

# Mapeamento de QTL para conteúdos de proteína e óleo em soja

Josiane Isabela da Silva Rodrigues<sup>(1)</sup>, Fábio Demolinari de Miranda<sup>(2)</sup>, Adésio Ferreira<sup>(2)</sup>, Leandro Luiz Borges<sup>(3)</sup>, Marcia Flores da Silva Ferreira<sup>(2)</sup>, Pedro Ivo Vieira Good-God<sup>(3)</sup>, Newton Deniz Piovesan<sup>(1)</sup>, Everaldo Gonçalves de Barros<sup>(3)</sup>, Cosme Damião Cruz<sup>(3)</sup> e Maurilio Alves Moreira<sup>(1)</sup>

<sup>(1)</sup>Universidade Federal de Viçosa (UFV), Departamento de Bioquímica e Biologia Molecular, Instituto de Biotecnologia Aplicada à Agropecuária, CEP 36571-000 Viçosa, MG. E-mail: josianeisabela@gmail.com, piovesan@ufv.br, moreira@ufv.br <sup>(2)</sup>Universidade Federal do Espírito Santo, Departamento de Produção Vegetal, CEP 29500-000 Alegre, ES. E-mail: fademolinari@yahoo.com.br, adesioferreira@gmail.com, mflouressf@gmail.com <sup>(3)</sup>UFV, Departamento de Biologia Geral, Instituto de Biotecnologia Aplicada à Agropecuária, CEP 36570-000 Viçosa, MG. E-mail: leandrodkufv@yahoo.com.br, pivggod@yahoo.com.br, ebarros@ufv.br, cdcruz@ufv.br

Resumo – O objetivo deste trabalho foi detectar e mapear locos de caracteres quantitativos (QTL) que afetam os conteúdos de proteína e óleo em soja (*Glycine max* L. Merr.). Plantas F<sub>2</sub>, derivadas do cruzamento entre a linhagem CS3032PTA276 e a variedade UFVS2012, foram cultivadas em casa de vegetação e forneceram as folhas para extração e análise de DNA. Quarenta e oito marcadores microssatélites (SSR) polimórficos foram avaliados na população F<sub>2</sub>. A avaliação dos fenótipos foi realizada em 207 famílias das progêneses F<sub>2,3</sub>, em um delineamento em blocos ao acaso, com três repetições, conduzido em Viçosa, MG, em 2006. Foram detectados quatro QTL associados ao conteúdo de proteína, nos grupos de ligação D1a, G, A1, e I, e três QTL associados ao conteúdo de óleo, nos grupos A1, I e O. A variação fenotípica explicada pelos QTL variou de 6,24 a 18,94% e 17,26 a 25,93%, respectivamente, para os conteúdos de proteína e óleo. Foram detectados novos QTL associados aos conteúdos de proteína e óleo, além dos previamente relatados em outros estudos. Regiões distintas das atualmente conhecidas podem estar envolvidas no controle genético do teor de proteína e óleo na soja.

Termos para indexação: *Glycine max*, grupos de ligação, loci para características quantitativas, marcadores moleculares.

## Mapping QTL for protein and oil content in soybean

Abstract – The objective of this study was to detect and map quantitative trait loci (QTL) affecting soybean (*Glycine max* L. Merr.) protein and oil contents. F<sub>2</sub> plants, derived from the cross between the CS3032PTA276 line and the variety UFVS2012, were grown in a greenhouse and provided the leaves for DNA extraction and analysis. Forty-eight polymorphic microsatellite markers (SSR) were evaluated in the F<sub>2</sub> population. Evaluation of the phenotype was performed in 207 families from F<sub>2,3</sub> progenies, in a complete block design with three replicates, carried out in Viçosa, MG, Brazil, in 2006. Four QTL associated with protein content, in linkage groups D1a, G, A1, and I, and three QTL for oil content in groups A1, I and O were identified. Phenotypic variation for protein and oil explained by QTL ranged from 6.24 to 18.94% and 17.26 to 25.93%, respectively. New QTL associated with protein and oil contents were detected, besides those previously reported in other studies. Other regions may be involved in the genetic control of protein and oil contents in soybean besides those already known.

Index terms: *Glycine max*, linkage groups, quantitative trait loci, molecular markers.

## Introdução

Fatores genéticos e ambientais afetam a composição química das sementes, como os conteúdos de proteína e óleo no grão, que são compostos determinantes da qualidade da soja. O aumento nesses conteúdos agrega valor ao grão e garante a competitividade da soja brasileira no mercado mundial. Em média, as cultivares de soja apresentam em torno de 40% de proteína e 20%

de óleo, e esses teores correlacionam-se negativamente. O teor de proteína da soja também tem associação negativa com a produtividade de grãos, o que dificulta ganhos adicionais no conteúdo de proteína por meio dos métodos de melhoramento tradicionais. Com o teor de óleo, no entanto, esta associação é positiva e possibilita ganhos simultâneos para os dois caracteres (Bonato et al., 2000).

O uso de marcadores moleculares nos programas de melhoramento, por meio da seleção assistida por marcadores (SAM), pode otimizar o processo de seleção, especialmente quanto aos caracteres de baixa herdabilidade, muito influenciados pelo ambiente. O emprego da SAM pode permitir a obtenção de novas cultivares num período de tempo menor do que o uso de métodos tradicionais de melhoramento. No entanto, o uso eficiente da SAM depende do conhecimento dos genes que determinam a herança dos caracteres de interesse e de suas contribuições para o fenótipo, entre outros fatores.

Nos últimos anos, pelo menos 86 QTL que envolvem o controle genético dos teores de proteína e óleo na soja foram mapeados (Sebolt et al., 2000; Csanádi et al., 2001; Specht et al., 2001; Chapman et al., 2003; Chung et al., 2003; Tajuddin et al., 2003; Fasoula et al., 2004; Hyten et al., 2004; Zhang et al., 2004; Panthee et al., 2005; Nichols et al., 2006; Reinprecht et al., 2006; Chen et al., 2007; Soares et al., 2008). Entretanto, a maioria destes locos não foi confirmada (Soybase, 2010), e há pouca informação sobre a eficiência de seleção com o uso desses QTL.

Nas cultivares de soja plantadas no Brasil, o controle genético dos teores de proteína e óleo é pouco conhecido. Locos quantitativos associados ao conteúdo de proteína no grão foram mapeados em cultivares brasileiras unicamente por Soares et al. (2008), que avaliaram 118 linhagens endogâmicas recombinantes (RIL), obtidas do cruzamento entre as cultivares BARC-8 e Garimpo, em dois locais nas regiões Sul e Sudeste do país, e mapearam cinco QTL nos grupos de ligação (GL) C2, E, G e N ( $R^2 = 7,34-14,37\%$ ). Com base na média dos ambientes, dois QTL, mapeados nos grupos E e L, explicaram 9,9 e 7,11% da variação do fenótipo, respectivamente. Em relação ao conteúdo de óleo, até o momento não foram mapeados QTL em populações de soja brasileiras, apesar de o teor de óleo ser prioridade dos programas de melhoramento. Para a utilização de marcadores moleculares no melhoramento da qualidade da soja, é de fundamental importância a caracterização precisa dos QTL e sua validação em diferentes populações.

O objetivo deste trabalho foi detectar e mapear QTL associados aos conteúdos de proteína e óleo na soja, em uma população  $F_{2,3}$  derivada do cruzamento entre a linhagem CS3032PTA276 e a variedade UFVS2012

de soja, genótipos com alto teor de proteína e óleo, respectivamente.

## Material e Métodos

A população de estudo consistiu de 207 progênes  $F_{2,3}$ , obtidas do cruzamento entre a linhagem CS3032PTA276 e a variedade UFVS2012 de soja. A linhagem CS3032PTA276 tem alto teor de proteína (43,7% de teor protéico e 18,9% de teor de óleo) e é derivada da cultivar BARC-8, desenvolvida pelo Departamento de Agricultura dos EUA. A variedade UFVS2012, desenvolvida pelo Programa de Melhoramento Genético da Soja, da Universidade Federal de Viçosa, MG, tem teor normal de proteína (37,6%) e alto conteúdo de óleo (21,6%).

As progênes  $F_2$  foram plantadas em casa de vegetação, e folhas jovens de cada indivíduo foram coletadas para extração e análise de DNA. Para a avaliação fenotípica dos conteúdos de proteína e óleo, foi realizado um experimento no Campo Experimental Diogo Alves de Mello, em Viçosa, MG (20°45'S, 42°52'W, altitude de 650 m, precipitação anual de 1.340 mm), no ano agrícola de 2006, tendo-se utilizado o delineamento experimental de blocos ao acaso, com três repetições. Foram utilizadas 207 plantas  $F_2$  para a obtenção do mesmo número de famílias  $F_{2,3}$ , sendo que cada planta  $F_2$  autofecundada forneceu cerca de 45 sementes, que foram separadas em três repetições e plantadas no campo em linhas de um metro. Como testemunhas, foram plantadas, em cada bloco, linhas das cultivares BARC-8 e Monarca e dos dois progenitores. Sementes de cinco plantas das famílias  $F_{2,3}$  foram colhidas em cada bloco para obtenção dos conteúdos de proteína e óleo. O conteúdo de proteína e óleo nas sementes foi determinado com um espectrômetro infravermelho FT-NIR, Agrosystem, modelo Instalab 600 product analyzer (Dickey John Corporation, Auburn, Illinois, USA). As sementes foram trituradas em um moinho para proporcionar granulometria adequada para a leitura do espectro no equipamento FT-NIR.

As análises de variância foram realizadas pela avaliação em blocos ao acaso, com testemunhas adicionais, com os progenitores CS3032PTA276 e UFVS2012 e as cultivares BARC-8 e Monarca utilizadas como testemunhas. O modelo estatístico utilizado foi o de ensaio de famílias com testemunhas intercalares, e o esquema da análise de variância está

ilustrado na Tabela 1. A herdabilidade ( $h^2$ ), no sentido amplo, e o coeficiente de variação experimental ( $CV_e$ ) foram obtidos pelos seguintes estimadores:

$$h^2 = (QMF - QMR)/QMF; \text{ e } CV_e = 100(QMR)^{0,5}/\mu;$$

em que QMF e QMR representam o quadrado médio de famílias e do resíduo, respectivamente, e  $\mu$  é a média geral. Os componentes de variância fenotípico, genético e relativo às testemunhas foram estudados, respectivamente, por meio dos estimadores:

$$\hat{\sigma}^2 = QMR; \hat{\sigma}^2_g = (QMF - QMR)/r; \text{ e } \hat{\phi}_g = (QMTe - QMR)/r,$$

em que QMTe é o quadrado médio das testemunhas e r o número de repetições.

Folhas das 207 plantas  $F_2$  foram coletadas, congeladas em nitrogênio líquido e armazenadas a  $-80^\circ\text{C}$  para extração de DNA e análise de marcadores moleculares. A extração de DNA foi baseada no protocolo de Doyle & Doyle (1990), com modificações. As reações de amplificação foram realizadas nas seguintes condições: Tris-HCl (pH 8,3); 50 mmol  $\text{L}^{-1}$  de KCl; 2 mmol  $\text{L}^{-1}$  de  $\text{MgCl}_2$ ; 100  $\mu\text{mol L}^{-1}$  de cada um dos desoxinucleotídeos dATP, dTTP, dGTP e dCTP; 0,6  $\mu\text{mol L}^{-1}$  de cada iniciador microssatélite (Cregan et al., 1999) (Research Genetics, Huntsville, AL, EUA); uma unidade de Taq polimerase e 30 ng de DNA, em um volume total de 15  $\mu\text{L}$ . As reações de PCR foram realizadas em um termociclador Perkin Elmer modelo 9600 (Applied Biosystems, Norwalk, CA, USA), programado para 30 ciclos de 1 min a  $94^\circ\text{C}$ , 1 min a  $50^\circ\text{C}$ , e 2 min a  $72^\circ\text{C}$ , seguidos de uma etapa de 7 min a  $72^\circ\text{C}$ . Os produtos amplificados foram separados por eletroforese em gel de poli(acrilamida) a 10%, tendo-se utilizado tampão TAE 1X (Tris-borato 90 mmol  $\text{L}^{-1}$ , acetato 100 mmol  $\text{L}^{-1}$  e EDTA 2 mmol  $\text{L}^{-1}$ ).

**Tabela 1.** Análise de variância pelo modelo de ensaios de famílias com testemunhas intercalares.

Fonte de variação	gl	Quadrados médios <sup>(1)</sup>	Esperança do quadrado médio	Teste F
Bloco	r-1	QMB	-	-
Tratamento	t-1	QMT	-	-
Famílias	f-1	QMF	$\sigma^2 + \sigma^2_g$	QMF/QMR
Testemunhas	e-1	QMTe	$\sigma^2 + r\phi_g$	QMTe/QMR
Família x testemunha	1	QMF vs Te	-	QMF vs Te/QMR
Resíduo	t(r-1)	QMR	$\sigma^2$	-

<sup>(1)</sup>QMB, quadrado médio de bloco; QMT, quadrado médio de tratamento; QMF, quadrado médio de famílias; QMTe, quadrado médio de testemunhas; QMF vs Te, quadrado médio de família versus testemunhas; e QMR quadrado médio do resíduo.

Após a eletroforese, os géis foram corados com brometo de etídio e fotografados sob luz ultravioleta, pelo sistema de fotodocumentação Eagle Eye II (Stratagene, La Jolla, CA, USA).

A segregação individual de cada marcador foi testada por meio do teste qui-quadrado tendo utilizado a correção de Bonferroni a 5% de probabilidade, seguida pela análise de marca simples. Os grupos de ligação (GL) foram estabelecidos com os critérios LOD mínimo de 3 e percentagem de recombinação máxima de 30%. As análises de detecção e mapeamento de QTL foram realizadas pelos métodos de intervalo simples (Lander & Botstein, 1989) e intervalo composto (Zheng, 1994). Para este último, a seleção de cofatores foi feita pela regressão stepwise, tendo-se adotado os níveis de significância de 5 e 10%, para entrada e saída das variáveis independentes, respectivamente. As estimativas dos valores aditivos e de dominância, o coeficiente de determinação do QTL, correspondente ao pico de maior significância estatística do QTL, e a posição do QTL foram declarados quando os valores da razão da máxima verossimilhança (likelihood ratio, LR) excederam os valores críticos de corte ( $\alpha=0,05$ ), em cada grupo de ligação. Os valores de LR críticos foram determinados pela realização de mil permutações. Todas as análises foram realizadas com o programa Gqmol (Cruz & Schuster, 2004).

## Resultados e Discussão

Os conteúdos de proteína e óleo apresentaram distribuição contínua normal na população, pelo teste de Lilliefors ( $p \leq 0,01$ ), conforme o esperado pelo padrão complexo de herança dos caracteres. As distribuições dos valores fenotípicos e os valores médios das características obtidas para cada genitor podem ser observados na Figura 1. Nas duas características avaliadas, foram observados segregantes transgressivos na população, tanto para o aumento quanto para a redução do valor fenotípico.

Os resultados das análises de variância evidenciaram a variabilidade genética da população quanto aos conteúdos de proteína e óleo (Tabela 2). Baixos coeficientes de variação experimental e valores elevados da razão  $CV_g/CV_e$ , obtidos nas análises das duas características, indicaram, respectivamente, boa precisão experimental e que grande parte da variação total observada é consequência da variância genética.

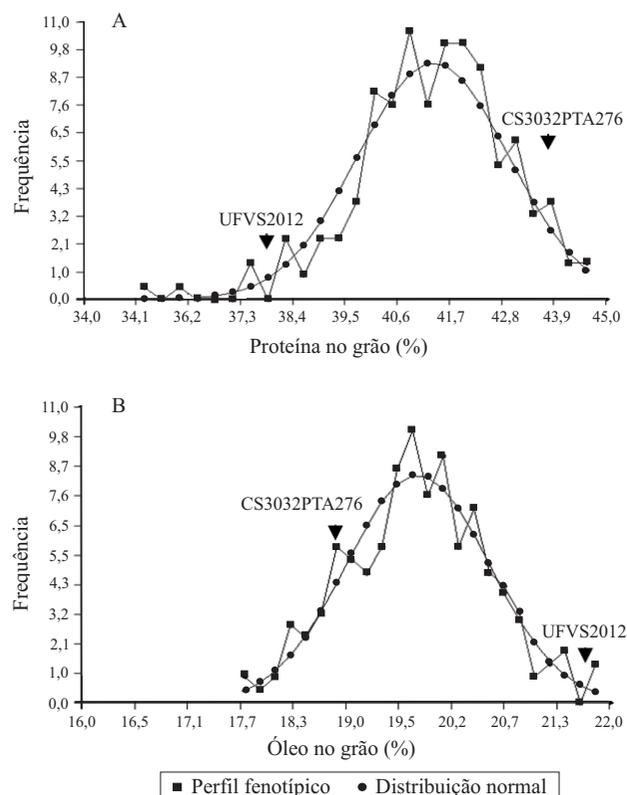
As estimativas de herdabilidade, no sentido amplo, obtidas para os conteúdos de proteína e óleo da população, foram elevadas e apontam o potencial da população para o mapeamento genético, bem como a possibilidade de seleção de genótipos superiores para o melhoramento destes caracteres. Os coeficientes de correlação genotípico, fenotípico e ambiental, verificados entre os conteúdos de proteína e óleo, foram de -0,89, -0,78 e -0,51, respectivamente. A associação observada foi negativa, ou seja, para aumento no conteúdo de proteína há redução no de óleo, resultado similar ao encontrado por Bonato et al. (2000). Além disso, a correlação genotípica foi superior à fenotípica, situação em que pelo menos um dos caracteres tem alto valor de herdabilidade.

Trezentos e cinquenta e sete pares de iniciadores microssatélites foram testados nos genitores da população, e aproximadamente 100 apresentaram fragmentos polimórficos. Destes, 48 marcadores foram amplificados, tendo sido escolhidos por estarem próximos a regiões do genoma da soja em que os QTL de teor de proteína e óleo foram previamente reportados

e pela qualidade da amplificação. Marcadores que não segregaram na proporção esperada de 1:2:1 (Satt429, Satt545) foram desconsiderados da análise.

Nove GL foram obtidos pelo agrupamento de 25 marcadores, e cobriram segmentos genômicos dos grupos A1, B1, D1a, G, I, M e O, de acordo com o mapa consenso da soja (Song et al., 2004) (Figura 2). Para o agrupamento dos marcadores, foram utilizados como critério valores de LOD = 3 e  $r = 0,30$ . Entretanto, para oito dos nove GL, o agrupamento ocorreu também com LOD = 5. Vinte e três marcadores, fisicamente distantes dos demais, não foram agrupados, o que é resultado esperado de acordo com o mapa consenso.

A associação marcador-característica foi determinada pela análise de marca simples, por meio de regressão linear (Tabela 3). Quinze marcadores apresentaram associação significativa com conteúdo de proteína, 9 a 1% de probabilidade, e os demais a 5% de probabilidade. O percentual da variação do teor de proteína, explicada individualmente por esses marcadores, variou entre 3,1 a 23,7%. Quanto ao conteúdo de óleo, 16 marcadores tiveram associação significativa, 9 a 1% e os demais a 5% de probabilidade, o que explica entre 3,1 e 25,4% do valor fenotípico da característica. Dos 16 marcadores com associações significativas, 13 foram comuns para as duas características, o que



**Figura 1.** Conteúdos de proteína (A) e óleo (B) no grão das progênies  $F_{2:3}$ , derivadas do cruzamento entre a linhagem CS3032PTA276 e a variedade UFVS2012.

**Tabela 2.** Análise de variância, médias observadas, coeficientes de variação experimental (CVe), genotípico (CVg) e ambiental (CVa), e parâmetros genéticos das características avaliadas nas progênies  $F_{2:3}$ , derivadas do cruzamento entre UFVS2012 e CS3032PTA276, bem como nas testemunhas.

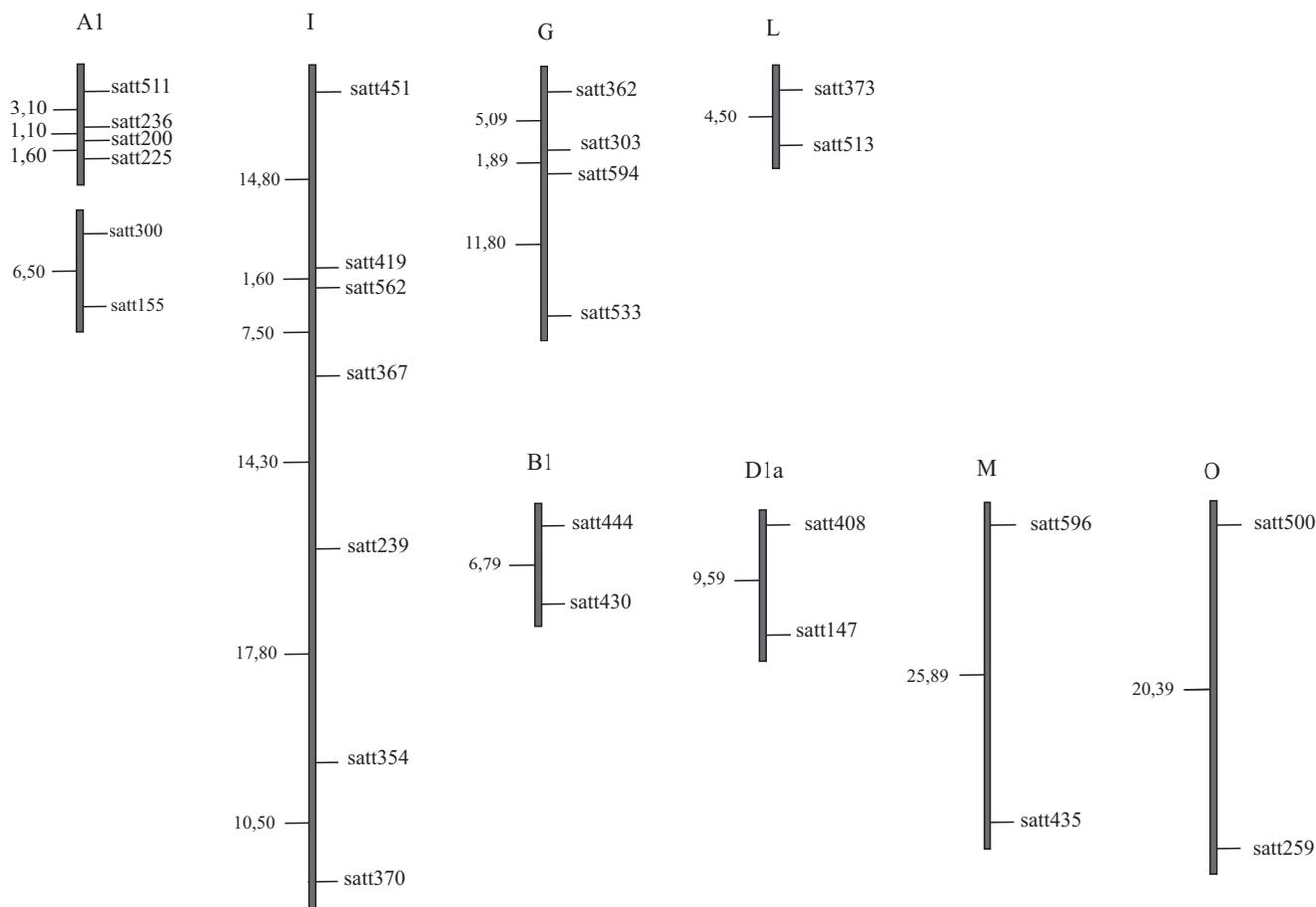
Fonte de variação	gl	Proteína	Óleo
Quadrados médios			
Bloco	2	48,47	2,39
Tratamento	210	8,50**	2,12**
Famílias $F_3$	206	7,51**	1,99**
Testemunhas	3	74,08**	11,35**
Famílias x testemunhas	1	15,01*	1,40 <sup>ns</sup>
Resíduo	420	2,82	0,56
Médias observadas (%)			
Média geral	-	41,38	19,81
Médias das famílias	-	41,36	19,81
Médias das testemunhas	-	42,49	19,47
Parâmetros genéticos			
$\sigma^2_g$	-	1,56	0,47**
CVg	-	3,01	3,47
Herdabilidade (%)	-	62,36	71,61
CV (%)	-	4,06	3,80
CVg/CVe	-	0,74	0,91

\* e \*\* Significativo a 5 e 1% de probabilidade, respectivamente.  $\sigma^2_g$ , variância genotípica.

evidencia a possibilidade de que regiões genômicas comuns atuem simultaneamente no controle das duas características. As associações altamente significativas ( $p < 0,01$ ) envolveram marcadores dos grupos A1, D1a, G e I. Entre essas, oito evidenciaram marcadores em associação simultânea com os dois caracteres. Os maiores valores de  $R^2$ , tanto para proteína quanto para óleo, foram verificados para o mesmo marcador, o Satt239 do GL I. Neste mesmo GL, os marcadores Satt367, Satt562, Satt419, Satt354 e Satt451 também mostraram alta significância e  $R^2$  relativamente elevado para os teores de proteína e óleo. Esse resultado indica, conforme relatado na literatura (Sebolt et al., 2000; Csanádi et al., 2001; Chung et al., 2003; Reinprecht et al., 2006), que essa

região apresenta QTL de grande efeito para os teores de proteína e óleo.

Dois marcadores altamente significativos (Satt436 e Satt370), tanto com relação ao teor de proteína quanto ao de óleo, explicaram cada um aproximadamente 10% da variação fenotípica, embora esses marcadores não tenham sido agrupados. Resultado semelhante foi encontrado para o marcador Satt191, que foi significativo para os dois caracteres ( $p < 0,05$ ) e explicou em torno de 4% da variação dos conteúdos de proteína e óleo na população. Pelo método de intervalo composto, foram detectados quatro QTL associados ao conteúdo de proteína, nos grupos D1a, G, A1, e I, e três associados ao conteúdo de óleo, localizados nos grupos A1, I e O (Tabela 4). A variação fenotípica dos teores de proteína e óleo, explicada por meio



**Figura 2.** Grupos de ligação definidos em uma população de 207 progênies  $F_{2,3}$ , obtidas do cruzamento entre a linhagem CS3032PTA276 (alto teor de proteína) e a variedade UFVS2012 (alto teor de óleo). Foram utilizados como critérios: 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.

de cada QTL, variou entre 6,2–18,9 e 17,3–25,9%, respectivamente. Os QTL detectados para teor de proteína (grupo A1, D1a, G e I) e um dos QTL para óleo (grupo A1) tiveram efeitos aditivos positivos. Os outros dois QTL para óleo tiveram estimativas negativas. Também foram observadas estimativas de dominância, tanto positivas quanto negativas, para os QTL detectados.

Um QTL, que explicou 6,2% da variação fenotípica do conteúdo de proteína, foi identificado no grupo D1a, no intervalo entre o Satt408 e Satt147, marcadores localizados nas posições 106,69 e 108,89 cM respectivamente, no mapa consenso da soja (Song et al., 2004). Esses marcadores também apresentaram associação significativa ( $p < 0,05$ ) com o teor de óleo, pela análise de marca simples, e explicaram 5,3 e 4,4% da variação de óleo, respectivamente.

**Tabela 3.** Análise de marca simples dos conteúdos de proteína e óleo, nas 207 progênies  $F_{2:3}$ , obtidas do cruzamento entre a linhagem CS3032PTA276 e a variedade UFVS2012.

Marca	Grupo de ligação	Posição <sup>(1)</sup> (cM)	Regressão	
			R <sup>2</sup> (%)	Prob. (F)
Teor de proteína				
Satt200	A1	92,89	3,07	0,046*
Satt300	A1	30,93	4,58	0,012*
Satt236	A1	93,23	3,16	0,048*
Satt408	D1a	106,69	5,57	0,016*
Satt436	D1a	70,69	10,51	0,000**
Satt370	D1a	60,99	9,16	0,001**
Satt147	D1a	108,89	3,95	0,025*
Satt191	G	96,57	3,26	0,038*
Satt533	G	56,52	9,57	0,007**
Satt367	I	27,98	15,78	0,000**
Satt239	I	36,94	23,69	0,000**
Satt562	I	22,84	14,52	0,000**
Satt419	I	21,90	16,59	0,000**
Satt451	I	20,34	9,84	0,001**
Satt354	I	46,22	13,78	0,000**
Teor de óleo				
Satt200	A1	92,89	3,10	0,044*
Satt225	A1	95,16	6,05	0,018*
Satt511	A1	94,20	3,84	0,026*
Satt236	A1	93,23	5,17	0,007**
Satt408	D1a	106,69	5,27	0,013*
Satt436	D1a	70,69	7,75	0,000**
Satt370	D1a	60,99	12,14	0,000**
Satt147	D1a	108,89	4,44	0,015*
Satt191	G	96,57	4,37	0,012*
Satt367	I	27,98	17,08	0,000**
Satt239	I	36,94	25,35	0,000**
Satt562	I	22,84	20,61	0,000**
Satt419	I	21,90	19,32	0,000**
Satt451	I	20,34	8,58	0,002**
Satt354	I	46,22	19,36	0,000**
Satt259	O	39,82	3,49	0,035*

<sup>(1)</sup>Posição de acordo com o mapa consenso de Song et al. (2004); \* e \*\* Significativo a 5 e 1% de probabilidade.

Considerando-se as posições dos marcadores no mapa consenso, essa região ainda não havia sido descrita em estudos anteriores. Outra região genômica envolvida no controle dos dois caracteres, nesse mesmo GL, foi evidenciada pela detecção de associações significativas ( $p \leq 0,01$ ) para as duas características, com o uso dos marcadores Satt436 (70,69 cM) e Satt370 (60,99 cM). Locus quantitativos associados às duas características, próximos à região desses marcadores, já foram relatados (Brummer et al., 1997; Csanádi et al., 2001; Specht et al., 2001). Outros QTL foram relatados no grupo D1a, para conteúdo de óleo e proteína, por Brummer et al. (1997) e por Hyten et al. (2004), respectivamente. Um novo QTL para conteúdo de proteína foi detectado no grupo G, na região do marcador Satt533 (56,52 cM), e explicou 13,9% da variação fenotípica da característica na população. Próximo a esta região genômica, Brummer et al. (1997) relataram associação do conteúdo de proteína com os RFLPs A816 e A890-1, e Lee et al. (1996) mapearam um QTL que explicou 23% da variação de proteína no cruzamento entre PI97-100 e Coker 237. Apesar de não ter sido agrupado, o microsátelite Satt191 (96,57 cM, grupo G) também apresentou associação com o teor de proteína ( $R^2 = 3,3\%$ ,  $p \leq 0,05$ ), e foi localizado na mesma região de um QTL para proteína mapeado por Brummer et al. (1997). Ainda sobre este GL, Panthee et al. (2005) reportaram que o Satt570 explicou 20,2% da variação do conteúdo de proteína, em uma população de 101 RIL derivada do cruzamento N87-984-16 x TN93-99.

No grupo A1, entre as marcas Satt300 e Satt155, foi identificado outro QTL para conteúdo de proteína,

**Tabela 4.** Locus de caracteres quantitativos (QTL) detectados para os conteúdos de proteína e óleo nas progênies  $F_{2:3}$ , derivadas de CS3032PTA276 e UFVS2012, pelos métodos de mapeamento por intervalo simples e composto.

Grupo de ligação <sup>(1)</sup>	Intervalo/região	LR	R <sup>2</sup> (%) <sup>(2)</sup>	a <sup>(3)</sup>	d <sup>(4)</sup>
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
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

<sup>(1)</sup>Conforme Song et al. (2004). <sup>(2)</sup>Proporção da variação fenotípica explicada pelo QTL. <sup>(3)</sup>Efeito aditivo estimado para o QTL. <sup>(4)</sup>Efeito da dominância estimado para o QTL. \* e \*\*, valores de razão de verossimilhança (LR) excederam os valores críticos de corte, a 5 e 1% de probabilidade, respectivamente. Os valores de LR críticos foram determinados pela realização de mil permutações.

que explicou 12,7% da variação fenotípica. Mansur et al. (1996), nessa mesma região, mapearam um QTL para proteína, em uma população de RIL derivada de Minsoy e Noir 1. Soares et al. (2008) também identificaram o Satt471, que explicou 17,43% da variação do conteúdo de proteína. Outros marcadores deste GL foram relatados por Mansur et al. (1996), Brummer et al. (1997) e Specht et al. (2001).

Neste mesmo grupo de ligação (A1), no intervalo entre as marcas Satt200 (92,89 cM) e Satt225 (95,16 cM), foi detectado um QTL para conteúdo de óleo que explicou 17,3% da variação da característica. O marcador Satt236 (93,23 cM) também mostrou associação com o conteúdo de óleo ( $R^2 = 5,2\%$ ,  $p < 0,01$ ). Locos quantitativos quanto ao teor de óleo, nessa região, foram relatados por Mansur et al. (1996), Brummer et al. (1997) e Specht et al. (2001), em diferentes localidades e backgrounds genéticos, o que indica que a região consiste em um ou mais QTL conservados no controle do conteúdo de óleo na soja, além de estar envolvida no controle do teor de proteína, mas com efeito menor. Os resultados obtidos neste GL validam, portanto, os QTL maiores e os menores reportados para as duas características.

O loco que explicou a maior proporção da variação do teor de proteína e de óleo foi detectado na mesma região do grupo I, próximo ao marcador Satt239 (36,94 cM), e explicou 18,9 e 25,3% da variação para proteína e óleo, respectivamente. O grupo I tem sido associado ao controle dessas características por diversos autores. Chung et al. (2003) mapearam um QTL que explicou 45,3 e 27,7% da variação do teor de proteína, nos anos de 1996 e 1999, respectivamente. O mesmo intervalo explicou 27,8 e 24,2% da variação do teor de óleo em cada ano. Na região do Satt239, Tajuddin et al. (2003) utilizaram 157 RIL e mapearam QTL que explicaram, em dois anos, 20,9 e 29,7% da variação do teor de proteína, e 8,6 e 6,8% da variação do teor de óleo. Nichols et al. (2006) estudaram dois grupos de populações de retrocruzamento e verificaram que o QTL principal que controla variações no teor de proteína e óleo no grupo I encontra-se na região de 3 cM, entre o microssatélite Satt239 e o AFLP ACG9b. Em estudo na mesma região, Soares et al. (2008) utilizaram 95 RIL obtidas do cruzamento entre BARC-8 e Garimpo e detectaram um QTL para proteína total no intervalo entre Satt239 e Satt105 (49,34 cM), que foi consistente em dois ambientes,

e explicou 14,74 e 11,34% da variação do conteúdo de proteína.

Outro loco associado ao teor de óleo foi detectado no grupo O, entre os marcadores Satt500 (14,17 cM) e Satt259 (39,82 cM), e explicou 18,51% da variação da característica. Nesse grupo, QTL para óleo foram mapeados por Panthee et al. (2005). Locos associados a teores de proteína e óleo nos grupos A1, D1a e G foram mapeados por Mansur et al. (1996), Brummer et al. (1997), Orf et al. (1999), Csanádi et al. (2001) e Specht et al. (2001). Na região do marcador Satt239 do grupo I, locos para proteína e óleo foram relatados por diversos autores, em diferentes ambientes e backgrounds genéticos (Chung et al., 2003; Tajuddin et al., 2003; Nichols et al., 2006; Soares et al., 2008), o que indica a presença de genes conservados na região. Além disso, os mesmos intervalos neste GL, que afetam tanto o conteúdo de proteína como o de óleo, revelam efeitos pleiotrópicos de um ou mais genes presentes na região. A análise criteriosa dos marcadores na região próxima ao Satt239 tem grande relevância no mapeamento genético do conteúdo de proteína e óleo, no sentido de se determinar com precisão a localização dos QTL. A confirmação desses intervalos pode contribuir para a clonagem posicional e a elucidação da estrutura e função dos genes, bem como seus papéis na determinação dos conteúdos de proteína e óleo. A consistência dos efeitos dos QTL mapeados sobre o controle dos teores de proteína e óleo na soja deve ser avaliada em trabalhos futuros, no sentido de se implementar o uso da SAM no melhoramento genético dessas características e praticar, simultaneamente, a seleção para teor de proteína e óleo, uma estratégia que permitirá obter ganhos em longo prazo para as duas características.

## Conclusões

1. Este estudo confirma QTL de efeito maior e menor para proteína e óleo, além de detectar novos QTL para essas características.

2. A região próxima ao Satt239, no grupo de ligação I, determina grandes variações tanto no conteúdo de proteína quanto no de óleo, consistentes em diferentes ambientes e populações.

3. Um novo QTL de pequeno efeito para teor de proteína no grupo de ligação D1a (Satt408–Satt147)

também apresenta associação com o teor de óleo nessa população.

4. Um novo QTL foi mapeado no grupo O (Satt500–Satt259) da soja, para o teor de óleo.

5. Locos quantitativos que controlam o teor de proteína nos grupos de ligação A1 (Satt300–Satt155) e G (Satt533), mapeados neste estudo, estão em regiões próximas de QTL descritos anteriormente, e a associação dos marcadores Satt436 e Satt370, no grupo D1a, com conteúdo de proteína e óleo valida um QTL de pequeno efeito nesta região.

### Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais, pelo apoio financeiro.

### Referências

- BONATO, E.R.; BERTAGNOLLI, P.F.; LANGE, C.E.; RUBIN, S. de A.L. Teor de óleo e de proteína em genótipos de soja desenvolvidos após 1990. **Pesquisa Agropecuária Brasileira**, v.35, p.2391-2398, 2000.
- 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.
- CHAPMAN, A.; PANTALONE, V.R.; USTUN, A.; ALLEN, F.L.; LANDAU-ELLIS, D.; TRIGIANO, R.N.; GRESSHOFF, P.M. Quantitative trait loci for agronomic and seed quality traits in an F<sub>2</sub> and F<sub>4,6</sub> soybean population. **Euphytica**, v.129, p.387-393, 2003.
- CHEN, Q.-S.; ZHANG, Z.-C.; LIU, C.-Y.; XIN, D.-W.; QIU, H.-M.; SHAN, D.-P.; SHAN, C.-Y.; HU, G.-H. QTL analysis of major agronomic traits in soybean. **Agricultural Sciences in China**, v.6, p.399-405, 2007.
- CHUNG, J.; BABKA, H.L.; GRAEF, G.L.; STASWICK, P.E.; LEE, D.J.; CREGAN, P.B.; SHOEMAKER, R.C.; SPECHT, J.E. The seed protein, oil, and yield QTL on soybean linkage group I. **Crop Science**, v.43, p.1053-1067, 2003.
- CREGAN, P.B.; JARVIK, T.; BUSH, A.L.; SHOEMAKER, R.C.; LARK, K.G.; KAHLER, A.L.; KAYA, N.; VANTOAI, T.T.; LOHNES, D.G.; CHUNG, J.; SPECHT, J.E. An integrated genetic linkage map of the soybean. **Crop Science**, v.39, p.1464-1490, 1999.
- CRUZ, C.D.; SCHUSTER, I. **GQMOL**: aplicativo computacional para análise de dados moleculares e de suas associações com caracteres quantitativos. Versão 2.1. Viçosa: UFV, 2004. Disponível em: <www.ufv.br/dbg/gqmol/gqmol.htm>. Acesso em: 02 nov. 2009.
- 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.
- DOYLE, J.J.; DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**, v.12, p.13-15, 1990.
- 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.
- HYTEN, D.L.; PANTALONE, V.R.; SAMS, C.E.; SAXTON, A.M.; LANDAU-ELLIS, D.; STEFANIAK, T.R.; SCHMIDT, M.E. Seed quality QTL in a prominent soybean population. **Theoretical and Applied Genetics**, v.109, p.552-561, 2004.
- LANDER, E.S.; BOTSTEIN, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, v.121, p.185-199, 1989.
- LEE, S.H.; BAILEY, M.A.; MIAN, M.A.R.; CARTER JUNIOR, T.E.; SHIPE, E.R.; ASHLEY, D.A.; PARROTT, W.A.; HUSSEY, R.S.; BOERMA, H.R. RFLP loci associated with soybean seed protein and oil content across populations and locations. **Theoretical and Applied Genetics**, v.93, p.649-657, 1996.
- 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.
- 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: I. Comparison of three related recombinant inbred populations. **Crop Science**, v.39, p.1642-1651, 1999.
- PANTHEE, D.R.; PANTALONE, V.R.; WEST, D.R.; SAXTON, A.M.; SAMS, C.E. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. **Crop Science**, v.45, p.2015-2022, 2005.
- REINPRECHT, Y.; POYSA, V.W.; YU, K.; RAJCAN, I.; ABLETT, G.R.; PAULS, K.P. Seed and agronomic QTL in low linolenic acid, lipoxygenase - free soybean (*Glycine max* (L.) Merrill) germplasm. **Genome**, v.49, p.1510-1527, 2006.
- SEBOLT, A.M.; SHOEMAKER, R.C.; DIERS, B.W. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. **Crop Science**, v.40, p.1438-1444, 2000.
- SOARES, T.C.B.; GOOD-GOD, P.I.V.; MIRANDA, F.D. de; SOARES, Y.J.B.; SCHUSTER, I.; PIOVESAN, N.D.; BARROS, E.G. de; MOREIRA, M.A. QTL mapping for protein content in soybean cultivated in two tropical environments. **Pesquisa Agropecuária Brasileira**, v.43, p.1533-1541, 2008.
- SONG, Q.J.; MAREK, L.F.; SHOEMAKER, R.C.; LARK, K.G.; CONCIBIDO, V.C.; DELANNAY, X.; SPECHT, J.E.; CREGAN, P.B. A new integrated genetic linkage map of soybean. **Theoretical and Applied Genetics**, v.109, p.122-128, 2004.

SOYBASE. **Soybase and the soybean breeder's toolbox**. Disponível em: <soybeanbreederstoolbox.org>. Acesso em: 05 mar. 2010.

SPECHT, J.E.; CHASE, K.; MACRANDER, M.; GRAEF, B.L.; CHUNG, J.; MARKWELL, J.P.; GERMANN, 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.

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 (*Glycine max* L. Merr.) genetic map and their association with EST markers. **Theoretical and Applied Genetics**, v.108, p.1131-1139, 2004.

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

---

Recebido em 23 de março de 2010 e aprovado em 30 de abril de 2010