Crop Breeding and Applied Biotechnology 6:30-39, 2006
Brazilian Society of Plant Breeding. Printed in Brazil

Prediction of genetic variability through AFLP-based measure of genetic distance in soybean

Ana Lídia Variani Bonato[1*], Eberson Sanches Calvo[2], Carlos A. Arrabal Arias[3], José Franscisco de Ferraz Toledo[3], and Isaias Olívio Geraldi[4]

Received 16 June 2005

Accepted 23 March 2006

ABSTRACT - *Molecular markers have been used to predict the genetic variance of segregating soybean populations by measuring the genetic distance between parents. For this purpose we analyzed the genetic distance among six pairwise combinations of four soybean lines grown in Southern Brazil. The AFLP-based distances (AFD) were estimated with 21 primer-pair combinations. Additive genetic variance (D) was estimated in 100 advanced inbred lines for each cross for seven traits, including grain yield, for four sowing dates, and four years. Results showed a low but significant overall correlation (r=0.21*) between grain yield variance and AFD in the average of the four years and sowing dates. Considering the sowing dates the highest correlation was in October (r=0.41*). Significant correlations were also found for five out of the remaining traits evaluated. We found no significant correlations between D and parentage coefficient for all traits evaluated, except for 100 seed weight (r=0.32*).*

Key words: *Glycine max*, molecular markers, genetic distance, coefficient of parentage, and additive genetic variance.

INTRODUCTION

The choice of parents to be used for crossing is one of the most critical steps in a plant breeding program. Ideally, breeders would like to reduce the number of crosses and populations by working only with those crosses which would maximize the genetic variance. It is generally accepted that, given the same mean between the parents, crosses between unrelated genotypes will maximize the number of segregating alleles resulting in a larger genetic variance of the progeny (Cox et al. 1985). Therefore, the previous knowledge of the genetic distance of the parents could be a valuable piece of information for cross planning in plant breeding programs. The use of pedigree-based measures of genetic distance, such as the coefficient of parentage (CP or f; Malécot 1948) has been proposed (Kemptorne 1973). This measure has been widely used to study genetic diversity in soybean germplasm (Vello et al. 1988, Gizlice et al. 1994).

Nevertheless, the use of CP may pose several problems that have motivated the search for other measure types. Molecular have been proposed as a direct, DNA-based measure of genetic distance in animals and plants including soybean (Williams et al. 1990, Thompson and Nelson 1998, Mohammadi and Prasanna 2003).

Regardless of the type of measure in use, the correlation between the genetic distance of the parents and the genetic variance in the progeny has led to mixed results. In oat, Cowen and Frey (1987) found a significant positive correlation between genetic distance, measured as 1-CP, and the genetic variance in the progeny for only straw yield and plant height. In an attempt to correlate RFLP-based measures (RFLP-GD) using the same parents

[1] Embrapa Trigo, C. P. 451, 99.001-970, Passo Fundo, RS, Brasil. *E-mail: analidia@cnpt.embrapa.br
[2] TMG - Tropical Melhoramento e Genética Ltda, Rodovia Celso Garcia Cid, Km 87, 86.183-600 , Cambé, PR, Brasil
[3] Embrapa Soja, C. P. 231, 86.001-970, Londrina, PR, Brasil
[4] Departamento de Genética, Escola Superior de Agricultura "Luiz de Queiroz", USP, C. P. 83, 13.400-970, Piracicaba, SP, Brasil

resulted in significant correlation for plant height only (Moser and Lee 1994). In a recent review, Dias et al. (2004) discussed articles about *a priori* choice based on parental distances by means of agronomic and molecular data, which is still a controversial procedure.

In soybean, attempts to correlate the genetic variance in populations and genetic distance between parents have also been reported. Manjarrez-Sandoval et al. (1997) observed that, although both measures correctly identified the population with the highest genetic variance, 1-CP was a better measure than RFLP-GD for the prediction of genetic variance of five different populations. Working with a larger set of population, Kisha et al. (1997) concluded that both measures can identify groups of crosses that give rise to mean populations with higher genetic variances with RFLP data that seemingly have a better performance. Powell et al. (1996) found only moderate correlations when comparing genetic distances in soybean measured by four different types of molecular markers. Furthermore, Helms et al. (1997) reported the lack of correlation between genetic variance and genetic distance measured by RAPD markers.

Our objective was to investigate the potential of AFLP-based genetic distance measures in the prediction of genetic variability in soybean populations.

MATERIAL AND METHODS

Plant material

Seeds from the four soybean genotypes (BR85-29009, FT-2, BR-13, and Ocepar 8) used in the analysis were obtained from single selfed plants. Apart from BR85-29009, which is an elite breeding line, the other three genotypes are cultivars adapted to southern Brazil and all belong to the same maturity group. They were previously crossed in a diallel mating design originating six different soybean populations as shown in Table 1 and reported by Triller and Toledo (1996). The additive genetic variance (D) among families for F_7, F_8, F_9, and F_{10} generations was based on previous information from completely randomized field experiments using hill plots carried out during the 1991/92, 1992/93, 1993/94, and 1994/95 growth seasons. During the 1991/92 and 1992/93 seasons three sowing dates (October, November and December) were evaluated. In the subsequent years the September sowing date was also included in the experiments. A total of 100 families with four plants each were evaluated in each generation, essentially as described by Triller and Toledo (1996).

Table 1 - Soybean populations and AFLP-based genetic distances (AFD) and genealogical distances (GD)

Population	Cross	AFD	GD
1	BR 85-29009 x FT-2	0.49	0.70
2	BR 85-29009 x BR-13	0.37	0.90
3	BR 85-29009 x OCEPAR-8	0.72	0.70
4	FT-2 x BR-13	0.51	0.89
5	FT-2 x OCEPAR-8	0.60	0.50
6	BR-13 x OCEPAR-8	0.62	0.89

Additive genetic variances and their significances were estimated with the least square method described by Mather and Jinks (1982). Seven quantitative agronomic traits were evaluated: grain yield (g plant^{-1}), days to flowering, days to maturity, plant height at flowering (cm), plant height at maturity (cm), number of nodes, and 100 seed weight (g). All experiments were conducted at Embrapa Soybean in Londrina, PR, Brazil.

DNA extraction and quantification

The molecular analysis was performed from 1999 to 2000. DNA was obtained from greenhouse grown plants at the V2 stage. The extraction was performed essentially as described by Saghai-Maroof et al. (1984). DNA was visually quantified in agarose gels using undigested lambda phage DNA as standards. The samples were diluted to approximately 30 ng uL^{-1} and used in AFLP analysis.

AFLP analysis

All AFLP analyses were carried out with the "AFLP Analyses Kit I" (Gibco-LifeTechnologies, Rockville, MD) essentially as described in the kit manual. All amplifications were conducted in a Perkin-Elmer Gene Amp 9600 (Perkin-Elmer Corp., Norwalk, CT) thermocycler. AFLP products were fractionated on 5% polyacrylamide sequencing gels, dried and exposed to autoradiography film. A total of 21 *Eco*RI-*Mse*I primer pair combinations were used and are listed in Table 2.

Polymorphisms were scored in a binary system, that is, presence or absence of the same band in all possible pairwise genotypic combinations. The matrix obtained was used to estimate the AFLP-based genetic similarity (AFS) calculated by means of the Nei and Li (1979) coefficient. AFLP-based genetic diversity (AFD) was obtained as 1-AFS.

Bootstrap analysis

The bootstrap analysis was employed to verify whether the number of polymorphisms used for the AFLP-

based measures was enough. The analysis methodology was described in depth by Barroso et al. (2003).

Coefficient of parentage (CP)

CP values were calculated from pedigree records basically as described by Vello et al. (1988). The genealogic distance was estimated as GD= 1-CP.

Correlation between additive genetic variance (D) and genetic distances

Spearman's correlations between D and the genetic distances AFD or GD as well as their significance were calculated for each trait and each sowing date using software SAS. The efficiency of the prediction was further calculated as the probability of correctly identifying the top three populations with the highest additive genetic variance by selecting the top three most divergent crosses.

RESULTS

AFLP analysis

Table 2 summarizes the results of the AFLP analysis. The 21 primer pair combinations generated a total of 963 bands. We were able to unambiguously score 165 polymorphisms (17.1%). Therefore, each primer pair yielded a mean of 7.86 polymorphisms. The primer pair combinations *Eco*RI-AAC/*Mse*I-CTG, *Eco*RI-AAC/*Mse*I-CTA, *Eco*RI-AAG/*Mse*I-CTT and *Eco*RI-ACT/*Mse*I-CAT resulted in the highest number of polymorphisms (Table 2).

Table 2 - *Eco*RI/*Mse*I primer-pair combinations used in the AFLP analysis and their respective number of polymorphic and non-polymorphic bands

*Eco*RI	*Mse*I	Number of bands	Polymorphic bands
AAC	CTG	50	16
AAC	CTC	37	5
AAC	CTA	62	14
AAC	CAC	42	4
AAC	CAG	34	3
AAC	CAA	34	4
AAC	CAT	86	9
AAG	CTG	45	12
AAG	CTC	48	10
AAG	CTA	65	7
AAG	CTT	72	13
AAG	CAC	40	5
AAG	CAA	34	8
AAG	CAT	42	5
ACT	CTG	32	8
ACT	CTA	44	9
ACT	CTT	37	5
ACT	CAC	27	5
ACT	CAG	41	6
ACT	CTC	26	4
ACT	CAT	65	13
Total		963	165

Bootstrap analysis

A mean variation coefficient (VC) of 9.2% was obtained by bootstrap analysis with the 165 markers used in the AFLP analysis. The VC decreased as the sample size increased, indicating that the accuracy of genetic similarity estimates increased with the increase in the number of polymorphic loci. Nevertheless, results showed that this increase dwindled when more than 100 polymorphic markers were used.

Estimates of genetic distances and correlation between AFD and GD

In spite of the small number of populations evaluated (Table 1) the AFD covered a good range of variation (0.37 to 0.72). The analysis identified cross BR 85 29009 X Ocepar 8 as the most divergent. On the other hand, three other crosses presented identical genetic distances when the GD was calculated, which suggested a poor correlation between the two measures beforehand. In fact, the correlation between the two measures was negative (r=-0.39).

Correlations between genetic distances (GD and AFD) and additive genetic variance (D)

The estimates of additive genetic variances for each sowing date averaged for the different years are presented in Table 3. The significances of the variances tested using software Genfit (Toledo 1991) manifested significant genetic variation for all evaluated traits in all populations (Table 3). The mean correlations between GD or AFD and D for the four years are shown in Table 4. Except for the traits days to flowering in September and 100 seed weight in October and December, there were no significant correlations between GD and D. Besides the low magnitudes, several traits were negatively correlated.

On the other hand, there were significant correlations between AFD and D for five different traits: grain yield, number of nodes, days to flowering, days to maturity, and plant height at maturity (Table 4). The trait number of nodes attained the highest overall correlation (r=0.50**).

The only significant correlation between D and AFD for grain yield was observed for the October sowing date

(0.41*). A graphical distribution of the additive genetic variance (D) for yield vs AFD (Figure 1) enabled us to easily visualize that some points had low AFD values but high additive genetic variances (marked by ◆ in Figure 1A-1D). A closer look at these data-points confirmed that they invariably came from the same population (BR 85-29009 x BR 13). Owing to this consistency we decided to test this cross as an outlier by removing the data from the sowing dates which resulted in additive genetic variance above the curve. In Table 4, the exclusion of these data greatly increased the overall correlations for yield as well as for other traits with exception of 100 seed weight which still had no correlation. Moreover, the most significant correlations with grain yield were obtained for the sowing dates October (0.70**) and November (0.52**), once more reinforcing the significance of the correlations. The graphical changes can be seen in Figure 1E-1H.

We were able to detect a significant and positive correlation between AFD and D, in order to get a clearer picture of the significance of these correlations from a plant breeding perspective. We measured the prediction efficiency as the probability of correctly identifying the top three crosses (those with the largest D values during the four years within each sowing date) by selecting the three (50% selection rate) crosses with the largest AFD values. Table 5 shows the results of the estimate of prediction efficiency for the seven different traits. The values ranged from 33% to 100% while the efficiencies for grain yield were 66.6%; 83.3%; 83.3%; and 58.3%, for September, October, November and December, respectively.

DISCUSSION

AFLP analysis

The results of polymorphism analysis with AFLP markers obtained here confirmed the ability of the AFLP technique to quickly generate a large number of markers. Greater polymorphism averages such as 18 and 14 have

Table 3 - Estimates of additive genetic variance (D) for seven agronomic traits of six soybean populations of four sowing dates[a]

Season	Population	Grain yield	Number of nodes	Days to flowering	Plant height at flowering	Days to maturity	Plant height at maturity	100 seed weight
September	1	49.91	0.91	6.96	13.69	30.08	35.86	2.81
	2	153.32	1.49	20.52	26.13	102.98	63.07	2.68
	3	146.82	1.78	16.04	27.24	121.59	79.11	1.07
	4	32.24	0.63	10.35	9.88	60.10	22.88	2.41
	5	67.21	0.85	8.63	9.58	80.15	28.87	1.54
	6	126.24	1.07	21.85	26.33	123.68	62.90	2.26
October	1	69.09	1.08	9.05	38.43	35.32	79.56	2.62
	2	211.13	1.89	25.64	61.60	60.53	127.59	3.65
	3	390.25	2.06	41.97	116.08	105.48	159.94	2.21
	4	97.44	1.33	10.26	27.86	65.04	77.13	3.57
	5	187.04	1.21	11.23	46.51	87.91	68.51	1.92
	6	216.84	2.17	16.46	50.14	87.34	130.05	3.22
November	1	66.44	0.62	3.44	40.86	23.24	67.15	1.96
	2	76.54	0.83	12.17	59.46	37.23	107.88	2.17
	3	118.27	2.30	21.92	87.56	60.20	140.34	2.31
	4	43.39	0.78	9.63	38.25	49.43	52.68	1.77
	5	85.51	1.17	11.22	27.32	57.03	72.70	2.10
	6	90.66	1.32	13.59	31.38	52.11	76.43	3.04
December	1	40.78	0.64	2.10	28.10	12.91	43.62	1.99
	2	58.21	0.86	7.42	47.03	24.13	74.58	2.10
	3	60.15	1.73	12.98	60.87	45.61	84.52	1.59
	4	26.40	1.09	7.77	29.56	21.71	60.41	2.17
	5	44.96	1.25	9.24	40.12	31.72	60.47	1.46
	6	53.41	1.34	9.50	33.38	31.35	68.04	2.84

[a] Significant at 1% probability by the t test

Table 4 - Correlations between additive genetic variance (D) and genealogical distance (GD), and additive genetic variance (D) and AFLP-based genetic distance (AFD) with and without outlier cross, for seven soybean agronomic traits, in four sowing dates

Traits	Sowing Dates	GD x D	AFD X D with outliers	AFD X D without outliers
Grain yield	General	0.01	0.21*	0.43**
	September	0.25	0.20	0.74*
	October	-0.02	0.41*	0.70**
	November	-0.19	0.37	0.52**
	December	0.03	0.11	0.29
Number of nodes	General	-0.04	0.50**	0.64**
	September	0.11	0.10	0.44
	October	0.35	0.35	0.62**
	November	-0.22	0.65**	0.72**
	December	-0.22	0.70**	0.75**
Days to flowering	General	0.14	0.35**	0.56**
	September	0.69**	0.20	0.79**
	October	0.22	0.29	0.49*
	November	0.06	0.54**	0.84**
	December	-0.16	0.69**	0.75**
Plant height at flowering	General	0.09	0.16	0.33**
	September	0.47	0.15	0.56
	October	0.04	0.24	0.35
	November	0.10	0.08	0.27
	December	0.07	0.19	0.59**
Days to maturity	General	-0.05	0.42**	0.58**
	September	0.23	0.56*	0.93**
	October	-0.16	0.68**	0.76**
	November	-0.15	0.55**	0.60**
	December	-0.08	0.52**	0.73**
Plant height at maturity	General	0.17	0.22*	0.41**
	September	0.21	0.35	0.49
	October	0.30	0.21	0.43*
	November	0.08	0.26	0.53**
	December	0.17	0.30	0.57**
100 seed weight	General	0.32**	-0.11	-0.03
	September	0.24	-0.44	-0.44
	October	0.56**	-0.38	-0.26
	November	0.11	0.33	0.53*
	December	0.49**	-0.12	-0.02

*, ** Significant at 5 and 1% probability, respectively

been reported for hop varieties (Hartl and Seefelder 1998) and wheat lines (Barret and Kidwel 1998), respectively. In soybean, Maughan et al. (1996) and Ude et al (2003) reported as much as 18 and 34 polymorphisms per primer pair, respectively. This difference can be explained by the fact that we looked only at a small set (four) of cultivated germplasm, unlike the studies of Maughan et al. (1996) which compared accessions of both *Glycine max* and *Glycine soja*, and Ude et al (2003) who determined the level of genetic distance within and between Asian and north American cultivars.

Considering the small set of genotypes involved in the analysis, these primer combinations are a good choice for future studies on soybean genetic diversity or mapping studies.

Bootstrap analysis

In Bonato et al. (2006) a mean variation coefficient (VC) of 7.7% was obtained by bootstrap analysis with the 78 markers used in the AFLP analysis showing that this number of AFLP markers could be considered sufficient to characterize the soybean cultivars for genetic similarity. Pejic et al. (1998) performed a comparative analysis of genetic similarity in maize measured with RFLP, RAPD, AFLP and SSR markers and obtained variation coefficients ranging from 5 to 10% for the four marker types using 150 polymorphisms per marker type. In our bootstrap study we reached a VC of 10% with approximately 135 markers. No VC reduction was attained with 165 markers suggesting that, at least with AFLP markers in soybean, the minimum acceptable marker number would also be over 100.

Estimates of genetic distances and correlation between AFD and GD

With respect to the negative correlation found between AFD and GD, previous studies have reported a lack of consistent correlation between diversity based on molecular markers and GD in soybean (Abdelnoor et al. 1995, Helms et al. 1997). One could argue that the low correlation could result from the small number of crosses used in the present as well as in previous studies. Although this fact may play an important role in defining the negative nature of correlation, we have recently conducted an extensive evaluation of the Brazilian soybean germplasm (100 soybean varieties) with AFLP's and also found, although positive, a low correlation (r=0.12) between AFS and CP (Bonato et al. 2006).

Lower correlations between AFD and GD have also been observed for other plant species such as wheat (Almanza-Pinzón et al. 2003) and oats (Vieira et al. 2005), which seems to be a general characteristic of autogamous plants (Bohn et al. 1999). There are several possible explanations for this lack of correlation and the distinct nature of the two measure types is certainly an important one. Molecular marker-based measures reflect the genes "identical by state" and genealogical distances reflect the genes 'identical by descent' and consequently they are subject to different types of errors (Melchinger 1999). While CP may be based on incorrect records and ignores the effects of selection, genetic drift and relatedness among the ancestor lines the molecular marker-based measures assume that all co-migrating bands are identical. One must be aware that, unlike the bias of AFD, the bias of CP estimates is expected to accumulate over generations. At this point we interpret the lack of correlation as a consequence of CP limitations.

Correlations between genetic distances (GD and AFD) and additive genetic variance (D)

Due to the low or negative correlations found between D and GD we concluded that GD was not a reliable predictor of D. This low correlations could be explained if the genotypes used in the study did not fulfill the requirements of CP, that is, the absence of relationship (f=0) among the ancestor lines and the equal contribution of the parents after a biparental cross (Vello et al. 1988). Moreover, due to the low number of genotypes used in the analysis a single error in genealogy record could lead to a considerable distortion in the GD measures.

The magnitude of the correlation observed in October (including the outlier cross), although low (0.41*), represents the first significant correlation reported to date for grain yield in soybean. In wheat, Burkhamer et al. (1998) and Bohn et al. (1999) used AFLP in the prediction of genetic variance, specifically. Neither attempts could detect any significant correlations between the two measures. Dias et al. (2004) discussed several causes that influence the results *a priori* choice among them the divergence-heterosis or variance genetic associations. More recently we applied the same technique to a larger set of population (28 different crosses) based on data from a single year and a single sowing date and found an overall correlation of 0.42**.

In relation to the outlier detection with conflicting values between AFD and D, Kisha et al. (1997) also observed the existence of an exceptional cross that

produced a large genetic variance in soybean. According to these authors, some specific crosses differ for many genes controlling the same traits and by chance these differences could be undetected with the markers used. We believe that a similar effect can result from one major gene that controls adaptation. Indeed, it is possible that BR 85-29009 may have such a gene for long juvenile period (Dr. Romeu Kiihl, personal communication).

From a soybean breeding perspective the most important parameter is not the correct ranking of the populations but rather the overall efficiency of using the AFD measure as a tool for selecting the group of crosses that will yield the largest genetic variance. We measured the prediction efficiency and suggest that if a plant breeder is using the AFD information to reduce the number of crosses to be made and consequently the number of populations to be advanced and evaluated by half, he would have an 83% success rate with the normal sowing dates. Therefore, we concluded that AFD measure can be a valuable tool for predicting the additive genetic variance in soybean breeding programs and can advantageously replace the coefficient of parentage (CP).

ACKNOWLEDGEMENTS

The authors acknowledge the financial support received from Fundação Mato Grosso (FMT) and Embrapa Soybean. ALV Bonato was holer of a doctorate scholarship from CNPq.

Table 5 - Prediction efficiency (%) based on AFD for seven soybean agronomic traits and four sowing dates, in two or three years

Trait	September	October	November	December
Grain yield	66.6	83.3	83.3	58.3
Number of nodes	66.6	66.6	74.9	91.6
Days to flowering	66.6	58.3	83.3	83.3
Plant height at flowering	66.6	66.6	66.6	58.3
Days to maturity	66.6	100.0	83.3	100.0
Plant height at maturity	66.6	66.6	66.6	66.6
100 seed weight	33.3	41.6	58.3	33.3

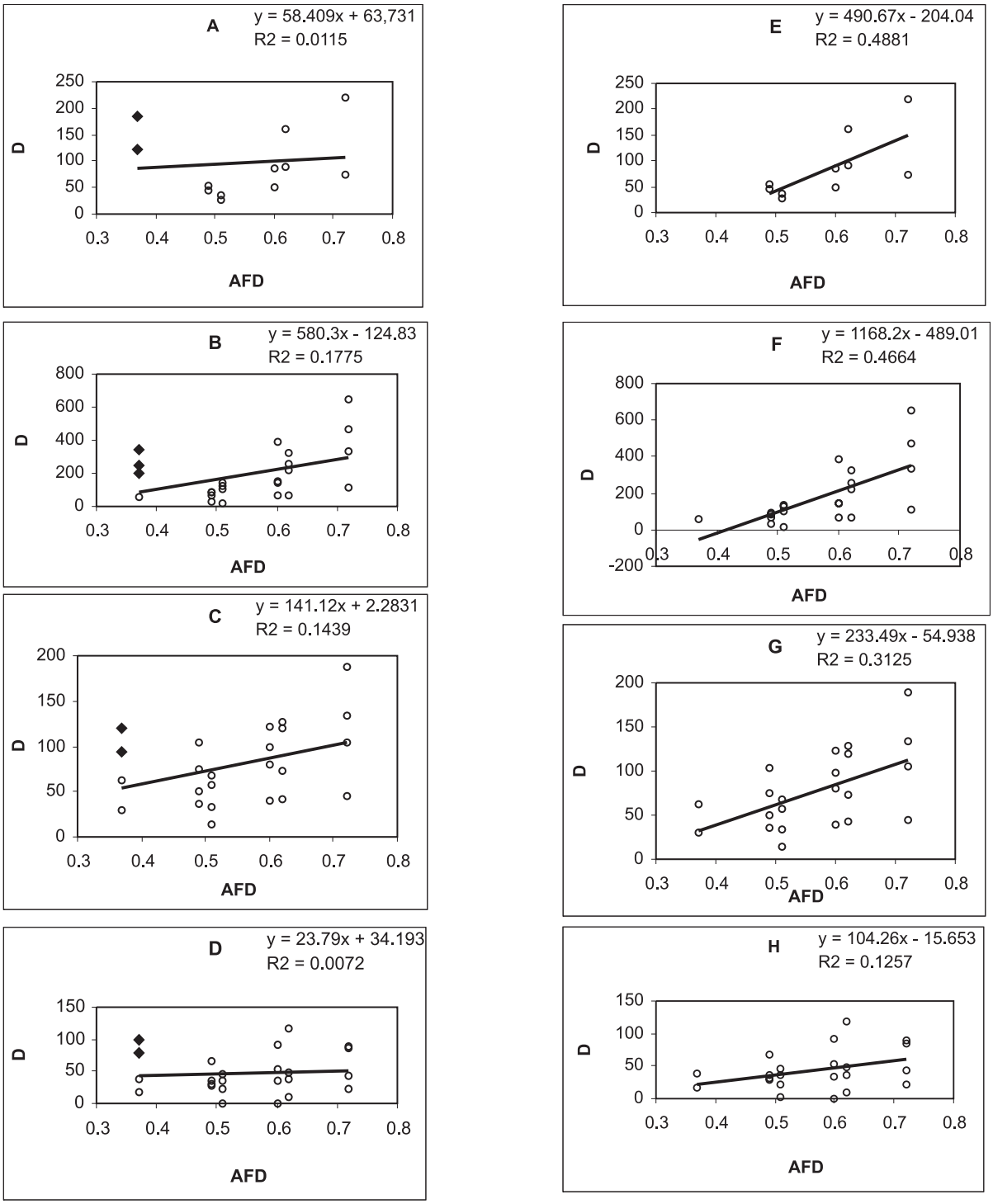

Figure 1 - Relationship between additive variance (D) and AFLP-based genetic distance (AFD) for grain yield. Mean estimates of additive variance covered four years (excluding sowing date September which was based on two years only). A, B, C, and D show the relationship for the sowing dates September, October, November, and December, respectively, with the outlier data (labeled ◆). E, F, G, and H present the same relationship though without outliers

Predição da variabilidade genética através da distância genética baseada em AFLP em soja

RESUMO - *Marcadores moleculares têm sido usados para predizer a variância genética de populações segregantes de soja medindo a distância genética entre os pais. Para este propósito foi analisada a distância genética entre seis cruzamentos de quatro genótipos de soja no Sul do Brasil. As distâncias baseadas em AFLP (AFD) foram calculadas com 21 pares de* primers. *A variância genética aditiva (D) foi calculada em 100 linhas avançadas para cada cruzamento para sete características, inclusive rendimento, em quatro épocas de semeadura e quatro anos. Observou-se baixa, mas significante, correlação geral (r=0,21*) entre variância de rendimento e AFD na média dos quatro anos e época de semeadura. Considerando as épocas de semeadura a correlação maior foi em outubro (r=0,41*) e também foram encontradas correlações significativas em cinco características restantes avaliadas. Não foi encontrada correlação significante entre D e coeficiente de parentesco para todas as características avaliadas, exceto para 100 peso de semente (r=0,32*).*

Palavras-chaves: *Glycine max*, marcadores moleculares, distância genética, coeficiente de parentesco e variância genética aditiva.

REFERENCES

Abdelnoor RV, Barros EG and Moreira MA (1995) Determination of genetic diversity within Brazilian soybean germplasm using random amplified polymorphic DNA techniques and comparative analysis with pedigree data. **Revista Brasileira de Genética 18**: 265-273.

Almanza-Pinzón MI, Khairallah M, Fox PN, Warburton ML (2003) Comparison of molecular markers and coefficients of parentage for the analysis of genetic diversity among spring bread wheat accessions. **Euphytica 130**: 77-86.

Barrett BA and Kidwell KK (1998) AFLP-based genetic diversity among wheat cultivars from the Pacific Northwest. **Crop Science 38**: 1261-1271.

Barroso PAV, Geraldi IO, Vieira, MLC, Pulcinelli, CE, Vencosvky R and Dias CTS (2003) Predicting performance of soybean populations using genetic distances estimated with RAPD markers. **Genetic and Molecular Biology 26**: 343-348.

Bohn M, Utz HF and Melchinger AE (1999) Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSrs and their use for predicting progeny variance. **Crop Science 39**: 228- 237.

Bonato ALV, Calvo ES, Geraldi IO and Arias CAA (2006) AFLP marker genetic similarity of soybean cultivars released in Brazil. **Genetic and Molecular Biology** (in press).

Burkhamer RL, Lanning SP, Martens RJ, Martin JM and Talbert LE (1998) Predicting progeny variance from parental divergence in hard red spring wheat. **Crop Science 38**: 234-248.

Cowen NM and Frey KJ (1987) Relationships between genealogical distance and breeding behavior in oats (*Avena sativa L.*) **Euphytica 36**: 413-424.

Cox TS, Kiang YT, Gorman MB and Rodgers DM (1985) Relationship between coefficient of parentage and genetic similarity indices in the soybean. **Crop Science 25**: 529-532.

Dias LAS, Picoli EAT, Rocha RB and Alfenas AC. (2004) *A priori* choice of hybrid parents in plants. **Genetics and Molecular Research 3**: 356-368

Gizlice Z, Carter Jr TE and Burton JW (1994) Genetic basic for North American public soybean cultivars released between 1947 and 1988. **Crop Science 34**: 1143-1151.

Hartl L and Seefelder S (1998) Diversity of selected hop cultivars detected by fluorescent AFLPs. **Theoretical and Applied Genetics 96**: 112-116.

Helms T, ORF J, Vallad G and McClean P (1997) Genetic variance, coefficient of parentage, and genetic distance of six soybean populations. **Theoretical and Applied Genetics 94**: 20-26.

Kemptorne O (1973) **An Introduction to Genetic Statistics**. 2nd ed., Iowa State University Press, Ames, 545p.

Kisha TJ, Sneller CH and Diers BW (1997) Relationship between genetic distance among parents and genetic variance in populations of soybean. **Crop Science 37**: 1317-1325.

Malécot G (1948) **Les mathématiques de l'hérédité**. Masson et Cie, Paris, 80p.

Manjarrez- Sandoval P, Carter Jr TE, Webb DM and Burton JW (1997) RFLP Genetic similarity estimates and coefficient of parentage as genetic variance predictors for soybean yield. **Crop Science 37**: 698-703.

Mather K and Jinks JL (1982) **Biometrical Genetics**. 3rd ed., Chapman and Hall, London, 396p.

Maughan PJ, Saghai-Maroof MA and Buss GR (1996) Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance and near-isogenic line analysis. **Theoretical and Applied Genetics 93**: 392-401.

Melchinger AE (1999) Genetic diversity and heterosis. In: Coors JC and Pandey S (eds.) **The Genetics and Exploitation of Heterosis in Crops**. ASA-CSSA, Madison, p. 99-118.

Mohammadi SA and Prasanna BM (2003) Analysis of genetic diversity in crop plants – salient statistical tools and considerations. **Crop Science 43**: 1235-1248.

Moser H and Lee M (1994) RFLP variation and geneological distance, multivariate distance, heterosis, and genetic variance in oats. **Theoretical and Applied Genetics 87**: 947-956.

Nei M and Li WH (1979) Mathematical model for study in genetic variation in terms of restriction endonucleases. **Proceedings of the National Academy of Sciences from USA 76**: 5269-5273.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S and Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germoplasm analysis **Molecular Breeding 2**: 225-238.

Saghai-Maroof MA, Sloman KM, Jorgensen RA and Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and populatoin dynamics. **Proceedings of the National Academy of Sciences from USA 81**: 8014-8018.

Thompson JA and Nelson RL (1998) Utilization of diverse germplasm for soybean yield improvement. **Crop Science 38**: 1362-1368.

Toledo JFF (1991) Programa de computador para estimar parâmetros genéticos, componentes de médias e variâncias pelo método dos quadrados mínimos ponderados. **Pesquisa Agropecuária Brasileira 26**: 1023-1029.

Triller C and Toledo JFF (1996) Using the F_3 generation for predicting the breeding potential of soybean crosses. **Revista Brasileira de Genética 19**: 289-294.

Ude GN, Kenworthy WJ, Costa JM, Cregan PB and Alvernaz J (2003) Genetic diversity of soybean cultivars from China, Japan, North America, and North American ancestral lines determined by amplified fragment length polymorphism. **Crop Science 43**: 1858-1867.

Vello NA, Hiromoto DM and Azevedo-Filho AJBV (1988) Coefficient of parentage and breeding of Brazilian soybean germplasm. **Revista Brasileira de Genética 11**: 679-697.

Vieira EA, Carvalho FIF, Oliveira AC, Giovani B, Zimmer PD, Silva JAG, Martins AF, Bertan I, Silva GO and Schmidt DAM (2005) Comparação entre medidas de distância genealógica, morfológica e molecular em aveia em experimentos com e sem a aplicação de fungicida. **Bragantia 64**: 51-60.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA and Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Research 18**: 6531-6535.

Young WP, Schupp JM and Keim P (1999) DNA methylation and AFLP marker distribution in the soybean genome. **Theoretical and Applied Genetics 99**: 785-790.