

UNIVERSIDADE ESTADUAL DE MARINGÁ  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E MELHORAMENTO

**MARCELO BERWANGER DE OLIVEIRA**

**Análise de desequilíbrio de ligação e diversidade genética em soja  
utilizando marcadores moleculares**

MARINGÁ  
PARANÁ – BRASIL  
MARÇO – 2014

**MARCELO BERWANGER DE OLIVEIRA**

**Análise de desequilíbrio de ligação e diversidade genética em soja  
utilizando marcadores moleculares**

Tese apresentada à Universidade Estadual de Maringá, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do Título de Doutor.

Orientador: Prof. Dr. Ivan Schuster.

MARINGÁ  
PARANÁ – BRASIL  
MARÇO – 2014

Aos meus pais, Paulo Roberto de Oliveira (*in memoriam*) e Carmem Bárbara  
Berwanger dos Santos.

Aos meus irmãos, Anderson A. dos Santos e Jefferson A. dos Santos.

A todos os meus amigos.

## AGRADECIMENTOS

Agradeço a Deus, acima de todas as coisas, pela minha vida, por me proporcionar tudo que tenho e estar sempre comigo nas conquistas e nas derrotas.

À Universidade Estadual de Maringá (UEM) e ao Programa de Pós-Graduação em Genética e Melhoramento (PGM), pela oportunidade de realização do Curso de Doutorado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela concessão da Bolsa de Doutorado.

À Cooperativa Central de Pesquisa Agrícola (Coodetec), por oferecer todas as condições necessárias para a realização deste trabalho.

Ao professor doutor Ivan Schuster, pela competente orientação, por sua confiança, pelo apoio, incentivo e dedicação neste trabalho, pelos ensinamentos, não apenas de nível técnico, mas também para a vida.

Aos meus Coorientadores, Ronald José Barth Pinto e Carlos Alberto Scapim, pelos ensinamentos e contribuições para este trabalho.

Aos professores do Programa de Pós-Graduação em Genética e Melhoramento (PGM), pelos ensinamentos que proporcionaram a base para a realização do trabalho.

Aos membros das bancas de qualificação e defesa pelas críticas e sugestões que elevam a qualidade desta Tese.

Aos meus pais, que me proporcionaram as condições necessárias para a evolução do meu conhecimento e pelo incentivo que me deram em todas as etapas da minha vida.

Aos meus irmãos, pelos momentos de conforto e pelo apoio de sempre.

Aos colegas e amigos que fiz na Coodetec, que muito me ajudaram ao garantir o excelente ambiente de trabalho que dividimos.

Aos meus amigos, que sempre estiveram presentes na minha vida, me ajudando e apoiando nos momentos difíceis.

Aos meus colegas e amigos do PGM, que me acompanharam nessa jornada, participando direta ou indiretamente deste trabalho.

## **BIOGRAFIA**

MARCELO BERWANGER DE OLIVEIRA, filho de Paulo Roberto de Oliveira e de Carmem Bárbara Berwanger dos Santos, nasceu em 09 de julho de 1985, na cidade de Porto Alegre, estado do Rio Grande do Sul.

Em dezembro de 1999, concluiu o Ensino Fundamental, no Colégio Marista de Cascavel, na cidade de Cascavel, estado do Paraná.

Concluiu o Ensino Médio em dezembro de 2002, no Colégio Marista de Cascavel, também em Cascavel, estado do Paraná.

Ingressou no Curso de Graduação em Ciências Biológicas, em fevereiro de 2003, na Universidade Paranaense(UNIPAR), na cidade de Cascavel, estado do Paraná, obtendo o título de biólogo em fevereiro de 2007.

Em março de 2007, ingressou no curso de Mestrado de Biotecnologia Aplicada à Agricultura, na Universidade Paranaense(UNIPAR), na cidade de Umuarama, estado do Paraná, obtendo o título de Mestre em fevereiro de 2009.

Participou, em março de 2009, do projeto GenoSoja – Consórcio Nacional para Estudos do Genoma da Soja – na Cooperativa Central de Pesquisa Agrícola – Coodetec, como bolsista DTI-3 do CNPq.

Em março de 2010, ingressou no Curso de Doutorado do Programa de Pós-Graduação em Genética e Melhoramento (PGM), da Universidade Estadual de Maringá (UEM), em Maringá, estado do Paraná, Brasil.

## SUMÁRIO

<b>1. INTRODUÇÃO GERAL.....</b>	<b>1</b>
<b>2. REVISÃO DE LITERATURA .....</b>	<b>3</b>
2.1. Importância econômica da soja e sua produção no Brasil.....	3
2.2. Fatores que afetam a produtividade da soja .....	3
2.3. Diversidade genética na cultura da soja brasileira .....	4
2.4. Seleção Assistida por Marcadores (SAM).....	5
2.4.1. Seleção assistida por marcadores no Brasil .....	7
2.5. Tipos de marcadores moleculares .....	8
2.5.1. Microssatélites (SSR) .....	9
2.5.2. Polimorfismo de um único nucleotídeo (SNPs).....	11
2.6. Plataformas de genotipagem.....	12
2.6.1. Plataforma Illumina Bead Array .....	12
2.6.2. OpenArray®.....	14
2.6.3. KASP™ (“Kompetitive” Allele Specific PCR).....	15
2.7. Mapeamento genético em plantas.....	15
2.7.1. Mapeamento por recombinação .....	16
2.7.2. Mapeamento por associação.....	18
<b>3. REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>20</b>
<b>CAPÍTULO 1 - Análise de diversidade genética em soja utilizando marcadores</b>	
<b>SNPs e microssatélites .....</b>	<b>28</b>
<b>1. INTRODUÇÃO .....</b>	<b>28</b>
<b>2. MATERIAL E MÉTODOS .....</b>	<b>30</b>
2.1. Material genético.....	30
2.2. Genotipagem de SNPs .....	31
2.3. Genotipagem dos microssatélites .....	32
2.4. Estimativa da diversidade genética.....	34
<b>3. RESULTADOS E DISCUSSÃO .....</b>	<b>35</b>
<b>4. CONCLUSÕES .....</b>	<b>43</b>
<b>5. REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>44</b>
<b>CAPÍTULO 2 - Desequilíbrio de ligação em soja tropical utilizando análise</b>	
<b>ampla do genoma com marcadores SNPs .....</b>	<b>48</b>

<b>1. INTRODUÇÃO .....</b>	<b>48</b>
<b>2. MATERIAL E MÉTODOS .....</b>	<b>50</b>
2.1. Material genético.....	50
2.2. Genotipagem dos marcadores SNPs.....	50
2.3. Análise dos blocos de desequilíbrio de ligação.....	50
<b>3. RESULTADOS E DISCUSSÃO .....</b>	<b>53</b>
<b>4. CONCLUSÕES .....</b>	<b>58</b>
<b>5. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>59</b>

## RESUMO

OLIVEIRA, Marcelo Berwanger de, D. Sc. Universidade Estadual de Maringá, fevereiro de 2014. **Análise de desequilíbrio de ligação e diversidade genética em soja utilizando marcadores moleculares.** Orientador: Ivan Schuster. Conselheiros: Carlos Alberto Scapim e Ronald José Barth Pinto.

O sucesso do melhoramento genético de plantas depende principalmente da escolha de progenitores divergentes e da seleção de genótipos com características de interesse agrônomo ao longo das gerações. A seleção assistida por marcadores (SAM) é uma técnica biotecnológica que visa a auxiliar o melhorista na seleção de genótipos com base no DNA das plantas. A eficiência da SAM é diretamente influenciada pela precisão do mapeamento genético. Entre as metodologias de mapeamento, o mapeamento por associação tem se destacado pela sua elevada precisão. Esta metodologia é baseada no desequilíbrio de ligação entre o marcador e o gene que controla a característica. O estudo dos blocos de desequilíbrio de ligação em uma espécie pode fornecer informações importantes para a realização do mapeamento por associação. Este trabalho teve como objetivo analisar a diversidade genética e o desequilíbrio de ligação em 153 cultivares de soja brasileiras, utilizando 4938 marcadores SNPs e 20 marcadores microssatélites. A porcentagem de heterozigosidade das cultivares atingiu uma média de 8%, variando de 3 a 41%. As distâncias genéticas variaram de 0,0052 a 0,4216, com média de 0,32, indicando a baixa diversidade genética presente no germoplasma brasileiro. A análise de agrupamento pelo método UPGMA formou 14 grupos, sendo que cultivares adaptadas para a mesma região de cultivo pertenciam ao mesmo grupo. Além disso, foi observado em algumas cultivares de diferentes programas de melhoramento mais de 95% de similaridade. Na análise de desequilíbrio de ligação, foram obtidos 567 blocos em desequilíbrio e a média de blocos por cromossomo foi de 28,35, variando de 20 a 41. A quantidade de SNPs por bloco variou de 2 a 15, obtendo-se uma média de 2,69. Aproximadamente, 65% dos blocos identificados continham apenas dois marcadores e menos de 2% continham sete ou mais marcadores. A média do tamanho dos blocos foi de 410,68 kb, variando de 1,173 kb a 13.692,9 kb. Mais de 70% dos blocos em desequilíbrio tiveram menos de 200 kb. Cerca de 20% do genoma das cultivares avaliadas apresentaram blocos em desequilíbrio de ligação. Este trabalho demonstra a grande similaridade genética em



cultivares de soja brasileira e a necessidade da introdução de cultivares oriundas de outros germoplasmas para manutenção da diversidade do germoplasma brasileiro. A análise do desequilíbrio de ligação em soja tropical pode fornecer informações importantes para o mapeamento por associação de características de interesse agronômicas, aumentando a eficiência e a utilização da SAM pelos programas brasileiros de melhoramento.

Palavras-chave: Mapeamento por associação, marcadores SNPs, genotipagem de alto rendimento.

## ABSTRACT

OLIVEIRA, Marcelo Berwanger de, D. Sc. Universidade Estadual de Maringá, February, 2014. **Linkage disequilibrium analysis and genetic diversity in soybean using molecular markers.** Advisor: Ivan Schuster. Committee Members: Carlos Alberto Scapim and Ronald José Barth Pinto.

The success of plant breeding depends mainly on the choice of parents genetically divergent and selection of genotypes with desirable agronomic characteristics over generations. The marker-assisted selection (MAS) is a biotechnology technique that aims to help the breeder in the selection of genotypes based on plant DNA. The efficiency of MAS is directly influenced by the accuracy of genetic mapping. Among the methodologies for mapping, association map stands out for its high accuracy. This methodology is based on the linkage disequilibrium between the marker and the gene that controls the trait. The study of blocks of linkage disequilibrium in culture can provide important information to perform an association mapping. This study aimed to analyze the genetic diversity and linkage disequilibrium in 153 Brazilian soybean cultivars using a set of 4938 SNP markers and 20 microsatellite markers. The percentage of heterozygosity of the cultivars had an average of 8 %, ranging from 3 to 41 %. Genetic distances ranged from 0.0052 to 0.4216, with a mean of 0.32, indicating low genetic diversity present in the Brazilian germplasm. Cluster analysis by UPGMA formed 14 groups, which cultivars adapted to the same growing region were in the same group. It was also observed that some cultivars from different breeding programs had more than 95 % of similarity. . In the analysis of linkage disequilibrium were observed 567 blocks of disequilibrium, where the average blocks per chromosome was 28.35, ranging from 20 to 41. The number of SNPs per block ranged from 2 to 15, with an average of 2.69. Approximately 65% of identified blocks contained only two markers and less than 2 % containing seven or more markers. The average size of the blocks was 410.68 kb, ranging from 1.173 kb to 13692.9 kb. Over 70 % of the disequilibrium blocks had less than 200 kb. About of 20 % of the cultivars genome had blocks in linkage disequilibrium. This work demonstrates the great genetic similarity in Brazilian soybean cultivars and the need for the introduction of cultivars from other germplasm for maintaining the diversity of Brazilian germplasm. The analysis of linkage disequilibrium in tropical soybean may

provide important information for association mapping of important agronomic traits, increasing the efficiency and utilization of MAS by the Brazilian breeding programs.

**Keywords:** Association mapping, SNPs markers, High-throughput genotyping.

## 1. INTRODUÇÃO GERAL

A soja (*Glycine max* (L.) Merrill) é uma cultura com grande importância econômica no Brasil. A relevância brasileira no mercado mundial como um dos principais produtores desta cultura deve-se principalmente à expansão da soja para todo o território nacional, incluindo áreas de cerrado. Segundo Priolli et al. (2004), isso só foi possível graças ao melhoramento genético da soja, que permitiu uma grande adaptabilidade para as diferentes regiões do país, além de cultivares mais resistentes aos patógenos e a estresses ambientais.

Embora os programas de melhoramento tenham obtido êxito na obtenção de cultivares adaptadas, inúmeros trabalhos têm relatado que a diversidade genética entre as cultivares de soja no Brasil é baixa, devido principalmente à sua estreita base genética. Segundo Hironoto e Vello (1986) e Gizlice et al. (1993), as cultivares de soja utilizadas no Brasil nas décadas de 1980 e 1990 derivavam de um pequeno número de cultivares ancestrais. Isso causa uma limitação na variabilidade genética, podendo ocasionar maior susceptibilidade ao ataque de pragas, aos patógenos e a estresses ambientais. A partir da metade da década de 2000, foram introduzidas no Brasil diversas cultivares provenientes do germoplasma norte-americano, vindas da Argentina, e que apresentaram boa adaptação às condições de cultivo no país. O real impacto das cultivares modernas de soja em uso no Brasil ainda não foi adequadamente estudado, ainda que vários trabalhos tenham sido publicados sobre a diversidade genética do germoplasma brasileiro (Priolli et al., 2002; Priolli et al., 2004; Bonato et al., 2006; Priolli et al., 2010). Além disso, os trabalhos que abordaram a diversidade genética de cultivares brasileiras de soja a nível molecular foram baseados em uma pequena quantidade de marcadores, quando comparados ao número de marcadores disponíveis atualmente, avaliando poucos locos.

Com o sequenciamento do genoma da soja, tornaram-se disponíveis informações sobre a localização de milhares de locos de SNPs (*single nucleotide polymorphism*). Estes locos contendo SNPs podem ser genotipados em larga escala com as modernas plataformas de genotipagem atualmente disponíveis. Esta genotipagem em larga escala pode fornecer uma quantidade muito maior de informação sobre a estrutura do genoma da soja e a variabilidade existente em uma determinada coleção de germoplasma. A partir dos dados de sequenciamento do

genoma da soja, um *chip* de marcadores SNPs de 6K (seis mil marcadores) foi desenvolvido para a soja, para ser utilizada na plataforma *iScan*<sup>TM</sup> da Illumina<sup>TM</sup>.

O objetivo deste trabalho foi avaliar a diversidade genética e a estrutura do genoma da soja brasileira, utilizando um grande número de marcadores SNPs. Foram genotipadas 169 variedades de soja, utilizando o *chip* 6K e a plataforma *iScan*<sup>TM</sup> (Illumina<sup>TM</sup>), e 20 marcadores microssatélites. A partir dos dados de genotipagem foi realizada uma análise de diversidade genética e de agrupamento entre as variedades. Também foi realizada uma análise da estrutura do genoma da soja, por meio da identificação dos blocos de desequilíbrio de ligação entre os marcadores SNPs. Este é o primeiro trabalho de avaliação da diversidade genética em soja tropical, utilizando genotipagem em larga escala de marcadores SNPs, abrangendo uma maior quantidade de regiões do genoma e explorando melhor a variabilidade. Também é o primeiro trabalho de análise estrutural do genoma da soja tropical, por meio da identificação de blocos de desequilíbrio de ligação, utilizando marcadores moleculares. Os resultados poderão contribuir para futuros trabalhos de mapeamento por associação que podem ser utilizados para aumentar a precisão da seleção assistida por marcadores moleculares em variedades tropicais. Além disso, os resultados podem ser utilizados de imediato pelos programas de melhoramento genético, uma vez que apresentam a relação existente entre as cultivares de soja em uso no Brasil e também em relação a algumas cultivares historicamente importantes e usadas com frequência em programas de cruzamentos.

## 2. REVISÃO DE LITERATURA

### 2.1. Importância econômica da soja e sua produção no Brasil

A soja (*Glycine max* (L.) Merril), planta leguminosa autógama originária da China, é atualmente a principal commodity agrícola do Brasil. Após ser domesticada, por volta do século XI a. C., a cultura foi disseminada para a América do Norte, Europa e América do Sul (Priolli et al., 2004). Posteriormente à sua introdução no Brasil, em 1882, a soja passou a ser cultivada em todo o país, adquirindo importância a partir da década de 60 do século XX, especialmente na Região Sul, onde apresentou melhor adaptação (Bonetti, 1981; Priolli et al., 2004). A expansão do cultivo da soja para outras regiões do país, principalmente no cerrado, possibilitou a ascensão do Brasil de uma posição inexpressiva para a de segundo maior produtor de soja mundial (Priolli et al., 2004).

O aumento da produtividade se deve principalmente pelo aprimoramento das técnicas utilizadas no melhoramento genético, com a obtenção de cultivares mais adaptadas para cada uma das diferentes regiões produtoras (Priolli et al., 2002). A criação de cultivares resistentes aos patógenos contribui para diminuir as perdas na produção da cultura, aumentando a produtividade (Alcântara Neto, 2001; Priolli et al., 2002).

A área plantada de soja no Brasil, na safra 2012/2013, ultrapassou 27 milhões de hectares e a produção rendeu mais de 81 milhões de toneladas. Mais da metade da produção brasileira tem se destinado à exportação. Em 2012, quase 49 milhões de toneladas de soja foram exportadas. A China, a União Européia, a Tailândia e a Coreia do Sul estão entre as principais importadoras da soja brasileira (Conab, 2013).

### 2.2. Fatores que afetam a produtividade da soja

Existem dois tipos de fatores que podem reduzir a produtividade: abióticos e bióticos. Dentre os abióticos, destacam-se: a falta ou excesso de água em determinados estádios de desenvolvimento da cultura; temperaturas muito elevadas ou muito baixas; falta de nutrientes no solo; danos físicos; entre outros (Oliveira,

2008). Os vírus, bactérias, fungos e nematóides são fatores bióticos que podem causar grandes perdas na produção da soja (Oliveira, 2008). Segundo Silva (2006), as perdas anuais de produção por doenças que afetam a cultura são estimadas em 15 a 20%. Para determinadas doenças, essa perda pode chegar a quase 100% (Silva, 2006).

Na safra de soja de 1997/98 do Brasil e Paraguai e da Argentina e Bolívia, as perdas por doenças somaram 7.990.100 toneladas ou o equivalente a US\$ 1.761.839.100,00, considerando o preço de US\$ 220,50 por tonelada. Somente no Brasil, as perdas foram estimadas em 5.881.000 toneladas (Yorinori, 2002).

Entre as doenças economicamente importantes que se destacam na cultura da soja, podemos citar: a antracnose (*Colletotrichum truncatum*), o cancro da haste (*Diaporthe phaseolorum f. sp. meridionalis*), o crestamento foliar (*Cercospora kikuchii*), a ferrugem (*Phakopsora pachyrhizi* e *P. meibomiae*), amancha olho-de-rã (*Cercospora sojina*), a mancha parda (*Septoria glycines*), o nematóide de cisto (*Heterodera glycines*), os nematóides de galha (*Meloidogyne incognita* e *M. javanica*) e a podridão branca da haste (*Sclerotinia sclerotiorum*) (Andrade, 2005; Brogin, 2005; Henning et al., 2010).

### **2.3. Diversidade genética na cultura da soja brasileira**

Grandes perdas na cultura da soja também podem ser causadas pela baixa variabilidade genética encontrada nas cultivares brasileiras, o que aumenta a vulnerabilidade ao ataque de pragas, patógenos e estresses ambientais (Hiromoto e Vello, 1986; Priolli et al., 2004). Por derivarem de poucos ancestrais comuns, as cultivares brasileiras possuem uma estreita base genética, provocando uma limitação na variabilidade genética da cultura (Yamanaka et al., 2007).

Hiromoto e Vello (1986) citaram que, na safra de 83/84, as cultivares recomendadas para plantio eram derivadas de apenas 26 ancestrais, sendo nove destes ancestrais responsáveis por 80% do conjunto gênico total. Esses pesquisadores também encontraram grande semelhança entre o germoplasma brasileiro e o germoplasma do sul dos Estados Unidos, indicando que os principais ancestrais das cultivares brasileiras provinham daquela região.

Pouca variabilidade genética também foi encontrada em pesquisas feitas por Abdelnoor et al. (1995), quando avaliaram 38 cultivares brasileiras de soja com

marcadores RAPD, encontrando distâncias genéticas que variaram de 0% a 37%, com média de 17,4%. Bonato et al. (2006) obtiveram resultados semelhantes, avaliando 317 cultivares de soja brasileiras, utilizando marcadores AFLP. Neste último trabalho, os coeficientes de similaridade genética tiveram uma média de 0,61, variando de 0,17 a 0,97, sendo a maioria desses coeficientes maiores que 0,6, o que demonstrou que as cultivares de soja brasileiras possuem muita semelhança entre si.

Priolli et al. (2004) analisaram a variabilidade genética dos programas de melhoramento brasileiros utilizando um grupo de 184 cultivares de soja e doze pares de marcadores microssatélites. Esses pesquisadores avaliaram a diversidade genética entre as cultivares e entre diferentes períodos da cultura no Brasil (década de 70, 80 e 90) e concluíram que, nesses 30 anos, a variabilidade genética do germoplasma brasileiro se manteve constante. Em outro trabalho, Priolli et al. (2010), utilizando 18 pares de microssatélites, avaliaram a relação genética de 168 cultivares brasileiras e obtiveram um coeficiente de diversidade genética médio de 0,58.

Vários estudos de diversidade genética do germoplasma brasileiro foram realizados. Entretanto, estes trabalhos avaliaram somente uma pequena porção do genoma da soja, utilizando-se poucos marcadores. Com o avanço da biotecnologia e o surgimento de novas técnicas de genotipagem de alto desempenho, é possível analisar amplamente o genoma da soja com a finalidade de demonstrar a atual realidade do germoplasma brasileiro nos programas de melhoramento e contribuir para a manutenção da diversidade genética nos bancos de germoplasma.

#### **2.4. Seleção Assistida por Marcadores (SAM)**

O melhoramento genético tem como objetivo aumentar a frequência de genes associados às características agrônômicas de interesse ao longo das gerações, utilizando-se de cruzamentos e de seleção (Jangarelli et al., 2010). Em programas de melhoramento convencional, a seleção de genótipos que contêm tais características é feita com base na avaliação fenotípica dos indivíduos e esta seleção necessita de uma planta formada. Essa técnica é bem-sucedida quando a característica de interesse é controlada por um ou poucos genes e apresenta elevada herdabilidade. Quando a característica é quantitativa, ou seja, controlada



por vários genes, o aumento do número de classes fenotípicas diminui a diferença entre as classes, dificultando a seleção. Além disso, quanto maior o número de genes que controlam a característica, maior a influência ambiental na expressão do fenótipo, o que também prejudica a seleção.

Com o advento dos marcadores moleculares, surgiu uma nova metodologia de seleção baseada no genótipo das plantas, chamada Seleção Assistida por Marcadores (SAM). Esse método permite a seleção em qualquer estágio da planta e não sofre interferências ambientais, podendo ser utilizada tanto para características de herança simples como para características quantitativas (Schuster, 2011<sub>a</sub>).

Em programas de melhoramento de plantas são várias as aplicações da SAM, como, por exemplo, acelerar a recuperação de genitores recorrentes e fixar alelos recessivos em programas de retrocruzamento (Alzate-Marin et al., 2005; Schuster, 2011<sub>a</sub>). Após um “*fingerprint*” do genoma do genitor recorrente, as linhagens provenientes do cruzamento são comparadas com o genótipo do recorrente e as mais similares, que contenham também a característica de interesse, avançam nos próximos ciclos de retrocruzamento (Alzate-Marin et al., 2005). Tajima et. al. (2002) introduziram o gene de resistência ao vírus Y da batata (*va/va*) em tabaco da variedade Kerti Nº 1 para a variedade comercial Coker 319, utilizando a seleção assistida por marcadores em retrocruzamento. Dez marcadores RAPD (*Random Amplified Polymorphic DNA*) e 76 marcadores AFLP (*Amplified Fragment Length Polymorphic*) foram utilizados na caracterização do genoma da variedade Coker 319. Dois marcadores ligados ao gene *va/va* foram utilizados na seleção para a característica de resistência. A utilização da SAM acelerou a obtenção de uma variedade similar à Coker 319, contendo o gene de resistência ao vírus Y da batata. Foram necessários apenas três retrocruzamentos e uma autofecundação para a obtenção desta variedade.

Outra aplicação é para características com baixa herdabilidade, difíceis de serem avaliadas em campo e que sofrem muita influência do ambiente (Rodrigues et al., 2010). Segundo Lande e Thompson (1990), quanto menor a herdabilidade da característica, maior a eficiência relativa da SAM, isso se uma grande parte da variância genética aditiva estiver associada aos marcadores. A seleção fenotípica de plantas para características com baixa herdabilidade possui grandes chances de falhar, se aplicada nas primeiras gerações de cruzamentos, ao contrário da seleção

por marcadores moleculares, onde a avaliação é baseada no genótipo e não sofre interferência ambiental (Schuster et al., 2001).

A SAM se torna uma alternativa quando os custos ou o tempo para avaliação fenotípica são elevados ou quando a característica é de difícil mensuração, como no caso da resistência a nematóides (Alzate-Marin et al., 2005; Schuster, 2011a). Silva (2006) realizou o mapeamento e a validação de QTLs (*Quantitative Trait Loci*) de resistência ao nematóide de cisto da soja (NCS), obtendo uma eficiência de seleção (ES) de 100%, utilizando marcadores microssatélites para as raças 3 e 14. Para a raça 3, a combinação dos marcadores Satt309 e GMENOD2B proporcionou 100% de ES. Similarmente, para a raça 14, a combinação de dois marcadores, Satt309 e Satt356, resultou na ES de 100%.

A piramidação de alelos de resistência, características qualitativas ou de vários QTLs de uma característica com herança complexa pode ser auxiliada pelo uso da SAM (Alzate-Marin et al., 2005; Schuster, 2011a). Os alelos a serem piramidados podem ser monitorados por marcadores moleculares ao longo do processo de piramidação, otimizando tempo e evitando avaliações fenotípicas para cada característica. Viganó (2012) utilizou marcadores moleculares microssatélites na SAM para a piramidação de genes de resistência à ferrugem asiática da soja (FAS). Após vários cruzamentos entre diversas fontes de resistência, obteve-se uma planta com quatro dos cinco genes de resistência à FAS descobertos até o momento. Segundo a autora, a piramidação destes genes de resistência somente pôde ser concretizada devido ao uso da seleção assistida por marcadores.

#### **2.4.1. Seleção assistida por marcadores no Brasil**

Embora comprovada a grande importância da SAM para auxiliar o melhoramento genético de plantas, esta metodologia é pouco utilizada por empresas brasileiras. O fato se deve principalmente à falta de infraestrutura necessária para avaliar as plantas molecularmente em larga-escala. Os equipamentos necessários para a utilização da SAM em largaescala têm um custo elevado, além de necessitar de mão-de-obra especializada. Outro grande problema é a falta de trabalhos de mapeamento para características de interesse, utilizando germoplasma local. A maioria dos marcadores moleculares ligados a características de interesse agrônômico provém de estudos realizados a partir de populações obtidas por

cruzamentos biparentais, sendo necessária a validação desses marcadores para qualquer população diferente das que foram criadas. Quando é possível utilizar esses marcadores para diversas populações, geralmente, eles estão ligados a características de herança simples, limitando a utilização da SAM. Enquanto grandes empresas multinacionais instaladas no Brasil utilizam amplamente a SAM, enviando suas amostras a laboratórios próprios no exterior, as empresas brasileiras, quando utilizam a seleção assistida por marcadores, se limitam à recuperação de genótipos recorrentes em programas de retrocruzamento e à seleção de plantas para características de herança simples. Isso representa uma enorme desvantagem para o Brasil e para as empresas locais, em comparação com os países mais desenvolvidos e as empresas globais. A diminuição dos custos da genotipagem de alto rendimento e o desenvolvimento de trabalhos de mapeamento com germoplasma local pode tornar possível o uso em larga-escala da SAM por programas de melhoramento brasileiros (Schuster, 2011b).

## **2.5. Tipos de marcadores moleculares**

Marcadores moleculares podem ser definidos como qualquer característica molecular proveniente de regiões expressas ou não do DNA, capazes de diferenciar dois ou mais indivíduos, e que são herdadas geneticamente sem sofrer interferência ambiental (Ferreira e Grattapaglia, 1996; Milach, 1998).

O uso destes marcadores em programas de melhoramento tem desempenhado um papel importante no desenvolvimento de variedades com elevada produtividade, adaptadas a diversos ambientes e resistentes às mais variadas doenças (Marcelino et al., 2007).

Os marcadores moleculares podem ser divididos em dois grupos, conforme a metodologia aplicada para obtê-los: 1) marcadores baseados na técnica de hibridização, como marcadores RFLP (*Restriction Fragment Length Polymorphism*); e 2) marcadores baseados na técnica de amplificação de DNA, conhecida como PCR (*Polymerase Chain Reaction*), como os marcadores RAPD (*Random Amplified Polymorphic DNA*); AFLP (*Amplified Fragment Length Polymorphism*); SSR (*Simple Sequence Repeats*), também conhecidos como microssatélites; e SNP (*Single Nucleotide Polymorphisms*) (Alzate-Marin et al., 2005; Caixeta et al., 2006; Marcelino et al., 2007). A metodologia do primeiro grupo consiste em isolar o DNA total dos

indivíduos e clivá-lo em sítios específicos com enzimas de restrição. A clivagem do DNA gera fragmentos que são separados por eletroforese e transferidos para uma membrana de náilon ou nitrocelulose. A membrana contendo os fragmentos é imersa em uma solução com sondas específicas marcadas e a hibridização ocorrerá nos fragmentos que contêm sequências complementares às sondas. A visualização da posição dos fragmentos que sofreram hibridização pode ser feita por autorradiografia. As variações nas sequências primárias dos sítios de restrição ou na mudança de suas posições, devido a inserções e/ou deleções, causam o polimorfismo entre os diferentes indivíduos (Alzate-Marin et al., 2005).

Os marcadores baseados na amplificação de DNA são os mais utilizados atualmente (Marcelino et al., 2007). A metodologia consiste na amplificação de fragmentos específicos de DNA, delimitados por iniciadores (também chamados de *primers*, em inglês), utilizando a reação em cadeia da polimerase (PCR) (Caixeta et al., 2006; Marcelino et al., 2007). O polimorfismo entre indivíduos pode ser causado por mutações no(s) sítio(s) de anelamento dos *primers*, impedindo a amplificação; ou devido à inserções/deleções e/ou erros na replicação do DNA, gerando fragmentos amplificados de tamanhos diferentes (Alzate-Marin et al., 2005).

Segundo Kumpatla et al. (2012), os marcadores moleculares também podem ser classificados conforme a metodologia e a capacidade de análise de cada método, formando três grupos principais: (1) de baixo rendimento, como é o caso dos RFLPs; (2) de rendimento médio: RAPD, AFLP e SSR; e (3) de alto rendimento: marcadores SNPs.

Entre os marcadores disponíveis atualmente, os mais utilizados são os microssatélites e os SNPs, devido a abundância desses marcadores no genoma, o baixo custo de genotipagem, a reprodutibilidade e principalmente a automatização da técnica.

### **2.5.1. Microssatélites (SSR)**

Os microssatélites, também conhecidos como SSR (*Simple Sequence Repeats*) ou STR (*Short Tandem Repeats*), correspondem a regiões comuns ao longo do genoma dos eucariotos com sequências de um a seis nucleotídeos repetidos em tandem (Fuganti et al., 2004; Caixeta et al., 2006). As sequências de DNA que flanqueiam os microssatélites geralmente são conservadas entre os

indivíduos de uma mesma espécie, permitindo a amplificação do DNA repetitivo, por meio de *primers* específicos. As regiões que contêm essas sequências podem ser amplificadas individualmente via PCR utilizando *primers* com cerca de 20 bases. Esses *primers* são complementares às sequências repetidas que flanqueiam os microssatélites (Caixeta et al., 2006; Schuster e Cruz, 2008). O polimorfismo entre os indivíduos é observado pela variação no tamanho das sequências amplificadas, onde cada diferença representa um alelo diferente do mesmo loco (Caixeta et al., 2006). Essa variação pode ser causada por um *crossing-over* desigual, por erros na DNA polimerase durante a replicação ou por substituições/inserções/deleções nas regiões flangeadoras dos microssatélites (Levinson e Gutman, 1987; Macaubas et al., 1997).

De acordo com o número de nucleotídeos repetidos, os microssatélites podem ser agrupados em diferentes classes, das quais três se destacam: Dinucleotídeos (repetição de duas bases em tandem), trinucleotídeos (repetição de três bases em tandem) e tetranucleotídeos (repetição de quatro bases em tandem) (Varshney et al., 2005). Embora haja outras classes, elas são menos frequentes no genoma. Segundo Gao et al. (2003) e Varshney et al. (2005), a classe mais frequente é de trinucleotídeos, seguida de dinucleotídeos e tetranucleotídeos. Na análise de ESTs (*Expressed Sequence Tags*) em soja, conforme o trabalho de Gao et al. (2003), as classes de SSR mais frequentes são  $(AG)_n$ ,  $(AT)_n$  e  $(AAG)_n$ .

Os locos microssatélites são altamente informativos, devido a sua característica codominante (capacidade de distinguir indivíduos heterozigotos dos homozigotos) e ao multialelismo (vários alelos por loco). Além disso, esses marcadores apresentam elevada reprodutibilidade, requerem uma quantidade pequena de DNA, têm baixo custo (depois de desenvolvidos) e a técnica pode ser automatizada. Por esses motivos, este tipo de marcador vem sendo utilizado para vários estudos, como mapeamento genômico, estudos de ligação, identificação de genótipos, proteção de cultivares, avaliação de pureza de sementes, utilização e conservação de germoplasma, estudos de diversidade, análise gênica e de locos quantitativos (QTLs), análise de *Pedigree*, seleção assistida por marcadores e análise de bibliotecas para clonagem de genes (Caixeta et al., 2006).

Por outro lado, marcadores microssatélites apresentam algumas desvantagens operacionais como a ocorrência de alelos nulos, causada por uma falha na amplificação devido a mutações no sítio de anelamento dos *primers*

(Fernando et al., 2001; Pompanon et al., 2005); homoplasia de tamanho, ou seja, obtenção de fragmentos de mesmo tamanho, mas não necessariamente com a mesma sequência, causada por inserções/deleções próximas ao marcador microssatélite (Pompanon et al., 2005); ocorrência de competição na reação de PCR, levando a uma amplificação desigual de alelos e comprometendo a observação de heterozigotos em amostras de DNA em bulk (Heckenberger et al., 2002; Pompanon et al., 2005); inconsistência na identificação de tamanhos de alelos, devido à diversidade de sequenciadores automáticos utilizados, com diferentes softwares, corantes e géis (Vignal et al., 2002); falsos alelos devido a artefatos na reação de PCR, como erro na Taq DNA polimerase (Fernando et al., 2001; Pompanon et al., 2005).

### **2.5.2. Polimorfismo de um único nucleotídeo (SNPs)**

Os marcadores de polimorfismo de um único nucleotídeo ou *Single Nucleotide Polymorphisms* (SNPs) se referem a uma substituição, inserção ou deleção (InDels) de um único nucleotídeo entre fragmentos homólogos de DNA (Zhu et al., 2003). Os SNPs são a fonte mais abundante de variação no genoma de eucariotos (Yoon et al., 2007). Para que esta variação seja considerada um SNP, ela tem que ocorrer em pelo menos 1% da população (Vignal et al., 2002; Caixeta et al., 2006).

Estes marcadores passaram a ter maior importância com o sequenciamento do genoma humano, estimado em aproximadamente 10 milhões de SNPs, representando 90% do polimorfismo encontrado no genoma (Caixeta et al., 2006; International HapMap Project). O Consórcio Internacional HapMap genotipou mais de 3,1 milhões de SNPs em 270 indivíduos de quatro populações geograficamente diferentes, em que a densidade dos SNPs foi de aproximadamente um a cada quilobase (kb) (The International HapMap Consortium, 2007). Em soja convencional, foi estimado um número de 4 a 5 milhões de SNPs baseado na média de um SNP a cada 270 bases, tendo como base o trabalho de Zhu et al. (2003), que analisaram um segmento de aproximadamente 76,3 kb e encontraram 280 SNPs (Yoon et al., 2007; Deulvot et al., 2010).

Os marcadores SNPs são codominantes e bi-alélicos, não sendo tão informativos por loco quanto os microssatélites (Vignal et al., 2002). Entretanto, essa

perda de informatividade é compensada pela maior densidade destes marcadores no genoma, obtendo ampla cobertura (Mammadov et al., 2012). Além disso, apresentam outras vantagens que tornaram os SNPs os principais marcadores da atualidade, como uma baixa taxa de mutação comparada aos SSR e a existência de vários métodos automatizados de alto rendimento com menores custos, menos erros de genotipagem maior rapidez na obtenção dos resultados (Brinkmann et al., 1998; Nachman e Crowell, 2000; Kondrashov, 2002; Yoon et al., 2007). Os marcadores SNPs podem tornar possíveis mapas genéticos saturados ou até mesmo supersaturados, obtendo um mapeamento fino em regiões alvo, uma rápida associação dos marcadores com a característica e acelerar a clonagem do gene/QTL de interesse (Mammadov et al., 2012).

Por outro lado, o uso de marcadores SNPs pode apresentar algumas barreiras, como: (1) um custo mais elevado, por utilizar um maior número de marcadores para alcançar o nível de polimorfismo de outros marcadores; (2) a existência de uma pequena porcentagem de SNPs polimórficos em plantas poliploides; (3) a falta de uma capacidade computacional que permita o processamento rápido de informações e o armazenamento de grande quantidade de arquivos de sequenciamento (Mammadov et al., 2012).

## **2.6. Plataformas de genotipagem**

Os marcadores SNPs são atualmente os mais utilizados em genética e melhoramento de plantas, várias tecnologias automatizadas de alto rendimento para sequenciamento e genotipagem têm sido desenvolvidas, diminuindo os custos de dados por amostra (*data point* – em inglês) (Schuster, 2011a; Kumpatla et al., 2012). A escolha da plataforma dependerá do número de SNPs, do comprimento das sequências moldes, número de amostras, tempo e recursos do melhorista (Kumpatla et al., 2012). Neste trabalho, serão apresentadas algumas plataformas de alto rendimento para genotipagem disponíveis atualmente.

### **2.6.1. Plataforma Illumina Bead Array**

Esta plataforma é baseada em esferas sílicas de três micrometros separadas, por aproximadamente 5,7 micrometros e inseridas em micro poços em

dois tipos de matrizes: feixes de fibra ótica ou lâminas de sílica planas. Cada esfera contém milhares de cópias de oligonucleotídeos específicos, que agem como sequências de captura em dois tipos de metodologias utilizadas pela Illumina: GoldenGate e Infinium. Um scanner de alta resolução, chamado *iScan*, lê as matrizes e gera dados de intensidade que são convertidos em dados, utilizando um software específico chamado GenomeStudio (Kumpatla et al., 2012).

A metodologia GoldenGate permite a genotipagem de 384 até 1536 SNPs, gerando até 147.456 genótipos em uma única reação, contendo 96 amostras (Fan et al., 2006; Illumina, 2013a). Esse método demora três dias e envolve nove etapas: (1) ativação do DNA: as amostras de DNA são ativadas por biotilação; (2) hibridização: os SNPs específicos, tampão de hibridização e partículas paramagnéticas são combinadas com o DNA ativado. Para cada loco SNP, existem três oligonucleotídeos, sendo dois alelos-específicos (ASOs) capazes de distinguir o SNP, e um loco específico (LSO); (3) extensão e ligação dos oligonucleotídeos: após várias lavagens para remover excessos, ocorre a extensão dos ASOs complementares aos SNPs pela DNA polimerase e então a DNA ligase une os ASOs ao LSO, formando moldes para o PCR; (4) amplificação: amplificação via PCR utilizando três primers universais, P1, P2 e P3, corados com Cy3, Cy5 e biotina, respectivamente; (5) preparação do produto de PCR: produtos de PCR dupla fita são imobilizados em partículas paramagnéticas onde são desnaturados formando DNAs de fita simples; (6) hibridização na matriz: os produtos de PCR de fita simples corados são hibridizados com as sequências IllumiCode complementares em suas respectivas esferas; (7) lavagem das placas; (8) leitura: é feita uma análise do sinal de fluorescência utilizando o BeadArray Reader; (9) genotipagem automática: baseado na intensidade das duas fluorescências possíveis para os alelos de cada SNP, é feita a genotipagem automática das amostras por um software (Fan et al., 2006; Illumina, 2013b).

Maior quantidade de locos SNPs pode ser avaliada com a metodologia Infinium, chegando a milhões de SNPs por amostra em uma reação (Kumpatla et al., 2012; Illuminab). Dependendo do tipo de SNP, duas metodologias podem ser utilizadas: Infinium I e Infinium II. Enquanto o Infinium I é próprio para SNPs (A/G) e (T/C), utilizando dois tipos de esferas por SNP (um alelo por esfera), o Infinium II é designado para todos os SNPs e requer apenas um tipo de esfera para ambos alelos (Kumpatla et al., 2012). A primeira etapa na metodologia Infinium I é a



amplificação de todo o genoma. Em seguida, acontece a hibridização nas esferas de sondas de captura que contêm sequências locais específicas com um alelo-específico na base 3' terminal. Uma reação de extensão com *primers* alelo-específicos é usada para incorporar nucleotídeos de biotina marcados. As etapas iniciais são as mesmas no Infinium II, com a diferença que a reação de extensão do *primer* é uma única base, que utiliza uma combinação de biotinas marcadas (ddCTP e ddGTP) e 2,4-dinitrofenol (DNP) marcado (ddATP e ddUTP) (Ragoussis, 2006). As intensidades das fluorescências são detectadas pelo BeadArray Reader e genotipados automaticamente pelo software da Illumina (Illumina<sub>b</sub>).

### 2.6.2. OpenArray®

Essa plataforma da Life Technologies é baseada no ensaio TaqMan end-point, utilizando um equipamento de PCR em tempo-real denominado QuantStudio™ 12K Flex. A placa OpenArray é oito vezes menor que uma placa de 384 poços convencional, mas contém uma capacidade oito vezes maior, por ser dividida em 48 sub-arranjos e cada subarranjo dividido em 64 poços, nos quais os volumes de reação são carregados. As placas são cobertas com compostos hidrofílicos e hidrofóbicos para manter os reagentes nos poços por tensão superficial (Kumpatla et al., 2012; Life Technologies, 2013). Estas placas são pré-carregadas com os reagentes do experimento pelo fornecedor e posteriormente encaminhadas ao usuário para carregar as amostras de DNA (Kumpatla et al., 2012). O equipamento tem a capacidade de analisar quatro placas OpenArray por corrida, sendo que cada corrida gera um total de 12.000 *data points*. Se for incorporado ao equipamento, o sistema AccuFill, a capacidade aumenta para 110.000 ou mais *data points* em oito horas (Life Technologies, 2013). Uma grande vantagem desse equipamento é a possibilidade de utilizar diferentes combinações de SNPs x amostra, incluindo formatos de 32 x 96, 64 x 48, 128 x 24, 192 x 16, 256 x 12. Isso permite avaliar pequenos grupos de SNPs em uma larga escala de amostras, sendo mais atrativo para SAM em programas de melhoramento, nos quais o gene ou a região do QTL requer poucos marcadores para seleção em uma grande quantidade de amostras (Kumpatla et al., 2012).

### 2.6.3. KASP™ (“Kompetitive” Allele Specific PCR)

Embora essa plataforma da LGC Genomics seja robusta e não permita multiplex, ela tem sido amplamente utilizada para genotipagem de SNPs (Kumpatla et al., 2012). Essa plataforma, quando acoplada ao conjunto de instrumentos denominados SNPLine da LGC Genomics, torna-se de alto rendimento, sendo capaz de gerar mais de 500.000 *data points* por dia, além de permitir diferentes configurações que possibilitam analisar qualquer número de amostra (de uma a milhares) para diferentes SNPs (LGC Genomics<sub>a</sub>). A metodologia é baseada na extensão de *primers* alelos-específicos e na transferência de energia de ressonância por fluorescência (FRET) (Kumpatla et al., 2012). A reação de PCR é composta por dois mix. O primeiro mix (KASP Primer Mix) contém *primers forward* alelos-específicos e um *primer reverse* loco específico comum. Cada *primer forward* apresenta uma sequência final, que correspondente a um dos cassetes universais de FRET (fluorescência FAM™ ou HEX™). O segundo mix (KASP Master Mix) é uma solução tampão otimizada que contém cassetes universais FRET, DNA polimerase, nucleotídeos livres (dNTPs) e MgCl<sub>2</sub>. A diferenciação dos alelos ocorre na reação de PCR, onde os *primers forward* competem pelo sítio de anelamento da região alvo. Para um indivíduo homocigoto, somente um dos tipos de *primers forward* consegue se anelar e amplificar, emitindo somente um tipo de fluorescência (FAM™ ou HEX™). Para o heterocigoto, os dois *primers forward* podem se anelar, emitindo ambas as fluorescências (LGC Genomics<sub>b</sub>).

### 2.7. Mapeamento genético em plantas

Os primeiros mapas genéticos eram baseados em características morfológicas e citológicas, o que limitava a obtenção de mapas genéticos a espécies que continham maiores informações genéticas, como o milho, o tomate e a ervilha. Com o descobrimento das isoenzimas, o número de espécies tornou-se potencialmente ilimitado. Entretanto, os mapas genéticos tinham pouca cobertura genômica, devido à pouca quantidade de locos avaliados, que geralmente não passavam de 30. O advento dos marcadores moleculares possibilitou a obtenção de mapas genéticos a todas as espécies com ampla saturação de todo o genoma (Carneiro e Vieira, 2002; Petrolí, 2013).

O mapeamento genético corresponde ao ordenamento de locos linearmente em grupos de ligação (Carneiro e Vieira, 2002; Petroli, 2013). Os mapas genéticos podem ser utilizados para várias aplicações, como análises filogenéticas, estudos de sintenia, localização e clonagem de genes e/ou QTLs de interesse (Carneiro e Vieira, 2002; Pereira e Pereira, 2006; Vieira et al., 2006). Para características quantitativas, é possível o estudo dos efeitos gênicos e da ação dos genes de maneira isolada, interativa e cumulativa (Pereira e Pereira, 2006). Os resultados obtidos pelo mapeamento genético fornecem uma base para a seleção assistida por marcadores, sendo selecionados os marcadores moleculares mais próximos da região do DNA que representa a característica de interesse, podendo ser governada por um ou por vários genes. Por esse motivo, a precisão do mapeamento genético influencia diretamente a eficiência da SAM.

A utilização de grande quantidade de marcadores moleculares para o mapeamento diminui a distância entre marcadores adjacentes, aumentando a resolução do mapeamento e garantindo a cobertura adequada do genoma (Jangarelli, 2010). Além disso, a metodologia empregada no mapeamento pode aumentar ou diminuir a precisão do mapa. De uma maneira geral, é possível classificar dois tipos de mapeamento: mapeamento por recombinação e mapeamento por associação, também chamado de mapeamento por desequilíbrio de ligação.

### **2.7.1. Mapeamento por recombinação**

Esse método se baseia na frequência de recombinação entre locos segregantes em uma população de mapeamento previamente gerada a partir de cruzamentos bi-parentais de indivíduos contrastantes. Carneiro e Vieira (2002) estabeleceram algumas premissas básicas na construção desses mapas, como a seleção da população apropriada, a escolha do tipo de marcador, a verificação da distorção da segregação mendeliana, o cálculo da frequência de recombinação entre pares de marcadores e funções de mapeamento, o estabelecimento dos grupos de ligação e a determinação da ordem dos marcadores nestes grupos.

A escolha da população apropriada deve seguir duas condições: o máximo de polimorfismo entre os genitores e a produção de gerações com locos em desequilíbrio de ligação (Carneiro e Vieira, 2002; Pereira e Pereira, 2006). O

desequilíbrio de ligação é fundamental para se identificar a co-segregação dos marcadores de DNA com as características fenotípicas de interesse (Pereira e Pereira, 2006). Os genitores devem ser geneticamente distantes e contrastantes fenotipicamente, resultando em maior proporção de marcas polimórficas e maior cobertura genômica, obtendo, assim, maior eficiência na identificação, no mapeamento e na análise dos genes de interesse (Carneiro e Vieira, 2002; Pereira e Pereira, 2006). Diversos tipos de populações de mapeamento podem ser utilizados, desde populações F<sub>2</sub>, populações derivadas de F<sub>2</sub> por autofecundação, linhagens endogâmicas recombinantes (*Recombinant Inbred Lines* – RILs), duplo-haplóides e retrocruzamento (Carneiro e Vieira, 2002; Schuster e Cruz, 2008). Geralmente, é utilizada a população F<sub>2</sub>, por ser fácil de ser obtida e apresentar máximo desequilíbrio de ligação (Carneiro e Vieira, 2002; Pereira e Pereira, 2006).

O tipo de marcador molecular também é fundamental no mapeamento, tanto no nível de informatividade de cada marcador, quanto no tempo necessário para genotipagem, custo e praticidade de cada um (Carneiro e Vieira, 2002). Os marcadores codominantes, como é o caso de marcadores RFLP, SSR e SNPs, apresentam maior nível de informação por loco, por diferenciar indivíduos heterozigotos dos homozigotos. Já marcadores dominantes, como RAPD e AFLP, identificam a presença de apenas um alelo, não sendo possível a distinção do homozigoto para o heterozigoto, diminuindo o nível de informação por loco (Carneiro e Vieira, 2002; Caixeta et al., 2006).

Após a obtenção da população do mapeamento a partir de pais contrastantes e a escolha dos marcadores que serão utilizados no mapeamento, é feita a avaliação fenotípica e genotípica dos indivíduos da população. Primeiramente, um grande número de marcadores é utilizado nos genitores contrastantes a fim de detectar polimorfismos entre eles, técnica chamada de *screening*. Os marcadores moleculares, que forem polimórficos para os genitores, são testados com a população de mapeamento. Em seguida, é feito um teste de aderência para verificar se os marcadores estão segregando na proporção esperada. Locos que apresentam segregação distorcida significativa devem ser removidos, pois podem gerar falsos positivos. Posteriormente, é verificado se estes marcadores, tomados par a par, estão ligados pelo teste  $\chi^2$ . Se os locos estiverem ligados, é estimada a frequência de recombinação entre pares de marcadores, podendo ser utilizados vários métodos, dos quais se destaca o método da máxima

verossimilhança. A próxima etapa é a formação dos grupos de ligação, onde são estabelecidos critérios (geralmente LOD score acima de 3 e frequência de recombinação máxima de 30%) para considerar quais marcadores ligados pertencem a determinado grupo de ligação. Por fim, é estabelecida a ordem de cada marcador dentro de cada grupo de ligação, onde diversas metodologias combinadas com funções de mapeamento podem ser utilizadas, destacando-se as funções de Haldane e Kosambi (Carneiro e Vieira, 2002; Bhering, 2008; Schuster e Cruz, 2008).

### **2.7.2. Mapeamento por associação**

O mapeamento por associação, também conhecido como mapeamento por desequilíbrio de ligação, detecta e localiza QTLs baseado na intensidade da correlação entre marcadores moleculares mapeados e as características fenotípicas, sem a necessidade de utilizar populações de mapeamento (Mackay e Powell, 2007). Esse tipo de mapeamento utiliza uma coleção de cultivares, linhagens ou variedades para detectar blocos em desequilíbrio de ligação, utilizando grande densidade de marcadores. O desequilíbrio de ligação entre dois locos é em função do tempo (número de gerações) decorrido, desde que se iniciaram as gerações de recombinação e a frequência de recombinação entre os locos. Após várias gerações de recombinação, em uma população não estruturada, apenas correlações entre QTLs e marcadores proximamente ligados devem permanecer, facilitando o mapeamento mais fino. Os blocos de desequilíbrio de ligação são pequenos e o genoma precisa estar bastante saturado para que seja possível identificar estes blocos, os quais contêm genes ou QTLs de interesse (Mackay e Powell, 2007).

Por utilizar uma coleção de cultivares, esta metodologia apresenta diversas vantagens sobre o mapeamento, utilizando cruzamentos bi-parentais: uma ampla variação genética é representada por diversos haplótipos, não sendo as marcas ou QTLs limitados apenas aos dois genitores; o mapeamento de associação tem uma maior resolução por acumular todas as meioses ao longo das gerações; os dados fenotípicos que caracterizam cada cultivar podem ser utilizados, não havendo a necessidade de testá-los com populações especiais de mapeamento, diminuindo o tempo para obtenção do mapa (Buckler IV e Thornsberry, 2002; Sorkheh et al., 2008).

Plataformas de genotipagem de alto rendimento e marcadores espalhados por todo o genoma, como é o caso dos SNPs, tornam possível a obtenção de um mapeamento por associação mais fino, o que pode garantir a precisão da seleção assistida por marcadores em programas de melhoramento genético.

### 3. REFERÊNCIAS BIBLIOGRÁFICAS

ALBDELNOOR, 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. **Revista Brasileira de Genética**, 18:265-273, 1995.

ALCÂNTARA NETO, F. **Marcadores microssatélites na identificação de cultivares de soja**. Viçosa: Universidade Federal de Viçosa, 2001. 46p. Dissertação (Mestrado em Fitotecnia).

ALZATE-MARIN, A.L.; CERVIGNI, G.D.L.; MOREIRA, M.A.; BARROS, E.G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. **Fitopatologia Brasileira**, 30:333-342, 2005.

ANDRADE, V.A. **Isolamento e caracterização de sequências expressas similares a genes de resposta de defesa a patógenos em soja**. Viçosa: Universidade Federal de Viçosa, 2005. 68p. Dissertação (Mestrado em Bioquímica Agrícola).

BHERING, L.L. **Mapeamento genético em famílias simuladas de irmãos completos**. Viçosa: Universidade Federal de Viçosa, 2008. 150p. Tese (Doutorado em Genética e Melhoramento).

BONATO, A.L.V.; CALVO E.S.; GERALDI, I.O.; ARIAS, C.A.A. Genetic similarity among soybean (*Glycine max* (L) Merrill) cultivars released in Brazil using AFLP markers. **Genetics and Molecular Biology**, 29:692-704, 2006.

BONETTI, L.P. Distribuição da Soja no Mundo. In: MIYASAKA, S.; MEDINA J.C. (eds.). **A soja no Brasil**. Campinas: ITAL, 1981. cap. 1, Item 1, p.1-6.

BRINKMANN, B.; KLINTSCHAR, M.; NEUHUBER, F.; HÜHNE, J.; ROLF, B. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. **The American Journal of Human Genetics**, 62:1408-1415, 1998.

BROGIN, R.L. **Mapeamento de genes de resistência à ferrugem e de QTLs envolvidos na resistência à septoriose em soja**. Piracicaba: Escola Superior de Agronomia Luiz Queiroz, 2005. 93p. Tese (Doutorado em Genética e Melhoramento).

BUCKLER IV, E.S.; THORNSBERRY, J.M. Plant molecular diversity and application to genomics. **Current Opinion in Plant Biology**, 5:107-111, 2002.

CAIXETA, E.T.; OLIVEIRA, A.C.B.; BRITO, G.G.; SAKIYAMA, N.S. Tipos de marcadores moleculares. In: BORÉM, A.; CAIXETA, E.T. (eds.). **Marcadores moleculares**. 1. ed. Viçosa: Universidade Federal de Viçosa, 2006. p. 09-78.

CARNEIRO, M.S.; VIEIRA, M.L.C. Mapas genéticos em plantas. **Bragantia**, 61:89-100, 2002.

CONAB. **Acompanhamento da safra brasileira: grãos – safra 2013/2014**. Companhia Nacional de Abastecimento. Brasília: Conab. 2013. Disponível em: [http://www.conab.gov.br/OlalaCMS/uploads/arquivos/13\\_10\\_09\\_17\\_39\\_08\\_boletim\\_portugues\\_outubro\\_2013.pdf](http://www.conab.gov.br/OlalaCMS/uploads/arquivos/13_10_09_17_39_08_boletim_portugues_outubro_2013.pdf). Acesso em: 18, outubro, 2013.

DEULVOT, C.; CHARREL, H.; MARTY, A.; JACQUIN, F.; DONNADIEU, C.; LEJEUNE-HÉNAULT, I.; BURSTIN, J.; AUBERT, G. Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. **BMC Genomics**, 11:1-10, 2010.

FAN, J.B.; GUNDERSON, K.L.; BIBIKOVA, M.; YEAKLEY, J.M.; CHEN, J.; GARCIA, E.W.; LEBRUSKA, L.L.; LAURENT, M.; SHEN, R.; BARKER, D. Illumina universal bead arrays. **Methods in Enzymology**, 410:57-73, 2006.

FERNANDO, P.; EVANS, B.J.; MORALES, J.C.; MELNICK, D.J. Electrophoresis artifacts – a previously unrecognized cause of error in microsatellite analysis. **Molecular Ecology Notes**, 1:325-328, 2001.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Brasília: Embrapa-Cenargen, 1996. 220p.



FUGANTI R.; BENEVENTI, M.A.; SILVA, J.F.V.; ARIAS, C.A.A.; MARIN, S.R.R.; BINNECK, E.; NEPOMUCENO, A.L. Identificação de marcadores moleculares de microssatélites para seleção de genótipos de soja resistentes a *Meloidogyne javanica*. **Nematologia Brasileira**, 28:125-130, 2004.

GAO, L.F.; TANG, J.; LI, H.; JIA, J. Analysis of microsatellites in major crops assessed by computational and experimental approaches. **Molecular Breeding**, 12:245-261, 2003.

GIZLICE, Z.; CARTER, T.E.; BURTON, J.W. Genetic diversity in North American soybean: I. multivariate analysis of founding stock and relation to coefficient of parentage. **Crop Science**, 33:614-620, 1993.

HECKENBERGER, M.; BOHN, M.; ZIEGLE, J.S.; JOE, L.K.; HAUSER, J.D.; HUTTON, M.; MELCHINGER, A.E. Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties. I genetic and technical sources of variation in SSR data. **Molecular Breeding**, 10:181-191, 2002.

HENNING, A.A.; ALMEIDA, A.M.R.; GODOY, C.V.; SEIXAS, C.D.S; YORINORI, J.T.; COSTAMILAN, L.M.; FERREIRA, L.P.; MEYER, M.C.; SOARES, R.M.; DIAS, W.P. **Manual de identificação de doenças em soja**. Londrina: Embrapa Soja. 2010.74p.

HIROMOTO, D.M.; VELLO, N.A. The genetic base of Brazilian soybean (*Glycine max* (L.) Merrill) cultivars. **Revista Brasileira de Genética**, 9:295-306, 1986.

ILLUMINA. **GoldenGate<sup>®</sup> Assay workflow**. Disponível em: [http://res.illumina.com/documents/products/workflows/workflow\\_goldengate\\_asass.pdf](http://res.illumina.com/documents/products/workflows/workflow_goldengate_asass.pdf). Acesso em: 09, novembro, 2013a.

ILLUMINA. **Infinium<sup>®</sup> II Assay workflow**. Disponível em: [http://res.illumina.com/documents/products/workflows/workflow\\_infinium\\_ii.pdf](http://res.illumina.com/documents/products/workflows/workflow_infinium_ii.pdf). Acesso em: 09, novembro, 2013b.

INTERNATIONAL HAPMAP PROJECT. **About the International HapMap Project**. Disponível em: <http://hapmap.ncbi.nlm.nih.gov/abouthapmap.html>. Acesso em: 03, novembro, 2013.

JANGARELLI, M.; EUCLYDES, R.F.; CRUZ, C.D.; CECON, P.R.; CARNEIRO, A.P.S. Análise de agrupamento de diferentes densidades de marcadores no mapeamento genético por varredura genômica. **Revista Ceres**, 57:695-700, 2010.

KONDRASHOV, A.S. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. **Human Mutation**, 21:12-27, 2002.

KUMPATLA, S.P.; BUYYARAPU, R.; ABDURAKHMONOV, I.Y.; MAMMADOV, J.A. Genomics-Assisted Plant Breeding in the 21st Century: Technological Advances and Progress. In: ABDURAKHMONOV, I.Y. (ed.). **Plant Breeding**. Rijeka – Croácia: InTech, 2012. p. 131-184.

LANDE, R.; THOMPSON, R. Efficiency of Marker-Assisted Selection in the improvement of quantitative traits. **Genetics**, 124:743-756, 1990.

LEVINSON, G.; GUTMAN, G.A. Slipped-Strand mispairing: a major mechanism for DNA sequence evolution. **Molecular Biology and Evolution**, 4:203-221, 1987.

LGC GENOMICS. **SNPline – the flexible genotyping solution**. Disponível em: [http://www.lgcgenomics.com/genotyping/snpline/?download\\_file=58\\_1\\_snpline\\_-\\_data\\_sheet.pdf&download\\_cat=downloads](http://www.lgcgenomics.com/genotyping/snpline/?download_file=58_1_snpline_-_data_sheet.pdf&download_cat=downloads). Acesso em: 10, novembro, 2013a.

LGC GENOMICS. **The KASP genotyping assay**. Disponível em: [http://www.lgcgenomics.com/genotyping/kasp-genotyping-reagents/?download\\_file=20\\_1\\_kasp\\_brochure.pdf&download\\_cat=downloads](http://www.lgcgenomics.com/genotyping/kasp-genotyping-reagents/?download_file=20_1_kasp_brochure.pdf&download_cat=downloads). Acesso em: 10, novembro, 2013b.

LIFE TECHNOLOGIES. **QuantStudio™ 12K Flex Real-Time PCR System**. Disponível em: [http://tools.lifetechnologies.com/content/sfs/brochures/cms\\_098456.pdf](http://tools.lifetechnologies.com/content/sfs/brochures/cms_098456.pdf). Acesso em: 10, novembro, 2013.

MACAUBAS, C.; JIN, L.; HALLMAYER, J.; KIMURA, A.; MIGNOT, E. The complex mutation pattern of a microsatellite. **Genome Research**, 7:635-641, 1997.

MACKAY, I.; POWELL, W. Methods for linkage disequilibrium mapping in crops. **Trends in Plant Science**, 12:57-63, 2007.

MAMMADOV, J; AGGARWAL, R.; BUYYARAPU, R.; KUMPATLA, S. SNP markers and their impact on plant breeding. **International Journal of Plant Genomics**, 2012:1-11, 2012.

MARCELINO, F.C.; BINNECK, E.; ABDELNOOR, R.V.; NEPOMUCENO, A.L. **Ferramentas biotecnológicas aplicadas à cultura da soja**. Londrina: Embrapa, 2007.5p. (Circular Técnica n. 47).

MILACH, S.C.K. **Marcadores moleculares em plantas**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 1998. 141p.

NACHMAN, M.W.; CROWELL, S.L. Estimate of the mutation rate per nucleotide in humans. **Genetics**, 156:297-304, 2000.

OLIVEIRA, M.A.R. **Mapeamento de genes de resistência ao vírus da necrose da haste utilizando marcadores moleculares microssatélites**. Londrina: Universidade Estadual de Londrina, 2008. 69p. Tese (Doutorado em Agronomia).

PEREIRA, M.G.; PEREIRA, T.N.S. Marcadores moleculares no pré-melhoramento de plantas. In: BORÉM, A.; CAIXETA, E.T. (eds.). **Marcadores moleculares**. 1 ed. Viçosa: Universidade Federal de Viçosa, p. 85-106, 2006.

PETROLI, C.D. **Caracterização genômica de marcadores DArT com base em mapeamento genético e físico e detecção de QTLs em *Eucalyptus***. Brasília: Universidade de Brasília, 2013. 126p. Tese (Doutorado em Ciências Biológicas).

POMPANON, F.; BONIN, A.; BELLEMAIN, E.; TABERLET, P. Genotyping errors: causes, consequences and solutions. **Nature Reviews Genetics**, 6:847-859, 2005.

PRIOLLI, R.H.G.; MENDES-JUNIOR, C.T.; ARANTES, N.E.; CONTEL, E.P.B. Characterization of Brazilian soybean cultivars using microsatellite markers. **Genetics and Molecular Biology**, 25:185-193, 2002.

PRIOLLI, R.H.G.; MENDES-JUNIOR, C.T.; SOUSA, S.M.B.; SOUSA, N.E.A.; CONTEL, E.B.P. Diversidade genética da soja entre períodos e entre programas de melhoramento no Brasil. **Pesquisa Agropecuária Brasileira**, 39:967-975, 2004.

PRIOLLI, R.H.G.; PINHEIRO, J.B.; ZUCCHI, M.I.; BAJAY, M.M.; VELLO, N.A. Genetic diversity among Brazilian soybean cultivars based on SSR loci and pedigree data. **Brazilian Archives of Biology and Technology**, 53:519-531, 2010.

PRITCHARD, J.K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, 155:945-959, 2000.

RAGOSSIS, J. Genotyping technologies for all. **Drug Discovery Today: Technologies**, 3:115-122, 2006.

RODRIGUES, J.I.S.; MIRANDA, F.D.; FERREIRA, A.; BORGES, L.L.; FERREIRA, M.S.S.; GOOD-GOD, P.I.V.; PIOVESAN, N.D.; BARROS, E.G.; CRUZ, C.D.; MOREIRA, M.A. Mapeamento de QTL para conteúdos de proteína e óleo em soja. **Pesquisa Agropecuária Brasileira**, 45:472-480, 2010.

SCHUSTER, I. Marker-assisted selection for quantitative traits. **Crop Breeding and Applied Biotechnology**, S1:50-55, 2011a.

SCHUSTER, I. Seleção assistida por marcadores moleculares para QTLs no melhoramento genético. In: CARDOSO, D.L.; LUZ, L.N.; SANTANA, E.N. (eds.). **Estratégias em melhoramento de plantas**. Viçosa: Arka, p. 67-80, 2011b.

SCHUSTER, I.; ABDELNOOR, R.V.; MARIN, S.R.R.; CARVALHO, V.P.; KIIHL, R.A.S.; SILVA, J.F.V.; SEDIYAMA, C.S.; BARROS, E.G.; MOREIRA, M.A. Identification of a new major QTL associated with resistanceto soybean cyst nematode (*Heterodera glycines*). **Theoretical and Applied Genetics**, 102:91-96, 2001.

SCHUSTER, I.; CRUZ, C.D. **Estatística genômica aplicada a populações derivadas de cruzamentos controlados**. Viçosa: Universidade Federal de Viçosa, 2008. 568p.

SILVA, M.F. **Mapeamento e validação de QTLs de resistência ao nematóide de cisto da soja**. Viçosa: Universidade Federal de Viçosa, 2006. 110p. Tese (Doutorado em Genética e Melhoramento).

SORKHEH, K.; MALYSHEVA-OTTO, L.V.; WIRTHENSOHN, M.G.; TARKESH-ESFAHANI, S.; MARTÍNEZ-GÓMEZ, P. Linkage disequilibrium, genetic association mapping and gene localization in crop plants. **Genetics and Molecular Biology**, 31:805-814, 2008.

TAJIMA, T.; NOGUCHI, S.; TANOUE, W.; NEGISHI, H.; NAKAKAWAJI, T.; OHNO, T. Background selection using DNA markers in backcross breeding program for Potato Virus Y resistance in Tobacco. **Breeding Science**, 52:253-257, 2002.

THE INTERNATIONAL HAPMAP CONSORTIUM. A second generation human haplotype map of over 3.1 million SNPs. **Nature**, 449:851-861, 2007.

VARSHNEY, R.K.; GRANER, A.; SORRELLS, M.E. Genic microsatellite markers in plants: features and applications. **Trends in Biotechnology**, 23:48-55, 2005.

VIEIRA, E.A.; NODARI, R.O.; CARVALHO, F.I.F.; FIALHO, J.F. **Mapeamento genético de caracteres quantitativos e sua interação com o ambiente**. Planaltina: Embrapa Cerrados. 28p. 2006.

VIGANÓ, J. **Piramidação de genes de resistência à ferrugem asiática da soja assistida por marcadores moleculares**. Maringá: Universidade Estadual de Maringá, 2012. 92. Tese (Doutorado em Genética e Melhoramento).

VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M.; EGGEN, A. A review on SNP and other types of molecular markers and their use in animal genetics. **Genetics Selection Evolution**, 34:275-305, 2002.

YAMANAKA, N.; SATO, H.; YANG, Z.; XU, D.H.; CATELLI, L.L.; BINNECK, E.; ARIAS, C.A.A.; ABDELNOOR, R.V.; NEPOMUCENO, A.L. Genetic relationships between Chinese, Japanese, and Brazilian soybean gene pools revealed by simple sequence repeat (SSR) markers. **Genetics and Molecular Biology**, 30:85-88, 2007.

YOON, M.S.; SONG, Q.J.; CHOI L.Y.; SPECHT, J.E.; HYTEN, D.L.; CREGAN, P.B. BARCSoySNP23: a panel of 23 selected SNPs for soybean cultivar identification. **Theoretical and Applied Genetics**, 114:885-899, 2007.

YORINORI, J.T. Situação atual das doenças potenciais no cone sul. In: II CONGRESSO BRASILEIRO DE SOJA, Foz do Iguaçu. 2002. **Resumos...** Londrina: Embrapa Soja, 2002. p. 171-186.

ZHU, Y.L.; SONG, Q.J.; HYTEN, D.L.; VAN TASSELL, C.P.; MATUKUMALLI, L.K.; GRIMM, D.R.; HYATT, S.M.; FICKUS, E.W.; YOUNG, N.D.; CREGAN, P.B. Single-Nucleotide Polymorphisms in soybean. **Genetics**, 163:1123-1134, 2003.

# **CAPÍTULO 1**

## **Análise de diversidade genética em soja utilizando marcadores SNPs e microssatélites**

### **1. INTRODUÇÃO**

As cultivares de soja introduzidas no Brasil derivam de um número limitado de genótipos, provavelmente provenientes do sul dos Estados Unidos (Hiromoto e Vello, 1986). Na safra de 1983/84, as cultivares recomendadas para plantio derivavam de apenas 26 ancestrais, das quais nove representavam 80% do conjunto gênico total (Hiromoto e Vello, 1986). Wysmierski e Vello (2013) analisaram 444 cultivares brasileiras com base no coeficiente de parentesco e encontraram 60 ancestrais, dos quais apenas quatro deles contribuíram para 55,3% da base genética. Um número reduzido de ancestrais ocasiona uma base genética estreita, limitando a variabilidade genética da cultura (Yamanaka et al., 2007).

O uso extensivo de cultivares semelhantes pode resultar no aumento da vulnerabilidade a pragas e doenças. Além disso, a probabilidade de se produzir genótipos únicos aumenta proporcionalmente à quantidade de genes que eles diferem (Van Esbroeck e Bowman, 1998). Portanto, o conhecimento da diversidade genética entre diferentes genótipos é importante para o sucesso dos programas de melhoramento (Bonato et al., 2006).

A baixa variabilidade genética das cultivares brasileiras foi demonstrada em diversos trabalhos, utilizando marcadores RAPD (Addelnoor et al., 1995); AFLP (Bonato et al., 2006); e SSR (Priolli et al., 2002; Priolli et al., 2004; Priolli et al., 2010). Entretanto, estes trabalhos analisaram a diversidade genética com uma quantidade pequena de marcadores, avaliando apenas alguns locos.

Com o advento de novas tecnologias de genotipagem de alta performance e o desenvolvimento de marcadores SNPs, é possível utilizar milhares de marcadores, permitindo a avaliação de várias regiões do genoma.

Neste trabalho, foi avaliada pela primeira vez a diversidade genética da soja brasileira, utilizando uma plataforma de genotipagem de alto rendimento, com milhares de marcadores SNPs, a fim de avaliar com maior precisão a variabilidade

genética do germoplasma brasileiro. Também foram avaliados alguns marcadores SSR relacionados a características de interesse previamente citados na literatura.



## 2. MATERIAL E MÉTODOS

### 2.1. Material genético

Sementes de 169 cultivares oriundas de obtentores, produtores e de Germoplasma de soja da Coodetec foram empregadas para genotipagem (Quadro 1).

Para extração de DNA, foram moídas 50g de sementes de cada variedade e, após homogeneização, utilizados 50mg na extração de DNA. A extração de DNA foi realizada utilizando-se o kit DNAeasy Plant Kit e o equipamento Biosprint 96, conforme instruções do fabricante (Qiagen, 2012).

Quadro 1 – Lista do material genético utilizado neste estudo

5D 660 RR	BRS MT Pintado	CD 240 RR	Engopa 304	Msoy 7901
5D 690 RR	CAC-1	CD 241 RR	FT Abyara	MSoy 8001
5D 711 RR	Capinópolis	CD 242 RR	Ft-Estrela	NA 4990
5G 770 RR	CD 201	CD 243 RR	Ft-Guaíra	NA 5909 RR
5G 830 RR	CD 202	CD 244 RR	Fundacep 33	NA A 4725 RG
A 6001 RR	CD 204	CD 245 RR	Fundacep 38	NA A 6411 RR
A 8000	CD 205	CD 246	Fundacep 39	NA A 7321 RG
Anta	CD 206	CD 247 RR	Fundacep 53 RR	NK 3358
BMX Apollo RR	CD 206 RR	CD 248 RR	Fundacep 55 RR	NK 7059
BMX Ativa RR	CD 208	CD 249 RR STS	Fundacep 56 RR	Ocepar 14
BMX Energia RR	CD 213 RR	CD 250 RR STS	Fundacep 58 RR	Ocepar 3-Primavera
BMX Magna RR	CD 215	CD 251 RR	Fundacep 59 RR	P98Y11
BMX Potência RR	CD 216	CD 252	Fundacep 61 RR	P98Y51
BMX Titan RR	CD 217	CD 253	Fundacep 63 RR	P98Y70
Bragg	CD 218	CD 254 RR	Garimpo	R7
BRS 184	CD 219 RR	CD 2630 RR	IAS 5	Spring 5.3
BRS 185	CD 221	CD 2721 RR	Igra RA 518 RR	Tabarana
BRS 213	CD 224	CD 2737 RR	Igra RA 628 RR	TMG 103 RR
BRS 230	CD 225 RR	CD 2792 RR	Liderança	TMG 1066 RR
BRS 232	CD 226 RR	CD 2800	M 6009 RR	TMG 1067 RR
BRS 243 RR	CD 228	CD 2840	M 6707 RR	TMG 115 RR
BRS 245 RR	CD 229 RR	CD 2860	M 7211 RR	TMG 1161 RR
BRS 246 RR	CD 230 RR	CD 5807	M 7578 RR	TMG 123 RR
BRS 256 RR	CD 231 RR	CD 5969	M 7908 RR	TMG 4001 RR
BRS 258	CD 232	CD Fapa	M 9144 RR	TMG 7161 RR
BRS 262	CD 233 RR	Celeste	M7839 RR	Tropical R
BRS 268	CD 235 RR	Conquista	Matrinchá	Tucunaré
BRS 283	CD 236 RR	Dom Mario 7.0RR	Mercedes 70	USP 1
BRS 284	CD 237 RR	Embrapa 48	MSoy 2001	Vencedora
BRS Lambari	CD 238 RR	Embrapa 59	Msoy 6101	
BRS MT Crixás	CD 239 RR	Engopa 302	Msoy 7201	

## 2.2. Genotipagem de SNPs

A genotipagem foi realizada utilizando-se a plataforma Illumina *iScan* e o painel *Infinium iSelect HD Custom Genotyping BeadChips 6k* (Illumina Inc., San Diego, CA, EUA). Foram utilizados oito *BeadChips*, contendo 24 amostras cada, totalizando 192 amostras, das quais 169 foram utilizadas neste trabalho. Cada amostra foi avaliada com 6000 marcadores SNPs. A genotipagem dos SNPs foi realizada na empresa Deoxi Biotecnologia Ltda ®, em Araçatuba/SP. A genotipagem foi realizada em seis etapas descritas a seguir, conforme as instruções do fabricante.

Etapa 1 – amplificação: as amostras de DNA foram diluídas a 75 ng/mL, desnaturadas e neutralizadas com NaOH para posterior amplificação, submetidas a um processo de amplificação isotérmico a 36 °C e incubadas no equipamento por aproximadamente 16 horas. Essa etapa resulta em aumento na concentração de DNA centenas de vezes, comparada à reação de PCR (do inglês, *Polymerase Chain Reaction*), melhorando, assim, a qualidade dos resultados.

Etapa 2 – fragmentação: após 16 horas de incubação, prosseguiu-se com a fragmentação enzimática. Foi adicionado isopropanol no DNA e centrifugado a 4°C, descartando em seguida o isopropanol. O DNA foi ressuspensionado em tampão de hibridização. O produto de amplificação foi fragmentado por um processo enzimático do tipo *end-point*, isto é, os sítios de fragmentação são determinados por atributos específicos da sequência de DNA, o que evita a ocorrência de fragmentação excessiva.

Etapa 3 – hibridização: o DNA amplificado e fragmentado se anela (hibridiza) à sequência de 50 bases covalentemente ligadas à nano partícula *bead*, sendo que cada *bead* corresponde a um loco específico do DNA. O chip contendo as amostras foi incubado por 16 a 24 horas em forno com temperatura controlada de 47°C.

Etapa 4 – extensão de base única: Após a hibridização, foi realizada a extensão de uma única base nucleotídica da sequência, para incorporar os corantes que permitiram a determinação dos genótipos.

Etapa 5 – coloração: primeiramente, os chips foram lavados sem remover o DNA hibridizado. Uma sequência de reagentes e corantes foi aplicada nos 69 chips, possibilitando a emissão de fluorescência para a determinação dos diferentes genótipos. Todo o procedimento ocorreu em equipamento com umidade e

temperatura controladas. Após a coloração, os chips foram acondicionados a vácuo até completa secagem.

Etapa 6 – genotipagem: após a secagem, os chips foram levados para o equipamento Illumina iScan®. Um conjunto de laser foi utilizado para excitar os fluoróforos que estavam ligados aos *beads* na lâmina, permitindo que o sinal de fluorescência fosse captado e analisado. A informação obtida do SNP para cada genótipo testado é determinada pela intensidade de fluorescência emitida pela reação físico-química que ocorre no sítio de hibridização específico para cada SNP, que é captada por diferentes tipos de plataformas de detecção. As intensidades determinaram a formação de três possíveis grupos que representam os genótipos hipotéticos AA, AB e BB.

### 2.3. Genotipagem dos microssatélites

Na genotipagem com microssatélites foram utilizados 20 marcadores SSR ligados a genes de resistência a doenças e a QTLs relacionados a características de interesse agrônomo (Quadro 2), obtidos no Soybase ([www.soybase.org](http://www.soybase.org)).

As reações de PCR foram realizadas em microtubos de 0,2 mL, sendo o volume total da reação de 20 µL. A reação foi constituída de 30 ng de DNA, 3 mM de MgCl<sub>2</sub>, tampão 1X (2 mM de Tris e 5 mM de KCl), 250 µM de dNTP, 0,4 µM de cada *primer* (senso e antisenso) e uma unidade de Taq DNA Polimerase. As amplificações foram realizadas em termociclador programado para um ciclo a 94 °C, por 3 minutos; 35 ciclos a 94° C, por 30 segundos, 55 °C, por 30 segundos, e 72 °C, por 45 segundos; e um ciclo final de 72 °C, por 20 minutos.

A genotipagem das amostras foi realizada por eletroforese capilar em *multiplex* e as reações de PCR foram realizadas em *singleplex*. O *multiplex* foi montado pela combinação de quatro marcadores em uma única amostra, combinando 5 µL de cada produto da amplificação individual com quatro diferentes fluorescências (6FAM, PET, VIC e NED). Da mistura do *multiplex* foi retirado 1 µL e diluído em 49 µL de água ultrapura. Em seguida, 1 µL foi transferido para uma placa de PCR, adicionando uma mistura composta por 8,80 µL de formamida (Hi-di) e 0,20 µL do padrão de tamanho molecular (*size standard*) 500LIZ de coloração laranja (Life Technologies, 2013). Os fragmentos amplificados foram separados em capilares

Quadro 2- Marcadores microssatélites selecionados conforme a característica de interesse

Característica	Origem (variedade)	Marcador	G. L.	Distância (cM)	R <sup>2</sup>	Referência
Floração	IX132	Satt286	C2	Satt286 = 101,75 <sup>1</sup>	36,07%	TASMA et al., 2001.
Fotoperíodo	IX136	Satt286	C2	Satt286 = 101,75 <sup>1</sup>	20,77%	TASMA et al., 2001.
Lipoxigenase	Pureunkong	Sat_074	F	8,6		KIM et al., 2004.
Mancha olho-de-rã	Davis	Satt547	J	0,6		MIAN et al., 1999.
Maturação	IX136	Satt276	A1	17,16 <sup>1</sup>	9,83-12,13%	TASMA et al., 2001.
Maturação	IX136	Satt286	C2	Satt286 = 101,75 <sup>1</sup>	21,72-28,65%	TASMA et al., 2001.
Mofo branco da soja	Corsoy79 / Dassel / DSR173 / S19-90 / Vinton	Satt114	F	18,7 <sup>2</sup>		ARAHANA et al., 2001.
Mofo branco da soja	IA2053	Satt185-Satt263	E	QTL posição = 45cM	15,70%	GUO et al., 2008.
NCS raça 3	Essex	Satt309	G	4,53 <sup>1</sup>	42-97%	MEKSEM et al., 1999.
Necrose da Haste	BRS133	Sat_308	G	28,87		OLIVEIRA, 2009.
Nematóide de galha	PI595099	Satt114	F	63,68 <sup>1</sup>	46,00%	FUGANTI et al., 2004.
Nematóide de galha	PI96354	Satt492-Satt358	O	3,1	55,80%	LI et al., 2001.
Nematóide Reniforme	PI 437654	Sat_168	G	3,9 <sup>1</sup>	8,50%	HA et al., 2007.
Nematóide Reniforme	PI 437654	Satt513	L	106,37 <sup>1</sup>	18,10%	HA et al., 2007.
Peso do grão	Young X PI416937	Satt263	E	45,4 <sup>1</sup>	18,00%	FASOULA et al., 2004.
Peso do grão	Young X PI416937	Satt303	G	53,41 <sup>1</sup>	8,00%	FASOULA et al., 2004.
Podridão da raiz da soja	Clark	Satt152 ( <i>Rps1, Rps1b, Rps</i> )	N	0		DEMIRBAS et al., 2001.
Podridão parda da Haste	BSR101	Satt244	J	65,04 <sup>1</sup>		KLOS et al., 2000.
Tamanho da semente	A97-775006 <sup>1</sup> / S12-49 <sup>2</sup>	Sat_099	L	78,23 <sup>1</sup>	33,70%	HOECK et al., 2003.
Tamanho da semente	A97-775026 <sup>1</sup> / A96-492058 <sup>2</sup>	Satt006	L	92 <sup>1</sup>	32,50%	HOECK et al., 2003.
Teor de óleo	N87-984-16 X TN93-99	Satt420	O	49,71 <sup>2</sup>	15,00%	PANTHEE et al., 2005
Teor de óleo	N87-984-16 X TN93-99	Satt479	O	54,20 <sup>2</sup>	12,00%	PANTHEE et al., 2005
Teor de proteína	Young X PI416937	Satt398	L	30,58 <sup>1</sup>	8,00%	FASOULA et al., 2004.
Teor de proteína	N87-984-16 X TN93-99	Satt570	G	12,21 <sup>2</sup>	20,20%	PANTHEE et al., 2005
Vírus do mosaico da Soja	J05	Satt114	F	5,42		SHI et al., 2008.

<sup>1</sup>Posição do marcador com base no mapa gmcomposite2003 da soybase.org.

<sup>2</sup>Posição do marcador com base no mapa de Cregan et al. (1999).

<sup>3</sup>Posição do marcador com base no mapa de Song et al. (2004).

contendo polímero POP7 (Life Technologies, 2013), utilizando-se o sequenciador automático ABI3130xl, de acordo com a recomendação do fabricante. Os dados foram codificados com auxílio do programa Gene Mapper versão 4.0 (Life Technologies, 2013).

#### **2.4. Estimativa da diversidade genética**

As relações genéticas entre as cultivares foram avaliadas por meio de uma matriz de dissimilaridade, construída utilizando-se o complemento do índice de similaridade para dados codominantes e multialélicos de Schuster e Cruz (2003):

$$D_{sc} = 1 - \frac{AC}{AT}$$

Em que:  $D_{sc}$  é a dissimilaridade entre dois indivíduos, AC é o número de alelos comuns entre os indivíduos e AT é o número de alelos totais avaliados.

Como a soja é uma espécie diplóide, o número total de alelos avaliados em cada indivíduo é duas vezes o número de locos avaliados, pois cada indivíduo possui dois alelos em cada loco.

Com as estimativas de dissimilaridade descritas por Schuster e Cruz (2003), foi gerado um dendrograma com base no método hierárquico de agrupamento UPGMA (em inglês - Unweighted Pair-Group Mean Average). As cultivares também foram agrupadas, com base na matriz de dissimilaridade, utilizando-se o método de Tocher.

Todas as análises de diversidade genética foram realizadas utilizando o programa GENES (Cruz, 2008).

### 3. RESULTADOS E DISCUSSÃO

A genotipagem de 169 cultivares na plataforma *iScan* produziu resultados para 5353 marcadores SNPs (89% do total). Pela análise exploratória dos dados, foram retiradas as cultivares que apresentaram menos de 90% de dados válidos, e também os marcadores que apresentaram menos de 90% de dados válidos. Nesta seleção, foram excluídas da análise 16 cultivares (9,5%) e 415 marcadores (7,7% dos marcadores que amplificaram). Destes, 357 marcadores foram monomórficos entre as 153 cultivares restantes (6,6% dos marcadores válidos) e 58 marcadores foram eliminados por apresentarem menos de 90% de dados válidos (1% dos marcadores). A análise completa foi realizada, então, com 153 cultivares e 4938 marcadores SNPs.

Uma ampla cobertura genômica foi obtida pelo elevado número de marcadores e o número de marcadores por cromossomo teve uma média de 246,9, variando de 199 (cromossomo 1) a 320 (cromossomo 8) (Figura 1). A utilização de uma grande quantidade de marcadores permite que seja feita uma comparação entre as cultivares em várias regiões do genoma, aumentando a precisão da análise de diversidade genética.

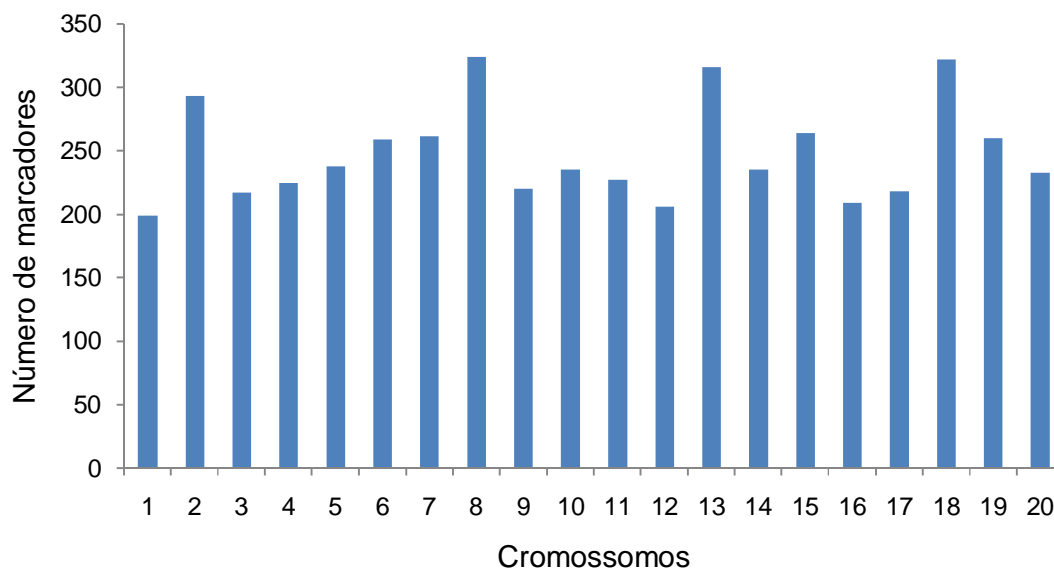


Figura 1 – Distribuição dos 4938 marcadores SNPs nos 20 cromossomos de soja.

Na análise com marcadores microssatélites, foram observados 124 alelos, variando entre 2 a 11 alelos por loco, com média de 6,2. Resultado semelhante foi encontrado por Narvel et al. (2000), que avaliaram a diversidade genética em 39 cultivares elite e 40 PI's, utilizando 74 marcadores microssatélites e observaram uma variação de 2 a 11 alelos por loco, com média de 5,4. Priolli et al. (2010), ao avaliar a diversidade genética utilizando 17 marcadores SSR em 168 cultivares brasileiras, observaram 91 alelos, variando de dois a oito alelos por loco, com média de 5,06.

Em soja, espera-se que todos os locos estejam em homozigose, devido ao modo de reprodução autógama e diversas gerações de autofecundação para obtenção de uma cultivar. Entretanto, todas as cultivares apresentaram locos heterozigotos. A porcentagem de heterozigosidade por cultivar variou de 3%(BRSMT Crixás, CD 205, P98Y70 e Celeste) a 41% (BMX Titan RR), com uma média de 8%. Sessenta e nove por cento das cultivares tiveram menos de 5% de heterozigotos e cerca de 18% das cultivares tiveram mais de 10% de heterozigosidade (Figura 2).

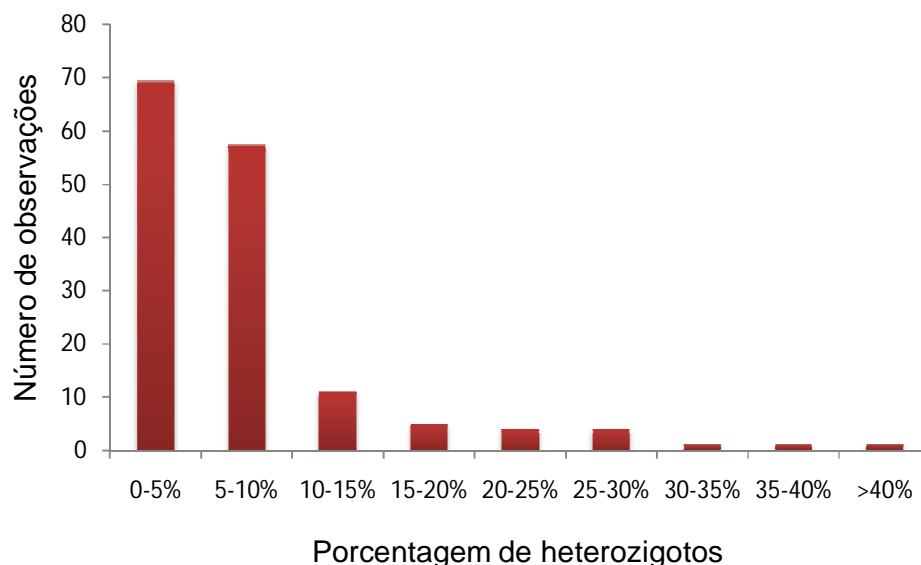


Figura 2 – Frequência de heterozigosidade observada em 153 cultivares de soja, utilizando 4938 marcadores SNPs e 20 SSR.

A heterozigosidade observada não significa a presença de plantas heterozigotas e sim a mistura de plantas homozigotas com genótipos diferentes, o que produz, no bulk de plantas utilizadas para a análise, a observação de dois alelos em um determinado loco. Por este motivo, é utilizado o termo heterozigosidade e

não heterozigose. A observação de heterozigosidade em cultivares de soja pode ser devida, principalmente, a uma mistura de linhas puras no processo de produção de sementes genéticas. Há ainda a possibilidade de mistura de cultivares na produção de sementes, embora esta mistura normalmente seja detectada nos procedimentos de controle de qualidade, na certificação de sementes. Outro fator que pode contribuir, embora muito pouco frequente, é a presença de mutação em determinados locos. Existe também a possibilidade de erro de genotipagem pela presença de alelos “fantasmas” no caso da avaliação com locos microssatélites, que podem indicar mais de um alelo por loco. Entretanto, os locos SSR representam apenas 0,04% do total de marcadores, não sendo suficiente para explicar o alto nível de heterozigosidade encontrado nas cultivares.

As estimativas de distâncias genéticas variaram de 0,0052 (P98Y11 e P98Y51) a 0,4216 (BMX ENERGIA e CD 206), com uma média de 0,32. Mais de 50% das distâncias genéticas se concentraram no intervalo de 0,31 a 0,35 (Figura 3), demonstrando que a grande maioria das cultivares diferem entre si em aproximadamente 1/3 do genoma. Menos de 6% das distâncias genéticas estão no intervalo de 0 a 0,25; 23,33% estão no intervalo de 0,26 a 0,30; 17,53% no intervalo de 0,36 a 0,40; e 0,64% das distâncias foram maiores que 0,41. Resultados diferentes foram encontrados por Priolli et al. (2010), que observaram uma média de diversidade genética de 0,58, variando de 0,07 a 0,826. A maioria das 168 cultivares avaliadas pelos autores apresentou uma diversidade genética acima de 0,43. Essa diferença nos resultados pode ser devida à amplitude do genoma avaliado. Enquanto Priolli et al. (2010) avaliaram 18 locos microssatélites, no presente trabalho, foram avaliados 20 marcadores microssatélites e 4938 SNPs, o que garante maior cobertura do genoma e maior confiabilidade dos resultados, trazendo-os mais próximos da realidade. Outro fator é o conjunto de diferentes cultivares avaliadas entre os dois trabalhos, o que resulta em diversidades genéticas diferentes.

Pela análise de agrupamento baseada no método hierárquico UPGMA (Figura 4) e com um ponto de corte no dendrograma estabelecido em uma análise paralela pelo método de Tocher (dados não mostrados), foi possível formar 14 grupos: um grupo com 29 cultivares, um grupo com 25 cultivares, três grupos com 20 cultivares, um grupo com 12 cultivares, um grupo com 10 cultivares, dois grupos



com cinco cultivares, dois grupos com duas cultivares e três grupos apresentando apenas uma cultivar (Quadro 3).

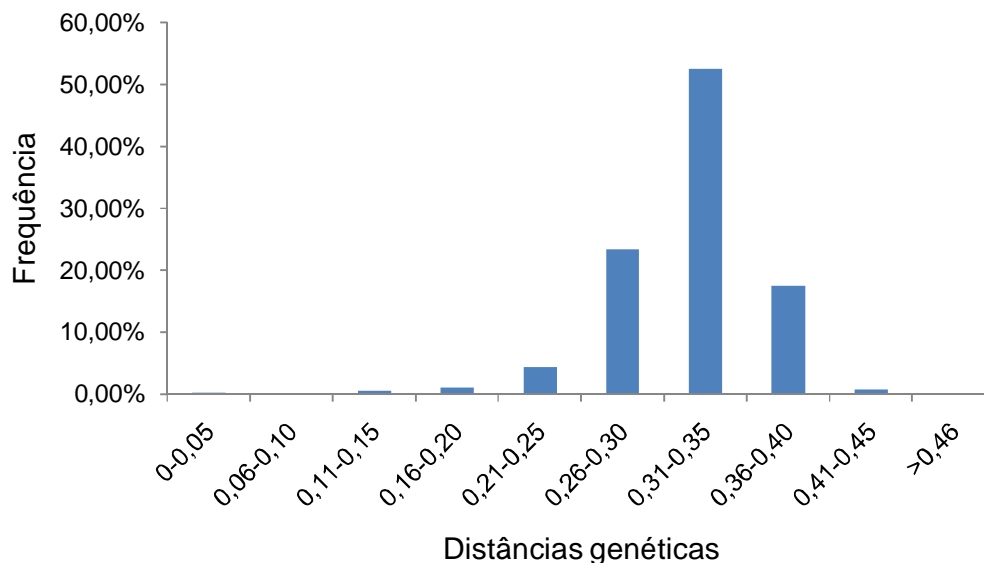


Figura 3 – Distribuição de frequência das distâncias genéticas em 153 cultivares de soja, utilizando 4938 marcadores SNPs e 20 marcadores microssatélites.

Observa-se, pela análise do dendrograma e no Quadro 3, que alguns grupos concentram cultivares do mesmo programa de melhoramento. As três cultivares pertencentes ao programa de melhoramento da Pioneer estão no grupo 1. A maioria das cultivares da Embrapa está no grupo 4. O grupo 10 é constituído, em sua maioria, por cultivares de origem do Germoplasma americano/argentino, introduzido recentemente no Brasil. Isso demonstra que este germoplasma mais recente é diferente do germoplasma tradicional do Brasil.

Algumas cultivares do mesmo programa de melhoramento apresentaram similaridade maior que 95%, tais como: P98Y51 e P98Y11, BRS MT Pintado e Tucunaré, BRS 230 e BRS 232, BRS 245 e BRS 246, CD 201 e CD 208, CD 218 e CD 238 RR, CD 206 e CD 206 RR, CD 217 e CD 237 RR, Fundacep 39 e Fundacep 59 RR, BMX Apollo RR e BMX Energia RR. As cultivares CD 201 e CD 208 são derivadas da cultivar Ocepar-4 Iguazu. A cultivar CD 238 RR deriva de retrocruzamentos da CD 218, assim como a CD 206 RR e CD 206; e CD 237 RR e CD 217. Isso explica a elevada similaridade dessas cultivares.

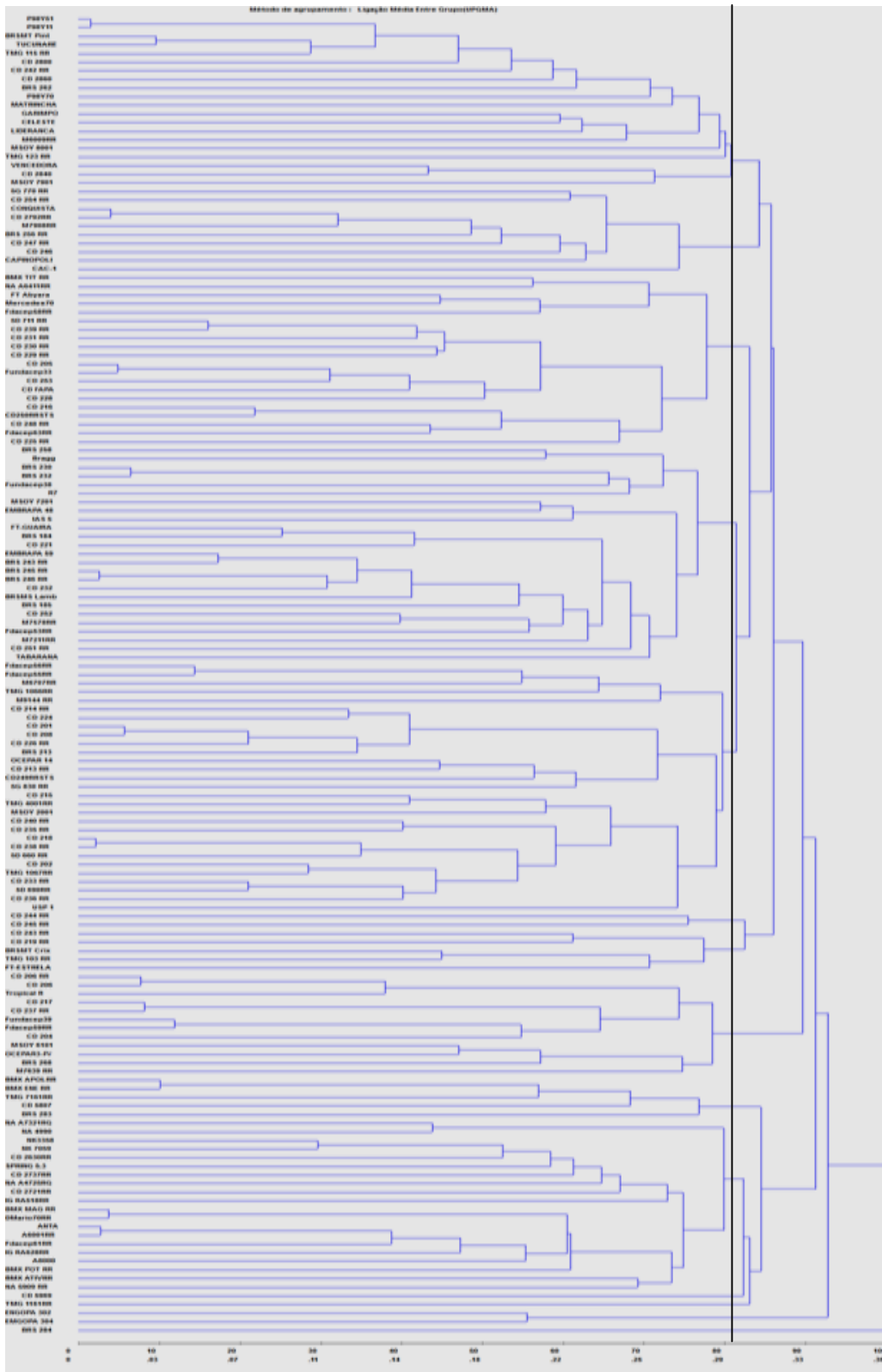


Figura 4 – Análise de agrupamento de 153 cultivares de soja pelo método UPGMA, a partir das distâncias genéticas geradas por 4938 marcadores SNPs e 20 SSR. A linha de corte na posição da distância 0,30 foi obtida a partir da análise de agrupamento de Tocher, comparando-se os grupos formados.

Quadro 3 – Agrupamento formado pelo dendrograma da Figura 4 em 153 cultivares de soja brasileiras utilizando 4938 marcadores SNPs e 20 SSR

Grupos		Cultivares			
		BRS MT			
1	P98Y51	P98Y11	Pintado	Tucunaré	TMG 115 RR
	CD 2800	CD 242 RR	CD 2860	BRS 262	P98Y70
	Matrinchã	Garimpo	Celeste	Liderança	M 6009 RR
	MSoy 8001	TMG 123 RR	Vencedora	CD 2840	Msoy 7901
2	5G 770 RR	CD 254 RR	Conquista	CD 2792 RR	M 7908 RR
	BRS 256 RR	CD 247 RR	CD 246	Capinópolis	CAC-1
3	BMX Titan				
	RR	NA A 6411 RR	FT Abyara	Mercedes 70	Fundacep 58 RR
	5D 711 RR	CD 239 RR	CD 231 RR	CD 230 RR	CD 229 RR
	CD 205	Fundacep 33	CD 253	CD Fapa 220	CD 228
		CD 250 RR		Fundacep 63	
4	CD 216	STS	CD 248 RR	RR	CD 225 RR
	BRS 258	Bragg	BRS 230	BRS 232	Fundacep 38
	R7	Msoy 7201	Embrapa 48	IAS 5	Ft-Guaíra
	BRS 184	CD 221	Embrapa 59	BRS 243 RR	BRS 245 RR
	BRS 246 RR	CD 232	BRS Lambari	BRS 185	CD 252
		Fundacep 53			
5	M 7578 RR	RR	M 7211 RR	CD 251 RR	Tabarana
	Fundacep 56	Fundacep 55		TMG 1066	
	RR	RR	M 6707 RR	RR	M 9144 RR
	CD 241 RR	CD 224	CD 201	CD 208	CD 226 RR
				CD 249 RR	
	BRS 213	Ocepar 14	CD 213 RR	STS	5G 830 RR
	CD 215	TMG 4001 RR	MSoy 2001	CD 240 RR	CD 235 RR
	CD 218	CD 238 RR	5D 660 RR	CD 202	TMG 1067 RR
CD 233 RR	5D 690 RR	CD 236 RR	USP 1		
6	CD 244 RR	CD 245 RR			
7	BRS MT				
	CD 243 RR	CD 219 RR	Crixás	TMG 103 RR	Ft-Estrela

Quadro 3, Cont.

8	CD 206 RR	CD 206	Tropical R	CD 217	CD 237 RR
		Fundacep 59			Ocepar 3-
	Fundacep 39	RR	CD 204	Msoy 6101	Primavera
	BRS 268	M7839 RR			
9	BMX Apollo	BMX Energia			
	RR	RR	TMG 7161 RR	CD 5807	BRS 283
10	A 7321 RG	NA 4990	NK 3358	NK 7059	CD 2630 RR
	Spring 5.3	CD 2737 RR	NA A 4725 RG	CD 2721 RR	RA 518 RR
	BMX Magna	Dom Mario			
	RR	7.0RR	Anta	A 6001 RR	Fundacep 61 RR
			BMX Potência	BMX Ativa	
	RA 628 RR	A 8000	RR	RR	NA 5909 RR
11	CD 5969				
12	TMG 1161				
	RR				
13	Engopa 302	Engopa 304			
14	BRS 284				

Outras cultivares se mostraram bastante similares entre programas de melhoramento: CD 205 e Fundacep 33, BMX Magna RR e Dom Mario 7.0 RR, Anta e A6001 RR. Priolli et al. (2004) avaliaram a diversidade genética entre programas de melhoramento de soja no Brasil utilizando doze marcadores microssatélites em 184 cultivares e obtiveram diferenças significativas apenas em cinco das quinze comparações realizadas entre os programas de melhoramento brasileiro. Além disso, constataram que havia maior variabilidade genética dentro dos programas de melhoramento do que entre eles. Isso indica a estreita base genética das cultivares brasileiras nos programas de melhoramento de soja, que pode ser comprovada pela baixa diversidade genética encontrada neste trabalho.

Grande parte das cultivares com a mesma região de adaptação para cultivo estão alocadas no mesmo grupo, como é o caso dos grupos 1, 2, 6, 7 e 13, formados, quase que exclusivamente, por cultivares adaptadas à região Central do Brasil. No grupo 3, 85% das cultivares são adaptadas à região Sul. Das 36 cultivares presentes nos grupos 8, 9 e 10, apenas uma é adaptada somente à região Central.

No grupo 4, cerca de 67% das cultivares são adaptadas à região Sul; 21% para a região Central e 12% de ampla adaptação. Aproximadamente 76% das cultivares do grupo 5 são adaptadas somente à região Sul. A elevada similaridade do agrupamento das cultivares com as regiões de cultivo novamente indica a utilização de poucos ancestrais para a obtenção de novas cultivares em programas de melhoramento brasileiros. Isso pode indicar maior vulnerabilidade das cultivares à doenças em determinadas regiões do Brasil e, portanto, um elevado parentesco por região pode facilitar a disseminação da doença e conseqüentemente aumentar os prejuízos da cultura.

Este trabalho avaliou pela primeira vez o Germoplasma da soja brasileira, utilizando uma ampla cobertura com marcadores SNPs e microssatélites. Foi observada baixa variabilidade entre as cultivares que, na sua maioria, compartilham 2/3 do genoma. A baixa variabilidade genética observada está em linha com as avaliações da base genética da soja brasileira. A estreita base genética, resultante do uso de poucos ancestrais, é responsável pela baixa diversidade do Germoplasma comercial de soja no Brasil.

As recentes introduções de cultivares originadas do Germoplasma americano/argentino podem contribuir para a ampliação da base genética e, conseqüentemente, da variabilidade das futuras cultivares. Isso ficou evidente nos resultados, uma vez que estas cultivares originárias do Germoplasma americano/argentino agruparam-se, em sua maioria, em um grupo único. Mesmo assim, para que a variabilidade genética e, conseqüentemente, as possibilidades de ganhos genéticos em diversas características possam aumentar, ainda é necessário introduzir novas fontes de Germoplasma, além de intensificar os programas de pré-melhoramento.

#### 4. CONCLUSÕES

a) A ampla cobertura do genoma com marcadores SNPs apresentou resultados de similaridade genética compatíveis com as informações da estreita base genética da soja no Brasil.

b) As cultivares brasileiras apresentam, em média, aproximadamente 70% de similaridade genética.

c) Cultivares da mesma região de adaptação de cultivo geralmente pertencem ao mesmo grupo de similaridade.

d) Para aumentar a variabilidade genética do germoplasma brasileiro existe a necessidade da introdução de cultivares de outros germoplasmas.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

ALBDELNOOR, 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. **Revista Brasileira de Genética**, 18:265-273, 1995.

ARAHANA, V.S.; GRAEF, G.L.; SPECHT, J.E.; STEADMAN, J.R.; ESKRIDGE, K.M. Identification of QTLs for resistance to *Sclerotinia sclerotiorum* in soybean. **Crop Science**, 41:180-188, 2001.

BONATO, A.L.V.; CALVO E.S.; GERALDI, I.O.; ARIAS, C.A.A. Genetic similarity among soybean (*Glycine max* (L) Merrill) cultivars released in Brazil using AFLP markers. **Genetics and Molecular Biology**, 29:692-704, 2006.

BONETTI, L.P. Distribuição da Soja no Mundo. In: MIYASAKA, S.; MEDINA J.C. (eds.). **A soja no Brasil**. Campinas: ITAL, 1981. cap. 1, Item 1, p. 1-6.

CREGAN, P.B.; JARVIK, T.; BUSH, A.L.; SHOEMAKER, R.C.; LARK, K.G.; KAHLER, A.L.; KAYA, N.; VANTOAI, T.T.; LOHNES, D.G.; CHUNG, J.; SPECHT, J.E. An integrated genetic linkage map of the soybean. **Crop Science**, 39:1464-1490, 1999.

CRUZ, C.D. **Programa Genes - Diversidade genética**. Viçosa: UFV, 2008. 278p.

DEMIRBAS, A.; RECTOR, B.G.; LOHNES, D.G.; FIORITTO, R.J.; GRAEF, G.L.; CREGAN, P.B.; SHOEMAKER, R.C.; SPECHT, J.E. Simple Sequence Repeat markers linked to the soybean Rps genes for *Phytophthora* resistance. **Crop Science**, 41:1220-1227, 2001.

FASOULA, V.A.; HARRIS, D.K.; BOERMA, R. Validation and Designation of Quantitative Trait Loci for Seed Protein, Seed Oil, and Seed Weight from Two Soybean Populations. **Crop Science**, 44:1218-1225, 2004.

FUGANTI R.; BENEVENTI, M.A.; SILVA, J.F.V.; ARIAS, C.A.A.; MARIN, S.R.R.; BINNECK, E.; NEPOMUCENO, A.L. Identificação de marcadores moleculares de

microssatélites para seleção de genótipos de soja resistentes a *Meloidogyne javanica*. **Nematologia Brasileira**, 28:125-130, 2004.

GUO, X.; WANG, D.; GORDON, S.G.; HELLIWELL, E.; SMITH, T.; BERRY, S.A.; MARTIN, S.K.S.; DORRANCE, A.E. Genetic Mapping of QTLs Underlying Partial Resistance to *Sclerotinia sclerotiorum* in Soybean PI 391589A and PI 391589B. **Crop Science**, 48:1129-1139, 2008.

HA, B.K.; ROBBINS, R.T.; HAN, F.; HUSSEY, R.S.; SOPER, J.F.; BOERMA, H.R. SSR Mapping and Confirmation of Soybean QTL from PI 437654 Conditioning Resistance to Reniform Nematode. **Crop Science**, 47:1336-1343, 2007.

HIROMOTO, D.M.; VELLO, N.A. The genetic base of Brazilian soybean (*Glycine max* (L.) Merrill) cultivars. **Revista Brasileira de Genética**, 9:295-306, 1986.

HOECK, J.A.; FEHR, W.R.; SHOEMAKER, R.C.; WELKE, G.A.; JOHNSON, S.L.; CIANZIO, S.R. Molecular marker analysis of seed size in soybean. **Crop Science**, 43:68-74, 2003.

KIM, M.Y.; HA, B.K.; JUN, T.H.; HWANG, E.Y.; VAN, K.; KUK, Y.I.; LEE, S.H. Single nucleotide polymorphism discovery and linkage mapping of lipoxygenase-2 gene (Lx2) in soybean. **Euphytica**, 135:169-177, 2004.

KLOS, K.L.M.; PAZ, M.M.; MAREK, L.F.; CREGAN, P.B.; SHOEMAKER, R.C. Molecular markers useful for detecting resistance to Brown Stem Rot in soybean. **Crop Science**, 40:1445-1452, 2000.

LI, Z.; JAKKULA, L.; HUSSEY, R.S.; TAMULONIS, J.P.; BOERMA, H.R. SSR mapping and confirmation of the QTL from PI96354 conditioning soybean resistance to southern root-knot nematode. **Theoretical and Applied Genetics**, 103:1167-1173, 2001.

LIFE TECHNOLOGIES. **3130xl Genetic Analyzer**. Disponível em: <http://www.lifetechnologies.com/order/catalog/product/3130XL>. Acesso em: 22, janeiro, 2015.



MEKSEM, K.; DOUBLER, T.W.; CHANCHAROENCHAI, K.; NJITI, V.N.; CHANG, S.J.C.; ARELLI, A.P.R.; CREGAN, P.E.; GRAY, L.E.; GIBSON, P.T.; LIGHTFOOT, D.A. Clustering among loci underlying soybean resistance to *Fusarium solani*, SDS and SCN in near-isogenic lines. **Theoretical and Applied Genetics**, 99:1131-1142, 1999.

MIAN, M.A.R.; WANG, T.; PHILLIPS, D.V.; ALVERNAZ, J.; BOERMA, H.R. Molecular mapping of the Rcs3 gene for resistance to Frogeye Leaf Spot in soybean. **Crop Science**, 39:687-1691, 1999.

NARVEL, J.M.; FEHR, W.R.; CHU, W.C.; GRANT, D.; SHOEMAKER, R.C. Simple sequence repeat diversity among soybean plant introductions and elite genotypes. **Crop Science**, 40:1452-1458, 2000.

OLIVEIRA, M.A.R. **Mapeamento de genes de resistência ao vírus da necrose da haste utilizando marcadores moleculares microssatélites**. Londrina: Universidade Estadual de Londrina, 2008. 69p. Tese (Doutorado em Agronomia).

PANTHEE, D.R.; PANTALONE, V.R.; WEST, D.R.; SAXTON, A.M.; SAMS, C.E. Quantitative Trait Loci for seed protein and oil concentration, and seed size in soybean. **Crop Science**, 45:2015-2022, 2005.

PRIOLLI, R.H.G.; MENDES-JUNIOR, C.T.; ARANTES, N.E.; CONTEL, E.P.B. Characterization of Brazilian soybean cultivars using microsatellite markers. **Genetics and Molecular Biology**, 25:185-193, 2002.

PRIOLLI, R.H.G.; MENDES-JUNIOR, C.T.; SOUSA, S.M.B.; SOUSA, N.E.A.; CONTEL, E.B.P. Diversidade genética da soja entre períodos e entre programas de melhoramento no Brasil. **Pesquisa Agropecuária Brasileira**, 39:967-975, 2004.

PRIOLLI, R.H.G.; PINHEIRO, J.B.; ZUCCHI, M.I.; BAJAY, M.M.; VELLO, N.A. Genetic diversity among Brazilian soybean cultivars based on SSR loci and pedigree data. **Brazilian Archives of Biology and Technology**, 53:519-531, 2010.

PROGRAMA GENES. **Aplicativo computacional em genética e estatística**. Versão 2013.5.1. Disponível em: [www.ufv.br/dbg/genes/genes.htm](http://www.ufv.br/dbg/genes/genes.htm). Acesso em: 10, fevereiro, 2014.

QIAGEN. DNeasy Plant Kit. 2012. Disponível em: <http://www.qiagen.com/br/resources/resourcedetail?id=6b9bcd96-d7d4-48a1-9838-58dbfb0e57d0&lang=en>. Acesso em: 22, janeiro, 2015.

SCHUSTER, I.; CRUZ, C.D. Similaridade genética obtida por um coeficiente de coincidência simples para dados codominantes e multi-alélicos. In: 49º CONGRESSO BRASILEIRO DE GENÉTICA. Águas de Lindóia-SP, 2003. **Resumos...** Ribeirão Preto: SBG, 2003. p. 689.

SHI, A.; CHEN, P.; ZHENG, C.; HOU, A.; ZHANG, B. A PCR-based marker for the Rsv1 locus conferring resistance to Soybean Mosaic Virus. **Crop Science**, 48:262-268, 2008.

SONG, Q.J.; MAREK, L.F.; SHOEMAKER, R.C.; LARK, K.G.; CONCIBIDO, V.C.; DELANNAY, X.; SPECHT, J.E.; CREGAN, P.B. A new integrated genetic linkage map of the soybean. **Theoretical and Applied Genetics**, 109:122-128, 2004.

TASMA, I.M.; LORENZEN, L.L.; GREEN, D.E.; SHOEMAKER, R.C. Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. **Molecular Breeding**, 8:25-35, 2001.

VAN ESBROECK, G.; BOWMAN, D.T. Cotton germplasm diversity and its importance to cultivar development. **The Journal of Cotton Science**, 2:121-129, 1998.

WYSMIERSKI, P.T.; VELLO, N.A. The genetic base of Brazilian soybean cultivars: evolution over time and breeding implications. **Genetics and Molecular Biology**, 36:547-555, 2013.

YAMANAKA, N.; SATO, H.; YANG, Z.; XU, D.H.; CATELLI, L.L.; BINNECK, E.; ARIAS, C.A.A.; ABDELNOOR, R.V.; NEPOMUCENO, A.L. Genetic relationships between Chinese, Japanese, and Brazilian soybean gene pools revealed by simple sequence repeat (SSR) markers. **Genetics and Molecular Biology**, 30:85-88, 2007.

## **CAPÍTULO 2**

### **Desequilíbrio de ligação em soja tropical utilizando análise ampla do genoma com marcadores SNPs**

#### **1. INTRODUÇÃO**

A Seleção Assistida por Marcadores (SAM) é uma técnica biotecnológica que permite a seleção direta ou indireta de genótipos de interesse, utilizando marcadores moleculares ligados a genes que expressam a característica em seleção. Esta técnica vem sendo amplamente utilizada em programas de melhoramento por apresentar diversas vantagens em relação à seleção fenotípica: (1) a seleção com base em marcadores moleculares não sofre interferências ambientais; (2) pode ser realizada em qualquer estágio da planta; (3) avalia características difíceis de serem avaliadas em campo (características com baixa herdabilidade); (4) acelera a recuperação de genótipos recorrentes e fixa alelos recessivos em programas de retrocruzamento; (5) acelera a piramidação de alelos de resistência, de características qualitativas ou de vários QTLs (Schuster et al., 2001; Alzate-Marin et al., 2005; Rodrigues et al., 2010; Schuster, 2011).

O sucesso da seleção assistida por marcadores depende de vários fatores: (1) um mapa genético com marcadores moleculares ligados a genes que controlam características de interesse agrônomico; (2) uma associação significativa entre os marcadores e os genes ou QTLs; (3) baixa recombinação entre o marcador associado à característica e ao restante do genoma; (4) a possibilidade de avaliar um grande número de plantas em pouco tempo a um custo razoável (Holland, 2004; Schuster, 2011).

Basicamente, duas estratégias podem ser utilizadas para o mapeamento: mapeamento por recombinação e por associação. O mapeamento por recombinação utiliza cruzamentos biparentais de indivíduos contrastantes para a obtenção do mapa, diferentemente do mapeamento por associação, no qual um conjunto de cultivares/linhagens é utilizado para realizar o mapeamento (Carneiro e Vieira, 2002; Mackay e Powell, 2007). Esta última metodologia é mais eficiente para o uso da SAM, pois a variabilidade genética é representada por diversos haplótipos e não se

limita a apenas dois genitores. Além disso, a associação entre o marcador e a região que controla a característica de interesse é elevada pelo fato dessa metodologia se basear na detecção de blocos de desequilíbrio de ligação, ou seja, o marcador e a região que controla a característica estão em desequilíbrio de ligação sendo herdados juntos (Ruane e Sonnino, 2007; Sorkheh et al., 2008).

Como os blocos de desequilíbrio de ligação são pequenos, somente é possível detectá-los por meio da saturação do genoma com marcadores moleculares (Mackay e Powell, 2007, Xie et al., 2008). O advento dos marcadores SNPs e tecnologias de genotipagem de alto rendimento tornaram possíveis a avaliação do genoma de diversas espécies com milhares de marcadores, cobrindo grande parte do genoma e possibilitando a detecção destes blocos e, conseqüentemente, a obtenção de mapas de associação. A análise do desequilíbrio de ligação em plantas pode elucidar a estrutura genômica das plantas e também facilitar o uso da SAM e a clonagem posicional de genes para características de difícil seleção (Sorkheh et al., 2008). Diversos trabalhos vêm sendo realizados com o objetivo de analisar o desequilíbrio de ligação em plantas, como em *Arabidopsis* (Nordborg et al., 2002), alface (Simko et al., 2009), milho (Xie et al., 2008), soja (Zhu et al., 2003) e tomate (Labate et al., 2009).

Este trabalho teve como objetivo analisar o desequilíbrio de ligação em cultivares brasileiras de soja utilizando marcadores SNPs e genotipagem de alto rendimento. Trata-se do primeiro trabalho de desequilíbrio de ligação em soja tropical.

## 2. MATERIAL E MÉTODOS

### 2.1. Material genético

Foram utilizadas 169 cultivares/linhagens de soja em uso comercial no Brasil, incluindo algumas que foram importantes no passado, como genitores nos programas de melhoramento. As sementes foram adquiridas dos obtentores das cultivares, de produtores de sementes que produzem sementes certificadas, e do banco de Germoplasma de soja da Coodetec. As cultivares foram cultivadas em parcelas experimentais, a campo, no verão 2011/12, na Coodetec, em Cascavel, PR. Foram coletados discos foliares de 10 plantas de cada cultivar, diretamente em tubos de coleta que foram imediatamente congelados em nitrogênio líquido. As amostras foram armazenadas em ultrafreezer, a  $-80^{\circ}\text{C}$ , até serem utilizadas para extração de DNA.

Para extração de DNA, foi utilizado o kit DNAeasy Plant Kit, da Qiagene, e o equipamento Biosprint 96, conforme instruções do fabricante.

### 2.2. Genotipagem dos marcadores SNPs

Foi utilizada a plataforma Illumina *iScan* e o painel *Infinium iSelect HD Custom Genotyping BeadChips 6k* (Illumina Inc., San Diego, CA, EUA). Foram customizados oito *BeadChips*, contendo 24 amostras cada, totalizando 192 amostras, das quais 169 foram utilizadas neste trabalho. Cada amostra foi avaliada com 6000 marcadores SNPs. A genotipagem dos SNPs foi realizada na empresa Deoxi Biotecnologia Ltda ®, em Araçatuba - SP.

### 2.3. Análise dos blocos de desequilíbrio de ligação

Todas as análises de estruturação dos blocos de desequilíbrio de ligação foram realizadas pelo software SNP & Variation Suite (SVS), versão 7.0, da empresa Golden Helix, disponível em [http://www.goldenhelix.com/SNP\\_Variation](http://www.goldenhelix.com/SNP_Variation).

Os dados gerados pela plataforma da Illumina foram filtrados no software SVS, removendo os marcadores que apresentaram menos de 90% de dados validados, SNPs monomórficos e cultivares com mais de 10% de dados perdidos.

Para a obtenção dos blocos de desequilíbrio de ligação, foram utilizados dois estimadores conjuntamente. O primeiro, chamado de Método de Haplótipo Composto (*Composite Haplotype Method – CHM*), é baseado no coeficiente genotípico de desequilíbrio de ligação ( $\Delta_{AB}$ ), de Weir (1996). A estimativa de  $\Delta_{AB}$  envolve o cálculo da frequência entre dois locos. No caso de dois locos bi- alélicos, essas frequências são estimadas por:

$$n_{AB} = 2n_{AABB} + n_{AABb} + n_{AaBB} + \frac{n_{AaBb}}{2}$$

Em que:  $n_{AaBb}$  é, por exemplo, o número de indivíduos com genótipo Aa/Bb, e  $n$  é o número total de amostras. O desequilíbrio composto é definido como a soma de componentes entre e intragaméticos:

$$\Delta_{AB} = D_{AB} + D_{A/B} = P_{AB} + P_{A/B} - 2p_A p_B$$

Sob análise aleatória de locos (equilíbrio de ligação),  $P_{A/B} = p_A p_B$ , e então  $\Delta_{AB}$  é uma estimativa tendenciosa do parâmetro de desequilíbrio de ligação,  $D_{AB}$ . No método CHM, não é necessário separar componentes entre e intragaméticos e o termo é dado por:

$$\rho_{AB} = \frac{P_{AB} + P_{A/B}}{2}$$

O desequilíbrio de ligação foi estimado também pelo quadrado da correlação das frequências alélicas ( $r^2$ ) (Weir, 1996);

$$r^2 = \chi^2 / n$$

$$\text{sendo: } \chi^2 = [n(n_{N1}n_{D2} - n_{N2}n_{D1})^2] / (n_1 n_2 n_N n_D)$$

Em que:

$n$  = número total de indivíduos;

$n_{N1}$  = número de indivíduos com alelo 1 no loco 1 e alelo 1 no loco 2;

$n_{D2}$  = número de indivíduos com alelo 1 no loco 1 e alelo 2 no loco 2;

$n_{N2}$  = número de indivíduos com alelo 2 no loco 1 e alelo 1 no loco 2;

$n_{D1}$  = número de indivíduos com alelo 2 no loco 1 e alelo 2 no loco 2;

$n_1$  = número total de indivíduos com alelo 1 no loco 1;

$n_2$  = número total de indivíduos com alelo 2 no loco 1;

$n_N$  = número total de indivíduos com alelo 1 no loco 2;

$n_D$  = número total de indivíduos com alelo 2 no loco 2.

A estimativa do desequilíbrio de ligação a partir de  $r^2$  pode ter vantagens sobre  $D_{AB}$ , uma vez que, na estimativa a partir de  $r^2$ , leva-se em conta a diferença na frequência dos alelos entre dois locos, enquanto a estimativa a partir de  $D_{AB}$  considera apenas a recombinação.

Os blocos de desequilíbrio foram extraídos manualmente a partir do gráfico gerado pelo software SVS, incluindo somente os blocos com desequilíbrio próximo de um. Estes blocos aparecem unidos no gráfico pela coloração vermelha, indicando o desequilíbrio de ligação naquela região. Quanto mais forte a tonalidade do vermelho, maior o desequilíbrio. Blocos formados por vermelho claro foram retirados da análise, bem como os blocos azuis, que representam equilíbrio de ligação.

### 3. RESULTADOS E DISCUSSÃO

Na genotipagem com o painel *Infinium iSelect HD Custom Genotyping BeadChips 6k* (Illumina Inc., San Diego, CA, EUA), foram obtidos genótipos para 5353 marcadores. Foi realizada uma filtragem destes marcadores, dos quais 391 foram retirados da análise, sendo 333 que foram monomórficos e 58 marcadores que apresentaram menos de 90% de dados validados. Das 169 cultivares analisadas, 153 cultivares foram mantidas e 16 cultivares foram retiradas por terem mais de 10% de dados perdidos.

O número de marcadores por cromossomo variou de 199 (cromossomo 1) a 320 (cromossomo 8), com média de 246,9 marcadores por cromossomo (Quadro 1). Para cada cromossomo, foi calculada a relação entre o número de SNPs e o tamanho em cM de cada cromossomo. Em média, existe um marcador SNP a cada 0,48 cM, variando de 0,33 cM/SNP (cromossomo 4) a 0,60 cM/SNP (cromossomo 17), demonstrando a ampla cobertura genômica do experimento.

Quadro 1 - Distribuição dos marcadores SNPs e dos blocos de desequilíbrio de ligação nos 20 cromossomos de soja

Cromossomo <sup>1</sup>	GL <sup>1</sup>	Tamanho (cM) <sup>1</sup>	Número de SNPs	cM/SNP	Blocos em deseq.
1	D1a	109,318	199	0,549337	25
2	D1b	143,608	293	0,49013	31
3	N	106,12	217	0,489032	25
4	C1	75,06	225	0,3336	27
5	A1	96,47	236	0,408771	30
6	C2	147,5	257	0,57393	28
7	M	134	260	0,515385	25
8	A2	156,882	320	0,490256	28
9	K	102,956	220	0,467982	26
10	O	139,36	235	0,593021	26
11	B1	135,09	227	0,59511	20
12	H	120,18	202	0,59495	21
13	F	144,67	316	0,457816	41
14	B2	106,11	235	0,451532	32
15	E	104,43	263	0,397072	41
16	J	91,619	209	0,438368	20
17	D2	128,4	213	0,602817	29
18	G	110,91	322	0,344441	36
19	L	111,59	260	0,429192	34
20	I	124,34	229	0,542969	22
<b>TOTAL</b>		<b>2388,613</b>	<b>4938</b>	<b>0,488286</b>	<b>567</b>

<sup>1</sup>Fonte: Soybase (www.soybase.org).



O elevado número de SNPs distribuídos ao longo do genoma da soja permitiu que fossem identificados 567 blocos em desequilíbrio de ligação (Figura 1). A média do número de blocos por cromossomo foi de 28,35, variando de 20 (cromossomos 11 e 16) a 41 (cromossomos 13 e 15) (Quadro 1). Não houve relação entre a diminuição da distância entre os marcadores SNPs e o aumento no número de blocos de desequilíbrio de ligação, demonstrando que estes blocos estão de forma aleatória no genoma. Entretanto, somente um número elevado de marcadores torna possível a detecção de pequenos blocos em desequilíbrio.

A quantidade de SNPs em desequilíbrio de ligação em cada bloco variou de 2 a 15, com média de 2,69 SNPs por bloco. Entre os blocos contendo SNPs em desequilíbrio de ligação, 64,9% tinham apenas dois marcadores e menos de 2% dos blocos continham sete ou mais marcadores (Figura 2). Isso indica a grande quantidade de marcadores SNPs necessários para identificar blocos de desequilíbrio de ligação em soja. Para um mapeamento por associação eficiente, a quantidade de marcadores deve ser próxima ou superior à quantidade utilizada neste trabalho.

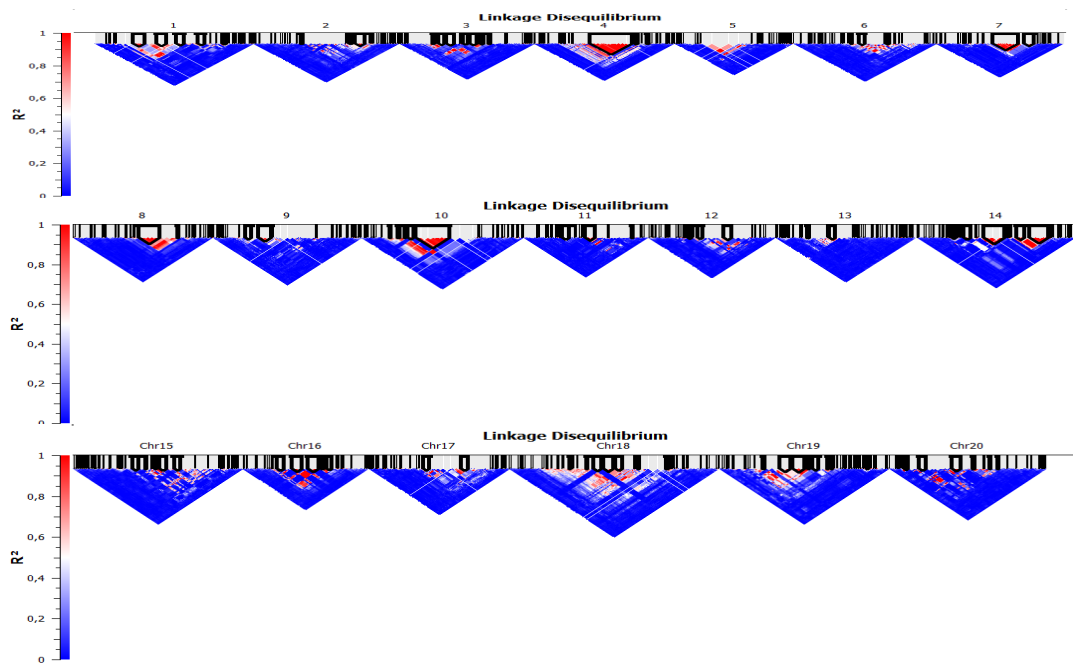


Figura 1 – Distribuição dos blocos de desequilíbrio de ligação ao longo dos 20 cromossomos de soja. O quadrado da correlação das frequências alélicas ( $R^2$ ) está representado por uma escala de cores. A coloração vermelha representa regiões de desequilíbrio de ligação ( $R^2 = 1$ ) e a coloração azul regiões de equilíbrio ( $R^2 = 0$ ). As marcações em preto nos cromossomos representam blocos de desequilíbrio de ligação.

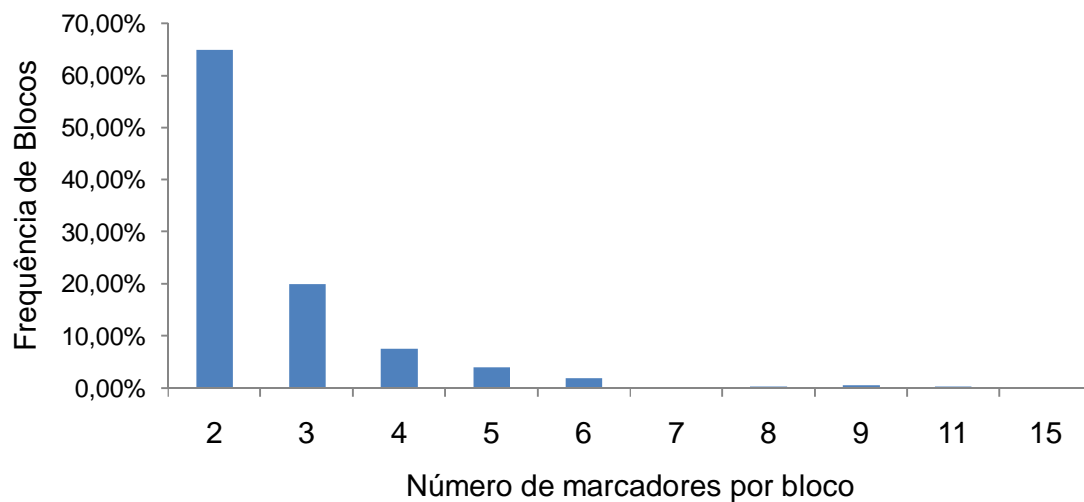


Figura 2 – Frequência do número de marcadores SNPs em desequilíbrio de ligação por bloco.

A quantidade de SNPs por bloco foi proporcional ao tamanho dos blocos. A média dos tamanhos dos blocos foi de 410,68 kb, variando de 1,173 a 13.692,9 kb (ambos os blocos contidos no cromossomo 4) (Figura 3).

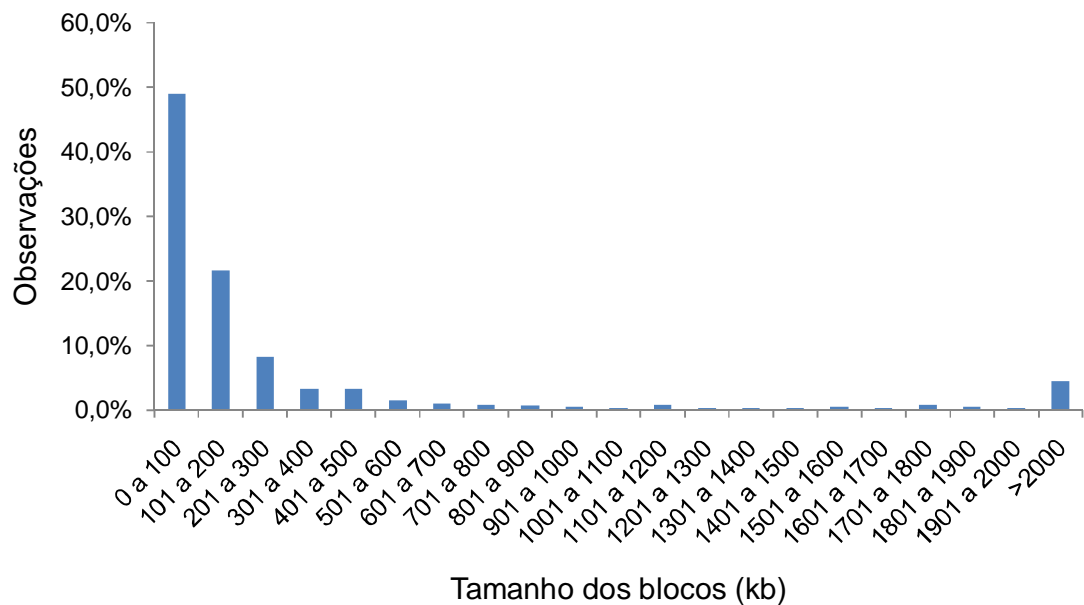


Figura 3 – Distribuição da frequência do tamanho dos blocos de desequilíbrio de ligação.

Pode-se observar, na Figura 3, que mais de 70% dos blocos de desequilíbrio de ligação possuem tamanhos menores que 200 kb. Hyten et al. (2007) avaliaram o desequilíbrio de ligação em soja, a partir de três regiões de sequências contíguas, utilizando marcadores STS (*Sequence Tagged Site*). Estes autores analisaram quatro tipos de populações: *Glycine soja*, Landraces, Ancestrais Norte-americanos e cultivares Elite. O tamanho dos blocos de desequilíbrio de ligação nestas três regiões variou conforme a população. O tamanho dos blocos da população *G. soja* variou de 25 kb a menos de 1 kb, com uma média de 4,8 kb por bloco. A população Landraces teve uma média de 21,5 kb por bloco e os blocos variaram de 186 kb a menos de 1 kb. O maior bloco da população de ancestrais norte-americanos foi de 89 kb, com média de 34,9 Kb por bloco. A maior média encontrada entre as populações foi no grupo das cultivares Elite, no qual a média de tamanho dos blocos foi de 80,1, mais que o dobro da população de ancestrais norte-americanos. Esses autores atribuíram a maior média à pressão de seleção dentro ou próximo a essas regiões avaliadas, ao efeito gargalo e ao fotoperíodo, no qual foi observado que as cultivares deste grupo criaram subdivisões entre maturação precoce e tardia.

Os resultados encontrados por Hyten et al. (2007) e no presente trabalho possuem grandes diferenças. Enquanto o trabalho daqueles autores avaliou regiões específicas do genoma da soja com tamanho médio de 474,33 kb, com 74 marcadores STS (média de um marcador a cada 6,4 kb), o presente trabalho avaliou todo o genoma da soja, com 4938 marcadores SNPs. Segundo Arumuganathan e Earle (1991), o genoma da soja possui cerca de 1,1 a 1,15 gb, o que significa que neste trabalho foi utilizado em média um marcador a cada 222 kb. Além disso, no presente trabalho, foi verificado que existe variação entre regiões do genoma quanto ao número de blocos e ao tamanho deles, o que significa que, no trabalho de Hyten et al. (2007), pode ter sido analisada uma parte do genoma que contém blocos menores. Outra diferença é em relação aos diferentes pools gênicos estudados. Conforme Yamanaka et al. (2007), o pool gênico brasileiro é diferente do chinês e do japonês, os quais serviram como base para duas populações avaliadas no trabalho de Hyten et al. (2007). As outras duas populações derivaram de cultivares norte-americanas, as quais podem apresentar semelhanças de pool gênico com as cultivares brasileiras, pelo fato de as cultivares brasileiras de soja derivarem de poucos ancestrais norte-americanos (Hiromoto e Vello, 1986).

A soma do tamanho dos blocos em desequilíbrio de ligação foi de 226.341,72 kb, o que representa 20% do genoma da soja, considerando que o genoma apresenta 1,1 gb. No trabalho de Hyten et al. (2007), essa proporção de blocos em desequilíbrio com o tamanho do genoma avaliado variou de 18% (população *G. soja*) a 51% (cultivares Elite). Isso demonstra que diferentes pools gênicos possuem diferentes níveis de queda de desequilíbrio de ligação, sendo que, quanto maior a queda no desequilíbrio de ligação, menor o tamanho dos blocos. A queda do desequilíbrio de ligação é menor em espécies autógamas do que em alógamas, pois nas autógamas a recombinação é menos efetiva, devido à elevada homozigose das plantas. Além da taxa efetiva de recombinação, outros fatores podem afetar a queda de desequilíbrio de ligação, como a domesticação, seleção, subdivisão populacional, estratificação da população e o efeito gargalo (Flint-Garcia et al., 2003).

Este é o primeiro trabalho de avaliação do genoma estrutural de soja tropical, utilizando análise de desequilíbrio de ligação de marcadores SNPs em genotipagem de larga escala. Utilizando esta estratégia, foi possível apresentar como estão distribuídos os blocos de desequilíbrio de ligação nos 20 cromossomos de soja, considerando o pool gênico de soja tropical. Este resultado pode servir como base para futuros estudos de mapeamento por associação para características de interesse agrônômico, tanto na utilização de SNPs quanto no tamanho dos blocos possíveis de serem detectados, desde blocos menores ~1 kb até blocos maiores de ~13.000 kb. Esta informação pode ser útil para determinar o número de marcadores a serem utilizados na análise e assim garantir um mapeamento de maior resolução e conseqüentemente uma maior eficiência na utilização da SAM.

#### 4. CONCLUSÕES

a) As cultivares de soja tropicais apresentam grande variação no tamanho dos blocos de desequilíbrio de ligação, tendo, em média, blocos de 410,68 kb, com variação de aproximadamente 1kb até mais de 13kb.

b) Setenta por cento dos blocos observados em desequilíbrio de ligação apresentam tamanho menor do que 200kb.

c) O germoplasma de soja tropical avaliado apresentou, para o conjunto de marcadores SNPs avaliado, uma média de 28,35 blocos em desequilíbrio de ligação por cromossomo, variando de 20 a 41 blocos.

d) O pool gênico avaliado possui aproximadamente 20% do genoma com blocos em desequilíbrio de ligação no conjunto de marcadores SNPs utilizados.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

- ALZATE-MARIN, A.L.; CERVIGNI, G.D.L.; MOREIRA, M.A.; BARROS, E.G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijão e soja. **Fitopatologia Brasileira**, 30:333-342, 2005.
- ARUMUGANATHAN, K.; EARLE, E.D. Estimation of nuclear DNA content of plants by flow cytometry. **Plant Molecular Biology Reporter**, 9:229-241, 1991.
- CARNEIRO, M.S.; VIEIRA, M.L.C. Mapas genéticos em plantas. **Bragantia**, 61:89-100, 2002.
- FLINT-GARCIA, S.A.; THORNSBERRY, J.M.; BUCKLER IV, E.S. Structure of linkage disequilibrium in plants. **Annual Review of Plant Biology**, 54:357-374, 2003.
- GOLDEN HELIX. **SNP & variation suite manual**. Disponível em: <http://doc.goldenhelix.com/SVS/latest/>. Acesso em: 25, janeiro, 2014.
- HIROMOTO, D.M.; VELLO, N.A. The genetic base of Brazilian soybean (*Glycine max* (L.) Merrill) cultivars. **Revista Brasileira de Genética**, 9:295-306, 1986.
- HOLLAND, J. Implementation of molecular markers for quantitative traits in breeding programs – change and opportunities. In: 4<sup>th</sup> INTERNATIONAL CROP SCIENCE CONGRESS - New directions for a diverse planet, Brisbane, 2004. **Proceedings...** Gosford: The Regional Institute Ltd, 2004.p.1-13.
- HYTEN, D.L.; CHOI, I.Y.; SONG, Q.; SHOEMAKER, R.C.; NELSON, R.L.; COSTA, J.M.; SPECHT, J.E.; CREGAN, P.B. Highly variable patterns of linkage disequilibrium in multiple soybean populations. **Genetics**, 175:1937-1944, 2007.
- LABATE, J.A.; ROBERTSON, L.D.; BALDO, A.M. Multilocus sequence data reveal extensive departures from equilibrium in domesticated tomato (*Solanum lycopersicum* L.). **Heredity**, 103:257-267, 2009.
- MACKAY, I.; POWELL, W. Methods for linkage disequilibrium mapping in crops. **Trends in Plant Science**, 12:57-63, 2007.

NORDBORG, M.; BOREVITZ, J.O.; BERGELSON, J.; BERRY, C.C.; CHORY, J.; HAGENBLAD, J.; KREITMAN, M.; MALOOF, J.N.; NOYES, T.; OEFNER, P.J.; STAHL, E.A.; WEIGEL, D. The extent of linkage disequilibrium in *Arabidopsis thaliana*. **Nature Genetics**, 30:190-193, 2002.

RODRIGUES, J.I.S.; MIRANDA, F.D.; FERREIRA, A.; BORGES, L.L.; FERREIRA, M.S.S.; GOOD-GOD, P.I.V.; PIOVESAN, N.D.; BARROS, E.G.; CRUZ, C.D.; MOREIRA, M.A. Mapeamento de QTL para conteúdos de proteína e óleo em soja. **Pesquisa Agropecuária Brasileira**, 45:472-480, 2010.

RUANE, J.; SONNINO, A. Marker-assisted selection as a tool for genetic improvement of crops, livestock, forestry and fish in developing countries: an overview of issues. In: GUIMARÃES, E.P.; RUANE, J.; SCHERF, B.D.; SONNINO, A.; DARGIE, J.D. (eds.). **Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish**. Roma: Food and Agriculture Organization of United Nations (FAO), p. 03-14, 2007.

SCHUSTER, I. Marker-assisted selection for quantitative traits. **Crop Breeding and Applied Biotechnology**, S1:50-55, 2011.

SCHUSTER, I.; ABDELNOOR, R.V.; MARIN, S.R.R.; CARVALHO, V.P.; KIIHL, R.A.S.; SILVA, J.F.V.; SEDIYAMA, C.S.; BARROS, E.G.; MOREIRA, M.A. Identification of a new major QTL associated with resistance to soybean cyst nematode (*Heterodera glycines*). **Theoretical and Applied Genetics**, 102:91-96, 2001.

SIMKO, I.; PECHENICK, D.A.; McHALE, L.K.; TRUCO, M.J.; OCHOA, O.E.; MICHELMORE, R.W.; SCHEFFLER, B.E. Association mapping and marker-assisted selection of the lettuce dieback resistance gene *Tvr1*. **BMC Plant Biology**, 9:1-16, 2009.

SORKHEH, K.; MALYSHEVA-OTTO, L.V.; WIRTHENSOHN, M.G.; TARKESH-ESFAHANI, S.; MARTÍNEZ-GÓMEZ, P. Linkage disequilibrium, genetic association mapping and gene localization in crop plants. **Genetics and Molecular Biology**, 31:805-814, 2008.

WEIR, B. **Genetics data analysis II.** Sunderland, Massachusetts: Sinauer Associates, Inc., 1996. 376p.

XIE, C.; WARBURTON, M.; LI, M.; LI, X.; XIAO, M.; HAO, Z.; ZHAO, Q.; ZHANG, S. An analysis of population structure and linkage disequilibrium using multilocus data in 187 maize inbred lines. **Molecular Breeding**, 21:407-418, 2008.

YAMANAKA, N.; SATO, H.; YANG, Z.; XU, D.H.; CATELLI, L.L.; BINNECK, E.; ARIAS, C.A.A.; ABDELNOOR, R.V.; NEPOMUCENO, A.L. Genetic relationships between Chinese, Japanese, and Brazilian soybean gene pools revealed by simple sequence repeat (SSR) markers. **Genetics and Molecular Biology**, 30:85-88, 2007.

ZHU, Y.L.; SONG, Q.J.; HYTEN, D.L.; VAN TASSELL, C.P.; MATUKUMALLI, L.K.; GRIMM, D.R.; HYATT, S.M.; FICKUS, E.W.; YOUNG, N.D.; CREGAN, P.B. Single-Nucleotide Polymorphisms in soybean. **Genetics**, 163:1123-1134, 2003.