

GENETIC DISTANCES IN SOYBEAN BASED ON RAPD MARKERS ⁽¹⁾

RONAN XAVIER CORRÊA ⁽²⁾, RICARDO VILELA ABDELNOOR ⁽³⁾,
FÁBIO GELAPE FALEIRO ⁽²⁾, COSME DAMIÃO CRUZ ⁽⁴⁾,
MAURILIO ALVES MOREIRA ^(2,5) & EVERALDO GONÇALVES DE BARROS ^(2,4)

ABSTRACT

Four methods were applied to determine pairwise genetic distances among five soybean genotypes which are potential genitors for a mapping population. Additionally, individual plants from the most divergent pair of genotypes were evaluated by the RAPD technique to determine their degree of homozygosity. Genetic distances based on RAPD data were calculated by the modified Rogers' distance, and also by the following arithmetical complements of similarity: simple match, Nei and Li, and Gower. These genetic distances were similar, presenting a correlation coefficient ranging from 0.99 to 1.00. In all four methods lines UFV 91-717 and Ichigowase were the most divergent ones (4.53 to 21.43%). DNA samples from five plants from each of the two most divergent genotypes were amplified with 28 different primers. Among the amplified products, only five were polymorphic in each group (2.10%), demonstrating their high intragroup degree of homozygosity. These homozygosity were maintained when DNA samples from 12 plants from each of the two most divergent genotypes were amplified. These parameters were extremely useful for the confirmation of the chosen pair of genitors to generate a mapping population.

Index terms: soybean, genetic diversity, RAPD markers, estimation methods.

⁽¹⁾ Received for publication in December 24, 1997 and approved in November 27, 1998.

⁽²⁾ Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa (UFV), 36571-000 (MG).

⁽³⁾ Embrapa-Soja, Caixa Postal 231, 86001-970 Londrina (PR).

⁽⁴⁾ Departamento de Biologia Geral, UFV, 36571-000 Viçosa (MG).

⁽⁵⁾ Departamento de Bioquímica e Biologia Molecular, UFV, 36571-000 Viçosa (MG).

RESUMO

DISTÂNCIAS GENÉTICAS EM SOJA COM BASE EM MARCADORES RAPD

Aplicaram-se quatro métodos para determinar as distâncias genéticas entre cinco cultivares de soja, que são genitores potenciais para uma população de mapeamento genético. Adicionalmente, o grau de homozigose do par de genótipos mais divergente foi avaliado por meio da técnica de RAPD. Calcularam-se as distâncias genéticas fundadas em dados obtidos pela técnica de RAPD pela distância modificada de Rogers e pelos seguintes complementos aritméticos de similaridade: distância simples; Nei e Li, e Gower. As distâncias genéticas foram similares, apresentando valores de coeficiente de correlação de 0,99 a 1,00. Nos quatro métodos, as linhagens UFV 91-717 e Ichigowase foram as mais divergentes (4,53 to 21,43%). Amostras de DNA de cinco plantas de cada uma dessas linhagens foram amplificadas com 28 *primers* e apenas cinco (2,10%) entre os 238 produtos amplificados foram polimórficos dentro de cada grupo, demonstrando o alto grau de homozigose dessas linhagens. Esse resultado foi mantido quando amostras de DNA de doze indivíduos de cada grupo foram amplificadas. Esses parâmetros permitiram confirmar a escolha do par de genitores a fim de gerar uma população para mapeamento.

Termos de indexação: soja, diversidade genética, marcadores RAPD, métodos de estimação.

1. INTRODUCTION

It is well established that previous knowledge about the genetic relationships among breeding materials is crucial for the efficient use of the germ plasm in a breeding program (Nienhuis et al., 1994). In addition to the traits of interest, genetic distances based on the alleles distributed along the genome may be used to select genitors with maximum genetic divergence. There are several examples in crop plants showing that genetic distances based on molecular markers may be an important element to aid breeding programs (Hu & Quiros, 1991; Tinker et al., 1993; Abdelnoor et al., 1995). This should be useful not only for breeding but also for the development of segregating populations for mapping purposes.

Several types of molecular markers have been used to determine genetic divergence within and among plant species (Vierling & Nguyen, 1992; Abdelnoor et al., 1995). DNA-based markers have received increased attention due to the almost unlimited number, to the easiness of DNA extraction and its ready availability. Genetic distances based on molecular markers have been determined in both auto-

gamous (common bean) and allogamous (*Eucalyptus*) species, in cultures with broad as well as with narrow genetic bases (Paterson, 1998). However, in some cases, little or no attention has been paid to the suitability of the method used for genetic distance calculations.

We have used the RAPD technique (Williams et al., 1990) to determine the genetic distances among five soybean genotypes to choose the most divergent pair to constitute a population for soybean mapping. In addition, the degree of homozygosity within groups of plants of the most divergent pair was determined. Four different methods used to estimate the genetic distances were also tested.

2. MATERIAL AND METHODS

2.1. Genetic material

Soybean seeds from cultivars Paraná, Tropical, Cometa, and from lines UFV 91-717 (F7) and Ichigowase were sown in the greenhouse in 1994. Plants used in all experiments were derived from seeds collected from a single pod from each genotype.

2.2. DNA extraction and amplification

Genomic DNA was extracted from leaves by a CTAB method described by Abdelnoor et al. (1995). Amplification reactions (25 µL) contained the following components: 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2.8 mM MgCl₂, 0.1 mM of each dNTP, 0.4 µM of primer (Operon Technologies, Alameda, CA, USA), 1 unit of Taq DNA polymerase (Promega), and 25 ng DNA. Reactions were performed in a thermocycler (Perkin Elmer-Cetus 9600), programmed to run two cycles at 94°C for 2 min, 35°C for 1 min and 72°C for 1 min, two cycles at 94°C for 1 min, 35°C for 30's and 72°C for 1 min, followed by 40 cycles at 94°C for 15 s, 35°C for 30 s and 72°C for 1 min, and one final step at 72°C for 5 min. Amplification products were analyzed by electrophoresis in 1.2% agarose gels containing ethidium bromide.

2.3. Determination of genetic distances

Genetic distances between genotypes *i* and *j* (*Dij*) were based on the presence (1) or absence (0) of amplification products and were determined by four (A to D) different methods as defined below:

Sij = similarity coefficient;

a = number of matches 1,1;

b = number of matches 1,0;

c = number of matches 0,1; and

d = number of matches 0,0.

A) Arithmetical complement of simple matching coefficient (Sneath & Sokal, 1973):

$$D_{ij} = 1 - S_{ij} = 1 - \frac{a + d}{a + b + c + d}$$

B) Arithmetical complement of Nei and Li's similarity coefficient (Dice, 1945; Nei & Li, 1979):

$$D_{ij} = 1 - S_{ij} = 1 - \frac{2a}{2a + b + c}$$

C) Arithmetical complement of Gower's similarity coefficient (Gower, 1971) also cited as Jaccard's coefficient:

$$D_{ij} = 1 - S_{ij} = 1 - \frac{a}{a + b + c}$$

D) Modified Rogers' distance, MRD (Rogers, 1972; Sokal & Rohf, 1962):

$$MRD_{ij} = \sqrt{RD_{ij}}$$

where,

$$RD_{ij} = \frac{\sum_{k=1}^c (p_{ik} - p_{jk})^2}{2n}$$

and,

RDij = Rogers' dissimilarity coefficient;

pik = frequency of allele *k* in line *i*;

pjk = frequency of allele *k* in line *j*;

c = number of alleles in locus *k*; and

n = number of loci analyzed.

For comparison of lines with high degree of homozygosity, MRD_{ij} equals the square root of the complement of simple matching coefficient:

$$D_{ij} = \sqrt{\frac{b + c}{a + b + c + d}}$$

The genetic distances determined by different methods were compared by the Spearman correlation analysis (Cruz & Regazzi, 1994), using the statistical computer program Genes (Cruz, 1997).

2.4. Homozygosity analysis

To determine the degree of homozygosity within the genotypes, DNA samples from five randomly picked plants from each of the most divergent genotype pair were amplified with 28 different primers, and the genetic distances were calculated. Three independent DNA extractions were performed for each plant. In addition, DNA samples from 12 plants from each of the most divergent genotype pair were amplified with eight primers.

3. RESULTS

3.1. Genetic distances by four different methods

Polymorphisms among genotypes Paraná, Tropical, Cometa, UFV 91-717, and Ichigowase were detected by 36 out of 70 random primers. Figure 1 shows the amplification patterns obtained with two of these primers. Each primer amplified 1 to 12 bands; the number of polymorphic bands per primer varied between 1 and 5.

The 57 polymorphic and 133 monomorphic bands obtained were used to estimate the genetic distances among the genotypes. The largest genetic distance, by the four methods used, was between lines UFV 91-717 and Ichigowase, while the shortest one

was between Paraná and FT-Cometa. The Spearman correlation coefficients among the distances calculated by the four methods ranged from 0.99 to 1.00 (Table 1).

3.2. Homozygosity within lines UFV 91-717 and Ichigowase

DNA amplification of five plants each from lines UFV 91-717 and Ichigowase with 28 primers revealed 238 bands. Only five (2.10%) among all these bands revealed intra-varietal polymorphisms. These polymorphisms were maintained when amplifications were performed with DNA samples of three independent extractions from the same plant. The level of intra-varietal polymorphisms were maintained when DNA amplification of twelve plants each from lines

Table 1. Pairwise genetic distances among soybean genotypes estimated by the methods: simple match (A); Nei and Li (B); Gower (C); and MRD (D) and correlation values

Genotypes ⁽¹⁾	Methods			
	A	B	C	D
1 x 2	0.2053	0.1200	0.2143	0.0453
1 x 3	0.1263	0.0750	0.1395	0.0355
1 x 4	0.1684	0.0988	0.1798	0.0410
1 x 5	0.1105	0.0646	0.1214	0.0332
2 x 3	0.1632	0.0948	0.1732	0.0404
2 x 4	0.1632	0.0937	0.1713	0.0404
2 x 5	0.1579	0.0904	0.1657	0.0397
3 x 4	0.1579	0.0904	0.1685	0.0397
3 x 5	0.0895	0.0520	0.0988	0.0299
4 x 5	0.1105	0.0634	0.1193	0.0332
Correlations ⁽²⁾		$r_{AB} = 1.00$	–	–
		$r_{AC} = 0.99$	$r_{BC} = 0.99$	–
		$r_{AD} = 1.00$	$r_{BD} = 1.00$	$r_{CD} = 0.99$

⁽¹⁾ 1: UFV 91-717; 2: Ichigowase; 3: Paraná; 4: Tropical, and 5: Cometa.

⁽²⁾ (r_{ij}) Spearman ranking correlation.

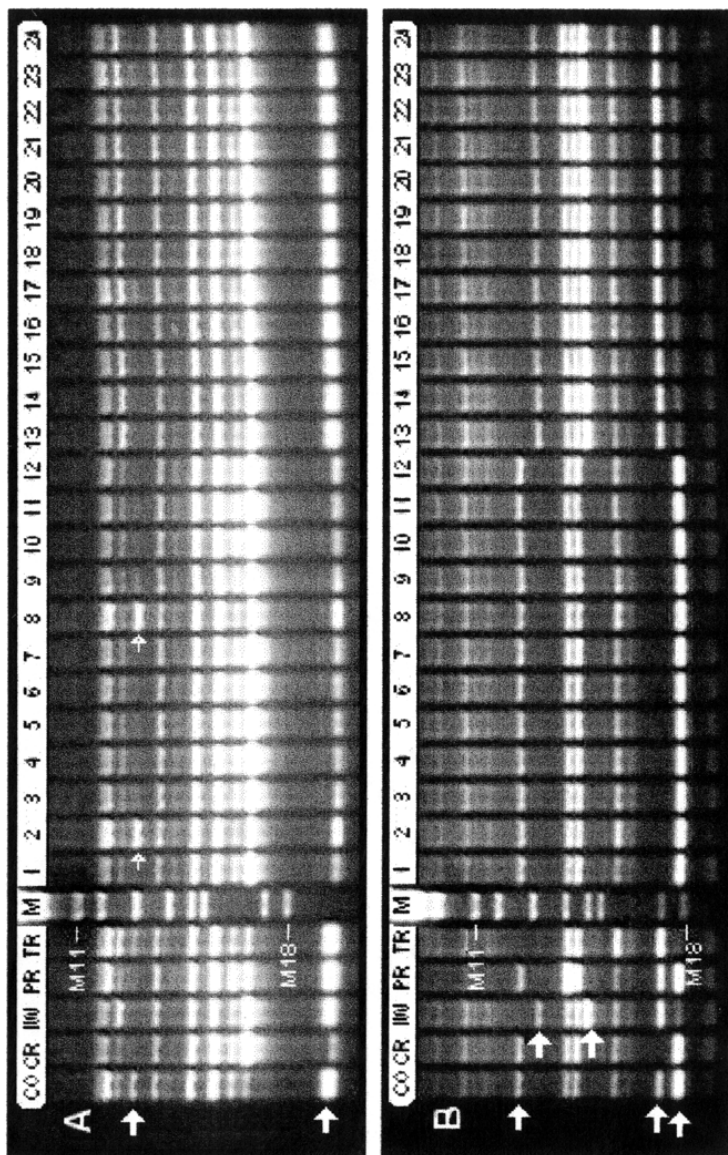


Fig. 1 Soybean RAPDs obtained with primers OPB10 (A) and OPI04 (B). Lanes are as follows: CO - Cometa, CR - UFV 91-717, IW - Ichigowase, PR - Paraná, TR - Tropical. M refers to lambda phage DNA digested with *EcoRI*, *HindIII*, and *BamHI* enzymes (size markers), M11 = 1,584 bp and M18 = 493 bp. Lanes 1-12 (UFV 91-717), 13-24 (Ichigowase). Arrows indicate intravarietal (small arrow) and intervarietal (large arrow) heteromorphisms.

UFV 91-717 and Ichigowase were performed with eight primers (only one intra-varietal polymorphic band among all 47 bands generated by these primers). Figure 1 shows the amplification patterns obtained with two of these primers.

4. DISCUSSION

4.1. Genetic distance

The basic difference between the four procedures used to estimate genetic distances is the inclusion of (0,0) as indicative of similarity in the simple match and MRD methods. Additionally, unlike the other three methods, in the method of Dice matchings of the type (1,1) have a greater weighting factor.

Estimation of genetic distances by the MRD method requires allelic information and, according to Skroch et al. (1992), this method should not be used for RAPD data. However, for comparison of autogamous species, with a high degree of homozygosity, it can be assumed that the allelic frequency is either 1 or 0, and MRD may be calculated by the square root of the simple match coefficient.

In general, the genetic distances among the five soybean genotypes calculated by the four procedures were similar as indicated by the Spearman correlation coefficient (0.99 to 1.00) (Table 1). This is in agreement with mathematical expectations: the simple match statistic and the Roger's distance (MRD) can be easily transformed one to the other when allelic frequencies are 1 or 0; Gower's statistic as compared to Nei and Li differ only in the weight assigned to situations in which both varieties lack a marker. These correlations seem to be valid only for materials with a short genetic base such as cultivated soybean. However, this situation does not seem to hold for cross pollinating species which normally present a broader genetic base. For instance, according to Skroch et al. (1992) matchings (0,0) can be considered for determination of intra-specific genetic distances if the number of markers is sufficiently high, because these matchings are normally indicative of similarity.

According to Dudley (1994) these matchings can be considered only when two alleles exist in a locus, and both alleles are present in the materials being compared. For this reason, the simple matching and MRD methods should be preferred for the determination of intra-specific genetic distances. On the other hand, for determination of interspecific genetic distances the methods of Nei and Li, and Gower should be used. For cultures with broad genetic bases the method of Nei and Li should be preferred because matchings (1,1) have a greater weighting factor which makes differences more evident.

Although some authors have used the simple matching coefficient to calculate interspecific genetic distances (Abo-Elwafa et al., 1995), we propose that in these cases the methods of Nei and Li, and Gower should be used once these procedures disregard (0,0) as indicative of similarity.

4.2. Homozygosity within lines UFV 91-717 and Ichigowase

Despite the low number of individuals used, our data indicated that the genotypes analyzed are highly homozygous. The number of individuals is comparable to the one used by Spooner et al. (1996) in a similar type of analysis. Our results are in accordance with what is expected from the genealogy of the genetic material used (Abdelnoor et al., 1995). Several investigations have demonstrated the usefulness of RAPD for cultivar identification and genetic diversity determination. Hu & Quiros (1991), for instance, showed that four primers could be used to distinguish 14 brocoli and 12 cauliflower cultivars. Tinker et al. (1993) were able to distinguish barley lines with parent coefficient of 0.92 with 19 RAPD primers, demonstrating that similarity prediction by molecular markers can be more informative than relationships deduced by pedigree analysis only.

It is normally recommended that during determination of interpopulational genetic diversity bulks of individuals should be used (Spooner et al., 1996). However, in our case, due to the high degree of homozygosity of the genotypes, which was

demonstrated by the molecular analysis, one single individual should be enough to represent the whole line and consequently differences between single individuals constitute an accurate measure of the distances among genotypes.

5. CONCLUSIONS

1. The largest genetic distance, by the four methods used, was between lines UFV 91-717 and Ichigowase, while the shortest one was between Paraná and FT-Cometa;

2. Lines UFV 91-717 and Ichigowase were chosen to generate the segregating population to construct a soybean intraspecific genetic map;

3. Lines UFV 91-717 and Ichigowase presented high degree of intra-varietal homozygosity. Therefore, one single individual should be enough to represent the whole line; and

4. All four methods were equivalent to calculate genetic distances among and within soybean lines.

ACKNOWLEDGMENTS

This research was supported by a grant from PADCT/FINEP. R.X.C. was the recipient of a fellowship from CNPq.

REFERENCES

- ABDELNOOR, R.V.; BARROS, E.G. & MOREIRA, M.A. Determination of genetic diversity within Brazilian soybean germplasm using random amplified polymorphic DNA techniques and comparative analysis with pedigree data. *Brazilian Journal of Genetics*, Ribeirão Preto, **18**:265-273, 1995.
- ABO-ELWAFI, A.; MURAI, K. & SHIMADA, T. Intra- and inter-specific variations in Lens revealed by RAPD markers. *Theoretical and Applied Genetics*, Berlin, **90**:335-340, 1995.
- CRUZ, C.D. *Programa Genes: aplicativo computacional em genética e estatística*. Viçosa, Universidade Federal de Viçosa, 1997.
- CRUZ, C.D. & REGAZZI, A.J. *Modelos biométricos aplicados ao melhoramento genético*. Viçosa, Universidade Federal de Viçosa, 1994.
- DICE, L.R. Measures of the amount of ecologic association between species. *Ecology*, Massachusetts, **26**:297-302, 1945.
- DUDLEY, J.W. Comparison of genetic distance estimators using molecular markers data. In: ANALYSIS OF MOLECULAR MARKER DATA. Corvallis, Oregon, 1994. *Proceedings*. Corvallis, American Society for Horticultural Science and Crop Science Society of America, 1994. p. 3-7.
- GOWER, J.C. A general coefficient of similarity and some of its properties. *Biometrics*, Washington, **27**:857-874, 1971.
- HU, J. & QUIROS, C.F. Identification of broccoli and cauliflower cultivars with RAPD markers. *Plant Cell Report*, Berlin, **10**:505-511, 1991.
- LINK, W.; DIXKENS, C.; SINGH, M.; SCHWALL, M. & MELCHINGER, A.E. Genetic diversity in European and Mediterranean faba bean germ plasm revealed by RAPD markers. *Theoretical and Applied Genetics*, Berlin, **90**:27-32, 1995.
- NEI, M. & LI, W.H. Mathematical model for studying genetic variations in terms of restriction endonucleases. *Proceedings of the National Academy of Science*, Washington, **76**:5269-5273, 1979.
- NIENHUIS, J.; TIVANG, J. & SKROCH, P. Analysis of genetic relationships among genotypes based on molecular markers data. In: ANALYSIS OF MOLECULAR MARKER DATA. Corvallis, Oregon, 1994. *Proceedings*. Corvallis, American Society for Horticultural Science and Crop Science Society of America, 1994. p. 8-14.
- PATERSON, A.H. *Molecular dissection of complex traits*. New York, CRC, 1998. 305p.
- ROGERS, J.S. *Measures of genetic similarity and genetic distance: studies in genetics*. VII. University of Texas, Texas, **7213**:145-153, 1972.
- SKROCH, P.; TIVANG, J. & NIENHUIS, J. Analysis of genetic relationships using RAPD marker data. In: APPLICATIONS OF RAPD TECHNOLOGY TO PLANT BREEDING. Minneapolis, 1992. *Proceedings*. Minneapolis, Crop Science Society of America, American Society for Horticultural Science and American Genetic Association, 1992. p.26-30.

- SNEATH, P.H.A. & SOKAL, R.R. *Numerical taxonomy*. San Francisco, Freeman, 1973.
- SOKAL, R.R. & ROHF, F.J. The comparison of dendrograms by objective methods. *Taxon*, Washington, **11**:33-40, 1962.
- SPOONER, M.M.; TIVANG, J.; NIENHUIS, J.; MILLER, J.T.; DOUCHES, D.S. & CONTRERAS-M.A. Comparison of four molecular markers in measuring relationships among the wild potato relatives *Solanum* section *Etuberosum* (subgenus Potatoe). *Theoretical and Applied Genetics*, Berlin, **92**:532-540, 1996.
- TINKER, N.A.; FORTIN, M.G. & MATHER, D.E. Random amplified polymorphic DNA and pedigree relationships in spring barley. *Theoretical and Applied Genetics*, Berlin, **85**:976-984, 1993.
- VIERLING, R.A. & NGUYEN, H.T. Use of RAPD markers to determine the genetic diversity of diploid wheat genotypes. *Theoretical and Applied Genetics*, Berlin, **84**:835-838, 1992.
- WILLIAMS, J.G.K.; KUBELIK, A.R.; LIVAK, K.J.; RAFALSKI, J.A. & TINGEY, S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, Oxford, **18**:6531-6535, 1990.