

FRANCISCO DE ALCÂNTARA NETO

MARCADORES MICROSSATÉLITES NA IDENTIFICAÇÃO DE CULTIVARES
DE SOJA

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitotecnia, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2001

Aos meus pais Newton e Edinez.

À minha esposa Ana Patricia.

À minha filha Gabriela.

Aos meus irmãos, Mônica, Newton José, Tatiana e Danielle.

Aos sobrinhos Mylena e Mateus.

A todos os meus familiares.

Aos meus amigos.

AGRADECIMENTO

À Deus, que me concedeu vitória nas lutas enfrentadas e possibilitou-me alcançar mais este importante objetivo.

À Universidade Federal de Viçosa, especialmente ao Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO) e ao Departamento de Fitotecnia, pela oportunidade de realização deste trabalho.

Ao Conselho Nacional de Pesquisa e Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de mestrado.

Ao povo brasileiro, que contribui pela manutenção da Universidade pública e gratuita.

Aos meus queridos pais Newton e Edinez, exemplo de amor e dedicação aos filhos.

Ao professor Maurilio Alves Moreira, pela competente orientação, pela confiança e amizade, pelo apoio e incentivo desde o período de minha Iniciação Científica, por sua colaboração em todas as etapas dessa minha formação e dos meus trabalhos.

Aos professores Everaldo Gonçalves de Barros e Aluizio Borém, pelas sugestões, pelo zelo incansável com as condições de pesquisa, pelo exemplo

de dedicação à UFV e pela eficiência no planejamento e na sugestão da pesquisa científica.

Ao Dr. Ivan Schuster, pela amizade, pelos ensinamentos, pelos conselhos, pela atenção dispensada e pelas valiosas sugestões.

A todos os professores que ministraram as disciplinas do curso, pelos ensinamentos.

Aos colegas de trabalho: Rita, Dario, Newton, Geraldo, Marcio Antônio, Valéria Guimarães, Valéria Delaretti, Jay, Cynthia, Márcia, Gerardo, Wagner, Cinthya Rocha, Abelmon, Pedro Ivo, Maria Luiza, João Paulo, Thiago e funcionários do BIOAGRO/UFV pela ajuda e agradável convivência.

A todos os amigos e colegas do curso que certamente contribuíram, direta ou indiretamente, para a conclusão deste trabalho.

BIOGRAFIA

FRANCISCO DE ALCÂNTARA NETO, filho de Newton Alves de Alcântara e Edinez Lins de Alcântara, nasceu em Maceió - AL, em 08 de abril de 1972.

Cursou o primário de primeira a quarta série no Colégio Monsenhor Luís Barbosa, em Maceió - AL, e quinta a sétima série no Colégio São Francisco de Assis, em Arapiraca - AL.

Cursou a oitava série do primeiro grau ao terceiro ano científico no Colégio Marista de Maceió, em Maceió - AL.

Em 1994, iniciou o curso superior na Universidade Federal de Viçosa, colando grau em março de 1999 como Engenheiro Agrônomo. Na UFV desenvolveu atividades de Iniciação Científica na área de Genética Molecular Aplicada ao Melhoramento Genético Vegetal (1995/96 e 1997/98).

Em abril de 1999, iniciou o curso de Mestrado em Fitotecnia, na área de Melhoramento de Plantas e Biotecnologia, na UFV, defendendo tese em 26 de março de 2001.

ÍNDICE

	Página
RESUMO	viii
ABSTRACT	x
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	5
2.1. A base genética da soja	5
2.2. Propriedade intelectual dos obtentores vegetais	7
2.3. Descrição de variedades baseada em dados morfológicos.....	8
2.4. Descrição de variedades com base no DNA	9
2.5. Uso de descritores moleculares: visão da UPOV	9
2.6. Pureza genética.....	10
2.7. Marcadores moleculares	11
2.7.1. Marcadores RFLP.....	11
2.7.2. A técnica de PCR	12
2.8. Marcadores moleculares baseados em PCR.....	13
2.8.1. Marcadores RAPD.....	13
2.8.2. Marcadores AFLP.....	14
2.8.3. Marcadores microssatélites (SSRs).....	15

3. MATERIAL E MÉTODOS	18
3.1. Genótipos de soja.....	18
3.2. Extração de DNA	19
3.3. Condições de amplificação e separação dos fragmentos.....	20
3.4. Seqüências dos <i>primers</i> de microssatélites.....	21
3.5. Cálculo da diversidade genética do loco	22
3.6. Cálculo do coeficiente de coincidência simples	23
3.7. Análise de agrupamento.....	23
4. RESULTADOS E DISCUSSÃO	25
4.1. Separação dos fragmentos.....	25
4.2. Determinação do número de alelos	27
4.3. Determinação do tamanho dos fragmentos	29
4.4. Informatividade dos <i>primers</i>	31
4.5. Caracterização das variedades	32
4.6. Divergência genética entre as variedades avaliadas.....	34
4.7. Análise de agrupamento.....	34
5. RESUMO E CONCLUSÕES	39
6. REFERÊNCIAS BIBLIOGRÁFICAS	42

RESUMO

Alcântara, Francisco Neto, M.S., Universidade Federal de Viçosa, março de 2001. **Marcadores microssatélites na identificação de cultivares de soja.** Orientador: Maurilio Alves Moreira. Conselheiros: Aluizio Borém de Oliveira e Everaldo Gonçalves de Barros.

A soja cultivada apresenta base genética bastante restrita, com baixo nível de polimorfismo fenotípico, o que dificulta sobremaneira a identificação inequívoca de cultivares por meio de descritores morfológicos. Entre os marcadores moleculares, os mais indicados para a caracterização genética da soja são os microssatélites, por sua natureza co-dominante e multi-alélica, além de serem extremamente conservados dentro da espécie. Este trabalho teve como objetivo estabelecer um protocolo, de fácil aplicação em análises de rotina, para a identificação molecular de genótipos de soja. Foram utilizados 32 genótipos de soja, sendo 30 cultivares elites e 2 acessos norte-americanos. Foram utilizados 22 pares de *primers* de microssatélites de soja, dos quais 19 evidenciaram polimorfismos. O uso desses *primers* permitiu a detecção de 61 alelos em diferentes locos nos 32 genótipos estudados. A diversidade genética nos 22 locos estudados variou de 0 a 0,73, com média de 0,41. O número de

alelos por loco variou de 1 a 5, com média de 2,77. Com a combinação dos pares de *primers* SATT186, SATT094, SATT070 e SATT197 foi possível identificar 28 dos 32 genótipos. Mantendo-se os 4 *primers* acima, quatro diferentes combinações de 5 pares de *primers* identificaram todos os genótipos, bastando incluir um dos seguintes pares de *primers*: SATT115, SATT184, SATT309 ou SATT536. Com base nos dados relativos à frequência de alelos comuns nos 32 genótipos estudados, foram determinadas as dissimilaridades entre eles. As medidas de dissimilaridade variaram de 9 a 67%, com média de 43%. Apenas dois pares apresentaram dissimilaridade inferior a 10%. A maioria apresentou índices de dissimilaridade superiores a 20%. Visando classificar os 32 cultivares de soja em grupos homogêneos, a matriz de dissimilaridade foi utilizada para a análise de agrupamento dos indivíduos pelo método aglomerativo. Foram utilizados 3 métodos: UPGMA, vizinho mais próximo e vizinho mais distante. O método UPGMA (“unweighted pair-group method using an arithmetic average”) foi o que melhor representou o inter-relacionamento entre os genótipos, com correlação cofenética de 70,5%. O método do vizinho mais próximo e vizinho mais distante tiveram correlação cofenética com os dados de dissimilaridade de 58,6% e 56,9%, respectivamente. Para obter grupos mutuamente exclusivos, foi utilizado o método de otimização de Tocher, que produziu 6 grupos, sendo o grupo 1 formado por 25 indivíduos (divididos em 8 subgrupos), os grupos 2 e 3 formados por 2 indivíduos cada um, e os outros três grupos, formados por um único indivíduo cada um.

ABSTRACT

Alcântara, Francisco Neto, M.S., Universidade Federal de Viçosa, March 2001.

Microsatellite markers in identification of soybean cultivars. Adviser: Maurilio Alves Moreira. Committee Members: Aluizio Borém de Oliveira and Everaldo Gonçalves de Barros.

The cropped soybean presents a genetic base quite restricted with low level of phenotypic polymorphism, what excessively hinders the unequivocal identification of cultivars by morphologic describers. Among the molecular markers, the most indicated for soybean genetic characterization are the microsatellites due to their co-dominant and multi-allelic nature, besides being extremely preserved in the species. The objective of this study was to establish an protocol that could be easily applied on routine analyses for molecular identification of soybean genotypes. Thirty two soybean genotypes were used, being 30 elite cultivars and 2 North American accesses. Twenty two *primers* pairs of soybean microsatellite were used, from which 19 showed polymorphism. The use of those *primers* allowed for detection of 61 alleles at

different loci in all studied genotypes. The genotype diversity in the 22 studied loci varied from 0 to 0.73, reaching an average of 0.41. The number of alleles per locus varied from 1 to 5 with an average of 2.77. By combining the *primers* pairs SATT186, SATT094, SATT070 and SATT197 it was possible to identify 28 from the 32 genotypes. By maintaining these 4 *primers*, four different combinations of 5 *primers* pairs identified all studied genotypes, being enough to include one of the following *primers* pairs: SATT115, SATT184, SATT309 or SATT536. Based on the data relative to the frequency of common alleles in the 32 studied genotypes, the dissimilarities among them were determined. The dissimilarity measures varied from 9 to 67% and reached an average of 43%. Only two pairs presented a dissimilarity below 10%. Most pairs presented dissimilarity indexes above 20%. In order to classify all 32 soybean cultivars into homogeneous groups, the dissimilarity matrix was used to analyze the individuals grouping by agglomerative method. Three methods were used: the UPGMA, the Single Linkage and the Complete Linkage methods. The UPGMA method ("unweighted pair-group method using an arithmetic average") was the one representing better the inter-relationship among genotypes, with a cophenetic correlation of 70.5%. Both the Single Linkage and the Complete Linkage methods showed a cophenetic correlation with dissimilarity data of 58.6% and 56.9%, respectively. To obtain mutually exclusive groups the Tocher optimization method was used, that produced 6 groups, being group 1 formed by 25 individuals (divided into 8 subgroups) and groups 2 and 3 formed by 2 individuals each one, whereas the other three groups were formed by only an individual in each one.

1. INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill), leguminosa originária da China, é uma cultura muito importante em todo o mundo. No Brasil, é responsável pela criação de um segmento econômico da mais alta expressão, ocupando lugar de destaque na pauta de exportação do País.

O cultivo da soja alcança uma ampla faixa de latitude, abrangendo todas as regiões do território brasileiro.

O Brasil é atualmente o segundo maior produtor de soja, superado apenas pelos Estados Unidos. A produtividade média de cerca de 2.400 kg/ha é das maiores do mundo, e consegue-se facilmente produzir mais de 3.000 kg/ha. Esse bom desempenho deve-se principalmente à grande quantidade de cultivares melhorados existentes, adaptados a todas as regiões do país.

Apesar do grande número de variedades comerciais de soja existentes no Brasil, há pouca variabilidade genética entre elas, devido principalmente ao fato de serem originárias de um mesmo conjunto de poucos ancestrais.

A caracterização genética, tanto das variedades quanto de acessos, constitui uma grande contribuição aos programas de melhoramento, facilitando a escolha das melhores combinações de progenitores.

Nos países em que já vigoram as leis de proteção de cultivares, a caracterização de variedades por meio de marcadores de DNA tem sido de grande importância na proteção do direito intelectual do melhorista, sendo utilizada como prova legal em processos jurídicos.

Com a aprovação no Brasil da lei de proteção de cultivares (Lei nº 9.456, de 25 de abril de 1997), o Serviço Nacional de Proteção de Cultivares (SNPC), órgão vinculado ao Ministério da Agricultura, ficou responsável pelo registro e proteção de novas variedades. Além disso, com a adesão à convenção internacional para proteção de novas variedades vegetais, em sua versão revisada de 1978, o Brasil tornou-se o 43^o país a fazer parte da UPOV. Na prática, isso significa que o SNPC obriga-se a acatar os requerimentos de pedidos de proteção de cultivares estrangeiras ou procedentes do exterior para as espécies em proteção no Brasil, ou seja, algodão, arroz, batata, cana-de-açúcar, feijão, milho, soja, sorgo e trigo, e, em contrapartida, os obtentores nacionais terão direito ao requerimento de proteção de suas cultivares nos demais países membros da UPOV.

Para requerer a proteção de uma variedade é necessário demonstrar que ela é diferente de qualquer outra variedade da mesma espécie, ou seja, que ela apresente distinguibilidade, homogeneidade e estabilidade.

Atualmente a proteção de cultivares é feita com base em descritores morfológicos, tais como, coloração de flor, hipocótilo, hilo e pubescência, tipo de folha, hábito de crescimento, grupo de maturação, altura e ciclo, além de resistência a doenças, entre outros. Em culturas que possuem uma base genética estreita, como é o caso da soja, em que as variedades são obtidas pela hibridação entre um grupo elite de parentais geneticamente semelhantes, as novas variedades tendem a ser muito semelhantes entre si e muitas vezes indistinguíveis com base nestas características. Além disso, com o aumento do número de variedades protegidas no banco de dados do SNPC, torna-se cada vez mais difícil distinguir uma nova variedade daquelas já existentes.

Além destes aspectos, marcadores morfológicos apresenta a desvantagem de serem somente identificados, em sua maioria, na planta inteira ou adulta. Por outro lado, marcadores baseados em DNA, como RFLP ("Restriction Fragment Length Polymorphism"), RAPD ("Random Amplified Polymorphic DNA), AFLP ("Amplified Fragment Length Polymorphisms") e SSR

(microssatélites) podem ser utilizados em qualquer estágio de desenvolvimento da planta, desde que quantidades suficientes de DNA possam ser obtidas.

Além da proteção, a caracterização de variedades é importante também na análise de mistura varietal. No caso de análise de sementes para certificação, a presença de sementes de outras cultivares em quantidade superior à tolerada, condena todo o lote de sementes. Em muitos casos, a identificação destas sementes gera dúvidas, sendo necessário um método mais preciso de análise.

Tanto para processos de proteção de cultivares, quanto para análises de pureza genética de sementes, descritores mais informativos e estáveis são necessários, a fim de diferenciar claramente os genótipos. Marcadores de DNA podem ser a solução, pois produzem um padrão exclusivo “fingerprinting” para cada variedade.

Entre os marcadores moleculares os mais indicados para a caracterização dos genótipos de soja são os microssatélites, por sua natureza co-dominante e multi-alélica, além de serem extremamente conservados dentro da espécie. Além disso, existe no mercado, uma série de *primers* de microssatélites desenvolvidos para a soja, o que facilita a utilização desta ferramenta. A informatividade destes *primers*, dada em termos de porcentagem de heterozigosidade (ou diversidade genética do loco), está disponível para variedades americanas, mas nada se conhece em relação às variedades brasileiras. Esta informação é importante, pois representa o poder discriminatório de cada *primer* ou loco de microssatélites.

O principal objetivo deste trabalho foi o de estabelecer um protocolo, de fácil aplicação em análises de rotina, para a identificação molecular de genótipos de soja.

Os objetivos específicos foram:

- Comparar a eficiência dos diversos sistemas de eletroforese disponíveis na discriminação dos alelos de microssatélites.
- Estimar a informatividade dos locos de microssatélites comumente utilizados em estudos de diversidade genética de soja no BIOAGRO/UFV.

- Selecionar um conjunto de *primers* de microssatélites capazes de diferenciar um grupo de genótipos representativo das variedades de soja cultivadas no Brasil.

2. REVISÃO DE LITERATURA

2.1. A base genética da soja

A soja (*Glycine max* (L.) Merrill) é uma espécie de autopolinação originária da Ásia, muito provavelmente do nordeste da China (antiga Manchuria). Ela foi introduzida nos Estados Unidos em 1950 após ter adquirido importância econômica. No Brasil, a soja começou a ser cultivada em grandes extensões somente depois de 1970. Hoje, o Brasil ocupa o segundo lugar no ranking mundial em produção de soja (HIROMOTO e VELLO, 1986).

Embora o número de cultivares existentes hoje no Brasil seja elevado, a base genética que constitui o germoplasma é bastante restrita, situação semelhante ao que ocorre nos Estados Unidos. Segundo KEIM et al. (1989), existem duas possíveis explicações para essa baixa diversidade genética no germoplasma da soja americana: a) por tratar-se de uma espécie autógama, em que os indivíduos são altamente homocigotos, mutações deletérias que contribuiriam para a diversidade genética são eliminadas por seleção; b) a maioria dos cultivares existentes nos Estados Unidos apresentam ancestrais

comuns, limitando, consideravelmente, a diversidade genética. Segundo DELANNAY et al. (1983), a similaridade dos germoplasmas brasileiros e americanos confirmam a hipótese de que o germoplasma brasileiro foi formado a partir de linhagens introduzidas dos Estados Unidos.

Segundo HIROMOTO e VELLO (1986), os quatro ancestrais com maior contribuição para o germoplasma de soja brasileiro são os mesmos que mais contribuem para o germoplasma do sul dos Estados Unidos, indicando que os cultivares brasileiros são formados principalmente por genótipos vindos daquela região.

Marcadores moleculares vêm sendo bastante utilizados como ferramenta para estudo da variabilidade genética em plantas. ABDELNOOR et al. (1995) avaliaram a diversidade genética em 38 cultivares de soja, utilizando marcadores RAPD. Foram encontrados cinco subgrupos distintos A1, A2, B1, B2 e B3, no nível de 76% de distância genética relativa. Os resultados estão de acordo com o coeficiente de parentesco. Os 14 cultivares do grupo A apresentaram, pelo menos, 50% do conjunto gênico derivado da PI 60406 e Tanloxi, ao passo que os 24 cultivares do grupo B não evidenciaram contribuição desses progenitores. No subgrupo B1, composto de 14 cultivares, três ancestrais (Roanoke, Tokyo e PI 54610) contribuíram com pelo menos 37% do conjunto gênico desse subgrupo. Quatro, dos seis cultivares que compõem o subgrupo B2 apresentaram a PI 240664 como ancestral comum. Quatro, dos seis cultivares que compõem o subgrupo B3, apresentaram a PI 37335 como ancestral comum.

TINGEY et al. (1991), trabalhando com marcadores moleculares RFLP, por intermédio da análise de 33 genótipos de soja comprovaram a baixa diversidade do germoplasma norte-americano. Foi verificado que 46% dos locos estudados apresentavam o mesmo alelo, e que dos locos restantes, considerados polimórficos, aproximadamente 25% apresentaram um tipo alélico presente em mais de 90% dos genótipos estudados.

MAUGHAN et al. (1996) investigaram 23 acessos de soja cultivada e silvestre pela técnica de AFLP. A diversidade genética determinada por AFLP foi maior nos acessos silvestres do que na soja cultivada. A análise de agrupamento claramente separou os acessos de *G. soja* dos acessos de *G. max*. Dentro do grupo de *G. max*, as variedades de soja cultivada se

agruparam mais densamente, demonstrando a pequena diversidade existente na soja cultivada.

MACHADO (1999), trabalhando com marcadores AFLP para avaliar a diversidade genética presente em 263 acessos de soja, sendo 31 *G. soja*, 10 *G. max* adaptados e 222 *G. max* não-adaptados, constatou que a soja cultivada apresenta uma base genética muito estreita e que existe uma grande diversidade genética nos materiais exóticos.

POWELL et al. (1996), trabalhando com uma combinação de marcadores microssatélites nucleares e de cloroplastos para investigar os níveis e padrões de variabilidade detectados em genótipos de *G. soja* e *G. max*, constataram que apesar da pequena base genética da soja atualmente cultivada, existe ainda grande variabilidade genética nas populações silvestres.

2.2. Propriedade intelectual dos obtentores vegetais

Nos países onde vigoram as Leis de Proteção de Cultivares, o uso de marcadores moleculares tem tido grande importância na proteção do direito intelectual do melhorista (LANZA et al., 2000).

Com a aprovação em 25 de abril de 1997 da Lei nº 9.456 de Proteção de cultivares, o Brasil passou a contar com um novo mecanismo legal para proteger as variedades de plantas que apresentam, concomitantemente, características de novidade, distiguibilidade, homogeneidade e estabilidade genética (NETO, 1999).

A Lei de Proteção de Cultivares fornece, finalmente, os mecanismos legais para que o obtentor tenha seu direito intelectual reconhecido, possa recuperar razoavelmente os investimentos realizados e, ainda, sinta-se mais motivado a continuar com o processo criativo, com a possibilidade de reinvestimento na pesquisa. Além disso, com a adesão à Convenção Internacional para Proteção de Novas Variedades Vegetais, em sua versão revisada de 1978, o direito ao requerimento de proteção de cultivares estende-se aos demais países membros da UPOV (SNPC, 1999).

Segundo SMITH e HELENTJARIS (1996), três tendências atuais salientaram a necessidade crescente de se caracterizar e proteger as variedades de plantas:

- 1) a constatação de redução na segurança do germoplasma devido à apropriação indevida.
- 2) os avanços dos marcadores moleculares e das possibilidades de transformação genética, incrementaram a velocidade e a precisão com as quais os melhoristas podem lidar com o germoplasma. A transformação e o retrocruzamento auxiliado por marcadores moleculares estão se tornando mais comuns nos programas de melhoramento do setor privado. Algumas mudanças genéticas podem ser apenas superficiais e representar tentativas de plagiar ou falsificar as variedades protegidas. Os sistemas de marcadores moleculares também oferecem várias oportunidades para que se revelem as diferenças entre as variedades, embora muitas diferenças possam ser insignificantes em termos fenotípicos; e,
- 3) tem havido aumento na competitividade, à medida em que novas empresas têm iniciado seus programas de melhoramento de plantas, em grande parte por causa das novas oportunidades tecnológicas. Conseqüentemente, os melhoristas de plantas têm buscado o aperfeiçoamento das bases técnicas e legais de proteção.

Ainda, segundo SMITH e HELENTJARIS (1996), metodologias detalhadas para caracterização de variedades são úteis para fortalecer a Propriedade Intelectual de Plantas por meio de sua capacidade de fornecer identificação positiva do germoplasma de interesse e para verificar a autenticidade dos “pedigrees”.

2.3. Descrição de variedades baseada em dados morfológicos

A descrição de variedades é pré-requisito para a concessão de proteção. A fim de atender tal exigência, as descrições devem ser discriminativas, livres de efeitos ambientais, em condições de serem geneticamente interpretadas e que evidenciem seu pedigree e sua constituição genética (SMITH e HELENTJARIS, 1996).

Os dados morfológicos fornecem a base das descrições para o registro e a identificação de variedades (MILACH, 1998). Muitos consideram que esses dados não podem ser substituídos, já que eles representam o produto da expressão dos genes, o que os torna ótimos indicadores. Entretanto, os efeitos das interações genótipo x ambiente, as dificuldades crescentes em mostrar discriminação (para algumas culturas), os custos crescentes para a coleta de dados, o tempo necessário para coletar dados e a ausência de conhecimento de controle genético estão levando os examinadores da Lei de Proteção de Cultivares e os melhoristas a fazer uso de dados moleculares. Os dados morfológicos geralmente não permitem que as distâncias genéticas sejam calculadas; conseqüentemente, esses dados são de pouca utilidade na determinação do status de variedade essencialmente derivadas (SMITH e HELENTJARIS, 1996).

2.4. Descrição de variedades com base no DNA

Os métodos de descrição que utilizam diretamente o DNA, potencialmente, poderiam suprir todas as limitações apresentadas pelos dados morfológicos. As classificações baseadas no DNA não são afetadas pelos efeitos das interações genótipo x ambiente e a base de herdabilidade pode ser compreendida. As classificações podem ser interpretadas em termos de presença ou ausência de alelos específicos. Os marcadores com base no DNA podem ser encontrados abundantemente em todo o genoma. Podem haver excelentes condições de discriminação, mesmo entre variedades muito próximas. Os métodos podem ser padronizados e automatizados. Os custos e a velocidade com a qual é possível a obtenção de dados são aceitáveis para a identificação de variedade (SMITH e HELENTJARIS, 1996).

2.5. Uso de descritores moleculares: Visão da UPOV

Tanto os melhoristas de plantas quanto os examinadores da lei de proteção de cultivares desejariam conduzir o processo de registro com finalidade de proteção, de maneira mais rápida e mais barata. Embora os

métodos de descrição de variedade que utilizam o DNA possam atingir esses objetivos, argumenta-se que os dados advindos da maioria das tecnologias de DNA não atendem às exigências da UPOV (UPOV, 1992), que no artigo 1º da Convenção da UPOV de 1991, descreve uma variedade como sendo:

- i) “definida pela expressão das características resultantes de um dado genótipo e de combinação de genótipos” , e,
- ii) “distinta de qualquer outro grupo de plantas pela expressão de pelo menos uma das características mencionadas”.

Portanto, os dados de seqüência de DNA de regiões não codificadoras do genoma são eliminados dos atos de definição e distinção de variedades se o termo “expressão” for interpretado especificamente e unicamente para produtos genéticos. Por exemplo, os componentes da maior parte dos RFLP, *primers* arbitrários e marcadores dos locos de mini-satélite e microssatélites não são produtos de expressão genética. Esses dados, entretanto, são expressões visíveis do genoma, sendo, conseqüentemente, uma evidência das diferenças entre variedades. Além disso, o DNA de regiões não codificadoras poderia desempenhar importante papel na regulação, afetando, desta forma, a expressão do genótipo. A partir daí, os dados de marcadores moleculares poderiam ser incluídos dentro do grupo de características que definem as variedades para registro. Para muitos, parece incoerente que as descrições do DNA humano possam ser aceitas como evidência em casos de disputas forenses, mas que os dados de fontes molecularmente equivalentes não sejam aceitos como evidência de identidade e distinção em variedades de plantas (SMITH e HELENTJARIS, 1996).

2.6. Pureza genética

A caracterização molecular de determinadas variedades, além de poder ser utilizada nos processos de proteção de cultivares, constitui-se em ferramenta auxiliar muito poderosa na análise de pureza genética de sementes, aplicável sempre que a caracterização visual permitir dúvidas. Desta forma, bastará analisar as sementes atípicas com os *primers* que caracterizam a

variedade e comparar os alelos obtidos com aqueles característicos da variedade (LANZA et al., 2000).

2.7. Marcadores moleculares

Os métodos tradicionais utilizados na identificação de genótipos de plantas baseiam-se na morfologia e características agronômicas, sendo capazes de discriminar apenas um pequeno número de genótipos. A principal razão é o baixo nível de polimorfismo e a influência de diferentes parâmetros ambientais.

Com o advento das técnicas de Biologia Molecular, surgiram marcadores genéticos com base no DNA, destacando-se o RFLP, RAPD, AFLP e microssatélites, que se baseiam na variação natural da seqüência de base do DNA (FERREIRA e GRATTAPAGLIA, 1996).

O avanço principal que essas técnicas trazem é a possibilidade de acessar diretamente o genótipo de um indivíduo, evitando, assim, a expressão do fenótipo e a influência do ambiente sobre este. O fato de haver um número ilimitado de descritores de DNA disponíveis e em razão de possibilitarem o acesso à variabilidade genética em diversas espécies vegetais, faz com que eles sejam de grande importância para caracterização de cultivares (MILACH, 1998).

2.7.1. Marcadores RFLP

Os marcadores RFLP “Restriction Fragment Length Polymorphisms” foram os primeiros marcadores de DNA a serem utilizados na construção de mapas genéticos na espécie humana (BOTSTEIN et al., 1980). Posteriormente, os RFLPs foram amplamente empregados nas mais variadas áreas da genética animal e vegetal.

A técnica do RFLP é baseada na digestão do DNA genômico com enzimas de restrição, separação dos fragmentos em gel de agarose e transferência destes para membranas de nitrocelulose, onde são hibridizados com sondas de DNA marcadas. O polimorfismo é obtido quando ocorre, por mutação de ponto, perda ou surgimento de sítios de restrição, que são

seqüências específicas de 4 a 6 nucleotídeos reconhecidas pelas endonucleases, ou quando ocorrem inserções ou deleções entre os dois sítios de restrição que flanqueiam uma região homóloga à sonda de DNA. Os marcadores RFLP apresentam herança co-dominante, sendo possível a identificação de indivíduos homozigóticos e heterozigóticos na população.

Embora o enorme potencial e a utilidade dos marcadores RFLP sejam indiscutíveis, a técnica em foco apresenta algumas limitações, como o número excessivo de passos, impossibilidade de automação e capacidade restrita de detectar polimorfismo em espécies com estreita base genética como a soja. Outros estudos relatam a inabilidade dos marcadores RFLP em criar um perfil único de DNA em genótipos de soja devido ao baixo número de alelos por loco (KEIM et al., 1989; KEIM et al., 1992). KEIM et al. (1989) trabalhando com marcadores RFLP para caracterizar 58 genótipos de soja, encontraram apenas 2 alelos em 15 dos 17 locos estudados. Nos 2 locos restantes foram encontrados apenas 3 alelos.

2.7.2. A técnica de PCR

O advento da técnica de PCR (“Polymerase Chain Reaction”) e seus posteriores avanços, utilizando uma enzima *DNA polimerase* termoestável e termocicladores programáveis com elevada capacidade de processamento, imprimiram grande automatização à síntese *in vitro* de DNA. A técnica de PCR consiste na amplificação simultânea das duas fitas de DNA de um segmento específico, usando-se dois oligonucleotídios (*primers*) de sequência conhecida, que são complementares às extremidades do segmento a ser amplificado (SAIKI et al., 1985). *Primers* são seqüências curtas de DNA, que pareiam ao DNA-molde e servem de iniciadores para a síntese *in vitro* de uma nova fita de DNA. Na técnica de PCR, os *primers* flanqueiam o segmento de DNA a ser amplificado e as reações ocorrem em ciclos alternados de temperatura, sendo que cada ciclo do PCR envolve três etapas: desnaturação, anelamento e extensão. Na primeira, ocorre a desnaturação da fita dupla de DNA, posteriormente, os *primers* pareiam-se com as seqüências complementares específicas que flaqueiam o sítio-alvo, e então a nova fita de DNA é sintetizada

a partir das extremidades 3' dos *primers* por meio da enzima DNA polimerase. Como cada ciclo é repetido várias vezes, a amplificação do DNA-alvo ocorre em progressão geométrica, requerendo uma quantidade muito pequena de DNA-molde.

A facilidade, a rapidez e a sensibilidade dessa técnica possibilitaram o surgimento de uma nova geração de marcadores moleculares baseados em PCR.

2.8. Marcadores moleculares baseados em PCR

2.8.1. Marcadores RAPD

A técnica denominada RAPD (“Random Amplified Polimorfic DNA”) (WILLIAMS et al., 1990; WELSH E McCLELLAND, 1990) é uma variação da técnica de PCR que possibilita a síntese de DNA *in vitro*, a partir de um DNA molde. É utilizado um *primer* de sequência aleatória de 10 pares de base (pb), que vai se ligar a regiões complementares no DNA-alvo. A amplificação ocorrerá quando o mesmo *primer* estiver flaqueando um segmento de DNA em direção oposta e em sítios distanciados de 500 a 2000 pb, o que possibilita a síntese desse segmento pela *Taq* DNA polimerase. Com isso, são amplificados fragmentos de DNA, distribuídos ao acaso no genoma, sem a necessidade do conhecimento prévio da sequência de DNA. Repetidos ciclos de desnaturação do DNA-alvo, ligação do *primer* e extensão dos mesmos produzem os fragmentos de DNA, que são visualizados em gel de agarose. O número e o tamanho dos fragmentos amplificados depende da sequência do *primer* usado e da sequência de bases do DNA molde (TINGEY & DEL TUFO, 1993).

O polimorfismo evidenciado pela técnica de RAPD é devido a mutações de ponto, ou deleções no sítio de pareamento do *primer*, ou inserções entre sítios de pareamento, deixando-os a uma distância tal que impossibilita a sua amplificação. O RAPD é uma técnica de fácil execução, de custo reduzido e aplicável a qualquer tipo de organismo. No entanto, existem problemas inerentes à reprodutibilidade dos padrões de amplificação, além do baixo

conteúdo de informação genética por loco, uma vez que são marcadores dominantes. Este tipo de marcador possibilita a detecção de apenas um alelo por loco, sendo que a presença de uma banda no gel identifica indivíduos homocigotos dominantes (AA) ou heterocigotos (Aa), não permitindo a distinção entre eles. O homocigoto recessivo (aa) é identificado pela ausência da banda. Ausência de banda pode indicar ainda a presença de qualquer outro alelo, que não aquele que produz a banda, em locos multi-alelicos.

2.8.2. Marcadores AFLP

O AFLP ("Amplified Fragment Length Polymorphisms") é uma classe de marcadores que alia a especificidade dos sítios de restrição do RFLP à praticidade da amplificação do PCR, apresentando-se como uma ferramenta poderosa na caracterização de genomas e no mapeamento genético (VOS et al., 1995). Os marcadores são fragmentos de restrição detectados após uma amplificação seletiva por PCR. O DNA genômico é digerido simultaneamente com duas enzimas de restrição, sendo *EcoRI* e *MseI* as mais usadas. A seguir, os fragmentos de DNA gerados são ligados a adaptadores nos sítios de *EcoRI* e de *MseI* e amplificados por PCR, utilizando-se *primers* que contêm seqüências complementares aos adaptadores. A seletividade é alcançada pela adição de nucleotídeos arbitrários à extremidade 3' do *primer*, causando aumento na seletividade do *primer* e resultando numa redução na complexidade dos produtos de amplificação por PCR. A detecção dos fragmentos polimórficos é feita em gel de seqüenciamento, utilizando um dos *primers* seletivos marcados com radioatividade ou fluorescência.

Uma das vantagens do AFLP é o grande poder de detecção de variabilidade genética, uma vez que a técnica explora polimorfismos de restrição e de amplificação. Como os *primers* utilizados nas etapas de amplificação são longos, em torno de 20 pb, a especificidade da reação é aumentada significativamente, quando comparada com o RAPD. Uma das limitações do AFLP é o baixo conteúdo de informação genética por loco que, à semelhança do RAPD, também são marcadores essencialmente dominantes. Como o AFLP é realizado em várias etapas, incluindo digestão do DNA com

enzimas de restrição, a técnica torna-se mais complexa. Outro fato que contribui para aumentar esta complexidade é que a resolução dos polimorfismos precisa ser realizada em géis de seqüenciamento, utilizando radioatividade ou fluorescência, e nem todos os laboratórios estão capacitados para trabalhar com tais metodologias (FERREIRA e GRATTAPAGLIA, 1996).

2.8.3. Marcadores microssatélites (SSRs)

Os genomas de organismos eucariotos são caracterizados pela ocorrência de grandes quantidades de seqüências repetidas de nucleotídeos de várias classes, dentre elas, se destaca uma classe de seqüências de DNA curtas (2 a 4 bp), repetidas em tandem, denominadas seqüências simples repetitivas (SSR) ou microssatélites (LITT& LUTY, 1989).

As seqüências de microssatélites têm sido detectadas em diversos organismos. Em geral, no DNA de plantas a seqüência (AT)_n está presente em maior abundância (Quadro 1).

Regiões contendo seqüências simples repetitivas são amplificadas individualmente por meio de PCR utilizando-se um par de *primers* específicos (de 20 a 30 bases) complementares a seqüências únicas que flanqueiam o microssatélite. A amplificação ocorrerá quando dois *primers* estiverem flaqueando um segmento de DNA em direção oposta e distanciados de 100 a 300 pb, o que possibilita a síntese desse segmento pela *Taq* DNA polimerase.

Quadro 1 – Frequência de microssatélites (SSR) em plantas

Tipo de SSR	Número de SSR	Porcentagem
Mono	21	16,2
Di		
AT	49	37,7
AG	10	7,7
AC	3	2,3
Tri	29	22,3
Tetra	18	13,8
Total	130	-

Fonte: WANG et al. (1994)

Tendo em vista a expressão co-dominante e o multialelismo, os marcadores SSR são os que possuem o mais elevado conteúdo de informação de polimorfismo por loco. Todas essas características têm feito dos microssatélites o marcador mais apropriado para o desenvolvimento de *fingerprints* genéticos únicos na discriminação de cultivares (CREGAN et al., 1994; THOMAS et al., 1994; RUSSEL et al.; 1997; DIWAN e CREGAN, 1997).

A identificação de locos microssatélites exige a confecção de uma biblioteca de DNA digerido, com fragmentos de aproximadamente 500 pb. Estes fragmentos são clonados em plasmídios específicos e utilizados na transformação de *Escherichia coli*. Colônias transformadas são hibridizadas com sondas de nucleotídeos radioativos tais como G(AT)₁₁ e CCC (ATT)₁₀. O DNA dos plasmídios que hibridizarem é isolado e sequenciado para identificação de uma seqüência com repetições curtas (características de microssatélite).

Uma vez obtidos os *primers*, marcadores de microssatélites são mais fáceis de utilizar do que RFLP, pois são baseados na técnica de PCR, que permite a avaliação de grande número de amostras em curto espaço de tempo. Para soja, já estão disponibilizados no mercado cerca de 650 pares de *primers* de microssatélites pela Research GeneticsTM, o que facilita grandemente a utilização desta técnica como rotina, pois não é necessário identificar os locos microssatélites. Além disso, os SSRs são mais baratos e requerem menos trabalho do que os marcadores RFLPs, apresentando em soja alto grau de polimorfismo por loco (AKKAYA et al., 1992).

KEIM et al. (1989), utilizaram 17 sondas de RFLP para caracterizar 58 genótipos de soja, incluindo variedades adaptadas e silvestres. Das 17 sondas utilizadas, 15 apresentaram apenas dois alelos, e as outras duas apresentaram 3 alelos. Em um trabalho posterior, KEIM et al. (1992) utilizaram 132 sondas de RFLP para caracterizar 38 genótipos de soja, incluindo também variedades adaptadas e silvestres. Apenas três sondas apresentaram mais de dois alelos.

RONGWEN et al. (1995), utilizando 7 pares de *primers* de microssatélites para caracterizar 96 genótipos de soja, encontraram de 11 a 26 alelos por loco. Apenas 4 pares de *primers* foram suficientes para discriminar 95 dos 96 genótipos avaliados.

KRAIC et al. (1998), comparando a capacidade de diferenciação dos marcadores RAPD e SSR em cultivares comerciais de cevada (*Hordeum vulgare* L.), verificaram que de 828 reações de RAPD (23 cultivares x 36 *primers* RAPD utilizados) resultou na seleção de quinze *primers* geradores de polimorfismo. Em contrapartida, somente 115 reações de amplificação (23 cultivares x 5 *primers*) foram necessárias para se ter uma diferenciação completa de todos os 23 cultivares por meio de *primers* de SSR. Deste modo, a efetividade de diferenciação de *primers* de SSR foi 7,2 vezes maior quando comparado com os *primers* de RAPD.

RUSSELL et al. (1997), estudando 11 locos de SSR em 24 genótipos de cevada, encontraram três combinações de quatro SSR, que identificaram todos os genótipos avaliados. Estes autores conseguiram distinguir inclusive duas linhas irmãs, que possuíam distância nula, com base na genealogia.

3. MATERIAL E MÉTODOS

3.1. Genótipos de soja

Os experimentos foram conduzidos no Instituto de Biotecnologia Aplicada à Agropecuária – BIOAGRO/UFV. As 30 variedades utilizadas neste trabalho foram selecionadas de modo a representar o germoplasma brasileiro, após uma amostragem prévia (divididas por origem, ciclo e local de cultivo) de um grupo de variedades utilizadas comercialmente. Além das variedades comerciais, dois acessos norte-americanos (Quadro 2) utilizados como fontes doadoras de genes nos programas de melhoramento, também foram incluídos. As sementes foram obtidas junto à EMBRAPA - Soja.

Trinta sementes de cada genótipo foram semeadas em bandejas, em casa de vegetação. Após o surgimento do primeiro trifólio, as folhas de cada genótipo foram colhidas em “bulk” e submetidas a extração do DNA.

Quadro 2 – Variedades comerciais de soja (*G. max*) , origem, ciclo e local de cultivo.

	Variedades e acessos	Origem	Ciclo	Local de cultivo
1	BR 4	EMBRAPA	Precoce/Semiprecoce	PR/SP/GO/DF
2	Bela Vista	EMGOPA	Tardio	GO/DF/TO
3	BRS 137	EMBRAPA	Precoce/Semiprecoce	RS/SC
4	Uirapuru	FUNDAÇÃO MT	Semitardio/Tardio	MT/RO/BA
5	DM 247	2 MARCOS	Médio/Precoce	MG/GO/DF/MT
6	DM Rainha	2 MARCOS	Precoce/Semiprecoce	MG/GO/DF
7	Hartwig	EUA	Precoce	-
8	KI-S 801	INDUSEM	Semitardio/Precoce/semiprecoce	PR/MT
9	Pioneira	FUNDAÇÃO MT	Médio	MT
10	Emgopa 316	EMGOPA	Precoce	SP/GO/DF
11	BR/IAC 21	EPAMIG	Precoce/Médio/Semitardio	TO/GO/DF/MG
12	Emgopa 305	EMGOPA	Médio	TO
13	Ipagro 21	FEPAGRO	Médio	RS
14	FT Estrela	MONSOY	Precoce/Semiprecoce/Tardio	MS/MG/GO/DF/SP/PR
15	FT Abyara	MONSOY	Médio/Semiprecoce	PR/SP/MS
16	RS-7 Jacuí	FEPAGRO	Médio	RS
17	UFV-16	UFV	Médio	MG
18	IAS 4	EMBRAPA	Médio	RS/SC
19	IAS 5	EMBRAPA	Precoce	PR/SP/MS/GO/DF
20	CD 205	COODETEC	Médio/Semiprecoce	PR/MG
21	BARC-8	EUA	Precoce	-
22	FT Guaira	MONSOY	Precoce	PR/SP
23	Pequi	FUNDAÇÃO MT	Precoce/médio	MS
24	Emgopa 312	EMGOPA	Precoce	MA
25	Planalto	FEPAGRO	Médio	RS/SC/PR
26	Embrapa 59	EMBRAPA	Simiprecoce	PR/SP
27	DM Nobre	2 MARCOS	Tardio	SP/GO/MT/TO/NE/NO
28	UFV-6	UFV	Tardio	ES
29	Pintado	FUNDAÇÃO MT	Médio	MT
30	FT 8	MONSOY	Médio	PR
31	IAC 8	IAC	Médio	SP
32	BR 1	EMBRAPA	Médio/Precoce	PR/RS

Fontes: EMBRAPA-Soja, FUNDAÇÃO MT e ABRASEM

3.2. Extração de DNA

A extração do DNA de folhas de soja foi realizada com base no protocolo descrito por DOYLE e DOYLE (1990), com algumas modificações. Cerca de 200 a 300 mg de folhas foram trituradas na presença de N₂ líquido, sendo o pó resultante transferido para tubos eppendorf. Após a adição de 650-800 µl de tampão de extração, constituído de Tris-HCl (pH 8,0) 50mM, EDTA (pH8,0) 50mM, NaCl 0,7M, CTAB (1% p/v) e β-mercaptoetanol (1% v/v), sendo este último adicionado separado dos demais componentes, as amostras foram,

então, incubadas em banho-maria a 65°C por, aproximadamente, uma hora. Após a incubação, as proteínas foram removidas pela extração com igual volume de clorofórmio, álcool-isoamílico (24:1) e centrifugação a 14.000 RPM em centrífuga eppendorf modelo 5415C. Os ácidos nucleicos foram, então, precipitados da fase aquosa, pela adição de 2/3 do volume obtido de isopropanol gelado, lavados em etanol (70% e 95%) gelado por 20 minutos, secados ao ambiente e ressuspensos em TE, pH 8,0 (10mM de Tris-HCl e 1mM de EDTA), seguindo-se um tratamento com RNase A, na concentração final de 40 µg/mL a 37°C, por 30 minutos. O DNA foi recuperado por precipitação, mediante a adição de um volume de etanol 95%, seguido de centrifugação a 14.000 RPM em centrífuga eppendorf modelo 5415C, por 15 minutos. Após a lavagem do precipitado com etanol (70% e 95%) gelado, e secagem ao ambiente, o DNA foi ressuspendido em 200-300 µL de TE.

A concentração do DNA foi estimada espectrofotometricamente por leitura da absorbância a 260nm, sendo que cada unidade de absorbância corresponde à concentração de 50 µg/mL de DNA fita dupla (SAMBROOK et al., 1989).

A integridade do DNA foi determinada em gel de agarose 0,8%, fotografado sob luz ultravioleta.

3.3. Condições de amplificação e separação dos fragmentos

Foram utilizados os 22 pares de *primers* de microssatélites de soja, de um total de 650, disponibilizados pela Research GeneticsTM. Os *primers* foram selecionados com base em trabalhos científicos realizados com microssatélites e também a partir de resultados de outros trabalhos realizados no BIOAGRO/UFV. Aqueles *primers* que mostraram-se bastante polimórficos, foram testados nas variedades em estudo.

As reações de amplificação de microssatélites foram realizadas em um volume total de 15 µL, contendo 12,5 mM de Tris-HCl (pH 8,3), 62,5 mM de KCl, 2,5 mM de MgCl₂, 125 µM de cada um dos deoxinucleotídeos (dATP, dTTP, dGTP e dCTP), 0,6 µM do *primer*, uma unidade da enzima Taq DNA polimerase e 18 ng de DNA. As amplificações foram realizadas em

termociclador programado para uma etapa inicial de 7 minutos a 72°C, seguida de 30 ciclos de 1 minuto a 94°C, 1 minuto a 50°C e 2 minutos a 72°C. Por fim, uma etapa de 7 minutos a 72°C.

Foram testadas várias condições de separação de fragmentos de DNA em gel de agarose e poliacríalmida de concentração variando de 4-12%, nativo ou desnaturante, com ou sem o polímero Spreadex™. Este polímero facilita a separação dos fragmentos, retendo os fragmentos maiores na parte superior do gel e empurrando os fragmentos menores para a parte inferior. Os géis de poliacrilamida nativos foram corados com brometo de etídio e os géis desnaturantes foram corados com prata. Também foram testados diferentes tamanhos de géis com dimensões de 10,2cm x 7,3cm x 1mm, 16,5cm x 14,5cm x 4mm e 19,5cm x 38cm x 5mm, respectivamente, para selecionar o método mais eficiente na identificação dos alelos em cada loco, e que fosse ao mesmo tempo econômico e de fácil aplicação em análises de rotina.

3.4. Seqüências dos *primers* de microssatélites

No quadro 3 pode ser visto as seqüências de *primers* de microssatélites que foram utilizados no estudo de diversidade genética e na identificação dos 32 genótipos.

Quadro 3 – Seqüência dos *primers* dos locos microssatélites que foram usados no estudo de diversidade genética e na caracterização dos 32 genótipos de soja avaliados

Loco	<i>Primer</i>	
	Senso	Anti-senso
Satt070	TAAAAATTTAAATACTAGAAGACAAC	TGGCATTAGAAAATGATATG
Satt094	CCAAGTGCCAATGAAG	ATCCATGGTTTTTTTGATG
Satt186	GGGAAGTTATAAGCAGAT	GGGATTCCATTCCCTGATGAGT
Satt197	CACTGCTTTTTCCCCTCTCT	AAGATACCCCAACATTATTTGTAA
Satt115	GGTTCGTTTTTTTATTGATG	ACGACGAAATTGATGATAA
Satt307	GCGCTGGCCTTTAGAAC	GCGTTGTAGGAAATTTGAGTAGTAAG
Satt184	GCGCTATGTAGATTATCCAATTACGC	GCCACTTACTGTTACTCAT
Satt536	GCGCCACAGAAATTCCTTTTTCTA	GCGCCATAAGGTGGTTACCAAAG

Continua ...

Loco	Senso	Primer	Anti-senso
Sat_085	GGTTTTAGATCCTTAAATTTGT	GGGGAAGCAAGTAGCT	
Satt002	TGTGGGTAAAATAGATAAAAAT	TCATTTTGAATCGTTGAA	
Satt005	TATCCTAGAGAAGAAGACTAAAAA	GTCGATTAGGCTTGAAATA	
Satt009	CCAACCTGAAATTACTAGAGAAA	CTTACTAGCGTATTAACCCTT	
Satt079	AGTCGAAGATACACAATTAGAT	CTTTTAGACACAAATTTATCACT	
Satt082	AATTCATTTAGGGAGTTGAT	CTAGCCAATGTCATATGACT	
Satt136	CCCCATATATTCAACATATCTTCA	AACTAAGCCAAATCTTTTCCTACTA	
Satt156	CGCACCCCTCATCCTATGTA	CCAACCTAATCCCAGGGACTTACTT	
Satt162	GGGAAGAAGTATATGCTACATCAA	GGGTTAATTTTTATCTTCTAATAGTTT	
Satt166	TTGCACAGTTGATTTTTGTTT	GCATCGAATTTCTGGATTTAC	
Satt172	AGCCTCCGGTATCACAG	CCTCCTTTCTCCCATTTT	
Satt179	GGGATTAGTTTTATGGAAGTTTATTAT	GGGTCATTAACGATCAGTAAGA	
Satt282	TGCGGGACAATTTTAGATTTAT	AGATGGATGAAAGTTGGAGAAAG	
Satt537	CGTGTACATCAAATGGTGTATTC	CGGGGAGGAACTTTGTCTCAGTAAT	

Satt indica que o core do microssatélite é (ATT)_n. Sat_ indica que o core do microssatélite é (AT)_n

3.5. Cálculo da diversidade genética do loco

Para todos os *primers* testados foi calculada a diversidade genética do loco na identificação dos genótipos. WEIR (1990) sugeriu o termo *gene diversity* para descrever o valor relativo de uma sonda de RFLP com respeito a quantidade de polimorfismo que é detectada em um grupo de indivíduos homozigotos de uma espécie autógama. A mesma expressão se aplica para *primers* de microssatélites. Neste trabalho, o termo foi traduzido para diversidade genética do loco, que é definida como:

$$\text{Diversidade genética do loco} = 1 - \sum P_{ij}^2,$$

onde P_{ij} é a frequência do alelo j para o *primer* ou loco i . O valor da diversidade genética do loco é muito similar a heterozigosidade que é comumente utilizada para descrever a informatividade de um marcador molecular em espécies alógamas. ANDERSON et al (1993) indicaram que a diversidade genética do loco é essencialmente a mesma da informação do conteúdo de polimorfismo (PIC – *polymorphism information content*) como utilizado por BOTSTEIN et al. (1980). KEIM et al. (1992) referem-se à diversidade genética do loco como um índice de polimorfismo.

3.6. Cálculo do coeficiente de coincidência simples

O coeficiente de coincidência simples foi calculado para cada possível combinação dos 32 cultivares utilizando os alelos de cada loco microssatélite avaliado. O coeficiente de coincidência simples é igual ao número de alelos idênticos dividido pelo número de comparações, isto é, o número de pares de *primers* testados.

3.7. Análise de Agrupamento

Esta análise envolve, basicamente, duas etapas, sendo a primeira referente à estimação de uma medida de similaridade (ou dissimilaridade) entre os materiais a serem agrupados e a segunda, à adoção de uma técnica de agrupamento para formação dos grupos (CRUZ, 1990).

Visando classificar os 32 cultivares de soja em grupos homogêneos, foi gerada uma matriz de distância e, posteriormente, realizado o agrupamento dos indivíduos pelo método hierárquico aglomerativo. Foram utilizados os métodos UPGMA (*unweighted pair-group using an arithmetic average*, ou método das médias das distâncias), vizinho mais próximo (*single linkage*) e vizinho mais distante (*complete linkage*).

A eficiência do método de agrupamento foi avaliado através do coeficiente de correlação cofenética, observando a concordância entre a matriz de dissimilaridade original e o dendrograma. O cálculo da correlação cofenética (r_{cof}) foi realizado através da fórmula:

$$r_{cof} = r_{DC} = \frac{C\hat{O}V(D, C)}{\sqrt{\hat{V}(D)\hat{V}(C)}}$$

Em que D representa a matriz de distâncias, e C a matriz cofenética, obtida a partir do dendrograma.

COV = Covariância

V = Variância

Outro método de agrupamento utilizado foi o método de otimização de Tocher. Neste método, adota-se o critério de que a média das medidas de dissimilaridade dentro de cada grupo deve ser menor que as distâncias médias entre quaisquer grupos.

O método requer a obtenção da matriz de dissimilaridade, sobre a qual é identificado o par de genótipos mais similar, que constituirão o grupo inicial. Em seguida, é avaliada a possibilidade de inclusão de novos indivíduos no grupo inicial.

A inclusão de um novo genótipo sempre aumenta o valor médio da distância dentro de um determinado grupo, podendo-se, assim, decidir sobre a inclusão deste novo genótipo em um outro grupo por meio da comparação entre o acréscimo no valor médio da distância dentro do grupo e um nível máximo permitido, que pode ser estabelecido arbitrariamente, ou adotar, como tem sido geralmente feito, o valor máximo da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada indivíduo (CRUZ e REGAZZI, 1994).

4. RESULTADOS E DISCUSSÃO

4.1. Separação dos fragmentos

Experimentos foram realizados com a finalidade de padronizar a metodologia de separação e análise dos fragmentos de DNA por eletroforese. Em ordem de praticidade, rapidez e simplicidade de adoção em análises de rotina, encontram-se o gel de agarose, gel de poliacrilamida nativo (ambos corados com brometo de etídio) e por último o gel de poliacrilamida desnaturante corado com prata. Entretanto, o gel de agarose possui pouco poder de resolução, quando comparado com os géis de poliacrilamida.

Na Figura 1, pode ser observado que os géis de poliacrilamida apresentaram uma maior eficiência na separação dos fragmentos em relação ao gel de agarose, sendo portanto este último inadequado para este tipo de análise. O gel de poliacrilamida na concentração de 8% em relação ao gel de poliacrilamida 6%, neste experimento, teve melhor eficiência na separação em virtude dos tamanhos dos fragmentos serem pequenos, e da menor porosidade do gel à 8% que aliado ao efeito do polímero SpreadexTM, possibilitaram a

retenção dos fragmentos maiores na parte superior do gel e avanço dos menores para a parte inferior.

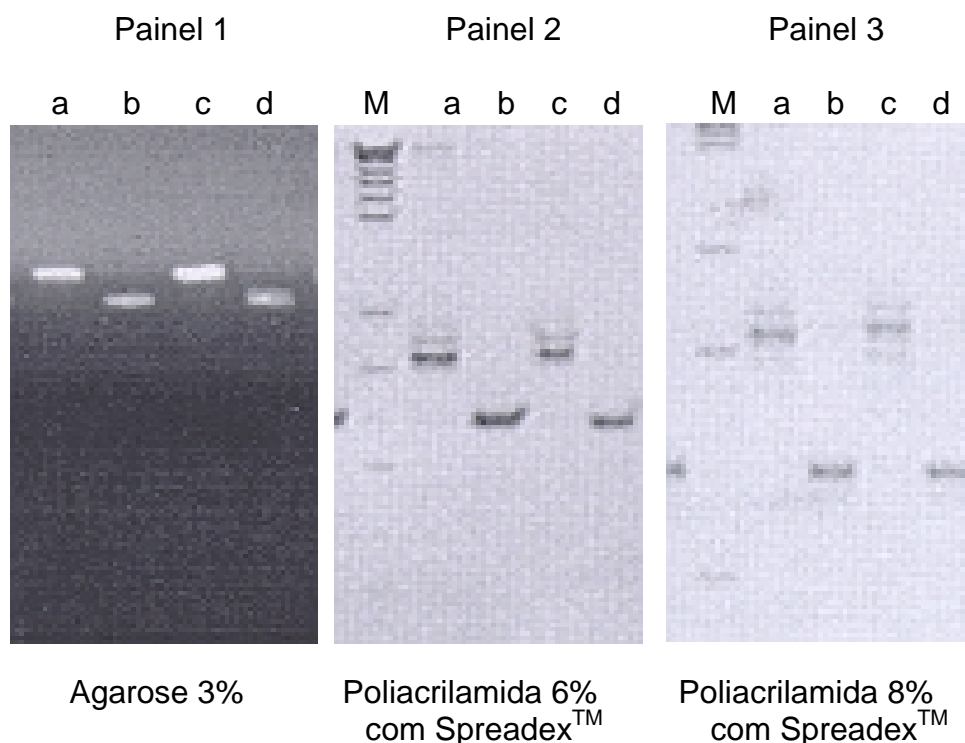


Figura 1 – Comparação de três métodos de separação de fragmentos de 194 e 240pb. Os três painéis contêm as mesmas amostras de DNA amplificadas com o par de *primers* SATT141. (a) e (c) refere-se ao fragmento de 240 pb e (b) e (d), ao fragmento de 194 pb. M refere-se ao marcador pUC18/*M*spl.

Géis nativos de poliacrilamida podem ser feitos em diversos tamanhos e espessura. Dentre os géis de poliacrilamida que oferecem maior facilidade em termos de confecção e manuseio, está o gel de dimensões: 10,2cm x 7,3cm x 1mm. Para aumentar o poder de resolução destes géis, cuja corrida é de apenas 7,3 cm, o polímero Spreadex™ foi utilizado juntamente com a poliacrilamida. A utilização conjunta destes dois polímeros aumentou a capacidade de separação de fragmentos de pequeno tamanho, de forma que os fragmentos de tamanho maior ficaram acumulados na parte superior do gel.

Desta forma, géis de poliacrilamida na concentração de 8%, utilizando Spreadex™, mostraram ser eficientes na separação de fragmentos com tamanho até 250 pb (Figura 1). Para separar fragmentos maiores, géis de 6% e até de 4% de poliacrilamida devem ser utilizados, pois quanto menor a

concentração do gel, maior é sua porosidade e, conseqüentemente, maior facilidade para separar fragmentos acima de 250 pb. A utilização de géis 8% para separar fragmentos grandes, resultará no acúmulo destes na parte superior do gel, sem que haja separação. No entanto, é possível utilizar géis de 6% de poliacrilamida para separar fragmentos com tamanho até 250 pb, desde que se adeque o tempo de corrida ao tamanho do fragmento que se deseja visualizar.

4.2. Determinação do número de alelos

A análise de fragmentos de DNA baseada unicamente nos resultados obtidos por programas computacionais pode levar a erros, quando ocorre fragmentos que diferem em poucos pares de bases. Na Figura 2 pode ser visto a análise de um gel feita pelo programa One Dscan. Por este resultado, 8 alelos que diferem por mais de três pares de bases (173, 178, 182, 194, 209, 216, 220, 233 pb) seriam obtidos no loco amplificado pelo par de *primers* SAT_085. Uma análise visual em cada painel, no entanto, demonstra a presença de apenas 4 ou 5 alelos. Observa-se que os tamanhos dos fragmentos de DNA calculados variam de um painel para outro, levando a um total de 8 alelos (4 no gel A e 4 no gel B da Figura 2).

Para determinar o verdadeiro número de alelos amplificados a partir do par de *primers* SAT_085, DNA de indivíduos contendo cada um dos 8 possíveis alelos definidos pelo programa One Dscan foram amplificados novamente, e aplicados em um gel, por ordem crescente de tamanho dos fragmentos identificados pelo programa. Na Figura 3 pode ser visto o resultado deste teste, demonstrando que só existem 5 alelos neste loco para os indivíduos analisados. Os indivíduos 16 e 1 (gel A da Figura 2) apresentaram fragmentos de DNA de tamanhos de 173 e 178 pb. Conforme mostrado na Figura 3, esta diferença calculada de 5 pb não resulta de alelos diferentes, mas da precisão da estimativa quando duas canaletas distantes entre si são comparadas por algum algoritmo computacional. O genótipo 18 (gel B da Figura 2) teve tamanho de fragmento calculado em 182 pb, apresentando diferença de 4 pb para o genótipo 1, e 9 pb para o genótipo 16 (gel A da Figura 2). Estes 3

genótipos apresentam, na verdade, o mesmo alelo, cujo tamanho calculado na Figura 3 foi de 174 pb.

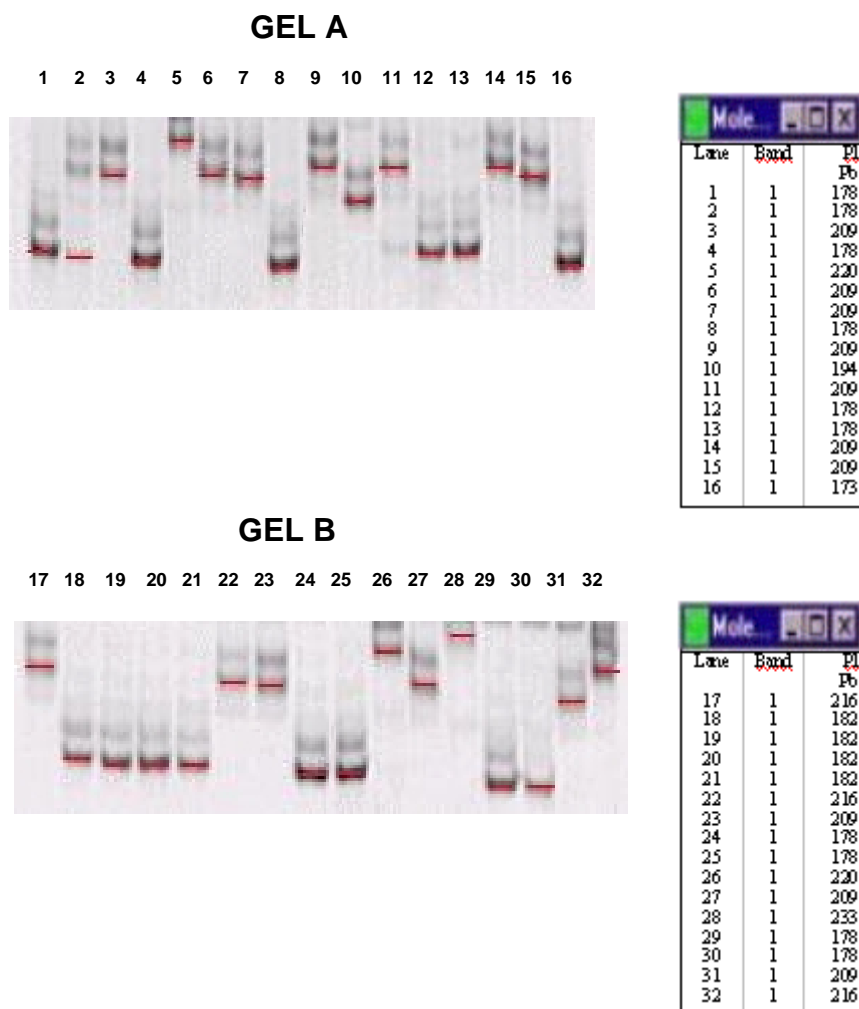


Figura 2 – Loco SAT_085 amplificado nos 32 genótipos avaliados. Cada painel apresenta 16 indivíduos. Foi utilizado o marcador pUC-18/MspI na determinação dos tamanhos dos fragmentos. As tabelas ao lado de cada gel apresentam os tamanhos calculados pelo programa One Dscan, dos fragmentos contidos em cada canaleta. Em cada canaleta, as bandas de microssatélites analisadas estão marcadas. As bandas mais fracas localizadas acima das marcadas é um fenômeno típico que ocorre em amplificações de microssatélites (“bandas fantasmas”). Os números, acima do gel, correspondem aos genótipos descritos no Quadro 2.

Como pode ser observado, alelos idênticos podem apresentar tamanhos calculados de fragmentos de DNA diferentes de um gel para o outro. Os genótipos 3 (gel A da Figura 2) e 17 (gel B da Figura 2) apresentaram tamanhos de fragmentos calculados de 209 e 216 pb. Novamente uma diferença de 7 pb entre dois fragmentos de DNA que representam o mesmo

alelo, mas que foram realizados scoring em géis diferentes. No gel da Figura 3, este fragmento foi calculado como possuindo 195 pb. O alelo do genótipo 10 apresentou 194 pb pelo gel A da Figura 2 e 188 pb pelo gel da Figura 3. O alelo do genótipo 5 apresentou 220 e 208 pb pelas Figuras 2 e 3, respectivamente. O alelo do genótipo 28 apresentou 233 pb na Figura 2 e 217 pb na Figura 3.

Estes resultados indicam que a avaliação visual é muito importante no momento de analisar o gel, e que a melhor forma de solucionar as dúvidas em relação à identificação dos alelos, é aplicar as amostras duvidosas em um novo gel, lado a lado. Desta forma a comparação é mais precisa. A interpretação visual foi levada em consideração nas análises dos géis do presente trabalho. Para os 22 locos avaliados, foram identificados 61 alelos.

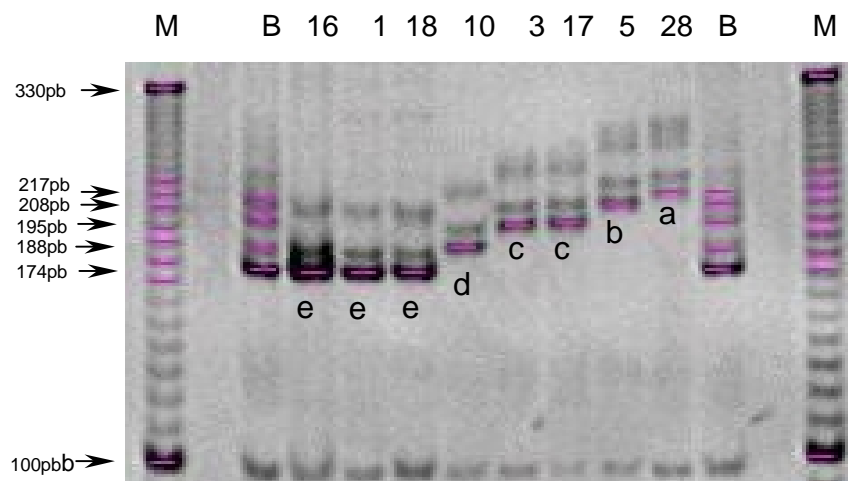


Figura 3 – Loco microssatélite Sat_085 apresentando 5 alelos. As setas indicam o tamanho dos alelos (em pares de base) dos genótipos. As canaletas indicadas com a letra B é o marcador específico feito através do “bulk” dos alelos. As canaletas indicadas pela letra M contém o marcador de 10pb da GIBCO. Os números de cada canaleta correspondem aos mesmos genótipos da Figura 2.

4.3. Determinação do tamanho dos fragmentos

A determinação precisa do tamanho de fragmentos de DNA amplificados pela técnica de PCR (tais como microssatélites) é obtida pelo seu seqüenciamento. A determinação realizada neste trabalho, por meio de gel

nativo, comparando-se com marcadores contendo fragmentos de tamanhos conhecidos, não foi muito preciso (Figuras 2 e 3).

Aplicando-se as mesmas amostras em dois géis de composições idênticas, e calculando-se o tamanho dos fragmentos com auxílio do programa One Dscan, obtiveram-se tamanhos diferentes para os mesmos fragmentos (Figura 4).

A determinação do tamanho dos fragmentos de DNA em gel, é comumente feita em gel desnaturante. Embora a confecção do gel desnaturante seja rápida, a coloração com solução de prata é bastante trabalhosa e demanda muito tempo, pois envolve muitas etapas até que o gel fique pronto para ser fotografado. Além disso, o tempo de corrida no gel desnaturante quando comparado com o gel nativo, é muito maior.



Figura 4 – Quantificação, realizada pelo programa One Dscan, dos tamanhos dos fragmentos de uma mesma amostra, em dois géis de mesma composição. As canaletas com mesma numeração nos painéis A e B, correspondem ao mesmo alelo, embora as tabelas indiquem tamanhos diferentes de um painel para o outro.

Para calcular o tamanho dos fragmentos, marcadores moleculares contendo fragmentos de tamanhos próximos devem ser utilizados. Marcadores tais como o pUC18 clivado com *MspI* não são adequados, pois os fragmentos observados neste marcador são de tamanhos muito distantes. O marcador de 10 pb é o mais indicado, pois apresenta fragmentos equidistantes a cada 10 pb.

Uma alternativa para identificar os alelos é reunir em um mesmo *bulk*, o DNA de indivíduos representando cada um dos alelos identificados por um *primer*. Desta forma, para um *primer* para o qual foram identificados 5 alelos, faz-se um *bulk* de DNA extraído de 5 indivíduos, cada um possuindo um dos alelos. Este *bulk*, amplificado e analisado simultaneamente com a população sob análise, serve de marcador para a identificação dos alelos presentes na população. Se algum indivíduo da população apresentar um novo alelo, o mesmo passará a fazer parte do *bulk* nas análises posteriores. Quando o número de alelos for muito grande, pode-se dividir os indivíduos em 2 *bulks* e proceder da mesma forma. Trabalhos preliminares indicaram que 5 a 7 alelos constitui um bom número para constituir um *bulk*. Deve ser construído um *bulk* para cada *primer*.

Concluindo, o tamanho dos fragmentos deve ser calculado, utilizando-se géis de poliacrilamida em condições desnaturantes. A análise da população pode ser feita em géis nativos, utilizando os marcadores específicos para cada *primer* formados pelos *bulks* de alelos (conforme mostrado na Figura 3).

4.4. Informatividade dos *primers*

Os valores da diversidade genética associados a cada loco microssatélite variou de 0 a 0,73, com média 0,41 (Quadro 4). Os valores de diversidade genética do loco (que corresponde à estimativa da informatividade do *primer*) encontrados foram bastante satisfatórios, uma vez que 30 dos 32 cultivares utilizados são variedades elites do germoplasma brasileiro, e são, geneticamente, bastante próximas entre si. Resultados similares, para a informatividade de *primers* microssatélites, foram encontrados por outros pesquisadores. NARVEL et al. (2000), estudando a diversidade entre variedades elites de soja e alguns acessos utilizando 74 marcadores SSR, obtiveram resultados semelhantes ao encontrado neste trabalho. Os valores da diversidade genética associados a cada marcador microssatélite encontrados por estes autores variaram de 0 a 0,84, com média 0,56 nos acessos utilizados e de 0 a 0,79, com média 0,50 nas variedades elites. RONGWEN et al. (1995), utilizando 7 pares de *primers* de microssatélites para caracterizar 96 genótipos

de soja bastante divergentes, encontraram valores de diversidade genética do loco extremamente altos que variaram de 0,71 a 0,95. Os valores elevados, encontrados por estes autores, podem ser explicados pela utilização de germoplasma constituído por introduções de plantas (PIs) e por genótipos de *Glycine soja* e *Glycine max*.

Quadro 4 – N^o de alelos e valores de diversidade genética do loco para 22 locos microssatélites em 32 genótipos de soja.

Locos	Todos os genótipos	
	N ^o alelos	Diversidade Genética do loco
Satt002	1	0,00
Satt005	1	0,00
Satt009	1	0,00
Satt179	2	0,12
Satt162	2	0,26
Satt166	2	0,26
Satt136	3	0,28
Satt282	3	0,28
Satt537	3	0,40
Satt184	3	0,44
Satt536	2	0,45
Satt172	3	0,46
Satt115	3	0,48
Satt156	2	0,48
Satt307	3	0,56
Satt082	3	0,60
Sat_085	5	0,63
Satt079	3	0,64
Satt197	4	0,64
Satt070	4	0,65
Satt094	4	0,70
Satt186	4	0,73

Satt indica que o core do microssatélite é (ATT)_n. Sat_ indica que o core do microssatélite é (AT)_n

4.5. Caracterização das variedades

Com os *primers* SATT186, SATT094, SATT070 e SATT197 foi possível identificar 28 dos 32 genótipos. Esses *primers* apresentaram maior valor para diversidade genética do loco. Os pares 22/25 e 18/30 representados pelos genótipos FT-Guaíra e Planalto, IAS4 e FT-8, respectivamente, não foram distinguidos com estes 4 *primers*. Com a inclusão aos quatro *primers* acima de

apenas um dos seguintes *primers* : SATT115, SATT184, SATT307 OU SATT 536, foi possível indentificar individualmente todos os genótipos.

No Quadro 5 pode ser visto a identificação dos 32 genótipos. Neste quadro, cada genótipo apresenta um único padrão. A variedade IAS-4 apresentou 2 alelos no loco Satt070. Isto indica que esta variedade não é uma linha pura ou, alternativamente, que a amostra avaliada continha impureza de outra variedade.

Quadro 5 – Identificação de 32 genótipos de soja com os 5 pares de *primers* de microsatélites.

Genótipos	Satt094	Satt186	Satt070	Satt197	Satt307
1 BR 4	c	a	c	b	b
2 Bela Vista	d	c	a	c	b
3 BRS 137	d	b	b	*	a
4 Uirapuru	d	c	c	c	b
5 DM 247	b	d	d	d	b
6 DM Rainha	b	b	c	c	b
7 Hartwig	a	d	b	b	c
8 KI-S 801	b	c	c	c	a
9 MT/BR49-Pioneira	a	d	d	b	b
10 Emgopa 316	b	b	d	c	c
11 BR/IAC 21	b	d	d	b	c
12 Emgopa 305	b	d	d	c	c
13 Ipagro 21	a	d	b	c	b
14 FT Estrela	b	c	c	d	b
15 FT Abyara	c	b	d	c	b
16 RS-7 Jacuí	c	b	b	b	b
17 UFV-16	c	d	d	d	c
18 IAS 4	c	d	c,d	c	a
19 IAS 5	b	b	c	c	a
20 CD 205	d	c	d	d	c
21 BARC-8	b	d	c	c	*
22 FT Guaira	c	a	d	c	c
23 Pequi	a	d	d	d	b
24 Emgopa 312	d	b	c	c	b
25 Planalto	c	a	d	c	b
26 Embrapa 59	c	a	d	b	b
27 DM Nobre	c	a	d	a	b
28 UFV-6	d	a	b	d	c
29 Pintado	c	c	b	b	b
30 FT 8	c	d	d	c	b
31 IAC 8	c	d	d	b	c
32 BR 1	c	c	b	d	b

* Dados perdidos

Nas colunas, os genótipos seguidos da mesma letra possuem alelos comuns. As letras foram atribuídas a partir do alelo de maior para o de menor tamanho, sendo o alelo de tamanho maior representado pela letra "a", e os demais representados pelas letras "b", "c", "d" e "e", respectivamente.

4.6. Divergência genética entre as variedades avaliadas

Com base nos dados relativos à frequência de alelos comuns nos 32 genótipos estudados, foram determinadas as dissimilaridades genéticas entre eles. A medida de dissimilaridade genética utilizada foi o complemento do coeficiente de coincidência simples. No Quadro 6 estão as medidas de dissimilaridades entre os 32 genótipos. As medidas de dissimilaridade variaram de 9 a 67%, com média 43%. Estes resultados indicam a eficiência dos marcadores microsatélites na medida da dissimilaridade, uma vez que foram utilizados indivíduos geneticamente muito próximos entre si.

4.7. Análise de agrupamento

As correlações cofenéticas entre os dados de dissimilaridade e as matrizes fenéticas obtidas pelos métodos UPGMA, vizinho mais próximo e vizinho mais distante foram de 70,5%, 58,6% e 56,9% respectivamente. O método UPGMA (Figura 5) apresentou alta eficiência em representar as dissimilaridades entre os genótipos avaliados.

Os resultados demonstram que os indivíduos são, na maioria, bastante divergentes entre si, considerando-se que a soja cultivada é uma espécie com base genética bastante estreita. Apenas os pares IAS-5 e KI-S 801 (dissimilaridade = 0,05) e IAC-8 e BR/IAC21 (dissimilaridade = 0,09) apresentaram dissimilaridade inferior a 10%. A maioria dos pares de genótipos ficou com índices de dissimilaridade superiores a 20%.

Para obter grupos mutuamente exclusivos, foi utilizado o método de otimização de Tocher, que identificou 6 grupos (Quadro 7), sendo o grupo 1 formado por 25 indivíduos, os grupos 2 e 3 formados por 2 indivíduos cada um, e os outros 3 grupos formados por um único indivíduo em cada grupo.

Por este resultado, verifica-se que as variedades Emgopa 316, Bela Vista, BARC 8, DM Rainha, FT Estrela, DM 247 e Embrapa 59 são divergentes das demais. Os genótipos restantes, que foram agrupados no primeiro grupo, embora sejam geneticamente mais similares entre si do que os genótipos dos outros grupos, apresentam ainda um grau considerável de divergência dentro

Quadro 6 – Medidas de dissimilaridade entre os 32 genótipos estudados, estimadas a partir do coeficiente de coincidência simples.

Genótipos	BR 4	Bela Vista	BRS 137	Uirapuru	DM 247	Rainha	Hartwig	KI-S 801	Pioneira	EMG 316	BR/IAC21	ENG 305	IPAGRO21	FT Estrela	FT Abyara	RS7Jacuí	UFV-16
BR 4	0,00																
Bela Vista	0,55	0,00															
BRS 137	0,40	0,31	0,00														
Uirapuru	0,32	0,32	0,31	0,00													
DM 247	0,55	0,55	0,60	0,50	0,00												
Rainha	0,41	0,59	0,45	0,36	0,55	0,00											
Hartwig	0,41	0,45	0,31	0,45	0,55	0,55	0,00										
KI-S 801	0,41	0,50	0,40	0,23	0,41	0,41	0,50	0,00									
Pioneira	0,41	0,41	0,40	0,36	0,36	0,45	0,18	0,45	0,00								
EMG 316	0,55	0,59	0,50	0,45	0,41	0,36	0,50	0,36	0,45	0,00							
BR/IAC21	0,41	0,45	0,36	0,36	0,41	0,41	0,23	0,36	0,18	0,27	0,00						
ENG 305	0,45	0,41	0,45	0,32	0,36	0,45	0,36	0,27	0,32	0,27	0,14	0,00					
IPAGRO 21	0,27	0,45	0,31	0,27	0,50	0,36	0,27	0,36	0,32	0,50	0,36	0,32	0,00				
FT Estrela	0,45	0,59	0,55	0,41	0,45	0,23	0,50	0,45	0,45	0,59	0,45	0,55	0,45	0,00			
FT Abyara	0,45	0,32	0,45	0,36	0,45	0,45	0,45	0,50	0,27	0,50	0,41	0,36	0,45	0,55	0,00		
RS-7Jacuí	0,34	0,45	0,33	0,34	0,61	0,48	0,30	0,52	0,39	0,50	0,45	0,50	0,34	0,57	0,39	0,00	
UFV-16	0,45	0,32	0,40	0,50	0,36	0,59	0,27	0,50	0,27	0,45	0,23	0,23	0,41	0,55	0,32	0,45	0,00
IAS 4	0,41	0,45	0,40	0,23	0,36	0,50	0,45	0,23	0,32	0,41	0,36	0,32	0,36	0,55	0,32	0,39	0,36
IAS 5	0,41	0,55	0,36	0,27	0,41	0,36	0,50	0,05	0,45	0,32	0,36	0,27	0,36	0,50	0,45	0,48	0,50
CD 205	0,39	0,43	0,33	0,30	0,43	0,57	0,36	0,39	0,41	0,43	0,30	0,30	0,34	0,48	0,55	0,43	0,30
BARC-8	0,62	0,52	0,58	0,57	0,52	0,57	0,48	0,43	0,57	0,52	0,57	0,43	0,52	0,57	0,62	0,50	0,52
FT Guáira	0,50	0,41	0,50	0,41	0,55	0,50	0,45	0,55	0,36	0,45	0,32	0,36	0,45	0,45	0,27	0,50	0,32
Pequi	0,52	0,39	0,38	0,43	0,36	0,52	0,34	0,48	0,20	0,52	0,30	0,30	0,39	0,48	0,34	0,55	0,30
EMG 312	0,36	0,50	0,36	0,18	0,45	0,27	0,50	0,23	0,41	0,27	0,41	0,36	0,32	0,50	0,45	0,34	0,55
Planalto	0,39	0,48	0,55	0,34	0,43	0,48	0,55	0,39	0,36	0,43	0,43	0,39	0,34	0,52	0,32	0,43	0,43
EMBR 59	0,50	0,59	0,64	0,50	0,32	0,64	0,55	0,59	0,36	0,50	0,50	0,59	0,64	0,59	0,41	0,48	0,55
DM NOBRE	0,39	0,45	0,33	0,48	0,43	0,57	0,43	0,48	0,39	0,50	0,41	0,50	0,43	0,57	0,48	0,45	0,41
UFV-6	0,41	0,50	0,31	0,41	0,59	0,59	0,27	0,50	0,41	0,50	0,36	0,41	0,36	0,59	0,55	0,43	0,36
Pintado	0,32	0,50	0,40	0,36	0,55	0,59	0,32	0,45	0,41	0,64	0,45	0,50	0,32	0,41	0,55	0,25	0,45
FT 8	0,36	0,55	0,48	0,36	0,55	0,45	0,50	0,50	0,45	0,59	0,45	0,41	0,23	0,45	0,45	0,43	0,45
IAC 8	0,41	0,41	0,40	0,41	0,41	0,50	0,23	0,45	0,14	0,36	0,09	0,23	0,41	0,50	0,32	0,41	0,14
BR 1	0,50	0,41	0,48	0,36	0,36	0,59	0,50	0,36	0,45	0,45	0,41	0,36	0,45	0,55	0,41	0,45	0,36

Continua...

Quadro 6 – Continuação

Genótipos	IAS 4	IAS 5	CD205	BARC-8	FT Guaíra	Pequi	EMG 312	Planalto	EMBR 59	DM NOBRE	UFV-6	Pintado	FT 8	IAC 8	BR 1
IAS 4	0,00														
IAS 5	0,23	0,00													
CD 205	0,43	0,43	0,00												
BARC-8	0,48	0,43	0,57	0,00											
FT Guaíra	0,32	0,55	0,43	0,67	0,00										
Pequi	0,48	0,48	0,41	0,45	0,48	0,00									
EMG 312	0,27	0,18	0,39	0,43	0,55	0,48	0,00								
Planalto	0,20	0,39	0,45	0,62	0,25	0,52	0,34	0,00							
EMBR 59	0,36	0,59	0,55	0,62	0,41	0,55	0,55	0,36	0,00						
DM NOBRE	0,43	0,48	0,34	0,50	0,45	0,34	0,43	0,39	0,39	0,00					
UFV-6	0,55	0,50	0,27	0,62	0,50	0,43	0,45	0,55	0,59	0,43	0,00				
Pintado	0,45	0,50	0,34	0,52	0,50	0,52	0,45	0,48	0,55	0,43	0,36	0,00			
FT 8	0,36	0,50	0,39	0,48	0,36	0,48	0,45	0,30	0,50	0,43	0,55	0,36	0,00		
IAC 8	0,27	0,45	0,34	0,62	0,23	0,34	0,45	0,34	0,41	0,41	0,41	0,41	0,45	0,00	
BR 1	0,41	0,41	0,43	0,62	0,50	0,39	0,41	0,34	0,45	0,45	0,45	0,45	0,50	0,41	0,00

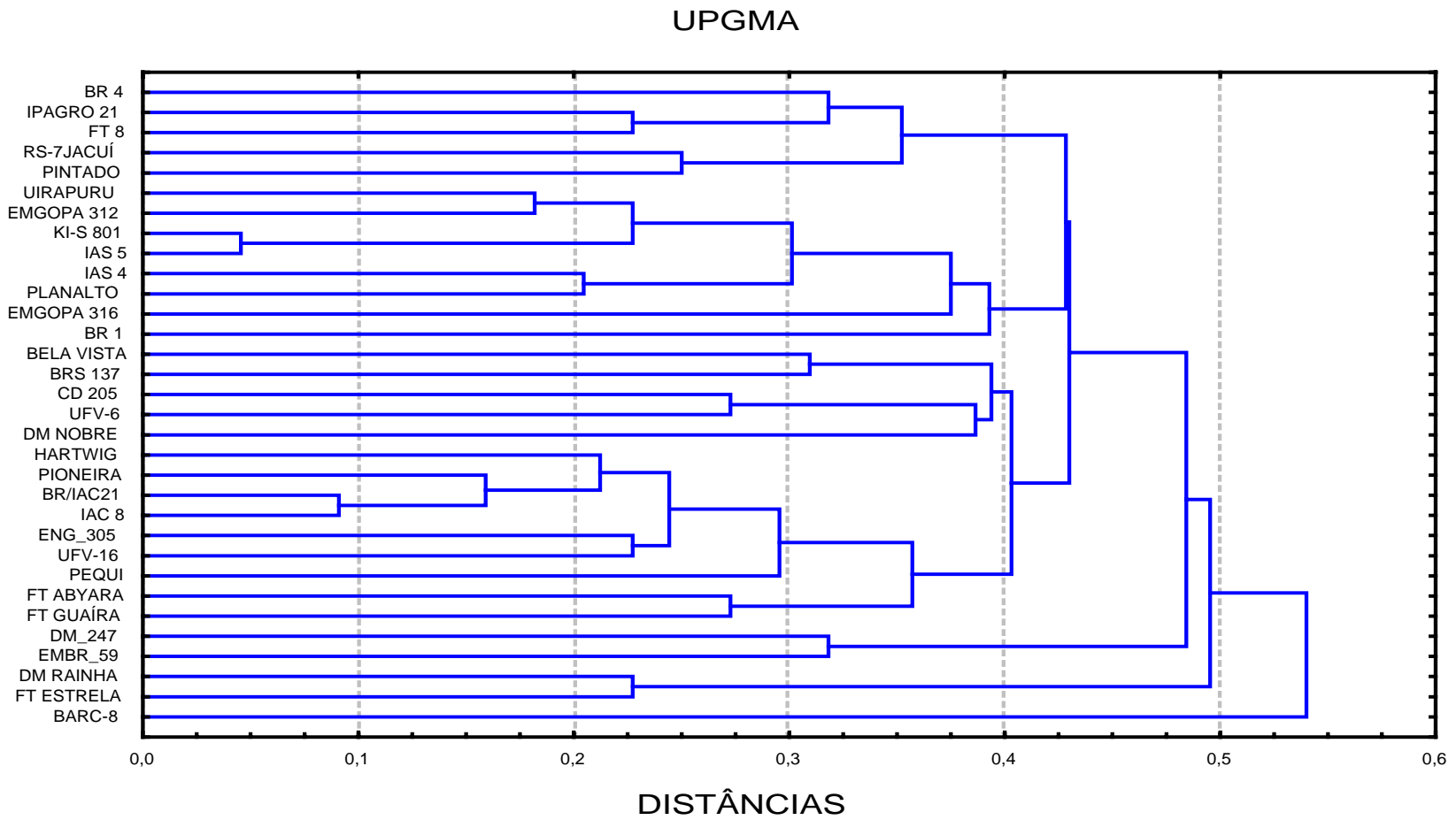


Figura 5 – Dendrograma construído com base nas medidas de dissimilaridade do Quadro 6, com uso do programa Statistics. UPGMA (Unweighted pair-group using a mean average, ou método das médias das distâncias)

do grupo. Por este motivo, este grupo foi dividido, produzindo oito subgrupos (Quadro 7).

Quadro 7 – Agrupamento pelo método de Tocher dos 32 genótipos estudados

Grupo	Subgrupo	Genótipo
1	1.1	KI-S 801; IAS 5; Emgopa 312; Uirapuru; IAS 4; Emgopa 305; BR/IAC 21; IAC 8; Pioneira
	1.2	Ipagro 21; FT 8; BR 4; Pintado; RS-7 Jacuí
	1.3	FT Guaíra; Planalto; FT Abyara
	1.4	Hartwig; UFV-16; Pequi
	1.5	CD 205; BRS 137
	1.6	UFV-6
	1.7	DM Nobre
	1.8	BR 1
2	-	DM Rainha; FT Estrela
3	-	DM 247; Embrapa 59
4	-	Emgopa 316
5	-	Bela Vista
6	-	BARC 8

Comparando os dados da análise de agrupamento com a origem, ciclo e local de cultivo (Quadro 2), pode-se observar que dentro de cada grupo, ou subgrupo, existem genótipos de várias origens, com ciclo variável, e região de recomendação diversa. Esta independência entre as características dos genótipos e o agrupamento genético revela que os diversos programas de melhoramento utilizam genótipos com a mesma base genética, pois possuem variedades representadas nos vários grupos e subgrupos. Cada grupo (ou subgrupo) possui genótipos de diversas empresas de melhoramento, indicando que as empresas utilizam germoplasmas semelhantes entre si.

Além de permitir a identificação precisa dos genótipos, a metodologia utilizada neste trabalho permitiu a comparação da distância genética entre os cultivares de soja, o que pode auxiliar na escolha de progenitores em programas de melhoramento. Cruzamentos entre indivíduos geneticamente mais divergentes produzem populações segregantes com maiores variabilidades, permitindo novas combinações genéticas, e, seleção de indivíduos superiores em programas baseados no método genealógico. Cruzamentos entre indivíduos geneticamente mais próximos permitem a recuperação mais rápida do genoma recorrente em programas de melhoramento baseados em retrocruzamentos.

5. RESUMO E CONCLUSÕES

Com o crescente número de pedidos de proteção de cultivares junto ao SNPC, o interesse pela caracterização de cultivares tem aumentado significativamente no Brasil. Marcadores microssatélites, por sua natureza co-dominante e multi-alélica, além de serem extremamente conservados dentro da espécie podem ser utilizados na caracterização genética, tanto para fins de proteção como também em análise de pureza genética em lotes de sementes, permitindo diferenciar claramente genótipos de soja bastante próximos.

Este trabalho teve como objetivos comparar a eficiência dos diversos sistemas de eletroforese disponíveis na discriminação dos alelos de microssatélites, estimar a informatividade dos locos de microssatélites comumente utilizados em estudos de diversidade genética de soja no BIOAGRO/UFV, selecionar um conjunto de *primers* de microssatélites capazes de diferenciar um grupo de genótipos representativo das variedades de soja cultivadas no Brasil, e por fim estabelecer um protocolo, de fácil aplicação em análises de rotina, para a identificação molecular de genótipos de soja.

Entre os diferentes géis utilizados, foi constatado que o gel de poliacrilamida apresentou melhor resolução na separação dos fragmentos, principalmente, com o uso do polímero SpreadexTM. A adequação do tamanho do fragmento com a concentração do gel e o tempo de corrida, tem grande importância no êxito dos resultados.

Este estudo também demonstrou que na determinação do número de alelos existentes num loco, a análise de fragmentos baseada unicamente nos resultados obtidos por programas computacionais pode levar a conclusões errôneas.

Foram utilizados 22 primers de microssatélites de soja, dos quais 19 foram polimórficos e 3 monomórficos, resultando em 61 alelos nos 32 genótipos estudados. A diversidade genética nos 22 locos estudados variou de 0 a 0,73 com média 0,41. O número de alelos variou de 1 a 5 com média 2,77.

Com a combinação dos primers SATT186, SATT094, SATT070 e SATT197 foi possível identificar 28 dos 32 genótipos. Mantendo os 4 primers acima, quatro diferentes combinações de 5 primers identificaram todos os genótipos, bastando incluir um dos seguintes primers: SATT115, SATT184, SATT309 ou SATT536.

As medidas de dissimilaridade, estimada pelo coeficiente de coincidência simples entre os genótipos avaliados, variaram de 9 a 67%, com média 43%. Embora haja pouca variabilidade genética entre os cultivares utilizados, os resultados indicam a eficiência dos marcadores microssatélites na medida da dissimilaridade.

O dendrograma obtido pelo método UPGMA representou bem o inter-relacionamento entre os genótipos. Apenas dois pares apresentaram dissimilaridade inferior a 10%. A maioria dos pares de genótipos ficou com índices de dissimilaridade superiores a 20%. Os resultados demonstram que os indivíduos utilizados neste trabalho ainda são, na maioria, bastante divergentes entre si, quanto aos locos microssatélites analisados, embora a soja cultivada seja uma espécie com base genética estreita.

Os resultados obtidos mostraram-se confiáveis, rentáveis em termos de análise, de fácil aplicação em análises de rotina, e que poderão ser utilizados como ferramenta auxiliar nos processos de proteção de cultivares, bem como

nos casos de análise de pureza genética de sementes em que a análise visual permitir dúvidas.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ABDELNOOR, R.V.; BARROS, E.G; MOREIRA, M.A. Determination of genetic diversity within Brazilian soybean germplasm using random amplified polymorphic DNA techniques and comparative analysis with pedigree data. **Rev. Brasil. Genet.**, v.18, n.2, p.265-273, 1995.
- AKKAYA, M.S.; BHAGWAT,A.A.; CREGAN, P.B. Length polymorphisms of simple sequence repeat DNA in soybean. **Genetics**, v.132, n.3, p.1131-1139, 1992.
- ANDERSON, J.A; CHURCHILL, G.A.; AUTRIQUE, J.E.; TANKSLEY, S.D.; SORRELS, M.E. Optimizing parental selection for genetic linkage maps. **Genome**, v.36, p.181-186, 1993.
- BOTSTEIN, D.; WHITE, R.L.; SKOLNICK, M.; DAVIS, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **Am. J. Hum. Genet.** v.32, n.1, p.314-331, 1980.

- CREGAN, P.B.; BHAGWAT, A.A.; AKKAYA, M.S.; ROGWEN, J. Microsatellite fingerprint and mapping of soybean. **Meth. Mol. Cell. Biol.** , v.5, p.49-61, 1994.
- 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.; SPECHT, J.E. An integrated genetic linkage map of the soybean genome. **Crop Sci.** v.39., p.1464-1490, 1999.
- CRUZ, C.D. **Aplicação de algumas técnicas multivariadas no melhoramento de plantas.** Piracicaba: ESALQ, 1990. 188p. (Tese DS).
- CRUZ, C.D; REGAZZI, A.J.; **Modelos biométricos aplicados ao melhoramento genético.** Viçosa, MG: UFV, 1994. 390p.
- DELANNAY, X.; ROGERS, D.M.; PALMER, R.G. Relative genetic contributions among ancestral lines to North American soybean cultivars. **Crop Science**, v.23, n.2, p.944-949, 1983.
- DIWAN, N.; CREGAN, P.B. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. **Theor. Appl. Genet.**, v.95, p.723-733, 1997.
- DOYLE, J.J. & DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**.v.12, p.13-15, 1990.
- EMBRAPA. Centro Nacional de Pesquisa de Soja (Londrina, PR). **Recomendações técnicas para a cultura da soja na região central do Brasil 2000/01.** Londrina, 2000. 245p.
- FERREIRA, M.E. & GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análises genéticas.** 2^{ed}. Brasília EMBRAPA-CENARGEN.220p, 1996.
- HIROMOTO, D.M.; VELLO, N.A. The genetic base of Brazilian soybean (*Glycine max* (L.) Merrill) cultivars. **Rev. Brasil. Genet.**, v.9, n.2, p.295-306, 1986.

- KEIM, P.; SHOEMAKER, R.C; PALMER, R.G. Restriction fragment length polymorphism diversity in soybean. **Theor. Appl. Genet.** v.77, p.786-792, 1989.
- KEIM,P.; BEAVIS, W.; SCHUPP, J.; FREESTONE, R. Evaluation of soybean RFLP markers diversity in adapted germplasm. **Theor. Appl. Genet.** v.85, p.205-212, 1992.
- KRAIC, J.; ZÁKOVÁ, M.; GREGOVÁ, E. Comparison of differentiation capability of RAPD and SSR markers in commercial barley (*Hordeum vulgare* L.) cultivars. **Cereal Research Communications**, v.26, n.4, p.375-382, 1998.
- LANZA, M.A.; SCHUSTER, I.; GUIMARÃES, C.T. Aplicação de marcadores moleculares no melhoramento genético. **Informe agropecuário**, v.21, n.204, p. 97-108, 2000.
- LITT, M. & LUTY, J.A. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. **Am. J. Hum. Genet.**, v.44, p. 937-1401, 1989.
- MACHADO, M.A. **Integração de mapas, mapeamento de QTLs e análise da diversidade genética em soja utilizando marcadores AFLP.** Viçosa, MG, 1999. 90p. (Tese MS).
- MAUGHAN, P.J., SAGHAI MAROOF,M.A., BUSS G. R. e HUESTIS, G.M. Amplified fragment length polymorphism (AFLP) in soybean: species diversity inheritance, and near isogenic line analysis. **Theor. Appl. Genet.**, v.93, n.2, p.392-401, 1996.
- MILACH, S.C. Uso de marcadores moleculares na caracterização de cultivares. In: **Biowork I**, 1998, A.BORÉM, M.P.DEL GIÚDICE, N.S. SAKIYAMA, M.A. MOREIRA, R.S. PORTUGAL. Viçosa, p. 43-58i, 1998.
- NARVEL, J.M.; FEHR, W.R.; CHU, W. C.; GRANT, D. e SHOEMAKER, R.C. Simple sequence repeat diversity among soybean plant introductions and elite genotypes. **Crop Science.** v.40, n.5, p.1452-1458, 2000.
- NETO, M.O.V., BORÉM, A., PORTUGAL, R.S. Lei de Proteção de Cultivares. **Melhoramento de espécies cultivadas.** 1 ed. Viçosa:UFV. p.669-798, 1999.

- POWELL, W., MORGANTE, M., DOYLE, J.J, McNICOL, J.W., TINGEY, S.V. e RAFALSKI, A.J. Genepool variation in genus *Glycine* sugenus *soja* revealed by polymorphic nuclear and cloroplast microsattellites. **Genetic**, v.144, n.2, p.793-803, 1996.
- RONGWEN, J.; AKKAYA, M.S.; BHAGWAT, A.A.; LAVI, U.; CREGAN, P.B. The use of microsatellite DNA markers for soybean genotype identification. **Theor. Appl.Genet**, v.90, n.1, p.43-48, 1995.
- RUSSELL, J.; FULLER, J.; YOUNG, G.; THOMAS, B.; TARAMINO, G.; MACAULAY, M.; WAUGH, R.; POWELL, W. Discriminating between barley genotypes using microsatellite markers. **Genome**, v. 40, p. 442-450, 1997.
- SAIKI, R.K.; SCHARF, S.; FALOONA, F.; MULLIS, B.K.; HORN, G.T.; ERLICH, H.A.; ARNHEIM, N. Enzymatic amplification of β -globin genomic sequence and restriction site analysis for diagnosis of sickle cell anemia. **Science**, v. 230, n.8, p. 1350-1354, 1985.
- SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. **Molecular cloning a laboratory manual**. 2. Ed. New York, Cold Spring Harbor Laboratory Prees, v.3, 1989.
- SMITH, S., HELENTJARIS, T. DNA fingerprinting and plant variety protection. In: **Genome Mapping in plants.**, 1996, A.H. PATERSON, Academic Press. p.95-110, 1996.
- SNPC. Ministério da Agricultura e do Abastecimento. In: **Informativo técnico**, Brasília,ano II, n.3, 6p., 1999.
- THOMAS, M.R., CAIN, P., SCOTT, N.S. DNA typing of grapevines: A universal methodology and database for describing cultivars and evaluating genetic relatedness. **Plant molecular Biology**, v.25, p.939-949,1994.
- TINGEY, S.V. & DEL TUFO, J.P. Genetic analysis with Random Amplified Polymorphic DNA markers. **Plant. Physiol.**, v.101, p.349-352, 1993.
- TINGEY, S.V.; RAFALSKI, J.A.; WILLIAMS, G.K. Soybean genome analysis: DNA polymorphisms are identified by oligonucleotide primers of arbitrary sequence. In: **Plant molecular biology 2**, 1990, R.G. HERMANN, R.G, B.A. LARKINS, Eds. New York. p.263-268, 1991.

UPOV. Essentially derived varieties. Sixth meeting with International Organizations, Geneva, Oct. 30, 1992. Union Pour le Protections d'Obtentions Vegetales Mimeo IOM/6/2, Geneva, Switzerland.

VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; LEE, T.; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M. e ZABEAU, M. AFLP: A new technique for DNA fingerprinting. **Nucleic. Acids Res.**, v.23, n.5, p.4407-4414, 1995.

WANG, Z.; WEBER, J.L.; ZHONG, G.; TANKSLEY, S.D. Survey of plant short tandem DNA repeats. **Theor. Appl. Genet.**, v.88, p.1-6, 1994.

WEIR, B.S. Genetics data analysis methods for discrete genetic data. **Sinauer Assoc.Inc. Sunderland**, Massachusetts – USA, 1990.

WELSH, J., McCLELLAND, M. Fingerprinting genomes using PCR with arbitrary primers. **Nucleic. Acids Res.**, v.18, n.7, p.7213-7218,1990.

WILLIAMS, J.G.K.; KUBELIK, A.R.; LIVAK,K.J.; RAFALSAKI, J.A.; TINGEY, S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic. Acids Res.**, v.18, n.6, p. 6531-6535, 1990.