

# ÍNDICE BASEADO EM RFLPs PARA SELEÇÃO DE LINHAGENS VISANDO SINTÉTICOS DE MILHO<sup>1</sup>

Glauce Cristina Ricardo Rumin<sup>2\*</sup>; Roland Vencovsky<sup>3</sup>

<sup>2</sup>Pós-Graduada do Depto. de Genética - USP/ESALQ.

<sup>3</sup>Depto de Genética - USP/ESALQ, C.P. 83 - CEP: 13400-970 - Piracicaba, SP.

\*Autor correspondente <grumin@terra.com.br>

**RESUMO:** Marcadores moleculares têm sido sugeridos como ferramentas úteis em programas de melhoramento de plantas. Assim, este trabalho teve como objetivo propor um índice de seleção de linhagens visando a geração de populações sintéticas de milho, baseado em marcadores moleculares e dados de campo. Para tanto foram utilizados valores fenotípicos e genotípicos, oriundos de experimentos convencionais de campo e através da genotipagem das plantas-mãe das linhagens por marcadores RFLP, respectivamente. O índice foi examinado quanto à sua capacidade de originar sintéticos com alta frequência de alelos favoráveis, em comparação com os obtidos baseando-se na seleção convencional. Utilizaram-se inicialmente 157 locos marcadores com distância média de 15,30 cM entre eles. Sessenta e oito linhagens S<sub>2</sub>, provenientes da geração F<sub>2</sub> do cruzamento entre duas linhagens homocigóticas foram cruzadas em *topcross* e avaliadas em quatro locais de Iowa, EUA, em 1996. Dos 157 locos foram selecionados 18 que explicaram 74,70% da variância genética entre progênies, na média dos locais. O índice proposto leva em conta o comportamento próprio (*per se*) das linhagens e a complementaridade genotípica entre elas, o que não é conseguido utilizando-se apenas dados dos *topcrosses*. O uso do índice proposto permite obter homólogo de tabela dialélica tendo-se apenas dados de médias de *topcrosses*. O índice levou à seleção de sintéticos com propriedades superiores às esperadas pela seleção baseada apenas nas médias das progênies *topcross*. Essa conclusão, no entanto, só é válida para os QTL's abrangidos pelas marcas. A metodologia proposta permite monitoramento muito adequado e útil das propriedades genéticas dos sintéticos a serem obtidos.

**Palavras-chave:** *Zea mays*, marcadores RFLP, linhagem endogâmica, índice de seleção, sintéticos

## RFLP MARKER INDEX FOR SELECTION OF INBRED LINES AIMED AT SYNTHETIC MAIZE POPULATIONS

**ABSTRACT:** Molecular markers have been suggested as useful tools in breeding programmes. The purpose of this work was to propose and apply an index for inbred line selection to be used for generating synthetic maize populations. It is based on field data and the RFLP genotypes of seed parents. Index usefulness was examined, considering its efficiency for generating synthetics with high QTL frequencies, in contrast with line selection based solely on topcross performance. Seed parents were genotyped for 157 RFLP marker loci with an average distance between loci of 15.30 cM. Sixty-eight S<sub>2</sub> lines, stemming from an F<sub>2</sub> population derived from a cross between two homozygous inbred lines, were topcrossed and evaluated in four locations in Iowa, USA, in 1996. From these 157 loci, 18 were identified as significantly associated with yield QTLs explaining 74.70 % of the genetic variance among topcross progenies. The proposed index included measures which are function of the *per se* line behaviour and the genetic complementarity between lines, and led to a diallel table homologue for all parental lines. These properties cannot be accomplished using only the topcross selection criterion. The index led to synthetic populations which are expected to be superior to the best synthetic population selected by topcross performance alone. However, this superiority is valid only for QTL regions covered by the markers used. This procedure allowed a useful monitoring of the genetic constitution of all possible synthetic populations.

**Key words:** *Zea mays*, RFLP marker, inbred lines, selection index, synthetic populations

## INTRODUÇÃO

A síntese de populações é uma etapa importante num programa de melhoramento de híbridos de milho. Em geral essas populações devem apresentar alta frequência de alelos favoráveis para servirem como fonte de extração de genótipos superiores. Entretanto, a avaliação dos genótipos e a confiabilidade dos dados experimentais num programa de

melhoramento de plantas é um grande desafio, principalmente para características de herança complexa, como a produtividade de grãos. Assim, esforços têm sido feitos para integrar algumas técnicas de biologia molecular aos procedimentos convencionais de melhoramento de plantas, uma vez que diversos estudos demonstraram a sua aplicabilidade e eficiência na avaliação desses caracteres.

<sup>1</sup>Parte da Dissertação de Mestrado da primeira autora apresentada à USP/ESALQ - Piracicaba, SP.

Os marcadores moleculares têm sido sugeridos como ferramenta útil em vários aspectos do melhoramento de plantas, como a descrição de variedades, a construção de mapas genéticos, a medição das distâncias genéticas entre as linhagens e o próprio processo de seleção.

Uma das possibilidades previstas com o advento dos marcadores moleculares foi a predição do desempenho de híbridos simples baseado na genotipagem das linhagens genitoras. Assim, vários estudos foram conduzidos com o objetivo de verificar correlações entre os genótipos dos pais, fornecidos pelos marcadores moleculares, e o comportamento de seus híbridos em testes de campo (Godshalk et al., 1990; Melchinger et al., 1990; Dudley et al., 1991; Messmer et al., 1993; Stuber, 1994; Dubreuil et al., 1996; Lanza et al., 1997). De uma forma geral, os autores não obtiveram sucesso na predição do comportamento do híbrido a partir do genótipo das linhagens, embora concordem que os marcadores moleculares são eficientes na determinação de distâncias genéticas entre linhagens e em sua ordenação em grupos heteróticos distintos.

Outra vantagem sugerida foi a diminuição do tempo necessário no processo de seleção devido ao enfoque direto nas regiões de interesse, o que teria importância ainda maior para caracteres de herdabilidade baixa. No caso de retrocruzamentos, a seleção assistida possibilita também o controle da incorporação de segmentos não desejáveis (Ragot et al., 1995). Diversos trabalhos foram conduzidos para avaliar a eficiência dos marcadores em auxiliar a seleção nos retrocruzamentos com resultados positivos (Hospital et al., 1992; Stuber, 1994; Beavis et al., 1994 e Openshaw et al., 1994). Selig et al. (1999) enfatiza que a transferência de germoplasma exótico num programa de melhoramento pode ser otimizada com o uso de marcadores moleculares.

Com relação à eficiência da seleção assistida por marcadores, deve-se destacar que a diminuição no desequilíbrio de ligação com o passar dos ciclos é um fato que prejudica o processo de seleção. Bearzoti (1997) trabalhou com seleção recorrente assistida por marcadores em espécies autógamas através de simulações, e verificou que esse tipo de seleção foi pouco eficiente por mais de um ciclo seletivo, apesar da eficiência constatada no primeiro ciclo de seleção. Num trabalho de seleção recorrente em milho, Edwards & Johnson (1994) compararam a seleção assistida por RFLPs com a seleção fenotípica convencional e observaram que, para caracteres de herdabilidade baixa, como é o caso da produtividade de grãos, a seleção assistida conduziu a progressos superiores em relação à seleção convencional. Porém, o inverso ocorreu para caracteres com alta herdabilidade, no caso a qualidade de grãos. Já Stromberg et al. (1994) não verificaram

diferenças significativas da seleção assistida por marcadores quando comparada ao melhoramento convencional, e uma das possíveis razões para este fato foi a baixa cobertura do genoma.

Outro ponto importante no uso da seleção assistida por marcadores é a identificação de associações específicas marcador - QTL significativas, tanto para características agrônômicas em milho (Goldman et al., 1994; Berke & Rocheford, 1995) como para resistência a doenças (Bubeck et al., 1993; Ogliari, 1999). Para que a seleção assistida por marcadores seja eficiente, é necessário inicialmente a saturação do genoma pelos marcadores, além de se considerar (1) o número de QTL's que afetam o caráter (Dudley, 1993), (2) o tamanho da população utilizada nos estudos (Moreau et al., 1999; Ajmone-Marsan et al., 1996), (3) o efeito do ambiente (Jansen et al., 1995; Schön et al., 1994) e (4) o tipo de população utilizada (Kerns et al., 1999).

Ao se testar a significância das associações marcador - QTL pode-se fazer uso de um teste estatístico em cada loco, como o teste t, ou se optar por outros métodos de abordagem, como o mapeamento por intervalo, em suas várias versões. Para Matioli<sup>1</sup> (1999) os métodos de abordagem múltipla são mais recomendados, pois o poder de análise é aumentado com a utilização de todos os locos simultaneamente. O nível de significância utilizado na detecção de associações significativas marcador - QTL varia conforme o autor e o caráter avaliada. Para maiores informações ver Godshalk et al. (1990) e Bubeck et al. (1993).

## MATERIAL E MÉTODOS

Os dados utilizados neste estudo foram gerados pelo departamento de pesquisa e desenvolvimento da empresa Garst Seeds, componente do grupo Zeneca, com base em Slater, Iowa, EUA, no ano agrícola de 1996. Inicialmente, foi produzida população base através do cruzamento de duas linhagens homocigóticas pertencentes ao grupo Iowa Stiff Stalk Synthetic, gerando os indivíduos da geração  $F_1$ , que foram cruzados ao acaso para originar a geração  $F_2$ . Da geração  $F_2$  obtida deste cruzamento, extraíram-se plantas que foram autofecundadas gerando as progênies  $S_1$ . Essas progênies  $S_1$  foram autofecundadas sem controle genealógico e geraram as 68 progênies  $S_2$  utilizadas neste estudo.

As progênies  $S_2$  foram avaliadas em *topcrosses* representando o genitor feminino. Como testador, ou genitor masculino, utilizou-se uma linhagem endogâmica pertencente ao grupo heterótico Lancaster. Os experimentos ainda continham híbridos simples comerciais da empresa Garst Seeds como testemunhas

<sup>1</sup>MATIOLI, S.R. (Instituto de Ciências Biológicas, Universidade de São Paulo). Anotações de Curso, 1999.

ou tratamentos comuns. Cada genitor feminino foi genotipado por RFLPs e, para isso, amostras de DNA foram coletadas das plântulas de milho no estágio de duas folhas.

No ano agrícola de 1996, as 68 progênies em *topcrosses* foram avaliadas em quatro locais diferentes, que se concentraram no estado de Iowa, EUA, na região do *Corn Belt* americano. O delineamento experimental utilizado foi em blocos incompletos com testemunhas comuns, como sugerido por Gomes (1990), por possibilitar a comparação entre progênies presentes nos diferentes blocos. Cada bloco conteve tratamentos regulares (cruzamentos entre as progênies S<sub>2</sub> e o testador) e tratamentos comuns ou testemunhas (híbridos simples comerciais da empresa Garst Seeds). Cada experimento teve duas repetições e três blocos incompletos com oito testemunhas cada um. O número de progênies presentes nos blocos foi de 30, 20 e 18, para os blocos um, dois e três, respectivamente, de cada repetição.

A característica medida nos experimentos e utilizada neste trabalho foi a produtividade de grãos, mensurada em kg ha<sup>-1</sup>. A densidade de plantio utilizada foi de 24.000 plantas por acre, o que corresponde a 59.304 plantas por hectare, de acordo com os padrões da empresa Garst Seeds.

As análises de variância, segundo o delineamento em blocos incompletos, foram executadas com o programa SAS (SAS Institute, 1992), com os valores de produtividade de grãos coletados por parcela em cada um dos quatro locais. Dessa forma foram obtidas as médias ajustadas das progênies para a análise conjunta. Essas foram utilizadas em todos os cálculos posteriores envolvendo as marcas.

Os procedimentos de laboratório foram realizados pelo departamento de pesquisas e desenvolvimento da empresa Garst Seeds, com base em Slater, Iowa, EUA, utilizando o protocolo GENTRA, patenteado pela empresa, e a genotipagem de RFLPs foi realizada para cada genitor feminino das linhagens. Para a combinação sonda - enzima, usaram-se as enzimas de restrição Eco RI e Hind III. As análises de eletroforese foram realizadas em gel de agarose a 1%, em instrumentação Pharmacia, a 85 V por duas horas, e os *southern blots* feitos imediatamente a seguir, em equipamento Pharmacia, transferindo-se o DNA para membranas Hybond-N. As sondas foram preparadas com marcação radioativa (<sup>32</sup>P). A fixação do DNA nas membranas, pré-hibridação, hibridação com as sondas marcadas e revelação foram realizadas segundo os procedimentos propostos em Current Protocols in Molecular Biology (Chory et al., 1996). A genotipagem foi realizada manualmente, atribuindo-se códigos de letras ao polimorfismo de RFLPs detectado. Os dados foram então transferidos a uma base de dados para que se executassem os cálculos relativos à análise genética.

Os alelos favoráveis associados às marcas foram identificados em cada loco marcador separadamente, considerando todas as progênies avaliadas. Inicialmente foi preciso definir a média de produtividade de grãos, em kg ha<sup>-1</sup>, das progênies que apresentavam o mesmo genótipo. Apenas os genitores homocigóticos foram considerados nestes cálculos, feitos com os dados da genotipagem e com as médias ajustadas das progênies. Definida a média em kg ha<sup>-1</sup> dos indivíduos homocigóticos para as marcas RFLP, passou-se à identificação dos homocigotos favoráveis. Para isso, foi feito um contraste entre as médias dos homocigotos, conforme a definição de Falconer & MacKay (1996). Considerando o loco A, com dois alelos marcadores A<sub>1</sub> e A<sub>2</sub>, o desvio genotípico é representado por:

$$"a" = (\bar{Y}_{A_1A_1} - \bar{Y}_{A_2A_2})/2 \quad \text{sendo,}$$

"a" = desvio genotípico para o loco marcador em questão,  $\bar{Y}_{A_1A_1}$  = média ajustada dos indivíduos com o genótipo marcador A<sub>1</sub>A<sub>1</sub>, e  $\bar{Y}_{A_2A_2}$  = média ajustada dos indivíduos com o genótipo marcador A<sub>2</sub>A<sub>2</sub>.

Quando o homocigoto contrastado apresentou um valor "a" positivo, foi considerado favorável. Em caso de valor "a" negativo, o homocigoto foi classificado como desfavorável. Uma vez definidos quais eram os homocigotos favoráveis e desfavoráveis, os locos foram codificados para as análises de regressão posteriores. Ao homocigoto favorável atribuiu-se o valor 1, e o homocigoto desfavorável recebeu o valor -1. Quando o genótipo apresentava apenas uma cópia do alelo favorável recebeu o valor 0.

O cálculo dos valores genotípicos foi executado conforme demonstrado acima nos locos que apresentavam apenas dois alelos, considerando-se como alelo favorável aquele cujo homocigoto apresentava um valor "a" positivo. Porém, cerca de 30% dos locos apresentavam alelos múltiplos, havendo desta forma mais de duas classes de homocigotos. Neste caso, o homocigoto que apresentou um valor maior em kg ha<sup>-1</sup> foi considerado como favorável, e recebeu o valor 1. Quando apenas uma cópia de seus alelos estava presente num indivíduo, este foi considerado heterocigoto para o alelo favorável e recebeu o valor 0. Os indivíduos que não apresentaram nenhuma cópia do alelo favorável receberam valor -1. Esse procedimento foi repetido em todos os 157 locos marcadores utilizados neste estudo.

Após a definição dos alelos favoráveis dos locos marcadores e a codificação das progênies em cada loco, foi possível identificar quais marcas apresentavam efeito significativo para o caráter produtividade de grãos, através de análises de regressão executadas pelo programa SAS (SAS Institute, 1992). Primeiramente, os locos marcadores foram submetidos ao procedimento de regressão múltipla *stepwise* para cada cromossomo separadamente, que identificou as marcas significativas

dentro dos níveis de significância  $P < 0,10$  e  $P < 0,15$ , para entrada e permanência no grupo, respectivamente. Essa seleção inicial por cromossomo foi feita devido à presença de um número maior de variáveis  $x$  que  $y$ , ou seja, de um número maior de marcas do que de progênies. O segundo passo foi executar uma análise de regressão múltipla reunindo todos os cromossomos, com as marcas selecionadas pelo processo *stepwise*, que forneceu os coeficientes parciais de regressão  $b_j$  e o coeficiente de determinação  $R^2$ , utilizado como medida da proporção da soma de quadrados entre *topcrosses* explicada pelos marcadores utilizados.

Na regressão múltipla foi empregado o modelo linear usual, ou seja,

$$Y_i = \mu + b_1 X_{i1} + b_2 X_{i2} + \dots + b_j X_{ij} + \dots + b_m X_{im} + d_{ij}, \text{ em que}$$

$Y_i$  é a média ajustada da produtividade de grãos da progênie (*topcross*)  $i$  ( $i=1,2,\dots,n$ );  $\mu$  é o intercepto;  $b_j$  é o efeito genotípico associado ao loco marcador  $j$  ( $j=1,2,\dots,m$ ), cuja participação no modelo foi decidida anteriormente pelo processo *stepwise* de regressão, aplicado por cromossomo, conforme já exposto;  $X_{ij}$  é uma variável matemática indicativa do estado genotípico da progênie  $i$ , no loco marcador  $j$ , valendo 1, 0 ou -1. Como já mencionado,  $X_{ij}$  valerá 1 se a progênie for homocigótica na marca  $j$  e tiver QTLs favoráveis associados a essa condição; será -1 para os homocigotos com QTLs desfavoráveis associados e será 0 se a progênie for heterocigótica nesse loco marcador  $j$ ;  $d_{ij}$  é um desvio em relação ao modelo e engloba o erro experimental ocorrido na observação  $Y_i$ , bem como qualquer insuficiência paramétrica do modelo.

Observa-se, nesse modelo, que os coeficientes  $b_j$  são equivalentes a efeitos aditivos, ou seja, efeitos aditivos associados às marcas. No caso, por exemplo, de um só QTL ( $Q$ ) ligado à marca  $M$  ( $M_1, M_2$ ) em que os efeitos genotípicos são dados por:  $QQ:\mu + a$ ;  $Qq:\mu + d$  e  $qq:\mu - a$  e a fração de recombinação entre  $Q$  e  $M$  é  $c$ , tem-se que  $b = a(1-2c)$ , se as frequências alélicas forem 0,5. Percebe-se que, nestas condições, há correspondência entre  $b$  e o desvio genotípico " $a$ ", mostrado anteriormente.

Para quantificar a proporção da variação explicada pelos marcadores moleculares, calculou-se o valor  $p$ , conforme proposto por Bearzoti & Vencovsky (1998). A composição de  $p$  pode ser vista a seguir.

$$p = \frac{\sigma_m^2}{\sigma_g^2} = \frac{\sigma_m^2}{\sigma_m^2 + \sigma_h^2} \text{ sendo,}$$

$p$  = proporção da variância genética explicada pelos marcadores;  $\sigma_g^2$  variância genética entre as progênies;  $\sigma_m^2$  variância genética explicada pelos marcadores;  $\sigma_h^2$  variância genética não explicada pelos marcadores.

O parâmetro  $p$  foi estimado a partir da análise conjunta dos locais, com base nas médias ajustadas por local. Nessa análise conjunta estimou-se a variância genética entre progênies ( $\sigma_g^2$ ) que foi desdobrada nas componentes  $\sigma_m^2$  e  $\sigma_h^2$  em função das marcas selecionadas pela regressão *stepwise*. Para isolamento dessas componentes foi necessário considerar as variâncias das interações  $\sigma_{mi}^2$  e  $\sigma_{hi}^2$  (marcas x local e desvios x local, respectivamente). Os detalhes dessa análise conjunta estão sendo objeto de pesquisa em andamento, para publicação em separado.

## RESULTADOS E DISCUSSÃO

A TABELA 1 apresenta a análise da variância para o conjunto dos locais. Verificando o valor do coeficiente de variação (CV), nota-se que a precisão dos experimentos foi muito boa. A produtividade também foi elevada, o que reflete condições favoráveis para o desenvolvimento das plantas. É preciso ressaltar a significância dos efeitos gerais de progênies bem como da interação progênies x locais. Tais resultados indicam boa qualidade dos dados para as finalidades desta pesquisa.

A distância entre as marcas utilizadas variou de 11,12 a 28,61 cM nos dez cromossomos do milho, com média de 15,30 cM para todos os cromossomos. Segundo Falconer & MacKay (1996), um espaçamento de aproximadamente 20 cM entre os marcadores é suficiente para uma boa cobertura do genoma quando se trata de identificação de QTLs em programas de melhoramento. Essa distância mínima foi de forma geral respeitada nos cromossomos, com exceção no cromossomo quatro, cujo valor médio de cobertura foi de 28,61 cM (dados não mostrados). Outro ponto importante diz respeito ao valor do intervalo de posicionamento entre sondas nos cromossomos. Verificou-se que em alguns casos a distância entre os intervalos foi maior que 20 cM, enquanto em outros essa distância foi bem menor. Ocorreu também o posicionamento de várias marcas numa mesma região do cromossomo. Para maiores informações sobre a cobertura do genoma pelos marcadores utilizados ver Rumin (2000). É importante que se evite a saturação de regiões cromossômicas em detrimento de outras, pois todas as regiões devem ser investigadas quanto à sua participação na manifestação do caráter, para que a seleção assistida possa apresentar vantagens em relação à seleção convencional. As marcas selecionadas pelo procedimento *stepwise* em cada cromossomo podem ser vistas na TABELA 2. A proporção da variância genética explicada pelos marcadores utilizados ( $p$ ) e o coeficiente de determinação  $R^2$  da análise de regressão múltipla podem ser observados na TABELA 3.

TABELA 1 - Análise conjunta da variância da produtividade de grãos (kg ha<sup>-1</sup>). *Topcrosses* e testemunhas de milho. Quatro locais, Iowa, EUA, 1996.

	FV	GL	QM	F
locais		3	63.953.769,67	66,93**
repetições / locais		4	545.100,11	-
blocos / repetições / locais		16	10.702.096,46	-
[tratamentos]		[75]	[3.033.829,45]	[3,20**]
[tratamentos x locais]		[225]	[951.106,20]	[1,50**]
progênes aj.		67	1.872.262,06	2,19 **
marcas (regressão)		18	4.380.854,08	4,60 **
desvios		49	950.738,46	1,11
testemunha		7	4.894.455,90	4,83**
progênes vs. testemunhas		1	67.834.459,50	9,55**
progênes x locais		201	852.907,07	1,35*
testemunhas x locais		21	1.012.500,64	1,60*
(progênes vs. testemunhas) x locais		3	7.100.686,81	11,21**
resíduo		358	633.256,71	
total		681		
CV		8,6		
média		8.228,7		

Conforme já mencionado, o objetivo deste estudo foi o de desenvolver um índice de seleção de linhagens para a montagem de sintéticos em milho, baseado nos valores de produtividade de grãos e nos dados da genotipagem de RFLPs. A seguir são dados os fundamentos relativos à construção do índice.

A composição do índice de seleção baseia-se nos coeficientes  $b_j$  de regressão e na constituição genotípica das linhagens. Inclui função do comportamento próprio (*per se*) de cada linhagem ( $L_i$ ) bem como uma medida da complementaridade genotípica de cada linhagem com as demais ( $D_{ii'}$ ).

Considerando a linhagem  $i$  e sendo  $m$  o número de locos marcadores pertinentes, a quantidade  $L_i = \left( \sum_j b_j X_{ij} \right) / m$  reflete o comportamento próprio dessa linhagem. Por sua vez  $L_{max} = \left( \sum_j |b_j| \right) / m$  é uma medida do comportamento da linhagem ideal, ou seja, aquela em que todos os valores  $X_{ij}$  são iguais a 1. Neste estudo, resolveu-se tomar  $L'_i$  como fração de  $L_{max}$ , o que reflete o afastamento da linhagem  $i$  em relação a esse máximo, ou seja  $L_i = L'_i / L_{max}$ . Da mesma forma, para a linhagem  $i'$ , tem-se  $L_{i'} = L'_{i'} / L_{max}$ .

Para a combinação das linhagens  $i, i'$ , portanto, o índice utilizado é dado por

$$I_{ii'} = L_{ii'} + D_{ii'} \quad (1)$$

sendo  $L_{ii'} = (L_i + L_{i'})/2$ .

O parâmetro  $D_{ii'}$ , por sua vez, é constituído de  $D_{ii'} = D'_{ii'} / D_{max}$ , com

$$D'_{ii'} = \sqrt{\sum_j [b_j (X_{ij} - X_{i'j})]^2 / m} \text{ e } D_{max} = 2\sqrt{\sum_j b_j^2 / m}.$$

A quantidade  $D'_{ii'}$  é uma distância euclidiana entre as linhagens  $i$  e  $i'$ , baseada nos valores genotípicos aditivos ( $b_j$ ) relativos a cada loco marcador. Se duas linhagens forem idênticas,  $X_{ij} = X_{i'j}$  e  $D_{ii'} = 0$ . Por sua vez, a distância entre a linhagem ideal ( $X_{ij} = 1$ ), e a pior linhagem possível ( $X_{ij} = -1$ ) é a distância máxima  $D_{max} = \sqrt{\sum_j b_j^2 [1 - (-1)]^2 / m} = 2\sqrt{\sum_j b_j^2 / m}$ . Conforme exposto,  $D_{ii'}$  também é tomado em relação a essa distância máxima,  $D_{max}$ .

O índice usado (expressão 1) portanto, tem uma constituição que se assemelha ao que é sugerido pelo modelo representativo de comportamento de um híbrido entre duas linhagens, ou seja,  $\bar{F}_1 = MP + h$ , em que  $MP$  é a média dos genitores e  $h$  a heterose, para o caráter estudado. Observa-se que a distância  $D_{ii'}$ , como medida de complementaridade genotípica entre as linhagens  $ii'$ , leva em conta todos os tipos de locos, independentemente do grau de dominância dos QTLs envolvidos. Nesse aspecto,  $D_{ii'}$  tende a ser superior à medida da heterose ou da capacidade específica de combinação as quais só detectam complementaridade nos locos com algum grau de dominância. Por outro lado,  $D_{ii'}$  tem como desvantagem óbvia o fato de ter sensibilidade ou eficiência apenas naqueles locos (de marcadores e QTL's) em que há segregação, no conjunto de linhagens sob estudo. O mesmo vale para

as medidas  $L_r$ . A eficiência de índices como o proposto, para auxiliar na seleção, depende muito também da cobertura genômica que se consegue com os marcadores.

O índice de seleção ( $I_{ij}$ ), conforme exposto no item anterior, foi obtido para todas as combinações, duas a duas, das 68 linhagens estudadas. Isso gerou uma tabela dialética com 2.278 valores. As linhagens candidatas a entrar na constituição de um sintético foram selecionadas com base na média marginal dos respectivos valores  $I_{ij}$ , ou seja, com base em  $\bar{I}_i = (\sum_j I_{ij}) / (n-1)$ , com  $i, i' = 1, 2, \dots, 68; i' \neq i$ .

As quantidades  $\bar{I}_i$ , que se assemelham a capacidades gerais de combinação, foram ordenadas de forma decrescente, selecionando-se como candidatas aquelas com maior valor.

Para fins de comparação, as linhagens também foram selecionadas com base nas médias de produtividade obtidas dos ensaios das progênies *topcross* ( $\bar{Y}_i$ ). Em ambos os casos, optou-se por

TABELA 2 - Detecção de marcas significativas por cromossomo. Método *stepwise* de regressão. Produtividade de grãos de milho (kg ha<sup>-1</sup>). Análise conjunta dos quatro locais, Iowa, EUA. 1996.

Cromossomos	Sondas selecionadas	Valor genotípico da marca associado à produtividade de grãos kg ha <sup>-1</sup>
1	S2	53,03
1	S4	18,84
1	S15	153,26
2	S27	209,30
2	S30	90,48
2	S34	155,65
3	S49	133,56
3	S57	60,69
5	S82	589,46
6	S98	-508,36
6	S105	79,24
7	S110	-12,85
8	S132	102,48
9	S141	58,44
9	S142	137,90
9	S143	135,31
9	S148	104,58
10	S152	144,75

TABELA 3 - Produtividade média, coeficiente de variação (CV) e valores calculados de  $\rho$  e  $R^2$  Análise conjunta, Iowa, EUA. 1996.

Produtividade kg ha <sup>-1</sup>	CV	$\rho$	$R^2$	Marcas
8.228,7	8,62	0,747	62,86	18

sintéticos de diferentes tamanhos. O menor sintético considerado foi o de tamanho 5. Depois deste, foram elaborados outros aumentando-se o número de linhagens uma a uma, sempre com base na respectiva ordenação, até se chegar ao sintético amplo, composto de todas as 68 linhagens.

Com a finalidade de confrontar o emprego do índice com o critério das médias dos *topcrosses* e de identificar o tamanho mais adequado dos sintéticos foram adotados dois critérios, a saber: (1) baseados na frequência do alelo marcador associado a alelos favoráveis em cada loco marcador selecionado ( $q_j$ ) e no conjunto desses locos ( $\bar{q}$ ) e (2) baseado no valor genotípico médio ( $\bar{b}$ ) inerente a um sintético.

No critério (1), dado no parágrafo anterior, a quantidade  $q_j$  do loco  $j$ , foi calculada através da soma das linhagens contendo o genótipo marcador favorável, dividido pelo número total de linhagens constituintes do sintético. Para o dado sintético, como um todo, a frequência média  $\bar{q}$ , foi calculada de modo ponderado, ou seja, através de  $\bar{q} = \frac{\sum q_j b_j}{\sum b_j}$ . Essa quantidade ( $\bar{q}$ ) reflete a riqueza de alelos marcadores associados a QTLs favoráveis, dando-se peso maior às sondas que detectam QTLs mais importantes na expressão do caráter.

No critério (2), que é complementar, considerou-se o coeficiente de regressão médio  $\bar{b} = \frac{\sum q_j b_j}{\sum q_j}$ , para cada tamanho de sintético. Agora o interesse foi de mensurar o valor genotípico (aditivo) médio detectado pelas marcas, com ponderação pela frequência com que os alelos marcadores favoráveis selecionados ocorrem, num dado sintético.

Neste trabalho foram comparados os resultados obtidos com os sintéticos segundo o critério do índice e segundo o critério do *topcross*. A seguir, na TABELA 4, observa-se a ordenação das 15 melhores linhagens segundo o critério do índice ( $\bar{I}_i$ ) e segundo o critério do *topcross* ( $\bar{Y}_i$ ).

Observa-se que a ordenação das linhagens feitas pelo índice de seleção, que leva em conta as distâncias genéticas entre os indivíduos, é diferente daquela estabelecida pelo *topcross*, onde foram tomados por base apenas os valores obtidos de produtividade de grãos das progênies. No confronto entre o critério do índice e o critério dos *topcrosses*, tendo como referencial as melhores 15 linhagens na média dos quatro locais, coincidiram pelos dois critérios as de número 63, 9, 62, 5, 47, 19, 54 e 39 (TABELA 4). Houve portanto uma coincidência de 8/15 linhagens ou de aproximadamente 50% delas. Essas últimas linhagens estão, sem margem de dúvida, entre as melhores do conjunto de 68 analisadas. A dúvida no processo de seleção surge, no entanto, diante das linhagens que se destacam por um critério mas não por outro. Assim, as linhagens 46, 60 e 18 estão entre as 15 superiores segundo o índice,

mas classificam-se como intermediárias pelas médias *topcross* (dados não mostrados). Inversamente, as de número 17, 64 e 44 despontam nos *topcrosses* e não se situam bem pelo índice (dados não mostrados). Essa discordância está no fato de o índice levar em conta a complementaridade genotípica entre as linhagens, e as médias *topcross* não. Lembrando que, para o par de linhagens  $ii'$ ,  $I_{ii'} = L_{ii'} + D_{ii'}$ , tem-se que a média marginal da tabela dialélica é  $\bar{I}_i = \bar{L}_i + \bar{D}_i$ ; a complementaridade média  $\bar{D}_i$  é pois levada em conta na seleção pelo índice. O artificialismo desse índice está em que  $\bar{L}_i$  e  $\bar{D}_i$  receberam pesos iguais para compor  $\bar{I}_i$ . Outros pesos poderiam ser investigados visando uma possível otimização desse índice. Um risco possível com o uso de  $\bar{I}_i$  está, por exemplo, na seleção de uma linhagem inferior para a qual  $\bar{D}_i$  deve ser alto, devido ao forte contraste entre as freqüências gênicas desta com as demais. Tendo essa linhagem inferior um valor baixo de  $\bar{L}_i$  ela, mesmo assim, poderá ser selecionada. Por essa razão utilizou-se como critério adicional de decisão sobre a composição de um sintético, a freqüência  $\bar{q}$ .

O tamanho do sintético considerado ideal foi o de 10 linhagens segundo o índice de seleção e 12 segundo o *topcross*. O critério utilizado para a definição do número ideal de linhagens na composição do sintético foi o conjunto de linhagens a partir do qual não haverá fixação de alelos desfavoráveis em nenhum dos locos envolvidos. Para a comparação entre os sintéticos formados por números diferentes de linhagens, o parâmetro da fixação alélica não é suficiente. É preciso verificar as freqüências dos alelos favoráveis associados às marcas, pois o objetivo do índice de

seleção foi o de obter sintéticos com maiores freqüências de alelos favoráveis quando comparados aos sintéticos construídos pelo *topcross*. As comparações foram feitas entre sintéticos de diferentes tamanhos formados pelo mesmo método, e entre sintéticos de igual tamanho formados por métodos diferentes.

Quanto às freqüências de marcas associadas aos alelos favoráveis,  $\bar{q}$ , os sintéticos formados pelo índice apresentaram valores superiores aos dos formados pelo *topcross*. Quando o parâmetro de avaliação foi  $\bar{b}$ , o índice levou a valores superiores em comparação aos dos obtidos com os *topcrosses*, indicando a seleção de locos com valores genotípicos altos associados à produtividade de grãos em milho, como pode ser observado na TABELA 5.

Uma contribuição interessante, dada neste estudo, está na possibilidade de monitorar melhor as propriedades genéticas de um sintético qualquer. É preciso lembrar que  $\bar{q}$  é uma medida da freqüência de linhagens em que as marcas estão associadas a QTLs favoráveis. Sendo todas as linhagens homozigóticas esse parâmetro corresponde à freqüência média de alelos das marcas ligados a alelos superiores nos QTLs. Considerando o sintético amplo, esse valor foi de  $\bar{q} = 0,22$  (TABELA 5). Nos melhores sintéticos, de tamanho 10 e 12, os valores correspondentes são  $\bar{q} = 0,38$  e  $\bar{q} = 0,37$  respectivamente. Ora, essa diferença  $(0,38-0,22)=0,16$  ou  $(0,37-0,22)=0,15$  é considerável em se tratando de um caráter quantitativo como o mostrado. Da mesma forma, não é desprezível a diferença entre  $\bar{q}$ , confrontando-se os sintéticos organizados com base no índice e com base nas médias *topcross* (TABELA 5). Como pode ser visto, o índice foi competente em conduzir o sintético a valores elevados de  $\bar{q}$ .

O presente trabalho permitiu chegar às seguintes conclusões:

O índice proposto pode ser útil para o melhorista por levar a um homólogo de uma tabela dialélica, sem que seja necessário obter todos os cruzamentos entre os genitores;

Pela sua natureza, o índice mensura a complementaridade genotípica entre os genitores em todos os locos segregantes, independentemente do grau de dominância dos alelos envolvidos. Nesse aspecto, tende a ser superior às medidas de capacidade específica de combinação e heterose.

A seleção de linhagens pelo índice, visando sintéticos, foi mais efetiva que o critério das médias *topcross*, em elevar a freqüência dos alelos favoráveis associados às marcas. Tal conclusão só é válida para os locos de QTLs abrangidos pelas marcas.

O uso de um índice como o proposto só é viável se houver cobertura abrangente do genoma pelas marcas utilizadas.

TABELA 4 - Ordenação das linhagens segundo os valores  $\bar{I}_i$ , obtidos da tabela dialélica, e segundo a produtividade de grãos no *topcross* ( $\bar{Y}_i$ ) - Dados conjuntos dos quatros locais para as 15 linhagens em cada caso.

Conjunta			
linhagens	$\bar{I}_i$	linhagens	$\bar{Y}_i$
63	0,5542	62	10461,19
9	0,5329	9	10156,32
62	0,5309	17	9832,01
14	0,4266	54	9801,43
12	0,4118	63	9794,90
5	0,4077	39	9762,42
47	0,3887	61	9638,14
19	0,3830	64	9636,56
54	0,3803	5	9624,33
52	0,3787	47	9535,13
46	0,3780	4	9533,57
39	0,3663	13	9520,48
60	0,3617	31	9498,07
21	0,3571	19	9473,39
18	0,3220	44	9473,06

TABELA 5 - Valores de  $q$ ,  $\bar{q}$  e  $\bar{b}$  para avaliação dos melhores sintéticos com diferentes tamanhos – Topcrosses de milho - Análise conjunta de quatro locais - Iowa, EUA, 1996.

número de linhagens*	Índice					número de linhagens*	Topcross				
	menor $q$	maior $q$	média	$\bar{q}$	$\bar{b}$		menor $q$	maior $q$	média	$\bar{q}$	$\bar{b}$
5	0,00	0,80	0,36	0,42	178,51	5	0,00	0,80	0,31	0,36	174,25
6	0,00	0,83	0,33	0,40	184,02	6	0,00	0,83	0,31	0,35	176,71
7	0,00	0,71	0,34	0,40	180,73	7	0,00	0,71	0,30	0,35	179,03
8	0,00	0,75	0,32	0,39	185,37	8	0,00	0,63	0,28	0,33	176,27
9	0,00	0,78	0,31	0,38	186,15	9	0,00	0,67	0,28	0,33	180,88
10	0,10	0,80	0,32	0,38	184,69	10	0,00	0,60	0,29	0,34	178,58
11	0,09	0,82	0,30	0,37	191,96	11	0,00	0,64	0,29	0,32	170,64
12	0,08	0,83	0,30	0,37	191,85	12	0,08	0,67	0,28	0,31	170,25
13	0,08	0,85	0,29	0,37	191,49	13	0,08	0,69	0,29	0,32	167,54
14	0,07	0,79	0,29	0,37	191,69	14	0,07	0,64	0,28	0,32	171,10
15	0,07	0,80	0,28	0,36	196,99	15	0,07	0,60	0,27	0,30	168,74
68	0,01	0,44	0,20	0,22	174,84	68	0,01	0,44	0,20	0,22	174,84

\*Tamanho do sintético conforme o número de linhagens.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AJMONE-MARSAN, P.; MONFREDINI, G.; BRANDOLINI, A.; MELCHINGER, A. E.; GARAY, G.; MOTTO, M. Identification of QTL for grain yield in an elite hybrid of maize: repeatability of map position and effects in independent samples derived from the same population. **Maydica**, v.41, p.49-57, 1996.
- BEARZOTI, E. Simulação de seleção recorrente assistida por marcadores moleculares em espécies autógamas. Piracicaba, 1997. 230p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- BEARZOTI, E.; VENCOSKY, R. Estimation of the proportion of genetic variance explained by molecular markers. **Genetics and Molecular Biology**, v.21, p.557-566, 1998.
- BEAVIS, W.D.; SMITH, O.S.; GRANT, D.; FINCHER, R. Identification of quantitative trait loci using a small sample of topcrossed and  $F_4$  progeny of maize. **Crop Science**, v.34, p.882-896, 1994.
- BERKE, T.G.; ROCHEFORD, T.R. Quantitative trait loci for flowering, plant and ear height, and kernel traits in maize. **Crop Science**, v.35, p.1542-1549, 1995.
- BUBECK, D.M.; GOODMAN, M.M.; BEAVIS, W.D.; GRANT, D. Quantitative trait loci controlling resistance to gray leaf spot in maize. **Crop Science**, v.33, p.838-847, 1993.
- CHORY, J.; MOORE, D.; REICHARDT, M.; RIBALDO, R.K.; RICHARDS, E.; ROGERS, S. Preparation and Analysis of DNA. In: AUSEBEL, F.M.; KINGSTON, R. E.; MOORE, D.D.; PRENT, R.; SEIDMAN, J.G.; SMITH, J.A.; STRUHL, K. (Ed.) **Current protocols in molecular biology**. Boston: John Wiley & Sons, 1996. v.1, cap.2, p.1-12.
- DUBREUIL, P.; DUFOUR, P.; KREJCI, E.; CAUSSE, M.; VIENNE, D. de; GALLAIS, A.; CHARCOSSET, A. Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. **Crop Science**, v.36 p.790-799, 1996.
- DUDLEY, J.W. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. **Crop Science**, v.33, p.660-668, 1993.
- DUDLEY, J.W.; SAGHAI MAROOF, M.A.; RUFENER, G.K. Molecular markers and grouping of parents in maize breeding programs. **Crop Science**, v.31, p.718-723, 1991.
- EDWARDS, M.; JOHNSON, L. RFLPs for rapid recurrent selection. In: PROCEEDINGS OF THE SYMPOSIUM ANALYSIS OF MOLECULAR MARKER DATA, Corvallis, 1994. **Proceedings**. Corvallis: American Society for Horticultural Science, 1994. p.33-40.
- FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4.ed. Essex: Longman, 1996. 464p.
- GODSHALK, E.B.; LEE, M.; LANKEY, K.R. Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize. **Theoretical and Applied Genetics**, v.80, p.273-280, 1990.
- GOLDMAN, I.L.; ROCHEFORD, T.R.; DUDLEY, J.W. Molecular markers associated with maize kernel oil concentration in an Illinois high protein x Illinois low protein cross. **Crop Science**, v.34, p.908-915, 1994.
- GOMES, F.P. **Curso de estatística experimental**. 13.ed. São Paulo: Livraria Nobel, 1990. 467p.
- HOSPITAL, F.; CHEVALET, C.; MULSANT, P. Using markers in gene introgression breeding programs. **Genetics**, v.132, p.1199-1210, 1992.
- JANSEN, R.C.; VAN OOIJEN, J.W.; STAM, P.; LISTER, C.; DEAN, C. Genotype-by-environment interaction in genetic mapping of multiple quantitative trait loci. **Theoretical and Applied Genetics**, v.91, p.33-37, 1995.
- KERNS, M. R.; DUDLEY, J. W.; RUFENER II, G. K. Tester and type of progeny affect QTL detection in maize. **Maydica**, v.44, p.69-83, 1999.
- LANZA, L.L.B.; SOUZA JR., C.L.; OTTOBONI, L.M.M.; VIEIRA, M.L.C.; SOUZA, A.P. Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. **Theoretical and Applied Genetics**, v.94, p.1023-1030, 1997.
- MELCHINGER, A.E.; LEE, M.; LAMKEY, K.R.; WOODMAN, W.L. Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize inbreds. **Crop Science**, v.30, p.1033-1040, 1990.
- MESSMER, M.M.; MELCHINGER, A.E.; HERRMANN, R.G.; BOPPENMAIER, J. Relationships among early european maize inbreds: II. Comparison of pedigree and RFLP data. **Crop Science**, v.33, p.944-950, 1993.

- MOREAU, L.; MONOD, H.; CHARCOSSET, A.; GALLAIS, A. Marker-assisted selection with spatial analysis of unreplicated field trials. **Theoretical and Applied Genetics**, v.98, p.234-242, 1999.
- OGLIARI, J. B. Identificação e localização de um gene de resistência de milho a *Exserohilum turcicum* (Pass.) Leonard & Suugs através do uso de marcadores microssatélites. Piracicaba, 1999. 115p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- OPENSHAW, S.J.; JARBOE, S.G.; BEAVIS, W.D. Marker-assisted selection in backcross breeding. In: SYMPOSIUM ANALYSIS OF MOLECULAR MARKER DATA, Corvallis, 1994. **Proceedings**. Corvallis: American Society for Horticultural Science, 1994. p.41-43.
- RAGOT, M.; SISCO, P.H.; HOISINGTON, D.A.; STUBER, C.W. Molecular-marker-mediated characterization of favorable exotic alleles at quantitative trait loci in maize. **Crop Science**, v.35, p.1306-1315, 1995.
- RUMIN, G.C.R. Índice para a seleção de linhagens endogâmicas de milho (*Zea mays* L.) visando populações sintéticas baseado em marcadores genéticos. Piracicaba, 2000. 201p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- SAS INSTITUTE. Release 6.12, (software). Cary: Statistical Analysis System Institute, 1992.
- SCHÖN, C.C.; MELCHINGER, A.E.; BOPPENMAIER, J.; BRUNKLAUS-JUNG, E.; HERRMANN, R.G.; SEITZER, J.F. RFLP mapping in maize: quantitative trait loci affecting testcross performance of elite european flint lines. **Crop Science**, v.34, p.378-389, 1994.
- SELIG, L.; LAMBERT, R.J.; ROCHEFORD, T.R.; SILVA, W.J. da. RFLP and cluster analysis of introgression of exotic germplasm into U.S. maize inbreds. **Maydica**, v.44, p.85-92, 1999.
- STROMBERG, L.D.; DUDLEY, J.W.; RUFENER, G.K. Comparing conventional early generation selection with molecular marker assisted selection in maize. **Crop Science**, v.34, p.1221-1225, 1994.
- STUBER, C.W. Enhancement of grain yield in maize hybrids using marker-facilitated introgression of QTLs. In: SYMPOSIUM ANALYSIS OF MOLECULAR MARKER DATA, Corvallis, 1994. **Proceedings**. Corvallis: American Society for Horticultural Science, 1994. p.44-46.

---

Recebido em 24.04.00