



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA
CENTRO DE CIÊNCIAS BIOMÉDICAS
CURSO DE CIÊNCIAS BIOLÓGICAS**

**BASE GENÉTICA EM LINHAGENS DE SOJA [*Glycine max* (L.) Merr.]
POR MEIO DA ANÁLISE DE MARCADORES MICROSSATÉLITES**

Kátia Bernardeli

Orientador: Prof. Dr. Luiz Ricardo Goulart Filho

Monografia apresentada à
Coordenação do Curso de Ciências
Biológicas da Universidade Federal
de Uberlândia para a obtenção do
grau de Bacharel em Ciências
Biológicas.

Uberlândia - MG
Dezembro - 1999



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
CENTRO DE CIÊNCIAS BIOMÉDICAS
CURSO DE CIÊNCIAS BIOLÓGICAS

**BASE GENÉTICA EM LINHAGENS DE SOJA [*Glycine max* (L.) Merr.] POR MEIO
DA ANÁLISE DE MARCADORES MICROSSATÉLITES**

Kátia Bernardeli

Aprovada pela banca examinadora em 17/12/99

Nota 100.00

Prof. Dr. Luiz Ricardo Goulart Filho
Orientador

Dr. Glóverson Lamego Moro
Có-Orientador

MSc. Cristhiane Regina Abegg Bothona
Có-Orientadora

Uberlândia, 17 de dezembro de 1999

*Aos meus pais, Mauro e Lúcia, que sempre me incentivaram
na busca dos meus objetivos;*

*À minha irmã Larissa, a dádiva de Deus que cruzou
meu caminho;*

Ao meu primo Eduardo, pelos inúmeros favores.

Dedico

AGRADECIMENTOS

Em especial, agradeço aos meus orientadores, Prof. Luiz Ricardo Goulart Filho, pela confiança e incentivo depositados; Dr. Glóverson Lamego Moro, pelos ensinamentos e exemplo de pesquisador; MSc. Cristhiane Abegg Bothona, pelo dinamismo e sadia convivência que experimentei durante os últimos seis meses. Confesso que não há distinção entre essas pessoas na elaboração e execução deste trabalho. Todos foram, verdadeiramente, orientadores!

Agradeço ainda, à Empresa *Novartis Seeds*, juntamente com seus funcionários, pela oportunidade oferecida na forma de estágio, o que contribuiu com meu crescimento profissional; em especial à Beatriz, César, Michelle, Karolina e Vagner Mateus, que me proporcionaram muitos momentos alegres.

Ao Dr. Howard Gabe, Dr. Césio Humberto Brito, Dr. Ademir Luiz Capelaro e Prof^a. Rosana de Cássia Oliveira, obrigada pelos momentos de discussão, foram valiosos na conclusão deste estudo.

Aos amigos Ana Paula e Francis, não há como expressar toda a minha gratidão... Fica aqui os agradecimentos e a saudade que sentirei dos momentos bons e ruins que passamos juntos!

À Tutora do PET/BIOLOGIA Prof^a. Ana Maria Bonetti, obrigada pelo exemplo de educadora e mãe, do qual convivi durante 4 anos e espero ter aprendido um pouco dos valores que sempre transmitiu.

Aos amigos Petianos: Bárbara, Braynner, Bruninho, Carlinhos, Cristiane, Daniel, Elisângela, Fernando Biase, Fernando Lourenço, Jaqueline, Kaila, Laiena, Luciana Londe, Luciana Paiva, Magno, Priscila, Raquel, Rodrigo e Wilson, vocês foram responsáveis por me ajudar a conhecer um lado mágico da vida: a arte da convivência.

Aos amigos da 42^a Turma de Ciências Biológicas, em especial, Fabiane, Tatyana, Francislene e Andreia, formamos um grupo capaz de saber diferir horas de estudo e humor, que na verdade sempre se misturavam!

À CAPES, pelo apoio financeiro.

A DEUS, pela oportunidade de agradecer...

ABREVIações

%	por cento (porcentagem)
µg	micrograma (s)
µL	microlitro (s)
µM	micromolar (s)
2n	diplóide
CTAB	hexadecyl trimethyl ammonium bromide
DNA	ácido desoxirribonucléico
dNTP	desoxinucleotídeo (s)
EDTA	ácido etilenodiaminotetra acético sal dissódico
F ₁	primeira geração entre dois cruzamentos
h	hora (s)
HCl	ácido clorídrico
M	molar
mg	miligrama (s)
MgCl ₂	cloreto de magnésio
min	minuto (s)
mL	mililitro (s)
mM	milimolar
NaCl	cloreto de sódio
ng	nanograma (s)
nm	nanômetro (s)
°C	grau (s) Celsius
OD	densidade ótica
pb	pares de bases
PCR	reação em cadeia da polimerase
pH	potencial hidrogeniônico
PIC	polymorphism information content
RNA	ácido ribonucléico
RNase	ribonuclease A
rpm	rotações por minuto
seg	segundo (s)
TBE	tris borato EDTA
TE	Tris EDTA
Tris HCl	hydroxymethyl aminomethane hydrochloride
U	unidade (s) de enzima
UPGMA	unweighted pair-group method using arithmetic averages
UV	ultra-violeta
V	volts
X	fator de concentração de solução

RESUMO

A soja, [*Glycine max* (L.) Merr.], cultura autógama, com $2n = 40$ cromossomos, apresenta baixo nível de polimorfismo, explicado não só pelo seu modo de propagação, mas também pela enorme seleção artificial a qual tem sido submetida ao longo do seu melhoramento. Por ser a classe de marcadores mais informativa, a técnica de microssatélites (SSR) tem sido usada para estudos de DNA *fingerprinting*, desenvolvimento de marcadores e mapas genéticos desta cultura. Este trabalho teve como objetivo caracterizar, geneticamente, 16 linhagens de soja, dos quais três acessos foram constituídos de soja selvagem e o restante comercial, por meio de marcadores tipo microssatélites, visando acessar sua variabilidade genética e fornecer subsídios aos programas de melhoramento. Para análise do material, o protocolo MINIPREP foi usado para a extração de DNA, seguido de amplificação com 30 pares de *primers*. Os produtos amplificados foram separados por eletroforese horizontal em gel de agarose *metaphor*® 3% e as imagens digitalizadas por meio do *software Multi-Analyst*®. Para cada marcador utilizado no estudo foi calculado o valor de PIC (*Polymorphism Information Content*), as distâncias genéticas foram calculadas pelo método de Porcentagem de Desacordo (PD) e a análise de *cluster* por UPGMA. Da análise de microssatélites resultaram 109 alelos, com tamanho variando entre 20 e 500 pares de bases. O número de alelos detectados variaram de 2 a 5 e os valores de PIC, 0.12 a 0.73, de acordo com o marcador. O nível de informatividade (PIC) é uma medida importante na escolha de marcadores, pois demonstram, além do número de alelos expressos, a capacidade discriminatória dos mesmos entre as linhagens de interesse. O polimorfismo de *loci* SSR detectados por meio de agarose *metaphor*® mostrou-se eficiente, pois separa fragmentos distintos em até 10 pares de bases, além de ser uma técnica eficiente para laboratórios que buscam resultados em pouco tempo. A distância genética limite calculada entre os 16 genótipos de soja analisados foi de 42%. Genótipos selvagens formaram um grupo separado com relação às linhagens comerciais no dendrograma. As informações geradas no estudo de DNA *fingerprinting* de soja serão importantes num programa de melhoramento, pois permitirão guiar os cruzamentos, de forma a evitar aqueles endogâmicos, desenvolver marcadores de cruzamento, caracterizar novas cultivares comerciais e na conversão de linhagens.

SUMÁRIO

ABREVIACÕES	iv
RESUMO	v
INTRODUÇÃO	
1.1 - A cultura da Soja.....	01
1.2 - Marcadores Moleculares.....	02
1.3 - DNA <i>fingerprinting</i>	05
OBJETIVO.....	07
MATERIAL E MÉTODOS	
3.1. Material biológico.....	08
3.2. Extração de DNA.....	09
3.3. Quantificação e qualificação do DNA.....	10
3.4. Amplificação do DNA.....	11
3.5. Separação de fragmentos e visualização dos produtos amplificados.....	14
3.6. Análise de dados.....	14
RESULTADOS E DISCUSSÃO.....	16
CONCLUSÕES.....	24
REFERÊNCIAS BIBLIOGRÁFICAS.....	26

INTRODUÇÃO

1.1 - A cultura da Soja

A soja, [*Glycine max* (L.) Merr.], é uma das mais importantes leguminosas cultivadas no mundo. Originária da região leste da China, sua domesticação ocorreu por volta do século XI a.C.. A partir daí, foi introduzida em outras regiões do oriente e sudeste da Ásia. Sua introdução no ocidente deu-se a partir do século XII e, particularmente, no Brasil, em 1882 (EMBRAPA, 1996). Graças as características de elevado teor de proteínas (20%) e óleo (40%) das sementes, a soja representa 51% da produção mundial total das principais oleaginosas (ROESSING *et al.*, 1993). Isso gerou um vasto complexo agro-industrial destinado ao processamento de derivados dessa cultura, tornando-a uma das principais *commodities* do mundo (EMBRAPA, 1998). Graças ao melhoramento genético desenvolvido, realizado para adaptar a cultura às condições locais, o Brasil tornou-se o 2º produtor mundial (SOUZA, 1999). O país, hoje, participa com aproximadamente 18% da produção mundial, embora seja o país com maior potencial de expansão de área e produtividade (EMBRAPA, 1998). Na safra 1997/1998, a área cultivada desta cultura superou os 13.135.400 hectares, com uma produção ao redor de 31.355.800 milhões de toneladas (CONAB, 1998).

Botanicamente, a soja é classificada na família Fabaceae, subfamília Papilionoideae, tribo Phaseoleae, gênero *Glycine*, no qual há dois subgêneros: *Glycine* e *Soja*. É neste último que se encontram as espécies: [*Glycine max* (L.) Merr.], a soja cultivada e [*Glycine soja* (L.) Sieb. e Zucc.], sua provável progenitora. Planta autógama, com $2n = 40$ cromossomos, apresenta uma base genética estreita que é explicada não só pelo seu modo de propagação, mas também pela enorme seleção artificial a qual tem sido submetida ao longo do seu melhoramento (HADLEY e HYMOWITZ, 1973).

Visando obter uma maior produtividade, estabilidade, resistência a doenças e pragas e alta qualidade fisiológica das sementes, os melhoristas executam cruzamentos artificiais entre os genótipos de interesse. Tais cruzamentos são monitorados por meio marcadores morfológicos, como cor de flor e/ou hilo, para confirmação da geração F_1 (SEDIYAMA, 1986). Entretanto, o pequeno número de marcadores morfológicos distintos em uma mesma linhagem reduz a probabilidade de se encontrar associações significativas entre estes marcadores com características de importância econômica da cultura (FERREIRA e GRATTAPAGLIA, 1996). Assim, marcadores moleculares podem tornar a técnica de seleção eficiente e extremamente útil no melhoramento desta cultura, em conjunto com a tradicionalmente feita por meio da análise do fenótipo.

1.2 - Marcadores Moleculares

As limitações advindas pelos métodos de avaliação fenotípica puderam ser reduzidas a partir do advento da Biologia Molecular, que trouxe técnicas disponíveis para a detecção de polimorfismo diretamente ao nível de DNA, pontos de referência nos cromossomos denominados marcadores moleculares (FERREIRA e GRATTAPAGLIA, 1996). A análise de polimorfismos de DNA tem

se mostrado particularmente útil para a caracterização genética de fontes biológicas, pois estes são fenotipicamente neutros e não sujeitos aos efeitos do ambiente (FAIRBANKS *et al.*, 1993 *apud* BELELE, 1997). Métodos moleculares são importantes em estudos de variabilidade genética entre linhagens, na seleção de pais para cruzamentos, no reconhecimento das variedades para fins legais e no controle de pureza no processo de produção de sementes (DUDLEY, 1993).

A utilização de marcadores moleculares na análise de variabilidade tem um grande número de aplicações no melhoramento de plantas. Ela permite o entendimento e organização da base genética das linhagens de interesse, facilitando a definição das estratégias de cruzamento e auxiliando na detecção de prováveis marcadores para genes de interesse agrônômico (BARBOSA NETO *et al.*, 1998).

Desde o início de sua utilização, as técnicas de marcadores moleculares passaram por intensa evolução. Os primeiros marcadores moleculares utilizados foram as isoenzimas, as quais têm a vantagem de fornecer respostas rápidas a baixos custos. Porém, são de uso limitado devido ao pequeno número de *loci* detectáveis para a análise, ao baixo polimorfismo detectado, à influência do ambiente na atividade isoenzimática e à dependência do estágio de desenvolvimento da planta (HELENTJARIS, 1992 *apud* TORRES, 1996). Posteriormente, surgiram os marcadores de DNA, capazes de detectar diferenças na informação genética entre indivíduos. Dentre eles destacam-se o RFLP (*Restriction Fragment Length Polymorphism*), RAPD (*Random Amplified Polymorphic DNA*) e AFLP (*Amplified Fragment Length Polymorphism*). As metodologias, vantagens e limitações destas técnicas foram descritas por FERREIRA e GRATTAPAGLIA (1996).

Marcadores baseados em *loci* hipervariáveis de microssatélites (*Simple Sequence Repeats* - SSR) têm ganhado cada vez mais importância. No início dos anos 80, foi comprovado por meio de vários experimentos que uma grande parte do genoma dos eucariotos é rico em regiões repetitivas, que consistem de

pequenas seqüências com 1 a 4 nucleotídeos de comprimento, repetidas em *tandem*. O termo microssatélites ou seqüências simples repetidas, SSR, foi introduzido em 1989 para definir tais *loci*. Estas seqüências representam o maior componente de DNA satélite e da heterocromatina, e têm sido caracterizadas e descritas em várias espécies animais e vegetais. A presença de número variável de nucleotídeos nos microssatélites geram polimorfismo, o qual pode ser detectado a partir da amplificação via técnica PCR (*Polymerase Chain Reaction*), utilizando *primers* específicos, de 20 a 30 pares de bases, complementares a seqüências únicas que flanqueiam essas seqüências simples repetidas. Cada segmento amplificado possui tamanho diferente, podendo ser visualizado, eletroforicamente, em gel de alta resolução (FERREIRA e GRATTAPAGLIA, 1996).

A utilidade dos microssatélites encontra-se em dois fatores principais: 1 - possuem o mais elevado nível de informação de polimorfismo detectado (PIC - *Polymorphism Information Content*), graças à característica multialélica e 2 - são capazes de distinguir indivíduos próximos, que estejam segregando para poucas características, porque são freqüentes e distribuídos ao acaso, permitindo a mais completa cobertura de qualquer genoma eucarioto (POWELL *et al.*, 1996). Por isso, marcadores microssatélites têm sido amplamente utilizados no melhoramento genético de culturas de importância econômica. Eles têm se mostrado eficientes para o estudo de diversidade genética, determinação da relação de parentesco, mapeamento de genomas e identificação de variedades, cultivares e clones de espécies como milho, arroz, sorgo, girassol, trigo e tomate (MOHAN *et al.*, 1997). Marcadores SSR também estão sendo utilizados para identificar genótipos resistentes ao nematóide do cisto, uma das maiores causas de baixa produtividade da cultura de soja (PABHU *et al.*, 1999).

Em soja, vários grupos estão desenvolvendo marcadores e mapas genéticos baseados em SSR por serem a classe de marcadores mais informativa. Isto deve-se não somente à grande expressão e valor comercial desta cultura no

mundo, mas também aos baixos níveis de diversidade genética no DNA encontrados por outros marcadores em germoplasma cultivado, o que tem tornado a construção de mapas genéticos baseados em RFLP ou mesmo RAPD difícil e tediosa (FERREIRA e GRATTAPAGLIA, 1996; MILACH, 1998).

Outro campo potencial do emprego de microssatélites é na caracterização genética de variedades para patenteamento. A Lei de Proteção de Cultivares (*Plant Variety Protection Office* - PVPO) visa assegurar aos melhoristas de plantas a propriedade sobre as cultivares lançadas comercialmente. Desta forma, é fundamental o estabelecimento de critérios objetivos para a identificação de novos genótipos. Por identificar diferenças no DNA, determinando um padrão específico para cada material, técnicas moleculares poderão ser uma eficiente alternativa na comparação de diferentes variedades com reduzida margem de erro (BARBOSA NETO *et al.*, 1998). A determinação deste padrão, chamado DNA *fingerprinting*, foi, originalmente, definido como o método que detecta as diferenças na seqüência característica de nucleotídeos entre os indivíduos por hibridização de sondas específicas multialélicas (JEFFREYS *et al.*, 1985).

1.3 - DNA *fingerprinting*

No início da década de 90, várias técnicas aprimoraram os meios úteis para detectar o DNA *fingerprinting*, sendo a mais importante delas o surgimento da reação em cadeia da polimerase (PCR), que resultou em determinar o padrão de bandas por amplificação de um fragmento específico de DNA (WEISING *et al.*, 1995). Nas análises dos genomas de plantas, DNA *fingerprinting* foi introduzido em 1988, ao demonstrar-se padrões de fragmentos diferentes de DNA entre duas variedades de cevada, *Hordeum vulgare*, por meio da digestão do DNA com *HaeIII* e hibridização com a sonda M13 (RYSKOV *et al.*, 1988).

Em soja, cultura que possui baixos níveis de polimorfismo, SSR são especialmente úteis para análise da base genética. Para caracterizar 58 genótipos de soja cultivada, [*Glycine max* (L.) Merr.], e soja selvagem, [*Glycine soja* (L.) Sieb. e Zucc.], 17 sondas de RFLP foram utilizadas. Apesar do número relativamente alto de sondas, somente 7 genótipos foram identificados. Em 15 de 17 *loci* RFLP, somente dois alelos foram encontrados (KEIM *et al.*, 1989). Adotando a técnica de microssatélites, 96 genótipos de soja foram identificados. Entre 11 e 26 alelos foram encontrados para 7 *primers* utilizados. A informatividade desses marcadores foi, em média, de 0.87 (RONGWEN *et al.*, 1995). Na determinação de um mapa de ligação genética em soja utilizando marcadores RFLP, SSR, isoenzimas e morfológicos de genótipos do cruzamento de *Glycine max* X *Glycine soja*, microssatélites constituem a mais poderosa classe de marcadores da atualidade, por constituírem marcadores multialélicos e co-dominantes, resultando no mais alto polimorfismo (CREGAN *et al.*, 1999).

Diante do exposto, o uso de marcadores moleculares tipo microssatélites em soja é uma ferramenta fundamental para estudos de DNA *fingerprinting*, pois trata-se de um trabalho básico, no qual as informações obtidas são de fundamental importância para o sucesso de um programa de melhoramento.

OBJETIVO

Caracterizar, geneticamente, linhagens de soja, [*Glycine max* (L.) Merr.], por meio de marcadores tipo microssatélites, visando acessar sua variabilidade genética e fornecer subsídios aos programas de melhoramento dessa cultura.

MATERIAL E MÉTODOS

Este experimento foi conduzido no Laboratório de Genética Molecular da Empresa Novartis Seeds, centro de pesquisa em Uberlândia-MG.

3.1 - Material biológico

Dentre os materiais do Banco de Germoplasma da Empresa Novartis Seeds, 16 genótipos de soja [*Glycine max* (L.) Merr.] foram cedidos para este estudo, dos quais três acessos são constituídos de soja selvagem e o restante das linhagens é comercial (Tabela 1).

As sementes de cada genótipo foram semeadas em casa de vegetação. Cerca de cinco dias após a germinação, foram coletadas folhas das plantas para extração do DNA.

TABELA 1: Genótipos utilizados no trabalho e seus respectivos *Pedigree*

Genótipos	<i>Pedigree</i>
LA SUPREMA	FT-Cristalina///Hood//Hill/PI 494454
HARTWIG	Forrest*4/PI 437654
ESSEX	Lee/S 5-7075
PEKING	(Silvestre)
CENTENNIAL	D64-4636/Pickett 71
SPRING	Williams/Essex
HA 200	F81-5344/Santa Rosa
B 412113	Spring/A4715
PI 90763	(Silvestre)
PI 88788	(Silvestre)
BEDFORD	Forrest*2//D68-18/PI 88788
COKER 6738	Braxton/Coker 368
S 51-00	Pio9471/Coker 6973
S 57-11	A5474/Braxton
S 80-J2	A6297/H6686
S 83-30	Braxton/Coker 368

3.2 - Extração de DNA

Para extração de DNA das plantas foi utilizado o protocolo MINIPREP (DOYLE e DOYLE, 1990) com modificações. Primeiramente, o tecido vegetal foi congelado em nitrogênio líquido. Em tubos *ependorf* de 1.5 mL com bolinhas de vidro, as folhas foram maceradas com a utilização de um *grinder*, que permitiu o rompimento mecânico das paredes e membranas celulares. O pó obtido foi ressuspendido em 400 μ L de tampão de extração (1.4 M NaCl, 100 mM Tris HCl pH 8.0, 20 mM E.D.T.A., 1% CTAB, 0.1% β -Mercaptoetanol e 1% Polivinilpirrolidone - PVP 40) com pH entre 8.0 e 9.0, favorável à atuação de enzimas. As amostras foram, então, incubadas a 65°C por um período de 40 min,

com agitação intermitente a cada 10 min, permitindo homogeneização da suspensão.

Após incubação, os tubos foram resfriados à temperatura ambiente durante 10 min. Em seguida, adicionou-se 400 μ L de solvente orgânico clorofórmio: álcool isoamílico (24:1), sendo o produto centrifugado a 14.000 rotações por minuto (rpm) por 15 min. Nesta etapa, proteínas e fragmentos celulares foram separados do DNA e o sobrenadante foi transferido para outro tubo. Esta etapa foi repetida para se apurar a qualidade do DNA. A precipitação foi feita com 400 μ L de isopropanol previamente resfriado. Novamente, os tubos foram centrifugados nas mesmas condições descritas acima. O sobrenadante foi descartado e o *pellet* lavado por duas vezes com etanol 70% e uma com etanol 95%, utilizando o mesmo procedimento de centrifugação. O *pellet* foi mantido em temperatura ambiente para secar e depois ressuscitado em 100 μ L de T.E. (10 mM Tris HCl pH 7,5 e 1 mM E.D.T.A.), contendo RNase na concentração final de 40 μ g/mL e incubado a 37°C por 30 minutos. Em seguida, foi adicionado 10 μ L de NaCl 5M juntamente com 65 μ L de isopropanol gelado para precipitar, novamente, o DNA, que foi ressuscitado em 100 μ L de T.E., após sucessivas lavagens em etanol.

3.3 - Quantificação e qualificação do DNA

Após a extração, a qualidade do DNA foi analisada eletroforéticamente em gel de agarose comum 1% com uma voltagem constante de 80 V durante 60 min. As amostras de DNA foram, então, quantificadas em espectrofotômetro, para leitura de absorbância a 260 nm. Calculou-se a concentração de DNA a partir da fórmula:

$$[\text{DNA}] \text{ ng}/\mu\text{L} = \text{OD}_{260} \times 50 \times \text{Fator de Diluição},$$

na qual: **OD260** é o valor da leitura da absorvância em onda 260;

50 é o fator de conversão para DNA dupla-fita;

Fator de diluição é a razão volume total (198 μL de água + 2 μL de DNA) / volume da amostra de DNA (2 μL).

Após a quantificação, as amostras de DNA foram diluídas para uma concentração final de 15 ng/ μL , que constituiu a coleção de soluções de trabalho para a PCR (Reação em Cadeia da Polimerase).

3.4 - Amplificação do DNA

As reações para amplificação foram feitas para um volume final de 15 μL e submetidas a PCR em termocicladores PTC 200, por 1 ciclo de 2 min a 95°C, 40 ciclos de 15 seg a 94°C, 0 seg a 69°C e 45 seg a 47°C e 1 ciclo de 2 min a 72°C. Na tabela 2, estão apresentados os reagentes que constituíram tais reações.

TABELA 2: Constituintes das reações de amplificação

SOLUÇÃO ESTOQUE	VOLUME DE CADA REAÇÃO	CONCENTRAÇÃO FINAL
H ₂ O estéril	7,28 μL	–
Tampão da PCR 10X (20 mM MgCl ₂)	1,5 μL	1X
dNTPs (5 mM)	0,75 μL	250 μM
Primers RF (25 μM)	0,3 μL	0,50 μM
Taq DNA polimerase (5 U/ μL)	0,17 μL	0,05 U/ μL
DNA genômico (15ng/ μL)	5,0 μL	75 ng
	Total 15,0 μL	

Foram utilizados 30 pares de *primers* de, em média, 20 pares de bases, distribuídos nos grupos de ligação do genoma de [*Glycine max* (L.) Merr.], cujas seqüências foram obtidas em banco de dados de domínio público (<http://129.186.26.94/default.html>) (Tabela 3).

TABELA 3: *Primers* utilizados para amplificar o DNA dos 16 genótipos e suas respectivas seqüências de nucleotídeos

<i>Primer</i>	Grupo de Ligação	Seqüência	
		<i>Forward</i> (5' → 3')	<i>Reverse</i> (3' → 5')
SATT - 142	H	GGACAACAACAGCGTTTTTAC	TTTGCCACAAAGTTAATTAATGTC
SATT - 147	D1a + Q	CCATCCCTTCCTCCAAATAGAT	CTTCCACACCCTAGTTTAGTGACAA
SATT - 156	L	CGCACCCCTCATCCTATGTA	CCAACATAATCCCAGGGACTTACTT
SATT - 168	B2	CGCTTGCCCAAAAATTAATAGTA	CCATTCTCCAACCTCAATCTTATAT
SATT - 184	D1a + Q	GCGCTATGTAGATTATCCAAATTACGC	GCCACTTACTGTTACTCAT
SATT - 185	E	GCGCATATGAATAGGTAAGTTGCACTAA	GCGTTTTCTACAATAATATTTTCAT
SATT - 190	C1	GGGAGTGTGAACCTTACATTGTCT	GGGCCTTGAATTTTGTGCTAT
SATT - 231	E	GCGTGTGCAAAATGTTTCATCATCT	GGCAGGAATCAACATCAAACTTC
SATT - 242	K	GCGTTGATCAGGTCGATTTTTATTTGT	GCGAGTGCCAACTAACTACTTTTATGA
SATT - 260	K	GCGCCAAATGTATACTTTAAATTCTT	GCGGGTTTAGCTAAAATAGTTCGTGC
SATT - 276	A1	GCGGACGGTAAGGACTATTTATGATA	GCGTCAGATGAAAAAAAATAAGATAC
SATT - 277	C2	GGTGGTGGCGGGTACTATTACT	CCACGCTTCAGTTGATTCTTACA
SATT - 281	C2	AAGCTCCACATGCAGTTCAAAAC	TGCATGGCAGCAGAAAGAAGTA
SATT - 285	J	GCGACATATTGCATTA AAAACATACTT	GCGGACTAATTCTATTTTACACCAACAAC
SATT - 311	D2	GGGGGAACCACAAAAATCTTAATC	GTTGAAGCTCAGGCTGTGATGAAT
SATT - 339	N	TAATATGCTTTAAGTGGTGTGGTTATG	GTTAAGCAGTTCCTCTCATCACG
SATT - 353	H	CATACACGCATTGCCTTTCTGAA	GCGAATGGGAATGCCTTCTTATTCTA
SATT - 371	C2	TGCAAACAACTGGATTCACCTCA	GAGATCCCGAAATTTTAGTGTAACA
SATT - 373	L	TCCGCGAGATAAATTCGTA AAAAT	GGCCAGATACCCAAGTTGTACTTGT
SATT - 385	A1	AATCGAGGATTCACCTTGAT	CATTGGGCCACACAACAAC
SATT - 417	K	TCTTGCTAATTGCTTCATTTTCAT	AATTGCTTGGGATTTTCATTT
SATT - 436	D1a + Q	GCGTATAAAGAAAAACGAGCATATCAT	GCGCTTATAAAGGCTTGTGAAAGACACT
SATT - 440	I	TGAGAACGTTTGAAAAGAGAT	GAAGAGATTAAGCATAAAGAATACTT
SATT - 458	D2	TTGGGTTGACCGTGAGAGGGGAGAA	GCGAACCACAAACAACAATCTTCA
SATT - 463	M	TTGGATCTCATATTCAAAC TTTCAAG	CTGCAAATTTGATGCACATGTGTCTA
SATT - 472	G	GCGAATACATAAAACTCAAATTCAAATCATA	GCGTTCTATAAATTTCAATTCATAGTTTCAAT
SATT - 487	O	ATCACGGACAGTTCATTTGTA	TGAACCGCGTATTCTTTTAATCT
SATT - 509	B1	GCGCTACCGTGTGGTGGTGTGCTACCT	GCGCAAGTGGCCAGCTCATCTATT
SATT - 534	B2	CTCCTCCTGCGCAACAACAATA	GGGGGATCTAGGCCATGAC
SATT - 537	D1b + W	GCGTGTACATCAAATGGTGTATTC	GCGGGAGGAACTTTGTCTCAGTAAT

3.5 – Separação de fragmentos e visualização dos produtos amplificados

Os produtos amplificados foram separados por eletroforese horizontal em gel de agarose *metaphor*® 3%, usando tampão Tris-Borato EDTA (TBE) 0,5X. Para a visualização das bandas de DNA, foi utilizado brometo de etídio em uma concentração de 0,5 µg/mL. Após 120 min de corrida nas cubas de eletroforese a 125 V, os padrões de bandas foram visualizados em transiluminador UV e as imagens digitalizadas por meio do *software Multi-Analyst*®.

3.6 - Análise de dados

Os alelos obtidos de cada *primer* foram registrados como presença (1) ou ausência (0) para cada genótipo avaliado, gerando uma matriz de valores binários, que foi analisada pelo programa STATISTICA 4.5 A (1993), para o cálculo de distância genética e análise de *cluster*.

As distâncias genéticas foram calculadas pelo método de Porcentagem de Desacordo (PD), que é dado pela fórmula N'_{AB} / N_T , na qual N'_{AB} é o número total de bandas polimórficas entre os genótipos comparados e N_T é o número total de bandas (PUTERKA et al, 1993 *apud* CABRAL, 1997).

A análise de *cluster* foi feita pelo Método Não-ponderado de Agrupamento aos Pares utilizando médias aritméticas (UPGMA), o qual agrupa, inicialmente, indivíduos mais similares e assim, sucessivamente, até indivíduos ou grupos mais distantes.

Para cada *primer* utilizado no estudo foi calculado o valor de PIC (*Polymorphism Information Content*) usando a fórmula:

$$\text{PIC} = 1 - \sum^n f_i^2 \quad i = 1,$$

na qual: f_i^2 é a frequência do i^{th} alelo.

O valor de PIC proporciona uma estimativa do poder discriminatório de um *locus*, não somente pelo número de alelos que expressa, mas também pela frequência desses alelos entre os genótipos analisados. Os valores de PIC oscilam entre 0 (monomórfico) e 1 (altamente discriminatório, com muitos alelos em frequências iguais) (SMITH *et al.*, 1997).

RESULTADOS E DISCUSSÃO

No banco de dados disponível sobre [*Glycine max* (L.) Merr.], de domínio público (<http://129.186.26.94/default.html>), são listados 500 pares de *primers* SSR distribuídos ao longo do genoma da espécie. Dentre estes, os *primers* utilizados neste estudo foram escolhidos de forma que garantissem uma maior cobertura dos grupos de ligação do genoma de soja, descritos por CREGAN *et al.* (1999), e considerando sua disponibilidade no acervo do laboratório. Isso foi levado em consideração para que a base genética dos genótipos analisados fosse amostrada de forma abrangente.

Foram obtidos produtos amplificados em todas as reações envolvendo os 30 pares de *primers* utilizados neste estudo. Da análise de microssatélites resultaram 109 alelos, com tamanho variando entre 20 e 500 pares de bases. O número de alelos detectados variaram de 2 a 5, dependendo do *primer* utilizado. Os valores de PIC variaram de 0.12 a 0.73 de acordo com o marcador (Tabela 4). Microssatélites que apresentaram o maior número de alelos expressados, possuem o maior valor de PIC, como pode ser observado com os microssatélites SATT 260 e SATT 353, que apresentaram 5 alelos e o maior valor de PIC, 0.73. Os microssatélites SATT 142 e SATT 417 apresentaram 2 alelos e nível de informatividade baixo, 0.12 e 0.28, respectivamente. Em contrapartida, a Figura 1, apresenta o padrão de amplificação de fragmentos de DNA do microssatélite SATT 277, que, apesar de expressar 5 alelos, possui PIC 0.50, ou seja, o nível

discriminatório desse marcador é relativamente baixo, o que é explicado pela frequência desigual dos 5 alelos entre os genótipos avaliados. Já o microsatélite SATT 231 apresenta o valor de PIC relativamente alto, 0.66, apesar de expressar somente 3 alelos. Tais dados ressaltam a importância de se avaliar previamente marcadores a serem aplicados nos programas de melhoramento, para que os mais informativos sejam utilizados. A análise baseada somente no número de alelos expressos não é suficiente, visto que o poder discriminatório é determinado pela frequência desses alelos entre os genótipos avaliados. O alto nível de polimorfismo observado nos locos microsatélites comprovou a eficiência da técnica para estudos de DNA *fingerprinting* em espécies vegetais, o que havia sido observado anteriormente (RONGWEN *et al.*, 1995). Em soja, cultura de importância econômica, marcadores microsatélites tornam-se ferramenta ideal para estudos de DNA *fingerprinting*, já que possui baixa variabilidade genética.

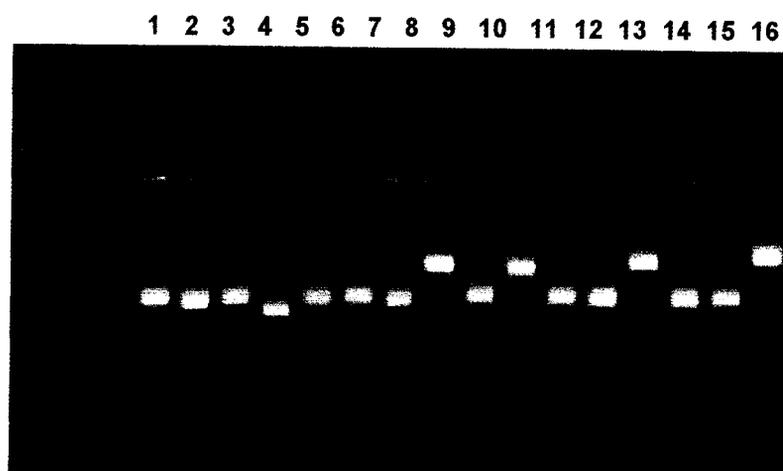


Figura 1: Padrão de amplificação de fragmentos de DNA obtidos do microsatélite SATT 277 para os 16 genótipos analisados em gel de agarose *metaphor*® 3%. Seqüência dos genótipos: 1 - LA SUPREMA, 2 - HARTWIG, 3 - ESSEX, 4 - PEKING, 5 - CENTENNIAL, 6 - SPRING, 7 - HA 200, 8 - B 412113, 9 - PI 90763, 10 - PI 88788, 11 - BEDFORD, 12 - COKER 6738, 13 - S 51-00, 14 - S 57-11, 15 - S 80-J2, 16 - S 83-30

Tabela 4: Relação do número de alelos e PIC para cada microsatélite usado

Microsatélite	Número de Alelos	PIC
SATT - 142	2	0.12
SATT - 147	4	0.66
SATT - 156	2	0.49
SATT - 168	3	0.59
SATT - 184	3	0.45
SATT - 185	3	0.29
SATT - 190	3	0.59
SATT - 231	3	0.66
SATT - 242	4	0.68
SATT - 260	5	0.73
SATT - 276	4	0.65
SATT - 277	5	0.50
SATT - 281	4	0.69
SATT - 285	5	0.72
SATT - 311	5	0.71
SATT - 339	3	0.63
SATT - 353	5	0.73
SATT - 371	4	0.68
SATT - 373	4	0.67
SATT - 385	4	0.62
SATT - 417	2	0.28
SATT - 436	2	0.38
SATT - 440	4	0.71
SATT - 458	4	0.49
SATT - 463	3	0.53
SATT - 472	4	0.49
SATT - 487	3	0.63
SATT - 509	4	0.68
SATT - 534	4	0.72
SATT - 537	4	0.65

Para a detecção das variantes alélicas, a técnica eletroforética foi utilizada. Como o polimorfismo detectado por marcadores microssatélites difere em poucos pares de bases de um alelo para outro, agarose *metaphor*® foi empregada. Por possuir uma 'malha' mais intensa, permite a separação de fragmentos em até 10 pares de bases, além de ser mais fácil de ser preparada numa rotina de laboratório que busca volume e rapidez nos resultados. Os géis podem ser reutilizados, compensando o alto custo dessa agarose especial. Entretanto, alguns marcadores SSR apresentaram 2 alelos, como o SATT 417 (Figura 2). Para uma melhor visualização da separação desses fragmentos seria melhor a aplicação das reações em géis de poliacrilamida. Visto que tais géis não são práticos na obtenção rápida de resultados, a não ser em sistemas automáticos, *primers* que apresentam essa particularidade são substituídos por outros que se localizam no mesmo grupo de ligação, garantindo completa cobertura do genoma, e que possuem alto valor de PIC. Acessar a base genética dos materiais utilizados nos programas de melhoramento é de fundamental importância, pois permite selecionar os melhores marcadores na avaliação dos genótipos, sem comprometer tempo com a obtenção de resultados.

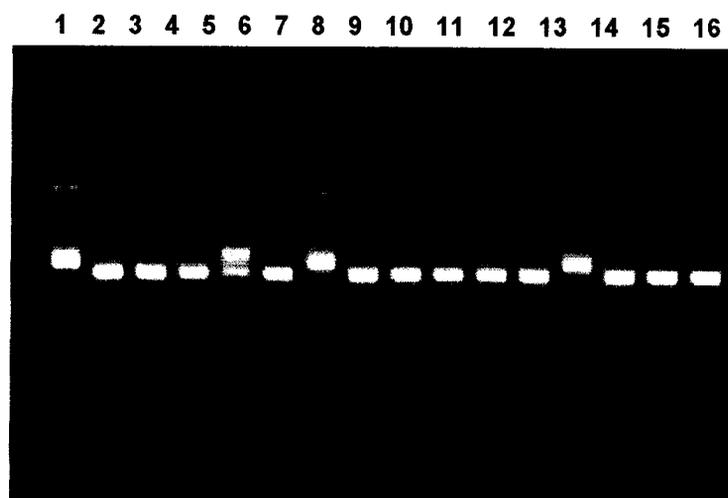


Figura 2: Padrão de amplificação de fragmentos de DNA obtidos do microsatélite SATT 417 para os 16 genótipos analisados em gel de agarose *metaphor*® 3%. Seqüência dos genótipos: 1 - LA SUPREMA, 2 - HARTWIG, 3 - ESSEX, 4 - PEKING, 5 - CENTENNIAL, 6 - SPRING, 7 - HA 200, 8 - B 412113, 9 - PI 90763, 10 - PI 88788, 11 - BEDFORD, 12 - COKER 6738, 13 - S 51-00, 14 - S 57-11, 15 - S 80-J2, 16 - S 83-30

A análise da variação genética baseada em marcadores microssatélites permitiu a determinação das distâncias genéticas entre as cultivares, que foram utilizadas para a construção de um dendrograma (Figura 3).

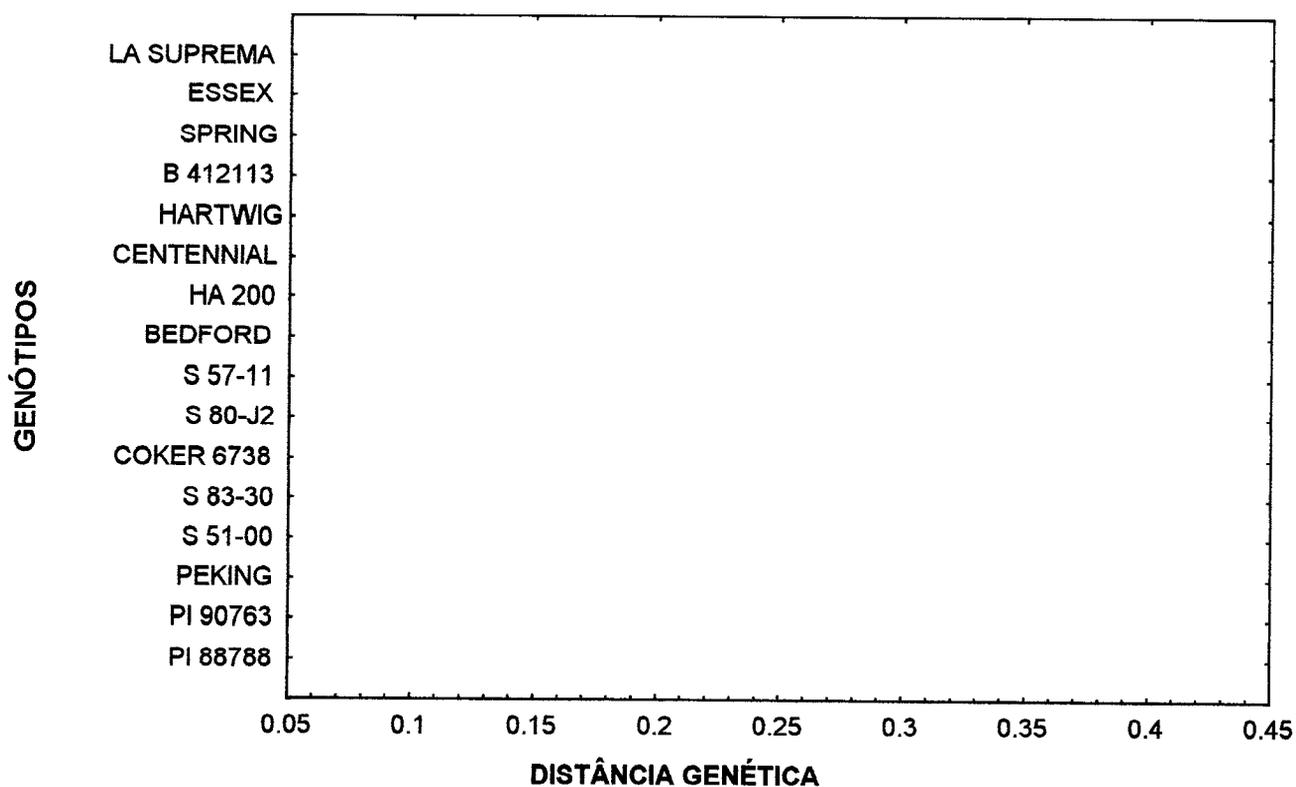


Figura 3: Dendrograma representativo da Distância Genética por Porcentagem de Desacordo e Agrupamento pelo Método de UPGMA entre os 16 genótipos de soja por meio de 109 alelos obtidos de 30 microssatélites

A distância genética limite calculada entre os genótipos analisados foi de 42%, como pode ser observado na análise de *cluster*. É interessante ressaltar que os genótipos que formaram grupos próximos no dendrograma possuem linhagens em comum na sua genealogia. Os genótipos *Essex* e *Spring* formaram o grupo mais similar dentro da análise, apresentando somente 13% de distância genética. Este resultado pode ser explicado pelo grau de parentesco entre as duas linhagens, já que *Essex* foi uma das linhagens utilizadas para desenvolver a variedade *Spring* (Tabela 1). O segundo *cluster* mais similar é constituído pelo grupo *Coker 6738* e *S 8330*, com cerca de 15% de distância genética. São materiais que apresentam a mesma genealogia (*Braxton/Coker 368*).

Num outro extremo, os genótipos silvestres *PI 90763* e *PI 88788* formaram o *cluster* mais distinto geneticamente entre os obtidos, com cerca de 25% de distância genética. Juntamente com *Peking*, a distância genética passa a ser de 33%. Sabendo que se tratam de linhagens selvagens é previsível tais resultados entre elas e entre os demais genótipos comerciais. A linhagem *La Suprema*, proveniente da Colômbia, também formou um *cluster* separado dos demais, assim como as silvestres. Tal fato confirma que indivíduos selvagens ou distantes geneticamente são fontes de variabilidade num programa de melhoramento. A pouca similaridade genética em relação às variedades comerciais são úteis para mapeamento, pois em cruzamentos com linhagens comerciais podem gerar muito polimorfismo. A realização de cruzamentos entre tipos geneticamente distintos podem contribuir para a ampliação da variabilidade genética em populações segregantes (MESSMER *et al.* 1993).

Em contrapartida, os genótipos *Bedford* e *S 57-11*, assim como *Hartwig* e *Centennial* formaram *cluster* similares geneticamente, embora na genealogia apresentada não apresentam parentes próximos entre si. Provavelmente, em cruzamentos anteriores possuíram linhagens em comum. Num banco de germoplasma nem sempre a genealogia das linhagens é conhecida, por isso as

informações obtidas neste estudo por meio de marcadores são importantes. Isso pode reduzir tempo e custo num programa de melhoramento, pois permitem aos melhoristas guiar os cruzamentos entre esses genótipos, evitando endogamia, no qual, provavelmente, não vão ser obtidas variedades com características desejáveis.

O conhecimento do germoplasma também é importante para caracterizar geneticamente variedades que são lançadas no mercado. No caso de soja, e outras espécies, a caracterização de linhagens baseava-se em atributos morfológicas para distinguir uma nova cultivar. Como as novas variedades são resultados de cruzamentos entre genótipos-elite, cada vez mais torna-se difícil sua distinção, sendo necessário a caracterização por meio de *DNA fingerprinting*.

Os padrões de marcadores microssatélites obtidos para os genótipos em estudo servirão para desenvolver marcadores que confirmam cruzamentos. Isso é importante para confirmar pureza genética do material que está trabalhando. Dentre os 30 *primers* utilizados, cerca de 18 (60%) apresentam valor de PIC acima de 0.60, o que comprova o poder discriminatório dos marcadores SSR entre genótipos de soja.

Na seleção assistida por marcadores - *Marker Assisted Selection (MAS)*, *DNA fingerprinting* é uma técnica importante na determinação de marcadores ligados a características não detectáveis fenotipicamente. A variabilidade obtida a nível molecular dos genótipos analisados neste estudo fornece base para estudos de mapeamento, visando associar marcadores moleculares a características de importância agrônômica.

DNA fingerprinting também é importante na conversão de linhagens. Com a mais nova ferramenta do melhoramento, as plantas transgênicas, a técnica é utilizada na escolha das linhagens mais próximas geneticamente com o material transgênico. Isso permite reduzir o número de cruzamentos executados na conversão.

CONCLUSÕES

Com base nos resultados apresentados neste estudo, pode-se concluir que:

- Os 30 pares de *primers* SSR utilizados foram suficientes para amostrar de forma abrangente a base genética dos genótipos de soja avaliados.
- O nível de informatividade (PIC) de cada *primer* é uma medida importante na escolha de marcadores para selecionar genótipos num programa de melhoramento de plantas, pois demonstram, além do número de alelos expressos, a capacidade discriminatória dos mesmos entre as linhagens de interesse.
- Marcadores microssatélites apresentaram-se eficientes para estudos de DNA *fingerprinting* em soja, pois são co-dominantes, multialélicos e distribuídos por todo o genoma. Tais características tornam-se, particularmente, importantes devido a baixa variabilidade genética dessa cultura.

- O polimorfismo de *loci* SSR detectados por meio de agarose *metaphor*® é eficiente, pois separa fragmentos distintos em até 10 pares de bases, além de ser uma técnica eficiente para laboratórios que buscam resultados em pouco tempo.
- A distância genética limite calculada entre os 16 genótipos de [*Glycine max* (L.) Merr.] analisados foi de 42%.
- Genótipos selvagens formaram um grupo separado com relação às linhagens comerciais no dendrograma.
- O estudo em genótipos de soja por meio de marcadores microsatélites, aqui apresentado, será importante num programa de melhoramento, pois as informações geradas permitirão aos melhoristas guiar os cruzamentos, de forma a evitar aqueles endogâmicos, desenvolver marcadores de cruzamento, caracterizar novas cultivares comerciais e na conversão de linhagens.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARBOSA NETO, J. F. e BERED, F. 1998. *Marcadores moleculares e diversidade genética no melhoramento de plantas*. In: MILACH, S. C. K. (editora). *Marcadores Moleculares em Planta*. Porto Alegre-RS, Universidade Federal do Rio Grande do Sul, 29-40.
- BELELE, C. L. 1997. Análise bioquímica e molecular da diversidade genética entre cultivares mesoamericanos e andinos de feijoeiro comum (*Phaseolus vulgaris* L.) *Dissertação de Mestrado*. Universidade Federal de Uberlândia, 12-19.
- CABRAL, T. A. T. 1997. Divergência genética entre genótipos de algodoeiro (*Gossypium hirsutum* L.) estimada por análise de marcadores RAPD. *Dissertação de Mestrado*. Universidade Federal de Uberlândia, p20.
- CONAB. Companhia Nacional de Abastecimento. 1998. *Indicadores da agropecuária*. Brasília, v. 2.

- CREGAN, P. B.; JARVIK, T.; BUSH, A. L.; SHOEMAKER, R. C.; LARK, K. G.; KAHLER, A. L.; KAYA, N.; VAN TOAI, T. T.; LOHNES, D. G.; CHUNG, J. e SPECHT, J. E. 1999. An integrated genetic linkage map of the soybean genome. *Crop Science* 39 (5): 1464-1490.
- DOYLE, J. J. e DOYLE, J. L. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12:13-15.
- DUDLEY, J. M. 1993. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. *Crop Science* 33 (4): 660-668.
- EMBRAPA. Centro Nacional de Pesquisa de Soja (Londrina, PR). 1996. *Banco ativo de germoplasma de soja*. Londrina-PR, EMBRAPA, Documentos, 104, p 169.
- EMBRAPA. Centro Nacional de Pesquisa de Soja (Londrina, PR). 1998. *Recomendações técnicas para a cultura da soja na região Central do Brasil*. Londrina-PR, EMBRAPA, Documentos, 98, 157-160.
- FERREIRA, M. E. e GRATTAPAGLIA, D. 1996. *Introdução ao uso de marcadores moleculares em análise genética*. 2ª ed. Brasília-DF, EMBRAPA-CENARGEN, 220p.
- HADLEY, H. H. e HYMOWITZ, T. 1973. *Speciation and cytogenetics*. In: CALDWELL, B. E. (editor). *Soybeans: improvement, production and uses*. Wisconsin-USA, Publisher, 97-116.
- JEFFREYS, A. J.; WILSON, V. e THEIN, S. L. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature (London)* 314: 67-73.

- KEIM, P.; SHOEMAKER, R. C. e PALMER, R. G. 1989. Restriction fragment length polymorphism diversity in soybean. *Theor Appl Genet* 77: 786-792.
- MESSMER, M. M.; MELCHINGER, A. E.; HERMANN, R. G. e BOPPENMAIER, J. 1993. Relationships among early european maize inbreds: I. Comparision of pedigree and RFLP data. *Crop Science* 33: 944-950.
- MILACH, S. C. K. 1998. *Principais tipos de marcadores moleculares e suas características*. In: MILACH, S. C. K. (editora). *Marcadores Moleculares em Planta*. Porto Alegre-RS, Universidade Federal do Rio Grande do Sul, 17-28.
- MOHAN, M.; NAIR, S.; BHAGWAT, A.; KRISHNA, T. G.; YANO, M.; BHATIA, C. R. e SASAKI, T. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding* 3: 87-103.
- POWELL, W.; MACHRAY, G. C. e PROVAN, J. 1996. Polymorphism revealed by simple sequence repeats. *Trends in plant science* 1 (7): 215-222.
- PRABHU, R. R.; NJITI, V. N.; BELL-JOHNSON, B.; JOHNSON, J. E.; SCHMIDT, M. E.; KLEIN, J. H. e LIGHTFOOT, D. A. 1999. Selecting soybean cultivars for dual resistance to soybean cyst nematode and sudden death syndrome using two DNA markers. *Crop Science* 39: 982-987.
- ROESSING, A. C. e GUEDES, L. C. A. 1993. *Aspectos econômicos do complexo soja: sua participação na economia brasileira e evolução na região do Brasil Central*. In: ARANTES, N. E. e SOUZA, P. I. M. (editores). *Simpósio sobre cultura da soja nos cerrados*. Piracicaba-SP, POTAFOS, 1-69.

- RONGWEN, J.; AKKAYA, M. S.; BHAGWAT, A. A.; LAVI, U. e CREGAN, P. B. 1995. The use of microsatellite DNA markers for soybean genotype identification. *Theor Appl Genet* 90: 43-48.
- RYSKOV, A. P.; JINCHARADZE, A. G.; PROSNYAK, M.I.; IVANOV, P. L. e LIMBORSKA, S. A. 1988. M13 phage DNA as a universal marker for DNA fingerprinting of animals, plants and microorganisms. *FEBS Lett* 233: 388-392.
- SEDIYAMA, T.; ALMEIDA, L. A.; MIYASAKA, S. e KIIHL, R. A. S. 1986. *Genética e melhoramento. In: A soja no Brasil Central*. 3ª ed. vol 1. Campinas-SP, Fundação Cargill, 21-74.
- SMITH, J. S. C.; CHIN, E. C. L.; SHU, H.; SMITH, O. S.; WALL, S. J.; SENIOR, M. L.; MITCHELL, S. E.; KRESOVICH, S. e ZIEGLE, J. 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPS and pedigree. *Theor Appl Genet* 95: 163-173.
- SOUZA, P. I. M. 1999. Diretrizes para o melhoramento de soja no Brasil. Anuário ABRASEM 99. Brasília-DF, 55-59.
- TORRES, G. A. 1996. Identificação de regiões cromossômicas do milho associadas com tolerância à toxidez de alumínio por meio de RFLP. *Dissertação de Mestrado*. Universidade Federal de Lavras, 1-16.
- WEISING, K.; NYBOM, H.; WOLFF, K e MEYER, W. 1995. *DNA fingerprinting in plants and fungi*. USA, CRC Press, 1-200.