Efficiency of genomic selection with models including dominance effect in the context of perennial crop breeding.

EUCARPIA XVth meeting - Biometrics in Plant Breeding Stuttgart Hohenheim, Germany 5-7 September, 2012

Denis Marie[1], Bouvet Jean-Marc [1] (1) Cirad-Research unit AGAP "Genetic improvement and adaptation of Mediterranean and tropical plants" International campus of Baillarguet TA A-108/C 34398 Montpellier Cedex 5 FRANCE

▪Introduction

In perennial plants, varieties can be produced by clones or elite full-sib families, where both additive and non-additive effects are taken into account in the selection process. Although this point is crucial in perennial crops, very few studies have analysed the value of including non-additive effects in the Genomic Selection (GS) model (Meuwissen et al. 2001) and there is a growing interest in testing new models.
We developed a simulation study to test the efficiency of GS in the case of perennials crop breeding with the example of *Eucalyptus* one of the most used forest tree genera in plantation. We simulated a recurrent selection scheme for clone production over four breeding cycles. The objectives of this work were (1) to assess the performance of GS when both additive and dominance effects are included in models to predict the breeding and genotypic values, (2) to examine the performance of GS when the training set is updated at each cycle, and (3) to analyse the genetic gain per unit time in the context of *Eucalyptus* breeding.

▪Material and methods

Simulation studies : Our analyses are based on data simulated with R software version 2.13.0 and the HaploSim package, developed by Coster and Bastiaansen (2009).The simulated genome for each individual consisted of two chromosomes each represented by two haplotypes and a genome size of 2 morgans. To avoid high linkage disequilibrium between QTLs due to the small genome size, only 44 QTLs were randomly assigned to the loci of each individual of generation 1,000 (Fig. 1) Two variables "a" and "k" were associated with each QTL at generation 1,000 and were used to define the total genotypic, breeding (additive) and dominance values. The breeding value of an individual represented by a biallelic locus is equal to $-2p\alpha$ for bb, $(q-p)\alpha$ for bB and $2q\alpha$ for BB, α being the substitution effect calculated by $\alpha=a(1+k(q-p))$, where q and p are the allele frequencies in the population. The deviation due to dominance was equal to $-2p^2ak$ for bb, $2pqak$ for bB and $-2q^2ak$ for BB

Scenario tested : In the case of *Eucalyptus*, simulations were developed in the frame of a recurrent selection scheme for clone production over four breeding cycles (Fig. 1). Scenarios crossing two broad sense heritabilities ($H^2=0.6$ and 0.1), three dominance to additive variance ratios (R=0.1; 0.5 and 1) and different training population structure were compared.

Estimation of marker effect, breeding and genotypic value : We used the Bayesian least absolute shrinkage and selection operator (Bayesian LASSO) developed by de los Campos et al. (2009) in the genomic selection context to estimate marker effects. The breeding or the total genotypic value (clone value) of each individual in the candidate population was calculated by adding the marker effects over all loci. The other method for selecting trees used the observed phenotype and is referred to as phenotypic selection (PS). The BLUP was estimated for parents and clones using the classic mixed linear model (Mrode and Thompson 2005).

Fig.1 : Illustration of the recurrent selection scheme based on the full-sib family progeny test with the production of clones. This scheme is based on classic practices in Eucalyptus breeding and was conducted over four cycles. The scheme in cycles 2, 3 and 4 is similar to cycle 1

▪Results-discussion

• Performance of GS over breeding cycles

When the training population was made up of the first cycle progeny tests and the candidate populations were the progeny tests of three successive cycles, GS accuracy decreased with breeding cycles (e.g. from 0.9 to 0.4 with $H^2=0.6$ and R=0.1), whereas PS presented constant performances (accuracy of 0.8 with $H^2=0.6$ and R=0.1), illustration with Figures 2a to d. When the training population was constituted of the reference population of C1, the superiority of the model with dominance was not demonstrated by our analysis for the breeding population , whatever the R ratio (Fig. 1 a and b) . For the clone population, the superiority of the model including the dominance effect was established for $H^2=0.6$ and $R\geq0.5$ (Fig. 1c and d).

•Performance of GS by updating the training population

The accuracy of GS was tested with different numbers of generations included in the training population. GS accuracy increased as the number of cycles constituting the training data set increased(Figs 3a and 3b). This trend was observed whatever the heritability and the variance ratio and was more pronounced for the selection on total genotypic value with the clone population. The model including the dominance effect performs better when updating the training population by combination of different cycles.
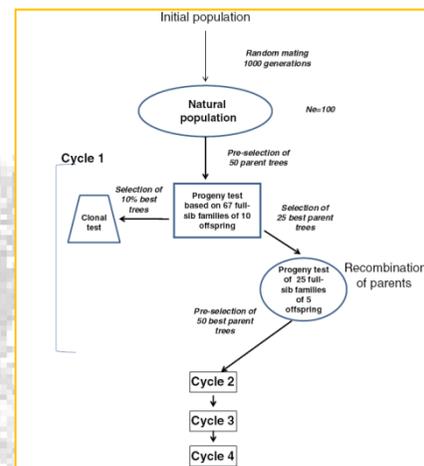
•Genetic gain per unit time

In terms of gain per unit time, GS led to better results for breeding population, with two or three times greater genetic gain per unit time than with PS (Figs. 4a). The superiority of GS in genetic gain per unit time was less pronounced for the clone population as the duration of clone evaluation based on phenotype is added (Fig 4b). Updating the training set slightly improved GS in terms of absolute gain (Figs. 4 a and b). This was due to the time required to update the training data set.

Fig.2 : Change in selection accuracy over cycles for (a) breeding population and (b) clone population without updating the training set: illustration with $H^2=0.6$ and R=1.00. For GS, the training population is constituted of the individuals of the cycle 1 progeny tests to predict the candidate populations of subsequent cycles.

Fig. 3: Change in selection accuracy over cycles for (a) breeding and (b) clone population with updating of the training set: illustration with heritability $H^2=0.6$ and R=1. GS without updating: training population in cycle 1 to predict subsequent cycles.GS-updating the training population: training population made of cycles C1+C2 to predict candidate population in C3 and training population made of cycles C1+C2+C3 to predict candidate population in C4
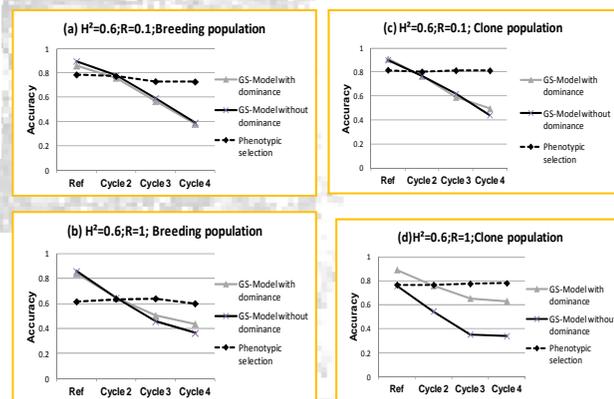
▪Conclusion

Our study shows that using a model with the dominance effect is pertinent for selection based on total genotypic value (e.g. selection of clones), for high dominance effect ($\sigma^2_D/\sigma^2_A \approx 1$) and a large training population size.

Bibliography : Coster A, Bastiaansen J (2009) Haplosim: Haplosim R package version 1.8
de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, et al. (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigrees. Genetics 182:375–385
Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829
Mrode RA, Thompson R (2005) Linear models for the prediction of animal breeding values, 2nd edn. CABI Publishing, Cambridge USA
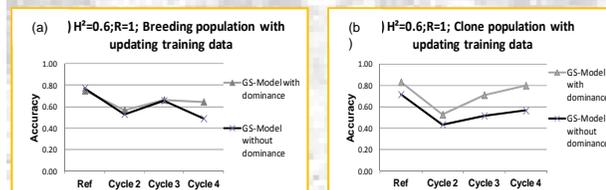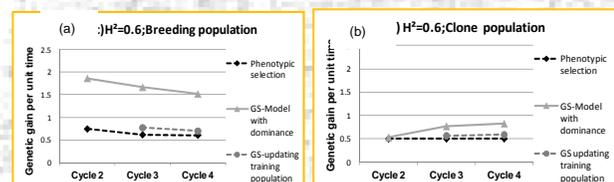
Fig. 4 Change in genetic gain per unit time over generations averaged over the three additive to dominance ratios for the breeding and clonal population comparing PS and GS methods : illustration with heritability $H^2=0.6$.