

DETERMINAÇÃO DO NÚMERO DE MARCADORES RAPD PARA ESTUDOS DA DIVERSIDADE GENÉTICA EM SOJA UTILIZANDO O MÉTODO BOOTSTRAP

BOOTSTRAP DETERMINATION OF THE NUMBER OF RAPD MARKERS FOR STUDIES OF GENETIC DIVERSITY IN SOYBEAN

Susete Araújo PEQUENO¹; José Baldin PINHEIRO^{2*}; Maria Imaculada ZUCCHI³; Roland VENCovsky⁴; Alexandre Siqueira Guedes COELHO⁵; Maria da Glória TRINDADE⁶

RESUMO: Vinte e seis cultivares de soja pertencentes ao grupo de maturação tardio foram avaliados com base em 85 locos RAPD, obtidos de 27 *primers*, submetendo-se os dados obtidos à técnica de bootstrap para determinar o número mínimo de locos necessários para se ter uma boa precisão nos estudos de diversidade genética. O procedimento de *bootstrap* dos locos foi utilizado para estimar o coeficiente de variação (CV) do índice de similaridade de Jaccard, utilizando-se 1000 permutações. O número mínimo de locos foi encontrado através da expressão $x_c = [a^2b^2(2b-1)/(b-2)]^{1/(2-2b)}$ onde x_c é o ponto de máxima curvatura da função, e corresponde à quantidade mínima de locos a ser amostrado no genoma. Foi obtido um coeficiente de determinação de 0,98, indicando um bom ajuste da curva estimada aos dados observados. Para este conjunto de dados, o ponto de curvatura máxima foi $x_c = 12,41$, que corresponde a um CV de 33 %, e indica que o uso de 12 a 13 locos marcadores proporcionaria a precisão obtida para as estimativas. O parâmetro *b* da equação representa uma medida da heterogeneidade entre bandas. O valor encontrado neste estudo, de $b = 0,426$, mostra uma alta heterogeneidade entre bandas, já que o valor máximo esperado para este parâmetro é de 0,5.

UNITERMOS: RAPD, *Glycine max*, Diversidade genética

INTRODUÇÃO

Atualmente existem diversas técnicas moleculares disponíveis para a realização de estudos da diversidade genética, tanto de espécies nativas quanto de espécies cultivadas. Nestas, o estudo com base em marcadores moleculares é particularmente útil, pois a maioria das espécies cultivadas vem apresentando um estreitamento da base genética, tornando-se cada vez maior a necessidade de utilização de técnicas que detectem uma quantidade maior de polimorfismos. No caso da soja esse estreitamento da base genética deveu-se à utilização repetida de pequeno número de genótipos

como genitores nas hibridações e à subsequente endogamia sofrida ao longo das sucessivas gerações de autofecundação. De acordo com Hiromoto e Vello (1986), esses eventos podem acarretar problemas como a vulnerabilidade genética e a dificuldade em melhorar o nível atual de produtividade de grãos.

A escolha do marcador a ser utilizado em estudos dessa natureza depende de vários critérios, como a praticidade e custo da técnica, bem como do objetivo do estudo. Neste sentido, os marcadores moleculares do tipo RAPD têm-se constituído numa ferramenta importante e amplamente utilizada para estes estudos.

Existem vários estudos na literatura utilizando

¹ Eng. Agr^o, Mestranda, Escola de Agronomia-UFG, Campus Samambaia, susetepequeno@bol.com.br

² Eng. Agr^o, Dr, Professor Adjunto, Escola de Agronomia-UFG, Campus Samambaia, baldin@agro.ufg.br. Autor correspondente.

³ Bióloga, Ms., Doutoranda, ESALQ-USP, Departamento de Genética – ESALQ/USP, mizucchi@carpa.ciagri.usp.br

⁴ Eng. Agr^o, Dr, Professor Visitante, Escola de Agronomia-UFG, rvencovs@esalq.usp.br

⁵ Eng. Agr^o, Doutorando, Professor Assistente, ICB-UFG, Instituto de Ciências Biológicas, Universidade Federal de Goiás –Campus Samambaia, acoelho@icb1.ufg.br

⁶ Eng. Agr^o, Mestranda Escola de Agronomia-UFG, Campus Samambaia, mgloria@agro.ufg.br

Received: 30/08/02 Accept: 20/12/02

marcadores RAPD para a análise da variabilidade e caracterização genética de cultivares (FRANCO et al., 2001; MÜHLEN et al., 2000). No entanto, um aspecto que deve ser considerado é a quantidade de marcas necessárias para amostrar satisfatoriamente o genoma da espécie em questão. Um número pequeno de marcas pode comprometer a precisão das inferências a serem feitas, entretanto, a utilização de um número muito elevado de locos marcadores pode inviabilizar o estudo de uma quantidade grande de materiais, além de aumentar os custos da pesquisa. Para se determinar o número mínimo de locos marcadores que irá fornecer uma boa precisão das estimativas obtidas a partir dos mesmos, têm-se utilizado, mais recentemente, o método *bootstrap*. Este método baseia-se na reamostragem de dados reais para revelar algum padrão neles presente. Desta forma, o objetivo deste trabalho foi determinar o número mínimo de locos marcadores RAPD para avaliar a diversidade genética em cultivares de soja de ciclo tardio.

MATERIAL E MÉTODOS

Vinte e seis cultivares de soja pertencentes ao grupo de maturação tardio foram avaliados com base em 85 locos RAPD, obtidos de 27 *primers*, submetendo-se os dados obtidos à técnica de *bootstrap* para determinar o número mínimo de locos necessários para se ter uma boa precisão nos estudos de diversidade genética. A extração de DNA foi realizada de acordo com o método CTAB descrito por Murray e Thompson (1980) e Rogers e Bendich (1985), com modificações. O DNA foi precipitado com isopropanol após lavagem com etanol e ressuspendido em tampão TE. O DNA extraído foi quantificado através de eletroforese em gel de agarose e comparado com DNA de fago λ de concentração conhecida e diluído à concentração de 10 ng/ μ L. Foi estabelecido um protocolo utilizando-se nas reações de amplificação 25 μ L contendo: 10 mM Tris-HCl pH 8,3; 50 mM de KCl; 5,0 mM MgCl₂; 1,25 mM de cada dNTP; 30 ng de *primer* (*Operon Technologies*), 30 ng de DNA molde, 1 unidade de Taq Polimerase e H₂O q.s.p.. As reações foram submetidas a amplificação utilizando-se uma etapa inicial de 5 minutos a 94°C, seguida de 48 ciclos da sequência: 30 segundos a 92°C, 1 minuto e meio a 37°C e 1 minuto e meio a 72°C. Ao final dos 48 ciclos foi realizada uma etapa final de extensão de 5 minutos a 72°C. Todos os indivíduos foram genotipados quanto à presença (1) ou ausência (0) de banda para todos os *primers* utilizados.

O procedimento de *bootstrap* dos locos foi utilizado para estimar o coeficiente de variação (CV) do índice de similaridade de Jaccard, utilizando-se 1000 permutações, de acordo com procedimento descrito por Tivang et al. (1994). A análise foi realizada através do software Dboot, desenvolvido por Coelho (2000). Encontrou-se uma função exponencial do tipo $y = ax^b$ que melhor se ajustasse aos dados, e o número mínimo de locos foi encontrado através da expressão $x_c = [a^2b^2(2b-1)/(b-2)]^{1/(2-2b)}$ onde x_c é o ponto de máxima curvatura da função, e corresponde à quantidade mínima de locos a ser amostrado no genoma, para o conjunto de genótipos analisados, porém nem sempre considerada ideal.

RESUSTADOS E DISCUSSÃO

Na Figura 1 estão representados os valores de CV observados, bem como a curva ajustada de função $y = 97,237 x^{-0,426}$. Foi obtido um coeficiente de determinação de 0,98, indicando um bom ajuste da curva estimada aos dados observados. Pode-se observar, através da Figura 1, que à medida que o número de locos aumenta, diminui o CV, como resultado de uma melhor amostragem do genoma. No entanto, chega-se a um ponto em que o aumento do número de marcas proporciona uma diminuição muito pequena no coeficiente de variação, contribuindo pouco para a melhoria da precisão das estimativas baseadas nesses marcadores. Isto deve-se ao comportamento assintótico da curva em relação ao eixo X. Uma maneira de se determinar o número mínimo de marcas necessárias para se obter uma precisão satisfatória das estimativas obtidas a partir desses dados é o cálculo do ponto de curvatura máxima da função. Para este conjunto de dados, o ponto de curvatura máxima foi $x_c = 12,41$, que corresponde a um CV de 33 %, e indica que o uso de 12 a 13 locos marcadores proporcionaria a precisão obtida para as estimativas. A obtenção da função que melhor se ajusta ao conjunto de dados permite a escolha, por parte do pesquisador, de um número de marcas que lhe proporcione uma precisão desejada. Assim, para se obter um CV de 15% seria necessário o uso de 80 locos RAPD, e para um CV de 10% seriam necessários 209 locos, considerando-se este conjunto de dados obtidos. O parâmetro *b* da equação representa uma medida da heterogeneidade entre bandas. O valor encontrado neste estudo, de $b = 0,426$, mostra uma alta heterogeneidade entre bandas, já que o valor máximo esperado para este parâmetro é de 0,5.

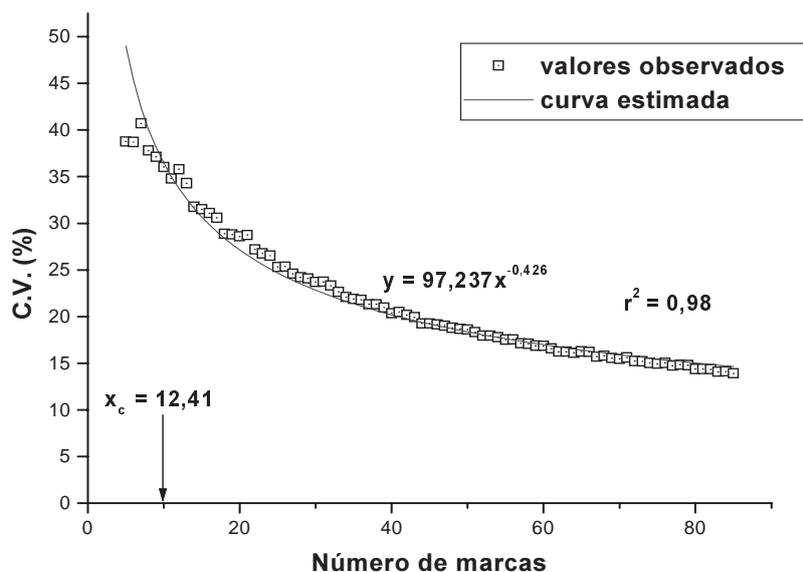


Figura 1. Valores observados e estimados de CV (%) em função no número de marcas RAPD em soja. x_c representa o ponto de máxima curvatura da função.

CONCLUSÕES

Esse tipo de estudo permitiu um dimensionamento adequado do esforço e dos recursos a serem empregados na obtenção dos dados moleculares, uma vez que a obtenção de um número muito elevado de marcas poderia não levar a um aumento significativo na precisão, aumentando-se apenas os custos da pesquisa. Por outro lado, a utilização de um número insuficiente de marcas

pode não amostrar satisfatoriamente o genoma da espécie em questão, resultando em pouca precisão das estimativas obtidas a partir dos dados moleculares.

AGRADECIMENTOS

Os autores agradecem ao CNPq pelo suporte financeiro a esta pesquisa.

ABSTRACT: Twenty-six soybean cultivars belonging to the late maturity group were investigated on the basis of 85 RAPD loci, obtained from 27 primers, submitting the data obtained to the bootstrap technique to determine the minimum number of loci necessary to have a good precision in genetic diversity studies. The bootstrap procedure over loci was used to estimate the coefficient of variation (CV) of Jaccard's similarity index through 1000 permutations. The minimum number of loci was found with the expression $x_c = [a^2b^2(2b-1)/(b-2)]^{1/(2-2b)}$ where x_c is the point of maximum curvature of the function, and corresponds to the minimum number of loci to be sampled in the genome. A coefficient of determination of 0.98 was obtained indicating a good adjustment of the estimated curve to the observed data. For this group of data, the point of maximum curvature was $x_c = 12.41$, which corresponds to a CV of 33%, indicating that the use from 12 to 13 locus markers would provide the precision obtained for the estimates. Parameter b of the equation represents a measure of the heterogeneity among bands. The value found in this study, $b = 0.426$, shows a high heterogeneity among bands, since the expected maximum value for this parameter is 0.5.

UNITERMS: RAPD, *Glycine max*, Genetic diversity

REFERÊNCIAS BIBLIOGRÁFICAS

COELHO, A. S. G. **Dboot**: avaliação dos erros associados a estimativas de distâncias/similaridades genéticas através do procedimento de bootstrap com número variável de marcadores (software). [S.l.]: Laboratório de Genética Vegetal, Instituto de Ciências Biológicas, UFG, 2000.

FRANCO, M. C.; CASSINI, S. T. A.; OLIVEIRA, V. R.; TSAI, S. M. Caracterização da diversidade genética em feijão por meio de marcadores RAPD. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 36, n. 2, p. 381-385, fev. 2001.

HIROMOTO, D. M.; VELLO, N. A. The genetic base of Brazilian soybean (*Glycine max* (L.) Merrill) cultivars. **Revista Brasileira de Genética**, Ribeirão Preto, v. 9, n. 2, p. 295-306, jun. 1986.

MÜHLEN, G. S.; MARTINS, P. S.; ANDO, A. Variabilidade genética de etnovarietades de mandioca, avaliada por marcadores de DNA. **Scientia Agricola**, Piracicaba, v. 57, n. 2, p. 319-328, abr./jun. 2000.

MURRAY, M. G.; THOMPSON, W. F. Rapid isolation of high molecular weight plant DNA. **Nucleic Acids Research**, Oxford, v. 8, n. 19, p. 4321-4325, oct. 1980.

ROGERS, S. O.; BENDICH, A. J. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plants tissues. **Plant Molecular Biology**, Dordrecht, v. 5, n. 2, p. 69-76, june. 1985.

TIVANG, J. G.; NIENHIUS, J.; SMITH, O. S. Estimation of sampling variance of molecular marker data using the bootstrap procedure. **Theoretical Applied Genetics**, Berlim, v. 89, n. 2/3, p. 259-264, oct. 1994.