Although it yielded promising results, the genetic characteristics of the simulated breeding populations differed from those of real breeding populations, thus necessitating an empirical study. Their simulation assumed a base breeding population formed by the selfing of a hybrid between two inbred lines, whereas real breeding populations are more complex: all the oil palm breeding populations originated from a small number of founders that were first subjected to mass selection (with the number of generations, intensity and traits of interest depending on the population) and they were subsequently submitted to RRS.

Our objective here was to assess the potential of GS in the context of current RRS oil palm breeding by obtaining the first empirical estimate of GS accuracy using the largest EBV and genotype datasets for the species. The GS strategy investigated involved the use of progeny tested individuals of the Deli population and Group B as

two training sets for within population GS (see Figure 1 for details). The data records used to calibrate the GS model were the individual EBV obtained from progeny tests between the two populations. The GS accuracy was assessed when predicting the EBV of individuals with no data records (ie not progeny tested). Specifically, we aimed to study the effects of four parameters on the GS accuracy: (1) the relationship between training and test sets: we used three methods to define the training and test sets in order to obtain a range of relationships between them; (2) the traits: we studied eight yield traits, assuming this would cover a broad range of genetic architectures; (3) the statistical method used to estimate the GEBV: we compared five statistical methods known to behave differently depending on the genetic architecture of the traits; and (4) the population: our study included Deli and Group B populations, assuming that their contrasted history would lead to genetic differences like LD profile and genetic architecture of traits.

### **3 MATERIALS AND METHODS**

The data available (i.e. individuals with both EBV and genotypes) represented 131 Deli and 131 Group B individuals. Individuals were genotyped with 265 SSR.

#### **3.1 Populations and molecular data**

All individuals belonged to families from the commercial oil palm breeding program of PalmElit ([www.palmelit.com](http://www.palmelit.com)). The Deli population originated from four ancestral oil palms planted in 1848 in Indonesia and was selected for yield at least from the early 20<sup>th</sup> century (Corley and Tinker 2003). The 131 Group B individuals included 93 La Mé (Côte d'Ivoire), 24 Yangambi (Democratic Republic of the Congo), 5 La Mé x Yangambi, 7 La Mé x Sibiti (Democratic Republic of the Congo, related to Yangambi) and 2 Nigeria individuals. The base of African populations was also formed by few founders, collected during the first half of the 20<sup>th</sup> century (Cochard et al. 2009). African populations were also submitted to selection for yield and inbreeding. Most of the 131 Deli and 131 B individuals were tested in the second RRS cycle and the remaining was selected at the end of the first cycle. The individuals were genotyped with 265 SSR (Billotte et al. 2005; Tranbarger et al. 2012). The number of polymorphic SSR markers was 220 in Deli and 260 in Group B, leading to marker densities of one SSR per 7.9 and 6.7 cM, respectively, based on a genome length of 1.743 cM (Billotte et al. 2005). Molecular coancestry (i.e. kinship) was calculated according to Eding and Meuwissen (2001) and was on average 0.58 in Deli (range 0.42 - 0.96) and 0.39 in Group B (0.12 - 0.92). The heat maps of the molecular coancestry matrices *G* indicated that the populations were highly structured (not showed).

#### **3.2 Estimation of breeding values from field experiments**

The estimated breeding values (EBV) of the individuals were obtained through progeny tests conducted in a large-scale experiment at Aek Loba (Sumatra). The mating design consisted of 492 Deli x Group B crosses carried out using an incomplete factorial design. The crosses were evaluated in 28 trials planted between 1995 and 2000.

Eight traits were studied. The bunch number (BN) and average bunch weight (ABW) were measured on palms from ages 6 to 11. The fruit to bunch (F/B), pulp to fruit (P/F), kernel to fruit (K/F) and oil to pulp (O/P) ratios, the number of fruits per bunch (NF) and the average fruit weight (FW) were measured on two bunches at

ages five and six on a sample of at least 40 palms per cross. The bunch production data concerned 30,872 palms and bunch quality data concerned 21,525 palms.

EBV were computed as BLUP predictors of the random effects  $\mathbf{a}_A$  and  $\mathbf{a}_B$ , using a mixed model of the form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a}_A + \mathbf{Z}_2\mathbf{a}_B + \mathbf{Z}_3\mathbf{b} + \mathbf{Z}_4\mathbf{c} + \mathbf{Z}_5\mathbf{p} + \mathbf{Z}_6\mathbf{k} + \mathbf{e}$$

where  $\mathbf{y}$  is the vector of data records for the trait being analyzed,  $\boldsymbol{\beta}$  the vector of fixed effects (general mean, trial and block within trial),  $\mathbf{a}_A$  and  $\mathbf{a}_B$  vectors of general combining ability of Deli  $\sim N(0, \mathbf{A}_{Deli} \sigma^2_{Deli})$  and Group B individuals  $\sim N(0, \mathbf{A}_B \sigma^2_B)$ , respectively,  $\mathbf{b}$  the vector of the incomplete block within block and trial effects  $\sim N(0, \mathbf{I}\sigma^2_b)$ ,  $\mathbf{c}$  the vector of specific combining ability of single crosses  $\sim N(0, \mathbf{D}\sigma^2_c)$ ,  $\mathbf{p}$  the vector of permanent environmental effects used to take repeated measures into account  $\sim N(0, \mathbf{I}\sigma^2_p)$ ,  $\mathbf{k}$  the vector of elementary plot effects  $\sim N(0, \mathbf{I}\sigma^2_k)$  and  $\mathbf{e}$  the vector of residual effects  $\sim N(0, \mathbf{I}\sigma^2_e)$ .  $\mathbf{X}$ ,  $\mathbf{Z}_1 - \mathbf{Z}_6$  are incidence matrices.  $\mathbf{A}_{Deli}$  and  $\mathbf{A}_B$  are matrices of additive relationships among Deli and Group B individuals, respectively, computed from pedigrees,  $\mathbf{D}$  is the matrix of dominance relationships among crosses computed from the pedigree and  $\mathbf{I}$  is an identity matrix. For BN and ABW, the model also included a fixed age effect and a random age within cross effect  $\mathbf{a} \sim N(0, \mathbf{I}\sigma^2_a)$ . Estimates of the narrow-sense heritability ( $h^2$ ) of each trait were obtained at the experimental design level as the ratio of additive variance ( $\sigma^2_{Deli}$  and  $\sigma^2_B$  for Deli and Group B, respectively) to the total phenotypic variance of crosses. Estimates of  $h^2$  ranged from 0.21 (O/P in Deli) to 0.57 (ABW in Group B). The EBV accuracy was computed for each individual from additive variances, standard error reported with the BLUP and inbreeding coefficient. The EBV accuracy was high, ranging from  $0.86 \pm 0.06$  (SD) for O/P in Deli to  $0.93 \pm 0.04$  for K/F in Group B. Prior to use in GS modeling, EBV were transformed into deregressed estimated breeding values (DEBV) using the approach described in Garrick *et al.* (2009).

### 3.3 Definition of training and test sets

In order to investigate the GS accuracy range that could be achieved within a given population, we used three strategies to define training and test sets: (1) K-means clustering was used to separate the individuals into five subpopulations. This method minimizes the relationships between training and test sets and maximizes the relationship within training sets (Saatchi *et al.* 2011). It was expected to give the lower bound in the accuracy range; (2) A within family strategy with random partition of each full-sib family into five groups, hence each individual in the test set having full-sibs in the training set. The aim was to achieve high accuracy associated with a high relationship between the training and test sets; and (3) the so called ‘‘CDmean’’ in Rincent *et al.* (2012). This defined a training set optimized from marker data so as to achieve the highest GS accuracy when using the remaining individuals as the test set.

In all cases, the GS model was fitted using the training individuals, and the fitted model was used to obtain the GEBV of the test individuals. The K-means clustering and Within-Family strategy allowed a five-fold cross-validation. Each combination of four groups was used in turn as a training set to estimate the GEBV on individuals in the fifth group, which was used as the test set. Consequently for K-means clustering and Within-Family strategies, five GS accuracy values were obtained for each population and trait. With CDmean, only one accuracy value was obtained for each population and trait as this method yields a single optimized sample of the genotyped individuals.

The K-means clustering strategy (Saatchi et al. 2011) uses a dissimilarity matrix between individuals computed from the additive relationship matrices ( $A$ ) of each population. Five clusters were made in each population using the Hartigan and Wong algorithm.

The CDmean method (Rincent et al. 2012) optimizes sampling of the training set among the genotyped individuals. The method allocated the individuals into training (80% of the individuals) or test sets based on their genotype, in a way that maximizes the expected accuracy of GS for the dataset. The optimization criterion is the mean of the generalized coefficients of determination (CD) of contrasts between each non-phenotyped individual and the population mean. The optimization algorithm is a simple exchange algorithm.

The relationship between the training and test sets was measured by the maximum additive genetic relationship between individuals in the test and training sets ( $a_{max}$ ) (Saatchi et al. 2011). In all populations, CDmean gave a high  $a_{max}$ , the Within-Family method gave intermediate values and clustering led to low  $a_{max}$ , with one replicate with  $a_{max}$  close to zero.

### 3.4 Genomic selection statistical methods and control pedigree-based model

We used five GS statistical methods to obtain the GEBV of individuals with masked EBV present in the test sets. For comparative purposes, we also used a control pedigree-based model (PBLUP) to check the usefulness of marker information. PBLUP was applied in the same way as GS statistical methods, except that PBLUP used a pedigree-based additive relationship matrix.

The GS methods were: GBLUP (Henderson 1975; Eding and Meuwissen 2001), Bayesian Lasso regression (BLR) (Park and Casella 2008; de los Campos et al. 2009), Bayesian random regression (BRR) (Pérez et al. 2010), BayesB (Meuwissen et al. 2001; Habier et al. 2011; Pérez and de los Campos 2013) and BayesC (Habier et al. 2011; Pérez and de los Campos 2013). GBLUP and BRR methods assume a common variance  $\sigma_m^2$  for all markers (actually alleles here, as SSR are multiallelic). BL estimates a variance specific to each allele. In BayesB and BayesC, priors of allele effects include null effects with probability  $\pi$  and non-null effects associated either with allele-specific variance (BayesB) or variance common to all alleles (BayesC) with probability  $(1-\pi)$ .

As the aim of this study was to predict DEBV, we only fitted the additive effects of each allele in our models.

For GBLUP, the following model was used:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{g} + \mathbf{e}$$

where  $\mathbf{y}$  is the vector of DEBV,  $\boldsymbol{\mu}$  is the overall mean,  $\mathbf{g}$  is the vector of random additive values of individuals (GEBV),  $\mathbf{g} \sim N(0, \mathbf{G}\sigma_g^2)$  with  $\sigma_g^2$  the additive variance,  $\mathbf{G}$  the molecular coancestry matrix computed according to Eding and Meuwissen (2001) and  $\mathbf{e} \sim N(0, \sigma_e^2)$ .

For BLR, BRR, BayesB and BayesC the following model was used:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}\mathbf{m} + \mathbf{e}$$

where  $\mathbf{m}$  is the vector of allele effects,  $\mathbf{Z}$  is the incidence matrix with elements  $Z_{ij} = 0, 1$  or  $2$  depending on the number of alleles  $j$  for individual  $i$ , and  $\mathbf{e}$  is the vector of residual effects. Using estimated allele effects, the GEBV of individual  $i$  was given by:

$$\hat{g}_i = \sum_{j=1}^n Z_{ij} \hat{m}_j$$

where  $n$  is the total number of alleles and  $\hat{m}_j$  is the estimated posterior mean effect of allele  $j$  over the post burn-in iterations.

For BRR,  $\sigma_m^2$  and  $\sigma_e^2$  had scaled inverse chi-square priors with specific degrees of freedom and scales and  $\mathbf{m}$  had a normal prior  $N(0, \sigma_m^2)$ . For BLR,  $\sigma_e^2$  followed a scaled inverse chi-square prior distribution,  $m_j$  followed a normal prior  $N(0, \tau_j^2 \sigma_e^2)$  with variance specific to each allele  $j$ ,  $\tau_j^2$  followed an exponential prior with rate  $\lambda^2 / 2$  where the regularization parameter  $\lambda^2$  followed a gamma prior. For BayesB,  $\pi$  followed a beta prior,  $\sigma_e^2$  a scaled inverse chi-square prior, the variance specific to each allele  $\sigma_{mj}^2$  followed a scaled inverse chi-square prior with probability  $(1 - \pi)$  and a null value with probability  $\pi$ , and  $m_j$  followed a normal prior  $N(0, \sigma_{mj}^2)$  with probability  $(1 - \pi)$  and a null value with probability  $\pi$ . For BayesC,  $\pi$  followed a beta prior,  $\sigma_e^2$  a scaled inverse chi-square prior,  $\sigma_m^2$  followed a scaled inverse chi-square prior with probability  $(1 - \pi)$  and a null value with probability  $\pi$ , and  $\mathbf{m}$  followed a normal prior  $N(0, \sigma_m^2)$  with probability  $(1 - \pi)$  and a null value with probability  $\pi$ . For all Bayesian methods, we used 50,000 iterations and the first 12,500 iterations were discarded as burn-in.

The control pedigree-based model (PBLUP) was similar to GBLUP, except that it used the  $A$  matrix of additive relationship computed from the pedigrees, instead of  $G$ . As PBLUP only used pedigrees to model genetic covariances between individuals, it did not account for Mendelian sampling term, giving identical EBV to full-sibs in the test set. Thus, PBLUP only differentiated families, not individuals within families. Consequently, we expected GS to reach a higher accuracy than PBLUP by accounting for both family effects and Mendelian sampling terms. In order to check whether the GBLUP accuracy was higher than PBLUP, we carried out one-tailed paired sample t-tests for each of population-trait combination.

We used R-ASReml (Butler et al. 2009) for GBLUP and PBLUP and the BGLR R package (de los Campos et al. 2013) for BL, BRR, BayesB and BayesC.

### 3.5 Prediction accuracy of GEBV

Given that the true breeding values (TBV) were unknown, it was not possible to estimate the GS accuracy, which is the correlation between GEBV and TBV. Instead, we estimated the prediction accuracy, which is the correlation between GEBV and DEBV. However, as the accuracy of EBV was high (mean 0.90, ranging from  $0.86 \pm 0.06$  (SD) for O/P in Deli to  $0.93 \pm 0.04$  for K/F in Group B) the prediction accuracy was expected to be close to theoretical GS accuracy.

When investigating the correlation between the accuracy and  $a_{max}$ , a box-cox transformation was applied to  $(accuracy + 1)$  using  $\lambda = 3$  to achieve the normality of residuals. In order to identify the factors affecting the GS accuracy, an analysis of variance (ANOVA) was performed using box-cox transformed accuracy. The factors included in the ANOVA were the GS statistical methods, the methods to define training sets, the populations, the traits, the interactions between traits and populations and the replicates (within traits and methods to define the training sets).

## 4 RESULTS

### 4.1 Effect of the GS statistical method on accuracy of GEBV

ANOVA indicated that there was no effect of the GS statistical method on accuracy. There were almost perfect positive linear correlations between the accuracies of the five statistical methods used for genomic predictions, with Pearson correlations ranging from 0.982 to 0.995. Therefore, all the methods yielded similar

accuracy regardless of the population, trait and training set definition method. Consequently, we only considered the results of the GBLUP method in the rest of the study.

#### **4.2 GBLUP accuracy compared to the control pedigree-based (PBLUP) model**

In Group B population, GBLUP accuracy was significantly higher than that of PBLUP for three traits (ABW, BN and FW) (Figure 2). For those traits, the accuracy gain with GBLUP with respect to PBLUP ranged from 22% (FW) to 89% (ABW). This superiority could be explained by the fact that GBLUP accounted for both family effects and Mendelian sampling terms. Therefore, GBLUP used the LD between markers and QTL to model segregation in realized additive relationships within full-sib families. For the other traits, GBLUP and PBLUP accuracies were similar, indicating that markers failed to capture Mendelian sampling differences and revealed only family effects. To illustrate that the ability of GBLUP to capture Mendelian sampling depended on the trait, we chose two examples: the replicate 5 of K-means clustering with ABW and the replicate 3 of K-means clustering with F/B. In the first example, two large full-sib families were present in the test set and the within-family GBLUP accuracy was high in both families (accuracy of 0.508 in the selfing of individual LM2T with 20 individuals and 0.562 in the LM2T x LM5T cross with 14 individuals). This indicated that the LD between QTL and markers allowed GBLUP to account for Mendelian sampling terms. Finally, in this example, the GBLUP accuracy reached 0.588 in the whole test set and outperformed PBLUP (accuracy -0.123). The results differed for the second example, in replicate 3 of the K-means clustering with F/B in Group B. In this test set, there was one large full-sib family for which the GBLUP accuracy was null (accuracy of 0.016 in the selfing of LM5T on 10 individuals) and GBLUP accuracy in the whole test set was not higher (0.433) than the PBLUP accuracy (0.506).

In the Deli population, GBLUP failed to outperform PBLUP for all traits. Even when the mean GBLUP accuracy was higher than PBLUP (F/B, K/F, P/F), this was not significant. Therefore, markers only estimated, at best, the family effects in Deli. We hypothesized that this could be explained by the fact that the Deli population combined the lowest within-family phenotypic variance (on average 49% lower in Deli than in Group B, ranging from 71% lower for O/P to 11% lower for F/B), and the lowest marker density.

The superiority of GBLUP over PBLUP increased when  $a_{max}$  decreased (not shown) as PBLUP could not perform well when the genetic covariances between individuals were too small (i.e. when  $a_{max}$  was small), while GBLUP could.

The population effect on the GBLUP accuracy was not significant. On average, over all traits, the GBLUP accuracy was 0.50 in Deli and 0.55 in Group B. However, the population affected the PBLUP accuracy, which was the lower in Group B (0.47) than in Deli (0.54).

#### **4.3 Factors affecting the GBLUP accuracy**

There was marked variation in the GBLUP accuracy, which ranged from negative (-0.41) to very high positive values (0.94), depending on the method to define the training set, replicates, traits and traits within populations. ANOVA showed that the method to define the training set had the strongest effect on accuracy ( $F=155.1$ ), followed by interactions between traits and populations ( $F=7.0$ ), trait ( $F=5.7$ ) and replicates ( $F=3.0$ ) ( $P<0.001$  for all factors).

The effect of the method to define the training set and replicates actually reflected the effect of the relationship between training and test sets. A significant positive correlation between the accuracy of GBLUP and the maximum additive genetic relationship  $a_{max}$  was found for almost all population-trait combinations (Figure 3). The highest accuracies were obtained when the training set was optimized with CDmean. They reached 0.79 on average, ranging from 0.49 (P/F in Group B) to 0.94 (FW in Group B).

The trait-population interaction was noted mostly because the O/P accuracies in Deli (0.29) was much lower than other accuracy values and because the FW accuracy in Group B was much higher (0.71). The trait effect was due to the accuracy of O/P (mean 0.42) significantly lower than the accuracy of BN (mean 0.60).

There was a significant positive correlation between accuracy and  $h^2$  in Group B, although weak ( $P=0.020$ ,  $R^2=0.62$ ). It was not significant in Deli. This was consistent with the findings of Grattapaglia (2014), who indicated that although  $h^2$  affected the GS accuracy, its effect was actually secondary. Moreover, we used DEBV as records and the\_deregression process reduces the effect of  $h^2$  on GS accuracy (Saatchi et al. 2011).

## 5 DISCUSSION

We found that, for one of the parental populations currently used in conventional reciprocal recurrent selection (Group B), genomic selection (GS) gave accuracies at least comparable or superior depending on traits to those from pedigree-based model (PBLUP) when predicting the EBV of individuals with no data records (ie not progeny tested). For Deli population, however, results were not as conclusive, with no detectable differences between accuracies between the two evaluation methods across targeted traits. In any case, GS appeared to be a valuable method for oil palm breeding.

Therefore, like the simulation study of Wong and Bernardo (2008), we confirmed the usefulness of GS for this species. Wong and Bernardo (2008) concluded that the genetic gain per year of GS would be higher than that of phenotypic selection if the training set had more than 50 individuals. Such a small training set was detrimental to the GEBV accuracy, but as the length of the breeding cycle with selection on markers alone was shortened to six years, the genetic gain per year ultimately increased. A novel aspect brought by our analysis is the effect of a real (complex) breeding population when assessing GS for the species. Indeed, we showed that reducing the need of progeny tests only to the generation used to train the GS model would be more difficult than in simulations of Wong and Bernardo (2008), where training was done over the result of single crosses. Some of the critical points regarding the performance of GS highlighted by our analyses are developed in the following sections.

### 5.1 Information captured by markers

The contrasted history of the two breeding populations likely explained the differences in performance of GBLUP relative to PBLUP among traits and populations as well as trait by population interactions on the GBLUP accuracy. There were differences in the number, location and polymorphism of SSR markers among populations, which indicated differences at the level of QTL. Each population suffered from different bottleneck events, were subjected to independent selection

regimes and distinct drift effects. The Deli population combined a narrow genetic base of four founders and the longest artificial selection and inbreeding history among the two populations. This might explain the fact that Deli had the lowest within-family phenotypic variance and narrow Mendelian sampling terms. In addition, Deli pedigree showed over several generations a high differentiation between families. In this context of family differentiation and low within-family variance, the potential advantage of GBLUP over PBLUP was therefore smaller in Deli than that observed in Group B population. Another consequence of the Deli population history is that it had the lowest marker density (due to lower polymorphism). The two populations had similar small effective sizes ( $N_e < 10$ ) (Cros et al. 2014) and therefore the lower marker density in Deli led to a lower LD between markers and QTL compared to Group B. Finally, the marker density in Deli appeared to be insufficient to allow GBLUP to generate good estimates of Mendelian sampling terms for individuals with no phenotypic records, and therefore GBLUP did not perform better than PBLUP. By contrast, the Group B population had higher within-family phenotypic variance and higher marker density than Deli, indicating that GBLUP could have a very marked advantage over PBLUP in Group B.

GS utilizes the additive genetic relationship between training and test sets and LD between markers and QTL to estimate GEBV, which accounts for both family effects and Mendelian sampling terms (Habier et al. 2007; Habier et al. 2010; Daetwyler et al. 2013). The proportion of GS accuracy coming from relationship and LD varies depending in particular on the marker density and training set size. Jannink *et al.* (2010) showed that when a small training size (400 individuals) was combined with a small number of markers (400 SNP), a large part of the GBLUP accuracy came from the relationship. This is what we observed empirically. LD information is of greater interest for the practical application of GS as it is more persistent than the relationship over generations (Habier et al. 2007). The challenge is thus to increase the proportion of accuracy due to LD. This could be achieved by increasing the training set size and marker number.

The highest superiority of GBLUP over PBLUP was obtained when  $a_{max}$  was small, i.e. when, according to the pedigree, the training and test sets were loosely related or unrelated. However, the pedigrees sometimes were not deep enough as to end up with unrelated founders, allowing for some individuals to appear erroneously as unrelated. In such cases, marker information brought advantages to GS, as they could capture hidden relationships between individuals, as well as possible identical-by-state QTL and markers between individuals.

Surprisingly, the PBLUP accuracy could be high, in particular when optimizing the training set with CDmean. The high accuracies obtained with PBLUP were due to the ability of the pedigree to model this structure. With GS, if high accuracies are obtained solely as a result of family differences, only selection between families can be carried out, with no possibility of selecting within families. This would lead to a marked increase in inbreeding and reduce future genetic progress. Therefore, in order to be useful for practical breeding, GS must account for two parts of breeding values, i.e. family effects and Mendelian sampling terms.

We studied eight traits, assuming there should be variations in genetic architecture among them, in particular in the number of QTL, as some traits could be less complex than others. Several authors using real data reported that there was no effect of the statistical method used to estimate GEBV (Heslot et al. 2012; Kumar et al. 2012; Daetwyler et al. 2013). We assumed that the results we obtained with five methods were similar due to the limited number of individuals.

## 5.2 Definition of training sets

Using K-means clustering, within-family and CDmean to define the training and test sets gave more valuable information on the GS accuracy than simple replicates with random assignation, as the different methods substantially affected the relationship between the training and test sets. We observed a marked decrease in GS accuracy with decreasing maximum additive genetic relationships ( $a_{max}$ ) between the training and test sets. This was similar to the results obtained by Habier *et al.* (2010) in Holstein cattle.

The use of the optimization algorithm, based on a CDmean maximized relationship between training and test sets and a minimized relationship within the training set, yielded the highest GS accuracies. CDmean therefore appeared to be the best method. In a practical use of GS, all individuals in the generation(s) used to calibrate the model would be genotyped at juvenile stage and CDmean would be applied to identify the subset of individuals to progeny test. Finally, selection would be made based on GEBV among all individuals, either both genotyped and progeny-tested or only genotyped. This subset would make an optimized training population, i.e. the one maximizing the GS accuracy. In our study, we defined an optimized training set specific to each trait using the corresponding heritability ( $h^2$ ) values. Obviously, for practical application, it would be necessary to use a mean value of  $h^2$  over traits that must be selected. This should have a negligible effect on the accuracy, as Rincent *et al.* (2012) showed that the CDmean method is robust to  $h^2$  variation.

## 5.3 Practical aspects of GS in oil palm

In the perspective of an optimal use of GS that would allow making selection on markers alone and limiting the use of progeny tests to the training of the GS model, oil palm breeding should evolve toward a reciprocal recurrent genomic selection breeding scheme integrating marker data to increase the selection intensity and decrease the length of breeding cycles (Figure 1). In this scheme, GS could be applied among individuals that have not been progeny tested and that belong to the same generation as the training individuals or to the following generation(s). As less effort would be required for genotyping candidate individuals than progeny testing them, GS could increase the selection intensity as compared to conventional breeding. In addition, if the GS accuracy is high enough to conduct selection solely on markers in the generation(s) following training, the length of the breeding cycle would decrease, as progeny tests would only be made in the generation used to train the model. However, this would only be possible if the GS accuracy were high enough for all the yield components. In Group B and La Mé, the accuracy for some key oil yield components (especially average bunch weight [ABW] and bunch number [BN]) in the test sets was higher with GS models than with the pedigree-based control model (PBLUP). The markers could thus be used for preselection before progeny tests by identifying genetically superior individuals for ABW and BN, which would subsequently be progeny tested to finalize selection on these two traits (as the accuracy of EBV from conventional progeny tests is higher than the GEBV accuracy), and for phenotypic-based selection on the other yield components with lower GBLUP accuracy. This would increase the intensity of selection on ABW and BN, thus increasing the rate of genetic gain for yield. Obviously, this would not tap the full potential of GS, which could only be achieved if GS reduced the need for progeny tests. This will not be possible as far as there is not a clear-cut advantage of the GS models over pedigree-based models for all yield traits. Considering that the

new scheme would alternate one generation of progeny tests to calibrate the GS model with one generation of selection on markers alone, the length of two cycles would be only 60% of the current length. This new breeding scheme will be a credible alternative when, for all yield components, GS will be able to account for the Mendelian sampling terms and will have a mean accuracy over two cycles higher than 60% of the accuracy of current reciprocal recurrent selection (RRS), i.e. higher than 0.51.

In order to validate our new breeding scheme integrating GS, the first points to investigate are the effects on accuracy of larger training sets and a larger number of markers, to identify how many individuals and markers are required for GS to outperform pedigree information for all traits and populations. The increase in the number of markers could be achieved by genotyping all individuals with next generation sequencing or with a SNP chip, which could be developed using the whole genome sequence (Singh et al. 2013). Another crucial question to be addressed is the decrease in GS accuracy when applying the model in the generation following training. The first results of progeny tests of the next breeding cycle will be available within a few years. They will be used to estimate the effect of a larger training set and a larger number of markers on the GS accuracy, as well as the decrease in accuracy when applying GS models in a test set generated by the crossing of individuals selected in the training generation.

## 6 ACKNOWLEDGMENTS

We acknowledge SOCFINDO (Indonesia) and CRAPP (Benin) for planning and carrying out the field trials with CIRAD (France) and authorizing use of the phenotypic data for this study. This research was partly funded by a grant from PalmElit SAS. We thank P. Sampers, C. Carrasco-Lacombe, A. Manez and S. Tisé for help in genotyping and L. Dedieu for reviewing the manuscript.

## 7 REFERENCES

- Billotte N, Marseillac N, Risterucci AM et al. (2005) Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). *Theor Appl Genet* 110:754-765
- Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) Mixed Models for S language environments: ASReml-R reference manual (Version 3). Queensland Department of Primary Industries and Fisheries
- Cochard B, Adon B, Rekima S et al. (2009) Geographic and genetic structure of African oil palm diversity suggests new approaches to breeding. *Tree Genetics & Genomes* 5:493-504
- Corley RHV (2009) How much palm oil do we need? *Environmental Science & Policy* 12:134-139
- Corley RHV, Tinker PB (2003) Selection and breeding. The oil palm, 4th edn. Blackwell Science Ltd Blackwell Publishing, Oxford, pp 133-199
- Cros D, Sánchez L, Cochard B et al. (2014) Estimation of genealogical coancestry in plant species using a pedigree reconstruction algorithm and application to an oil palm breeding population. *Theor Appl Genet* 127:981-994
- Daetwyler HD, Calus MPL, Pong-Wong R, de los Campos G, Hickey JM (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193:347-365

- de los Campos G, Naya H, Gianola D et al. (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182:375-385
- de los Campos G, Pérez P, Vazquez A, Crossa J (2013) Genome-enabled prediction using the BLR (Bayesian linear regression) R-Package. In: Gondro C, van der Werf J, Hayes B (eds) *Genome-wide association studies and genomic prediction*. Humana Press, pp 299-320
- Eding H, Meuwissen THE (2001) Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *Journal of Animal Breeding and Genetics* 118:141-159
- Garrick D, Taylor J, Fernando R (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution* 41:55
- Gascon JP, de Berchoux C (1964) Caractéristique de la production d'*Elaeis guineensis* (Jacq.) de diverses origines et de leurs croisements. Application à la sélection du palmier à huile. *Oleagineux* 19:75-84
- Grattapaglia D (2014) Breeding forest trees by genomic selection: current progress and the way forward. In: Tuberosa R, Graner A, Frison E (eds) *Genomics of Plant Genetic Resources*. Springer Netherlands, pp 651-682
- Habier D, Fernando R, Kizilkaya K, Garrick D (2011) Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186
- Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389-2397
- Habier D, Tetens J, Seefried F-R, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution* 42:5
- Henderson C (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423-447
- Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic selection in plant breeding: a comparison of models. *Crop Sci* 52:146-160
- Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics* 9:166-177
- Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, Carlisle C (2012) Genomic selection for fruit quality traits in apple (*Malus domestica* Borkh.). *Plos One* 7:e36674
- Lorenz AJ, Chao S, Asoro FG et al. (2011) Genomic selection in plant breeding: Knowledge and prospects. *Advances in Agronomy*. D. L. Sparks, pp 77-123
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829
- Park T, Casella G (2008) The Bayesian LASSO. *Journal of the American Statistical Association* 103:681-686
- Pérez P, de los Campos G (2013) BGLR: a statistical package for whole genome regression and prediction
- Pérez P, de los Campos G, Crossa J, Gianola D (2010) Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. *Plant Genome* 3:106-116
- Rincent R, Laloe D, Nicolas S et al. (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics* 192:715-728
- Saatchi M, McClure M, McKay S et al. (2011) Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genetics Selection Evolution* 43:40

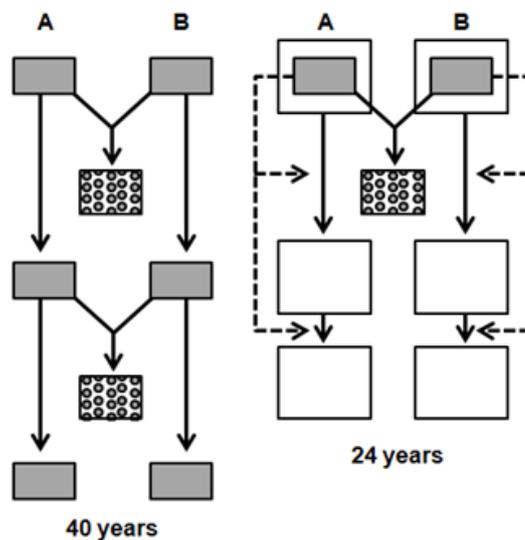
Singh R, Ong-Abdullah M, Low E-TL et al. (2013) Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature advance online publication

Tranbarger T, Kluabmongkol W, Sangsrakru D et al. (2012) SSR markers in transcripts of genes linked to post-transcriptional and transcriptional regulatory functions during vegetative and reproductive development of *Elaeis guineensis*. BMC Plant Biology 12:1

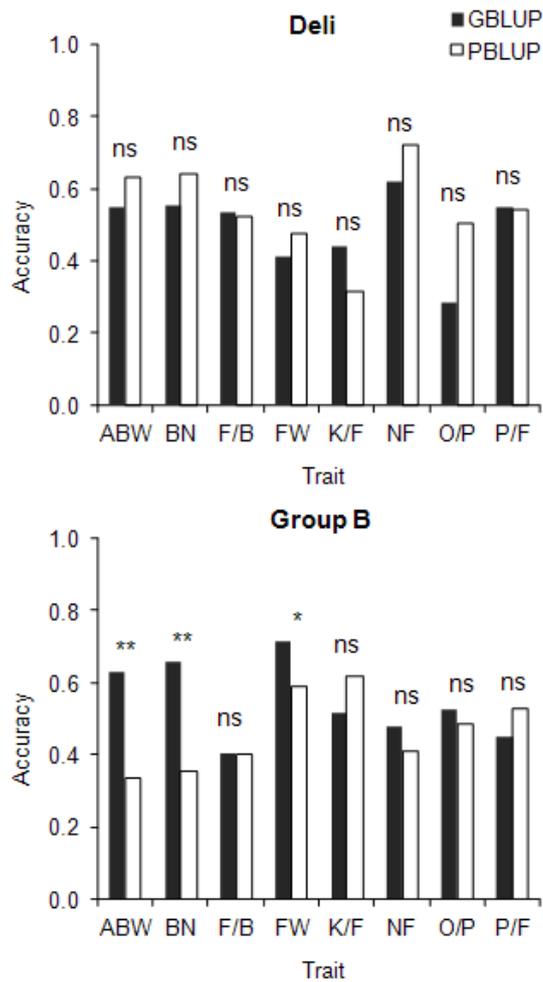
USDA (2013) Oilseeds: world market and trade. Foreign Agricultural Service, Circular Series

Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116:815-824

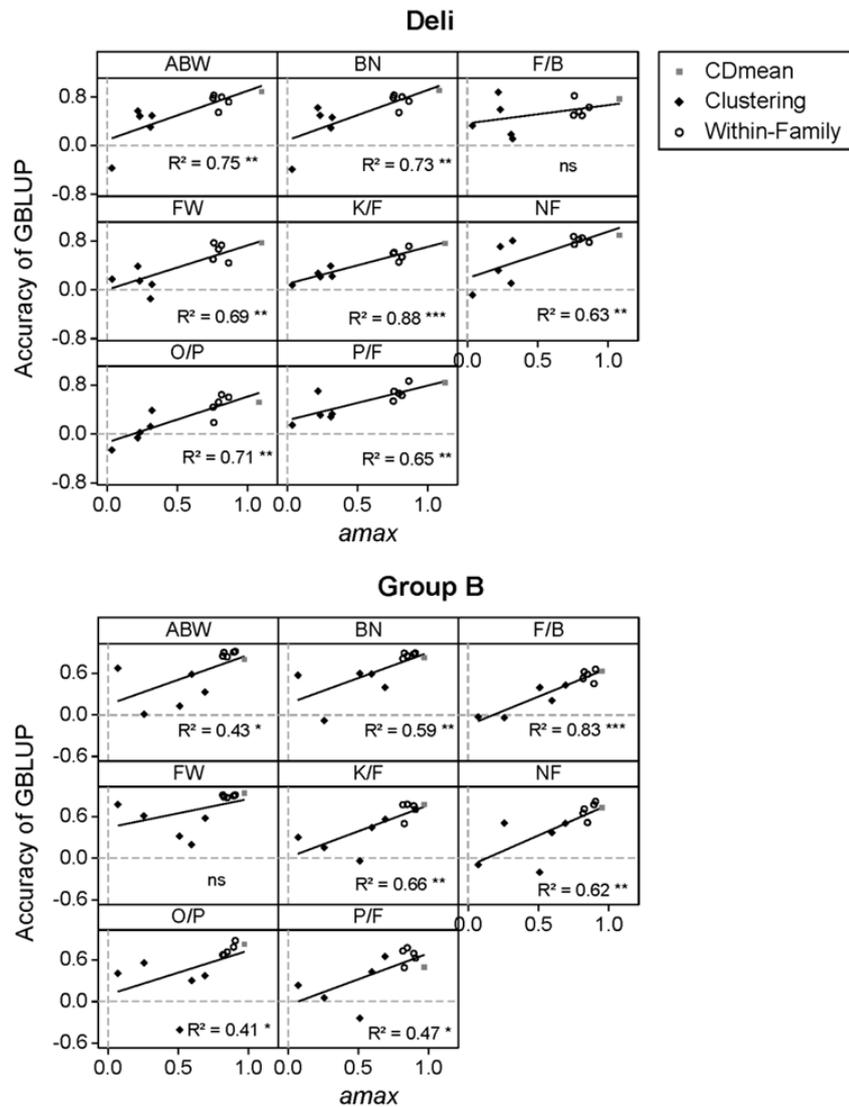
## 8 TITLES AND FIGURE LEGENDS



**Figure 1** Reciprocal recurrent selection (RRS, left) versus reciprocal recurrent genomic selection (GS, right). One cycle of conventional RRS requires 20 years due to preselection before progeny tests made on the most heritable traits, progeny tests and recombination between selected individuals. For GS, 24 years are enough to complete two cycles, with 18 years for the first cycle used to calibrate the GS model (preselection on heritable traits is no longer necessary) and 6 years to complete the second cycle with selection on markers alone. For GS, selection could be made among individuals that have not been progeny tested and that belong either to the same generation as the training individuals or to the following generation(s). Filled blocks: individuals progeny tested (RRS) or progeny tested and genotyped (GS). Dashed blocks: phenotyped individuals (genetic trials). Blanked blocks: individuals genotyped but not progeny tested. Dashed lines: application of GS.



**Figure 2** Mean accuracy of the GS model (GBLUP) and control pedigree-based model (PBLUP) in Deli and Group B (n=11). One-tailed paired sample t-tests were performed to check whether the accuracy of GBLUP > PBLUP. Significance of t-tests: \*  $0.05 > P \geq 0.01$ , \*\*  $0.01 > P \geq 0.001$ , ns = not significant.



**Figure 3** Accuracy of GBLUP versus the maximum additive genetic relationship ( $a_{max}$ ) according to the population (Deli and Group B) and trait (ABW: average bunch weight, BN: bunch number, FW: fruit weight, NF: number of fruits per bunch, F/B: fruits to bunch ratio, P/F: pulp to fruit ratio, O/P: oil to pulp ratio and K/F: kernel to fruit ratio). Each dot indicates the accuracy value obtained in one test set. The symbols of the dots indicate the method used to define the training and test sets (K-means clustering, Within-Family and CDmean). The  $R^2$  values are the coefficients of determination of the linear regression between box-cox transformed accuracy of GBLUP and  $a_{max}$ . Accuracy of GBLUP was box-cox transformed prior to regression analysis. Significance of the correlation: ns: not significant, \*  $0.05 > P \geq 0.01$ , \*\*  $0.01 > P \geq 0.001$ , \*\*\*  $0.001 > P$ .