

FÁBIO DEMOLINARI DE MIRANDA

USO DE MARCADORES RAPD PARA MAPEAMENTO DE QTLS QUE  
DETERMINAM TEOR DE PROTEÍNA EM SOJA

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

VIÇOSA

MINAS GERAIS – BRASIL

2002

**Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV**

T

M672u  
2002

Miranda, Fábio Demolinari de, 1978-

Uso de marcadores RAPD para mapeamento de QTLs  
que determinam teor de proteína em soja / Fábio Demolinari  
de Miranda. – Viçosa : UFV, 2002.  
56p. : il.

Orientador: Everaldo Gonçalves Barros  
Dissertação (mestrado) – Universidade Federal de  
Viçosa

1. Soja - Mapeamento genético. 2. Soja - Melhora -  
mento genético. 3. Soja - Genética molecular. 4. Marca  
dores RAPD. 5. Proteínas de soja. I. Universidade Federal  
de Viçosa. II. Título.

CDD 19.ed. 633.34

CDD 20.ed. 633.34

FÁBIO DEMOLINARI DE MIRANDA

USO DE MARCADORES RAPD PARA MAPEAMENTO DE QTLs QUE  
DETERMINAM TEOR DE PROTEÍNA EM SOJA

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

APROVADA: 12 de agosto de 2002.

---

Prof. Maurílio Alves Moreira  
(Conselheiro)

---

Prof<sup>a</sup> Elza Fernandes Araújo  
(Conselheira)

---

Prof. João Marcos de Araújo

---

Prof<sup>a</sup> Mara Garcia Tavares

---

Prof. Everaldo Gonçalves de Barros  
(Orientador)

A Deus, por sempre iluminar meu caminho.

À minha mãe, Rosa Demolinari pelo exemplo de luta e esforço para garantir educação à mim e meus irmãos. Pelo pai e mãe que sempre foi. A meu pai Gastão

Pires de Miranda (*in memoriam*)

A meus irmãos Ivan, Luiz Renato e Meiriele.

## **AGRADECIMENTO**

A minha avó Maria Penha pelo carinho e por tudo mais.

À Universidade Federal de Viçosa (UFV), pela oportunidade de realizar o curso de Ciências Biológicas.

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão de bolsas de estudo.

A todos os amigos que conviveram comigo no PET, os quais tiveram um papel decisivo para minha formação como profissional e como pessoa.

Ao Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), pela infra-estrutura adequada ao desenvolvimento de trabalhos científicos.

Ao professor Everaldo Gonçalves de Barros, pela orientação, pelo apoio e pela amizade.

Ao professores Maurilio Alves Moreira e Elza Fernandes de Araújo, pela orientação, e pelo apoio e pela disponibilização de recursos.

Ao professor Lúcio Antônio de Oliveira Campos, pelo apoio e pelo exemplo de ética e honestidade.

Ao Dario Alves de Oliveira pela pelos conselhos, pela amizade.

Ao amigo Gerardo pela amizade e pelo apoio fundamental no desfecho deste trabalho.

À Rita por toda amizade, pelo apoio, pelas dicas e sugestões, por ser um exemplo de pesquisadora.

Ao Newton, pela amizade, pelas críticas e sugestões.

Ao Ivan Schuster, pelo apoio e pelas sugestões.

À Taís pela amizade pelas dicas e conselhos.

Aos amigos de laboratório: Inês, Lucinete, Lucimara, Cândida, Maria Fernanda, Fernanda, Gustavo, Arlindo, Antônio Vieira, Francisco, Jay, Luciano, Thiago, Marcelo, Carlos André, Telma, Márcia, Valéria, Ana Lília, Marta, Pedro Ivo, Marcinho.

Aos funcionários do BIOAGRO: Aloísio, Jander, Fausto, Tadeu, Gláucia, Sandra, Eugênio, Edson, Alessandra, Márcio, Hamilton, Expedito, a todos pelo convívio.

À Cynthia e Francismar, por serem grandes amigas e até mesmo conselheiras.

Aos funcionários da secretaria do curso de Genética e Melhoramento Conceição, Rita e Sr Paulo.

A todos os colegas dos demais laboratórios pelo apoio.

A todos os meus amigos de Piraúba.

A todos os meus amigos de república; antigos (Guilherme, Marcos André, Marcos Matoso, Alexandre, Igor) e atuais, (Breno e Túlio) pela convivência.

Ao “amigo irmão” Torão pela amizade, convivência na república. Mesmo tendo trocado o excelente 486 por um péssimo “pentium 4”, e que tenha dormido todos os dias até o meio dia. Muito obrigado!

Ao amigo Vagner pela amizade, trocas de experiências, nas mais diversas áreas e pela parceria em muitas situações nem sempre fáceis.

Aos camaradas do laboratório, João Paulo e Reginaldo (Naldo) pela amizade.

As minhas vizinhas da república Convento, Fernanda, Bruna e Tassiana, pela amizade, convivência e por tudo mais.

A todos aqueles que colaboraram e torceram pelo meu sucesso.

À força que me protege, guia e direciona.

## **BIOGRAFIA**

FÁBIO DEMOLINARI DE MIRANDA, filho de Gastão Pires de Miranda e Rosa Demolinari, nasceu em 18 de janeiro de 1978, em Piraúba, Estado de Minas Gerais, Brasil.

Em outubro do ano de 2000, formou se em Ciências Biológicas – Bacharelado e Licenciado – pela Universidade Federal de Viçosa, em Viçosa, Minas Gerais.

Em outubro do ano de 2000, ingressou no Curso de Mestrado em Genética e Melhoramento da Universidade Federal de Viçosa, em Viçosa, Minas Gerais, submetendo se à defesa de tese em agosto de 2002.

## CONTEÚDO

RESUMO.....	viii
ABSTRACT .....	ix
1. INTRODUÇÃO .....	1
2. REVISÃO DE LITERATURA .....	6
2.1. Proteínas de reserva de soja .....	6
2.2. A soja na alimentação humana .....	7
2.3. Herdabilidade da característica teor de proteínas .....	8
2.4. Marcadores moleculares .....	10
2.4.1. Marcadores RAPD .....	10
2.4.2. Marcadores microssatélites .....	11
2.5. Mapeamento genético .....	12
2.5.1. Construção do mapa genético .....	12
2.6. Identificação de QTL .....	14
3. MATERIAL E MÉTODOS .....	16
3.1. Material genético .....	16
3.2. Determinação de proteínas .....	16
3.2.1. Método Kjeldahl .....	16
3.3. Análise de variância .....	18
3.3.1. Análise de variância individual .....	18
3.3.2. Análise de variância conjunta .....	20
3.4. Mapeamento visando a identificação de QTLs .....	25

3.4.1. Extração de DNA .....	25
3.5. Amplificação e análise de DNA pela técnica de RAPD .....	26
3.6. Amplificação do DNA e análise de microssatélites .....	26
3.7. Construção do mapa de ligação genética .....	27
3.8. Identificação e mapeamento de QTLs .....	27
3.9. Determinação do ponto de corte para mapeamento de QTLs .....	29
4. RESULTADOS E DISCUSSÃO .....	31
4.1. Detecção de polimorfismo e seleção dos “primers” de RAPD .....	31
4.2. Mapa de ligação genética .....	35
4.3. Análise do teor de proteína .....	39
4.4. Associação entre marcadores e QTLs .....	40
4.5. Mapeamento de QTLs .....	42
5. CONCLUSÕES .....	48
6. REFERÊNCIAS BIBLIOGRÁFICAS .....	49
7. ANEXO .....	54

## RESUMO

MIRANDA, Fábio Demolinari, Universidade Federal de Viçosa, agosto de 2002.  
**Uso de marcadores RAPD para mapeamento de QTLs que determinam teor de proteína em soja.** Orientador: Everaldo Gonçalves de Barros. Conselheiros: Maurílio Alves Moreira e Elza Fernandes Araújo.

O presente trabalho teve como objetivo o aumento do número de marcas no mapa de ligação da soja construído pelo programa de melhoramento da qualidade da soja do Bioagro/UFV e também a identificação de QTLs (*Quantitative Trait Loci*) associados à determinação do conteúdo de proteínas em sementes de soja. Para isso, foram acrescentados ao mapa original, marcadores do tipo RAPD e marcadores microssatélites não mapeados anteriormente. Foram utilizadas 118 linhagens recombinantes endogâmicas (RILs) obtidas do cruzamento entre a variedade norte-americana BARC 8 (genótipo com alto teor protéico) e a variedade brasileira Garimpo (genótipo com teor normal de proteínas). Foram testados inicialmente 1200 “primers” RAPD, dos quais 127 evidenciaram polimorfismo entre os progenitores, dos quais somente 65 mostraram polimorfismos na população de RILs segregando na proporção mendeliana esperada de 1:1, pelo teste do qui-quadrado. Foram obtidos 24 grupos de ligação pouco saturados, contendo 75 marcadores, além de 70 marcas não ligadas. Nas análises de regressão e mapeamento por intervalo composto para associação entre marcadores e a característica “teor de proteína”, foram identificados 11 marcadores e mapeados três QTLs, nos grupos de ligação MGL D2, MGL L e MGL C22 os quais explicam 7,8% e 8,1% e 7,4%, respectivamente, para as famílias cultivadas em Cascavel. Para as famílias cultivadas em Viçosa foram identificados nove marcadores e mapeado um QTL, no grupo de ligação MGL 3, que explica aproximadamente 16,7% da característica. Estudos posteriores deverão ser conduzidos, visando aumentar o grau de saturação do mapa. Isto poderá permitir a identificação de novos QTLs que determinem um maior porcentagem da expressão da característica.

## ABSTRACT

MIRANDA, Fábio Demolinari, Universidade Federal de Viçosa, August, 2002. **Use of RAPD markers for mapping QTLs that control protein content in soybean.** Adviser: Everaldo Gonçalves de Barros. Committee Members: Maurílio Alves Moreira and Elza Fernandes Araújo.

The present work aimed at increasing the number of markers in the soybean linkage map built by the Bioagro/UFV breeding program for soybean quality. It also aimed at identifying QTLs (Quantitative Trait Loci) governing protein accumulation in the seed. RAPD molecular markers and microsatellites which had not been mapped before were added to the original map. One hundred and eighteen recombinant inbred lines (RILs) derived from a cross between the north american variety BARC-8 (with high protein content) and the Brazilian variety Garimpo (with normal protein content) were used. Initially 1,200 RAPD primers were tested. One hundred and twenty seven of them showed polymorphism between the progenitors and 65 of these showed polymorphism among the RILs. All 65 markers segregated according to the expected 1:1 ratio as indicated by the chi-square test. Twenty four linkage groups with a low saturation level (75 markers) were obtained. Seventy other markers were not mapped in the linkage groups. Regression analyses and composed interval mapping identified 11 markers and three QTLs associated with “protein content” were mapped. These QTLs were located in the linkage groups MGL D2, MGL L and MGL C22 and explained 7.8, 8.1 and 7.4% of variation of this trait, respectively in the lines grown in Cascavel (state of Paraná). For the lines grown in Viçosa (state of Minas Gerais) nine markers were identified and one QTL was mapped to linkage MGL 3, and it explained 16.7% of the trait variation. Further studies should be conducted to increase the saturation level of the map. This should allow the identification of new QTLs which might explain a higher percentage of the variation of the protein content in soybean seeds.

# 1. INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill) é um alimento rico em proteínas e lipídeos, consumida pela população mundial na forma de óleo e seus derivados, como extrato hidrossolúvel (leite de soja), proteína texturizada, além de servir como matéria prima para diversos produtos industrializados.

Esta leguminosa apresenta excelente desempenho produtivo, estando dentre os principais produtos agrícolas brasileiros. Do ponto de vista econômico, o farelo é o subproduto mais importante, por ser um dos componentes protéicos mais baratos, sendo utilizado em larga escala na produção de rações (EMBRAPA, 1998).

A maioria dos cultivares de soja apresenta de 30 a 45% de proteínas, 15 a 25% de lipídeos, 20 a 35% de carboidratos e cerca de 5% de cinzas (MOREIRA et al., 1979). Teoricamente todas essas características podem ser modificadas geneticamente pela combinação de genes adequados provenientes do germoplasma dessa espécie.

Um aspecto referente à qualidade do grão de soja é o teor e a qualidade da proteína. A qualidade protéica está diretamente associada ao teor de aminoácidos sulfurados e também às propriedades funcionais da proteína. Duas grandes frações compõem as proteínas de reserva da soja: 7S ( $\beta$  - conglucina) e 11S (glicina), correspondendo a cerca de 70% das proteínas da semente. Essas frações diferem entre si no seu teor de metionina, sendo este maior na glicina (MOREIRA et al., 1979; MOREIRA et al., 1990).

Em estudo de caracterização dos polipeptídeos das subunidades que compõem a globulina 11S, foram identificados seis polipeptídeos ácidos que possuem uma seqüência N-terminal rica em fenilalanina, leucina, isoleucina e arginina, além de quatro polipeptídeos básicos ricos em metionina (MOREIRA et al., 1979).

Dentre as proteínas de reserva da soja, a glicina apresenta maior teor de aminoácidos sulfurados, enquanto a  $\beta$ -conglucina é pobre em tais aminoácidos.

Nas subunidades da  $\beta$ -conglucina a subunidade  $\beta$  é a mais deficiente, contendo apenas um resíduo de cisteína, não possuindo metionina (COATES et al., 1985).

Apesar de ser considerada, de modo geral, uma leguminosa de alto valor nutritivo, a fração protéica da soja e seus derivados apresentam restrição quanto ao seu consumo por humanos no mundo ocidental. Essa restrição está associada às características de qualidade do grão, particularmente no que diz respeito ao sabor e à presença de oligossacarídeos nessa fração que podem causar desconforto intestinal e diarreia (YOUNG et al., 1979).

Quanto a este problema, mais recentemente os avanços no setor de tecnologia de alimentos, têm resultado no surgimento de uma série de produtos comestíveis, derivados de soja, com características que possibilitam uma melhor aceitação por populações ocidentais.

Muitas das características desejáveis, em vários grupos vegetais, incluindo soja, resultam da ação conjunta de mais de um gene e das interações destes com o ambiente. Tais características exibem fenótipos com distribuição contínua que são difíceis de serem analisados, sendo determinadas por mais de um loco no genoma, os chamados QTLs (*Quantitative Trait Loci*). O desenvolvimento mais eficiente de procedimentos de melhoramento depende do melhor entendimento do tipo de ação gênica e herança dos caracteres quantitativos. Diversos programas de melhoramento de soja têm dado maior ênfase no desenvolvimento de variedades com alta produção e alto teor de óleo e proteína. Portanto, para atender objetivos tão diversos é imprescindível que se disponha de uma ampla variabilidade genética (BRIM et al., 1973).

Estudos realizados por FEHR et al. (1978), levaram à conclusão de que altura de plantas, produção de sementes, resistência ao acamamento, porcentagem de óleo e de proteína e tamanho das sementes são características quantitativas em soja.

WEBER et al. (1968), citados por VERNETTI et al. (1983), ao estudar o conteúdo de proteínas nas sementes até a geração F3, em um cruzamento interespecífico de *Glycine max* e *Glycine ussuriensis* concluíram que essa

característica mostrou herdabilidade de 70% e que, apesar de não apresentar herança simples, como “dias para maturação”, mostrou não ter herança tão complexa como “tamanho de semente”.

Melhoramento para características quantitativas não é uma tarefa fácil porque os efeitos de genes individuais que controlam essas características não podem ser facilmente separados ou identificados.

O uso da seleção artificial em programas de melhoramento potencializa a habilidade do melhorista em determinar os efeitos genéticos, e assim poder distinguí-los da porção fenotípica determinada pela ação ambiental. Nestes casos, o uso de marcadores genéticos (marcas genéticas herdáveis e detectáveis que estão associadas a genes no genoma de um organismo), podem ser usadas como importantes ferramentas por geneticistas (DARVASI et al., 1994).

A seleção de características agrônomicas monitorada por marcadores moleculares, baseia-se no princípio de que se um gene, ou um bloco de genes, encontra-se ligado a um marcador genético de fácil identificação, então, esse marcador pode ser usado para selecionar a característica de interesse em um programa de melhoramento. Segundo HAYWARD et al. (1994), o método mais eficaz de detectar este tipo de associação envolve a produção de mapas genéticos detalhados nos quais são localizadas regiões contendo locos para características quantitativas (QTLs). Em um estudo de mapeamento de tais características, uma boa seleção para múltiplos locos ou QTLs, usando marcadores moleculares, será efetiva se houver uma significativa associação entre os locos que determinam a característica e o marcador.

Ultimamente, grande esforço vem sendo feito no sentido de se obter um mapa de ligação saturado para a soja, com o auxílio de marcadores moleculares. Os marcadores moleculares mais utilizados no estudo de mapeamento tem sido aqueles que envolvem a técnica de reação da polimerase em cadeia (PCR). Estes se destacam pelo poder de detecção advindo do PCR, além da praticidade e facilidade da aplicação desta técnica em muitos tipos de análises.

Em relação a estes marcadores, a técnica de RAPD é uma modificação da reação da polimerase em cadeia (PCR), desenvolvida por WILLIAMS et al. (1990). Os marcadores RAPD (“Random Amplified Polymorphic DNA”) originam-se a partir de uma metodologia que se baseia na amplificação de segmentos de DNA utilizando “primers” únicos de seqüências de nucleotídeos tomadas ao acaso. Na técnica de RAPD, fragmentos de DNA do genoma são amplificados pelo uso de tais “primers”. Os fragmentos obtidos podem ser separados de acordo com o seu tamanho por eletroforese.

As vantagens da técnica de RAPD são: o grande número de marcas originadas da amplificação, a possibilidade de se avaliar regiões únicas ou repetidas do genoma, além da simplicidade dos passos envolvidos na técnica.

No início da década de noventa, os marcadores RAPD, passaram a ser bastante utilizados na obtenção de mapas genéticos (HAYWARD et al., 1994). Além do uso dos marcadores apropriados para cada situação, a construção de um mapa genético, envolve outras etapas de suma importância, como: a escolha de uma população mais apropriada para o mapeamento, o cálculo das frequências de recombinação entre as marcas, o estabelecimento de grupos de ligação, as estimativas das distâncias de mapa, a determinação da ordem dos marcadores em questão e finalmente, a estimativa das distâncias entre as marcas genéticas e os locos de interesse. Para facilitar a análise dos dados, foram desenvolvidos programas computacionais, tais como o MAPMAKER (LANDER et al., 1987), QTL Cartografer (BASTEN et al., 1999) e QMol (<http://www.ufv.br/dbg>). Esses programas usam os dados obtidos a partir das populações segregantes, para estimar as frequência de recombinação e arranjo linear dos marcadores genéticos bem como as estimativas das distâncias entre os marcadores e regiões genômicas de interesse.

Este trabalho teve como objetivos aumentar o número de marcas do mapa de ligação da soja construído no programa de melhoramento da qualidade da soja do Bioagro/UFV (SOARES, 2000) e também a identificação de QTLs associados à determinação do conteúdo de proteínas em sementes de soja. Para isso, foram

acrescentados ao mapa original, marcadores do tipo RAPD e marcadores microssatélites não mapeados anteriormente. Foram utilizadas 118 linhagens recombinantes endogâmicas (RILs) obtidas do cruzamento entre o acesso norte-americano BARC 8 (genótipo com alto teor protéico) e a variedade brasileira Garimpo (genótipo com teor normal de proteínas).

## 2. REVISÃO DE LITERATURA

### 2.1. Proteínas de reserva de soja

A proteína de reserva de soja,  $\beta$ -conglucina de soja é uma glicoproteína, uma das mais abundantes proteínas de reserva no grão, se acumula durante a formação da semente. Requerida como proteína estrutural na composição das características físicas da semente, é hidrolizada durante a germinação funcionando como reserva de carbono e nitrogênio para o embrião. A proteína oligomérica completa, com massa molecular estimada em 150-170 KDa é composta por três subunidades distintas, designadas  $\alpha$ ,  $\alpha'$  e a subunidade  $\beta$  da  $\beta$ -conglucina (HILL et al., 1974).

Foram identificados pelo menos quinze genes que codificam para as subunidades da  $\beta$ -conglucina, espalhados pelo genoma da soja, constituindo uma família multigênica. Destes, pelo menos três codificam para as subunidades  $\alpha/\alpha'$  designados *Cg-1*, 2 e 3, seis para a subunidade  $\beta$  sendo designados *Cg-4*, 8, 11, 12, 13, e 15 e quatro genes que apresentam homologia com ambas as subunidades sendo designados *Cg-5*, 6, 7 e 14. Grande quantidade de tais proteínas de reserva, se acumulam durante a formação de sementes de leguminosas, estas funcionam como fonte de carbono e nitrogênio utilizadas durante a germinação e início do desenvolvimento do embrião (MOREIRA et al., 1979).

Outra importante proteína de reserva presente no grão da soja, é uma globulina, a glicinina, que apresenta massa molecular de aproximadamente 350 KDa. Estudos realizados por MOREIRA et al. (1979), demonstraram que a glicinina é uma proteína oligomérica composta por seis subunidades similares. Cada uma destas subunidades é composta por dois polipeptídeos unidos por ligações bissulfídicas, um polipeptídeo com ponto isoelétrico ácido e outro com ponto isoelétrico básico.

Em estudos de caracterização desta proteína foram isolados seis polipeptídeos ácidos (A1a, A1b, A2, A3, A4 e A5) e cinco básicos (B1a, B1b, B2, B3 e B4), que se encontram associados de modo específico, formando as subunidades A1aB2, A1bB1b, A2B1a, A3B4 e A5B3 (MOREIRA et al., 1979; 1981). O polipeptídeo A4 está associado ao B3, por meio de ligações não covalentes, formando a subunidade A5A4B3 (STASWICK et al., 1981). Atualmente, essas subunidades são denominadas G1 (A1aB2), G2 (A1bB1b), G3 (A2B1b), G4 (A5A4B3) e G5 (A3B4) e seus genes *Gy1*, *Gy2*, *Gy3*, *Gy4* e *Gy5*.

As proteínas de reserva, glicinina (11S) e  $\beta$ -conglucina (7S), representam em torno de 70% das proteínas do grão de soja (HILL et al., 1974). Como as globulinas 11S e 7S constituem uma grande porção da proteína total da semente e representam uma importante característica nutricional e de valorização do grão. Esforços vêm sendo aplicados em programas de melhoramento no sentido de obter-se variedades que contenham altos teores de proteína.

## **2.2. A soja na alimentação humana**

As “características de qualidade” ligadas diretamente à composição do grão da soja podem ser modificadas para o desenvolvimento de variedades mais adequadas para uso na indústria de alimentos.

Além de ser muito utilizada na produção de ração animal, a soja possui reconhecidamente um importante papel na dieta de populações humanas do sudeste da Ásia. Mais recentemente os avanços no setor de tecnologia de alimentos, tem resultado no surgimento de uma série de produtos comestíveis, derivados de soja, que possibilitam uma melhor aceitação desta leguminosa por populações ocidentais, possibilitando a manutenção de forma mais barata da dieta protéica.

Tal dieta protéica é indispensável devido à necessidade da ingestão de aminoácidos essenciais pelo homem ou outros animais, além de possibilitar um balanço positivo de nitrogênio o qual é fundamental para manutenção de uma série

de funções do organismo (YOUNG et al., 1979). MOREIRA et al. (1979), em estudo de caracterização dos polipeptídeos das subunidades que compõe a globulina 11S, identificaram a presença de seis polipeptídeos ácidos que possuem uma seqüência N-terminal rica em resíduo de fenilalanina, leucina, isoleucina e arginina. Além de quatro polipeptídeos básicos ricos em metionina.

### **2.3. Herdabilidade da característica teor de proteínas**

Além do conhecimento da magnitude da variância genética de um caráter, bem como a de seus componentes, estimativas de herdabilidade e de correlações entre caracteres agronômicos são de extrema utilidade nos processos de seleção. Em programas de melhoramento de qualquer espécie, tais conhecimentos possibilitam a avaliação da potencialidade da população para o melhoramento e facilitam as decisões de escolha do método de seleção mais eficiente a ser empregado.

Segundo DUDLEY et al. (1969), o termo herdabilidade em um sentido geral, é a razão entre a variância genética total e a variância fenotípica. Em um sentido mais restrito, herdabilidade representa a razão entre a variância genética aditiva e a variância genotípica. Variância fenotípica é o total da variância dentre os fenótipos dos indivíduos desenvolvidos sob determinadas condições. A variância genética total é a parte da variância fenotípica que pode ser atribuída as diferenças genotípicas existente entre os fenótipos.

Para BRIM et al. (1973), o desenvolvimento mais eficiente de procedimentos de melhoramento depende do entendimento do tipo de ação gênica e herança dos caracteres avaliados. Assim, para atender a tais objetivos é imprescindível um profundo conhecimento quanto à natureza e ao número de genes que determinam uma característica além de dispor de uma ampla variabilidade genética.

Em relação a este primeiro ponto, FEHR et al. (1978), avaliando diversas característica agronômicas em soja, concluíram que altura de plantas, produção de

sementes, resistência ao acamamento, percentagem de óleo e de proteína e tamanho das sementes em soja são características quantitativas.

JOHNSON et al. (1963), estudando herdabilidade para várias características em soja, dentre elas teor de proteína, mostraram que em gerações precoces (F<sub>2</sub>) a herdabilidade da característica “teor de proteína” apresenta baixos valores, quando comparados com gerações mais avançadas (F<sub>4</sub>). Isto leva à dificuldades para se identificar e selecionar genótipos superiores em gerações precoces, no caso F<sub>2</sub>, resultando em pouco progresso esperado por seleção.

Na mesma linha de pensamento, SHANNON et al. (1972), estudando estimativas de ganhos por seleção em seis populações na geração F<sub>4</sub>, derivado dos cruzamentos entre quatro progenitores, dois destes com alto teor de proteína, encontraram valores similares entre as populações e elevados valores para herdabilidade.

Os estudos de correlações entre características em soja, principalmente teor de proteína e produção, indicam que não há nenhuma limitação consistente ao progresso genético, visando o aumento simultâneo das duas características, como demonstra o trabalho desenvolvido por WILCOX et al. (1995), utilizando retrocruzamentos, obtendo-se ao final progênies com alto teor de proteína e alta produção.

FEHR et al. (1978), obtiveram uma correlação positiva da gravidade específica e tamanho das sementes, com o teor de proteína. A associação destas duas características pode ser suficiente para ser usada como critério de seleção para aumentar o teor de proteínas nas sementes, pois sementes grandes e de alta gravidade possuem um maior teor de proteína.

## **2.4. Marcadores moleculares**

### **2.4.1. Marcadores RAPD**

Um dos aspectos mais importantes da revolução possibilitada pela técnica do PCR foi a chance de se gerar grandes quantidade de DNA, de segmentos específicos do genoma.

Um grande avanço na área de marcadores moleculares baseados em PCR, ocorreu em 1990 com a publicação dos trabalhos de dois grupos independentes, que o chamaram de RAPD (“Random Amplified Polymorphic DNA”) (WILLIAMS et al., 1990) ou AP – PCR (“Arbitrary Primed” PCR) (WELSH et al., 1990), sendo que a primeira denominação tornou-se mais conhecida. Neste método são usados “primers” curtos e de seqüência arbitrária, eliminando-se a necessidade do conhecimento prévio da seqüência a ser amplificada.

Na utilização desta técnica, um fragmento específico é amplificado quando um “primer” se “anela” em dois locais nas fitas opostas do DNA molde, sendo que o espaço entre esses dois pontos não deve ser superior a 2 ou 3 Kb. Os fragmentos de DNA obtidos podem ser separados de acordo com o seu tamanho e visualizados em gel de eletroforese. O polimorfismo detectado pode ser devido a mutações nos sítios de pareamento, ou inserções e deleções na região flanqueada por estes sítios. Os “primers” geralmente são oligonucleotídios, com 10 nucleotídios, de seqüência arbitrária e uma freqüência de G + C superior a 50% (WILLIAMS et al., 1990).

As vantagens da técnica de RAPD são: o grande número de marcas originadas da amplificação, a possibilidade de se avaliar regiões únicas ou repetidas do genoma, além da simplicidade dos passos envolvidos na técnica. Por estes motivos, os marcadores RAPD tem sido amplamente usados em estudos de mapeamento, identificação de locos, clonagem baseada em mapas, caracterização de cultivares (“fingerprinting”) e seleção assistida por marcadores (SAM).

Estes marcadores, quando empregados na avaliação da diversidade genética, podem fornecer subsídios úteis ao melhorista para seleção de populações básicas a serem utilizadas em programas de melhoramento. Isso porque os marcadores RAPD geram grande quantidade de caracteres adicionais, que combinados com características fenotípicas, fornecem um quadro mais completo para o agrupamento de genótipos e o planejamento de cruzamentos (FERREIRA et al., 1996).

Porém, existem alguns pontos que limitam a sua aplicação, como o baixo conteúdo gerado na amplificação por loco, uma vez que normalmente apenas um alelo é detectado na amplificação, sendo assim considerados marcadores dominantes e não permitindo a distinção entre genótipos homozigoto e heterozigoto. Um segundo problema apresentado é a baixa reprodutibilidade dos dados. Entretanto, estes pontos negativos podem ser superados, por exemplo, padronizando-se as condições da técnica o que permite uma perfeita reprodutibilidade (YU et al., 1993).

BURR et al. (1988), em estudos de avaliação da eficiência de marcadores moleculares para trabalhos de mapeamento em populações de RILs (“Recombinant Inbred Lines”), determinaram que marcadores dominantes fornecem a mesma informação obtida com marcadores codominantes, se a população escolhida para o trabalho em questão for adequada.

#### **2.4.2. Marcadores microssatélites**

Microssatélites ou “Single Sequence Repeats (SSR)” consistem de agrupamentos de seqüências repetidas em série presentes no genoma de diversos organismos, sendo que o número de seqüências repetidas varia para cada organismo e para cada região do genoma. Os microssatélites são flanqueados por seqüências de DNA extremamente conservadas. Assim, pode se usar “primers” que sejam complementares às regiões flanqueando os microssatélites, para se amplificar um determinado loco, via reação de polimerase em cadeia (PCR). As variações existentes nas regiões internas flanqueadas pelo “primer”, resultam de um variado

número de repetições das seqüências microssatélites, podendo ser identificadas por análise eletroforética, através de variações nos tamanhos dos fragmentos gerados. Os marcadores microssatélites apresentam como características principais: serem abundantes e bem distribuídos no genomas dos organismos; podem apresentar polimorfismos; são codominantes; identificados via PCR, sendo bastante acessíveis visto as recentes publicações de “primers” de diversos organismos (WEBER et al., 1990; SAGHAI et al., 1994).

CONDIT et al. (1991), verificaram a abundância de dinucleotídeos repetidos em milho e cinco espécies arbóreas tropicais. AKKAYA et al. (1992), realizaram o primeiro estudo de variação de microssatélites (SSR) em soja. Foram analisadas variações de SSR em 43 linhagens de soja identificando de seis a oito alelos para três locos microssatélites. MORGANTE et al. (1993), publicaram os dados de seqüências de DNA, identificando a presença abundante de repetições de dinucleotídeos e trinucleotídeos em 34 espécies com uma freqüência média de uma repetição a cada 50 kb. WANG et al. (1994), também analisaram seqüências publicadas de 54 espécies de planta para a presença de possíveis microssatélites, neste estudo foi detectada que a repetição  $(AT)_n$  estava repetida de forma mais abundante seguida por  $(A)_n$ ,  $(AG)_n$ ,  $(AAT)_n$ ,  $(ATT)_n$ ,  $(AAC)_n$ ,  $(AGC)_n$ ,  $(AAG)_n$ ,  $(AATT)_n$ ,  $(AAAT)_n$ , e  $(AC)_n$ . Em média uma única seqüência repetida foi encontrada a cada sessenta e quatro quilobases em monocotiledôneas e a cada vinte e um quilobases em dicotiledôneas.

## **2.5. Mapeamento genético**

### **2.5.1. Construção do mapa genético**

A seleção de indivíduos com características de interesse, assistida por marcadores moleculares, pode ajudar na obtenção de variedades com uma ou mais características em um menor espaço de tempo, otimizando o programa de

melhoramento (BRUMMER et al., 1997). Assim, o desenvolvimento de ferramentas moleculares tem facilitado o trabalho de identificar regiões cromossômicas relacionadas a características particulares (BRUMMER et al., 1997). Além disso, o melhor entendimento das bases genéticas de características como teor de proteína e óleo, podem ajudar na elucidação de estratégias de melhoramento, aumentando a possibilidade de manutenção destas características durante as gerações em uma variedade (BRUMMER et al., 1997).

A construção de mapas genéticos relacionando características de qualidade a marcadores de DNA é considerada uma das aplicações de maior impacto da tecnologia de marcadores, na análise genética de espécies e potencialmente no melhoramento de plantas, principalmente em se tratando de característica de herança complexa.

Para a maioria das características quantitativas, poucas informações existem sobre o número, posição cromossômica, magnitude do efeito e interações dos QTL que controlam a sua expressão. A manipulação de características de herança contínua é mais complexa do que aquela de herança simples. Mapas genéticos de marcadores moleculares oferecem a possibilidade de estudar a arquitetura de características quantitativas e, potencialmente, manipular estes fatores durante os procedimentos de seleção e recombinação genética (FERREIRA et al., 1996).

Estudos de mapeamento genético molecular, de características de qualidade em soja, têm resultado na obtenção de mapas genéticos detalhados (LARK et al., 1993; SHOEMAKER et al., 1992), inicialmente com a identificação de QTLs para um variado número de características (KEIM et al., 1990; DIERS et al., 1992; MANSUR et al., 1993), além da avaliação de variações alélicas entre os cultivares (KEIM et al., 1998) e organização genômica (SHOEMAKER et al., 1994).

## 2.6. Identificação de QTL

Muitas características economicamente importantes em plantas e animais são determinadas por muitos genes. Estes são, há bastante tempo, um importante alvo de estudo por geneticista e melhoristas, que buscam identificar e determinar qual a proporção da variância fenotípica é determinada por estes genes, entre populações ou espécies em estudo. Com o advento da tecnologia dos marcadores moleculares, foi possível lançar mão de técnicas como de marcadores moleculares, para se obter informações de mapeamento, posição e efeito, de possíveis QTLs existentes em um genoma (PATERSON et al., 1988; JACOB et al., 1991).

Em contraste com a clássica detecção de ligação para características determinadas por genes únicos, diferentes estratégias têm sido sugeridas, para a identificação (detecção e localização) de um único QTL. Estas estratégias objetivam identificar o maior nível da variância genética total que contribui para a variação de uma característica. Estas estratégias diferem principalmente devido ao tipo e número de marcadores que estão sendo avaliados quando se busca estimar o grau de ligação entre um marcador e um possível QTL. Testes para avaliação da associação QTLs/fenótipo podem envolver a avaliação de um único marcador por vez, dois marcadores por loco simultaneamente, ou considerar todos os possíveis locos ao mesmo tempo.

Neste sentido, pode-se dizer que a habilidade para se encontrar o QTL e determinar seu efeito sobre certa característica é função do tamanho e tipo da população em estudo e da frequência de recombinação entre o marcador e o QTL. A análise baseada em um único marcador ligado ao QTL, pode não ser totalmente eficiente para detecção do mesmo com significância estatística ou seu efeito poderá ser desconsiderado.

ZENG et al. (1994), desenvolveram um método para mapeamento de QTL que combina intervalos de mapeamento com funções de regressão múltipla. Assim, o teste estatístico formulado considera o efeito da presença de um QTL

(influenciando no fenótipo da característica de interesse) em um intervalo e é analisado independente do efeito de um QTL próximo, presente em outro intervalo. Assim o processo seleciona especificamente QTLs únicos ligados a marcas específicas, isolando o efeito deste, de outros QTLs que por ventura estejam próximos.

A partir daí, a presença de um possível QTL, em um intervalo específico, pode ser determinada por testes de permutação como relatado por DEORGE et al. (1996). O teste é realizado misturando se os valores da característica quantitativa em questão e repetindo se a análise. Tal procedimento é realizado até mil vezes, encontrando-se valores de LOD ou razão de verossimilhança no mesmo número das repetições realizadas. Adota-se agora um determinado “ponto de corte” que é o valor mínimo de significância para que a probabilidade da presença de um QTL seja considerada em um intervalo determinado.

### **3. MATERIAL E MÉTODOS**

#### **3.1. Material Genético**

Foi utilizada uma população de 118 RILs “Recombinant Inbred Lines” no mapeamento de QTLs ligados ao teor protéico. Estes indivíduos foram obtidos do cruzamento entre o acesso norte-americano BARC 8, obtida da “Beltsville Agricultural Experimental Station” – USDA, Beltsville, MA, EUA (genótipo com alto teor protéico, em torno de 50%) e a variedade brasileira Garimpo (genótipo com teor normal de proteínas, em torno de 36%). As RILs foram obtidas pelo método SSD (“Single Seed Descent”), a partir da geração F<sub>2</sub>, tomando-se uma única semente de cada planta até a geração F<sub>5</sub>. Na geração F<sub>6</sub>, sementes das 118 linhagens foram plantadas em duas localidades diferentes: Viçosa – Minas Gerais e Cascavel – Paraná. Os experimentos foram plantados no delineamento de famílias com testemunhas intercaladas, sendo as testemunhas, os genitores. Para as famílias não houve repetição local, entretanto, para as testemunhas as repetições foram plantadas a cada 10 famílias.

Foram coletadas folhas de cinco plantas em cada família, no experimento de Viçosa para que estas fossem avaliadas em análise de mapeamento.

#### **3.2. Determinação de proteínas**

##### **3.2.1. Método Kjeldahl**

O teor de proteínas das sementes de cada família foi determinado tomando-se uma amostra de cinco plantas por família. A determinação foi feita utilizando o método Kjeldahl modificado, segundo as normas analíticas do Instituto Adolfo Lutz que requer um mínimo de três gramas (10 sementes) de soja moída. O princípio do método consiste em oxidar uma quantidade da amostra com ácido sulfúrico e

catalisadores à quente. No processo de oxidação, o nitrogênio presente nas moléculas é convertido em sais de amônio, que permanecem no digerido. Em outra etapa, os sais de amônio são alcalinizados pela adição de hidróxido de sódio, resultando na formação de amônia. Pela destilação por arraste de vapor, a amônia é recolhida em solução de ácido bórico adicionada de indicador de pH; forma-se então, o metaborato de amônia, responsável pela mudança de pH. Em seguida, o metaborato de amônia é titulado por uma solução diluída de ácido clorídrico, até nova mudança de cor do indicador de pH. A quantidade de ácido gasta na titulação corresponde ao teor de amônia, que por sua vez representa o teor de nitrogênio da amostra. A partir do teor de nitrogênio calcula-se a porcentagem de proteínas totais da amostra, empregando-se o fator de nitrogênio para o material analisado.

Para o cálculo da porcentagem de proteínas totais pode se usar a seguinte fórmula:

$$\%PT = \left[ \frac{(V_A - V_B) \times N \times f \times 1,4}{P} \right] \times FCN$$

Onde:

% PT = porcentagem de proteínas totais.

$V_A$  = mL de HCl gastos na titulação da amostra.

$V_B$  = mL de HCl gastos na titulação do branco.

N = normalidade do HCl.

F = fator de correção da normalidade do HCl.

P = peso (em g) ou volume (em mL) empregado da amostra.

FCN = fator de conversão de nitrogênio para proteína.

### 3.3. Análise de variância

#### 3.3.1 Análise de variância individual

As análises de variância de cada local foram realizadas pelo procedimento de avaliação de famílias com testemunhas intercalares, sendo os progenitores BARC-8 e Garimpo, utilizados como testemunhas. Este modelo de análise de variância é útil quando não se dispõe de repetições nas parcelas a serem avaliadas. Dessa forma, empregam-se testemunhas adicionais para estimar os efeitos ambientais (resíduo) dos experimentos. Nas testemunhas, o modelo adotado foi o seguinte:

$$Y_{ij} = \mu + T_i + e_{ij}$$

em que,

$Y_{ij}$  = valor da característica para a  $i$ -ésima testemunha na  $j$ -ésima repetição;

$\mu$  = média geral das testemunhas;

$T_i$  = efeito da  $i$ -ésima testemunha ( $i=1,2,\dots,t$ ); e

$e_{ij}$  = erro aleatório que incide sobre as testemunhas, sendo  $e_{ij} \sim NID(0, \sigma^2)$ .

Nas famílias, foi adotado o seguinte modelo:

$$y_i = \mu_f + F_i + e_i$$

em que,

$y_i$  = valor da característica na  $i$ -ésima família;

$\mu_f$  = média geral das famílias;

$F_i$  = efeito da  $i$ -ésima família ( $i=1,2,\dots,f$ ); e

$e_i$  = erro aleatório que incide sobre as famílias, sendo  $e_i \sim NID(0, \sigma^2)$ .

O esquema de análise de variância é ilustrado no Quadro 1.

Quadro 1 - Esquema de análise de variância de cada ambiente, adotando-se o modelo de ensaios de famílias com testemunhas intercalares

FV	GL	SQ	QM	E(QM)	F
Famílias	f-1	SQF	QMF	$\sigma^2 + \sigma^2 g$	QMF/QMR
Testemunhas	t-1	SQTe	QMTe	$\sigma^2 + r \sigma^2 t$	QMTe/QMR
Resíduo	t(r-1)	SQR	QMR	$\sigma^2$	

As somas de quadrados das testemunhas e do resíduo foram obtidas de forma usual para ensaios inteiramente ao acaso, ou seja:

$$SQTo = \sum_{i=1}^t \sum_{j=1}^r \hat{a} \hat{a} Y_{ij}^2 - C$$

$$SQTe = 1/r \sum_{i=1}^t \hat{a} Y_i^2 - C$$

$$SQR = SQTo - SQTe$$

$$C = Y^2/rt$$

A soma de quadrados para famílias foi obtida pela expressão:

$$SQF = \sum_{i=1}^f \hat{a} Y_i^2 - 1/f (\sum_{i=1}^f \hat{a} y_i)^2$$

A herdabilidade no sentido amplo e o coeficiente de variação experimental foram obtidos pelas seguintes expressões:

$$h^2 = s^2 g / QMF$$

$$CV = \frac{100 \sqrt{QMR}}{\hat{\mu}}$$

### 3.3.2. Análise de variância conjunta

Na análise conjunta dos dois ambientes, foram utilizados dois modelos de análise de variância, sendo um para as testemunhas e outro para as famílias. Nas testemunhas, o modelo de análise de variância consistiu em uma análise fatorial, quantificando-se a variação entre elas e a variação devido à interação com o ambiente. Na análise de variância das famílias em dois ambientes, foi adotado um modelo de análise similar à análise em blocos ao acaso, em que cada ambiente correspondeu a um bloco. Nesse tipo de análise, a SQR foi interpretada como a interação família x ambiente. O resíduo da análise conjunta foi obtido pela análise das testemunhas. Esta análise gerou, ainda, uma soma de quadrados de grupos (SQGr) que testa o contraste entre as médias das famílias e das testemunhas avaliadas.

Nas testemunhas, o modelo adotado foi o seguinte:

$$Y_{ijk} = \mu + T_i + A_k + T_i A_k + e_{ijk}$$

em que,

$Y_{ijk}$  = valor da característica para a i-ésima testemunha na j-ésima repetição, no k-ésimo ambiente;

$\mu$ . = média geral das testemunhas;

$T_i$  = efeito da i-ésima testemunha ( $i=1,2,\dots,t$ );

$A_k$  = efeito do k-ésimo ambiente ( $j = 1,2$ );

$T_i A_k$  = efeito da interação entre a i-ésima testemunha e o k-ésimo ambiente; e

$e_{ijk}$  = erro aleatório que incide sobre as testemunhas, sendo  $e_{ijk} \sim \text{NID}(\theta, \sigma^2)$ .

Nas famílias foi adotado o seguinte modelo:

$$Z_{ik} = \mu + F_i + A_k + F_i A_k + e_{ik}$$

em que,

$Z_{ik}$  = valor da característica para a  $i$ -ésima família no  $k$ -ésimo ambiente;

$\mu$ , = média geral das famílias;

$F_i$  = efeito da  $i$ -ésima família ( $i=1,2,\dots,f$ );

$A_k$  = efeito do  $k$ -ésimo ambiente ( $k=1,2$ );

$F_i A_k$  = efeito da interação entre a  $i$ -ésima família e o  $k$ -ésimo ambiente; e

$e_{ik}$  = erro aleatório que incide sobre as famílias, sendo  $e_{ik} = e_{ijk} \sim \text{NID}$

$(0, \sigma^2)$ .

O esquema de análise de variância conjunta é ilustrado no Quadro 2.

Quadro 2 - Esquema de análise de variância conjunta de vários ambientes, nos quais foram avaliadas, em cada ambiente, testemunhas com repetição e famílias sem repetição.

FV	GL	SQ	QM	F
Ambiente (A)	$a - 1$	SQA	QMA	QMA/QMR
Testemunha (Te)	$t-1$	SQTe	QMTe	QMTe/QMR
Te X A	$(a-1)(t-1)$	SQTe x A	QMTe x A	QMTe x A/QMR
Famílias (F)	$f-1$	SQF	QMF	QMF/QMR
FXA	$(f-1)(a-1)$	SQF x A	QMF x A	QMF x A/QMR
Grupo	1	SQGr	QMGr	QMGr/QMR
Resíduo	*	SQR	QMR	
Total	$(rat+af)-1^{\#}$	SQTo		

\* Obtido por diferença.

# Quando o número de repetições das testemunhas é diferente, o GL total pode ser obtido pelo número total de parcelas avaliadas menos 1.

As somas de quadrados foram obtidas pelas seguintes expressões:

$$SQT_o = \sum_{i=1}^t \sum_{j=1}^r \sum_{k=1}^a Y_{ijk}^2 - C$$

$$C = Y/N$$

em que,

$Y_{...}$  = total geral; e

$N = rat + af$  ou número total de parcelas.

$$SQR = \sum_{k=1}^a SQR_K$$

$$SQA = (1/r \sum_{k=1}^a Y_{..k}^2 / t + \sum_{k=1}^a Z_K^2 / f) - C$$

em que,

$Y_{..k}$  = total das testemunhas no ambiente k; e

$Z_{.k}$  = total das famílias no ambiente k.

$$SQTe = 1/ra \sum_{i=1}^t Y_{i..}^2 - Ct$$

sendo  $Ct = Y^2_{...}/ra$ ;  $Y_{i..}$  é o total das testemunhas em cada ambiente e  $Y_{...}$ , o total das testemunhas.

$$SQTexA = SQTe, A^* - (SQTe + SQA^*)$$

Sendo

$$SQA^* = 1/rt \sum_{K=1}^a Y_{..k}^2 - Ct$$

sendo  $Y_{..k}$  o total de cada testemunha nos vários ambientes.

$$SQTe, A^* = 1/r \sum_{I=1}^t \sum_{K=1}^a Y_{i.k}^2 - Ct$$

sendo  $Y_{i.k}$  o total de cada testemunha em cada ambiente.

$SQA^*$  é a soma de quadrados de ambiente, considerando-se apenas as testemunhas (não envolve informações de famílias).

$$SQGr = Y^2.../rat + Z^2../fa - C$$

$$SQF = 1/a \sum_{i=1}^f Z_i^2 - Cf$$

em que  $Cf = Z^2../fa$ , sendo  $Z_{..}$ , o total de famílias nos vários ambientes e  $Z_{i.}$ , o total de cada família, somando-se os vários ambientes.

$$SQA^{**} = 1/f \sum_{K=1}^a Y_{.k}^2 - Cf$$

sendo  $Z_{.k}$  o total de cada ambiente, considerando-se apenas as famílias.

$$SQF,A^{**} = \sum_{I=1}^f \sum_{K=1}^a \dot{a} \dot{z}_{i,k}^2 - Cf$$

$$SQFxA = SQF,A^{**} - (SQF + SQA^{**})$$

$SQA^{**}$  e  $SQTo^{**}$  são as somas de quadrados de ambiente e total, considerando se apenas as famílias.

$$CV = \frac{100\sqrt{QMR}}{\hat{\mu}}$$

A variância genética e a herdabilidade no sentido amplo foram obtidas por

$$s^2_g = \frac{.2 QMF - QMR}{a}$$

$$h^2 = \frac{s^2_g}{s^2_g + s^2}$$

em que,

$$s = QMR$$

sendo essas estimativas obtidas com base nas médias das famílias.

### **3.4. Mapeamento visando a identificação de QTLs**

Para mapeamento genético e identificação de QTLs que determinam o teores de proteína em soja, foram utilizadas amostras de DNA de plantas de 118 RILs. Foram utilizados marcadores RAPD. Para a extração do DNA, cinco folhas de cada família foram coletadas, embaladas em alumínio, congeladas em N<sub>2</sub> líquido e armazenadas a - 80°C.

#### **3.4.1. Extração de DNA**

A extração do DNA de folhas de soja foi feita com base no protocolo descrito por KEIM et al. (1998), com algumas modificações. Cerca de 200 a 300 mg de folhas foram triturados na presença de N<sub>2</sub> líquido, sendo o pó resultante transferido para tubos do tipo *Eppendorf*. Após a adição de 650-800 µL de tampão de extração, constituído de Tris-HCl 50mM (pH 8,0), EDTA 50mM (pH 8,0), NaCl 0,5M, β-mercaptoetanol 1% (v/v) e CTAB 1% (p/v) - sendo este último adicionado separado dos demais componentes – as amostras foram, então, incubadas em banho-maria a 65°C por, aproximadamente, uma hora.

Após a incubação, as proteínas foram removidas pela adição com igual volume de clorofórmio:álcool-isoamílico (24:1) e centrifugação a 14.000 rpm em centrífuga *Eppendorf* modelo 5415C. Os ácidos nucleicos foram, então, precipitados da fase aquosa, pela adição de isopropanol gelado (2/3 do volume da fase aquosa), lavados em etanol 70% gelado por 20 minutos, secos à temperatura ambiente e ressuspendidos em TE (Tris-HCl 10mM, EDTA 1mM, pH 8,0), seguindo-se um tratamento com RNase A, na concentração final de 40 µg/mL, a 37°C, por 30 minutos. O DNA foi recuperado por precipitação, mediante a adição de um volume de isopropanol 95%, seguido de centrifugação a 14.000 rpm, por 15 minutos. Após a lavagem do precipitado com etanol 90% gelado, e secagem ao ambiente, o DNA foi ressuspendido em 200-300 µL de TE.

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

### **3.5. Amplificação e análise do DNA pela técnica de RAPD**

Amostras de DNA foram amplificadas pela técnica de RAPD, de acordo com WILLIAMS et al., (1990), com “primers” decâmeros (“Operon Technologies”, Alameda, CA, EUA), em termociclador Perkin-Elmer, modelo 9600.

As condições de amplificação foram as seguintes: uma etapa de desnaturação do DNA (94°C por 15 segundos), uma etapa de pareamento do “primer” ao DNA-molde (35°C por 30 segundos) e uma etapa de extensão do fragmento (72°C por 1 minuto). Após 40 ciclos, foi efetuada uma última etapa de extensão a 72°C, por sete minutos. Os produtos de amplificação foram separados por eletroforese em gel de agarose 1,2% contendo 10 mg/ml de brometo de etídio, imerso em tampão TBE (tris-borato 90 mM, EDTA 1 mM, pH 8,0). As bandas de DNA foram visualizadas sob luz ultravioleta e fotodocumentadas com o sistema Eagle Eye II (Stratagene).

### **3.6. Amplificação do DNA e análise de microssatélites**

As reações de amplificação de microssatélites foram feitas em um volume total de 15 µL, contendo 1,5 µL de Tris-KCl 50 mM pH 8,3, 1,8 mM de MgCl<sub>2</sub>, 100 µM de cada um dos desoxinucleotídios (dATP, dTTP, dGTP e dCTP), 0,6 µM de cada primer (“Reserach Genetics”, Huntsville, EUA), uma unidade de Taq-polimerase e 30 ng de DNA. As amplificações foram realizadas em termociclador programado para uma etapa inicial de sete minutos a 72°C, seguida de 30 ciclos de um minuto a 94°C, um minuto a 50°C e dois minutos a 72°C e, ao final dos 30 ciclos, uma etapa de sete minutos a 72°C. Os fragmentos amplificados de microssatélites foram separados por eletroforese em gel de agarose 3,0% contendo 6

$\mu$ L de brometo de etídio 10mg/mL, imerso em tampão TBE 1X (Tris-borato 90 mM e EDTA 2 mM, pH 7,0), a 100 volts, por cerca de 3 horas. Terminada a corrida, os géis foram fotodocumentados sob luz ultravioleta, pelo sistema de Eagle Eye II (Stratagene).

### **3.7. Construção do mapa de ligação genética**

Os marcadores microssatélites foram codificados como homozigotos para os alelos vindos de Garimpo (A), homozigotos para os alelos vindos de BARC-8 (B) ou falhas (-) das bandas amplificadas em cada um dos 118 indivíduos da população F<sub>6</sub>. Como os dados foram codificados para RIL's, heterozigotos eventuais foram descartados. A segregação de cada fragmento polimórfico foi testada na população por meio do teste qui-quadrado. Apenas os fragmentos cuja segregação se ajustou à proporção esperada 1:1 ( $P < 0,05$ ) foram utilizados nas análises posteriores.

O mapeamento dos marcadores microssatélites e RAPD e o estabelecimento dos grupos de ligação foram feitos com o uso do programa QMol (<http://www.ufv.br/dbg>). Os marcadores foram agrupados, utilizando-se *LOD score* mínimo de 3,0 e frequência máxima de recombinação de 0,40. A distância de mapeamento de Kosambi foi utilizada para a conversão da frequência de recombinação para centiMorgans (cM).

### **3.8. Identificação e mapeamento de QTL's**

As informações do mapa de ligação foram utilizadas para mapear os QTL's ao longo dos grupos de ligação por meio do mapeamento por intervalo composto (ZENG et al., 1993; 1994), empregando-se o programa QTL Cartographer (BASTEN et al., 1999). Esse tipo de metodologia diferencia-se do mapeamento por intervalo clássico, por isolar o efeito de outros QTL's ligados ao intervalo. O

mapeamento por intervalo composto é um híbrido entre o mapeamento por intervalos e a técnica de regressão múltipla, sendo adotado o seguinte modelo:

$$y_j = \mu + \alpha X_{ij}^* + \sum_k \beta_k X_{jk} + e_i$$

em que,

$y_j$  = valor da característica no j-ésimo indivíduo;

$\mu$  = média do modelo;

$\alpha$  = efeito do QTL, expresso como a diferença entre os dois homozigotos;

$X_{ij}^*$  = uma variável indicadora assumindo valores -1 ou 1, com probabilidade dependendo do genótipo dos marcadores i e j e da posição que está sendo testada no intervalo entre esses marcadores (Quadro 3) ignorando-se recombinações duplas dentro do intervalo);

$\beta_k$  = coeficiente de regressão parcial entre o fenótipo y e o marcador k;

$X_{jk}$  = um coeficiente conhecido para o k-ésimo marcador, no j-ésimo indivíduo, com valores -1 e 1, dependendo do genótipo dos marcadores; e  $e_i$  = variável aleatória.

Quadro 3 - Especificações da variável indicadora  $x^*$  para o modelo de mapeamento por intervalo composto, utilizando-se RIL's

Genótipo do marcador			Nº de indivíduos	$x^*$
Grupo	i	i+1		
1	+	+	n1	1
2	+	-	n2	$\left\{ \begin{array}{l} 1 \text{ com probabilidade } 1-p \\ -1 \text{ com probabilidade } p \end{array} \right.$
3	-	+	n3	
4	-	-	n4	-1

(+) indica genótipo homozigoto para o progenitor 1; (-) indica homozigoto para o progenitor 2.

$P = r_{iq}/r_{i(i+1)}$ , em que  $r_{iq}$  é a frequência de recombinação entre o marcador i e a posição que está sendo testada no intervalo, e  $r_{i(i+1)}$  é a frequência de recombinação entre o marcador i e o marcador i+1; recombinações duplas no intervalo entre esses marcadores são ignoradas.

Os coeficientes  $k$  e  $x_{jk}$  são utilizados para controlar o "background" genético, eliminando-se os efeitos de outros possíveis QTL's sobre  $x^*$ .

### 3.9. Determinação do ponto de corte para o mapeamento de QTL's

A significância de um QTL é atribuída pela razão de verossimilhança entre a hipótese de existência do QTL e a hipótese de que nenhum QTL existe em determinada posição do mapa de ligação

O ponto de corte do LOD para considerar a presença de um QTL pelo método de mapeamento por intervalo composto foi obtido a partir de um teste de permutação como relatado por DEORGE et al. (1996). O teste é realizado misturando se os valores da característica quantitativa em questão, repetindo se a

análise. Tal procedimento é realizado até mil vezes, encontrando-se valores de LOD ou razão de verossimilhança no mesmo número das repetições realizadas. Adota-se agora um determinado “ponto de corte” que é o valor mínimo de significância para que a probabilidade da presença de um QTL seja considerada em um intervalo determinado.

No mapeamento por intervalo composto, o  $\chi^2$  possui um grau de liberdade, pois o denominador da expressão de verossimilhança possui um parâmetro a mais do que o numerador.

Foi utilizado um nível de significância de 5% para o todo o mapa.

## **4. RESULTADOS E DISCUSSÃO**

### **4.1. Detecção de polimorfismos e seleção dos “primers” de RAPD**

Os “primers” de RAPD foram, inicialmente, testados nos progenitores, com a finalidade de selecionar aqueles que gerassem fragmentos polimórficos e que, conseqüentemente, pudessem ser utilizados para mapeamento na população segregante.

Foram testados 1200 “primers”, dos quais 127 evidenciaram polimorfismo entre os progenitores, isto é, cerca de 10,6%.

A baixa diversidade genética da soja cultivada dificulta a obtenção de um grande número de marcas polimórficas em uma população derivada de cruzamento intra-específico. Estudos anteriores evidenciaram que o polimorfismo entre dois genótipos de soja é freqüentemente muito baixo (SHOEMAKER et al., 1994). Em estudos de mapeamento de QTLs relacionados com o conteúdo de proteína e óleo em oito populações de soja, BRUMMER et al. (1997), constataram que 34% dos marcadores RFLP testados não eram polimórficos em nenhuma das oito populações, sendo que nenhum marcador era polimórfico em todas essas populações. AKKAYA et al. (1992), selecionaram um total de 40 “primers” microssatélites, que apresentaram um único produto de amplificação polimórficos entre os genitores, para serem utilizados no mapeamento de uma população de soja constituída de 60 plantas F2 obtidas do cruzamento entre NIL's (linhagens quase isogênicas) dos cultivares Clark e Harosoy. SOARES (2000), trabalhando da mesma forma com uma população de RIL, obteve 65 “primers” microssatélites polimórficos na população, a partir de uma avaliação inicial de aproximadamente 500 “primers”.

Dos 127 “primers” RAPD que mostraram polimorfismos entre os genitores, somente 65 mostraram polimorfismos na população RIL.

A maior parte dos 65 “primers” RAPD que evidenciaram polimorfismos entre os genitores e que produziram bandas monomórficas na população RIL exibiram o

alelo do progenitor BARC-8 em toda a população. Provavelmente, a ausência de polimorfismos nesses locos seja devida ao fato de ter sido utilizada uma isolinha do cultivar Garimpo, Garimpo RCH (resistente ao cancro-da-haste), nos testes de polimorfismo. Uma vez que não estavam disponíveis sementes da variedade Garimpo, que foi usada para gerar as RILs, assumimos que poderíamos utilizar uma isolinha de Garimpo. Os dados obtidos indicam que Garimpo RCH não é uma isolinha de Garimpo e que durante a sua criação, o arraste de material genético do progenitor doador do gene de resistência ao cancro-da-haste foi expressivo. O número de retrocruzamentos efetuados pode não ter sido suficiente para a completa recuperação do genoma da variedade Garimpo na variedade Garimpo RCH, podendo haver divergência genética entre esses dois genótipos. Uma outra hipótese, que não pode ser descartada, é a de que a variedade Garimpo não seja geneticamente pura, o que também poderia contribuir para esse tipo de comportamento dos marcadores.

A segregação dos marcadores RAPD foi testada por meio do teste qui-quadrado ( $P < 0,05$ ). Todos os 65 que levaram à amplificação de fragmentos polimórficos segregaram na proporção mendeliana esperada de 1:1.

Esta mesma população havia sido mapeada com 65 marcadores microssatélites, (SOARES, 2000). Os dados de 15 novos marcadores microssatélites, testados na população e segregando na proporção de 1:1, os quais não haviam ainda sido avaliados foram incluídos nas análise de mapeamento (comunicação pessoal de SOARES).

Os dados dos marcadores RAPD polimórficos identificados neste trabalho, dos 65 marcadores microssatélites mapeados por SOARES (2000) e dos 15 marcadores microssatélites até então não avaliados, estão apresentados nos Quadros 4, 5 e ANEXO (Quadro 8), respectivamente.

Quadro 4 – Teste qui-quadrado da segregação dos marcadores RAPD polimórficos identificados no presente trabalho. A letra A, representa o alelo presente em Garimpo e a letra B, o alelo presente em BARC-8.

“Primer”	Alelo		$\chi^2$	P(%)	“Primer”	Alelo		$\chi^2$	P(%)
	A	B				A	B		
OPN14	64	54	0,84	35,72	OPAI15a	60	58	0,03	85,39
OPP09	58	60	0,03	85,39	OPAI15b	54	64	0,84	35,72
OPP12	59	59	0	100	OPAI01	57	61	0,13	71,27
OPW08	56	62	0,30	58,07	OPAN03	62	56	0,30	58,07
OPY02	65	53	1,22	26,92	OPAN09	65	53	1,22	26,92
OPY07	59	59	0	100	OPAN20	56	62	0,30	58,07
OPY13a	62	56	0,30	58,07	OPAO06	58	60	0,03	85,39
OPY13b	60	58	0,03	85,39	OPAP04a	62	56	0,30	58,07
OPY14	57	61	0,13	71,27	OPAP04b	56	62	0,30	58,07
OPY17	62	56	0,30	58,07	OPAP04c	61	57	0,13	71,27
OPZ06a	59	59	0	100	OPAP04d	58	60	0,03	85,39
OPZ06b	61	57	0,13	71,27	OPAP07a	59	58	0,09	92,63
OPZ18	56	62	0,30	58,07	OPAP07b	61	57	0,13	71,27
OPAA03	58	60	0,03	85,39	OPAR08	66	52	1,66	19,74
OPAA14	58	60	0,03	85,39	OPAS04	65	53	1,22	26,92
OPAA15	62	56	0,30	58,07	OPAS07	50	68	2,74	9,752
OPAB08a	56	62	0,30	58,07	OPAS08	54	64	0,84	35,72
OPAB08b	64	54	0,84	35,72	OPAT19	60	58	0,03	85,32
OPAB08c	57	61	0,13	71,27	OPAU02a	53	65	1,22	26,92
OPAB11	59	59	0	100	OPAU02b	65	53	1,22	26,92
OPAB13	65	53	1,22	26,92	OPAU04	62	56	0,30	58,07
OPAB15	51	56	0,06	80,90	OPAU19	59	58	0,09	92,63
OPAB18a	53	65	1,22	26,92	OPAW09a	64	53	1,03	30,91
OPAB18b	65	52	1,44	22,94	OPAW09b	60	58	0,03	85,39
OPAB19	67	51	2,16	14,07	OPAW18	57	61	0,13	71,27
OPAB20	68	50	2,74	9,750	OPAX12	53	65	1,22	26,92
OPAC01a	63	55	0,54	46,14	OPAX20	59	59	0	100
OPAC01b	50	68	2,74	9,753	OPAZ02	58	60	0,03	85,39
OPAC02	63	55	0,54	46,14	OPAZ07	55	63	0,54	46,14
OPAC04	59	58	0,09	92,63	OPBD16	67	51	2,16	14,07
OPAI11a	61	57	0,13	71,27	OPBE04	63	55	0,54	46,14
OPAI11b	61	57	0,13	71,27	OPBE13	56	62	0,30	58,07
					OPBH02	53	57	0,02	90,33

Quadro 5 – Teste qui-quadrado da segregação dos 15 marcadores microssatélites polimórficos identificados por SOARES (comunicação pessoal), até então não avaliados. A letra A, representa o alelo presente em Garimpo e a letra B, o alelo presente em BARC-8

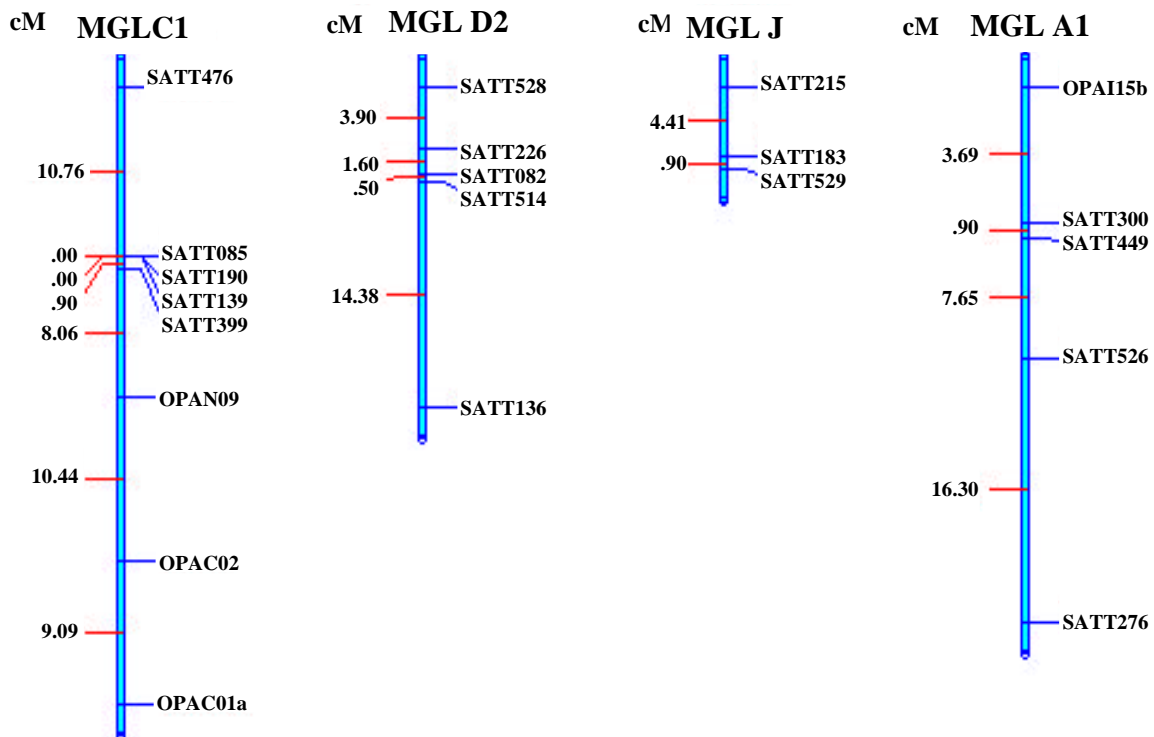
“Primer”	Alelo		$\chi^2$	P(%)
	A	B		
Satt282	58	54	0,14	70,54
Satt215	52	56	0,14	70,03
Satt242	57	58	0,09	92,57
Satt471	58	58	0	100
Satt518	53	59	0,32	55,52
Satt082	54	61	0,42	51,39
Satt537	48	63	2,02	5,70
Satt536	56	56	0	100
Satt350	64	46	2,94	8,61
Satt286	57	59	0,03	85,26
Satt279	47	65	2,89	8,89
Satt454	53	59	0,32	5,70
Satt338	51	60	0,73	39,29
Satt594	54	58	0,14	70,54
Satt130	60	55	0,21	64,10

Em uma população  $F_6$  espera-se encontrar 1,56% dos locos em heterozigose. Essa proporção foi confirmada na população RIL utilizada (dados não mostrados). Como o programa que gera o mapa não reconhece locos em heterozigose, quando os dados são codificados como provenientes de RILs, tais dados foram eliminados da análise com os “primers” microssatélites avaliados no estudo inicial (SOARES, 2000). Como os marcadores RAPD não permitem a distinção no indivíduos entre locos homozigotos e heterozigotos e como estes últimos estariam presentes na população em uma frequência muito baixa, sua presença foi desconsiderada nas análises.

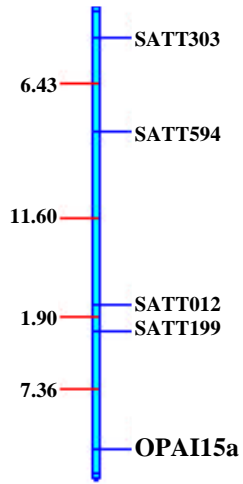
## 4.2. Mapa de ligação genética

Para o agrupamento dos marcadores e construção do mapa de ligação, foram inicialmente definidos o “LOD score” (LOD = 3,0) e a máxima frequência de recombinação ( $r = 0,40$ ). Com os grupos definidos, foi determinada a ordem mais provável dos marcadores em cada grupo de ligação. Foram obtidos 24 grupos de ligação pouco saturados, contendo 75 marcadores como pode ser observado na Figura 1, além de 70 marcas não ligadas.

Figura 1 - Mapa genético da soja com base em uma população RIL constituída por 118 linhagens obtidas do cruzamento entre as variedades BARC 8 e Garimpo. Os grupos de ligação foram obtidos adotando-se os valores de LOD = 3,0 e  $r = 0,40$ . Os valores à esquerda representam as distâncias entre os marcadores (em cM), e à direita está a designação dos marcadores.



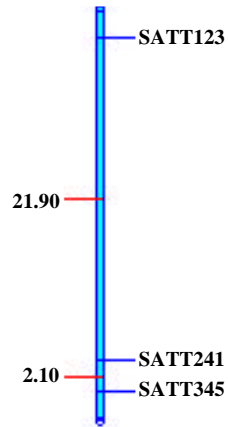
cM MGL G



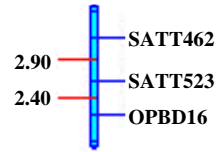
cM MGL 3



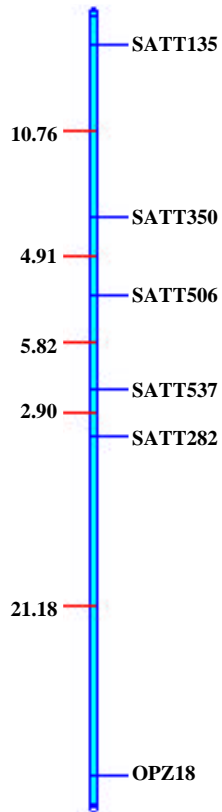
cM MGL O



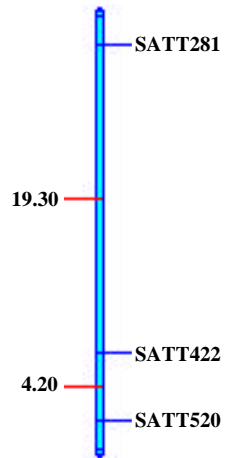
cM MGL L



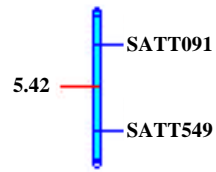
cM MGLD1b+W



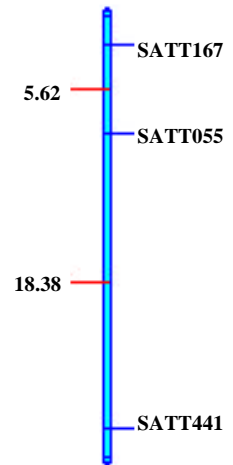
cM MGL C2

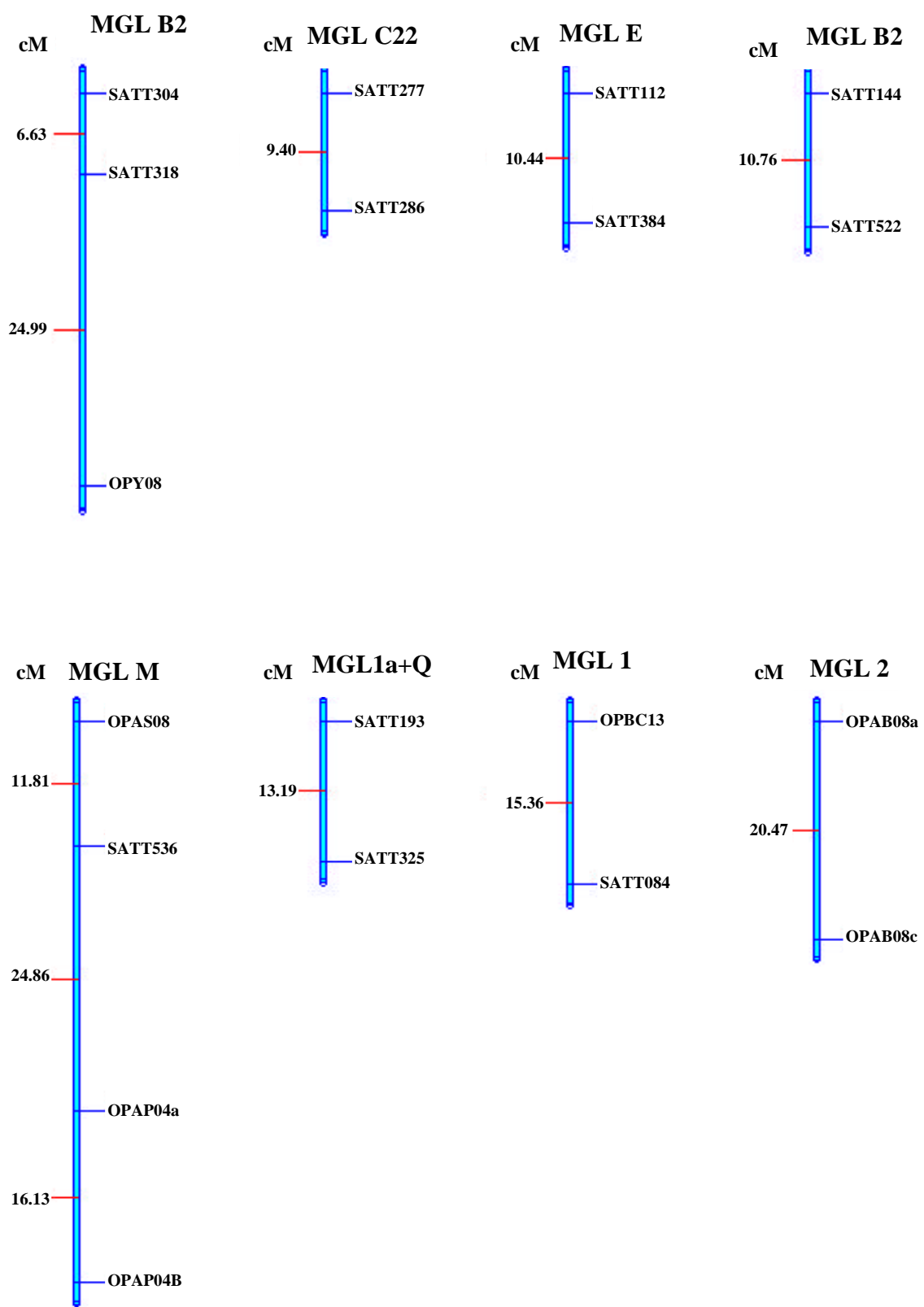


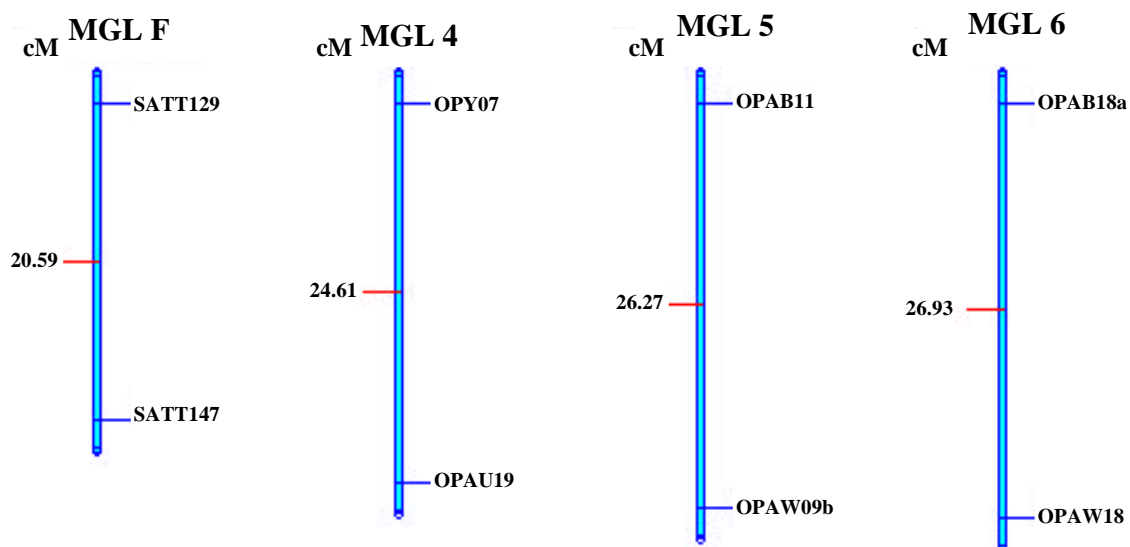
cM MGL N



cM MGL K







Em estudos iniciais, na mesma população, utilizando marcadores microssatélites, SOARES (2000), havia mapeado 16 grupos de ligação contendo 40 marcadores. No presente trabalho, foram acrescentados àquele mapa 34 marcadores, dos quais 21 são RAPD e 14 são microssatélites não alocados em grupos de ligação anteriormente.

Os seguintes marcadores foram mapeados em 10 dos 16 grupos de ligação definidos por SOARES (2000): OPAN09, OPAC02, OPAC01a (MGL C1), Satt082 e Satt186 (MGL D<sub>2</sub>), Satt215 (MGL J), OPAI15b (MGL A1), OPAI15a (MGL G), Satt123 (MGL O), Satt350, Satt537, Satt282 e OPZ18 (MGL D1b + W), OPBD16 (MGL L), e OPY02 (MGL B2). Ainda foram formados oito novos grupos de ligação: MGL C2, MGL M, MGL 1, MGL 2, MGL 3, MGL 4, MGL 5 e MGL 6, formados por marcadores RAPD e microssatélites não mapeados anteriormente.

A grande maioria dos marcadores mapeados por SOARES (2000), bem como aqueles mapeados no presente trabalho, foram posicionados nos mesmos grupos de ligação definidos por CREGAN et al. (1999), trabalhando com populações F<sub>2</sub> do cruzamento interespecífico *Glycine soja* x *Glycine Max*, o que demonstra a boa representabilidade dos marcadores utilizados. Poucas alterações na ordem dos marcadores em um dado grupo de ligação ou no seu posicionamento em diferentes

grupos foram encontrados comparando-se o mapa aqui obtido e aquele obtido por CREGAN et al. (1999). As poucas alterações encontradas podem, muito provavelmente, serem explicadas pelos diferentes tipos de populações utilizadas nos dois trabalhos, pelos tipos de progenitores usados para gerar as populações de mapeamento, bem como pelo fato de uma população ser derivada de cruzamento interespecífico e outra de cruzamento dentro da mesma espécie.

### **4.3. Análise do teor de proteína**

Como já determinado por SOARES (2000), o conteúdo de proteínas nas 118 famílias apresentou distribuição aproximadamente normal, tanto no experimento de Viçosa quanto no de Cascavel ANEXO [Figura 4 (A) e (B)]. Esse tipo de distribuição indica que vários genes estão envolvidos na determinação do conteúdo de proteínas em soja. A média dessa característica foi maior no experimento de Cascavel ANEXO [(Figura 4 (B)]. Esse fato, provavelmente, seja devido ao efeito diferencial de fatores ambientais como temperatura, solo, altitude, entre outros, que podem ter influenciado no acúmulo de proteínas de reserva nas sementes nos dois locais.

As análises de variância para cada local ANEXO (Quadros 9 e 10) evidenciaram que a precisão experimental foi grande, tanto no experimento de Viçosa quanto no de Cascavel. Os coeficientes de variação foram de 3,28%, em Viçosa e 2,93% em Cascavel. Tanto as famílias quanto os progenitores apresentaram grande variabilidade quanto ao teor de proteína. A análise de variância conjunta dos dois locais ANEXO (Quadro 11) indicou que fatores ambientais estão envolvidos na regulação da expressão gênica de proteínas de reserva nos dois ambientes (Viçosa e Cascavel), pois a interação genótipo x ambiente foi significativa tanto nas testemunhas quanto nas famílias (SOARES, 2000).

A herdabilidade da característica teor de proteína obtida por componentes de variância foi alta nos dois locais e também na análise conjunta. Isso significa que a

maior parte da variação observada é devida a causas genéticas e reflete também a precisão experimental. Altos valores de herdabilidade para o teor de proteína, utilizando-se componentes de variância, também foram obtidos por PIOVESAN (2000), em vários cruzamentos com progenitores contrastantes quanto ao teor de proteína.

#### **4.4. Associação entre marcadores e QTLs**

O melhoramento genético de espécies baseado em seleção artificial depende da habilidade do pesquisador em estimar a proporção de uma característica de interesse que é determinada por fatores genéticos herdáveis. É importante, da mesma forma, distinguir este componente da porção do fenótipo determinada por fatores ambientais. No caso de características de herança complexa, nas quais o componente ambiental é importante, o melhorista pode encontrar grande dificuldade para exercer uma seleção adequada.

Nesses casos, a seleção artificial pode ser facilitada pelo uso de marcadores de DNA que estejam próximos, ou mesmo, que flanqueiem os QTLs de interesse. Usando o programa computacional JUMP, Versão 3.1.6.2 (“SAS Institute Inc.”, 1990) e realizando-se análises de regressão simples para cada marcador e múltiplas para os marcadores flanqueando o QTL, foram identificados 11 marcadores mapeados na população cultivada em Cascavel-PR, com associação significativa com a característica “teor de proteína” (Quadro 6).

Na análise dos marcadores associados a essa característica para as famílias cultivadas em Viçosa-MG foi encontrada associação significativa com probabilidade ( $P < 0,05$ ) para nove marcadores (Quadro 7).

Quadro 6 - Análises de regressões simples e múltiplas entre os marcadores moleculares e a característica teor de proteína das famílias cultivadas em Cascavel-PR.

Marcadores	Grupos de Ligação	R <sup>2</sup>	R <sup>2</sup> <sub>ajustado</sub>	Regressão Simples		Regressão Múltipla	
				F	P > F	F	P > F
OPAU19	MGL 5	4,60	3,82	4,965	0,0189		
OPAW18	MGL 8	3,77	2,93	4,094	0,0206		
OPAI15a	MGL G	3,66	2,79	3,980	0,0244		
OPBD16	MGL L	5,52	4,68	5,730	0,0159	8,325	0,0003
Satt277	MGL C2	3,57	2,71	3869	0,0257	6,237	0,0023
Satt012	MGL G	4,85	3,68	5,907	0,0166		
Satt523	MGL L	4,64	3,80	5,149	0,0136	7,954	0,0004
Satt082	MGL D2	3,60	2,40	3,258	0,0312	9,235	0,0009
Satt514	MGL D2	3,25	2,39	3,762	0,0208	8,5431	0,0004
Satt286	MGL C2	5,77	4,96	7,104	0,0088	4,322	0,0012
Satt528	MGL A1	2,18	1,50	0,678	0,4120		

\*R<sup>2</sup> = 31,36%    \*R<sup>2</sup><sub>ajustado</sub> = 25,64%

Quadro 7 - Análises de regressões simples e múltiplas entre os marcadores moleculares e a característica teor de proteína das famílias cultivadas em Viçosa-MG.

Marcadores	Grupos de Ligação	R <sup>2</sup>	R <sup>2</sup> <sub>ajustado</sub>	Regressão Simples		Regressão Múltipla	
				F	P > F	F	P > F
OPAAB08a	MGL 2	2,64	1,80	3,146	0,0787		
Satt144	MGL K	2,80	1,97	3,346	0,0699		
Satt146	MGL 3	5,03	4,21	6,142	0,0146	6,311	0,0023
Satt167	MGL J	3,00	2,16	3,587	0,0607		
OPAC01b	MGL 3	5,47	4,65	6,712	0,0108	7,544	0,0015
Satt276	MGL C2	3,36	2,80	4,375	0,0387		
Satt384	MGL F	4,50	3,65	5,437	0,2140		
Satt084	MGL C1	3,21	2,38	3,852	0,0521		
Satt286	MGL C2	2,62	1,90	3,321	0,0674		

\*R<sup>2</sup> = 31,28%    \*R<sup>2</sup><sub>ajustado</sub> = 25,55% viçosa

A predição do valor de uma marca genética a ser usada como ferramenta de seleção artificial em programas de melhoramento depende de vários fatores. Repetibilidade do processo e proximidade de ligação entre o marcador genético e o gene de interesse aparecem como pontos de suma importância. A existência de uma ligação estreita entre um QTL e um marcador genético (<10 cM) poderá permitir que este marcador venha a ser utilizado como uma eficiente ferramenta de seleção.

#### 4.5. Mapeamento de QTLs

A análise de regressão pelo programa JUMP, Versão 3.1.6.2 (“SAS Institute Inc.”, 1990) apenas informou se os marcadores mapeados estavam associados à característica “teor de proteína”. No entanto, esse tipo de análise não informa a respeito do posicionamento relativo entre o marcador e um possível QTL. Com o intuito de aumentar o poder de detecção de QTLs e melhorar a precisão das estimativas do seu efeito e posição, foi utilizado o método de mapeamento por intervalo composto (ZENG et al., 1993). Este método é baseado na separação do QTL presente no intervalo analisado. O procedimento desconsidera os possíveis QTLs posicionados em outros intervalos, que não estejam sendo analisados, de forma a isolar o efeito de um QTL do efeito de outros possíveis QTLs presentes em intervalos adjacentes.

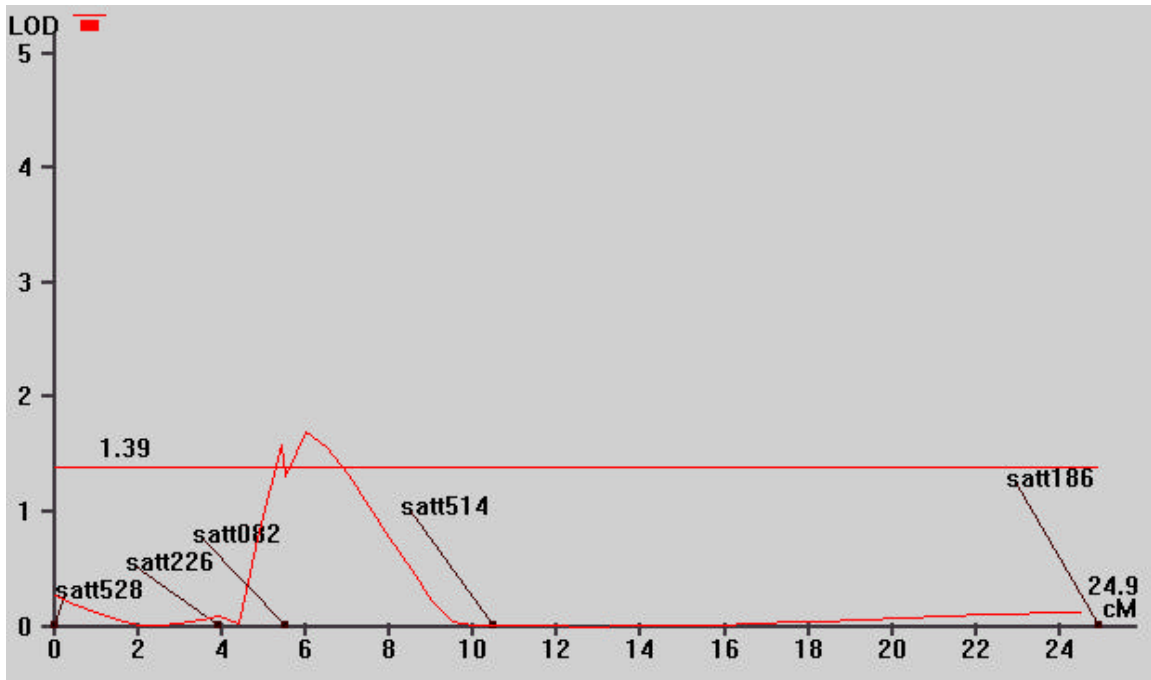
O ponto de corte para mapeamento de possíveis QTLs, dentro de um intervalo testado, foi obtido por teste de permutação (DEORGE et al., 1996). Adotando-se um nível de significância de 5% para o mapa completo, sendo este 6,4 e 12,8 para as famílias cultivadas em Cascavel-PR e Viçosa-MG respectivamente.

Empregando-se este nível de significância, foram mapeados três QTLs, associados com teor de proteína do grão, nos grupos de ligação MGL D2, MGL L e MGL C22 para as famílias cultivadas em Cascavel-PR, os quais explicam 7,8% e 8,1% e 7,4% da característica, respectivamente. Um QTL foi mapeado, associado com a característica, no grupo de ligação MGL 3, para as famílias cultivadas em Viçosa-MG que explica aproximadamente 16,7% da característica. Empregando o nível de significância de 10%, SOARES (2000), identificou dois QTLs na população de Cascavel-PR, nos grupos de ligação MGL C1, MGL C2.

Para confirmar os dados obtidos com mapeamento por intervalo composto, foi efetuada a regressão múltipla, pelo programa JUMP. Para as plantas cultivadas em Cascavel-PR, os marcadores Satt082 e Satt514 (grupo de ligação MGL D2) (Figura 2A), Satt523 e OPBD16 (grupo de ligação MGL L) (Figura 2B) e Satt277 e

Satt286 (grupo de ligação MGL C22) (Figura 2C) explicam cerca de 9,4%, 10,1% e 9,8% do teor total de proteína presente no grão, respectivamente. Os marcadores OPAC01b e Satt146 (grupo de ligação MGL3) (Figura 3) explicam juntos aproximadamente 19,7% do teor total de proteína do grão para as plantas cultivadas em Viçosa-MG.

### Grupo de ligação MGL D2 (A)



### Grupo de ligação MGL L (B)

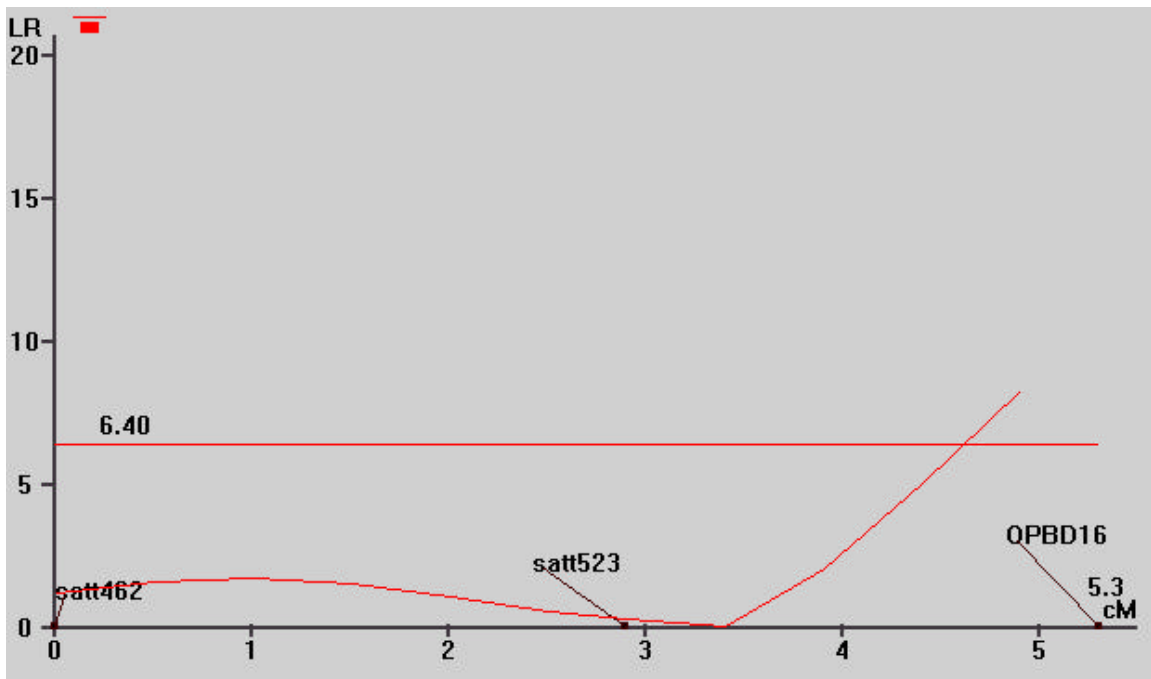


Figura 2 - QTL's encontrados nas análises das famílias cultivadas em Cascavel-PR presentes no grupo de ligação MGL D2 (A) e MGL L (B). Foi utilizado um nível de significância de 5%.

### Grupo de ligação MGL C22 (C)

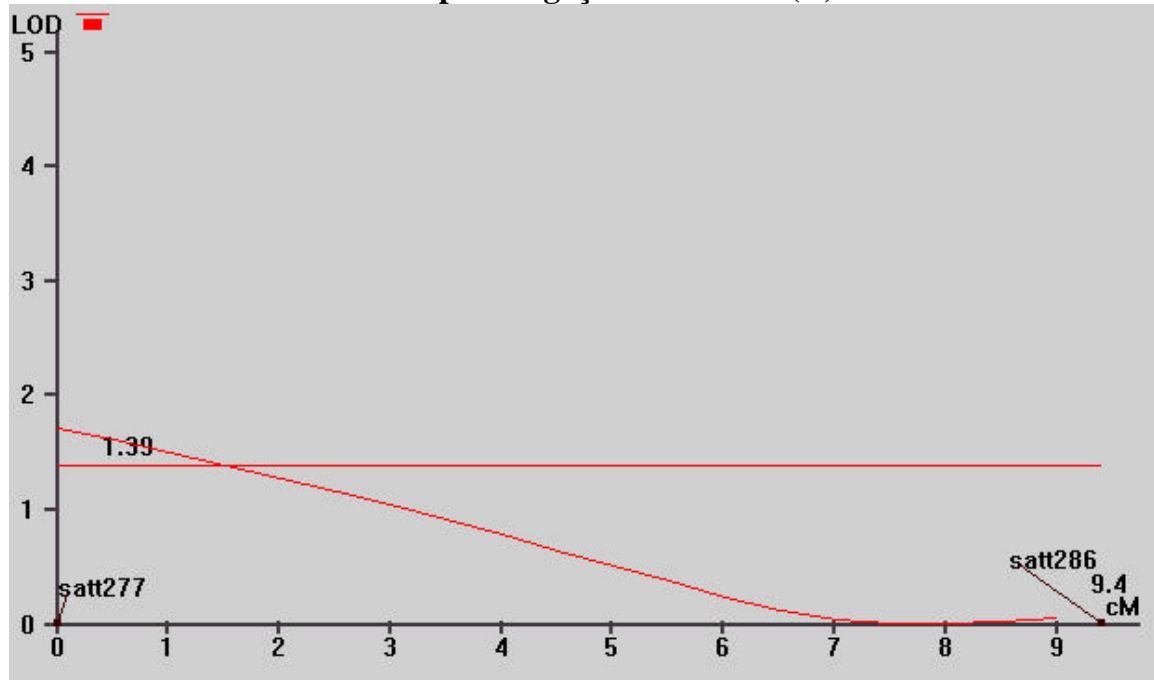


Figura 2 - QTL encontrado nas análises das famílias cultivadas em Cascavel-PR presentes no grupo de ligação MGL C22 (C). Foi utilizado um nível de significância de 5%.

### Grupo de ligação MGL3

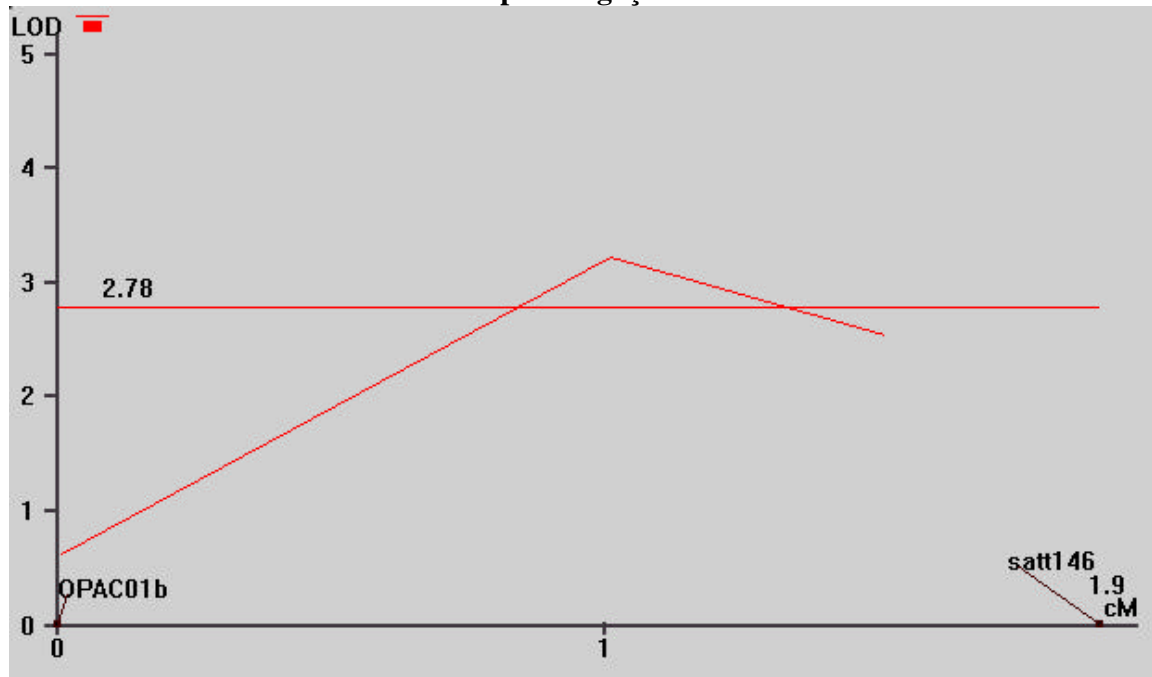


Figura 3 - QTL encontrado nas análises das famílias cultivadas em Viçosa-MG presente no grupo de ligação MGL 3. Foi utilizado um nível de significância de 5%.

Diversos genes relacionados a teor de proteína em soja foram mapeados (DIERS et al., 1992). MANSUR et al. (1993), encontraram marcadores RFLP proximamente associados a genes que determinam teor de proteína em soja no grupo de ligação A2.

Os três possíveis QTLs identificados e mapeados nas linhagens cultivadas em Cascavel, juntos explicam aproximadamente 24% da característica “teor de proteína” do grão de soja. Da mesma forma, o único QTL identificado e mapeado para as linhagens cultivadas em Viçosa explica cerca de 16,7% da mesma característica. Em ambos os casos, a estreita ligação obtida, bem como o posicionamento de QTLs, flanqueados por marcadores RAPD ou microssatélites, poderá permitir com maior eficiência o monitoramento da transmissão e a determinação da presença destes QTLs em linhagens que estejam sendo submetidas a programas de melhoramento em Viçosa ou Cascavel.

Estudos posteriores deverão ser conduzidos visando aumentar o grau de saturação do mapa. Isto poderá permitir a identificação de novos QTLs que expliquem uma maior porcentagem da expressão da característica.

## 5. CONCLUSÕES

Dos 1200 “primers” RAPD testados nos progenitores BARC-8 e Garimpo, 127 evidenciaram polimorfismos, isto é, cerca de 10,6%. Esse dado confirma a estreita base genética da soja cultivada.

Dos 127 “primers” RAPD que mostraram polimorfismos entre os progenitores, somente 65 evidenciaram polimorfismos na população de 118 “Recombinant inbred lines” (RILs) derivada do cruzamento entre BARC-8 e Garimpo. Isto ocorreu provavelmente, devido ao fato de ter sido utilizada uma isolinha do cultivar Garimpo, Garimpo RCH (resistente ao cancro-da-haste), nos testes de polimorfismo dos progenitores.

Todos os 65 que levaram à amplificação de fragmentos polimórficos segregaram na proporção mendeliana esperada de 1:1, pelo teste do qui-quadrado.

Na análise de agrupamento dos marcadores e construção do mapa de ligação, foram obtidos 24 grupos de ligação pouco saturados, contendo 75 marcadores.

Nas análises de regressão simples e múltipla foram identificados 11 marcadores mapeados na população cultivada em Cascavel-PR, e nove marcadores para as famílias cultivadas em Viçosa-MG os quais apresentaram associação significativa com probabilidade ( $P < 0,05$ ) para a característica “teor de proteína”.

Empregando-se o método de mapeamento por intervalo composto, foram mapeados três QTLs associados com teor de proteína do grão, nos grupos de ligação MGL D2, MGL L e MGL C22, para as famílias cultivadas em Cascavel, os quais explicam 7,8% e 8,1% e 7,4% da característica, respectivamente. Foi também mapeado um QTL para teor de proteína, no grupo de ligação MGL3, para as famílias cultivadas em Viçosa, o qual explica aproximadamente 16,7% da característica.

Estudos posteriores deverão ser conduzidos, visando aumentar o grau de saturação do mapa. Isto poderá permitir a identificação de novos QTLs que determinem um maior porcentagem da expressão da característica.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- AKKAYA, M. S.; BHAGWAT, A. A.; CREGAN, P. B. Length polymorphism of single sequence repeat DNA in soybean. **Genetics**, 132: 1132 – 1139, 1992.
- BRIM, C. A.; CALDWELL, B. E. Quantitative genetics and breeding in soybeans. Improvement, production, and uses. **American Society of Agronomy**, Inc. Publishers, Inc., Madison, 155 – 186 1973.
- BASTEN, C. J.; WEIR, B. S.; ZENG, Z. B. **QTL cartographer**; version 1.13. Raleigh, NC: Department of Statistics; North Carolina State University, 132p, 1999.
- BRUMMER, E. C.; GRAEF, G. L.; ORF, J.; WILCOX, R. J.; SHOEMAKER, R. C. Mapping QTL for seed protein and oil content in eight soybean populations. **Crop Science**, 37: 370 - 378, 1997.
- BURR, B.; BURR, F. A.; THOMPSON, K. H.; ALBERTSON, M.; STUBER, C. W. Gene mapping with recombinant inbred lines in maize. **Genetics**, 118, 4: 519 – 526, 1988.
- CONDIT, R.; HUBBELL, S. P. Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. **Genome**, 34: 66 – 71, 1991.
- COATES, J. B.; MEDEIROS, J. S.; THANH, V. H.; NIELSEN, N. C. Characterization of the subunits of  $\beta$ -conglycinin. **Archives of Biochemistry and Biophysics**, 243: 184 - 194, 1985.
- CREGAN, P. B.; JARVIK, T.; BUSH, A. L.; SHOEMAKER, K. C.; LARK, K. G.; KAHLER, A. L.; KAYA, N.; VAN TOAI, T. T.; LOHNES, D. G.; CHUNG, J.; SPECHT, J. E. An integrated genetic linkage map of the soybean genome. **Crop Science**, 39: 1464 – 1490, 1999.
- DARVASI, A.; SOLLER, M. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. **Theoretical Applied Genetics**, 85: 353 – 359, 1994.
- DEORGE, R. W.; CHURCHILL, G. A. Permutation tests for multiple loci affecting a quantitative character. **Genetics**, 142, 1: 285 – 294, 1996.

- DIERS, B. W.; KEIM, P.; FEHR, W. R.; SHOEMAKER, R. C. RFLP analysis of soybean seed protein and oil content. **Theoretical Applied Genetics**, 83: 608 - 612, 1992.
- DUDLEY, J. W.; MOLL, R. H. Interpretation and use of estimates of heritability and genetic variances in plant breeding. **Crop Science**, 3, 9: 257 - 262, 1969.
- EMBRAPA. Centro Nacional de Pesquisa de Soja (Londrina, PR). **Recomendações técnicas para a cultura da soja na região central do Brasil 1998/99**. Londrina, 182p. (Embrapa-Soja. Documentos, 120). 1998.
- FEHR, W. R.; NORMAN, A. G. Breeding In: Soybean physiology, agronomy, and utilization. **Academic Press New York**, 119 - 155, 1978.
- FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análises genéticas**. 2ed. Brasília EMBRAPA-CENARGEN. 220p, 1996.
- HAYWARD, M. D.; McADAM, N. J.; JONES, J. G.; EVANS, C.; EVANS, G. M.; FORSTER, J. W.; USTIN, A.; HOSSIN, K. G.; QUADER, B.; STAMMERS, M.; WILL, J. K. Genetic markers and the selection of quantitative traits in forage grasses. **Euphytica**, 77: 269 - 75, 1994.
- HILL, J. E.; BREIDENBACH, R. W. Protein of soybean seeds. **Plant Physiology**, 53: 742 - 746, 1974.
- JACOB, H. J.; LINDPAINNER, K.; LINCO, S. E.; KUSUMI, K.; BUNKER, R. K.; GANTEN, D.; DZAU, V. J.; LANDER, E. S. Genetics mapping of gene causing hypertension in the stroke-prone spontaneously hypertensive rat. **Cell**, 67, 1: 213 - 224, 1991.
- JOHNSON, H. W.; BERNARD, R. L. Soybean genetics and breeding. **The Soybean. Academy Press, New York**, 34: 1 - 73, 1963.
- KEIM, P.; DIERS, B. W.; SHOEMAKER, R. C. Genetic analysis of soybean hard seediness with molecular markers. **Theoretical Applied Genetics**, 79: 465 - 469, 1990.
- KEIM, P.; OLSONN, T. C.; SHOEMAKER, R. C. A rapid protocol for isolating soybean DNA. **Soybean Genetics Newsletter**, 15: 150 - 152, 1998.

- LARK, K. G.; WEISEMANN, J. M.; MATTHEWS, B. F.; PALMER, R.; CHASE, K.; MACALMA, T. A genetic map of soybean (*Glycine max L.*) using an intraspecific cross of two cultivars: “Minsoy” and ‘Noir 1’. **Theoretical Applied Genetics**, 86: 901-906, 1993.
- LANDER, E. S.; GREEN, P.; ABRAHAMSON, J.; BARLOW, A.; DALY, M. J.; LINCOLN, S. E.; NEWBURG, L. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. **Genomics**, 1: 174 – 181, 1987.
- MANSUR, L. M. L.; LARK, K. G.; KROSS, H.; OLIVEIRA, A. Interval mapping of quantitative trait locus for reproductive, morphological and seed traits of soybean (*Glycine max L. Merr*) **Theoretical Applied Genetics**, 86: 907 - 913, 1993.
- MOREIRA, M. A.; HERMADSON, M.; LARKINS, B. A. Partial characterization of the acidic and basic polypeptides of glycinin. **Journal Biological Chemistry**, 254: 9921 - 9926, 1979.
- MOREIRA, M. A.; HERMODSON, M.; LARKINS, B. A.; NIELSEN, N. C. Comparison of the primary structure of the acidic polypeptides of glycinin. **Archives of Biochemistry and Biophysics**, 210: 636 - 642, 1981.
- MOREIRA, M. A.; REZENDE, S. J.; SEDIYAMA, C. S.; BARROS, E. G. Obtenção de cultivares de soja de sabor agradável e com sementes de alta qualidade fisiológica. In: TORRES, A. C.; CALDAS, L. S. (Eds.). **Técnicas e aplicações da cultura de tecidos de plantas**. Brasília: ABCTP, 417 – 426, 1990.
- MORGANTE, M.; OLIVIERI, A. M. PCR-amplified microsatellites as markers in plant genetics. **Plant Journal**, 3:175 – 182, 1993.
- PATERSON, A. H.; LANDER, E. S.; HEWITT, J. D.; PATERSON, S.; LINCOLN, E. E.; TANKSLEY, S. D. Resolution of quantitative trait into a mendelian factors by using a couplet linkage map of restriction fragment length polymorphism. **Nature**, 335, 6193: 721 – 726, 1988.
- PIOVESAN, N. D. **Aplicação de cruzamentos dialélicos no melhoramento genético do teor protéico em soja**. Viçosa, MG: UFV, 2000. 84p. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 2000.

- SAMBROOK, J.; FRITSCH, E. F.; MANIATIS, T. **Molecular cloning: a laboratory manual**. 2.ed. New York: Cold Spring Harbor Laboratory, 3, p.8B 1989.
- SAGHAI, M.; BIYASHEV, R. B.; YANG, G. P.; ZHANG, Q.; ALLARD, R. W. Extraordinarily polymorphic microsatellites DNA in barley: species diversity, chromosomal locations, and population dynamics. **Proc. National Academy Science. U.S.A.** 91: 5466 – 5470, 1994.
- SHANNON, J. G.; WILCOX, J. R.; PROBST, A. H. Estimated gains from selection for protein and yield in the F4 generation of six soybean populations. **Crop Science**, 12: 824 - 826, 1972.
- SHOEMAKER, R. C.; GUFFY, R. D.; LORENZEN, L. L.; SPECHT, J. E. Molecular genetic mapping of soybean: map utilization. **Crop Science**, 32: 1091 - 1098, 1992.
- SHOEMAKER, R. C. RFLP map of soybean. In: PHILLIPS, R. L.; VASIL, I. K., eds. DNA- based markers in plants. s.1, **Kluwer Academic Publishers**, 2990 - 3009 1994.
- SOARES, T. C. B. **Mapeamento de locos que controlam o conteúdo de proteína em soja**. Viçosa, MG: UFV, 2000. 58p. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 2000.
- STASWICK, P. E.; HERMODSON, M. A.; NIELSEN, N. C. Identification of the acidic and basic subunits complexes of glycinin. **Agricultural and Biological Chemistry**, 256: 8752 - 8755, 1981.
- VERNETTI, F. J. **Soja - genética e melhoramento**. Fundação Cargill. Vol. 2. Campinas, 465 – 990, 1983.
- WANG, Z.; WEBER, J. L.; ZHONG, G.; TANKSLEY, S. D. Survey of plant short tandem DNA repeats. **Theoretical Applied Genetics**, 88: 1 – 6, 1994.
- WEBER, C. R. Mass selection by seed size and specific gravity in soybean population. **Crop Science**, 8: 551 - 554, 1968.
- WEBER, J. J. Informative ness of human (dC-dA)<sub>n</sub> (dG-dT)<sub>n</sub> polymorphism's. **Genomics**, 7: 524 – 530, 1990.

- WELSH, J.; McCLELLAND, M. Fingerprinting genomes using PCR with arbitrary primer. **Nucleic Acids Research**, 18, 24: 7213 – 7218, 1990.
- WILLIAMS, J. G. K.; KUBELIK, A. R.; LIVAK, K. J.; RAFALSAKI, J. A.; TINGEY, S. V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Research**, 18: 6531 - 6535, 1990.
- WILCOX, J. R.; CAVINS, J. F. Backcrossing higher seed protein to a soybean cultivar. **Crop Science**, 35: 1036 - 1041, 1995.
- YOUNG, V. R.; SCRIMSHAW, N. S. Soybean protein in human nutrition an overview. **Journal of the American Oil Chemistry Society**, 56: 110 - 120, 1979.
- YU, K. F.; DEYNZE, A. V.; PAULS, K. P. Random, amplified polymorphic DNA (RAPD) analysis. **Methods in plant molecular biology and biotechnology**. New York :CRC, 287 – 301, 1993.
- ZENG, Z. B. Theoretical basis for separation of multiple linked gene effect in mapping quantitative trait loci. **Proc. National Academy Science. USA**, 90, 20: 10972 – 10976, 1993.
- ZENG, Z. B. Precision mapping for quantitative trait loci. **Genetics**, 136, 9: 1457 – 1468, 1994.

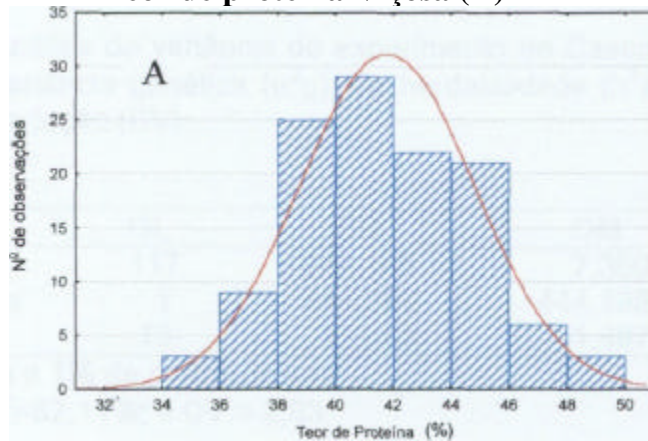
## 7. ANEXO

Quadro 8 – Teste qui-quadrado da segregação dos marcadores microsatélites polimórficos identificados por SOARES (2000). A letra A, representa o alelo presente em Garimpo e a letra B, o alelo presente em BARC-8.

“Primer”	Alelo		$\chi^2$	P(%)	“Primer”	Alelo		$\chi^2$	P(%)
	A	B				A	B		
Satt177	45	66	0,99	31,89	Satt300	59	56	0,02	88,88
Satt144	47	63	0,58	44,56	Satt304	59	53	0,08	77,68
Satt055	48	63	0,51	47,65	Satt325	62	54	0,14	71,03
Satt186	55	53	0,01	92,32	Satt333	55	63	0,14	71,27
Sat-091	46	67	0,98	32,33	Satt345	65	51	0,42	51,93
Sat-135	67	48	0,78	34,57	Satt369	66	52	0,42	51,57
Sat-105	55	60	0,05	81,57	Satt318	57	60	0,02	88,97
Sat-084	43	70	1,161	20,41	Satt324	54	56	0,036	84,87
Sat-085	55	59	0,04	85,14	Satt384	58	56	0,035	85,14
Sat-112	54	53	0,00	96,14	Satt353	57	58	0,08	92,57
Satt190	54	56	0,01	92,40	Satt377	57	58	0,08	92,57
Satt164	64	48	0,57	44,97	Satt399	55	61	0,08	78,06
Satt167	48	62	0,45	50,45	Satt422	55	61	0,08	78,06
Satt183	55	61	0,08	78,06	Satt441	56	59	0,02	88,88
Satt153	68	46	1,06	30,29	Satt449	53	62	0,704	40,13
Satt146	46	61	0,53	46,84	Satt462	59	58	0,009	92,63
Satt181	55	55	0,00	100	Satt468	59	58	0,00	96,31
Satt184	58	47	0,29	59,14	Satt475	59	53	0,321	57,07
Satt012	63	50	0,37	54,09	Satt476	59	57	0,034	85,26
Satt129	55	62	0,10	74,63	Satt495	63	55	0,14	71,27
Satt123	60	42	0,79	37,29	Satt506	59	59	0	100
Satt147	53	58	0,06	81,08	Satt514	48	62	0,44	50,83
Satt139	52	72	0,877	31,24	Satt520	62	54	0,11	74,41
Satt193	62	52	0,22	63,96	Satt522	55	61	0,08	78,66
Satt199	57	54	0,02	88,68	Satt523	51	60	0,18	66,93
Satt216	51	56	0,06	80,90	Satt526	54	62	0,14	71,03
Satt226	53	52	0,010	92,22	Satt527	52	58	0,08	77,48
Satt232	53	52	0,010	92,22	Satt528	58	57	0,09	92,570
Satt241	59	41	0,81	36,81	Satt529	59	59	0	100
Satt276	53	55	0,01	92,33	Satt531	68	49	0,77	37,98
Satt277	67	47	0,88	34,90	Satt538	61	57	0,03	85,39
Satt281	56	54	0,01	92,40	Satt549	52	65	0,36	59,79
Satt303	49	57	0,15	69,76					

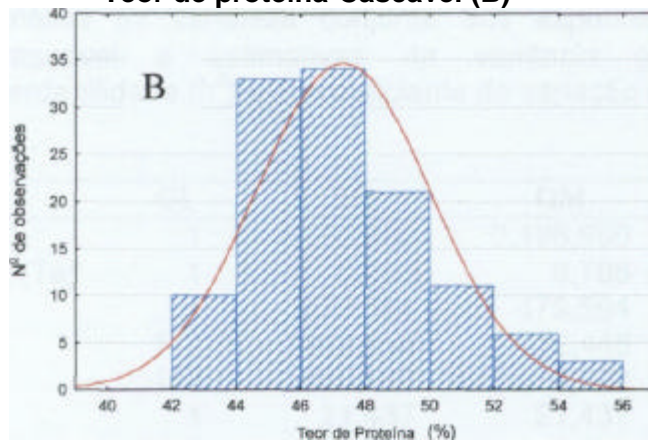
## Análise do teor de proteína

### Teor de proteína-Viçosa (A)



— Normal  
esperado

### Teor de proteína-Cascavel (B)



— Normal  
esperado

Figura – 4 Distribuição dos teores de proteína em 118 linhas endogâmicas recombinantes (RILs) cultivadas em Viçosa-MG (A) e Cascavel-PR (B). Determinado por SOARES, 2000

Quadro 9 - Análise de variância do experimento de Viçosa e estimativas da variância genética ( $\sigma^2g$ ), da herdabilidade ( $h^2$ ) e do coeficiente de variação (CV). Resultados apresentados por SOARES, 2000

FV	GL	SQ	QM	F
Famílias	117	1065,879	9,110	3,769**
Testemunhas	1	1079,238	1079,238	446,520**
Resíduo	15	36,261	2,417	

\*\*Significativo a 1% de probabilidade.

$\sigma^2 = 6,693$ ;  $h^2 = 73,4\%$ ;  $CV = 3,28$ .

Quadro 10 - Análise de variância do experimento de Cascavel e estimativas da variância genética ( $\sigma^2g$ ), da herdabilidade ( $h^2$ ) e do coeficiente de variação (CV)

FV	GL	SQ	QM	F
Famílias	117	861,145	7,360	4,916**
Testemunhas	1	444,398	444,398	296,859**
Resíduo	13	19,463	1,497	

\*\*Significativo a 1% de probabilidade.  $\sigma^2g = 6,043$ ;  $h^2 = 82,11\%$ ;  $CV = 2,93$ .

Quadro 11 - Análise de variância conjunta dos experimentos de Viçosa e Cascavel e estimativas da variância genética ( $\sigma^2g$ ), da herdabilidade ( $h^2$ ) e do coeficiente de variação (CV)

FV	GL	SQ	QM	F
Ambiente (A)	1	2195,980	2195,980	1143,14**
Testemunha (Te)	1	8,766	8,766	4,56*
TeXA	1	479,594	479,594	249,66**
Famílias (F)	117	1456,436	12,448	3,095**
FXA	117	470,587	4,022	2,09*
Grupo	1	21,437	21,437	11159**
Resíduo	29	55,724	1,921	
Total	267			

\*Significativo a 5% de probabilidade. \*\*Significativo a 1% de probabilidade.  $CV=3,10\%$ ;  $\sigma^2=5,236$ ; e  $h^2=73,16\%$ .