

Marcadores moleculares e suas aplicações no melhoramento genético

Claudia Teixeira Guimarães¹
Jurandir Vieira de Magalhães²
Marcelo Abreu Lanza³
Ivan Schuster⁴

Resumo - O melhoramento genético tem sido um dos grandes responsáveis pelos avanços na agricultura, com o desenvolvimento de cultivares superiores, quer pela maior produtividade, quer pela melhor adaptação aos ambientes adversos. O sucesso de um programa de melhoramento genético depende, fundamentalmente, de algumas etapas como a escolha de genitores e a seleção de genótipos superiores. Desde o surgimento dos marcadores RFLP, na década de 1980, as metodologias para a exploração dos polimorfismos de DNA têm sido alvo de grandes avanços na automação e na geração de dados em larga escala, fornecendo uma amostragem do genoma, cada vez mais ampla, a um custo menor. Assim, com a tendência do aumento crescente no volume de dados, associada à redução nos custos, os marcadores moleculares firmam-se como estratégias sólidas para auxiliar o melhoramento genético, assim como estudos sobre clonagem e função de genes, filogenia, diversidade e estrutura genética em espécies cultivadas e silvestres. Serão apresentados marcos importantes na evolução das tecnologias de marcadores de DNA e algumas das suas aplicações, com ênfase nas análises de diversidade, mapeamento genético e seleção assistida em plantas.

Palavras-chave: Marcador molecular. Modelo estatístico. Polimorfismo. Genoma.

INTRODUÇÃO

Com o advento das técnicas de biologia molecular, tornou-se possível a manipulação do ácido desoxirribonucleico (DNA), culminando com o surgimento dos marcadores moleculares na década de 1980. Desde então, esses marcadores têm acompanhado os avanços da era genômica, beneficiando-se do grande volume de informações de sequências de DNA disponível. Além das vantagens apresentadas sobre os marcadores morfológicos, as tecnologias existentes fornecem um número praticamente ilimitado de polimorfismos

distribuídos aleatoriamente ao longo de todo o genoma, de forma automatizada e a custos cada vez mais reduzidos. O avanço das técnicas moleculares tem sido acompanhado de perto pelo grande desenvolvimento nas áreas da bioinformática, da estatística e da genética quantitativa. Com isso, encontram-se disponíveis informações genômicas ancoradas a mapas físicos e genéticos altamente saturados, além dos inúmeros genes diferencialmente expressos, que são facilmente integrados por meio da bioinformática. Assim, o mapeamento de caracteres quantitativos – *quantitative*

trait locus (QTLs) e de QTLs expressos tem dado suporte para a identificação de genes e para a seleção assistida, trazendo um vasto conteúdo de informação genética para a maioria das espécies vegetais. A aplicação dessas metodologias, para acelerar e monitorar os programas de melhoramento genético, tem implicado em avanços no desenvolvimento de variedades melhoradas.

As limitações em termos de saturação do genoma, número de indivíduos analisados e precisão na detecção de QTLs tendem a ser minimizadas com a revolução trazida

¹Eng^a Agr^a, D.Sc. Pesq. Embrapa Milho e Sorgo, Caixa Postal 151, CEP 35701-970 Sete Lagoas-MG. Correio eletrônico: claudia@cnpms.embrapa.br

²Eng^a Agr^a, D.Sc. Pesq. Embrapa Milho e Sorgo, Caixa Postal 151, CEP 35701-970 Sete Lagoas-MG. Correio eletrônico: jurandir@cnpms.embrapa.br

³Eng^a Agr^a, D.Sc. Pesq. U.R. EPAMIG TP, Caixa Postal 351, CEP 38001-970 Uberaba-MG. Correio eletrônico: mlanza@epamig.br

⁴Eng^a Agr^a, D.Sc. Pesq. COODETEC, Caixa Postal 301, CEP 85813-450 Cascavel-PR. Correio eletrônico: ivan@coodetec.com.br

pelo sequenciamento em altíssima escala. Tais avanços têm disponibilizado um grande número de polimorfismos de base única—*single-nucleotide polymorphism* (SNP), aleatórios no genoma ou em genes expressos, que, por sua vez, têm possibilitado a construção de plataformas para análise de centenas de milhares de marcadores em paralelo, a um custo reduzido. Essas condições combinadas com modelos estatísticos, propostos para seleção com base na genotipagem de genomas completos, irão contribuir cada vez mais para que a seleção assistida assumam um papel efetivo no desenvolvimento de cultivares superiores, mais adaptadas e com características especiais.

MARCADORES MOLECULARES

RFLP, RAPD e AFLP

Os marcadores do tipo *restriction fragment length polymorphism* (RFLP), *random amplified polymorphic DNA* (RAPD) e *amplified fragment length polymorphism* (AFLP) foram os marcadores de DNA mais utilizados nas décadas de 80 e 90, sendo ainda aplicados para alguns estudos genéticos. Esses marcadores foram responsáveis pelo desenvolvimento dos primeiros mapas genéticos de muitas espécies cultivadas, além da descoberta de genes e de inúmeros estudos de diversidade genética. Cada um deles apresenta vantagens e limitações, cobrindo um amplo espectro de aplicações.

Microsatélites ou SSR

Os genomas eucarióticos apresentam várias classes de sequências repetidas e uma delas consiste de repetições em tandem de 1 a 6 nucleotídeos, sendo denominadas microsatélites ou *simple sequence repeat* (SSR). As regiões genômicas que flanqueiam as sequências repetidas são conservadas dentro de uma espécie, permitindo o desenho de pares de *primers* entre 20 e 30 pb, que são utilizados para amplificar os microsatélites. Os polimorfismos

no tamanho dos fragmentos amplificados devem-se às diferenças no número de elementos simples repetidos, necessitando de géis de alta resolução para a separação dos fragmentos, uma vez que estes diferem por poucos pares de base. Em plantas, os microsatélites foram rapidamente implementados em várias espécies e são utilizados até hoje na construção de mapas genéticos, nos estudos de diversidade e no melhoramento assistido.

Dentre as vantagens que tornaram esses marcadores largamente utilizados podem-se destacar a alta frequência e a distribuição ao acaso nos genomas eucariotos, além de serem co-dominantes e multialélicos, fornecendo um elevado nível de informação genética por loco. A técnica de SSR é de fácil execução e altamente reproduzível, quando comparada com outras metodologias, facilitando o intercâmbio de informações entre diferentes grupos de pesquisa. Uma limitação da técnica é a obtenção dos *primers* específicos que flanqueiam os microsatélites, requerendo a construção de bibliotecas genômicas enriquecidas, o sequenciamento e o desenho dos *primers*. Tal limitação tem sido superada com os avanços nas técnicas moleculares, incluindo o sequenciamento em grande escala. Com isso, existe uma grande quantidade de pares de *primers* SSR com sequências publicamente disponíveis para várias espécies vegetais e animais. Apesar das possibilidades de automatização e da combinação de vários *primers* em sequenciadores automáticos, o número de locos analisados por reação ainda é reduzido diante dos novos desafios.

ISSR

Os marcadores *inter simple sequence repeat* (ISSR) utilizam um único *primer* desenhado com base nas sequências repetidas dos microsatélites na extremidade 5' com alguns nucleotídeos extras na extremidade 3'. Assim, os *primers* anelam-se dentro das repetições e amplificam as regiões genômicas entre os SSRs, cujos tamanhos dos fragmentos são limitados pela própria técnica da

polymerase chain reaction (PCR). Assim, o ISSR assemelha-se ao RAPD, uma vez que amplifica sequências aleatórias no genoma, embora seja mais reproduzível que o RAPD, considerando que os *primers* são maiores e necessitam de temperatura de anelamento mais elevada nas reações de PCR. A técnica surgiu em função da dificuldade de gerar os *primers* específicos para amplificar os SSR (ZIETKIEWICZ; RAFALSKI; LABUDA, 1994), sendo também muito usado em plantas e fungos.

DArTs

Com o surgimento dos microarranjos de DNA, Jaccoud et al. (2001) propuseram uma adaptação dessa tecnologia para a análise de polimorfismos de DNA com a criação dos *diversity arrays technology* (DArTs), onde fragmentos de DNA, que representam o genoma de um organismo ou de uma população, são digeridos com duas enzimas de restrição, clonados e impressos em uma lâmina de microarranjo (Fig. 1). Os polimorfismos são detectados pela presença ou ausência do fragmento após a hibridização com o DNA genômico de cada indivíduo digerido e marcado com fluoróforos. Assim, a técnica oferece uma ampla capacidade de amostragem de milhares de marcadores aleatórios no genoma sem necessidade do conhecimento prévio das sequências alvo e de forma automatizada.

Apesar do grande progresso no sequenciamento de genomas com a descoberta de SNPs, esses avanços não atingem todas as espécies cultivadas. Nesse aspecto, os DArTs oferecem soluções para a genotipagem em larga escala em espécies de órfãs como mandioca, feijão-de-corda (*Cajanus cajan*), cevada e quinoa, além de ser uma fonte valiosa para a geração de marcadores até mesmo para espécies cujos genomas já foram sequenciados como o sorgo (MACE et al., 2008). Uma lista das espécies, que possuem arranjos de DArTs disponíveis para a genotipagem, encontra-se em Diversity Arrays Technology (200—).

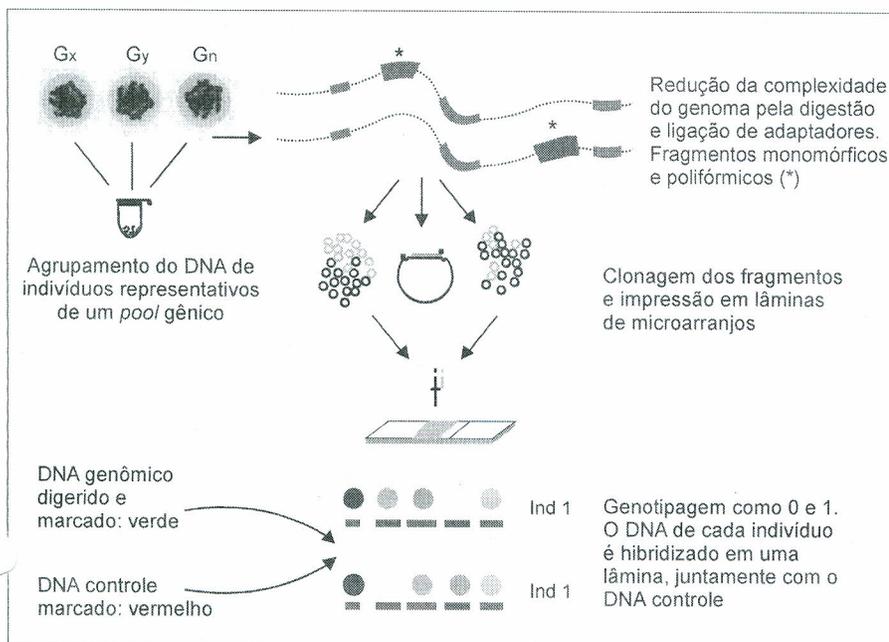


Figura 1 - Esquema da técnica de DArTs

NOTA: DArTs – Diversity arrays technology; DNA – Ácido desoxirribonucleico.

FONTE: Dados básicos: Jaccoud et al. (2001).

Marcadores com base em informações de sequências

O sequenciamento de DNA passou por duas grandes revoluções: a tecnologia dos dideoxinucleotídeos, que viabilizou a automatização do processo e, mais recentemente, o surgimento das novas tecnologias, também denominado sequenciamento de próxima geração, gerando bilhões de etiquetas de DNA (454/Roche; Genome Analyzer/Illumina® e SOLiD/Applied Biosystems™⁵). Maiores detalhes podem ser obtidos no artigo “Decifrando o genoma em grande escala”. Tais avanços viabilizaram os projetos de sequenciamento de genomas completos e expressos, gerando um volume enorme de informações de sequências de DNA. Com o surgimento dos sequenciadores automáticos de DNA, os marcadores RFLP, RAPD e AFLP puderam ser convertidos em *primers* específicos, para utilização em reações de PCR gerando marcadores polimórficos diretamente nas regiões de interesse. Sondas genômicas

de RFLP utilizadas no mapeamento do genoma humano foram os primeiros marcadores a serem sequenciados para amplificação com *primers* específicos, sendo denominados *sequence tagged site* (STS) (OLSON et al., 1989). A partir de então, a denominação STS tem sido aplicada para qualquer marcador que tenha sido convertido em *primers* de PCR por meio do sequenciamento das regiões alvo.

Os produtos de amplificação com *primers* específicos podem ser sequenciados, buscando polimorfismos que gerem ou eliminem sítios de restrição. O polimorfismo é detectado pela digestão dos produtos de amplificação com enzimas de restrição adequadas, gerando então o marcador *cleaved amplified polymorphic sequences* (CAPS) (KONIECZNY; AUSUBEL, 1993). Tais marcadores apresentam as vantagens práticas da técnica do PCR, mantendo a natureza co-dominante do RFLP. Uma alternativa para a genotipagem de SNP é a técnica de *derived cleaved amplified polymorphic sequence* (dCAPS), proposta

por Neff et al. (1998), onde a técnica de PCR é utilizada para introduzir diferenças de nucleotídeos e, assim, criar um polimorfismo com base na mutação alvo. Essa técnica é uma alternativa interessante nos casos em que o SNP a ser genotipado não cria ou elimina sítios de restrição, inviabilizando a utilização da técnica de CAPS. Neff et al. (1998) desenvolveram o *software* dCAPs Finder, que pode ser utilizado para marcadores do tipo dCAPS.

A conversão de marcadores RAPD em pares de *primers* específicos foi inicialmente realizada por Paran e Michelmore (1993), sendo denominada *sequence characterized amplified regions* (SCAR), com a vantagem da especificidade da amplificação para a região alvo em relação ao RAPD. Fragmentos gerados pela técnica de AFLP têm sido convertidos em STS polimórficos e específicos para locos de interesse.

SNPs e InDels

Os polimorfismos de nucleotídeos únicos ou SNPs são as variações mais frequentes no genoma de qualquer organismo. Estima-se que a frequência desse polimorfismo seja de uma substituição a cada 31 pares de base em regiões não codificadoras e a cada 124 pares de base em regiões codificadoras em milho (CHING et al., 2002). Em soja, estima-se a frequência de um SNP a cada 2.038 pares de bases em regiões codificadoras, e um SNP a cada 191 pares de bases em regiões não codificadoras (ZHU et al., 2003). Por sua abundância no genoma e pela lenta taxa de mutação dentro de gerações, os SNPs prometem ser a futura geração de marcadores genéticos de grande aplicação em processos de genotipagem em larga escala (USECHE et al., 2001). Outra classe de polimorfismo que tem sido muito explorada é a presença de pequenas inserções e deleções, que variam de um a três nucleotídeos, conhecidas na literatura como *InDels*.

⁵Disponíveis respectivamente nos sites: <http://www.454.com>; <http://www.illumina.com>; <http://www3.appliedbiosystems.com>

Os SNPs e os *InDels* são as bases genéticas da maioria das variações alélicas e podem ser tratados como marcadores bialélicos, possuindo amplas aplicações no mapeamento genético de alta resolução e em testes diagnósticos. Tais marcadores podem ser identificados *in silico*, explorando os bancos de dados públicos e mais recentemente pelo ressequenciamento de genomas completos. Useche et al. (2001) identificaram 2.439 SNPs e 822 *InDels* no banco público contendo 68 mil ESTs de milho, utilizando algoritmos com interfaces específicas entre diferentes bancos de dados⁶. Assim, tais marcadores têm gerado dados em larga escala, de maneira automatizada e a custos reduzidos, sendo úteis para a construção de mapas genéticos de alta densidade, para estudos de associação genética e para seleção assistida com base no genoma completo.

Uma vez identificados *in silico*, os marcadores precisam ser convertidos em pares de *primers* específicos (STS), que amplifiquem apenas a região que contém a mutação, para posterior validação e mapeamento destes. No caso de *InDels*, as inserções/deleções de nucleotídeos podem ser reveladas diretamente em géis de sequenciamento dos produtos de amplificação, utilizando um dos *primers* marcados com fluorescência ou radioatividade. Para a detecção das substituições de base, que ocorrem nos SNPs, têm sido desenvolvidas várias estratégias, normalmente utilizando-se etapas adicionais de extensão de *primers* que ancoram a uma base de distância do SNP, após a amplificação do fragmento de DNA que contém a sequência alvo. Várias técnicas têm sido adaptadas para revelar os SNPs, como a extensão de *primers*, que utiliza a detecção por espectroscopia de massa e PCR em tempo real.

No entanto, é importante o desenvolvimento de sistemas para a visualização dos SNPs, que sejam mais simples e de

baixo custo, envolvendo uma única reação de PCR seguida de eletroforese em géis de agarose ou poliacrilamida. Essas características permitem a sua aplicação em programas de melhoramento genético vegetal, em laboratórios com estruturas mais simples. Dentre estas, destaca-se a técnica *amplification refractory mutation system* (ARMS-PCR), descrita por Ye et al. (2001). O sistema ARMS-PCR consiste da amplificação com uma combinação de dois *primers* alelo-específicos, um para cada alelo SNP, e dois *primers* externos e comuns (Fig. 2). Os dois amplicons alelo-específicos são gerados pela combinação de cada um dos *primers* externos com cada um dos *primers* alelo-específicos. A especificidade alélica é conferida por uma diferença entre a base no terminal 3'

dos *primers* internos e a base do alelo alternativo do SNP, como mostrado na Figura 2. Para aumentar a especificidade, um segundo erro é adicionado deliberadamente na posição -2, a partir do terminal 3' (indicado por um asterisco na Fig. 2). O sistema pode ser desenhado para ensaios co-dominantes, com a amplificação simultânea dos dois alelos (sistema de quatro *primers*) ou amplificação em separado para cada um dos alelos (sistema de três *primers*) (BUNDOCK et al., 2006). No sistema ARMS, é importante notar que o desenho de *primers* compatíveis para amplificação simultânea é, muitas vezes, desafiador. Programas, com base em *web*⁷ estão disponíveis para essa finalidade e auxiliam na escolha dos *primers* ideais, para o sistema de três e quatro *primers*.

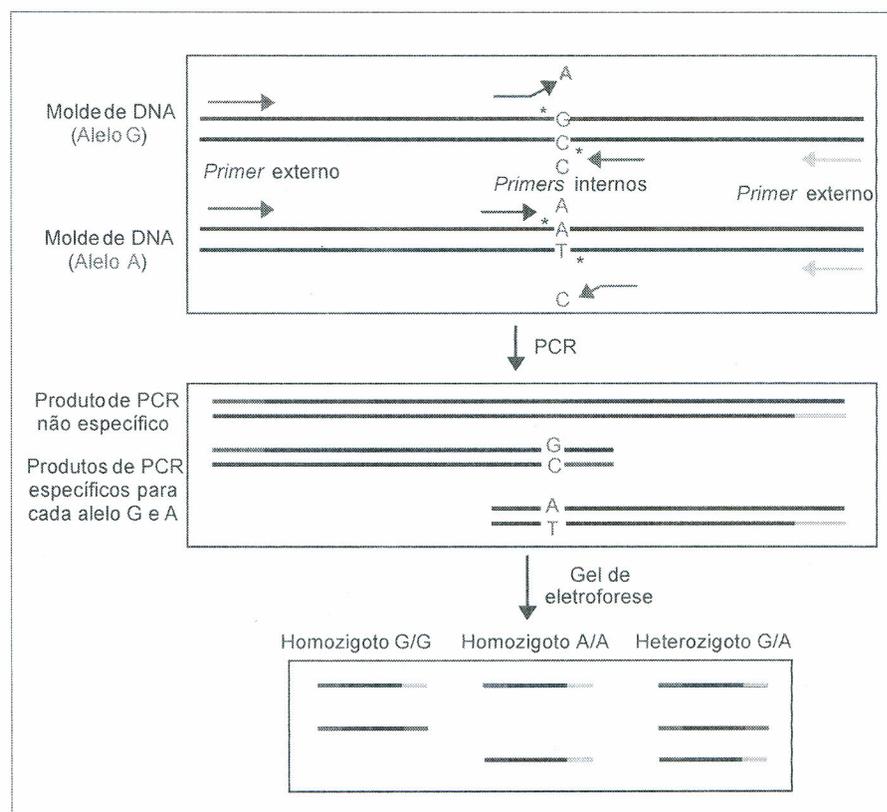


Figura 2 - Esquema de amplificação pela técnica ARMS utilizando quatro *primers*
 FONTE: Ye et al. (2001).

NOTA: DNA – Ácido desoxirribonucleico; PCR – Polymerase chain reaction.

⁶Ests – Expressed sequence tags.

⁷Disponível no site: <http://wheat.pw.usda.gov/demos/BatchPrimer3/>

As tecnologias de sequenciamento de próxima geração têm disponibilizado uma quantidade muito grande de informações de SNPs, que tem recebido plataformas para genotipagem simultânea de milhares de marcadores a custos reduzidos, como Golden Gate da Illumina® e OpenArray da Applied Biosystems™.

SFP

Os microarranjos de DNA ou *chips*, que contêm oligonucleotídeos desenhados para estudos de expressão gênica, estão sendo usados para a geração de polimorfismos para genotipagem em grande escala, denominados *single feature polymorphism* (SFP). Tais marcadores têm sido identificados com sucesso desde genomas simples como arroz (KUMAR et al., 2007), até genomas complexos como trigo (BERNARDO et al., 2009), os quais apresentam elevada capacidade de amostragem genômica, utilizando a mesma estrutura e os *chips* de DNA aplicados em análise da expressão global.

Cada polimorfismo é avaliado pela presença ou ausência de sinais derivados da hibridização do DNA genômico, marcado com os oligonucleotídeos impressos nos *chips*, como os da Affymetrix®. Além dos microarranjos já disponíveis, existe a possibilidade de desenvolver *chips* para genotipagem por SFP pelo alinhamento das sequências genômicas e de ESTs públicas, culminando com desenho de sondas de 70 pares de base que cobrem todo o genoma da espécie alvo.

APLICAÇÕES DOS MARCADORES MOLECULARES NO MELHORAMENTO GENÉTICO

Diversidade genética e seleção de genitores

Os diferentes tipos de marcadores moleculares disponíveis apresentam uma ampla capacidade de amostragem do ge-

noma, sendo de grande potencial para a avaliação da diversidade genética, tanto para aplicações filogenéticas e evolutivas, quanto para fins práticos em programas de melhoramento genético e na manutenção de bancos de germoplasma. Os genótipos são avaliados por meio dos marcadores, as bandas comuns a todos os indivíduos são interpretadas como semelhanças genéticas e as bandas não-comuns, como diferenças. Os resultados são codificados, a fim de gerar uma matriz de similaridade ou dissimilaridade, interpretada por meio de análise de agrupamento ou multivariada.

A escolha dos genitores e o planejamento dos cruzamentos são etapas de fundamental importância para o sucesso de um programa de melhoramento. Essas etapas podem ser auxiliadas por marcadores moleculares e fornecem aos melhoristas informações genéticas adicionais e mais detalhadas dos genótipos, aumentando a probabilidade de obtenção de cultivares superiores. A seleção de genitores superiores é mais crítica em espécies perenes, como fruteiras e essências florestais, uma vez que o tempo para obtenção de uma cultivar melhorada é bastante longo.

Oliveira (2009) obteve uma correlação de 63% entre as distâncias genéticas conseguidas por marcadores microssatélites e as distâncias obtidas pela genealogia de 32 cultivares brasileiras de soja, originárias de um único programa de melhoramento. A análise de agrupamento realizada com dados de marcadores microssatélites coincidiu perfeitamente com a origem das cultivares, com base em sua genealogia. Esse autor ressalta, ainda, que os dados de genealogia nem sempre estão disponíveis, enquanto as informações dos marcadores sempre podem ser obtidas com acurácia.

A caracterização de linhagens de milho em grupos heteróticos é de fundamental importância na escolha dos genitores para a obtenção de híbridos com elevada *performance* agronômica. Aguiar et al. (2008), ao utilizarem marcadores microssa-

télites para formação de grupos heteróticos em milho, obtiveram formação de grupos semelhantes àqueles obtidos por meio de cruzamentos testes. No entanto, o agrupamento que utilizou marcadores moleculares foi realizado em um mês, enquanto que a obtenção dos resultados de avaliação dos cruzamentos testes necessitou de dois ciclos da cultura do milho. Além disso, na avaliação fenotípica, muitos dados foram perdidos, em função da instabilidade climática durante os períodos de avaliação. Um conjunto de 770 linhagens de milho, representando germoplasmas melhorados do México, China e Brasil, foi avaliado com 1.034 SNPs, sendo possível separar claramente as linhagens temperadas e tropicais (LU et al., 2009). Entre as linhagens da China e do Brasil foi possível definir os agrupamentos em função da genealogia, sendo também definido um conjunto de SNPs mais informativos entre as coleções de germoplasma de diferentes origens.

DNA fingerprinting e proteção de cultivares

A caracterização de variedades, linhagens ou híbridos, por meio de marcadores de DNA, tem sido de grande importância na proteção intelectual do obtentor, sendo utilizada como prova legal em processos jurídicos nos países onde vigoram as leis de proteção de cultivares. Com a aprovação, no Brasil, da Lei de Proteção de Cultivares (BRASIL, 1997), o Serviço Nacional de Proteção de Cultivares (SNPC), vinculado ao Ministério da Agricultura, Pecuária e Abastecimento (MAPA), ficou responsável pelo registro e proteção de novas variedades. Para que uma variedade seja protegida, é necessário demonstrar que é diferente de qualquer outra variedade da mesma espécie. Apesar de o processo para a proteção de cultivares ser realizado com base em descritores morfológicos, os marcadores moleculares têm sido estimulados e aceitos nos processos para a identificação de cultivares.

®Consultar o site: <http://www.affymetrix.com>

Um sistema com base em marcadores de DNA permite a identificação de um padrão único de combinação alélica, como uma impressão digital (*fingerprinting*) para cada cultivar, facilitando sua proteção e assegurando os direitos de propriedade intelectual. A capacidade de discriminação dos métodos de *fingerprinting* molecular depende do número de marcadores utilizados e da frequência alélica para cada loco dentro de uma população representativa da espécie. Ao avaliar um conjunto de 32 cultivares de soja de um mesmo programa de melhoramento, portanto com grande probabilidade de ser estreitamente relacionadas, Oliveira (2009) selecionou um conjunto de 11 marcadores microssatélites, que identificaram todas as cultivares com no mínimo 99,9999% de probabilidade de exclusão. Foram distinguidas inclusive duas cultivares que possuíam 99,22% de similaridade genética, com base na genealogia.

Análise de pureza genética de sementes

Da mesma forma como os marcadores de DNA são úteis na identificação de cultivares, tornam-se também uma importante ferramenta na identificação das misturas varietais em análises de pureza genética de sementes, com a finalidade de certificação ou fiscalização de sementes. Comumente, misturas varietais são identificadas com base em *grow out* no campo ou em características morfológicas de sementes e plântulas ou, em alguns casos, pela eletroforese de proteínas da semente. Em soja, por exemplo, a análise de pureza genética de sementes é feita comumente pela cor do hilo e do hipocótilo e reação à peroxidase. Esses caracteres são pouco informativos, pois tanto a cor do hipocótilo, quanto a reação à peroxidase só permitem dois resultados possíveis, sendo que a combinação de ambos só permite classificar as variedades em quatro grupos distintos.

Cor de hilo apresenta um número maior de resultados possíveis, representados pelas diversas cores e tonalidades existentes, que podem ser alteradas em função de condições ambientais. Por exemplo, variedades com hilo preto imperfeito podem ser confundidas com variedades de hilo preto ou mesmo marrom. Análises moleculares têm demonstrado que nem sempre a diferença na cor do hilo significa mistura varietal. Em alguns casos, sementes de hilo preto imperfeito encontradas em lotes de sementes de hilo marrom são geneticamente idênticas.

Marcadores de DNA são uma ferramenta auxiliar e poderosa na análise de pureza genética de sementes e outras fontes de propagação, tais como mudas, estolões, estacas, borbulhas, etc., aplicável sempre que a caracterização visual gerar dúvidas. Essa análise tem sido feita explorando-se ao máximo a representatividade do genoma, pela utilização de *primers* bem distribuídos em mapas genéticos da espécie. Schuster et al. (2004) descrevem uma metodologia prática para a utilização de marcadores microssatélites na avaliação de pureza genética de cultivares de soja que pode ser facilmente estendida para as demais culturas.

Mapeamento de genes e características complexas

Os avanços nas áreas dos marcadores moleculares e das metodologias estatísticas implementadas em programas computacionais têm aumentado a saturação dos mapas genéticos e, conseqüentemente, a precisão na identificação de marcadores associados com as características de interesse. Para um marcador ser utilizado no mapeamento genético, este deve ser polimórfico entre os indivíduos parentais e apresentar uma segregação mendeliana esperada na prole. Vários tipos de populações segregantes podem ser utilizadas na construção de mapas genéticos, como F_2 , linhagens recombinantes endogâmicas – *recombinant*

inbred lines (RILs) e de retrocruzamentos, tendo cada uma vantagens e limitações. Vários programas computacionais estão disponíveis para a construção de mapas de ligação, sendo alguns públicos como o MapMaker (LANDER et al., 1987), One-Map (MARGARIDO; SOUZA; GARCIA, 2007) e GQMOL⁹.

O grande objetivo da construção de mapas genéticos é identificar marcadores moleculares associados às características de interesse. Há uma infinidade de trabalhos que reportam ao mapeamento de genes e ao QTLs em diferentes espécies. Resistência a doenças tem sido uma característica alvo para esse tipo de estudo, uma vez que a identificação de marcadores ligados a genes de resistência permite monitorar a introgressão e a piramidação de dois ou mais genes de resistência em uma cultivar ou linhagem comercial.

Garcia et al. (2008) e Silva et al. (2008) mapearam os genes *Rpp5*, *Rpp2* e *Rpp4*, que conferem resistência à ferrugem da soja. Adicionalmente, mapeamento de genes é um passo importante para a clonagem de genes.

A maioria das características de importância econômica está sob controle genético complexo, envolvendo a ação de vários genes, o que torna difícil sua manipulação e compreensão. Regiões genômicas contendo locos gênicos associados a tais caracteres quantitativos são denominados QTLs. Marcadores moleculares analisados por metodologias estatísticas implementadas em programas de mapeamento de QTLs têm possibilitado a dissecação dos caracteres complexos em fatores mendelianos simples. O mapeamento de QTLs possibilita estimar o número e a localização de genes que controlam a variação fenotípica de um caráter, a magnitude dos seus efeitos e as interações com outros QTLs. A habilidade de detectar QTLs por meio de marcadores moleculares é função da magnitude do efeito do QTL, do tamanho e do tipo da população e do nível de saturação do mapa genético.

⁹<http://www.ufv.br/dbg/gqmol/gqmol.htm>

O mapeamento de QTLs baseia-se em testes de associação entre marcadores moleculares e os dados fenotípicos por meio de várias metodologias estatísticas, cujos procedimentos podem ser vistos com mais detalhes em vários livros especializados como, Schuster e Cruz (2008), dentre outros. De forma bastante resumida, os métodos para mapeamento de QTLs avançaram desde a associação com marcas simples, passando pelo mapeamento por intervalo simples, mapeamento por intervalo composto, por múltiplos QTLs, até o mapeamento por intervalos múltiplos. Programas específicos para o mapeamento de QTLs, contendo algoritmos apropriados para implementar vários dos procedimentos estatísticos descritos estão disponíveis para acesso público, como QTLCartographer¹⁰ e QTMOL.

Métodos de mapeamento de populações derivadas de cruzamentos biparentais amostram apenas uma pequena fração de todos os possíveis alelos em uma determinada espécie. Por esse motivo, marcadores moleculares associados a QTLs geralmente só podem ser usados nas populações em que foram desenvolvidos. Recentemente, a utilização do desequilíbrio de ligação, ou seja, a associação não ao acaso entre alelos em locos ligados, tem sido utilizada para realizar o mapeamento por associação. Diferentemente do mapeamento de ligação (equilíbrio de ligação), o mapeamento por desequilíbrio de ligação infere a associação entre genótipos (haplótipos) e a característica, pela avaliação do polimorfismo genético que foi gerado em diferentes *backgrounds* genéticos e por meio de muitas gerações de recombinação (DEKKERS; HOSPITAL, 2002).

O desequilíbrio de ligação é definido como a associação não ao acaso, de alelos em locos diferentes do mesmo cromossomo. Se for considerado que um determinado loco possui os alelos *A* e *a* com frequências de p_A e $1-p_A$, e um segundo loco possui

os alelos *B* e *b*, em frequências de p_B e $1-p_B$, então, no equilíbrio, mesmo que os locos estejam ligados, a frequência esperada de um determinado haplótipo é o produto das frequências dos alelos constituintes desse haplótipo. Por exemplo, a frequência esperada do haplótipo AB, no equilíbrio, é $p_{AB} = p_A \times p_B$. O desequilíbrio de ligação é a diferença $D = p_{AB} - (p_A \times p_B)$. No equilíbrio, $D = 0$ (MACKAY; POWELL, 2007).

O mapeamento por desequilíbrio de ligação, também conhecido como mapeamento de associação, detecta e localiza QTLs com base na intensidade da correlação entre marcadores moleculares mapeados e as características fenotípicas. O desequilíbrio de ligação entre dois locos é função do tempo (número de gerações) decorrido, desde que se iniciaram as gerações de recombinação e a frequência de recombinação entre os locos. Após várias gerações de recombinação, em uma população não estruturada, apenas correlações entre QTLs e marcadores proximamente ligados devem permanecer, facilitando o mapeamento mais fino (MACKAY; POWELL, 2007).

Para o mapeamento por desequilíbrio de ligação, não há necessidade de preparar uma população de mapeamento e o genoma inteiro é avaliado para identificar regiões associadas a um fenótipo em particular. O primeiro obstáculo para essa análise de associação é a presença, dentro da população, de subgrupos com distribuição desigual de alelos. Nesses casos, falsas associações positivas entre marcadores e fenótipo podem ser identificadas, mesmo que não haja ligação física entre os marcadores e os genes (BUCKLER; THORNSBERRY, 2002). Por esse motivo, a primeira etapa no mapeamento por associação (desequilíbrio de ligação) é identificar a existência de subgrupos. Isso pode ser feito pela construção de uma matriz de similaridade genética, estimada a partir dos marcadores moleculares.

Melhoramento genético assistido por marcadores moleculares

Retrocruzamento assistido

A utilização de marcadores moleculares em programas de introgressão de genes por meio de retrocruzamento é um dos exemplos mais concretos de melhoramento assistido por marcadores. Germoplasmas não-adaptados têm sido utilizados em programas de melhoramento com o objetivo de introduzir características de herança simples, como resistência a doenças ou transgenes. A seleção pode ser feita para monitorar o gene de interesse ou o genoma recorrente, visando à redução do tempo para obtenção de linhagens convertidas e aumento da eficiência do programa, principalmente quando se objetiva a pirâmidação de dois ou mais genes de interesse.

Diferentes estratégias de retrocruzamento assistido podem ser aplicadas dependendo dos objetivos, sendo que o custo com as genotipagens pode ser compensado pela redução significativa no número de gerações para recuperar o genoma recorrente, principalmente com o alto valor agregado de um evento transgênico (GUIMARÃES et al., 2009). Ainda, considerando a significativa redução dos custos das novas tecnologias de marcadores, em situações onde o fenótipo é de avaliação complexa, a seleção com marcadores moleculares fornece a vantagem adicional na redução dos custos e na seleção precoce dos indivíduos de interesse.

SAM

A seleção de características simples com o auxílio de marcadores moleculares tem grande utilidade, quando a característica é de difícil medição ou quando se deseja selecionar para diversas características simultaneamente, como no caso de pirâmidação de genes. No entanto, a seleção assistida por marcadores (SAM) tem sua maior contribuição na seleção de caracte-

¹⁰Disponível no site: <http://statgen.ncsu.edu/qtlcart/cartographer.html>

rísticas quantitativas. Associações entre alelos de marcadores moleculares e alelos de QTLs podem ser usadas para selecionar, indiretamente, alelos favoráveis de QTLs. Para caracteres de baixa herdabilidade, a seleção fenotípica é realizada em gerações mais avançadas, pois a herdabilidade e a acurácia estatística da estimação das médias das progênies aumentam com o aumento do número de repetições, gerações, locais e anos de teste, o que leva a um aumento muito grande do número de plantas a ser avaliado. Assim, os melhoristas devem optar entre produzir estimativas mais precisas da média de um número menor de progênies ou amostrar um grande número de progênies, com estimativas menos precisas. A SAM apresenta-se como uma alternativa eficiente para permitir a seleção precoce de características de baixa herdabilidade. A seleção precoce diminui o número de plantas a ser analisado por família, o que permite a avaliação de um número muito maior de progênies. A SAM é mais efetiva em gerações iniciais de seleção entre progênies de cruzamentos entre linhas endogâmicas. Nessas gerações, a herdabilidade é baixa, pois o número de repetições é limitado e o desequilíbrio de ligação é alto. O paradoxo é que o poder de mapeamento de QTLs decresce, à medida que a herdabilidade decresce, e é menor justamente para características que teoricamente sofreriam o maior impacto pela SAM (LANDE; THOMPSON, 1990). Para evitar a detecção de falsas associações entre marcadores e QTLs, um nível de significância bastante alto é utilizado para se declarar associação entre um marcador e um QTL. A utilização de níveis de significância bastante estridentes na escolha dos marcadores aumenta a acurácia da SAM e produz ganhos significativos desta sobre a seleção fenotípica. Trabalhos mais recentes tendem a utilizar a regressão múltipla do fenótipo aos marcadores, como o método de identificar a associação entre marcadores e QTLs, e estimar o efeito dos marcadores (LANDE; THOMPSON, 1990). Regressão múltipla é um procedimento

computacionalmente mais simples do que o método da máxima verossimilhança, produzindo resultados muito similares (LANDE; THOMPSON, 1990).

Os resultados de estudos teóricos e experimentais permitem concluir que a SAM pode ser utilizada para aumentar substancialmente a eficiência dos programas de melhoramento (EDWARDS; STUBER; WENDEL, 1987; STUBER; EDWARDS; WENDEL, 1987; LANDE; THOMPSON, 1990). A eficiência da seleção assistida por marcadores, comparada com a seleção fenotípica, é afetada por diversos fatores, tais como, herdabilidade da característica, cobertura do genoma pelos marcadores, identificação de associação entre marcadores e QTLs, tamanho e número de famílias, tipo de população e escolha do método de seleção assistida por marcadores. A seleção que utiliza marcadores flanqueando o QTL também é mais eficiente do que a seleção que se baseia em um único marcador. LANDE; THOMPSON (1990) propuseram a teoria do índice de seleção, utilizando marcadores, que maximizam a taxa de ganhos genéticos pela combinação das informações do polimorfismo nos locos dos marcadores genéticos com dados da variação fenotípica entre os indivíduos. Esse índice pode ser usado tanto para a seleção individual, como para a seleção de linhas. Porém, a utilização desse índice requer testes de progênies com repetições, para a obtenção dos valores fenotípicos. A seleção que se baseia apenas em marcadores pode ser utilizada e possui eficiência semelhante ao índice, quando $p > 0,5$ para características de baixa herdabilidade, sendo p a proporção da variância genética aditiva da característica que é associada com o loco marcador.

Muitas características de importância econômica são afetadas por vários QTLs. Além de, em muitos casos, os QTLs para diferentes características apresentarem-se ligados (STUBER; EDWARDS; WENDEL, 1987; EDWARDS; STUBER; WENDEL, 1987). Como na maioria dos programas de melhoramento, o objetivo

é melhorar a *performance* para diversas características simultaneamente, o número de QTLs envolvidos em um programa de melhoramento facilitado por marcadores moleculares deve ser grande. Pela natureza intrínseca dos QTLs e pelo fato de caracteres quantitativos serem controlados por um grande número de genes de pequeno efeito, a seleção recorrente assistida por marcadores moleculares é uma escolha apropriada para o melhoramento de populações. Por meio de ciclos sucessivos de intercrossamentos e seleção, a seleção recorrente favorece a quebra de ligações genéticas, o que facilita a recombinação entre genes e aumenta a frequência de alelos favoráveis nas populações de melhoramento. O uso da seleção assistida por marcadores moleculares permite a identificação de combinações favoráveis de alelos, permitindo maiores ganhos com a seleção. Xie; Xu (1998) apresentaram uma comparação teórica entre diversos métodos de seleção recorrente com base em famílias e seleção massal, utilizando SAM e seleção fenotípica. A seleção recorrente assistida por marcadores foi duas vezes mais eficiente do que a seleção fenotípica, com famílias de tamanho pequeno, baixa herdabilidade e alta proporção da variação fenotípica explicada pelos marcadores (p). Com o aumento do tamanho da população, aumento da herdabilidade e diminuição de p , a vantagem da seleção assistida por marcadores diminuiu. Seleção assistida por marcadores é mais competitiva com a seleção fenotípica tradicional, quando a característica possui baixa herdabilidade. Se o objetivo do programa de melhoramento é maximizar a resposta por unidade de custo, a SAM pode ser inferior à seleção fenotípica, quando a herdabilidade é superior a 30% e o custo da avaliação fenotípica por indivíduo é menor do que a genotipagem do loco marcador.

O número de artigos científicos publicados anualmente com o termo *marker-assisted selection* ultrapassou a barreira de mil, no ano de 2003 (XU; CROUCH, 2008). No entanto, a quase totalidade desses artigos procura demonstrar a apli-

cação potencial da SAM nos programas de melhoramento e não sua aplicação prática. Assim, o uso efetivo da SAM no desenvolvimento de variedades em programas de melhoramento tem-se limitado, em sua maior parte, às grandes empresas multinacionais, que têm desenvolvido ferramentas genômicas para as espécies de maior interesse comercial, tais como milho, soja, canola, algodão e girassol. Os programas de melhoramento que utilizam essas ferramentas têm desenvolvido estratégias para gerar um genótipo ideal, a partir da seleção de um mosaico de segmentos cromossômicos favoráveis (XU; CROUCH, 2008).

Utilizando essas ferramentas, tais programas de melhoramento têm relatado taxas de ganhos genéticos duas vezes superiores aos ganhos realizados com base na seleção fenotípica (XU; CROUCH, 2008). Estima-se que, nos Estados Unidos, a partir do ano de 2010, 12% das variedades comerciais sejam desenvolvidas por meio do melhoramento molecular (FRALEY, 2006).

Converter as informações publicadas na literatura científica em aplicações práticas de larga escala nos programas de melhoramento requer algumas considerações de ordem prática, logística e genética. Primeiramente, os marcadores moleculares publicados devem ser validados em um grande número de populações que sejam representativas do material de melhoramento. Em seguida, é necessário desenvolver um procedimento técnico simples, rápido e econômico para as etapas de amostragem de tecidos, extração de DNA, genotipagem e coleta de dados que seja viável e preciso quando aplicado rotineiramente em larga escala. Além disso, os melhoristas precisam desenvolver um sistema integrado, com rastreabilidade dos dados e sistemas de controle que assegurem a integração da genotipagem nos programas de melhoramento. Por fim, é fundamental delinear um sistema de melhoramento que permita otimizar ferramentas de tomada de decisão, para auxiliar melhoristas na seleção de forma rápida e precisa (XU; CROUCH, 2008).

REFERÊNCIAS

- AGUIAR, C.G. et al. Heterotic groups in tropical maize germplasm by test crosses and simple sequence repeat markers. **Genetics and Molecular Research**, v.7, n.4, p.1233-1244, 2008.
- BERNARDO, A.N. et al. Discovery and mapping of single feature polymorphisms in wheat using Affymetrix arrays. **BMC Genomics**, v.10, p.251, 2009.
- BRASIL. Lei nº 9.456, de 25 de abril de 1997. Institui a Lei de Proteção de Cultivares e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, 28 abr. 1997. Disponível em: <http://www.planalto.gov.br/ccivil_03/Leis/L9456.htm>. Acesso em: 2009.
- BUCKLER IV, E.S.; THORNSBERRY, J.M. Plant molecular diversity and applications to genomics. **Current Opinion in Plant Biology**, v.5, n.2, p.107-111, Apr. 2002.
- BUNDOCK, P.C. et al. Robust allele-specific polymerase chain reaction markers developed for single nucleotide polymorphisms in expressed barley sequences. **Theoretical and Applied Genetics**, v. 112, n.2, p.358-365, Jan. 2006.
- CHING, A. et al. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. **BMC Genetics**, v.3, p.19, Oct. 2002.
- DEKKERS, J.C.M.; HOSPITAL, F. Multifactorial genetics: the use of molecular genetics in the improvement of agricultural populations. **Nature Reviews Genetics**, v.3, n.1, p.22-32, Jan. 2002.
- DIVERSITY ARRAYS TECHNOLOGY. **Diversity arrays technology Pty Ltda (DART P/L)**. [S.l., 200-]. Disponível em: <<http://www.diversityarrays.com>>. Acesso em: 2009.
- EDWARDS, M.D.; STUBER, C.W.; WENDEL, J.F. Molecular-marker-facilitated investigations of quantitative-trait loci in maize - I: numbers, genomic, distribution and type of gene action. **Genetics**, v.116, n.1, p.113-125, May 1987.
- FRALEY, R. **Monsanto european investor day**. Saint Louis: Monsanto, 2006. Disponível em: <http://www.monsanto.com/dpf/investors/2006/EU_RobbFraley_1110006.pdf>. Acesso em: 30 abr. 2008.
- GARCIA, A. et al. Molecular mapping of soybean rust (*Phakopsora pachyrhizi*) resistance genes: discovery of a novel locus and alleles. **Theoretical and Applied Genetics**, v.117, n.4, p. 545-553, Aug. 2008.
- GUIMARÃES, C.T. et al. Aplicação de marcadores moleculares no melhoramento de plantas. In: BOREM, A.; CAIXETA, E. (Ed.). **Marcadores moleculares**. Viçosa, MG: UFV, 2009. p.129-175.
- JACCOUD, D. et al. Diversity arrays: a solid state technology for sequence information independent genotyping. **Nucleic Acids Research**, v.29, n.4, p.e25, Feb. 2001.
- KONIECZNY, A.; AUSUBEL, F.M. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. **Plant Journal**, v.4, p. 403-410, 1993.
- KUMAR, R. et al. Single feature polymorphism discovery in rice. **PLoS ONE**, v.2, n.3, p.e284, 2007.
- LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits genetics. **Genetics**, v.124, n.3, p.743-756, Mar. 2007.
- LANDER, E.S. et al. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. **Genomics**, v.1, n.2, p.174-181, Oct. 1987.
- LU, Y. et al. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. **Theoretical and Applied Genetics**, v.120, n.1, p.93-115, Dec. 2009.
- MACE, E.S. et al. DArT markers: diversity analyses and mapping in *Sorghum bicolor*. **BMC Genomics**, v.9, p.26, 2008.
- MACKAY, I.; POWELL, W. Methods for linkage disequilibrium mapping in crops. **Trends in Plant Science**, v.12, n.2, p.57-63, Feb. 2007.
- MARGARIDO, G.R.A.; SOUZA, A.P.; GARCIA, A.A.F. OneMap: software for genetic mapping in outcrossing species. **Hereditas**, v.144, n.3, p.78-79, July 2007.
- NEFF, M.M. et al. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. **Plant**

Journal, v.14, n.3, p.387-392, May 1998.

OLIVEIRA, M.B. **Caracterização molecular de cultivares de soja utilizando marcadores microssatélites genotipados em sequenciador automático**. 2009. 88p. Dissertação (Mestrado) – Universidade Paranaense, Umuarama.

OLSON, M. et al. A common language for physical mapping of the human genome. *Science*, v.254, p.1434-1435, 1989.

PARAN, I.; MICHELMORE, R.W. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics*, v.85, n.8, p.985-993, Feb. 1993.

SCHUSTER, I.; CRUZ, C.D. **Estatística genômica aplicada a populações derivadas de cruzamentos controlados**. 2. ed. Viçosa, MG: UFV, 2008. 586p.

_____. et al. Determinação da pureza varietal de sementes de soja com auxílio de marcadores moleculares microssatélites. *Pesquisa Agropecuária Brasileira*, Brasília, v.39, n.3, p.247-253, Mar. 2004.

SILVA, D.C.G. et al. Molecular mapping of two loci that confer resistance to Asian rust in soybean. *Theoretical and Applied Genetics*, v.117, n.1, p.57-63, June 2008.

STUBER, C.W.; EDWARDS, M.D.; WENDEL, J.F. Molecular marker-facilitated investigations of quantitative trait loci in maize - II: factors influencing yield and its component traits. *Crop Science*, v.27, n.4, p.639-648, July 1987.

USECHE, F.J. et al. High-throughput identification, database storage and analysis of SNPs in EST sequences. *Genome Informatics*, v.12, p.194-203, 2001.

XIE, C.; XU, S. Strategies of marker-aided recurrent selection. *Crop Science*, v.38, n.6, p.1526-1535, Dec. 1998.

XU, Y.; CROUCH, J.H. Marker-assisted selection in plant breeding: from publications to practice. *Crop Science*, v.48, n.2, p.391-407, Apr. 2008.

YE, S. et al. An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Research*, v.29, n.17, p.e88, Sept. 2001.

ZHU, Y.L. et al. Single-nucleotide polymorphisms in soybean. *Genetics*, Austin, v.163, n.3, p.1123-1134, Mar. 2003.

ZIETKIEWICZ, E.; RAFALSKI, A.; LABUDA, D. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics*, v.20, n.2, p.176-183, Mar. 1994.

186 mm x 140 mm