



ARTICLE

Breeding self-pollinated plants

Magno Antônio Patto Ramalho^{1*} and Lilian Cristina Andrade de Araújo¹

Received 4 April 2011

Accepted 19 May 2011

ABSTRACT - *This presentation aims to discuss strategies to increase the efficiency of selection in self-pollinated plants. One of the strategies is the use of mixed models in order to appropriate genealogy information on the method of pedigree and through the sequential analysis in evaluation tests of progeny in several environments/generation. It is also discussed the use of multiline as an alternative to extend the life of a cultivar.*

Key words: *plant breeding, quantitative genetics, multiline, use of mixed models.*

INTRODUCTION

The results of genetic breeding have been excellent for most self-pollinated species under improvement in Brazil. The evidence is the gain in productivity over the past 30 years. The mean gain of some of the most important species in the country was 4.85 % per year. Assuming that at least 50 % of this progress must be attributed to genetic improvement, the achievement of genetic gain is easily demonstrated. It should be emphasized that the evaluation of the breeding success should not only be based on the yield increase, but, mainly on the improved sustainability of self-pollinated species grown in Brazil. This increase in sustainability can be measured by the availability of cultivars adapted to new growing conditions, more tolerant to biotic and abiotic stresses.

Among the self-pollinated species there are some as soybean that attract the attention of large multinational seed companies. The case of tomato is also very peculiar, because attention is focused on the production of hybrid seeds and although there are public institutions dedicated to tomato improvement, most of the seed stock is produced by multinational companies. Contrastingly, in the case of

some species such as common bean, the development of new cultivars is controlled by the public sector. In essence, the improvement process in different species is rather similar, but there are differences in the resources invested in breeding, especially in terms of the crosses number, progenies evaluated annually and the number of test sites of new lines.

The commonly used breeding methods of self-pollinated plants were developed over a century ago in Europe. No great subsequent changes were made. Basically, the selection procedure of the best progenies was optimized, particularly with the development of agricultural experimentation based on the publications of Fisher. In the past 25 years, the techniques of molecular biology opened up the prospect of the possibility of direct selection of a genotype. The emphasis of global research was focused mainly on the use of molecular DNA markers. Abundant basic information was compiled, but evidently the phenotypic analysis in the field, as accurately as possible, is still indispensable.

Information was accumulated that indicated that the traits the breeders work on are mostly complex, in other words, probably controlled by several genes. This

¹ Universidade Federal de Lavras (UFLA), Departamento de Biologia, C.P. 37, 37.200-000, Lavras, MG, Brazil. *E-mail: magnoapr@ufla.com

observation became even patent because the improvement consists, in summary, of an accumulation of advantages through successive selection cycles, recurrent selection, that involves intercrosses of the best lines/cultivars available in each cycle. The objective is therefore to identify the plants/progenies with the highest genotypic merit in each selection cycle. Another goal is to develop cultivars that can be used for several years.

Numerous studies have focused on improving breeding efficiency, especially in Brazil. This presentation will comment some of these studies targeting an increased breeding efficiency of self-pollinated plants by the identification of the progenies/lines with highest genotypic merit and comment about strategies to obtain cultivars with a longer lifetime.

USE OF MIXED MODELS TO IMPROVE SELECTION EFFICIENCY

The mixed model methodology was proposed by the American statistician Charles Roy Henderson in 1949, and has since been routinely used, especially in animal breeding (Henderson 1984). In plant breeding, this statistical technique only became known in the late twentieth century. A series of advantages are attributed to its use (Resende 2002, Piepho et al. 2008). Some of the applications will be discussed in this section.

Use of mixed models in the pedigree method

One of the advantages of the pedigree method is the possibility to record the genealogy of all plants/progenies. Although this advantage is often cited, its usefulness is limited. One possibility of exploiting information on the genealogy was pointed out by Nunes et al. (2008). For this purpose, a genetic model without epistasis was considered and one trait controlled by 20 segregating genes with independent distribution. Equal gene effects and an allelic frequency of 0.5 were assumed. In the simulation scenarios, different heritability (h^2) values were considered for selection based on the mean of F_4 progenies. For each heritability, 1000 segregating populations with the pedigree method were simulated, with records of the genealogy.

The pedigree method is based on a segregating F_2 population consisting of 64 plants, which resulted in 64 $F_{2:3}$ progenies, each with 40 plants. From each $F_{2:3}$ progeny, two plants were randomly taken, resulting in 128 $F_{3:4}$ progenies. The process was repeated in the following generation in order to obtain 256 $F_{4:5}$ progenies, each with 40 plants.

The phenotypic data from 256 $F_{4:5}$ progenies of each population were then analyzed in a completely randomized design, with two replications and plots of 20 plants. The mixed model approach was used for this analysis, considering the following models: GA model, which includes information on parentage among the families, and the GI model, in which this information of genetic relationships is ignored. From the above models, the predictions BLUP (Best Linear Unbiased Prediction) with (BLUPA) and without parentage (BLUPI) were determined. It is commonly comment that BLUP-predicted breeding values are free from estimation errors. The simulation showed that the correlation between the predicted additive genetic values (AGV) and the true AGV varied according to the h^2 of the trait, i.e., the greater the h^2 , the higher the correlation (Figure 1). It is understood, however, that the predictor considering the genealogy (BLUPA) was, independent from h^2 , which best predicted the true AGV, i.e., a higher correlation between BLUPA and AGV. The difference between the correlations of predictors BLUPA and BLUPI with AGV were more expressive at lower h^2 (Figure 2). On the other hand, this result shows that at heritability values of more than 50 %, the genetic values associated with the progenies are already well-determined by phenotypic means, which is equivalent to the BLUPI in terms of selection in balancing condition (Duarte and Vencovsky 2001).

In general, the simulation showed that the inclusion

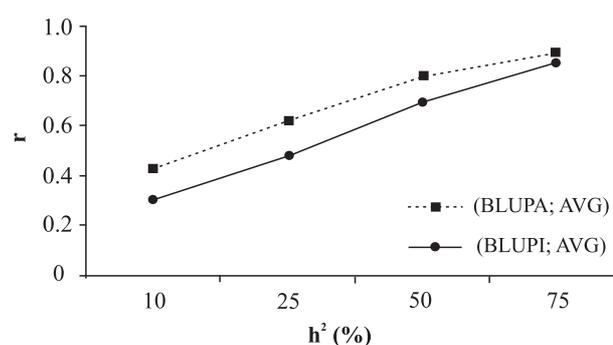


Figure 1. Mean values of the Spearman correlations between the additive genetic values (AGV) and BLUP predictions/estimates considering (BLUPA) and ignoring (BLUPI) AGV parentage between progenies $F_{4:5}$, carry by the genealogical/pedigree method, for different heritability values in the progeny means (h^2). Source: Adapted from Nunes et al. (2008).

of genetic similarity among progenies in BLUPA resulted in higher gain with selection (GS) than with BLUPI for all pre-established heritability values and proportion of selection (Figure 2). The relative GS increased with the

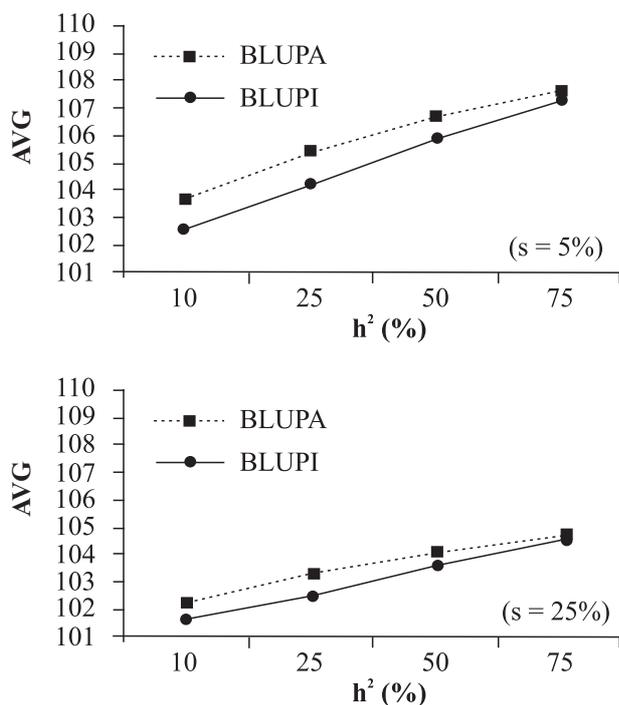


Figure 2. Additive genetic values (AGV) progeny means $F_{4,5}$ selected by BLUP considering (BLUPA) and ignoring (BLUPI) a information on additive parentage due to the broad-sense heritability in the progeny means (h^2) and selected proportions (s) of 5 % and 25 %. Source: Adapted from Nunes et al. (2008).

reduction in heritability and selection intensity, i.e., the weight of relatedness was more pronounced at lower heritability. Thus it is evident that the effort of recording the genealogy of the segregating population should be compensated by using this information in the selection of progenies in the application of mixed models, which can result in additional gain from selection.

Sequential analysis of the evaluation progeny tests

A recurrent selection program usually begins with a large number of $F_{2,3}$ or $S_{0,1}$ progenies that decreases in every generation. At the end of the process, five to ten plants are selected to participate in a competition experiment of lines for two to six seasons, in various environments, where lines are then identified to participate in tests of VCU (Value for Cultivation and Use). In this situation there are at least two stages of selection, i.e., of the segregating $S_{0,1}$, $S_{0,2}$, $S_{0,3}$, $S_{0,4}$, $S_{0,5}$ progenies. Often only the generation of reference is considered, i.e., in the selection of $S_{0,4}$, for example, the performance of the progenies in the previous generation is not considered.

In the analyses of a decreasing number of progenies, the sequential analysis, which includes all progenies evaluated by BLUP, is suggested as a good alternative (Piepho and Möhring 2006). To quantify the improvement in the process efficiency through sequential analysis, was used an evaluation experiment of progenies of two cycles of a recurrent selection program, of the Universidade Federal de Lavras (UFLA) since 1990 (Ramalho et al. 2005). The $S_{0,1}$ to $S_{0,10}$ progenies were evaluated in these experiments. The number of progenies ranged from 322 in $S_{0,1}$ to 10 in $S_{0,10}$. Two controls were used in all experiments.

Firstly, the analysis of variance per generation was performed. The subsequent analyses were based on the sequential procedure, recovering the information from the evaluated generations. For this purpose, three analysis strategies were adopted: a) sequential analysis considering a combination of the generations $S_{0,1}$ to $S_{0,3}$, $S_{0,1}$ to $S_{0,4}$ and $S_{0,1}$ to $S_{0,5}$, using all progenies; b) sequential method comprising the 23 best progenies in common in the $S_{0,1}$ to $S_{0,5}$ generations; c) sequential analysis involving the 10 progenies of the $S_{0,1}$ - $S_{0,10}$ evaluations. Analyses were performed using Proc Mixed of SAS (SAS 2000) considering the progeny effect as random (BLUP) and fixed (BLUE). To quantify the efficiency of the analysis procedures the coincidence of progenies selected by the strategies was contrasted with the standard adopted by the breeding program of the institution. The correspondence in the ranking of the progenies in the different analysis strategies was compared as well.

The coincidence between the progenies selected by the traditional method, based only on the mean of the last generation evaluated, was compared with the BLUP considering all generations from $S_{0,1}$, called BLUPT in this study (Table 1). For the 119 progenies selected in $S_{0,2}$, selected by the mean, coincidence by BLUPT would be 84 in the $S_{0,1}$ and $S_{0,2}$ generations. In the same cycle, now considering the 62 progenies selected in $S_{0,3}$, the coincidence by BLUPT ($S_{0,1}$, $S_{0,2}$ and $S_{0,3}$ generations) would be 39 progenies. The data in Table 2 clearly show that there was little correspondence between the sequential analysis involving all progenies, compared to the traditional procedure.

The percentage of this discrepancy in the progeny classification decreases as the number of generations involved increases. In the traditional process, the progenies \times generations (seasons) interaction is not considered because only environments or locations within a same generation are considered. Part of this no coincidence

Table 1. Number of selected progenies in the reference generation by the mean and that would coincide considering sequential analysis by BLUPT

Reference generation	Number of selected progenies in the mean of the reference generation	Number of progenies identified by BLUPT concerning mean
S _{0:2} ^{1/}	119	84 (71%) ^{2/}
S _{0:3}	62	39 (63%)
S _{0:4}	23	9 (40%)

¹ BLUPT analysis considering from S_{0:1} to the reference generation;

² Percentage of coincidence in relation to the number of selected progenies in the reference generation.

should therefore be attributed to the progenies x generations interaction.

In the S_{0:5} generation, 23 progenies were evaluated that originated the 10 that were evaluated up to S_{0:10}. Using some analysis strategies the coincidence of the progeny classification based on the standard procedure was compared with the BLUP and BLUE analyses involving only the 23 progenies, but including information from S_{0:1}. It was stated that the coincidence between these two procedures was very high. This shows that the influence of considering a fixed or random progeny effect on the classification was rather low (Table 2). The same observation was reported in a study of Piepho and Möhring (2006), in which the authors identified a correlation between the results of analysis with BLUP and BLUE in all cases ranging from $r = 0.994$ to $r = 1.0$.

A breeding program should be focused mostly on the identification of the best line(s) in the final evaluations.

Table 2. Ranking of progenies evaluated considering the reference generation (mean), analysis com information recovery for random effect involving the progenies in common of the generations (BLUP), analysis with information recovery for random effect involving all progenies in all generations (BLUPT) and analysis with information recovery for fixed effect (BLUE) for the 10 progenies evaluated in the elite assays

Progenies	Mean	Ranking		
		BLUE	BLUP	BLUPT
M-85-17	1	5	5	9
M-55-14	2	3	3	4
M-55-3	3	1	1	1
M-39-18	4	6	6	6
M-85-1	5	4	4	-
M-215-10	6	2	2	3
M-119-4	7	9	9	-
M-85-11	8	7	7	-
M-39-24	9	8	8	-
M-45-5	10	10	10	-

Ten lines were evaluated for five generations (seasons), totaling 13 environments. If one considers the selection of the five best lines for VCU, the correlation was four of five, or 80 %. Interestingly, even when only 10 progenies were considered and in the mean all 13 environments, the coincidence with BLUP and BLUE was greater, but still there was a loss of efficiency.

When considering BLUPT, two of the five best progenies coincided. It should be emphasized, however, that the progenies M-55-14 and M-55-3, classified on average as second and third, would be covered in all procedures (Table 2).

It was concluded that in the traditional analysis, based on the means only, efficiency is reduced. This loss is more pronounced in the early generations up to S_{0:5}, since the progeny x crop interaction is not taken into account. In the later generations, that is S_{0:6} up to S_{0:8} or S_{0:10}, when the number of progenies is lower and the analyses are based on all evaluated generations, the efficiency loss is lower. The use of any statistical analysis procedures, with the computer facilities available nowadays, does not entail any additional onus. It is therefore recommended that, in the future, the sequential analysis should be used in recurrent selection programs due to its greater selection accuracy, as suggested by Resende (2002).

Comparison of BLUP with selection among and within progenies and mass selection to select plants of a segregating population

There are some alternatives for selection in a segregating population of self-pollinated plants, including the Bulk method and Bulk within F₂ progenies. By these methods, the populations are advanced for a few generations, usually to F₄ or F₅, when plants are selected and their progenies derived are evaluated more detailed in replicated experiments.

There are several options that can be used to identify the plants/progenies in the subsequent selection process, for example, mass selection and selection among and within families, the latter when using progenies from F₂ plants. Another not much explored possibility in common bean breeding is the BLUP procedure. Mendes et al. (2011) compared alternatives to identify the best plants in the populations by the method Bulk within F₂ progenies using the population of cross CVIII8511 x RP-26. Of this population, 51 F_{2:3} progenies were obtained and evaluated in the F_{2:4} generation in a randomized block design, with 20 replications and one plant per plot. Yield data per plant

and the grade of architecture on a 1 to 9 scale were obtained. The data for each trait were standardized, resulting in a Z-index, and subsequently, the ΣZ , underlying the simultaneous evaluation of the two traits. For this analysis, the software Selegen – Computer Program for Genetic Selection (Resende 2002), model 59 was used, by which individual genotypic values can be calculated. The 100 best and 100 worst plants were identified, leading to the $F_{4:5}$ progenies. The $F_{2:4}$ generation was also analyzed by the least squares (LS) method, in a randomized block design with 20 replications and one plant per plot. With data from the individual genotypic values, obtained by BLUP, the following strategies were compared for selection in the $F_{2:4}$ generation: a) Mass selection: the 100 best and worst plants were selected in the $F_{2:4}$ generation as a function of ΣZ , regardless of the progeny or replication to which they belonged; b) Stratified mass selection: the plants were divided into strata and each stratum was considered a replication. Thus, each stratum contained one plant of each progeny, with a total of 51 plants per stratum. The five best and worst plants of each replication were selected as a function of ΣZ , resulting in 100 progenies from each group in the 20 replications; c) Selection among and within progenies: from the analysis of variance using the least squares (LS) method, the means of 51 progenies and the six best and worst, respectively, were selected (selection intensity among progenies of 11.7 %) and within these, the 16 best of the 20 existing plants. The same was done for the group of the worst, totaling 96 plants of each group.

The individual BLUP analysis provides an estimate that involves all model variables, for example, the merit of a progeny, the individual in the progeny, and even the replication it belongs to (Resende 2002). It was therefore expected that the selection among and within progenies would largely coincide with the BLUP for plants to be selected, which was actually confirmed. Taking as reference the ΣZ , the coincidence was higher than 80 % (Table 3). When comparing the coincidence of the plants that would be identified by mass or stratified mass selection with those selected by BLUP, the proportion was lower than 40 %.

The results indicate that mass selection was less efficient than BLUP. However, it has been argued that selection by different strategies does not necessarily identify the same progeny/individual, but rather similar individual in terms of performance. In this condition, the efficiency of mass selection strategies would be underestimated. To demonstrate this, the differential of selection (ds) index of the different strategies was

Table 3. Coincidence (in %) of the plants selected in the $F_{2:4}$ generation by different selection strategies compared to BLUP considering ΣZ

Strategy	best	worst	ds ^{1/}
Mass	44	36	3.63
Stratified mass	42	39	3.50
Among and within progenies	83	84	3.14
BLUP	100	100	2.60

Source: Mendes et al. (2011).

¹ differential of selection.

estimated (Table 3). It was observed that the 'ds' was higher in the strategy of mass selection. It could be argued, on this basis, that mass selection was more efficient. The shrinkage effect of BLUP should however be emphasized, which reduces the extreme values and consequently the ds estimate. Moreover, to calculate the gain with selection (GS), the 'ds' must be multiplied by the heritability (h^2). In the case of the BLUP, $h^2 = 1.0$, so, the estimated GS would be the ds estimate. In the selection among and within families on the other hand, the h^2 among and h^2 within should be weighted. Especially the latter should be lower. In the case of mass selection, $GS = ds \times h_1^2$, h_1^2 is the heritability for selection at the plant level, a value that is usually not high (Moreto et al. 2007). It can be inferred that the gain expected from mass selection and BLUP would be similar if h^2 was 71.6 %, i.e., $2.6 / 3.63 = 0.7162$. Although this h_1^2 estimate was not obtained, it would be difficult reach this value. However, the efficiency of mass selection should not be assumed as only 40 % of BLUP in view of the coincidence of the selected plants.

Multilines as strategy to increase durability of a cultivar

Most pathogens that infect cultivated plants have several races and new ones often appear. Therefore, the cultivars with resistance to pathogens quickly become susceptible and eventually the cultivar is withdrawn from the market. An option for a more durable resistance is the pyramidization of genes. But although rather time-consuming, pyramidization is no guarantee against the appear of a new race that can break the resistance. Another option, widely acclaimed in the past, are multilines. To evaluate the effectiveness of multilines, Botelho et al. (2011) evaluated seven lines of carioca bean separately and in combination or in a multiline. The lines had different resistance alleles to *Colletotrichum lindemuthianum*, the causal agent of anthracnose (Table 4). The experiments were conducted in two generation. When the third trifoliolate

appeared, the plants were inoculated with a mixture of five races of the pathogen.

The estimated of area under the disease progress curve (AUDPC) was greatest for the most susceptible lines (CI - 107 and Carioca). The most significant result, especially in 2007, was that the performance of the multiline was the same as of the resistant lines, bearing in mind that 2/7 of the mixture consisted of very susceptible lines. Apparently, the resistant plants in the mixture acted as a physical barrier against the spread of inoculum, limiting the amount of inoculum reaching the sensitive components of the mixture. The inoculum was also diluted due to its

decreased production in resistant plants, and less infective inoculum on susceptible hosts due to the deposition of some of the spores on resistant plants.

For yield, the multiline performed very similar to the lines with best performance (Table 4). Again this result is very significant, because the two lines Carioca and CI-107 were part of the mixture, with low average grain yields, mainly due to the pathogen attack. Therefore, it can be concluded that the multiline increased stability.

Table 4. Lines of carioca bean type with the respective number of the race of *C. lindemuthianum* to which it is resistant, estimate of the area under the disease progression curve (AUDPC) and grain yield (g plot⁻¹) of these lines and the multiline

Line	Pathogen race	AUDPC		Grain yield (g.plot ⁻¹)	
		2007	2008	2007	2008
Carioca	susceptible ^{1/}	35.00 a ^{2/}	39.25 a	306.67 c	121.67 c
Talismã	89	28.42 b	27.83 b	723.33 b	270.00 b
RC-1-8	337	14.42 c	13.33 d	1043.33 a	276.67 b
MA-11-8	81	11.67 c	12.00 d	1030.33 a	431.67 a
MA-11-16	87	16.17 c	18.75 c	1058.33 a	298.33 b
MA-11-22	65	7.50 c	7.00 d	1108.33 a	275.00 b
CI-107	susceptible	36.67 a	28.75 b	301.67 c	73.33 c
Multilinha	-	10.83 c	19.17 c	906.67 b	283.33 b

¹ Susceptible to all inoculated races;

² Means followed by the same letter belong to the same group by the test of Scott and Knott (1974) test at 5 % significance. Source: Botelho et al. (2011).

ACKNOWLEDGEMENTS

The authors wish to thank to the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for research studentships of the authors.

Melhoramento de plantas autógamas

RESUMO - Essa apresentação tem por objetivo discutir estratégias para aumentar a eficiência da seleção no melhoramento de plantas autógamas. Entre as estratégias está o emprego de modelos mistos visando o aproveitamento das informações de genealogia no método do pedigree e na análise seqüencial nos testes de avaliação de progênie em vários ambientes/geração. É discutido também o emprego de multilinhas como alternativa para aumentar a vida útil de uma cultivar.

Palavras-chave: melhoramento de plantas, genética quantitativa, multilinha, emprego de modelos mistos.

REFERENCES

- Botelho FBS, Ramalho MAP, Abreu AFB and Rosa HJA (2011) Multiline as a strategy to reduce damage caused by *Colletotrichum lindemuthianum* in common bean. **Journal of Phytopathology** **159**: 175-180.
- Duarte JB and Vencovsky R (2001) Estimação por modelo linear misto com ênfase na ordenação de médias de tratamentos genéticos. **Scientia Agrícola** **58**: 109-117.
- Henderson CR (1984) **Applications of linear models in animal breeding**. University of Guelph, Guelph, 462p.
- Mendes MP, Ramalho MAP and Abreu AFB (2011) Strategies for selecting individuals in common bean breeding programs. **Bean Improvement Cooperative** **54**: 68-69.
- Moreto AL, Ramalho MAP and Abreu AFB (2007) Estimação dos componentes da variância fenotípica em feijoeiro utilizando o método genealógico. **Ciência e Agrotecnologia** **31**: 1035-1042.
- Nunes JAR, Ramalho MAP and Ferreira DF (2008) Inclusion of genetic relationship information in the pedigree selection method using mixed models. **Genetics and Molecular Biology** **31**: 73-78.
- Piepho HP and Möhring J (2006) Selection in cultivar trials – is it ignorable? **Crop Science** **46**:192-201.
- Piepho HP, Möhring J, Melchinger AE and Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica** **161**: 209-228.
- Ramalho MAP, Abreu AFB and Santos JB (2005) Genetic progress after four cycles of recurrent selection for yield and grain traits in common bean. **Euphytica** **144**: 23-29.
- Resende MDV (2002) **Genética biométrica e estatística no melhoramento de plantas perenes**. EMBRAPA Informação Tecnológica, Brasília, 975p.
- SAS Institute (2000) **SAS/STAT user's guide, version 8**. SAS, Cary.
- Scott AJ, Knott MA (1974) Cluster analysis method for grouping means in the analysis of variance. **Biometrics** **30**: 507-512.