

TAÍS CRISTINA BASTOS SOARES

MAPEAMENTO DE LOCOS QUE CONTROLAM O CONTEÚDO DE
PROTEÍNA EM SOJA

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

VIÇOSA
MINAS GERAIS – BRASIL
NOVEMBRO – 2000

TAÍS CRISTINA BASTOS SOARES

MAPEAMENTO DE LOCOS QUE CONTROLAM O CONTEÚDO DE
PROTEÍNA EM SOJA

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

APROVADA: 11 de fevereiro de 2000.

Prof. Everaldo Gonçalves de Barros
(Conselheiro)

Prof^a Maria Goreti de Almeida Oliveira
(Conselheira)

Dr. Ivan Schuster

Prof. Luiz Orlando de Oliveira

Prof. Maurilio Alves Moreira
(Orientador)

A Deus, a quem “pertence o poder e retribui a cada um, segundo sua obra”.
Aos meus pais e aos meus irmãos Yaska e Thiago.

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 Mestrado.

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.

À professora Maria Goreti de Almeida Oliveira, pela amizade e pelas sugestões.

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

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

À amiga Analina, pelo incentivo desde a minha iniciação científica, pela amizade, pela dedicação e pelos ensinamentos.

Às minhas amigas Aline, Ariane, Roberta, Shirley, Dili, Sandra, Denise Bazzolli e Andréa, 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, que me orientou durante a minha graduação; a ele toda a minha gratidão e o meu respeito, pela confiança depositada em mim.

Aos meus colegas Lucinete, Lucimara, Valéria, Anna, Rita, Tatiana, Inês, Carlos, Newton, Fábio, Joselito, Cynthia, Cristiano, Vagner, Valéria Dellaretti, Márcio, Pedro, Fran, Malu, Márcia, Abelmon 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.

Aos meus colegas do Curso de Agroquímica, pela agradável convivência.

Aos funcionários do BIOAGRO Jander, Márcia, 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, submetendo-se à defesa de tese em fevereiro de 2000.

CONTEÚDO

	Página
EXTRATO	viii
ABSTRACT.....	x
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.....	7
2.4. Marcadores moleculares.....	10
2.5. Mapeamento genético.....	12
2.5.1. Construção de mapas genéticos.....	13
2.5.2. Mapeamento de caracteres de herança quantitativa.....	15
3. MATERIAL E MÉTODOS	17
3.1. Material genético	17
3.2. Análise do conteúdo de proteína.....	18
3.3. Análise de variância	18
3.3.1. Análise de variância individual.....	18
3.3.2. Análise de variância conjunta.....	20
3.4. Extração de DNA.....	24
3.5. Condições de amplificação e de eletroforese.....	25

	Página
3.6. Construção do mapa de ligação genética.....	26
3.7. Identificação e mapeamento de QTL's.....	26
3.7.1. Determinação do ponto de corte para mapeamento de QTL's.....	28
4. RESULTADOS E DISCUSSÃO.....	30
4.1. Detecção de polimorfismos e seleção dos <i>primers</i> de microssatélites.....	30
4.2. Mapa de ligação genética.....	32
4.3. Análise do teor de proteína.....	36
4.4. Análise de QTL's para teor de proteína.....	43
4.5. Mapeamento de QTL's.....	46
5. RESUMO E CONCLUSÕES	50
REFERÊNCIAS BIBLIOGRÁFICAS	52

EXTRATO

SOARES, Taís Cristina Bastos, M. S., Universidade Federal de Viçosa, novembro de 2000. **Mapeamento de locos que controlam o conteúdo de proteína em soja**. Orientador: Maurilio Alves Moreira. Conselheiros: Everaldo Gonçalves de Barros e Maria Goreti de Almeida Oliveira.

Uma população de 118 RIL's (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 F₆ foram abertas linhas, que constituíram o material genético utilizado neste trabalho. Esse material foi analisado, utilizando-se marcadores microssatélites. Foi feita também a análise do teor de proteínas para essa população cultivada em dois ambientes distintos: Cascavel, PR, e Viçosa, MG. Análise de regressão simples e múltipla e mapeamento por intervalo composto foram utilizados para detectar e mapear as regiões genômicas associadas com alto teor de proteína. A proporção da variância fenotípica explicada por marcador variou entre 3,25% e 6,37% em Cascavel e entre 2,92 % e 12,43% em Viçosa. Na análise de regressão múltipla, os marcadores Satt190, Satt384, Satt422 e Sat-105 explicaram aproximadamente 23% do total da variação do teor de proteínas (Cascavel), e os marcadores Satt190, Satt384, Satt012, Satt304, Satt369 e Sat-105 explicaram, juntos, aproximadamente 31% do total da variação do teor de proteínas em Viçosa. Utilizando-se as médias dos

teores de proteína dos dois locais, dois marcadores mostraram-se estáveis nos dois ambientes: Satt384 e Sat-105. Esses marcadores estão associados a genes conservados, cujas funções estão ligadas à determinação do conteúdo de proteínas na semente. No mapeamento por intervalo composto, foram encontrados dois QTL's associados a teor de proteína nos grupos de ligação C1 e C2, para as famílias cultivadas em Cascavel, que explicam, respectivamente, 11,13% e 12,19% da variação do teor de proteínas. Outras regiões foram encontradas nos grupos de ligação B2, C1, G, K (Viçosa) e E (Cascavel), mas a ligação entre essas regiões e o conteúdo de proteínas não foi significativa nas condições do teste de ligação efetuado.

ABSTRACT

SOARES, Taís Cristina Bastos, M.S., Universidade Federal de Viçosa, November, 2000. **Mapping loci that control protein content in soybean.** Adviser: Maurilio Alves Moreira. Committee Members: Everaldo Gonçalves de Barros and Maria Goreti de Almeida Oliveira.

A soybean population of 118 RILs (recombinant inbred lines) obtained from the cross between the accession BARC - 8 (genotype with high protein content, around 50%) and the brazilian commercial variety Garimpo (genotype with normal protein content, around 36%) was used for the construction of a molecular linkage map, with SSR markers, and for identifying QTLs which control protein level in soybean seeds. F₆ generations were grown in two distinct locations (Viçosa, MG and Cascavel, PR) in the summer of 1998/99. Protein contents in the seeds from each line cultivated in the two locations were determined by the Kjeldahl method. DNA samples purified from leaves from each line of the population were amplified with 567 soybean SSR primers. Sixty five SSR markers were polymorphic and segregated in a 1:1 ratio in the population. Single and multiple regression and composed interval mapping analysis were used to detect genomic regions associated with high protein content. The SSR markers were also used to construct a molecular genetic map containing 16 linkage groups. The proportion of the phenotypic variance explained by the markers varied from 3,25% to 6,37% in Cascavel, and from

2,92% to 12,43% in Viçosa. Multiple regression analysis identify that markers Satt190, Satt384, Satt422 and Sat-105 explained 23% of the total protein content variation in the RILs cultivated in Cascavel, PR, and Satt190, Satt384, Satt012, Satt304, Satt369 and Sat-105 explained 31% of the total protein content variation in the RILs grown in Viçosa, MG. By using mean values of protein content of the two places, markers Satt384 and Sat-105 shown to be stable in the two environment. It is proposed that these markers are associated with conserved genes, whose functions are very important in protein synthesis and deposition in the seed. Composited interval mapping analysis identified two QTL's associated with protein content in linkage groups C1 and C2 for RIL's cultivated in Cascavel, which explained, respectively, 11,13% and 12,19% of protein content variation. Other genomic regions were detected in linkage groups B2, C1, G, K (Viçosa) and E (Cascavel), but with no significative correlation with protein content based on the test used.

1. INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill) constitui um dos produtos agrícolas de maior importância para o Brasil, ocupando lugar de destaque na pauta de exportação do país, 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 e obtenção de novos cultivares são de fundamental importância para a manutenção desse potencial produtivo nacional.

O interesse mundial pelo cultivo da soja deve-se à excelente combinação que ela apresenta entre produtividade e conteúdo de óleo e proteína.

O desenvolvimento de procedimentos mais eficientes de melhoramento depende do melhor entendimento do tipo de ação gênica e herança dos caracteres envolvidos. Atualmente, os programas de melhoramento de soja têm dado ênfase ao desenvolvimento de variedades mais produtivas contendo, ao mesmo tempo, altos teores de óleo e proteína. Portanto, para atender a objetivos tão diversos, é imprescindível que se disponha de ampla variabilidade genética.

As primeiras variedades de soja cultivadas no Brasil no início dos anos 70 foram introduções de variedades americanas ou foram derivadas de intercruzamentos dessas variedades. Sabe-se também que essas variedades foram selecionadas principalmente para produção de grãos e teor de óleo. Pelo

fato de esses dois caracteres serem negativamente correlacionados com teor de proteína, este tem-se mantido em níveis mais baixos (em torno de 40%) do que potencialmente pode ser conseguido (em torno de 50% em materiais não-adaptados).

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. Foi construído um primeiro mapa genético 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).

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 PI437.654. 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. Foram definidos 20 grupos de ligação-consenso.

Até o momento, os mapas disponíveis para a soja são, 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).

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 (QTL's), como: altura de plantas, produção de sementes, resistência ao acamamento, percentagem de óleo e de proteína e tamanho das sementes de soja (FEHR, 1978).

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 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, através de uma série de retrocruzamentos, entre dois progenitores contrastantes com relação a essas duas características, verificaram correlação negativa 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 objetivo identificar marcadores do tipo de microssatélites ligados a QTL's que controlam o teor de proteína em soja, bem como dar início à construção de um mapa genético para soja (*G. max* x *G. max*), utilizando-se genótipos tropicais.

2. REVISÃO DE LITERATURA

2.1. Base genética da soja

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

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 devida ao fato de a soja ser uma espécie autógama, com indivíduos altamente homocigotos; 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 (CIZLICE et al., 1993; MAUGHAN et al., 1996). CIZLICE et al. (1993) citaram que, nos Estados Unidos, a base genética das variedades recomendadas é derivada de cerca 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

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).

Essa proteína 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 bissulfídricas 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) e seus 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 seqüência 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 -conglucina 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 -conglucina, 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 -conglucina 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).

2.3. Herdabilidade do teor de proteínas

Herdabilidade é definida como a fração da variância fenotípica de uma unidade 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 progenitores. A seleção é efetuada, fundamentalmente, nessa semelhança.

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 e F_2 , 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.

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.

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.

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* \times *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.

2.4. Marcadores moleculares

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 se baseiam 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 quais se baseiam em características como cor de flor e cor de semente, dentre outras. 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; e, finalmente, esses marcadores não apresentam efeitos pleiotrópicos (Evola et al., 1986, Beckman e Soller, 1986, e 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 seqüência arbitrária e conteúdo de CG maior que 50%, que se ligam ao DNA genômico em dois sítios, em fitas opostas do DNA-molde. Se esses sítios

estiverem separados por uma distância amplificável (geralmente 200 a 2000 pb), produtos de amplificação são formados. O número e o tamanho dos fragmentos amplificados dependem da seqüência do *primer* usado e da seqüência 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 se baseiam 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 seqüência de DNA. No entanto, essa técnica não permite a distinção entre indivíduos heterozigotos e homozigotos, sendo os marcadores, portanto, considerados de herança dominante.

Marcadores AFLP é uma classe recente de marcadores, que alia a especificidade dos sítios de restrição do RFLP com a praticidade da amplificação do 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 seqüência 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* que contêm a seqüência dos *primers* pré-seletivos acrescida de dois nucleotídios 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 gerado por essa técnica é preferencialmente dominante, sendo possível, no entanto, identificar indivíduos heterozigotos pela diferença de intensidade das bandas, se esta for realizada com boa precisão.

Os genomas de organismos eucariotos são caracterizados pela ocorrência de grandes quantidades de seqüências repetidas de nucleotídios 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 cobre 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).

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).

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 originais 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 apresentam 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 (RIL's), 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 a estimativa do mecanismo de ação gênica e das interações com outros locos. Em populações recombinantes endogâmicas (RIL's), 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 RIL's, o genótipo dos indivíduos é perpetuado; 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.

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 foram avaliados pelo teste qui-quadrado, e aqueles que apresentarem segregação mendeliana foram submetidos às análises da população segregante. 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) e JoinMap (STAM, 1993).

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 QTL's 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 QTL's, 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 QTL's (XIAO et al., 1996).

KEIM et al. (1990a) mapearam 150 marcadores RFLP numa população de 60 indivíduos F₂ 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 QTL's 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 RIL's.

Os pontos de corte para determinar a significância de um possível QTL dentro de um intervalo testado têm sido obtidos por teste de permutação (DEORGE 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) se encontram 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 se 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. MATERIAL E MÉTODOS

3.1. Material genético

Uma população de 118 RIL's 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 RIL's foram obtidas pela descendência, a partir da geração F_2 , de uma única semente de cada planta (SSD – "single seed descent"). Na geração F_6 foram abertas linhas, que constituíram o material genético utilizado neste trabalho.

Essas RIL's foram plantadas em duas localidades diferentes: Viçosa, MG, e Cascavel, PR. Os experimentos foram plantados no delineamento de famílias com testemunhas intercaladas, sendo as testemunhas os progenitores. Para as famílias, não houve repetição em cada local, ao passo que para as testemunhas as repetições foram plantadas a cada 10 linhas de famílias.

Foram coletadas folhas de cinco plantas de cada família, no experimento de Viçosa.

3.2. Análise do conteúdo de proteína

O teor de proteínas das sementes de cada família foi determinado, tomando-se uma amostra de cinco plantas por família. Essa determinação foi pelo método Kjeldahl modificado, segundo as normas analíticas do INSTITUTO ADOLFO LUTZ (1985), que requer um mínimo de três gramas (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.

3.3. Análise de variância

3.3.1. Análise de variância individual

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

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

em que

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

μ = média geral das testemunhas;

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

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

Nas famílias, foi adotado o seguinte modelo:

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

em que

y_i = valor da característica na i-ésima família;

μ_f = média geral das famílias;

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

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

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

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

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

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

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

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

$$SQR = SQT_o - SQTe$$

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

A soma de quadrados para famílias foi obtida pela expressão

$$SQF = \sum_{i=1}^f y_i^2 - \frac{1}{f} \left(\sum_{i=1}^f y_i \right)^2$$

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

$$h^2 = \frac{\hat{S}_g^2}{QMF}$$

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

3.3.2. Análise de variância conjunta

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

uma soma de quadrados de grupos (SQGr) que testa o contraste entre as médias das famílias e das testemunhas avaliadas.

Nas testemunhas, o modelo adotado foi o seguinte:

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

em que

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

μ = média geral das testemunhas;

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

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

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

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

Nas famílias foi adotado o seguinte modelo:

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

em que

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

μ = média geral das famílias;

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

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

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

e_{ik} = erro aleatório que incide sobre as famílias, sendo $e_{ik} = e_{ijk} \sim \text{NID}(0, \sigma^2)$.

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

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

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

* Obtido por diferença.

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

As somas de quadrados foram obtidas pelas seguintes expressões:

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

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

em que

Y... = total geral; e

N = rat+af ou número total de parcelas.

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

$$SQA = \left(\frac{1}{r} \sum_{k=1}^a \frac{Y_{..k}^2}{t} + \sum_{k=1}^a \frac{Z_k^2}{f} \right) - C \quad Z_{.k?}$$

em que

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

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

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

sendo $Ct = \frac{Y_{...}^2}{rta}$; $Y_{i..}$ é o total das testemunhas em cada ambiente e

$Y_{...}$, o total das testemunhas.

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

Sendo

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

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

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

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

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

$$SQGr = \frac{Y_{...}^2}{rat} + \frac{Z_{..}^2}{fa} - C$$

$$SQF = \frac{1}{a} \sum_{i=1}^f Z_{i.}^2 - Cf$$

em que $Cf = \frac{Z_{..}^2}{fa}$, sendo $Z_{..}$ o total de famílias nos vários ambientes e $Z_{i.}$, o

total de cada família, somando-se os vários ambientes.

$$SQA^{**} = \frac{1}{f} \sum_{k=1}^a Z_{.k}^2 - Cf$$

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

$$SQF, A^{**} = \sum_{i=1}^f \sum_{k=1}^a Z_{ik}^2 - Cf$$

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

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

O coeficiente de variação experimental foi obtido pela expressão

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

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

$$\hat{S}_g^2 = \frac{QMF - QMR}{a}$$

$$h^2 = \frac{\hat{S}_g^2}{\hat{S}_g^2 + \hat{S}^2}$$

em que

$$\hat{S} = QMR$$

sendo essas estimativas obtidas com base nas médias das famílias.

3.4. Extração de DNA

Para a extração do DNA, cinco folhas de cinco plantas de cada família foram coletadas, embaladas em papel-alumínio, congeladas em N_2 líquido e armazenadas a $-80^\circ C$. 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 50mM, pH 8,0, EDTA 50 mM, pH 8,0, NaCl 0,7 M, β-mercaptoetanol 1% v/v e CTBA 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%.

3.5. Condições de amplificação e de eletroforese

As reações de amplificação de microssatélites foram feitas em um volume total de 15 µl, contendo 100 mM/500 mM de Tris-HCl, pH 8,3, 50 mM de KCl, 2 mM de MgCl₂, 100 µM de cada um dos desoxinucleotídeos (dATP, dTTP, dGTP e dCTP), 0,6 µM de cada *primer* (“Reserach 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, finalmente, de 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 µL de brometo de etídio 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).

3.6. Construção do mapa de ligação genética

Os marcadores microssatélites foram codificados como homozigotos para os alelos vindos de Garimpo (A), homozigotos para os alelos vindos de BARC-8 (B) e falhas (–) das bandas amplificadas em cada um dos 118 indivíduos da população F₆. Como os dados foram codificados para RIL's, heterozigotos eventuais foram descartados. A segregação de cada fragmento polimórfico foi testada na população por meio do teste qui-quadrado. Apenas os fragmentos cuja segregação se ajustou à proporção esperada 1:1 (P < 0,05) foram utilizados nas análises posteriores.

O mapeamento dos microssatélites e o estabelecimento dos grupos de ligação foram feitos com a utilização do programa MAPMARKER (LINCOLN et al., 1993). Os marcadores foram agrupados, utilizando-se LOD mínimo de 4,0 e frequência máxima de recombinação de 0,30. A distância de mapeamento de Kosambi foi utilizada para a conversão da frequência de recombinação para centiMorgans (cM).

3.7. Identificação e mapeamento de QTL's

Todos os marcadores com segregação mendeliana 1:1 foram avaliados para determinar associações com a característica “teor de proteína”, utilizando-se um modelo de regressão simples. Os marcadores associados significativamente com o teor de proteína (P < 0,05) foram analisados por modelos de regressão múltipla, utilizando-se o procedimento de eliminação

“stepwise”. Para essa análise, os dados foram codificados como -1 (alelos vindos de Garimpo) e 1 (alelos vindos de BARC-8). Os dados perdidos (-) foram codificados por valores entre -1 e 1, com o auxílio dos marcadores próximos, no grupo de ligação. A análise de regressão foi realizada com o auxílio do programa Jump (SAS Institute Inc., Cary, NC, EUA). Assim, foi determinada a proporção da variância fenotípica explicada por marcador e pelo conjunto de marcadores que foram significativos na regressão múltipla a $P < 0,05$.

As informações do mapa de ligação foram utilizadas para mapear os QTL's ao longo dos grupos de ligação por meio do mapeamento por intervalo composto (ZENG, 1993 e 1994; JANSEN, 1993), empregando-se o programa QTL Cartographer (BASTEN et al., 1994 e 1999). Esse tipo de metodologia diferencia-se do mapeamento por intervalo clássico, por isolar o efeito de outros QTL's ligados ao intervalo. O mapeamento por intervalo composto é um híbrido entre o mapeamento por intervalos e a técnica de regressão múltipla, sendo adotado o seguinte modelo:

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

em que

y_j = valor da característica no j-ésimo indivíduo;

β_0 = média do modelo;

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

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

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

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

ε_i = variável aleatória.

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

Grupo	Genótipo do marcador		Nº de indivíduos	x^*
	i	i+1		
1	+	+	n1	1
2	+	-	n2	{ 1 com probabilidade 1-p -1 com probabilidade p
3	-	+	n3	{ 1 com probabilidade p -1 com probabilidade 1-p
4	-	-	n4	-1

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

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

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

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

A significância de um QTL é atribuída pela razão de verossimilhança entre a hipótese de existência do QTL e a hipótese de que nenhum QTL existe em determinada posição do mapa de ligação. A razão de verossimilhança possui aproximação com o χ^2 pela seguinte expressão:

$$-2\ln \Lambda \cong \chi^2 \alpha$$

em que Λ representa a razão de verossimilhança e α , o nível de significância para o χ^2 .

A razão de verossimilhança é representada, na maioria das vezes, por seu logaritmo natural (LOD). O ponto de corte do LOD para considerar a

presença de um QTL pelo método de mapeamento por intervalo composto foi obtido a partir da seguinte expressão:

$$\text{LOD} = \chi^2_{(2)}\alpha/2\ln(10).$$

No mapeamento por intervalo composto, o χ^2 possui dois graus de liberdade, pois o denominador da expressão de verossimilhança possui dois parâmetros a mais do que o numerador.

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):

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

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

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 II. Por esse motivo, foi utilizado um nível de significância de 10% para o mapa, que leva a um nível de significância de 0,00656% para cada grupo de ligação, considerando-se 16 grupos.

4. RESULTADOS E DISCUSSÃO

4.1. Detecção de polimorfismos e seleção dos *primers* de microssatélites

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.

Foram testados um total de 567 pares de *primers* de microssatélites (“Research Genetics”, Huntsville, AL, EUA). Desse total, 106 exibiram polimorfismo nos progenitores, mas somente 66 apresentaram fragmentos polimórficos na população segregante (Figura 1). Nesta figura, pode-se observar que a família 13 apresenta tanto o alelo de Garimpo quanto o de BARC-8, representando um exemplo de heterozigose. Dos 40 restantes, 27 não segregaram na população e 13 não tiveram boa qualidade de amplificação, o que não permitiu a análise dos fragmentos gerados.

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 RIL's foi a variedade Garimpo, e o genótipo utilizado para se detectarem os polimorfismos entre os progenitores foi a variedade Garimpo RCH (resistente

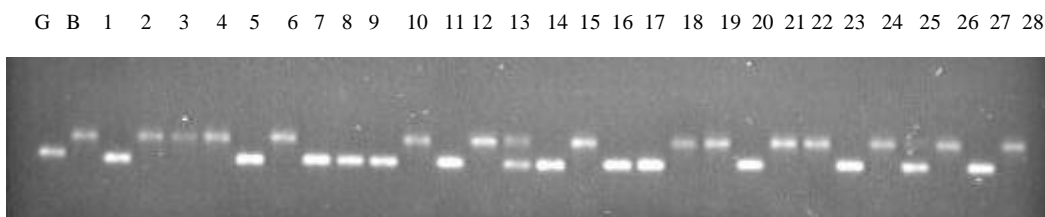


Figura 1 – Gel de agarose 3% fotografado sob luz ultravioleta apresentando fragmentos polimórficos amplificados, utilizando-se *primers* de microssatélites nos pais Garimpo (G) e BARC-8 (B) e em parte da população segregante (1 a 28).

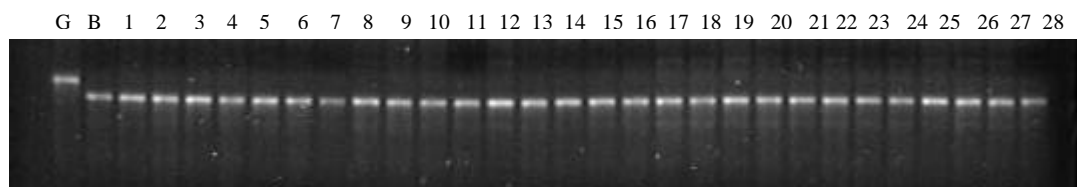


Figura 2 – Gel de agarose 3% fotografado sob luz ultravioleta apresentando fragmentos amplificados, utilizando-se *primers* de microssatélites polimórficos nos progenitores Garimpo (G) e BARC-8 (B) e que não segregaram na população parcialmente representada (1 a 28).

ao cancro-da-haste). A variedade Garimpo RCH foi obtida através de retrocruzamentos tendo como progenitor recorrente a variedade Garimpo. O número de retrocruzamentos efetuados pode não ter sido suficiente para a completa recuperação do genoma da variedade Garimpo na variedade Garimpo RCH, podendo haver divergência genética entre esses dois genótipos. Propõe-se que, numa próxima etapa, sejam analisadas amostras de DNA de plantas dos genótipos Garimpo e Garimpo RCH com esses 27 *primers* de microssatélites, com a finalidade de concluir a respeito dessa suposição. Outra suposição é a de que a variedade Garimpo utilizada não seja de uma linhagem geneticamente pura, o que também poderia contribuir com esse tipo de comportamento entre os marcadores.

A segregação dos marcadores foi testada por meio do teste qui-quadrado ($P < 0,05$), e, dos 66 *primers*, 65 apresentaram fragmentos

polimórficos, que segregaram na proporção mendeliana esperada de 1:1 após descartados os locos heterozigotos (Quadro 4). Em uma população F₆, esperase encontrar 1,56% 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 RIL's, 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 QTL's 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) selecionaram um total de somente 40 *primers* SSR, que apresentaram um único produto de amplificação e polimórficos entre os progenitores para serem utilizados no mapeamento de uma população de soja constituída de 60 plantas F₂ obtidas do cruzamento entre NIL's (linhagens quase isogênicas) dos cultivares Clark e Harosoy.

4.2. Mapa de ligação genética

A primeira etapa do processo foi a definição dos parâmetros de LOD e máxima freqüência de recombinação (r) para o agrupamento dos marcadores. Foram adotados os valores de LOD = 5,0 e de $r = 0,40$, 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 16 grupos de ligação pouco saturados, contendo 43 marcadores, e mais 23 marcas não-ligadas, como pode ser observado na Figura 3.

Quadro 4 – Teste qui-quadrado (χ^2) para a segregação dos marcadores. A letra A representa o alelo presente em Garimpo e a B, alelo presente em BARC-8

<i>Primer</i>	Alelo			χ^2	P%	<i>Primer</i>	Alelo			χ^2	P%
	A	B	Total				A	B	Total		
Sat-117	45	66	111	0,99	31,89	Satt300	59	56	115	0,02	88,88
Satt144	47	63	110	0,58	44,56	Satt304	59	53	112	0,08	77,68
Satt055	48	63	111	0,51	47,65	Satt325	62	54	116	0,14	71,03
Satt186	55	53	108	0,01	92,32	Satt333	55	63	118	0,14	71,27
Sat-091	46	67	113	0,98	32,33	Satt345	66	52	118	0,42	51,93
Sat-135	67	48	115	0,78	37,57	Satt369	65	51	116	0,42	51,57
Sat-105	55	60	115	0,05	81,57	Satt318	57	60	117	0,02	88,97
Sat-084	43	70	113	1,61	20,41	Satt384	54	56	110	0,01	92,40
Sat-085	55	59	114	0,04	85,14	Satt353	58	56	114	0,01	92,54
Sat-112	54	53	107	0,00	96,14	Satt377	57	58	115	0,00	96,28
Satt190	54	56	110	0,01	92,40	Satt399	55	61	116	0,08	78,06
Satt164	64	48	112	0,57	44,97	Satt422	55	61	116	0,08	78,06
Satt167	48	62	110	0,45	50,45	Satt441	50	64	114	0,43	51,21
Satt183	55	61	116	0,08	78,06	Satt449	56	59	115	0,02	88,88
Satt153	68	46	114	1,06	30,29	Satt462	53	62	115	0,18	67,48
Satt146	46	61	107	0,53	46,84	Satt468	59	58	117	0,00	96,31
Satt181	55	55	110	0	100	Satt475	51	65	116	0,42	51,57
Satt184	58	47	105	0,29	59,14	Satt476	59	53	112	0,08	77,68
Satt012	63	50	113	0,37	54,09	Satt495	59	57	116	0,01	92,60
Satt129	55	62	117	0,10	74,63	Satt506	59	59	118	0,00	100
Satt123	60	42	102	0,79	37,29	Satt514	49	63	112	0,44	50,83
Satt147	52	57	109	0,06	81,08	Satt520	61	54	115	0,11	74,41
Satt139	53	58	111	0,06	81,24	Satt522	55	61	116	0,08	78,06
Satt193	62	52	114	0,22	63,96	Satt523	51	60	111	0,18	66,93
Satt199	57	54	111	0,02	88,68	Satt526	54	62	116	0,14	71,03
Satt216	51	56	107	0,06	80,90	Satt527	52	58	110	0,08	77,48
Satt226	53	52	105	0,00	96,11	Satt528	58	57	115	0,00	96,28
Satt232	53	52	105	0,00	96,11	Satt529	59	59	118	0,00	100
Satt241	59	41	100	0,81	36,81	Satt531	68	49	117	0,77	37,98
Satt276	53	55	108	0,01	92,33	Satt538	61	57	118	0,03	85,39
Satt277	67	47	114	0,88	34,90	Satt549	52	65	117	0,36	59,79
Satt281	56	54	110	0,01	92,40	Satt594	63	55	118	0,14	71,27
Satt303	49	57	106	0,15	69,76						

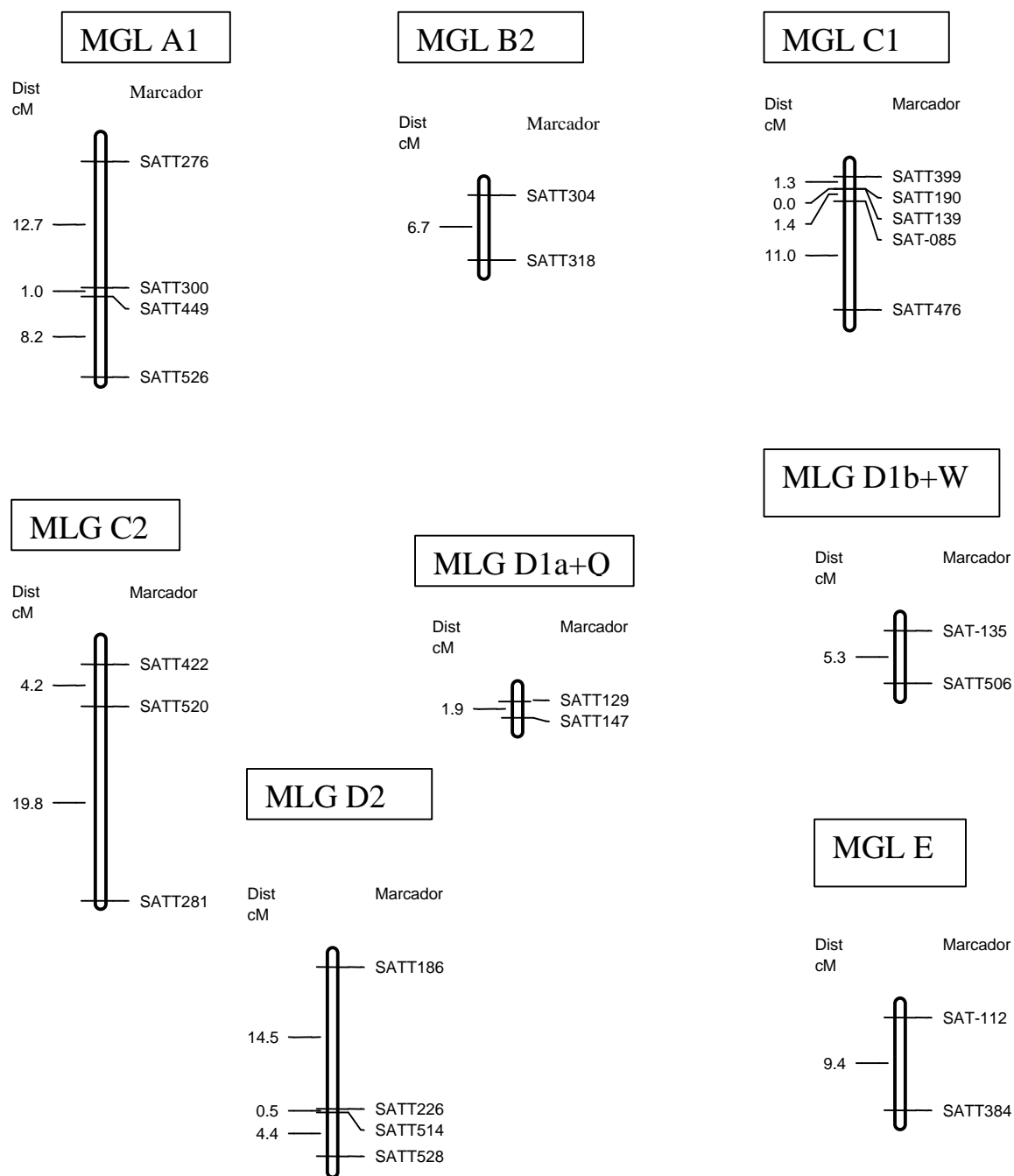
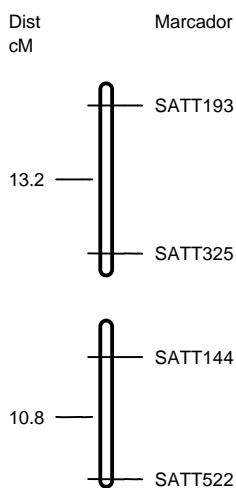
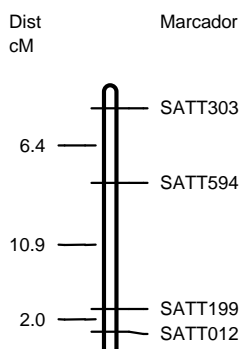


Figura 3 – Mapa genético com 16 grupos de ligação definidos, utilizando-se 118 RIL's obtidas do cruzamento entre as variedades BARC8 x Garimpo. Os grupos de ligação foram obtidos, adotando-se os valores de LOD = 5,0 e $r = 0,40$. Os valores à esquerda representam as distâncias entre os marcadores (em cM), e à direita estão os nomes dos marcadores. A nomenclatura utilizada para os grupos de ligação é a mesma adotada por CREGAN et al. (1999). As letras entre parênteses para os marcadores não-ligados indicam os grupos onde estes marcadores estão mapeados no mapa integrado de CREGAN et al. (1999).

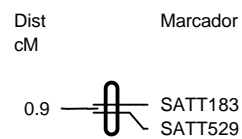
MGL F



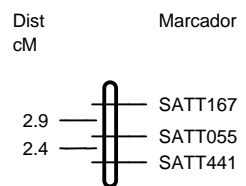
MGL G



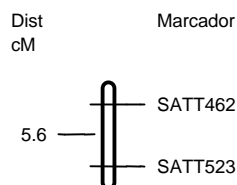
MGL J



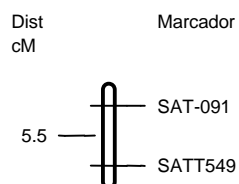
MGL K



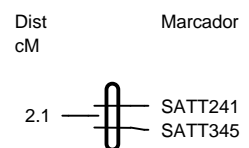
MGL L



MGL N



MGL O



Marcadores não-ligados: Sat_117(G), Sat_105(I), Sat_084(N), Satt164(C1), Satt153(O), Satt146(F), Satt181(H), Satt184(D₁a), Satt123(O), Satt216(W), Satt232(L), Satt277(C₂), Satt333(A₂), Satt369(E), Satt324(G), Satt353(H), Satt377(A₂), Satt468(D₁a), Satt475(K), Satt495(ñ lig.), Satt527(L), Satt531(D₁a) e Satt538(A₂)

Figura 3, Cont.

Os 16 grupos de ligação apresentaram correlação 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 é, 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 (Figura 4). 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). Esses marcadores poderão ser importantes em programas de mapeamento futuros.

Essa é a primeira etapa do mapeamento desses QTL's que controlam teor de proteína em soja. Devido à pouca variabilidade intra-específica da soja, os grupos de ligação obtidos no presente trabalho estão ainda pouco saturados. Em trabalhos futuros serão utilizadas outras populações RIL's e outros marcadores, e os mapas poderão, então, ser integrados.

4.3. Análise do teor de proteína

O teor de proteína nas 118 famílias apresentou distribuição aproximadamente normal, tanto no experimento de Viçosa quanto no de Cascavel, como pode ser observado na Figura 5, indicando que vários genes estão envolvidos com o controle do teor de proteína em soja. A média do teor de proteínas foi maior no experimento de Cascavel, Figura 5 (B). 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.

As análises de variância de cada local (Quadros 5 e 6) evidenciaram que a precisão experimental foi grande, tanto no experimento de Viçosa quanto no de Cascavel, como pode ser observado pelos coeficientes de variação (3,28% em Viçosa e 2,93% em Cascavel). Tanto as famílias quanto os progenitores apresentaram grande variabilidade quanto ao teor de proteína. A análise de variância conjunta dos dois locais (Quadro 7) indicou que fatores

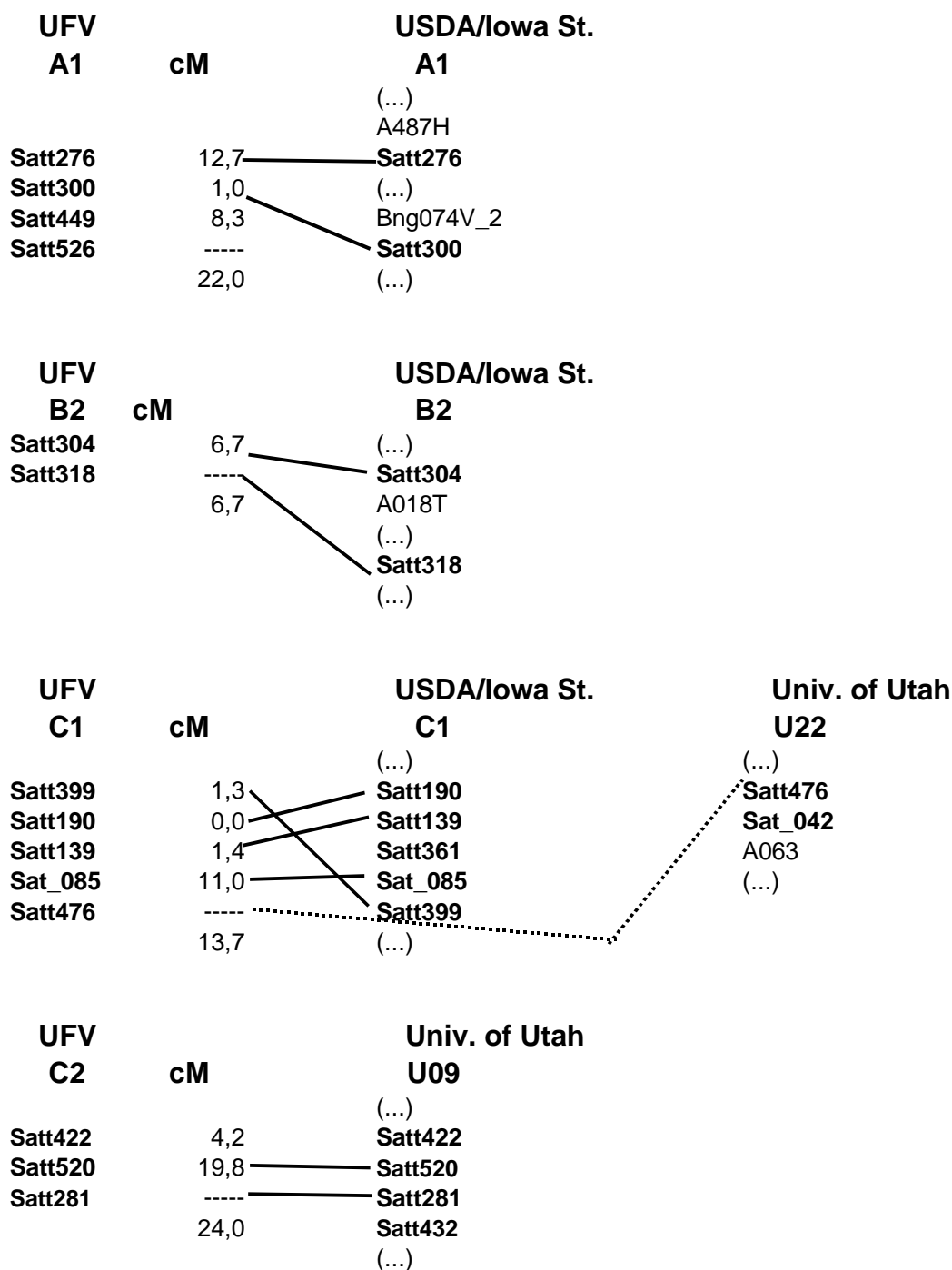


Figura 4 – Integração do mapa obtido neste trabalho (UFV) com o mapa de CREGAN et al.(1999). A distância (cM) entre os marcadores obtidos está representada à direita do marcador. Os mapas integrados por CREGAN estão representados parcialmente, sem os valores das distâncias entre os marcadores e a saturação dos mapas. O símbolo (...) representa as regiões desses mapas, que estão ausentes.

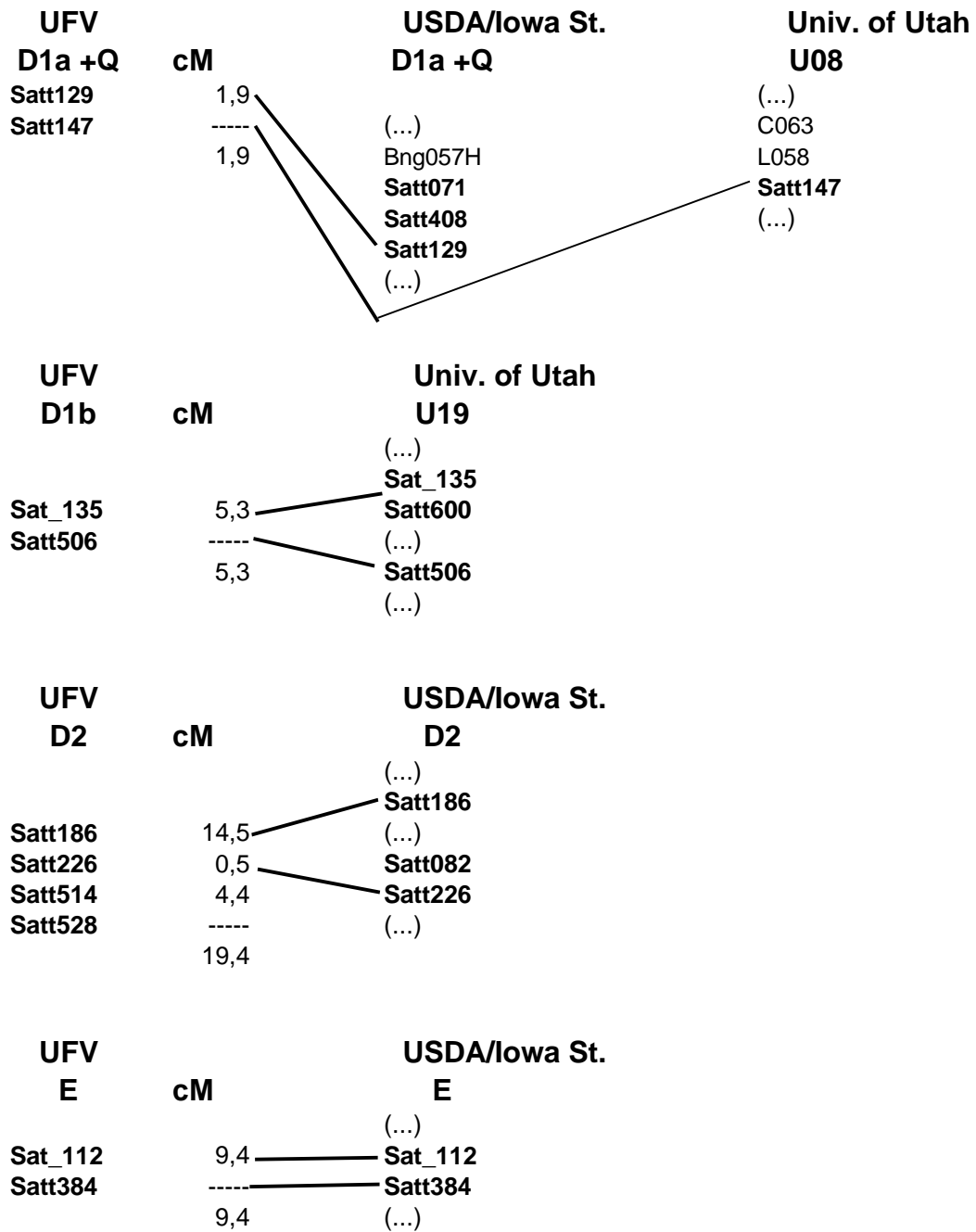


Figura 4, Cont.

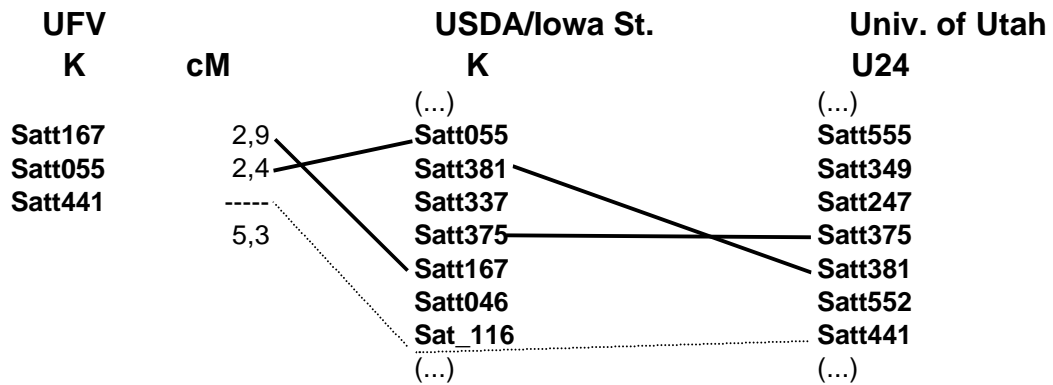
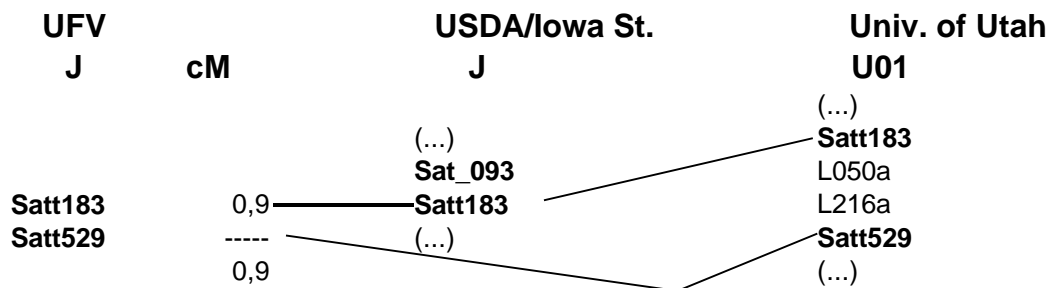
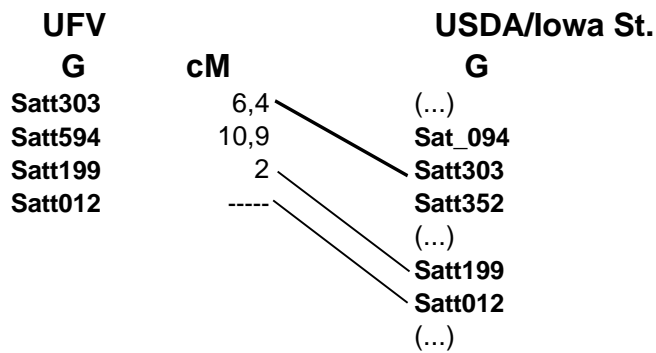
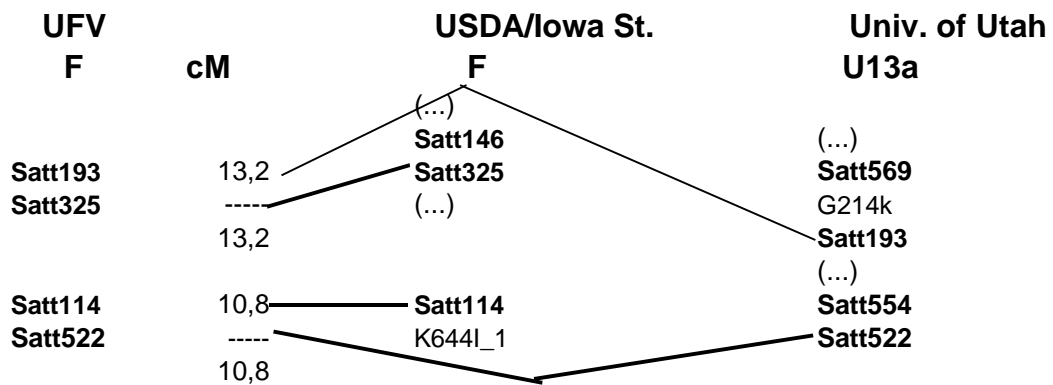


Figura 4, Cont.

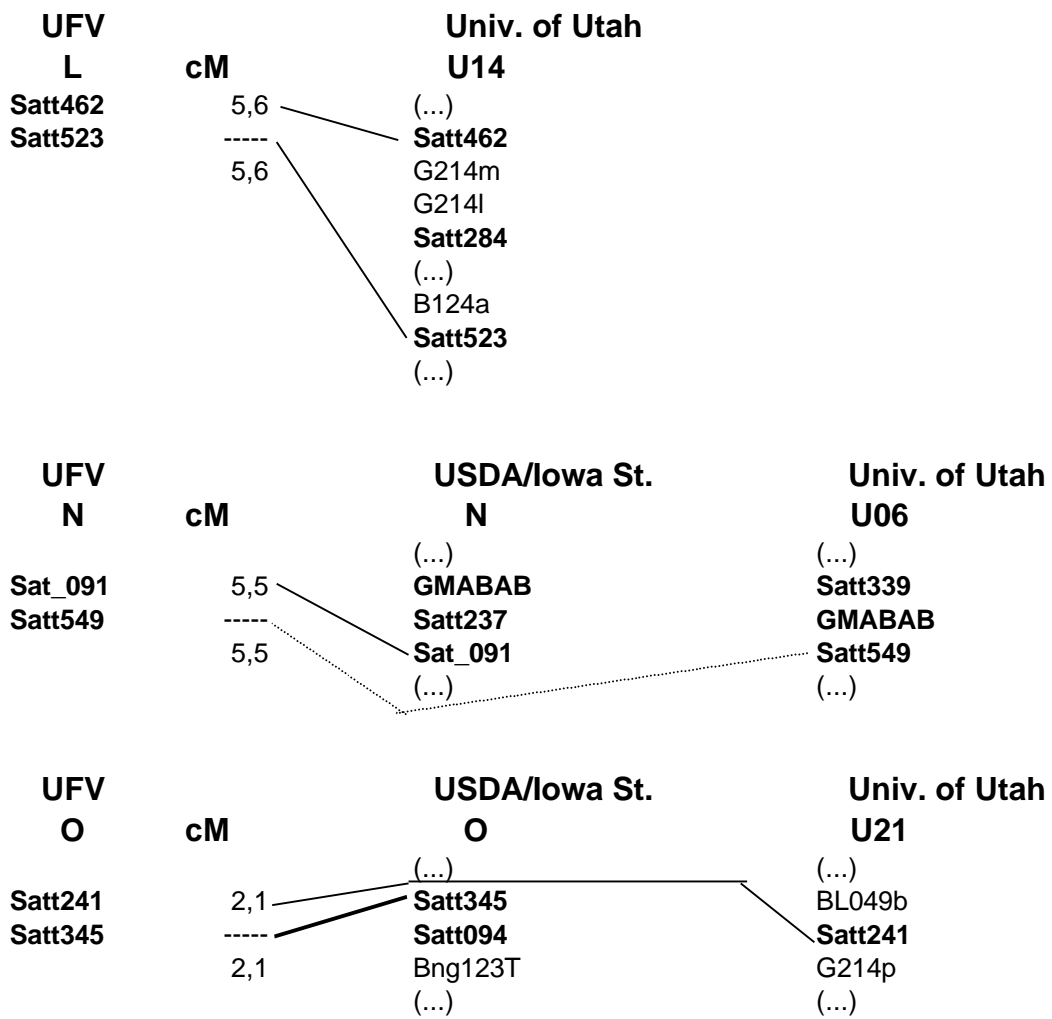


Figura 4, Cont.

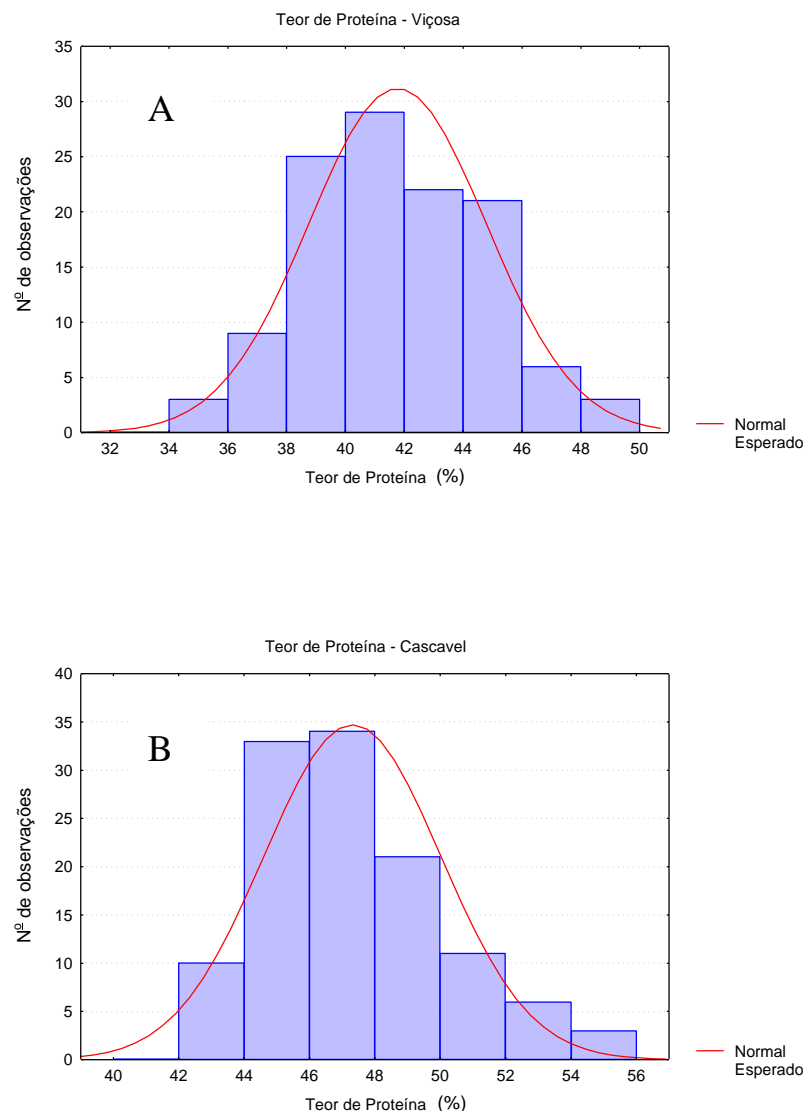


Figura 5 – Distribuição dos teores de proteína em 118 linhas endogâmicas recombinantes (RIL's) cultivadas em Viçosa (A) e Cascavel (B).

Quadro 5 – Análise de variância do experimento de Viçosa e estimativas da variância genética (σ^2g), da herdabilidade (h^2) e do coeficiente de variação (CV)

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

**Significativo a 1% de probabilidade.

$\sigma^2g=6,693$; $h^2=73,47\%$; e $CV = 3,28$.

Quadro 6 – Análise de variância do experimento de Cascavel e estimativas da variância genética (σ^2g), da herdabilidade (h^2) e do coeficiente de variação (CV)

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

**Significativo a 1% de probabilidade.

$\sigma^2g=6,043$; $h^2=82,11\%$; e $CV = 2,93$.

Quadro 7 – Análise de variância conjunta dos experimentos de Viçosa e Cascavel e estimativas da variância genética (σ^2g), da herdabilidade (h^2) e do coeficiente de variação (CV)

FV	GL	SQ	QM	F
Ambiente (A)	1	2.195,980	2.195,980	1.143,14**
Testemunha (Te)	1	8,766	8,766	4,56*
Te X A	1	479,594	479,594	249,66**
Famílias (F)	117	1456,436	12,448	3,095**
F X A	117	470,587	4,022	2,09*
Grupo	1	21,437	21,437	11,159**
Resíduo	29	55,724	1,921	
Total	267			

*Significativo a 5% de probabilidade.

**Significativo a 1% de probabilidade.

$CV=3,10\%$; $\sigma^2g=5,236$; e $h^2=73,16\%$.

ambientais estão envolvidos na regulação da expressão gênica de proteínas de reserva nos dois ambientes (Viçosa e Cascavel), pois a interação genótipo x ambiente foi significativa tanto nas testemunhas quanto nas famílias.

A herdabilidade da característica teor de proteína obtida por componentes de variância foi alta nos dois locais e também na análise conjunta. Isso significa que a maior parte da variação observada é devida a causas genéticas e reflete também a precisão experimental. Altos valores de herdabilidade para o teor de proteína, utilizando-se componentes de variância, também foram obtidos por PIOVESAN (2000), em vários cruzamentos com progenitores contrastantes quanto ao teor de proteína.

4.4. Análise de QTL's do teor de proteína

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

Utilizando as análises de regressão simples, das famílias cultivadas em Cascavel, foram identificados nove marcadores com associação significativa ($P < 0,05$) com o teor de proteína: Sat-085, Satt139, Satt190, Satt399 (pertencentes ao grupo de ligação C1), Satt384 (grupo de ligação E), Satt281, Satt520, Satt422 (grupo de ligação C2) e Sat-105 (não-ligado, pertencente ao grupo de ligação I no mapa de CREGAN et al., 1999), que explicaram cada um, de 3,25% a 6,37% da variação no teor de proteína. Na análise de regressão múltipla, os marcadores Satt190, Satt384, Satt422 e Sat-105 explicaram aproximadamente 23% do total de variação do teor de proteínas (Quadro 8). O menor número de marcadores no modelo de regressão múltipla se deveu ao fato de alguns marcadores estarem ligados ao mesmo QTL e, portanto, tendo o mesmo significado. Já nas famílias de Viçosa foram encontrados 13 marcadores com associação significativa ($P < 0,05$) com o teor de proteína: Satt441, Satt167 (grupo de ligação K), Satt476, Sat-085, Satt139, Satt190, Satt399 (grupo de ligação C1), Satt384 (grupo de ligação E), Satt012, Satt594 (grupo de ligação G), Satt369 e Sat-105 (não-ligados), que explicaram cada um, de 2,92% a 12,43% (Quadro 9). Na regressão múltipla, os marcadores

Quadro 8 – Análises de regressões simples e múltipla entre os marcadores moleculares e o teor de proteína das famílias cultivadas em Cascavel

Marcadores	Grupo de ligação	R ² (%)	R ² (%) ajustado	Regressão simples		Regressão múltipla*	
				F	P > F	F	P > F
Sat-085	C1	4,30	3,45	5,075	0,0262		
Satt139	C1	5,84	5,01	7,014	0,0092		
Satt190	C1	6,54	5,72	7,912	0,0058	9,419	0,0027
Satt399	C1	5,95	5,12	7,148	0,0086		
Satt384	E	7,19	6,37	8,760	0,0038	4,013	0,0055
Satt281	C2	4,10	3,25	4,827	0,0301		
Satt520	C2	4,96	4,12	5,896	0,0168		
Satt422	C2	6,54	5,71	7,908	0,0058	11,613	0,0009
Sat-105	não-ligado	5,52	4,68	6,600	0,0115	8,402	0,0045

*R²=25,75; e *R² ajustado=23,02.

Quadro 9 – Análises de regressões simples e múltipla dos marcadores moleculares e do teor de proteína das famílias cultivadas em Viçosa

Marcadores	Grupo de ligação	R ² (%)	R ² (%) ajustado	Regressão simples		Regressão múltipla*	
				F	P > F	F	P > F
Satt441	K	3,75	2,92	4,515	0,0357		
Satt167	K	4,41	3,59	5,353	0,0224		
Satt476	C1	3,82	2,99	4,610	0,0339		
Sat-085	C1	4,45	3,63	5,401	0,0219		
Satt139	C1	6,14	5,34	7,594	0,0068		
Satt190	C1	6,50	5,69	8,060	0,0053	5,894	0,0168
Satt399	C1	5,76	4,95	7,091	0,0088		
Satt384	E	5,29	4,48	6,483	0,0122	10,255	0,0018
Satt012	G	4,59	3,76	5,577	0,0199	3,848	0,0523
Satt594	G	3,82	2,99	4,608	0,0339		
Satt304	B2	6,27	5,46	7,763	0,0062	6,359	0,0131
Satt369	não-ligado	5,06	4,24	6,187	0,0143	6,587	0,0116
Sat-105	não-ligado	13,18	12,43	17,605	0,0001	14,974	0,0002

*R²=25,75; e *R² ajustado=23,02.

Satt190, Satt384, Satt012, Satt304, Satt369 e Sat-105 explicaram, juntos, aproximadamente 31% do total da variação do teor de proteínas. Na análise de regressão utilizando as médias dos teores de proteína dos dois locais, dois marcadores se mostraram estáveis nos dois ambientes: Satt384 e Sat-105 (Quadro 10).

Quadro 10 – Análise de regressões simples e múltipla dos marcadores moleculares e do teor de proteína das médias das famílias cultivadas em Viçosa e Cascavel

Marcadores	Grupo de ligação	R ² (%)	R ² (%) ajustado	Regressão simples		Regressão múltipla*	
				F	P > F	F	P > F
Satt384	E	9,13	8,34	15,741	0,0001	15,666	0,0001
Sat-105	não-ligado	11,95	11,19	11,65	0,0009	11,608	0,0009

*R²=25,75; e *R² ajustado=23,02.

Os marcadores associados ao teor de proteína 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 F₂ de tomate (*Lycopersicon esculentum*), QTL's para algumas características mostraram-se variáveis através 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 QTL's associados à produção de grãos em milho (*Zea mays* L.).

Alguns grupos de ligação contendo QTL's para teor de proteína foram iguais nos dois locais (grupos C1 e E), indicando que os locos associados a esses grupos de ligação são conservados e, possivelmente, tenham função importante no desenvolvimento normal da planta.

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 QTL's: 1) QTL's estáveis a fatores ambientais (no mínimo um QTL para proteína em cada população, totalizando nove QTL's), detectados em dois ou três anos e na média dos três anos em que foram

conduzidos os estudos; e 2) treze QTL's para proteínas sensíveis a fatores ambientais, que foram identificados nos grupos de ligação A₁, A₂, B₂, C₂, E, G, I, J e L.

4.5. Mapeamento de QTL's

Os modelos lineares simples, como ANOVA e regressão linear, não necessitam das informações dos mapas de ligação, e a distribuição dos valores fenotípicos é analisada 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 e 1994; JANSEN, 1993), com os intuitos de aumentar o poder de detecção de QTL's e melhorar a precisão das estimativas do efeito e da posição deles mesmos. Esse método é baseado na segregação de marcadores adjacentes ao QTL, utilizando-se marcadores em outros intervalos para controlar o "background" genético, ou seja, isolar o efeito de outros possíveis QTL's se estiverem ligados. 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 0,00656% 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.

Empregando-se esse nível de significância, foram encontrados dois QTL's associados a teor de proteína nos grupos de ligação C₁ e C₂ (Figura 6) para o experimento de Cascavel, que explicam, respectivamente, 7,8% e 12,22% da variação do teor de proteínas. Foi encontrado também um possível QTL no grupo de ligação E (experimento de Cascavel) que explicou 10,16% da variação do teor de proteína, que se aproximou do nível de significância, mas não foi significativo. Dentre os QTL's estáveis encontrados por BRUMMER et al. (1997), um pertence ao mesmo grupo de ligação (C1) do QTL identificado

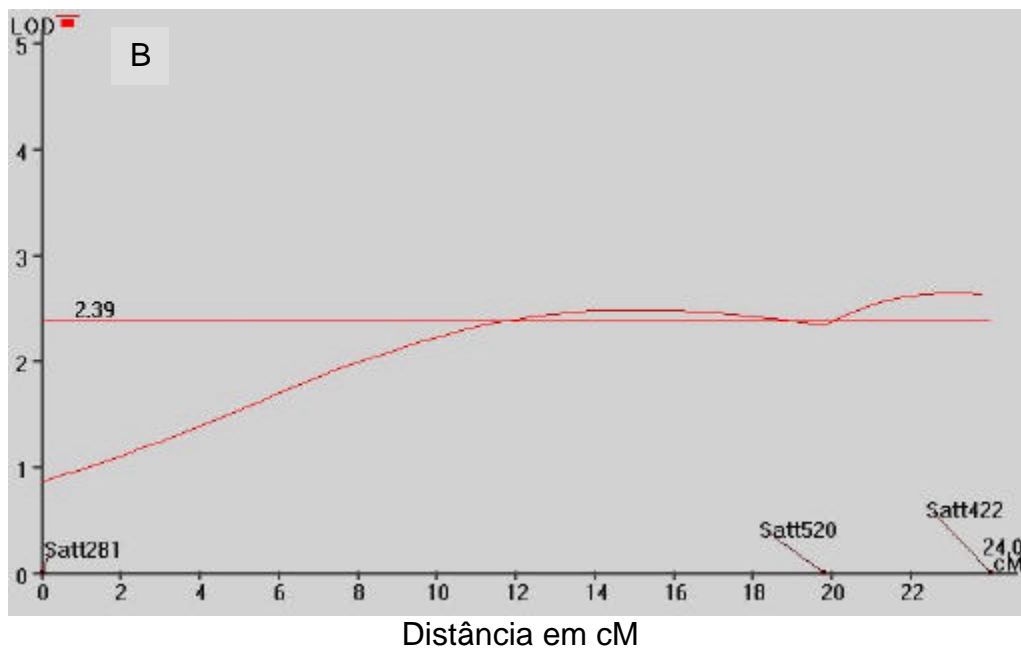
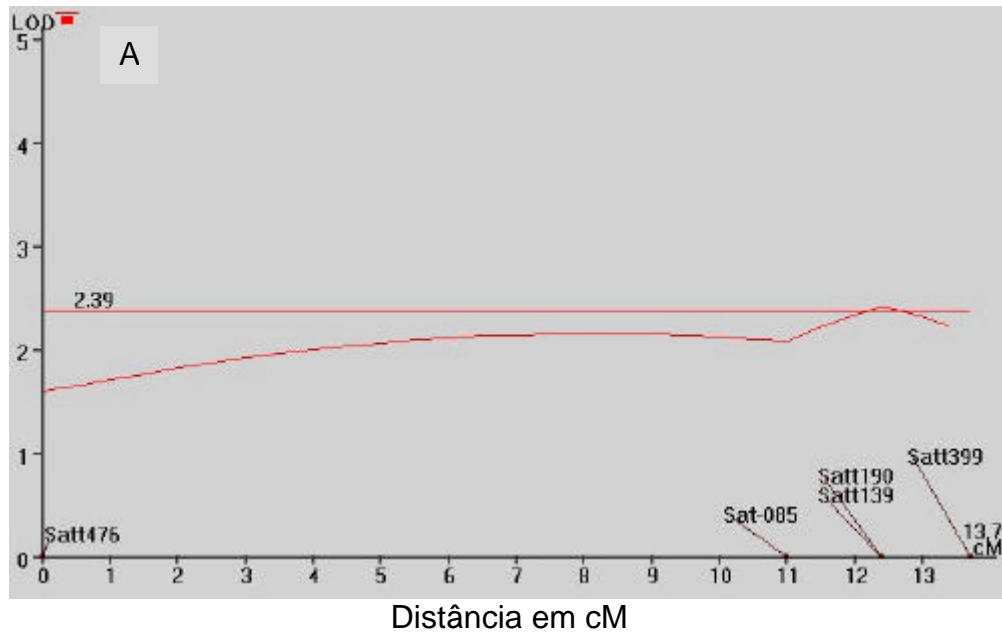


Figura 6 – QTL's associados a teor de proteína nos grupos de ligação C1 (A) e C2 (B), encontrados no experimento de Cascavel. Foi utilizado um nível de significância de 10% para o mapa completo, que resultou em 0,00656% para cada cromossomo.

neste trabalho e foi estável na análise de regressão feita por eles ($P < 0,05$), explicando 19,5% e 15,3% da variação do teor de proteína nas populações 5 e 11, respectivamente. Esses autores encontraram também outro QTL estável no grupo de ligação E, que explicou 12,3% da variação do teor de proteína.

Nas RIL's cultivadas em Viçosa foram encontrados possíveis QTL's nos grupos de ligação B2, C1, G e K, mas, devido ao rigor do teste, não foram significativos (Figura 7). O aumento na saturação do mapa deverá melhorar essas estimativas.

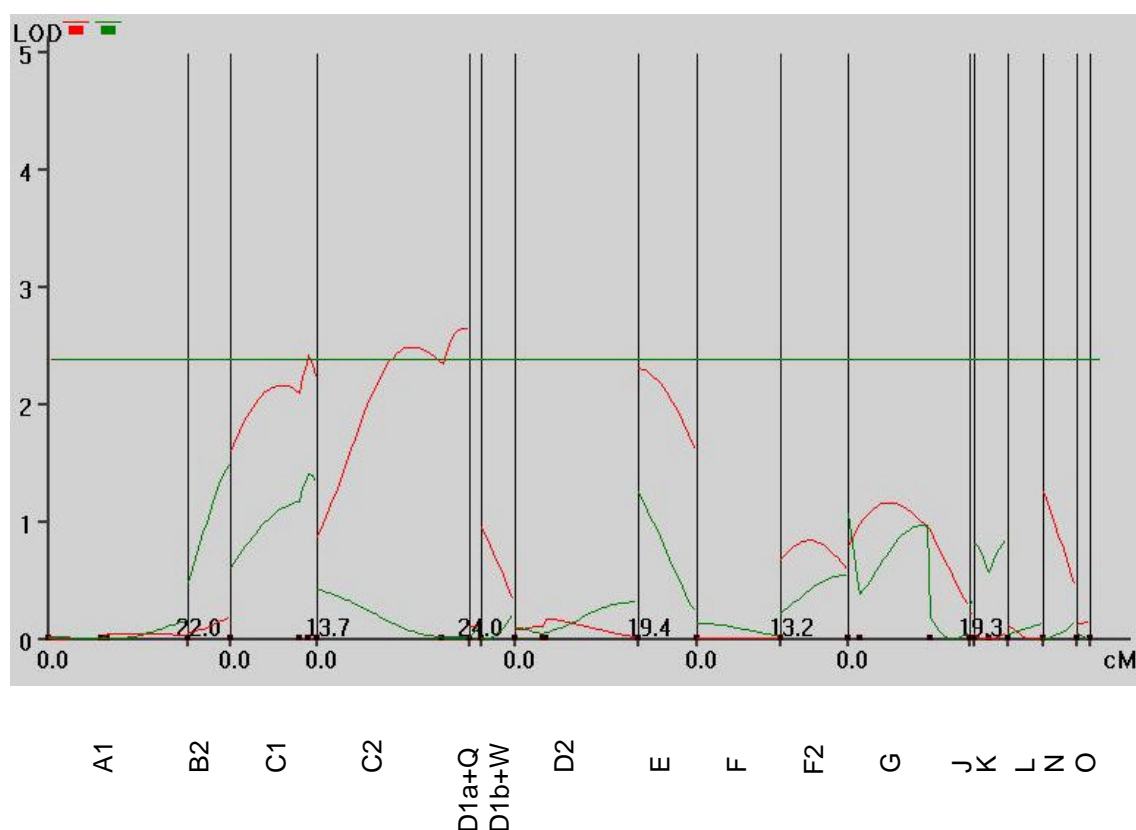


Figura 7 – Picos pequenos que foram encontrados nos grupos de ligação B2, C1, E, G, K (Viçosa - verde) e E (Cascavel - vermelho), mas que, devido ao rigor do teste, não foram significativos.

Os resultados obtidos, tanto pelas análises de variância e regressão quanto pelo mapeamento por intervalo composto, indicaram que o teor de proteína é controlado por dois tipos de QTL's: estáveis (expressos em vários ambientes) e ambiente-específicos (expressos apenas em determinado ambiente). Esse resultado é muito útil aos programas de melhoramento, pois as variedades de soja com alto teor de proteína devem possuir, preferencialmente, aqueles QTL's estáveis nos vários ambientes em que são cultivadas. Os marcadores identificados podem ser utilizados na seleção assistida, em programas de melhoramento que visem aumentar o teor de proteína nos grãos de soja.

Os resultados apresentados se referem a um ano e a dois locais. Avaliações em mais locais e mais anos devem ser incluídas a esses dados, para confirmar os locos contendo os QTL's estáveis, além de identificar outros. Além disso, a utilização de géis de poliacrilamida para obtenção de maior número de polimorfismos e a inclusão de outras classes de marcadores deverão aumentar a saturação do mapa, melhorando a resolução do mapeamento de QTL's.

5. 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 faz necessário para o desenvolvimento de procedimentos mais eficientes de melhoramento.

No presente trabalho se propuseram identificar marcadores do tipo microssatélites ligados a QTL's que controlam o teor de proteína em soja e dar início à construção de um mapa genético para soja, utilizando genótipos tropicais. Para isso, utilizaram-se 118 RIL's (linhagens endogâmicas recombinantes), que se encontravam na geração F_6 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 RIL's foram cultivadas em dois ambientes distintos: Viçosa, MG, e Cascavel, PR. O mapeamento genético foi realizado com marcadores de microssatélites.

Os resultados obtidos neste trabalho permitiram apontar algumas conclusões:

1. Os teores de proteína obtidos pelo método de Kjeldahl para as 118 famílias apresentaram distribuição aproximadamente normal, tanto no experimento de Viçosa quanto no de Cascavel.
2. Fatores ambientais estão envolvidos na regulação da expressão de genes de proteínas de reserva nos dois ambientes (Viçosa e Cascavel), pois a

interação genótipo x ambiente foi significativa tanto para testemunhas quanto para famílias.

3. A herdabilidade da característica teor de proteína foi alta nos dois locais e também na análise conjunta. Isso significa que a maior parte da variação observada foi devida a causas genéticas e refletiu também a precisão experimental.
4. Foram identificados nove marcadores com associação significativa ($P < 0,05$) ao teor de proteína para as famílias cultivadas em Cascavel, PR: Sat-085, Satt139, Satt190, Satt399 (pertencentes ao grupo de ligação C1), Satt384 (grupo de ligação E), Satt281, Satt520, Satt422 (grupo de ligação C2) e Sat-105 (não-ligado, pertencente ao grupo de ligação I no mapa de CREGAN et al., 1999), os quais explicaram, cada um, de 3,25% a 6,37% da variação no teor de proteína.
5. Foram encontrados 13 marcadores com associação significativa ($P < 0,05$) com o teor de proteína, para as famílias cultivadas em Viçosa, MG: Satt441, Satt167 (grupo de ligação K), Satt476, Sat-085, Satt139, Satt190, Satt399 (grupo de ligação C1), Satt384 (grupo de ligação E), Satt012, Satt594 (grupo de ligação G), Satt369 e Sat-105 (não-ligados), os quais explicaram, cada um, de 2,92% a 12,43% da variação no teor de proteína.
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. Alguns grupos de ligação contendo QTL's para teor de proteína foram iguais nos dois locais (grupos C1 e E), indicando que os locos associados a esses grupos de ligação são conservados.
8. Utilizando-se um nível de significância de 10% para o mapa, foram encontrados dois QTL's associados ao teor de proteína nos grupos de ligação C1 e C2, no experimento de Cascavel, que explicaram, respectivamente, 11,13% e 12,19% da variação no teor de proteínas.

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.
- 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 Science**, v.35, n.4, p.1439-1445, 1995.
- BASTEN, C.J., WEIR, B.S., ZENG, Z.-B. **QTL cartographer; version 1.13**. Raleigh, NC: Department of Statistics; North Carolina State University, 1999. 132p.
- BASTEN, C.J., WEIR, B.S., ZENG, Z.-B. Zmap—a QTL cartographer. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 5, 1994, Guelph, Ontario, Canadá. **Proceedings ...** Guelph, Ontario, Canadá: The Organizing Committee, 1994. v.22, p.65-66.
- 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.
- 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 Science**, v.37, n.1, p.370-378, 1997.

- 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.
- 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.
- CIZLICE, 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 Science**, v.33, n.3, p.614-620, 1993.
- COATES, J.B., MEDEIROS, J.S., THANH, V.H., NIELSEN, N.C. Characterization of the subunits of β -conglucinin. **Archives of Biochemistry and Biophysics**, 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 Science**, v.39, n.5, p.1464-1490, 1999.
- DEORGE, 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.
- DOYLE, J.J., DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**, v.12, n.1, p.13-15, 1990.
- DUDLEY, J.M., MOLL, R.H. Interpretation and use of estimates of heritability and genetics variances in plant breeding. **Crop Science**, Madison, v.9, n.3, p.257-262, 1969.
- ERICKSON, L.R., VOLDENG, H.D., BEVERSDORF, W.D. Early generations selection for protein in *Glycine max* x *Glycine soja* crosses. **Canadian Journal of Plant Science**, Ottawa, v.6, n.4, p.901-908, 1981.
- FALCONER, D.S. **Introdução à genética quantitativa**. 1ª ed. (Trad. de Silva, M.A.). Viçosa: UFV, 1987. 279p.
- 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**. 2.ed. Brasília: EMBRAPA; CENARGEN, 1996. 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 Physiology**, v.76, n.3, p.840-842, 1984.
- FREY, K.J., HORNER, T. Heritability in standard units. **Agron. J.**, v.49, n1, p.59-62, 1957.
- 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., 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.
- 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.
- HELMS, T.C., ORF, J.H. Protein, oil, and yield of soybean lines selected for increased protein. **Crop Science**, v.38, n.4, p.707-710, 1998.
- HILL, J.E., BREIDENBACH, R.W. Protein of soybean seeds. **Plant Physiology**, v.53, n.5, p.742-746, 1974.
- 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.
- 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., 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 Science**, 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.
- KWON, S.H., TORRIE, J.H. Heritability and interrelationships among traits of two soybean populations. **Crop Science**, 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.
- LINCOLN, S., DALY, M., LADER, E. **Constructing genetic maps with MAPMARKER/EXP 3.0: a tutorial and reference manual**. 2.ed. Cambridge, MA: Whitehead Institute, Technical Report, 1993. Paginação irregular.
- 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.
- LYNCH, M., WALSH, B. **Genetics and analysis of quantitative traits**. Massachusetts: Sinauer Associates, 1998. 980p.
- 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.
- 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. **Archives of Biochemistry and Biophysics**, 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.
- PATERSON, A.H., DAMON, S., HEWITT, J.D., ZAMIR, D., RABINOWITZ, 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 proteico 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.
- 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.2990-309.
- SHOEMAKER, R.C., GUFFY, R.D., LORENZEN, L.L., SPECHT, J.E. Molecular genetic mapping of soybean: map utilization. **Crop Science**, v.32, n.5, p.1091-1098, 1992.

- 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.
- SMITH, J.D., KINMAN, M.L. The use of parent-offspring regression as an estimator of heritability. **Crop Science**, v.5, n.6, p.595-596, 1965.
- 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. **Agricultural and Biological Chemistry**, v.256, n.6, p.8752-8755, 1981.
- 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.
- 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.
- THANH, V.H., SHIBASAKI, K. Beta-conglycinin from soybean proteins. Isolation and immunological and physicochemical properties of the monomeric forms. **Biochimica et Biophysica Acta**, v.490, n.2, p.370-384, 1977.
- TINGEY, S.V., DELTUSO, J.P. Genetic analysis with Random Amplified Polymorphic DNA markers. **Plant Physiology**, 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 Science**, v.35, n.4, p.1036-1041, 1995.

- WILLIAMS, J.G.K., KUBELIK, A.R., LIVAK, K.J., RAFALSAKI, J.A., TINGEY, S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Res.**, v.18, n.13, p.6531-6535, 1990.
- 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.
- 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.