

# Divergência em QTLs e variância genética para teores de proteína e óleo em soja

Josiane Isabela da Silva Rodrigues<sup>(1)</sup>, Klever Márcio Antunes Arruda<sup>(1)</sup>, Cosme Damião Cruz<sup>(1)</sup>, Newton Deniz Piovesan<sup>(1)</sup>, Everaldo Gonçalves de Barros<sup>(2)</sup> e Maurílio Alves Moreira<sup>(3)</sup>

<sup>(1)</sup>Universidade Federal de Viçosa, Instituto de Biotecnologia Aplicado à Agropecuária, Avenida Peter Henry Rolfs, Campus Universitário, CEP 36570-000 Viçosa, MG, Brasil. E-mail: josianeisabela@gmail.com, kleverantunes@yahoo.com.br, cdcruz@ufv.br, piovesan@ufv.br

<sup>(2)</sup>Universidade Católica de Brasília, SGAN 916, Módulo B, Bloco C, Sala 213, CEP 70790-160 Brasília, DF, Brasil. E-mail: everaldodebarros@gmail.com <sup>(3)</sup>In memoriam.

**Resumo** – O objetivo deste trabalho foi avaliar a relação entre os parâmetros de divergência em regiões de QTLs e a variância genética em genótipos de soja, quanto aos teores de proteína e óleo nos grãos. Dois grupos de genótipos foram avaliados, em diferentes ambientes, quanto aos teores de proteína e óleo e genotipados com marcadores moleculares de regiões de QTLs. A partir de cada grupo, estabeleceram-se subgrupos por critérios pré-definidos e avaliou-se a relação entre os parâmetros, tendo-se comparado a divergência média e a variância genética entre os subgrupos. Os subgrupos foram definidos com base nos critérios de diferença em divergência média, homogeneidade e heterogeneidade nos subgrupos e proximidade em uma projeção tridimensional da matriz de distância. As percentagens de concordância entre maiores valores de divergência média e de variância genética para o total de subgrupos de cada grupo inicial foram de 72,5 e de 73,4%, respectivamente. Portanto, nestes genótipos, há relação positiva entre as estimativas de divergência em regiões de QTL e variância genética para os teores de proteína e de óleo dos grãos. As distâncias genéticas com base nos marcadores moleculares de regiões de QTLs são eficientes para a predição da variabilidade genética em genótipos de soja para os teores de proteína e de óleo dos grãos.

**Termos para indexação:** *Glycine max*, distância genética, marcadores moleculares, variabilidade genética.

## Divergence on QTLs and genetic variance for protein and oil contents in soybean

**Abstract** – The objective of this work was to evaluate the relationship between the parameters of divergence on QTL regions and the genetic variance in soybean genotypes for protein and oil contents. Two groups of genotypes were evaluated for protein and oil contents in different environments, and genotyped with molecular markers of QTL regions. Subgroups were established by predefined criteria from each group, and the relationship between the parameters was evaluated by comparing mean divergence and genetic variance between the subgroups. The subgroups were defined based on the criteria of difference of mean divergence, homogeneity and heterogeneity in the subgroups, and proximity in a three-dimensional projection of the distance matrix. Percentage agreement between the greater values for mean divergence and genetic divergence for the subgroup total of each initial group was 72.5 and 73.4%, respectively. Therefore, there is a positive relationship between the estimates on QTL regions and genetic variance for protein and oil contents in grains in these genotypes. The genetic distances based on molecular markers of QTL regions are efficient to predict the genetic variability in soybean genotypes, for protein and oil contents.

**Index terms:** *Glycine max*, genetic distance, molecular markers, genetic variability.

### Introdução

Medidas de divergência genética entre genitores têm sido utilizadas para prever médias e variâncias em populações de plantas (Bonato et al., 2006). A hipótese de que a divergência reflete a variabilidade nos cruzamentos fundamenta-se no fato de que o número de genes segregantes aumenta pela complementaridade

gênica e de que o encontro gamético, em processos sexuais, possibilita o aparecimento de novos genótipos (Cruz et al., 2011).

Porém, em estudos com plantas, poucas vezes foi possível prever a variância genética a partir de medidas de distância genética com base em marcadores moleculares (Dias et al., 2004). A falta de correlação entre as estimativas de distância entre

os genitores e a variabilidade genética é atribuída, entre vários fatores, à baixa cobertura do genoma em razão do reduzido número de marcadores moleculares e à ausência de ligação destes aos locos das características quantitativas (quantitative trait loci, QTLs) envolvidos com as características agrônômicas (Rodrigues et al., 2013). Entre outras razões, a falta de correlação dos parâmetros pode estar associada à baixa repetibilidade da variância genética, associada a um elevado erro-padrão, ao erro ou viés da estimativa da divergência genética ou à baixa correlação entre os valores reais (Gumber et al., 1999).

O objetivo deste trabalho foi avaliar a relação entre os parâmetros de divergência em QTLs e a variância genética em genótipos de soja, quanto aos teores de proteína e óleo nos grãos.

### Material e Métodos

Dois grupos de genótipos, com linhagens avançadas e cultivares comerciais de soja, foram avaliados. Os genótipos são oriundos do banco de germoplasma do Programa de Melhoramento da Qualidade da Soja, da Universidade Federal de Viçosa, Viçosa, MG. O primeiro grupo (Grupo 1) consistiu de 29 genótipos com alto teor proteico. O segundo grupo (Grupo 2) consistiu de 22 genótipos com alto teor de óleo (Tabela 1).

Os dois grupos de genótipos de soja foram cultivados no Estado de Minas Gerais, nos seguintes municípios: Viçosa (20°45'S 42°52'W, a 663 m de altitude), em dezembro de 2009; Visconde do Rio Branco (21°00'S 42°50'W, a 358 m de altitude), em outubro de 2010; e São Gotardo (19°18'S 46°02'W, a 1.152 m de altitude), em fevereiro e outubro de 2010. Utilizou-se para esses cultivos o delineamento de blocos ao acaso, com três repetições. O clima predominante nestes locais, segundo a classificação de Köppen-Geiger, é do tipo Cwa ou mesotérmico úmido, caracterizado por verões quentes e úmidos e invernos frios e secos, com precipitação média anual de 1.200 mm, e médias de temperaturas máxima e mínima de 26,1 e 14,0°C respectivamente. Em cada ensaio, semearam-se 15 sementes por fileira de 1 m, e o espaçamento entre as fileiras foi de 0,5 m. Em cada experimento, a colheita foi realizada manualmente e, após o beneficiamento, os grãos foram triturados em um moinho industrial MA020 (Marconi Equipamentos

para Laboratórios Ltda., Piracicaba, SP). O farelo de soja foi analisado quanto aos teores de proteína e óleo, por meio de espectrometria do infravermelho, com um espectrômetro FT-NIR Analyser Antaris II (Thermo Fisher Scientific Brasil Instrumentos de Processo Ltda., São Paulo, SP), de acordo com a metodologia proposta por Choung et al. (2001).

Realizou-se a análise de variância conjunta com base no arranjo fatorial, em que se estudou o efeito dos fatores genótipo e ambiente, por meio do modelo estatístico proposto por Cruz (2006) como  $Y_{ijk} = \mu + G_i + A_j + GA_{ij} + B/A_{jk} + \varepsilon_{ijk}$ , em que:  $Y_{ijk}$  é a observação do k-ésimo bloco, avaliado no i-ésimo genótipo e j-ésimo ambiente;  $\mu$  é a média geral do ensaio;  $B/A_{jk}$  é o efeito do bloco k, dentro do ambiente j;  $G_i$  é o efeito do tratamento ou genótipo i;  $A_j$  é o efeito do ambiente j;  $GA_{ij}$  é o efeito da interação entre o genótipo i e o ambiente j; e  $\varepsilon_{ijk}$  é o erro aleatório associado à observação ijk. O efeito de genótipos ( $G_i$ ) foi considerado fixo, e o de ambientes ( $A_j$ ), aleatório.

Os mesmos genótipos avaliados em campo foram mantidos em vasos, em casa de vegetação, para a coleta de quatro folhas jovens por genótipo, para a extração de DNA. As folhas foram armazenadas a -80°C. A purificação das amostras de DNA foi realizada com o kit Wizard SV Genomic DNA Purification System (Promega, Madison, WI, USA) (Promega, 2014). A concentração de DNA foi estimada em gel de agarose, tendo-se utilizado quantidades conhecidas do DNA do bacteriófago lambda, conforme metodologia em Oliveira et al. (2010). Cada amostra de DNA foi diluída em água para a concentração de 10 ng  $\mu\text{L}^{-1}$ . Os procedimentos de amplificação por reação em cadeia da polimerase (PCR) e eletroforese vertical foram realizados conforme Santana et al. (2014). No PCR, utilizou-se o volume final de 15  $\mu\text{L}$  e as seguintes condições: Tris-HCl, 10 mmol  $\text{L}^{-1}$ ; pH, 8,3; KCl, 50 mmol  $\text{L}^{-1}$ ;  $\text{MgCl}_2$ , 2 mmol  $\text{L}^{-1}$ ; Triton X100, 0,1%; 100  $\mu\text{mol L}^{-1}$  de cada um dos desoxinucleotídeos; 0,3  $\mu\text{mol L}^{-1}$  de cada iniciador; uma unidade de Taq polimerase (Phoneutria Biotecnologia e Serviços Ltda., Belo Horizonte, MG); e 30 ng de DNA. As reações de PCR tiveram passo inicial de 94°C por 4 min, 30 ciclos de 94°C por 1 min, 55°C por 1 min, 72°C por 2 min, e uma etapa final de 72°C por 7 min. Os produtos de amplificação foram separados por eletroforese em géis verticais de poliacrilamida a 10%, com tampão TAE 1X (Tris-acetato a 40 mmol  $\text{L}^{-1}$  e EDTA a 1 mmol  $\text{L}^{-1}$ ) e

período de corrida de três horas a 140 volts. Os géis foram corados com nitrato de prata a 2%, conforme Creste et al. (2001), e registrados com o auxílio de um equipamento fotodocumentador L-PIX EX (Loccus Biotechnology, Cotia, SP).

Trinta e nove pares de iniciadores de marcadores microsatélites, localizados em QTLs relacionados aos teores de proteína e óleo em soja, foram amplificados a partir das amostras de DNA dos 29 genótipos de soja com alto teor proteico (grupo 1). Os seguintes marcadores moleculares – dos grupos de ligação A1, B2, C1, C2, D1a, D2, E, F, I, K, L e M – foram

analisados neste grupo de genótipos: Satt166; Satt127; Satt178; Satt567; Satt077; Satt567; Satt196; Satt239; Satt310; Satt301; Satt384; Satt281; Satt156; Satt496; Satt239; Satt251; Satt373; Satt460; Satt335; Satt539; Satt486; Satt570; Satt389; Satt571; Satt419; Satt270; Satt263; Satt151; Satt117; Satt445; Satt257; Satt173; Satt581; Satt476; Satt471; Satt286; Satt518; e Satt164.

Outros 33 pares de iniciadores de marcadores microsatélites, em QTLs relacionados aos teores de proteína e óleo, foram amplificados a partir das amostras de DNA dos 22 genótipos de soja com alto teor de óleo (grupo 2). Os seguintes marcadores

**Tabela 1.** Análise de variância conjunta e percentuais médios de proteína e óleo de dois grupos de genótipos de soja, em base seca, em quatro ambientes.

Grupo 1	Proteína (%)	Óleo (%)	Grupo 2	Proteína (%)	Óleo (%)
1, BRSMG Garantia	38,82	19,74	1, BRSMG Garantia	38,82	19,74
2, FMT Tucunaré	37,56	20,82	2, FMT Tucunaré	37,56	20,82
3- BRSGO Luziânia	39,16	19,42	3, BRSGO Luziânia	39,16	19,42
4,Capinópolis (UFV-16)	38,47	19,85	4, MG/BR46 (Conquista)	38,54	19,47
5, CS3032PTA182	41,84	17,91	5, Suprema	34,25	23,01
6,BRS Sambaíba	37,55	19,60	6, BARC-8	45,18	18,27
7, CS94731	40,71	18,27	7, CS303TNKCA	38,36	20,86
8, M-SOY8914	37,26	20,57	8, BRSMG 68	37,51	19,79
9, UFV 18 (Patos de Minas)	37,72	20,43	9, PI181544	41,72	18,26
10, M-SOY8400	39,26	18,81	10, PI371611	39,13	20,66
11, DM – 339	39,26	18,76	11, PI371610	38,09	21,50
12, UFVS 2012	37,15	20,74	12, CD 225RR	39,46	20,22
13, UFVS 2011	37,90	19,61	13, CD 224RR	36,91	20,71
14, M-SOY8001	37,19	19,83	14, CD 219RR	36,61	21,61
15, UFVS 2001	39,71	19,12	15, PI235347	39,54	19,48
16, MG/BR46 (Conquista)	38,54	19,47	16, CD 226RR	39,62	20,13
17, Suprema	34,25	23,01	17, UFV 20 (Floresta)	37,53	19,67
18, M-SOY6101	35,96	21,90	18, CD 222	36,04	20,63
19, PI417360	41,01	17,19	19, CD01RR8376	37,13	21,42
20, CS3032PTA190-5-1	43,63	17,30	20, CD983321RR	39,15	19,84
21, CS3032PTA137-4-10	43,55	16,48	21, BR8014887	44,71	17,28
22, B3PTA382-2-10	43,80	17,67	22, CD01RR8384	36,85	22,91
23, CS3032PTA276-3-4	44,31	16,84	-	-	-
24, B3PTA216-1-9	43,48	18,47	-	-	-
25, BARC-8	45,18	18,27	-	-	-
26, CS3032PTA276-1-2	43,75	17,13	-	-	-
27, CS3032PTA167-1-2	41,40	18,41	-	-	-
28, B3PTA213-3-4	43,23	18,50	-	-	-
29, BR8014887	44,71	17,28	-	-	-
Média geral	40,22	19,01	-	38,72	20,25
F	15,10**	7,05**	-	14,20**	8,29**
CV (%)	3,713	5,01	-	3,74	4,73
Maior QMR/Menor QMR	4,78	4,01	-	3,79	5,55
Razão CVg/CVe	1,9408	1,50	-	1,68	1,35

\*\*Significativo a 1% de probabilidade, pelo teste F. Grupos 1 e 2 referem-se, respectivamente, aos 29 e 22 genótipos de soja avaliados quanto ao teor de proteína e óleo, em quatro ambientes: Viçosa, MG, dezembro de 2009; Visconde do Rio Branco, MG, fevereiro de 2010; e São Gotardo, MG, fevereiro de 2010 e outubro de 2011.

moleculares dos grupos de ligação – A1, B2, C2, D1, D1a, D1b, D2, E, F, H, I, K, L, M, N e O – foram analisados no segundo grupo de genótipos: Satt006; Satt468; Satt510; Satt020; Satt196; Satt562; Satt239; Satt308; Satt250; Satt156; Satt496; Satt313; Satt398; Satt184; Satt458; Satt166; Satt229; Satt540; Satt274; Satt317; Satt420; Satt479; Satt277; Satt212; Satt571; Satt144; Satt257; Satt551; Satt141; Satt355; Satt384; Satt200; e Satt259. Os respectivos QTLs relacionados aos teores de proteína e óleo, na região dos marcadores moleculares mapeados em diferentes populações de soja, são individualmente listados por Rodrigues et al. (2013) e disponíveis no banco de dados Soybase (Grant et al., 2010).

Com os dados moleculares, as distâncias genéticas foram estimadas por meio do complemento do índice de similaridade ponderado, como apresentado a seguir:

$$D = 1 - S_{ii}; \quad S_{ii} = \frac{1}{2} \sum_{j=1}^L p_j c_j; \quad p_j = a_j / A;$$

em que:  $p_j$  é o peso associado ao loco  $j$ ;  $a_j$  é o número total de alelos do loco  $j$ ;  $A$  é o número total de alelos estudados; e  $c_j$  é o número de alelos comuns entre os pares de acessos  $i$  e  $i'$ . No primeiro e segundo grupos de genótipos, investigou-se a relação entre as estimativas de divergência e de variância genética, em relação a cada característica quantitativa individualmente. Para avaliar essa relação, cada grupo de genótipos foi dividido inicialmente em dois e, depois, em três subgrupos, conforme as quatro metodologias mencionadas a seguir.

A primeira metodologia define subgrupos a partir de um grupo inicial, de modo a existir diferença em relação à divergência média e igual número de genótipos por subgrupo. A partir das distâncias genéticas, calcularam-se os valores de divergência média de cada genótipo em relação aos acessos restantes e, depois, ordenaram-se os genótipos, de maneira crescente, pelos valores de divergência média. Os genótipos foram, então, divididos em subgrupos com menor e maior divergência média e igual número de genótipos por subgrupo.

A segunda metodologia estabelece subgrupos de modo a existir homogeneidade dentro dos subgrupos e heterogeneidade entre os subgrupos. A partir das distâncias genéticas pelo complemento do índice de similaridade ponderado, os genótipos foram agrupados

pelo método de ligação média entre grupos (UPGMA) (Cruz et al., 2011) e, a partir desse dendrograma, definiram-se subgrupos em que a similaridade era máxima.

Na terceira metodologia, os subgrupos são definidos de modo a existir divergência dentro dos subgrupos. Pelas estimativas de similaridade genética, pelo índice de similaridade ponderado, os genótipos foram agrupados pelo método UPGMA e, pelo dendrograma, definiram-se subgrupos em que a divergência era máxima.

Na quarta metodologia, para a definição de subgrupos, utilizam-se os critérios de proximidade dos acessos em uma projeção bi- ou tridimensional da matriz de distância e igual número de genótipos por subgrupo. Com a projeção tridimensional da matriz de distância, calculada com base no complemento do índice de similaridade ponderado, definiram-se os subgrupos pela proximidade na projeção, tendo-se mantido o mesmo número de genótipos em cada subgrupo. Cada uma das metodologias se baseia na diferença de divergência média entre os subgrupos, na homogeneidade ou heterogeneidade dentro dos subgrupos e na proximidade dos acessos em uma projeção da matriz de distância, respectivamente.

Para o estudo da relação entre os parâmetros de divergência e variância genética, as estimativas dessas variáveis em cada par de subgrupos foram analisadas, e o número de ocorrências do evento de maior variância, em um subgrupo com maior divergência média, foi contabilizado em cada análise. Após a comparação destas estimativas, em todos os possíveis pares de subgrupos, contabilizou-se a frequência de concordância entre elas, em cada análise individual, tendo-se, em seguida, considerado as análises conjuntamente. Na contagem dos eventos de concordância entre as maiores estimativas dos parâmetros, analisaram-se as variáveis teor de proteína e teor óleo nos grãos.

Para cada subgrupo de genótipos, calcularam-se a divergência média e a variância genética das características fenotípicas em cada ambiente. A divergência média nos subgrupos foi calculada pela média aritmética das distâncias genéticas, a partir dos coeficientes da matriz de distância. A variância genética foi obtida após análise de variância, de acordo com o modelo hierárquico, conforme Cruz (2006):  $Y_{ijk} = \mu + A_i + G/A_{ij} + B_k + \epsilon_{ijk}$ , em que:  $Y_{ijk}$  é a observação

na parcela do k-ésimo bloco, referente ao genótipo  $j$ , dentro do  $i$ -ésimo grupo;  $\mu$  é a média geral do ensaio;  $A_i$  é o efeito fixo do  $i$ -ésimo grupo ( $i=1,2,\dots,a$ );  $G/A_{ij}$  é o efeito fixo do  $j$ -ésimo genótipo, dentro do  $i$ -ésimo grupo ( $j=1,2,\dots,g_i$ );  $B_k$  é o efeito aleatório do  $k$ -ésimo bloco ( $k=1,2,\dots,r$ ); e  $\varepsilon_{ijk}$  é o erro aleatório atribuído à parcela  $ijk$ . Assim, a variância genética pôde ser estimada a partir da variação em cada ambiente, por meio do estimador  $\sigma_g^2 = (QMGenot/Grupo - QMR)/r$ , em que:  $\sigma_g^2$  é a variância genética;  $QMGenot/Grupo$  é o quadrado médio de genótipo dentro do grupo; e  $QMR$  é o quadrado médio do resíduo. Para a análise estatística utilizou-se o programa Genes (Cruz, 2013).

## Resultados e Discussão

Os marcadores moleculares de QTLs para os teores de proteína e óleo foram eficazes em distinguir os 29 genótipos de soja do Grupo 1, que exibiram ampla variação quanto ao teor de proteína. Resultado similar foi observado no Grupo 2, de 22 genótipos de soja, em que se observou ampla variação quanto ao teor de óleo (Tabela 1). No Grupo 1, a variação nos percentuais de proteína, ao longo dos ambientes, foi de 33,86 a 47,35%, enquanto, no Grupo 2, a variação do teor de óleo foi de 16,00 a 24,20%. Na análise conjunta, houve variação significativa dos teores destes componentes nos dois grupos ( $p < 0,001$ ). Outros autores também relatam variabilidade do teor de proteína e óleo em soja. Rodrigues et al. (2014), em estudo com 49 genótipos, relatam teores de proteína entre 32,07 e 48,57% e, de óleo, entre 14,40 e 24,20%. Quanto à divergência genética, as distâncias genéticas estimadas entre os 29 genótipos do Grupo 1 variaram entre 0,04 e 0,85 e apresentaram média de 0,57. Entre os 22 genótipos do Grupo 2, as distâncias genéticas variaram entre 0,06 e 0,81 e apresentaram média de 0,61. Esses valores indicam razoável nível de divergência em cada grupo. A mesma metodologia foi utilizada em outros trabalhos com cultivares elite e acessos de germoplasma de soja. Nesses trabalhos, os autores relatam intervalos de valores de distância genética de 0,19 a 0,96 (Mulato et al., 2010) e de 0 a 0,94 (Dong et al., 2014). As estimativas de distância nos Grupos 1 e 2 têm, portanto, valores de variação próximos aos da maioria dos intervalos citados, embora as distâncias genéticas não possam ser diretamente comparadas, em razão da variação dos índices. Ainda assim, pode-se afirmar que

a divergência nos grupos é ampla e próxima àquela vista em outros trabalhos, em que o número de acessos de germoplasma é ainda maior.

A partir do Grupo 1, pela metodologia 1, nos dois e três subgrupos com diferença em divergência média e igual número de genótipos por subgrupo, houve concordância entre os maiores valores de divergência média e variância genética em 87,5% (7/8) e 79,2% (19/24) das estimativas nos ambientes, respectivamente (Tabela 2). Nos dois e três subgrupos com similaridade máxima e número variável de genótipos, na análise pela metodologia 2, os maiores valores de divergência média foram observados em subgrupos com maiores valores de variância genética em 75% (6/8) e 62,5% (10/16) das estimativas nos ambientes, respectivamente (Figura 1, Tabela 3). Nos subgrupos com máxima divergência, na análise pela metodologia 3, houve concordância entre maiores valores em 93,75% (15/16) e 31,3% (5/16) das estimativas, respectivamente (Figura 2, Tabela 3). Nos subgrupos com maior proximidade na projeção tridimensional de distâncias e igual número de genótipos, na análise pela metodologia 4, houve concordância entre os maiores valores de divergência média e variância genética em 62,5% (5/8) e 83,3% (20/24) das estimativas, com a divisão dos genótipos em dois e três subgrupos, respectivamente (Figura 3, Tabela 4). Então, para o total de subgrupos, em cada análise, a concordância das estimativas foi de 81,25% (26/32), 66,7% (16/24), 62,5% (20/32) e 78,13% (25/32), respectivamente. Finalmente, o percentual geral de concordância das estimativas de divergência

**Tabela 2.** Estimativas de divergência média e variância genética, nos dois e três subgrupos de 29 genótipos de soja do Grupo 1, avaliados quanto ao teor de proteína e óleo em quatro ambientes, com base na diferença em divergência média.

N	SG	DG	A1PTN	A1OL	A2PTN	A2OL	A3PTN	A3OL	A4PTN	A4OL
2	1	0,46	8,23	3,00	7,26	2,10	6,40	1,39	7,96	2,89
	2	0,64	12,40	5,10	13,43	4,52	14,15	3,04	10,25	2,78
3	1	0,34	7,03	3,89	8,89	2,46	4,91	1,53	5,71	2,12
	2	0,52	8,80	0,25	4,92	0,86	8,37	1,29	8,98	2,62
	3	0,64	11,98	5,90	11,74	3,46	9,55	2,76	8,24	2,93

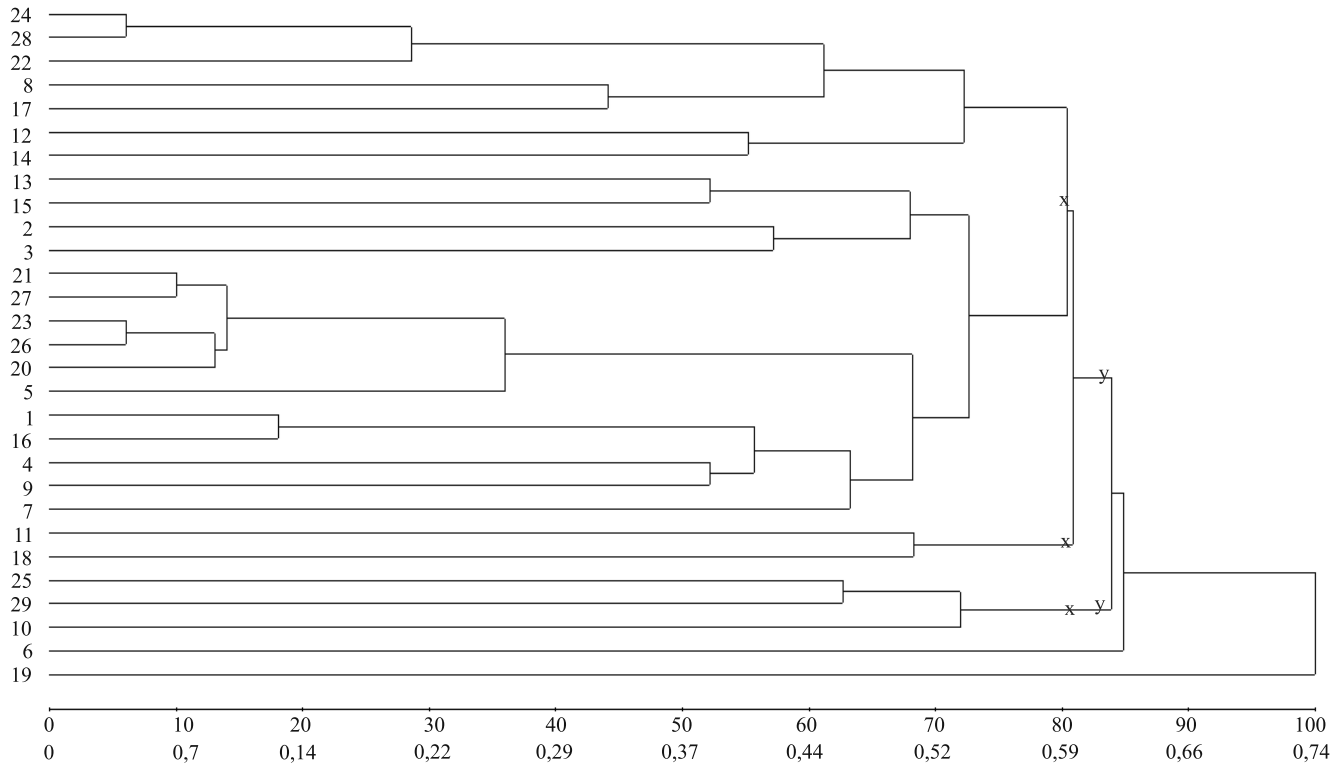
N, número de subgrupos; SG, numeração do subgrupo; DG, divergência média calculada; PTN, teor de proteína; e OL, teor de óleo. Ambientes no Estado de Minas Gerais, A1 a A4: A1, Viçosa, dezembro de 2009; A2, Visconde do Rio Branco, fevereiro de 2010; A3, São Gotardo, fevereiro de 2010; e A4, São Gotardo, outubro de 2011.

e variância genética, contabilizando-se o total de análises a partir do Grupo 1, foi de 72,5% (87/120), o que evidencia a relação direta entre as estimativas.

A relação entre a divergência e a variância genética de algumas características agrônômicas na mesma espécie é estudada em outros trabalhos na literatura. Kisha et al. (1997), pela primeira vez, relatam valores elevados de correlação entre os parâmetros ( $r=0,73$ ,  $r=0,82$ ), embora, somente em uma característica – altura de plantas –, e em apenas um entre os grupos de populações avaliados. Manjarrez-Sandoval et al. (1997), relatam, quanto aos mesmos parâmetros, a correlação positiva para o caráter produtividade de grãos ( $r=0,58$ ) em outro grupo de populações. Barroso et al. (2003) relatam ausência de correlação entre os parâmetros, quanto ao caráter produtividade, e baixa correlação, quanto ao caráter altura de plantas, em que os coeficientes variaram entre 0,10 e 0,44 nas gerações avaliadas. Bonato et al. (2006) observaram coeficientes de correlação entre 0,16 e 0,5, entre sete componentes de produção de grãos, avaliados para seis populações cultivadas em quatro épocas de semeadura.

Estes trabalhos investigam a mesma relação entre os parâmetros de divergência e variância genética. No entanto, é baixa a repetibilidade dos coeficientes de correlação, tanto entre os estudos, como também em cada estudo em que se avalia mais de uma população ou época de cultivo. Ainda assim, a relação positiva entre os parâmetros de divergência e variância no Grupo 1 concorda com alguns coeficientes de correlação relatados em populações nos trabalhos de Kisha et al. (1997) e Manjarrez-Sandoval et al. (1997).

No presente estudo, a relação positiva entre os parâmetros de divergência e variância genética pode ser atribuída à amostragem de QTLs relacionados à variabilidade das características em estudo. Outra razão é a natureza codominante do marcador molecular utilizado, que tem informatividade genética superior àqueles de natureza dominante em trabalhos anteriores (Barroso et al., 2003; Bonato et al., 2006), e provável acurácia na estimativa da divergência genética, o que contribui para a correlação entre os parâmetros, por reduzir o viés. Outra razão pode ser atribuída às



**Figura 1.** Definição de dois (letra y) e três (letra x) subgrupos de genótipos com máxima similaridade, a partir do Grupo 1 de 29 genótipos de soja.

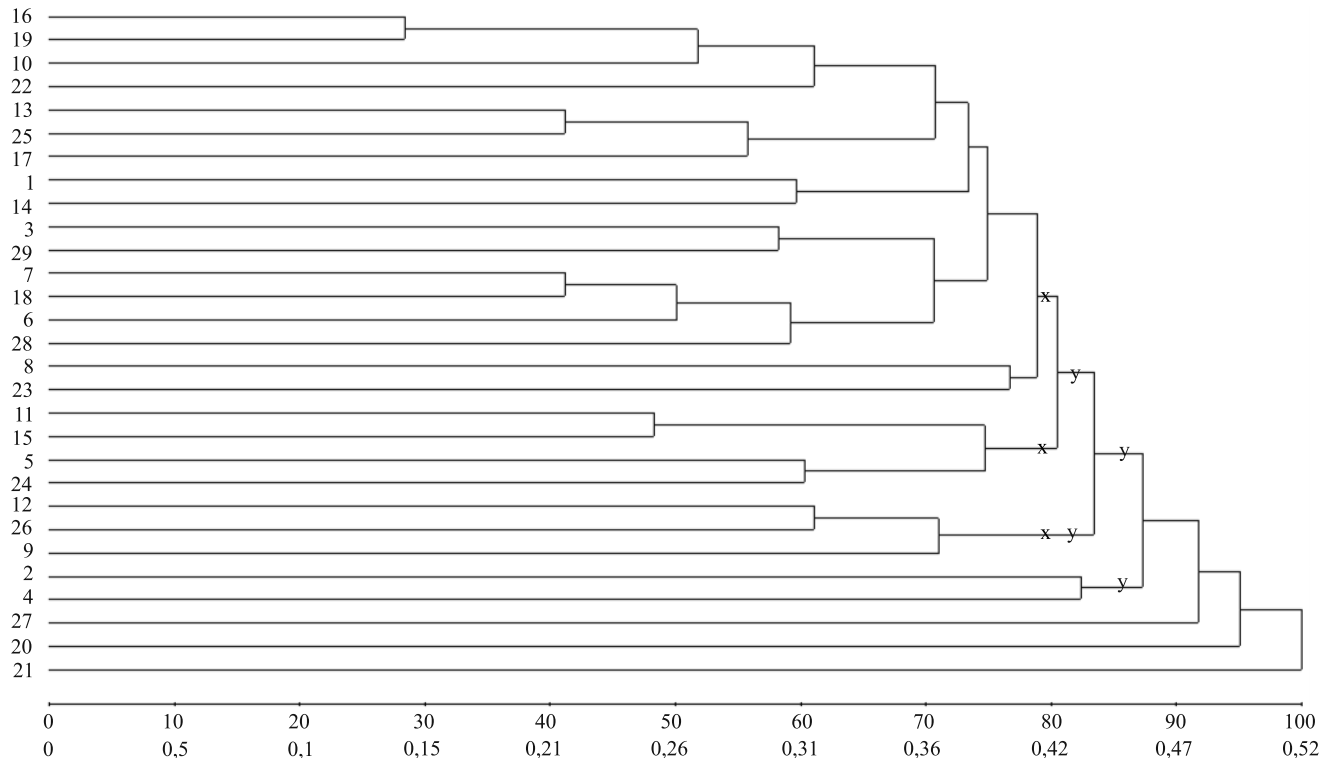
características agronômicas estudadas. Espera-se, para a variância genética quanto ao teor de proteína e óleo, maior repetibilidade e menor desvio-padrão em relação

à variância para os componentes de produção de grãos, uma vez que a produtividade e seus componentes têm expressão mais complexa e influenciada por outros

**Tabela 3.** Estimativas de divergência média e variância genética nos subgrupos de 29 genótipos de soja do Grupo 1, avaliados quanto ao teor de proteína e óleo em quatro ambientes, com máxima e mínima similaridade.

N	SG	DG	A1PTN	A1OL	A2PTN	A2OL	A3PTN	A3OL	A4PTN	A4OL
Máxima similaridade										
2	1	0,54	12,51	4,94	7,53	2,22	7,08	1,89	8,80	2,65
	2	0,51	9,75	0,67	3,03	0,05	20,26	0,07	10,21	1,59
3	1	0,52	11,76	4,23	7,47	1,98	7,23	1,72	8,63	2,79
	2	0,51	7,17	8,12	18,01	9,33	1,13	4,12	0,43	0,47
	3	0,51	9,75	0,67	3,03	0,05	20,26	0,07	10,21	1,59
Máxima divergência										
2	1	0,60	11,46	4,20	10,09	3,18	11,43	2,29	9,55	2,65
	2	0,57	1,41	1,14	0,35	0,65	3,87	1,70	-0,05	2,42
2	1	0,61	11,70	4,40	10,51	3,29	11,41	2,26	9,49	2,44
	2	0,64	14,70	4,38	11,72	3,58	14,37	3,58	14,25	5,52
3	1	0,62	12,38	4,93	12,70	3,88	13,55	2,83	10,29	2,82
	2	0,64	10,96	2,35	0,30	0,64	2,70	-0,18	7,63	1,13
	3	0,64	14,70	4,38	11,72	3,58	14,37	3,58	14,25	5,52

N, número de subgrupos; SG, numeração do subgrupo; DG, divergência média calculada; PTN, teor de proteína; e OL, teor de óleo. Ambientes no Estado de Minas Gerais: A1, Viçosa, dezembro de 2009; A2, Visconde do Rio Branco, fevereiro de 2010; A3, São Gotardo, fevereiro de 2010; e A4, São Gotardo, outubro de 2011.

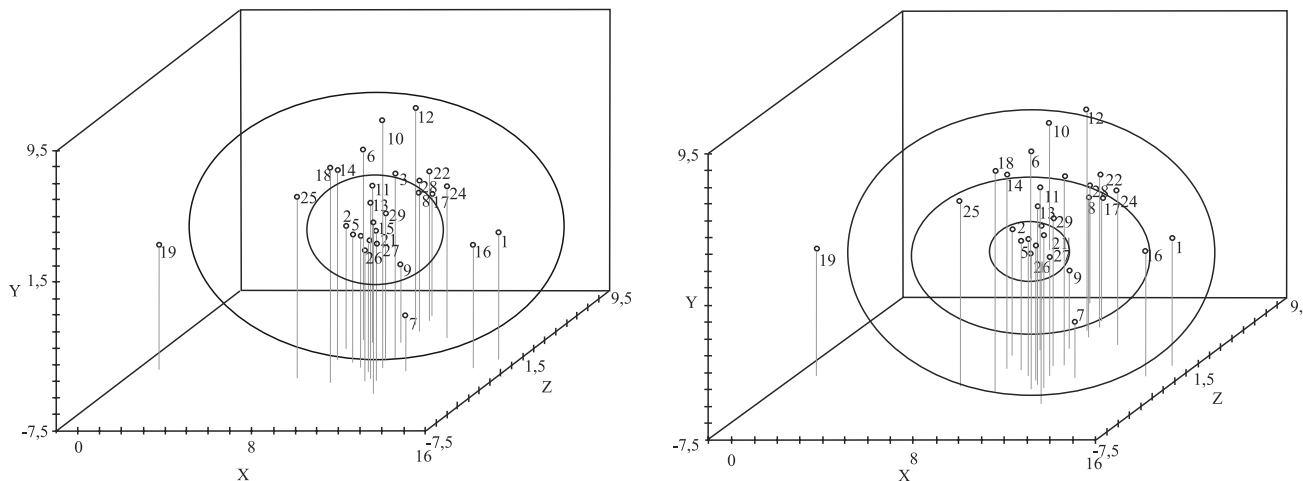


**Figura 2.** Definição de dois (letra y) e três (letra x) subgrupos de genótipos com máxima dissimilaridade, a partir do Grupo 1 de 29 genótipos de soja.

efeitos além do genótipo (Fritsche-Neto & Borém, 2011).

A partir do Grupo 2 e metodologia 1, houve concordância entre maiores valores de divergência média e variância genética em 87,5% (7/8) e 83,3% (20/24) das estimativas nos ambientes, nas análises com dois e três subgrupos respectivamente (Tabela 4).

Nos subgrupos com similaridade máxima, pela metodologia 2, os maiores valores de divergência média foram observados em subgrupos com maiores valores de variância genética em 75% (6/8) e 79,2% (19/24) das estimativas nos ambientes (Figura 4 A, Tabela 5). Nos subgrupos com máxima divergência pela metodologia 3, ocorreu concordância entre



**Figura 3.** Definição de subgrupos de genótipos, com base na proximidade na projeção tridimensional de distâncias, a partir do Grupo 1 de 29 genótipos de soja. A e B referem-se aos dois e três subgrupos, respectivamente, definidos com as circunferências a partir do Grupo 1.

**Tabela 4.** Estimativas de divergência média e variância genética, nos subgrupos de 29 e 22 genótipos de soja dos Grupos 1 e 2, respectivamente, avaliados quanto ao teor de proteína e óleo em quatro ambientes, com proximidade na projeção tridimensional da matriz de distância.

N	SG	DG	A1PTN	A1OL	A2PTN	A2OL	A3PTN	A3OL	A4PTN	A4OL
Grupo 1										
2	1	0,48	9,22	3,62	9,39	2,62	7,26	1,26	8,20	3,05
	2	0,59	13,90	4,18	8,75	1,42	11,87	2,26	10,29	2,21
3	1	0,35	8,35	3,66	8,60	2,35	3,49	1,21	5,10	2,17
	2	0,61	12,74	4,20	13,77	3,10	15,37	2,23	13,01	3,96
	3	0,57	12,13	2,85	7,06	0,97	8,52	1,62	8,18	1,50
Grupo 2										
2	1	0,54	2,27	0,15	8,81	0,83	8,40	1,61	5,77	1,11
	2	0,66	9,73	2,74	7,12	4,11	8,79	4,35	6,42	2,90
3	1	0,51	0,31	0,37	3,27	0,61	2,03	1,37	1,05	0,87
	2	0,59	8,34	0,84	17,60	1,82	20,15	2,56	10,90	2,44
	3	0,63	9,07	1,94	5,64	4,42	4,79	3,86	4,27	1,94

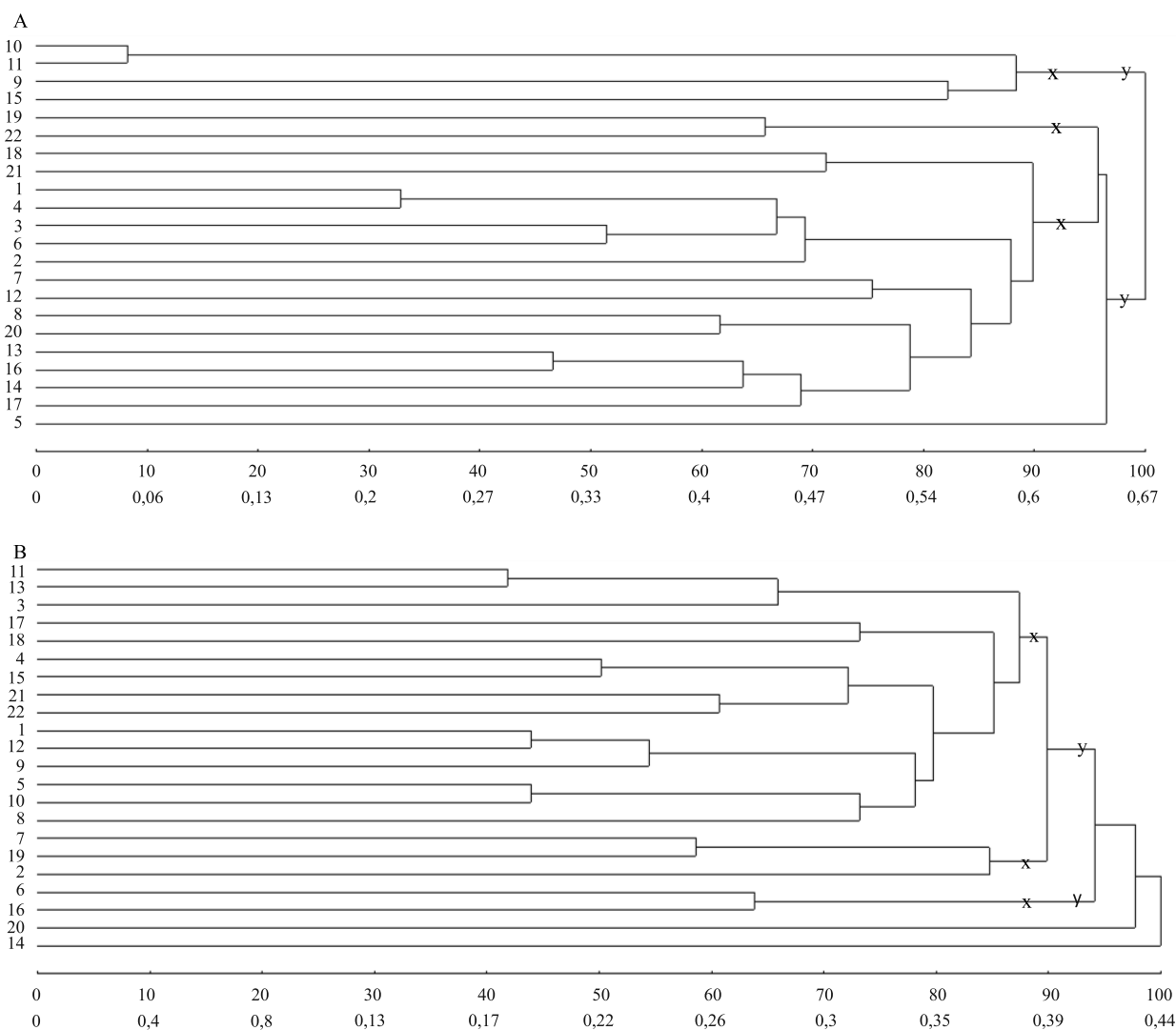
N, número de subgrupos; SG, numeração do subgrupo; DG, divergência média calculada; PTN, teor de proteína; e OL, teor de óleo. Ambientes no Estado de Minas Gerais: A1, Viçosa, dezembro de 2009; A2, Visconde do Rio Branco, fevereiro de 2010; A3, São Gotardo, fevereiro de 2010; e A4, São Gotardo, outubro de 2011.



os maiores valores dos parâmetros, em 75% (6/8) e 41,6% (10/24) das estimativas nos ambientes (Figura 4 B, Tabela 5). Nos subgrupos de genótipos com proximidade na projeção tridimensional pela metodologia 4, houve concordância entre os maiores valores de divergência média e variância genética, em 62,5% (5/8) e 87,5% (21/24) das estimativas nos ambientes (Figura 5, Tabela 6). Assim, no total de subgrupos, em cada análise, a concordância das estimativas foi de 84,38% (27/32), 78,13% (25/32), 50% (16/32) e 81,25% (26/32). O percentual geral de concordância entre os maiores valores de divergência média e variância genética, contabilizando-se o total

de análises a partir do Grupo 2, foi de 73,4% (94/128), o que indica novamente a relação positiva entre os parâmetros.

Essa relação positiva entre os parâmetros de divergência e variância genética, contudo, foi poucas vezes observada nos trabalhos anteriores com a mesma espécie (Kisha et al., 1997; Barroso et al., 2003; Bonato et al., 2006). Mas, em ao menos dois trabalhos, é também observada relação positiva entre os mesmos parâmetros (Kisha et al., 1997; Manjarrez-Sandoval et al., 1997). Todavia, não é clara a relação entre os parâmetros de divergência e variância genética nos trabalhos anteriormente publicados, em que se relata



**Figura 4.** Definição de dois (letra y) e três (letra x) subgrupos de genótipos com máxima similaridade (A) e máxima dissimilaridade (B), a partir do Grupo 2 de 22 genótipos de soja.

tanto correlação positiva como ausência de correlação (Kisha et al., 1997; Barroso et al., 2003; Bonato et al., 2006). No presente estudo, no entanto, a relação positiva entre os parâmetros de divergência e variância genética é constatada mais de uma vez, a partir de diferentes populações de genótipos. A relação entre os parâmetros é também muito próxima nos Grupos 1 e 2 (0,72, 0,73), o que indica consistência da relação observada.

No entanto, há na literatura indícios de relação positiva entre os parâmetros de heterose e divergência

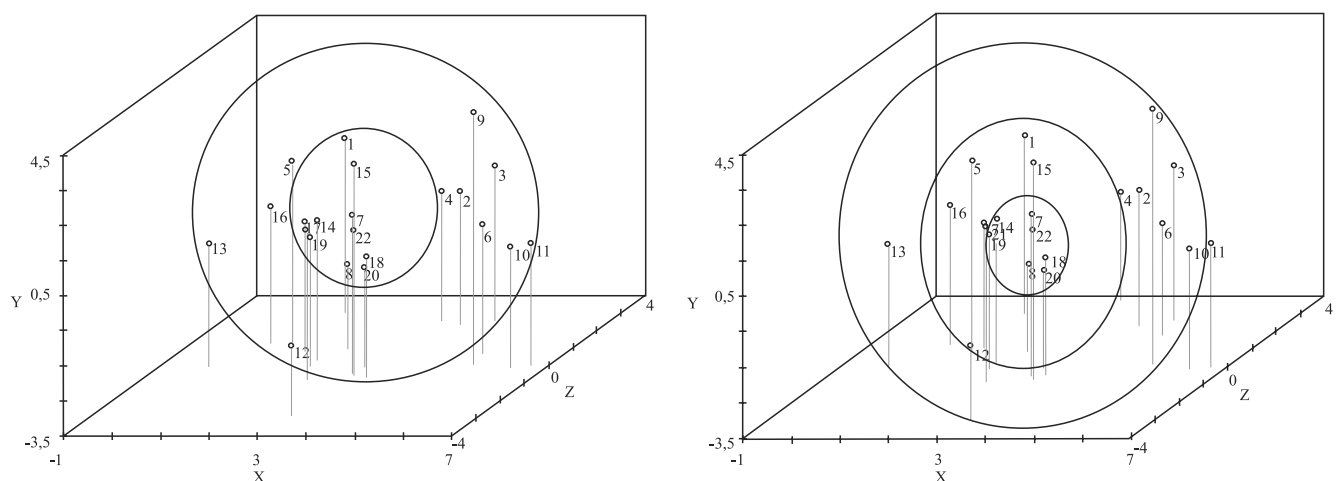
genética. Em ao menos três trabalhos, com espécies de trigo (Krystkowiak et al., 2009), soja (Colombori Filho et al., 2010) e *Brassica napus* L. (Riaz et al., 2001), a divergência entre os genitores, com base em marcadores moleculares, apresentou correlação positiva com a expressão da heterose.

A correlação nos trabalhos citados, entre heterose e divergência, por sua vez, concorda com a relação observada entre os parâmetros de variância e divergência genética nos Grupos 1 e 2, uma vez que

**Tabela 5.** Estimativas de divergência média e variância genética, nos subgrupos de 22 genótipos de soja do Grupo 2, avaliados quanto ao teor de proteína e óleo em quatro ambientes, com máxima e mínima similaridade.

N	SG	DG	A1PTN	A1OL	A2PTN	A2OL	A3PTN	A3OL	A4PTN	A4OL
Máxima similaridade										
2	1	0,5	3,53	1,69	7,71	7,35	2,29	4,15	-0,06	0,21
	2	0,59	6,77	2,15	8,66	1,58	9,11	2,56	7,12	2,28
3	1	0,5	3,65	1,69	7,77	7,37	2,31	4,17	-0,07	0,21
	2	0,44	-0,06	0,77	-1,32	1,34	-0,08	0,98	-0,24	0,36
	3	0,56	4,03	0,52	10,22	1,35	9,88	1,79	6,05	1,53
Máxima divergência										
2	1	0,63	5,98	2,52	7,01	2,56	5,86	2,87	4,54	1,98
	2	0,72	6,39	0,16	16,69	2,59	26,99	5,58	12,41	-0,04
3	1	0,64	6,86	2,85	7,95	2,92	6,51	3,06	4,94	2,09
	2	0,66	2,78	1,47	1,11	0,05	-0,29	0,19	4,00	-0,03
	3	0,72	6,39	0,16	16,69	2,59	26,99	5,58	12,41	-0,04

N, número de subgrupos; SG, numeração do subgrupo; DG, divergência média calculada; PTN, teor de proteína; e OL, teor de óleo. Ambientes no Estado de Minas Gerais: A1, Viçosa, dezembro de 2009; A2, Visconde do Rio Branco, fevereiro de 2010; A3, São Gotardo, fevereiro de 2010; e A4, São Gotardo, outubro de 2011.



**Figura 5.** Definição de subgrupos de genótipos, com base na proximidade na projeção tridimensional de distâncias, a partir do Grupo 2 de 22 genótipos de soja. A e B referem-se aos dois e três subgrupos, respectivamente, definidos com as circunferências a partir do Grupo 2.

**Tabela 6.** Estimativas de divergência média e variância genética, nos subgrupos de 22 genótipos de soja do Grupo 2, avaliados quanto ao teor de proteína e óleo em quatro ambientes, com proximidade na projeção tridimensional da matriz de distância.

N	SG	DG	A1PTN	A1OL	A2PTN	A2OL	A3PTN	A3OL	A4PTN	A4OL
2	1	0,61	3,87	1,36	8,45	1,81	7,70	3,54	5,86	2,25
	2	0,63	8,86	2,97	8,11	3,13	9,15	2,48	6,15	1,75
3	1	0,59	1,45	1,40	3,27	0,60	0,94	2,01	2,49	0,92
	2	0,63	11,58	2,94	9,94	2,44	9,28	2,85	9,89	3,74
	3	0,62	6,56	1,81	10,97	3,26	12,70	2,56	5,50	0,96

N, número de subgrupos; SG, numeração do subgrupo; DG, divergência média calculada; PTN, teor de proteína; e OL, teor de óleo. Ambientes no Estado de Minas Gerais: A1, Viçosa, dezembro de 2009; A2, Visconde do Rio Branco, fevereiro de 2010; A3, São Gotardo, fevereiro de 2010; e A4, São Gotardo, outubro de 2011.

a heterose é também resultado da variabilidade e complementaridade alélica no cruzamento, assim como a variância genética (Cruz, 2005). A heterose, do mesmo modo que a variância genética, relaciona-se com a liberação de variabilidade genética no cruzamento, e se espera que a variância genética esteja relacionada à divergência no cruzamento, tal como observado nos Grupos 1 e 2. Assim, uma vez que a relação entre os parâmetros de divergência e variância genética é positiva nos Grupos 1 e 2, a capacidade preditiva dos marcadores moleculares é também evidenciada em relação à variância genética quanto ao teor de proteína e óleo no grão de soja.

### Conclusões

1. Há relação positiva entre as estimativas de divergência e variância genética para os teores de proteína e de óleo dos grãos nos genótipos avaliados.

2. As distâncias genéticas com base nos marcadores moleculares de locos de características quantitativas (QTLs) são eficientes para a predição da variabilidade genética em genótipos de soja para os teores de proteína e de óleo dos grãos.

### Agradecimento

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, processo 141627/2008-5), pelo apoio financeiro.

### Referências

BARROSO, P.A.V.; GERALDI, I.O.; VIEIRA, M.L.C.; PULCINELLI, C.E.; VENCOSKY, R.; DIAS, C.T.S. Predicting performance of soybean populations using genetic distances

estimated with RAPD markers. **Genetics and Molecular Biology**, v.26, p.343-348, 2003. DOI: 10.1590/S1415-47572003000300020.

BONATO, A.L.V.; CALVO, E.S.; ARIAS, C.A.A.; TOLEDO, J.F. de F.; GERALDI, I.O. Prediction of genetic variability through AFLP-based measure of genetic distance in soybean. **Crop Breeding and Applied Biotechnology**, v.6, p.30-39, 2006. DOI: 10.12702/1984-7033.v06n01a05.

CHOUNG, M.G.; BAEK, I.Y.; KANG, S.T.; HAN, W.Y.; SHIN, D.C.; MOON, H.P.; KANG, K.H. Determination of protein and oil contents in soybean seed by near infrared reflectance spectroscopy. **Korean Journal of Crop Science**, v.46, p.106-111, 2001.

COLOMBORI FILHO, J.M.; GERALDI, I.O.; BARONA, M.A.A. Heterose e distâncias genéticas moleculares para a produção de grãos em soja. **Ciência e Agrotecnologia**, v.34, p.940-945, 2010. DOI: 10.1590/S1413-70542010000400020.

CRESTE, S.; TULMANN NETO, A.; FIGUEIRA, A. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. **Plant Molecular Biology Reporter**, v.19, p.299-306, 2001. DOI: 10.1007/BF02772828.

CRUZ, C.D. GENES - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, v.35, p.271-276, 2013. DOI: 10.4025/actasciagron.v35i3.21251.

CRUZ, C.D. **Princípios de genética quantitativa**. Viçosa: Ed. da UFV, 2005. 394p.

CRUZ, C.D. **Programa Genes: estatística experimental e matrizes**. Viçosa: Ed. da UFV, 2006. 285p.

CRUZ, C.D.; FERREIRA, F.M.; PESSONI, L.A. **Biometria aplicada ao estudo da diversidade genética**. Viçosa: Suprema, 2011. 620p.

DIAS, L.A.S.; PICOLI, E.A.T.; ROCHA, R.B.; ALFENAS, A.C. A priori choice of hybrid parents in plants. **Genetics and Molecular Research**, v.3, p.356-368, 2004.

DONG, D.; FU, X.; YUAN, F.; CHEN, P.; ZHU, S.; LI, B.; YANG, Q.; YU, X.; ZHU, D. Genetic diversity and population structure of vegetable soybean (*Glycine max* (L.) Merr.) in China as revealed by SSR markers. **Genetic Resources and Crop Evolution**, v.61, p.173-183, 2014. DOI: 10.1007/s10722-013-0024-y.

- FRITSCHÉ-NETO, R.; BORÉM, A. **Melhoramento de plantas para condições de estresses abióticos**. Viçosa: Ed. da UFV, 2011. 250p.
- GRANT, D.; NELSON, R.T.; CANNON, S.B.; SHOEMAKER, R.C. SoyBase, the USDA-ARS soybean genetics and genomics database. **Nucleic Acids Research**, v.38, p.D843-D846, 2010. DOI: 10.1093/nar/gkp798.
- GUMBER, R.K.; SCHILL, B.; LINK, W.; KITTLITZ, E.V.; MELCHINGER, A.E. Mean, genetic variance, and usefulness of selfing progenies from intra- and inter-pool crosses in faba beans (*Vicia faba* L.) and their prediction from parental parameters. **Theoretical and Applied Genetics**, v.98, p.569-580, 1999. DOI: 10.1007/s001220051106.
- KISHA, T.J.; SNELLER, C.H.; DIERS, B.W. Relationship between genetic distance among parents and genetic variance in populations of soybean. **Crop Science**, v.37, p.1317-1325, 1997. DOI: 10.2135/cropsci1997.0011183X003700040048x.
- KRYSTKOWIAK, K.; ADAMSKI, T.; SURMA, M.; KACZMAREK, Z. Relationship between phenotypic and genetic diversity of parental genotypes and the specific combining ability and heterosis effects in wheat (*Triticum aestivum* L.). **Euphytica**, v.165, p.419-434, 2009. DOI: 10.1007/s10681-008-9761-y.
- MANJARREZ-SANDOVAL, P.; CARTER, T.E.; WEBB, D.M.; BURTON, J.W. RFLP genetic similarity estimates and coefficient of parentage as genetic variance predictors for soybean yield. **Crop Science**, v.37, p.698-703, 1997. DOI: 10.2135/cropsci1997.0011183X003700030002x.
- MULATO, B.M.; MÖLLER, M.; ZUCCHI, M.I.; QUECINI, V.; PINHEIRO, J.B. Genetic diversity in soybean germplasm identified by SSR and EST-SSR markers. **Pesquisa Agropecuária Brasileira**, v.45, p.276-283, 2010. DOI: 10.1590/s0100-204x2010000300007.
- OLIVEIRA, T.S.; HOFFMANN, L.V.; ALVES, P.F.; LUCENA, V.S.; SILVA FILHO, J.L. da. Validação de métodos laboratoriais aplicados a análises com marcadores microssatélites. **Revista Ciência Agronômica**, v.41, p.279-284, 2010. DOI: 10.1590/S1806-66902010000200016.
- PROMEGA. **Wizard genomic DNA purification kit**: technical manual. Madison: Promega Corporation, 2014. 18p. Available at: <www.promega.com/protocols/>. Accessed on: 10 July 2015.
- RIAZ, A.; LI, G.; QURESH, Z.; SWATI, M.S.; QUIROS, C.F. Genetic diversity of oilseed *Brassica napus* inbred lines based on sequence-related amplified polymorphism and its relation to hybrid performance. **Plant Breeding**, v.120, p.411-415, 2001. DOI: 10.1046/j.1439-0523.2001.00636.x.
- RODRIGUES, J.I. da S.; ARRUDA, K.M.A.; CRUZ, C.D.; PIOVESAN, N.D.; BARROS, E.G. de; MOREIRA, M.A. Associação de marcadores microssatélites com teores de óleo e proteína em soja. **Pesquisa Agropecuária Brasileira**, v.48, p.255-262, 2013. DOI: 10.1590/S0100-204X2013000300003.
- RODRIGUES, J.I. da S.; ARRUDA, K.M.A.; CRUZ, C.D.; PIOVESAN, N.D.; BARROS, E.G. de; MOREIRA, M.A. Biometric analysis of protein and oil contents of soybean genotypes in different environments. **Pesquisa Agropecuária Brasileira**, v.49, p.475-482, 2014. DOI: 10.1590/S0100-204X2014000600009.
- SANTANA, F.A.; SILVA, M.F. da; GUIMARÃES, J.K.F.; FERREIRA, M.F. da S.; PEREIRA, W.D.; PIOVESAN, N.D.; BARROS, E.G. de. Marker-assisted selection strategies for developing resistant soybean plants to cyst nematode. **Crop Breeding and Applied Biotechnology**, v.14, p.180-186, 2014. DOI: 10.1590/1984-70332014v14n3a27.

---

Recebido em 27 de novembro de 2014 e aprovado em 19 de agosto de 2015