

JOSIANE ISABELA DA SILVA RODRIGUES

**MAPEAMENTO DE QTLs ASSOCIADOS A CONTEÚDO DE PROTEÍNA,
ÓLEO E COMPONENTES DE PRODUÇÃO EM SOJA**

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

**VIÇOSA
MINAS GERAIS – BRASIL
2008**

JOSIANE ISABELA DA SILVA RODRIGUES

**MAPEAMENTO DE QTLs ASSOCIADOS A CONTEÚDO DE PROTEÍNA E
ÓLEO E COMPONENTES DE PRODUÇÃO EM SOJA**

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

APROVADA: 26 de fevereiro de 2008

Prof. Everaldo Gonçalves de Barros
(Co-orientador)

Prof. Cosme Damião Cruz
(Co-orientador)

Prof. Marcia Flores da Silva

Prof. Adésio Ferreira

Prof. Maurilio Alves Moreira
(Orientador)

Aos meus pais Ana Rosa e João, razões da minha vida

Ao meu noivo João Henrique, razão da minha felicidade

Aos meus irmãos, João Henrique e Júlio César, pelos momentos felizes.

AGRADECIMENTOS

À Universidade Federal de Viçosa

Ao Programa de Pós-graduação em Genética e Melhoramento, pela oportunidade.

Ao CNPq, pela concessão da bolsa de estudos.

Ao Professor Maurilio Alves Moreira, pela orientação e apoio.

Ao Professor Everaldo Gonçalves de Barros, pela dedicação.

Ao Professor Cosme Damião Cruz, pelos ensinamentos.

Ao casal Márcia Flores e Adésio Ferreira, pela ajuda imprescindível.

Ao Newton Deniz Piovesan, pela colaboração.

Ao Leandro, Fernanda, Loreta e Julierme, pela grande ajuda.

Ao Fábio Demolinari, pela contribuição.

Aos colegas do laboratório, pelo companheirismo.

Aos meus amigos Ariadney, Rosilene e Jairo, pela forte amizade.

Aos meus pais, João e Ana Rosa, pelo amor e incentivo.

Aos meus avós, Jacira e José Venâncio, pelo apoio.

Ao meu noivo João Henrique, pelo amor e carinho.

A Deus.

SUMÁRIO

RESUMO.....	vi
ABSTRACT.....	viii
1. INTRODUÇÃO.....	1
2. REVISÃO DE LITERATURA.....	3
2.1. Base genética e aspectos genéticos da soja.....	3
2.1.1. Base genética e centro de origem da soja.....	3
2.1.2. Genoma e aspectos genéticos da soja.....	4
2.2. Marcadores moleculares.....	6
2.2.1. Marcadores moleculares no melhoramento de plantas.....	6
2.2.2. Evolução das técnicas moleculares.....	7
2.3. Mapeamento de QTLs associados a conteúdo de proteína e óleo e produtividade em soja.....	10
3. MATERIAL E MÉTODOS.....	15
3.1. Material genético.....	15
3.2. Amplificação e avaliação de marcadores microsatélites.....	15
3.3. Análise de associação marcador - QTL.....	15
4. RESULTADOS E DISCUSSÃO.....	20
4.1. Dados fenotípicos.....	20
4.2. Mapeamento de QTL.....	21
4.2.1. Segregação dos marcadores.....	21
4.2.2. Mapeamento.....	22
4.2.3. Análise de locos associados ao conteúdo de proteína e óleo.....	27
4.2.4. Análise de locos associados a componentes de produção de grãos.....	36
5. RESUMO E CONCLUSÕES.....	45
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	47

RESUMO

RODRIGUES, Josiane Isabela da Silva, M. Se., Universidade Federal de Viçosa, fevereiro de 2008. **Mapeamento de QTLs associados a conteúdo de proteína, óleo e componentes de produção em soja.** Orientador: Maurilio Alves Moreira. Co-orientadores: Everaldo Gonçalves de Barros e Cosme Damião Cruz.

Os programas de melhoramento têm se preocupado, cada vez mais, com o desenvolvimento de variedades produtivas e com altos conteúdos de proteína e óleo. A utilização de marcadores moleculares em estudos de mapeamento genético tem levado a identificação de QTLs envolvidos com o controle genético de diversas características de interesse em soja, sobretudo, produtividade e conteúdo de proteína e óleo do grão. Apesar do número considerável de QTLs mapeados para tais caracteres disponíveis na literatura, existe ainda informação limitada e inconsistente sobre a confirmação desses QTLs. Este trabalho teve como objetivos a identificação de QTLs associados a conteúdo de proteína (PTN) e óleo (OLEO) e componentes de produção em soja (peso de sementes por planta - PRO ; peso de cem sementes - PCS, número de vagens por planta - NVP, número de sementes por planta - NSP). O estudo de mapeamento de QTL foi realizado a partir de 206 indivíduos F_2 obtidos do cruzamento entre a linhagem de soja CS3035PTA276-1-5-2 (alto teor de proteína e baixo teor de óleo) e a variedade UFVS2012 (baixo teor de proteína e alto teor de óleo). Plantas F_3 foram avaliadas fenotipicamente para 11 características da soja em um experimento que constou de três repetições ou blocos, de forma que cada família F_3 foi semeada em três repetições. Os resultados das análises de variância e das estimativas dos parâmetros genéticos das características indicaram a existência de variabilidade genética na população para as onze características ao nível de significância de 1%. Além disso, foi evidenciada a existência de contrastes entre as médias dos genitores para a maioria das características. Com base nesses resultados, foi confirmado o potencial genético da população para o estudo de mapeamento de QTL. Quarenta e oito marcadores microssatélites tiveram sua segregação avaliada nos 206 indivíduos F_2 . Foram obtidos nove grupos de ligação, formados pelo agrupamento de 25 marcadores. As análises de associação de marca simples e mapeamento por intervalo simples e composto foram utilizadas para detectar e mapear regiões genômicas associadas as características em questão. As análises de QTL

foram conduzidas por grupo de ligação separadamente e direcionadas as características do grão e aos componentes de produção, foco principal do trabalho. Foram identificados quatro QTLs associados ao conteúdo de proteína nos grupos de ligação A1, D1a, G e I que explicaram entre 6,24% e 18,94% da variação fenotípica da característica. Três QTLs para conteúdo óleo foram detectados nos grupos A1, I e O que explicaram entre 17,26% e 25,93% da variação da característica. Para produção de grãos foram identificados dois QTLs nos grupos de ligação A1 e D1a que explicaram 12,32% e 9,03% da variação fenotípica, respectivamente. No grupo de ligação A1 foram ainda detectados QTLs associados ao número de vagens por planta e número de sementes por planta. Estes QTLs explicaram 9,43% e 7,19% da variação para estes fenótipos, respectivamente. A população utilizada neste trabalho demonstrou grande potencial para o mapeamento de QTL. E o número considerável de associações significativas detectadas entre marcadores e características pressupõe a existência de outros QTLs que poderão ser caracterizados pela análise de mais marcadores em determinadas regiões.

ABSTRACT

RODRIGUES, Josiane Isabela da Silva, M. Se., Universidade Federal de Viçosa, february, 2008. **QTL mapping associated to protein content, oil and production components in soybean.** Adviser: Maurilio Alves Moreira. Co-Advisers: Everaldo Gonçalves de Barros and Cosme Damião Cruz.

The improvement programs have been concerned, more and more, with the development of productive varieties and with high protein content and oil. The use of molecular markers in studies of genetic mapping has been taking the identification of QTLs involved with the genetic control of several characteristics of interest in soybean, above all, productivity and protein content and oil of the grain. In spite of the considerable number of QTLs mapped for such characters available in the literature, it still exist limited and inconsistent information about the confirmation of these QTLs. This work had as objectives the identification of QTLs associated to protein content (PTN) and oil (OIL) and production components in soybean (weight of seeds for plant - W_S ; weigh of a hundred seeds - W_H , number of nodules for plant - N_N , number of seeds for plant - N_S). The study of QTL mapping was permormed from 206 F_2 individuals obtained of the crossing among the soybean line CS3035PTA276-1-5-2 (high protein and low oil content) and the variety UFVS2012 (low protein and high oil content). F_3 plants were phenotypic evaluated for 11 characteristics of the soybean in an experiment that consisted of three repetitions or blocks, so that each F_3 family was sowed in three repetitions. The results of the variance analyses and of the estimates of the genetic parameters of the characteristics indicated the existence of genetic variability in the population for the eleven characteristics at the level of significance of 1%. Besides, the existence of contrasts among the averages of the genitors was evidenced for most of the characteristics. With base in these results, the genetic potential of the population was confirmed for the study of QTL mapping. Forty eight microsatellite markers had his segregation assessed in the 206 F_2 individuals. Were obtained nine linkage groups, formed by the grouping of 25 markers. The analyses of association of simple mark and simple and composed interval mapping were used to detect and to map genomic areas associated to the characteristics in subject. The analyses of QTL were driven by linkage group separately and addressed to the characteristics of the grain and to the production

components, main focus of the work. Were identified four QTLs associated to the protein content in the linkage groups A1, D1a, G and I that explained among 6,24% and 18,94% of the phenotypic variation of the characteristic. Three QTLs for content oil were detected in the groups A1, I and O what explained among 17,26% and 25,93% of the variation of the characteristic. For production of grains were identified two QTLs in the linkage groups A1 and D1a that explained 12,32% and 9,03% of the phenotypic variation, respectively. In the linkage group A1 were still detected QTLs associated to the number of nodules by plant and number of seeds for plant. These QTLs explained 9,43% and 7,19% of the variation to these phenotypes, respectively. The population used in this work demonstrated great potential for the QTL mapping. And the considerable number of significant associations detected between markers and characteristics presupposes the existence of others QTLs that can be characterized by the analysis of more markers in certain areas.

1. INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill) é o principal produto agrícola de exportação brasileiro, sendo cultivada praticamente em todo território nacional. No período de 2006/2007 foram produzidos 56.316.000 toneladas deste grão, que corresponde a quase metade da produção total de grãos do país (Conab, 2007). O grande interesse no cultivo da soja deve-se à composição típica do grão, que apresenta em torno de 40% de teor protéico e 20% de teor de óleo, além da alta produtividade de seus cultivares.

Além de excelente fonte de proteínas e óleo a importância nutricional da soja é também atribuída à presença de diversos constituintes funcionais que conferem benefícios à saúde e associados à prevenção de algumas doenças. A soja é também rica em minerais e vitaminas do complexo B. Tais fatores contribuem para a elevada demanda mundial no consumo de soja, em torno de 217 milhões de toneladas do grão (<http://www.fas.usda.gov/psd>), e a crescente utilização desta oleaginosa como matéria prima na indústria. Com isso, um complexo agro-industrial foi gerado destinado ao processamento de derivados da soja, tornando-a uma das principais *commodities* do mundo (Silva, 2006).

Ao longo do tempo, os programas de melhoramento da soja de maneira geral têm buscado o desenvolvimento de cultivares mais produtivas e estáveis em ambientes variados. Outro fator de interesse desses programas é o melhoramento das características de qualidade do grão, como aumentos nos teores de proteína e óleo, e alteração no perfil de ácidos graxos da semente.

Desde sua introdução, a soja alcançou grandes ganhos em produtividade por meio de estratégias de melhoramento clássico. Tradicionalmente têm sido utilizadas linhagens parentais produtivas, com características agronômicas de interesse, para a obtenção de novas cultivares com alta produtividade (Reyna e Sneller, 2001). Esse processo ao longo do tempo, no entanto, parece ter estreitado a diversidade genética das populações de soja, dificultando cada vez mais a obtenção de ganhos adicionais em produtividade. Estimativas mostram, por exemplo, que os aumentos alcançados na produtividade da soja nos EUA têm sido muito pequenos, aproximadamente 30 Kg ha⁻¹ ano⁻¹ ou em torno de 1% ao ano (Wilcox, 2001). De maneira similar, algumas dificuldades têm também limitado a obtenção de aumentos nos conteúdos de proteína e óleo no grão. Embora as herdabilidades (sentido restrito) para as duas características sejam altas, geralmente maiores que 50%, a variação ambiental prejudica

consideravelmente a eficiência de seleção para as mesmas (Brummer *et al.*, 1997). Outras dificuldades encontradas constituem a correlação negativa existente entre conteúdo de proteína e produtividade, bem como, a correlação negativa entre os conteúdos de proteína e óleo no grão.

O controle genético das características de interesse em soja envolve geralmente vários genes e tem interação significativa com o ambiente. A utilização de marcadores moleculares tem levado a identificação e localização de inúmeros locos para tais características. Neste sentido, os marcadores moleculares apresentam-se como uma poderosa ferramenta para o melhoramento de plantas, aumentando potencialmente sua eficiência por meio de estratégias de seleção assistida por marcadores (SAM).

Através da SAM, torna-se possível a seleção indireta da característica de interesse, com base no genótipo de plantas em estágio juvenil, a partir de gerações precoces. A partir de populações segregantes, os alelos desfavoráveis podem ser eliminados ou bastante reduzidos durante os estágios iniciais de desenvolvimento das plantas, conduzindo-se a seleção no campo de reduzido número de plantas maduras. Outra aplicação dos marcadores moleculares é a possibilidade de introgressão de alelos favoráveis a partir de germoplasma exótico ou não adaptado nos *backgrounds* genéticos das cultivares comerciais (Concibido *et al.*, 2003). A procura por QTLs para produtividade em germoplasma exótico não é um conceito novo em soja. Um grande número de pesquisadores tem reconhecido *Glycine soja* como fonte de genes úteis para a soja (Wang *et al.*, 2004).

Nos últimos tempos, um grande esforço vem sendo feito para obtenção de um mapa de ligação saturado para a soja e para a identificação e validação de QTLs associados a características de interesse em soja, tais como produção de grãos e conteúdo de proteína e óleo na semente. E o grande número de trabalhos envolvendo mapeamento molecular para tais características demonstra o interesse na utilização dos marcadores moleculares no melhoramento genético da soja.

Este trabalho teve como objetivos a identificação de QTLs associados a conteúdo de proteína e óleo e componentes de produção de grãos em soja. Para isso, foi utilizada uma população F₂ de 206 indivíduos derivada do cruzamento entre a linhagem CS3032PTA276 (alto teor de proteína e baixo teor de óleo) e a variedade UFVS2012 (baixo teor de proteína e alto teor de óleo). A genotipagem dos marcadores foi realizada a partir dos indivíduos F₂, enquanto a fenotipagem realizada em famílias F₃.

2. REVISÃO DE LITERATURA

2.1. Base genética e aspectos genéticos da soja

2.1.1. Base genética e centro de origem da soja

A soja cultivada hoje é muito diferente dos seus ancestrais, que eram plantas rasteiras que se desenvolviam na costa leste da Ásia. Sua evolução começou com o aparecimento de plantas derivadas de cruzamentos naturais entre duas espécies de soja selvagens que foram domesticadas na China. O centro de origem da soja encontra-se na região centro-sul da China, enquanto a Manchúria, região onde a soja foi domesticada, constitui seu centro secundário (Xu *et al.*, 1989). Apesar de ser consumida pela civilização oriental por milhares de anos, só foi introduzida na Europa no final do século XV.

No Brasil, a soja foi introduzida na Bahia em 1982 e em 1890 foi testada como produtora de forragem em Pelotas, RS. O primeiro registro do cultivo comercial no Brasil ocorreu em 1914 no município de Santa Rosa, RS. Mas somente a partir dos anos 40, o cultivo da soja adquiriu alguma importância econômica. E em 1960, a soja estabeleceu-se como cultura economicamente importante para o país (Hasse, 1996).

Atualmente, o Brasil é o segundo maior produtor mundial de soja. Na safra 2006/2007, a cultura ocupou uma área de 20,687 milhões de hectares, o que totalizou uma produção de 58,4 milhões de toneladas. A produtividade média da soja brasileira é de 2823 kg ha⁻¹, chegando a alcançar cerca de 3000 kg ha⁻¹ no estado de Mato Grosso, o maior produtor brasileiro de soja. Além disso, a soja tem uma importante participação nas exportações brasileiras, tendo movimentado no ano de 2006 US\$ 9,3 bilhões, o que representou 6,77% do total exportado (Embrapa, 2008).

Diversos programas de melhoramento contribuíram para o desenvolvimento de cultivares de alto rendimento e adaptadas às diferentes condições agroclimáticas do país. De acordo com Paludzysz Filho *et al.* (1993), inicialmente foram introduzidas no país linhagens desenvolvidas no sul dos EUA, com o posterior desenvolvimento de cultivares mais adaptadas.

Estimativas sobre a variabilidade genética da cultura têm destacado que o germoplasma brasileiro provém de base genética restrita, originado a partir de poucas linhagens ancestrais (Priolli *et al.*, 2004). Bonetti (1983) estimou que cerca de 70% das cultivares desenvolvidas para o Rio Grande do Sul naquela data eram descendentes das cultivares americanas Hill, Hood ou ambas. Em 1986, Hiromoto e Vello relataram que todas as cultivares recomendadas naquele ano eram derivadas de apenas 26 genótipos ancestrais, quatro dos quais responsáveis por quase 50% daquele conjunto gênico.

Este cenário não foi diferente nos anos seguintes. No estudo de Abdelnnor *et al.* (1995) utilizando marcadores RAPDs, não foi encontrado grande variação nas distâncias genéticas entre 38 cultivares de soja brasileiras. As distâncias estimadas estiveram entre 14,2% e 20,5%. Priolli *et al.* (2002), através de um estudo de caracterização de 186 cultivares de soja brasileiras utilizando marcadores microssatélites, revelaram que variedades norte-americanas contribuíram efetivamente para a formação deste grupo de cultivares. Santa Rosa, D4929I (linhagem irmã de Lee), Hill, Davis e Hood foram as mais utilizadas como genitores, sendo identificadas direta ou indiretamente na maioria dos *clusters* do dendograma consenso construído.

2.1.2. Genoma e aspectos genéticos da soja

A soja contém 20 cromossomos que cobre uma extensão de aproximadamente 1,115 milhões de pares de bases por genoma haplóide (Shoemaker *et al.*, 2003). Em torno de 55% do genoma consiste de seqüências repetitivas (Danesh *et al.*, 1998), e mais de 35% do material genético encontra-se sob a forma de heterocromatina. O genoma da soja é caracterizado como tetraplóide parcialmente diploidizado (Shultz *et al.*, 2006). Dados de mapeamento utilizando marcadores RFLPs levaram a identificação de vários segmentos duplicados do genoma da soja. Foram observados segmentos presentes no genoma em até seis cópias, com 2,25 duplicações por segmento em média. Entre as sondas de RFLP testadas neste trabalho, 90% detectaram locos duplicados e cerca de 60% detectaram três ou mais locos (Shoemaker *et al.*, 1996).

Embora a soja tenha um genoma complexo, várias metodologias têm sido desenvolvidas com o objetivo de elucidar aspectos estruturais e funcionais do genoma. Com esse propósito, inúmeros estudos em soja têm empregado técnicas que incluem: a construção de mapas de ligação saturados; mapeamento de QTL; desenvolvimento de bibliotecas genômicas; análises de coleções de ESTs; mapeamento físico, entre outras.

A detecção e localização de QTL no mapa de ligação da soja têm grande aplicação para a seleção assistida por marcadores moleculares em programas de melhoramento, além disso, é de grande importância para a clonagem posicional de genes (*map-based cloning*), facilitando a identificação de locos associados a diversas características de interesse em soja (Harada e Xia, 2004).

No mapa de Cregan *et al.* (1999a), inúmeros genes da soja já encontravam-se integrados, por exemplo, os locos para cor da flor (*W1*), cor dos cotilédones (*D1*, *D2*), pigmentação foliar (*Y9*), florescência da raiz (*Fr1*, *Fr2*), entre muitos outros. Vários locos associados a resistência à patógenos foram mapeados, entre eles, os genes de resistência ao nematóide de cisto da soja (NCS). O gene *rhg4*, associado a resistência à raça 3 do NCS, foi mapeado no grupo de ligação A2 (Matthews *et al.*, 1998) e o gene *rhg1*, que confere resistência parcial às diferentes raças do NCS, mapeado no grupo de ligação G (Cervigni, *et al.*, 2004). Um grande número de trabalhos têm sido realizado envolvendo o mapeamento de QTLs para diversas características agronômicas da soja, sobretudo características relacionadas a produção e composição dos grãos.

Com o propósito de elucidar características do genoma da soja como um todo, grande quantidade de informação tem sido gerada a partir do sequenciamento de clones de cDNA, geração e análise de ESTs. Através do projeto público Soja EST dos EUA, mais de 80 bibliotecas de cDNA foram geradas representando transcritos de vários genótipos, em diferentes estádios de desenvolvimento, órgãos, tecidos e condições de estresse (Harada e Xia, 2004). Shoemaker *et al.* (2002) analisaram mais de 120.000 ESTs gerados a partir de mais de 50 bibliotecas de cDNA. Em um trabalho similar, Tian *et al.* (2004) analisaram 314.254 ESTs da soja, incluindo 29.540 de seu laboratório e 284.714 provenientes do *GenBank*. A partir desses estudos foram identificadas inúmeras seqüências descritas como genes específicos da soja, envolvidas principalmente com desenvolvimento nodular e síntese de proteínas de armazenamento da semente. A análise dessas seqüências não revelou alinhamento com seqüências do genoma de *Arabidopsis* e nem de outros organismos.

Outra metodologia para estudo do genoma constitui a obtenção de clones contendo grandes fragmentos de DNA, essencial para a geração de mapas físicos de alta resolução, bem como, para a clonagem posicional de genes e construção de mapas físicos de genomas inteiros. Várias bibliotecas BAC já foram construídas em soja (Marek e Shoemaker, 1997; Danesh *et al.*, 1998, Salimath e Bhattacharyya, 1999; Tomkins *et al.*, 1999; Meksem *et al.*, 2000; Tomkins *et al.*, 2000). Uma das maiores

contribuições desses trabalhos foi o isolamento por clonagem posicional e caracterização dos genes de resistência ao NCS *rhg1* e *rhg4* (Mecksem *et al.*, 2002).

2.2. Marcadores Moleculares

2.2.1. Marcadores moleculares no melhoramento de plantas

Acredita-se que a demanda mundial por alimentos aumente consideravelmente nos próximos anos. Segundo relatório da ONU divulgado no ano de 2006, a população mundial aumentará em 2,5 mil milhões de habitantes nos próximos 43 anos, passando dos atuais 6,7 mil milhões para 9,2 mil milhões de pessoas, um aumento de 37,3% (Agencia Lusa, 2008).

No passado, aumentos na produtividade agrícola foram conseguidos pela combinação do melhoramento genético, utilização de insumos agrícolas e extensão da área cultivada. Para os próximos tempos, acredita-se que ganhos adicionais na produção para as principais espécies cultivadas dependam quase que integralmente do melhoramento genético de plantas (Tanksley e McCouch, 1997).

O melhoramento de plantas, na sua forma convencional, é baseado na seleção fenotípica de genótipos superiores dentro de populações segregantes obtidas a partir de cruzamentos. Uma estratégia comum é o cruzamento de cultivares existentes, deficientes em alelos favoráveis que controlam uma ou mais características, com linhagens que possuam o alelo determinante do padrão fenotípico desejável. A seleção dos recombinantes superiores é então realizada entre vários produtos da segregação. Esse procedimento encontra dificuldades principalmente relacionadas a interações genótipo x ambiente. Além disso, requer vários cruzamentos, várias gerações e a seleção cuidadosa dos fenótipos, um processo laborioso e demorado. Os custos são bastante elevados e para determinadas características a avaliação fenotípica é muitas vezes impraticável, como por exemplo, a seleção para tolerância a estresse abiótico. A forte ligação de genes de interesse com locos indesejáveis constitui outro importante fator que dificulta o alcance do objetivo na maioria das vezes (Kumar, 1999).

A utilização de marcadores moleculares em programas de melhoramento é, portanto, uma alternativa para o melhoramento genético das principais culturas de interesse econômico, no sentido de aumentar a eficiência e proporcionar ganhos

genéticos maiores. A seleção assistida por marcadores moleculares, nesse sentido, supera grande parte das dificuldades dos métodos tradicionais, pois altera o critério de seleção do fenótipo para a seleção de genes, direta ou indiretamente. Além de aprimorar estratégias de seleção, os marcadores moleculares têm sido utilizados para monitorar a variação na sequência do DNA dentro de espécies, para a criação de novas fontes de variação genética visando à introdução de características desejáveis.

2.2.2 *Evolução das técnicas moleculares*

O desenvolvimento de técnicas moleculares para análises genéticas aumentou o entendimento da estrutura e comportamento do genoma das plantas. Nos últimos anos, diferentes sistemas de marcadores foram desenvolvidos para várias espécies vegetais. A aplicação desses marcadores em estudos genéticos tem sido bastante diversa, as principais em estudos com cereais incluem: análise de diversidade genética, caracterização de germoplasma, *fingerprinting*, mapeamento de QTL e seleção assistida por marcadores moleculares (Korzun, 2003).

Até meados da década de 60, os marcadores utilizados em estudos de genética e melhoramento eram controlados por genes associados a caracteres morfológicos, em geral fenótipos de fácil identificação visual. Além da variação ambiental, a expressão dos marcadores morfológicos é alterada por interações epistáticas e pleiotrópicas. A disponibilidade desses marcadores é restrita a poucas espécies de plantas e, na maioria dos casos, seus alelos interagem de maneira dominante, o que impossibilita a distinção entre indivíduos heterozigotos e homozigotos dominantes (Kumar, 1999). Esses fatores acabaram limitando consideravelmente o emprego destes marcadores pelos programas de melhoramento genético das principais espécies cultivadas.

Os primeiros marcadores de natureza molecular estabelecidos foram as isoenzimas. O termo isoenzima, introduzido por Markert e Moller (1959), designa formas moleculares múltiplas de enzimas com a mesma especificidade enzimática. Enzimas com variações na estrutura protéica, causadas por substituições na sequência de aminoácidos, podem ser distinguidas através da eletroforese em gel nativo de acordo com diferenças no tamanho e carga elétrica (Schlotterer, 2004). O controle genético das isoenzimas mais comum em plantas apresenta segregação monogênica e co-dominância, ocorrendo expressão dos dois produtos alélicos no heterozigoto. Os híbridos, desta forma, podem ser eficientemente diferenciados dos seus genitores

(Borém e Caixeta, 2006). No entanto, o uso destes marcadores tem sido restrito pelo número limitado entre as espécies, como também, por estarem sujeitos a modificações pós-traducionais. Além disso, o número de locos informativos é pequeno para o uso das isoenzimas em estudos de associação e mapeamento genético (Hillis *et al.*, 1996).

Com a descoberta das endonucleases de restrição e avanço das técnicas moleculares, uma nova classe de marcadores genéticos surgiu, os marcadores genéticos de DNA, destacando-se os marcadores RFLP, RAPD, AFLP e SSR. Em relação aos marcadores morfológicos, os marcadores genéticos de DNA apresentam grandes vantagens, não são influenciados por fatores ambientais já que o polimorfismo é baseado na variação de bases na seqüência do DNA, são geralmente em número elevado e apresentam alto grau de polimorfismo (Soares, 2004).

Os primeiros destes marcadores formam os RFLPs (*Restriction fragment length polymorphisms*). Esses marcadores baseiam-se na variação no DNA devido a substituições de base em sítios de reconhecimento das enzimas de restrição, alterando o padrão de fragmentos de restrição gerados (Schlotterer, 2004). O polimorfismo é obtido quando ocorre, por mutação de ponto, perda ou aparecimento de sítios de restrição. São marcadores de herança co-dominante e permitem a análise de regiões codificadoras e não codificadoras, sendo capazes de identificar alterações gênicas silenciosas. Inicialmente, os RFLPs foram desenvolvidos para o mapeamento do genoma humano. Mais tarde, foram empregados para o mapeamento do genoma de várias espécies vegetais, inclusive da soja (Keim *et al.*, 1990; Shoemaker *et al.*, 1992; Cregan *et al.*, 1999).

A facilidade, rapidez e sensibilidade da reação de polimerização em cadeia possibilitaram o surgimento dos marcadores moleculares baseados em PCR *primers* específicos (RAPD, AFLP, SSR, entre outros). Os RAPDs (*Random amplified polymorphic DNA*) são fragmentos de DNA amplificados por PCR a partir de um único *primer* de seqüência arbitrária e tamanho em torno de nove a dez nucleotídeos. São amplificados fragmentos de DNA distribuídos ao acaso no genoma, sem necessidade do conhecimento prévio da seqüência do DNA. Os RAPDs são marcadores dominantes e o polimorfismo revelado geralmente tem origem em mutações de ponto no sítio de pareamento do *primer*. Os RAPDs têm como limitação a baixa reprodutibilidade, em razão da sua alta sensibilidade a pequenas variações experimentais.

Marcadores AFLPs (*Amplified fragment length polymorphism*) constituem outra classe de marcadores baseada na combinação da especificidade dos sítios de restrição

dos RFLPs com a praticidade da amplificação por PCR (Vos *et al.*, 1995). O polimorfismo é preferencialmente dominante sendo possível, no entanto, identificar indivíduos heterozigotos pela diferença de intensidade das bandas geradas quando a técnica é realizada com boa precisão.

Um marcador de DNA baseado na variação no número de repetições em *tandem* de seqüências curtas de nucleotídeos foi sugerido simultaneamente por três grupos de pesquisa (Litt e Luty, 1989; Tautz, 1989; Waber e May, 1989). Por meio de pesquisas no *Genbank*, Cregan (1992) relatou a presença de DNA microssatélite em várias espécies vegetais e sugeriu o potencial da utilização desses marcadores em estudos com plantas.

Os microssatélites constituem uma classe de seqüências curtas de DNA de dois a quatro pares de bases, repetidas em *tandem*. São abundantes e ocorrem frequentemente e de forma aleatória no DNA genômico de todos eucariotos estudados, inclusive em regiões do DNA expressas (Gupta e Varshney, 2000). Em um trabalho com 54 espécies de plantas, Wang *et al.* (1994) observaram que as seqüências (AT)_n são as mais abundantes, seguidas por (A)_n, (AG)_n, (AAT)_n, (AAC)_n, (AGC)_n, (AAG)_n, (AATT)_n, (AAAT)_n e (AC)_n. Regiões microssatélites são também encontradas no genoma de cloroplastos em vários sistemas vegetais, esses microssatélites têm sido comumente utilizados em estudos de diversidade genética em plantas. Em soja, por exemplo, microssatélites do DNA cloroplastidial têm sido utilizados em estudos sobre origem, linhagem maternal e grau de introgressões na soja (Xu *et al.*, 2002).

O princípio da técnica consiste na amplificação por PCR de regiões microssatélites individualmente, utilizando pares de *primers* específicos com tamanho em torno de 20 a 30 bases, complementares a seqüências únicas que flanqueiam regiões microssatélites particulares. Esses marcadores têm expressão co-dominante, são multialélicos, e apresentam, de maneira geral, alto grau de polimorfismo. Além da alta reprodutibilidade, possuem alto conteúdo de informação genética.

Nos últimos anos foram desenvolvidos novos sistemas de marcadores mais eficientes e não baseados na análise de géis de eletroforese, passíveis de automatização e adequados para análises genéticas em grande escala (Gupta *et al.*, 2001). Os SNPs são uma nova classe de marcadores genéticos baseados na detecção de polimorfismos resultantes da variação de um único nucleotídeo no genoma. São marcadores de natureza bi-alélica, entretanto, são extremamente abundantes nos genomas. Podem ser encontrados em regiões codificadoras e não codificadoras, seqüências regulatórias e

regiões de DNA repetitivo. A genotipagem dos SNPs, diferente de outros marcadores, não se baseia nas diferenças de tamanho dos alelos, o que dispensa a utilização de géis.

Inicialmente, as abordagens de detecção dos SNPs consistiam na amplificação e sequenciamento de fragmentos de regiões específicas do DNA de indivíduos da mesma espécie, e comparação das seqüências identificando variações (Borém e Caixeta, 2006). Hoje existem disponíveis diversas metodologias para a identificação de novos SNPs, como também, vários métodos de genotipagem de indivíduos para locos SNPs.

2.3. Mapeamento de QTLs associados a conteúdo de proteína e óleo e produtividade em soja

A soja é cultivada principalmente como fonte de proteína e óleo. No entanto, o valor da soja para o produtor é baseado na sua produtividade, característica que tem sido o foco principal da maioria dos programas de melhoramento. Entretanto, fatores de qualidade do grão, como conteúdo de proteína e óleo, têm recebido grande atenção pelos melhoristas nos últimos tempos (Nichols *et al.*, 2006).

Além da produtividade, as características de importância agrônômica da soja geralmente são de herança quantitativa, e sua manipulação é complexa em relação as características de herança simples. O controle genético da variação quantitativa dessas características envolve geralmente muitos locos, de composição na maioria das vezes desconhecida. Em 1975, Geldermann designou tais locos como QTL ou *Quantitative trait loci*.

A construção de mapas genéticos, muitas vezes, fornece informação genética para o melhoramento das características agrônômicas. Vários mapas genéticos da soja foram construídos desde 1990 (Keim *et al.*, 1990; Keim *et al.*, 1997; Muehlbauer *et al.*, 1991; Lark *et al.*, 1993; Akkaya *et al.*, 1995; Shoemaker e Specht, 1995; Zhang *et al.*, 1997; Liu *et al.*, 2000; Wu *et al.*, 2001, Song *et al.*, 2004). No entanto, para muitas características pouca informação consistente existe sobre o número, posição, magnitude do efeito e interações dos QTLs.

A disponibilidade dos marcadores moleculares, neste sentido, tem facilitado o estudo das características quantitativas. E sua utilização em estudos de mapeamento tem levado a identificação de QTLs para as principais características de interesse em soja. São inúmeros os estudos sobre mapeamento de QTL disponíveis na literatura, envolvendo o controle genético de diversas características, tais como: características

morfológicas (Keim, *et al.*, 1990; Mansur *et al.*, 1996; Lee *et al.*, 1996; Zhang *et al.*, 2004); reprodutivas (Keim *et al.*, 1990; Mansur *et al.*, 1996); conteúdo de proteína e óleo no grão (Diers *et al.*, 1992; Lee *et al.*, 1996; Brummer *et al.*, 1997, Orf *et al.*, 1999; Qiu *et al.*, 1999; Sebolt *et al.*, 2000; Casanadi *et al.*, 2001; Chung *et al.*, 2003; Nichols *et al.*, 2006); resistência à doenças (Concibido *et al.*, 1994; Webb *et al.*, 1995; Qiu *et al.*, 1999; Yuan *et al.*, 2002); produtividade (Wang *et al.*, 2004; Chung *et al.*, 2003; Concibido *et al.*, 2003), entre outras.

Concibido *et al.* (2003) identificaram um QTL de *Glycine soja* para aumento da produtividade. A população utilizada foi uma RC₂ de 265 indivíduos, obtida do cruzamento HS-1 x PI 407305 (acesso de *G. soja* portador de locos para produção). O QTL, presente no grupo de ligação (LG) B2, teve seu efeito testado em dois anos e em vários locais e explicou um aumento no rendimento de 9,4%. Os autores também testaram o efeito do QTL em seis *backgrounds* genéticos elite da soja e em dois destes os efeitos do QTL foram consistentes.

No ano seguinte, Zhang *et al.* (2004) utilizaram um grupo de 184 RILs derivadas do cruzamento Kejeng N^o1 x Nanong 1138-2 e construíram um mapa genético que cobriu 3.595,9 cM do genoma da soja, com 452 marcadores distribuídos em 21 grupos de ligação. Neste trabalho foram mapeados 61 QTLs associados a dez características agrônômicas da soja. Apenas para produtividade foram mapeados sete QTLs em três grupos de ligação, B1, C2 e M. Entre eles, o QTL de maior efeito explicou 12,6% da variação da característica. Os outros seis locos detectados explicaram entre 6% e 12,3% da variação fenotípica.

Wang *et al.* (2004) obtiveram cinco populações constituídas por linhagens RC₂F₄, utilizando o cultivar *G. max* IA2008 como genitor recorrente e *G. soja* PI 468916 como genitor doador. Quatro QTLs para rendimento foram identificados nos grupos de ligação C2, E, K e M. O QTL do grupo de ligação C2 foi consistente em uma das populações para três ambientes. Esse QTL explicou 40% da variação para produtividade nesta população. Em outra população foram mapeados QTLs nos grupos de ligação E e M, que explicaram 28% e 29% da variação da característica, respectivamente. O QTL do grupo de ligação E foi consistente em três dos quatro ambientes, e o QTL do grupo de ligação M foi confirmado em todos os ambientes. O QTL do grupo de ligação K, por sua vez, foi identificado em outras duas populações e foi consistente em todos ambientes testados. Esse QTL, localizado no intervalo entre os

marcadores Satt137 e Satt178, explicou em uma população 40% e em outra 38% da variação para produtividade.

Panthee *et al.* (2007) realizaram um estudo de mapeamento de QTL para características agrônômicas da soja, envolvendo uma população de 101 RILs obtidas do cruzamento entre N98-984 x TN93-99, avaliadas em três localidades e por dois anos. Três QTLs próximos aos marcadores Satt042 (LG A1, $R^2=14,6\%$), Satt412 (LG G; $R_2=13,3\%$) e Satt076 (LG G; $R^2=13,8\%$) foram associados com produtividade nesta população. No entanto, apenas o QTL próximo ao Satt076 apresentou significância em todos ambientes testados. Juntos, os três QTLs explicaram aproximadamente 41% da variação total para produtividade nesta população.

Além da produtividade, existem também muitos trabalhos de mapeamento genético para conteúdo de proteína e óleo em grãos de soja envolvendo diferentes tipos de populações (RILs, NILs, F_2 , F_3 , F_n e RC) e diferentes *backgrounds* genéticos. De acordo com Hyten *et al.* (2004), existem atualmente na literatura pelo menos 61 QTLs associados ao conteúdo de proteína em sementes de soja e a maioria deles não foram confirmados.

Um dos primeiros trabalhos de identificação de QTL controlando concentração de proteína foi desenvolvido por Diers *et al.* (1992), utilizando uma população de 60 famílias F_3 obtidas do cruzamento entre um genótipo não adaptado com alto teor de proteína (*Glycine soja*) e uma variedade comercial. Foram genotipados 252 locos RFLPs que se distribuíram em 31 grupos de ligação. Os autores localizaram duas regiões nos grupos de ligação I e E de *G. soja* determinando elevado teor de proteína e reduzido teor de óleo no grão. Os locos de maior significância para cada grupo de ligação foram associados com aumentos na concentração de proteína de 17 g Kg^{-1} para o LG E, e 24 g Kg^{-1} para o GL I.

Sebolt *et al.* (2000) testaram os efeitos destes dois QTLs para proteína retrocruzando *G. soja*, doador dos locos para alto teor de proteína, e o genitor recorrente A81-356022. Uma população RC_3 foi obtida e avaliada para performance agrônômica e conteúdo de proteína e óleo em dois locais e por dois anos. Os autores confirmaram o efeito do QTL do grupo de ligação I para teor de proteína, no entanto, não identificaram efeito significativo para o QTL do grupo de ligação E, reportado por Diers *et al.* (1992). Entre os marcadores do LG I testados, Satt127 obteve o maior valor de R^2 para teor de proteína, enquanto A144H-1 obteve o maior R^2 para teor de óleo. Os alelos de *G. soja* para os locos Satt127 e A144H foram associados com aumento de proteína e redução de

óleo. Para proteína, o valor de R^2 para BARC-Satt127 foi de 0,8 no ano de 1996 e 0,41 em 1997, explicando uma variação de 30 g Kg⁻¹ e 12 g Kg⁻¹ nos anos de 1996 e 1997, respectivamente.

Csanádi *et al.* (2001) construíram um mapa genético com 23 grupos de ligação utilizando 120 marcadores e 82 indivíduos de uma população F₂. A partir dos dados fenotípicos de linhagens F₃ avaliadas em cinco ambientes, os autores identificaram 15 marcadores associados com conteúdo de proteína em pelo menos um ambiente ($P < 0,01$). Destes, apenas quatro (Satt77, Sct_028, Satt567, Satt196) tiveram associação significativa com conteúdo de proteína nos cinco ambientes. Juntos, estes marcadores explicaram 23% da variação da característica.

No trabalho de Chung *et al.* (2003) também foi mapeado um QTL para concentração de proteína no grupo de ligação I a partir da PI437088A (acesso de *G. max* com alto conteúdo de proteína), na mesma região reportada por Sebolt *et al.* (2000). Eles observaram um aumento de 18 g Kg⁻¹ na concentração de proteína entre as linhagens que continham o alelo da PI437088A em relação às outras.

Nichols *et al.* (2006) realizaram um estudo de mapeamento de QTL direcionado ao grupo de ligação I, utilizando dois grupos de populações. Um grupo consistia de três populações de linhagens RC₄F₃ desenvolvidas a partir da população RC₃ descrita por Sebolt *et al.* (2000). O outro incluía quatro populações RC₅. Nas populações do primeiro grupo, QTLs para produtividade, concentração de proteína e óleo, maturidade e tamanho da semente foram mapeados no intervalo de 11cM entre as marcas Satt614 e Satt354. Os dados relativos as populações do segundo grupo redefiniram a posição do QTL para teor de proteína e óleo no intervalo de 3 cM entre os marcadores Satt239 e ACG9b.

Apesar do número considerável de QTLs mapeados para características relacionadas a produtividade e conteúdo de proteína e óleo no grão, existe ainda informação limitada e inconsistente sobre a confirmação dos QTLs reportados na literatura.

Neste contexto, os marcadores moleculares podem oferecer novas possibilidades para o melhoramento genético de plantas, principalmente por meio do mapeamento de QTLs que controlam a variação genética de características complexas de interesse, e da possibilidade de seleção assistida de genótipos para QTLs identificados (Bouchez *et al.*, 2002).

3. MATERIAL E MÉTODOS

3.1. *Material genético*

Foi utilizada para o mapeamento genético uma população de 206 indivíduos F_2 obtida do cruzamento entre a linhagem CS3035PTA276-1-5-2 (43,7% de teor protéico e 18,9% de teor de óleo) e a variedade UFVS2012 (37,6% de teor protéico e 21,6% de teor de óleo). As progênies da geração F_2 foram plantadas em casa de vegetação e folhas jovens de cada indivíduo foram coletadas para a extração e análise de DNA.

A avaliação do fenótipo foi realizada em famílias F_3 em um experimento montado no ano agrícola de 2006, que constou de três repetições ou blocos, de forma que cada família de plantas F_3 foi semeada em três repetições. Em cada família de cada bloco foram coletados valores fenotípicos de cinco indivíduos. A população foi plantada no campo experimental Diogo Alves de Mello, localizado no Campus da Universidade Federal de Viçosa, Zona da Mata do Estado de Minas Gerais. A sede do município situa-se aproximadamente a 650m de altitude e o solo nessa localidade é classificado como latossolo vermelho-amarelo distrófico, com textura argilosa. Como testemunhas, foram plantadas em cada bloco, linhas das cultivares BARC-8 e Monarca, bem como dos genitores (CS3035PTA276-1-5-2 e UFVS2012). A população de genótipos de soja, bem como, a avaliação fenotípica das características em estudo, foi desenvolvida por Miranda (2006).

As seguintes características foram avaliadas na geração F_3 : número de dias para florescimento (NDF); número de dias para maturação (NDM); altura da planta na maturação (APM); altura da primeira vagem (APV); número de nós na maturação (NNM); número de vagens por planta (NVP); número de sementes por planta (NSP); peso de sementes por planta (PRO); peso de cem sementes (PCS); conteúdo de proteína (PTN) e conteúdo de óleo (OLEO).

3.2. *Amplificação e Avaliação de Marcadores Microsatélites*

A extração de DNA dos 206 indivíduos F_2 foi realizada por Miranda (2006). A concentração de DNA foi estimada espectrofotometricamente por leitura da absorbância a 260 nm, a cada unidade de absorbância foi atribuído uma concentração de $50 \mu\text{g mL}^{-1}$

de DNA fita dupla. E a integridade do DNA foi verificada em gel de agarose 0,8% fotografado sob luz ultravioleta.

As reações de amplificação dos microsátélites foram feitas em um volume de reação de 15 μ L, contendo 1,5 μ L de tampão IA *Phoneutria* (KCl 500 mM, Tris-HCl 100 mM, Triton X-100 1%, MgCl₂ 7,5 mM), 100 μ M de cada um dos desoxinucleotídeos, 0,3 μ M de cada *primer*, uma unidade de *Taq* polimerase e 30 ng de DNA. As amplificações foram conduzidas em termociclador Perkin Elmer[®] modelo 9600 programado para uma etapa inicial de 4 min a 94°C, seguida de 30 ciclos de 1 min a 94°C, 1 min a 50°C e 2 min a 72°C, e por fim, uma etapa de 7 min a 72°C. Os fragmentos amplificados foram separados por eletroforese em gel de poliacrilamida 10% utilizando tampão TAE 1X (Tris-base 90 mM, Acetato 100 mM e EDTA 2mM), a 140 V. Terminada a eletroforese, os géis foram corados com brometo de etídio (1 mg/mL) e fotografados sob luz ultravioleta pelo sistema de fotodocumentação *Eagle Eye II* (Stratagene).

3.3. Análise de associação marcador - QTL

A segregação individual dos marcadores foi testada por meio do teste qui-quadrado utilizando o critério de Proteção Bonferroni ($P < 0,05$). Os marcadores foram agrupados nos grupos de ligação utilizando-se LOD score mínimo de 3,0 e frequência máxima de recombinação de 30%.

A associação entre marcadores e características foi determinada através da análise de associação pelo método da ANOVA e por meio de regressão linear. Para a análise de QTL, no entanto, é necessário o conhecimento das distâncias em frequência de recombinação entre os marcadores de um intervalo. Portanto é preciso dispor, antes da análise de associação com o QTL, de informações dos grupos de ligação dos marcadores.

As análises de QTL foram feitas pelos métodos de intervalo simples (Lander e Botstein, 1989) e composto (Zheng, 1994; Jansen, 1993). No mapeamento por intervalo simples é testada a ocorrência do QTL em uma série de posições dentro de um único intervalo entre dois marcadores, sem considerar a influência de outras regiões. Essa análise é, portanto, muitas vezes afetada pelo efeito de outros QTLs do mesmo grupo de ligação ou até de outros grupos, e o poder do teste é reduzido. Pelo método de intervalo

composto, no entanto, são adicionados co-fatores na análise que aumentam o poder de detecção e a precisão da estimativa da posição do QTL. Diferente do método de intervalo simples, este método não é afetado pelos efeitos de outros QTLs localizados em outros intervalos.

A seleção dos co-fatores para análise de mapeamento por intervalo composto, foi baseada na análise de regressão stepwise, adotando-se o nível de significância de 5% e 10%, respectivamente, para entrada e saída das variáveis independentes. Para alguns intervalos foram ainda utilizados co-fatores do próprio grupo de ligação, marcadores com associação significativa com a característica em estudo pela análise de marca simples.

O mapeamento por intervalo composto é um método baseado em análise de regressão, sendo adotado o seguinte modelo:

$$y_j = \mu + ax_j^* + dz_j^* + \varepsilon_j$$

em que

y_j = valor da característica no j-ésimo indivíduo;

μ = média da característica na população.

a = efeito aditivo do loco que está sendo estudado sobre a característica.

d = efeito da dominância do loco que está sendo estudado sobre a característica.

x_j^* = variáveis condicionadoras e dependentes dos genótipos dos marcadores

que flanqueiam o QTL, no indivíduo j; e

ε_j = erro aleatório.

A significância de um QTL é atribuída pela razão de verossimilhança entre a hipótese de existência do QTL sobre a hipótese de não existência do QTL em um determinado intervalo no mapa de ligação. A razão de verossimilhança possui aproximação com o χ^2 pela seguinte expressão:

$$LR = -2n \ln \frac{SQD_{Reduzido}}{SQD_{Completo}}$$

em que

n = número de indivíduos avaliados;

$SQD_{Reduzido}$ = soma de quadrado do desvio do modelo reduzido, dado por

$$y_j = \beta_0 + \sum \beta_k x_{jk} + \varepsilon_i ; e$$

$SQD_{Completo}$ = soma de quadrado do desvio do modelo.

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

$$\alpha_I \cong \frac{\alpha_C}{M}$$

em que

C = Número de grupos de ligação; e

M = Número de marcas dentro de um grupo de ligação.

Uma alternativa para declarar o valor crítico para declarar a existência de um QTL associado ao marcador é o teste de permutações (Churchil e Deorge, 1994). Para realização deste teste indexam-se os indivíduos de 1 até n . Os valores fenotípicos das características são tomados ao acaso e atribuídos aos indivíduos conforme forem sendo retirados. O primeiro valor retirado da amostra de dados fenotípicos é atribuído ao indivíduo 1, e assim sucessivamente, até que todos os indivíduos tenham um valor associado. Os dados permutados são então submetidos a análise para detectar a presença de QTL. Os resultados desta análise são armazenados, e o procedimento todo é repetido N vezes.

Dessa forma qualquer associação entre marcadores e valores fenotípicos é destruída, e as análises realizadas correspondem aos resultados esperados sob a hipótese nula. Quanto maior o número de permutações, maior a precisão do valor crítico. É conhecido que para os níveis de significância 5% e 10% são necessárias 1000 permutações, já para o nível de significância de 1% o número de permutações deverá ser de 10.000 (Deorge e Rebai, 1996).

O mapeamento por intervalo composto incorpora a informação de locos múltiplos de cada indivíduo. No entanto a inclusão de marcadores que não estão associados com a característica pode levar a um aumento na variância residual, em vez de sua redução. Dessa forma são utilizados apenas um subconjunto de marcadores selecionados como co-fatores na análises, de forma que se tenha o modelo:

$$y = \beta_0 + \sum_{k \neq i, i+1}^c \beta_K x_{kj}$$

O número de marcadores que devem ser utilizados como co-fatores na análise não é definido. A princípio todos os marcadores não ligados que apresentarem associação significativa com a característica pela análise de regressão simples devem

ser usados. No entanto, a inclusão de muitos co-fatores pode reduzir muito o poder do teste. E se diversos marcadores ligados no mesmo grupo de ligação mostrarem associação significativa, a utilização apenas dos marcadores de maior efeito é suficiente para reter a variância devida ao QTL localizado neste grupo.

Todas as análises foram realizadas utilizando o programa GQMOL (<http://www.ufv.br/dbg/gqmol/gqmol/htm>).

4. RESULTADOS E DISCUSSÃO

4.1. Dados fenotípicos

As análises estatísticas e a determinação dos parâmetros genéticos para as 11 características em estudo foram realizadas por Miranda (2006). Foram estimados os componentes da variação genética, as herdabilidades em sentido amplo e as correlações para as seguintes características da soja: número de dias para florescimento (NDF); número de dias para maturação (NDM); altura da planta na maturação (APM); altura da primeira vagem (APV); número de nós na maturação (NNM); número de vagens por planta (NVP); número de sementes por planta (NSP); peso de sementes por planta (PRO); peso de cem sementes (PCS); conteúdo de proteína (PTN) e conteúdo de óleo (OLEO).

Os resultados das análises de variância e das estimativas dos parâmetros genéticos das características indicaram a existência de variabilidade genética na população para as onze características a 1% de significância. Com base nestes resultados, foi confirmado o potencial genético da população para estudo de mapeamento de QTL. Sua viabilidade para fins de mapeamento genético foi ainda evidenciada pela existência de contrastes de grande magnitude entre as médias dos genitores para a maioria das características, valores descritos na Tabela 1.

Para as características NDF, NDM e PCS, embora não exista diferença entre os valores médios dos genitores, foi verificada segregação transgressiva. Os segregantes transgressivos correspondem aos indivíduos com valores fenotípicos superiores aos dos genitores para uma dada característica.

Quanto à variância genética dos valores relativos a cada caráter, o teste F a 1% de probabilidade indicou a existência de significância para as onze características avaliadas nas famílias F₃. A razão CV_g/CV_e observada foi elevada para todas as características, indicando que boa parte da variação total é devido à variância genética. Os coeficientes de variação apresentaram valores aceitáveis, sendo para proteína e óleo estes valores iguais a 3,01 e 3,47, respectivamente. Os valores reduzidos do coeficiente de variação observados indicam boa precisão experimental.

Os valores de herdabilidade observados para as características avaliadas variaram entre 44,92% e 86,59%, com as maiores estimativas verificadas para as

características NNM e PCS, superiores a 80%. Os valores de herdabilidade estimados para PTN e OLEO foram de 62,36% e 71,51%, respectivamente. E as estimativas observadas para os componentes de produção foram de 56,55% e 86,59% para PRO e PCS, respectivamente. Este parâmetro refere-se a proporção da variação fenotípica devido a causa genética. Valores elevados de herdabilidade estão relacionados com alta eficiência de seleção fenotípica, já que grande parte da variância fenotípica é explicada pela variância genética.

Sobre a correlação entre as características em estudo foi verificado que as maiores correlações genéticas com PRO foram encontradas para as características NDM (0,74) e NVP (0,80). Os maiores coeficientes de correlação com PCS e NSP, por sua vez, foram verificados para as características PRO (0,51) e NVP (0,96), respectivamente. Para a característica teor de proteína do grão (PTN), todas as correlações com outras características apresentaram valores negativos. E para a característica teor de óleo, os valores de correlação com a maioria das características mostraram-se positivos, no entanto, relativamente baixos.

Tabela 1. Análise descritiva das características em estudo da geração F₃ e dos genitores

Caráter	Unidade	População			Média dos Genitores	
		Média	DP ^a	Mín - Máx ^b	CS303	UFVS2012
NDF	dias	64,12	1,28	60,5-67,6	64,3	64,6
NDM	dias	138,15	7,52	88,5-141,8	137,5	138,3
APM	cm	87,1	11,53	53,4-147,04	102,3	76,3
APV	cm	16,34	2,51	7,0-23,2	15	19,3
NNM	-	14,95	1,40	6,5-17,5	18	13
NVP	-	112,1	18,40	57,4-158,0	100,3	112,3
PRO	g	196,43	36,11	109,9-291,9	195,4	179,3
PCS	g	17,02	1,54	8,2-20,8	16,3	15
NSP	-	214,45	38,83	129,1-394,5	226	203,6
PTN	g	41,36	1,58	35,1-44,8	43,7	37,6
OLEO	g	19,81	0,81	17,8-21,9	18,9	21,6

DP^a = desvio padrão da característica. Min/Max^b = valor mínimo e máximo para cada característica. Número de dias para florescimento (NDF); número de dias para maturação (NDM); altura da planta na maturação (APM); altura da primeira vagem (APV); número de nós na maturação (NNM); número de vagens por planta (NVP); peso de sementes por planta (PRO); peso de cem sementes (PCS); número de sementes por planta (NSP); conteúdo de proteína (PTN) e conteúdo de óleo (OLEO).

4.2. Mapeamento de QTL

4.2.1. Segregação dos marcadores

Foi realizado um *screening* com 357 *primers* microssatélites testados nos genitores. Desses, 107 *primers* apresentaram fragmentos polimórficos. Entre os marcadores polimórficos selecionados, 48 tiveram sua segregação avaliada na população. Estes microssatélites foram selecionados entre os marcadores polimórficos com base em informações sobre QTLs associados ao conteúdo de proteína e óleo reportados na literatura. Apenas marcadores localizados em regiões próximas a QTLs envolvidos com o controle gênico do conteúdo de proteína e óleo em soja foram amplificados a partir das amostras de DNA dos indivíduos da população. Foi considerado ainda a qualidade de amplificação dos *primers*.

A segregação de cada marcador foi testada por meio do teste qui-quadrado utilizando o critério de Proteção Bonferroni ($P < 0,05$). Três dos 48 marcadores apresentaram distorção na segregação (Satt352, Satt429 e Satt454). Estes marcadores não foram excluídos, sendo utilizados tanto para a obtenção dos grupos de ligação, quanto para as análises de intervalo, já que os mesmos não formaram intervalos com associação com qualquer uma das características.

Na construção de mapas genéticos a partir de dados provenientes de populações derivadas de cruzamentos controlados é esperado um padrão típico de segregação para cada tipo de população segregante. Quando ocorre a presença de marcadores que não apresentam a segregação mendeliana clássica esperada, diz-se ter ocorrido distorção de segregação (Ferreira, 2006). Segundo Ferreira (2006), a decisão de incluir ou não marcadores com distorção de segregação é discutível. Para alguns autores é recomendável a exclusão desses locos, capazes de afetar os testes de detecção de ligação gênica, alterar as estimativas de distância e a ordem dos marcadores de grupo de ligação particular, comprometendo, portanto, a qualidade do mapa (Oliveira et al., 2004 e Frisch et al., 2004). Por outro lado, alguns trabalhos de mapeamento genético tem utilizado dados de marcadores com distorção na razão de segregação, algumas das vezes indicando-se estes marcadores no grupo de ligação com o nível de probabilidade utilizado no teste de segregação individual, sem saber os efeitos da utilização de tais marcadores (Costa e Silva et al., 2004).

4.2.2. Mapeamento

Foram obtidos nove grupos de ligação definidos pelo agrupamento de 25 marcadores microssatélites, como pode ser observado na Figura 1. Os grupos obtidos constituem segmentos genômicos dos grupos de ligação A1, B1, D1a, G, I, M e O, e o conjunto de marcadores de cada grupo coincidiu com o mapa consenso da soja, exceto para um marcador, o Satt370, que foi agrupado no grupo de ligação I. De acordo com o mapa consenso de Song *et al.* (2004), este marcador localiza-se no grupo de ligação D1a. Duas inversões foram observadas sobre a disposição dos marcadores nos grupos de ligação obtidos.

Vinte e três marcadores não foram agrupados em razão da distância entre as marcas, na maioria das vezes superior a 30 cM, de acordo com o mapa consenso de Song *et al.* (2004). Outra justificativa para algumas das marcas não ligadas, é que para alguns grupos de ligação apenas um marcador foi avaliado nos respectivos grupos. As regiões contendo marcadores não ligados e associados significativamente a alguma das características de interesse, devem ser exploradas pela avaliação da segregação de mais marcadores, possibilitando a análise de mapeamento por intervalos.

O agrupamento dos marcadores nos grupos de ligação foi verificado utilizando-se os critérios de $LOD = 3$ e $r = 0,30$. Para oito dos nove grupos de ligação, o agrupamento também ocorreu utilizando-se o valor de $LOD = 5$. O LOD é considerado um teste de significância para se testar a hipótese de ligação entre dois locos. Quando o LOD score é maior que 3 conclui-se que os locos estão ligados (Schuster e Cruz, 2004). A obtenção dos grupos de ligação foi pré-requisito para a análise de QTL pelos métodos de mapeamento por intervalo simples e composto.

A associação entre marcadores e características foi determinada através da análise de associação por meio de regressão linear. O efeito significativo da regressão pode indicar a existência da associação entre um marcador e a característica, e essa associação é dada pela ligação genética entre o marcador e o QTL (Schuster e Cruz, 2004). As associações detectadas por este método podem ser atribuídas a QTLs de grande efeito e relativamente distantes do marcador, como também, a QTLs de pequeno efeito e próximos ao marcador.

Considerando todas as características avaliadas, 33 marcadores apresentaram associação com pelo menos uma característica e o número de marcadores significativos

para cada característica variou de 1 a 16. Os resultados das análises de marca simples para as onze características em estudo são apresentados na Tabela 2.

Alguns marcadores mostraram associação com mais de uma característica, notavelmente aqueles associados a PTN e OLEO. Treze marcadores comuns associados a essas características foram verificados, possivelmente, devido aos efeitos pleiotrópicos no controle gênico da expressão de tais caracteres.

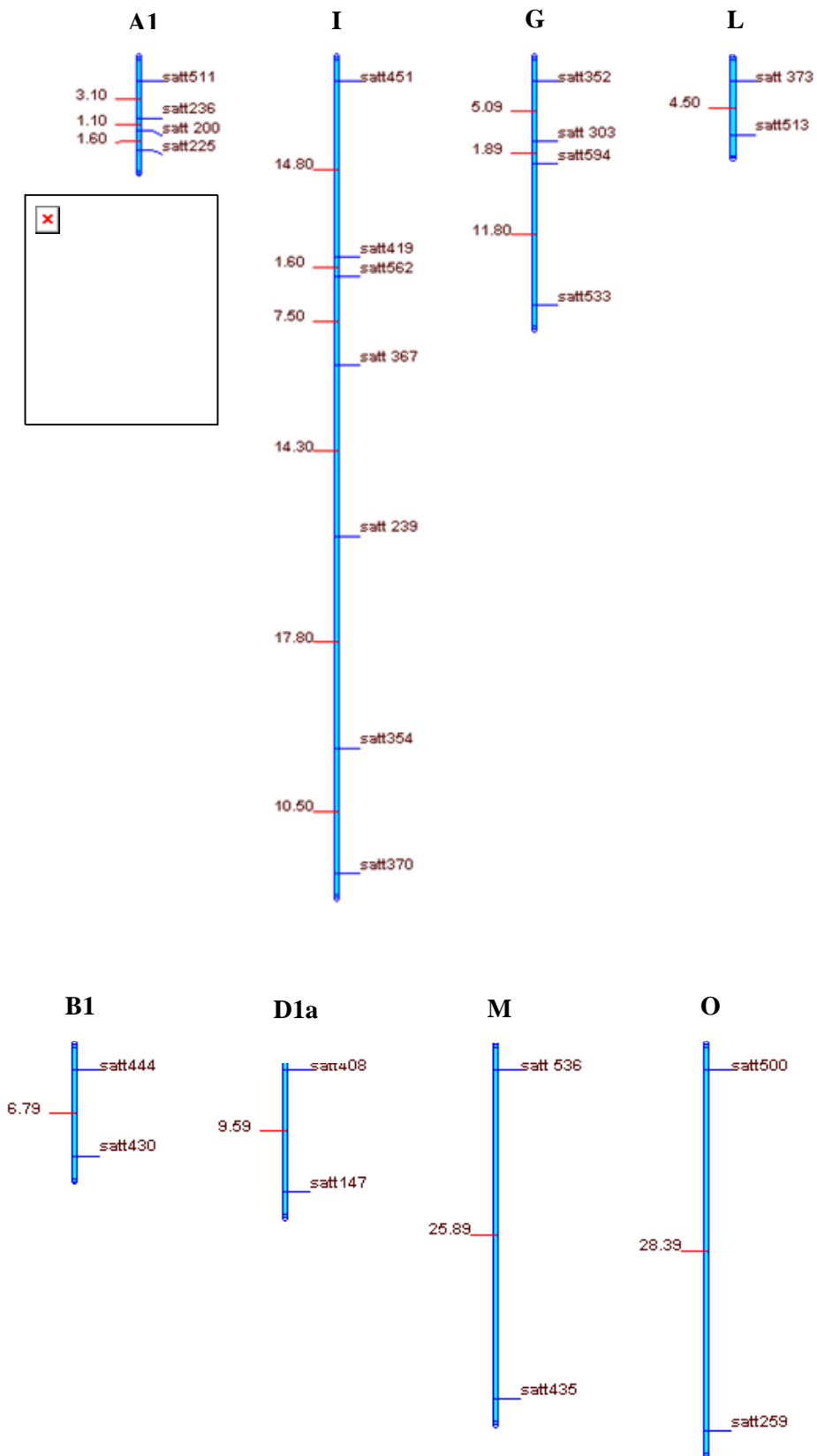


Figura 1. Grupos de ligação definidos utilizando-se uma população de 206 indivíduos F_2 obtidos a partir do cruzamento entre CS3035PTA276-1-5-2 e UFVS2012. Os grupos de ligação foram obtidos adotando-se os critérios de LOD = 3,0 e $r = 0,30$. Os valores da esquerda representam as distâncias entre os marcadores (frequência de recombinação) e à direita estão os nomes dos marcadores. * Marcadores que apresentaram distorção de segregação.

Tabela 2. Análise de associação dos marcadores moleculares com onze características da soja avaliadas na geração F₃ obtida do cruzamento entre CS3035PTA276-1-5-2 (2) e UFVS2012 (0)

PTN						
Marca	Prob (F)	Média (2)	Média (1)	Média (0)	R² (%)	GL
Satt200	0,046*	41,749	41,079	41,389	3,071	A1
Satt300	0,012*	41,333	41,693	40,870	4,581	A1
Satt236	0,048*	41,699	41,066	41,510	3,165	A1
Satt408	0,016*	41,805	41,501	40,854	5,572	D1a
Satt436	0,0**	42,040	41,414	40,625	10,515	D1a
Satt370	0,001**	42,115	41,289	40,750	9,165	D1a
Satt147	0,025*	41,577	41,557	40,912	3,947	D1a
Satt191	0,038*	41,723	41,424	40,957	3,258	G
Satt533	0,0065**	41,719	40,746	41,135	9,566	G
Satt367	0,0**	42,363	41,176	40,650	15,785	I
Satt239	0,0**	42,350	41,295	40,256	23,687	I
Satt562	0,0**	42,184	41,355	40,443	14,523	I
Satt419	0,0**	42,340	41,349	40,537	16,589	I
Satt451	0,0007**	42,186	41,228	40,955	9,84	I
Satt354	0,0004**	42,079	41,454	40,336	13,779	I
OLEO						
Satt200	0,044*	19,705	19,987	19,689	3,105	A1
Satt225	0,018*	19,625	19,875	19,449	6,047	A1
Satt511	0,026*	19,681	20,010	19,700	3,842	A1
Satt236	0,007**	19,673	20,035	19,632	5,173	A1
Satt408	0,0131*	19,553	19,74	20,019	5,275	D1a
Satt436	0,0**	19,502	19,848	20,115	7,749	D1a
Satt370	0,0**	19,292	19,830	20,110	12,137	D1a
Satt147	0,015*	19,611	19,799	20,040	4,444	D1a
Satt147	0,015*	19,611	19,799	20,040	4,444	D1a
Satt191	0,012*	19,596	19,784	20,059	4,372	G
Satt367	0,0**	19,303	19,891	19,277	17,08	I
Satt239	0,0**	19,265	19,837	20,363	25,355	I
Satt562	0,0**	19,288	19,837	20,351	20,613	I
Satt419	0,0**	19,290	19,832	20,301	19,318	I
Satt451	0,0019**	19,417	19,877	20,022	8,582	I
Satt354	0,0**	19,227	19,604	20,225	19,357	I
Satt259	0,035*	19,793	19,673	20,032	3,49	O
NDF						
Satt191	0,033*	138,164	137,447	134,628	3,386	G
Satt594	0,038*	138,518	137,505	138,947	12,785	G
R8_NDM						
Satt191	0,033*	138,164	137,447	134,628	3,386	G
Satt594	0,038*	138,518	137,505	138,947	12,785	G
APM						
Satt454	0,0494*	79,569	86,501	86,813	6,011	A1
Satt436	0,003**	89,885	85,819	82,155	5,745	D1a
Satt408	0,0192*	89,139	87,278	82,81	4,823	D1a
Satt191	0,046*	87,940	86,940	82,831	3,054	G
Satt354	0,048	85,66	86,066	93,301	6,001	I
APV						
Satt300	0,023*	15,290	16,245	16,274	3,943	A1

Tabela 2. Continuação

Marca	Prob (F)	Média (2)	Média (1)	Média (0)	R² (%)	GL
Satt429	0,001**	14,260	16,674	16,129	17,109	A2
Satt436	0,028*	15,423	15,978	16,768	3,637	D1a
Satt263	0,044*	15,643	15,766	16,799	3,191	E
Satt367	0,002**	16,733	15,448	16,847	6,7	I
Satt562	0,003**	16,993	15,506	16,260	5,861	I
Satt419	0,022*	16,767	15,573	16,148	3,903	I
Satt354	0,022*	16,555	15,679	17,349	7,076	I
Satt536	0,004**	17,090	15,483	16,059	6,7	M
Satt512	0,031*	15,863	15,371	16,983	6,27	-
NNM						
Satt454	0,016*	14,364	15,136	14,912	4,54	A1
Satt436	0,026*	15,101	14,968	14,400	3,712	D1a
Satt367	0,033*	14,411	14,997	15,115	3,611	I
Satt562	0,003**	14,286	15,069	15,077	5,795	I
Satt451	0,0292*	14,347	15,085	15,038	4,954	I
Satt373	0,0**	15,255	14,999	14,246	7,724	L
Satt513	0,001**	15,822	15,128	14,255	19,156	L
NVP						
Satt300	0,005**	114,957	113,039	104,574	5,467	A1
Satt454	0,001**	102,027	119,254	119,401	13,217	A1
Satt155	0,004**	113,941	112,883	103,422	5,515	A1
Satt251	0,005**	107,456	110,795	118,983	5,496	B1
Satt301	0,009**	111,988	105,140	96,329	11,023	D2
Satt529	0,049*	110,975	109,600	117,381	2,987	J
Satt536	0,016*	108,958	116,464	108,494	4,296	M
PRO						
Satt300	0,0**	205,464	200,736	178,380	9,759	A1
Satt454	0,0002**	175,282	211,663	214,926	15,87	A1
Satt155	0,0**	201,599	201,621	177,287	8,32	A1
Satt429	0,043*	205,774	178,528	187,373	8,503	A2
Satt147	0,045*	199,828	202,437	187,526	3,313	D1a
Satt301	0,039*	197,070	183,369	170,175	7,699	D2
Satt263	0,033*	191,433	203,992	189,669	3,468	E
Satt562	0,009**	182,372	201,088	199,987	4,833	I
Satt529	0,038*	197,958	190,766	206,762	3,232	J
Satt500	0,009**	177,315	200,872	209,133	10,942	O
PCS						
Satt444	0,0387*	16,669	16,684	17,278	3,511	B1
Satt370	0,001**	16,329	16,946	17,576	9,409	D1a
Satt154	0,0**	17,420	17,164	15,858	11,836	D2
Satt263	0,0**	17,553	17,077	16,173	10,782	E
Satt573	0,002**	16,646	16,802	17,642	6,571	E
Satt303	0,037*	16,614	17,244	17,213	3,742	G
Satt191	0,011*	16,641	16,851	17,483	4,473	G
Satt367	0,002**	16,349	17,257	17,190	6,642	I
Satt239	0,0**	16,239	17,169	17,507	9,496	I
Satt562	0,009**	16,401	17,203	16,973	4,805	I
Satt419	0,004**	16,385	17,265	16,874	5,648	I
Satt529	0,041*	16,451	17,106	17,246	3,157	J
Satt536	0,034*	16,674	17,299	16,746	3,557	M
NSP						
Satt300	0,013*	220,925	217,502	201,346	4,502	A1

Tabela 2. Continuação

Marca	Prob (F)	Média (2)	Média (1)	Média (0)	R² (%)	GL
Satt454	0,0009**	196,754	230,963	230,601	13,379	A1
Satt155	0,015*	218,578	217,510	199,305	4,276	A1

*, **, significativo a 5% e 1% de probabilidade, respectivamente. O número 2 representa o alelo proveniente da linhagem CS3035PTA276-1-5-2, 0 representa o alelo proveniente da variedade UFVS2012 e 1 representa o heterozigoto. Conteúdo de proteína (PTN); conteúdo de óleo (OLEO); número de dias para florescimento (NDF); número de dias para maturação (NDM); altura da planta na maturação (APM); altura da primeira vagem (APV); número de nós na maturação (NNM); número de vagens por planta (NVP); peso de sementes por planta (PRO); peso de cem sementes (PCS) e número de sementes por planta (NSP).

4.2.3. Análise de locos associados ao conteúdo de proteína e óleo

Para a característica PTN foram identificados 15 marcadores com associação significativa, nove dos quais foram altamente significativos. As associações a 1% de probabilidade envolveram marcadores dos grupos de ligação I, D1a, e G. A variação do teor de proteína explicada pelos marcadores variou entre 3,07% e 23,69% (Tabela 2).

Para a característica OLEO foram encontrados 17 marcadores com associação significativa, nove dos quais tiveram associação com a característica a 1% de probabilidade. Os marcadores altamente significativos foram verificados para os grupos de ligação I, D1a e A1. A variação no teor de óleo explicada pelo conjunto de marcadores variou entre 3,11% e 25,36% (Tabela 2).

Considerando apenas as associações dos marcadores a 1% de probabilidade para as características PTN e OLEO, foram encontrados oito marcadores comuns, todos localizados no grupo de ligação I. Os maiores valores de R² verificados tanto para PTN quanto para OLEO foram obtidos para o mesmo marcador, o Satt239 do grupo de ligação I. Para esse marcador, o alelo 2 (CS3035PTA276-1-5-2) determinou teores elevados de PTN e teores reduzidos de OLEO, enquanto o alelo 0 (UFVS2012) determinou a situação inversa.

A análise de mapeamento de QTL foi realizada pelos métodos de intervalo simples e composto, por grupo de ligação separadamente. No método por intervalo composto são incluídos marcadores adicionais como co-fatores na análise, o que aumenta a precisão das estimativas da posição e do efeito do QTL. A seleção dos co-fatores foi baseada na análise de regressão stepwise. A inclusão desses na análise reduz os efeitos provocados por outros QTLs fora do intervalo, o que aumenta consideravelmente o poder do teste (Schuster e Cruz, 2004).

Pelo método de intervalo composto foram detectados quatro QTLs associados ao conteúdo de proteína, nos grupos de ligação D1a, G, A1, e I, e três QTLs associados ao conteúdo de óleo, localizados nos grupos A1, I e O. A variação fenotípica dos teores de proteína e óleo explicada por estes QTLs variou entre 6,24-18,94% e 17,26-25,93%, respectivamente.

Um QTL associado a PTN, que explica 6,24% da variação da característica, foi identificado no intervalo entre os marcadores Satt408 e Satt147 do grupo de ligação D1a. Outro QTL associado a PTN foi detectado na região do marcador Satt533, no grupo de ligação G. Este QTL explica 13,89% da variação do teor de proteína (Figura 2).

No grupo de ligação A1, no intervalo entre os marcadores Satt300 e Satt155, foi identificado um QTL que explica 12,71% da variação do conteúdo de proteína. (Figura 3). Neste mesmo grupo de ligação, no intervalo entre os marcadores Satt200 e Satt225, foi detectado um QTL para OLEO que explica 17,26% da variação da característica (Figura 4).

O QTL que explica a maior proporção da variação do fenótipo para PTN foi identificado no grupo de ligação I próximo ao marcador Satt239, que explica 18,94% da variação da característica. Nesta mesma região do grupo de ligação I foi detectado um QTL para OLEO que explica 25,93% da variação da característica. As análises gráficas destes QTLs são apresentadas nas Figuras 5 e 6.

Outro QTL associado a OLEO foi detectado no grupo de ligação O entre os marcadores Satt500 e Satt259, que explica 18,51% da variação da característica (Figura 7). Os resultados das análises de QTL para PTN e OLEO são apresentados na Tabela 3 e os co-fatores utilizados nessas análises são descritos na Tabela 4.

Brummer *et al.* (1997) detectaram três marcadores associados a QTLs para conteúdo de óleo no grupo de ligação A1 em três populações distintas, avaliadas por três anos. Um QTL, ligado ao marcador A975, foi confirmado nos três anos e explicou entre 7,5% e 12,9% da variação fenotípica. Outros dois QTLs para conteúdo de óleo próximos aos marcadores K400-1 e A104 foram consistentes em dois anos de avaliação. O QTL ligado ao K400-1 explicou 11,5% e 13,7%, enquanto o QTL ligado ao A104 explicou 17,6% e 13,7% da variação da característica em cada ano. Estes dois marcadores foram também associados a QTLs para conteúdo de proteína. No entanto, estes locos não se apresentaram estáveis aos fatores ambientais. O QTL próximo ao marcador K400-1 teve efeito significativo apenas em um ano e explicou 20,3% da

variação do teor de proteína. Já o QTL próximo ao marcador A104 teve efeito significativo em dois anos e explicou 9,4% e 15,4% da variação da característica, em cada ano.

QTLs associados aos conteúdos de proteína e óleo no grupo de ligação A1 foram também mapeados em outros trabalhos (Mansur *et al.*, 1996; Orf *et al.*, 1999; Specht *et al.*, 2001). Outros autores também identificaram QTLs para teor de proteína no grupos de ligação D1a (Brummer *et al.*, 1997; Csanádi *et al.*, 2001) e G (Diers *et al.*, 1992; Brummer *et al.*, 1997).

O grupo de ligação I têm sido associado a QTLs para conteúdo de proteína e óleo em grãos de soja por muitos autores. Chung *et al.* (2003) mapearam um QTL para conteúdo de proteína neste grupo de ligação, no intervalo entre os marcadores Satt239 e Satt496, utilizando 76 RILs obtidas do cruzamento entre PI43088A (acesso com alto teor de proteína) e Asgrow (cultivar com alta produtividade) avaliadas por dois anos. Este QTL explicou 45,3% e 27,7% da variação do teor de proteína nos anos de 1996 e 1999, respectivamente. Neste mesmo intervalo foi mapeado um QTL para óleo, que explicou 27,8% e 24,2% da característica em cada ano.

Na região do marcador Satt239 também foram mapeados QTLs para conteúdo de proteína e óleo em um experimento com 157 RILs avaliadas por dois anos. Estes QTLs explicaram em cada ano, respectivamente, 20,9% e 29,7% da variação do teor de proteína, e 8,6% e 6,8% da variação do teor de óleo (Tajuddin *et al.*, 2003).

Soares (2004) utilizando 95 RILs obtidas do cruzamento entre BARC8 e Garimpo, avaliadas em duas localidades, identificaram um QTL para proteína total no grupo de ligação I, no intervalo entre os marcadores Satt239 e Sat_105. O efeito deste QTL foi consistente entre os ambientes testados e explicou 14,74% e 11,34% da variação da característica nas localidades de Viçosa e São Gotardo (MG), respectivamente.

QTLs para conteúdo de proteína e óleo na região do microsatélite Satt239 foram ainda mapeados por Nichoes *et al.* (2006) utilizando dois grupos de populações de retrocruzamento. A partir dos resultados das análises de um grupo de populações, foi sugerido que o QTL que controla teor de proteína e óleo no grupo de ligação I encontra-se no intervalo de 11 cM entre os marcadores Satt614 e Satt354. No entanto, os resultados obtidos a partir do segundo grupo de populações redefiniram que o QTL contido nesta região pode estar delimitado no intervalo de 3 cM entre o microsatélite Satt239 e o AFLP ACG9b.

Locos detectados para conteúdo de proteína em regiões próximas ao marcador Satt239 do grupo de ligação I, por diversos autores e envolvendo diferentes *backgrounds* genéticos, indicam a presença de genes conservados nestas regiões, possivelmente envolvidos no controle da expressão de proteínas. A existência de marcadores comuns neste grupo associados a QTLs que influenciam teores de proteína e óleo, possivelmente, deve-se aos efeitos pleiotrópicos de genes nestas regiões.

Com base nestes fatores, torna-se fundamental a análise de mais marcadores na região próxima ao Satt239 em estudos de mapeamento de locos para conteúdo de proteína e óleo, no sentido de determinar com precisão a localização destes QTLs. A certeza destes intervalos poderá ser de grande importância para a clonagem destes locos e elucidação da estrutura e função dos genes, bem como, seus papéis na determinação dos conteúdos de proteína e óleo em soja.

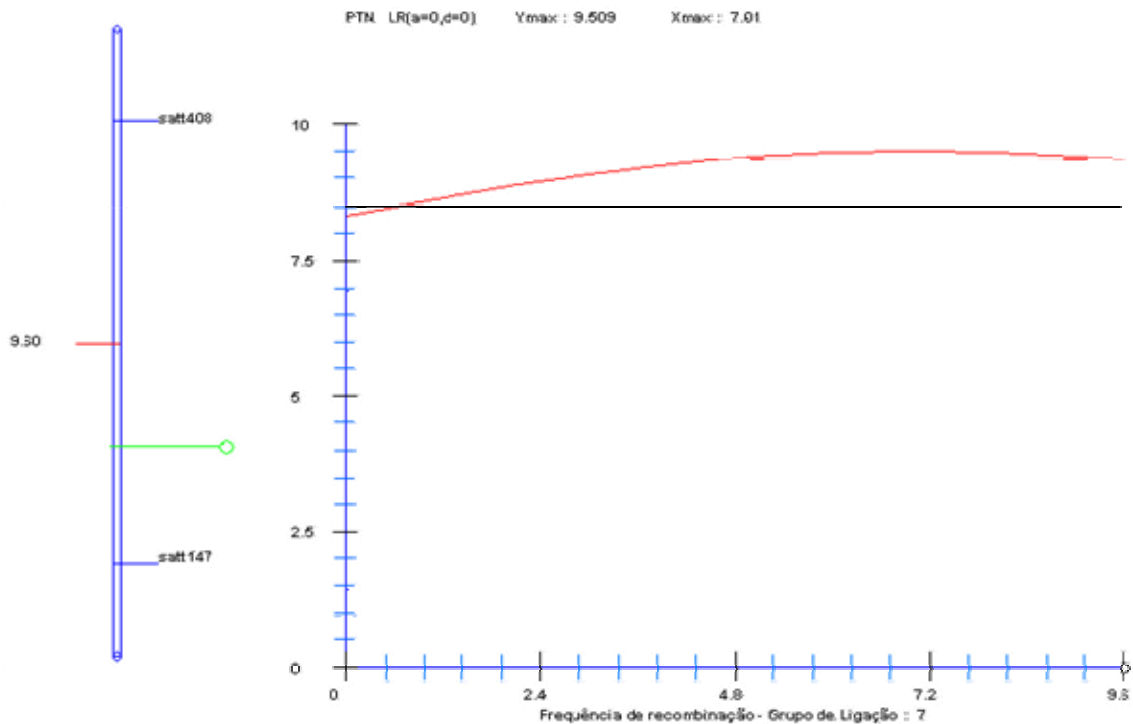


Figura 2. QTL associado ao conteúdo de proteína que explica 6,4% da variação da característica. Segmento do grupo de ligação D1a representado à esquerda e curva da razão de verossimilhança à direita. O ponto de corte refere-se ao valor de LR crítico 8,99 obtido a partir de 1000 simulações, ao nível de significância de 5%.O ponto mais alto da curva representa o valor de LR de 9,51 distante 7,01 cM do marcador Satt408.

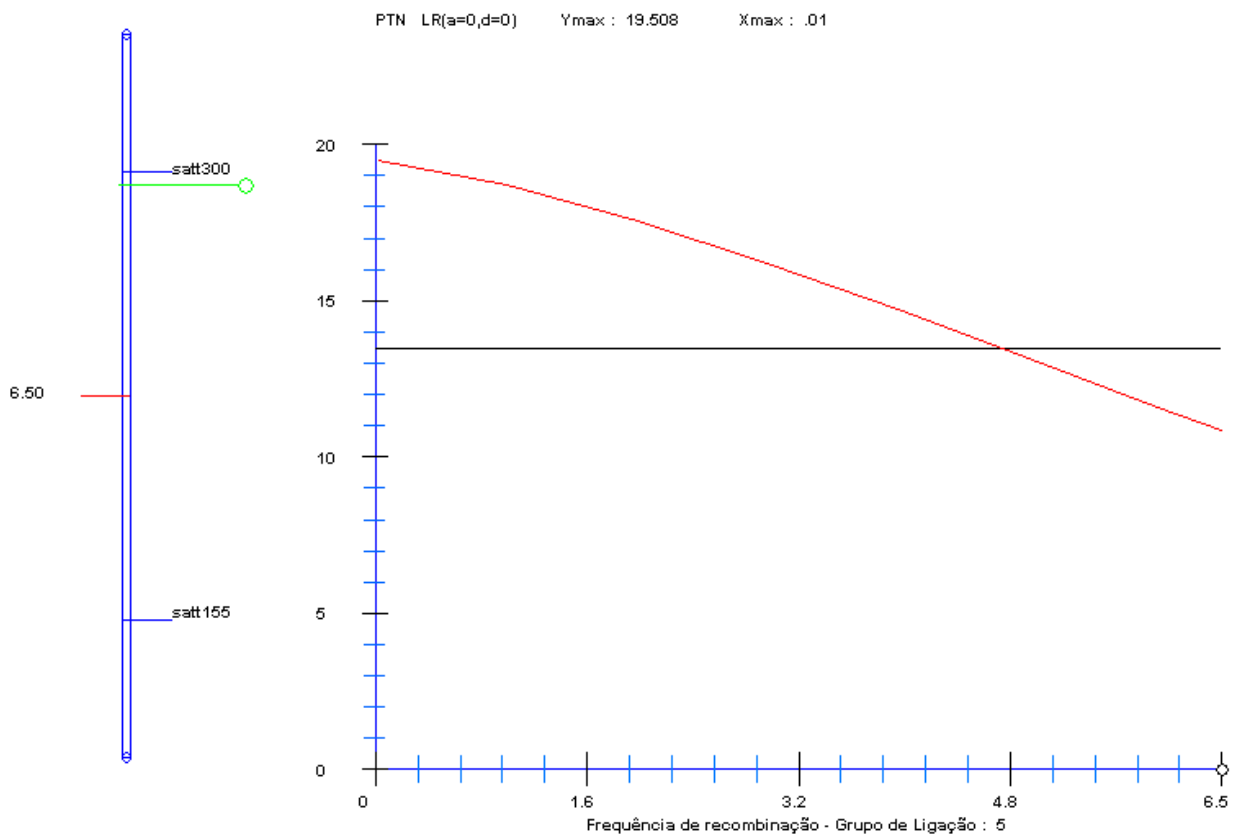


Figura 3. QTL associado ao conteúdo de proteína que explica 12,71% da variação da característica. Segmento do grupo de ligação A1 representado à esquerda e curva da razão de verossimilhança à direita. O ponto de corte refere-se ao valor de LR crítico 13,68 obtido pelo teste com 1000 simulações, ao nível de significância de 1%. O ponto mais alto da curva representa o valor de LR de 19,51 distante 0,01 cM do marcador Satt300.

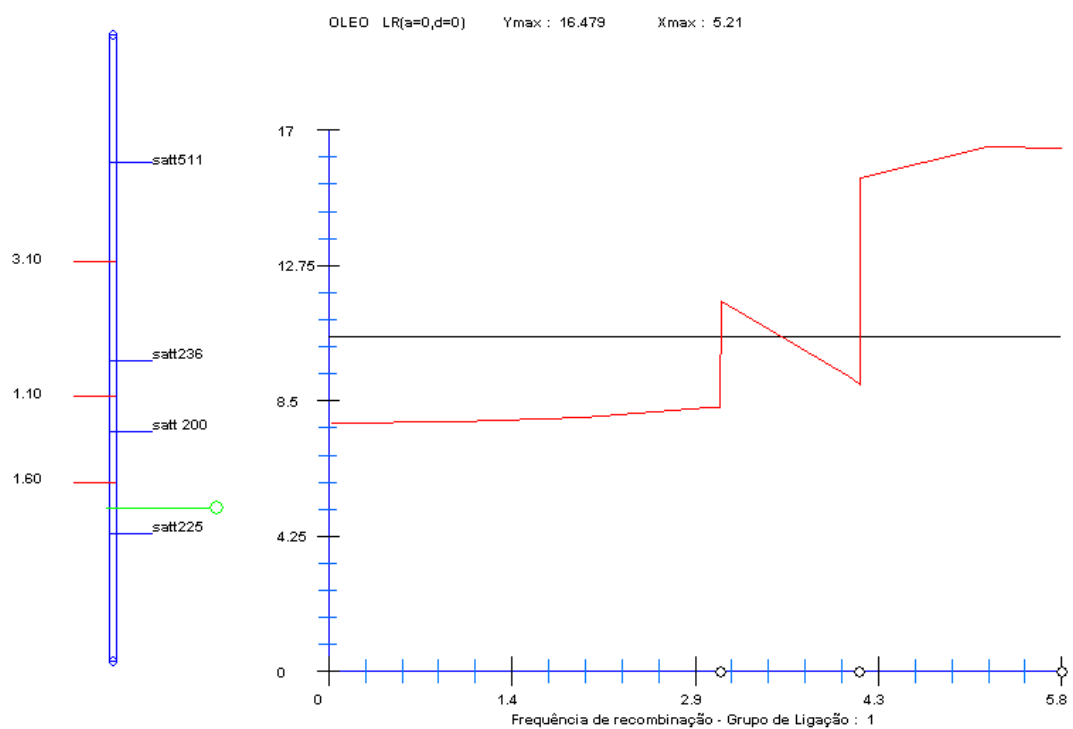


Figura 4. QTL associado ao conteúdo de óleo que explica 17,26% da variação da característica. Segmento do grupo de ligação A1 representado à esquerda e curva da razão de verossimilhança à direita. O ponto de corte refere-se ao valor de LR crítico 12,91 obtido pelo teste com 1000 simulações, ao nível de significância de 1%. O ponto mais alto da curva representa o valor de LR de 16,48 distante 5,21 cM do marcador Satt511.

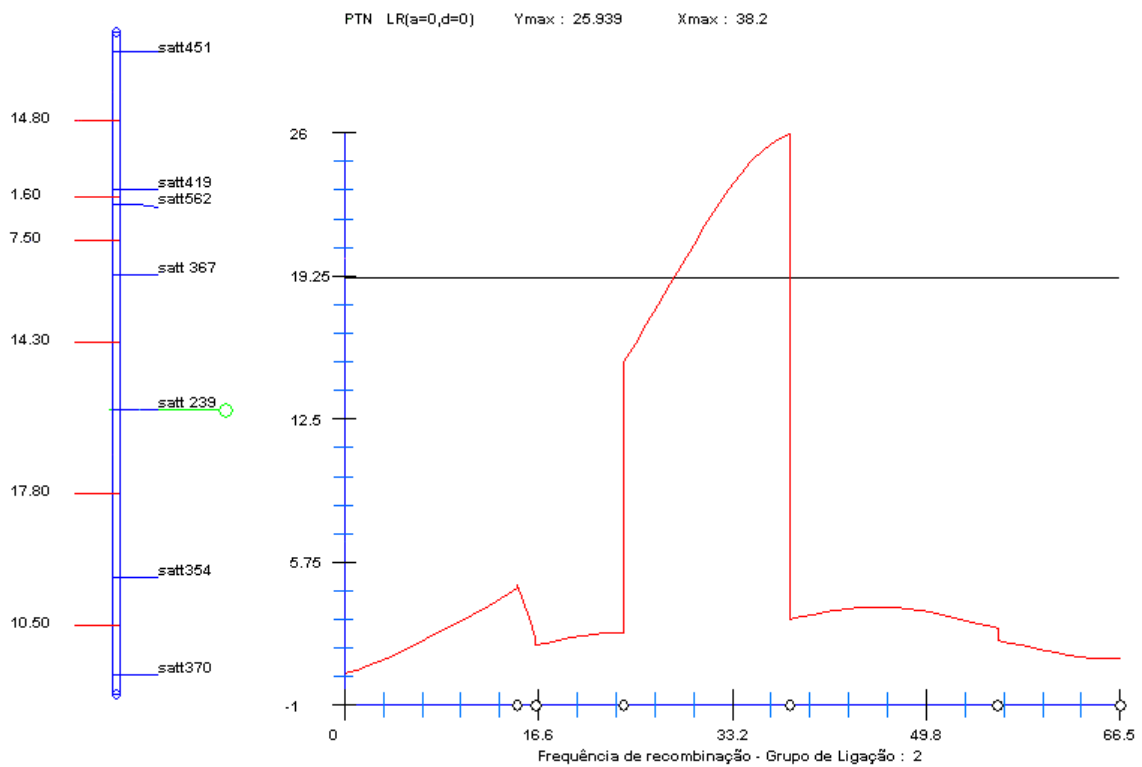


Figura 5. QTL associado ao conteúdo proteína que explica 18,94% da variação da característica. Segmento do grupo de ligação I representado à esquerda e curva da razão de verossimilhança à direita. O ponto de corte refere-se ao valor de LR crítico 12,91 obtido pelo teste com 1000 simulações, ao nível de significância de 1%. O ponto mais alto da curva representa o valor de LR de 25,94 distante 38,2 cM do marcador Sattu451.

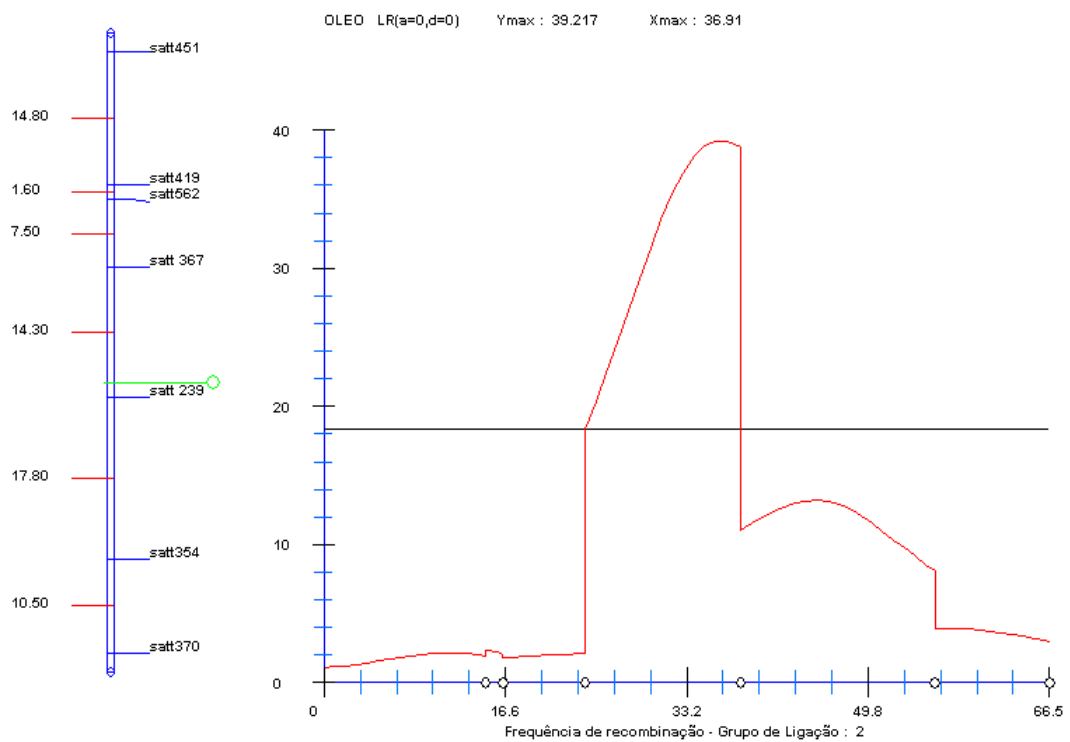


Figura 6. QTL associado ao conteúdo de óleo. Segmento do grupo de ligação I representado à esquerda e curva da razão de verossimilhança à direita. O ponto de corte refere-se ao valor de LR crítico 21,06 obtido pelo teste com 1000 simulações, ao nível de significância de 1%. O ponto mais alto da curva representa o valor de LR de 39,22 distante 36,91 cM do marcador Satt451.

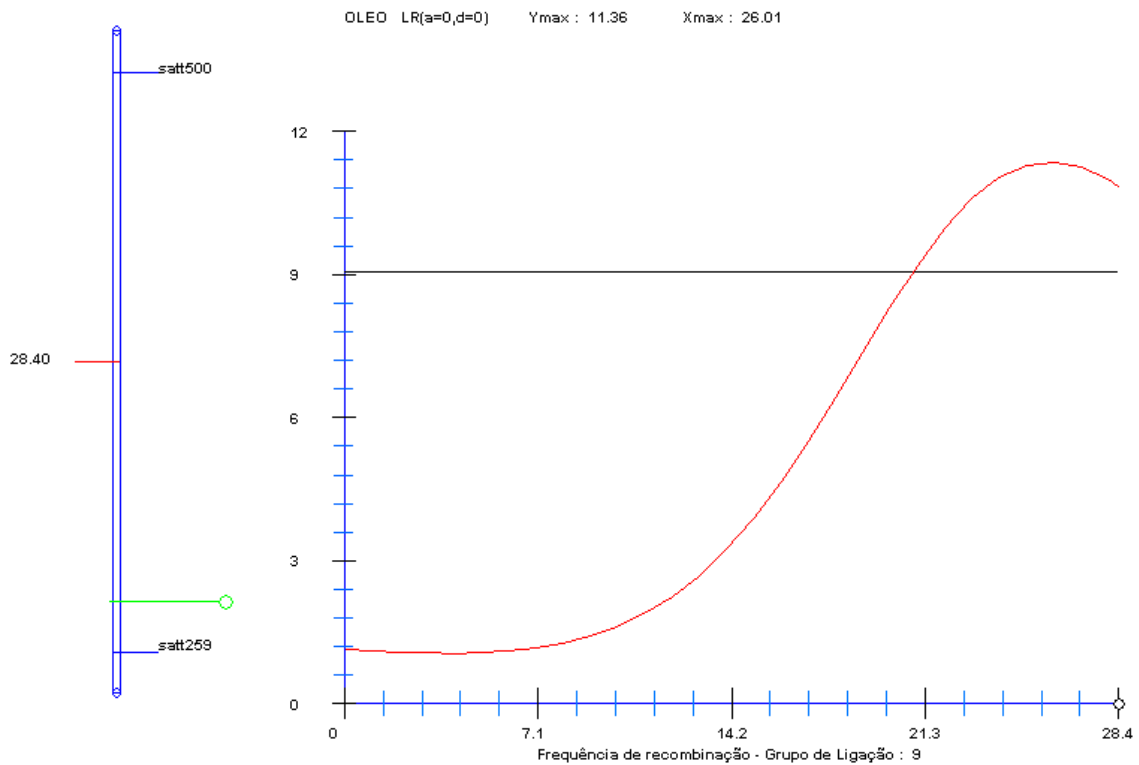


Figura 7. QTL associado ao conteúdo de óleo. Segmento do grupo de ligação O representado à esquerda e curva da razão de verossimilhança à direita. O ponto de corte refere-se ao valor de LR crítico 21,06 obtido pelo teste com 1000 simulações, ao nível de significância de 1%. O ponto mais alto da curva representa o valor de LR de 39,22 distante 36,91 cM do marcador Satt451

Tabela 3. QTLs associados aos conteúdos de proteína e óleo no grão de soja

Característica	GL	Intervalo/Região	LR	R ² %	a	d
PTN	A1	Satt300-Satt155	19,51**	12,71	0,434	1,670
	D1a	Satt408-Satt147	9,51*	6,24	0,477	0,606
	G	Satt533	12,58*	13,89	0,234	-1,815
	I	Satt239	25,94**	18,94	0,862	-0,814
OLEO	A1	Satt200-Satt225	164,48**	17,26	0,056	0,963
	I	Satt239-Satt367	39,22**	25,93	-0,562	-0,034
	O	Satt500-Satt259	11,36*	18,51	-0,215	-1,442

GL = Grupos de ligação denominados de acordo Song *et al.* (2004). R²% = proporção da variação fenotípica explicada pelo QTL; a = efeito aditivo estimado para o QTL. d = efeito da dominância estimado para o QTL. * e ** valores de LR excederam os valores críticos de corte a 5% e 1%, respectivamente. Os valores de LR críticos foram determinados pela realização de 1000 permutações.

Tabela 4. Seleção de co-fatores para a análise de QTL por intervalo composto para as características conteúdo de proteína e óleo no grão

Característica	GL	Co-fatores
PTN	A1	Satt239 - Satt594 - Satt536
	D1a	Satt200
	G	Satt200 - Satt147 - Satt536
	I	Satt200 - Satt594 - Satt408 - Satt536 - Satt367 - Satt239
OLEO	A1	Satt239 - Satt594 - Satt147 - Satt430 - Satt259
	I	Satt200 - Satt594 - Satt430 - Satt147 - Satt239 - Satt367 - Satt259
	O	Satt200 - Satt239 - Satt594 - Satt147

GL = Grupos de ligação denominados de acordo Song *et al.* (2004). Conteúdo de proteína (PTN) e conteúdo de óleo (OLEO).

4.2.4. Análise de locos associados a componentes de produção de grãos

Na análise de marca simples para os componentes de produção PRO e PCS foram identificados 10 e 12 marcadores com associação significativa, respectivamente. Os valores de R² encontrados para esses marcadores variaram entre 3,23-15,87% para PRO e 3,16-11,84% para PCS. Para os caracteres NVP e NSP foram identificados 7 e 3 associações significativas, respectivamente. Para os marcadores associados a NVP e NSP, estes valores variaram entre 2,99-13,27% e 4,28-13,78%, respectivamente.

Na análise de QTL para PRO foram identificados dois QTLs nos grupos de ligação A1 e D1a que explicam 12,32% e 9,03% da variação da característica, respectivamente. As análises gráficas destes QTL são apresentadas na Figuras 8 e 9.

Um QTL associado a PCS foi identificado no grupo de ligação I. Este QTL foi localizado na região do marcador Satt239 e explica 13,47% da variação da característica (Figura 10).

Outros dois QTLs associados a NVP e NSP foram detectados no grupo de ligação A1, no intervalo entre os marcadores Satt300 e Satt155. A variação do fenótipo explicada por estes QTLs foi de 9,43% para NVP e 7,19% para NSP. As análises gráficas destes QTLs são apresentadas nas Figuras 11 e 12. Os resultados das análises de QTL para os componentes de produção de grãos são descritos na Tabela 5 e os cofatores utilizados nessas análises são descritos na Tabela 6.

Orf *et al.* (1999) identificaram um QTL no grupo de ligação A1 para peso de sementes em uma população de RILs derivada do cruzamento entre Minsoy e Noir 1. Este QTL explicou 7% da variação da característica e foi localizado próximo ao marcador Satt174. Para esta população, foram mapeados mais seis QTLs nos grupos de ligação A2, G, D1a, M e L, que explicaram entre 6% a 8% da variação da característica. Outros nove QTLs de pequeno efeito foram ainda identificados a partir de duas outras populações. Um QTL associado a peso de sementes no grupo de ligação D1a foi também identificado por estes autores a partir de duas populações de RILs derivadas dos genitores Minsoy e Archae.

Um QTL associado a peso de sementes foi também identificado no grupo de ligação I na região próxima ao marcador Satt562 por Csanádi *et al.* (2001). Este QTL explicou 11,6% da variação do peso de mil sementes.

Quatro QTLs associados a peso de sementes foram mapeados por Zhang *et al.* (2004) a partir de 184 RILs. Dois QTLs do grupo de ligação D2, localizados nos intervalos A611D-B146H e B146H-Satt458, explicaram 9,2% e 11,4% para PCS, respectivamente. Outros dois QTLs foram detectados nos grupos de ligação A2 e B1, que explicaram 6,7% e 10,2% da variação da característica.

Outros três QTLs para peso de sementes foram detectados por Vieira *et al.* (2006) utilizando 118 RILs obtidas do cruzamento entre as variedades BARC-8 e Garimpo. O QTL detectado no intervalo Satt475-OPAW09a, do grupo de ligação K, explicou 9,4% da variação da característica. Outros dois QTLs de pequeno efeito foram localizados nos grupos de ligação G e O, que explicaram 6,8% e 7% da variação para PCS. Nestes grupos foram ainda detectados QTLs associados ao número de sementes por planta, que explicaram em torno de 7% da variação da característica.

As associações significativas entre os marcadores avaliados e as demais características podem ser visualizadas na Tabela 7. Para grande parte das associações não houve o agrupamento dos marcadores, impossibilitando a análise de QTL. No entanto, a significância dos marcadores pressupõe a sua associação com QTLs. Mais marcadores nestas regiões devem ser, portanto, analisados. A existência destes QTLs poderá, então, ser verificada pelas análises de mapeamento por intervalo simples e composto.

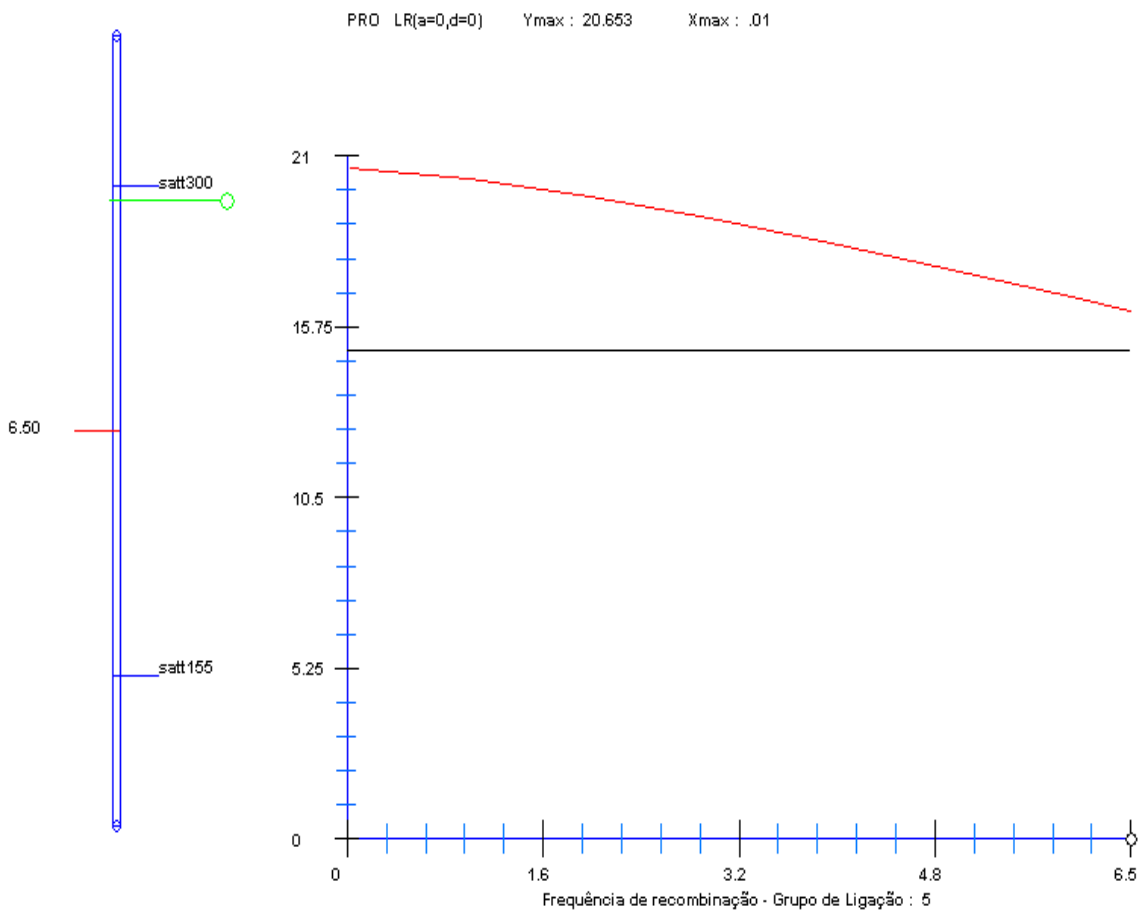


Figura 8. QTL associado ao peso de sementes por planta que explica 12,32% da variação do fenótipo. Segmento do grupo de ligação A1 representado à esquerda e curva da razão de verossimilhança à direita. O ponto de corte refere-se ao valor de LR crítico 17,19 obtido pelo teste com 1000 simulações, ao nível de significância de 1%. O ponto mais alto da curva representa o valor de LR de 20,65 na posição 0,01 cM próximo ao marcador Satt300.

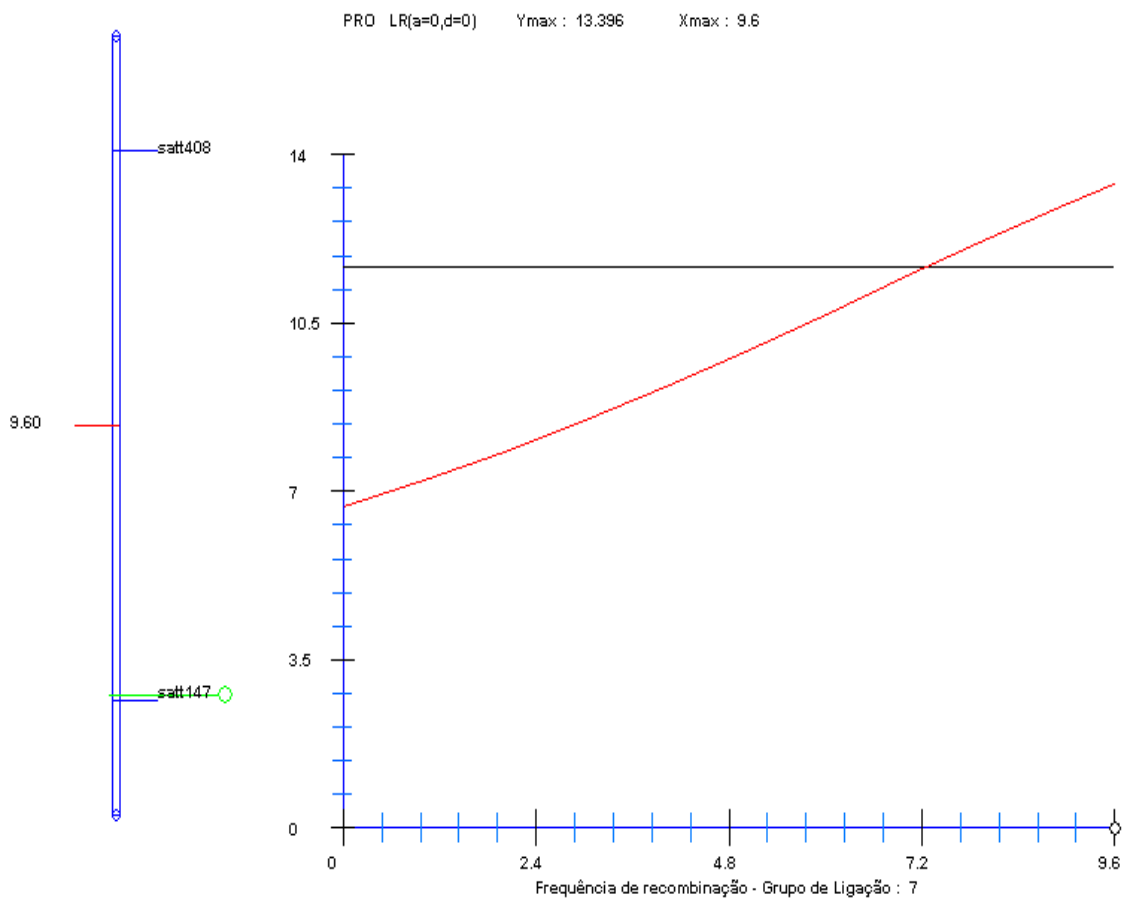


Figura 9. QTL associado ao peso de sementes por planta que explica 9,03% da variação do fenótipo. Segmento do grupo de ligação D1a representado à esquerda e curva da razão de verossimilhança à direita. O ponto de corte refere-se ao valor de LR crítico 13,4 obtido pelo teste com 1000 simulações, ao nível de significância de 1%. O ponto mais alto da curva representa o valor de LR de 13,4 distante 9,6 cM do marcador Satt408.

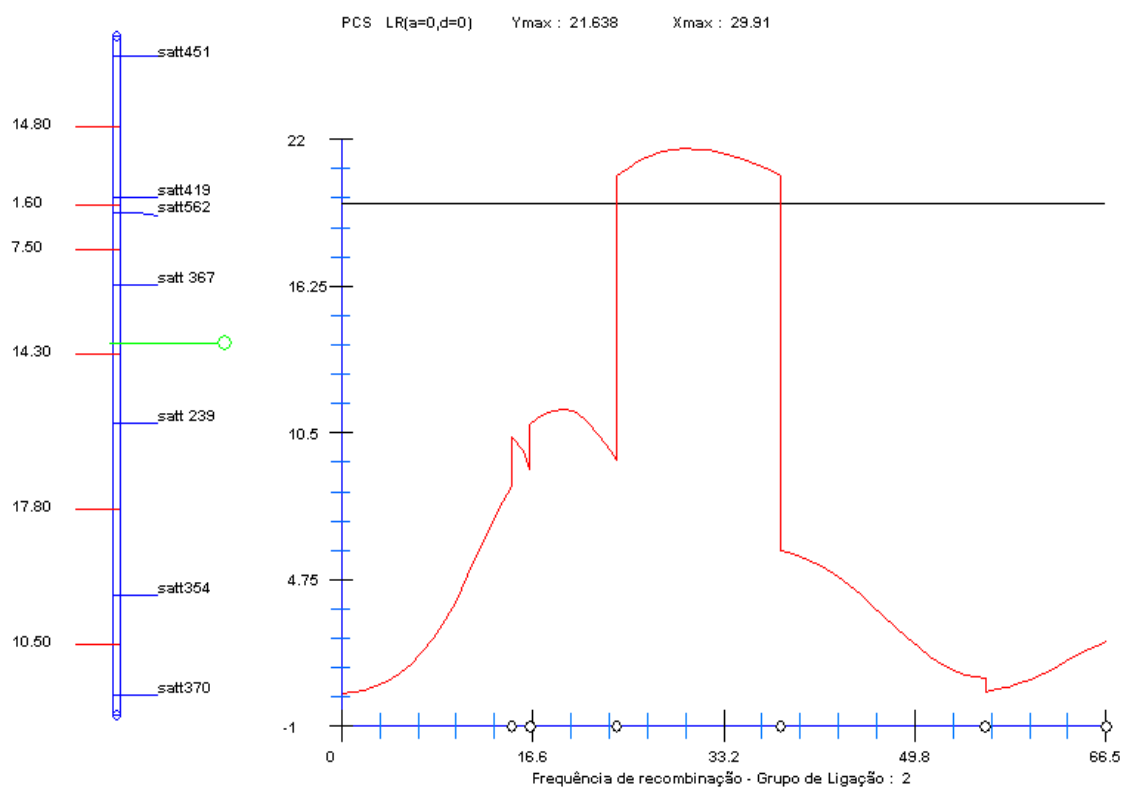


Figura 10. QTL associado a peso de cem sementes que explica 13,47% da variação do fenótipo. Segmento do grupo de ligação I representado à esquerda e curva da razão de verossimilhança à direita. O ponto de corte refere-se ao valor de LR crítico 20,02 obtido pelo teste com 1000 simulações, ao nível de significância de 1%. O pico da curva representa o valor de LR de 21,64 distante 29,91 cM do marcador Satt451.

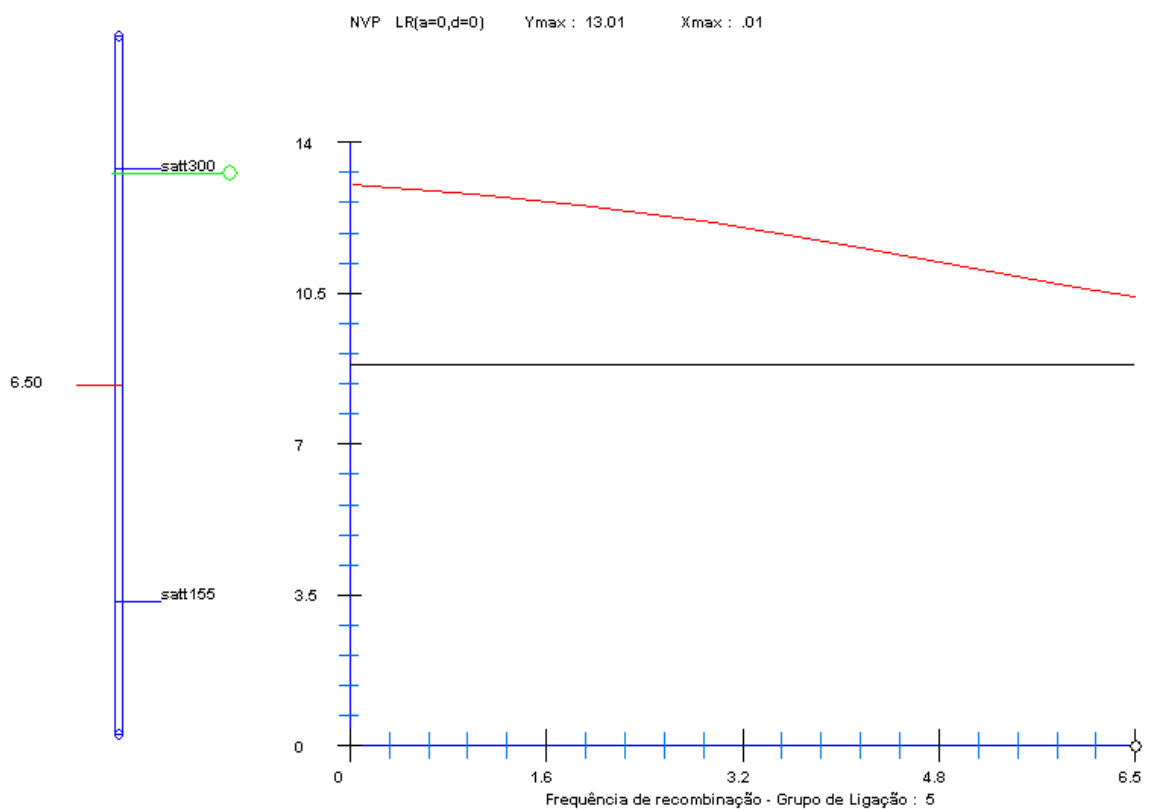


Figura 5. QTL associado ao número de vagens por planta que explica 9,4% da variação do fenótipo. Segmento do grupo de ligação A1 representado à esquerda e curva da razão de verossimilhança à direita. O ponto de corte refere-se ao valor de LR crítico 12,3 obtido pelo teste com 1000 simulações, ao nível de significância de 1%. O pico da curva representa o valor de LR de 13,01 na posição 0,01 cM do marcador Satt300.

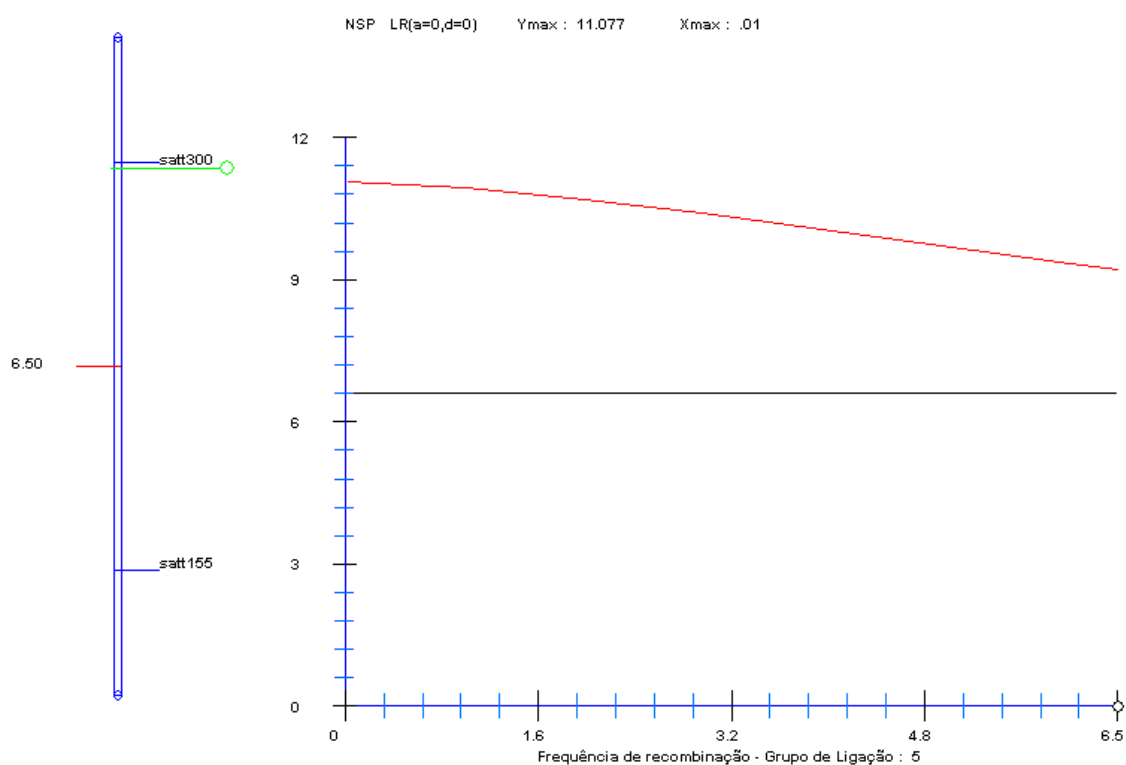


Figura 6. QTL associado ao número de sementes por planta que explica 7,19% da variação do fenótipo. Segmento do grupo de ligação A1 representado à esquerda e curva da razão de verossimilhança à direita. O ponto de corte refere-se ao valor de LR crítico 9,04 obtido pelo teste com 1000 simulações, ao nível de significância de 5%. O pico da curva representa o valor de LR de 11,08 na posição 0,01 cM do marcador Satt300.

Tabela 5. QTLs associados aos componentes de produção de grãos em soja

Característica	GL	Intervalo/Região	LR	R²%	a	d
PRO	A1	Satt300-Satt155	20,65**	0,123	15,511	20,403
	D1a	Satt408-Satt147	13,40**	9,03	10,051	28,281
PCS	I	Satt367-Satt239	21,63**	13,47	-0,602	1,242
NVP	A1	Satt300-Satt155	13,01**	0,094	7,167	7,292
NSP	A1	Satt300-Satt155	11,08*	7,19	13,38	19,387

GL = Grupos de ligação denominados de acordo Song *et al.* (2004). R²% = proporção da variação fenotípica explicada pelo QTL; a = efeito aditivo estimado para o QTL. d = efeito da dominância estimado para o QTL. * e ** valores de LR excederam os valores críticos de corte a 5% e 1%, respectivamente. Os valores de LR críticos foram determinados pela realização de 1000 permutações.

Tabela 6. Seleção de co-fatores para a análise de QTL por intervalo composto para os componentes de produção de grãos em soja

Característica	GL	Co-fatores
PRO	A1	Satt536 – Satt 147
	D1a	Satt200 – Satt562
PCS	I	Satt594 – Satt367 – Satt239 – Satt536
NVP	A1	Satt373 – Satt352 – Satt536 – Satt147
NSP	A1	Satt200 – Satt536 – Satt147

GL = Grupos de ligação denominados de acordo Song *et al.* (2004). Peso de sementes por planta (PRO); peso de cem sementes (PCS); número de vagens por planta (NVP) e número de sementes por planta (NSP).

Tabela 7. Associação marcador – característica para as onze características da soja

Marca	GL	PTN	OLEO	PRO	PCS	NDF	NDM	APM	APV	NNM	NVP	NSP	Posição ¹
Satt454	A1			**				*			**	**	28,08
Satt300	A1	*		**					*		**	*	30,93
Satt155	A1			**							**	*	32,68
Satt200	A1	*	*										92,89
Satt236	A1	*	**										93,23
Satt511	A1		*										94,20
Satt225	A1		*										95,16
Satt429	A2			*						**			162,03
Satt251	B1					*					**		36,48
Satt444	B1				*								85,92
Satt370	D1a	**	**		**								60,99
Satt436	D1a	**	**					**	*	*			70,69
Satt408	D1a	*	*					*					106,69
Satt147	D1a	*	*	*		*							108,89
Satt301	D2			*							**		93,71
Satt154	D2				**								57,07
Satt573	E				**								35,79
Satt263	E			*	**				*				45,4
Satt533	G	**											56,52
Satt191	G	*	*		*		*	*					96,57
Satt451	I	**	**								*		20,34
Satt419	I	**	**		**				*				21,90
Satt562	I	**	**	**	**				**	**			22,84
Satt367	I	**	**		**				**	*			27,98
Satt239	I	**	**		**								36,94
Satt354	I	**	**					*	*				46,22
Satt529	J			*	*						*		41,9
Satt373	K									**			107,24
Satt513	L									**			106,37
Satt536	M				*				**		*		62,14
Satt500	O			**									14,17
Satt259	O		*										39,82
Satt512	-								*				-

* e **significativos a 5% e 1% de probabilidade. Posição¹ de acordo com o mapa consenso de Song *et al.* (2004). Conteúdo de proteína (PTN); conteúdo de óleo (OLEO); peso de sementes por planta (PRO); peso de cem sementes (PCS); número de dias para florescimento (NDF); número de dias para maturação (NDM); altura da planta na maturação (APM); altura da primeira vagem (APV); número de nós na maturação (NNM); número de vagens por planta (NVP) e número de sementes por planta (NSP).

5. CONCLUSÕES

- i. A segregação de 48 marcadores foi avaliada em 206 indivíduos F₂ obtidos do cruzamento entre a linhagem CS3035PTA276-1-5-2 (alto teor de proteína e baixo teor de óleo) e a variedade UFVS2012 (baixo teor de proteína e alto teor de óleo). Foram obtidos 9 grupos de ligação formados pelo agrupamento de 25 marcadores.
- ii. Considerando todas as características avaliadas, 33 marcadores apresentaram associação com pelo menos uma característica e o número de marcadores significativos para cada característica variou de 1 a 16.
- iii. Para conteúdo de proteína (PTN) e óleo (OLEO) foram verificados 15 e 16 marcadores com associação significativa, respectivamente. Considerando apenas as associações altamente significativas, para cada característica foram encontrados nove marcadores, seis dos quais foram comuns. A variação nos teores de proteína e de óleo explicada por estes marcadores variou entre 3,07-23,69% e 3,11-25,36%, respectivamente.
- iv. Para produção de grãos (PRO) e peso de cem sementes (PCS) foram verificados 10 e 12 marcadores com associação significativa, respectivamente. A variação fenotípica para PRO e PCS explicada pelos grupos de marcadores variou entre 3,23-15,87% e 3,16-11,84%, respectivamente.
- v. Para os outros componentes de produção de grãos NSP (número de sementes por planta) e NVP (número de vagens por planta) foram identificados 7 e 3 marcadores com associação significativa, respectivamente. A variação fenotípica para NSP e NVP explicada por estes marcadores variou entre 2,99-13,27% e 4,28-13,78%, respectivamente.
- vi. No grupo de ligação A1 foram identificados QTLs associados as características PTN, OLEO, PRO, NVP e NSP. Os QTLs detectados no intervalo entre os marcadores Satt300-Satt155 explicaram 8,4% da variação para PTN, 9,43% da variação para NVP e 7,19% da variação para NSP . Esse intervalo foi também associado a um QTL para PRO, que explicou 12,32% da variação da característica. E no intervalo entre os marcadores Satt200-Satt225, neste mesmo grupo de ligação, foi detectado um QTL para óleo que explicou 17,26% da variação da característica.

- vii. No grupo de ligação I foram identificados três QTLs próximos ao marcador Satt239 associados as características PTN, OLEO e PCS. Estes QTLs explicaram 18,94%, 25,93% e 13,47% da variação fenotípica para tais características, respectivamente.
- viii. No grupo de ligação D1a foi detectado um QTL para PTN que explicou 6,24% da variação da característica, no intervalo entre os marcadores Satt408 e Satt147. Neste mesmo intervalo, foi detectado um QTL para PRO que explicou 9,03% da variação fenotípica.
- ix. Um QTL para OLEO foi ainda detectado no grupo de ligação O, no intervalo entre os marcadores Satt500 e Satt259. Este QTL explicou 18,51% da variação no conteúdo de óleo.
- x. Pode-se concluir que marcadores próximos ao microssatélite Satt239 estão ligados a locos envolvidos com o controle genético dos conteúdos de proteína e óleo. Estes locos são conservados entre diferentes *backgrounds* genéticos e exercem efeitos pleiotrópicos, influenciando simultaneamente os teores de proteína e óleo no grão de soja.
- xi. A população utilizada neste trabalho mostrou-se adequada para o estudo de mapeamento de QTL. O número considerável de associações significativas, detectadas entre os marcadores avaliados e as características em estudo, pressupõe a existência de outros QTLs e demonstra o potencial desta população para fins de mapeamento genético. Estes fatores justificam a continuidade do trabalho. Portanto, mais marcadores devem ser avaliados no sentido de mapear outros locos associados ao conteúdo de proteína e óleo, características de maior interesse do programa de melhoramento da qualidade da soja, desenvolvido pelo BIOAGRO/Universidade Federal de Viçosa.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ABDELNNOOR, R.V.; BARROS, E.G.; MOREIRA, M.A. Determination of diversity within Brazilian soybean germoplasm using random amplified polymorphic DNA techniques and comparative analysis with pedigree data. **Brazilian Journal of Genetics**, v.18, p.265-273, 1995.

Agencia Lusa (2008) <<Disponível em: <http://www.agencialusa.com.br>>> Acesso em 10/01/08.

AKKAYA, M.S.; SHOEMAKER, R.C.; SPECHT, J.E.; BHAGWAT, A.A.; CREGAN, P.B. Integration of simple sequence repeat DNA markers into a soybean linkage map. **Crop Science**, v.35, p.1439-1445, 1995.

ARANTES, N.E.; SOUZA, P.I.M. (Ed.) **Cultura da soja nos cerrados**. Piracicaba: Potafos. p.535. 1993.

BONETTI, L.P. **Cultivares e seu melhoramento genético**. In: Vernetti, F.J. Soja – Genética e Melhoramento. Campinas: Fundação Cargill, p.741-800, 1983.

BORÉM, A.; CAIXETA, E.T. **Marcadores Moleculares**. Viçosa: UFV. 374p. 2006.

BOUCHEZ, A.; HOSPITAL, F.; CAUSSE, M.; GALLAIS, A.; CHARCOSSET, A. Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. **Genectis Society of América**, v.162, p.1945-1959, 2002.

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, p.370-378, 1997.

CERVIGNI, G.D.L.; SCHUSTER, I.; BARROS, E.G.; MOREIRA, M.A. Two microsatellite markers flanking a dominant gene for resistance to soybean cyst nematode race 3. **Theoretical and Applied Genetics**, v.135, p.99-105, 2004.

CHURCHILL, G.A.; DOERGE, R.W. Empirical threshold values for quantitative trait mapping. **Genetics**, v.138, p.963-971, 1994.

CONAB - Companhia Nacional de Abastecimento (Levantamento 2007) <<Disponível em: <http://www.conab.gov.br>>> Acesso em 13/02/2008

CONCIBIDO, V.C.; VALLE, B.L.; MCLAIRD, P.; PINEDA, N.; MEYER, J.; HUMMEL, L.; YANG, J.; WU, K.; DELANNAY, X. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. **Theoretical and Applied Genetics**, v.106, p.575-582, 2003.

CONCIBIDO, V.C.; DENNY, R.L.; BOUTIN, S.R.; HOUTEA, R.; ORF, J.H.; YOUNG, N.D. DNA marker analysis of loci underlying resistance to soybean cyst nematode (*Heterodera glycines Ichinohe*). **Crop Science**, v.34, p.240-246, 1994.

CONCIBIDO, V.C.; DOUGLAS, L.; DENNY, R.L.; ORF, J.H.; YOUNG, N.D. Genome mapping of soybean cyst nematode resistance genes in Peking, PI 90763 and PI 88788 using DNA markers. **Crop Science**, v.37, p.258-264, 1997.

CHUNG, J.; BABKA, H.L.; GRAEF, G.L.; STASWICK, P.E.; LEE, D.J.; CREGAN, P.B.; SHOEMAKER, R.C.; SPEECH, J.E. The seed protein, oil, and yield QTL on soybean linkage group I. **Crop Science**, v.43, p.1053-1067, 2003.

CREGAN, P.B. Simple sequence repeat DNA length polymorphisms. **Probe Spring**, p.18-22, 1992.

CREGAN, P.B.; JARVIK, T.; BUSH, A.L.; SHOEMAKER, R.C.; LARK, K.G.; KAHLER, A.L.; KAY, N.; VANTO, T.T.; LOHNESS, D.G.; CHUNG, J.; SPEECH, J.E. An Integrated Genetic Linkage Map of the Soybean Genome. **Crop Science**, v.39, p.1464-1490, 1999.

CREGAN, P.; MUDGE, J.; FICHUS, E.W.; DENNY, D.R.; YOUNG, N.D. Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. **Theoretical and Applied Genetics**, v.99, p.811-818, 1999.

SCHUSTER, I.; CRUZ, C. D. **Estatística genômica aplicada a populações derivadas de cruzamentos controlados**. Viçosa: UFV. 568p. 2004.

CSANÁDI, G.; VOLLMANN, J.; STIFT, G.; LELLEY, T. Seed Quality QTLs Identified in a Molecular Map of Early Maturing Soybean. **Theoretical and Applied Genetics**, v.103, p.912-919, 2001.

DANESH, D.; PENUELA, S.; MUDGE, J.; DENNY, R.L.; NORDSTROM, H.; MARTINEZ, J.P.; YOUNG, N.D. Bacterial artificial chromosome library for soybean and identification of clones near a major cysts nematode resistance gene. **Theoretical and Applied Genetics**, v.96, p.196-202, 1998.

DEGEORGE, R.W.; REBAI, A. Significance threshold for QTL interval mapping tests. *Hereditas*, v.76, p.59-64, 1996.

DIERS, B.W.; KEIM, P.; FEHR, W.R.; SHOEMAKER, R.C. RFLP analysis of soybean seed protein and oil content. **Theoretical and Applied Genetics**, v.83, p.608-612, 1992.

EMBRAPA (2008) <<Disponível em: <http://www.cnpso.embrapa.br/>>> Acesso em 03/01/2008.

FASOULA, V.A.; HARRIS, D.K.; BOERMA, H.R. Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. **Crop Science**, v.44, p.1218-1225, 2004.

GELDERMANN, H. Investigations of quantitative characters in animals by gene markers. **Theoretical and Applied Genetics**, v.43, n.7, p.319-330, 1975.

GUPTA, P.K.; VARSHNEY, The development and use of microsatellite markers for analysis and plant breeding with emphasis on bread wheat. **Euphytica**, v.113, p.163-185, 2000.

GUPTA, P.K.; ROY, J.K.; PRASAD, M. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. **Current Science**, v.80, n.4, p.224-235, 2001.

GQMOL. Viçosa: 15.02.2008. (<http://www.ufv.br/dbg/gqmol/gqmol/htm>).

JANSEN, R.C. Interval mapping of multiple quantitative trait loci. **Genetics**, v.135, p.205-211, 1993.

HASSE, G. **O Brasil da soja: abrindo fronteiras, semeando cidades**. Porto alegre, RS, L&PM Editores. 256p. 1996.

HARADA, K.; XIA, Z. Soybean Genomics: Efforts to reveal the complex genome. **Breeding Science**, v.54, p.215-224, 2004.

HARRIS, H. Enzyme polymorphism in man. **Proc. Natl. Acad. Sci. USA**, v.164, p.298-310, 1966.

HILLIS, D.M.; MORTIZ, C.; MABLE, B.K. **Applications of Molecular Systematics: The state of the field and a look to the future**. In *Molecular Systematics*, 2nd ed., Hillis, D.M., Moritz, C.; Mable, B.K. (eds.), p. 515-543, 1996.

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

HYTEN, D.L.; PANTALONE, V.R.; SAXTON, A. M.; SCHMIDT, M.E.; SAMS, E.C. Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. **JAOCs**, v.81, n.12, p.1115-1118, 2004.

KEIM, P.; DIERS, B.W.; OLSON, T.C.; SHOEMAKER, R.C. RFLP mapping in soybean: association between marker loci and variation in quantitative trait. **Genetics**, v.126, p.735-742, 1990.

KEIM, P.; SCHUPP, J.M.; TRAVIS, S.E.; CLAYTON, K.; ZHU, T.; SHI, L.; FERREIRA, A.; WEBB, D.F. A high-density soybean genetic map based on AFLP markers. **Crop Science**, v.37, p.537-543, 1997.

LANDER, E.S.; BOTSTEIN, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, v.121, n.1, p.185-199, 1989.

KORZUN, V. Molecular markers and their applications in cereals breeding. Marker assisted selection: a fast track to increase gain in plant and animal breeding? 2003. <<Disponível em <http://www.fao.org/biotech/Torino.htm>

KUMAR, L.S. DNA markers in plant improvement: An overview. **Biotechnology Advances**, v.17, p.143-182, 1999.

LARK, K.G.; WEISEMANN, J.M.; MATTHEWS, B.F.; PALMER, R.; CHASE, K.; MACALMA, T. A genetic map of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: Minsoy and Noir 1. **Theoretical and Applied Genetics**, v.86, p.901–906, 1993.

LEE, S.H.; BAILEY, M.A.; MIAN M.A.R.; SHIPE, E.R. ASHLEY, D.A.; PARROT W.A.; HUSSEY, R.S.; BOERMA, H.R. Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. **Theoretical and Applied Genetics**, v.92, p.516-523, 1996.

LITT, M.; LUTY, J.A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American Journal of Human Genetic**, v.44, p.397-401, 1989.

LIU, F.; ZHUANG, B.C.; ZHANG, J.S.; CHEN, S.Y. Construction and analysis of soybean genetic map. **Acta Genetica Sinica**, v.27, p.1018–1026, 2000.

MANSUR, L.M.; ORF, J.H.; CHASE, K.; JARVIK, T.; CREGAN, P.B.; LARK, K.G. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. **Crop Science**, v.36, p.1327-1336, 1996.

MANSUR, L.M.; LARK, K.G.; OLIVEIRA, L. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean. **Theoretical and Applied Genetics**, v.86, p.907-913, 1993.

MAREK, L.P.; SHOEMAKER, R.C. BAC contig development by fingerprintint analysis in soybean. **Genome**, v.40, p.420-427, 1997.

MARKERT, C.L.; MOLLER, F. Multiple forms of enzymes tissue, ontogenetic and species specific patterns. **Proceeding of the Nacional Academy of Sciences**, v.45, p.735-762, 1959.

MATTHEWS, B.S.; MACDONALD, M.H.; GEBHARDT, J.S.; DEVINE, T.E. Molecular marker residing close to the Rhg4 locus conferring resistance to soybean cyst nematode race 3 on linkage group A of soybean. **Theoretical and Applied Genetics**, v.97, p.1047-1052, 1998.

MEKSEM, K.; ZOBRIK, K.; RUBEN, E.; HYTEN, D.; QUANZHOU, T.; ZHANG, H.B.; LIGHTFOOT, D.A. Two Large-insert soybean genomic libraries constructed in a binary vector: applications in chromosome walking and genome wide physical mapping. **Theoretical and Applied Genetics**, v.101, p.747-755, 2000.

MEKSEM, K.; JAMAI, A.; RUBEN, E.; ZOBRIK, K.; TRIWITAKOM, K.; ARELLI, P.; LIGHTFOOT, D.A. Resistance to soybean cyst nematode : an Xa21 like gene family that requires protein dimerization. **Plant, Animal and Microbe Genome**, v.105, p.111, 2002.

MIRANDA, F.D. 2006. Produção, conteúdo de proteína e óleo no grão da soja: herdabilidades, correlações e seleção de genótipos superiores. Viçosa, MG:UFV, 2006. 76p. Dissertação (Doutorado em Genética e melhoramento) – Universidade Federal de Viçosa, 2006.

MUDGE, J.; CREGAN, P.B.; KENWORTHY, J.W.; ORF, J.H.; YOUNG, N.D. Two microsatellite markers that flank the major soybean cyst nematode resistance locus. **Crop Science**, v.37, n.3, p.1611-1615, 1997.

MUEHLBAUER, G.J.; STASWICK, P.E.; SPECHT, J.E.; GRAEF, G.L.; SHOEMAKER, R.C.; KEIM, P. RFLP mapping using near-isogenic lines in the soybean [glycine max (L.) Merr.]. **Theoretical and Applied Genetics**, v.81, p.189-198, 1991.

NICHOLS, D.M.; GLOVER, K.D.; CARLSON, S.R.; SPECHT, J.E.; DIERS, B.W. Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. **Crop Science**, v.46, p.834-839, 2006.

ORF, J.H.; CHASE, K.; JARVIK, T.; MANSUR, L.M.; CREGAN, P.B.; ADLER, F.R.; LARK, K.G. Genetics of soybean agronomic traits: comparison of three related recombinant inbred populations. **Crop Science**, v.39, n.6, p.1642-1651, 1999.

ORF, J.H.; CHASE, K.; ADLER, F.R.; MANSUR, L.M.; LARK, K.G. Genetics of soybean agronomic traits: Interactions between yield quantitative trait loci in soybean. **Crop Science**, v.39, n.6, p.1652-1657, 1999.

PALUDZYSZ FILHO, E.; KIIHL, R.A.S.; ALMEIDA, L.A. **Desenvolvimento de cultivares de soja na região norte e nordeste do Brasil**. In: Simpósio sobre a cultura da soja nos cerrados. Anais. Piracicaba: POTAFOS, p.255-265, 1993.

PANTHEE, D.R.; PANTALONE, V.R.; SAXTON, A.M.; WEST, D.R.; SAMS, C.E. Quantitative trait loci for agronomic traits in soybean. **Plant Breeding**, v.126, p.51-57, 2007.

PRIOLLI, R.H.G.; MENDES JUNIOR, C.T.; SOUSA, S.M.B.; SOUSA, N.E.A.; CONTEL, E.P.B. Characterization of Brazilian soybean cultivars using microsatellite markers. **Genetics and Molecular Biology**, v.25, n.2, p.185-193, 2002.

PRIOLLI, R.H.G.; MENDES JUNIOR, C.T.; SOUSA, S.M.B.; SOUSA, N.E.A.; CONTEL, E.P.B. Diversidade Genética da soja entre períodos e entre programas de melhoramento no Brasil. **Pesquisa Agropecuária Brasileira**, v.39, n.10, p.967-975, 2004.

QIU, B.X.; ARELLI, P.R.; SLEPOR, D.A. RFLP markers associated with soybean cyst nematode resistance and seeds composition in a “Peking” x “Essex” population. **Theoretical and Applied Genetics**, v.98, p.356-364, 1999.

REYNA, N.; SNELLER, C.H. Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. **Crop Science**, v.41, p.1317-1321, 2001.

SALIMATH, S.; BHATTACHARYYA, S. Generation of a soybean BAC library, and identification of a DNA sequences tightly linked to the Rps1-K disease resistance gene. **Theoretical and Applied Genetics**, v.98, p.712-720, 1999.

SCHLOTTERER, C. The evolution of molecular markers, just a matter of fashion? **Genetics**, v.5, p.63-69, 2004.

SEBOLT, A.M.; SHOEMAKER, R.C; DIERS, B.W. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein. **Crop Science**, v.40, p.1438-1444, 2000.

SHOEMAKER R.C.; SPECHT J.E. Integration of the soybean molecular and classical genetic linkage groups. **Crop Science**, v.35, p.436–446, 1995.

SHOEMAKER, R.C.; SCHLUETER, J.A.; CREGAN, P.; VOLKIN, L. The status of soybean genomics and its role in the development of soybean biotechnologies. **AgBioforum**, v.6, p.4-7, 2003.

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, p.329-338, 1996.

SHOEMAKER, R. C.; GUFFY, R.D.; LOREZEN, L.L.; SPEECH, J.E. Molecular genetic mapping of soybean: map utilization. **Crop Science**, v.32, n.5, p.1901-1098, 1992.

SHOEMAKER, R. C.; KEIM, P.; VODKIN, L.; RETZEL, E.; CLIFTON, S.W.; WATERSTON, R.; SMOLLER, D.; CORYELL, V.; KHANNA, A.; ERPELDING, J.; GAI, X.; BRENDLE, V.; SCHIMIDT, C.; SHOOP, E.G.; SCHAMTZ, M.; PAPE, D.; BOWERS, Y.; THEISNIG, B.; MARTIN, J.; DANTE, M.; WYLIE, T.; GRANGER, C. A compilation of a soybean ESTs: generation and analysis. **Genome**, v.45, p.329-338, 2002.

SHULTZ, J.L.; KURUNAM, D.; SHOPINSKI, K.; IQBAL, M.J.; KAZI, S.; ZOBRIST, K.; BASHIR, R.; YAEGASHI, S.; LAVU, N.; AFZAL, A.J.; YESUDAS, C.R.; KASSEM, M.A. The Soybean Genome Database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrates physical and genetic maps of *Glycine max*. **Nucleic Acids Research**, v.34, p.758-765, 2006.

SILVA, M.F. Mapeamento e validação de QTLs de resistência ao nematóide de cisto da soja. Viçosa, MG: UFV, 2006. 110p. Dissertação (Doutorado em Genética e melhoramento) – Universidade Federal de Viçosa, 2006.

SINGH, R.J.; MOWITZ, T. The genomic relationship between *Glycine max* (L.) Merrill and *G. soja*. **Theoretical and Applied Genetics**, v.76, p.705-711, 1998.

SOARES, T.C.B. Mapeamento de locos associados ao conteúdo de proteínas de reserva em soja. Viçosa, MG: UFV, 2004. 113p. Dissertação (Doutorado em Bioquímica Agrícola) – Universidade Federal de Viçosa, 2004.

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

SPECHT, J.E.; CHASE, K.; MACRANDER, M.; GRAEF, B.L.; CHUNG, J.; MARKWELL, J.P.; GERMAN, M.; ORF, J.H.; LARK, K.G. Soybean response to water: a QTL analysis of drought tolerance. **Crop Science**, v.41, p.493-509, 2001.

TAJUDDIN, T.; WATANABE, S.; YAMANAKA, N.; HARADA, K. Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines. **Breeding Science**, v.53, p.133-140, 2003.

TANKSLEY, S.D. Mapping polygenes. **Annual Review Genetics**, v.27, p.205-233, 1993.

TANKSLEY, S.D.; MCCOUCH, S.R. Seed banks and molecular maps: unlocking genetic potential from the wild. **Science**, v.77, p.1063-1066, 1997.

TAUTZ, D. Hypervariability of simple sequence as general source of polymorphic DNA markers. **Nucleic Acids Research**, v.17, p.6463-6471, 1989.

TIAN, A.; WANG, J.; CUI, P.; HAN, Y.; XU, H.; CONG, L.; HUANG, X.; WANG, X.; JIAO, Y.; WANG, B.; ZHANG, J.; CHEN, S. Characterization of soybean genomic features by analysis of its expressed sequence tags. **Theoretical and Applied Genetics**, v.108, p.903-913, 2004.

TOMKINS, J.P.; FRISCH, D.A.; BYRUM, J.R.; WING, R.A. Construction and characterization of a BAC library for the soybean cultivar A3244. **Soybean Genetics**, 2000. <<disponível em: <http://soy.ohio-state.edu/sng/articles/SGN2000-002.htm>.

TOMKINS, J.P.; MAHALINGAM, R.; SMITH, H.; GOICOECHEA, J.L.; KNAP, H.T.; WING, R.A. A bacterial artificial chromosome library for soybean PI 437654 and identification of clones associates with cyst nematode resistance. **Plant Molecular Biology**, v.42, p.25- 32, 1999.

VIEIRA, A.J.D.; OLIVEIRA, D.A.; SOARES, T.C.B.; SCHUSTER, I.; PIOVESAN, N.D.; MARTINEZ, C.A.; BARROS, E.B.; MOREIRA, M.A. Use of the QTL approach to the study of soybean trait relationships in two populations of recombinant inbred lines at the F₇ and F₈ generations. **Brazilian Journal Plant Physiology**, v.18, n.2, p.281-290, 2006.

VOS, P.; HOGERS, R.; BLEEKER, M.; REIJMANS, M.; VAN DE LEE, T.; HORNES, M.;FRJTTTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M.; ZABEAY, M. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Research**, v.23, p.4407-4414, 1995.

XU, D.H., ABE, J.; GAI, J.Y., SHIMAMOTO, Y. Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. **Theoretical and Applied Genetics**, v.105, p.645-653, 2002.

XU, B.; ZHEIN, H.; LU, Q.; ZHAO, S. Three new evidences of original area of soybean. In: Conferencia mundial de investigation en soja, 4, Buenos Aires: AASoja, t1, p.124-128, 1989.

ZENG, Z.B. Precision mapping of quantitative trait loci. **Genetics**, v.136, n.9, p.1457-1468, 1994.

ZHANG, W.K.; WANG, Y.J.; LUO, G.Z.; ZHANG, J.S.; HE, C.Y.; WU, X.L.; GAI, J.Y.; CHEN, S.Y. QTL mapping of ten agronomic traits on the soybean genetic map and their association with EST markers. **Theoretical and Applied Genetics**, v.108, p.1131-1139, 2004.

ZHANG, D.S.; DONG, W.; HUI, D.W.; CHEN, S.Y.; ZHUAN, G.B.C. Construction of a soybean linkage map using an F2 hybrid population from a cultivated variety and a semi-wild soybean. **Chin. Sci. Bull.**, v.42, p.1326–1330, 1997.

WANG, Z.; WEBER, J.L.; ZHONG, G; TANKSLEY, S.D. Survey of plant short tandem repeats. **Theoretical and Applied Genetics**, v.88, p.1–6, 1994.

WANG, D.; GRAEF, G.L.; PROCOPIUK, A.M.; DIERS, B.W. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. **Theoretical and Applied Genetics**, v.108, p.458-467, 2004.

WABER, J.L.; MAY, P.E. Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. **American Journal Human Genetic**, v.44, p.388-396, 1989.

WEEB, D.M.; BALTAZAR, B.M.; RAO-ARELLI, A.P.; SCHUPP, J.; CALYTON, K.; KEIM, P.; BEAVIS, W.D. Genetic mapping of soybean cyst nematode race-3 resistance loci in soybean PI437654. **Theoretical and Applied Genetics**, v.91, p.574-581, 1995.

WEISEMANN, J.M.; MATTHEWS, B.S.; DEVINE, T.E. Molecular markers located to the soybean cyst nematode resistance gene, Rhg4. **Theoretical and Applied Genetics**, v.85, n.1, p.136-138, 1992.

WILCOX, J.R. Sixty years of improvement in publicity developed elite soybean lines. **Crop Science**, v.49, p.1711-1716, 2001.

WU, X.L.; HE, C.Y.; WANG, Y.J.; ZHANG, Z.Y.; DONGFANG, Y.; ZHANG, J.S.; CHEN, S.Y.; GAI, J.Y. Construction and analysis of a genetic linkage map of soybean. **Acta Genet. Sin.**, v.28, p.1051–1061, 2001.

YUAN, J.; NJITI, V.N.; MEKSEM, K.; IQBAL, M.J.; TRIWITAYKORN, K.; KASSEN, M.A.; DAVIS, G.T.; SCHIMIDT, M.E.; LIGHHFOOT, D.A. Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. **Crop Science**, v.42, p.271-277, 2002.