

Algoritmo computacional para determinar o perfil mínimo de marcadores moleculares que discriminam um conjunto de cultivares

SILLA, P. R.¹; PASSIANOTTO, A. L. L.²; CAMARGO-BRUNETTO, M. A. O.¹; GONELA, A.²; MARCELINO, F. C.³; BINNECK, E.³

¹Universidade Estadual de Londrina, ²Universidade Estadual de Maringá,

³Embrapa Soja, Caixa Postal, 231, 86001-970, Londrina, Paraná,

e-mail: paulo.silla@gmail.com

Introdução

A identificação de cultivares é uma ferramenta essencial para que os fundamentos da lei de proteção de cultivares, instituída no Brasil através da lei nº 9.456, de 1997, sejam praticáveis, reconhecendo a propriedade intelectual dos obtentores de genótipos comerciais. Um problema recorrente para a aplicação da lei de cultivares é o número insuficiente de descritores disponíveis, não só para a identificação, mas também para o monitoramento da pureza genética e a descrição de novas cultivares. Por exemplo, os cerca de 38 descritores morfológicos, fisiológicos e bioquímicos, entre os obrigatórios e os adicionais, atualmente utilizados para diferenciar cultivares de soja, são ainda insuficientes para distinguir todas as cultivares de maneira inequívoca, o que impõe a necessidade de ampliar a lista de descritores utilizados (Nogueira et al., 2008). Entre o rol de descritores disponíveis para ampliar essa lista, os marcadores moleculares de DNA são descritores ideais por várias razões, entre elas: possibilitam acessar diretamente a informação genética dos materiais, podem ser utilizados em qualquer estágio de desenvolvimento da planta, além de possibilitarem a determinação de um padrão exclusivo *fingerprint* para cada cultivar. Por isso, hoje, a utilização de marcadores moleculares já é uma realidade na identificação de genótipos, sejam eles cultivares, linhagens ou acessos de banco de germoplasma. Isso porque além da aplicação direta na identificação de cultivares, os marcadores moleculares são de grande valia na seleção das progênies promissoras ao melhoramento, facilitando a escolha das melhores combinações de progenitores bem como a segregação dos acessos e linhagens. Dentre os marcadores moleculares disponíveis, os microssatélites ganham destaque por sua natureza co-dominante e multialélica, por serem extremamente conservados dentro da espécie e por haver um grande número de marcadores disponíveis para cada espécie. Embora exista um grande número de microssatélites conhecidos (por exemplo, mais de mil para a soja), por questões de praticidade, economicidade e rapidez nas análises, busca-se o menor número de marcadores capazes de discriminar de maneira inequívoca o conjunto completo de materiais estudados. No entanto, o trabalho manual de selecionar esse conjunto mínimo de marcadores nem sempre é uma tarefa fácil. Nesse contexto, o presente trabalho visa estabelecer um método computacional automático para determinar o perfil mínimo de marcadores moleculares microssatélites que seja capaz de identificar unicamente cada material em um conjunto de genótipos.

Material e Métodos

A abordagem foi desenvolvida com o uso da linguagem de programação Python (disponível para *download* gratuito em www.python.org) e recebe como entrada um arquivo texto, com os dados referentes aos picos observados pelas leituras, de modo que as cultivares estão dispostas em linhas e os marcadores em colunas, como pode ser observado na Tabela 1. O algoritmo considera conflito a existência de valores iguais para cultivares distintas em um mesmo marcador.

Tabela 1. Exemplo da formatação do arquivo de entrada: m cultivares e n marcadores.

	Sat 1	Sat 2	...	Sat n
Cultivar 1	10*	10	...	12
Cultivar 2	12	15	...	14
...
Cultivar m	10*	11	...	13

* Existência de conflito.

Em seguida, são criados vetores de pontuação para todos os marcadores, através da comparação dos valores observados entre as cultivares. Cada posição do vetor de pontuação representa a existência ou não de um conflito entre a cultivar de mesmo índice e alguma outra. Para representar um conflito foi determinado o valor -1 . O valor $+1$ indica a não existência de conflitos.

Os pesos de cada vetor de pontuação são obtidos através da soma dos pontos definidos no passo anterior. Também é criado um vetor de pontuação, denominado vetor resposta, para armazenar os resultados obtidos em cada iteração. O método converge para uma solução quando todos os elementos do vetor resposta possuem o valor $+1$.

Cada iteração do método criado possui os seguintes passos: 1) os vetores de pontuação são inseridos em uma lista ligada em ordem decrescente de peso; 2) o vetor com maior peso (primeiro na lista ordenada) é escolhido como parte da resposta; 3) as posições onde não existe conflito são marcadas com $+1$ no vetor resposta e em todos os outros vetores de pontuação que continuam na lista; 4) uma verificação é realizada a fim de descobrir se o método convergiu ou se ainda existem vetores na lista ligada; 5) caso o método não tenha obtido a resposta e ainda existem vetores na lista ligada, os pesos são novamente calculados, a lista é reordenada e o método retorna ao passo dois.

Para a validação do método, foram utilizados dois conjuntos de dados: o primeiro (D01) consiste de valores hipotéticos, contendo 10 cultivares e 15 marcadores e o segundo (D02) construído a partir de leituras realizadas no laboratório de biotecnologia vegetal da Embrapa Soja para 22 cultivares com o uso de 19 marcadores microsatélite. Para tal, foram observados os picos gerados pelo *software* GeneMapper® fornecido pela empresa Applied Biosystem®. Considerou-se, nesse estudo, apenas as leituras com valores de intensidade acima de 1000.

Resultados e Discussão

O conjunto de dados D01 foi construído manualmente de modo a necessitar de cinco dos 15 marcadores definidos para identificar unicamente as 10 cultivares existentes. O método desenvolvido apresentou o resultado esperado, ou seja, cinco marcadores para discriminar unicamente as cultivares.

Ao executar o método utilizando o conjunto de dados reais (D02), obtém-se que nove marcadores microsatélite conseguem discriminar unicamente as 22 cultivares.

Conclusões

A abordagem desenvolvida mostrou-se eficiente para determinar uma lista mínima de marcadores moleculares microsatélites capaz de identificar um conjunto de genótipos. O método também mostrou-se robusto para indicar os casos em que o conjunto de marcadores estudados são insuficientes em número ou não são polimórficos o suficiente para discriminar todos os genótipos.

Referências

NOGUEIRA, A. P. O. ; SEDIYAMA T.; CRUZ, C. D.; REIS, M. S.; PEREIRA, D. G.; JANGARELLI. M. Novas características para diferenciação de cultivares de soja pela análise discriminante. **Ciência Rural**, n° 38, p. 2427-2433, 2008.