

TAÍS CRISTINA BASTOS SOARES

MAPEAMENTO DE LOCOS ASSOCIADOS AO CONTEÚDO DE PROTEÍNAS
DE RESERVA EM SOJA

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

VIÇOSA
MINAS GERAIS – BRASIL
2004

TAÍS CRISTINA BASTOS SOARES

MAPEAMENTO DE LOCOS ASSOCIADOS AO CONTEÚDO DE PROTEÍNAS
DE RESERVA EM SOJA

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

APROVADA: 24 de março de 2004.

Prof. Cosme Damião Cruz
(Conselheiro)

Prof. Everaldo Gonçalves de Barros
(Conselheiro)

Dra. Cláudia Teixeira Guimarães

Dr. Ivan Schuster

Prof. Maurílio Alves Moreira
(Orientador)

A Deus.

Aos meus pais, aos meus irmãos Yaska e Thiago, e à Danizinha.

AGRADECIMENTO

A Deus, por todos os meus momentos felizes.

À Universidade Federal de Viçosa (UFV) e ao Departamento de Bioquímica e Biologia Molecular, pela oportunidade de realização do Programa de Doutorado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela concessão da bolsa de estudo.

Ao professor Maurilio Alves Moreira, pela orientação, pelo incentivo nos momentos difíceis e pela amizade.

Ao professor Everaldo Gonçalves de Barros, pela amizade e pelos ensinamentos.

Ao professor Cosme Damião Cruz, pela amizade, pelos valiosos ensinamentos e sugestões.

Ao pesquisador Ivan Schuster, pela dedicação, pelos ensinamentos e pela amizade.

À pesquisadora Cláudia Teixeira Guimarães, pelos ensinamentos e pela amizade.

À minha família, minha gratidão, minha admiração e meu respeito, por tudo.

Ao meu querido Friedemann Berger, pela compreensão, pelo incentivo constante e pelo silêncio nos momentos difíceis.

Ao Newton Deniz Piovesan, por tudo.

À Lucinete Colombo, pela amizade, pelo respeito, e pela nossa Mariana.

Ao casal Márcia e Adésio, pela amizade e pela ajuda nas análises estatísticas.

Ao Luciano, pela amizade e pela troca de conhecimentos.

Às minhas amigas Analina, Aline, Ariane, Roberta, Shirley, Dili, Sandra, Denise Bazzolli, Adriane e Isabela, pelo apoio, mesmo que a distância, e pela amizade inabalável.

Aos meus grandes amigos Vanderlei e Marcílio, pela palavra amiga nos momentos difíceis.

Ao professor Elio Hideo Babá, pela confiança depositada em mim.

Aos meus colegas Valéria, Anna, Rita, Telma, Andréia, Inês, Cândida, Francisco, Simone, Maria Fernanda, Klever, Demerson, Vilmar, Magda, Raniele, Carlos, Fábio, Gerardo, Vagner, Márcio, Pedro Ivo, Fran, Márcia, Thiago, Josie e a todos os meus amigos do Instituto de Biotecnologia Aplicada à Agropecuária (Bioagro), pelo convívio.

Aos meus amigos de Ouro Preto e à República TOKA, pelo apoio nos momentos difíceis.

A todos os professores e colegas do Curso de Bioquímica Agrícola, pela agradável convivência.

Aos funcionários do BIOAGRO Jander, João Paulo, Reginaldo, Sandra, Aloísio, Gláucia e Fausto, pelo apoio.

A todos aqueles que, de alguma forma, contribuíram para a realização deste trabalho.

BIOGRAFIA

TAÍS CRISTINA BASTOS SOARES, filha de João Lopes Soares e Ana Maria Bastos Soares, nasceu em 27 de março de 1974, em Manhuaçu, Estado de Minas Gerais, Brasil.

Em janeiro de 1998, formou-se em Farmácia – Habilitação Análises Clínicas pela Universidade Federal de Ouro Preto, em Ouro Preto, Minas Gerais.

Em março de 1998, ingressou no Programa de Mestrado em Agroquímica da Universidade Federal de Viçosa, em Viçosa, MG, tendo defendido tese em fevereiro de 2000.

Em março de 2000, ingressou no Programa de Doutorado em Bioquímica Agrícola da Universidade Federal de Viçosa, em Viçosa, MG, submetendo-se à defesa de tese em março de 2004.

CONTEÚDO

	Página
RESUMO	x
ABSTRACT.....	xii
1. INTRODUÇÃO.....	1
2. REVISÃO DE LITERATURA	5
2.1. Base genética da soja.....	5
2.2. Proteínas de soja.....	6
2.3. Herdabilidade do teor de proteínas.....	8
2.4. Marcadores moleculares.....	10
2.5. Mapeamento genético.....	13
2.5.1. Construção de mapas genéticos.....	15
2.5.2. Mapeamento de caracteres de herança quantitativa.....	16
3. REFERÊNCIAS BIBLIOGRÁFICAS.....	19
CAPÍTULO 1	
ANÁLISE QUANTITATIVA DO CONTEÚDO DE PROTEÍNA TOTAL E DAS PRINCIPAIS PROTEÍNAS DE RESERVA (11S E 7S) EM SOJA PARA FINS DE MAPEAMENTO GENÉTICO.....	25
RESUMO.....	26
1. INTRODUÇÃO.....	28
2. MATERIAL E MÉTODOS	31
2.1. Material genético	31
2.2. Análise do conteúdo de proteína.....	32
2.2.1. Análise do conteúdo de proteína total	32
2.2.2. Análise das proteínas 11S e 7S e suas respectivas subunidades.....	32
2.3. Testes de Normalidade.....	35
2.3.1. Teste de assimetria.....	35
2.3.2. Teste de curtose.....	35
2.3.3. Teste de Lilliefors.....	36
2.4. Análise de variância.....	37
2.4.1. Análise de variância individual.....	37

2.4.2. Análise de variância conjunta.....	39
2.5. Correlação genotípica.....	41
3.RESULTADOS.....	43
3.1. Análise do teor de proteína total e das proteínas 11S e 7S e suas respectivas subunidades.....	43
3.2. Análise de variância.....	49
3.2.1. Análise de variância individual.....	49
3.2.2. Análise de variância conjunta.....	51
3.3. Correlações genotípicas.....	54
4. RESUMO E CONCLUSÕES	58
5.REFERÊNCIAS BIBLIOGRÁFICAS	60
CAPÍTULO 2	
MAPEAMENTO DE LOCOS QUE CONTROLAM O CONTEÚDO DE PROTEÍNA TOTAL E DAS PRINCIPAIS PROTEÍNAS DE RESERVA EM SOJA.....	
RESUMO.....	63
1. INTRODUÇÃO.....	64
2. MATERIAL E MÉTODOS.....	67
2.1. Material genético.....	71
2.2. Extração de DNA.....	72
2.3. Condições de amplificação e de eletroforese	73
2.4. Construção do mapa de ligação genética	74
2.5. Identificação e mapeamento de QTL's.....	75
2.5.1. Determinação do ponto de corte para o mapeamento de QTL's.....	77
3. RESULTADOS.....	78
3.1. Detecção de polimorfismos e seleção dos <i>primers</i> de microsatélites, RAPD e AFLP	78
3.2. Mapa de ligação genética.....	82
3.3. Análise de QTL's.....	86
3.3.1. Análise de marca simples.....	86
3.3.2. Mapeamento por intervalo composto.....	94
4. RESUMO E CONCLUSÕES	108

5.REFERÊNCIAS BIBLIOGRÁFICAS	111
------------------------------------	-----

RESUMO

SOARES, Taís Cristina Bastos, D. S., Universidade Federal de Viçosa, março de 2004. **Mapeamento de locos associados ao conteúdo de proteínas de reserva em soja.** Orientador: Maurilio Alves Moreira. Conselheiros: Cosme Damião Cruz e Everaldo Gonçalves de Barros.

Uma população de 118 RILs (linhagens recombinantes endogâmicas) foi obtida por meio do cruzamento entre o acesso BARC-8 (com alto teor de proteína) e a variedade brasileira Garimpo (com teor de proteína normal). Na geração F_9 foram abertas linhas, que constituíram o material genético utilizado neste trabalho. Noventa e cinco dessas RILs, cultivadas em dois ambientes distintos: Viçosa, MG e São Gotardo, MG, foram analisadas para as seguintes características: teor de proteína total em sementes de soja, teor das proteínas de reserva 11S e 7S e de suas respectivas subunidades, e da relação proteína total menos proteínas de reserva (PT-(11S+7S)). Quase todas estas características apresentaram uma distribuição aproximadamente normal ($P < 0,01$), com exceção de proteína total em Viçosa. As herdabilidades das características avaliadas foram altas nos dois locais, entretanto, na análise conjunta, as herdabilidades foram menores devido ao efeito da interação genótipo X ambiente. Amostras do DNA das 118 RILs foram amplificadas com *primers* microssatélites, RAPD e AFLP. Análise de associação de marcas simples, regressão múltipla e mapeamento por intervalo composto foram utilizados para detectar e mapear as regiões genômicas associadas com estas características. Foram encontrados QTLs associados ao conteúdo de proteína total (GL I, que explica 14,74% da variação desta característica), ao conteúdo da subunidade α (GL G e GL C2, que explicam em conjunto 28,83% da variação desta característica), ao conteúdo da subunidade β (GL G e dois no grupo K, que explicam em conjunto 32,62% da variação desta característica), ao conteúdo de PT-(11S+7S) (GL D1b, que explica 13,71% da variação desta característica) e ao conteúdo de 7S (GL G e GL E) que explicam em conjunto 21,86% da variação desta característica no experimento de Viçosa. No experimento de São Gotardo, foram encontrados QTLs associados ao conteúdo de proteína total (GL 3 e I, que explicam em conjunto 11,42% da variação desta característica), ao conteúdo da subunidade α' (GL A2, que

explica 10,64% da variação desta característica), ao conteúdo das subunidades ácidas (GL 3, que explica 12,06% da variação desta característica), ao conteúdo de PT-(11S+7S) (GL I, que explica 11,34% da variação desta característica) e ao conteúdo de 11S (dois QTLs em diferentes fragmentos do GL K) que explicam em conjunto 20,88% da variação desta característica. Os marcadores associados ao teor de proteína foram diferentes de um local para outro, o que indica interação entre o genótipo e o ambiente. O grupo de ligação I foi associado a QTLs para proteína total em Viçosa e em São Gotardo. Este fato sugere que os locos associados presentes neste grupo de ligação contêm regiões com genes que são expressos em diferentes ambientes.

ABSTRACT

SOARES, Taís Cristina Bastos, D. S., Universidade Federal de Viçosa, March, 2004. **Mapping loci associated with storage protein content of soybeans.** Adviser: Maurilio Alves Moreira. Committee members: Cosme Damião Cruz and Everaldo Gonçalves de Barros

A 118 soybean RILs (recombinant inbred lines) population was obtained by crossing the BARC-8 access (cultivar with high protein content) and the Brazilian variety Garimpo (with normal protein content). Lines from the F_3 generation constituted the genetic material used in this work. Ninety-five RILs cultivated in two different environments, Viçosa - MG and São Gotardo - MG, were analyzed following the characteristics: total protein content, storage protein content 11S and 7S and their respective subunits and relation of total protein minus storage proteins (PT-(11S+7S)). Almost all these characteristics presented a distribution approximately normal ($P < 0,001$), except total protein in Viçosa. The heritability values of the characteristics were high in both places, however in a combined analysis; the heritability values were lower due to the effect of genotype x environment interaction. DNA samples of the 118 RILs were amplified with microsatellite primers, RAPD and AFLP. Analysis of association of simple markers, multiple regression and composed interval mapping were used to detect and map the genomic regions associated with these traits. QTLs (quantitative trait loci) associated with total protein content were found in GL I, that explains 14,7 % of the variation of this characteristic, with content of the α subunit in GL G and GL C2, that together explain 28,83% of the variation of this characteristic, with content of the β subunit in GLG and two in group K, that together explain 32,62 % of the variation of this characteristic, with PT-(11S+7S) in GL D1b, that explains 13,71 % of the variation of this characteristic, and with content of 7S in GL G and GL E that together explain 21,86 % of the variation of this characteristic in the experiment conducted in Viçosa. QTLs associated with total protein content were found in the experiment conducted in São Gotardo in GL 3 and GL I, that together explain 11,42% of the variation of this characteristic, with content of the α' subunit in GL A2, that explain 10,64 % of the variation of this characteristic, with

content of acid subunits in GL 2, that explain 12,06 % of the variation of this characteristic, with PT-(11S+7S) in GL I, that explain 11,34 % of the variation of this characteristic and with the 11S content two QLTs in different GLK fragments that explain 20,88 % of the variation of this characteristic. The markers associated with protein content varied in different environment, indicating genotype and environment interaction. The linkage group I was associated with QTLs to total protein in Viçosa and São Gotardo. This suggests that total protein content loci associated with this linkage group contain regions with genes that express in different environments.

1. INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill) constitui atualmente um dos produtos agrícolas de maior importância no Brasil, ocupando lugar de realce na pauta de exportação do país e sendo cultivada praticamente em todo o território nacional. Isso se deve, principalmente, ao grande número de cultivares disponíveis e adaptados a todas as regiões brasileiras. Assim, técnicas que venham auxiliar na caracterização desses cultivares e direcionar a obtenção de novos outros, são de fundamental importância para a manutenção deste potencial produtivo nacional.

O grande interesse mundial pelo cultivo da soja deve-se à excelente combinação que ela apresenta em termos de produtividade (geralmente superior a 2.000 Kg/ha) e porcentagens de óleo e proteínas, em média, 20 e 40% do peso das sementes, respectivamente (VELLO,1985).

Historicamente, a soja foi melhorada geneticamente com o intuito de aumentar a produtividade e ao conteúdo de óleo nos grãos. Recentemente, alguns programas de melhoramento de soja têm dado ênfase ao desenvolvimento de variedades mais produtivas, contendo altos teores de proteína.

A fração protéica e seus derivados apresentam restrição quanto ao seu consumo pela população brasileira. Essa barreira alimentar aos derivados protéicos da soja está associada à baixa qualidade do grão utilizado no processamento industrial. Um aspecto referente à qualidade do grão de soja é o teor e a qualidade de proteína. A qualidade protéica refere-se ao teor de aminoácidos sulfurados e também às propriedades funcionais da proteína.

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

Muitos caracteres de importância agronômica apresentam baixas herdabilidades ou são de difícil medição e identificação, podendo a eficiência

de sua seleção ser aumentada pelo uso de uma característica auxiliar de alta herdabilidade que esteja correlacionada com o caráter principal.

A possibilidade de prever ganhos por intermédio de estratégias de seleção constitui-se em uma das mais importantes contribuições da Genética Quantitativa para o melhoramento genético, pois as informações obtidas permitem decidir, com base científica e sem custo adicional, sobre as técnicas de seleção, tornando mais eficiente o processo de obtenção de novos cultivares (MIRANDA, 2002).

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. O primeiro mapa genético foi construído com número de grupos de ligação bem próximo ao de cromossomos, no qual foram definidos 21 grupos contendo 81 marcadores RFLP e 13 marcadores RAPD (WILLIAMS *et al.*, 1990). Mais recentemente foi desenvolvido um mapa com 24 grupos de ligação contendo 359 marcadores RFLP, no qual foram mapeadas nove regiões ligadas a características quantitativas e dez ligadas a caracteres qualitativos (SHOEMAKER, 1994).

Os primeiros mapas disponíveis para a soja foram, na sua maioria, feitos a partir de cruzamentos interespecíficos entre *Glycine max* e *Glycine soja*. O primeiro trabalho de mapeamento intra-específico de *Glycine max* foi apresentado em 1993, tendo sido construído a partir de marcadores RFLP (LARK *et al.*, 1993). Embora o cruzamento interespecífico seja interessante para a construção de mapas em si, não o é para os programas de melhoramento que se baseiam exclusivamente em genótipos de *G. max* (MANSUR *et al.*, 1993).

CREGAN *et al.* (1999) construíram um mapa integrado para a soja utilizando dados de três mapas diferentes: USDA/Iowa State (*G. max* x *G. soja*), Univ. de Utah (Minsoy x Noir) e da Univ. de Nebraska (Clark x Harosoy). Esse mapa integrado foi definido, utilizando-se um total de 606 *primers* SSR e outros marcadores (RFLP, RAPD, AFLP, isoenzimas e marcadores clássicos), no total de 1.423 marcadores. Foram definidos 20 grupos de ligação-consenso.

A seleção de características agrônômicas 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. Segundo HAYWARD *et al.* (1994), o método mais eficaz para este tipo de associação envolve a produção de mapas genéticos detalhados nos quais são localizadas regiões contendo locos para características quantitativas (QTLs).

Segundo FEHR (1978), altura de plantas, produção de sementes, resistência ao acamamento, percentagem de óleo e de proteína e tamanho das sementes são características quantitativas.

WILCOX & CAVINS (1995), estudando correlações entre teor de proteína e produção de grãos, por meio de uma série crescente de retrocruzamentos entre dois progenitores contrastantes para estas duas características, verificaram correlação negativa decrescente entre teor de proteína e produção de grãos, à medida que se avançaram os retrocruzamentos, obtendo ao final do 3º retrocruzamento linhagens produtivas e com alto teor de proteína.

As sementes de soja contêm células especiais que produzem e armazenam as proteínas de reserva na semente. As sementes produzem centenas de vacúolos de reserva de proteínas. Mas, poucas proteínas principais são armazenadas na semente de soja. Duas proteínas de reserva, a β -conglícinina e glicínina constituem cerca de 70% das proteínas de reserva em soja. Geralmente, a glicínina é a mais abundante (MOREIRA *et al.*, 1979; COATES *et al.*, 1985).

A quantidade destas duas proteínas pode variar nas diferentes sementes de soja, dependendo da constituição genética da semente e do ambiente onde elas foram cultivadas. Onde a aplicação de produtos de proteína requer maior quantidade de aminoácidos sulfurados, um maior conteúdo de glicínina é preferido. Recentemente, a fração proteína de soja tem se tornado mais importante que a fração óleo. Têm sido encontradas diferenças nos conteúdos dos componentes da β -conglícinina e glicínina em linhagens de soja com alto teor de proteína. Estas diferenças podem ser devidas a alguns mecanismos de regulação. Esta informação pode ser usada para aumentar seletivamente o tipo de proteína de reserva para usos específicos. Soja utilizada na indústria de suínos deve conter mais β -

conglucina e a soja utilizada na indústria de aves deve conter mais glicina (YAKLICH *et al.*, 1999).

Pesquisas biotecnológicas envolvendo a manipulação genética dos níveis destas duas proteínas estão sendo empregadas com o intuito de aumentar os níveis de aminoácidos essenciais nas formulações de rações tanto para consumo animal quanto para consumo humano (KWANYUEN *et al.*, 1997).

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 MIRANDA, 2002), e também identificar QTLs associados às características: conteúdo de proteína total em sementes de soja, conteúdo das proteínas de reserva 11S e 7S e de suas respectivas subunidades, e da relação proteína total menos proteínas de reserva (PT-(11S+7S)). Para isso, foram acrescentados ao mapa original, marcadores microssatélites não mapeados anteriormente e marcadores AFLP. 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. Base genética da soja

A soja (*Glycine max* (L.) Merrill), leguminosa originária da China, é uma espécie autógama, com 20 pares de cromossomos (SHOEMAKER *et al.*, 1992).

Os relatos da introdução da soja no Brasil datam de 1882, na Bahia, e de 1890, em Pelota-RS, quando foi testada como produtora de forragem (HASSE, 1996). No Brasil, já foram desenvolvidas variedades adaptadas ao cultivo em diferentes latitudes, solos e condições climáticas. Entretanto, vários fatores limitam a obtenção de altos rendimentos por estes cultivares (CERVIGNI, 2003).

Embora o número de cultivares existentes no Brasil seja bastante elevado, a base genética que constitui o germoplasma é bastante restrita, situação semelhante ao que ocorre nos Estados Unidos (KEIM *et al.*, 1989). Provavelmente, essa baixa diversidade genética seja devido ao fato da soja ser uma espécie autógama, com indivíduos altamente homozigotos; e porque a maioria dos cultivares apresenta ancestrais comuns. A baixa diversidade do germoplasma brasileiro foi confirmada por ABDELNOOR *et al.* (1995).

A diversidade genética em soja tem sido estudada com a aplicação de diferentes informações, como características morfológicas, coeficiente de parentesco, isoenzimas e marcadores moleculares do tipo RFLP, RAPD, AFLP e microssatélites (GIZLICE *et al.*, 1993; MAUGHAN *et al.*, 1996). GIZLICE *et al.* (1993) citaram que, nos Estados Unidos, a base genética das variedades recomendadas é derivada de 15 progenitores.

De modo geral, os trabalhos de diversidade genética em soja têm como objetivos caracterizar variedades e buscar informações sobre a constituição genética de germoplasmas e, de modo prático, são conduzidos para verificação de genealogias, no planejamento de cruzamentos e na exploração da heterose (PIOVESAN, 2000).

2.2. Proteínas de soja

Proteínas de reserva são definidas como sendo aquelas proteínas que possuem como única função conhecida servirem como reserva de aminoácidos, as quais, durante a germinação da semente, são degradadas por enzimas proteolíticas para fornecerem os aminoácidos necessários ao desenvolvimento do embrião (TEIXEIRA, 2003).

As proteínas de soja constituem cerca de 35 a 45% da matéria seca da semente. As proteínas de reserva, glicinina (11S) e β -conglucina (7S), representam em torno de 70% da proteína do grão de soja (HILL e BREIDENBACH, 1974).

A glicinina apresenta massa molecular de aproximadamente 350 kDa, sendo constituída de polipeptídios de caráter ácido e de caráter básico, que diferem da β -conglucina em seus conteúdos de metionina (MOREIRA *et al.*, 1979; COATES *et al.*, 1985). É constituída de seis polipeptídios ácidos (A1a, A1b, A2, A3, A4 e A5) e cinco básicos (B1a, B1b, B2, B3 e B4), que se encontram associados, de modo específico, por meio de ligações bissulfídicas entre um polipeptídio ácido e um básico, formando as subunidades A1aB2, A1bB1b, A2B1a, A3B4 e A5B3 (MOREIRA *et al.*, 1979, 1981). O polipeptídio 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) sendo os seus respectivos genes, *Gy1*, *Gy2*, *Gy3*, *Gy4* e *Gy5*. Os polipeptídios A1a, A1b, A2, B1 e B2 possuem de três a seis vezes mais metionina que A3, A4, A5, B3 e B4 (MOREIRA *et al.*, 1979). Em glicinina, têm sido observadas variações na composição polipeptídica, como mutantes naturais com ausência do complexo G4 (FONTES *et al.*, 1984).

As cinco subunidades da glicinina podem ser divididas em dois grupos baseados na homologia da sequência de aminoácidos (NIELSEN, 1989). G1, G2 e G3 são as subunidades do Grupo 1 e G4 e G5, as subunidades do Grupo 2. A homologia das seqüências entre subunidades de um mesmo grupo está em torno de 90%, mas está em torno de somente 50% quando membros de grupos diferentes são comparados.

O Grupo 1 é codificado por três genes presentes em dois domínios cromossomais distintos. Um domínio contém dois genes de glicinina ligados (*Gy1* e *Gy2*) e o outro, o gene *Gy3*. Os dois domínios têm alto grau de homologia e contêm, no mínimo, cinco genes cada, que são expressos em embriões ou em folhas de plantas maduras. Dois outros genes foram identificados e designados como glicinina-relacionados, podendo codificar outras famílias de subunidades de glicinina que se acumulam em menor quantidade nas sementes. Os genes *Gy1*, *Gy2* e *Gy3* não estão ligados a outros genes de proteína de semente, mas estão ligados a outros genes significativamente divergentes dos Grupos 1 e 2 (NIELSEN, 1989). DIERS *et al.* (1994) mapearam os genes *Gy4* e *Gy5* da glicinina em uma população obtida do cruzamento interespecífico de *Glycine max* x *G. soja*, verificando que esses genes estão localizados nos grupos de ligação “O” e “F”, respectivamente.

A β -conglícinina é uma glicoproteína com massa molecular estimada em 150-170 kDa e possui três subunidades, α' , α e β , sendo codificadas pelos genes *Cgy1*, *Cgy2* e *Cgy3*, respectivamente (THANH e SHIBASAKI, 1977). As subunidades α e α' possuem dois e três resíduos de metionina, respectivamente, e a subunidade β não possui esse aminoácido (COATES *et al.*, 1985). Existe polimorfismo genético na composição polipeptídica de β -conglícinina, tendo sido identificado genótipo com ausência de α' (KITAMURA *et al.*, 1984). Subunidades menores dessa globulina, designadas β' , γ e δ , também têm sido descritas (COATES *et al.*, 1985). HARADA *et al.* (1989) concluíram que a família de genes da β -conglícinina contém no mínimo 15 membros que estão distribuídos entre seis regiões distintas do genoma. Três dessas regiões contêm genes múltiplos (região A, cinco genes; região B, dois; região C, três; e regiões D, E e F, um cada).

A proteína da soja é inferior em qualidade quando comparada com a proteína animal no que tange ao teor de aminoácidos sulfurados metionina e cisteína, que estão presentes nesta leguminosa em quantidades limitantes. O nível de metionina na soja situa-se entre 1,0 a 1,9 %, representando cerca de 43% do nível padrão considerado pela FAO (WOLF e COWAN, 1975).

Com o objetivo de determinar se um aumento do acúmulo de proteínas de reserva pode ser detectado utilizando a eletroforese em gel de poliacrilamida em condições desnaturantes (SDS-PAGE), acompanhada da densitometria e “immunoblotting” estimando-se a quantidade das subunidades das proteínas β -conglucina e glicina, YAKLICH (2001) estudou linhagens de soja com teores normal e alto de proteína. Os dados apresentados neste estudo mostraram que existem diferenças significativas nas subunidades das proteínas 7S e 11S das linhagens contendo alto teor de proteína. Algumas das linhagens estudadas apresentaram um acúmulo das subunidades α , α' e β da β -conglucina, enquanto outras apresentaram um acúmulo da subunidade A3 da glicina. Os polipeptídeos básicos da glicina também se acumularam na maioria das linhagens. Baseado nestas observações, o autor concluiu que linhagens com alto teor de proteína contêm maiores concentrações de β -conglucina e glicina do que linhagens normais.

A mesma metodologia foi usada também por FEHR *et al.* (2003), visando estudar a influência do genótipo, do ambiente e interações genótipo x ambiente sobre a composição de subunidades das proteínas de reserva da soja. Estes autores, trabalhando com 14 cultivares, alguns de uso amplo e outros de uso específico, como cultivares livre de lipoxigenases e cultivares com alto teor de proteína, que foram cultivados em 8 localidades diferentes por 3 anos consecutivos, concluíram que a concentração de 7S, 11S e a razão 11S/7S são influenciadas pelo ambiente. Por outro lado, ressaltaram que não são esperadas alterações significativas nestes parâmetros de um ano para o outro a nível industrial, em virtude dos grãos colhidos em localidades próximas serem misturados nas indústrias de processamento de alimentos.

2.3. Herdabilidade do teor de proteínas

Herdabilidade é definida como a fração da variância fenotípica a ser transmitida para a progênie ou, em termos de unidade de seleção, a fração do diferencial de seleção esperado pelo ganho quando a seleção é praticada em uma referida unidade. A estimativa do coeficiente de herdabilidade atende a

duas finalidades básicas no melhoramento genético, a de demonstrar a relativa facilidade com que diferentes caracteres são selecionados em determinado esquema de melhoramento e a de permitir que estimativas dos progressos esperados com a seleção sejam obtidas (HANSON e WEBER, 1962).

Para FALCONER (1987), a herdabilidade de um caráter métrico é uma das mais importantes propriedades de uma população, porque expressa a proporção da variância total que é atribuída ao efeito médio dos genes, o que determina o grau de semelhança entre aparentados. A seleção é efetuada, fundamentalmente, nessa semelhança.

STRICKBERGER (1985), destaca que a importância das estimativas de herdabilidade reside em sua relação com a seleção, uma vez que quanto maior for a herdabilidade, maior será o sucesso de seleção para dada característica, o que resulta numa relação positiva entre alta herdabilidade e eficiência de seleção. Para os progenitores transmitirem certas características aos seus descendentes que sejam de algum modo predizíveis, é necessário que a variância ambiental seja baixa e variância genética seja alta.

Na literatura, são descritas várias metodologias para obtenção de estimativas do coeficiente de herdabilidade. A utilização da variabilidade entre as plantas na geração F_2 em relação à variabilidade entre as plantas dos progenitores foi utilizada por MAHAMUD e KRAMER (1951) para obtenção do coeficiente de herdabilidade. WARNER (1952) propôs uma metodologia para estimar a herdabilidade, com base nos progenitores, nas gerações F_1 , F_2 e dos dois retrocruzamentos (recorrente e doador). HANSON e WEBER (1962) determinaram estimativas do coeficiente de herdabilidade com base nos componentes de variância. FREY e HORNER (1957) propuseram estimar a herdabilidade em unidades-padrão, as quais são obtidas, calculando-se a regressão sobre os dados codificados em termos de desvio-padrão. SMITH e KINMAN (1965) utilizaram fatores de correção do coeficiente de regressão pai-filho para estimar a herdabilidade, usando progenitores endogâmicos para algumas gerações de autofecundação.

DUDLEY e MOLL (1969) discutiram que a estimativa de herdabilidade refere-se a uma característica e a uma população e é específica para as

condições experimentais nas quais os genótipos foram estudados, sendo difícil generalizar estimativas de uma população para outra ou em diferentes condições experimentais.

KWON e TORRIE (1964), procurando estimar a herdabilidade para caracteres de interesse em soja, estudaram duas populações derivadas da seleção de 64 plantas F_2 para cada cruzamento. Avaliaram essas populações na geração F_3 em um ano e um local e as gerações F_4 e F_5 em dois anos e dois locais. Os parâmetros genéticos extraídos da análise de variância proporcionaram estimativas de herdabilidade livre dos efeitos da interação genótipo $\times$ ambiente. Esses autores verificaram que a interação genótipo $\times$ ano foi maior que a interação genótipo $\times$ local, em muitas características, principalmente produção de sementes. A herdabilidade para produção de sementes foi mais baixa devido à maior interação genótipo $\times$ ano, ficando na ordem de 3 a 10%. Para os caracteres teor de óleo e teor de proteína, a herdabilidade só foi estimada para um dos cruzamentos, apresentando valores de 51 e 57%, respectivamente.

Objetivando estudar os componentes genéticos envolvidos no controle do teor de proteína e em outros 8 caracteres agronômicos, PIOVESAN (2000) encontrou, com base nos componentes de variância, estimativas de herdabilidade, para teor de proteína nas 6 populações F_2 estudadas, variando de 22 a 81%.

2.4. Marcadores moleculares

Até meados da década de 60, os marcadores utilizados em estudos de genética e melhoramento eram controlados por genes associados a caracteres morfológicos, em geral fenótipos de fácil identificação visual, como nanismo, deficiência clorofítica, cor de pétala ou morfologia foliar. Entretanto, só ocasionalmente é que marcadores morfológicos ligados a genes de importância econômica eram identificados, limitando o seu emprego em programas de melhoramento. Além disso, a disponibilidade destes marcadores era

essencialmente restrita às poucas espécies de plantas utilizadas como sistemas modelos para o estudo da genética (FERREIRA e GRATAPAGLIA, 1998).

A revolução neste quadro iniciou-se com o desenvolvimento de marcadores isoenzimáticos. O número de marcadores genéticos foi grandemente ampliado e a aplicabilidade da técnica passou a incluir potencialmente todos os programas de melhoramento (TANKSLEY e HEWITT., 1988).

Com o advento das técnicas de biologia molecular, surgiram marcadores genéticos do DNA, destacando-se RFLP ("Restriction Fragment Length Polymorphisms"), RAPD ("Random Amplified Polymorphic DNA"), AFLP ("Amplified Fragment Length Polymorphism") e microssatélites. Eles baseiam-se na variação natural da seqüência de bases do DNA, possuem segregação mendeliana e tornam possível a construção de mapas de ligação genética de alta densidade.

Esses marcadores apresentam várias vantagens com relação aos marcadores morfológicos. Os marcadores moleculares apresentam-se em número elevado e com alto grau de polimorfismo. Além disso, não são influenciados por condições ambientais, pois o polimorfismo é verificado diretamente no DNA (Evola *et al.*, 1986; Beckman e Soller, 1986; Gallais, 1984, citados por MALUF, 1990).

Os marcadores RFLP foram os primeiros marcadores de DNA a serem utilizados na construção de mapas genéticos da espécie humana (BOTSTEIN *et al.*, 1980). Posteriormente, os RFLPs foram amplamente empregados nas mais variadas áreas das genéticas animal e vegetal. A técnica do RFLP é baseada na digestão do DNA genômico com enzimas de restrição, que clivam o DNA em regiões específicas, na separação dos fragmentos em gel de agarose e transferência destes para membranas de nitrocelulose, onde são hibridizados com sondas de DNA marcadas. O polimorfismo é obtido quando ocorre, por mutação de ponto, perda ou surgimento de sítios de restrição ou quando ocorrem inserções ou deleções entre os dois sítios de restrição que flanqueiam uma região homóloga à sonda de DNA. Os marcadores RFLP apresentam herança co-dominante, sendo possível a identificação de indivíduos homozigotos e heterozigotos na população.

Os marcadores RAPD são fragmentos discretos de DNA, amplificados pela reação da polimerização em cadeia a partir de *primers* únicos de sequência arbitrária e conteúdo de CG maior que 50%, que se ligam ao DNA genômico em dois sítios e em fitas opostas do DNA-molde. Se esses sítios estiverem separados por uma distância amplificável (geralmente 200 a 2000 pb), os produtos de amplificação são formados. O número e o tamanho dos fragmentos amplificados dependem da sequência do *primer* usado e da sequência de bases do DNA-molde (TINGEY e DELTUSO, 1992). Os polimorfismos resultam de mudanças na molécula de DNA que alteram os sítios de ligação dos *primers*, diminuindo ou aumentando a distância entre eles. Essas mudanças podem alterar o comprimento do fragmento compreendido entre os dois sítios ou mesmo impedir que a amplificação ocorra. Os marcadores RAPD baseiam-se na reação de PCR ("Polymerase Chain Reaction"), o que torna o uso deste tipo de marcadores muito vantajoso, permitindo automação e não havendo necessidade do conhecimento prévio da sequência de DNA. No entanto, essa técnica não permite a distinção entre indivíduos heterozigotos e homozigotos, sendo os marcadores considerados de herança dominante.

Marcadores AFLP constituem uma classe de marcadores, que alia a especificidade dos sítios de restrição do RFLP com a praticidade da amplificação por PCR, apresentando-se como ferramenta poderosa na caracterização de genomas e no mapeamento genético (VOS *et al.*, 1995). A técnica baseia-se na digestão simultânea do DNA genômico com duas enzimas de restrição, sendo *Eco* RI e *Mse* I as mais usadas. Adaptadores específicos com terminais complementares às extremidades coesivas dos sítios de restrição são ligados aos fragmentos de DNA digeridos, sendo utilizado um adaptador específico para cada sítio de restrição. Esses fragmentos são submetidos a uma reação de PCR com *primers* pré-seletivos de sequência complementar aos adaptadores, acrescidos de um nucleotídeo arbitrário na sua extremidade 3'. Dessa forma, apenas os fragmentos que possuem os sítios de restrição de *Eco* RI e *Mse* I serão amplificados, aumentando a proporção dos fragmentos de interesse. A população de fragmentos pré-amplificados constituirá o DNA-molde para as reações subseqüentes de amplificação seletiva, utilizando-se *primers*

com a mesma seqüência dos *primers* pré-seletivos acrescida de dois nucleotídeos arbitrários na sua extremidade 3'. Os produtos amplificados são separados em gel de seqüenciamento, com resolução de um nucleotídeo. A detecção dos fragmentos amplificados é feita pela marcação radioativa ou fluorescente de um dos *primers* seletivos. O polimorfismo evidenciado por essa técnica é preferencialmente dominante, sendo possível, no entanto, identificar indivíduos heterozigotos pela diferença de intensidade das bandas, se esta técnica for realizada com boa precisão.

Os genomas de organismos eucariotos são caracterizados pela ocorrência de grandes quantidades de seqüências repetidas de nucleotídeos de várias classes, dentre elas uma classe de seqüências curtas de DNA (2 a 4 pb), repetidas em “tandem”, denominadas seqüências simples repetitivas (SSR) ou microssatélites (LITT e LUTY, 1989). Regiões contendo SSR são amplificadas, individualmente, por meio de PCR, utilizando-se um par de *primers* específico (de 20 a 30 bases), complementares a seqüências únicas que flanqueiam o microssatélite. Tendo em vista a expressão co-dominante e o multialelismo, os marcadores SSR são os que possuem o mais elevado conteúdo de informação genética. Por esse motivo, essencialmente toda e qualquer população segregante pode ser utilizada como população-referência para estudos de ligação e mapeamento genético (MORGANTE *et al.*, 1994). A técnica para a utilização dos marcadores microssatélites é menos trabalhosa do que a de marcadores RFLP, e os SSR constituem a classe mais informativa de marcadores moleculares (TAUTZ, 1989; WEBER e MAY, 1989). Além disso, apresentam alto grau de polimorfismo em soja (AKKAYA *et al.*, 1992). A alta informação contida em loco SSR, aliada à sua transmissão co-dominante, potencial para automação e requerimento de pequena quantidade de DNA, fazem com que os SSRs sejam considerados marcadores genéticos ideais (RAFALSKI e TINGEY, 1993).

2.5. Mapeamento genético

O mapeamento genômico de uma espécie consiste na produção de um conjunto ordenado de informações sobre as seqüências de DNA que cobrem todo o genoma. O mapeamento genético é baseado na hipótese de que a co-transmissão de dois marcadores reflete a proximidade entre eles, criando a possibilidade de se ordenar linearmente a informação genética ao longo dos grupos de ligação e determinar a distância entre os genes ligados. A determinação da distância e da ordem é feita com base na taxa de recombinação genética.

Mapas genéticos de marcadores moleculares oferecem a possibilidade de estudar a arquitetura de características quantitativas, ou seja, identificar, mapear e medir a magnitude do efeito dos principais fatores genéticos envolvidos no controle dessas características e, potencialmente, manipular esses fatores em base individual durante os procedimentos de seleção e recombinação genética. A construção de um mapa baseia-se na existência de variação na constituição genética de uma população. Os marcadores genéticos são ferramentas para conhecer e manipular essas variações, que são estudadas a partir de cruzamentos apropriados. Esses cruzamentos possibilitam a ocorrência de *crossing-over* entre marcadores genéticos e são usados para gerar as populações segregantes. Uma etapa comum da aplicação desses marcadores no mapeamento do genoma de uma espécie tem sido a integração de diversos tipos de marcadores, incluindo os morfológicos e isoenzimáticos, bem como os RAPD, RFLP e AFLP (CHAPARRO *et al.*, 1994; KEIM *et al.*, 1997). Outro marcador molecular que vem sendo muito usado na construção de mapas genéticos são os microssatélites (MORGANTE *et al.*, 1994).

KEIM *et al.* (1990a) mapearam 150 marcadores RFLP numa população de 60 indivíduos F_2 de soja, identificando 25 grupos de ligação, num total de 1.200 cM. Esse mapa foi expandido para 252 marcadores RFLP, em 31 grupos de ligação, perfazendo um total de 2.147 cM (DIERS *et al.*, 1992). Posteriormente, esse mapa foi aumentado para 359 marcadores RFLP, com 38 QTLs e 10 locos para características monogênicas (SHOEMAKER, 1994). KEIM (1997) obteve um mapa com 840 marcadores, sendo 165 RFLP, 25 RAPD e 650

AFLP, em 28 grupos de ligação, num total de 3.441 cM, partindo-se de uma população inicial de 300 RILs.

A integração entre mapas, que se baseiam em marcadores moleculares, com o mapa clássico da soja, vem sendo feita para uma série de características, com base em diversos tipos de populações (KEIM *et al.*, 1990a,b; SHOEMAKER, 1994).

CREGAN *et al.* (1999) construíram um mapa integrado para a soja utilizando dados de três mapas diferentes: USDA/Iowa State (*G. max* x *G. soja*), Univ. de Utah (Minsoy x Noir) e da Univ. de Nebraska (Clark x Harosoy). Esse mapa integrado foi definido, utilizando-se um total de 606 *primers* SSR e outros marcadores (RFLP, RAPD, AFLP, isoenzimas e marcadores clássicos), num total de 1.423 marcadores. Foram definidos 20 grupos de ligação-consenso.

2.5.1. Construção de mapas genéticos

A construção de um mapa de ligação envolve desde a aplicação de técnicas de biologia molecular até os conceitos clássicos de herança genética demonstrados por Mendel. Os requisitos básicos para a construção de um mapa genético são a obtenção de populações segregantes e a utilização de vários marcadores que sejam polimórficos entre os progenitores e que apresentem segregação mendeliana na progênie.

Mapas de ligação são construídos pela análise da segregação dos marcadores em uma população segregante, isto é, em desequilíbrio de ligação gênica. Tais populações podem ser oriundas de retrocruzamento, de F_2 ou do cruzamento entre indivíduos heterozigotos. Linhagens recombinantes endogâmicas (RILs), obtidas por autofecundações sucessivas de indivíduos F_2 até atingirem níveis desejáveis de homozigose, são também utilizadas na construção de mapas (BURR *et al.*, 1988). A escolha da população depende do hábito reprodutivo da espécie e do objetivo do estudo. A população F_2 apresenta indivíduos com todas as possíveis combinações alélicas (AA, Aa e aa), possibilitando o entendimento do mecanismo de ação gênica e das interações com outros locos. Em populações RILs, os indivíduos apresentam elevado índice

de homozigose. A ausência de heterozigose é, em parte, compensada pelos ciclos sucessivos de autofecundação, que criam maiores oportunidades de ocorrer recombinação, em comparação com populações F_2 . Nas populações RILs, os genótipos dos indivíduos são perpetuados; assim, um mesmo genótipo pode ser avaliado em locais diferentes, o que possibilita a quantificação da interação genótipo x ambiente, além do intercâmbio entre informações genéticas geradas por diferentes grupos de pesquisa.

O número de marcadores moleculares necessários para construir um mapa genético depende do tamanho do genoma, do número de cromossomos e da frequência de recombinação genética. Um mapa pode ser considerado completo quando o número de grupos de ligação obtidos pela análise dos marcadores for igual ao número de cromossomos gaméticos do organismo e quando todos os marcadores genéticos mapeados estiverem ligados, indicando que todas as regiões do genoma estão representadas (GUIMARÃES e MOREIRA, 1999).

No contexto do mapeamento de características quantitativas, os progenitores são selecionados conforme os respectivos fenótipos contrastantes para as características de interesse. Os marcadores que são capazes de detectar alelos diferentes entre as linhagens genitoras são avaliados pelo teste qui-quadrado, e aqueles que apresentarem segregação mendeliana são submetidos às análises de co-segregação. Esses marcadores são agrupados pela análise de dois pontos e, posteriormente, ordenados em cada grupo de ligação pela análise de três pontos através de vários algoritmos para construção de mapas genéticos como o MapMaker (LANDER *et al.*, 1987), GMendel (LIU e KNAPP, 1992), JoinMap (STAM, 1993) e GQMOL (www.ufv.br/dbg/gqmol/gqmol.htm).

2.5.2. Mapeamento de caracteres de herança quantitativa

Muitos caracteres de importância agrônômica estão sob o controle genético complexo, envolvendo vários genes. Esses caracteres são denominados poligênicos ou quantitativos, apresentando variação contínua do

fenótipo em vez de classes fenotípicas discretas, o que dificulta a manipulação e compreensão dos efeitos desses genes.

Para a maioria das características quantitativas, poucas informações existem sobre o número, a posição cromossômica, a magnitude do efeito e as interações dos locos que controlam a sua expressão. Esses locos são denominados QTL ("Quantitative Trait Loci"). A determinação de ligação genética entre marcadores e alelos do QTL depende da existência de desequilíbrio de ligação entre os alelos no loco marcador e alelos do QTL. A maioria das estratégias de mapeamento de QTLs está baseada no uso de populações segregantes tradicionalmente utilizadas para a construção de mapas genéticos (FERREIRA e GRATTAPAGLIA, 1996).

O uso de marcadores moleculares facilitou grandemente o estudo de características herdadas quantitativamente e tem tornado possível analisar poligenes para cada característica em fatores mendelianos individuais. Utilizando ligação molecular de mapas genéticos e a tecnologia de mapeamento de QTLs, é possível estimar o número de locos controlando a variação genética em uma população segregante e caracterizar esses locos com respeito a sua posição no genoma, efeitos fenotípicos, efeitos pleiotrópicos e interações epistáticas com outros QTLs (XIAO *et al.*, 1996).

A identificação de QTLs por meio de marcadores moleculares é conceitualmente simples. Do cruzamento entre dois genitores que contrastem para a maioria das características de interesse, é obtida uma população segregante (F_2 , retrocruzamento, RIL, dentre outras). Uma progênie (superior a 100 indivíduos) avaliada para as características de interesse e o fenótipo do marcador molecular obtido a intervalos regulares (5-20 cM), ao longo do genoma. A procura dos QTLs realiza-se mediante associações entre os marcadores segregantes e os valores fenotípicos das características. Se tal associação é identificada, significa que existe uma ligação fatorial entre o marcador e o QTL que afeta a característica (CERVIGNI, 2003).

Uma das dificuldades em detectar QTL é determinar a sua significância na metodologia estatística. Os pontos de corte para determinar a significância de um possível QTL dentro de um intervalo testado têm sido obtidos por teste de

permutação (DOERGE e CHURCHILL, 1996). Esse teste é realizado, embaralhando-se os valores da característica quantitativa que está sendo analisada e repetindo a análise. Repetindo-se tal procedimento, muitas vezes (1.000 vezes ou mais) são encontrados os valores de LOD ou razão de verossimilhança em número igual ao de testes realizados. Utiliza-se então, como ponto de corte, o valor encontrado, que se situa na posição do nível de significância adotado. Por exemplo, se forem realizadas 1.000 permutações, obtém-se um quadro com 1.000 valores de LOD. Se for adotado o nível de significância de 95%, o ponto de corte será o valor de LOD na posição 950, desde que os valores estejam agrupados em ordem crescente de valor. Isso significa que, em somente 5% das possíveis combinações de valores da característica, o valor de LOD será igual ou superior àquele adotado como ponto de corte. Portanto, a probabilidade de o marcador estar ligado ao QTL é grande. No entanto, a dificuldade computacional é elevada, pois são necessários muitos cálculos para estimar o ponto de corte, e este é válido apenas para o experimento que está sendo avaliado.

A razão de verossimilhança possui uma aproximação com o χ^2 , e um valor de χ^2 em um nível de significância escolhido pode ser adotado como ponto de corte.

3. REFERÊNCIAS BIBLIOGRÁFICAS

- ABDELNOOR, R.V., BARROS, E.G., MOREIRA, M.A. Determination of genetic diversity within Brazilian soybean germplasm using random amplified polymorphic DNA techniques and comparative analysis with pedigree data. **Rev. Brasil. Genet.**, v.18, n.2, p.265-273, 1995.
- AKKAYA, M.S., BHAGWAT, A.A., CREGAN, P.B. Length polymorphisms of simple sequence repeat DNA in soybean. **Genetics**, v.132, n.4, p.1131-1139, 1992.
- BOTSTEIN, D., WHITE, R.L., SKOLNICK, M., DAVIS, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **Am. J. Hum. Genet.**, v.32, n.2, p.314-331, 1980.
- BURR, B., BURR, F.A., THOMPSON, K.H., ALBERTSON, M., STUBER, C.W. Gene mapping with recombinant inbreds in maize. **Genetics**, v.118, n.4, p.519-526, 1988.
- CERVIGINI, G.D.L. **Análise genômica da resistência às raças 3 e 9 do nematóide de cisto e avaliação da variabilidade genética de caracteres agronômicos em soja.** Viçosa, MG: UFV, 2003. 82p. Dissertação (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, 2003.

- CHAPARRO, J.X., WERNER, D.J., D'MARLLEY, D., SEDEROFF, R.R. Targeted mapping and linkage analysis of morphological, isozyme, and RAPD markers in peach. **Theor. Appl. Genet.**, v.87, n.7, p.801-15, 1994.
- COATES, J.B., MEDEIROS, J.S., THANH, V.H., NIELSEN, N.C. Characterization of the subunits of β -conglucinin. **Arch. of Bioch.and Bioph.**, v.243, n.1, p.184-194, 1985.
- CREGAN, P.B., JARVIK,T., BUSH, R.C., SHOEMAKER, R.C., LARK, K.G., KAHLER, A.L., KAYA, N., VANTOAI, T.T., LOHNES, D.G., CHUNG, J., SPEXHT, J.E. An integrated genetic linkage map of the soybean genome. **Crop Sci.**, v.39, n.5, p.1464-1490, 1999.
- DOERGE, R.W., CHURCHILL, G.A. Permutation tests for multiple loci affecting a quantitative character. **Genetics**, v.142, n.1, p.285-294, 1996.
- DIERS, B.W., BEILINSON, V., NIELSEN, N.C., SHOEMAKER, R.C. Genetic mapping of the Gy4 e Gy5 glycinin genes in soybean and the analysis of a variant of Gy4. **Theor. Appl. Genet.**, v.89, n.2, p.297-304, 1994.
- DIERS, B.W., KEIM, P., FEHR, W.R., SHOEMAKER, R.C. RFLP analysis of soybean seed protein and oil content. **Theor. Appl. Genet.**, v.83, n.5, p.608-612, 1992.
- DUDLEY, J.M., MOLL, R.H. Interpretation and use of estimars of heritability and genetics variances in plant breeding. **Crop Sci.**, Madison, v.9, n.3, p.257-262, 1969.
- FALCONER, D.S. **Introdução à genética quantitativa**. 1^a ed. (Trad. de Silva, M.A.). Viçosa: UFV, 1987. 279p.
- FEHR, W.R., HOECK, J.A., JOHNSON, S.L., MURPHY, J.D.N., PADILLA, G.I., WELKE, G.A. Genotype and environment influence on protein components of soybean. **Crop Science**, v.43, p.511-514, 2003.
- FEHR, W.R. Breeding. In: NORMAN, A.G. **Soybean physiology, agronomy, and utilization**. New York: Academy Press, 1978. p.119-155.
- FERREIRA, M.E., GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análises genéticas**. 3.ed. Brasília: EMBRAPA; CENARGEN, 1998. 220p.
- FONTES, E.P.B., MOREIRA, M.A, DAVIES, C.S. , NIELSEN, N.C. Urea-elicited changes in relative electrophoretic mobility of certain glycinin and β -conglycinin subunits. **Plant Physiol.**, v.76, n.3, p.840-842, 1984.
- FREY, K.J., HORNER, T. Heritability in standard units. **Agron. J.**, v.49, n.1, p.59-62, 1957.

- GIZLICE, Z., CARTER, JR., T.E., BURTON, J.W. Genetic diversity in North American soybean: I. Multivariate analysis of founding stock and relation to coefficient of parentage. **Crop Sci.**, v.33, n.3, p.614-620, 1993.
- GUIMARÃES, C.T. e MOREIRA, M.A. Genética molecular aplicada ao melhoramento de plantas. In: BORÉM, A. (Ed.). **Melhoramento de espécies cultivadas**. Viçosa: Editora UFV, 1999. p. 715-740.
- HANSON, D.W., WEBER, C.R. Analysis of genetic variability from generations of plant-progeny lines in soybeans. **Crop Science**, v.2, n.1, p.63-67, 1962.
- HARADA, J.J., BARKER, S.J., GOLDBERG, R.B. Soybean β -conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. **The Plant Cell**, v.1, n.4, p.415-425, 1989.
- HASSE, G. **O Brasil da soja: abrindo fronteiras, semeando cidades**. Porto Alegre, RS, L&PM Editores. 256p. 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**, v.77, n.3, p.269-75, 1994.
- HILL, J.E., BREIDENBACH, R.W. Protein of soybean seeds. **Plant Physiol.**, v.53, n.5, p.742-746, 1974.
- JANSEN, R.C. Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm. **Biometrics**, v.49, n.1, p.227-231, 1993.
- KEIM, P., DIERS, B.W., OLSON, T.C., SHOEMAKER, R.C. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. **Genetics**, v.126, n.3, p.735-742, 1990a.
- KEIM, P., DIERS, B.W., SHOEMAKER, R.C. Genetic analysis of soybean hard seedness with molecular markers. **Theor. Appl. Genet.**, v.79, n.2, p.465-469, 1990b.
- KEIM, P., SCUPP, J.M., TRAVIS, S.E., CLAYTON, K., WEBB, D.M. A high-density soybean genetic map based upon AFLP markers. **Crop Sci.**, v.37, n.2, p.537-543, 1997.
- KEIM, P., SHOEMAKER, R.C., PALMER, R.G. Restriction fragment length polymorphism diversity in soybean. **Theor. Appl. Genet.**, v.77, n.6, p.786-92, 1989.

- KITAMURA, K., DAVIES, C.S., NIELSEN, N.C. Inheritance of alleles for *Ggy1* and *Gy4* storage protein genes in soybean. **Theor. Appl. Genet.**, v.68, n.3, p.253-254, 1984.
- KWANYUEN, P., PANTALONE, V.R., BURTON, J.W., WILSON, R.F. A new approach to genetic alteration of soybean protein composition and quality. **JAOCs**, v. 74, n.8, p.983-987, 1997.
- KWON, S.H., TORRIE, J.H. Heritability and interrelationships among traits of two soybean populations. **Crop Sci.**, v.4, n.2, p.196-198, 1964.
- 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**, v.1, n.1, p.174-181, 1987.
- LARK, K.G., WEISEMANN, J.M., MATTHEWS, B.F., PALMER, R., CHASE, K., MACALMA, T. A genetic map of soybean (*Glycine max* L. Merr) using an intraspecific cross of two cultivars: 'Minosy' and 'Noir 1'. **Theor. Appl. Genet.**, v.86, n.8, p.901-906, 1993.
- LITT, M., LUTY, J.A. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. **Am. J. Hum. Genet.**, v.44, n.3, p.937-1401, 1989.
- MAHAMUD, I., KRAMER, H.H. Segregation for yield, height, and maturity following a soybean cross. **Agron. J.**, v.43, n.3, p.605-609, 1951.
- MALUF, W.R. Perspectivas da aplicação da biologia molecular no melhoramento de plantas: o uso dos RFLPs. In: TORRES, A.C., CALDAS, L.S. (Eds.). **Técnicas e aplicações da cultura de tecidos de plantas**. Brasília: ABCTP; EMBRAPA; CNPH, 1990. p.381-389.
- MANSUR, L.M.L., LARK, K.G., KROSS, H., OLIVEIRA, A. Interval mapping of quantitative trait loci for reproductive, morphological and seed traits of soybean (*Glycine max* L. Merr.). **Theor. Appl. Genet.**, v.86, n.8, p.907-913, 1993.
- MAUGHAN, P.J., SAGHAI MAROOF, M.A., BUSS, G.R., UESTIS, G.M. Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near isogenic line analysis. **Theor. Appl. Genet.**, v.93, n.3, p.392-401, 1996.
- MIRANDA, F. D. **Uso de marcadores RAPD para mapeamento de QTLs que determinam teor de proteína em soja**. Viçosa, MG: UFV, 2002 56p. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 2002

- MOREIRA, M.A., HERMADSON, M., LARKINS, B.A. Partial characterization of the acidic and basic polypeptides of glycinin. **J. Biol. Chem.**, v.254, n.20, p.9921-9926, 1979.
- MOREIRA, M.A., HERMADSON, M., LARKINS, B.A., NIELSEN, N.C. Comparison of the primary structure of the acidic polypeptides of glycinin. **Arch. of Bioch. and Bioph.**, v.210, n.2, p.636-642, 1981.
- MORGANTE, M., RAFALSKI, A., BIDDLE, P., TINGEY, S., OLIVIERI, A.M. Genetic mapping and variability of seven soybean simple sequence repeat loci. **Genoma**, v.37, n.5, p.763-769, 1994.
- NIELSEN, N.C., DICKINSON, C.D., CHO, T.J, THANH, V.H., SCALLON, B.J., FISCHER, R.L., SIMS, T.L., DREWS, G.N., GOLDEBERG, R.B. Characterization of the glycinin gene family in soybean. **The Plant Cell** , v.1, n.3, p.313-328, 1989.
- 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.
- RAFALSKI, J.A., TINGEY, S.V. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. **Trends Genet.**, v.9, n.2, p.275-280, 1993.
- SHOEMAKER, R.C. RFLP map of soybean. In: PHILLIPS,R.L., VASIL, I.K., (Eds.). **DNA-based markers in plants**. [s.l.]: Kluwer Academic, 1994. p.2990-309.
- SHOEMAKER, R.C., GUFFY, R.D., LORENZEN, L.L., SPECHT, J.E. Molecular genetic mapping of soybean: map utilization. **Crop Sci.**, v.32, n.5, p.1091-1098, 1992.
- SMITH, J.D., KINMAN, M.L. The use of parent-offspring regression as an estimator of heritability. **Crop Sci.**, v.5, n.6, p.595-596, 1965.
- SOARES, T. C. B **Mapeamento de locos que controlam o conteúdo de proteína em soja**. Viçosa, MG: UFV, 2000. 58p. Dissertação (Mestrado em Agroquímica) – Universidade Federal de Viçosa, 2000.
- STAM, P. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. **Plant J.**, v.3, n.5, p.739-744, 1993.
- STASWICK, P.E., HERMODSON, M.A., NIELSEN, N.C. Identification of the acidic and basic subunits complexes of glycinin. **Agric. and Biol. Chem.**, v.256, n.6, p.8752-8755, 1981.
- STRICKBERGER, M.W. **Genetics**. New York: The McMillan, 1985. 842p.

- TANKSLEY, S.D., HEWITT, J. Use of molecular markers in breeding for soluble solids content in tomato- A re-examination. **Theor. Appl. Genet.**, v.75, n.5, p.811-823, 1988.
- TAUTZ, D. Hypervariability of simple sequences as a general source of polymorphic DNA markers. **Nucleic Acids Res.**, v.17, n.13, p.6463-6471, 1989.
- TEIXEIRA, A.I. **Herdabilidades e correlações entre concentrações de proteína em soja avaliadas por diferentes metodologias**. Viçosa, MG: UFV, 2003. 53p. Dissertação (Mestrado em Bioquímica Agrícola) – Universidade Federal de Viçosa, 2003.
- THANH, V.H., SHIBASAKI, K. Beta-conglycinin from soybean proteins. Isolation and immunological and physicochemical properties of the monomeric forms. **Bioch. et Biophys. Acta**, v.490, n.2, p.370-384, 1977.
- TINGEY, S.V., DELTUSO, J.P. Genetic analysis with Random Amplified Polymorphics DNA markers. **Plant Physiol.** v.101, n.3, p.349-352, 1992.
- VOS, P., HOGERS, R., BLEEKER, M., REIJANS, M., LEE, T.V., HORNES, M., FRIJTERS, A., POT, J., PELEMAN, J., KUIPER, M., ZABEU, M. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Res.**, v.23, n.21, p.4407-4414, 1995.
- WARNER, J.N. A method for estimating heritability. **Agron. J.**, v.44, n.2, p.427-430, 1952.
- WEBER, J.K., MAY, P.E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. **Am. J. Hum. Genet.**, v.44, n.4, p.388-397, 1989.
- WILCOX, J.R., CAVINS, J.F. Backcrossing higher seed protein to a soybean cultivar. **Crop Sci.**, v.35, n.4, p.1036-1041, 1995.
- WILLIAMS, J.G.K., KUBELIK, A.R., LIVAK, K.J., RAFALSKI, J.A., TINGEY, S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Res.**, v.18, n.13, p.6531-6535, 1990.
- WOLF, W.J., COWAN, J.C. Soybeans as a food source. Review. Ed. CRC Press, Cleveland, OH, 1975.
- XIAO, J., LI, J., YUAN, L., TANKSLEY, S.D. Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a subspecific rice cross. **Theor. Appl. Genet.**, v.92, n.2, p.230-244, 1996.
- YAKLICH, R.W. Beta-Conglycinin and glycinin in high-protein soybean seeds, **J. Agric. and Food Chem.**, v.49, n.2, 2001, p 729-735.

YAKLICH, R.W., HELM ,R.M., COCKRELL, G., HERMAN, E.M. Analysis of the distribution of the major seed allergens in a core collection of *Glycine max* accessions. **Crop Sci.**, v.39, n.2, p.144-147, 1999.

CAPÍTULO 1

ANÁLISE QUANTITATIVA DO CONTEÚDO DE PROTEÍNA TOTAL E DAS PRINCIPAIS PROTEÍNAS DE RESERVA (11S E 7S) EM SOJA PARA FINS DE MAPEAMENTO GENÉTICO

RESUMO

SOARES, Taís Cristina Bastos, D. S., Universidade Federal de Viçosa, março de 2004. **Mapeamento de locos associados ao conteúdo de proteínas de reserva em soja.** Orientador: Maurilio Alves Moreira. Conselheiros: Cosme Damião Cruz e Everaldo Gonçalves de Barros.

Uma população de 95 RILs (linhagens recombinantes endogâmicas) foi obtida do cruzamento entre o acesso BARC-8 (com alto teor de proteína) e a variedade brasileira Garimpo (com teor de proteína normal), na geração F_9 foram abertas linhas, que constituíram o material genético utilizado neste trabalho. Esse material foi analisado para as características: teor de proteína total em sementes de soja, teor das proteínas de reserva 11S e 7S e de suas respectivas subunidades, e da relação proteína total menos proteínas de reserva (PT-(11S+7S)). Essa população foi cultivada em dois ambientes distintos: Viçosa, MG e São Gotardo, MG. O teor de proteínas de reserva 7S e 11S, bem como os de suas respectivas subunidades apresentaram distribuição aproximadamente normal ($P < 0,01$), tanto no experimento de Viçosa, quanto no de São Gotardo, o que permitirá que estas características sejam analisadas para a detecção da existência de possíveis QTLs envolvidos na expressão das mesmas, uma vez que o método de detecção de QTLs pressupõe que a característica apresente esse tipo de distribuição. O teor de proteína total em São Gotardo, também apresentou distribuição aproximadamente normal. O teor de proteína total em Viçosa não apresentou uma distribuição aproximadamente normal, apesar do grau de simetria e de curtose terem sido próximos do esperado para distribuição normal. Caso sejam detectados QTLs relacionados a esta característica, eles deverão ser analisados com ressalva. Fatores ambientais estão envolvidos na regulação da expressão das características avaliadas nos dois locais (Viçosa e São Gotardo), pois a interação RIL x local foi significativa. 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. As herdabilidades das outras características avaliadas também foram altas nos dois locais. Entretanto, as herdabilidades destas características na análise conjunta foram

menores, devido ao efeito da interação genótipo x ambiente. Isto indica que os estudos moleculares referentes às análises de QTLs para estas características devam ser realizados para cada ambiente específico. As subunidades α' , α e β apresentaram correlação positiva com a proteína 7S, este resultado era esperado, uma vez que estas são as subunidades constituintes desta proteína. Da mesma forma, as subunidades de caráter ácido e básico da proteína 11S apresentaram correlação positiva e alta com esta proteína. A proteína 11S e o total de proteína de reserva (11S+7S) apresentaram correlação positiva e alta com o conteúdo de proteína total. A proteína 7S, assim como as suas subunidades α e β apresentaram uma correlação baixa com o conteúdo de proteína total, a subunidade α' apresentou uma correlação próxima de zero. Uma vez que um maior conteúdo de 11S em relação à 7S é desejado, o fato de 11S ter apresentado uma correlação positiva e alta com proteína total torna-se benéfico no processo de seleção, já que selecionando genótipos com alto teor de proteína, serão selecionados genótipos com alto teor de 11S simultaneamente.

1. INTRODUÇÃO

As sementes de soja contêm mais proteína do que qualquer outra espécie vegetal utilizada comercialmente. Aproximadamente 40% da matéria seca das sementes de soja é constituído de proteínas e 20%, de óleo.

Historicamente, a soja tem sido melhorada geneticamente com o intuito de aumentar a produtividade e conteúdo de óleo nos grãos. Recentemente, alguns programas de melhoramento de soja têm dado ênfase ao desenvolvimento de variedades mais produtivas, contendo altos teores de proteína. (YAKLICH, 2001).

Atualmente, o farelo de soja é um dos principais produtos das indústrias de esmagamento no Brasil, sendo que a maior parte dele destina-se à exportação. O teor de proteína no farelo influencia a determinação de seu valor comercial, assim, torna-se necessário o desenvolvimento de variedades produtivas e com alto teor de proteína (PIOVESAN, 2000).

A fração protéica e seus derivados apresentam restrição quanto ao seu consumo pela população. Essa barreira alimentar aos derivados protéicos da soja está associada à baixa qualidade nutricional do grão utilizado no processamento industrial. Para o melhoramento do grão de soja dois aspectos devem ser considerados: o teor e a qualidade de proteína, sendo que a qualidade protéica refere-se ao teor de aminoácidos sulfurados e também às propriedades funcionais da proteína (WILCOX e SHIBLES, 2001).

As sementes de soja contêm células especiais que produzem e armazenam as proteínas de reserva na semente. As sementes produzem centenas de vacúolos de estoque de proteínas, mas poucas proteínas principais são armazenadas na semente de soja. Duas proteínas de reserva, a β -conglícinina e glicinina, constituem cerca de 70% das proteínas de reserva em soja. Geralmente a glicinina é a mais abundante (MOREIRA *et al.*, 1979; COATES *et al.*, 1985).

A glicinina (11S) contém 3 a 4 vezes mais cisteína e metionina que a β -conglícinina (7S) (KOSHIYAMA, 1968, COATES *et al.*, 1985, HARADA *et al.*, 1989). A concentração destas duas proteínas poderá variar nas diferentes variedades de soja, dependendo da constituição genética da semente e do ambiente onde elas são cultivadas. Uma vez que 11S contém mais

aminoácidos sulfurados, uma maior razão 11S:7S leva a mudanças benéficas na qualidade nutricional da soja.

Foi observado também que a relação da quantidade destas duas proteínas é importante na fabricação de produtos alimentícios. Por exemplo, mudanças na razão β -conglucina/glicina resultam em diferenças nas propriedades físicas do *tofu*, também chamado “queijo de soja”, resultante da coagulação do extrato aquoso de sementes de soja. As variedades que apresentam menor razão β -conglucina/glicina têm mostrado serem melhores para a produção deste importante produto à base de soja (YAKLICH *et al.*, 1999).

O fato de ocorrerem variações genéticas naturais na composição de subunidades nas duas principais proteínas de reserva da soja, e uma vez que estas variações podem estar associadas aos mecanismos de regulação da expressão dos genes que codificam estas subunidades. O conhecimento destes mecanismos poderá ser usado para aumentar seletivamente o tipo de proteína de reserva para usos específicos (YAKLICH *et al.*, 1999). Pesquisas envolvendo a manipulação genética dos níveis dessas duas proteínas estão sendo empregadas com o intuito de aumentar os níveis de aminoácidos essenciais nas formulações de rações tanto para consumo animal quanto para consumo humano (KWANYUEN *et al.*, 1997).

As estimativas de herdabilidade, os ganhos esperados com a seleção e as correlações das características selecionadas são extremamente importantes para orientar e planejar programas de melhoramento (FALCONER, 1997).

ERICKSON *et al.* (1981) estimaram a herdabilidade para teor de proteína em quatro populações obtidas a partir de quatro pares de cruzamentos biparentais interespecíficos (*G. max* x *G. soja*). Para o cálculo da herdabilidade no sentido restrito, foi utilizada a regressão das médias de famílias F_3 sobre as plantas F_2 ; para a herdabilidade no sentido amplo, o cálculo foi feito utilizando-se os componentes de variância das médias de famílias F_3 . A estimativa de herdabilidade para teor de proteína, no sentido restrito, foi de 27% e, no sentido amplo, de 78%, com base em 100 famílias F_3 avaliadas em dois locais.

HELMS e ORF (1998), trabalhando com 10 diferentes populações na geração F_5 , testadas em sete ambientes diferentes, calcularam a herdabilidade,

considerando-se todos os ambientes para teor de proteína, e encontraram valores elevados para a maioria das populações.

WILCOX (1998) conduziu um programa de seleção recorrente para aumentar o teor de proteína nas sementes em duas populações de soja. Estimativas de herdabilidade foram calculadas a cada ciclo de seleção recorrente e variaram de 55 a 89% entre o primeiro e o sétimo ciclo. Houve também aumento da variabilidade para teor de proteína à medida que se aumentava o número de ciclos de seleção recorrente. Segundo HANSON (1963), a estimativa do coeficiente de herdabilidade atende a duas finalidades básicas no melhoramento genético: a primeira é a de demonstrar a relativa facilidade com que diferentes caracteres são selecionados em um determinado esquema de melhoramento, e a segunda é a de permitir que estimativas dos progressos esperados com a seleção sejam obtidas.

Para CRUZ e REGAZZI (1997), o conhecimento da correlação entre caracteres é de grande importância no melhoramento de plantas, especialmente se alguma das características avaliadas apresentarem baixa herdabilidade e, ou, tiver problemas de medição e identificação.

Este trabalho faz parte do programa de melhoramento de qualidade da soja desenvolvido pelo BIOAGRO/UFV, que visa o desenvolvimento de variedades produtivas e com alto teor protéico e teve como objetivos: (i) estimar a herdabilidade da concentração de proteína total em soja, avaliada pelo método de Kjeldahl modificado; (ii) estimar a herdabilidade das concentrações das proteínas de reserva da soja (11S e 7S) e de suas respectivas subunidades, avaliadas pela eletroforese em gel de poliacrilamida em condição desnaturante (SDS-PAGE) acompanhada de densitometria, (iii) estimar as correlações genotípicas de cada uma destas características, (iv) avaliar a interação genótipo x ambiente para essas características.

2. MATERIAL E MÉTODOS

2.1. Material genético

Uma população de 95 RILs foi obtida do cruzamento entre dois progenitores contrastantes para teor de proteínas: a linhagem BARC-8, obtida da "Beltsville Agricultural Experimental Station" – USDA, Beltsville, MA, USA (genótipo com alto teor de proteína, em torno de 50%) e a variedade comercial Garimpo (genótipo com teor normal de proteínas, em torno de 36%), sendo BARC-8 o progenitor feminino. As RILs foram obtidas pela descendência, a partir da geração F_2 , de uma única semente de cada planta (SSD – "single seed descent"). Na geração F_9 foram abertas linhas, que constituíram o material genético utilizado neste trabalho.

Essas RILs foram plantadas em duas localidades: Viçosa e São Gotardo, ambas no estado de Minas Gerais. Os experimentos foram instalados no delineamento em blocos casualizados com três repetições. Cada parcela foi constituída de uma fileira de 2 metros, com espaçamento de 0,5 metros entre fileiras. O plantio foi realizado em Viçosa na segunda quinzena de novembro de 2001, e em São Gotardo, na primeira quinzena de dezembro de 2001.

Em Viçosa, o experimento foi instalado no Campo experimental Diogo Alves de Mello, localizado no Campus da Universidade Federal de Viçosa, na Zona da Mata do Estado de Minas Gerais, cujas coordenadas geográficas são 20°45'20"S e 42°51'W. A sede do município situa-se aproximadamente a 650m de altitude. Em São Gotardo, o experimento foi instalado na Estação experimental da Cooperativa Agropecuária do Alto Paranaíba (COOPADAP), na região do Alto Paranaíba do Estado de Minas Gerais, cujas coordenadas geográficas são 19°29'59"S e 46°06'05"W. A sede do município situa-se a 1055m de altitude. O solo, encontrado nas duas localidades é classificado como Latossolo Vermelho-Amarelo distrófico, textura argilosa.

2.2. Análise de proteína

2.2.1. Análise do conteúdo de proteína total

O teor de proteínas das sementes de cada família foi determinado, tomando-se uma amostra representativa de todas as plantas da parcela. Essa determinação foi realizada pelo método Kjeldahl modificado, segundo as normas analíticas do INSTITUTO ADOLFO LUTZ (1985), que requer um mínimo de trezentos miligramas (cerca de 10 sementes) de soja moída. Na fase de digestão, após a obtenção do material aparentemente digerido com ácido sulfúrico, foi adicionado o peróxido de hidrogênio 30%, sendo a mistura levada ao aquecimento por mais 30 minutos. Na fase de destilação, a amônia liberada foi recolhida em solução de ácido bórico 4%. Obteve-se o teor de nitrogênio pela titulação da amônia com ácido clorídrico 0,05%. A partir do teor de nitrogênio, foi calculada a porcentagem de proteínas totais da amostra, empregando-se o fator de nitrogênio 6,25 no material analisado. Os resultados dos teores de proteína foram expressos em porcentagem, com base na matéria seca obtida pela secagem em estufa a 105°C por 24 horas.

2.2.2. Análise das concentrações das proteínas 11S e 7S e de suas respectivas subunidades

As proteínas de reserva glicinina (11S) e β -conglucina (7S) das RILs foram extraídas com tampão fosfato salino (tampão fosfato de sódio 0,05 M, pH 7,6; NaCl 0,4 M; β -mercaptoetanol 0,28%). Foram utilizados aproximadamente 50 mg de soja moída e 1,5 mL do tampão de extração em tubos de 1,5 mL. Os tubos ficaram 45 minutos em banho de ultrassom (Branson 1210), após o qual foram centrifugados durante 15 minutos a 14.000 rpm em microcentrífuga eppendorf 3415C. Duas alíquotas, uma de 800 μ L (que foi reservada para análise posterior) e outra de 30 μ L foram retiradas do sobrenadante. À alíquota de 30 μ L foram adicionados 15 μ L do tampão de amostra 3X (Tris 0,1875 M; SDS 6,9%; glicerol 30%; pH 6,8), 10 μ L do corante azul de bromofenol 0,05% e 10 μ L de β -mercaptoetanol. Posteriormente, os tubos foram colocados em banho-maria a 100 °C por 3 minutos e 25 μ L da solução foram aplicados em

gel de poliacrilamida 12,5%. A corrida foi conduzida inicialmente a 60 Volts, até a banda do corante atingir a superfície do gel de separação (~1,0 h), quando a voltagem foi aumentada para 80 Volts, permanecendo assim até o final da corrida (~5h), que foi determinado como sendo 30 minutos após o corante sair na extremidade inferior do gel. Terminada a corrida, os géis foram retirados da placa e colocados em solução corante (1,5 g de Coomassie Brilliant Blue G, 90 mL de ácido acético, 450 mL de metanol e 460 mL de água) por 12 horas e depois em solução descorante (75 mL de ácido acético, 250 mL de metanol e 675 mL de água) por 24 horas. Os géis, depois de descorados, foram armazenados em solução de glicerol 10% até o momento da análise por densitometria. Em todas as análises foi usada a mesma amostra padrão (BARC-8) que foi extraída, aplicada no gel e quantificada junto com as outras amostras. Esta amostra, que foi denominada de padrão de extração, foi adicionada nas análises para corrigir as possíveis diferenças de extração e das condições de eletroforese que podem ocorrer entre uma análise e outra.

A análise densitométrica foi realizada no Personal Densitometer SI Molecular Dynamics (Sunnyvale, CA), pertencente ao Núcleo de Biologia Aplicada da Embrapa Milho e Sorgo - Sete Lagoas, MG. O software ImageQuant para a integração do volume referente à banda foi usado na determinação da porcentagem de cada subunidade das proteínas de reserva presentes nos géis em relação ao total de proteína de cada canaleta do gel. A densidade ótica (OD) aparente de cada proteína foi obtida pela subtração da OD média do “background” local, da OD total das bandas das proteínas. As subunidades das proteínas de reserva foram identificadas com base nos padrões eletroforéticos constantes em trabalhos publicados na literatura (FONTES *et al*, 1984; YAKLICH, 2001).

Com o intuito de transformar os valores das porcentagens das subunidades obtidas, após as análises de densitometria, em porcentagem em 100 mg de proteína, procedeu-se a quantificação das alíquotas de 800 μ L, reservadas após a extração, pelo método de Kjeldahl modificado.

A porcentagem final de cada subunidade α' , α e β da β -conglucina, e das subunidades ácidas e básicas da glicina foram obtidas por meio da seguinte forma:

1- As porcentagens das subunidades foram obtidas de todas as amostras analisadas, inclusive dos padrões de extração:

$$\%Subunidade = \frac{\%S \times PE}{100}$$

em que

$\%Subunidade$ = porcentagem da subunidade em questão em 100 mg de proteína

$\%S$ = porcentagem da subunidade presente no gel

PE = porcentagem de proteína do extrato obtida pelo método de Kjeldahl modificado.

2 - Cada subunidade dos padrões de extração foi corrigida a partir da média dos padrões:

$$E = e_p - e_{\bar{x}}$$

em que

E = Desvio de cada subunidade para cada gel;

e_p = % de cada subunidade no padrão de extração para cada gel; e

$e_{\bar{x}}$ = % média de cada subunidade dos padrões de extração

3 - Cada subunidade das proteínas das famílias das RILs foi obtida a partir de E

$$\%final = \%Subunidade - E$$

4 - Os valores de porcentagem final de cada subunidade foram corrigidos para 70%, o que corresponde ao valor médio esperado para as proteínas de reserva em soja, sendo também o valor observado na média dos 48 padrões de gel.

2.3. Testes de normalidade

2.3.1. Teste de assimetria

O coeficiente de assimetria fornece o grau de desvio, ou afastamento da simetria, de uma distribuição. A estimativa do coeficiente de assimetria (s) de uma variável x é dada por:

$$\hat{s} = \frac{n}{(n-1)(n-2)} \cdot \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s(x)} \right)^3$$

onde:

n = número de observações;

$s(x)$ = desvio padrão da amostra de dados;

x_i = valor máximo da classe i ;

$\bar{x}$ = média geral da amostra de dados.

Nesta expressão, se o resultado for zero, a distribuição é simétrica; se o resultado for negativo, a distribuição é assimétrica negativa (inclinada para a esquerda) e se o resultado for positivo, a distribuição é assimétrica positiva (inclinada para a direita).

As hipóteses $H_0: s=0$ e $H_a: s \neq 0$ foram testadas pelo teste t :

$$t = \frac{\hat{s}}{\sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}}$$

O valor de t calculado foi comparado ao de t tabelado, ao nível α de probabilidade, com infinitos graus de liberdade. Se $|t_{cal}| \geq t_{tab}$, rejeita-se H_0 . (SCHUSTER e CRUZ, 2004).

2.3.2. Teste de curtose

Curtose é o grau de achatamento de uma distribuição, considerado usualmente em relação à distribuição normal. Segundo SCHUSTER e CRUZ

(2004), para se estimar o grau de curtose (k) de uma variável X, utiliza-se a seguinte fórmula:

$$\hat{k} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \cdot \sum_{i=1}^n \left(\frac{x_i - \bar{X}}{s(X)} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Se o resultado for zero, a distribuição de freqüências é a própria distribuição normal, sendo chamada de mesocúrtica, se o resultado for negativo, a distribuição é achatada (alta variabilidade) e chamada de planicúrtica e se o resultado for positivo, a distribuição é concentrada em torno da média (alta homogeneidade) e chamada de leptocúrtica.

Para testar a hipótese $H_0: k=0$ vs $H_a: k \neq 0$ tem-se o teste t:

$$t = \frac{\hat{k}}{\sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}}$$

Este valor de t calculado foi comparado ao valor de t tabelado, a nível α de probabilidade, com infinitos graus de liberdade. Se $|t_{cal}| \geq t_{tab}$, rejeita-se H_0 .

2.3.3. Teste de Lilliefors

O teste de Lilliefors apresenta o mesmo princípio que o teste de qui-quadrado (χ^2), ou seja, dada uma amostra de tamanho n, agrupam-se os dados em k classes, podendo-se formar uma distribuição de freqüência com intervalos de classes iguais a 0,25, 0,50, 0,75, 1,0 do desvio padrão. As freqüências esperadas (FE_i) de cada classe são encontradas de acordo com a distribuição normal reduzida z, multiplicando-se a área de cada classe pelo número total de observações (n) e as freqüências observadas (FO_i) são os números de observação de cada classe (n_i). Para isso, obtêm-se os valores de z com a introdução de uma modificação, ampliando o seu uso para os casos em que a média e a variância não são conhecidas, mas sim estimadas por meio dos dados da amostra, como segue:

$$z_i = \frac{X_i - \bar{X}}{s}, \text{ em que}$$

z_i = valor da variável normal padronizada da classe i ;

X_i = valor máximo da classe i ;

$\bar{X}$ = média geral da amostra de dados;

s = desvio padrão da amostra de dados.

Na maioria das vezes, o teste de Lilliefors é mais poderoso que o χ^2 .

O teste de Lilliefors pressupõe o cálculo de todos os z_i , os quais devem ser ordenados para as seguintes considerações:

$F(z_i) = FE_i \leq z_i$ = valor da tabela de distribuição normal reduzida (área);

$S(z_i) = FO_i \leq z_i = n_i/n$, em que:

n_i = número de valores observados $\leq z_i$;

n = número total de observações da amostra.

D = Máximo $|F(z_i) - S(z_i)|$.

O teste é bilateral, como segue:

H_0 : é razoável estudar os dados através da distribuição normal;

H_a : não é razoável estudar os dados através da distribuição normal.

Deve-se lembrar, porém, que a não rejeição de H_0 indica apenas que esta é uma razoável aproximação da distribuição conhecida.

2.4. Análise de variância

2.4.1. Análise de variância individual

Para as análises de variância das características em cada local, os experimentos foram realizados pelo delineamento em blocos casualizados, com três repetições. O modelo adotado foi o seguinte:

$$Y_{ij} = \mu + G_i + B_j + e_{ij}$$

em que

Y_{ij} = valor da característica para o i -ésimo genótipo (RIL) no j -ésimo bloco;

μ = média geral;

G_i = efeito aleatório do i-ésimo genótipo (RIL), $i=1,2,\dots,g$, sendo $G_i \sim \text{NID}(0, \sigma_g^2)$

B_j = efeito aleatório do j-ésimo bloco, $j=1,2,\dots,b$, sendo $B_j \sim \text{NID}(0, \sigma_b^2)$ e

e_{ij} = erro aleatório associado à observação Y_{ij} , sendo $e_{ij} \sim \text{NID}(0, \sigma^2)$.

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

Quadro 1 – Esquema de análise de variância de cada ambiente e esperanças de quadrados médios para modelo aleatório, adotando-se o delineamento de blocos casualizados

FV	GL	SQ	QM	E(QM)	F
Blocos	b-1	SQB	QMB		
Genótipos	g-1	SQG	QMG	$\sigma^2 + b\sigma_g^2$	QMG/QMR
Resíduo	(b-1)(g-1)	SQR	QMR	σ^2	

$H_0: \sigma_g^2 = 0$ vs $H_a: \sigma_g^2 > 0$ com (g-1) e (b-1)(g-1) graus de liberdade.

As somas de quadrados dos blocos, dos genótipos e do resíduo foram obtidas de forma usual para ensaios em blocos casualizados, ou seja:

$$SQTo = \sum_{i=1}^g \sum_{j=1}^b Y_{ij}^2 - C$$

$$SQG = \frac{1}{b} \sum_{i=1}^g Y_{i.}^2 - C$$

$$SQB = \frac{1}{g} \sum_{j=1}^b Y_{.j}^2 - C$$

$$SQR = SQTo - SQB - SQG$$

$$C = \frac{Y_{..}^2}{gb}$$

Por se tratar de uma população de RILs foi possível estimar a herdabilidade no sentido restrito, de forma que a sua estimativa e a do

coeficiente de variação experimental foram obtidos pelas seguintes expressões:

$$h^2 = \frac{\hat{s}_g^2}{\hat{s}_f^2}, \text{ em que } \hat{s}_f^2 = \frac{QMG}{b} \text{ e } \hat{s}_g^2 = \frac{QMG - QMR}{b}$$

e

$$CV\% = \frac{100 \cdot \sqrt{QMR}}{\hat{m}}$$

2.4.2. Análise de variância conjunta

Na análise da variância conjunta (em Viçosa e São Gotardo) foi adotado o seguinte modelo:

$$Y_{ijk} = \mu + G_i + A_j + (GA)_{ij} + (B/A)_{jk} + e_{ijk}$$

em que:

μ = média geral;

G_i = efeito aleatório do i-ésimo genótipo (RIL) ($i = 1, 2, \dots, g$); $G_i \sim \text{NID}(0, \sigma_g^2)$;

A_j = efeito aleatório do j-ésimo ambiente ($j = 1, 2, \dots, a$); $A_j \sim \text{NID}(0, \sigma_a^2)$;

GA_{ij} = efeito aleatório da interação do i-ésimo genótipo com o j-ésimo ambiente; $GA_{ij} \sim \text{NID}(0, \sigma_{ga}^2)$;

B/A_{jk} = efeito aleatório do k-ésimo bloco dentro do j-ésimo ambiente ($k = 1, 2, \dots, b$); $B/A_{jk} \sim \text{NID}(0, \sigma_{b/a}^2)$;

e_{ijk} = erro aleatório, $e_{ijk} \sim \text{NID}(0, \sigma^2)$.

O esquema da análise de variância conjunta, segundo este modelo estatístico, é apresentado no Quadro 2:

Quadro 2 – Esquema da análise de variância conjunta e esperanças de quadrados médios para modelo aleatório, adotando-se o delineamento em blocos casualizados

FV	GL	QM	EM (QM)	F
Blocos/Locais	a(b-1)	QMB	$\sigma^2 + g\sigma_{b/a}^2$	
Locais (L)	a-1	QMA	$\sigma^2 + b\sigma_{ga}^2 + g\sigma_{b/a}^2 + gb\sigma_a^2$	$\frac{QMA + QMR}{QMB + QMGA}$
RILs (R)	g-1	QMG	$\sigma^2 + b\sigma_{ga}^2 + ab\sigma_g^2$	$\frac{QMG}{QMGA}$
R X L	(a-1)(g-1)	QMGA	$\sigma^2 + b\sigma_{ga}^2$	$\frac{QMGA}{QMR}$
Resíduo	a(b-1)(g-1)	QMR	σ^2	
Total	agb-1			

$H_{01}: \sigma_{ga}^2=0$ vs $H_{a1}: \sigma_{ga}^2>0$ com (a-1)(g-1) e a(b-1)(g-1) graus de liberdade

$H_{02}: \sigma_g^2=0$ vs $H_{a2}: \sigma_g^2>0$ com (g-1) e (a-1)(g-1) graus de liberdade

$H_{03}: \sigma_a^2=0$ vs $H_{a3}: \sigma_a^2>0$ com $n_1 = \frac{(QMA + QMR)^2}{\frac{(QMA)^2}{a-1} + \frac{(QMR)^2}{a(b-1)(g-1)}}$ e

$n_2 = \frac{(QMB + QMGA)^2}{\frac{(QMB)^2}{a(b-1)} + \frac{(QMGA)^2}{(a-1)(g-1)}}$ graus de liberdade.

As somas de quadrados foram obtidas pelas seguintes expressões:

$$C = Y^2 \dots / agb$$

$$SQB = \frac{1}{g} \sum_{j=1}^a \sum_{k=1}^b Y_{.jk}^2 - \frac{1}{gb} \sum_{j=1}^a Y_{.j.}^2$$

$$SQA = \frac{1}{gb} \sum_{j=1}^a Y_{.j.}^2 - C$$

$$SQG = \frac{1}{ab} \sum_{i=1}^g Y_{i..}^2 - C$$

$$SQGA = \frac{1}{b} \sum_{i=1}^g \sum_{j=1}^a Y_{ij.}^2 - \frac{1}{gb} \sum_{j=1}^a Y_{.j.}^2 - \frac{1}{ab} \sum_{i=1}^g Y_{i..}^2 + C$$

$$SQT = \sum_{i=1}^g \sum_{j=1}^a \sum_{k=1}^b Y_{ijk}^2 - C$$

$$SQR = SQT - (SQB + SQA + SQG + SQGA)$$

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

$$\hat{s}_g^2 = \frac{QMG - QMGA}{ab}$$

$$h^2 = \frac{\hat{s}_g^2}{\hat{s}_f^2} \text{ onde, } \hat{s}_f^2 = \frac{QMG}{ab}$$

em que:

h^2 é a herdabilidade no sentido restrito;

$\hat{s}_g^2$ é o estimador da variância genotípica na população, sendo a população estudada RILs, esta variância é de natureza aditiva;

$\hat{s}_f^2$ é o estimador da variância fenotípica.

2.5. Correlação genotípica

A correlação genotípica foi estimada segundo CRUZ e REGAZZI (1997):

$$r_{G_{XY}} = \frac{COV_G(X,Y)}{\sqrt{\hat{s}_{G_X}^2 \cdot \hat{s}_{G_Y}^2}}$$

em que:

$r_{G_{XY}}$ é o estimador do coeficiente de correlação genotípica entre as características X e Y.

Todas as análises estatísticas deste capítulo foram realizadas no programa GENES (CRUZ, 2001).

3. RESULTADOS E DISCUSSÃO

3.1. Análise do teor de proteína total e das proteínas 11S e 7S e suas respectivas subunidades

A análise por densitometria das subunidades das proteínas de reserva foi feita a partir dos géis de poliacrilamida 12,5% (Figura 1). O teor de proteína total em São Gotardo, e das proteínas de reserva 7S e 11S, bem como os de suas respectivas subunidades, tanto no experimento de Viçosa quanto no de São Gotardo, apresentaram distribuição aproximadamente normal ($P < 0,01$), como pode ser observado nas Figuras 2 (Viçosa) e 3 (São Gotardo). Este resultado indica a existência de variação contínua e a possibilidade de que vários genes estejam envolvidos no controle destas características em soja. No entanto, o teor de proteína total em Viçosa não apresentou uma distribuição normal (Figura 2). Uma vez que estas características estão sendo avaliadas para fins de mapeamento genético, caso sejam detectados QTLs relacionados com o teor de proteína total obtido em Viçosa, eles deverão ser analisados com ressalva, já que o método de estimação de QTLs pressupõe que a característica deva apresentar uma distribuição normal; e pelos testes de Lilliefors, de assimetria e de curtose, conclui-se que esta característica não segue uma distribuição normal (Quadro 3). Entretanto deve ser observado que o grau de assimetria (0,8565) e de curtose (4,1105) não são tão elevados quando comparados, por exemplo, com a característica 11S+7S, cujo teste de Lilliefors foi não significativo.

As médias de quase todos os teores obtidos foram maiores no experimento de São Gotardo (Quadro 3). Esse fato, provavelmente, seja devido ao efeito de fatores ambientais como temperatura, solo, altitude etc., que podem estar influenciando o conteúdo de proteínas da soja cultivada nesses locais.

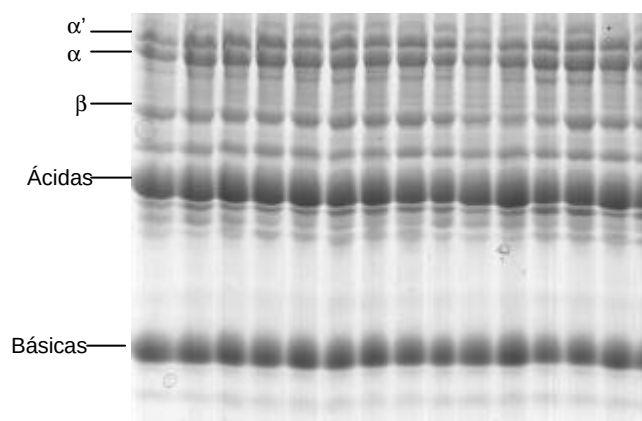


FIGURA 1 – Gel de poliacrilamida desnaturante (SDS-PAGE) de algumas RILs utilizadas neste trabalho. As subunidades da β -conglucina (7S) são α' , α e β , e as subunidades da glicinina (11S) são as subunidades ácidas e básicas.

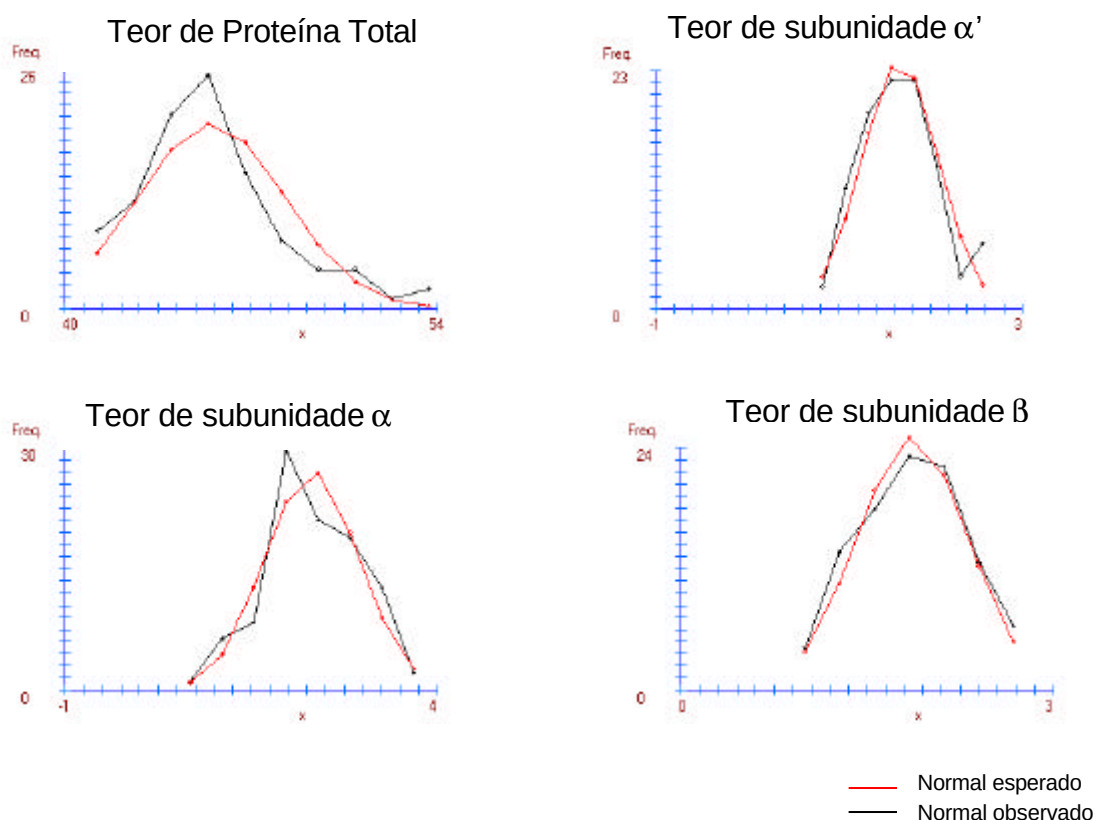
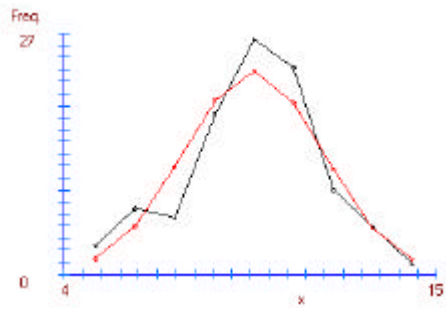
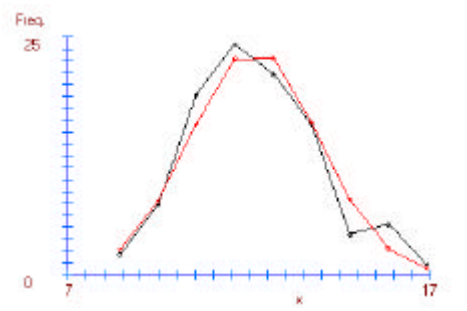


Figura 2 - Distribuição dos teores de proteína total, das proteínas de reserva 7S e 11S, de suas respectivas subunidades e das relações 11S+7S, 11S:7S, PT-(11S+7S), em 95 linhas endogâmicas recombinantes (RILs) cultivadas em Viçosa.

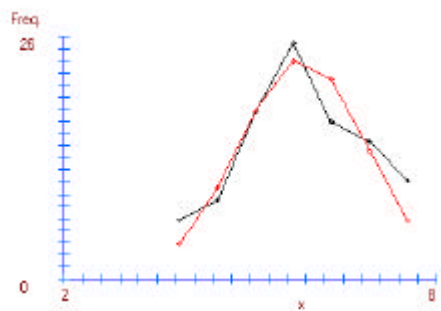
Teor de subunidades ácidas



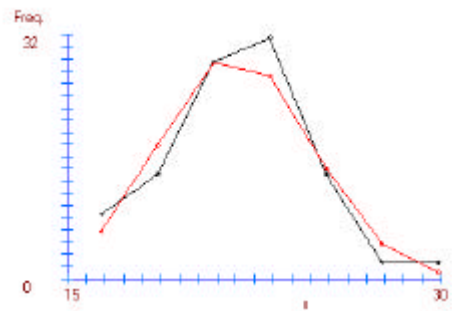
Teor de subunidades básicas



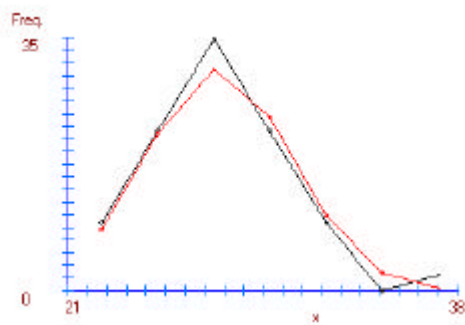
Teor de 7S



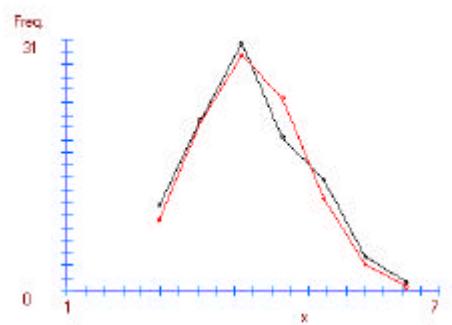
Teor de 11S



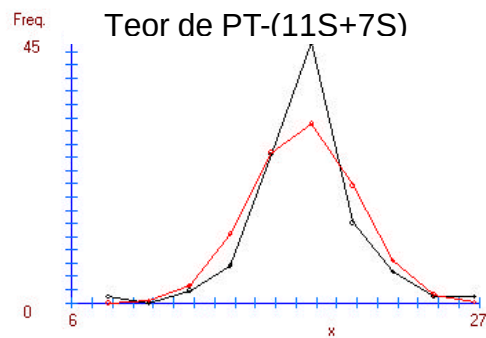
Teor de 11S+7S



Teor de 11S:7S



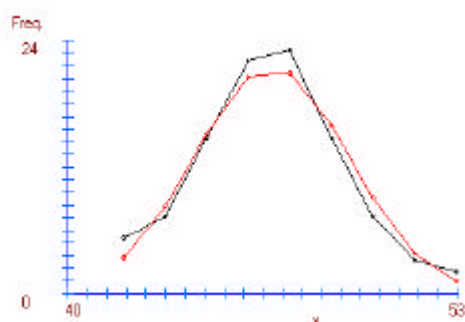
Teor de PT-(11S+7S)



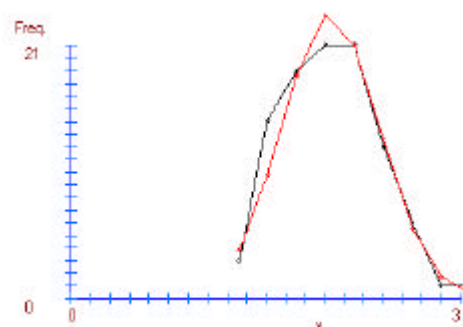
— Normal esperado
— Normal observado

FIGURA 2-Continuação...

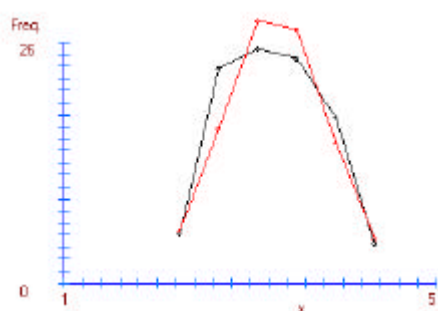
Teor de Proteína Total



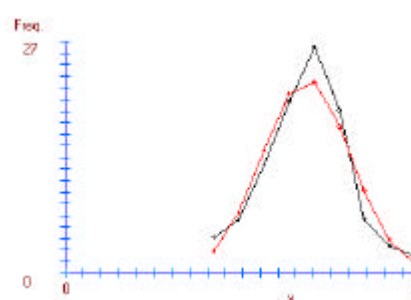
Teor de subunidade α'



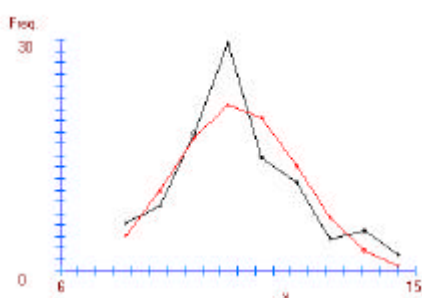
Teor de subunidade α



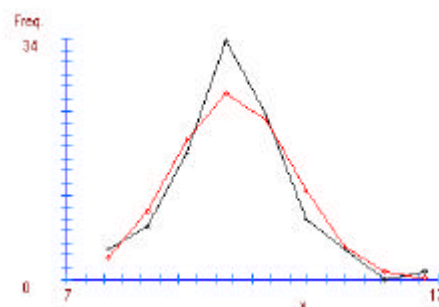
Teor de subunidade β



Teor de subunidades ácidas



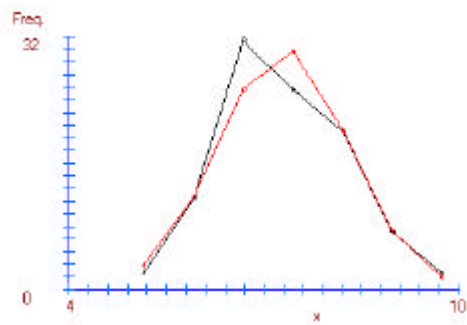
Teor de subunidades básicas



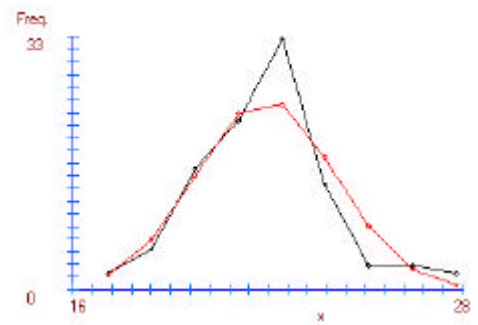
— Normal esperado
— Normal observado

Figura 3 - Distribuição dos teores de proteína total, das proteínas de reserva 7S e 11S, de suas respectivas subunidades e das relações 11S+7S, 11S:7S, PT-(11S+7S), em 95 linhas endogâmicas recombinantes (RILs) cultivadas em São Gotardo.

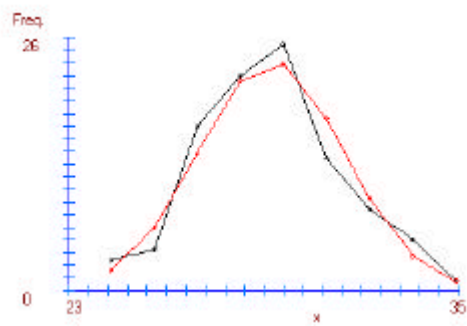
Teor de 7S



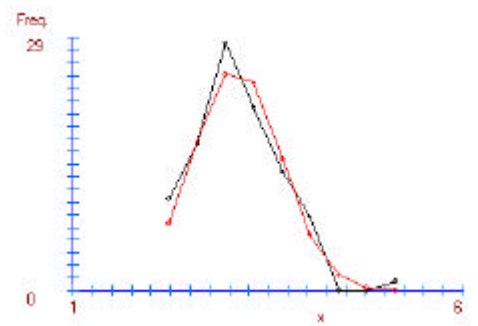
Teor de 11S



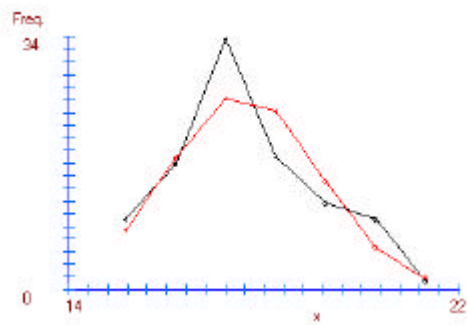
Teor de 11S+7S



Teor de 11S:7S



Teor de PT-(11S+7S)



— Normal esperado
— Normal observado

Figura 3– Continuação...

QUADRO 3 – Médias das características, testes de Lilliefors, de assimetria e de curtose, para as características avaliadas em Viçosa e São Gotardo.

Características	Média (%)	Teste de Lilliefors	Simetria	Curtose
Viçosa				
PT	45,49	0,1209**	0,8565*	4,1105**
α'	1,6657	0,0779 ^{ns}	0,3721 ^{ns}	2,976 ^{ns}
α	2,3133	0,0598 ^{ns}	-0,2014 ^{ns}	3,2192 ^{ns}
β	1,8651	0,0607 ^{ns}	-0,0002 ^{ns}	2,3739 ^{ns}
Ácidas	9,5848	0,0591 ^{ns}	-0,2496 ^{ns}	3,3076 ^{ns}
Básicas	12,1459	0,1197**	0,4132 ^{ns}	3,2435 ^{ns}
7S	5,8444	0,0563 ^{ns}	-0,0903 ^{ns}	2,4924 ^{ns}
11S	21,7307	0,0699 ^{ns}	0,2089 ^{ns}	3,8908 ^{ns}
11S+7S	27,5751	0,0702 ^{ns}	0,6037*	4,3714**
11S:7S	3,8898	0,0772 ^{ns}	0,2914 ^{ns}	3,26 ^{ns}
PT-(11S+7S)	17,917	0,0898 ^{ns}	-0,4054 ^{ns}	6,0791**
São Gotardo				
PT	46,79	0,0711 ^{ns}	0,1701 ^{ns}	3,1494 ^{ns}
α'	1,8994	0,0817 ^{ns}	0,3929 ^{ns}	3,1323 ^{ns}
α	3,2611	0,0808 ^{ns}	0,1547 ^{ns}	2,3661 ^{ns}
β	2,0389	0,0696 ^{ns}	-0,0305 ^{ns}	3,1882 ^{ns}
Ácidas	10,4586	0,1112**	0,4743 ^{ns}	3,319 ^{ns}
Básicas	11,4685	0,0586 ^{ns}	0,138 ^{ns}	4,1551*
7S	7,2984	0,0981 ^{ns}	0,3101 ^{ns}	3,1169 ^{ns}
11S	21,9271	0,0796 ^{ns}	0,3745 ^{ns}	3,9864*
11S+7S	29,2255	0,0589 ^{ns}	0,1366 ^{ns}	3,2143 ^{ns}
11S:7S	3,0992	0,1063 ^{ns}	0,7779*	4,5446*
PT-(11S+7S)	17,5672	0,1117**	0,4186 ^{ns}	2,6305 ^{ns}

**significativo a 1% de probabilidade.

* significativo a 5% de probabilidade.

^{ns} não significativo.

3.2. Análise de variância

3.2.1. Análise de variância individual

As análises de variância de cada local, apresentadas nos Quadros 4 (Viçosa) e 5 (São Gotardo) evidenciaram que a precisão experimental foi grande para teor de proteína total, tanto no experimento de Viçosa quanto no de São Gotardo, como pode ser observado pelos coeficientes de variação (3,10% em Viçosa e 2,64% em São Gotardo). Os valores dos coeficientes de variação estão de acordo com os encontrados por SOARES (2000) e TEIXEIRA (2003). Houve variabilidade genética significativa entre as RILs para todas as características avaliadas nos dois ambientes. As outras características: proteínas de reserva 7S e 11S, e suas respectivas subunidades, e as relações 11S+7S, 11S:7S e proteína total menos proteínas de reserva (PT-(11S+7S)) apresentaram coeficientes de variação mais altos, e estes variaram de 4,41% a 20,15% em Viçosa, e de 5,21% a 16,64% em São Gotardo.

Estudos de mapeamento de QTLs fundamentam-se na possibilidade de encontrar marcas moleculares, cuja presença e ausência determinam variação significativa na média dos caracteres avaliados. Assim, a existência de variabilidade genética, como constatado nestes ensaios, é indicativo da viabilidade de estudos moleculares futuros.

Quadro 4 - Resumo da análise de variância das características analisadas no experimento de Viçosa e estimativas da variância genética (σ^2_g), da herdabilidade (h^2) e do coeficiente de variação (CV)

FV	GL	QUADRADOS MÉDIOS										
		PT	α'	α	β	Ácidas	Básicas	7S	11S	11S+7S	11S:7S	PT-(11S+7S)
Blocos	2	66,7720	0,0125	1,1142	0,3263	14,4773	17,2097	0,8667	10,4385	6,0644	1,1478	90,8649
RILs	94	22,3617**	0,5076**	1,1488**	0,5623**	11,7362**	9,1916**	3,0769**	25,5745**	27,3436**	2,3727**	19,2339**
Resíduo	188	1,9935	0,1127	0,1866	0,1206	1,5608	4,3324	0,5681	5,8737	6,2792	0,4731	6,0019
$\hat{\sigma}^2_g$		6,7893	0,1316	0,3207	0,1472	3,3918	1,6197	0,8362	6,5669	7,0215	0,6332	13,6535
$h^2\%$		91,0851	77,787	83,7534	78,5598	86,7012	52,8657	81,5355	77,0331	77,0359	80,0599	68,7951
CV %		3,1036	20,1586	18,6730	18,6592	13,0640	17,1370	12,9063	11,1640	9,0959	17,6864	4,4107

** : significativo ao nível de 1% de probabilidade, pelo teste F.

Quadro 5 -Resumo da análise de variância das características analisadas no experimento de São Gotardo e estimativas da variância genética (σ^2_g), da herdabilidade (h^2) e do coeficiente de variação (CV)

FV	GL	QUADRADOS MÉDIOS										
		PT	α'	α	β	Ácidas	Básicas	7S	11S	11S+7S	11S:7S	PT-(11S+7S)
Blocos	2	2,7703	0,4161	1,8694	2,2187	20,0921	25,5331	2,0032	20,1848	12,0409	1,9419	6,9085
RILs	94	19,2721**	0,4079**	0,9013**	0,4118**	7,4156**	7,5692**	2,8520**	13,3894**	14,5994**	0,9206**	2,2759**
Resíduo	188	1,5329	0,1106	0,2903	0,0860	2,6950	2,9267	0,7886	2,0621	2,3161	0,2426	3,0725
$\hat{\sigma}^2_g$		5,9131	0,0991	0,2037	0,1086	1,5735	1,5475	0,6878	3,7758	4,0944	0,2260	1,3067
$h^2\%$		92,0458	72,8885	67,7928	79,1072	63,6569	61,3336	72,3496	84,599	84,1356	73,653	56,0612
CV %		2,6459	16,6409	16,5212	14,3869	15,6968	14,9171	12,1674	6,5489	5,2074	15,8909	9,9779

** : significativo ao nível de 1% de probabilidade, pelo teste F.

3.2.2. Análise de variância conjunta

A análise de variância conjunta dos dois locais (Quadro 6) indicou que fatores ambientais estão envolvidos na regulação da expressão gênica das características avaliadas em Viçosa e São Gotardo, pois a interação RILs x locais foi significativa ($P < 0,01$). Houve também efeito significativo de RILs e locais, para todos os caracteres ($P < 0,01$), com exceção do efeito de ambiente para as características 11S e PT-(11S+7S) que foram não significativos.

Acredita-se que, em qualquer material genético com o qual o melhorista esteja trabalhando, as interações genótipo x ambientes estejam presentes. A magnitude relativa dessas interações fornece subsídios ao melhorista quanto a melhor estratégia para selecionar genótipos de adaptação ampla ou de adaptação restrita a ambientes específicos (VENCOVSKY e BARRIGA, 1992). Estes mesmos subsídios podem ser utilizados pelo geneticista molecular, para a identificação de QTLs específicos para um ambiente ou que possam ser expressos em diferentes ambientes.

Em milho, sete caracteres agronômicos, incluindo produção de grãos, foram medidos na mesma população, em seis ambientes diferentes, nos Estados Unidos (STUBER *et al.*). QTLs detectados em um ambiente foram, freqüentemente, detectados em outros ambientes, sugerindo pequena interação QTLs por ambientes. Um estudo relatado em tomate avaliou uma população segregante em três diferentes ambientes (PATERSON *et al.*, 1991). Observou-se que 48% dos QTLs foram detectados em pelo menos dois ambientes e os outros 52% foram detectados em apenas um ambiente. Os QTLs que mostraram maiores efeitos em um ambiente foram também os mais prováveis de se detectar em outro ambiente. Os locais mais similares, com relação ao clima e práticas agrícolas, foram os que mostraram mais QTL em comum, condizente com o menor efeito da interação genótipos por ambientes. Estes estudos sugerem que grande proporção dos QTLs que afetam um caráter quantitativo em um ambiente pode estar ativa em outro ambiente, e isto é especialmente verdade para os QTLs com maior efeito (TANKSLEY, 1993). No entanto, PATERNIANI (1990) mostra que as variações ambientais no Brasil são mais drásticas do que nos Estados Unidos, principalmente quanto às épocas de cultivo, que normalmente apresentam variações climáticas muito

intensas, o que sugere um maior efeito da interação genótipos por ambientes. Desta forma, espera-se menor estabilidade dos QTLs identificados em nossas condições de cultivo.

A herdabilidade da característica teor de proteína obtida por componentes de variância foi alta nos dois locais (91,08% em Viçosa e 92,04% em São Gotardo) e também na análise conjunta (64,62%). Isso significa que a maior parte da variação observada é devido 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. SOARES (2000) e VIEIRA (2003), trabalhando com esta mesma população de RILs, também encontraram altos valores de herdabilidade para teor de proteína total.

As herdabilidades das outras características avaliadas também foram altas nos dois locais, variando de 52,86% a 86,70% em Viçosa (Quadro 4), e de 56,86% a 84,60% em São Gotardo (Quadro 5). Entretanto, as herdabilidades destas características na análise conjunta foram menores, variando de 7,41% a 56,56%, devido ao efeito da interação genótipo x ambiente (Quadro 6). Isto indica que a seleção praticada para estas características deve ser realizada para cada ambiente específico, e que posteriormente o cultivar de alto teor de proteína a ser recomendado deverá apresentar alta estabilidade e adaptabilidade.

JOHNSON e BERNARD (1963), estudando herdabilidade para várias características em soja, dentre elas o teor de proteína, mostraram que em gerações precoces (F_2) a herdabilidade apresenta baixos valores quando comparada com aquela obtida em gerações mais avançadas (F_4). Isso implica que é mais difícil identificar e selecionar genótipos superiores em gerações precoces (no caso F_2), resultando em pouco progresso por seleção.

Quadro 6 - Resumo da análise de variância conjunta das características analisadas nos experimentos de Viçosa e São Gotardo, e estimativas da variância genética (σ_g^2), da herdabilidade (h^2) e do coeficiente de variação (CV)

FV	GL	QUADRADOS MÉDIOS										
		PT	α'	α	β	Ácidas	Básicas	7S	11S	11S+7S	11S:7S	PT-(11S+7S)
Blocos/Locais	4	34,7711	0,2143	1,4918	1,2725	17,2847	21,3714	1,4349	15,3116	9,0527	1,5449	48,8867
RILs (R)	94	30,7771**	0,5207**	1,0927**	0,6524**	13,3516**	8,7030**	3,5284**	25,6730**	25,2760**	2,0679**	14,1905**
Locais (L)	1	241,0033**	15,766**	127,94**	4,5185**	144,3122**	65,381**	303,01**	6,7907 ^{ns}	400,5257**	88,89**	20,1499 ^{ns}
R X L.	94	10,8568**	0,3948**	0,9274**	0,3218**	5,8001**	8,0579**	2,4005**	13,2909**	16,6671**	1,2253**	12,0361**
Resíduo	376	1,7632	0,1117	0,2385	0,1033	2,1279	3,6296	0,6784	3,9679	4,2977	0,3578	4,5372
$\hat{\sigma}_g^2$		3,3200	0,02097	0,02255	0,0551	1,2586	0,1075	0,1879	2,0637	1,4348	0,1404	0,3591
$h^2\%$		64,7244	24,1688	12,3809	50,6737	56,5586	7,4128	31,9655	48,23	34,0596	40,7482	15,182
CV %		2,877778	18,2403	17,5193	16,4836	14,5717	16,1355	12,5376	9,1299	7,3029	17,1199	11,9969

** : significativo ao nível de 1% de probabilidade, pelo teste F.

3.3. Correlações genotípicas

As correlações genotípicas obtidas nas análises de Viçosa e São Gotardo podem ser observadas no Quadro 7, as correlações acima de 0,400 foram consideradas altas. Os resultados apresentados nos dois ambientes foram semelhantes. As subunidades α' , α e β apresentaram correlação positiva com a proteína 7S (0,6538; 0,8431; 0,5207, respectivamente, em Viçosa e 0,7783; 0,8943; 0,5485 em São Gotardo). Este resultado era esperado, uma vez que estas são as subunidades constituintes desta proteína. Da mesma forma, as subunidades ácidas e básicas, da proteína 11S, apresentaram correlação positiva e alta com esta proteína (0,8835; 0,7351, respectivamente, em Viçosa, e 0,7799, 0,7756 em São Gotardo). A proteína 11S e o total de proteína de reserva (11S+7S) apresentaram correlações de 0,6427 e 0,6807, respectivamente, com o conteúdo de proteína total em Viçosa, e de 0,9173 e 0,8841 em São Gotardo. A relação proteína total menos proteína de reserva (PT-(11S+7S)) apresentou correlações negativas com todas as características avaliadas; com exceção de proteína total no experimento de Viçosa, com a qual apresentou uma correlação de 0,3818. No experimento de São Gotardo, PT-(11S+7S) apresentou correlações positivas e altas com o conteúdo de proteína total e com a relação 11S:7S, 0,5622 e 0,5938, respectivamente. Nos casos em que se verificaram altas correlações genéticas, há expectativa de que as análises de QTLs venham indicar a existência de blocos gênicos controladores simultaneamente destas características ou mesmo a existência de genes pleiotrópicos.

A proteína 7S, assim como as suas subunidades α e β apresentaram correlações baixas e a subunidade α' apresentou uma correlação próxima de zero com esta característica. As correlações genotípicas entre as proteínas 11S e 7S foram próximas de zero nos dois locais.

Em estudos de genética quantitativa aplicados ao melhoramento de plantas, é necessário distinguir as causas da correlação fenotípica que podem ser de natureza genética ou ambiental. No caso da correlação genética, a explicação mais comum é a pleiotropia, em que um gene pode condicionar duas ou mais características, simultaneamente, embora esta não

necessariamente seja de fácil detecção. Outra possibilidade é a ocorrência de ligações gênicas, especialmente em populações originadas de cruzamentos entre linhagens divergentes. Já o ambiente pode influenciar na correlação fenotípica se a variação climática alterar simultaneamente duas características da planta (FALCONER e MACKAY, 1997).

FEHR (2003) avaliou a influência do ambiente nas interações entre as subunidades das proteínas 11S e 7S em 14 cultivares, em oito locais nos anos 1998, 1999 e 2000, encontrando que a proteína 7S tem uma correlação fenotípica de $-0,99$ ($P < 0,01$) com a razão 11S:7S, e que 11S tem uma correlação positiva de $0,96$ com esta razão. A correlação fenotípica entre as proteínas 7S e 11S foi de $-0,92$ ($P < 0,01$), indicando uma forte relação inversa entre estas duas proteínas. Os coeficientes de correlação fenotípica com o conteúdo de proteína total foram de $-0,34$ para 7S, $0,37$ para 11S e $0,33$ para a razão 11S:7S. Ele observou também que as proteínas 7S e 11S, e a razão 11S:7S foram influenciadas pelo ambiente.

KWANYUEN (1997), estudando 45 acessos de soja selecionados do USDA Soybean Germplasm Collection, determinou que cerca de 45% das variações de proteína entre os acessos estudados foram atribuídos a mudanças no conteúdo de 11S. Essencialmente, nenhuma correlação foi encontrada entre os níveis de 7S e proteína total.

MORAES (2003), estudando duas linhagens de alto teor de proteína obtidas no programa de melhoramento genético da soja para a agroindústria do BIOAGRO/UFV e a variedade UFVTN 105 (com ausência de lipoxigenases e teor normal de proteína) encontrou que o aumento do teor de proteína das isolinhas foi acompanhado, principalmente, pelo aumento na quantidade da proteína 11S. A soma 11S+7S aumentou quando houve aumento do teor de proteína. No entanto a quantificação individual mostra que este aumento ocorreu somente na proteína 11S, havendo até mesmo uma pequena redução de 7S.

Os resultados, encontrados neste trabalho, divergem em parte daqueles encontrados por YAKLICH (2001) que estudou linhagens de soja com teores normal e alto de proteína. Os dados apresentados por esse autor mostraram que existem diferenças significativas nas subunidades das proteínas 7S e 11S das linhagens contendo alto teor de proteína. Algumas das linhagens

estudadas apresentaram um acúmulo das subunidades α , α' e β da β -conglucina, enquanto outras apresentaram um acúmulo da subunidade A3 da glicina. Os polipeptídeos básicos da glicina também se acumularam na maioria das linhagens. Baseado nessas observações, o autor concluiu que linhagens com alto teor de proteína contêm maiores concentrações de 7S e 11S do que linhagens normais.

Uma vez que a proteína 11S contém mais aminoácidos essenciais (cisteína e metionina) que a 7S, uma maior razão entre 11S:7S deva levar a mudanças benéficas na qualidade nutricional da soja.

Quadro 7 – Estimativas de correlações genotípicas entre as 11 características avaliadas na análise de Viçosa (superior) e São Gotardo (inferior, e entre parênteses).

	α'	α	β	Ácidas	Básicas	7S	11S	11S+7S	11S:7S	PT-(11S+7S)
PT	-0,088 (0,255)	0,2048 (0,0949)	0,1893 (-0,1343)	0,5939 (0,8055)	0,4346 (0,6206)	0,1715 (0,008)	0,6427 (0,9173)	0,6807 (0,8841)	0,2042 (0,5397)	0,3818 (0,5622)
α'		0,4033 (0,6821)	0,0174 (0,0693)	-0,254 (-0,2413)	-0,0638 (0,0579)	0,6538 (0,7783)	-0,2142 (-0,1187)	0,0184 (0,205)	-0,6404 (-0,6497)	-0,1321 (-0,3085)
α			0,1522 (0,2294)	0,0622 (-0,2446)	-0,1671 (0,2168)	0,8431 (0,8943)	-0,0383 (-0,0191)	0,2539 (0,3482)	-0,7087 (-0,6785)	-0,0663 (-0,4144)
β				-0,0347 (-0,4831)	0,181 (0,2551)	0,5207 (0,5485)	0,0649 (-0,1486)	0,2425 (0,0821)	-0,3926 (-0,4769)	-0,0711 (-0,4311)
Ácidas					0,3318 (0,2098)	-0,0768 (-0,4167)	0,8835 (0,7799)	0,8279 (0,5781)	0,5055 (0,7375)	-0,3077 (0,6902)
Básicas						-0,0529 (0,2413)	0,7351 (0,7756)	0,6927 (0,8437)	0,3868 (0,3025)	-0,3347 (-0,1734)
7S							-0,0815 (-0,1145)	0,2663 (0,2999)	-0,8577 (-0,8053)	-0,1233 (-0,5139)
11S								0,939 (0,9134)	0,5554 (0,6698)	-0,3873 (0,3346)
11S+7S									0,2411 (0,3131)	-0,4171 (0,1107)
11S:7S										-0,0508 (0,5938)

4. RESUMO E CONCLUSÕES

O desenvolvimento de novas variedades de soja que associem alta produtividade, elevado teor protéico e melhor qualidade nutricional, tem sido um desafio para os melhoristas de soja de vários países. Alguns grupos de pesquisadores no Brasil e em outros países como Estados Unidos e Japão têm conseguido desenvolver variedades melhoradas que apresentam alto teor de proteína na semente e boa produtividade. Quanto à qualidade protéica do grão, principalmente no que se refere à concentração de aminoácidos sulfurados, pouco sucesso tem sido relatado na literatura. Alguns autores sugerem que maiores concentrações de aminoácidos sulfurados podem ser obtidas produzindo-se cultivares que apresentem uma maior relação 11S:7S, pois a fração 7S contém menor teor de metionina e cisteína do que a fração 11S.

Os resultados obtidos neste trabalho permitiram apontar algumas conclusões:

1. O teor das proteínas de reserva 7S e 11S, bem como os de suas respectivas subunidades apresentaram distribuição aproximadamente normal ($P < 0,01$), tanto no experimento de Viçosa quanto no de São Gotardo, o que permitirá que estas características sejam analisadas para a detecção da existência de possíveis QTLs envolvidos na expressão das mesmas, uma vez que o método de detecção de QTLs pressupõe que a característica apresente este tipo de distribuição. O teor de proteína total em São Gotardo também apresentou distribuição aproximadamente normal.
2. O teor de proteína total em Viçosa não apresentou uma distribuição aproximadamente normal, apesar do grau de simetria e de curtose terem sido próximos do esperado para distribuição normal. Caso sejam detectados QTLs relacionados a esta característica, eles deverão ser analisados com ressalva.
3. Fatores ambientais estão envolvidos na regulação da expressão das características avaliadas nos dois locais (Viçosa e São Gotardo), pois a interação RIL x local foi significativa.
4. 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.

5. As herdabilidades das outras características avaliadas também foram altas nos dois locais. Entretanto, as herdabilidades destas características na análise conjunta foram menores, devido ao efeito da interação genótipo x ambiente. Isto indica que os estudos moleculares referentes às análises de QTLs para estas características devam ser realizada para cada ambiente específico.
6. As subunidades α' , α e β apresentaram correlação positiva com a proteína 7S. Este resultado era esperado, uma vez que estas são as subunidades constituintes desta proteína. Da mesma forma, as subunidades ácidas e básicas, da proteína 11S, apresentaram correlação positiva e alta com esta proteína.
7. A proteína 11S e o total de proteína de reserva (11S+7S) apresentaram correlação positiva e alta com o conteúdo de proteína total. A proteína 7S, assim como as suas subunidades α e β apresentaram uma correlação baixa com o conteúdo de proteína total, a subunidade α' apresentou uma correlação próxima de zero. Uma vez que um maior conteúdo de 11S em relação à 7S é desejado, o fato de 11S ter apresentado uma correlação positiva e alta com proteína total torna-se benéfico no processo de seleção, já que selecionando variedades com alto teor de proteína, serão selecionadas variedades com alto teor de 11S simultaneamente.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- COATES, J.B., MEDEIROS, J.S., THANH, V.H., NIELSEN, N.C. Characterization of the subunits of β -conglycinin. **Arch. of Bioch. and Biophysics**, v.243, n.1, p.184-194, 1985.
- CRUZ, C.D. **Programa genes: aplicativo computacional em genética e estatística**. Viçosa, MG: UFV, 2001. 648p.
- CRUZ, C.D., REGAZZI, A.J. **Modelos Biométricos aplicados ao melhoramento genético**. Editora UFV. Viçosa-MG. 1997.390p.
- ERICKSON, L.R., VOLDENG, H.D., BEVERSDORF, W.D. Early generations selection for protein in *Glycine max* x *Glycine soja* crosses. **Canadian J. of Plant Sci.**, Ottawa, v.6, n.4, p.901-908, 1981.
- FALCONER, D.S., MACKAY, T.F.C. **Introducing to quantitative genetics**. Four Edition. Harlow: Longman, 1997. 464p.
- FEHR, W.R., HOECK, J.A., JOHNSON, S.L., MURPHY, J.D.N., PADILLA, G.I., WELKE, G.A. Genotype and Environment Influence on Protein Components of Soybean. **Crop Sci.**, v.43, p.511-514, 2003.
- FONTES, E.P.B., MOREIRA, M.A, DAVIES, C.S., NIELSEN, N.C. Urea-elicited changes in relative electrophoretic mobility of certain glycinin and β -conglycinin subunits. **Plant Phys.**, v.76, n.3, p.840-842, 1984.
- HANSON, D.W., WEBER, C.R. Analysis of genetic variability from generations of plant-progeny lines in soybeans. **Crop Sci.**, v.2, n.1, p.63-67, 1963.
- HARADA, J.J., BARKER, S.J., GOLDEBERG, R.B. Soybean β -conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. **The Plant Cell**, v.1, n.4, p.415-425, 1989.
- HELMS, T.C., ORF, J.H. Protein, oil, and yield of soybean lines selected for increased protein. **Crop Sci.**, v.38, n.4, p.707-710, 1998.
- INSTITUTO ADOLFO LUTZ. **Normas analíticas do Instituto Adolfo Lutz, Métodos químicos e físicos para análise de alimentos**. 3.ed. São Paulo: 1985. v.1, 533p.
- JOHNSON, H.W., BERNARD, R.L. Soybean genetics and breeding. In: NORMAN, A.G. (Ed.). **The soybean**. New York: Academy Press, 1963. p.1-73.
- KWANYUEN, P., PANTALONE, V.R., BURTON, J.W., WILSON, R.F. A new approach to genetic alteration of soybean protein composition and quality. **JAACS**, v.74, n.8, p.983-987, 1997.

- KOSHIYAMA, I. Chemical and physical properties in soybean globulins. **Cereal Chem.**, n. 45, p.394-404, 1968.
- MORAES, R.M.A. **Introgessão de alelos para alto teor de proteína em soja assistida por marcadores moleculares**. Viçosa, MG: UFV, 2003. 139p. Dissertação (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, 2003.
- MOREIRA, M.A., HERMADSON, M., LARKINS, B.A. Partial characterization of the acidic and basic polypeptides of glycinin. **J. Biol. Chem.**, v.254, n.20, p.9921-9926, 1979.
- PATERNIANI, E. Maize breeding in the tropics. **CRC Critical Review in Plant Science**, v.9, p.125-154, 1990.
- PATERSON, A.H., DAMON, S., HEWITT, J.D., ZAMIR, D., RABINOWITCH, H.D., LINCOLN, S.E., LANDER, E.S., TANKSLEY, S.D. Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. **Genetics**, v.127, n.1, p.181-197, 1991.
- 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.
- SCHUSTER, I., CRUZ, C. D. **Estatística genômica aplicada a populações derivadas de cruzamentos**. Editora UFV. Viçosa-MG. 2004. (no prelo)
- SOARES, T.C.B. **Mapeamento de locos que controlam o conteúdo de proteína em soja**. Viçosa, MG: UFV, 2000. 58p. Dissertação (Mestrado em Agroquímica) – Universidade Federal de Viçosa, 2000.
- STUBER, C.W., LINCOLN, S.E., WOLFF, D.W., HELENTJARIS, T., ANDER, E.S. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. **Genetics**, v.132, n.3, p.823-839, 1992.
- TANKSLEY, S.D. Mapping polygens. **Annu. Rev. Genet.** V.27, p. 205-233. Review. 1993.
- TEIXEIRA, A.I. **Herdabilidades e correlações entre concentrações de proteína em soja avaliadas por diferentes metodologias**. Viçosa, MG: UFV, 2003. 53p. Dissertação (Mestrado em Bioquímica Agrícola) - Universidade Federal de Viçosa, 2003.
- VENCOVSKY, R., BARRIGA, P. **Genética biométrica no fitomelhoramento**. Sociedade Brasileira de Genética. Ribeirão Preto-SP. 1992. 486p.
- VIEIRA, A.J.D. **Identificação de QTLs associados à fotossíntese e a características de crescimento em soja**. Viçosa, MG. UFV, 2003. 73p.

Dissertação (Doutorado em Fisiologia Vegetal) - Universidade Federal de Viçosa, 2003.

WILCOX, J.R. Increasing seed protein in soybean with eight cycles of recurrent selection. **Crop Sci.**, v.38, p.1536-1540, 1998.

WILCOX, J.R., SHIBLES, R.M. Interrelationships among seed quality attributes in soybean. **Crop Sci.**, v.41, p.11-14, 2001.

YAKLICH, R.W., HELM, R.M., COCKRELL, G., HERMAN, E. M. Analysis of the distribution of the major seed allergens in a core collection of *Glycine max* accessions. **Crop Sci.**, v.39, n.2, p.144-147, 1999.

YAKLICH, R.W. Beta-Conglycinin and glycinin in high-protein soybean seeds, **J. of Agric. and Food Chem.**, v.49, n.2, p 729-735, 2001.

CAPÍTULO 2

MAPEAMENTO DE LOCOS QUE CONTROLAM O CONTEÚDO DE PROTEÍNA TOTAL E DAS PRINCIPAIS PROTEÍNAS DE RESERVA EM SOJA

RESUMO

SOARES, Taís Cristina Bastos, D. S., Universidade Federal de Viçosa, março de 2004. **Mapeamento de locos associados ao conteúdo de proteínas de reserva em soja.** Orientador: Maurilio Alves Moreira. Conselheiros: Cosme Damião Cruz e Everaldo Gonçalves de Barros.

Uma população de 118 RILs (linhagens recombinantes endogâmicas) foi obtida através do cruzamento entre o acesso BARC-8 (com alto teor de proteína) e a variedade brasileira Garimpo (com teor de proteína normal). Na geração F_9 foram abertas linhas, que constituíram o material genético utilizado neste trabalho. Esse material foi analisado, utilizando-se marcadores microssatélites, RAPD e AFLP. Foram obtidos 29 grupos de ligação pouco saturados, contendo 95 marcadores, e mais 79 marcas não-ligadas. Foi feita também a análise das características: teor de proteína total em sementes de soja, teor das proteínas de reserva 11S e 7S e de suas respectivas subunidades, e da relação proteína total menos proteínas de reserva (PT-(11S+7S)), para essa população cultivada em dois ambientes distintos: Viçosa, MG, e São Gotardo, MG. Análise de associação de marcas simples, regressão múltipla e mapeamento por intervalo composto foram utilizados para detectar e mapear as regiões genômicas associadas com estas características. No experimento de Viçosa, foram encontrados 13 marcadores associados ao conteúdo de proteína total e explicam de 4,39% a 13,32% da variação do teor de proteína. Ao conteúdo da subunidade α' da proteína 7S, foram encontrados 9 marcadores associados, que explicam de 4,21% a 7,09% da variação desta característica. Dezesete marcadores mostraram associação com o conteúdo da subunidade α da proteína 7S e explicam de 4,07% a 13,16% da variação desta característica. Para o conteúdo da subunidade β foram encontrados 16 marcadores associados, e eles explicam de 4,78% a 21,80% da variação desta característica. Associados ao conteúdo das subunidades ácidas e básicas da proteína 11S, foram encontrados 17 e 8 marcadores, respectivamente, e eles explicam de 4,08% a 9,60% (subunidades ácidas) e de 5,49% a 11,10% (subunidades básicas) da variação destas características. Associados ao conteúdo da característica proteína total menos proteína de reserva ((PT-(11S+7S))) foram encontrados sete marcadores que explicam de 4,11% a

10,77% da variação desta característica. Dezenove marcadores mostraram associação com o conteúdo de 7S e eles explicam de 4,08% a 15,57% da variação desta característica. Dezesesseis marcadores apresentaram associação com o conteúdo de 11S e eles explicam de 4,17% a 9,69% da variação desta característica. No experimento de São Gotardo foram encontrados 13 marcadores associados a conteúdo de proteína total que explicam de 4,31% a 12,11% da variação do teor de proteína. Para a subunidade α' da proteína 7S, foram encontrados 13 marcadores que explicam de 4,34% a 10,42% da variação desta característica. Cinco marcadores mostraram associação com a subunidade α da proteína 7S e explicam de 4,32% a 6,17% da variação desta característica. Para a subunidade β foram encontrados 12 marcadores associados, e eles explicam de 4,09% a 9,11% da variação desta característica. Para as subunidades ácidas e básicas da proteína 11S, foram encontrados 13 e 17 marcadores respectivamente e eles explicam de 4,21% a 13,36% (subunidades ácidas) e de 4,12% a 8,12% (subunidades básicas) da variação destas características. Para a característica proteína total menos proteína de reserva ((PT-(11S+7S)) foram encontrados 6 marcadores associados que explicam de 4,39% a 7,92% da variação desta característica. Seis marcadores mostraram associação com o conteúdo de 7S e eles explicam de 5,16% a 8,11% da variação desta característica. Dezesesseis marcadores apresentaram associação com o conteúdo de 11S e eles explicam de 4,15% a 13,67% da variação desta característica. Foram encontrados QTLs associados: ao conteúdo de proteína total (GL I, que explica 14,74%), ao conteúdo da subunidade α (GL G e GLC2 que explicam em conjunto 28,83% da variação desta característica), ao conteúdo da subunidade β (GL G e dois no grupo K, que explicam em conjunto 32,62% da variação desta característica), ao conteúdo de PT-(11S+7S) (GL D1b, que explica 13,71%) e conteúdo de 7S (GLG e GLE) que explicam em conjunto 21,86% da variação desta característica no experimento de Viçosa. No experimento de São Gotardo foram encontrados QTLs associados ao conteúdo de proteína total (GL 3 e I, que explicam em conjunto 11,42%), ao conteúdo da subunidade α' (GL A2, que explica 10,64%), ao conteúdo da subunidades ácidas (GL 3, que explica 12,06%), PT-(11S+7S) (GL I, que explica 11,34%) e ao conteúdo de 11S (dois

QTLs em diferentes fragmentos do GL K) que explicam em conjunto 20,88% da variação desta característica. Os marcadores associados ao teor de proteína foram diferentes de um local para outro, o que indica interação entre o genótipo e o ambiente. O grupo de ligação I foi associado a QTLs para proteína total em Viçosa e em São Gotardo, este fato sugere que os locos associados a esse grupo de ligação são conservados. Os dois fragmentos do grupo K que apresentaram QTLs associados ao conteúdo da subunidade β de 7S no experimento de Viçosa, apresentaram QTLs associados ao conteúdo de 11S em São Gotardo.

1. INTRODUÇÃO

As primeiras variedades de soja cultivadas no Brasil, no início dos anos 70 foram introduções de variedades americanas ou derivadas de inter cruzamento destas variedades. Sabe-se também que os programas de melhoramento, principalmente nos Estados Unidos, priorizavam o desenvolvimento de variedades produtivas e com alto teor de óleo. Pelo fato destes dois caracteres serem negativamente correlacionados com teor de proteína, este tem se mantido em níveis mais baixos (em torno de 40%) do que podem ser conseguidos (em torno de 50%) (PIOVESAN, 2000).

Muito da expressão das características de importância agrônômica e econômica, em vários grupos vegetais, incluindo soja, resulta da ação conjunta de mais de um gene e das interações destes entre si e com o ambiente. Tais características exibem valores fenotípicos com distribuição contínua que são difíceis de serem analisados, sendo determinados vários locos 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 da 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 ampla variabilidade genética (BRIM *et al.*, 1973).

Melhoramento para características quantitativas não é 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, marcadores genéticos (marcas genéticas herdáveis e detectáveis que estão associadas a genes no genoma de um organismo) podem ser usados como importantes ferramentas por geneticistas (DARVASI e SOLLER, 1994).

O avanço das técnicas moleculares permitiu a obtenção de grande número de marcadores genéticos que exploram a variabilidade existente nas

seqüências de bases do DNA, possibilitando a elaboração de mapas genéticos detalhados para várias espécies vegetais. A cultura da soja, que é amplamente estudada, requer mapas genéticos que possam ser utilizados por diferentes grupos de pesquisa, visando ao intercâmbio de informações.

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. O primeiro mapa genético foi construído com número de grupos de ligação bem próximo ao de cromossomos, no qual foram definidos 21 grupos contendo 81 marcadores RFLP e 13 marcadores RAPD (WILLIAMS *et al.*, 1990). A soja é diplóide, apresentando em suas células somáticas $2n = 2x = 40$ cromossomos.

SHOEMAKER (1994) desenvolveu um mapa com 24 grupos de ligação contendo 359 marcadores RFLP, no qual foram mapeadas nove regiões ligadas a características quantitativas e 10 ligadas a caracteres qualitativos.

KEIM *et al.* (1997) construíram um mapa saturado de soja numa população de 300 linhagens endogâmicas recombinantes do cruzamento entre as variedades BSR-101 X PI 437.654. Primeiramente, foi construído um mapa-base com 165 marcadores RFLP, utilizando-se toda a população de 300 indivíduos. Quarenta e dois indivíduos foram empregados para a inclusão de 650 marcadores AFLP. O mapa final apresentou 840 marcadores, sendo 165 RFLP, 650 AFLP e 25 RAPD, abrangendo 3.441 cM em 28 grupos de ligação.

CREGAN *et al.* (1999) construíram um mapa integrado para a soja utilizando dados de três mapas diferentes: USDA/Iowa State (*G. max* x *G. soja*), Univ. de Utah (Minsoy x Noir) e da Univ. de Nebraska (Clark x Harosoy). Esse mapa integrado foi definido, utilizando-se um total de 606 *primers* SSR e outros marcadores (RFLP, RAPD, AFLP, isoenzimas e marcadores clássicos), no total de 1.423 marcadores, sendo definidos 20 grupos de ligação-consenso.

A seleção de características agronômicas monitorada por marcadores moleculares baseia-se no princípio de que um gene, ou um grupo de genes, encontra-se ligado a um marcador genético de fácil identificação. Dessa forma, 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 para determinar esse tipo de associação envolve a construção de mapas genéticos detalhados, nos quais podem ser localizadas regiões

contendo locos para características quantitativas (QTLs). Altura de plantas, produção de sementes, resistência ao acamamento, percentagem de óleo e de proteína e tamanho das sementes são características quantitativas (FEHR, 1978).

Uma característica que tem merecido considerável interesse pelos melhoristas é o teor de proteínas. Algumas das informações importantes sobre herdabilidade e as associações de caracteres com o teor de proteína já estão disponíveis na literatura.

JOHNSON e BERNARD (1963), estudando a herdabilidade de várias características de interesse em soja, demonstraram que, em gerações precoces (F_2), o teor de proteína apresenta baixos valores de herdabilidade, em comparação com gerações mais avançadas (F_4). Esse fato implica em dificuldade de identificar e selecionar genótipos superiores em gerações precoces, resultando em pouco progresso por seleção.

WILCOX e CAVINS (1995), estudando correlações entre teor de proteína e produtividade em uma série de retrocruzamentos entre dois progenitores contrastantes com relação a essas duas características, verificaram correlação negativa decrescente entre teor de proteína e produção de grãos, à medida que se avançaram os ciclos de retrocruzamentos. Dessa maneira, obtiveram, ao final do terceiro ciclo, linhagens produtivas e com alto teor de proteína.

Correlações genéticas entre teor de proteína e produção de grãos parecem não oferecer obstáculos para obtenção de novos genótipos de soja superiores quanto ao teor de proteína e à produção de grãos, desde que certas estratégias de genética, de métodos de melhoramento e até mesmo de uso de marcadores moleculares sejam usadas. A utilização de marcadores moleculares permite identificar ligações genéticas indesejáveis, como algumas existentes entre baixa produtividade e alto teor de proteína.

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 MIRANDA, 2002), e também identificar QTLs associados às características: teor de proteína total em sementes de soja, teor das proteínas de reserva 11S e 7S e de suas respectivas subunidades, e da relação proteína total menos proteínas de

reserva (PT-(11S+7S)). Para isso, foram acrescentados ao mapa original, marcadores microssatélites não mapeados anteriormente e marcadores AFLP. 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. MATERIAL E MÉTODOS

2.1. Material genético

Uma população de 118 RILs foi obtida do cruzamento entre dois progenitores contrastantes para teor de proteínas, a linhagem BARC-8, obtida da “Beltsville Agricultural Experimental Station” – USDA, Beltsville, MA, USA (genótipo com alto teor de proteína, em torno de 50%) e a variedade comercial Garimpo (genótipo com teor normal de proteínas, em torno de 36%), sendo o progenitor feminino BARC-8. As RILs foram obtidas pela descendência, a partir da geração F_2 , de uma única semente de cada planta (SSD – “single seed descent”). Na geração F_9 foram abertas linhas, que constituíram o material genético utilizado neste trabalho.

Essas RILs foram plantadas em duas localidades: Viçosa e São Gotardo, ambas no estado de Minas Gerais. Os experimentos foram instalados no delineamento em blocos casualizados com três repetições. Cada parcela foi constituída de uma fileira de 2 metros, com espaçamento de 0,5 metros entre fileiras. O plantio foi realizado em Viçosa na segunda quinzena de novembro de 2001, e em São Gotardo, na primeira quinzena de dezembro de 2001.

Em Viçosa, o experimento foi instalado no Campo experimental Diogo Alves de Mello, localizado no Campus da Universidade Federal de Viçosa, na Zona da Mata do Estado de Minas Gerais, cujas coordenadas geográficas são 20°45'20"S e 42°51'W. A sede do município situa-se aproximadamente a 650m de altitude. Em São Gotardo, o experimento foi instalado na Estação experimental da Cooperativa Agropecuária do Alto Paranaíba (COOPADAP), na região do Alto Paranaíba do Estado de Minas Gerais, cujas coordenadas geográficas são 19°29'59"S e 46°06'05"W. A sede do município situa-se a 1055m de altitude. O solo, encontrado nas duas localidades é classificado como Latossolo Vermelho-Amarelo distrófico, textura argilosa.

2.2. Extração de DNA

Para a extração do DNA, cinco folhas de cinco plantas de cada fileira foram coletadas, embaladas em papel-alumínio, congeladas em N₂ líquido e armazenadas a -80°C. A extração do DNA foi feita com base no protocolo descrito por DOYLE e DOYLE (1990), com algumas modificações. Cerca de 200 a 300 mg de folhas foram triturados na presença de N₂ líquido, sendo o pó resultante transferido para microtubos. Após a adição de 650-800 µL de tampão de extração constituído de Tris-HCl 50 mM, pH 8,0, EDTA 50 mM, pH 8,0, NaCl 0,7 M, β-mercaptoetanol 1% v/v e CTAB 1% p/v, este último adicionado separado dos demais componentes. As amostras foram, então, incubadas em banho-maria a 65°C por, aproximadamente, uma hora, sendo, posteriormente, centrifugadas por cinco minutos a 18.845 xg em centrífuga Beckman, AVANTI™ 30 e o sobrenadante, transferido para novos tubos.

Ao sobrenadante foi adicionado igual volume da mistura de clorofórmio:álcool isoamílico (24:1). Os tubos foram suavemente invertidos e centrifugados por cinco minutos a 18.845 xg. A fase superior foi transferida para outros tubos e o DNA, precipitado com a adição de 2/3 do volume obtido de isopropanol gelado. O isopropanol foi descartado, sendo adicionados 650 µL de etanol gelado (95%) ao precipitado. Após o descarte do etanol, o precipitado foi secado ao ar, ressuspenso em TE (10 mM de Tris-HCl, pH 8,0, 1 mM de EDTA) e incubado a 37°C, por 30 minutos, com RNase A na concentração final de 40 µg/mL. O DNA foi recuperado por precipitação, mediante a adição de 30 µL de NaCl 5 M e 200 µL de isopropanol gelado aos tubos, seguido de centrifugação a 18.845 xg por 15 minutos. Após a lavagem do precipitado com etanol 90% gelado e a secagem ao ambiente, o DNA foi ressuspenso em 200-300 µL de água.

A concentração do DNA foi estimada espectrofotometricamente por leitura da absorbância a 260 nm, em que cada unidade de absorbância correspondeu à concentração de 50 µg/mL de DNA fita dupla (SAMBROOK *et al.*, 1989). A integridade do DNA foi avaliada em gel de agarose 0,8%.

2.3. Condições de amplificação e de eletroforese

As reações de amplificação dos *primers* microssatélites foram feitas em um volume total de 15 μ L, contendo 10 mM de Tris, pH 8,3; 50 mM de KCl; 2 mM de $MgCl_2$; 100 μ M de cada um dos desoxinucleotídeos (dATP, dTTP, dGTP e dCTP), 0,6 μ M de cada *primer* ("Research Genetics", Huntsville, AL, 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 término destes 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 μ L de brometo de etídio 3 mg/mL em um tampão TBE 1 X (Tris-borato 90 mM e EDTA 2 mM, pH 7,0), a 100 volts. Terminada a corrida, os géis foram fotografados sob luz ultravioleta, pelo sistema de fotodocumentação Eagle Eye II (Stratagene, Lajolla, CA).

Amostras de DNA foram amplificadas pela técnica de RAPD, de acordo com WILLIAMS *et al.* (1990), com *primers* decâmeros ("Operon Technologies", Alameda, CA), 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 extensão final 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, Lajolla, CA).

Os marcadores AFLP foram obtidos segundo a metodologia proposta por VOS *et al.* (1995), com modificações sugeridas pelo AFLP Starter Primer Kit (Gibco-BRL, Alameda, CA). O DNA genômico de cada indivíduo foi clivado com as endonucleases de restrição *EcoRI* e *MseI*. Aos fragmentos de restrição, foram ligados os adaptadores específicos para as extremidades coesivas das duas enzimas *EcoRI* e *MseI*. Foi feita uma amplificação pré-seletiva utilizando

primers que possuem, além da sequência complementar a cada adaptador, um nucleotídeo na extremidade 3' (*EcoRI* +A e *MseI* +C). A segunda etapa de amplificação foi a amplificação seletiva, e utilizou os produtos da pré-amplificação diluídos como molde e *primers* seletivos. Estes *primers* seletivos são marcados com fluorescência, sendo que três *primers EcoRI* são marcados com a fluorescência JOE (E_1 , E_5 e E_8), três com a fluorescência TAMRA (E_2 , E_4 e E_7) e dois com 5-FAM (E_3 e E_6). A análise dos fragmentos foi realizada em gel de poliacrilamida desnaturante contendo 4,25% de poliacrilamida (acrilamida:bisacrilamida 19:1), 36% p/v de uréia em TBE 1X. Os fragmentos amplificados foram detectados e analisados pelo sequenciador automático ABI 377[®] (Applied Biosystems, Foster City, CA, EUA), utilizando um marcador de peso molecular GENESCAN 500-ROX (Applied Biosystems, Foster City, CA, EUA), em cada linha.

2.4. Construção do mapa de ligação genética

Os marcadores foram codificados como homozigotos para os alelos vindos de Garimpo (0), homozigotos para os alelos vindos de BARC-8 (2) e falhas (9) das bandas amplificadas em cada um dos 118 indivíduos da população F_9 . Como os dados foram obtidos de RILs, os 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, RAPD e AFLP e o estabelecimento dos grupos de ligação foram feitos com a utilização do programa GQMOL (www.ufv.br/dbg/gqmol/gqmol.htm). Os marcadores foram agrupados, utilizando-se *LOD score* mínimo de 3,0 e frequência máxima de recombinação de 30%. Para ordenar os marcadores nos respectivos grupos de ligação utilizou-se o método SARF, que se baseia na escolha da melhor ordem também pela menor soma das recombinações adjacentes. A função de mapeamento de Kosambi foi utilizada para a conversão da frequência de recombinação para centiMorgans (cM).

2.5. Identificação e mapeamento de QTLs

Primeiramente, foram realizadas análises de variâncias de todos os dados fenotípicos: teor de proteína total em sementes de soja, teor das proteínas de reserva 11S e 7S e de suas respectivas subunidades, e da relação proteína total menos proteínas de reserva (PT-(11S+7S)). Todos os marcadores com segregação mendeliana 1:1 foram avaliados para determinar associações com estas características pela análise de marca simples, utilizando o programa GQMOL, com um nível de significância de $P < 0,05$ para tentativa de detecção de QTLs. A análise de regressão linear também foi realizada com o intuito de comprovação dos resultados da ANOVA e, adicionalmente, para a determinação do R^2 , que neste caso é interpretado como a proporção da variância fenotípica explicada pelo marcador. Este procedimento permite a avaliação de todos os marcadores em relação à característica fenotípica, inclusive aqueles que não agruparam no mapa de ligação. A precisão desta metodologia é baixa por não permitir a detecção da melhor posição do QTL, não permite a estimação dos efeitos gênicos do QTL.

As informações do mapa de ligação foram utilizadas para mapear os QTLs ao longo dos grupos de ligação por meio do mapeamento por intervalo composto (ZENG, 1993; 1994; JANSEN, 1993), empregando-se o programa GQMOL (www.ufv.br/dbg/gqmol/gqmol.htm). Esse tipo de metodologia diferencia-se do mapeamento por intervalo clássico, por isolar o efeito de outros QTLs ligados e não 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:

$$y_j = \beta_0 + \beta^*x_j^* + \sum \beta_k x_{jk} + \varepsilon_i$$

em que

y_j = valor da característica no j-ésimo indivíduo;

β_0 = intercepto;

β^* = efeito aditivo do QTL, expresso como a diferença entre os dois homozigotos;

x_j^* = variável indicadora assumindo valores de probabilidade dependente do genótipo dos marcadores i e j e da posição que está sendo testada no intervalo entre esses marcadores (Quadro 1, ignorando-se recombinações duplas dentro do intervalo);

β_k = coeficiente de regressão parcial entre o fenótipo y e o marcador k;

x_{jk} = coeficiente conhecido para o k-ésimo marcador (ou cofator), no j-ésimo indivíduo, com valores -1 e 1, dependendo do genótipo dos marcadores; e

ε_i = variável aleatória.

Quadro 1 – Especificações da variável indicadora x^* para o modelo de mapeamento por intervalo composto, utilizando-se RILs

Grupo	Genótipo do marcador		N ^o de indivíduos	x^*
	i	i+1		
1	+	+	n1	$(1-r'_a-r'_b)/(1-r')$
2	+	-	n2	$(r'_b-r'_a)/(r')$
3	-	+	n3	$-(r'_a-r'_b)/(1-r')$
4	-	-	n4	$-(1-r'_a-r'_b)/(1-r')$

+ indica genótipo homozigoto para o progenitor 1, e - indica homozigoto para o progenitor 2.

$$\text{sendo: } r'_a = \frac{2r_a}{1+2r_a}, r'_b = \frac{2r_b}{1+2r_b} \text{ e } r' = \frac{2(r_a + r_b)}{(1-2r_a)(1+2r_b)}$$

em que

r_a : porcentagem de recombinação entre a marca i e o QTL

r_b : porcentagem de recombinação entre a marca i+1 e o QTL

r : porcentagem de recombinação entre as marcas i e i+1.

Os coeficientes β_k e x_{jk} são utilizados para controlar o “background” genético, eliminando-se os efeitos de outros possíveis QTLs sobre β^* .

2.5.1. Determinação do ponto de corte para o mapeamento de QTLs

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. A razão de verossimilhança (LR) possui aproximação com o χ^2 pela seguinte expressão:

$$LR = -2n \ln \frac{SQD_{\text{reduzido}}}{SQD_{\text{completo}}}$$

em que

n: número de indivíduos avaliados

SQD_{reduzido} : soma de quadrado do desvio do modelo reduzido, dado por

$$y_j = b_o + \sum b_k c_{jk} + e_j$$

SQD_{completo} : soma de quadrado de desvio do modelo.

A estatística LR está associada ao χ^2 com 1 grau de liberdade.

O nível crítico foi estabelecido considerando que α_M e α_C e α_I são os níveis de significância do mapa para o cromossomo e um intervalo particular, respectivamente. Tem-se a seguinte relação entre eles (ZENG, 1994):

$$\alpha_C = 1 - (1 - \alpha_M)^{1/C} \cong \alpha_M/C$$

$$\alpha_I \cong \alpha_C/M.$$

em que

C = número de cromossomos (grupos de ligação)

M = número de marcas dentro de um grupo de ligação

Dessa forma, mesmo um nível de significância não muito rigoroso para o mapa pode levar a um nível de significância muito drástico para cada intervalo, aumentando a chance de se cometer o erro tipo I. Por esse motivo, foi utilizado um nível de significância de 10% para o mapa, que leva a um nível de significância de 0,00345% para cada grupo de ligação, considerando-se 29 grupos.

3. RESULTADOS

3.1. Detecção de polimorfismos e seleção dos *primers* de microssatélites, RAPD e AFLP

Os *primers* de microssatélites foram, inicialmente, testados nos progenitores com a finalidade de selecionar aqueles que produzissem fragmentos polimórficos e que, conseqüentemente, pudessem ser mapeados na população segregante.

Foi testado um total de 650 pares de *primers* de microssatélites (“Research Genetics”, Huntsville, AL, EUA). Sendo que 567 destes pares haviam sido testados por SOARES (2000). Do total dos 650 pares, 134 exibiram polimorfismo nos progenitores, mas somente 94 apresentaram fragmentos polimórficos na população segregante. Dos 40 pares de *primers* restantes, 27 não segregaram na população e 13 não tiveram boa qualidade de amplificação, o que não permitiu a análise dos fragmentos gerados. Na Figura 1, pode-se observar que a família 13 apresenta tanto o alelo de Garimpo quanto o de BARC-8, representando um exemplo de heterozigose.

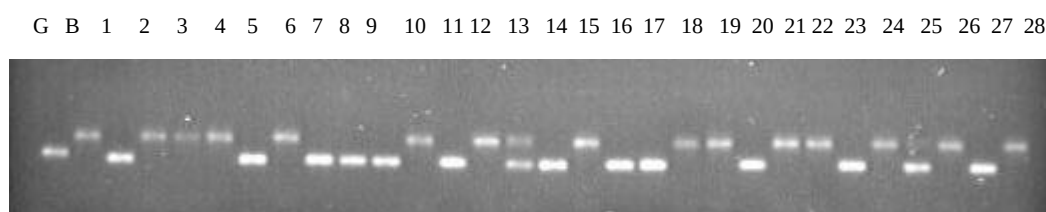


Figura 1 – Gel de agarose 3% fotografado sob luz ultravioleta apresentando fragmentos polimórficos amplificados, utilizando-se *primers* de microssatélites nos pais Garimpo (G) e BARC-8 (B) e em parte da população segregante (1 a 28).

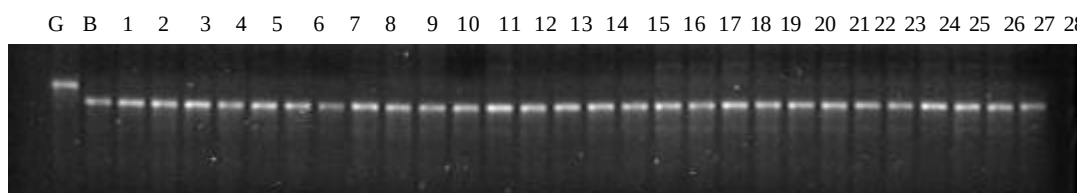


Figura 2 – Gel de agarose 3% fotografado sob luz ultravioleta apresentando fragmentos amplificados, utilizando-se *primers* de microssatélites polimórficos nos progenitores Garimpo (G) e BARC-8 (B) e que não segregaram na população parcialmente representada (1 a 28).

A maior parte dos 27 *primers* que foram polimórficos entre os progenitores e não apresentaram polimorfismos na população segregante exibiu o alelo do progenitor BARC-8 em toda a população (Figura 2). Provavelmente, essa ausência de polimorfismo na população seja devida ao fato de que um dos genótipos utilizados como progenitor para obtenção das RILs foi a variedade Garimpo, e o genótipo utilizado para se detectarem os polimorfismos entre os progenitores foi a variedade Garimpo RCH (resistente ao cancro-da-haste). A variedade Garimpo RCH foi obtida por meio de retrocruzamentos tendo como progenitor recorrente a variedade Garimpo. Os cinco retrocruzamentos efetuados podem não terem sido suficientes para a completa recuperação do genoma da variedade Garimpo na variedade Garimpo RCH, podendo haver divergência genética entre esses dois genótipos. Outra suposição é a de que a variedade Garimpo utilizada não seja geneticamente pura, o que também poderia contribuir com esse tipo de comportamento entre os marcadores.

O mapa obtido por SOARES (2000), para esta população foi saturado por MIRANDA (2002) com 64 *primers* RAPD.

Foram testadas também 64 combinações de *primers* AFLP. Destas, 40 foram polimórficas nos pais, mas somente 32 foram polimórficas na população.

A segregação de cada um dos 190 marcadores (microssatélites, RAPD e AFLP) foi testada por meio do teste qui-quadrado ($P < 0,05$), e 174 fragmentos segregaram na proporção mendeliana esperada de 1:1 depois de descartados os locos heterozigotos (Quadro 2). Em uma população F_9 , espera-se encontrar

0,195% de heterozigotos. Na população trabalhada, foi encontrada essa proporção de heterozigotos ($P < 0,05$ no teste qui-quadrado, dados não mostrados). Como o programa que gera o mapa não aceita heterozigotos quando os dados são codificados como RILs, uma vez que heterozigotos não são dados esperados para este tipo de população, tornou-se necessário eliminar os heterozigotos da análise.

A baixa diversidade genética da soja dificulta a obtenção de um número maior de marcadores que apresentam fragmentos polimórficos em uma população intra-específica. Estudos anteriores têm evidenciado que a incidência de polimorfismo entre dois genótipos de soja é freqüentemente muito baixa (SHOEMAKER *et al.*, 1996). Em estudo para mapear 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 e nenhum marcador era polimórfico em todas essas populações. AKKAYA *et al.* (1995), conseguiram selecionar somente 40 *primers* SSR, que apresentaram um único produto de amplificação e que eram polimórficos entre os progenitores, para serem utilizados no mapeamento de uma população de soja constituída de 60 plantas F_2 obtidas do cruzamento entre NILs (linhagens quase isogênicas) dos cultivares Clark e Harosoy.

Quadro 2 – Teste qui-quadrado (χ^2) para a segregação dos marcadores. O número 0 representa o alelo presente na variedade de soja Garimpo e o 2, alelo presente em BARC-8

Primer	0	2	χ^2	P%	Primer	0	2	χ^2	P%	Primer	0	2	χ^2	P%
OPN14	64	54	0,85	35,72 ^{ns}	OPAT19	54	64	0,85	35,72 ^{ns}	Satt303	49	57	0,60	43,71 ^{ns}
PP09	58	60	0,03	85,39 ^{ns}	OPAU02a	60	58	0,03	85,39 ^{ns}	Satt300	59	56	0,08	77,96 ^{ns}
OPP12	59	59	0,00	100,0 ^{ns}	OPAU02b	53	65	1,22	26,92 ^{ns}	Satt304	59	53	0,32	57,07 ^{ns}
OPW08	56	62	0,31	58,07 ^{ns}	OPAU04	65	53	1,22	26,92 ^{ns}	Satt325	62	54	0,55	45,76 ^{ns}
OPY02	65	53	1,22	26,92 ^{ns}	OPAU19	62	56	0,31	58,07 ^{ns}	Satt333	55	63	0,54	46,14 ^{ns}
OPY07	59	59	0,00	100,0 ^{ns}	OPAW09a	59	58	0,01	92,63 ^{ns}	Satt345	66	52	1,66	19,74 ^{ns}
OPY13a	62	56	0,31	58,07 ^{ns}	OPAW09b	64	53	1,03	30,91 ^{ns}	Satt369	65	51	1,69	19,36 ^{ns}
OPY13b	60	58	0,03	85,39 ^{ns}	OPAW18	60	58	0,03	85,39 ^{ns}	Satt318	57	60	0,08	78,15 ^{ns}
OPY14	57	61	0,14	71,27 ^{ns}	OPAX12	57	61	0,14	71,27 ^{ns}	Satt324	81	37	16,4	0,0051 *
OPY17	62	56	0,31	58,07 ^{ns}	OPAX20	53	65	1,22	26,92 ^{ns}	Satt384	54	56	0,04	84,87 ^{ns}
OPZ06a	59	59	0,00	100,0 ^{ns}	OPAZ02	59	59	0,00	100,0 ^{ns}	Satt353	58	56	0,04	85,14 ^{ns}
OPZ06b	61	57	0,14	71,27 ^{ns}	OPAZ07	58	60	0,03	85,39 ^{ns}	Satt377	57	58	0,01	92,57 ^{ns}
OPZ18	56	62	0,31	58,07 ^{ns}	OPBD16	55	63	0,54	46,14 ^{ns}	Satt399	55	61	0,31	57,74 ^{ns}
OPAA 03	58	60	0,03	85,39 ^{ns}	OPBE04	67	51	2,17	14,07 ^{ns}	Satt422	55	61	0,31	57,74 ^{ns}
OPAA14	58	60	0,03	85,39 ^{ns}	OPBE13	63	55	0,54	46,14 ^{ns}	Satt441	50	64	1,72	18,97 ^{ns}
OPAA15	62	56	0,31	58,07 ^{ns}	OPBH02	56	62	0,31	58,07 ^{ns}	Satt449	56	59	0,08	77,96 ^{ns}
OPAB08a	56	62	0,31	58,07 ^{ns}	Satt177	45	66	3,97	4,62 *	Satt462	53	62	0,70	40,13 ^{ns}
OPAB08b	64	54	0,85	35,72 ^{ns}	Satt144	47	63	2,33	12,71 ^{ns}	Satt468	59	58	0,01	92,63 ^{ns}
OPAB08c	57	61	0,14	71,27 ^{ns}	Satt055	48	63	2,03	15,45 ^{ns}	Satt475	51	65	1,69	19,36 ^{ns}
OPAB11	59	59	0,00	100,0 ^{ns}	Satt186	55	53	0,04	84,73 ^{ns}	Satt476	59	53	0,32	57,07 ^{ns}
OPAB13	65	53	1,22	26,92 ^{ns}	Sat-091	46	67	3,90	4,82*	Satt495	59	57	0,03	85,26 ^{ns}
OPAB15	71	47	4,88	2,71 *	Sat-135	67	48	3,14	7,64 ^{ns}	Satt506	59	59	0,00	100,0 ^{ns}
OPAB18a	53	65	1,22	26,92 ^{ns}	Sat-105	55	60	0,22	64,10 ^{ns}	Satt514	49	63	1,75	18,58 ^{ns}
OPAB18b	65	52	1,44	22,94 ^{ns}	Sat-084	43	70	6,45	1,10 *	Satt520	62	54	0,55	45,76 ^{ns}
OPAB19	67	51	2,17	14,07 ^{ns}	Sat-085	55	59	0,14	70,79 ^{ns}	Satt522	55	61	0,31	57,74 ^{ns}
OPAB20	68	50	2,75	9,75 ^{ns}	Sat-112	54	53	0,01	92,29 ^{ns}	Satt523	51	60	0,73	39,29 ^{ns}
OPAC01a	63	55	0,54	46,14 ^{ns}	Satt190	54	56	0,04	84,87 ^{ns}	Satt526	54	62	0,55	45,76 ^{ns}
OPAC01b	50	68	2,75	9,75 ^{ns}	Satt164	64	48	2,29	13,05 ^{ns}	Satt527	52	58	0,33	56,72 ^{ns}
OPAC02	63	55	0,54	46,14 ^{ns}	Satt167	48	62	1,78	18,19 ^{ns}	Satt528	58	57	0,01	92,57 ^{ns}
OPAC04	59	58	0,01	92,63 ^{ns}	Satt183	55	61	0,31	57,74 ^{ns}	Satt529	59	59	0,00	100,0 ^{ns}
OPAI11a	61	57	0,14	71,27 ^{ns}	Satt153	68	46	4,25	3,93 *	Satt531	68	49	3,09	7,8994 ^{ns}
OPAI11b	61	57	0,14	71,27 ^{ns}	Satt146	46	61	2,10	14,70 ^{ns}	Satt538	61	57	0,14	71,27 ^{ns}
OPAI15a	60	58	0,03	85,39 ^{ns}	Satt181	55	55	0,00	100,0 ^{ns}	Satt549	52	65	1,44	22,94 ^{ns}
OPAI15b	54	64	0,85	35,72 ^{ns}	Satt184	58	47	1,15	28,30 ^{ns}	Satt594	63	55	0,54	46,14 ^{ns}
OPAI01	57	61	0,14	71,27 ^{ns}	Satt012	63	50	1,50	22,13 ^{ns}	Satt282	54	58	0,14	70,54 ^{ns}
OPAN03	62	56	0,31	58,07 ^{ns}	Satt129	55	62	0,42	51,75 ^{ns}	Satt215	58	54	0,14	70,54 ^{ns}
OPAN09	65	53	1,22	26,92 ^{ns}	Satt123	60	42	3,18	7,47 ^{ns}	Satt242	52	56	0,15	70,03 ^{ns}
OPAN20	56	62	0,31	58,07 ^{ns}	Satt147	52	57	0,23	63,2 ^{ns}	Satt471	57	58	0,01	92,57 ^{ns}
OPAO06	58	60	0,03	85,39 ^{ns}	Satt139	53	58	0,23	63,50 ^{ns}	Satt518	58	58	0,00	100,0 ^{ns}
OPAP04a	62	56	0,31	58,07 ^{ns}	Satt193	62	52	0,88	34,89 ^{ns}	Satt082	53	59	0,32	57,07 ^{ns}
OPAP04b	56	62	0,31	58,07 ^{ns}	Satt199	57	54	0,08	77,58 ^{ns}	Satt537	54	61	0,43	51,39 ^{ns}
OPAP04c	61	57	0,14	71,27 ^{ns}	Satt216	51	56	0,23	62,88 ^{ns}	Satt536	48	63	2,03	15,45 ^{ns}
OPAP07a	58	60	0,03	85,39 ^{ns}	Satt226	53	52	0,01	92,22 ^{ns}	Satt350	56	56	0,00	100,0 ^{ns}
OPAP07b	59	58	0,01	92,63 ^{ns}	Satt232	53	52	0,01	92,22 ^{ns}	Satt286	64	46	2,95	8,61 ^{ns}
OPAR08	61	57	0,14	71,27 ^{ns}	Satt241	59	41	3,24	7,18 ^{ns}	Satt279	57	59	0,03	85,26 ^{ns}
OPAS04	66	52	1,66	19,74 ^{ns}	Satt276	53	55	0,04	84,7 ^{ns}	Satt454	47	65	2,89	8,89 ^{ns}
OPAS07	65	53	1,22	26,92 ^{ns}	Satt277	67	47	3,51	6,10 ^{ns}	Satt338	53	59	0,32	57,0 ^{ns}
OPAS08	50	68	2,75	9,75 ^{ns}	Satt281	56	54	0,04	84,87 ^{ns}	Satt130	51	60	0,73	39,29 ^{ns}

Primer	0	2	c ²	P%	Primer	0	2	c ²	P%	Primer	0	2	c ²	P%
Satt272	60	55	0,22	64,10 ^{ns}	ACTXCAG330	61	57	0,14	71,27 ^{ns}	AAGXCAC243	61	57	0,14	71,27 ^{ns}
Satt239	57	55	0,04	85,01 ^{ns}	ACTXCAG89	66	52	1,66	19,74 ^{ns}	AAGXCAC89	59	59	0,00	100,0 ^{ns}
Satt310	59	55	0,14	70,79 ^{ns}	ACTXCAG90	57	61	0,14	71,27 ^{ns}	AACXCTC93	57	61	0,14	71,27 ^{ns}
Satt210	46	69	4,60	3,19 *	AACXCTA81	36	82	17,9	0,002 *	AACXCTC87	47	71	4,88	2,71 *
Satt207	63	54	0,69	40,53 ^{ns}	AACXCTA108	51	67	2,17	14,07 ^{ns}	AACXCTG325	49	69	3,39	6,56 ^{ns}
Satt262	63	51	1,26	26,10 ^{ns}	AAGXCAG250	73	45	6,64	0,99 *	AAGXCAC113	70	48	4,10	4,284 *
Satt273	57	56	0,01	92,50 ^{ns}	ACTXCTA184	47	71	4,88	2,71 *	AAGXCAC126	54	64	0,85	35,72 ^{ns}
Satt348	66	52	1,66	19,74 ^{ns}	ACTXCTA132	49	69	3,39	6,56 ^{ns}	ACAXCTG242	74	44	7,63	0,575 *
Satt349	62	50	1,29	25,68 ^{ns}	ACTXCTA114	62	56	0,31	58,07 ^{ns}	ACAXCTG313	61	57	0,14	71,27 ^{ns}
Satt586	58	54	0,14	70,54 ^{ns}	ACTXCTA84	69	49	3,39	6,56 ^{ns}	ACAXCTG127	55	63	0,54	46,14 ^{ns}
Satt274	62	56	0,31	58,07 ^{ns}	ACTXCTA156	60	58	0,03	85,39 ^{ns}	ACAXCAC78	62	56	0,31	58,07 ^{ns}
Satt406	58	60	0,03	85,39 ^{ns}	ACTXCTA160	57	61	0,14	71,27 ^{ns}	ACAXCAC147	54	64	0,85	35,72 ^{ns}
Satt126	48	58	0,94	33,14 ^{ns}	AACXCTC130	67	51	2,17	14,07 ^{ns}	ACAXCAC344	75	43	8,68	0,3221 *
Satt437	68	50	2,75	9,75 ^{ns}	AAGXCAC189	51	67	2,17	14,07 ^{ns}	ACGXCAA78	58	60	0,03	85,39 ^{ns}
ACTXCAG190	52	66	1,66	19,74 ^{ns}	AAGXCAC239	53	65	1,22	26,92 ^{ns}	AGCXCAC147	70	48	4,10	4,28 *
ACTXCAG350	75	43	8,68	0,32 *	AAGXCAC243	61	57	0,14	71,27 ^{ns}					

3.2. Mapa de ligação genética

A primeira etapa do processo de mapeamento é a definição dos parâmetros de LOD e máxima frequência de recombinação (r) para o agrupamento dos marcadores. Foram adotados os valores de $LOD_{min} = 3,0$ e de $r_{max} = 30\%$, por apresentarem distribuição mais homogênea dos marcadores entre os grupos de ligação. Com os grupos definidos, foi determinada a ordem mais provável dos marcadores em cada grupo de ligação.

Foram obtidos 29 grupos de ligação pouco saturados, contendo 95 marcadores, e mais 79 marcas não-ligadas, como pode ser observado na Figura 3. Os 29 grupos de ligação apresentaram relação com os 20 grupos de ligação obtidos em uma integração de três mapas públicos de ligação gênica da soja feita por CREGAN *et al.* (1999). Isto significa que o conjunto de marcas pertencente a cada grupo de ligação obtido neste trabalho coincide com um conjunto de marcas presente em determinado grupo de ligação no mapa integrado, e poucas inversões entre os marcadores obtidos aqui e os do mapa integrado foram observadas. Foram mapeados ainda cinco novos marcadores: Satt449 e Satt526 (pertencentes ao grupo A1), Satt514 e Satt528 (grupo D2) e Satt594 (grupo G), que não estavam presentes em nenhum dos três mapas (USDA/Iowa State Univ., Univ. of Utah, Univ. of Nebraska) utilizados na

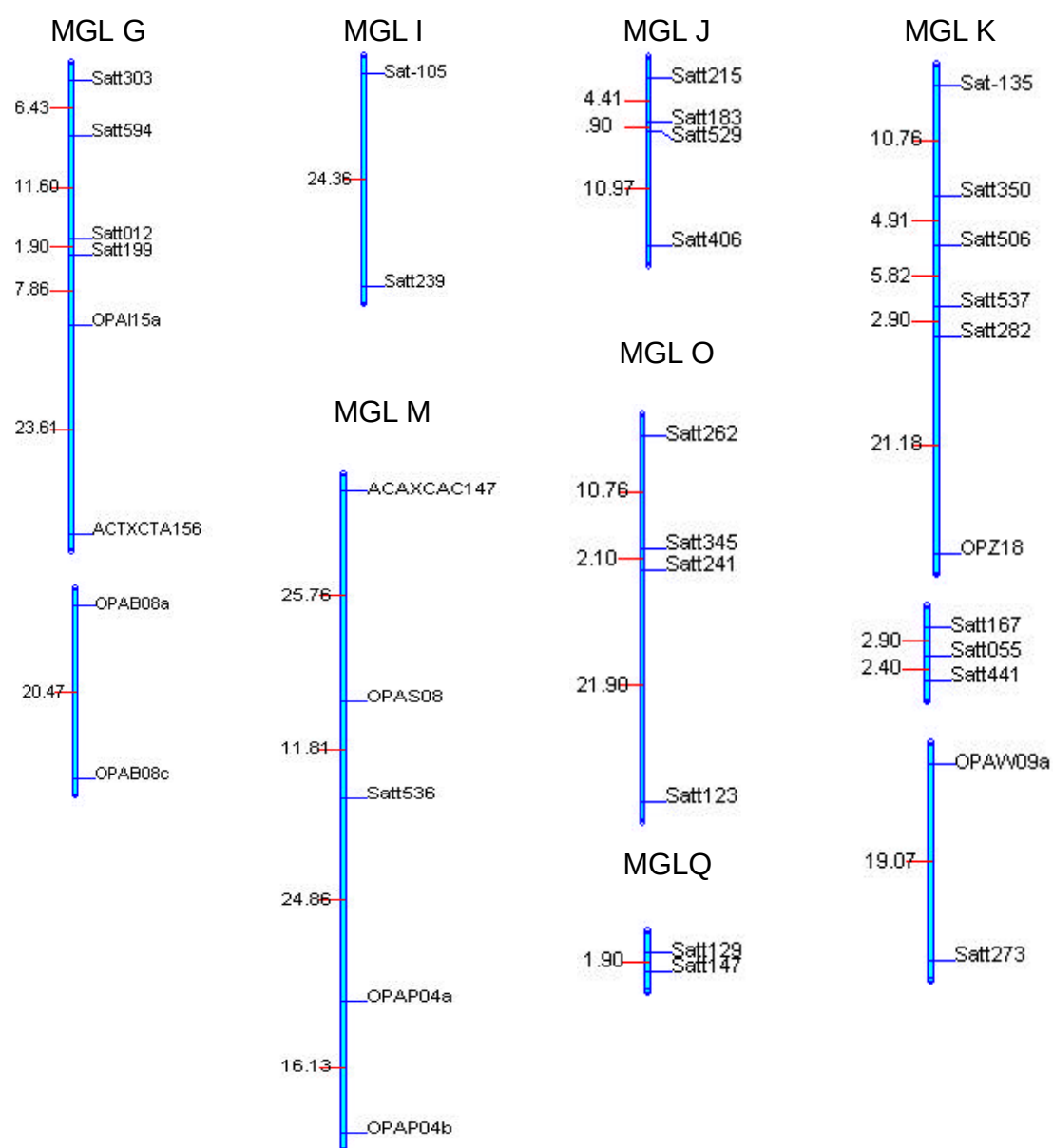
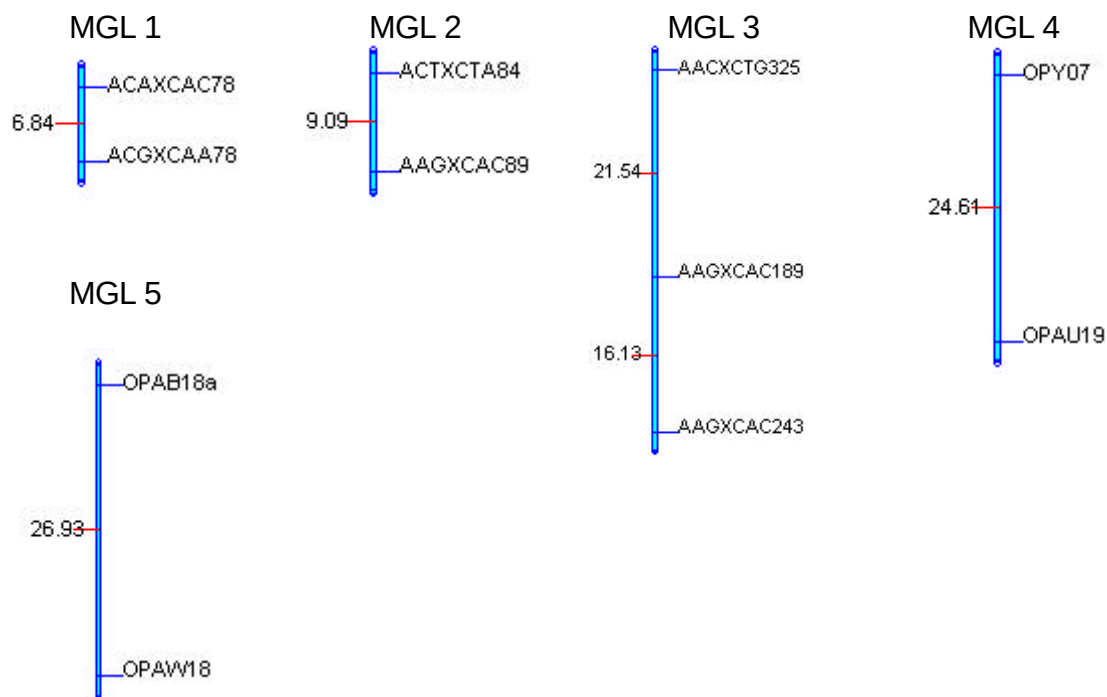


Figura 3 – Continuação...



Marcadores não ligados: OPN14, OPP12, OPW08, OPY13a, OPY13b, OPY14, OPY17, OPZ06a, OPZ06b, OPAA03, OPAA14, OPAA15, OPAB08b, OPAB13, OPAB18b, OPAB19, OPAB20, OPAC04, OPAI11a, OPAI11b, OPAI01, OPAN03, OPAN20, OPAO06, OPAP04c, OPAP07a, OPAP07b, OPAR08, OPAS04, OPAS07, OPAT19, OPAU02a, OPAU02b, OPAU04, OPAX12, OPAX20, OPAZ02, OPAZ07, OPBE04, OPBE13, OPBH02, Satt164, Satt181, Satt184, Satt216, Satt232, Satt369, Satt353, Satt377, Satt468, Satt475, Satt495, Satt527, Satt531, Satt538, Satt549, Satt242, Satt471, Satt518, Satt279, Satt338, Satt130, Satt272, Satt207, Satt349, Satt586, Satt274, Satt126, Satt437, ACTXCAG330, ACTXCAG89, AACXCTA108, ACTXCTA132, ACTXCTA114, ACTXCTA160, AACXCTC130, AACXCTC93, AAGXCAC126, ACAXCTG127,

Figura 3 – Continuação...

integração por CREGAN *et al.* (1999). Esses marcadores poderão ser importantes em programas de mapeamento futuros.

Essa é a primeira etapa de um processo que visa o mapeamento de QTLs que controlam as características analisadas. Devido a pouca variabilidade intra-específica da soja, os grupos de ligação obtidos no presente trabalho estão ainda pouco saturados. Em trabalhos futuros, em que serão utilizadas outras populações RILs e outros marcadores, os mapas poderão, então, ser integrados.

3.3. Análise de QTLs

3.3.1. Análise de marca simples

A capacidade de associar marcadores moleculares e QTLs depende, em grande parte, da magnitude dos efeitos dos QTLs, da estrutura e do tamanho da população de mapeamento e do nível de saturação do genoma com os marcadores moleculares.

Utilizando as análises de marca simples (Quadro 3), entre os marcadores e as características das RILs cultivadas em Viçosa, observa-se que 13 marcadores estão associados ao conteúdo de proteína total e explicam de 4,39% a 13,32% da variação do teor de proteína. Ao conteúdo da subunidade α' da proteína 7S, foram encontrados 9 marcadores associados, que explicam de 4,21% a 7,09% da variação desta característica. Dezesete marcadores mostraram associação com o conteúdo da subunidade α da proteína 7S e explicam de 4,07% a 13,16% da variação desta característica. Para o conteúdo da subunidade β foram encontrados 16 marcadores associados, e eles explicam de 4,78% a 21,80% da variação desta característica. Associados ao conteúdo das subunidades ácidas e básicas da proteína 11S, foram encontrados 17 e 8 marcadores, respectivamente, e eles explicam de 4,08% a 9,60% (subunidades ácidas) e de 5,49% a 11,10% (subunidades básicas) da variação destas características. Associados ao conteúdo da característica proteína total menos proteína de reserva ((PT-(11S+7S)) foram encontrados 7 marcadores que explicam de 4,11% a 10,77%

da variação desta característica. Dezenove marcadores mostraram associação com o conteúdo de 7S e eles explicam de 4,08% a 15,57% da variação desta característica. Dezesesseis marcadores apresentaram associação com o conteúdo de 11S e eles explicam de 4,17% a 9,69% da variação desta característica. Alguns destes marcadores estavam associados a mais de uma característica como pode ser observado no Quadro 4, indicando que há genes com efeitos pleiotrópicos no controle genético da expressão destas características.

De acordo com as análises de marca simples (Quadro 5), entre os marcadores e as características das RILs cultivadas em São Gotardo, foram observados 13 marcadores associados a conteúdo de proteína total que explicam de 4,31% a 12,11% da variação do teor de proteína. Para a subunidade α' da proteína 7S, foram encontrados 13 marcadores que explicam de 4,34% a 10,42% da variação desta característica. Cinco marcadores mostraram associação com a subunidade α da proteína 7S e explicam de 4,32% a 6,17% da variação desta característica. Para a subunidade β foram encontrados 12 marcadores associados, e eles explicam de 4,09% a 9,11% da variação desta característica. Para as subunidades ácidas e básicas da proteína 11S, foram encontrados 13 e 17 marcadores respectivamente e eles explicam de 4,21% a 13,36% (subunidades ácidas) e de 4,12% a 8,12% (subunidades básicas) da variação destas características. Para a característica proteína total menos proteína de reserva ((PT-(11S+7S)) foram encontrados 6 marcadores associados que explicam de 4,39% a 7,92% da variação desta característica. Seis marcadores mostraram associação com o conteúdo de 7S e eles explicam de 5,16% a 8,11% da variação desta característica. Dezesesseis marcadores apresentaram associação com o conteúdo de 11S e eles explicam de 4,15% a 13,67% da variação desta característica. Alguns destes marcadores também apresentaram associação com mais de uma característica como pode ser observado no QUADRO 6.

Quadro 3 – Análise de associação de marcas moleculares com características de proteínas avaliadas em RILs obtidas do cruzamento entre as variedades de soja Garimpo (0) e BARC-8 (2) no experimento de Viçosa.

Marca	Prob. F ^{1/}		Média (2)	Média (0) ^{2/}	R ² (%)	Grupo de ligação ^{3/}
Proteína Total						
OPAO06	0,0188	*	46,1257 a	44,8174 b	5,7961	NL
Sat-105	0,0029	**	46,2882 a	44,6387 b	9,2629	I
Satt129	0,0426	*	46,0135 a	44,8627 b	4,3944	Q
Satt276	0,0356	*	46,0308 a	44,801 b	5,0346	A ₁
Satt304	0,0415	*	46,0442 a	44,9487 b	4,5344	B ₂
Satt318	0,0314	*	46,0494 a	44,8462 b	4,8806	B ₂
Satt549	0,002	**	46,1252 a	44,3745 b	9,8934	NL
Satt242	0,0414	*	44,8508 a	46,0123 b	4,8599	NL
Satt279	0,0309	*	44,8555 a	46,07 b	4,9621	NL
Satt239	0,0004	**	46,6384 a	44,6174 b	13,3235	I
AAGXCAC239	0,0086	**	46,2133 a	44,7557 b	7,2007	E
AAGXCAC243	0,0189	*	46,1385 a	44,8321 b	5,7847	3
AAGXCAC126	0,0182	*	46,1562 a	44,8419 b	5,8547	NL
a'						
OPAR08	0,0251	*	1,5666 a	1,7549 b	5,2806	NL
OPAT19	0,046	*	1,5894 a	1,7581 b	4,2132	NL
OPBE13	0,0343	*	1,7538 a	1,5759 b	4,7257	NL
Satt144	0,0483	*	1,7334 a	1,5653 b	4,4583	F
Satt129	0,0301	*	1,7502 a	1,5651 b	5,0131	Q
Satt518	0,0215	*	1,7665 a	1,5723 b	5,6146	NL
Satt279	0,0095	**	1,7756 a	1,5566 b	7,0945	NL
Satt454	0,0394	*	1,7467 a	1,5679 b	4,6299	A ₁
Satt338	0,0217	*	1,5686 a	1,7639 b	5,7164	NL
a						
OPAA15	0,0499	*	2,1827 a	2,4314 b	4,0697	NL
OPAB20	0,0245	*	2,4954 a	2,2027 b	5,3222	NL
OPAI11b	0,0055	**	2,1486 a	2,4969 b	7,9829	NL
OPAI15a	0,0003	**	2,0732 a	2,521 b	13,1592	G
OPAP04c	0,0042	**	2,1286 a	2,4872 b	8,4775	NL
OPAU19	0,0267	*	2,1661 a	2,4464 b	5,1695	5
OPAX12	0,0287	*	2,4533 a	2,1768 b	5,0446	NL
OPBE04	0,0436	*	2,1731 a	2,4297 b	4,3064	NL
Sat-112	0,0256	*	2,2068 a	2,497 b	5,7922	E
Satt012	0,0424	*	2,1903 a	2,4494 b	4,5989	G
Satt199	0,004	**	2,1283 a	2,505 b	9,047	G
Satt277	0,0029	**	2,5391 a	2,1641 b	9,4176	C ₂
Satt281	0,0404	*	2,1438 a	2,4189 b	4,7947	D1b
Satt475	0,0319	*	2,1866 a	2,4643 b	4,9609	NL
Satt528	0,0223	*	2,4484 a	2,1561 b	5,6678	D ₂
Satt594	0,0161	*	2,1469 a	2,4514 b	6,0645	G
Satt286	0,0079	**	2,5387 a	2,1889 b	7,8323	C ₂

^{1/}** significativo a 1% e * a 5%.

^{2/} Médias seguidas de letras diferentes, na linha, diferem entre si pelo teste t a 5% de probabilidade.

^{3/} NL – Não ligado.

Quadro 3 – Continuação...

Marcas	Prob. F		Média (2)	Media (0)	R ² (%)	Grupo de ligação
b						
OPY13b	0,0331	*	1,9581 a	1,7695 b	4,7922	NL
OPAI15a	0,004	**	1,7252 a	1,9779 b	8,5616	G
OPAW09a	0,0	**	2,0515 a	1,6632 b	20,7174	K
Satt055	0,0001	**	1,7242 a	2,0632 b	15,0532	K
Satt167	0,0004	**	1,7331 a	2,0569 b	13,3664	K
Satt012	0,0035	**	1,705 a	1,9708 b	9,2888	G
Satt199	0,0014	**	1,6924 a	1,9738 b	10,9857	G
Satt300	0,0187	*	1,9716 a	1,7627 b	5,9942	A ₁
Satt384	0,0057	**	1,9705 a	1,7196 b	8,452	E
Satt441	0,0008	**	1,728 a	2,0274 b	11,7206	K
Satt475	0,0019	**	1,7435 a	2,0209 b	10,1042	NL
Satt527	0,0353	*	1,9533 a	1,759 b	5,0511	NL
Satt594	0,0042	**	1,723 a	1,9748 b	8,4716	G
Satt273	0,0	**	2,0583 a	1,6549 b	21,8059	K
Satt349	0,0395	*	1,9554 a	1,7628 b	4,7812	NL
AAGXCAC239	0,0074	**	1,9772 a	1,742 b	7,4547	E
Ácidas						
OPP12	0,0493	*	9,1691 a	9,9651 b	4,0914	NL
OPW08	0,0081	**	9,0481 a	10,1113 b	7,2939	NL
OPZ18	0,048	*	9,9508 a	9,1497 b	4,1402	D1b
OPAT19	0,0069	**	10,0553 a	8,9675 b	7,574	NL
Satt144	0,0329	*	9,1666 a	10,0901 b	5,1848	F
Satt186	0,0408	*	9,1106 a	9,9488 b	4,72	D ₂
Sat-105	0,0337	*	9,9669 a	9,0969 b	4,807	I
Satt146	0,0155	*	9,996 a	8,9272 b	6,7762	F
Satt181	0,034	*	9,1157 a	10,0061 b	5,1235	NL
Satt232	0,0462	*	8,9433 a	9,8172 b	4,7601	NL
Satt304	0,0413	*	9,9552 a	9,1229 b	4,5445	B ₂
Satt325	0,0242	*	9,1909 a	10,0656 b	5,4558	F
Satt520	0,0172	*	9,0511 a	10,0215 b	6,012	D1b
Satt549	0,0024	**	10,0654 a	8,8122 b	9,6053	NL
ACTXCTA132	0,0083	**	9,0604 a	10,1212 b	7,2482	NL
ACTXCTA160	0,0248	*	9,1147 a	10,0207 b	5,3007	NL
AAGXCAC243	0,0247	*	10,0114 a	9,1048 b	5,3081	3
Básicas						
Satt186	0,0142	*	11,6563 a	12,5522 b	6,7212	D ₂
Satt226	0,0037	**	11,5592 a	12,6779 b	9,605	D ₂
Satt514	0,0013	**	11,681 a	12,7926 b	11,1025	D ₂
Satt528	0,0212	*	11,6979 a	12,5006 b	5,7635	D ₂
Satt082	0,0165	*	11,7657 a	12,5875 b	6,3628	D ₂
Satt272	0,0131	*	11,6528 a	12,5615 b	6,6399	NL
Satt239	0,0263	*	12,5306 a	11,7057 b	5,4864	I
Satt349	0,0211	*	12,5485 a	11,6964 b	5,9596	NL
PT-(11S+7S)						
OPZ06a	0,0489	*	18,4591 a	17,4381 b	4,1072	NL

Quadro 3 - Continuação...

PT-(11S+7S) (continuação...)						
Marca	Prob. F		Média (2)	Média (0)	R ² (%)	Grupo de ligação
OPZ18	0,0263	*	17,3868 a	18,536 b	5,1993	D1b
Satt146	0,0351	*	17,6065 a	18,7486 b	5,1771	F
Satt422	0,0012	**	18,793 a	17,1685 b	10,7658	D1B
Satt520	0,0159	*	18,5903 a	17,333 b	6,1597	D1B
Satt527	0,0468	*	18,4901 a	17,4216 b	4,5191	NI
AAGXCAC89	0,0267	*	18,547 a	17,3998 b	5,1722	2
7S						
OPAB18a	0,0495	*	6,0214a	5,611b	4,0837	8
OPAI15a	0,0001	**	5,412a	6,209b	15,5692	A1
OPAP08	0,0192	*	5,585a	6,069b	5,7563	NL
OPAW09a	0,0235	*	6,085a	5,612b	5,4512	K
OPAX12	0,0291	*	6,068a	5,616b	5,0157	NL
OPBD16	0,0391	*	6,042a	5,614b	4,4988	D1b
OPBE04	0,0478	*	5,614a	6,027b	4,1446	NL
Satt012	0,0082	**	5,546a	6,098b	7,6862	G
Satt129	0,0207	*	6,062a	5,578b	5,6825	Q
Satt199	0,001	**	5,481a	6,183b	11,6732	G
Satt277	0,0024	**	6,230a	5,599b	9,8111	C2
Satt300	0,0263	*	6,079a	5,608b	5,3648	A1
Satt475	0,033	**	5,566a	6,183b	9,0969	NL
Satt594	0,005	**	5,454a	6,159b	12,1499	G
Satt286	0,0091	**	6,190a	5,616b	7,5724	C2
Satt279	0,0293	*	6,066a	5,611b	5,0578	NL
Satt273	0,0331	*	6,063a	5,610b	5,0022	K
AACXCTC130	0,0358	*	6,124a	5,674b	4,6512	NL
AAGXCAC239	0,009	**	6,106a	5,568b	7,1036	E
11S						
OPZ18	0,0215	*	22,372a	21,002b	5,5551	K
OPAB18a	0,0471	*	22,236a	21,042b	4,1704	8
OPAT19	0,005	**	22,462a	20,797b	8,147	NL
OPAU02b	0,0136	*	22,419a	20,952b	6,3741	NL
Satt186	0,003	**	20,766a	22,501b	9,6904	D2
Satt105	0,0094	**	22,443a	20,886b	7,0996	I
Satt167	0,0293	*	21,161a	22,532b	5,3421	K
Satt146	0,0352	*	22,208a	20,847b	5,1702	F
Satt226	0,0309	*	20,928a	22,311b	5,4251	D2
Satt514	0,0286	*	21,148a	22,437b	5,3256	D2
Satt528	0,0281	*	20,995a	22,260b	5,2452	D2
Satt549	0,0126	*	22,337a	20,815b	6,5735	NL
Satt239	0,0147	*	22,472a	20,947b	6,5722	I
Satt310	0,0447	*	20,969a	22,202b	4,451	D2
ACTXCTA132	0,0045	**	20,912a	22,593b	8,3573	L
AAGXCAC243	0,0308	*	22,3462a	21,057b	4,9195	3

Quadro 4 – Marcadores que explicam a variação de mais de uma característica para as RILs cultivadas em Viçosa

Marca	Característica	Marca	Característica
Satt129	1, 2 e 8	Satt012	3, 4 e 8
Satt279	1, 2 e 8	Satt199	3, 4 e 8
AAGXCAC239	1, 4 e 8	Satt475	3, 4 e 8
Sat-105	1, 5 e 9	Satt594	3, 4 e 8
Satt304	1 e 5	Satt528	3, 6 e 9
Satt549	1, 5 e 9	Satt527	4 e 7
AAGXCAC243	1, 5 e 9	Satt349	4 e 6
Satt239	1, 6 e 9	Satt186	5 e 6
OPAT19	2, 5 e 9	OPZ18	5, 7 e 9
Satt144	2 e 5	Satt146	5, 7 e 9
OPAI15a	3, 4 e 8	Satt520	5 e 7

Característica: 1 – Proteína Total, 2 – α' , 3– α , 4 – β , 5 – Ácidas, 6 – Básicas, 7 - PT-(11S+7S), 8 - 7S, 9 - 11S.

Quadro 5 - Análise de associação de marcas moleculares com características de proteínas avaliadas em RILs obtidas do cruzamento entre as variedades de soja Garimpo (0) e BARC-8 (2) no experimento de São Gotardo

Marca	Prob. F	Média (2)	Média (0)	R ² (%)	Grupo de ligação
Proteína Total					
OPP12	0,0117 *	46,1496 a	47,4493 b	6,643	NL
OPY14	0,0418 *	46,3229 a	47,3855 b	4,3814	NL
OPAS08	0,0433 *	46,3458 a	47,407 b	4,3188	M
Satt055	0,0031 **	46,2134 a	47,7683 b	9,3893	K
Sat-105	0,0067 **	47,4629 a	46,0567 b	7,7225	I
Satt167	0,0012 **	46,2837 a	47,9556 b	11,3594	K
Satt181	0,0479 *	46,082 a	47,0795 b	4,4749	NL
Satt441	0,0138 *	46,1817 a	47,4938 b	6,5459	K
Satt549	0,0006 **	47,4665 a	45,6569 b	12,1113	NL
Satt536	0,0492 *	46,264 a	47,3336 b	4,3722	M
Satt239	0,0334 *	47,4681 a	46,3285 b	5,0407	I
Satt437	0,0051 **	47,601 a	46,1521 b	8,1461	NL
AAGXCAC243	0,0024 **	47,5605 a	46,0084 b	9,4738	3
a'					
OPY13a	0,0154 *	1,9083a	2,0903b	6,1536	NL
OPAI15b	0,025 *	2,0703a	1,8995b	6,1775	A ₁
OPY17	0,0151 *	1,8981a	2,0813b	5,2875	NL
Satt277	0,012 *	2,1121a	1,9212b	6,8078	C ₂
Satt303	0,0461 *	1,9183a	2,0793b	4,5986	G
Satt300	0,0076 **	2,0908a	1,8887b	7,6523	A ₁
Satt333	0,0014 **	1,8884a	2,1259b	10,4272	A ₂
Satt353	0,0036 **	2,106a	1,8841b	9,0304	NL
Satt449	0,0242 *	2,0846a	1,9106b	5,5155	A ₁
Satt495	0,0027 **	1,8744a	2,0998b	9,446	NL

Quadro 5 – Continuação...

Marca	Prob. F		Média (2)	Média (0)	R ² (%)	Grupo de ligação
a' (continuação...)						
Satt594	0,0021	**	1,8726a	2,1024b	9,7217	G
ACTXCAG330	0,0419	*	1,9224a	2,0759b	4,3778	NL
ACTXCAG90	0,0428	*	1,9259a	2,0789b	4,3369	A ₂
a						
OPAB18a	0,0381	*	3,1577 a	3,3916 b	4,543	8
Satt353	0,0468	*	3,3687 a	3,1396 b	4,3189	NL
Satt495	0,0163	*	3,1203 a	3,389 b	6,1746	NL
Satt538	0,0274	*	3,1364 a	3,3832 b	5,1248	NL
ACTXCTA132	,03	*	3,3764 a	3,1331 b	4,9645	NL
b						
OPAB08a	0,0198	*	1,9555 a	2,1316 b	5,6966	4
OPBE13	0,0397	*	1,9618 a	2,1177 b	4,4689	NL
Sat-105	0,0279	*	1,9589 a	2,127 b	5,1466	I
Satt384	0,0045	**	2,1386 a	1,9155 b	8,8885	E
Satt422	0,0081	**	2,1361 a	1,9349 b	7,3771	D1b
Satt475	0,0057	**	1,9484 a	2,1612 b	8,1084	NL
Satt520	0,0031	**	2,1515 a	1,9287 b	9,1075	D1b
Satt549	0,0349	*	1,9764 a	2,1424 b	4,7478	NL
Satt273	0,0054	**	2,1422 a	1,9318 b	8,3826	K
Satt274	0,013	*	2,1336 a	1,9462 b	6,4563	NL
ACTXCTA114	0,0493	*	2,1209 a	1,9711 b	4,0912	NL
AACXCTC93	0,0168	*	2,1263 a	1,9458 b	5,9923	NL
Ácidas						
OPAB18b	0,012	*	10,8907 a	10,0982 b	6,6641	NL
OPAI11b	0,0347	*	10,1368 a	10,8161 b	4,7044	NL
OPAP04a	0,0209	*	10,8249 a	10,0844 b	5,605	M
Sat-105	0,0324	*	10,7902 a	10,0956 b	4,8797	I
Satt167	0,0383	*	10,2592 a	10,9528 b	4,8414	K
Satt276	0,0411	*	10,8347 a	10,1384 b	4,7625	A ₁
Satt304	0,0066	**	10,8627 a	10,0225 b	7,9111	B ₂
Satt549	0,0003	**	10,872 a	9,7117 b	13,3629	NL
Satt350	0,0482	*	10,8159 a	10,1577 b	4,3125	D1b
ACTXCAG190	0,046	*	10,7323 a	10,0822 b	4,2123	A ₂
AAGXCAC189	0,0112	*	10,8041 a	9,9834 b	6,7129	3
AAGXCAC243	0,0021	**	10,9419 a	9,9649 b	9,7547	3
AACXCTG325	0,042	*	10,7084 a	10,0303 b	4,3736	3
Básicas						
OPP12	0,034	*	11,1279 a	11,8163 b	4,7452	NL
OPY14	0,0306	*	11,1562 a	11,8626 b	4,9286	NL
OPAI11a	0,0486	*	11,1513 a	11,7925 b	4,116	NL
OPAW09a	0,007	**	11,9003 a	11,0255 b	7,6424	K
OPAW18	0,0114	*	11,064 a	11,8816 b	6,6935	8
Satt055	0,0084	**	11,056 a	11,9414 b	7,5391	K
Satt167	0,0096	**	11,0834 a	11,9723 b	7,4646	K

1/** significativo a 1% e * a 5%.

2/ Médias seguidas de letras diferentes, na linha, diferem entre si pelo teste t a 5% de probabilidade.

3/ NL – Não ligado.

Quadro 5 – Continuação...

Marca	Prob. F		Média (2)	Média (0)	R ² (%)	Grupo de ligação
Básicas (continuação...)						
Satt184	0,043	*	11,7902 a	11,0859 b	4,7286	NL
Satt129	0,0258	*	11,119 a	11,8445 b	5,2841	Q
Satt303	0,0439	*	11,8136 a	11,1066 b	4,6919	G
Satt441	0,0263	*	11,1329 a	11,8833 b	5,3629	K
Satt462	0,0482	*	11,7971 a	11,1347 b	4,2663	L
Satt468	0,0457	*	11,1189 a	11,7702 b	4,2246	NL
Satt239	0,0309	*	11,8333 a	11,1138 b	5,1832	I
Satt126	0,0086	**	11,9674 a	11,0629 b	8,1249	NL
Satt437	0,0095	**	11,9388 a	11,0958 b	7,0215	NL
ACTXCAG330	0,0205	*	11,0973 a	11,8476 b	5,6365	NL
PT-(11S+7S)						
OPAA 03	0,0125	*	17,9431 a	17,1667 b	6,5273	NL
OPAI11a	0,0209	*	17,9229 a	17,2039 b	5,6021	NL
Sat-105	0,006	**	17,985 a	17,1396 b	7,9227	I
Satt333	0,0111	*	17,9334 a	17,1427 b	6,7409	A ₂
Satt549	0,0427	*	17,7869 a	17,1396 b	4,3905	NL
Satt274	0,0392	*	17,2419 a	17,8858 b	4,4935	NL
7S						
Satt333	0,0235	*	7,0891a	7,541b	5,3989	A2
Satt353	0,0059	**	7,5717a	7,0153b	8,1058	NL
Satt475	0,0202	*	7,0999a	7,5756b	5,7846	NL
Satt538	0,0259	*	7,0743a	7,5178b	5,2249	NL
Satt594	0,02	*	7,0441a	7,5086b	5,683	G
Actxcta132	0,0269	*	7,5074a	7,0662b	5,1591	NL
11S						
OPP12	0,0097	**	21,3779a	22,4879b	6,9737	NL
OPAB18a	0,0332	*	22,4352a	21,5118b	4,837	NL
OPAI11b	0,0173	*	21,4411a	22,467b	5,9412	NL
OPAW09a	0,0327	*	22,3943a	21,4629b	4,8637	K
Satt055	0,0006	**	21,3256a	22,8258b	12,433	K
Sat-105	0,0487	*	22,3361a	21,4736b	4,156	I
Satt167	0,0004	**	22,3425a	22,9251b	13,6373	K
Satt199	0,0385	*	22,3229a	21,429b	4,7772	G
Satt303	0,0272	*	22,4178a	21,4382b	5,6103	G
Satt422	0,0285	*	21,4792a	22,4312b	5,1101	D1b
Satt441	0,0094	**	21,4086a	22,5695b	7,2583	K
Satt549	0,0064	**	22,4017a	21,1899b	7,7939	NL
Satt594	0,0236	*	22,4635a	21,4835b	5,3889	G
Satt239	0,0236	*	22,467a	21,4557b	5,6884	I
Satt437	0,011	*	22,54a	21,4413b	6,7425	NL
AAGXCAC243	0,0039	**	22,5379a	21,3032b	8,6297	3

Quadro 6 – Marcadores que explicam a variação de mais de uma característica para as RILs cultivadas em São Gotardo

Marca	Característica	Marca	Característica
OPP12	1, 6 e 9	Satt303	2, 6 e 9
OPY14	1 e 6	Satt333	2, 7 e 8
Satt055	1, 6 e 9	Satt353	2, 3 e 8
Sat-105	1, 4, 5, 7 e 9	Satt495	2 e 3
Satt167	1, 6 e 9	ACTXCAG330	2 e 6
Satt441	1, 6 e 9	Satt274	4 e 7
Satt549	1, 4, 5, 7 e 9	OPAI11a	6 e 7
Satt239	1, 6 e 9		
Satt437	1, 6 e 9		
AAGXCAC243	1, 6 e 9		

Característica: 1 – Proteína Total, 2 – α' , 3 – α , 4 – β , 5 – Ácidas, 6 – Básicas, 7 – PT-(11S+7S), 8 – 7S, 9 – 11S.

Os marcadores associados às características foram diferentes de um local para outro, indicando interação entre o genótipo e o ambiente, como já havia sido demonstrado pela análise de variância. Em estudos utilizando populações F_2 de tomate (*Lycopersicon esculentum*), QTLs para algumas características mostraram-se variáveis em função de fatores ambientais e “backgrounds” genéticos (TANKSLEY e HEWITT, 1988; PATERSON *et al.*, 1991). Entretanto, STUBER *et al.* (1992) não encontraram forte influência ambiental em QTLs associados à produção de grãos em milho (*Zea mays* L.).

3.3.2. Mapeamento por intervalo composto

Os modelos lineares simples, como ANOVA e regressão linear, não necessitam das informações dos mapas de ligação, e a distribuição dos valores fenotípicos são analisados para cada marcador separadamente. Portanto, por utilizarem apenas diferenças entre as médias dos marcadores, não é possível estimar a magnitude do efeito do QTL, nem a sua posição no genoma (LYNCH e WALSH, 1998). Desse modo, foi utilizado o método de mapeamento por intervalo composto (ZENG, 1993; 1994; JANSEN, 1993), com o intuito de aumentar o poder de detecção de QTLs e de melhorar a precisão das estimativas do efeito e da posição dos mesmos. Esse método é baseado na segregação de marcadores adjacentes ao QTL, utilizando marcadores em outros intervalos para isolar o efeito de outros possíveis QTLs que estiverem

próximos no mesmo grupo de ligação. Pelo grande número de permutações necessárias para obter o valor do ponto de corte e pelo fato de ter que ser repetido a cada experimento ou modelo testado, esta estratégia é computacionalmente muito lenta. Como a razão de verossimilhança possui relação com o qui-quadrado, o ponto de corte foi adotado conforme a significância do qui-quadrado, em nível de 10% do mapa completo, que resulta em 0,003627% para cada cromossomo. Os valores do ponto de corte foram similares aos obtidos pelo teste de permutação em um teste preliminar utilizando somente alguns grupos de ligação.

Pela análise de mapeamento por intervalo composto, feita por meio do programa GQMOL, que permite escolher os cofatores para cada característica, foram encontrados 16 QTLs associados com as características avaliadas. Os cofatores foram escolhidos com base na análise de regressão stepwise, e ainda foram adicionados cofatores que não foram confirmados nesta análise de regressão, mas que estavam associados à característica pela análise de marca simples.

No experimento de Viçosa foram encontrados QTLs associados às características: conteúdo de proteína total (GL I), conteúdo da subunidade α (GL G e GL C2), conteúdo da subunidade β (GL G e dois no grupo K), conteúdo de PT-(11S+7S) (GL D1b) e conteúdo de 7S (GLG e GLE) como pode ser observado nas Figuras 4, 5, 6, 7 e 8, respectivamente.

Para análise do QTL associado à proteína total (GL I, que explica 14,74% da característica) foram escolhidos os cofatores Satt549 (NL), Satt279 (NL), ambos foram confirmados na regressão stepwise. Para análise do QTL associado à subunidade α (GL G, que explica 12,41% da característica) os cofatores foram Satt277 (GL C2), Satt286 (GL C2), eles pertencem ao grupo C2 que apresentou um QTL associado a esta característica. Para a análise do QTL presente no grupo C2 e que explica 10,13% da variação da característica α , não foram utilizados cofatores. Os dois QTLs associados ao conteúdo da subunidade α explicam em conjunto 28,83% da variação desta característica. Os cofatores escolhidos para o grupo G, que apresentou um QTL associado à subunidade β , que explica 10,58% desta característica, foram OPA15a (GL G), e Satt273(GL K). O primeiro está presente no grupo G e explica 8,53% da

característica, o outro está presente em um dos fragmentos do grupo K. A subunidade β apresentou ainda dois QTLs cada um em um fragmento do grupo K. Para a análise de QTL, que explica 9,84% da característica β , presente no grupo K contendo o marcador Satt055, foram escolhidos os cofatores OPAI15a (GL G), que foi confirmado na regressão stepwise, e o Satt273 (GL K), que está presente no outro fragmento do grupo K e explica 21,80% da característica. Para a análise do QTL, que explica 25,06% da característica, e está presente no outro fragmento do grupo K foi escolhido como cofator a marca RAPD OPAI15a (GL G). Estes três QTLs associados à característica β explicam em conjunto 32,62% desta característica. O QTL associado com as outras proteínas presentes no extrato (PT-(11S+7S)) e que explica 13,71% da característica, foi analisado com os cofatores OPZ06a (NL), AAGXCAC89 (GL 2), OPZ18 (GL D1b). Estes cofatores estão associados à característica pela análise de marca simples. O marcador Satt422, que pertence ao grupo D1b, não foi escolhido como cofator porque se encontra muito próximo ao QTL. Para análise do QTL presente no grupo de ligação G associado ao conteúdo de 7S e que explica 15,64% da variação desta característica, foi escolhido como cofator a marca OPAI15a, que está presente neste grupo de ligação e explica 15,57% da variação desta característica. Esta marca foi confirmada na regressão stepwise. Para a análise do QTL associado ao conteúdo de 7S, presente no grupo E (que explica 8,86% da variação desta característica) foram escolhidos como cofatores as marcas: Satt286 que está presente no GLC2 e explica 7,57% da variação desta característica, pela análise de marca simples; Satt594 e ACTXCTA156 presentes no GL G (que contém o outro QTL). A marca Satt594 explica 12,15% da variação desta característica. Os dois QTLs associados ao conteúdo de 7S explicam em conjunto 21,86% da variação desta característica.

No experimento de São Gotardo foram encontrados QTLs associados às características: conteúdo de proteína total (GL 3 e I), conteúdo da subunidade α' (GL A2), conteúdo das subunidades ácidas (GL 3), conteúdo de PT-(11S+7S) (GL I) e teor de 11S (dois QTLs em diferentes fragmentos do GL K), como mostram as Figuras 9, 10, 11, 12 e 13, respectivamente.

Para a análise de QTL do grupo 3 associado ao conteúdo de proteína total, que explica 9,51%, foi utilizado como cofator a marca Satt239, que está presente no GL I (grupo que apresentou o outro QTL associado a esta característica) e explica 5,04% da variação desta característica. Para a análise do QTL associado ao conteúdo de proteína total presente no GL I (que explica 10,83% da variação desta característica) não foram utilizados cofatores. Estes dois QTLs associados à proteína total explicam em conjunto 11,42% desta característica. Para a análise do QTL presente no GL A2, associado ao conteúdo da subunidade α' , que explica 10,64%, foram escolhidos os cofatores Satt300 (GL A1), que foi confirmado pela regressão stepwise e Satt594 (GL G), que explica 9,72% da característica pela análise de marca simples. Os cofatores escolhidos para a análise do QTL presente no grupo 3, associado com as subunidades ácidas foram AACXCTG243 (GL 3), que explica 9,75% da variação desta característica e ACTXCAG190 (GL A2) confirmado pela regressão stepwise. Este QTL explica 12,06% da variação desta característica. Para a análise do QTL presente no grupo I associado com as outras proteínas presentes no extrato, e que explica 11,34%, foram utilizados os cofatores OPAA03 (NL) e OPAI11a (NL) estas marcas estão associadas à característica pela análise de marca simples explicando 6,52% e 5,60% respectivamente. Para a análise dos QTLs associados ao conteúdo de 11S presentes nos dois fragmentos do GL K e que explicam 14% (fragmento que contém a marca Satt441) e 10,89% (fragmento que contém a marca OPAW09a) da variação desta característica foram escolhidos como cofatores as marcas Satt422 (GL D1b) e Satt549 (NL), que explicam, respectivamente, 5,11% e 7,79% da variação desta característica. Estes duas marcas foram confirmadas na regressão stepwise. Os dois QTLs associados ao conteúdo de 11S explicam em conjunto 20,88% da variação desta característica.

O grupo de ligação I apresentou QTLs para conteúdo de proteína total em Viçosa e em São Gotardo. Este fato sugere que os locos associados a esse grupo de ligação são conservados e, possivelmente, tenham função importante no desenvolvimento normal da planta. Os dois fragmentos do grupo K que apresentaram QTLs associados ao conteúdo da subunidade β de 7S no experimento de Viçosa, apresentaram QTLs associados ao conteúdo de 11S em São Gotardo. Este fato não era esperado, uma vez que estas duas

características (conteúdo de 7S e 11S) apresentaram correlação genotípica próximas de zero nos dois ambientes. No experimento de Viçosa também foram encontrados um QTL associado à proteína 7S no grupo de ligação E, e este grupo não apresentou nenhum outro QTL associado às subunidades constituintes desta proteína, o que sugere que podem existir outros genes associados à expressão de 7S além dos genes que expressam as suas subunidades constituintes.

CHUNG *et al.* (2003), mapearam um QTL para conteúdo de proteína, óleo e produtividade em sementes de soja, em uma população de 76 RILs na geração F₅, obtida ao cruzamento entre um acesso com alto teor de proteína (PI 437088A) e uma cultivar com alta produtividade (Asgrow A3733), que foram avaliadas em seis tratamentos de irrigação em dois anos. O QTL foi mapeado próximo à marca RAPD OPAW13a em um pequeno intervalo flanqueado pelos microssatélites Satt496 e Satt239 no grupo de ligação I.

BRUMMER *et al.* (1997), ao mapearem QTL para conteúdo de óleo e proteína em sementes de soja em populações diferentes, utilizando RFLP, identificaram dois tipos de QTLs: 1) QTLs estáveis a fatores ambientais (no mínimo um QTL para proteína em cada população, totalizando nove QTLs), detectados em dois ou três anos e na média dos três anos em que foram conduzidos os estudos; e 2) treze QTLs para proteínas sensíveis a fatores ambientais, associados aos grupos de ligação A₁, A₂, B₂, C₂, E, G, I, J e L.

Os resultados obtidos, tanto pelas análises de variância e regressão quanto pelo mapeamento por intervalo composto, indicaram que as características avaliadas, com exceção de proteína total são controladas por QTLs ambiente-específicos (expressos apenas em determinado ambiente).

O QTL associado ao grupo de ligação I pode ser classificado como um QTL estável a fatores ambientais neste trabalho, o que diverge do trabalho de BRUMMER *et al.* (1997), uma vez que eles encontraram que este grupo de ligação está associado a QTLs sensíveis a fatores ambientais. Esse resultado é muito útil aos programas de melhoramento, pois as variedades de soja com alto teor de proteína devem possuir, preferencialmente, aqueles QTLs estáveis nos vários ambientes em que são cultivadas. Os marcadores identificados podem ser testados na seleção assistida, em programas de melhoramento que visem aumentar o teor de proteína nos grãos de soja. De acordo com os

resultados obtidos neste trabalho, aqueles marcadores que flanqueiam o QTL e os que apresentaram associação na análise de marca simples com a característica avaliada, seriam os marcadores com potencial para serem empregados na seleção assistida por marcadores (SAM). Estes marcadores podem ser utilizados preliminarmente à seleção fenotípica em programa de melhoramento que visam o desenvolvimento de variedades com alto teor protéico. A aplicação mais direta destes marcadores seria em programas de retrocruzamento, em que alelos favoráveis seriam introgridos em uma variedade comercial a partir do progenitor de alto teor protéico (BARC-8 ou uma RIL de alto teor protéico que apresente as marcas selecionadas), e os alelos desejáveis seriam monitorados em todas as gerações de retrocruzamento. Uma outra aplicação seria no desenvolvimento de novas variedades com alto teor protéico pelo método genealógico modificado, em que a seleção de alelos favoráveis seria feita nas etapas iniciais do programa. Estes marcadores poderiam, ainda, ser empregados no desenvolvimento de novas variedades utilizando seleção recorrente, onde alelos favoráveis de dois progenitores doadores de alto teor protéico poderiam ser reunidos em uma mesma progênie, e esta nova variedade passaria a ter um teor de proteína mais alto do que qualquer um dos pais.

Os resultados apresentados se referem a um ano e a dois locais. Avaliações em mais locais e mais anos devem ser adicionadas a esses dados, para confirmar os locos contendo os QTLs estáveis, além de identificar outros. Além disso, a saturação do mapa com a inclusão de novos marcadores será importante para melhorar a resolução do mapeamento de QTLs.

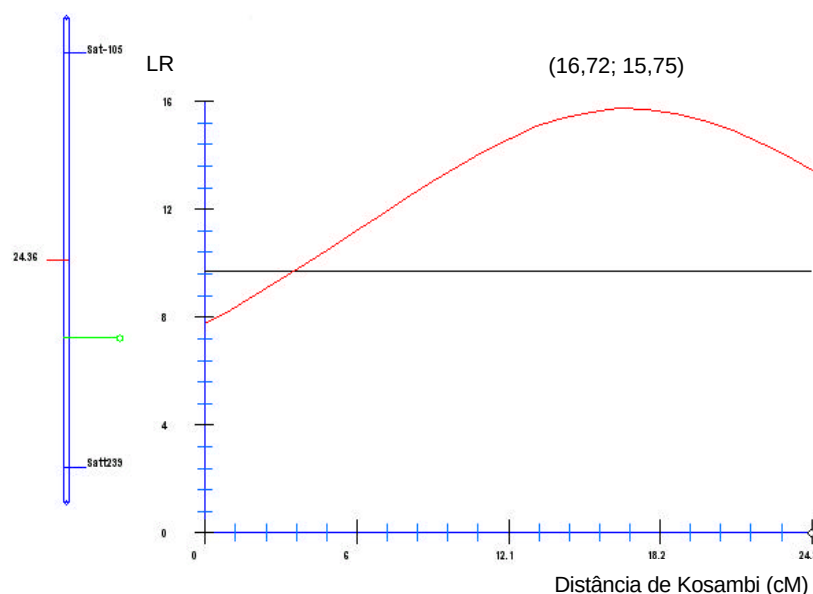


Figura 4 – QTL associado ao conteúdo de proteína total, no experimento de Viçosa. Grupo de ligação I representado à esquerda, e a curva da razão de verossimilhança representada no gráfico à direita. O pico da curva representa um valor de LR de 16,72 (eixo Y) na posição 15,75 cM (eixo X) representado no mapa entre os marcadores Satt239 e Sat-105. O valor de LR para o ponto de corte foi de 11,42.

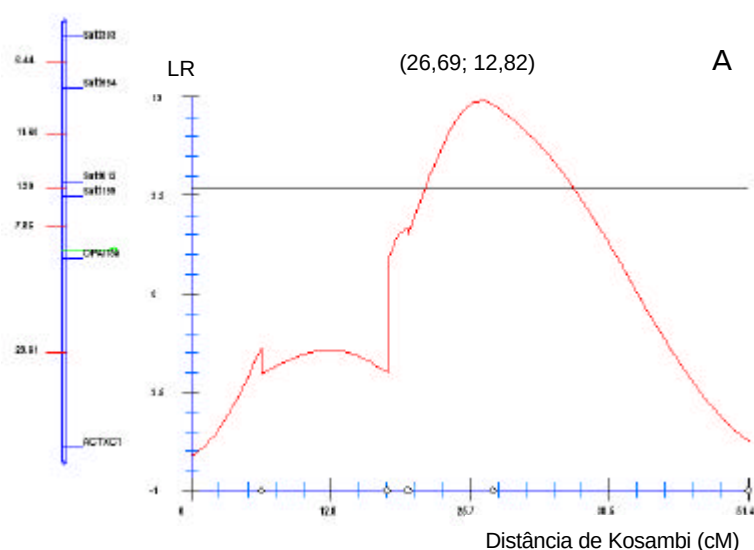


Figura 5 – QTLs associados ao conteúdo da subunidade α . Grupo de ligação G (5A) representado à esquerda, e a curva da razão de verossimilhança representada no gráfico à direita. O pico da curva representa um valor de LR de 12,82 (eixo Y) na posição 26,69 cM (eixo X) representado no mapa entre os marcadores Satt199 e OPA15a. Grupo de ligação C2 (5B) com um valor de LR de 10,47 e na posição 1,0 cM representado no mapa entre os marcadores Satt277 e Satt286. O valor de LR para o ponto de corte em ambos os casos foi de 9,73.

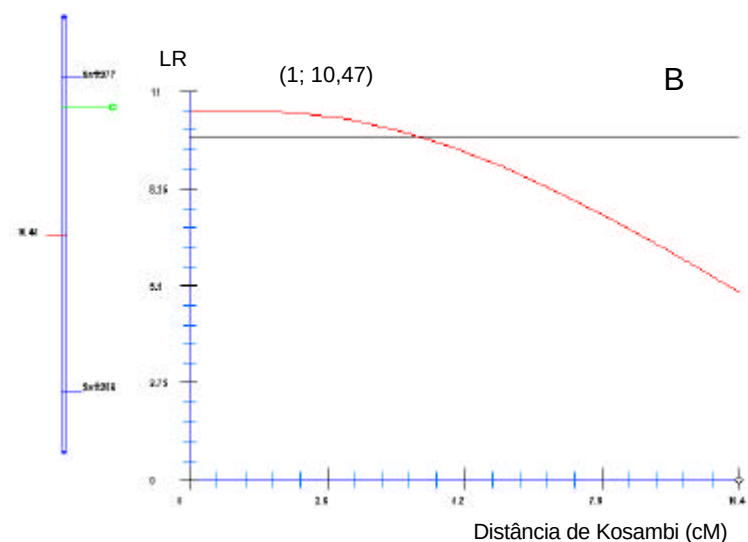


Figura 5- Continuação...

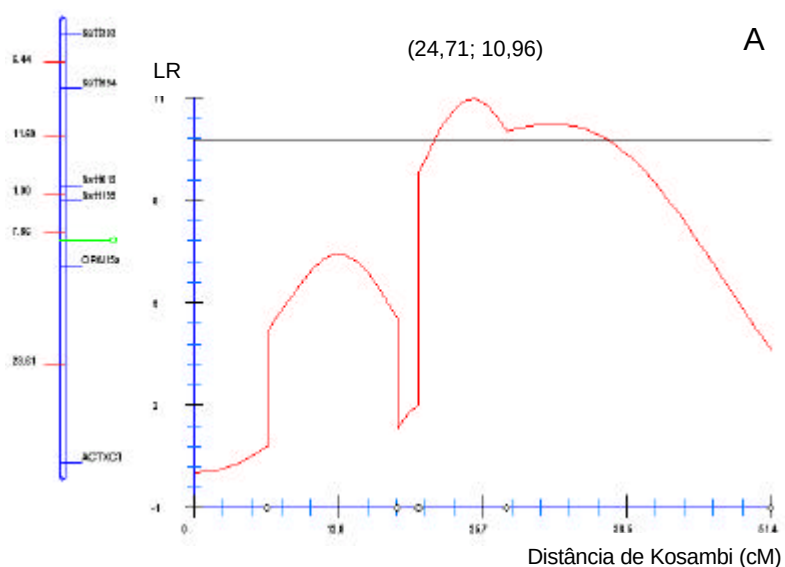


Figura 6 – QTLs associados ao conteúdo da subunidade β . Grupo de ligação G (6A) representado à esquerda, e a curva da razão de verossimilhança representada no gráfico à direita. O pico da curva representa um valor de LR de 10,96 (eixo Y) na posição 24,71 cM (eixo X) representado no mapa entre os marcadores Satt199 e OPAI15a. Fragmento do grupo de ligação K (6B), com um valor de LR de 9,94 na posição 287 cM entre os marcadores Satt187 e Satt055. Segundo fragmento do grupo K (6C), com um valor de LR de 30,62 na posição 12,09 cM entre os marcadores OPAW09a e Satt273. O valor de LR para o ponto de corte foi de 9,73 para os 3 QTLs.

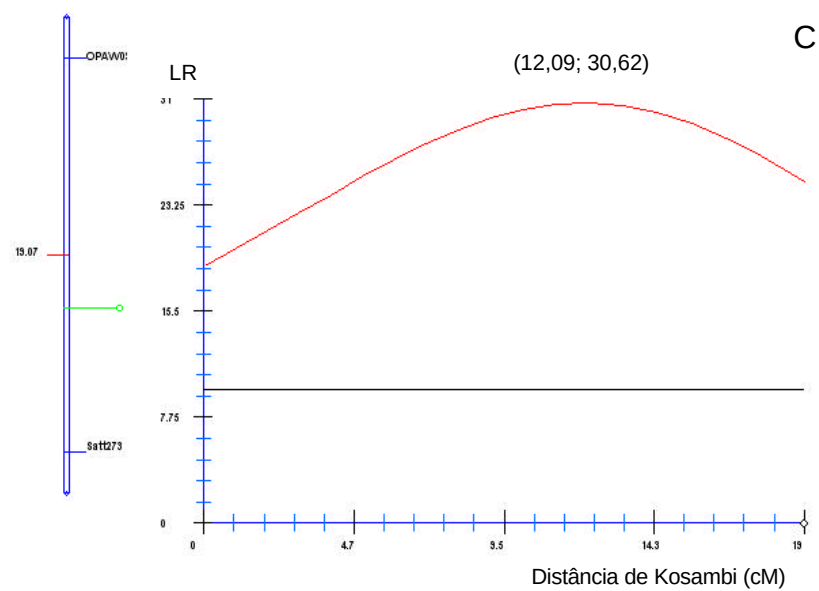
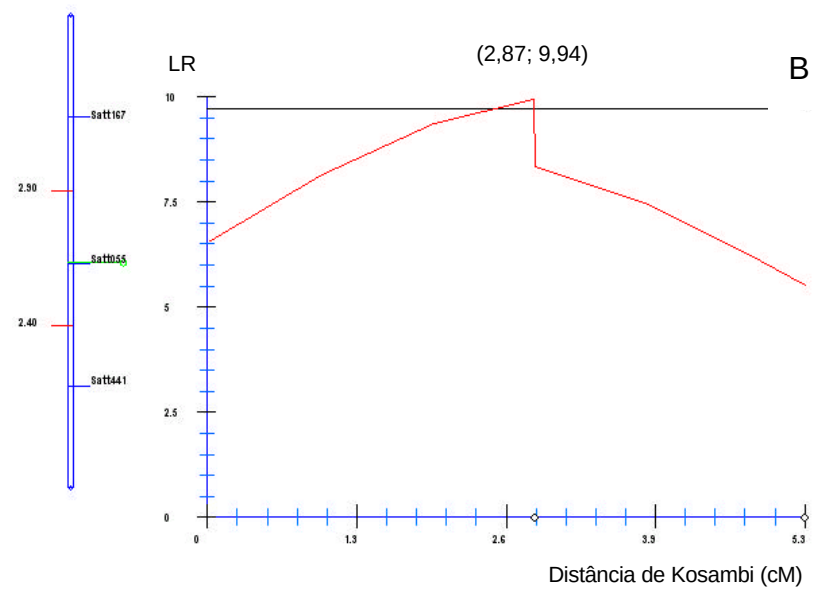


Figura 6 – Continuação...

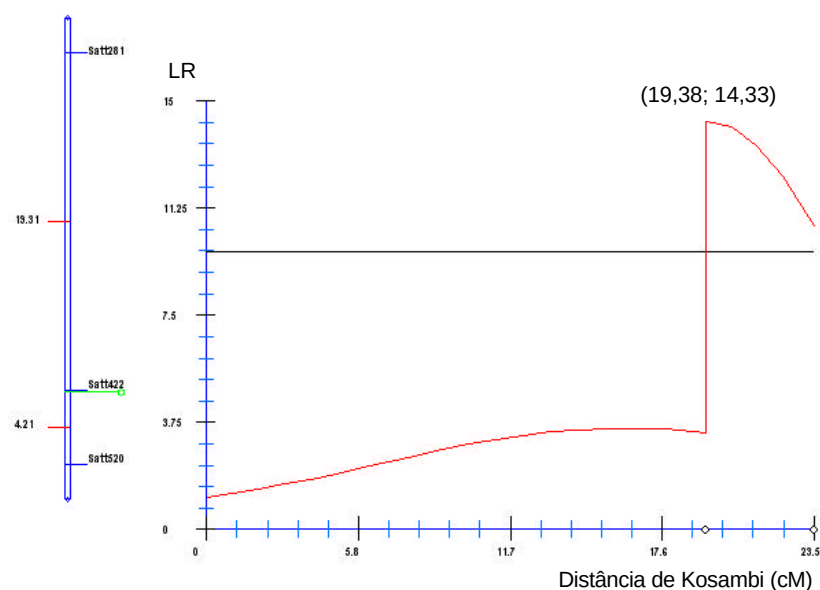


Figura 7 – QTL associado ao conteúdo das outras proteínas presentes no extrato PT-(11S+7S). Grupo de ligação D1b representado à esquerda, e a curva da razão de verossimilhança representada no gráfico à direita. O pico da curva representa um valor de LR de 14,33 (eixo Y) na posição 19,38 cM (eixo X) representado no mapa entre os marcadores Satt422 e Satt520. O valor de LR para o ponto de corte foi de 9,73.

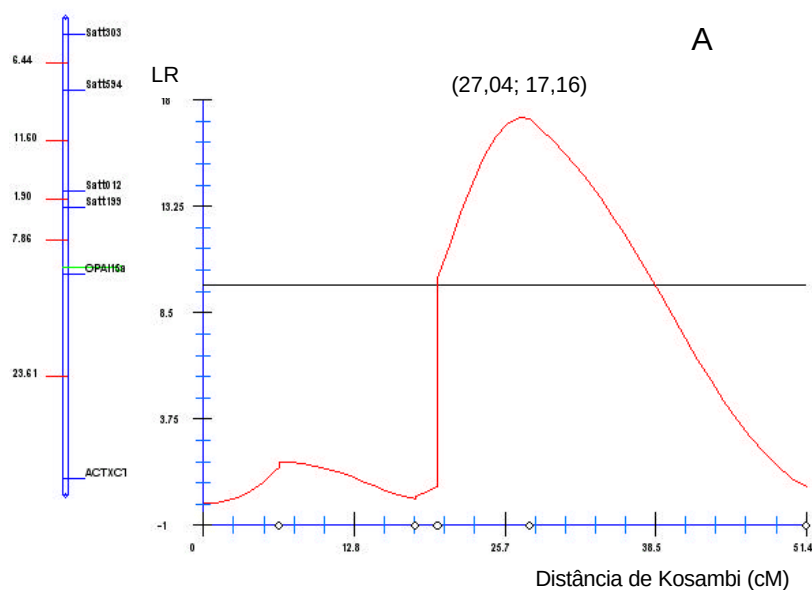


Figura 8 - QTL associado ao conteúdo de 7S. Grupo de ligação G (8A) representado à esquerda, e a curva da razão de verossimilhança representada no gráfico à direita. Os picos da curva representam valores de LR (eixo Y) de 17,16 na posição 27,04cM representado no mapa entre os marcadores Satt199 e OPA15a. Grupo de ligação E (8B), com LR de 10,46 na posição 33,29cM entre os marcadores Satt384 e AACXCAC239. O valor de LR para o ponto de corte foi de 9,73 para ambos os QTLs.

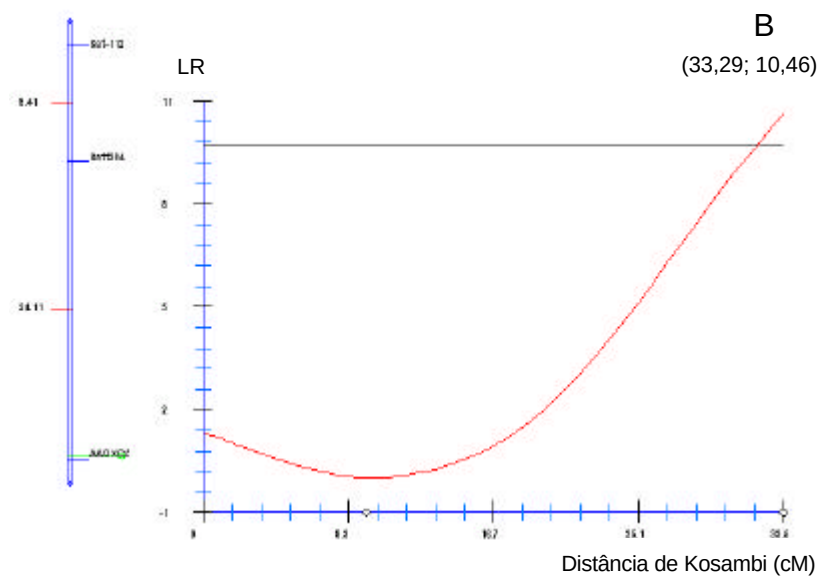


Figura 8- Continuação...

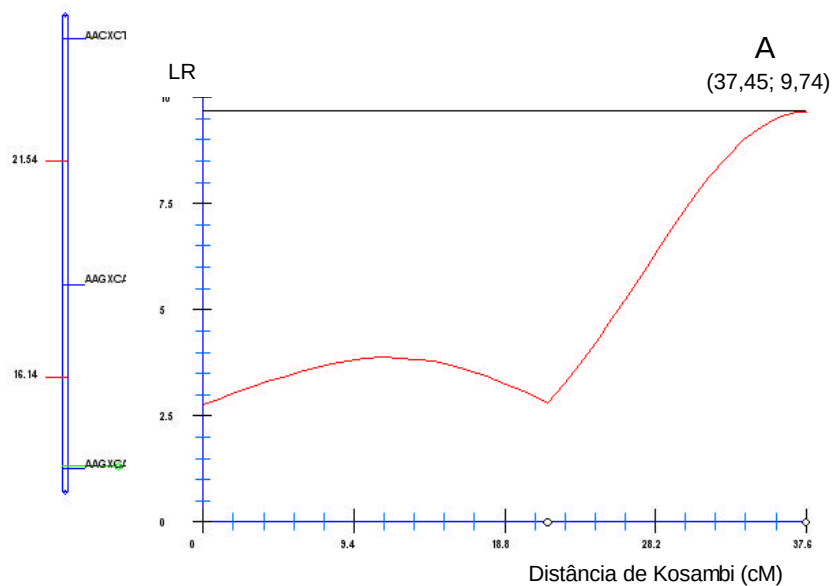


Figura 9 – QTL associado ao conteúdo de proteína total, experimento de São Gotardo. Grupo de ligação 3 (9A) representado à esquerda, e a curva da razão de verossimilhança representada no gráfico à direita. O pico da curva representa um valor de LR de 9,74 (eixo Y) na posição 37,45cM (eixo X) representado no mapa entre os marcadores AAGXCAC189 e AAGXCAC243. Grupo de ligação I (9B), com LR de 11,31 na posição 10,11 cM entre os marcadores Sat-105 e Satt239. O valor de LR para o ponto de corte foi de 9,73 em ambas as análises.

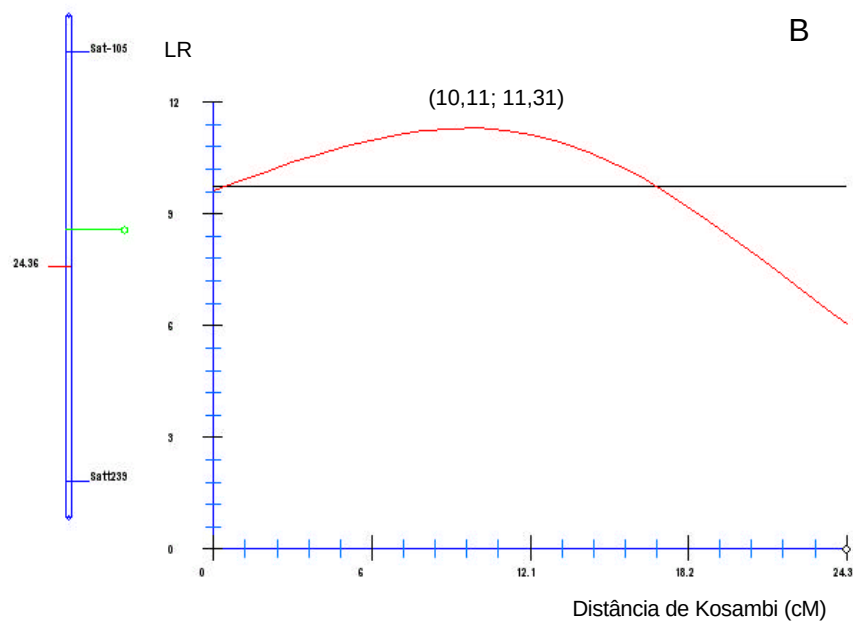


Figura 9 – Continuação...

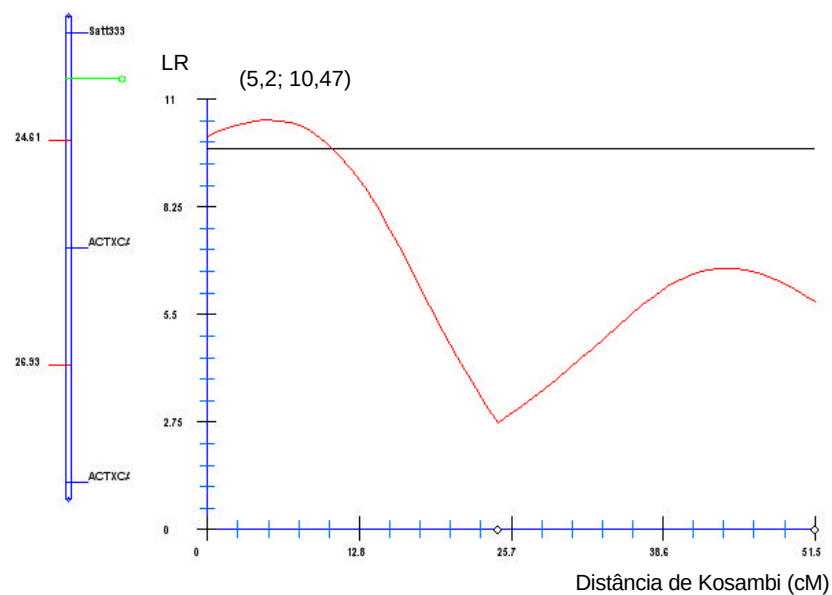


Figura 10 – QTL associado ao conteúdo da subunidade α' no experimento de São Gotardo. Grupo de ligação A2 representado à esquerda, e a curva da razão de verossimilhança representada no gráfico à direita. O pico da curva representa um valor de LR de 10,47 (eixo Y) na posição 5,2 cM (eixo X) representado no mapa entre os marcadores ACTXCAG190 e ACTXCAG90. O valor de LR para o ponto de corte foi de 9,73.

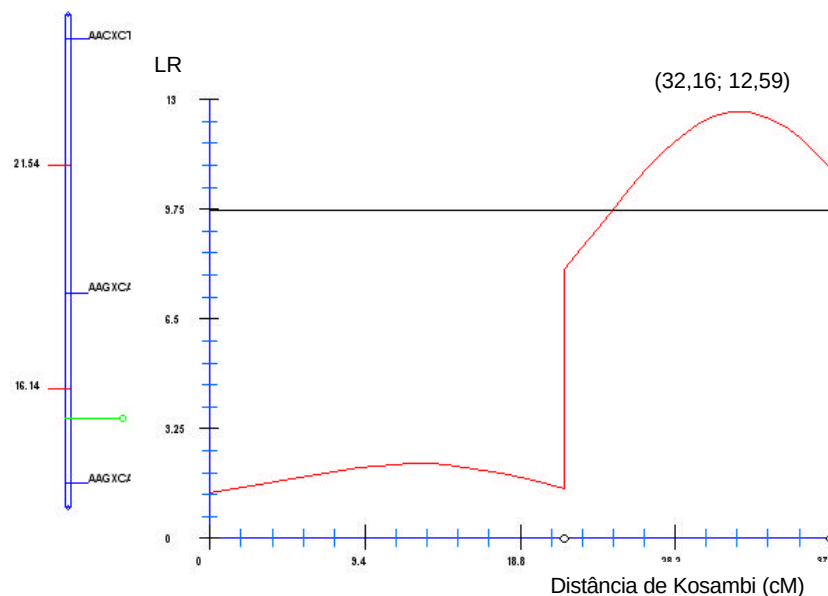


Figura 11 – QTL associado ao conteúdo das subunidades ácidas no experimento de São Gotardo. Grupo de ligação 3 representado à esquerda, e a curva da razão de verossimilhança representada no gráfico à direita. O pico da curva representa um valor de LR de 12,59 (eixo Y) na posição 32,16 cM (eixo X) representado no mapa entre os marcadores AACXCTG325 e AAGXCAC189. O valor de LR para o ponto de corte foi de 9,73.

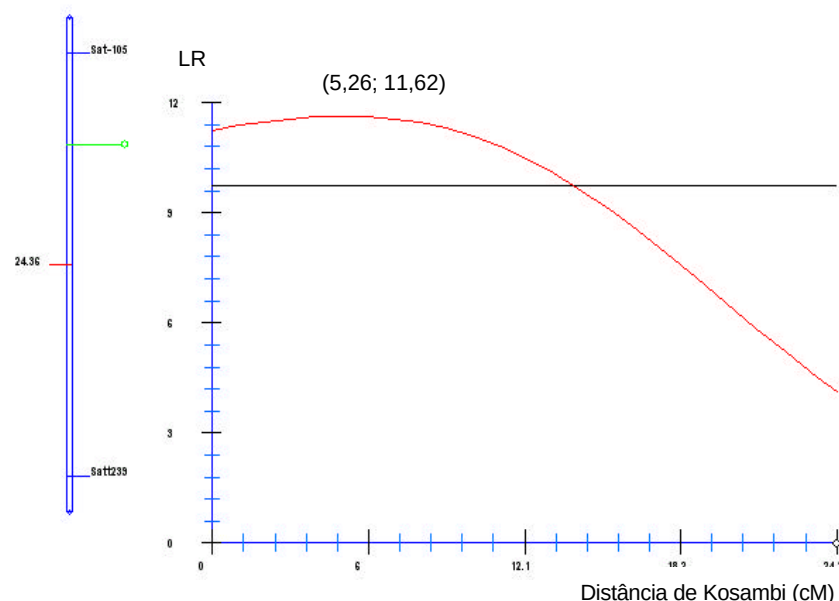


Figura 12 – QTL associado ao conteúdo das outras proteínas presentes no extrato, no experimento de São Gotardo. Grupo de ligação I representado à esquerda, e a curva da razão de verossimilhança representada no gráfico à direita. O pico da curva representa um valor de LR de 11,62 (eixo Y) na posição 5,26 cM (eixo X) representado no mapa entre os marcadores Sat-105 e Satt239. O valor de LR para o ponto de corte foi de 9,73.

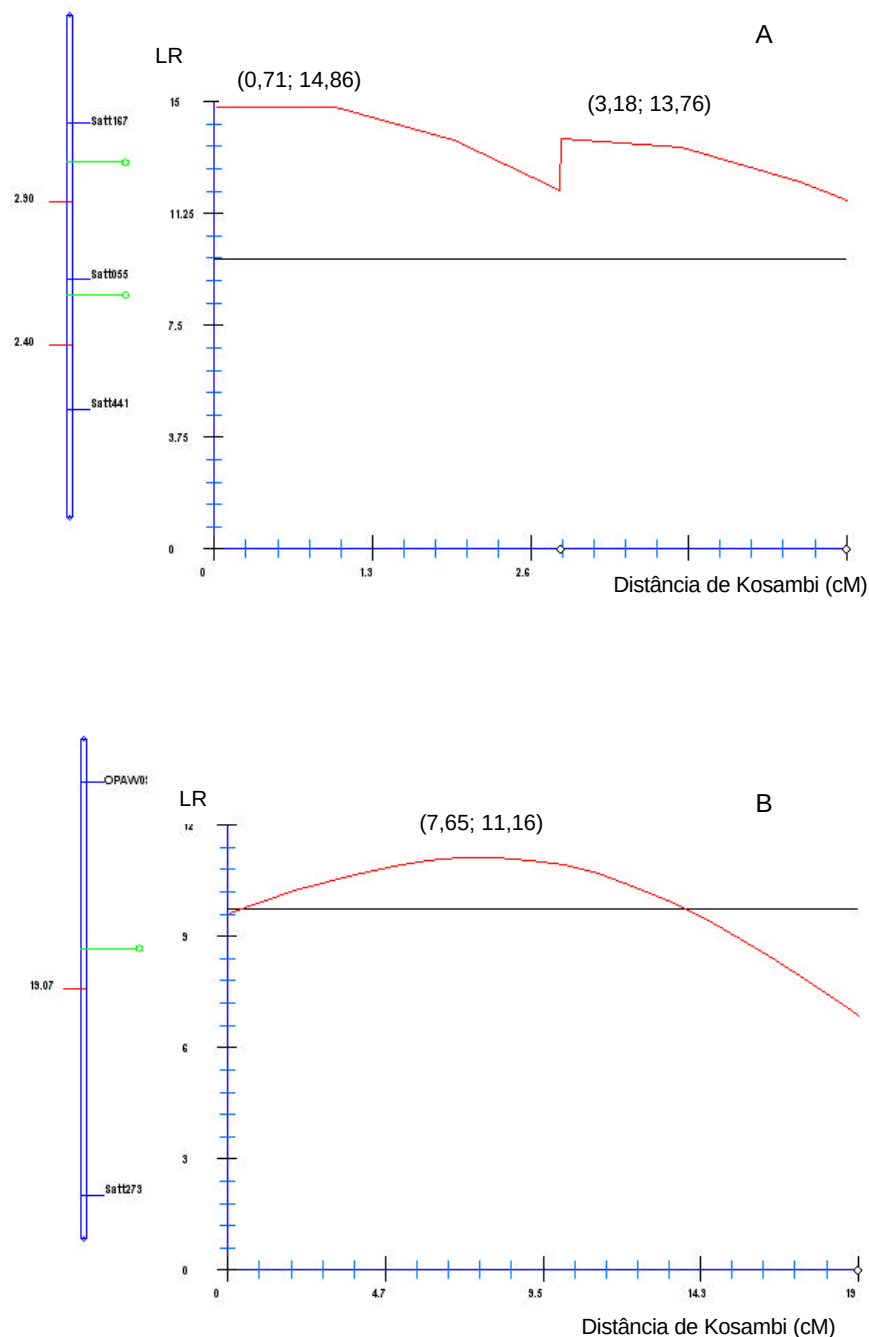


Figura 13 - QTL associado ao conteúdo de 11S, experimento de São Gotardo. Fragmento do grupo de ligação K (13A) representado à esquerda, e a curva da razão de verossimilhança representada no gráfico à direita. Os picos da curva representam valores de LR (eixo Y) de 14,36 na posição 0,71 cM representado no mapa entre os marcadores Satt167 e Satt055 (pico da esquerda), e LR de 13,76 na posição 3,18cM entre os marcadores Satt055 e Satt441 (pico da direita). Outro fragmento do GL K (13B), com LR de 11,16 na posição 7,65 cM entre os marcadores OPAW09a e Satt273. O valor de LR para o ponto de corte foi de 9,73 em ambas as análises.

4. RESUMO E CONCLUSÕES

Os programas de melhoramento de soja têm-se preocupado, mais recentemente, com o desenvolvimento de variedades produtivas e com altos teores de proteínas. Para atingir esse objetivo, um melhor entendimento do tipo de ação gênica e herança dos caracteres envolvidos se fazem necessários para o desenvolvimento de procedimentos mais eficientes de melhoramento.

No presente trabalho o objetivo foi identificar marcadores do tipo microssatélites, RAPD e AFLP ligados a QTLs que controlam as características teor de proteína total em sementes de soja, teor das subunidades das proteínas de reserva 11S e 7S, e da relação proteína total menos proteínas de reserva (PT-(11S+7S)). Além disto, o trabalho teve como meta 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), utilizando genótipos tropicais. Para isso, utilizaram-se 118 RILs (linhagens endogâmicas recombinantes), que se encontravam na geração F₉ e foram obtidas do cruzamento entre dois genótipos contrastantes para teor de proteínas: BARC-8 (alto teor protéico) e Garimpo (teor normal de proteínas). Essas RILs foram cultivadas em dois ambientes distintos: Viçosa, MG, e São Gotardo, MG.

Os resultados obtidos neste trabalho permitiram apontar algumas conclusões:

1. Foram obtidos 29 grupos de ligação pouco saturados, contendo 95 marcadores, e mais 79 marcas não-ligadas. Foram mapeados ainda cinco novos marcadores: Satt449 e Satt526 (pertencentes ao grupo A1), Satt514 e Satt528 (grupo D2) e Satt594 (grupo G), que não estavam presentes em nenhum dos três mapas (USDA/Iowa State Univ., Univ. of Utah, Univ. of Nebraska) utilizados na integração por CREGAN *et al.* (1999).
2. No experimento de Viçosa, foram encontrados 13 marcadores associados ao conteúdo de proteína total e explicam de 4,39% a 13,32% da variação do teor de proteína. Ao conteúdo da subunidade α' da proteína 7S, foram encontrados 9 marcadores associados, que explicam de 4,21% a 7,09% da variação desta característica. Dezesete marcadores mostraram associação

com o conteúdo da subunidade α da proteína 7S e explicam de 4,07% a 13,16% da variação desta característica. Para o conteúdo da subunidade β foram encontrados 16 marcadores associados, e eles explicam de 4,78% a 21,80% da variação desta característica. Associados ao conteúdo das subunidades ácidas e básicas da proteína 11S, foram encontrados 17 e 8 marcadores, respectivamente, e eles explicam de 4,08% a 9,60% (subunidades ácidas) e de 5,49% a 11,10% (subunidades básicas) da variação destas características. Associados ao conteúdo da característica proteína total menos proteína de reserva ((PT-(11S+7S))) foram encontrados sete marcadores que explicam de 4,11% a 10,77% da variação desta característica. Dezenove marcadores mostraram associação com o conteúdo de 7S e eles explicam de 4,08% a 15,57% da variação desta característica. Dezesseis marcadores apresentaram associação com o conteúdo de 11S e eles explicam de 4,17% a 9,69% da variação desta característica.

3. No experimento de São Gotardo foram encontrados 13 marcadores associados a conteúdo de proteína total que explicam de 4,31% a 12,11% da variação do teor de proteína. Para a subunidade α' da proteína 7S, foram encontrados 13 marcadores que explicam de 4,34% a 10,42% da variação desta característica. Cinco marcadores mostraram associação com a subunidade α da proteína 7S e explicam de 4,32% a 6,17% da variação desta característica. Para a subunidade β foram encontrados 12 marcadores associados, e eles explicam de 4,09% a 9,11% da variação desta característica. Para as subunidades ácidas e básicas da proteína 11S, foram encontrados 13 e 17 marcadores respectivamente e eles explicam de 4,21% a 13,36% (subunidades ácidas) e de 4,12% a 8,12% (subunidades básicas) da variação destas características. Para a característica proteína total menos proteína de reserva ((PT-(11S+7S))) foram encontrados 6 marcadores associados que explicam de 4,39% a 7,92% da variação desta característica. Seis marcadores mostraram associação com o conteúdo de 7S e eles explicam de 5,16% a 8,11% da variação desta característica. Dezesseis marcadores apresentaram

associação com o conteúdo de 11S e eles explicam de 4,15% a 13,67% da variação desta característica.

4. Foram encontrados QTLs associados: ao conteúdo de proteína total (GL I, que explica 14,74%), ao conteúdo da subunidade α (GL G e GLC2 que explicam em conjunto 28,83% da variação desta característica), ao conteúdo da subunidade β (GL G e dois no grupo K, que explicam em conjunto 32,62% da variação desta característica), ao conteúdo de PT-(11S+7S) (GL D1b, que explica 13,71%) e conteúdo de 7S (GLG e GLE) que explicam em conjunto 21,86% da variação desta característica no experimento de Viçosa.
5. No experimento de São Gotardo foram encontrados QTLs associados ao conteúdo de proteína total (GL 3 e I, que explicam em conjunto 11,42%), ao conteúdo da subunidade α' (GL A2, que explica 10,64%), ao conteúdo da subunidades ácidas (GL 3, que explica 12,06%), PT-(11S+7S) (GL I, que explica 11,34%) e ao conteúdo da de 11S (dois QTLs em diferentes fragmentos do GL K) que explicam em conjunto 20,88% da variação desta característica.
6. Os marcadores associados ao teor de proteína foram diferentes de um local para outro, o que indica interação entre o genótipo e o ambiente.
7. O grupo de ligação I foi associado a QTLs para proteína total em Viçosa e em São Gotardo, este fato sugere que os locos associados a esse grupo de ligação são conservados.
8. Os dois fragmentos do grupo K que apresentaram QTLs associados ao conteúdo da subunidade β de 7S no experimento de Viçosa, apresentaram QTLs associados ao conteúdo de 11S em São Gotardo. Este fato não era esperado, uma vez que estas duas características (conteúdo de 7S e 11S) apresentaram correlação genotípica próximas de zero nos dois ambientes.
9. No experimento de Viçosa também foram encontrados um QTL associado à proteína 7S no grupo de ligação E, e este grupo não apresentou nenhum outro QTL associado às subunidades constituintes desta proteína, o que sugere que podem existir outros genes associados à expressão de 7S além dos genes que expressam as suas subunidades constituintes.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- AKKAYA, M.S., SHOEMAKER, R.C., SPECHT, J.E., BHAGWAT, A.A., GREGAN, P.B. Integration of simple sequence repeat DNA markers into soybean linkage map. **Crop Sci.**, v.35, n.4, p.1439-1445, 1995.
- BRIM, C.A. Quantitative genetics and breeding. In: B.E. Caldwell (ED.) **Soybean: Improvement, production, and uses**. ASA, Madison, WI. 1973. p.155-186.
- BRUMMER, E.C., GRAEF, G.L., ORF, J., WILCOX, J.R., SHOEMAKER, R.C. Mapping QTL for seed protein and oil content in eight soybean populations. **Crop Sci.**, v.37, n.1, p.370-378, 1997.
- CHUNG, J., BABKA, H.L., GRAEF, P.E., STASWICK, D.J., LEE, D.J., CREGAN, P.B., SHOEMAKER, R.C., SPECHT, J.E. The seed protein, oil, and yield QTL on soybean linkage group I. **Crop Sci.**, v.43, p. 1053-1067, 2003.
- CREGAN, P.B., JARVIK, T., BUSH, R.C., 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 genome. **Crop Sci.**, v.39, n.5, p.1464-1490, 1999.
- DARVASI, A.; SOLLER, M. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. **Theor. Appl. Genet.**, 85: 353 – 359, 1994.
- DOYLE, J.J., DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**, v.12, n.1, p.13-15, 1990.
- FEHR, W.R. Breeding. In: NORMAN, A.G. **Soybean physiology, agronomy, and utilization**. New York: Academy Press, 1978. p.119-155.
- 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**, v.77, n.3, p.269-75, 1994.
- JANSEN, R.C. Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm. **Biometrics**, v.49, n.1, p.227-231, 1993.
- JOHNSON, H.W., BERNARD, R.L. Soybean genetics and breeding. In: NORMAN, A.G. (Ed.). **The soybean**. New York: Academy Press, 1963. p.1-73.
- KEIM, P., SCUPP, J.M., TRAVIS, S.E., CLAYTON, K., WEBB, D.M. A high-density soybean genetic map based upon AFLP markers. **Crop Sci.**, v.37, n.2, p.537-543, 1997.

- LARK, K.G., WEISEMANN, J.M., MATTHEWS, B.F., PALMER, R., CHASE, K., MACALMA, T. A genetic map of soybean (*Glycine max* L. Merr) using an intraspecific cross of two cultivars: 'Minosy' and 'Noir 1'. **Theor. Appl. Genet.**, v.86, n.8, p.901-906, 1993.
- LYNCH, M., WALSH, B. **Genetics and analysis of quantitative traits**. Massachusetts: Sinauer Associates, 1998. 980p.
- MANSUR, L.M.L., LARK, K.G., KROSS, H., OLIVEIRA, A. Interval mapping of quantitative trait loci for reproductive, morphological and seed traits of soybean (*Glycine max* L. Merr.). **Theor. Appl. Genet.**, v.86, n.8, p.907-913, 1993.
- MIRANDA, F. D. **Uso de marcadores RAPD para mapeamento de QTLs que determinam teor de proteína em soja**. Viçosa, MG: UFV, 2002. 56p. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, 2002.
- PATERSON, A.H., DAMON, S., HEWITT, J.D., ZAMIR, D., RABINOWITCH, H.D., LINCOLN, S.E., LANDER, E.S., TANKSLEY, S.D. Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. **Genetics**, v.127, n.1, p.181-197, 1991.
- 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, 1989. v.3, p.8B.
- SHOEMAKER, R.C. RFLP map of soybean. In: PHILLIPS, R.L., VASIL, I.K., (Eds.). **DNA- based markers in plants**. [s.l.]: Kluwer Academic, 1994. p.299-309.
- SHOEMAKER, R.C., POLZIN, K., LABATE, J., SPECHT, J., BRUMMER, E.C., OLSON, T., YOUNG, N., CONCIBIDO, V., WILCOX, J., TAMULONIS, J.P., KOCHERT, G., BOERMA, H.R. Genome duplication in soybean (*Glycine* subgenus *soja*). **Genetics**, v.144, n.1, p.329-338, 1996.
- SOARES, T. C. B. **Mapeamento de locos que controlam o conteúdo de proteína em soja**. Viçosa, MG: UFV, 2000. 58p. Dissertação (Mestrado em Agroquímica) – Universidade Federal de Viçosa, 2000.
- STUBER, C.W., LINCOLN, S.E., WOLFF, D.W., HELENTJARIS, T., ANDER, E.S. Identification of genetic factors contributing to heterosis in a hybrid from

- two elite maize inbred lines using molecular markers. **Genetics**, v.132, n.3, p.823-839, 1992.
- TANKSLEY, S.D., HEWITT, J. Use of molecular markers in breeding for soluble solids content in tomato- A re-examination. **Theor. Appl. Genet.**, v.75, n.5, p.811-823, 1988.
- VOS, P., HOGERS, R., BLEEKER, M., REIJANS, M., LEE, T.V., HORNES, M., FRIJTERS, A., POT, J., PELEMAN, J., KUIPER, M., ZABEU, M. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Res.**, v.23, n.21, p.4407-4414, 1995.
- WILCOX, J.R., CAVINS, J.F. Backcrossing higher seed protein to a soybean cultivar. **Crop Sci.**, v.35, n.4, p.1036-1041, 1995.
- WILLIAMS, J.G.K., KUBELIK, A.R., LVAK, K.J., RAFALSAKI, J.A., TINGEY, S.V. DNA polymorphisms amplified by arbitrary *primers* are useful as genetic markers. **Nucleic Acids Res.**, v.18, n.13, p.6531-6535, 1990.
- ZENG, Z.B. Precision mapping quantitative trait loci. **Genetics.**, v.136, n.9, p.1457-1468, 1994.
- ZENG, Z.B. Theoretical basis for separation of multiple linked gene effect in mapping quantitative trait loci. **Proc. Natl. Acad. Sci. USA**, v.90, n.20, p.10972-10976, 1993.