

ARTHUR MARTINS ALMEIDA BERNARDELI

**VALIDAÇÃO DE MARCADORES MOLECULARES SNP PARA ALTO
CONTEÚDO DE PROTEÍNA EM GRÃOS DE SOJA [*Glycine max* (L.) Merr.]**

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
2019

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

B522v
2019
Bernardeli, Arthur Martins Almeida, 1993-
Validação de marcadores *SNP* para alto conteúdo de
proteína em grãos de soja [*Glycine max* (L.) Merr.] / Arthur
Martins Almeida Bernardeli. – Viçosa, MG, 2019.
viii, 39 f. : il. (algumas color.) ; 29 cm.

Orientador: Aluizio Borém de Oliveira.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 35-39.

1. Soja - Melhoramento genético. 2. Soja - Seleção.
3. Marcadores genéticos. I. Universidade Federal de Viçosa.
Departamento de Fitotecnia. Programa de Pós-Graduação em
Genética e Melhoramento. II. Título.

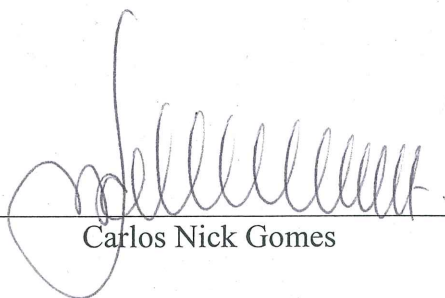
CDD 22. ed. 633.342

ARTHUR MARTINS ALMEIDA BERNARDELI

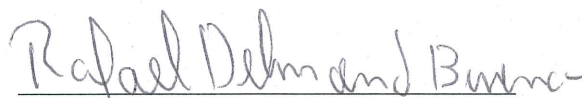
**VALIDAÇÃO DE MARCADORES MOLECULARES SNP PARA ALTO
CONTEÚDO DE PROTEÍNA EM GRÃOS DE SOJA [*Glycine max* (L.) Merr.]**

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*.

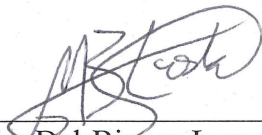
APROVADA: 22 de fevereiro de 2019.



Carlos Nick Gomes



Rafael Delmond Bueno



Maximiller Dal-Bianco Lamas Costa
(Coorientador)



Aluizio Borem de Oliveira
(Orientador)

*À minha Mãe Celida,
Ao meu Pai Vilson,
À minha irmã Carolinne,
Aos meus avós,
Aos meus demais familiares e amigos.*

DEDICO

AGRADECIMENTOS

Agradeço a Deus, por direcionar os meus passos e me proteger.

À minha mãe Celida, pelo apoio incondicional.

Ao meu pai Wilson, pela boa orientação.

À minha irmã pelo companheirismo.

À minha avó Nair e ao meu avô João (*in memoriam*), pelo carinho.

À minha Tia Celma, pelo zelo.

À minha namorada Juliane, pelo amor e ajuda.

Aos demais familiares, pelo apoio.

À Universidade Federal de Viçosa (UFV) e ao Programa de Pós-Graduação em Genética e Melhoramento, pela oportunidade.

Ao Laboratório de Bioquímica Genética de Plantas (LBGP – BioAgro-UFV) e ao Programa de Melhoramento de Qualidade de Soja, por terem me proporcionado um grande aprendizado.

Ao Laboratório de Genética Molecular de Plantas (BioAgro-UFV) pelo suporte.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Apoio à Pesquisa do Estado de Minas Gerais (FAPEMIG), pelo apoio financeiro.

Aos amigos Leonardo Risso, Vinícius Almeida, Leonardo Volpato, Rafael Aguiar, José Carlos e Reginaldo, pela amizade e companheirismo.

Aos amigos e equipe do Laboratório de Bioquímica Genética de Plantas pelo bom convívio durante este período.

Ao meu orientador, Prof. Aluizio Borem, e Coorientadores, Prof. Maximiller Dal-Bianco e Dr. Rafael Bueno, e ao amigo Dr. Newton, pelos bons conselhos e ensinamentos.

A Prof. Carlos Nick, pela disponibilidade de participação na banca.

Muito Obrigado!

BIOGRAFIA

ARTHUR MARTINS ALMEIDA BERNARDELI, filho de Celida Martins de Almeida e Vilson Bernardeli, nasceu em 13 de outubro de 1993, em Ituiutaba, Minas Gerais, porém viveu toda sua infância e adolescência em Capinópolis, Minas Gerais.

Concluiu o ensino médio no Colégio Santa Teresa em Ituiutaba no ano de 2010, e, meses depois, foi admitido no curso de Agronomia da Universidade Federal de Viçosa, dando início à realização de um sonho de quando ainda era criança.

Durante sua graduação, vivenciou experiências extracurriculares principalmente na área de melhoramento genético de plantas, tanto no Brasil quanto no exterior, com a finalidade de complementar o conhecimento adquirido em sala de aula. Em janeiro de 2017 obteve o título de Engenheiro Agrônomo pela Universidade Federal de Viçosa.

Em março de 2017, iniciou o curso de Mestrado em Genética e Melhoramento pela Universidade Federal de Viçosa, submetendo-se à defesa da dissertação em fevereiro de 2019.

SUMÁRIO

RESUMO	vii
ABSTRACT	viii
1. Introdução	1
2. Objetivo geral	3
2.1. Objetivos específicos	3
3. Revisão de literatura	3
3.1. Proteína em grãos soja: características genéticas	3
3.2. Utilidade dos marcadores moleculares no melhoramento de plantas	4
3.3. Validação de marcadores para conteúdo de proteína em grãos de soja	6
3.4. Seleção assistida e melhoramento convencional para alta proteína em soja	7
3.5. Retrocruzamento assistido por marcadores moleculares em soja	8
4. Material e Métodos	9
4.1. Material Vegetal	9
4.2. Ensaios: locais, épocas e tratos culturais	9
4.3. Delineamento experimental	9
4.4. Genotipagem	10
4.4.1. Extração de DNA	10
4.4.2. Seleção de marcadores	10
4.4.3. Metodologia de genotipagem	11
4.5. Fenotipagem	11
4.6. Análises genético-estatísticas	12
4.6.1. Estudo de associação marcador-característica	12
4.6.2. Cálculo da herdabilidade	13
4.6.3. Correlação entre conteúdos de proteína e óleo	14
4.7. Recursos computacionais	14
5. Resultados e discussão	14
5.1. Estatísticas do fenótipo	14
5.1.1. Análise de variância	14
5.1.2. Teste de normalidade	17
5.1.3. Correlação entre os caracteres	18
5.1.4. Características gerais da população	20
5.2. Associação de marcadores moleculares às características avaliadas	23
5.2.1. Marcadores associados	23

5.2.2. Combinação de loci favoráveis.....	26
5.2.3. Segregação dos loci associados em progênies superiores e inferiores	28
5.2.4. Contribuição da região cromossômica de ss190.....	31
5.3. Seleção de progênies	34
6. Conclusão.....	34
7. Refrências bibliográficas.....	35

RESUMO

BERNARDELI, Arthur Martins Almeida, M.Sc., Universidade Federal de Viçosa, fevereiro de 2019. **Validação de marcadores moleculares SNP para alto conteúdo de proteína em grãos de soja [*Glycine max* (L.) Merr.]** Orientador: Aluizio Borem de Oliveira. Coorientador: Maximiller Dal-Bianco Lamas Costa.

O objetivo deste estudo foi estimar o efeito de marcadores moleculares SNP sobre o fenótipo (r^2), determinar a herdabilidade dos conteúdos de proteína e óleo em grãos de soja e o coeficiente de correlação entre estes dois caracteres, e identificar as progênies superiores via seleção assistida e fenotipagem. Para isto, 271 RILs derivadas de F5 foram testadas em ensaios de campo em Capinópolis e Viçosa, em 2017, genotipadas, e fenotipadas para conteúdos de proteína e óleo via NIR. Os genótipos apresentaram diferenças significativas entre si e ao longo dos ambientes testados, sendo que os resultados de Capinópolis foram maiores para conteúdo de proteína das RILs (47,845 %) e os de Viçosa para o conteúdo de óleo das testemunhas (23,291 %). As herdabilidades foram maiores para a análise conjunta dos dados que para a análise individual, em que $h^2_{proteína} = 79,490\%$ e $72,810\%$, e $h^2_{óleo} = 84,190\%$ e $79,520\%$, para Capinópolis e Viçosa, respectivamente; e $h^2_{proteína} = 88,230\%$ e $h^2_{óleo} = 91,050\%$, para análise conjunta. Os coeficientes de Correlação de Pearson para dados fenotípicos foram de $r_{Capin.} = -0,664$ e $r_{Viçosa} = -0,587$. Dos marcadores moleculares avaliados, ss56, ss62, ss115 e ss190 associaram para o conteúdo de proteína em Capinópolis; e ss56, ss62 e ss190 para este mesmo caráter em Viçosa. Os marcadores ss62, ss115 e ss190 associaram para o conteúdo de óleo em ambas as localidades. O marcador ss190 destacou-se ao apresentar r^2 entre 25-29 % para ambos os conteúdos nas duas localidades. Efeito pleiotrópico de todos os marcadores foi observado, exceto para ss56, em que seu uso na seleção assistida promove o aumento do conteúdo de proteína e não onera o de óleo. Em termos práticos, as progênies superiores para alta proteína foram as que reuniram todos ou a maioria dos alelos favoráveis. As progênies 61-9, 78-38, 84-13 e 84-33 foram superiores para o conteúdo de proteína e em desempenho de campo, além de reunirem alelos favoráveis identificados via seleção assistida das marcas associadas neste estudo, justificando seu uso no programa de melhoramento.

ABSTRACT

BERNARDELI, Arthur Martins Almeida, M.Sc., Universidade Federal de Viçosa, February, 2019. **SNP validation for high seed protein content in soybean [*Glycine max* (L.) Merr.]**. Adviser: Aluizio Borem de Oliveira. Co-adviser: Maximiller Dal-Bianco Lamas Costa.

The objective of this research was to estimate the SNP effect over the phenotype (r^2), determine heritability related to seed protein and oil content and the coefficient of correlation between these two traits, as well as to identify the superior lines through marker assisted selection and phenotyping. 271 F5-derived RILs were submitted to field trials in Capinópolis and Viçosa in 2017, genotyped, and phenotyped for protein and oil contents through NIR. The genotypes presented significant difference among them and along the environments, in which the mean of protein content from Capinópolis was higher for the RILs (47,845 %) and mean from oil content from Viçosa was higher for the checks (23,291 %). Heritabilities were higher for the joint analysis than to individual ones, in which $h^2_{protein} = 79,490\%$ and $72,810\%$, and $h^2_{oil} = 84,190\%$ and $79,520\%$, for Capinópolis and Viçosa, respectively; and $h^2_{protein} = 88,230\%$ and $h^2_{oil} = 91,050\%$, for the joint data. Phenotypic Pearson Correlation were $r_{Capin.} = -0,664$ and $r_{Viçosa} = -0,587$. From the markers studied, ss56, ss62, ss115 and ss190 showed significant association to protein content in Capinópolis; and ss56, ss62 and ss190 for this same trait in Viçosa. The markers ss62, ss115 and ss190 showed significant association to oil content for both locations. SNP ss190 caught the attention by presenting r^2 in a range of 25-29 % for both contents in all locations. Pleiotropic effect of all markers was observed, except for ss56. Its use in marker assisted selection promotes the increase of protein content without hampering oil rates. The transgressive lines for high protein were the ones that gathered all or the most of favorable alleles. The lines 61-9, 78-38, 84-13 e 84-33 were superior for protein content and field performance, and gathered favorable alleles identified through marker assisted selection of the significantly associated SNPs, warranting their use in the breeding program.

1. Introdução

A soja [*Glycine max* (L.) Merr.] é a uma das principais espécies vegetais largamente cultivadas em todo o mundo. Atualmente, os Estados Unidos se consolidam como o principal produtor mundial de soja, seguido pelo Brasil. Quanto às exportações, o Brasil lidera com 79.5 milhões de toneladas, à frente dos Estados Unidos, com 51.02 milhões de toneladas (USDA, 2019).

Esta espécie é a cultura mais utilizada como fonte de proteína e possui conteúdo médio de 40% deste composto no grão (Zhang et al., 2015). Isto denota a importância da cultura para a agricultura, economia e segurança alimentar mundial, uma vez que os mercados consumidores buscam uma fonte de proteína segura e acessível.

Os valores médios para o conteúdo de proteína em grãos de soja estão em declínio, e implicam na indústria e comércio dos produtos derivados desta cultura. Como exemplo, um dos principais itens úteis à alimentação animal é o farelo de soja. No Brasil, em décadas passadas, obtinha-se o farelo concentrado de soja com conteúdo proteico entre 46 e 48 %, com menor gasto de matéria prima. Atualmente, para se atingir tal valor, é necessário um maior dispêndio de grãos para tal fim, visto que o conteúdo de proteína decresceu ao longo dos anos (Landgraf, 2015). Em contrapartida, empresas de processamento de grãos de soja estão trabalhando com recomendação de cultivares de alto conteúdo proteico, além de adoção de um mecanismo de compensação financeira ao produtor que entrega sua soja com maior conteúdo médio de proteína no momento da venda (Plume, 2018).

O caráter conteúdo de proteína em soja é de herança quantitativa e se correlaciona negativamente com a maioria das principais características levadas em consideração em um programa de melhoramento genético de soja. Os percentuais proteicos são difíceis de serem elevados, uma vez que a produtividade é influenciada ao objetivar incremento proteico, e os programas de melhoramento ainda priorizam potencializar a produtividade de grãos, por meio de tolerância à seca e resistência a insetos e doenças.

Coefficientes de correlação genotípicos negativos entre conteúdo de proteína e conteúdo de óleo e produtividade justificam seu decréscimo (Chung et al., 200; Bandillo et al., 2015; Patil et al., 2018; Zhang et al., 2018), apesar de sua herdabilidade variar entre 79 e 89 % em diferentes ambientes segundo Zhang et al. (2015).

A complexidade ao se melhorar o conteúdo de proteína em grãos de soja, devido à necessidade da reunião de combinações alélicas específicas nos vários loci que

codificam para altas médias do caráter é mitigada ao se reunir metodologias promissoras de aumentar o conteúdo proteico do grão de soja sem comprometer a produtividade.

Estratégias voltadas para a validação de marcadores moleculares em populações estruturadas e a seleção assistida por marcadores aliadas às metodologias de melhoramento tradicionais, já estabelecidas, são opções para ajudar na celeridade do processo, uma vez que a necessidade de aumentar o conteúdo de proteína em grãos de soja vem notadamente ganhando espaço e relevância nos programas comerciais e institucionais de melhoramento genético de soja em todo o mundo. Frente à demanda, foram descritos vários *QTLs* (*Quantitative trait loci*) de efeito importante para o conteúdo de proteína nos 20 cromossomos da espécie *Glycine max* (L.) Merr. ($2n=40$) (<https://www.soybase.org>).

De maneira geral, estudos de validação são utilizados para testar a associação entre marcadores moleculares e *QTLs* (Jun et al., 2008) e, uma vez constatada a associação, a seleção assistida por marcadores piramida alelos favoráveis de forma mais rápida.

Este processo se torna ainda mais complexo no melhoramento convencional à medida que o número de genes envolvidos aumenta, inviabilizando a rastreabilidade de todos os alelos dos genes envolvidos no estudo por meio da avaliação fenotípica, apenas. Por outro lado, o processo é facilitado ao se utilizar marcadores moleculares, com possibilidade de se omitir a fenotipagem em algumas etapas de seleção (Collard e Mackill, 2008).

A transferência de alelos via seleção assistida já se mostrou eficiente para o aumento do conteúdo de proteína em soja, aplicada no pré-melhoramento de genitores que resultaram em populações ou linhagens superiores, a exemplo dos resultados de Zhang et al., 2015. Seu estudo consistiu na validação de marcadores SSR e em dois ciclos de seleção assistida. Ao fim, obtiveram um genótipo superior em 9 % em comparação aos genitores para o conteúdo de proteína.

Desta forma, os parentais, adaptados e superiores, que ainda têm alelos desfavoráveis para alta proteína passam por retrocruzamento assistido por marcadores moleculares com *landraces* doadoras de alelos de alta proteína. A combinação das novas versões dos genitores resulta em populações transgressivas para o caráter.

Portanto, o incremento do conteúdo protéico no grão é um desafio para os programas de melhoramento de soja e é fundamental para a manutenção de sua oferta mundial.

2. Objetivo geral

Realizar a validação de 6 marcadores *SNPs* (*single nucleotide polymorphism*) e 271 *RILs* (*recombinant inbred lines*) de soja derivadas de F5 para o conteúdo de proteína no grão.

2.1. Objetivos específicos

- i. Validar os marcadores selecionados e estimar o efeito dos marcadores sobre o fenótipo (r^2) de forma individual e conjunta;
- ii. Estimar a herdabilidade relacionada ao conteúdo de proteína em grãos de soja para os ensaios executados;
- iv. Estimar o coeficiente de correlação entre conteúdo de proteína e óleo nos grãos de soja;
- v. Concluir quais as progênies de alta proteína reúnem o maior número de alelos favoráveis identificados via seleção assistida, em conjunto com altos valores fenotípicos.

3. Revisão de literatura

3.1. Proteína em grãos soja: características genéticas

Acessos de bancos de germoplasma de soja com alto conteúdo de proteína e alta produtividade são raros (Cober e Voldeng, 2000). Em 1964, Kwon e Torrie mostraram que proteína em grãos de soja se correlaciona de maneira fortemente negativa com conteúdo de óleo e produtividade dos grãos, florescimento, frutificação e maturação das plantas; e de maneira moderada com altura e acamamento de plantas. Esta evidência possibilitou inferir sobre a queda em ganhos genéticos para alta proteína em detrimento aos outros principais caracteres levados em consideração para o lançamento de uma cultivar de soja superior. Isto justifica a queda de proteína nos grãos ao longo dos ciclos de seleção nos programas de melhoramento em regiões altamente produtoras.

Além de proteína e óleo, a soja contém em torno de 30 % de carboidratos, principalmente os polissacarídeos não amídicos e oligossacarídeos. A dieta de animais monogástricos em confinamento se faz em boa parte por farelo de soja, no entanto, tais carboidratos podem ser fatores antinutricionais. Os polissacarídeos não amídicos solúveis causam sensação de saciedade, resultando em menor conversão de peso, e os oligossacarídeos (rafinose e estaquiose) implicando em desconforto abdominal e

flatulência. Segundo Bueno et al. (2018), rafinose e estaquiose, e polissacarídeos não amídicos se correlacionam positivamente com óleo (0,75 e 0,22, respectivamente) e negativamente com proteína (-0,86 e -0,35, respectivamente). Isto implica em um desafio para o melhoramento de qualidade da soja, pois ao longo das últimas décadas tem-se aumentado produtividade e a fração de óleo e carboidratos, e conseqüentemente diminuídos proteína e valor nutricional do grão.

Apesar do exposto acima, foram realizados alguns trabalhos acerca da introgressão de alelos favoráveis a alta proteína em soja, a exemplo do estudo de Dudley (1982). Sua conclusão foi a de que o método dos retrocruzamentos é eficiente para a transferência de alelos para materiais adaptados e que quanto maior a divergência genética entre os genitores, um maior número de retrocruzamentos tem de ser administrado.

Um resultado satisfatório foi obtido nos estudos de Wehrman et. al (1987), onde a seleção para alta proteína em soja após duas gerações de retrocruzamentos foi bem sucedida, matendo-se a produtividade. O mesmo ocorreu no trabalho de Wilcox e Cavins (1995), observado na terceira geração de retrocruzamentos. Presume-se que tais resultados foram de obtenção complexa, uma vez que existe uma moderada a forte correlação inversa entre proteína no grão e produtividade de soja, de estimativa $r = -0,23$ a $-0,86$ (Cober e Voldeng, 2000). Apesar do sucesso em tais trabalhos, Wilcox e Cavins (1995) levaram um total de 20 anos para realizar a fenotipagem completa entre todos os ciclos de retrocruzamento, a fim de selecionar aqueles indivíduos produtivos e de alta proteína.

O mesmo não se pode afirmar em trabalhos com progênies oriundas de populações obtidas por meio de cruzamentos simples biparentais, em que os genitores são superiores e adaptados, uma vez que os backgrounds genéticos de ambos reúnem alelos não favoráveis ao caráter, como pode ser observado em Cober e Voldeng (2000).

Desta forma, o melhoramento convencional mostra-se limitado para obtenção de progênies superiores e de alta proteína. Alternativamente e de forma paralela ao melhoramento clássico para este caráter, o uso de marcadores moleculares tem servido para tal fim, com adoção em diversos programas de melhoramento, onde resultados satisfatórios são publicados.

3.2. Utilidade dos marcadores moleculares no melhoramento de plantas

Existem três tipos fundamentais de marcadores: morfológicos, bioquímicos e os moleculares. Os marcadores morfológicos são representados pela diferença fenotípica

entre indivíduos ou espécies. Os bioquímicos incluem as variantes alélicas de enzimas, chamadas isozimas. Estas duas primeiras modalidades são limitadas em número e sofrem grande efeito ambiental a depender da característica avaliada (Winter & Kahl, 1995).

Já os marcadores moleculares são diferenças genéticas entre indivíduos ou espécies. Na maioria das vezes, eles não representam o gene em si, mas agem como um sinalizador para os mesmos, uma vez que estão em curta proximidade com os genes de interesse. Tais marcadores são abundantes, tem origem a partir de mutações de ponto, inserção ou deleção, ou erros na replicação (Paterson, 1966). Os marcadores, quando evidenciam diferenças entre indivíduos ou espécies, são ditos polimórficos.

Marcadores moleculares são aplicáveis ao melhoramento quando o fenótipo é selecionado baseando-se no genótipo das marcas avaliadas. Neste caso, os marcadores para seleção assistida têm de ser adequadamente escolhidos, caso contrário, não auxiliará na escolha correta do fenótipo. Assim, após a avaliação dos marcadores em estudos de mapeamento de QTL em uma população adequada e suficientemente grande, e validados também em populações nesta mesma conformação em ambientes representativos, estes estão prontamente disponíveis ao uso na seleção assistida.

A exemplo do presente trabalho, a validação ou estudo de associação de marcadores moleculares testa a confiabilidade da marca em prever o fenótipo, para então usá-la rotineiramente no programa de melhoramento. Por consequência, leva à economia de tempo em ensaios, seleção de genótipos em estágios de plântulas ou mesmo sementes, possibilidade de piramidação de genes de interesse e evita o *linkage drag* (arraste de ligação). No entanto, a adoção de marcadores para estudos de associação e seleção assistida tem de levar em consideração a relação custo benefício da genotipagem e fenotipagem, herança e herdabilidade do caráter e instalações adequadas (Collard et al., 2005).

O tipo de marcador adotado em programas de melhoramento também varia, principalmente devido ao custo, qualidade das amostras de DNA extraídas, nível de polimorfismo, procedimentos técnicos de genotipagem e distância entre o gene e a marca (Collard e Mackill, 2008).

Para estudos de seleção assistida para alta proteína em soja, os marcadores *SSRs* (*simple sequence repeats* ou microssatélites) são amplamente utilizados por serem baratos, co-dominantes e de boa reprodutibilidade. No entanto, necessitam do preparo do gel de géis e mostra apenas um locus por ensaio. A migração gradual para o uso de marcadores *SNPs* em trabalhos deste contexto é devido a abundância e especificidade de

tais marcadores, e ao desenvolvimento e sofisticação de técnicas de detecção de *SNPs* (Rafalski, 2002; Koebner e Summers, 2003). Zhu et al. (2003) define *SNPs* como uma variação em nucleotídeo simples em uma região específica do genoma.

Os trabalhos sobre detecção de *QTLs*, estudos de associação e seleção assistida por marcadores moleculares em soja são abundantes e relatados desde meados dos anos 80, envolvendo principalmente a resistência a patógenos, estresses abióticos e outros caracteres agronômicos.

A partir da observação da eficiência do uso dos marcadores para proteína no grão de soja, trabalhos relevantes foram desenvolvidos, a exemplo da pesquisa de Wang et al. (2014), que adotaram o uso tanto de marcadores *SSRs* quanto de *SNPs* e obtiveram resultados importantes sobre *QTLs* envolvidos com conteúdo de proteína e óleo e produtividade de grãos; Yan et al. (2014), que identificaram cinco *QTLs* associados a alta proteína utilizando um total de oito marcadores *SSR*; Warrington et al. (2015), os quais utilizaram uma abundância de marcadores *SSRs* e *SNPs* na maioria dos 20 cromossomos da espécie, em que os *QTLs* identificados poderão ser envolvidos em estudos de seleção assistida; e Zhang et al. (2015), que ao lançarem mão do uso de marcadores *SSRs* na seleção assistida, obtiveram uma progênie superior em 9 % em relação aos acessos parentais para o conteúdo de proteína.

3.3. Validação de marcadores para conteúdo de proteína em grãos de soja

Mapas de ligação que identificaram *QTLs* que controlam conteúdo de proteína em grãos de soja foram estabelecidos em estudos de Diers et al. (1992), Lee et al. (1996); Brummer et al. (1997), Orf et al. (1999), Casanadi et al. (2001), e Jun et al. (2007). Desde então, estudos de associação entre marcadores têm sido executados, como o de Jun et al. (2007), os quais identificaram dois novos marcadores *SSR* (*satt431* e *satt551*) que ainda não haviam sido reportados na literatura para tal fim.

Em outro trabalho de validação de marcadores em *QTLs* para alto conteúdo de aminoácidos específicos em grãos de soja, Warrington et al. (2015) obtiveram uma população de *RILs* derivadas de F5 a partir de um cruzamento simples entre os acessos Benning e Danbaekkong, que possuem 42 e 51 % de proteína nos grãos, respectivamente. Como resultado, encontraram quatro *QTLs* (*qProt_Gm14*, flanqueado por *BARC-018353* e *BARC-056587* no cromossomo 14; *qProt_Gm15*, flanqueado por *Sat_273* e *BARC-027786* no cromossomo 15; *qProt_Gm17* flanqueado por *BARC-025927* e *Satt256* no cromossomo 17; e *qProt_Gm20* flanqueado por *GSM0012* e *Satt354* no cromossomo 20) que explicavam de 5 a 55 % da variação total de proteína

crua na população, correspondendo a um efeito aditivo de até de 14%. Tais *QTLs* continham alelos favoráveis para aumento no conteúdo de proteína à semelhança daqueles presentes no genitor Danbaekkong.

Yan et al. (2014) objetivaram determinar o efeito genético de *QTLs* que controlam o conteúdo de proteína em soja, em duas populações derivadas de cruzamentos entre duas cultivares de *Glycine max* e uma de *Glycine soja*. No entanto, para 8 *QTLs* analisados, a contribuição para o fenótipo variou apenas entre 6,6 e 14,5 %, onde as maiores contribuições foram oriundas das cultivares da espécie *Glycine max*, ao contrário do que foi constatado por Diers et al. (1992), que atribuíram aos *QTLs* de *Glycine soja* maior contribuição para o este caráter. Portanto, o desenvolvimento de mais trabalhos a semelhança dos mencionados acima vem sendo de grande importância.

3.4. Seleção assistida e melhoramento convencional para alta proteína em soja

A seleção assistida por marcadores é um método de melhoramento complementar (Li et al., 2017), útil quando aliada às metodologias de melhoramento convencionais. Desta forma, é importante estabelecer uma fronteira entre os avanços obtidos na seleção de progênies de soja para alta proteína no grão, por meio das metodologias de melhoramento convencionais e aquelas assistidas por marcadores.

Zhang et al. (2015) objetivaram avaliar ciclos contínuos de seleção assistida por marcadores para conteúdo de proteína no grão de soja. Assim, realizaram um primeiro ciclo de seleção em duas populações. A primeira população (XG) contendo 147 *RILs* derivadas de F6, levadas pelo método *SSD* (*single seed descent*) até F5, foi oriunda do cruzamento simples entre genitores de alto e baixo conteúdo de proteína. A segunda população (WT), de estrutura e condução de gerações segregantes idênticas à anterior, foi obtida do cruzamento simples entre dois genitores de alta proteína no grão, resultando em 142 *RILs*. Tais progênies foram submetidas a ensaios de campo e fenotipadas, estimando seu conteúdo proteico. Assim, identificou-se um segregante transgressivo com alto conteúdo de proteína em cada uma das populações. Tais progênies foram genotipadas e constatou-se que ambas possuíam alelos divergentes para os três genes avaliados, nos *QTLs* Prot-08-1, Prot-14-1 e Prot-19-2, mas complementares. Tais progênies foram cruzadas, iniciando assim o segundo ciclo de seleção, gerando a população WX. As gerações subsequentes até F6 foram genotipadas e fenotipadas em ensaios de campo, onde a contribuição de cada marca para o caráter foi estimada pelo coeficiente de determinação (r^2) da análise de regressão para o estudo de seleção assistida. Ao final, houve um progresso genético de 9,32 % no conteúdo de

proteína na progênie selecionada com base nos alelos favoráveis sobre os quatro genitores iniciais, que possuíam o conteúdo na faixa de 35,35-44,83 %. A progênie final selecionada obteve 54,10 % de conteúdo proteico.

Já Cober e Voldeng (2000) obtiveram uma população X3585 por meio do cruzamento simples entre os genitores AC Proteus (alta proteína) e Maple Glen (de alta produtividade). Também, executaram o retrocruzamento de 60 indivíduos F1 da população X3585 com Maple Glen. Como resultado, em termos de Kg de semente por hectare e Kg de proteína por hectare, nenhuma das metodologias resultou em progênies superiores aos genitores Maple Glen e AC Proteus, evidenciando uma ineficiência de metodologias convencionais adotadas de forma única.

Conclui-se que a possibilidade de ganhos genéticos positivos ao longo dos ciclos de seleção é maior ao aliar o melhoramento convencional ao uso dos marcadores moleculares para este caráter.

3.5. Retrocruzamento assistido por marcadores moleculares em soja

O retrocruzamento assistido por marcadores vem se tornando uma ferramenta útil ao melhorar uma cultivar elite para um ou poucos caracteres, a depender da distância marca-gene, tamanho da população a cada geração de retrocruzamento e número de marcadores (Li et al., 2017).

Exemplos têm sido observados na introgressão de genes de resistência a doença e insetos, de resistência a herbicidas e de composição do grão em soja (Diers et al., 2005; Kim et al., 2008; Maroof et al., 2008; Landau-Ellis e Pantalone, 2009; King et al., 2016). Landau-Ellis e Pantalone (2009) retrocruzaram a cultivar CX1834-1-2 de baixo fitato com a cultivar elite 5601T.

Marcadores *SSR* monitoraram não apenas a presença de alelos recessivos para dois genes que controlam conteúdo do composto, bem como auxiliaram na seleção da progênie mais similar ao genitor recorrente. Este processo foi concluído com êxito após três gerações de retrocruzamento.

Para obtenção de progênies de soja produtivas e de alto percentual protéico, o preparo de genitores é um resultado do retrocruzamento assistido. Na China, proteína no grão de soja varia na faixa de 35 a 50 % na maioria das cultivares (Zhang et al., 2015), e nos Estados Unidos de 34,1 a 56,8 %. No entanto, as que possuem o maior percentual são aquelas pouco adaptadas ou ditas *landraces*, que ao serem combinadas, resultam em populações segregantes inferiores para os demais caracteres importantes para a obtenção de uma boa cultivar. No entanto, seus alelos favoráveis para o caráter podem ser

transferidos até as cultivares superiores pelo retrocruzamento assistido em um processo de pré-melhoramento. Posteriormente, a combinação de tais genitores preparados, tende a resultar em populações com maior frequência de segregantes superiores.

4. Material e Métodos

4.1. Material Vegetal

No ano de 2014, foi executado um cruzamento simples envolvendo dois acessos adaptados e de alto conteúdo de proteína no grão. Estes acessos pertencem ao banco de germoplasma do Programa de Melhoramento Genético de Qualidade da Soja BIOAGRO/UFV. Os genitores escolhidos foram PMQS80 e PMQS12, pertencentes aos grupos de maturidade 6,2 e 7,0, possuindo 47,007 % e 45,780 % de conteúdo de proteína, respectivamente. A partir dos indivíduos F₁, obteve-se, então, 271 sementes, e a condução de cada indivíduo ocorreu pelo método SSD desde a geração F₂ até F₅ em casa de vegetação.

Os indivíduos F₄ tiveram irrigação, fertilização e iluminação suplementar de forma com que cada planta proporcionasse quantidade elevada de sementes. Ainda em casa de vegetação, foi identificada a quantidade de dias até a maturação de cada indivíduo para tornar mais prática a aleatorização das progênies nos ensaios de campo. Posteriormente, cada planta foi colhida individualmente, dando origem a 271 RILs, as quais foram submetidas a ensaios de campo em anos consecutivos.

4.2. Ensaios: locais, épocas e tratos culturais

Os ensaios foram administrados em duas localidades, Capinópolis-MG (18° 40' 55" S 49° 34' 12" O) e Viçosa-MG (20° 45' 14" S 42° 52' 55" O), entre novembro, 2017 e março, 2018, correspondentes à safra de verão, propícia para o cultivo da espécie.

Os tratos culturais seguiram as recomendações para a cultura da soja, incluindo fertilização do solo no momento do plantio, controle químico de plantas daninhas, insetos-praga e fitopatógenos. Irrigação via aspersão foi administrada de acordo com a demanda até o momento em que as parcelas atingiram o ponto de maturação fisiológica.

4.3. Delineamento experimental

O delineamento adotado foi o de blocos casualizados a exemplo de Zhang et. al. (2015). Cada bloco possuiu 45 tratamentos, sendo 40 correspondentes às progênies e cinco correspondentes às testemunhas PMQS80, PMQS12, M7739, M8372 e

ANSC83022. A dimensão das parcelas foi de 1 m de comprimento, 0,5 m entre de largura entre parcelas e 1 m de intervalo entre blocos subseqüentes. Foram semeadas 20 sementes por metro, com posterior desbaste pós-emergência para uma população final de 12 plantas por metro. Após a maturidade fisiológica (R8), cada parcela foi colhida separadamente em bulk de sementes e amostradas para fenotipagem.

O modelo proposto foi $Y_{ijk} = \mu + G_i + (B/A)_{jk} + A_j + GA_{ij} + \varepsilon_{ijk}$, onde Y_{ijk} é a observação no k-ésimo bloco, avaliada no i-ésimo genótipo e j-ésimo ambiente; μ é a média geral do ensaio; G_i é o efeito aleatório do genótipo i; $(B/A)_{jk}$ é o efeito aleatório do bloco k dentro do ambiente j; A_j é o efeito aleatório do ambiente j; GA_{ij} é o efeito aleatório da interação entre o genótipo i e o ambiente j; e ε_{ijk} o erro aleatório associado à observação Y_{ijk} .

4.4. Genotipagem

4.4.1. Extração de DNA

Discos foliares foram coletados das plantas de soja do ensaio alocado no ambiente Viçosa em estágio V4, em janeiro de 2018. Dos tecidos foliares, foi extraído o DNA de cada progênie de acordo com Dellaporta et al. (1983). Após a verificação da qualidade do DNA extraído via eletroforese em gel de agarose 0,8 %, cada amostra foi quantificada por meio do espectrofotômetro NanoDrop (NanoDrop Technologies, Wilmington, DE, EUA) e diluída para a concentração de trabalho $10 \text{ ng.}\mu\text{L}^{-1}$ para a genotipagem dos SNPs.

4.4.2. Seleção de marcadores

Um projeto conduzido pelo Programa de Melhoramento da Qualidade da Soja (PMQS/BIOAGRO/DBB/UFV) selecionou 269 *SNPs* no banco de dados soybase (<https://www.soybase.org>), presentes em regiões associadas à *QTLs* que aumentaram o conteúdo de proteína, óleo e resistência à nematoides. Estes marcadores *SNPs* foram selecionados com base em: 1) *SNPs* publicadas em estudos de associação genômica ampla (*GWAS*) para estas características; 2) alguns *SNPs* próximas a estas anteriormente mencionadas e; 3) *SNPs* próximas a microssatélites sabidamente associados com as características mencionadas e já utilizados no programa. Posteriormente, foi realizada genotipagem dos *SNPs* selecionadas em 34 genótipos utilizados em nosso programa de melhoramento. Os dados de genotipagem foram

associados com os dados fenotípicos para o conteúdo de proteína das 34 cultivares e feito rodadas de seleção “*in silico*” para identificar os marcadores mais promissores.

Dentre marcadores associados no estudo de *GWAS*, foram selecionados aqueles polimórficos entre PMQS80 e PMQS12 para execução deste trabalho. Desta forma, a seleção assistida por marcadores foi executada em seu primeiro estágio, na seleção dos genitores, e posteriormente nas progênes deste projeto, assim como em Zhang et al. (2015). Os marcadores selecionados podem ser verificados na **Tabela 2**.

Tabela 2. Marcadores moleculares selecionados para o estudo de validação.

Marcador	Tipo	Cromossomo	GL	Trabalhos anteriores
ss46	<i>SNP</i>	1	D1a	Song et al., 2013
ss56	<i>SNP</i>	18	G	Song et al., 2013
ss62	<i>SNP</i>	20	I	Song et al., 2013
ss94	<i>SNP</i>	8	A2	Zhang et al., 2015
ss115	<i>SNP</i>	15	E	Bandillo et al., 2015
ss190	<i>SNP</i>	20	I	Bandillo et al., 2015

4.4.3. Metodologia de genotipagem

A genotipagem das 6 *SNPs* seguiu a metodologia de genotipagem KASP (LGC Genomics, 2013), executada no equipamento Applied Biosciences 7500 (AB7500). A reação de amplificação compreendeu 1 ciclo de 94 °C por 15 minutos; 10 ciclos de 94 °C por 20 segundos, com gradiente de 61-55 °C, decaindo 0,6 °C a cada 60 segundos; 30 ciclos de 94 °C por 20 segundos e 55 °C por 60 segundos; e um ciclo de 37 °C por 60 segundos. Cada reação consistiu em 2,5 uL de DNA a 10 ng.uL⁻¹, 2,5 uL de 2x Master Mix e 0,14 uL de Primer Mix. A discriminação alélica foi executada no programa computacional AB 7500 Software v.2.3.

4.5. Fenotipagem

Após colheita manual das parcelas nos ensaios de campo, uma quantidade mínima de 30 grãos por parcela foi moída em um moinho industrial modelo MA020 (Marconi Equipamentos para Laboratório, Piracicaba, SP, Brazil). O farelo obtido foi analisado para o conteúdo de proteína e óleo por espectrometria de infravermelho próximo, utilizando-se o equipamento Antaris II FT-NIR analyzer (Thermo Fisher Scientific Brasil Instrumentos de Processo Ltda., São Paulo, SP, Brasil), conforme o estudo realizado por Rodrigues et al. (2014).

Os valores de conteúdo obtidos foram convertidos para percentuais em base seca, por meio das fórmulas:

$$\%Prot_{base\ seca} = \frac{100 \times \%Prot}{100 - Umidade\ do\ farelo}, \text{ para conteúdo de proteína;}$$

$$\%Óleo_{base\ seca} = \frac{100 \times \%Óleo}{100 - Umidade\ do\ farelo}, \text{ para conteúdo de óleo.}$$

4.6. Análises genético-estatísticas

4.6.1. Estudo de associação marcador-característica

O estudo de associação marcador-característica ocorreu pela a análise de marca simples (Schuster e Cruz, 2008), por meio de regressão linear, utilizando-se o seguinte modelo.

$$Y_j = \beta_0 + \beta_1 X_{1j} + \varepsilon_j$$

em que:

Y_j = é o valor fenotípico da característica quantitativa avaliada no j-ésimo indivíduo da população;

X_j = é o código do marcador (AA = 2, Aa = 1 e aa = 0);

β_0 = intercepto da regressão (média da característica);

β_1 = inclinação da reta para efeito aditivo;

ε_j = erro aleatório manifestado na característica no j-ésimo indivíduo.

Assim, na avaliação de N indivíduos em cada geração os dados serão codificados conforme a **Tabela 1**.

Tabela 1. Codificação de acordo com os genótipos dos marcadores

Genótipo do marcador	Número de indivíduos	Valores da característica	Código de X (alelo)
AA	n ₁	$Y_{11}, Y_{21}, \dots, Y_{n1,1}$	2
Aa	n ₂	$Y_{12}, Y_{22}, \dots, Y_{n2,2}$	1
aa	n ₃	$Y_{13}, Y_{23}, \dots, Y_{n3,3}$	0

As matrizes utilizadas no modelo $Y_j = \beta_0 + \beta_1 X_{1j} + \varepsilon_j$ são descritas a seguir:

$$Y = \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \\ \dots \\ \gamma_{12} \\ \gamma_{22} \\ \dots \\ \gamma_{n3,3} \end{bmatrix} \quad X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \dots & \dots \\ 1 & 1 \\ 1 & 1 \\ \dots & \dots \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \text{e } \varepsilon = \begin{bmatrix} e_{11} \\ e_{21} \\ \dots \\ e_{12} \\ e_{22} \\ \dots \\ e_{n3,3} \end{bmatrix}$$

em que:

$$\hat{\beta} = (X'X)^{-1}(X'Y);$$

$$\widehat{\beta}_1 = \bar{Y};$$

$$\widehat{\beta}_1 = \frac{cov(X_1, Y)}{V(X_1)}.$$

O valor do coeficiente de determinação da regressão (r^2) é a proporção da variação dos valores fenotípicos explicados pelo marcador. Assim, a contribuição de cada locus é estimada pelo valor de r^2 da análise de regressão.

4.6.2. Cálculo da herdabilidade

Como proposto por Zhang et al. (2015), a herdabilidade (h^2) para o caráter conteúdo de proteína no grão foi calculada por:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma^2}{r}}, \text{ para ambientes individualizados;}$$

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ga}^2}{n} + \frac{\sigma^2}{ar}}, \text{ para os ambientes combinados;}$$

em que:

σ_g^2 é o componente de variância do efeitos aleatório dos genótipos;

σ_{ga}^2 é o componente de variância do efeito aleatório da interação genótipos por ambientes;

σ^2 é o componentete de variância associado ao erro;

a é o numero de ambientes;

r é o número de repetições.

4.6.3. Correlação entre conteúdos de proteína e óleo

Para determinar a magnitude e direção de respostas à seleção simultânea (Falconer e Mackay, 1996), foi feito um estudo de correlação fenotípica, genotípica e entre os caracteres conteúdo de proteína e óleo, e de correlação ambiental, por meio da estimação do coeficiente de correlação (r) entre as duas variáveis aleatórias X (conteúdo de proteína) e Y (conteúdo de óleo). A metodologia de obtenção pode ser verificada em Cruz et al. (2012), em que:

$$r = \frac{c\hat{\sigma}(X,Y)}{\sqrt{\hat{V}(X)\hat{V}(Y)}} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}};$$

$$x_i = X_i - \bar{X};$$

$$y_i = Y_i - \bar{Y};$$

$$c\hat{\sigma}(X, Y) = \frac{\sum_{i=1}^n x_i y_i}{n-1};$$

$$\hat{V}(X) = \frac{\sum_{i=1}^n x_i^2}{n-1};$$

$$\hat{V}(Y) = \frac{\sum_{i=1}^n y_i^2}{n-1}.$$

4.7. Recursos computacionais

As análises genético-estatísticas foram executadas nos aplicativos computacionais GENES (Cruz, 2013) e SELEGEN REML/BLUP (Resende, 2016).

5. Resultados e discussão

5.1. Estatísticas do fenótipo

5.1.1. Análise de variância

A análise de variância apresentou diferenças significativas ($P < 0,01$) para os conteúdos de proteína e óleo em ambos os ensaios ao se testar os componentes de variância dos genótipos (σ_g^2) e da interação genótipo x ambiente (σ_{ga}^2). A herdabilidade estimada pela esperança do quadrado médio da ANOVA conjunta dos dados foi de 91,050 % para o conteúdo de óleo, e de 88,230 % para o conteúdo de proteína, e assumiu valores de 84,190 % e 79,520 % para o conteúdo de óleo e 79,490 % e 72,810 % para o conteúdo de proteína, para Capinópolis e Viçosa, respectivamente (**Tabela 3**). Tal resultado sugere que a variação genética existente na população é responsável pela maior porção da variância fenotípica, uma vez que Zhang et al. (2015) e Wang et al.

(2014) também observaram um acréscimo da herdabilidade para estes caracteres ao avaliar ensaios conjuntamente. Estes mesmos autores também obtiveram valores de herdabilidade similares ao do presente estudo para os conteúdos de óleo e proteína.

Além dos resultados da análise de variância que denotam uma grande variabilidade genética da população de RILs derivadas de F5 para os conteúdos de proteína e óleo, valores elevados da razão CV_g/CV_e confirmam a sua grande contribuição para a variação total dos valores observados.

Os coeficientes de variação foram abaixo de 5% para os conteúdos de proteína e ligeiramente acima de 5% para os conteúdos de óleo em ambas as localidades. Da mesma forma, baixos valores da relação $\frac{>QMR}{<QMR}$ foram observados. Estas constatações revelam boa precisão e homogeneidade dos dados fenotípicos e das variâncias residuais, assim como no trabalho de Rodrigues et al (2014).

O resumo dos dados fenotípicos mostrados na **Tabela 3** evidencia o fato de que composição da semente de soja é significativamente afetada por condições ambientais e que a ocorrência de genótipos com valores extremos foi frequente, sendo necessária análise do padrão de distribuição dos dados fenotípicos.

Tabela 3. Resumo da análise fenotípica dos conteúdos de proteína e óleo (%) em soja em dois ambientes

Característica	Ambiente	Média ± desvio padrão			F_g^d	$F_{g \times e}^d$	DMS _(0.05) ^e	Número de RILs	CV% ^f	$\frac{CV_g}{CV_e}$	$\frac{> QMR}{< QMR}$	Herdabilidade	
		Parentais ^b	Testemunhas ^c	RILs								Simplex	Conjunta
% óleo	Capin ^a , 2017	18,389 ± 1,682	21,876 ± 1,753	18,917 ± 2,128	16,740**	11,181**	4,701	271	5,905	1,675	1,127	84,190	91,050
	Viçosa, 2017	19,295 ± 1,788	23,291 ± 2,178	19,405 ± 2,048			5,069	268	6,048	1,338		79,520	
% proteína	Capin, 2017	47,459 ± 1,814	43,396 ± 1,469	47,845 ± 2,912	11,233**	8,500**	7,168	271	3,581	1,437	1,144	79,490	88,230
	Viçosa, 2017	45,341 ± 1,944	39,911 ± 2,471	45,163 ± 2,827			7,948	268	4,073	1,150		72,810	

^a Capinópolis

^b PMQS80 e PMQS12

^c ANSC83022, M7739, M8372

^d F_g e $F_{g \times e}$ são valores de F para genótipo e interação genótipo x ambiente, respectivamente.

^e Diferença mínima de significância, em que $\alpha=0,05$

^f Coeficiente de variação

CV_g/CV_e = relação entre coeficiente de variação genotípico e residual.

QMR = quadrado médio do resíduo

* e **, valores significativos à $p<0.05$ e $p<0.01$

5.1.2. Teste de normalidade

A **Figura 1** mostra a frequência de ocorrência dos valores fenotípicos observados e esperados para os conteúdos de proteína e óleo nos ambientes testados. Os conteúdos de proteína e óleo avaliados neste estudo foram submetidos ao teste de Lilliefors ($p=0,01$) e mostraram uma distribuição contínua dos dados, seguindo o padrão esperado para uma herança quantitativa.

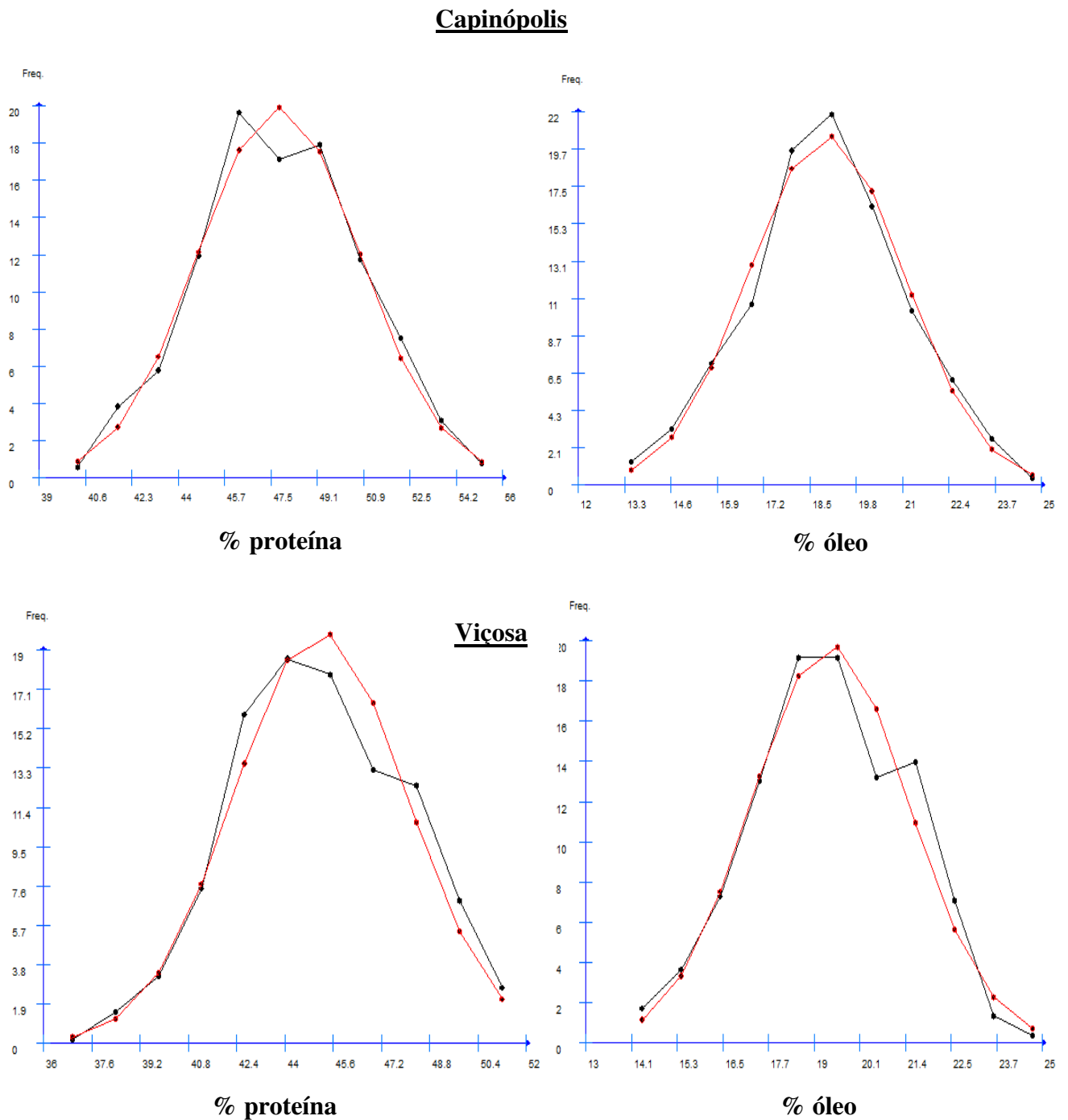


Figura 1. Distribuição normal dos conteúdos de proteína e óleo (%) para os ambientes Capinópolis e Viçosa.

Os valores baseados na média fenotípica das repetições de cada genótipo variaram de 39,366 % a 54,783 % e de 13,683 % a 24,200 % em ambientes individuais e de 41,104 % a 52,494 % e de 15,152 % a 23,595 % da análise conjunta de dois ambientes, para os percentuais de proteína e óleo, respectivamente. A população utilizada neste estudo se encontrava em elevado grau de recombinação para os loci avaliados. Desta forma, a segregação transgressiva foi constatada tanto para o aumento quanto para a redução do valor fenotípico, em especial para o sentido de maiores médias para o conteúdo de proteína no ambiente Capinópolis.

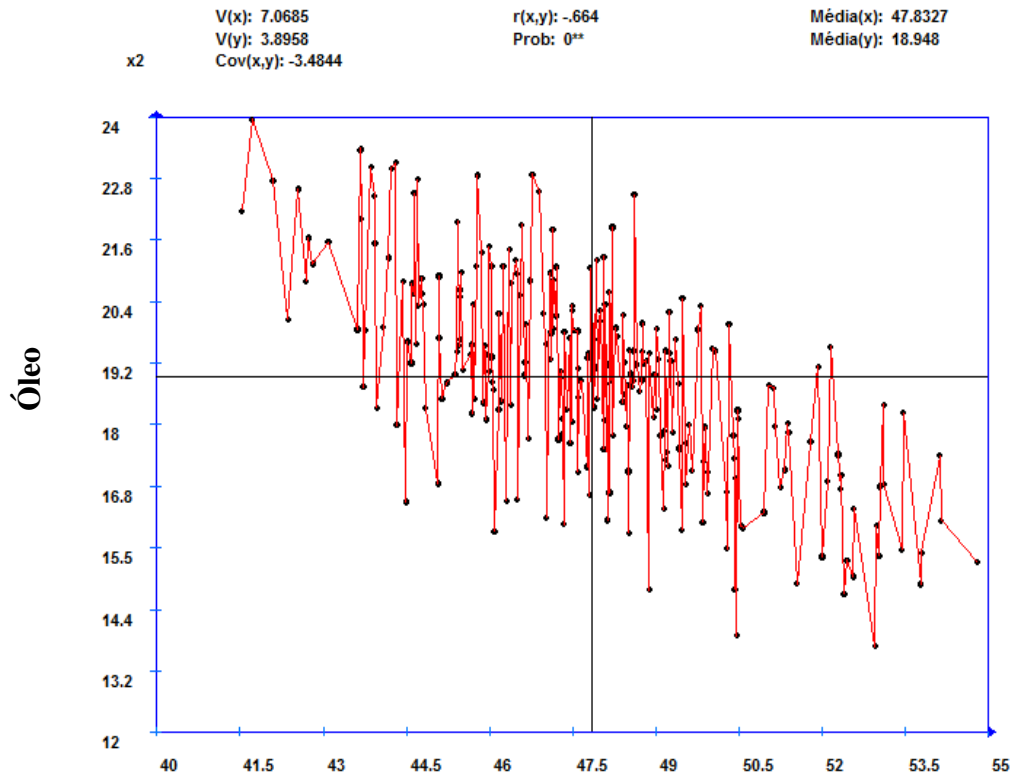
A segregação transgressiva observada por Zhang et al. (2015), com variação de 35,89 % a 49,10 % para conteúdo de proteína na população NJRSXG de *RILs* derivadas de F6, possuiu amplitude dos valores fenotípicos maior do que o do presente estudo. Pode-se atribuir tal fato aos genitores desta população que são mais divergentes para os conteúdos de proteína que os acessos PMQS80 e PMQS12. Os valores observados por Rodrigues et al. (2010) variaram de 32,2 % a 44,5 % e de 17,7 % a 22 % para os conteúdos de proteína e óleo no grão, respectivamente. Por ter adotado genitores contrastantes para conteúdo de proteína e óleo, a amplitude dos seus valores observados para conteúdo de óleo foi menor do que em comparação a este estudo. Em seu trabalho, Rodrigues et al. (2010) utilizou uma população $F_{2:3}$, em que a falta de ciclos de recombinação pode ter limitado uma maior frequência de valores mais extremos para o conteúdo de óleo na semente.

5.1.3. Correlação entre os caracteres

Os dados apresentados mostraram uma correlação negativa significativa ($p < 0,01$) entre os conteúdos de proteína e óleo (**Figura 2**), em que o coeficiente de correlação de Pearson foi de -0,664 e -0,587 para Capinópolis e Viçosa, respectivamente.

As correlações genéticas, as quais indicam a magnitude e direção de respostas à seleção simultânea (Falconer e Mackay, 1996), também foram calculadas, assumindo valores de -0,8083 e -0,7651 para Capinópolis e Viçosa, respectivamente. Este tipo de associação sugere a atuação de genes pleiotrópicos ou efeitos de ligação gênica, os quais dificultam o desenvolvimento de cultivares que combinam altos conteúdos de proteína e óleo no grão, segundo Li et al. (2017).

Capinópolis



Viçosa

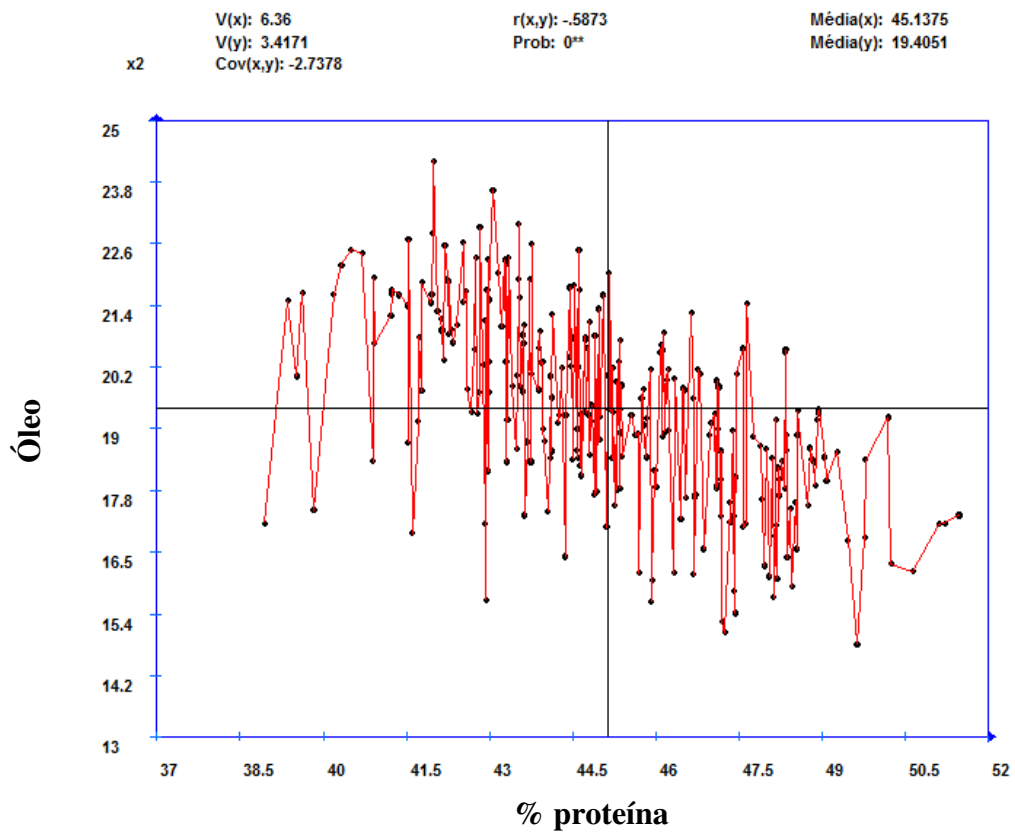


Figura 2. Correlação entre os conteúdos de proteína e óleo para os ambientes Capinópolis e Viçosa.

Para checar a consistência dos dados, um estudo de correlação entre os ambientes Capinópolis e Viçosa foi feito (**Figura 3**). Coeficientes de 0,557 e 0,643 foram significativos a $p < 0,01$, para os conteúdos de proteína e óleo, respectivamente. Tais valores indicam uma alta repetibilidade do desempenho dos caracteres (Zhang et al., 2016).

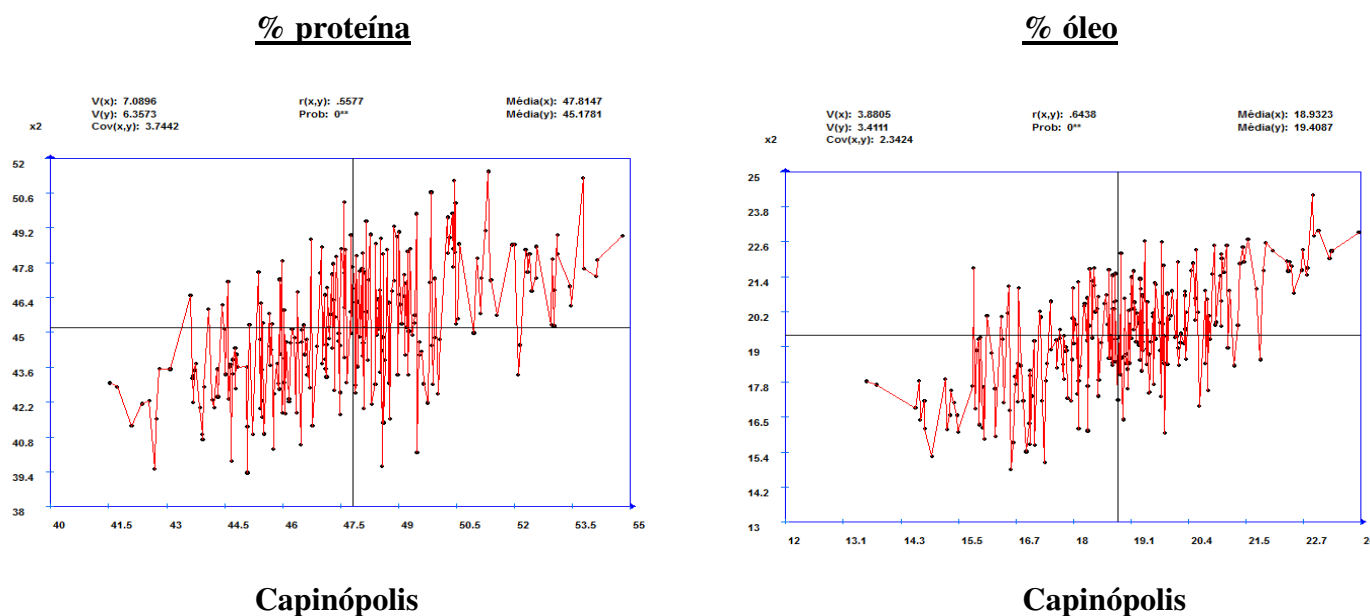


Figura 3. Correlação entre os ambientes Capinópolis e Viçosa para conteúdo de proteína e óleo (%)

Estas informações apontam que a relação inversa entre os caracteres apresentados tende a provocar um resultado indireto no sentido oposto para óleo, ao se selecionar para proteína, e vice-versa. Esta relação pode ser considerada um complicador para o processo de seleção de indivíduos e progênies.

5.1.4. Características gerais da população

O genitor PMQS80 apresentou média geral de 47% e 17,7% e PMQS12, 45,780% e 20,035%, para conteúdo de proteína e óleo, respectivamente. Tanto a média populacional quanto a dos indivíduos superiores sugere que esta é uma população superior para seleção para o caráter conteúdo de proteína no grão.

Esta mesma população foi submetida à avaliação de dias até a maturidade em casa de vegetação antes do plantio das *RILs* em campo, sendo classificadas como de ciclo precoce, médio e tardio. Os conteúdos de proteína e óleo não foram afetados em função da maturidade em nenhum ambiente de avaliação deste estudo, assim como constatado por Zhang et al. (2018), ao encontrar valores de correlação baixos e não

significativos para proteína ou óleo e dias até a maturidade ($r = 0,001$ e $r -0,013$, respectivamente).

O conteúdo de proteína foi maior nas *RILs* no ambiente Capinópolis, pois além desta população ser constituída de segregantes transgressivos que elevam a média geral e por ambiente, este caráter é favorecido por temperaturas mais elevadas (Piper e Boote, 1999).

O conteúdo de óleo foi maior dentro da média das testemunhas em Viçosa. Justifica-se tal fato por esta população não ser oriunda de genitores com alta média para o conteúdo de óleo, e por este caráter ser acrescido quadraticamente até 28°C, sendo ligeiramente decrescido em temperaturas acima desta (Piper e Boote, 1999); ou pelo acréscimo de óleo na semente sob altas temperaturas ocorrer em uma menor taxa de progresso que o de proteína (Wolf et al., 1982). Também, as testemunhas, que são cultivares modernas, foram melhoradas em processos que priorizaram o aumento da produtividade e que perderam alelos favoráveis para acúmulo de proteína, uma vez que a correlação genética entre proteína e produtividade foi de -0,58 no trabalho de Kwon e Torrie (1964). Da mesma forma, Yesudas et al., (2013) reportaram correlação positiva entre produtividade e conteúdo de óleo, e correlação negativa entre produtividade e conteúdo de proteína. A **Figura 4** ilustra o desempenho dos genitores, população e testemunhas para as características avaliadas nos dois ambientes.

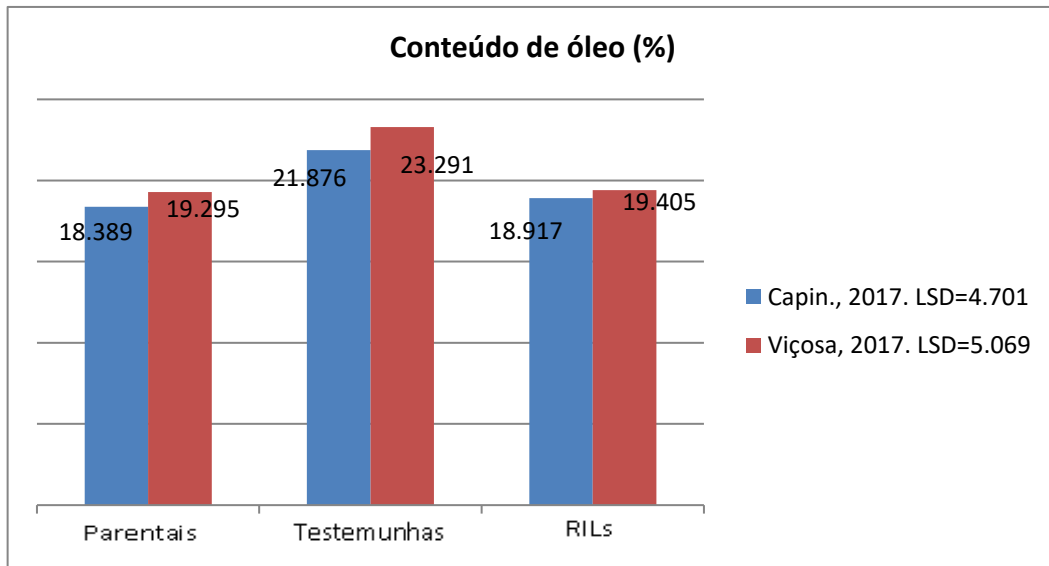
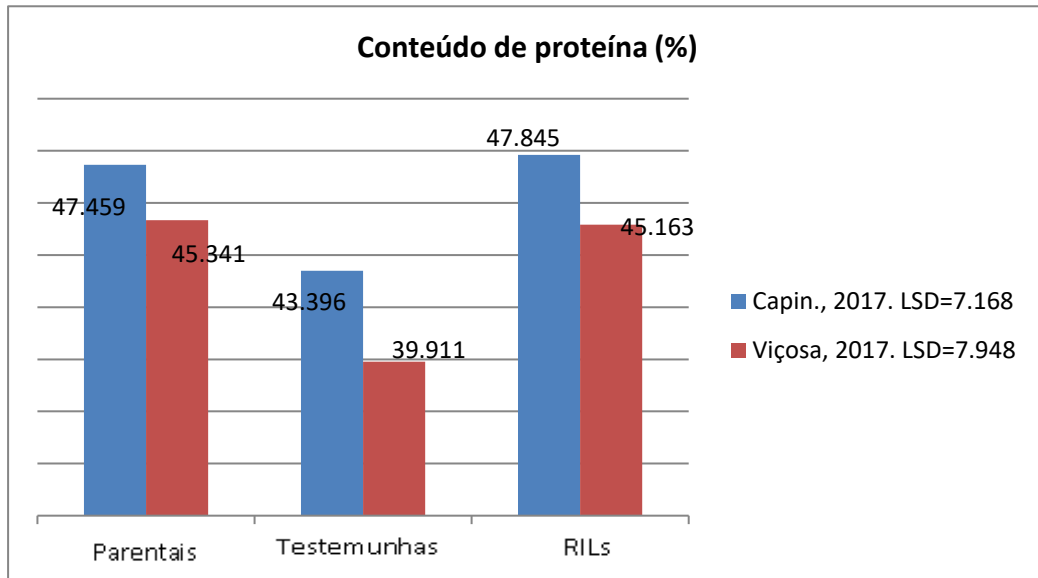


Figura 4. Conteúdo de proteína e óleo dos parentais, testemunhas e *RILs* nos ambientes Capinópolis e Viçosa.

5.2. Associação de marcadores moleculares às características avaliadas

5.2.1. Marcadores associados

Dentre os seis marcadores associados ao conteúdo de proteína no estudo de GWAS e polimórficos entre os genitores selecionados para este estudo (**Tabela 2**), quatro apresentaram associação significativa para o conteúdo de proteína em Capinópolis, sendo eles ss56, ss62, ss115, ss190; e três em Viçosa, sendo eles ss56, ss62 e ss190. A variação total do conteúdo de proteína foi explicada individualmente por estes marcadores em uma faixa de 2.096% a 29.203%. (**Tabela 4**). A estabilidade dos valores de r^2 e efeito aditivo ao longo dos ambientes indicam a relevância de uso destes marcadores (Zhang et al., 2018).

Para o conteúdo de óleo, os SNPs ss62, ss115 e ss190 associaram em ambas as localidades de ensaio, e explicaram a variação total dos dados numa faixa entre 4.845% e 29.394%.

Observou-se que dentre os marcadores associados, três se mostraram comuns tanto para conteúdo de proteína quanto para conteúdo de óleo na semente, indicando um efeito pleiotrópico dos genes próximo a esta marca, ou que seus *QTLs* estão fortemente ligados (Wang et al., 2014). A primeira hipótese é mais provável, uma vez que alto conteúdo de proteína e baixo conteúdo de óleo co-segregam, e que a tentativa de separar alto conteúdo de proteína de baixo conteúdo de óleo é complexa (Hwang et al., 2014). Em conjunto, estes marcadores explicaram em torno de 40% da variação de proteína ou óleo nas duas localidades. Este fenômeno resulta em uma correlação fortemente negativa entre estas duas características, segundo Patil et al. (2018).

O SNP ss190 foi a que mostrou maiores valores de associação para ambas as características nas duas localidades de ensaio envolvidas. Este SNP localizada no cromossomo 20 contribuiu cerca de três vezes mais que o somatório dos demais marcadores associados para o conteúdo de proteína, e o dobro da contribuição do marcador ss115 para conteúdo de óleo. Warrington et al. (2015) também verificaram que o *QTL* qProt_Gm20, também presente no cromossomo 20, contribuiu cerca de cinco vezes mais que o próximo *QTL* mais associado ao conteúdo de proteína, qProt_Gm15.

O SNP ss115 não se mostrou consistente para as localidades. Patil et al. (2018) justificaram fato similar em seus estudos pela natureza complexa de uma característica quantitativa, em que sua interação com o ambiente implica em instabilidade de *QTLs* de composição de semente de soja.

Apesar desta pesquisa envolver uma baixa quantidade de marcadores associados, pode-se observar a importância da complementaridade de genitores para resultarem em populações superiores para proteína. Os alelos favoráveis dos SNPs ss62, ss115 e ss190 foram doados pelo acesso PMQS80, enquanto o alelo favorável atribuído à SNP ss56 veio de PMQS12. Já no estudo de Warrington et al (2015), Danbaekong foi o genitor que contribuiu com todos os alelos favoráveis. De acordo com a **Tabela 4**, estes mesmos genitores não são complementares para o conteúdo de óleo.

Nosso trabalho sugere que uma atenção especial à *SNP* ss190 deve ser atribuída pelos seus resultados consistentes de alto efeito aditivo e associação aos caracteres avaliados, e que o efeito estável dos loci entre ambientes justificam a aplicabilidade dos marcadores em nosso programa de melhoramento da qualidade da soja.

Apenas uma associação influenciou o conteúdo de proteína, sem que este efeito fosse observado no conteúdo de óleo. O SNP ss56 contribuiu com valores absolutos em torno de 0.8% para o conteúdo de proteína. Tal associação sugere que seu uso na seleção assistida promove o aumento do conteúdo de proteína, manutenção do conteúdo de óleo e provavelmente um decréscimo do conteúdo de carboidratos. Assim, para programas que focam em aumento simultâneo dos conteúdos de proteína e óleo, esta é um SNP recomendável para uso. No entanto, as demais *SNPs* associadas neste trabalho são viáveis para pesquisas cujo foco é apenas o aumento do conteúdo de proteína ou óleo. Li et al. (2017) concluíram de forma similar, ao não recomendar a seleção de *QTNs* (*quantitative trait nucleotides*) co-localizados para programas que objetivam aumento simultâneo das duas características.

Há a possibilidade de se aumentar conteúdos de óleo para 20-22% e proteína para 41-43% de forma simultânea. Incremento no conteúdo de uma destas características acima desta faixa resultará em dedução dos valores médios da outra (Zhang et al., 2018), em razão da seleção de alelos de genes pleiotrópicos. No entanto, o foco dos programas de melhoramento pode variar entre o aumento do conteúdo de proteína, ou óleo, ou ambos, ao passo que exista mercado que remunere a produção de cultivares com estas características específicas, como as indústrias processamento.

Tabela 4. SNPs associadas aos conteúdos de proteína e óleo, identifiadas através de mapeamento por marca simples em RILs.

Característica	Ambiente	SNP id	Alelo ^a	Crom/GL ^b	R ² % ^d	p-valor	a ^e	% Variação ^f	Fonte de alelos favoráveis
Proteína	Capinópolis	ss56	G:A	18/G	2,096	0,0268	-0,768	41,689	PMQS12
		ss62	G:A	20/I	5,774	0,0003	1,301		PMQS80
		ss115	A:G	15/E	4,616	0,0009	1,164		PMQS80
		ss190	A:G	20/I	29,203	0	2,924		PMQS80
	Viçosa	ss56	G:A	18/G	2,170	0,0213	-0,802	33,928	PMQS12
		ss62	G:A	20/I	6,074	0,0001	1,250		PMQS80
ss190		A:G	20/I	25,684	0	2,579	PMQS80		
Óleo	Capinópolis	ss62	A:G	20/I	5,815	0,0003	-0,973	43,291	PMQS12
		ss115	G:A	15/E	12,383	0	-1,413		PMQS12
		ss190	G:A	20/I	25,093	0	-2,016		PMQS12
	Viçosa	ss62	A:G	20/I	4,845	0,0007	-0,818	40,946	PMQS12
		ss115	G:A	15/E	6,707	0	-0,966		PMQS12
		ss190	G:A	20/I	29,394	0	-2,016		PMQS12

SNPs: Single Nucleotide Polimorfism.

RILs: Recombinant Inbred Lines.

A: adenina; G: guanina; T: timina; C: citosina.

^a A forma alternativa representa o alelo favorável .

^b Crom/GL: Cromossomo/Grupo de Ligação.

^c Posição em pares de bases.

^d Proporção da variância fenotípica explicada por um marcador molecular ou QTL específico.

^e Efeito aditivo do marcador molecular em homozigose para os caracteres estudados. Valor positivo indica que o alelo favorável para a característica é herdado de PMQS80.

^f Proporção da variância fenotípica explicada por um conjunto de marcadores moleculares.

5.2.2. Combinação de loci favoráveis

Com intuito de avaliar o somatório da contribuição dos marcadores e seu efeito no conteúdo de proteína, os loci associados com o tal característica em ambas as localidades foram agrupados em todas as combinações possíveis. Desta forma, é possível verificar se o somatório da contribuição de alelos favoráveis sob determinadas combinações influenciou significativamente no conteúdo de proteína (**Tabela 5**).

A média dos indivíduos que reuniram todos os alelos favoráveis foi maior para as duas localidades. Em Capinópolis, a combinação dos quatro alelos obteve média estatisticamente igual àquela da combinação entre os alelos favoráveis de ss56, ss115 e ss190. Para amenizar a resposta sobre o conteúdo de óleo, dado o efeito pleiotropico de ss62, a opção da seleção assistida da combinação de três alelos se mostra mais vantajosa que a de quatro alelos, o que também implica na economia de genotipagem. Em Viçosa, a média baseada no alelo favorável de ss190 foi estatisticamente igual a qualquer uma combinação entre dois ou três alelos favoráveis. A opção pela seleção assistida para aumento do conteúdo de proteína baseado apenas em ss190 é desvantajoso em termos de ganhos genéticos, uma vez que ss56 e ss62, também associados à proteína nesta localidade, tiveram um somatório de efeito aditivo igual a 2,052%.

Em Capinópolis, a combinação dos alelos favoráveis dos SNPs ss56 e ss115 resulta em médias fenotípicas similares às de ss115. Já o alelo favorável desta mesmo SNP combinada aqueles de ss190 ou ss62, ou ambos simultaneamente, resulta em acréscimo da média fenotípica. O alelo favorável de ss56, combinado com os de ss190 ou ss62, ou ambos simultaneamente resultam em médias superiores para o conteúdo de proteína. Tal resultado sugere que os SNPs ss56, ss115, apesar de estarem em cromossomos diferentes, podem ter funções similares em uma mesma via metabólica. Este mesmo raciocínio não foi elaborado para o ambiente Viçosa, uma vez que ss115 não associou significativamente para o conteúdo de proteína nesta localidade. No entanto, a combinação dos três alelos favoráveis dos SNPs associadas à característica nesta localidade se mostra vantajosa para o aumento da média do caráter.

Portanto, a aplicabilidade do uso de vários marcadores para a seleção assistida do conteúdo de proteína em soja deve ser estudada previamente para a tomada de decisão sobre o uso ou não daqueles que apresentam resultados redundantes.

Tabela 5. Comparação do conteúdo de proteína em grãos de soja em função do número e combinação de loci favoráveis para a característica

Ambiente	Número de alelos favoráveis	SNP	% população	% proteína	R ^{2a}
Capinópolis	0	-	6,67	42,843 c	-
	1	ss56	44,31	48,229 b	2,096
		ss62	52,55	48,417 ab	5,774
		ss115	43,53	48,416 ab	4,616
		ss190	47,45	49,191 ab	29,203
	2	ss56, ss62	24,71	48,900 ab	7,870
		ss56, ss115	21,18	48,433 ab	6,712
		ss56, ss190	21,57	49,977 ab	31,299
		ss62, ss115	18,82	49,286 ab	10,390
		ss62, ss190	36,86	49,378 ab	34,977
		ss115, ss190	20,39	50,077 ab	33,819
	3	ss56, ss62, ss115	12,55	49,164 ab	12,486
		ss56, ss62, ss190	17,25	50,183 ab	37,073
		ss56, ss115, ss190	9,02	50,775 a	35,915
		ss62, ss115, ss190	16,47	50,399 ab	39,593
	4	ss56, ss62, ss115, ss190	8,63	50,813 a	41,689
Viçosa	0	-	13,58	43,531 b	-
	1	ss56	43,02	45,524 ab	2,170
		ss62	52,08	45,719 ab	6,074
		ss190	46,79	46,298 a	25,684
	2	ss56, ss62	24,15	46,183 a	8,244
		ss56, ss190	20,75	46,196 a	27,854
		ss62, ss190	36,60	46,527 a	31,758
	3	ss56, ss62, ss190	16,98	47,205 a	33,928

Proporção da variância fenotípica explicada por um marcador molecular ou QTL específico.

As médias seguidas pela mesma letra não diferem entre si pelo teste de Duncan à $p < 0,05$.

5.2.3. Segregação dos loci associados em progênies superiores e inferiores

Os valores fenotípicos (**Tabela 6**) e a estrutura genética de cinco segregantes transgressivos para alta e baixa proteína obtidos nos dois ambientes foi representado na forma alélica para os SNPs ss56, ss62, ss115, ss190 (**Tabela 7**).

Dentre as dez progênies superiores selecionadas no total, sete apresentaram pelo menos três alelos favoráveis. Adicionalmente todas continham o alelo favorável do SNP ss190. Dentre as inferiores, oito apresentaram até dois alelos favoráveis, e duas apresentaram três. Apenas dois genótipos apresentaram o alelo favorável do SNP ss190, os quais não correspondem àqueles que reúnem três alelos favoráveis. Desta forma, nota-se novamente a importância de ss190 para a característica. Nesta população, sua contribuição acresceu o conteúdo de proteína em 2.924% (Capinópolis) e 2.579% (Viçosa).

No entanto, nem todos os genótipos superiores que possuíram os quatro alelos favoráveis foram aqueles com maiores valores para o conteúdo de proteína. Da mesma forma, tal fato pode ser verificado para aqueles genótipos com menores conteúdos de proteína em relação ao número de alelos desfavoráveis. Zhang et al. (2015) também obtiveram progênies de alta performance para conteúdo de proteína com uma quantidade de alelos favoráveis menores quando comparadas às outras progênies da população. Justifica-se tal fato pela interação dos *QTLs* envolvidos com o ambiente, bem como à atuação de outros *QTLs* com o caráter, incluindo os de menor efeito.

Dentre 5 progênies superiores de cada ambiente, apenas uma foi verificada em ambos os grupos (81-20), apesar de todas apresentarem elevados valores de conteúdo de proteína em ambos (**Tabela 6**), sugerindo uma interação genótipos por ambientes e *QTLs* por ambientes significativos para esta característica.

Tabela 6. Desempenho dos 5 segregantes transgressivos para conteúdo de proteína.

<i>RILs</i>	Conteúdo de proteína (%)	
	Capinópolis ^a	Viçosa ^a
61-18	54.117	47.238
84-13	53.789	47.568
81-20	53.777	51.211
83-05	53.462	46.077
61-09	53.436	46.854
81-20	53.777	51.211
78-38	50.435	51.103
78-43	49.843	50.633
64-02	47.591	50.238
84-33	50.472	50.189

A fonte em negrito indica as 5 *RILs* com maior conteúdo de proteína para o respectivo ambiente.

RILs: Recombinant Inbred Lines.

^aLocalidades de ensaio

Algumas dessas progênies transgressivas superiores são complementares para alguns loci. Isto sugere que uma maior segregação transgressiva para o conteúdo de proteína pode ser obtida em populações oriundas de um possível cruzamento entre estes materiais. Este processo foi feito por Zhang et al. (2015) onde as progênies superiores e complementares do primeiro ciclo de seleção foram cruzadas e deram origem aos demais transgressivos para alta proteína em um segundo ciclo de seleção, em que todo o processo foi amparado pela seleção assistida.

Tabela 7. Segregação alélica dos 4 loci selecionados em 5 RILs com valores extremos para conteúdo de proteína nos ambientes de ensaio

Ambiente	RIL	% proteína	SNP e alelo favorável ^a			
			ss56 / G:A	ss62 / G:A	ss115 / A:G	ss190 / A:G
Capinópolis/2017	61-18	54,117 a	G	A	G	G
	84-13	53,788 a	A	A	G	G
	81-20	53,777 a	G	A	G	G
	83-05	53,461 a	G	A	A	G
	61-09	53,436 a	G	A	G	G
	78-48	42,089 b	A	A	G	A
	64-08	42,358 b	A	A	G	A
	78-36	42,551 b	G	G	A	A
	63-08	42,692 b	G	A	A	G
	83-10	42,727 b	G	A	A	A
Viçosa/2017	81-20	51,211 a	G	A	G	G
	78-38	51,102 a	A	G	A	G
	78-43	50,633 a	A	A	G	G
	64-02	50,237 a	G	A	A	G
	84-33	50,189 a	A	A	G	G
	63-08	39,516 b	G	A	A	G
	78-30	39,625 b	A	G	G	A
	84-19	39,827 b	G	G	G	A
	64-16	40,190 b	G	G	A	A
62-12	40,326 b	G	G	A	A	

^a A forma alternativa representa o alelo favorável.

A fonte em negrito indica as RILs com maior conteúdo de proteína.

As médias seguidas pela mesma letra não diferem entre si pelo teste de Duncan à $p < 0,05$.

A: adenina; G: guanina; T: timina; C: citosina.

5.2.4. Contribuição da região cromossômica de ss190

A importância do cromossomo 20 da soja mais uma vez é evidenciada nesta pesquisa, a exemplo do estudo de Patil et al. (2018) que destacou haplótipos em regiões entre 28.5 e 33.5 Mpb do cromossomo 20 como um *hotspot* para o conteúdo de proteína.

Este intervalo sobrepõe haplótipos (**Figura 7**) identificadas nos estudos de mapeamento de *QTL* e *GWAS* de Bandillo et al. (2015), Hwang et al. (2014); Vaughn et al. (2014), Bolon et al. (2010) e Warrington et al. (2015), e contém 4100 SNPs, em que 19 são associadas ao *QTL* qPro_20. A este *QTL*, são atribuídas mutações deletérias em genes envolvidos no acúmulo e armazenamento de proteína (Patil et al., 2018). Este *QTL* possui 51.41Kb e está localizado em um haplótipo de aproximadamente 1.5 Mpb, posicionado cerca de 600 Kpb à frente do SNP ss190.

Com a finalidade de estimar de forma mais precisa a contribuição da região cromossômica de ss190 e mesmo a proximidade de determinadas marcas ao provável gene envolvido no conteúdo de proteína, foi estabelecido um intervalo de aproximadamente 765 Kpb flanqueado por ss715637262 e ss715637294, e que engloba a recombinação verificada na população utilizada nesta pesquisa para o SNP ss190. Este intervalo contém vários genes que codificam para acúmulo de proteína e óleo no grão, os *QTLs* *Seed protein 4-g8* e *Seed oil 5-g8*, e aproximadamente 30 marcadores *SSR* (**Figura 5**). Da mesma forma, 765 Kpb se mostra um intervalo de dimensão razoável para uma melhor compreensão deste segmento cromossômico, uma vez que Bolon et al. (2010) delimitaram 8.4 Mpb, Hwang et al. (2014) 2.1 Mpb, Vaughn et al. (2014) 900 Kpb e Bandillo et al. (2015) 550 Kpb. Vale ressaltar que todas essas regiões contêm pelo menos um SNP associada significativamente para o conteúdo de proteína no grão e que a o intervalo flanqueado por ss715637262 e ss715637294 se encontra em proximidade com o de Bandillo et al. (2015) e sobreposta pelo bloco gênico definido por Vaughn et al. (2014), o qual possui 8 dentre 10 *SNPs* associadas no estudo de Bandillo et al. (2015).

Estes marcadores *SSR* são úteis para a execução do mapeamento fino da região. Desta forma, dentre 24 marcadores *SSR* selecionados neste intervalo, 14 se mostraram polimórficos entre PMQS80 e PMQS12 (**Figura 6**) e serão aplicados em estudos futuros do programa de melhoramento para aprofundar estudos na região envolvida.

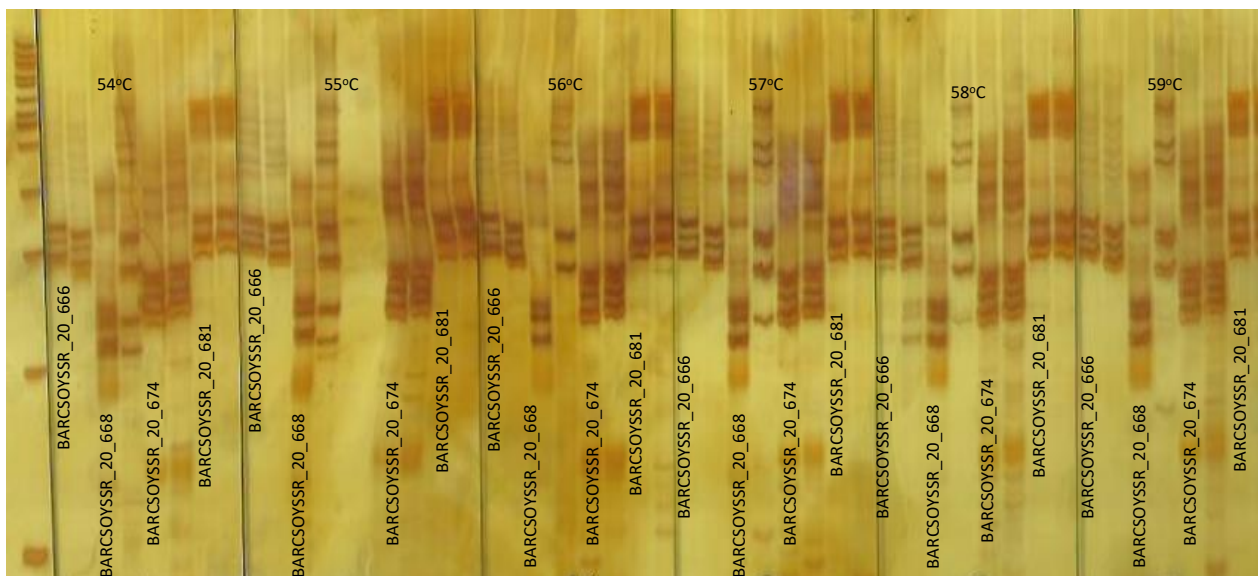


Figura 6. Ajuste das condições de reação de PCR e teste de polimorfismo entre PMQS80 e PMQS12 para quatro marcadores *SSR* flanqueados por *ss715637262* e *ss715637294*. O primeiro e segundo poços representam PMQS80 e PMQS12, respectivamente, para cada marcador *SSR*.

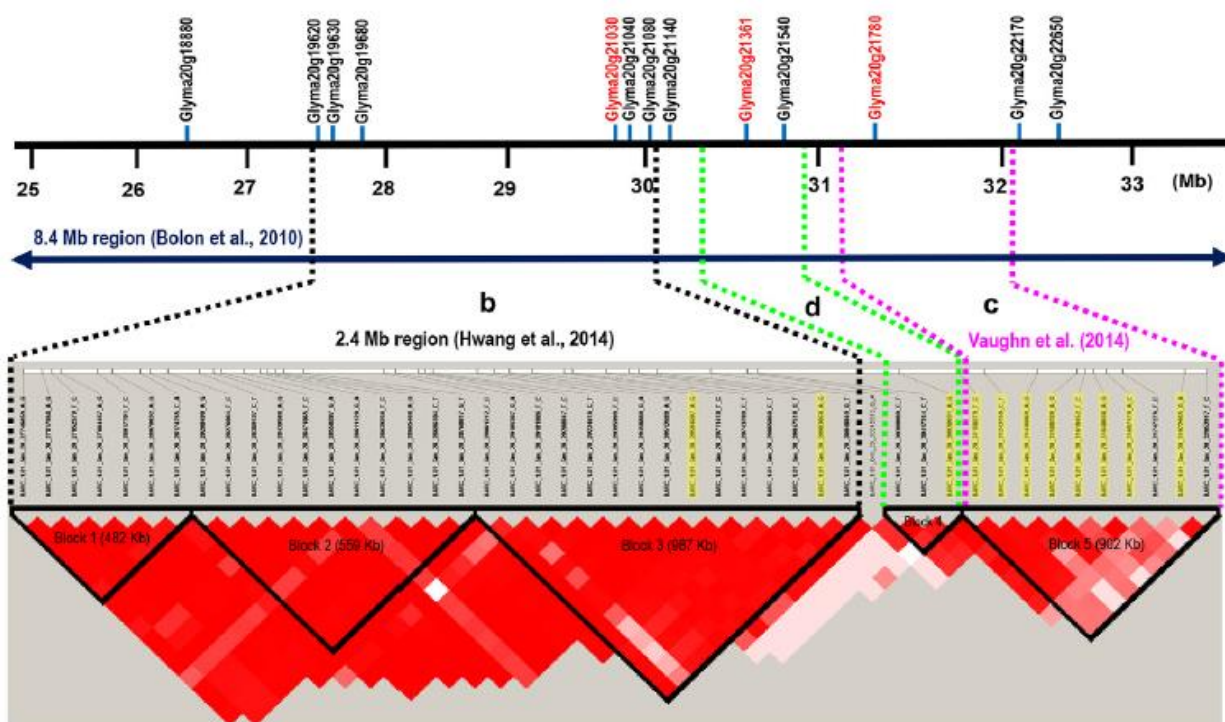


Figura 7. Redes de haplótipos e região do cromossomo 20 que contém *QTLs* de maior efeito para conteúdo de proteína e óleo em soja, segundo Bandillo et al., 2015.

5.3. Seleção de progênies

Foi feita a avaliação fenotípica visual para identificar o mérito agrônômico das 271 progênies envolvidas no estudo para os dois ambientes, que junto com os valores de fenotipagem de proteína e genotipagem dos marcadores associados, determinaram quais linhagens foram avançadas para ensaios de produtividade ou doadoras de alelos superiores para alta proteína. Foram atribuídas às características acamamento, resistência à doenças, maturidade, ramificação e quantidade total de vagens notas 1, 3, 5, 7 e 9, em que a nota 1 corresponde à classe superior para cada característica, e a 9 à inferior.

Pode-se verificar que as progênies 61-9, 61-13, 61-17, 61-28, 61-29, 62-1, 62-12, 63-29, 64-5, 64-17, 64-32, 78-6, 78-7, 78-27, 78-33, 78-38, 78-45, 78-50, 79-10, 79-20, 79-24, 79-34, 83-5, 83-14, 84-13, 84-16, 84-24, 84-27 e 84-33 tiveram maiores notas de acordo com a avaliação fenotípica. Adicionalmente, as progênies 61-9, 78-38, 84-13 e 84-33 também possuíam valores expressivos para o conteúdo de proteína, e quais reúnem de dois a quatro alelos favoráveis para o caráter, entre os quatro associados neste estudo, conforme a **Tabela 6**. Estes, então, são genótipos de uso potencial no programa de melhoramento.

6. Conclusão

Este trabalho comprovou o efeito do ambiente sobre a atuação dos genes e o efeito pleiotrópico de determinados loci, que dificultam a seleção em programas de melhoramento para conteúdo de semente de soja.

A alta associação de ss190 pode contribuir para futuros estudos que executem mapeamento fino da região envolvida no controle dos caracteres ou mesmo na clonagem posicional para elucidação do efeito do gene, e o uso de ss56 na seleção assistida não resulta em dedução estatisticamente detectável nos conteúdos de óleo.

Esta pesquisa foi útil tanto para a associação de loci relacionados ao conteúdo de proteína quanto à identificação de *RILs* superiores. Estes materiais poderão ser usados no programa de melhoramento como doadoras de alelos superiores em retrocruzamentos ou como genitores para obtenção de populações mais transgressivas para o conteúdo de proteína.

7. Refrências bibliográficas

- BOLON, Y.T. et al. Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. **BMC Plant Biol.** v. 10, n. 41, 2010
- BUENO, R.D. et al. Quantification of anti-nutritional factors and their correlations with protein and oil in soybeans. **Anais da Academia Brasileira de Ciências**, v. 90, n. 1, p.205-217, mar. 2018.
- BANDILLO, N. et al. A Population Structure and Genome-Wide Association Analysis on the USDA Soybean Germplasm Collection. **The Plant Genome**, v. 8, n. 3, p.1-13, 2015.
- BRUMMER, E.C. et al. Mapping QTL for seed protein and oil content in eight soybean populations. **Crop. Sci.**, v.37, p.370–378, 1997.
- CASANADI, G. et al. Seed quality QTLs identified in a molecular map of early maturing soybean. **Theor. Appl. Genet.**, v.103, p.912–919, 2001.
- CHUNG, J. et al. The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci.* v. 43, p. 1053 –1067, 2003.
- COBER, E.R.; VOLDENG, H.D. Developing High-Protein, High-Yield Soybean Populations and Lines. **Crop Science**, v. 40, n.1, p.39-42, 2000.
- COLLARD, B.C.Y. et al. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. **Euphytica**, v.142, p.169–196, 2005.
- COLLARD, B.C.Y.; MACKILL, D. J. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. **Philosophical Transactions of The Royal Society B: Biological Sciences**, v.363, n.1491, p.557-572, 2008.
- CRUZ, C.D. GENES – a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum**, v.35, n.3, p.271-276, 2013.
- CRUZ, C.D.; REGAZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 4ed., v.1, Viçosa:UFV, 2012. 514p.
- DELLAPORTA, S.L.; WOOD, J.; HICKS, J.B. A plant DNA miniprep: Version II. **Plant Molecular Biology Reporter**, 1983, v.1, n.4, p.19-21, 1983.
- DIERS, B.W. et al. RFLP analysis of soybean seed protein and oil content. **Theor. Appl. Genet.**, v.83, p.608–612, 1992.

- DIERS, B.W. et al. Registration of LDX01-1-65 soybean germplasm with soybean cyst nematode resistance derived from *Glycine soja*. **Crop. Sci.**, v.45, p.1671–1672, 2005.
- DUDLEY, J.W. Theory for transfer of alleles. **Crop Sci.**, v. 22, p.631–637, 1982.
- ELLUR, R.K. et al. Improvement of Basmati rice varieties for resistance to blast and bacterial blight diseases using marker assisted backcross breeding. **Plant Science**, v. 242, p.330-341, 2016.
- FALCONER, D.S. et al. **Introduction to quantitative genetics**. 4 ed. Trends Genet., 1996, 280p.
- HWANG, E. et al. A genome-wide association study of seed protein and oil content in soybean. **Bmc Genomics**, v. 15, n. 1, p.1-12, 2014.
- JUN, T. et al. Association analysis using SSR markers to find QTL for seed protein content in soybean. **Euphytica**, v.162, n.2, p.179-191, 2007.
- KIM, K.H. et al. Marker-assisted foreground and background selection of near isogenic lines for bacterial leaf pustule resistant gene in soybean. **J. Crop. Sci. Biotechnol.**, v.11, p.63–268, 2008.
- KING, Z.R. et al. Registration of four near-isogenic soybean lines of G00-3213 for resistance to Asian soybean rust. **J. Plant. Reg.**, v.10, p.189–194, 2016.
- KOEBNER, R. M. D.; SUMMERS, R. W. 21st century wheat breeding: plot selection or plate detection? **Trends Biotech.**, v.21, p.59–63. 2003.
- KWON, S.H.; TORRIE, J.H., Heritability and interrelationships among traits of two soybean populations. **Crop Science**, v.4, p.196-198, 1964.
- LANDAU-ELLIS, D.; PANTALONE V.R. Marker-assisted backcrossing to incorporate two low phytate alleles into the Tennessee soybean cultivar 5601T. Em: Shu, Q.Y. (ed) **Induced plant mutations in the genomics era**. Food and Agriculture Organization of the United Nations, Roma, p 316–318, 2009.
- LANDGRAF, L. **Soja sofre redução no conteúdo de proteína ao longo do tempo**. 2015. Disponível em: <<https://www.embrapa.br/busca-de-noticias/-/noticia/7693893/soja-sofre-reducao-no-teor-de-proteina-ao-longo-do-tempo>>. Acesso em: 24 abr. 2018.
- LEE, S.H. et al. RFLP loci associated with soybean seed protein and oil content across populations and locations. **Theor Appl Genet.**, v.93, p.649–657, 1996.
- LGC GENOMICS. **KASP genotyping chemistry User guide and manual. 2013**. Disponível em:

- <<https://www.lgcgroup.com/LGCGroup/media/PDFs/Products/Genotyping/KA-SP-genotyping-chemistry-User-guide.pdf>>. Acesso em: 24 abr. 2018.
- LI, Z. et al. Impact of Genomic Research on Soybean Breeding. **The Soybean Genome**, p.111-129, 2017.
- MANNUR, D. M. et al. Super Annigeri 1 and improved JG 74: two Fusarium wilt-resistant introgression lines developed using marker-assisted backcrossing approach in chickpea (*Cicer arietinum* L.). **Molecular Breeding**, v. 39, n. 1, p.1-13, dez. 2018.
- MAROOF, M.A.S. et al. Pyramiding of soybean mosaic virus resistance genes by marker-assisted selection. **Crop Sci.**, v.48, p.517–526, 2008.
- ORF, J.H. et al. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. **Crop Sci.**, v.39, p.1642–1651, 1999.
- PATERSON, A.H. Making genetic maps. Em: A.H. Paterson (Ed.), **Genome Mapping in Plants**, p.23–39. R. G. Landes Company, San Diego, California; Academic Press, Austin, Texas, 1996.
- PATIL, G. et al. Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. **Plant Biotechnology Journal**, v. 16, n. 11, p.1939-1953, 16 maio 2018.
- PLUME, K. **Protein plight: Brazil steals U.S. soybean share in China**. 2018. Disponível em: <<https://www.reuters.com/article/us-usa-soybeans-protein-insight/protein-plight-brazil-steals-u-s-soybean-share-in-china-idUSKBN1FE0FM>>. Acesso em: 24 abr. 2018.
- PIPER, E.L.; BOOTE, K.J. Temperature and Cultivar Effects on Soybean Seed Oil and Protein Concentrations. **Jaocs**, Gainesville, Florida, v. 76, n. 10, p.1233-1241, 25 maio 1999.
- RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Curr. Opin. Plant Biol.**, v.5, p.94–100, 2002.
- RESENDE, M.D.V. Software Selegen-REML/BLUP: a useful tool for plant breeding. **Crop Breeding and Applied Biotechnology**, v.16, n.4, p.330-339, 2016.
- RODRIGUES, J.I.S et al. Mapeamento de QTL para conteúdos de proteína e óleo em soja. **Pesquisa Agropecuária Brasileira**, Brasília, v. 45, n. 5, p.472-480, maio 2010.

- RODRIGUES, J.I.S et al. Biometric analysis of protein and oil contents of soybean genotypes in different environments. **Pesquisa Agropecuária Brasileira**, v. 49, n. 6, p.475-482, jun. 2014.
- SCHUSTER, I.; CRUZ, C. D. **Estatística Genômica**. 2. ed. Viçosa, Mg: Editora Ufv, 2008. 568 p.
- SONG, Q. et. al. Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. **PLoS ONE**, v.8, n.1, e54985, 2013.
- United States Department of Agriculture - Foreign Agricultural Service. **World Agricultural Production**, February, 2019.
- VAUGHN, J.N. et al. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. **G3 Genes Genom Genet.**, v.4, 2283–2294, 2014.
- WANG, X. et al. Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. **Molecular Genetics and Genomics**, v.289, n.5, p.935-949, 2014.
- WARRINGTON, C.V. et al. QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. **Theoretical and Applied Genetics**, v.128, n.5, p.839-850, 2015.
- WEHRMANN, V.K. et al. Transfer of high seed protein to high-yielding soybean cultivars. **Crop Sci.**, p.27, v.927-931. 1987.
- WILCOX, J.R.; J.F. CAVINS. Backcrossing high seed protein high-protein parent is improved germplasm to a soybean cultivar. **Crop Sci.**, v.35, p.1036-104, 1995.
- WINTER, P.; KAHL, G. Molecular marker technologies for plant improvement. **World Journal of Microbiology & Biotechnology**, v.11, p.438–448, 1995.
- WOLF, R.B. et al. Effect of Temperature on Soybean Seed Constituents: Oil, Protein, Moisture, Fatty Acids, Amino Acids and Sugars. **Jaocs**, Peoria, Illinois, v. 59, n. 5, p.230-232, maio 1982.
- YAN, L. et al. Identification of quantitative trait loci associated with soybean seed protein content using two populations derived from crosses between *Glycine max* and *Glycine soja*. **Plant Genetic Resources**, v.12, n.1, p.104-108, 2014.
- YESUDAS, C. R. et al. Identification of germplasm with stacked QTL underlying seed traits in an inbred soybean population from cultivars Essex and Forrest. **Molecular Breeding**, v. 31, n. 3, p.693-703, 5 jan. 2013.

- ZHANG, J. et al. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). **Theoretical And Applied Genetics**, v. 129, n. 1, p.117-130, 30 out. 2016.
- ZHANG, J. et al. Genome-wide Scan for Seed Composition Provides Insights into Soybean Quality Improvement and the Impacts of Domestication and Breeding. **Molecular Plant**, v. 11, n. 3, p.460-472, mar. 2018.
- ZHANG, Y.H. et al. Marker-assisted breeding for transgressive seed protein content in soybean [*Glycine max* (L.) Merr.]. **Theoretical and Applied Genetics**, v.128, n.6, p.1061-1072, 2015.
- ZHU, Y.L. et al. Single-Nucleotide Polymorphisms in Soybean. **Genetics**, v.163, n1123-1134, 2003.