

MID PARENT AND COEFFICIENT OF PARENTAGE AS PREDICTORS FOR SCREENING AMONG SINGLE CROSSES FOR THEIR INBREEDING POTENTIAL

José Francisco Ferraz de Toledo

ABSTRACT

A method that allows the early comparison of the proportions of the distributions of inbred lines derivable from single crosses scoring above or below given standards is presented. The method uses information usually already available to the breeder. The mean and genetic variability of the random inbreds are estimated from the mid parental values and the inverse of the Malécot's parentage coefficient, respectively. The mid parental values are available from the line or cultivar testing experiments and the Malécot's coefficients are easily calculated from the ancestry of the parents. The method was efficient for selecting superior crosses.

INTRODUCTION

Early prediction of the outcome of a plant breeding program can greatly improve its efficiency. In self-pollinating species the predictions of single, multiple or back cross inbreeding potential can be made using the distributional properties of the end or target populations. Generally, when the number of segregating loci controlling the character under study is large, the distributions of the random pure lines stemming from those crosses will be normal, with theoretically defined mean and variance (Jinks and Pooni, 1976; Toledo *et al.*, 1984; Pooni and Jinks, 1985). Therefore, the requirements for the predictions are the estimation of these means and variances, obtainable from segregating generations of the single, multiple or back crosses. These procedures have been tested and have proved to be reliable (Caligari *et al.*, 1985; Toledo, 1989).

In this paper an alternative and approximate method of assessing and comparing the crosses variability is presented. The relative breeding potential of single crosses is calculated using an adaptation of the formulae of Jinks and Pooni (1976). The crosses are then ranked according to their relative breeding potential for screening purposes.

THEORY

The method of Jinks and Pooni (1976) allows the prediction of the single cross potential to generate inbred lines scoring $\geq P_1$, $\geq F_1$, $\leq P_2$ or actually falling in any category one may care to specify. For a quantitative character under the additive and dominance genetic model, the probability that random inbred lines will score above P_1 , below P_2 or above F_1 is given by

$$\int_{P_1}^{\infty} f(x) dx, \quad \int_{-\infty}^{P_2} f(x) dx \quad \text{and} \quad \int_{F_1}^{\infty} f(x) dx$$

respectively, with the corresponding abscissa values of

$$\frac{P_1 - m}{\sqrt{D}}, \quad \frac{P_2 - m}{\sqrt{D}} \quad \text{and} \quad \frac{F_1 - m}{\sqrt{D}}$$

The proportion of inbred lines scoring $>$ std (any standard variety) is given by the abscissa value of

$$\frac{\text{std} - m}{\sqrt{D}}$$

In these formulae m and \hat{D} are the expected mean and variance of the F_{∞} inbred lines. The predictions are still valid under the more complicated genetic models, including epistasis, linkage disequilibrium and genotype \times environmental interactions (Toledo, 1984). The accuracy of the predictions will depend on the source of the estimators of m and D (Pooni *et al.*, 1977; Pooni and Jinks, 1979).

In this work the early estimate of m is obtained from the mid parental values calculated from data of line testing experiments (yield trials). Under the simple genetic model involving additive genetic and dominance effects the mid parental value is an unbiased estimate of m . Under the more complex model involving digenic epistasis the

mid parental value is an estimate of $(m + [i])$ and the predictions will be affected according to the relative magnitude of m and $[i]$ (Pooni *et al.*, 1977). Linkage disequilibrium will not affect the estimate of m in the absence of non-allelic interaction, and overall, its effects are not expected to significantly affect the predictions (Toledo, 1987). Genotype \times environmental effects on the estimate of m can be avoided, for practical purposes, by using data from trials from three or four locations.

To evaluate the variability of the cross, the genetic distance between the parental lines was used. Actually, scaled values of the inverse of the coefficient of parentage (Malècot, 1969), where the largest 1/Malècot's Coefficient value was made equal to one, were used to obtain the expected relative degree of variability between crosses. The square roots of these values were used in Jinks and Pooni formulae, substituting for D , and a rank of the relative breeding potential of the crosses was obtained.

Because no actual estimate of the genetic variance of a cross was possible without raising the segregating generations, the absolute values of the crosses potentials to generate superior lines could not be calculated. For screening purposes, however, the relative breeding value of each cross is all that is required, because it allows the ranking of crosses according to their potential to generate superior inbred lines.

MATERIALS AND METHODS

The experimental data for the predictions were obtained from a randomized complete block with three replications, involving eight soybean (*Glycine max* L. Merrill) cultivars (Pérola, Paraná, IAS-4, Davis, BR-5, Bragg, BR-1 and Bossier). The experimental plots were formed by three bordered rows, four meters long after end trimming. The experiment was carried out during the 1981/82 growing season, at Londrina, PR, and was originally planned for other purposes. The Malècot's coefficient of parentage between those eight cultivars were published by Vello *et al.* (1988). Cultivars BR-5 and Paraná were used as standards. The average yield of the parents and the variability estimator were used to calculate the breeding potential of each cross, which were then ranked according to their potential.

The second set of data was obtained from approximately 90 random F_4 families available for each of 28 crosses between the same eight soybean cultivars. The materials were grown in the experimental field at Londrina, PR, during the 1981/82 season. One meter bordered rows were used. For each five rows, one row of either BR-5 or Paraná was planted as standard. Each set of 90 random F_4 families plus standards formed one

experiment. Twenty-eight experiments were available in 1981/82. The number of F_4 progenies outyielding the check varieties was used to evaluate the observed inbreeding potential of the crosses. The genetic variance of the cross (D) was estimated from the between F_4 progeny variance which is equal to $(1/2 D + 1/64 H + E)$, as $D = 2(V_{1F_4} - E)$, assuming $1/64 H = 0$. E was calculated as the average of the check varieties (non-segregating material) variability.

The predictions were compared with the observed proportions of F_4 inbred lines outscoring the same two standards. Correlation analysis, between predicted and observed ranks of breeding potential and between the inverse of the Malécot's coefficient and the estimated D obtained from the between F_4 family variance, were performed (Snedecor and Cochran, 1967). The selection efficiency (Hamblin and Zimmermann, described by Abreu *et al.*, 1990) of the method was calculated.

RESULTS AND DISCUSSION

The mid parental values, the scaled inverse of the Malécot's coefficient of parentage used for the predictions and the observed value of D for each cross are shown in Table I. The calculated D value was negative for the crosses Pérola x IAS-4 and Pérola x Paraná, and we assumed the variability within these crosses to be zero. The R^2 value of the regression of \hat{D} over the inverse of the Malecot's coefficient of parentage was 0.38 (no intercept regression model). Table II shows the predicted and observed proportions of inbred lines outyielding the standards. The rank correlation between predictions and observations was 0.59, significant at the 1% level, for the proportions of lines outyielding the average of the standards. In Table III the results of selecting 20% and 50% of the best crosses (to generate random inbred lines scoring above the average of the standards), classified by the proposed prediction method, can be observed. Selecting the predicted best six and 14 crosses outyielding the average of the standards the 3rd, 5th, 6th, 8th, 11th and 13th and 1st, 2nd, 3rd, 5th, 6th, 7th, 8th, 9th, 10th, 11th, 13th, 15th, 17th and 24 th positioned crosses, respectively, were picked. Using the selection efficiency criteria of Hamblin and Zimmermann we found that it was 57.1% and 40.0% (for 50% and 20% selection pressure based on inbreds scoring above the average of the standards, respectively). These results show that the effects of selection are positive and far from random. For mild selection pressure the results were superior, indicating that the method may be most useful as a preliminary screening tool when the breeder has to decide among several alternative crosses.

Table I - Mid parental value, scaled inverse of Malécot's coefficient of parentage and additive genetic variance between F₄ lines for each cross.

Cross	Mid parent kg/ha	Inverse of Malécot's coefficient ^{/1}	D ²
Pérola x Paraná	3215	0.1426	2362
Pérola x IAS-4	3302	0.0571	0
Pérola x Davis	3525	0.0998	0
Pérola x BR-5	3483	0.0587	570
Pérola x Bragg	3089	0.2852	3078
Pérola x BR-1	3562	0.1110	10578
Pérola x Bossier	3513	0.2496	11796
Paraná x IAS-4	3197	0.0998	2196
Paraná x Davis	3420	0.0907	3236
Paraná x BR-5	3378	0.0434	2654
Paraná x Bragg	2983	0.1426	3274
Paraná x BR-1	3457	0.0624	2636
Paraná x Bossier	3408	0.1248	9134
IAS-4 x Davis	3507	0.0688	3936
IAS-4 x BR-5	3466	0.0540	2088
IAS-4 x Bragg	3071	0.0512	4596
IAS-4 x BR-1	3545	1.0000	5504
IAS-4 x Bossier	3496	0.2496	10140
Davis x BR-5	3688	0.0907	4420
Davis x Bragg	3293	0.1535	3982
Davis x BR-1	3767	1.0000	9142
Davis x Bossier	3718	0.2496	13796
BR-5 x Bragg	3252	0.1331	3556
BR-5 x BR-1	3726	0.0587	13412
BR-5 x Bossier	3677	0.0832	7522
Bragg x BR-1	3331	0.2496	16168
Bragg x Bossier	3282	0.0624	3742
BR-1 x Bossier	3756	0.1248	4500

/1 Largest value scaled to be equal to one.

/2 Obtained as 2 (V_{IF4} - E).

Table II - Predicted and observed rank of the genetic potential of the single crosses to generate inbred lines outyielding the standards.

Cross	Rank of genetic potential					
	≥ BR-5		≥ PR		≥ Mean	
	P	O	P	O	P	O
Pérola x Paraná	22	21	24	18	24	16
Pérola x IAS-4	25	24	16	27	21	27
Pérola x Davis	11	16	10	12	8	15
Pérola x BR-5	14	4	4	20	9	7
Pérola x Bragg	20	25	26	14	25	23
Pérola x BR-1	8	10	9	8	6	11
Pérola x Bossier	9	5	15	1	13	2
Paraná x IAS-4	24	28	25	23	26	21
Paraná x Davis	16	14	13	15	16	14
Paraná x BR-5	23	19	11	16	18	19
Paraná x Bragg	27	20	28	24	27	26
Paraná x BR-1	17	22	8	25	12	24
Paraná x Bossier	15	15	14	17	17	18
IAS-4 x Davis	12	12	7	9	7	10
IAS-4 x Bragg	28	26	27	13	28	22
IAS-4 x BR-1	7	27	22	28	15	28
IAS-4 x BR-5	18	18	5	22	11	17
IAS-4 x Bossier	10	3	17	2	14	1
Davis x BR-5	5	11	6	7	4	8
Davis x Bragg	19	17	20	26	20	20
Davis x BR-1	4	9	19	10	10	9
Davis x Bossier	3	6	12	6	5	5
BR-5 x Bragg	21	23	23	21	22	25
BR-5 x BR-1	1	13	1	11	1	13
BR-5 x Bossier	6	7	2	5	3	6
Bragg x BR-1	13	8	21	19	19	12
Bragg x Bossier	26	1	18	4	23	4
BR-1 x Bossier	2	2	3	3	2	3

Table III - Effect of selection based on the predicted rank of the potential of the single crosses to generate random inbred lines scoring higher than the average of the standards.

Selection of the Best	Actual position of the crosses selected
20% (6 out of 28)	3, 5, 6, 8, 11 and 13
50% (14 out of 28)	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 13, 15, 17 and 24

The proposed method is not exact, as can be readily noted from the correlation analysis and from some of the departures between predicted and observed data presented in Table II. The data do not allow a detailed evaluation of the genetical and sampling biases involved, but some comments can be made: Some departures arise because predictions and observations were obtained from different experiments. These differences, to a certain extent, mimic the real breeding situation where predictions and observations will necessarily occur on separate occasions. The other reasons for the departures are the genetical biases present in our estimates of the mean and genetical variability of the derivable inbred lines used for the predictions (Theory section). The presence of *i* type epistasis is known to cause problems for the predictions because the mid parental actually becomes an estimate of $(m + [i])$. Non-allelic interaction is considered the most important source of deviations between predictions and observations (Pooni *et al.*, 1977; Toledo *et al.*, 1984). Non-allelic interaction has been found in crosses between soybean lines or varieties (Brim and Cockerham, 1961; Toledo *et al.*, 1991) and is, therefore, expected to be present in several of our crosses. Genotype x environmental interaction, that is potentially a serious problem for selection purposes, was deliberately excluded from our results by predicting and selecting in the same environment. The estimator of the cross variability depends on the degree of certainty with which the parental coefficient represents the true genetic distance between the parents, i.e. how accurate the assumption that each parent contributes with 50% of each offspring's genetic architecture holds even in the presence of selection (St. Martin, 1982). Vello *et al.*, 1988 made pertinent comments on the reliability of the Malècot coefficient as a variability predictor, including the use of averages of the coefficients given a set of parents. Although both authors reported that the 50% assumption is correct for most situations the occasional departures from the 50% rule may cause departures between the predictions and observations. It is important to note, however, that any variability estimator used in the prediction formulae is square rooted, diminishing the influence of any bias present. A complete evaluation of the contribution of each source of bias to the observed departures between predictions and observations must await the results of the experiments specifically planned to study them.

The 28 crosses evaluated in this work, are not a particularly large sample, but are expected to represent a wide range of genetical situations. The results obtained were illustrative of the applicability of the method.

ACKNOWLEDGMENTS

This work was supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

RESUMO

A previsão dos resultados de um programa de melhoramento, quando este se encontra em sua fase inicial, é altamente desejável. Diversos métodos que prevêm a distribuição de frequência dos genótipos gerados nos programas foram apresentados em passado recente e mostraram ser confiáveis. Invariavelmente, entretanto, esses métodos requerem experimentação além daquelas previstas no âmbito dos programas. Neste trabalho, apresentamos uma metodologia que permite a comparação das proporções esperadas de linhas puras deriváveis de cruzamentos simples, que são superiores ou inferiores aos referenciais estabelecidos. Assume-se normalidade das distribuições das linhas puras para características poligênicas. A média e a variabilidade das distribuições são estimadas a partir da média e dos coeficientes de Malécot dos parentais, respectivamente. A média dos parentais pode ser calculada a partir de informações normalmente disponíveis dos testes de linhagens e os coeficientes de Malécot são obtidos das informações da genealogia dos parentais. O método demonstrou ser eficiente na seleção de cruzamentos superiores.

REFERENCES

- Abreu, A.F.B., Ramalho, M.A.P., Santos, J.B. and Pereira-Filho, I.A. (1990). Effect of genotype x environment interaction on estimation of genetic and phenotypic parameters of common beans. *Braz. J. Genetics* 13: 75-82.
- Brim, C. and Cockerham, C.C. (1961). Inheritance of quantitative characters in soybeans. *Crop. Sci.* 1: 187-190.
- Caligari, P.D.S., Powell, W. and Jinks, J.L. (1985). The use of doubled haploids in barley breeding. 2. An assessment of univariate cross prediction methods. *Heredity* 54: 353-358.
- Jinks, J.L. and Pooni, H.S. (1976). Predicting the properties of recombinant inbred lines derived by single seed descent. *Heredity* 36: 253-266.
- Malécot, G. (1969). *The Mathematics of Heredity*. English edition. W.H. Freeman and Company. San Francisco, U.S.A.
- Pooni, H.S., Jinks, J.L. and Cornish, M.A. (1977). The causes and consequences of non-normality in predicting properties of recombinant inbred lines. *Heredity* 38: 329-338.
- Pooni, H.S. and Jinks, J.L. (1979). Sources and biases of the predictors of the properties of recombinant inbreds produced by single seed descent. *Heredity* 42: 41-48.

- Pooni, H.S. and Jinks, J.L. (1985). Predicting the properties of first cycle inbreds and second cycle hybrids extractable from two, three and four parent crosses. *Heredity* 54: 397-411.
- Snedecor, C.W. and Cochran, W.G. (1967). *Statistical Methods*. 6th edition. Iowa University Press, Ames.
- St. Martin, S.K. (1982). Effective population size for the soybean improvement program in maturity group 00 to IV. *Crop. Sci.* 22: 151-152.
- Toledo, J.F.F. (1987). Predicting the inbreeding and the outcrossing potential of soybean (*Glycine max* (L) Merrill) varieties. *Braz. J. Genetics* 10: 543-558.
- Toledo, J.F.F. (1989). Quantitative Genetics in Soybean Breeding. *Proceedings of the World Soybean Research Conference IV* (Pascual, A.J., ed.). Buenos Aires, Argentina, pp. 909-914.
- Toledo, J.F.F., Pooni, H.S. and Jinks, J.L. (1984). Predicting the transgressive potentials of *Nicotiana rustica* crosses for superior recombinant inbreds and second cycle hybrids. *Proceedings of the Fifth Meeting of Biometrics Section of Eucarpia*. (Geiger, H.H. and Ruckebauer, P., eds.). Hohenheim, Stuttgart, pp. 151-156.
- Toledo, J.F.F., Rossini, M.C., Souza, R.F. and Leão, F.F. (1991). Genetical analysis of a soybean biparental cross and comparative model fitting to means and variances of two sets of generations. *Braz. J. Genet.* 14: 1041-1064.
- Vello, N.A., Hiroto, D.M. and Azevedo Filho, A.J.B.V. (1988). Coefficient of parentage and breeding of Brazilian soybean germplasm. *Braz. J. Genet.* 11: 679-697.

(Received December 20, 1990)