Theor Appl Genet (2015) 128:397–410
DOI 10.1007/s00122-014-2439-z

ORIGINAL PAPER

Genomic selection prediction accuracy in a perennial crop: case study of oil palm (*Elaeis guineensis* Jacq.)

David Cros · Marie Denis · Leopoldo Sánchez · Benoit Cochard · Albert Flori · Tristan Durand-Gasselin · Bruno Nouy · Alphonse Omoré · Virginie Pomiès · Virginie Riou · Edyana Suryana · Jean-Marc Bouvet

Received: 28 April 2014 / Accepted: 27 November 2014 / Published online: 7 December 2014

Abstract

Key message **Genomic selection empirically appeared valuable for reciprocal recurrent selection in oil palm as it could account for family effects and Mendelian sampling terms, despite small populations and low marker density.**

Abstract Genomic selection (GS) can increase the genetic gain in plants. In perennial crops, this is expected mainly through shortened breeding cycles and increased selection intensity, which requires sufficient GS accuracy in selection candidates, despite often small training populations. Our objective was to obtain the first empirical estimate of GS accuracy in oil palm (*Elaeis guineensis*), the major world oil crop. 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 within-population GS accuracies when predicting breeding values of non-progeny-tested individuals 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 GS could account for family effects and Mendelian sampling terms in Group B but only for family effects in Deli. Presumably, this difference between populations originated from their contrasting breeding history. The GS accuracy ranged from −0.41 to 0.94 and was positively correlated with the relationship between training and test sets. Training sets optimized with the so-called CDmean criterion gave the highest accuracies, ranging from 0.49 (pulp to fruit ratio in Group B) to 0.94 (fruit weight in Group B). The statistical methods did not affect the accuracy. Finally, Group B could be preselected for progeny tests by applying GS to key yield traits, therefore increasing the selection intensity. Our results should be valuable for breeding programs with small populations, long breeding cycles, or reduced effective size.

Communicated by Chris Carolin Schön.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2439-z) contains supplementary material, which is available to authorized users.

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

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

A. Omoré
INRAB, CRAPP, Pobè, Benin

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

Introduction

Genomic selection (GS) is a form of marker-assisted selection that can improve 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 marker. 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 gain per year is expected to 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, 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). The Mendelian sampling term originates from the random sampling of the parental gametes. It represents the deviation between the additive value of the individual and the average breeding value of its parents (Daetwyler et al. 2007, 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 plant species with long breeding cycles (>10 years) (see Grattapaglia 2014; Isik 2014 for reviews), and to our knowledge only Zapata-Valenzuela et al. (2012) assessed GS with a limited number of phenotyped individuals. Oil palm (*Elaeis guineensis*) is a diploid, monoecious, and allogamous perennial crop with high GS potential due to its conventional breeding system. It is the major world oil crop, with a production over 55 Mt (USDA 2013) which is expected to further increase substantially as demand for palm oil could be between 120 and 156 Mt in 2050 (Corley 2009). Currently, oil palm genetic improvement is generally based on the reciprocal recurrent selection (RRS) scheme designed in the 1950s (Gascon and de Berchoux 1964). It relies on two populations, the Deli (of Asian origin) and the Group B (a mixture of African populations), used as parents of the commercial hybrids. Phenotypically, these populations differ, with Deli producing a small number of large bunches and Group B a large number of small bunches. Also, they have different histories: Deli has fewer founders (4) than Group B (15–20) and was submitted to more generations of selection, inbreeding, and genetic drift, as Deli founders were planted in 1848 and Group B founders were collected in the first half of the Twentieth century. In addition, the mass selection that was applied in both populations differed in its intensity, traits of interest, etc. The RRS scheme 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. Candidate palms sampled from full-sib families in each of the two populations are progeny tested in Deli × Group B crosses and evaluated in extensive field trials, in order to get reliable estimated breeding values (EBV, with accuracy between 0.80 and 0.90 for all yield components). The best individuals are selected within each parental population to produce the following generation and commercial hybrids. Therefore, conventional breeding in oil palm is costly and time consuming, with a long breeding cycle (around 20 years, while sexual maturity is reached at around 3 years of age) and a limited number of tested individuals. The private oil palm breeding sector is thus seeking a practical implementation of GS that would increase the annual rate of genetic gain. In this species, the main GS challenge is currently to achieve accuracy of GEBV high enough to allow selecting among individuals that have not been progeny tested, despite the small training sets that are available (<200 progeny-tested individuals per population and generation). The growing number of transcriptomic studies (e.g., Tranbarger et al. 2011; Dussert et al. 2013; Tee et al. 2013) and the fact that the whole genome sequence is now available (Singh et al. 2013) will facilitate the development of large numbers of SNP markers, which in turn will boost GS applications. Oil palm could therefore become a model species for GS in plants, especially for species with a long breeding cycle and/or limited phenotypic records.

The only study in which the GS potential was investigated in oil palm is a simulation by Wong and Bernardo (2008), which yielded promising results. However, their results might not be easily generalized as the simulated breeding populations resulted from selfing a hybrid between two inbred lines, while real breeding populations are more complex. Therefore, an empirical study appeared necessary.

Our objective here was to assess the potential of GS in the context of current oil palm RRS breeding by obtaining the first empirical estimate of GS accuracy using the largest EBV and genotype datasets available for the species. Specifically, we investigated a within-population GS strategy for Deli and Group B populations (see Fig. 1 for details). For this purpose, we used individuals with microsatellite (SSR) genotypes and deregressed EBV (DEBV) that were obtained from interpopulation progeny tests. Within each population, cross-validation was performed in order to assess the prediction accuracy of GS, as the ability to predict the breeding value of individuals that were not progeny tested. We aimed at quantifying 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 on the basis of their genetic relationships; (2) the genetic architecture of the trait, for which we

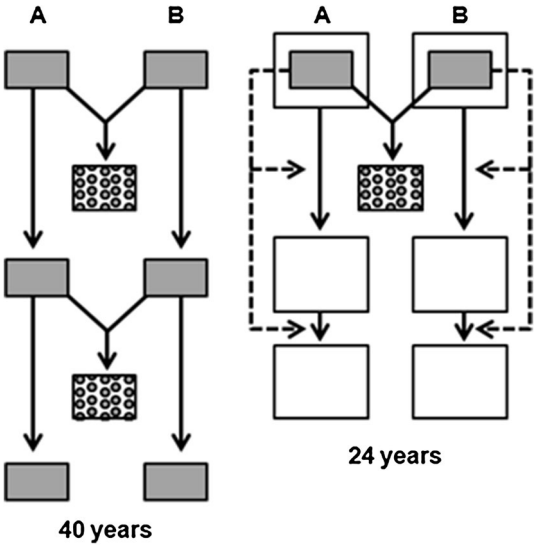

Fig. 1 Reciprocal recurrent selection (RRS, *left*) versus reciprocal recurrent genomic selection (GS, *right*). One cycle of conventional RRS requires 20 years due to preselection before progeny tests made on the most heritable traits, progeny tests, and recombination between selected individuals. For 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

studied eight yield traits; (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.

Materials and methods

The data available (i.e., individuals with both EBV and genotypes) represented 131 Deli and 131 Group B individuals. The progeny tests to obtain EBV required around 350 ha and 15 years of data records, illustrating the difficulty to build large training sets in oil palm. Individuals were genotyped with 265 SSR.

Populations and molecular data

All individuals belonged to families from the commercial oil palm breeding program of PalmElit, a leading oil palm breeding company (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 Twentieth century. Inbreeding was commonly used, by selfing or mating related selected individuals (Corley and Tinker 2003). The 131 Group B individuals included 94 La Mé (Côte d'Ivoire), 24 Yangambi (Democratic Republic of the Congo), 4 La Mé × Yangambi, 7 La Mé × 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 Twentieth century. In particular, the Congo population originated from around ten individuals, one of which being over 50 % represented, and La Mé originated from three individuals (Cochard et al. 2009). African populations were also submitted to inbreeding and selection for yield. The inbreeding effective population size (N_e) calculated with LDNE software (Waples and Do 2008) as described by Cros et al. (2014) was 5.0 ± 1.1 (SD) for Deli and 3.9 ± 0.8 for Group B. The 131 Deli and 131 B individuals spread over three generations. From the eldest to the most recent generation, the individuals were as follows: eight Deli and seven of Group B, 89 Deli and 99 of Group B, and 34 Deli and 25 of group B (see pedigrees in Supplementary Figs. S1 and S2). The 15 individuals of the eldest generation were selected at the end of the first RRS cycle and the others were tested in the second 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). The polymorphic SSR had 2.7 ± 0.8 alleles in Deli and 6.2 ± 2.2 in Group B. For GS analysis, alleles with a frequency of under 0.05 in the training set were excluded. BEAGLE 3.3.2 software (Browning and Browning 2007) was used for imputing sporadic missing SSR genotypes, which represented 1.74 % of the data in Deli and 2.90 % in Group B. Molecular coancestry (i.e., kinship) calculated according to Lynch (1988) and Li et al. (1993) 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 are presented in Fig. 2 and indicated that the populations were highly structured.

Estimation of breeding values used as data records for GS

Prior to the GS analysis, we calculated the estimated breeding values (EBV) of the 131 Deli and 131 Group B individuals. This was done using the traditional BLUP methodology (T-BLUP) (Henderson 1975), using their

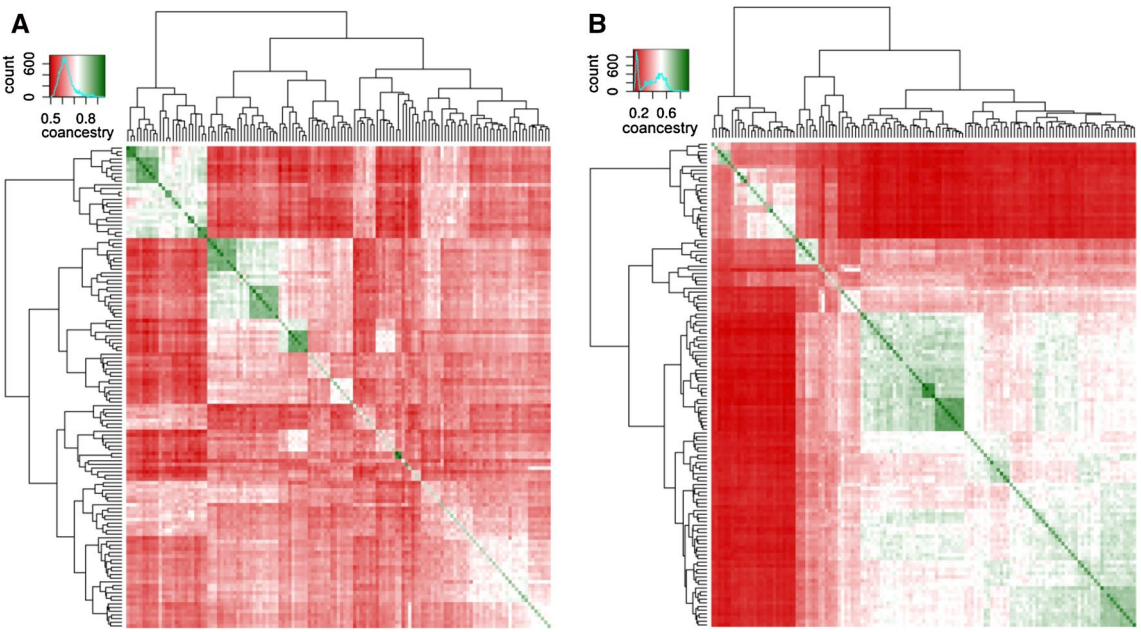

Fig. 2 Heat map of the molecular coancestry matrices of the **a** 131 Deli individuals obtained with 220 polymorphic SSR markers and the **b** 131 individuals of Group B obtained with 260 polymorphic SSR markers

pedigree and the data of their progeny tests, conducted in a large-scale experiment at Aek Loba (Sumatra). Eight traits were considered at adult age: bunch number (BN), average bunch weight (ABW), fruit-to-bunch (F/B), pulp-to-fruit (P/F), kernel-to-fruit (K/F), and oil-to-pulp (O/P) ratios, number of fruits per bunch (NF), and the average fruit weight (FW). The details about the computation of the EBV are given in Appendix. Estimates of the narrow-sense heritability (h^2) of each trait were obtained at the experimental design level from the T-BLUP analysis as the ratio of additive variances (σ^2_{Deli} and σ^2_B for Deli and Group B, respectively) to the total phenotypic variance of crosses. The EBV accuracy was computed from the prediction error variance reported with the BLUP of each individual, the additive variances and inbreeding coefficients (See Appendix). T-BLUP shrinks individual EBV toward the parental average, thus invalidating their use as records for GS or association studies. This shrinkage, however, can be corrected by deregressing the EBV. The use of deregressed EBV (DEBV) as data records for genomic selection has proved to be beneficial compared to the use of EBV (Ostersen et al. 2011; Gao et al. 2013). Deregressed EBV can be obtained directly from existing evaluations. It appears to be equivalent to the use of other indirect methods commonly used, like daughter-yield deviation in livestock (Thomsen et al. 2001). To transform EBV into DEBV we used the approach described in Garrick et al. (2009), as previously applied in eucalyptus (Resende et al. 2012).

Definition of training and test sets

In order to investigate the range of GS accuracy 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 had 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) using an optimization method, termed "CDmean" (Rincent et al. 2012), that maximized the expected accuracy of GS for the dataset. 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 DEBV and genotype of the training individuals, and the fitted model was used to obtain the GEBV of the test individuals from their genotype. The K-means clustering and Within-Family strategy allowed a fivefold 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 uses a dissimilarity matrix between individuals computed from the additive relationship matrices (***A***) of each population, according to Saatchi et al. (2011). Five clusters were made in each population using the Hartigan and Wong algorithm, implemented in the R software (R Core Team 2013).

The CDmean method (Rincent et al. 2012) optimizes sampling of the training set among the genotyped individuals. The method allocated the individuals into training 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 parameters used were the additive and residual variances obtained from the mixed model that produced the initial EBV, with 16,000 iterations and 80 % of the individuals assigned to the training set.

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 order to measure the relationships between individuals in a training set, $a_{\max}$ was also calculated within training sets ($a_{\max\ \mathrm{TRAINING}}$).

Table 1 summarizes the characteristics of the obtained training sets.

Table 1 Characteristics of the training sets used in each population (Deli population and Group B which is a mixture of various African populations)

	Population	
	Deli	Group B
Number of individuals per group: K-means clustering	25, 47, 14, 29, 16	12, 16, 32, 19, 52
Within-family	27, 26, 28, 25, 25	29, 25, 23, 31, 23
Mean size of training set (range)[a]	104.8 (84–117)	104.8 (79–119)
Mean number of polymorphic markers[a]	219.8 (209–223)	260.9 (259–263)
Mean number of alleles in training set (range)[a]	533.3 (504–544)	959.7 (794–1,158)

[a] Mean over 11 values (five for clustering, five for Within-Family and one for CDmean)

Genomic selection statistical methods and control pedigree-based model

We used five GS statistical methods to obtain the GEBV of test individuals. 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 instead of marker data to model the dependencies between training and test individuals.

The GS methods were the GBLUP, which is a linear mixed model (Henderson 1975) using a molecular additive relationship matrices ***G*** (Lynch 1988; Li et al. 1993), and four Bayesian methods: Bayesian Lasso regression (BLR) (Park and Casella 2008; de los Campos et al. 2009), Bayesian random regression (BRR) (Pérez et al. 2010), BayesCπ (Habier et al. 2011; de los Campos et al. 2013), and BayesDπ (Habier et al. 2011; de los Campos et al. 2013). GBLUP and BRR methods assume a common variance σ^2_{m} for all markers (actually alleles here, as SSR are multiallelic). BLR estimates a variance specific to each allele. In BayesCπ and BayesDπ, a priori an allele effect is zero with a probability π and non-zero either with variance common to all alleles (BayesCπ) or allele-specific variance (BayesDπ) with probability (1-π). In both approaches, the parameter π is considered unknown and estimated from the data. As the aim of this study was to predict DEBV, we only fitted the additive effects of each allele in our models. Due to the multiallelic nature of SSR markers, the molecular data were arranged into a matrix ***Z*** with alleles in columns (instead of markers when dealing with SNP) and individuals in rows, and elements $Z_{ij} = 0$, 1, or 2 depending on the number of alleles j for individual i. For all GS methods, we used an heterogeneous residual variance depending on the reliability of the EBV on the individual, as described in Garrick et al. (2009).

For GBLUP, the following model was used:

$$y = \mathbf{1}\mu + \boldsymbol{Xg} + \boldsymbol{e},$$

where y is the vector of DEBV, μ is the overall mean, $\mathbf{1}$ is a column vector of 1s, $\boldsymbol{g}$ is the vector of random additive values of individuals (GEBV) following $N(0, \boldsymbol{G}\sigma^2_g)$ with σ^2_g the additive variance and $\boldsymbol{G}$ the molecular relationship matrix, $\boldsymbol{X}$ is a diagonal design matrix and $\boldsymbol{e}$ is the vector of residual effects following $N(0, \sigma^2_e)$, with σ^2_e the residual variance. $\boldsymbol{G}$ contained the similarity indices of Lynch (1988) and Li et al. (1993), which can be applied to multiallelic markers and are unbiased estimators of coancestry when assuming founder alleles were unique (Eding and Meuwissen 2001). This is equivalent to $\boldsymbol{G} = \boldsymbol{Z}\,{}^{t}(\boldsymbol{Z})/\,4q$, with q the number of markers and ${}^{t}(\boldsymbol{Z})$ the transpose matrix of $\boldsymbol{Z}$.

The BRR, BLR, BayesCπ, and BayesDπ statistical methods estimated allele effects using the following model:

$$y = \mathbf{1}\mu + \boldsymbol{Zm} + \boldsymbol{e},$$

where $\boldsymbol{m}$ is the vector of allele 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, σ_m^2 and σ_e^2 had scaled inverse Chi-square priors with specific degrees of freedom and scales and $\boldsymbol{m}$ had a normal prior $N(0, \sigma_m^2)$. For BLR, σ_e^2 followed a scaled inverse Chi-square prior distribution, m_j followed a conditional Gaussian prior distribution $N(0, \tau_j^2\,\sigma_e^2)$ with variance specific at each allele j where τ_j^2 followed an exponential prior with rate $\lambda^2/2$ and the regularization parameter λ^2 followed a gamma prior. For BayesCπ, π followed a beta prior, σ_e^2 followed a scaled inverse Chi-square prior, the conditional prior distribution of m was a Gaussian distribution $N(0, \sigma_m^2)$ with probability $(1-\pi)$ and a null value with probability π, *and* σ_m^2 followed a scaled inverse Chi-square prior. For BayesDπ, π followed a beta prior, σ_e^2 followed a scaled inverse Chi-square prior, the conditional prior distribution of m_j was a Gaussian distribution $N(0, \sigma_{mj}^2)$ with probability $(1-\pi)$ and a null value with probability π, and the allele-specific variance σ_{mj}^2 followed a scaled inverse Chi-square prior with the scale parameter treated as unknown and following a Gamma(1,1) prior. For all Bayesian methods, we used 50,000 iterations with the first 12,500 as burn-in and a thinning interval of 10.

The control pedigree-based model (PBLUP) was similar to GBLUP, except that it used the $\boldsymbol{A}$ matrix of additive relationship computed from the pedigrees, instead of $\boldsymbol{G}$. 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 BLR, BRR, BayesCπ, and BayesDπ.

Prediction accuracy and bias 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 in oil palm progeny tests is high (between 0.80 and 0.90) 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).

The prediction bias was estimated by comparing the regression of DEBV on GEBV and its expected value of one. The slope of the regression of GEBV on DEBV was thus calculated for each trait using simple linear regression.

Results

Effect of the GS statistical method on accuracy and bias of GEBV

ANOVA indicated that there was no effect of the GS statistical method on accuracy. This point is illustrated by Supplementary Fig. S3, which shows 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. The same conclusion was reached with respect to the bias, which was similar for all the statistical methods (not shown). Consequently, we only considered the results of the GBLUP method in the rest of the study.

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

In Group B, GBLUP accuracy was significantly higher than that of PBLUP for three traits (ABW, BN and FW) (Fig. 3). For those traits, the accuracy gain with GBLUP 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 (individual deviations from family effects). For the other traits, GBLUP and PBLUP accuracies were similar, indicating that markers failed to capture Mendelian sampling differences and only revealed, at best, family effects. The ability of GBLUP to capture Mendelian sampling terms was also illustrated by the existence of significant correlations

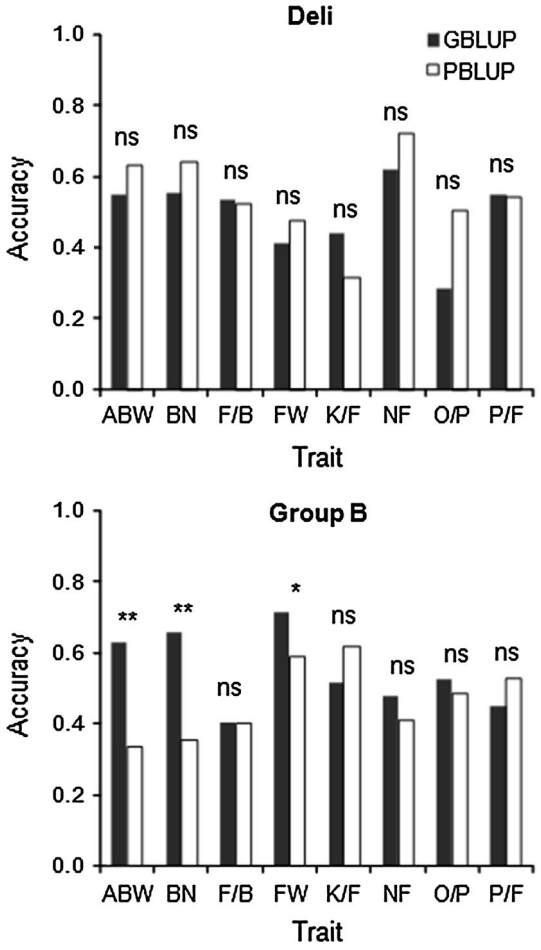

Fig. 3 Mean accuracy of the GS model (GBLUP) and control pedigree-based model (PBLUP) in Deli and Group B ($n = 11$). One-tailed paired sample t-tests were performed to check whether the accuracy of GBLUP > PBLUP. Significance of t tests: *0.05 > $P \geq 0.01$, **0.01 > $P \geq 0.001$, *ns* not significant. Values are means over 11 accuracy estimates (five for clustering, five for within-family, and one for CDmean)

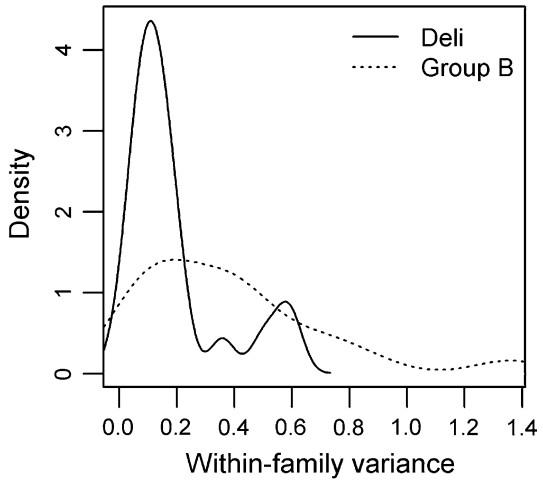

Fig. 4 Distribution of within-family variance for estimated breeding values of average bunch weight according to population. Mean within-family variance was 0.19 for Deli population and 0.33 for Group B. 15 full-sib families of Deli were used for this calculation and 14 of Group B

between GEBV and DEBV within-full-sib families. For example, in the replicate 5 of K-means clustering in group B, the within-family GBLUP accuracy was high for ABW in the two large full-sib families that were present in the test set, reaching 0.508 in the selfing of individual LM2T (20 individuals, $p < 0.05$) and 0.562 in the LM2T × LM5T cross (14 individuals, $p < 0.05$). In this example, the GBLUP accuracy reached 0.588 in the whole test set and outperformed PBLUP (accuracy −0.123). However, GBLUP was not able to estimate Mendelian sampling terms in all cases. Thus, in the replicate 3 of K-means clustering in group B, the GBLUP accuracy was null for F/B in the largest full-sib family that was present in the test set (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 in Deli test individuals, the markers (like the pedigree) only allowed estimating, at best, the family effects. The Deli population was also the one presenting the lowest within-family phenotypic variance, which was on average 40 % lower in Deli than in Group B, ranging from 69 % lower for O/P to 14 % lower for F/B (see example of ABW in Fig. 4), less polymorphic markers and lower marker density; and all these conditions could impair the advantage of GBLUP over that of PBLUP.

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).

Factors affecting the GBLUP accuracy

There was marked variation in the GBLUP accuracy, which ranged from negative (−0.41) to 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 ($F = 7.0$),

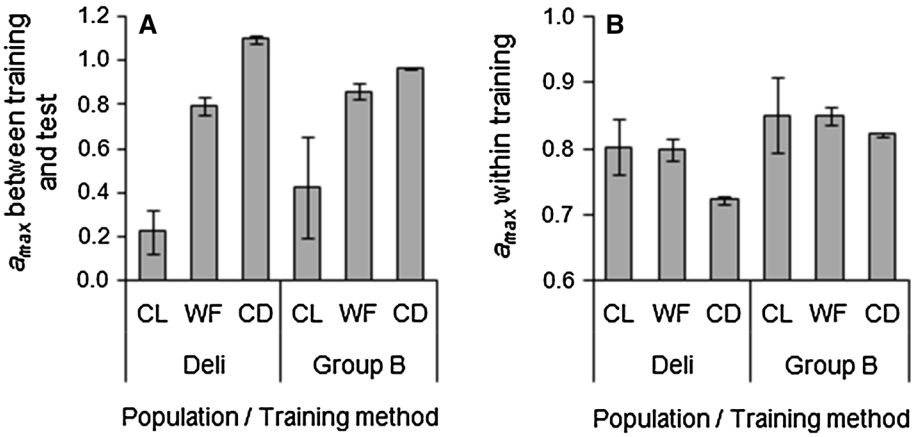

Fig. 5 Maximum additive genetic relationship (a_{max}) **a** between training and test sets and **b** within training sets, according to the population (Deli and Group B) and method to define the training set (*CL* K-means clustering, *WF* within-family, *CD* CDmean). *Bars* are SD. For CL and WF, SD were calculated between replicates ($n = 5$), while for CD it was calculated between traits ($n = 8$)

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. In all populations, CDmean gave a high maximum additive relationship between training and test set (a_{max}), the Within-Family method gave intermediate a_{max} and clustering led to low a_{max}, with one replicate with a_{max} close to zero (Fig. 5). The maximum additive relationship within training sets was also affected by the method to define the training sets, but to a lesser extent than a_{max}. A significant positive correlation between the accuracy of GBLUP and a_{max} was found for almost all population-trait combinations (Fig. 6). 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). When the training set was defined by K-means clustering, the accuracy was low, at 0.29 on average, ranging from 0.04 for O/P in Deli to 0.49 for FW in Group B. For some training sets defined with clustering (in particular for those with very small a_{max} with the training individuals), negative accuracies were found. We assumed this reflected different linkage phase between marker and QTL alleles for distantly related individuals present in the training and test sets.

A significant value for the trait-population interaction in the ANOVA analysis was obtained 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) (see Fig. S4 for the complete interaction diagram). The trait effect was due to the accuracy of O/P (mean 0.42) significantly lower than the accuracy of BN (mean 0.60).

Estimates of h^2 ranged from 0.21 (O/P in Deli) to 0.57 (ABW in Group B) (Supplementary Fig. S5). 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).

GS bias

A strong correlation was found between accuracy and bias, indicating that the higher the accuracy, the lower the bias. GEBV was unbiased from accuracies of around 0.6 and above (Supplementary Fig. S6).

Discussion

This paper presents the first experiment on genomic evaluation in a set of two oil palm breeding populations currently used in conventional reciprocal recurrent selection. We found that genomic selection (GS), in the conditions of this experiment, 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 (i.e., not progeny tested). Superiority in accuracies was attained in one of the populations (Group B) and for some traits, due to the ability of GS to estimate the Mendelian sampling term of individuals that were not progeny tested, as indicated by the significant correlations between GEBV and DEBV that could be observed in full-sib families. For the second population (Deli), 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, as it opens the door to reduce the load of phenotypic evaluation and the generation interval, both important constraints in the current breeding program.

The only study to date that focused on the feasibility and potential of GS for oil palm is the simulation work

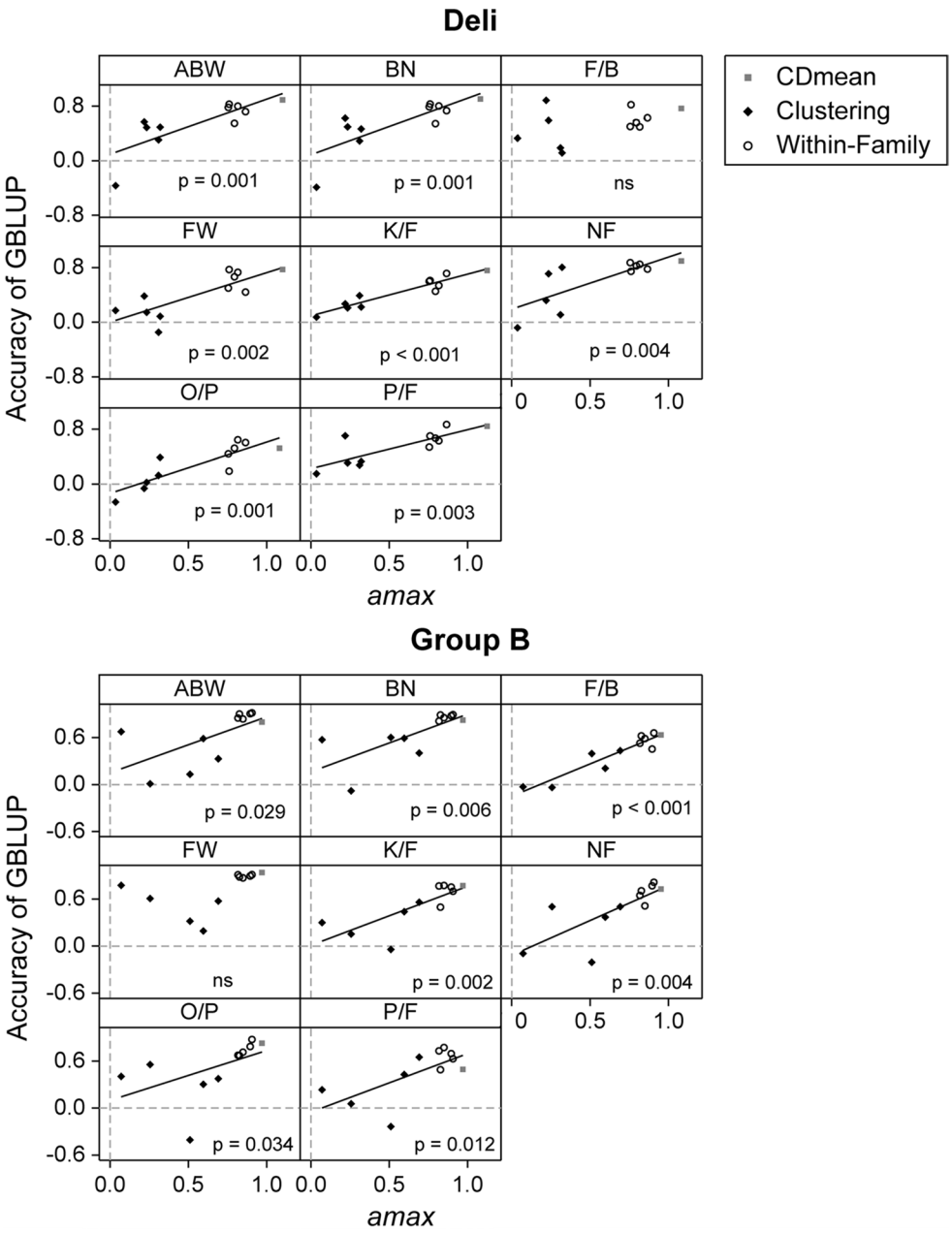

Fig. 6 Accuracy of GBLUP versus the maximum additive genetic relationship (a_{max}) according to the population (Deli and Group B) and trait (*ABW* average bunch weight, *BN* bunch number, *FW* fruit weight, *NF* number of fruits per bunch, *F/B* fruits-to-bunch ratio, *P/F* pulp-to-fruit ratio, *O/P* 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). Accuracy of GBLUP was box-cox transformed prior to regression analysis. Significance of the correlation: ns: not significant, *$0.05 > P \geq 0.01$, **$0.01 > P \geq 0.001$, ***$0.001 > P$

of Wong and Bernardo (2008). They 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 compared to that of conventional evaluation, but as the length of the breeding cycle with selection on markers alone was shortened to 6 years, the genetic gain per year ultimately increased. A novel aspect brought by our analysis is the assessment of GS in true breeding conditions, using real data from two selected populations that represent the complexity that can be found in the breeding programs for the species. 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 the fore cited simulations, 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.

The range of accuracy of GS we had in our study was comparable to the values obtained by Zapata-Valenzuela et al. (2012) in loblolly pine. They also studied the implementation of GS in a perennial crop with a small number of individuals (149), using a population with a low N_e (resulting from a structured mating design). Although they had a larger number of markers (3,406 SNP), they hypothesized that their GS accuracy relied more on familial linkage than

on historical LD between markers and QTL. In their case GS accuracy was similar to conventional phenotypic selection. This was not the case in our study as conventional phenotypic selection in parental populations of oil palm has a high accuracy (between 0.80 and 0.90).

Information captured by markers

We assumed that the differences in performance of GBLUP relative to PBLUP among traits and populations, as well as the effect of trait by population interactions on the GBLUP accuracy, resulted from different phenotypic variances among populations and traits and from the difference in marker informativeness among populations. These differences were likely a consequence of the contrasted history of the two populations. Each population suffered from different bottleneck events, were subjected to independent selection regimes and distinct drift effects. Compared to Group B, the Deli population had a narrower genetic base of four founders and a longer history of artificial selection, drift, and inbreeding. This likely explains the fact that Deli had the lowest within-family phenotypic variance and consequently, Mendelian sampling terms are expected to be of smaller magnitude than in the Group B. As another consequence of its history, the Deli had the lowest number of alleles per marker, which was on average 2.4 compared to 3.7 in Group B, and the lowest marker density (due to more monomorphic markers than in Group B). Finally, the markers used in this study were not informative enough for the Deli population to give good estimates of the realized additive relationships and this did not allow GBLUP to generate good estimates of Mendelian sampling terms for individuals that were not progeny tested; which lead to GBLUP not performing better than PBLUP. By contrast, the Group B had higher within-family phenotypic variance and higher total number of alleles than Deli, indicating that GBLUP could have a marked advantage over PBLUP in Group B. In other words, in Group B, compared to Deli, the Mendelian sampling terms of individuals not progeny tested were easier to estimate with GS as they had a higher magnitude and because the markers were more informative.

GS utilizes the additive genetic relationship between training and test sets and LD between markers and QTL to estimate GEBV, thus accounting for both family effects and Mendelian sampling terms (Habier et al. 2007, 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. Cochard (2008) showed that the LD was higher in the Deli than in the African populations used in this study for short distances (below 30–35 cM) and was lower for longer distances. He also found that the LD, measured by the correlation coefficient between SSR markers (r^2), decayed to less than 0.10 within approximately 17 cM in Deli, 10 cM in La Mé, and 7 cM in Yangambi. Consequently, given the marker density in our two parental populations, the LD between adjacent markers was higher in Deli than in Group B. However, as GBLUP could only estimate Mendelian sampling terms in Group B, this indicated that LD was actually not the key parameter in our dataset. 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 size of the training set and marker density.

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. One information to bring into consideration here is the fact that pedigrees were not deep enough as to reach the base of unrelated founders (for example in Deli the pedigree did not trace back to the four founders of 1848), allowing for some individuals to appear erroneously as unrelated according to pedigree records. 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. Obviously, this does not mean that progeny tests are useless, but it does indicate that there was a strong genealogical structure in our breeding populations, as a consequence of inbreeding and selection. The high accuracies obtained with PBLUP were due to the ability of the pedigree to model this structure. Using GS to select among individuals that were not progeny tested, 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 the two parts of breeding values, i.e., family effects and Mendelian sampling terms. Our results stress the need for a control pedigree-based method when evaluating the potential of GS, as it helps in assessing the ability of GS to account for 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). This could be due to the limited number of training individuals and markers, or could have resulted from the fact that the true genetic architecture actually involved large numbers of QTL for all traits.

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 assignment, 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 relationships (a_{max}) between the training and test sets. This was similar to the results obtained by Habier et al. (2010) in Holstein cattle with large training sets (2,096 and 1,408) and a large number of SNP (54,001).

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. This subset would make an optimized training population, i.e., the one maximizing the GS accuracy. Finally, selection would be made based on GEBV among all individuals, either both genotyped and progeny tested or only genotyped. 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, which we also observed here as the training sets were very similar among traits.

Practical prospects for 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 (Fig. 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). Using selection candidates highly related to the training set (for instance full-sibs) would correspond to the situation we studied with the Within-Family and CDmean strategies, which proved to be favorable in terms of accuracy. However, if selection candidates are loosely related to the training set (although from the same population), our results with the K-means strategy indicated that GS would fail, with accuracy very low, and possibly negative. This case could occur for example when companies exchange breeding material after several generations of independent selection. 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, 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 highest accuracy obtained currently in reciprocal recurrent selection, i.e., higher than 0.54.

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. Larger training sets could be achieved by

adding each new generation of progeny-tested individuals to the existing training set. 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 now available (Singh et al. 2013). Another crucial question to be addressed is the decrease in GS accuracy when applying the model in the subsequent generations following training. Moreover, our study used data that were collected in a single environment, which likely led to an upward accuracy bias due to a common error component in both GEBV and EBV (Lorenz et al. 2011). 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.

To our knowledge, this is the first empirical study of GS with SSR markers. In the near future we will rather use SNP markers, as this will make the analysis easier (due to the biallelic nature of SNP), decrease the cost per data point and allow faster genotyping. In a simulation, Solberg et al. (2008) concluded that two to three times more SNP were required to achieve the same accuracy as with SSR. In our oil palm breeding populations, the difference in the number of SNP and SSR necessary to reach a given accuracy will surely be smaller, as the polymorphism of SSR was low, with some markers actually being biallelic.

We used a two-step approach, first obtaining deregressed estimates of the additive value (DEBV) of the progeny-tested Deli and Group B parents and, second, using these values as data records in the GS model to measure the GS accuracy when predicting the DEBV of individuals not progeny tested. An alternative would have been to implement a single-step methodology, using the whole dataset (i.e., the phenotypic data of the progenies) and, for a given training set of parents, considering only the crosses made with these parents (i.e., discarding from the analysis the data of the progenies of the test parents) to directly predict the GEBV of the test parents. Although such an approach was appealing, it could not be implemented here. Indeed, the available data represented only one experimental design, and this had to be analyzed as a whole. Analyzing just one part of the experimental design would have lead to a highly unbalanced dataset, with parents and trials becoming disconnected from the rest of the experimental designs and biases appearing in the estimates of the non genetic effects. In real life situations, oil palm breeders would use the results of a whole experimental design to calibrate the GS model, therefore taking advantage of its qualities (connectedness between trials and between parents, balance in the number of crosses per parent, etc.). We chose a two-step procedure in order to mimic such a situation. Obviously, when data will become available from several experimental designs, we will likely adopt a single-step approach.

Authors contribution statement DC carried out data analysis and wrote the paper, with the contribution of MD, LS and JMB. BC and VP carried out genotyping work. TDG, ES, BC, AO, BN, AF, and VR made field experiments and data collection.

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. Tisné for their help in genotyping, L. Dedieu for reviewing the manuscript as well as two anonymous reviewers and C.C. Schön for their helpful comments.

Conflict of interest The authors declare no conflict of interest.

Appendix: Estimation of parental breeding values

The mating design of the progeny tests consisted of 445 Deli × Group B crosses made according to an incomplete factorial design. The crosses were evaluated in 26 trials planted between 1995 and 2000. The experimental designs of the trials were RCBD with five or six blocks and balanced lattices of rank four or five. The bunch production was measured on 30,872 palms and bunch quality on 21,525 palms. Eight traits were studied. The bunch number (BN) and average bunch weight (ABW) were measured every ten days on palms from ages 6 to 11. The annual cumulative BN and mean annual ABW were used in analysis. The median number of progenies with bunch production data was 169 per Deli parent (ranging from 25 to 743) and 141 (23–859) per Group B parent. 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 24 palms per cross. The median number of bunches analyzed was 327 per Deli parent (ranging from 69 to 1,358) and 309 per Group B parent (73–1,149).

EBV were computed as traditional pedigree-based BLUP (T-BLUP) predictors of the random effects $\boldsymbol{a_A}$ and $\boldsymbol{a_B}$, using a mixed model of the form:

$$y = \boldsymbol{X\beta} + \boldsymbol{Z_1 a}_{\text{Deli}} + \boldsymbol{Z_2 a}_B + \boldsymbol{Z_3 b} + \boldsymbol{Z_4 c} + \boldsymbol{Z_5 p} + \boldsymbol{Z_6} k + \boldsymbol{e},$$

where $\boldsymbol{y}$ is the vector of data records for the trait being analyzed, β the vector of fixed effects (general mean, trial and block within trial), $\boldsymbol{a}_{\text{Deli}}$ and $\boldsymbol{a_B}$ vectors of general combining ability of Deli ~$N(0,\ 0.5\boldsymbol{A}_{\text{Deli}}\ \sigma^2_{\text{Deli}})$ and

Group B individuals ~$N(0,\ 0.5\boldsymbol{A_B}\ \sigma_B^2)$, respectively, $\boldsymbol{b}$ the vector of the incomplete block within block and trial effects ~$N(0, \boldsymbol{I}\sigma_b^2)$, $\boldsymbol{c}$ the vector of specific combining ability of single crosses ~$N(0, \boldsymbol{D}\sigma_c^2)$, $\boldsymbol{p}$ the vector of permanent environmental effects used to take repeated measures into account ~$N(0, \boldsymbol{I}\sigma_p^2)$, $\boldsymbol{k}$ the vector of elementary plot effects ~$N(0, \boldsymbol{I}\sigma_k^2)$ and $\boldsymbol{e}$ the vector of residual effects ~$N(0, \boldsymbol{I}\sigma_e^2)$. $\boldsymbol{X}$, $\boldsymbol{Z_1} - \boldsymbol{Z_6}$ are incidence matrices. $\boldsymbol{A}_{\text{Deli}}$ and $\boldsymbol{A_B}$ are matrices of additive relationships among Deli and Group B individuals, respectively, computed from pedigrees. $\boldsymbol{D}$ is the matrix of dominance relationships among crosses computed from the pedigree, with value between crosses Deli × B and Deli′ × B′ equal to $f_{\text{Deli,Deli}'} \times f_{B,B'}$, where $f_{\text{Deli,Deli}'}$ and $f_{B,B'}$ are the coefficient of coancestry between the Deli and Group B parents. $\boldsymbol{I}$ is an identity matrix. For BN and ABW, the model also included a fixed age effect and a random age within cross effect $\boldsymbol{a}$ ~$N(0, \boldsymbol{D} \otimes \boldsymbol{I}\sigma_a^2)$. This model was based on the model of Stuber and Cockerham (1966) for hybrids between unrelated populations, as previously used in oil palm by Purba et al. (2001). The R-ASReml package (Butler et al. 2009) for R (R Core Team 2013) was used to obtain variance component estimates and EBV of all individuals.

The accuracy of the general combining ability a_i of an individual i (actually a_{Deli_i} or a_{B_i} depending on the population of origin of i) is given by $r_{a,\hat{a}_i} = \sqrt{\frac{1-PEV_{a_i}}{0.5(1+F_i)\sigma_a^2}}$, where PEV_{ai} is the prediction error variance associated with a_i, $0.5(1 + F_i)$ is the diagonal of the relationship matrix used in the mixed model (i.e., $0.5\boldsymbol{A}_{\text{Deli}}$ or $0.5\boldsymbol{A_B}$, depending on the population of origin of i), F_i is the inbreeding coefficient and σ_a^2 is the additive variance (i.e., σ_{Deli}^2 or σ_B^2, depending on the population). This formula was used to compute the mean accuracy of the general combining ability of the 131 Deli and 131 Group B parents used in the GS analysis, which was 0.89, ranging from 0.83 ± 0.06 (SD) for O/P in Deli to 0.93 ± 0.04 for K/F in Group B.

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

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 81:1084–1097

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 (2008) Etude de la diversité génétique et du déséquilibre de liaison au sein de populations améliorées de palmier à huile (*Elaeis guineensis* Jacq.). Montpellier SupAgro, Montpellier, pp 97–175

Cochard B, Adon B, Rekima S et al (2009) Geographic and genetic structure of African oil palm diversity suggests new approaches to breeding. Tree Genet Genomes 5:493–504

Corley RHV (2009) How much palm oil do we need? Environ Sci 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, Villanueva B, Bijma P, Woolliams JA (2007) Inbreeding in genome-wide selection. J Anim Breed Genet 124: 369–376

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, 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, New York, pp 299–320

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

Dussert S, Guerin C, Andersson M et al (2013) Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. Plant Physiol 162:1337–1358

Eding H, Meuwissen THE (2001) Marker-based estimates of between and within population kinships for the conservation of genetic diversity. J Anim Breed Genet 118:141–159

Gao H, Lund MS, Zhang Y, Su G (2013) Accuracy of genomic prediction using different models and response variables in the Nordic Red cattle population. J Anim Breed Genet 130:333–340

Garrick D, Taylor J, Fernando R (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet Sel Evol 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 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. Genet Sel Evol 42:5

Habier D, Fernando R, Kizilkaya K, Garrick D (2011) Extension of the bayesian alphabet for genomic selection. BMC Bioinform 12:186

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

Isik F (2014) Genomic selection in forest tree breeding: the concept and an outlook to the future. New Forest 45:379–401

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genom 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

Li CC, Weeks DE, Chakravarti A (1993) Similarity of DNA fingerprints due to chance and relatedness. Hum Hered 43:45–52

Lorenz AJ, Chao S, Asoro FG et al (2011) Genomic selection in plant breeding: knowledge and prospects. In: Sparks DL (ed) Advances in agronomy. Academic Press, San Diego, pp 77–123

Lynch M (1988) Estimation of relatedness by DNA fingerprinting. Mol Biol Evol 5:584–599

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Ostersen T, Christensen O, Henryon M, Nielsen B, Su G, Madsen P (2011) Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. Genet Sel Evol 43:38

Park T, Casella G (2008) The Bayesian LASSO. J Am Stat Assoc 103:681–686

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

Purba AR, Flori A, Baudouin L, Hamon S (2001) Prediction of oil palm (*Elaeis guineensis*, Jacq.) agronomic performances using the best linear unbiased predictor (BLUP). Theor Appl Genet 102:787–792

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Resende MDV, Resende MFR, Sansaloni CP et al (2012) Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol 194:116–128

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. Genet Sel Evol 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. Nat Adv. doi:10.1038/nature12309

Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE (2008) Genomic selection using different marker types and densities. J Anim Sci 86:2447–2454

Stuber CW, Cockerham CC (1966) Gene effects and variances in hybrid populations. Genetics 54:1279–1286

Tee S-S, Tan Y-C, Abdullah F, Ong-Abdullah M, Ho C-L (2013) Transcriptome of oil palm (*Elaeis guineensis* Jacq.) roots treated with *Ganoderma boninense*. Tree Genet Genom 9:377–386

Thomsen H, Reinsch N, Xu N et al (2001) Comparison of estimated breeding values, daughter yield deviations and de-regressed proofs within a whole genome scan for QTL. J Anim Breed Genet 118:357–370

Tranbarger TJ, Dussert S, Joët T et al (2011) Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. Plant Physiol 156:564–584

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 Biol 12:1

USDA (2013) Oilseeds: world market and trade. Foreign Agricultural Service, Circular Series, July 2013 http://www.fas.usda.gov/oilseeds_arc.asp

Waples RS, Do CHI (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. Mol Ecol Resour 8:753–756

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

Zapata-Valenzuela J, Isik F, Maltecca C et al (2012) SNP markers trace familial linkages in a cloned population of *Pinus taeda*: prospects for genomic selection. Tree Genet Genom 8:1307–1318