CROPS AND SOILS RESEARCH PAPER
Best linear unbiased prediction for genetic evaluation in reciprocal recurrent selection with popcorn populations

J. M. S. VIANA[1]*, G. B. MUNDIM[1], R. O. DELIMA[1], F. F. E. SILVA[2] AND M. D. V. DE RESENDE[3]

[1] *General Biology Department, Federal University of Viçosa, 36570-000 Viçosa, MG, Brazil*
[2] *Statistics Department, Federal University of Viçosa, 36570-000 Viçosa, MG, Brazil*
[3] *Forestry Science Department, Embrapa Forestry/Federal University of Viçosa, 36570-000 Viçosa, MG, Brazil*

SUMMARY

The objective of the present study was to present the theory and application of best linear unbiased prediction (BLUP) in reciprocal recurrent selection (RRS). Seven progeny tests from two RRS programmes with popcorn (*Zea mays* L. ssp. *mays* [syn. *Zea mays* L. ssp. *everta* (Sturtev.) Zhuk.]) populations were conducted and analysed for expansion volume and grain yield. The interpopulation half- and full-sib family models were fitted using ASReml software. Half-sib selection is equivalent to selection for the general combining ability (GCA) of the common parents. With inbred full-sib progeny and BLUP analysis, it is possible to predict the general and specific combining ability effects. The standard error of prediction of the progeny effect was lower than the standard deviation of the best linear unbiased estimation (BLUE) estimate. For half- and full-sib RRS, the BLUE and BLUP provided highly correlated estimates of progeny genotypic values. The coincidence between selected parents ranged from 64 to 95%. With inbred full-sib progeny, the correlations between the BLUE of progeny genotypic values and the BLUP of GCA effects were lower. Consequently, the coincidence between selected parents was lower, ranging from 0 to 57%. The percentage of common selected inbred progeny based on the BLUE and BLUP of the progeny genotypic value ranged from 57 to 100%.

INTRODUCTION

The conventional approach for estimating variance components and genotypic values in annual crop breeding is to use the analysis of variance (ANOVA) method and best linear unbiased estimation (BLUE). Recently, best linear unbiased prediction (BLUP) and Bayesian inference have emerged as alternative procedures for genetic evaluation by plant breeders (Piepho *et al.* 2008; Bauer *et al.* 2009). Best linear unbiased prediction was originally developed for predicting additive values in animal breeding (Henderson 1974) and it requires the parametric values of the variance components. As these values are unknown, predictions are commonly based on the restricted maximum likelihood (REML) estimates (Patterson & Thompson 1971). Best linear unbiased prediction and Bayesian inference allow the inclusion of pedigree

* To whom all correspondence should be addressed. Email: jmsviana@ufv.br

information in the analysis. The Bayesian approach allows incorporating prior knowledge about parameters of the model (Blasco 2001). Several studies have shown the efficacy of BLUP in annual crop breeding for predicting untested maize single crosses performance (Bernardo 1996) and breeding values in recurrent selection and in the development of pure and inbred lines (Bauer *et al.* 2006; Oakey *et al.* 2007; Viana *et al.* 2011).

In maize breeding, reciprocal recurrent selection (RRS) is one of the most commonly used methods. Comstock *et al.* (1949) proposed the evaluation of interpopulation paternal half-sib families and recombination of S_1 progeny. Hallauer (1967) suggested a modified method for continually assessing inbred interpopulation full-sib families to obtain hybrids. Full-sib RRS with S_1 progeny recombination was proposed by Hallauer & Eberhart (1970). Ordás *et al.* (2012) compared the efficiency of the full-sib RRS with a modification to the method that essentially applied

a selection index to include the S_2 progeny performance (FS-S_2-RRS). Both procedures efficiently improved the populations *per se* and the FS-S_2-RRS was more efficient than the standard method for developing high-yielding and stable crosses between the two populations. dos Reis *et al.* (2012) investigated the efficiency of RRS strategies by determining the expected and realized responses to selection as well as changes in heterosis, concluding that the method proposed by Souza (1987) showed the highest efficiency in improving the interpopulation cross but not the parental populations. This method also increased heterosis in the interpopulation hybrid after three selection cycles. Romay *et al.* (2011) evaluated the effect of full-sib RRS in two Spanish maize populations, observing that this method was effective for the improvement of grain yield in the interpopulation cross and in one of the parental populations. Furthermore, the grain yield heterosis increased significantly.

Reciprocal recurrent selection has proven to be a highly efficient procedure for improving the interpopulation cross regardless of the trait, populations, family or recombination unit (Penny & Eberhart 1971; Romay *et al.* 2011). Experimental and simulated results have also shown improvement in one or both of the two populations (Keeratinijakal & Lamkey 1993; Peiris & Hallauer 2005; Viana *et al.* 2013). Throughout the past 63 years of RRS use, identification of the best families has been made based on the BLUE of progeny genotypic values. Viana *et al.* (2013) showed that half-sib RRS is more efficient than full-sib RRS based on progeny genotypic value, but as efficient as full-sib RRS based on the general combining ability (GCA) effect. Unfortunately, as highlighted by Viana *et al.* (2013), with the standard full-sib RRS (evaluation of non-inbred full-sib progeny and S_1 recombination) it is not possible to estimate or predict the GCA effects of the parents. The GCA prediction is possible with the assessment of inbred full-sib families based on BLUP with pedigree information (additive and dominance relationship matrices different from an identity matrix). This is because with non-inbred progeny there is only one covariance between relatives (full-sibs), which depends on the three components of the genotypic variance (two GCA variances and the specific combining ability (SCA) variance), assuming no epistasis. Even assuming absence of dominance the covariance depends on two components of variance. Then the additive and dominance relationship matrices must include at least three distinct covariances between relatives.

Thus, the objective of the present study was to present theory and application of BLUP for genetic evaluation in RRS.

MATERIALS AND METHODS

Experimental data

Expansion volume (ml/g) and grain yield (kg/ha) of seven progeny tests, corresponding to two programmes of RRS with Viçosa and Beija-Flor popcorn (*Zea mays* L. ssp. *mays* [syn. *Zea mays* L. ssp. *everta* (Sturtev.) Zhuk.]) populations were analysed. The first programme was conducted with full-sib families of the Beija-Flor Cycle 1 (BF_{C1}) and Viçosa (V_{C0}) populations, the second with half- and full-sib families of the Beija-Flor Cycle 3 (BF_{C3}) and Viçosa Cycle 3 (V_{C3}) populations. All of the improved populations were obtained by half-sib selection. Both programmes are based on the procedure proposed by Hallauer (1967).

In Programme 1, 188 full-sib families from the first selection cycle ($S_0 \times S_0$ hybrids) were evaluated during the 2002/03 growing season in a 14×14 simple lattice at Maringá State University (23°25′S, 51°57′W, 596 m a.s.l.), Paraná state, Brazil. One hundred and thirty six inbred interpopulation full-sib families from the second cycle ($S_1 \times S_1$ hybrids) were evaluated during the 2004/05 growing season in a 12×12 lattice with three replications. This trial was conducted in an experimental area of Federal University of Viçosa (UFV) located in Coimbra (20°51′S, 42°48′W, 720 m a.s.l.), Minas Gerais state, Brazil. In Programme 2, half- and full-sib families from the first cycle were evaluated in three different trials during the 2007/08 growing season, in experimental areas of UFV at Coimbra and Ponte Nova (20°24′S, 42°54′W, 431 m a.s.l.), Minas Gerais. The full-sib families test was designed as a 13×13 lattice with three replications. The half-sib family tests were designed as 11×11 and 12×12 lattices with three replications, for the Viçosa Cycle 3 and Beija-Flor Cycle 3 progeny, respectively. In these trials, eight checks were inserted as additional treatments (one plot by replication). Inbred interpopulation full-sib families from the second and third cycles ($S_1 \times S_1$ and $S_2 \times S_2$ hybrids) were evaluated during the 2009/10 and 2011/12 growing seasons at Coimbra, and the trials were designed as 9×9 lattices

with three replications. These trials included 73 and 72 progenies, respectively.

Each trial plot corresponded to a 5 m row of 25 plants (ideal stand) spaced 0·9 m apart. Final stand and grain moisture (covariates) were measured in all plots of each trial. The expansion volume was assessed using 30 g samples in a 27-litre microwave oven (900 W). The only control variety common to all trials was the commercial hybrid IAC 112, which was developed by the Agronomic Institute of Campinas, São Paulo State, Brazil. With one exception (the test of $S_0 \times S_0$ progeny from Programme 1), all trials included the parental populations. The controls were not included in the BLUP and BLUE analyses for generating unbalanced datasets.

Genetic model

For two populations A and B in Hardy–Weinberg equilibrium, the genotypic mean of an interpopulation half-sib family of a plant from population A is

$$\overline{G}_{HSF(A \times B)} = M_H + GCA_{(A)}$$

where M_H is the mean of the interpopulation cross and $GCA_{(A)}$ is the GCA effect of the common parent (Viana *et al.* 2013). Thus, selection among half-sib families is equivalent to selection for the GCA of the common parents because for each gene the correlation between the GCA effect and the number of genes that increase trait expression is 1 or -1, depending on the degree of dominance and gene frequency. This correlation can only be negative with overdominance. Assuming 100 genes for determining a quantitative trait, the correlation is 0·95 (average of 100 simulations) for degrees of dominance between $-1·2$ and 1·2 (bidirectional dominance as evidenced for the expansion volume) and between 0 and 1·2 (positive dominance as evidenced for the grain yield). The correlation approaches 1 as the degree of dominance approaches 0.

The genotypic mean of an interpopulation full-sib family of plants from populations A and B is

$$\overline{G}_{FSF(A \times B)} = M_H + GCA_{(A)} + GCA_{(B)} + SCA$$

where $GCA_{(B)}$ is the GCA effect of the parent from population B and SCA is the SCA effect (Viana *et al.* 2013). The SCA effect primarily reflects the difference of gene frequencies between parents. Thus, selection among interpopulation full-sib families is equivalent to selection for general and specific combining ability. The model derived by Viana *et al.* (2013) is a

generalization of the model proposed by Bernardo (1994) for a factorial design involving inbred lines from two heterotic pools, for any inbreeding coefficient. Additionally, Viana *et al.* (2013) presented the parametric values of the GCA and SCA effects and variances.

With half- and full-sib RRS (assessment of half- or full-sib progeny with recombination of selected S_1) Hardy–Weinberg equilibrium is expected in both parental populations. With continuous assessment of inbred full-sib progeny, because the process suggested by Hallauer (1967) involves random crosses between S_{n+1} plants derived only from selected S_n parents, and not recombination of the selected inbred progeny (with this procedure Hardy–Weinberg equilibrium is also expected in both parental populations), Viana *et al.* (2013) assumed the following probabilities for AA, Aa and aa individuals: $p^2 + pqF$, $2pq(1-F)$ and $q^2 + pqF$, where p and q are allelic frequencies and F is the inbreeding coefficient. Viana *et al.* (2013) also assumed an independent assortment of genes.

BLUP analysis

The phenotypic values of the interpopulation half- and full-sib families can be expressed as:

$$P_{HSF} = M_H + GCA_{(A)} + \sum_f (\text{fixed effect})_f$$
$$+ \sum_r (\text{non-genetic random effect})_r + \text{error}$$

and

$$P_{FSF} = M_H + GCA_{(A)} + GCA_{(B)} + SCA$$
$$+ \sum_f (\text{fixed effect})_f$$
$$+ \sum_r (\text{non-genetic random effect})_r + \text{error}$$

Considering a multi-generation analysis to increase the prediction accuracy and defining trial and replications within trial as fixed effects, and blocks within replication within trial as a random factor, the models are, in matrix terms:

$$y = X\beta + Z_1 u_1 + Z_2 u_2 + \varepsilon,$$

for half-sib families and

$$y = X\beta + Z_1 u_1 + Z_2 u_2 + Z_3 u_3 + Z_4 u_4 + \varepsilon,$$

for full-sib families

where y is the vector of phenotypic values; X is the incidence matrix of the fixed effects; β is the vector of fixed effects; Z_1, Z_2, Z_3 and Z_4 are the incidence

matrices of the random effects; and ε is the residuals vector. For the interpopulation half-sib family model, u_1 is the vector of GCA effects of the common parents and u_2 is the vector of blocks within replication within trial effects. For the interpopulation full-sib family model, u_1 is the vector of GCA effects of the parents from population A; u_2 is the vector of GCA effects of the parents from population B; u_3 is the vector of SCA effects; and u_4 is the vector of blocks within replication within trial effects. All random vectors were assumed to be mutually independent.

The criterion used to obtain the BLUP of a random vector is the maximization of either the joint probability density function of y and the random vector(s) or the random vector(s) and ε, which obtains, assuming Normal distribution, the mixed model equations (MME) (Henderson 1974). Assuming $\mathrm{Var}(\varepsilon) = R$, the MME for these models are:

$$\mathrm{GCA}_{A_1A_2} = \left(\frac{q-p}{2}\right)\alpha_B = (\alpha_B/\alpha_A)\left(\frac{q-p}{2}\right)\alpha_A$$
$$= (\alpha_B/2\alpha_A)A_{A_1A_2}$$

and

$$\mathrm{GCA}_{A_2A_2} = -p\alpha_B = (\alpha_B/\alpha_A)(-p\alpha_A) = (\alpha_B/2\alpha_A)A_{A_2A_2}$$

where p and q are the allelic frequencies in the population A; α_A and α_B are the average effects of a gene substitution in the populations A and B, respectively; and A is the additive value.

Then:

$$\mathrm{Cov}(\mathrm{GCA}_i, \mathrm{GCA}_j) = (\alpha_B/2\alpha_A)^2 2r_{ij}\sigma^2_{A(A)} = 2r_{ij}(\alpha_B/2\alpha_A)^2$$
$$\times 2pq\alpha_A^2 = 2r_{ij}\left(\frac{1}{2}pq\alpha_B^2\right)$$
$$= 2r_{ij}\sigma^2_{\mathrm{GCA}(A)}$$

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z_1 & X'R^{-1}Z_2 \\ Z_1'R^{-1}X & Z_1'R^{-1}Z_1 + G_1^{-1} & Z_1'R^{-1}Z_2 \\ Z_2'R^{-1}X & Z_2'R^{-1}Z_1 & Z_2'R^{-1}Z_2 + G_2^{-1} \end{bmatrix} \begin{bmatrix} \beta^o \\ \tilde{u}_1 \\ \tilde{u}_2 \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z_1'R^{-1}y \\ Z_2'R^{-1}y \end{bmatrix} \quad \text{and}$$

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z_1 & X'R^{-1}Z_2 & X'R^{-1}Z_3 & X'R^{-1}Z_4 \\ Z_1'R^{-1}X & Z_1'R^{-1}Z_1 + G_1^{-1} & Z_1'R^{-1}Z_2 & Z_1'R^{-1}Z_3 & Z_1'R^{-1}Z_4 \\ Z_2'R^{-1}X & Z_2'R^{-1}Z_1 & Z_2'R^{-1}Z_2 + G_2^{-1} & Z_2'R^{-1}Z_3 & Z_2'R^{-1}Z_4 \\ Z_3'R^{-1}X & Z_3'R^{-1}Z_1 & Z_3'R^{-1}Z_2 & Z_3'R^{-1}Z_3 + G_3^{-1} & Z_3'R^{-1}Z_4 \\ Z_4'R^{-1}X & Z_4'R^{-1}Z_1 & Z_4'R^{-1}Z_2 & Z_4'R^{-1}Z_3 & Z_4'R^{-1}Z_4 + G_4^{-1} \end{bmatrix} \begin{bmatrix} \beta^o \\ \tilde{u}_1 \\ \tilde{u}_2 \\ \tilde{u}_3 \\ \tilde{u}_4 \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z_1'R^{-1}y \\ Z_2'R^{-1}y \\ Z_3'R^{-1}y \\ Z_4'R^{-1}y \end{bmatrix}$$

With interpopulation half-sib families, $\mathrm{Cov}(u_{1i}, u_{1j}) = \mathrm{Cov}(\mathrm{GCA}_i, \mathrm{GCA}_j) = 0$, as the common parents of the progenies i and j are not related. Then $\mathrm{Var}(u_1) = G_1 = I\sigma^2_{\mathrm{GCA}}$ where G_1 is the variance matrix of the GCA effects and σ^2_{GCA} is the variance of the GCA effects. Furthermore, $\mathrm{Var}(u_2) = G_2 = I\sigma^2_b$, where G_2 is the blocks within replication within trial variance matrix.

With full-sib families, the covariance between the GCA effects of two parents from population A is, for each gene:

$$\mathrm{Cov}(\mathrm{GCA}_i, \mathrm{GCA}_j) = \mathrm{Cov}[(\alpha_B/2\alpha_A)A_i, (\alpha_B/2\alpha_A)A_j]$$

since (Viana et al. 2013),

$$\mathrm{GCA}_{A_1A_1} = q\alpha_B = (\alpha_B/\alpha_A)q\alpha_A = (\alpha_B/2\alpha_A)A_{A_1A_1},$$

where $\sigma^2_{A(A)}$ is the additive variance in the population A.

Thus, $\mathrm{Var}(u_1) = G_1 = A_{(A)}\sigma^2_{\mathrm{GCA}(A)}$ where $A_{(A)} = \{2r_{ij}\}$ is the additive relationship matrix of the parents from population A and r_{ij} is the coefficient of coancestry between the parents i and j. Then, $\mathrm{Var}(u_2) = G_2 = A_{(B)}\sigma^2_{\mathrm{GCA}(B)}$ where $A_{(B)}$ is the additive relationship matrix of the parents from population B and $\sigma^2_{\mathrm{GCA}(B)} = (1/2)\sum rs\alpha_A^2$ where r and s are the allelic frequencies in the population B. Furthermore, defining i and j as the parents from population A and k and l as the parents from population B, $\mathrm{Cov}(\mathrm{SCA}_{ik}, \mathrm{SCA}_{jl}) = p_{(ik)(jl)}(1+F)^2\sigma^2_{\mathrm{SCA}}$, where $p_{(ik)(jl)}$ is the probability that the parents of the two progenies have genotypes identical by descent and F is the inbreeding coefficient. This probability is $p_{(ik)(jl)} = u_{ij}u_{kl} + u_{il}u_{kj} = u_{ij}u_{kl}$ where u_{xy} is the probability that the

individuals x and y have genotypes identical by descent. Thus, $\mathrm{Var}(u_3) = G_3 = D(1+F)^2\sigma^2_{SCA}$ where $D = \{p_{(ik)(jl)}\}$ is the dominance relationship matrix and $\sigma^2_{SCA} = \sum pqrsd^2$ is the variance of the SCA effects where d is the deviation due to dominance (Viana *et al.* 2013).

The accuracy (correlation between the true and predicted values) of the progeny, GCA and SCA effects were computed as $\sqrt{1 - (\mathrm{PEV}/v)}$ (Mrode 2005) where PEV is the prediction error variance and $v = \tilde{\sigma}^2_G$, $(1+F)\tilde{\sigma}^2_{GCA}$ or $(1+F)^2\tilde{\sigma}^2_{SCA}$, respectively, where $\tilde{\sigma}^2_G$, $\tilde{\sigma}^2_{GCA}$ and $\tilde{\sigma}^2_{SCA}$ are the REML estimates of the progeny, GCA and SCA variances. The ASReml v3.0 software (Gilmour *et al.* 2009) was used for all analyses (see code in the appendix). The additive and dominance relationship matrices were computed with a program developed in REALbasic 5.5 (REAL Software 2004).

BLUE analysis

Each progeny test was also analysed individually using the GLM procedure of the SAS System v.9.2 (SAS Institute 2007). Progeny and replication were defined as fixed factors and blocks within replication were regarded as a random factor (analysis with recovery of the interblock information). The estimated treatment genotypic means were used to compute the midparent heterosis and the realized interpopulation genetic gain in each selection cycle of each RRS programme. The realized interpopulation genetic gain was calculated as ΔM = mean of the $S_n \times S_n$ progeny – mean of the $S_{n-1} \times S_{n-1}$ progeny – (mean of IAC 112 in the $S_n \times S_n$ trial – mean of IAC 112 in the $S_{n-1} \times S_{n-1}$ trial).

Selection criteria

In each progeny test, 10 and 20% of the best progeny were selected adopting the following selection criteria: selection of the best $S_n \times S_n$ families based on the progeny genotypic value as estimated by BLUE or predicted by BLUP; and selection of the best parents for recombination of the S_{n+1} ($n \geqslant 0$) families based on either the BLUE of the progeny genotypic value or the BLUP of the GCA effect. Selection criteria were compared based on the correlation between estimated genotypic values and predicted GCA effects as well as the percentage of coincident $S_n \times S_n$ progeny and parents of the families selected.

RESULTS

In general, for both traits and RRS programmes there was evidence of genetic variability in the parental populations (Table 1). If there is no pedigree information, only the progeny genotypic variance can be estimated. The additive and additive plus dominance models were fitted only with pedigree information and data from inbred full-sib progeny. The combined analysis of half- and non-inbred full-sib progeny did not allow the inclusion of dominance in the model. The evidence of dominance for expansion volume and grain yield is also shown by the estimates of heterosis (Table 2). The average heterosis ranged from $-3\cdot7$ to $3\cdot0\%$ and from $-18\cdot1$ to $28\cdot5\%$ for expansion volume and grain yield, respectively. The maximum values of average heterosis were observed for the parental populations of Programme 1 ($\mathrm{BF}_{C1} \times \mathrm{V}_{C0}$). After the first cycle, positive average heterosis for grain yield ($12\cdot3\%$) occurred only with $S_1 \times S_1$ hybrids of Programme 2 ($\mathrm{BF}_{C3} \times \mathrm{V}_{C3}$). Interpopulation genetic gain for expansion volume ($9\cdot7\%$) and grain yield ($27\cdot3\%$) was observed only in the first cycle relative to Programme 2. After the second cycle of Programme 2, a decrease in interpopulation genetic gain was observed for both traits.

The standard error of prediction of the progeny effect was lower than the standard deviation of the BLUE estimate, even with no pedigree information, showing that BLUP analysis was superior to the BLUE approach (Table 3). For both traits, the higher accuracies of the progeny, GCA and SCA effects were observed in the $S_2 \times S_2$ test. The joint analysis of half- and non-inbred full-sib progeny did not increase the GCA prediction accuracy. The accuracy of the SCA effect was greater than the accuracy of the progeny and GCA effect. The accuracy of the GCA effect ranged from $0\cdot52$ to $0\cdot85$ for expansion volume and from $0\cdot40$ to $0\cdot68$ for grain yield.

Regardless of the trait analysed and RRS programme used, the coincidences between the best $S_0 \times S_0$ families selected according to the progeny genotypic values obtained through BLUP or BLUE ranged from 79 to 95% (Table 4). The percentage of common $S_n \times S_n$ progeny ($n \geqslant 1$) selected according to these criteria ranged from 57 to 100%, ignoring the cases of no coincidence due to absence of genetic variability (expansion volume, Programme 2, $S_1 \times S_1$ test). All correlations between genotypic values were, with one exception, $>0\cdot74$. With half-sib families, the coincidences between selected parents were also high for

Table 1. *Restricted maximum-likelihood estimates of the genotypic (G), general and specific combining ability and error variances, and the standard errors (S.E.) from the analyses with and without pedigree information, relative to expansion volume (EV; ml/g) and grain yield (GY; kg/ha) of progeny from two RRS programmes (BF = Beija-Flor and V = Viçosa)*

| Trait | Programme | Progeny | No pedigree information | | Pedigree information | | | |
			$\tilde{\sigma}^2_G$	$\tilde{\sigma}^2$	$\tilde{\sigma}^2_{GCA(BF)}$	$\tilde{\sigma}^2_{GCA(V)}$	$\tilde{\sigma}^2_{SCA}$	$\tilde{\sigma}^2$
EV	$BF_{C1} \times V_{C0}$	$S_0 \times S_0$	6·3 ± 1·16	8·2				
		$S_1 \times S_1$	3·2 ± 0·60	8·2	3·4 ± 0·92	0·1 ± 0·59		8·3
	$BF_{C3} \times V_{C3}$	$S_0 \times S_0$	0·8 ± 0·67	13·1				
		$S_1 \times S_1$	0·7 ± 1·11	14·2	1·3 ± 0·86	0·6 ± 0·77		13·5
		$S_2 \times S_2$	16·3 ± 3·66	13·4	1·4 ± 0·61	0·4 ± 0·40	1·4 ± 0·88	13·7
		Half-sib BF_{C3}			2·2 ± 1·07			16·9
		Half-sib V_{C3}				3·0 ± 1·13		14·8
		Joint analysis			1·0 ± 0·60	0·8 ± 0·61		15·5
GY	$BF_{C1} \times V_{C0}$	$S_0 \times S_0$	121 068 ± 33 025	288 058				
		$S_1 \times S_1$	248 158 ± 42 123	586 739	57 874 ± 45 600	94 963 ± 48 244		686 872
	$BF_{C3} \times V_{C3}$	$S_0 \times S_0$	324 177 ± 61 715	583 618				
		$S_1 \times S_1$	364 906 ± 149 031	1 219 120	8179 ± 60 939	0	499 911 ± 110 574	792 593
		$S_2 \times S_2$	216 725 ± 58 584	353 467	20 258 ± 24 634	0	191 209 ± 42 504	528 924
		Half-sib BF_{C3}			163 363 ± 54 791			732 213
		Half-sib V_{C3}				250 472 ± 75 688		826 892
		Joint analysis			126 880 ± 42 058	195 876 ± 47 726		772 438

Table 2. *Minimum, maximum and average heterosis, standard errors (S.E.) for progeny and for average heterosis and realized interpopulation genetic gain for expansion volume (EV; ml/g) and grain yield (GY; kg/ha) from two reciprocal recurrent selection programmes (BF = Beija-Flor and V = Viçosa)*

| Trait | Programme | Progeny | Heterosis | | | | | Genetic gain |
			Min.	Max.	S.E.	Average	S.E.	
EV	$BF_{C1} \times V_{C0}$	$S_0 \times S_0$	−7	13	2·4	1 (3·0%)	1·4	
		$S_1 \times S_1$	−*	−	−	−	−	−6·1 (−35·2%)
	$BF_{C3} \times V_{C3}$	$S_0 \times S_0$	−7	4	3·2	−1 (−2·7%)	1·9	
		$S_1 \times S_1$	−9	4	3·2	−1 (−3·7%)	1·9	3·1 (9·7%)
		$S_2 \times S_2$	−13	11	3·3	−1 (−3·5%)	1·9	−6·4 (−19·4%)
GY	$BF_{C1} \times V_{C0}$	$S_0 \times S_0$	−1393	1981	448·0	608 (28·5%)	260·0	
		$S_1 \times S_1$	−	−	−	−	−	−703·5 (−25·7%)
	$BF_{C3} \times V_{C3}$	$S_0 \times S_0$	−2006	1753	653·2	−396 (−18·1%)	379·3	
		$S_1 \times S_1$	−735	1421	571·1	305 (12·3%)	334·2	488·6 (27·3%)
		$S_2 \times S_2$	−1574	1798	505·5	−201 (−6·4%)	295·9	−374·3 (−13·4%)

* Trial without the parental populations as controls.

Table 3. *Standard error of prediction or standard deviation and accuracy (in parenthesis) for the effects predicted by BLUP or estimated by BLUE, relative to the expansion volume (EV; ml/g) and grain yield (GY; kg/ha) of progeny from two RRS programmes (BF = Beija-Flor and V = Viçosa)**

| Programme | Progeny | Parameter | BLUP | | BLUE | |
			EV	GY	EV	GY
$BF_{C1} \times V_{C0}$	$S_0 \times S_0$	G–M	1·6 (0·77)	260·1 (0·66)	2·8	644·1
	$S_1 \times S_1$	G–M	1·2 (0·73)	336·7 (0·74)	2·4	769·1
		GCA_{BF}	1·2 (0·85)	245·2 (0·55)		
		GCA_V	0·3 (0·08)	284·0 (0·66)		
$BF_{C3} \times V_{C3}$	$S_0 \times S_0$	G–M	0·8 (0·40)	360·4 (0·77)	3·2	660·0
	$S_1 \times S_1$	G–M	0·8 (0·35)	453·6 (0·66)	3·8	676·5
		GCA_{BF}	1·0 (0·67)	101·7 (0·40)		
		GCA_V	0·8 (0·56)	−		
		SCA	−	360·3 (0·94)		
	$S_2 \times S_2$	G–M	2·0 (0·87)	279·6 (0·80)	3·4	533·7
		GCA_{BF}	1·1 (0·74)	162·5 (0·50)		
		GCA_V	0·7 (0·52)	−		
		SCA	1·1 (0·85)	321·1 (0·91)		
	Half-sib BF_{C3}	GCA_{BF}	1·3 (0·52)	316·9 (0·62)	3·5	710·6
	Half-sib V_{C3}	GCA_V	1·4 (0·61)	368·7 (0·68)	3·3	776·9
	Joint analysis	GCA_{BF}	0·9 (0·39)	306·8 (0·51)		
		GCA_V	0·8 (0·32)	353·5 (0·60)		

* G–M, progeny effect; GCA and SCA, general and specific combining ability effects.

both traits, ranging from 64 to 88%. The correlations ranged from 0·95 to 0·98. The best S_n ($n \geqslant 1$) parents for recombination of inbred progeny based on the genotypic values of the $S_n \times S_n$ progeny estimated by BLUE analysis differed from the selected parents based on the BLUP of the GCA effects of plants from each population. According to these two selection criteria, the percentage of coincident selected parents ranged from 0 to 57%. The difference between these selection criteria is also evidenced by the lower correlations between progeny and GCA effects, which ranged from −0·05 to 0·72. The higher coincidences were

Table 4. *Pearson's correlation between the estimated genotypic values and combining ability effects, and coincidence between selected $S_n \times S_n$ hybrids and parents according to different selection criteria and proportion of selected progeny**

Programme	Progeny	Parameter	Selection Criteria	Selection criteria, expansion volume			Selection criteria, grain yield		
				$\tilde{G}$	$G\tilde{C}A_{BF}$	$G\tilde{C}A_V$	$\tilde{G}$	$G\tilde{C}A_{BF}$	$G\tilde{C}A_V$
$BF_{C1} \times V_{C0}$	$S_0 \times S_0$	Correlation	$\hat{G}$	0·97			0·96		
		Coincidence		0·95/0·90			0·90/0·82		
	$S_1 \times S_1$	Correlation	$\hat{G}$	0·95	0·07	− 0·05	0·98	0·01	0·07
		Coincidence		0·64/0·71	0·14/0·32	0·70/0·24	0·79/0·87	0·0/0·18	0·70/0·29
$BF_{C3} \times V_{C3}$	$S_0 \times S_0$	Correlation	$\hat{G}$	0·97			0·99		
		Coincidence		0·82/0·79			0·94/0·91		
	$S_1 \times S_1$	Correlation	$\hat{G}$	− 0·09	− 0·12	0·04	0·74	0·04	–
		Coincidence		0·0/0·14	0·0/0·29	0·0/0·70	0·57/0·79	0·14/0·29	–
	$S_2 \times S_2$	Correlation	$\hat{G}$	0·90	0·72	0·54	0·97	0·65	–
		Coincidence		0·71/0·71	0·57/0·57	0·29/0·43	0·86/1·00	0·57/0·57	–
	Half-sib BF_{C3}	Correlation	$\hat{G}$		0·95			0·95	
		Coincidence			0·64/0·75			0·86/0·82	
	Half-sib V_{C3}	Correlation	$\hat{G}$			0·96			0·98
		Coincidence				0·67/0·83			0·83/0·88

* $\hat{G}$, selection based on the BLUE of the progeny genotypic value; $\tilde{G}$, selection based on the BLUP of the progeny genotypic value; $G\tilde{C}A$, selection based on the BLUP of the GCA effect of the parent (BF=Beija-Flor and V=Viçosa); coincidence for 10 and 20% of selected progeny.

associated with the lower selection intensity, equivalent to 20% of selected progeny.

DISCUSSION

In the last 63 years, some investigations based on simulations and many applied studies have shown the efficacy of RRS for the improvement of the interpopulation cross and populations. This selection procedure also determines increases in the heterosis of the populations, decreases in the inbreeding depression in the populations, population differentiation and decreases in genetic variability (Penny & Eberhart 1971; Peiris & Hallauer 2005; Solomon et al. 2010; Romay et al. 2011). Until now, analyses have employed the BLUE for progeny genotypic means. Why should plant breeders change their analyses to BLUP or Bayesian inference? This question does not have an easy answer.

Several studies have compared the BLUE, BLUP and Bayesian inference methods for estimating variance components and predicting breeding values. Mathew et al. (2012) observed that REML estimates were as accurate as Gibbs sampling estimates for barley lines data, but the Bayesian analysis provided more accurate estimates for simulated data. Littell (2002) recognized the contribution of the least square methods to

research in agriculture and other areas, but commented that the advantages of the residual maximum-likelihood method compared with ANOVA-based methods far exceed their disadvantages. Van Tassell & Van Vleck (1996) and Van Tassell et al. (1995) also verified that REML estimates were as accurate as Gibbs sampling estimates, especially for traits with high heritability. Harville & Carriquiry (1992) commented that when the ratio between the estimates of genetic and error variances is small and datasets are either limited in extent or are highly unbalanced, the results produced by BLUP can be markedly superior to those obtained by least-squares methods.

The inclusion of pedigree information, historical data and prior knowledge about parameters of the model tends to make BLUP and Bayesian inference superior to BLUE regarding precision and accuracy of breeding values (Atkin et al. 2009; Piepho & Williams 2006; Piepho & Möhring 2006). Forkman & Piepho (2013) compared these three methods in experiments with small number of treatments, based on simulation. The smallest root-mean-square error (RMSE) – a measure of precision of the difference between two treatment effects – was obtained with the Bayesian approach. The BLUP analysis provided consistently smaller RMSE compared with the BLUE analysis. Several experimental results have evidenced

equivalence between breeding values predicted by BLUP and Bayesian analysis (Schenkel *et al.* 2002; Waldmann *et al.* 2008; Bauer *et al.* 2009; Mathew *et al.* 2012). Furthermore, the breeding values presented the same accuracy. Harville & Carriquiry (1992) showed that BLUP and Bayesian predictions are highly correlated when the ratio between the estimates of genetic and error variances is greater than zero. Henderson (1986) commented that as covariance matrices estimated by ANOVA-based methods may be negative definite, very inaccurate predictions of breeding values can be obtained.

It should be emphasized that the success of BLUP and Bayesian inference in crop species breeding depends upon the availability of flexible software for data processing such as ASReml and Just Another Gibbs Sampler (JAGS) (Plummer 2012), the recognition of the advantages of these methods and the professional training to interpret results. These factors will determine the acceptance of BLUP and Bayesian inference by breeders of crop species, at least as alternative procedures for genetic evaluation or as the options for more complex models as highlighted by Blasco (2001). Although Bayesian inference is still infrequently used in plant breeding (Bauer *et al.* 2009; Mathew *et al.* 2012), genetic evaluation by BLUP has been used in recurrent selection programmes and for developing pure and inbred lines (Piepho *et al.* 2008).

In RRS programmes with inbred full-sib families ($S_n \times S_n$, $n \geqslant 1$), selection of parents for recombination of S_{n+1} progeny is more efficient when based on the GCA effect compared with the average genotypic value of the family. Viana *et al.* (2013) showed increases in the intra- and interpopulation genetic gains ranging from 12 to 194%, depending on both the parental populations and the trait. Selection based on GCA, however, is only possible with BLUP or Bayesian approaches that include pedigree information. For selection of the best $S_n \times S_n$ hybrids (genotypic value-based selection) and with standard half- and full-sib RRS (no related parents), these methods are at least as efficient as BLUE.

REFERENCES

ATKIN, F. C., DIETERS, M. J. & STRINGER, J. K. (2009). Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. *Theoretical and Applied Genetics* **119**, 555–565.

BAUER, A. M., REETZ, T. C. & LÉON, J. (2006). Estimation of breeding values of inbred lines using best linear unbiased prediction (BLUP) and genetic similarities. *Crop Science* **46**, 2685–2691.

BAUER, A. M., REETZ, T. C., HOTI, F., SCHUH, W. D., LÉON, J. & SILLANPÄÄ, M. J. (2009). Bayesian prediction of breeding values by accounting for genotype-by-environment interaction in self-pollinating crops. *Genetic Research, Cambridge* **91**, 193–207.

BERNARDO, R. (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Science* **34**, 20–25.

BERNARDO, R. (1996). Best linear unbiased prediction of maize single-cross performance. *Crop Science* **36**, 50–56.

BLASCO, A. (2001). The Bayesian controversy in animal breeding. *Journal of Animal Science* **79**, 2023–2046.

COMSTOCK, R. E., ROBINSON, H. F. & HARVEY, P. H. (1949). A breeding procedure designed to make maximum use of both general and specific combining ability. *Agronomy Journal* **41**, 360–367.

DOS REIS, M. C., GUEDES, F. L., ABREU, G. B. & SOUZA, J. C. (2012). Reciprocal recurrent selection in maize enhances heterosis and ears yield. *Euphytica* DOI: 10.1007/s10681-012-0762-5

FORKMAN, J. & PIEPHO, H. P. (2013). Performance of empirical BLUP and Bayesian prediction in small randomized complete block experiments. *Journal of Agricultural Science, Cambridge* **151**, 383–397.

GILMOUR, A. R., GOGEL, B. J., CULLIS, B. R. & THOMPSON, R. (2009). *ASReml User Guide Release 3.0*. Hemel Hempstead, UK: VSN International Ltd.

HALLAUER, A. R. (1967). Development of single-cross hybrids from two-eared maize populations. *Crop Science* **7**, 192–195.

HALLAUER, A. R. & EBERHART, S. A. (1970). Reciprocal full-sib selection. *Crop Science* **10**, 315–316.

HARVILLE, D. A. & CARRIQUIRY, A. L. (1992). Classical and Bayesian prediction as applied to an unbalanced mixed linear model. *Biometrics* **48**, 987–1003.

HENDERSON, C. R. (1974). General flexibility of linear model techniques for sire evaluation. *Journal of Dairy Science* **57**, 963–972.

HENDERSON, C. R. (1986). Recent developments in variance and covariance estimation. *Journal of Animal Science* **63**, 208–216.

KEERATINIJAKAL, V. & LAMKEY, K. R. (1993). Responses to reciprocal recurrent selection in BSSS and BSCB1 maize populations. *Crop Science* **33**, 73–77.

LITTELL, R. C. (2002). Analysis of unbalanced mixed model data: a case study comparison of ANOVA versus REML/GLS. *Journal of Agricultural, Biological and Environmental Statistics* **4**, 472–490.

MATHEW, B., BAUER, A. M., KOISTINEN, P., REETZ, T. C., LÉON, J. & SILLANPÄÄ, M. J. (2012). Bayesian adaptive Markov chain Monte Carlo estimation of genetic parameters. *Heredity* **109**, 235–245.

MRODE, R. A. (2005). *Linear Models for the Prediction of Animal Breeding Values*, 2nd edn, Wallingford, UK: CABI Publishing.

OAKEY, H., VERBYLA, A. P., CULLIS, B. R., WEI, X. & PITCHFORD, W. S. (2007). Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. *Theoretical and Applied Genetics* **114**, 1319–1332.

ORDÁS, B., BUTRÓN, A., ALVAREZ, A., REVILLA, P. & MALVAR, R. A. (2012). Comparison of two methods of reciprocal recurrent selection in maize (*Zea mays* L). *Theoretical and Applied Genetics* **124**, 1183–1191.

PATTERSON, H. D. & THOMPSON, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58**, 545–554.

PEIRIS, B. L. & HALLAUER, A. R. (2005). Comparison of half-sib and full-sib reciprocal recurrent selection and their modifications in simulated populations. *Maydica* **50**, 25–37.

PENNY, L. H. & EBERHART, S. A. (1971). Twenty years of reciprocal recurrent selection with two synthetic varieties of maize (*Zea mays* L). *Crop Science* **11**, 900–903.

PIEPHO, H. P. & MÖHRING, J. (2006). Selection in cultivar trials – is it ignorable? *Crop Science* **46**, 192–201.

PIEPHO, H. P. & WILLIAMS, E. R. (2006). A comparison of experimental designs for selection in breeding trials with nested treatment structure. *Theoretical and Applied Genetics* **113**, 1505–1513.

PIEPHO, H. P., MÖHRING, J., MELCHINGER, A. E. & BÜCHSE, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* **161**, 209–228.

PLUMMER, M. (2012). *JAGS: Just Another Gibbs Sampler v.3.3.0.* Available from: http://mcmc-jags.sourceforge.net/ (verified 12 November 2012).

REAL Software (2004). *REALbasic 5.5. Version 5.5.4.* Austin, TX: REAL Software Inc.

ROMAY, M. C., ORDÁS, B., REVILLA, P. & ORDÁS, A. (2011). Three cycles of full-sib reciprocal recurrent selection in two Spanish maize populations. *Crop Science* **51**, 1016–1022.

SAS Institute (2007). *The SAS System for Windows, Version 9.2.* Cary, NC: SAS Institute Inc.

SCHENKEL, F. S., SCHAEFFER, L. R. & BOETTCHER, P. J. (2002). Comparison between estimation of breeding values and fixed effects using Bayesian and empirical BLUP estimation under selection on parents and missing pedigree information. *Genetics Selection Evolution* **34**, 41–59.

SOLOMON, K. F., MARTIN, I. & ZEPPA, A. (2010). Temporal genetic structure patterns in tropical maize populations under reciprocal recurrent selection. *Euphytica* **176**, 239–249.

SOUZA, C. L. Jr. (1987). Reciprocal recurrent selection with half-sib progenies obtained alternately from non-inbred (S_0) and inbred (S_1) plants in maize. *Maydica* **32**, 19–31.

VAN TASSELL, C. P., CASELLA, G. & POLLAK, E. J. (1995). Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. *Journal of Dairy Science* **78**, 678–692.

VAN TASSELL, C. P. & VAN VLECK, L. D. (1996). Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. *Journal of Animal Science* **74**, 2586–2597.

VIANA, J. M. S., FARIA, V. R., SILVA, F. F. & RESENDE, M. D. V. (2011). Best linear unbiased prediction and family selection in crop species. *Crop Science* **51**, 2371–2381.

VIANA, J. M. S., DELIMA, R. O., MUNDIM, G. B., CONDÉ, A. B. T. & VILARINHO, A. A. (2013). Relative efficiency of the genotypic value and combining ability effects on reciprocal recurrent selection. *Theoretical and Applied Genetics* **126**, 889–899.

WALDMANN, P., HALLANDER, J., HOTI, F. & SILLANPÄÄ, M. J. (2008). Efficient Markov chain Monte Carlo implementation of Bayesian analysis of additive and dominance genetic variances in noninbred pedigrees. *Genetics* **179**, 1101–1112.

APPENDIX

ASReml code for genetic evaluation in RRS

```
BLUP analysis, RRS, 3 cycles (S₀×S₀, S₁×S₁ and S₂×S₂ progeny and checks) # title line
   trial 3 # progeny tests
   pop 16 # 1, 2 and 3 for progeny and 4–16 for the checks
   rep 3 # replications
   blo 13 # blocks
   ga 327 !i !sort # 314 gca effects of parents from population A
   gb 327 !i !sort # 314 gca effects of parents from population B
   sab 327 # 314 sca effects for progeny
   s # final stand – ideal stand
   m # grain moisture – 14.5
   y # grain yield
   ev # expansion volume
A1.grm # additive relationship matrix for population A
```

```
A2.grm # additive relationship matrix for population B
D.grm # dominance relationship matrix
s0s1s2.asd # data file
!continue # more than 10 iterations (default), if necessary
!maxit 100 # maximum number of iterations
y ~ mu pop trial pop.trial trial.rep s m !r trial.rep.blo ga gb sab !f mv # mv estimates missing values
#ev ~ mu pop trial pop.trial trial.rep !r trial.rep.blo ga gb sab !f mv
1 1 4 # one R structure and four G structures
0 0 id
trial.rep.blo 1
trial.rep.blo 0 id 1
ga 1
ga 0 giv1 1 # giv1 is the inverse of A1.grm
gb 1
gb 0 giv2 1 # giv2 is the inverse of A2.grm
sab 1
sab 0 giv3 1 # giv3 is the inverse of D.grm


BLUP analysis, RRS, 1st cycle and three progeny tests # title line
   trial 3 # progeny tests (half-sib families from populations A and B and full-sib families)
   rep 3
   blo 13
   ga 290
   gb 313
   sab 169
   s
   m
   y
   ev
s0.asd
!continue
!maxit 100
y ~ mu trial trial.rep s m !r trial.rep.blo -at(trial,2).ga ga -at(trial,1).gb gb -at(trial,1,2).sab sab !f mv
#ev ~ mu trial trial.rep !r trial.rep.blo -at(trial,2).ga ga -at(trial,1).gb gb -at(trial,1,2).sab sab !f mv
1 1 4
0 0 id
trial.rep.blo 1
trial.rep.blo 0 id 1
ga 1
ga 0 id 1
gb 1
gb 0 id 1
sab 1
sab 0 id 1
```