Selection via simulated individual BLUP based on family genotypic effects in sugarcane

Marcos Deon Vilela de Resende[1] and Márcio Henrique Pereira Barbosa[2]

[1]Embrapa Gado de Corte, Rod. BR 262, Km 4, Caixa Postal 154, CEP 79002-970 Campo Grande, MS, Brazil. E-mail: marcosdeon@cnpgc.embrapa.br
[2]Universidade Federal de Viçosa, Dep. de Fitotecnia, CEP 36570-000 Viçosa, MG, Brazil. E-mail: barbosa@ufv.br

Abstract – The objective of this work was to propose a new selection strategy for the initial stages of sugarcane improvement, based on the methodology 'simulated individual BLUP (BLUPIS)', which promotes a dynamic allocation of individuals selected in each full-sib family, using BLUP as a base for both the genotypic effects of the referred families and plot effects. The method proposed applies to single full-sib families or those obtained from unbalanced or balanced diallel crosses, half-sib families and self-pollinated families. BLUPIS indicates the number of individuals to be selected within each family, the total number of clones to be advanced, and the number of families to contribute with selected individuals. Correlation between BLUPIS and true BLUP was 0.96, by method validation. Additionally, BLUPIS allows the identification of which replication contains the best individuals of each family.

Index terms: BLUP/REML, BLUPIS, selection strategies, mixed models, sugarcane breeding.

Seleção via BLUP individual simulado baseado nos efeitos genotípicos de famílias em cana-de-açúcar

Resumo – O objetivo deste trabalho foi propor uma nova estratégia de seleção nos estádios iniciais do desenvolvimento da cana-de-açúcar, utilizando-se a metodologia BLUP individual simulado (BLUPIS) que promove a distribuição dinâmica dos indivíduos selecionados em cada família de irmãos-completos, usando BLUP como base para os efeitos genotípicos da família e para os efeitos de parcela. O método proposto se aplica a famílias de irmãos-completos simples ou obtidas de cruzamentos dialélicos desbalanceados ou balanceados, famílias de meios-irmãos e famílias de autofecundação. Por meio do BLUPIS, indica-se o número de indivíduos a ser selecionado por família, o número total de clones a ser avançado e o número de famílias a contribuir com indivíduos selecionados. A validação do método propiciou uma correlação de 0,96 entre o BLUPIS e o BLUP verdadeiro. Além disso, o BLUPIS permite identificar em qual repetição encontram-se os melhores indivíduos de cada família.

Termos para indexação: BLUP/REML, BLUPIS, estratégias de seleção, modelos mistos, melhoramento de cana-de-açúcar.

Introduction

Crosses between superior parents, followed by individual selection aiming at cloning, are the classical procedure adopted in the improvement of asexually propagated species. In many of these species, field experimentation is based on the evaluation of plot totals or means, without using data from individual plants. This is the case of crops such as sugarcane (Matsuoka et al., 2005), forages as *Brachiaria* spp., *Panicum* spp. and elephantgrass (Ferreira & Pereira, 2005) and potato (Barbosa & Pinto, 1998). In these species, individual selection is frequently practiced without using family information, in other words, it is a mass selection (Matsuoka et al., 2005). Otherwise, a moderate or weak selection intensity is performed among families, and mass selection is practiced within the selected families, meaning that family genotypic effect is not effectively used as a guide for individual selection.

In sugarcane breeding, individual selection in the initial stages has been based on mass selection methods (Mariotti et al., 1999; Matsuoka et al., 2005), Australian sequential selection (among families selection followed by mass selection) (McRae et al., 1998; Cox et al., 2000; Kimbeng & Cox, 2003), and modified sequential selection (Bressiani, 2001). The two latter methods use family information and are therefore superior to mass selection for characters presenting

heritability based on family means, higher than the heritability at individual level.

The optimum selection strategy would be through genotypic values predicted by individual BLUP (Resende, 2002b) that would use simultaneously information on family and individual for selection. However, this method has not been used in sugarcane breeding due to difficulties in obtaining data from individual plants.

Production of superior hybrid families allows individual selection efficiency to be increased. The genotypic value is the best parameter to describe superiority of a certain cross. Although genetic variance within families could be also another parameter, besides the mean, to infer the potential of a certain cross for generating superior individuals, its estimate presents errors greater than that for the variance among families and involves high additional cost that will result in low overall efficiency of selection.

Another issue is that about 90% of the individuals have been discarded on the grounds of restrictive characters of high heritability. In the Australian sugarcane improvement program, the proportion of selected individuals is around 8%, that is, about 2,800 clones obtained from a population of approximately 35 thousand seedlings (Cox et al., 2000). This discarded percentage, approximately 90%, has also been practiced by other sugarcane improvement programs around the world. In the South African program, approximately 4,000 clones have been selected from a population of 35 thousand seedlings. Therefore, additional recording data of individual plants for individual BLUP is not worth, since many genotypes would be discarded due to restrictive characters of high heritability.

In sugar cane, the alternative and nondestructive procedure to evaluate stalk production of individual plants would be the one used by Chang & Milligan (1992). Assuming that the stalks are perfect cylinders, ratoon stalks (p) or individual weight would have to be obtained by the expression $p = d.\pi.r^2.c.n$, in which d is the density considered equal to 1 g cm^{-3}, r is the stalk radius, c is the stalk length and n is the ratoon stalk number.

The whole work has to be carried out without previous trash burning. Besides, there is additional difficulty when genotypes tumble for expressive stalk development, or for its own decumbent growth habit or, also, for plant lodging due to the wind action. Harvesting the experiment becomes more effective when the whole family plot is weighed through the manual harvest system with previous burning or through mechanical harvester, as in Australia (Cox et al., 2000). Consequently, a practical procedure similar to individual BLUP is necessary for increasing sugarcane breeding efficiency.

The objective of this work was to propose a new selection strategy for the initial stages of sugarcane improvement, based on the methodology 'simulated individual BLUP (BLUPIS)', which promotes a dynamic allocation of individuals selected in each full-sib family, using BLUP as a base for both the genotypic effects of the referred families and for plot effects.

Material and Methods

Experimental details

Three experiments, each with eight blocks of 16 regular treatments and three controls, were arranged in an augmented block design and established in a same experimental area at Centro de Pesquisa e Melhoramento da Cana-de-Açúcar (CECA), Universidade Federal de Viçosa, in the municipal district of Oratórios, MG (20º25'S; 42º48'W; altitude 494 m; Rhodic Eutrudox soil).

Regular treatments were represented by 113 full-sib families from unbalanced diallel crosses. The common treatments consisted of three cultivars, RB72454, RB835486 and RB739359. Cultivar RB72454 was used as lateral border of the experiment. Soil fertilization used 500 kg ha^{-1} of a formula containing 5% of N, 25% P_2O_5 and 25% K_2O.

Crosses were performed by Copersucar, at Camamu, BA. To prevent self-fertilization, all the inflorescences used as female were emasculated with hot water (Machado Junior et al., 1995).

Seed germination took place in August 1999, and seedling transplanting to the field in November 1999. Families and cultivars were evaluated in double-row plots with ten plants each. The inter-row spacing was 1.40 m and intra-row spacing 0.5 m. In July 2000, all plants were manually cut with machete, as a mean to submit seedlings to natural selection for ratooning ability in unfavorable environmental conditions, that is, dry and cold seasons. In May 2001, data collection was carried out in the ratoon.

The traits appraised at plot level were the total number of millable stalks (NS) and weigh of 20 stalks, randomly sampled, with subsequent transformation for stalk mean weight (SMW). Tonnes of cane per hectare (TCH) were obtained by multiplying NS by SMW.

Data analysis procedures

Statistical analyses were performed with the genetics and statistics software program Selegen-REML/BLUP (Resende, 2002a). Mixed model equations (Resende, 2002b) were used to calculate BLUP of the genetic values, and the specific combining ability (SCA) of each family for NS, SMW and TCH, considering the relationship matrix described.

The mixed linear model used was $y = Xl + Za + Wc + Ub + e$, in which y, l, a, c, b and e are, respectively, the vectors of data of the fixed effects of the experiments, the random additive genetic effects, the random SCA effects, the random block effects, and random errors; X, Z, W, and U are the matrices of incidence of l, a, c, and b, respectively.

The following mixed model equations were used to calculate BLUP:

$$\begin{bmatrix} X'X & X'Z & X'W & X'U \\ Z'X & Z'Z+A^{-1}\lambda_1 & Z'W & Z'U \\ W'X & W'Z & W'W+I\lambda_2 & W'U \\ U'X & U'Z & U'W & U'U+I\lambda_3 \end{bmatrix} \begin{bmatrix} \hat{l} \\ \hat{a} \\ \hat{c} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \\ U'y \end{bmatrix},$$

in which:

$$\lambda_1 = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1 - h^2 - c^2 - b^2}{h^2}; \lambda_2 = \frac{\sigma_e^2}{\sigma_c^2} = \frac{1 - h^2 - c^2 - b^2}{c^2};$$

$$\lambda_3 = \frac{\sigma_e^2}{\sigma_b^2} = \frac{1 - h^2 - c^2 - b^2}{b^2};$$

$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_b^2 + \sigma_e^2}$ is the narrow sense individual heritability; $c^2 = \frac{\sigma_c^2}{\sigma_a^2 + \sigma_c^2 + \sigma_b^2 + \sigma_e^2}$ is the determination coefficient of the specific combining ability effects; and $b^2 = \frac{\sigma_b^2}{\sigma_a^2 + \sigma_c^2 + \sigma_b^2 + \sigma_e^2}$ is the correlation due to the common environment of the block.

To calculate the heritability estimates at the individual level and at the level of full-sib family means, iterative estimators of the variance components by REML via EM algorithm were obtained as follows:

$$\hat{\sigma}_e^2 = [y'y - \hat{l}'X'y - \hat{a}'Z'y - \hat{c}'W'y - \hat{b}'U'y]/[N - r(x)]$$

$$\hat{\sigma}_a^2 = [\hat{a}'A^{-1}\hat{a} + \hat{\sigma}_e^2 \, tr\,(A^{1}C^{22})]/q$$

$$\hat{\sigma}_c^2 = [\hat{c}'\hat{c} + \hat{\sigma}_e^2 \, tr\, C^{33}]/s_1;$$

$\hat{\sigma}_b^2 = [\hat{b}'\hat{b} + \hat{\sigma}_e^2 \, tr\, C^{44}]/s_2$, in which C^{22}, C^{33} and C^{44} come from the inverse of C; C is the matrix of the mixed model equations coefficients; tr is the trace of a matrix operator; r(x) is the rank of the X matrix; N, q, s_1, and s_2 are the total number of data, of parents, of crosses and of blocks, respectively.

The estimator of the component of dominance variance between families is given by $\hat{\sigma}_d^2 = \hat{\sigma}_c^2$; in other words, it is equal to the variance component associated to the specific combining ability. In this case, $\hat{\sigma}_d^2$ is equivalent to ¼ of the genetic variance of total dominance in the population.

Family genotypic effects were predicted by $(1/2)(\hat{a}_i + \hat{a}_j) + \hat{c}_{ij}$, where $\hat{a}_i$ and $\hat{a}_j$ are the addictive genetic values, predicted for the parents i and j, respectively, and $\hat{c}_{ij}$ is the specific combining ability of the cross between parents i and j.

Selection strategy based on the dynamic allocation of the number of individuals selected per family using simulated individual BLUP (BLUPIS)

The ideal procedure of individual selection for cloning, in the initial stages of the sugarcane improvement program, is individual BLUP considering simultaneously information on individual, family, experimental design and relationship between families and parents. However, information about the individual is not usually obtained when families are being evaluated, because these are estimated by total harvest of plots.

The real genotypic value, intrinsic or parametric of these non-evaluated individuals, considering the individual i from family j, is given by $u + g_{ij} = u + g_j + g_{i/j}$, in which u is the general mean; g_{ij} is the genotypic effect of the individual ij; g_j is the genotypic effect of the family j; and $g_{i/j}$ is the genotypic effect within family of the individual ij. This expression can be rewritten as $u + g_{ij} = u + g_j + h^2_{gd}(y_{ij} - g_j) = u + g_j(1 - h^2_{gd}) + h^2_{gd}y_{ij}$, in which y_{ij} is the phenotypic observation of the individual ij; and h^2_{gd} is the genotypic heritability within the full-sib family, whose numerator is given by $(1/2)\sigma_a^2 + (1/4)\sigma_d^2$. BLUP of $u + g_{ij}$ is given by $\hat{u} + \hat{g}_{ij} = \hat{u} + \hat{g}_j + h^2_{gd}(y_{ij} - \hat{g}_j) = \hat{u} + \hat{g}_j(1 - h^2_{gd}) + h^2_{gd}y_{ij}$, in which $\hat{g}_j$ is BLUP for full-sib families, obtained after considering the genetic relationship between families and between parents involved in the genetic evaluation; but as y_{ij} was not observed, such BLUP can not be calculated explicitly. However, the comparison between the BLUPs of two different individuals ij and lk, belonging

to families j and k, can be done. In this case, the individual from family j will be superior to the individual from family k, if $\hat{u} + \hat{g}_j (1 - h^2_{gd}) + h^2_{gd}y_{ij} > \hat{u} + \hat{g}_k (1 - h^2_{gd}) + h^2_{gd}y_{lk}$. The quantities $h^2_{gd}y_{ij}$ and $h^2_{gd}y_{lk}$, i.e., the fractions of y dictated by the heritability within family, are completely independent of the family genotypic values $\hat{g}_j$ and $\hat{g}_k$, and are completely random as they are effects of Mendelian segregation. Consequently, $h^2_{gd}y_{ij}$ and $h^2_{gd}y_{lk}$ present equal expected value $h^2_{gd}y$. Hence, in average or expected value, the individual from family j will be superior if $\hat{u} + \hat{g}_j (1 - h^2_{gd}) - h^2_{gd}y > \hat{u} + \hat{g}_k (1 - h^2_{gd}) + h^2_{gd}y$. In other words, if $\hat{g}_j > \hat{g}_k (1 - h^2_{gd})/(1 - h^2_{gd}) + h^2_{gd}y - h^2_{gd}y + \hat{u} - \hat{u}$, therefore if $\hat{g}_j > \hat{g}_k$, i.e., if $\hat{g}_j - \hat{g}_k > 0$ or still if $\hat{g}_j/\hat{g}_k > 1$. Thus, $\hat{g}_j/\hat{g}_k$ indicates the mean rate of superior individuals from family j, in relation to the individuals from family k. If $\hat{g}_j/\hat{g}_k = 1.2$ and 40 individuals are selected per family k, 48 individuals from family j should be selected, so that the worst individual selected from family j may have the same level of the worst individual selected from family k. In this case, these 88 individuals should coincide approximately with the 88 best individuals, which would have been selected by the BLUP applied in the selection of individuals belonging to these two families.

To sum up, the establishment of the number of individuals to be selected in each family, by using the relation among the genotypic effects of full-sib families, will simulate adequately the selection through individual BLUP. For this reason, such procedure will be denominated 'simulated individual BLUP (BLUPIS)', and the expression that will determine, in a dynamic way, the number of individuals n_k selected in each family k is given by $n_k = (\hat{g}_k/\hat{g}_j)n_j$, in which $\hat{g}_j$ refers to the genotypic value of the best family and n_j is equal to the number of individuals selected in the best family. The determination of n_j involves the concept of effective population size. Alternatively, such expression can be given by $n_k = [1 - (\hat{g}_j - \hat{g}_k)/(\hat{g}_j)]n_j = (\hat{g}_k/\hat{g}_j)n_j$. The latter expression shows that n_k depends on the differences among the genotypic effects of the two families, as a proportion of the best family's genotypic effect. The method eliminates automatically the families with negative genotypic effect, that is to say, those below the general mean of the experiment. This seems reasonable, considering the extremely low probability of obtaining a superior clone in these families.

Results and Discussion

The first step concerned with the definition of the number of individuals to be selected (included in clonal tests) within the best full-sib family, presented at Table 1, was to study the genetic representativeness of a full-sib family in terms of its effective population size (Ne). The maximum Ne of a full-sib family is 2, and the number of individuals needed to approach Ne = 2 is given by the expression Ne = [2n/(n+1)], as shown by Vencovsky (1978).

Table 2 shows the number of individuals per family needed to reach determined percentage of its maximum Ne. When n = 50 individuals, one gets 98% of the maximum family representativeness, and to reach 99% there is a need of twice as many, in other words, 100 individuals per family. Therefore increasing sampling within family, starting from n = 50, barely contributes to add different individuals in the sample. This means that many average individuals and a few extreme (including superior genotypes here) are added, when the sample is increased from n = 50. Hence, it is believed that 50 (at maximum) individuals of the best family, mass selected for several restrictive traits, are enough to hold the best individual in the progeny for productivity, which will be later identified by clonal test. It is important to state that half-sib progenies or polycross mating need 150 individuals to reach 98% of maximum representativeness (Ne = 4) of a family. Thus, with polycrosses, a larger number of individuals per family is recommended and, consequently, there will be a larger total number of clones to be evaluated. With full-sib families derived from related parents, the Ne maximum is smaller than 2, so that less than 50 individuals per family to reach 98% of the maximum Ne are needed.

When 50 individuals are taken from the best family, the number of individuals to be taken from the other families is a function of the relative proportion among genotypic effects of families predicted by BLUP. According to the simulated individual BLUP propositions, the number of individuals to be selected in the other families k is given by $n_k = (\hat{g}_k/\hat{g}_1)50$, where $\hat{g}_k$ and $\hat{g}_1$ refer to the predicted genotypic values of families k and the best family (number 1 in the ranking), respectively.

The number of individuals selected per family decreased progressively and slowly from 50 (for the best family) to 0 (for the average family), in the three traits evaluated (Table 1). These results reveal the importance of the dynamic allocation, dependent on the relative

difference among the genotypic effects of families in evaluation, of the number of individuals selected per family, to the detriment of the a priori acceptance of fixed proportions of selection within families, as proclaimed and practiced in Australia (McRae et al., 1998; Cox et al., 2000; Kimbeng & Cox, 2003).

The total number of clones to be advanced is also automatically determined by this methodology and it depends on the magnitude of the differences among the families in evaluation as well. This number was around 990 for stalk number and TCH, and 870 for stalk mean weight, when 50 individuals of the best family were selected (Table 1). That represented a selection proportion of approximately 14% for TCH and NS. When 30 individuals from the best family were selected, the total number of clones selected was 590, 599 and 522 for TCH, NS and SMW, respectively. In this case, the selection proportion dropped to approximately 9%. The percentages mentioned, 9 to 14%, are in agreement with the selection proportions applied to several sugarcane improvement programs in the world, as previously discussed.

The simulated individual BLUP indicated that individuals, in different proportions within the families,

Table 1. Genotypic effect (g_k) and number of individuals to be selected from the *k*th full-sib sugarcane family, via simulated BLUP for the traits stalk tons per hectare (TCH), stalk number (NS) and stalk mean weight (SMW)[(1)].

Family code	TCH			Family code	NS			Family code	SMW		
	g_j	n_j=30	n_j=50		g_j	n_j=30	n_j=50		g_j	n_j=30	n_j=50
	g_k	n_k	n_k		g_k	n_k	n_k		g_k	n_k	n_k
108	10.55	30	50	35	13.33	30	50	108	0.14	30	50
38	9.92	28	47	93	10.64	24	40	38	0.10	22	36
47	8.31	24	39	99	10.09	23	38	107	0.10	22	36
103	7.97	23	38	92	8.92	20	33	46	0.09	19	32
35	7.73	22	37	7	8.86	20	33	111	0.09	19	32
46	7.58	22	36	33	8.65	19	32	86	0.09	19	32
109	6.73	19	32	32	8.42	19	32	103	0.08	17	29
95	6.68	19	32	40	8.39	19	31	49	0.07	15	25
49	6.56	19	31	34	7.65	17	29	47	0.07	15	25
34	6.47	18	31	38	7.57	17	28	106	0.07	15	25
32	6.33	18	30	58	7.51	17	28	45	0.06	13	22
80	5.52	16	26	100	7.19	16	27	76	0.06	13	22
40	5.43	15	26	117	7.01	16	26	95	0.06	13	22
33	5.06	14	24	101	6.99	16	26	109	0.06	13	22
97	4.50	13	21	71	6.49	15	24	48	0.06	13	22
7	4.22	12	20	47	6.29	14	24	80	0.06	13	22
86	4.10	12	19	27	6.19	14	23	87	0.06	13	22
48	4.04	11	19	28	5.79	13	22	97	0.06	13	22
111	4.04	11	19	91	5.69	13	21	8	0.05	11	18
92	4.01	11	19	98	5.42	12	20	22	0.05	11	18
112	3.88	11	18	59	5.38	12	20	34	0.05	11	18
107	3.70	11	18	37	5.30	12	20	104	0.05	11	18
1	3.68	10	17	109	5.29	12	20	1	0.04	9	15
27	3.50	10	17	74	5.00	11	19	120	0.04	9	15
37	3.41	10	16	49	4.98	11	19	94	0.04	9	15
76	3.30	9	16	6	4.88	11	18	118	0.04	9	15
117	3.11	9	15	103	4.55	10	17	11	0.04	9	15
6	2.97	8	14	95	4.47	10	17	82	0.04	9	15
58	2.88	8	14	89	4.39	10	16	112	0.04	9	15
28	2.88	8	14	84	4.31	10	16	53	0.03	7	11
106	2.62	7	12	102	4.26	10	16	4	0.03	7	11
71	2.60	7	12	1	4.14	9	16	32	0.03	7	11
22	2.44	7	12	73	4.04	9	15	41	0.03	7	11
73	2.23	6	11	26	4.02	9	15	110	0.03	7	11
93	2.21	6	10	25	3.79	9	14	105	0.03	7	11
25	2.15	6	10	90	3.38	8	13	113	0.03	7	11
89	2.08	6	10	72	3.37	8	13	123	0.02	5	8
17	2.03	6	10	112	3.27	7	12	17	0.02	5	8
84	1.91	5	9	55	3.26	7	12	40	0.02	5	8

Continue.

Table 1. Continuation.

Family code	TCH			Family code	NS			Family code	SMW		
	g_j / g_k	n_j=30 / n_k	n_j=50 / n_k		g_j / g_k	n_j=30 / n_k	n_j=50 / n_k		g_j / g_k	n_j=30 / n_k	n_j=50 / n_k
104	1.87	5	9	96	3.24	7	12	37	0.02	5	8
83	1.85	5	9	108	3.05	7	11	83	0.02	5	8
120	1.84	5	9	66	2.96	7	11	60	0.02	5	8
4	1.79	5	8	80	2.74	6	10	35	0.02	5	8
8	1.73	5	8	62	2.27	5	9	75	0.02	5	8
45	1.67	5	8	83	1.79	4	7	85	0.02	5	8
75	1.52	4	7	85	1.40	3	5	42	0.02	5	8
62	1.44	4	7	46	1.32	3	5	18	0.02	5	8
11	1.42	4	7	97	1.31	3	5	6	0.01	3	4
91	1.32	4	6	56	1.24	3	5	81	0.01	3	4
85	1.24	4	6	122	1.24	3	5	33	0.01	3	4
94	1.24	4	6	75	1.02	2	4	56	0.01	3	4
99	1.06	3	5	76	0.82	2	3	73	0.01	3	4
87	0.99	3	5	48	0.78	2	3	12	0.00	0	1
56	0.94	3	4	116	0.67	1	2	7	0.00	0	1
26	0.88	3	4	3	0.63	1	2	13	0.00	0	1
53	0.81	2	4	120	0.33	1	1	63	0.00	0	1
82	0.65	2	3	60	0.29	1	1	21	0.00	0	1
110	0.63	2	3	13	0.02	0	0	25	0.00	0	1
105	0.61	2	3		-	-	-	119	0.00	0	1
74	0.56	2	3		-	-	-	27	0.00	0	1
96	0.51	1	2		-	-	-	28	0.00	0	1
41	0.42	1	2		-	-	-	79	0.00	0	1
100	0.38	1	2		-	-	-	84	0.00	0	1
3	0.29	1	1		-	-	-	62	0.00	0	1
60	0.29	1	1		-	-	-	92	0.00	0	1
102	0.11	0	1		-	-	-	89	0.00	0	1
55	0.08	0	0		-	-	-	126	0.00	0	1
Total		590	983			599	998			522	870

[(1)] g_j: genotypic effect of best family; g_k: genotypic effect of family k; n_j: number of individuals to be selected within best family; n_k: number of individuals to be selected within family k; to apply BLUPIS, the selection of 30 or 50 individuals from the family with the highest genotypic effect was considered.

Table 2. Effective size of a full-sib family (N_{ef}) and fraction of the maximum effective size of a family (N_{efmax}), as a function of the number of individuals sampled per family (N).

N	N_{ef}	N_{efmax}	Fraction of the N_{efmax}
1	1.000	-	0.500
5	1.667	-	0.833
7	1.750	-	0.875
10	1.818	-	0.910
12	1.846	-	0.923
15	1.875	-	0.938
18	1.895	-	0.947
20	1.905	-	0.952
25	1.923	-	0.962
30	1.935	-	0.968
40	1.951	-	0.976
50	1.961	-	0.980
60	1.967	-	0.984
100	1.980	-	0.990
∞	2.000	2.000	1.000

should be advanced, providing gains in selective efficiency by using family information. The proposed procedure should, therefore, be routinely used in the sugarcane improvement practice. It provides three types of important information: the number of individuals to be selected per family, the total number of clones to be advanced and the number of families that contribute to selected individuals. By means of this more appropriate methodology, a smaller number of better clones is advanced, increasing the efficiency of the selective process and reducing costs of the improvement program.

In the selection of 30 individuals from the family of largest genotypic effect for TCH, NS and SMW, about 90% of the selected individuals derived from about 38, 35 and 35% of the total of families, respectively (Table 1). Such numbers are similar to those practiced in other improvement programs (Cox et al., 2000).

Another useful aspect can be provided by the mean genotypic value of the experimental plots within each progeny. Such aspect indicates in which replication the best individuals of each family are.

Methods and strategies used for selection at the initial stages of sugarcane improvement are: mass or individual

selection, as used by Matsuoka et al. (2005) at UFSCar, and Mariotti et al. (1999) in Argentina (i); Australian sequential selection, according to McRae et al. (1998), Cox et al. (2000) and Kimbeng & Cox (2003) (ii); modified sequential selection, as proposed by Bressiani (2001) (iii); multi-effects index selection or BLUP using individual, family and experimental design information (Resende & Higa, 1994; Bressiani, 2001) (iv); selection by simulated individual BLUP as proposed in the present work (v).

All these methods take up implicitly (ii, iii and v) or explicitly (i and iv) some form of individual selection. Method iv or legitimate BLUP is theoretically the most efficient, because it provides higher selective accuracy. However, it demands data collection from individual plants, which, in some situations, would be prohibitive, or it allows the evaluation of a reduced number of individuals in other circumstances, jeopardizing selection intensity. Mass selection (i), on one hand, provides high selection intensity, but on the other hand it is less precise for low heritable traits.

The Australian sequential selection (ii) uses high accuracy for selection among families and allows selection within families guided by their own genotypic values. The selected families are separated in four groups, and around 32, 24, 16 and 8 individuals are selected within the best for the worst group of families, respectively, summing up a total of 2,800 clones of 140 families selected from 300 to 350, evaluated with around 80 to 90 seedlings per family (Cox et al., 2000). In this method, the selection proportions within family in each group are predetermined, and therefore the level of genotypic difference among families in each group is not considered. This question was accounted for by Bressiani (2001), when he proposed the modified sequential selection, attributing differentiated proportions of selection within each family separately. However, the basic principle of this selection is to take numbers of individuals per family, in such a way that the mean of individuals selected from each family is the same for all families (Bressiani, 2001). This implicates that selective process incorporates some individuals from the best family, which are worse than the worst individual from a family (the second family in the ranking) worse than the first. If this is not so, there would not be a way to reach the criterion of equal means. This is a disadvantage that was overcome by the simulated individual BLUP method, which has as principle to admit individuals of the best family, since such individual is equivalent or superior to the worst individual of an inferior family.

The selection of the referred numbers of individuals can be performed optionally: (a) in the second ratoon crop of the own family experiment, as in Australia (Cox et al., 2000); (b) in the field named T1 consisted of seedlings planted for mass selection, without any type of experimental design; (c) through new planting of the selected families and selection within them. Option (a) provides slight advantage over (b) and (c) in the family effect contribution for the individual selection, when family size in the experiment is small. This is due to the fact that individuals to be effectively selected have contributed to the family mean in the experiment. This efficiency is expressed by:

$$E_I = \frac{[(1/2)\,\sigma_a^2 + (1/4)\,\sigma_d^2] + [[(1/2)\,\sigma_a^2 + (3/4)\,\sigma_d^2]/(pb)]}{[(1/2)\,\sigma_a^2 + (1/4)\,\sigma_d^2]} =$$

$$1 - \frac{[[(1/2)\,\sigma_a^2 + (3/4)\,\sigma_d^2]}{[(1/2)\,\sigma_a^2 + (1/4)\,\sigma_d^2](pb)}.$$

Assuming low dominance $\hat{\sigma}_d^2$ tends to zero and $[(1/2)\,\sigma_a^2 + (3/4)\sigma_d^2] / [(1/2)\sigma_a^2 + (1/4)\sigma_d^2]$ tends to 1, so that the selection efficiency according to option (a) is given approximately by $E_1 = 1 + 1/(pb)$ in which p is the number of plants per plot and b is the number of replications; with pb = 60 similar to the present experiment, this efficiency is 1.02 or 2%, therefore low. Options (b) and (c) present advantages, for providing higher intensity of mass selection within families on several restrictive characters. But precision in selection is lower, because of the increased size of the environmental stratum for individual selection. Assuming that the stratum for selection within families has the size of a block in the experiment, the total variance within the stratum is $(\sigma_w^2 + \sigma_e^2)$, compared with only variation within plot (σ_w^2), in case of option (a), in which the efficiency for (a), in heritability terms, is given by $E_2\,(\sigma_w^2 + \sigma_e^2) / (\sigma_w^2) = 1 + (\sigma_e^2 / \sigma_w^2)$ and therefore it depends on the ratio $(\sigma_e^2 / \sigma_w^2)$ being though greater than 1.

Another favorable point to (a) is that BLUP predicts plot genotypic effects for each family, given by

$$\hat{g}_{parc_r} = (y - X\hat{I} - Z\hat{a}_i - Z\hat{a}_j - W\hat{c}_{ij} - U\hat{b})\,[(1/2)\,\sigma_a^2 + (3/4)\,\sigma_d^2]/p\sigma_e^2,$$

providing information about which plot or which replication the superior genotypes of each family are

found in. For instance, for the best family (108), plot genotypic effects for TCH were 10.67, 8.75, and 11.69 for the three replications, respectively. Thus, from the 30 best individuals selected in this family, 10, 8 and 11 genotypes should be selected in replications one, two and three, respectively. These values are given by $(\hat{g}_{parc_r}/\sum \hat{g}_{parc_r}$, in which $\hat{g}_{parc_r}$ is the genotypic effect predicted for plot r. Therefore, these advantages of (a) need to be counterbalanced by the advantage of higher selection intensity in (b) and (c). In case of not choosing (a), one should at least include in the selection the best individual of the plot, with the highest genotypic value for each selected family in the trial.

BLUPIS was validated by using real data (provided by Embrapa), referring to the evaluation of 140 full-sib eucalyptus families, obtained under an unbalanced diallel crossing scheme. Data were collected individually, and 50 individuals, in the experiment, represented each family initially. The results referring to the trait trunk circumference (individual narrow-sense heritability of 20% and individual broad-sense heritability of 30%) are presented in Figure 1. Optimum agreement rate is verified between the real and the simulated BLUP, in terms of coincidence of number of individuals selected. The correlation between the numbers of individuals by the two procedures (simulated individual BLUP and real BLUP) was of 0.9555. A total of 1,072 individuals were suggested for selection by BLUPIS within 69 families above the general mean. With real BLUP the best 1,072 individuals were selected in the same families. The effectiveness of BLUPIS was therefore validated. It is important to point out that the method should be applied

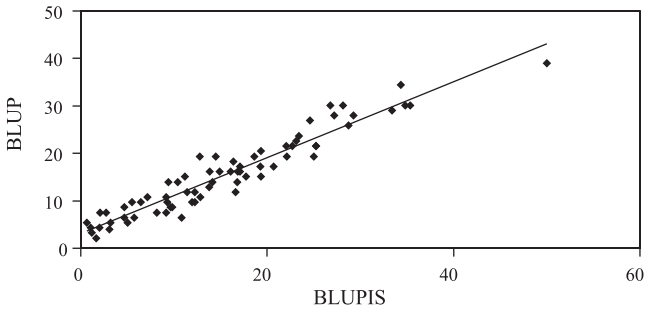

Figure 1. Number of individuals selected by the methodologies individual BLUP and BLUPIS, in 69 *Eucalyptus* full-sib families. The individual selection only took place in families presenting positive genotypic effect, from a total of 140 families evaluated.

using the predicted genotypic effects ($\hat{g}_j$), and not the genotypic values ($\hat{u} + \hat{g}_j$), and even less the phenotypic mean of each progeny. Besides being incorrect, the determination of n_k, based on the last two statistics, leads to the absence of selection among families, i.e., it leads to selection of individuals from all families, thus resembling mass selection.

BLUPIS is also suitable for all the annual self-pollinated species (such as soybean, rice, bean, wheat, oat, barley) and for Arabic coffee (perennial self-pollinated), which are evaluated at level of plot totals. In these species, the number of sister lines to be advanced, that is, the number of individuals to be selected and advanced within each line, can be determined by BLUPIS. In this case, n_j can be considered as 20, which represents 98% of Ne maximum of a family F3 or S1. Ne for families S1 is given by Ne = n/(n + 0,5).

The results of BLUPIS also indicated that the selection using information about families should be practiced, when individuals are being directed for clonal tests in species of vegetative propagation. In many of these species, such as rubber plant, cassava and several fruit species, the individuals go through mass selection in order to proceed to clonal tests. The use of individual BLUP or BLUPIS in these species will provide increase in clonal selection efficiency and reduction in the improvement program costs.

Conclusions

1. The simulated individual BLUP proposed presents high correlation (0.9555, in a validation study with eucalyptus), with true individual BLUP referring to the number of individuals to be selected per family.

2. BLUPIS indicated for the genetic improvement of species whose data recording at family level (total harvest of plot) is operationally easier than at individual level, being therefore suitable for improvement programs of sugarcane, forage and annual self-pollinated species, especially for low heritable traits.

3. BLUPIS has to be used with the predicted genotypic effects of families rather than with genotypic means.

4. BLUPIS determines the number of individuals to be selected per family, the total number of clones to be advanced, and the number of families to contribute with selected individuals.

5. BLUPIS allows the identification of which replication the best individuals of each family are in.

References

BARBOSA, M.H.P.; PINTO, C.A.B.P. Análise dialélica parcial entre cultivares de batata nacionais e introduzidas. **Pesquisa Agropecuária Brasileira**, v.33, p.307-320, 1998.

BRESSIANI, J.A. **Seleção seqüencial em cana-de-açúcar**. 2001. 134p. Tese (Doutorado) - Universidade de São Paulo, Piracicaba.

COX, M.; HOGARTH, D.M.; SMITH, G. Cane breeding and improvement. In: HOGARTH, D.M.; ALLSOPP, P.G. (Ed.). **Manual of cane growing**. Indooroopilly: BSES, 2000. p.91-108.

CHANG, Y.S.; MILLIGAN, S.B. Estimating the potential of sugarcane families to produce elite genotypes using univariate cross prediction methods. **Theoretical and Applied Genetics**, v.84, p.662-671, 1992.

FERREIRA, R.P.; PEREIRA, A.V. Melhoramento de forrageiras. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. Viçosa: UFV, 2005. p.781-812.

KIMBENG, C.A.; COX, M.C. Early generation selection of sugarcane families and clones in Australia: a review. **Journal of the American Society of Sugar Cane Technologists**, v.23, p.20-39, 2003.

MACHADO JUNIOR, G.R.; WALKER, D.I.; BRESSIANI, J.A.; SILVA, J.A.G. da. Emasculation of sugarcane tassels using hot water. In: INTERNATIONAL SOCIETY OF SUGAR CANE TECHNOLOGISTS CONGRESS, 22., 1995, Cali. **Proceedings**. Cartagena: ISSCT Tecnicaña, 1995. p.346-352.

McRAE, T.A.; ERQUIAGA, D.L.; JENSEN, L.F.; RATTEY, A.R.; STRINGER, J.K. BSES sugarcane breeding program in the Burdekin. In: AUSTRALIAN SOCIETY OF SUGAR CANE TECHNOLOGISTS, 1998, Brisbane. **Proceedings**. Watson Ferguson: ISSCT, 1998. p.196-203.

MARIOTTI, J.A.; CUENYA, M.I.; GARCÍA DE SALAS, M.B. Análisis de componentes familiares y intra-familiares en progenies derivadas de combinaciones biparentales de caña de azúcar (*Saccharum* spp.). **Revista Industrial y Agrícola de Tucumán**, v.76, p.55-60, 1999.

MATSUOKA, S.; GARCIA, A.A.F.; ARIZONO, H. Melhoramento da cana-de-açúcar. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. Viçosa: Ed. da UFV, 2005. p.225-274.

RESENDE, M.D.V. de. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002b. 975p.

RESENDE, M.D.V. de. **Software Selegen–REML/BLUP**. Curitiba: Embrapa Florestas, 2002a. 67p. (Embrapa Florestas. Documentos, 77).

RESENDE, M.D.V. de; HIGA, A.R. Maximização da eficiência da seleção em testes de progênies de *Eucalyptus* através da utilização de todos os efeitos do modelo matemático. **Boletim de Pesquisa Florestal**, v.28/29, p.37-55, 1994.

VENCOVSKY, R. Effective size of monoecious populations submitted to artificial selection. **Revista Brasileira de Genética**, v.1, p.181-191, 1978.

Received on November 11, 2004 and accepted on July 20, 2005