Selection between and within full-sib sugarcane families using the modified BLUPIS method (BLUPISM)

R.D. Castro¹, L.A. Peternelli¹, M.D.V. Resende², C.D. Marinho¹, P.M.A. Costa¹, M.H.P. Barbosa² and E.F.A. Moreira¹

¹Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, Brasil
²Embrapa Florestas, Colombo, PR, Brasil
³Departamento de Fitotecnia, Universidade Federal de Viçosa, Viçosa, MG, Brasil

Corresponding author: L.A. Peternelli
E-mail: peternelli@ufv.br

Genet. Mol. Res. 15 (1): gmr.15017334
Received July 29, 2015
Accepted October 6, 2015
Published January 8, 2016
DOI http://dx.doi.org/10.4238/gmr.15017334

ABSTRACT. The objective of this study was to assess the efficiency of a modification of the simulated individual best linear unbiased prediction (BLUPIS) procedure, which is used for the approximation of classic individuals (BLUPI) for selection between and within sugarcane families. A total of 110 full-sib families were employed in an experiment initiated in 2007 using a randomized block design with five replicates. The variable tons of stalks per hectare was measured from a plot containing 20 plants. The modified BLUPIS (BLUPISM) procedure showed a 0.98 correlation with BLUPI, thus demonstrating great efficiency in selecting individuals in sugarcane families during the initial phase of genetic breeding programs.

Key words: Sugarcane breeding; Mixed-models; BLUP; Family selection
INTRODUCTION

In the 2013-2014 harvest, sugarcane crops occupied an area of approximately 8.8 million hectares, and there was a total production of 658.8 million tons (CONAB, 2014). In total, 297.90 million tons of sugarcane were used for sugar production, which generated approximately 37.88 million tons of sugar. Additionally, 360.90 million tons of sugarcane were used for ethanol production, which generated 27.96 billion liters of ethanol.

To increase sugar and ethanol production without expanding croplands, cropping practices must be improved, and new and more productive varieties must be developed. The development of sugarcane varieties is considered the production component that can provide the highest financial return to producers.

The process of developing new varieties in sugarcane breeding involves the T1, T2, T3, and EP phases, which are the first, second, and third test phases and an experimental phase, respectively (Barbosa and Silveira, 2012). The clonal multiplication phase subsequently occurs in mills and distilleries participating in sugarcane-breeding programs.

A key stage in sugarcane (Saccharum spp) breeding corresponds to the initial phase (T1), during which thousands of heterozygous individuals resulting from crosses between previously selected parents are produced (Cesnik and Miocque, 2004). The individual selection of clones during the initial stage of breeding for the production of stalks is usually inefficient. However, when family productivity is measured, most of the variability among the mean values results from genetic effects, thus increasing the efficiency of selection between families. The individuals selected within these families are then cultivated in larger plots for clonal tests with replicates, and this leads to reduced environmental effects relative to genetic effects. Therefore, individual selection becomes more reliable (Stringer et al., 2011a).

Certain sugarcane-breeding programs have routinely practiced family selection prior to obtaining clones (Cox et al., 1996; Bressiani, 2001; Kimbeng and Cox, 2003; Stringer et al., 2011a; Barbosa et al., 2012), especially for traits that have higher family mean heritability compared to individual plant heritability (e.g., stalk production). When selection is practiced in families with high genotypic values, the likelihood of finding superior clones in their respective progenies is higher (Barbosa et al., 2005; Resende and Barbosa, 2005).

Resende (2002) showed that the optimal selection strategy in sugarcane is the prediction of genotypic values using the best linear unbiased predictor (BLUP) method at the individual level (BLUPI). This procedure simultaneously uses information for families and individuals within families for selection. However, this method is rarely used in breeding programs because of operational problems related to data collection at the plant level.

The initial breeding phase of sugarcane family selection is usually performed in accordance with all of the data corresponding to the plots, because they are collected as a whole rather than in accordance with the individuals of a family within each plot. To help select individuals within a plot based on all of the plot data, Resende and Barbosa (2006) proposed selection by simulated individual BLUP (BLUPIS), which is based on the genotypic effects of families tested in the field.

An advantage of this method includes the lack of requirement of the individual assessment of experimental plants. Furthermore, this method optimizes selection by indicating the number of individuals to be selected per family, the total number of clones to be evaluated, and the number of families involved with the individuals selected (Resende and Barbosa, 2006).

The BLUPIS method automatically eliminates families that have negative genotypic effects.
that are below the experimental overall average, which is reasonable when considering the extremely low likelihood of obtaining a superior clone in these families (Resende and Barbosa, 2005).

According to Stringer et al. (2011a), the Australian sugarcane genetic-breeding program tests approximately 1200 families annually. Presumably, a higher number of tested families will provide a higher likelihood of finding promising clones. The plant-cane selection method must be efficient and easily applied to increase the number of families tested in a sugarcane genetic-breeding program. BLUPIS is an alternative method of testing a large number of families, which is highly efficient (Oliveira et al., 2011) in producing correlations between the numbers of individuals selected per family relative to both methods (BLUPIS and BLUPI). However, BLUPIS tends to select far more individuals than BLUPI, and the number of individuals selected per family is different between the two methods. Moreover, differential selection of certain individuals also occurs with the two methods, and individuals in 50% of families are selected without excluding median families. To overcome the drawbacks of the BLUPIS procedure, the present study experimentally evaluated a method referred to as modified BLUPIS (BLUPISM) to further improve selection between and within sugarcane families relative to BLUPIS (Resende and Barbosa, 2006), a method currently practiced in certain programs.

The objective of this study was to describe the BLUPISM procedure and to show results that demonstrate its higher efficiency in selections between and within sugarcane families compared with the BLUPI procedure.

MATERIAL AND METHODS

Plant material and phenotypic assessment

The study used 110 full-sib families from crosses performed at Serra do Ouro Experimental Station of the Federal University of Alagoas in the municipality of Murici, Alagoas State, Brazil, in 2006. The crosses were controlled, named biparental, or single crosses. The seeds resulting from these crosses were collected and packed in suitable containers, and were subsequently sent to the Sugarcane Research and Breeding Center (CECA) of the Federal University of Viçosa. CECA is in the municipality of Oratórios, Minas Gerais State, Brazil, which is located at 20°25’S latitude and 42°48’W longitude. The site is at an altitude of 494 m, and it has dark red Latosol soil. The seeds were sown in trays with a substrate to produce seedlings, and were placed in a greenhouse for seedling germination and growth. The seedlings were subsequently subjected to 120 days of 50% light and 30 days without shading for acclimatization. The seedlings resulting from the respective two-parent crosses were sent to an experimental area of the CECA, where the family experiments were initiated after acclimatization.

The experiment was initiated in May 2007 using a randomized block design with five replicates. Each plot consisted of 20 plants distributed in two 5-m long furrows that were spaced 1.40 m apart. The plant-cane data were collected in June 2008, and the ratoon cane data were collected in June 2009.

In 2008, the total plot mass (TPM; expressed in kg) was determined by weighing all of the stalks in the plot using a dynamometer. Stalk productivity was expressed as tons of stalks per hectare (TSH), and it was calculated using the following equation:

\[ \text{TSH} = \frac{\text{TPM}}{\text{PA}} \times 10 \]  

(Equation 1)
where TPM is the total plot mass expressed in kg, and PA is the plot area expressed in m² (the plot was 14 m²). The constant corresponds to the conversion factor between units of measurement.

In 2009, the clump mass of the plot was determined at the ratoon cane stage by cutting all of the clumps and weighing them using a dynamometer. The TPM of each individual was determined using the following equation:

\[
TPM = \frac{MP \times PA}{IPA}
\]  
(Equation 2)

where IPA is the area occupied by an individual (expressed as m²). The IPA value was 0.7 m² in the present study, and TSH was determined as described above.

Data analysis procedures

The data were analyzed using restricted maximum likelihood/best linear unbiased prediction (REML/BLUP) mixed models. REML was used to estimate the components of variance, and BLUP was used to predict the genotypic values.

The genotypic values of each family were obtained through experiments that used a complete block design, and assessments were performed at the plot level. The statistical model was denoted as follows:

\[
Y = Xr + Zg + e
\]  
(Equation 3)

\[
g \sim N(0, \sigma_g^2); e \sim N(0, \sigma_e^2)
\]  
(Equation 4)

where \(y\) is the data vector; \(r\) is the repetition effect vector (assumed to be fixed) added to the overall mean; \(g\) represents the phenotypic effects (assumed to be random), vector \(e\) represents the random errors or residuals; and \(X\) and \(Z\) represent the incidence matrices for \(r\) and \(g\), respectively.

The conventional BLUPIS procedure estimates the number of individuals (\(n\)) to be selected from each family (\(k\)) based on the following equation from Resende and Barbosa (2006):

\[
\hat{n}_k = \left(\frac{\hat{g}_k}{\hat{g}}\right)\hat{n}_{j=1}
\]  
(Equation 5)

where \(\hat{g}_k\) corresponds to the genotypic value associated with family \(k\), and the subscript \(j = 1\) represents the family with the highest genotypic value. An improvement to BLUPIS, which was designated as modified BLUPIS or BLUPISM, considers the estimated genotypic value of the family (GV) rather than its genotypic effect \(g\), and it includes the term \(n_{k-1}\) instead of \(n_{j=1}\). Thus, when using BLUPISM:

\[
\hat{n}_k = (GV_k/GV_{j=1})n_{k-1}
\]  
(Equation 6)

\[
GV_k = \text{overall mean } + \hat{g}_k
\]  
(Equation 7)
where \( n_{k-1} \) refers to the number of individuals selected in the family immediately prior to the ranking, and \( n_{j=1} \) indicates the total number of individuals selected in the best family.

The following steps were used to conduct the BLUPISM analysis: i) identify superior families in plant-cane after harvesting the entire plot, and assess the GV associated with each family; ii) use BLUPISM to determine the number of individuals selected per family; iii) measure the mass of the ratoon cane and individual plant-canues exclusively within families with one or more individuals indicated by BLUPISM (approximately 30% of the families), and assess the remaining families by harvesting the entire plot and calculating the mean per plant in the plot; and iv) perform the individual BLUP predictions in the selected families, and rank the pool of families. The genotypic values of each individual \( i \) may be calculated using the following equation:

\[
\hat{g}_i = \hat{g}_{fam} + \hat{h}_{a,w}^2 \hat{e}_i
\]  
(Equation 8)

where \( \hat{g}_{fam} \) refers to the genotypic effects of the family in the simultaneous analyses of both harvests;

\[
\hat{e}_i = y_r - \hat{y}_r = y_r - (X\hat{b} + Z\hat{g}_{fam} + W\hat{p})
\]  
(Equation 9)

represents the residual associated with individual observations of the ratoon cane;

\[
\hat{h}_{a,w}^2 = \frac{\text{Var}(\hat{g}_{fam})}{\text{Var}(\hat{e}_i)}
\]  
(Equation 10)

represents the heritability within the progeny of the ratoon cane; \( y_r \) represents the individual measurements performed for the ratoon cane; and

\[
\hat{y}_p = X\hat{b} + Z\hat{g}_{fam} + W\hat{p}
\]  
(Equation 11)

represents individual phenotypes predicted by fitting the model in the simultaneous analyses of plant-cane and ratoon cane.

The statistical analyses were performed using the SELEGEN-REML/BLUP software (Resende, 2007). A comparison between BLUPISM and BLUPI was performed using Pearson’s correlation coefficient.

RESULTS

Table 1 outlines the estimates of variance components and the genetic parameters of the variable TSH analyzed at the plot level in 2008, and the results of the analysis with individual plants in 2009 are shown. The selective accuracy values (Table 1) for TSH were 0.90, which indicated high experimental accuracy according to the criteria defined by Resende and Duarte (2007).

The estimate of the mean broad-sense heritability of the family (\( h_{mf}^2 \)) was 0.82 for TSH. However, selecting individuals within a plot could be inefficient, because the additive heritability within a plot (\( h_{ad}^2 \)) was low (0.03; Table 1). The individual narrow-sense heritability (\( h_{a}^2 \)) of the TSH trait was 0.05, which was much lower than the mean broad-sense heritability of the families (\( h_{mf}^2 \)).
These values show that for TSH, the individual selection (mass selection) would be less efficient than family selection.

Table 2 outlines the values of genotypic effects of each family, the number of individuals to be selected within each family using the BLUPISM selection method in the 2008 harvest, and the number of individuals selected by BLUPI in the 2009 harvest for the TSH trait. The selection of individuals using the BLUPI procedure was performed by ranking the genotypic values of all 7345 assessed individuals and selecting the best 1596 as indicated by the BLUPIS method (Resende and Barbosa, 2006). A total of 29 families were selected, and they represented a selection rate of approximately 27% of the assessed families.

Theoretically, the likelihood of finding promising genotypes for TSH within the 27% best families was high. Of the 7345 assessed individuals, 370 were selected by BLUPISM (Barbosa et al., 2012). This value corresponds to a selection rate of approximately 5%.

Family 67 will be used to illustrate the calculations made to generate the data shown in Table 2. The selection of 38 individuals from family 67 was determined according to the BLUPISM method by dividing the value 105.04 by 110.65, and the result was then multiplied by the number of individuals selected in the previous family (40 individuals).

<table>
<thead>
<tr>
<th>Parameters(^1)</th>
<th>TSH (t/ha)</th>
<th>Parameters(^2)</th>
<th>TSH (t/ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma^2_e)</td>
<td>282.1307</td>
<td>(\sigma^2_g)</td>
<td>514.49</td>
</tr>
<tr>
<td>(\sigma^2_e)</td>
<td>306.5197</td>
<td>(\sigma^2_{plot})</td>
<td>42.07</td>
</tr>
<tr>
<td>(\sigma^2_f)</td>
<td>588.6504</td>
<td>(\sigma^2_{within})</td>
<td>19227.02</td>
</tr>
<tr>
<td>(h^2_g)</td>
<td>0.4792 ± 0.083</td>
<td>(\sigma^2_f)</td>
<td>19783.59</td>
</tr>
<tr>
<td>(h^2_{mf})</td>
<td>0.8214</td>
<td>(h^2_m)</td>
<td>0.72</td>
</tr>
<tr>
<td>Acfam</td>
<td>0.9063</td>
<td>(h^2_a)</td>
<td>0.052 ± 0.01</td>
</tr>
<tr>
<td>CVgi%</td>
<td>21.8591</td>
<td>CVe%</td>
<td>22.7844</td>
</tr>
<tr>
<td>CVr</td>
<td>0.9593</td>
<td>Overall mean</td>
<td>76.8407</td>
</tr>
<tr>
<td>Overall mean at the plant level</td>
<td>162.639</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)Genotypic variance (\(\sigma^2_g\)); residual variance (\(\sigma^2_e\)); individual phenotypic variance (\(\sigma^2_f\)); broad-sense heritability of individual plots (\(h^2_g\)); mean family heritability (\(h^2_{mf}\)); accuracy of family selection (Acfam); coefficient of genotypic variation (CVgi\%); coefficient of environmental variation (CVe\%); and coefficient of relative variation (CVr). \(^2\)Genotypic variance between full-sib progenies (\(\sigma^2_g\)); environmental variance between plots (\(\sigma^2_{plot}\)); residual variance within the plot (\(\sigma^2_{within}\)); individual phenotypic variance (\(\sigma^2_f\)); narrow-sense individual heritability (\(h^2_a\)); and additive heritability within the plot (\(h^2_m\)).

The five families with the highest genotypic effect values were 61, 67, 38, 28, and 25. Family 67 had the second highest genotypic effect value (28.20) in 2008, and the BLUPISM method classified 38 individuals. The BLUPI method classified 39 individuals from this family in 2009, which demonstrates the accuracy of the method. The BLUPISM method exhibited a 0.98 correlation with the BLUPI method, and the BLUPI and BLUPISM methods indicated that 750 to 800 clones from
the experiments should undergo clonal testing. A significantly higher number was indicated by the classic BLUPIS method that selected 1596 individuals and exhibited a correlation of only 0.85 between the methods (results not shown).

The number of individuals selected based on the genotypic value of the individual using the BLUPI (nBLUPI) method in 2009 for the tons of stalks per hectare (TSH) trait is also shown, and the values are based on the numbers of individuals selected in the best family (n_j = 40).

<table>
<thead>
<tr>
<th>Family</th>
<th>( \hat{g}_j )</th>
<th>( \mu + \hat{g}_j )</th>
<th>nBLUPISM</th>
<th>nBLUPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>33.8121</td>
<td>110.8528</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>67</td>
<td>28.2013</td>
<td>105.0420</td>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td>38</td>
<td>26.3037</td>
<td>103.1444</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>28</td>
<td>24.6590</td>
<td>101.4998</td>
<td>32</td>
<td>34</td>
</tr>
<tr>
<td>25</td>
<td>24.6574</td>
<td>101.4981</td>
<td>30</td>
<td>41</td>
</tr>
<tr>
<td>117</td>
<td>22.1469</td>
<td>98.9876</td>
<td>27</td>
<td>23</td>
</tr>
<tr>
<td>12</td>
<td>21.8906</td>
<td>98.7313</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>39</td>
<td>21.0215</td>
<td>97.8622</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>20.0570</td>
<td>96.8977</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>113</td>
<td>20.0127</td>
<td>96.8534</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>26</td>
<td>19.8697</td>
<td>96.7104</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>27</td>
<td>18.2497</td>
<td>95.0904</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>106</td>
<td>17.7798</td>
<td>94.6206</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>17.6878</td>
<td>94.5385</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>66</td>
<td>17.3822</td>
<td>94.2230</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>70</td>
<td>17.2656</td>
<td>94.1003</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>90</td>
<td>17.1013</td>
<td>93.9420</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>22</td>
<td>16.2667</td>
<td>93.7074</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>54</td>
<td>15.9496</td>
<td>92.7903</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>45</td>
<td>15.1774</td>
<td>92.0181</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>47</td>
<td>15.0344</td>
<td>91.8751</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>60</td>
<td>13.2748</td>
<td>90.1155</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>111</td>
<td>12.8065</td>
<td>89.6472</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>43</td>
<td>12.4763</td>
<td>89.3170</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>56</td>
<td>12.4303</td>
<td>89.2710</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>75</td>
<td>12.2413</td>
<td>89.0620</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>55</td>
<td>12.1263</td>
<td>88.9670</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>42</td>
<td>12.1230</td>
<td>88.9637</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>112</td>
<td>11.8667</td>
<td>88.7074</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

The number of individuals selected based on the genotypic value of the individual using the BLUPI (nBLUPI) method in 2009 for the tons of stalks per hectare (TSH) trait is also shown, and the values are based on the numbers of individuals selected in the best family (n_j = 40).

**DISCUSSION**

Coefficient of genetic variation (CV_g) values above 10% suggested the presence of variability with potential selection (Oliveira et al., 2005, 2008). In this study, CV_g = 21.85, which indicated the presence of genetic variability and potential selection for TSH.

Accuracy refers to the correlation between the true genetic value of the genetic treatment and the genetic value estimated or predicted based on experimental data. Accuracy values close to one (or 100%) are adequate, and smaller absolute deviations between the parametric genetic values and estimated or predicted genetic values indicate higher accuracy (Resende and Duarte, 2007).

The mean broad-sense heritability of the family estimate favors the selection of promising families. Pedrozo et al. (2011) also obtained high estimates of broad-sense heritability for TSH when assessing sugarcane families. According to Stringer et al. (2011a), most of the variability between families was due to genetic effects, and selection may occur between these families.
Bressiani (2001) and Barbosa et al. (2004) reported the importance of using family selection in sugarcane. These authors argued that the probability of finding promising individuals for traits with low individual heritability was higher in the best families at the early stages of genetic breeding.

According to Cox et al. (1996) and Kimbeng et al. (2000), family selection should be focused on the top 30 to 40% of families because they have the highest percentage of elite clones, which corroborates the results found in the present study.

When assessing the efficiency of the BLUPIS procedure for sugarcane in plant-cane and ratoon cane, Silva et al. (2015) obtained 16 and 32% selection rates when using 50 and 100 individuals selected from the best families ( ), respectively.

According to Shanthi et al. (2008), family selection is useful for identifying families that most likely have the highest ratio of elite clones. It can also be used to select superior clones, to keep or discard parents, and to plan future crosses. A method designed to utilize the potential of the best families includes cultivating a higher number of seedlings from these families and performing individual selection for TSH within the family.

The results of the BLUPISM method were highly correlated with those of the BLUPI method, demonstrating accuracy for selection in plant-cane-breeding programs. The high repeatability between assessments in plant-cane and ratoon cane (Pedrozo et al., 2011) validates this result.

BLUPI selects all of the individuals in the top 30% of the best families, and it tends to select almost all individuals from the best families. BLUPISM almost exactly reproduced BLUPI results, resulting in a 0.98 correlation between the two procedures. For instance, the total number (N) of individuals selected, the number of individuals selected per family (error smaller than four individuals per family), ~30% of the best families selected and individuals selected were the same by using the BLUPISM and BLUPI methods (Table 2). These results demonstrate the efficiency of BLUPISM.

The BLUPISM method is an important contribution in terms of the development of analytical tools that lead to more efficient sugarcane breeding. BLUPISM faithfully reproduced the classic individual BLUP (BLUPI), and it only required an assessment of individual plants from 30% of the best families. Other important contributions to sugarcane breeding were recently reported, including the use of more complete pedigree information (Atkin et al., 2009), models for simultaneous analyses of repeated measures in experiments at multiple locations (Smith et al., 2007), and simultaneous modeling of intergenotypic competition and spatial variability (Stringer et al., 2011b). All of these methods provide the necessary framework to maximize the efficiency of sugarcane production systems.

The BLUPISM method results are consistent with the results from the optimal selection procedure (BLUPI), which is based on selecting individuals at the plant level in the field. BLUPISM selected the same families and indicated nearly the same numbers of individuals for selection within the best families, which demonstrated similar accuracy to that of the classic BLUP method for selection in plant-cane crops. We recommend the use of BLUPISM for selection between and within sugarcane families, because the results were highly consistent with those from the optimal selection procedure (BLUPI). Moreover, the method does not require an extensive collection of field data.

Conflicts of interest

The authors declare no conflict of interest.
ACKNOWLEDGMENTS

This study obtained financial support from the National Council of Technological and Scientific Development (CNPq), the Research Support Agency of Minas Gerais (FAPEMIG), and the Sugarcane Genetic Breeding Program (PMGCA/RIDESA/UFV).

REFERENCES


