

JÉSSICA NAYARA BASILIO SILVA

**VALIDAÇÃO DE MARCADORES *SNP* ASSOCIADOS AO CONTEÚDO DE
ÓLEO EM SOJA [*Glycine max* (L.) Merr.]**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Aplicada, 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

S586v
2019
Silva, Jéssica Nayara Basilio, 1993-
Validação de marcadores *SNP* associados ao conteúdo de
óleo em Soja [*Glycine max* (L.) Merr.] / Jéssica Nayara Basilio
Silva. – Viçosa, MG, 2019.
vii, 30f. : il. (algumas color.) ; 29 cm.

Orientador: Maximiller Dal Bianco Lamas Costa.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 24-30.

1. Melhoramento genético. 2. Marcadores moleculares.
3. Soja - Seleção. I. Universidade Federal de Viçosa.
Departamento de Bioquímica e Biologia Molecular. Programa de
Pós-Graduação em Bioquímica Aplicada. II. Título.

CDD 22. ed. 631.52

JÉSSICA NAYARA BASILIO SILVA

**VALIDAÇÃO DE MARCADORES *SNP* ASSOCIADOS AO CONTEÚDO DE
ÓLEO EM SOJA [*Glycine max* (L.) Merr.]**

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


APROVADA: 25 de julho de 2019.



José Eustáquio de Souza Carneiro



Derly José Henriques da Silva



Rafael Delmond Bueno
(Coorientador)



Maximilner Dal Bianco Lamas Costa
(Orientador)

*À minha Mãe Maria Aparecida,
Ao meu pai Antônio,
Ao meu namorado Vinícius,
Aos meus demais familiares e amigos.*

DEDICO

AGRADECIMENTOS

Agradeço a Deus, por direcionar os meus passos.

Agradeço aos meus pais, Antônio e Maria Aparecida e ao meu irmão Jeferson, por todo o apoio e suporte que sempre me ofereceram.

Ao meu amigo e namorado Vinícius, por uma relação baseada no respeito e admiração.

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

Ao Laboratório de Bioquímica Genética de Plantas (LBGP – BIOAGRO-UFV) e ao Programa de Melhoramento de Qualidade de Soja (BIOAGRO/DBB/UFV), por todo o aprendizado.

Ao Laboratório de Genética Molecular de Plantas (BIOAGRO-UFV) e ao Prof. Cléberon Ribeiro, 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, João Victor, Bruna e Thaís pelo apoio e amizade de uma vida inteira.

Aos funcionários José Carlos e Reginaldo e aos estudantes Arthur, Yan, Rafael, Bruno, Lenon, Walter, Lucas e Guilherme do Laboratório de Bioquímica Genética de Plantas. Em especial, a Marina e Rodrigo, pelo auxílio na rotina de laboratório e pela amizade, confiança e companheirismo em todos os momentos.

Ao doutorando Arthur Bernardeli e o pesquisador Dr. Newton Deniz Piovesan, pelo acompanhamento nos experimentos.

Ao coorientador Dr. Luíz Cláudio Costa Silva, pelo auxílio em análises e conselhos.

Ao coorientador Dr. Rafael Delmond Bueno, pelos conselhos e aprendizados.

Ao meu orientador Prof. Maximiller Dal-Bianco Lamas Costa, pela forma carinhosa e atenciosa com que me recebeu em sua equipe de trabalho, pelo respeito, pela paciência infinita e pela confiança que sempre teve no meu trabalho.

Ao Prof. Derly José Henriques da Silva e Prof. José Eustáquio de Souza Carneiro, meus agradecimentos pela contribuição e disponibilidade de avaliar meu trabalho.

Muito Obrigada!

BIOGRAFIA

JÉSSICA NAYARA BASILIO SILVA, filha de Aparecido Antônio da Silva e Maria Aparecida Basilio da Silva, nasceu em 20 de agosto de 1993, em Bocaiúva, Minas Gerais.

Concluiu o ensino médio no Colégio Renovar, em Bocaiúva no ano de 2010.

Em fevereiro de 2011, ingressou no curso de Ciências Biológicas na Universidade Federal de Viçosa, Campus Rio Paranaíba. Em 2013, transferiu-se para a Universidade Estadual de Montes Claros, onde em maio de 2017, formou-se em Ciências Biológicas - Bacharelado.

Em agosto de 2017, iniciou o curso de Mestrado em Bioquímica Aplicada pela Universidade Federal de Viçosa, submetendo-se à defesa da dissertação em julho de 2019.

RESUMO

SILVA, Jéssica Nayara Basilio, M.Sc., Universidade Federal de Viçosa, julho de 2019. **Validação de marcadores SNP associados ao conteúdo de óleo em Soja [*Glycine max* (L.) Merr.]** Orientador: Maximiller Dal-Bianco Lamas Costa. Coorientadores: Rafael Delmond Bueno e Luíz Cláudio Costa Silva.

Os objetivos deste estudo foram identificar e estimar o efeito de marcadores moleculares *SNP* sobre o conteúdo de óleo em soja, estimar a herdabilidade e a correlação entre os conteúdos de proteína e óleo, e identificar genes na região do marcador associado à característica de interesse. Para isso, 476 *RILs* derivadas de F4 foram avaliadas em ensaios de campo em quatro diferentes ambientes, sendo Viçosa/2017, Viçosa/2018, Capinópolis/2018 e Viçosa/2019. A população F4 foi genotipada pela metodologia KASP e fenotipada para conteúdo de proteína e óleo por espectrometria de infravermelho próximo (NIR). Além disso, foi realizada uma análise de RNASeq com os genitores para gerar uma lista de genes diferencialmente expressos (DEGs) presentes na região alvo deste estudo. Observou-se efeito significativo de genótipos para conteúdo de óleo e proteína nos dois ambientes avaliados. O conteúdo de proteína foi maior nas *RILs* no ambiente de Viçosa/2018 (48,50%) e menor em Capinópolis/2018 (42,55%). Já o conteúdo de óleo foi maior nas *RILs* no ambiente de Capinópolis/2018 (23,72%) e menor em Viçosa/2018 (15,23%). A herdabilidade assumiu valores de 71,02% e 71,75% para o conteúdo de óleo e 73,51% e 80,12% para o conteúdo de proteína, para Viçosa/2018 e Capinópolis/2018, respectivamente. Foi encontrada uma forte correlação negativa significativa ($p < 0,01$) entre os teores de proteína e óleo, com um coeficiente de correlação de Pearson variando de -0,483 a -0,730. Dos marcadores moleculares avaliados, ss61, ss73, ss134 e ss185 associaram para o conteúdo de proteína em um dos ambientes. O marcador ss115 associou tanto para o conteúdo de óleo, quanto para proteína em todas as localidades, com provável efeito pleiotrópico. O marcador teve efeito aditivo e R^2 variando de 0,4081 a 1,285 e 1,7882 a 17,3147, respectivamente. Após filtragem dos genes mapeados no cromossomo 15 da soja por estudos de RNASeq, 35 genes diferencialmente expressos ($p < 0,05$) foram analisados. A aplicação de um limiar de FDR ($q \leq 0,01$) e $FC > 1.5$ reduziu o número para quatro genes candidatos (Glyma.15g057700, Glyma.15g063000, Glyma.15g064100 e Glyma.15g068700). Os dados de nosso trabalho validaram a utilização do marcador ss115 em programa de melhoramento de soja para o conteúdo de óleo e proteína assim como identificou genes que podem estar associados com estas características.

ABSTRACT

SILVA, Jéssica Nayara Basilio, M.Sc., Universidade Federal de Viçosa, July, 2019. **SNP Validation for markers associated with oil content in soybean [*Glycine max* (L.) Merr.]**. Adviser: Maximiller Dal-Bianco Lamas Costa. Co-advisers: Rafael Delmond Bueno and Luíz Cláudio Costa Silva.

The main goal of the present study was to identify and estimate the effect of *SNP* molecular markers on soybean oil content, determine heritability, correlation of protein and oil contents and identify genes around the marker region associated with the characteristic of interest. 476 F4-derived *RILs* were submitted to field trials in four different environments, Viçosa/2017, Viçosa/2018, Capinópolis/2018, and Viçosa/2019. The F4 population was genotyped by the KASP methodology and phenotyped for protein and oil content using Near Infrared Spectrometry (NIR). Besides, it was conducted an RNASeq of the parentals to generate a list of differentially expressed genes (DEGs). The genotypes presented significant differences among them and along with the environments. The mean of protein content was higher in *RILs* of Viçosa/2018 (48,49%) and lower in Capinópolis/2018 (42,54%). The oil content was higher in Capinópolis/2018 (23,72%) and lower in Viçosa/2018 (15,23%). The heritability assumed values of 71,02% and 71,75% for the oil content and 73,51% and 80,12% for the protein content, for Viçosa/2018 and Capinópolis/2018, respectively. A strong negative correlation ($p < 0,01$) between protein and oil contents was found, with a Pearson correlation coefficient ranging from -0,483 to -0,730. The ss61, ss73, ss134 and ss185 molecular markers were associated with the protein content in just one tested environment. The marker ss115 was associated with both oil and protein content in all localities with a probable pleiotropic effect. The marker had an additive effect and R^2 ranging from 0.4081 to 1.285 and 1.7882 to 17.3147, respectively. After filtering the genes mapped in the soybean chromosome 15 using RNASeq studies, 35 differentially expressed genes ($p < 0,05$) were analyzed. The application of an FDR threshold ($q \leq 0,01$) and $FC > 1.5$ reduced the number of genes to four (Glyma.15g057700, Glyma.15g063000, Glyma.15g064100, and Glyma.15g06870). Our study validated the use of ss115 marker on a breeding program with a focus in development varieties for oil and protein content as well as identifying genes that are probably associated with these characteristics.

SUMÁRIO

1. Introdução.....	1
2. Material e Métodos	3
2.1 Material vegetal.....	3
2.2 Ensaios: locais, épocas e tratos culturais	4
2.3 Delineamento experimental.....	4
2.4 Genotipagem	5
2.4.1 Extração de DNA.....	5
2.4.2 Seleção de marcadores	5
2.4.3 Metodologia de Genotipagem.....	6
2.5 Fenotipagem	6
2.6 Análises genético-estatísticas	6
2.6.1 Estudo de associação marcador-característica.....	6
2.6.2 Estimativa da herdabilidade	8
2.6.3 Correlação entre conteúdo de proteína e óleo.....	8
2.6.4 Análises Estatísticas	9
2.7 Mapeamento dos <i>reads</i> e Análise de expressão diferencial	9
3. Resultados.....	9
3.1 Parâmetros genéticos e fenotípicos.....	9
3.2 Características gerais da população	10
3.3 Associação de marcadores moleculares às características avaliadas	12
3.4 Genes diferencialmente expressos no cromossomo 15	15
4. Discussão.....	19
4.1 Parâmetros genéticos e fenotípicos.....	19
4.2 Associação de marcadores moleculares às características avaliadas.....	20
4.3 Genes diferencialmente expressos no cromossomo 15	21
5. Conclusão	23
6. Referências	24

1. Introdução

A soja (*Glycine max* (L.) Merrill) é a mais importante leguminosa cultivada no mundo devido ao seu alto conteúdo de proteína (40%) e óleo (20%), além da alta produtividade e baixo custo de produção (Clemente e Cahoon, 2009). No Brasil, o complexo soja tem destaque expressivo na pauta de exportações sob a forma de farelo, óleo e grãos, e também para o suprimento do mercado interno de óleos comestíveis, alimentação animal, demais derivados proteicos e produção de Biodiesel.

A cultura da soja é responsável por cerca de metade da produção de grãos do Brasil, o que corresponde a um terço de toda a produção mundial (Conab, 2019). Os Estados Unidos da América (EUA) permanecem como principal produtor, seguido por Brasil e Argentina. Entretanto, a produção e exportação do Brasil cresce rapidamente nos últimos anos e a perspectiva é de que esse se torne o principal produtor, devido aos aumentos de produtividade e a expansão da sojicultura para novas áreas (Soystats, 2019; Zafalon, 2019).

Um importante produto que advém dessa cultura é o óleo de soja, usado para consumo humano, bem como matéria-prima renovável para vários produtos industriais. A composição desse produto pode variar em função do genótipo e do ambiente de cultivo, sendo formado em média por 13% de ácido palmítico (16:0), 4% de ácido esteárico (18:0), 20% de ácido oleico (18:1), 55% de ácido linoleico (18:2) e 8% de ácido linolênico (18:3) (Pham et al. 2010; Goettel et al. 2014). Recentemente, tanto o conteúdo quanto a qualidade do óleo chamaram a atenção dos programas de melhoramento da soja, devido ao aumento da demanda por óleos vegetais de melhor qualidade (Clemente e Cahoon, 2009; Silva et al. 2018).

Do ponto de vista econômico, o óleo de soja é o segundo óleo vegetal mais consumido no mundo, representando quase um terço do mercado mundial (Conab, 2019). No Brasil, sua importância é ainda maior, correspondendo a mais de 80% do mercado de óleos vegetais na indústria alimentícia e mais de 80% da demanda total da fabricação de biodiesel (Aprosoja, 2019).

Nos últimos anos, os programas de melhoramento da soja têm enfatizado o desenvolvimento de cultivares produtivas com maior conteúdo de proteína e óleo (Li et al. 2018; Van et al. 2017; Mahmoud et al. 2006; Tripathi e Khare, 2016). Ambos os conteúdos são determinados em parte por genes com predominância de efeitos aditivos, em que os valores de herdabilidade variam de médio a alto (Jaureguy et al. 2011), com estimativas variando de 0,57 a 0,99 para proteína e 0,51 a 0,99 para óleo (Chung et al. 2003; Panthee et al. 2005).

O conteúdo de óleo correlaciona-se negativamente com o conteúdo de proteínas, rendimento e tamanho de sementes (Chung et al. 2003; Wilcox et al. 2001; Hyten et al. 2004; Clemente e Cahoon, 2009; Bandillo et al. 2015), o que dificulta a obtenção de linhagens com alto conteúdo de óleo. Como a associação entre essas características é difícil, os programas de melhoramento estão buscando formas de aumentar uma característica, sem que ocorra diminuição em outras, e ainda levando em consideração o aumento de produtividade (Akond et al. 2013).

Diferentes trabalhos vêm tentando identificar *QTLs* (*Quantitative trait loci*) que regem o acúmulo de proteína e óleo nas sementes de soja. Com essa finalidade, muitas populações foram mapeadas para duas principais regiões genômicas, consideradas *hotspots* para soja. Estudos conduzidos por Diers et al. (1992), Sebolt et al. (2010), Chung et al. (2003) e Bandillo et al. (2015) detectaram efeitos importantes nos cromossomos 20 (Gm20) e cromossomo 15 (Gm15), ambos exibindo correlação negativa para estas duas características. Observou-se que o cromossomo 20 tem efeito mais significativo, conforme destacado por Lee et al. (2019). Neste contexto, a crescente demanda por cultivares de soja superiores criou maior interesse em explorar seus mecanismos genéticos (Zhang et al. 2018). Embora tenha havido ganhos no rendimento da soja, mecanismos genéticos moleculares que controlam o rendimento e conteúdo de proteína e óleo são ainda desconhecidos e uma maior elucidação dessas características pode ajudar a identificar estratégias para o melhoramento (Patil et al. 2017).

Considerando o que foi anteriormente mencionado, a identificação de *locus* genéticos que governem essas características facilitariam estudos nesta área, assim como no desenvolvimento de linhagens (Chung et al. 2003; Eskandari et al. 2013; Phansak et al. 2016). Neste sentido, o emprego de marcadores moleculares é de grande importância, uma vez que possibilita a utilização da variabilidade genética para identificar genes responsáveis pelas características de interesse (Chung et al. 2003; Eskandari et al. 2013). De acordo com Kim et al. (2016) um total de aproximadamente 145 associações de marcadores significativos para concentração de proteína e 179 para concentração de óleo foram catalogadas no *Soybase*. Entretanto, desse total, apenas quatro *QTL* para óleo e dois para proteína foram confirmados baseadas nas regras estabelecidas pelo Comitê de Genética da Soja. Dentre esses estudos estão os de Fasoula et al. (2004), Nichols et al. (2006), Qi et al. (2011) e Pathan et al. (2013). Da mesma forma que esses dados revelam um futuro promissor no desenvolvimento desta área, também ressaltam a carência de informação para os programas de melhoramento desenvolverem cultivares para as características mencionadas.

Entre as classes de marcadores mais usadas atualmente estão os *SNPs* (*Simple Nucleotide Polymorphism*). São conhecidos como variações de nucleotídeos ou pequenas inserções/deleções (*indels*) em sequências de DNA homólogas (Collins et al. 1998) e ganharam interesse da comunidade científica por serem a forma de polimorfismo mais abundante em genomas (Zhu et al. 2003), além de grande facilidade de automatização e variedade de tecnologias atualmente disponíveis (Fan et al. 2006). Tornaram-se uma ferramenta potencial nos programas de melhoramento e é considerado o marcador ideal para o estudo do genoma da soja (Song et al. 2013). Neste contexto, os *SNPs* permitem melhorar as chances de sucesso em uma diversidade de aplicações em programas de melhoramento de soja, incluindo identificação de genes candidatos, clonagem posicional, análise de associação, mapeamento de *QTL* e a determinação de relações genéticas entre indivíduos (Choi et al. 2007; Patil et al. 2016).

Em soja, a frequência de *SNPs* é relativamente mais baixa em relação a outras espécies agrícolas cultivadas (Zhu et al. 2003; Hyten et al. 2006). Isso se deve provavelmente à estreita base genética dos genótipos utilizados para a obtenção das primeiras cultivares, da baixa variabilidade do ancestral selvagem *Glycine soja*, além do processo de seleção ao longo dos anos (Choi et al. 2007; Hyten et al. 2006). Ainda assim, já foram encontradas mais de 200 mil *SNPs* em soja, a partir de análises de quase 500 milhões de *reads* de 35-45 pb, em comparação com o genoma de referência Williams 82 (Song et al. 2013).

No presente estudo, um total de 476 *RILs* foram desenvolvidas e avaliadas em quatro diferentes ambientes. Os dados fenotípicos foram associados com a genotipagem de seis marcadores *SNP* com o objetivo de estimar os efeitos destes marcadores para aplicação em programa de melhoramento voltado para o conteúdo de óleo em soja e identificar genes candidatos presentes na região proximal ao marcador utilizado.

2. Material e Métodos

2.1 Material vegetal

A partir de dois acessos do banco de germoplasma do Programa de Melhoramento da Qualidade da Soja (BIOAGRO/DBB/UFV) foi realizado um cruzamento dos genitores PMQS-SUPR (36,32% de proteína e 24,54% de óleo) e PMQS-08 (45,08% de proteína e 21,21% de óleo). As plantas F1 foram autofecundadas gerando 476 sementes F2, que foram conduzidas pelo método *SSD* (*Single Seed Descendent*) até F4 em casa de vegetação. Os

indivíduos F4 tiveram irrigação, fertilização e iluminação suplementar de forma com que cada planta proporcionasse quantidade elevada de sementes para serem posteriormente submetidas a ensaios de campo em anos consecutivos.

2.2 Ensaios: locais, épocas e tratos culturais

Os ensaios foram conduzidos em duas localidades, Viçosa-MG (20° 45' 14" S 42° 52' 55" O) e Capinópolis-MG (18° 40' 55" S 49° 34' 12" O). Em Viçosa, o plantio foi realizado na Unidade de Ensino, Pesquisa e Extensão (UEPE) do Departamento de Fitotecnia, em duas diferentes épocas de semeadura, sendo elas novembro/2017 (geração F4) e abril/2018 (geração F_{4.5}), correspondentes às safras de verão e inverno, respectivamente. Em janeiro/2019 foi conduzido em casa de vegetação (geração F_{4.6}). Em Capinópolis, o plantio foi realizado na Central de Experimentação, Pesquisa e Extensão do Triângulo Mineiro (CEPET), em novembro/2018 (geração F_{4.6}). Dessa forma, os genótipos foram avaliados em quatro ambientes distintos. 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 os grãos atingiram o ponto de maturação fisiológica. A colheita das plantas foi realizada manualmente e o material armazenado em casa de vegetação por uma semana. Posteriormente, foi devidamente trilhado e organizado para análises de fenotipagem.

2.3 Delineamento experimental

O delineamento adotado foi o de blocos casualizados a exemplo de Zhang et al. (2015). A dimensão das parcelas foi de 1 m de comprimento, 0,5 m de largura entre parcelas e 1 m de intervalo entre blocos subsequentes. Foram semeadas 20 sementes por metro, com posterior desbaste pós-emergência para uma população 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 da abordagem 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} .

2.4 Genotipagem

2.4.1 Extração de DNA

Folhas jovens e sadias foram coletadas em estádio V4, em janeiro de 2018 das plantas F4. A extração de DNA foi feita de acordo com o protocolo de Dellaporta et al. (1983). Cada amostra foi quantificada por meio do espectrofotômetro NanoDrop (NanoDrop Technologies, Wilmington, DE, EUA). A verificação da qualidade do DNA foi realizada através de eletroforese em gel de agarose 0,8%. Posteriormente foi realizada diluição para a concentração de trabalho de $10 \text{ ng} \cdot \mu\text{L}^{-1}$ para a genotipagem.

2.4.2 Seleção de marcadores

Trabalhos anteriores do Programa de Melhoramento da Qualidade da Soja (PMQS/BIOAGRO/DBB/UFV) selecionaram 269 *SNPs* na literatura e na plataforma online *Soybase* (<https://www.soybase.org/>). A seleção foi realizada com base em *SNPs* publicados em estudos de associação genômica ampla (*GWAS*) e *SNPs* próximos a microssatélites com associação ao conteúdo de proteína, óleo e resistência à nematoides, principais linhas de pesquisa desenvolvidas pelo grupo de pesquisa. Posteriormente, foi realizada genotipagem dos *SNPs* selecionados em 34 genótipos utilizados no programa de melhoramento do BIOAGRO/UFV. Esta genotipagem foi realizada pela empresa LGC Genomics.

A fenotipagem dos 34 genótipos para proteínas foi realizada a partir do método de KJELDAHL, que determina o nitrogênio total em amostras. O teor de óleo foi analisado por espectrometria de infravermelho próximo (NIR), utilizando o equipamento Antaris II FT-NIR analyzer (Thermo Fisher Scientific Brasil Instrumentos de Processo Ltda., São Paulo, SP, Brasil). Esses dados de fenotipagem e genotipagem foram utilizados para a associação nos programas GQMOL (como marca simples) e TASSEL (como estudo de seleção genômica ampla-GWAS). A partir disso foram selecionados *SNPs* para serem utilizadas como possíveis marcadores para proteína e uma nova lista de *SNPs* obtida para conteúdo de óleo. Dentre estes marcadores para óleo, foram selecionados e analisados aqueles polimórficos entre PMQS-SUPR e PMQS-08.

2.4.3 Metodologia de Genotipagem

Para a validação dos marcadores, foi utilizada a população do cruzamento entre os dois genitores selecionados (PMQS-SUPR e PMQS-08). A genotipagem dos marcadores *SNP* selecionados seguiu a metodologia de genotipagem KASP (LGC Genomic, 2003), executada no equipamento Real Time PCR 7500 Systems (Applied Biosciences). 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 0,07 uL de Primer Mix. A discriminação alélica foi executada no programa computacional AB 7500 Software v.2.3.

2.5 Fenotipagem

Após colheita manual das plantas 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, Brasil), de forma a gerar granulometria adequada. O farelo obtido foi analisado para o conteúdo de proteína e óleo por espectrometria de infravermelho próximo (NIR), utilizando-se o equipamento Antaris II FT-NIR analyzer (Thermo Fisher Scientific Brasil Instrumentos de Processo Ltda., São Paulo, SP, Brasil). Os valores obtidos foram convertidos para percentuais em base seca, por meio das fórmulas:

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

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

2.6 Análises genético-estatísticas

2.6.1 Estudo de associação marcador-característica

O estudo de associação marcador-característica ocorreu pela análise de marca simples (Schuster; 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 foram 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
AA	n_1	$Y_{11}, Y_{21}, \dots, Y_{n_1,1}$	2
Aa	n_2	$Y_{12}, Y_{22}, \dots, Y_{n_2,2}$	1
aa	n_3	$Y_{13}, Y_{23}, \dots, Y_{n_3,3}$	0

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

$$Y = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \dots \\ Y_{12} \\ Y_{22} \\ \dots \\ Y_{n_3,3} \end{bmatrix} \quad X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \dots & \dots \\ 1 & 1 \\ 1 & 1 \\ \dots & \dots \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad e \quad \varepsilon = \begin{bmatrix} e_{11} \\ e_{21} \\ \dots \\ e_{12} \\ e_{22} \\ \dots \\ e_{n_3,3} \end{bmatrix}$$

em que:

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

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

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

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.

2.6.2 Estimativa da herdabilidade

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

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

em que:

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

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

r é o número de repetições.

2.6.3 Correlação entre conteúdo 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 teor de proteína e óleo, e de correlação ambiental, por meio da estimação do coeficiente de correlação (r) entre as duas variáveis aleatórias X (teor de proteína) e Y (teor de óleo). A metodologia de obtenção pode ser verificada em Cruz et al. (2012), em que:

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

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

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

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

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

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

2.6.4 Análises Estatísticas

As análises estatísticas realizadas foram realizadas com auxílio dos aplicativos computacionais GENES (Cruz, 2013), GQMOL (Cruz, C.D; Schuster, 2004) e TASSEL (Bradbury et al. 2007).

2.7 Mapeamento dos *reads* e Análise de expressão diferencial

Trabalhos anteriores desenvolvidos pelo grupo de pesquisa PMQS/UFV realizou uma comparação do perfil transcricional de três variedades contendo alta proteína e baixo óleo com três variedades contendo baixa proteína e baixo óleo no estágio de desenvolvimento R5 (Silva, 2018). Com o intuito de identificar genes diferencialmente expressos dentro da possível região ancorando um importante QTL no Gm15 para o aumento do conteúdo de óleo em sementes de soja, foram analisados os dados de RNASeq para as duas variedades utilizadas neste estudo (*Sequence Read Archive* (SRA) número SRP155599).

Os *reads* dos genitores PMQS-SUPR e PMQS-08 foram mapeados com base no genoma de referências da soja Williams 82 na região do cromossomo 15 utilizando o *software* STAR (Dobin et al. 2013). A contagem dos *reads* mapeados e análise de expressão diferencial foi realizada com o *software* R a partir das contagens de alinhamento com o genoma de referência. Foi utilizada uma comparação contrastando a linhagens de alto conteúdo de óleo (PMQS-SUPR) e baixo conteúdo de óleo (PMQS-08). Os genes diferencialmente expressos (DEGs) encontrados na região selecionada foram anotados e investigados nas plataformas *Soybase*, *Phytozoma* e PUBMED para a identificação dos prováveis candidatos.

3. Resultados

3.1 Parâmetros genéticos e fenotípicos

Observou-se diferenças significativas ($p < 0,01$) entre genótipos para os conteúdos de proteína e óleo nos dois ambientes. A interação genótipo x ambiente (G x A) foi altamente significativa ($p < 0,01$), indicando comportamento diferenciado nos ambientes avaliados. A herdabilidade assumiu valores entre 71,02% e 71,75% para o conteúdo de óleo e 73,51% e 80,12% para o conteúdo de proteína, para Viçosa/2018 e Capinópolis/2018, respectivamente. Além dos resultados da análise de variância, observa-se valores acima de 1 para a razão

CV_g/CV_e para os dois ambientes. Os coeficientes de variação (CV) foram abaixo de 4% para os conteúdos de proteína e entre 5 e 8% para os conteúdos de óleo. Da mesma forma, baixos valores da relação $>QMR/<QMR$ foram observados, sendo 2,289 para proteína e 1,06 para óleo. Foi encontrada uma forte correlação negativa significativa ($p<0,01$) entre os teores de proteína e óleo nos grãos, com um coeficiente de correlação de Pearson variando de -0,483 a -0,730 para os quatro ambientes analisados. O resumo das análises é apresentado na **Tabela 2**.

Tabela 2. Resumo da análise individual e conjunta, e estimativas de parâmetros genéticos para conteúdo de proteína e óleo. Componentes de variância, herdabilidade no sentido amplo (h^2), coeficiente de variação (CV%) e razão entre coeficiente de variação genético e ambiental ($\frac{CV_g}{CV_e}$) para teor de proteína e óleo em dois dos quatro ambientes avaliados.

Característica	Conjunta			Individual			
	Ambientes	Genótipo	G x A	Genótipo	h^2	CV	$\frac{CV_g}{CV_e}$
Proteína (%)	Viçosa/2018			4,71**	73,51	3,8	1,17
	Capin./2018	1,76**	2,08**	2,99**	80,12	2,8	1,41
Óleo (%)	Viçosa/2018			1,90**	71,02	8,1	1,10
	Capin./2018	1,49**	1,50**	1,84**	71,75	5,0	1,12

** significativo a $p<0,01$

3.2 Características gerais da população

O conteúdo de proteína foi maior nas *RILs* no ambiente Viçosa/2018, apresentando um valor de 48,50% e menor em Capinópolis/2018, com um valor de 42,55%. Nos outros dois ambientes a média foi de 44%. Já o conteúdo de óleo foi maior nas *RILs* no ambiente de Capinópolis/2018, apresentando valor de 23,72% e menor em Viçosa/2018, com 15,23%. Nos outros dois ambientes a média foi de 22%. A **Tabela 3** e a **Figura 1** ilustram o desempenho dos genitores e da população para os conteúdos de proteína e óleo nos quatro ambientes.

Tabela 3. Médias fenotípicas dos conteúdos de proteína e óleo (%) da população de *RILs* e genitores nos quatro ambientes avaliados.

Característica	Ambientes	Núm. de <i>RILs</i>	(PQMS-SUPR)	(PMQS-08)	<i>RILs</i>
Proteína (%)	Viçosa/2017	459	-	-	44,01 ± 1,96
	Viçosa/2018	452	43,93 ± 3,28	50,54 ± 1,43	48,50 ± 2,34
	Capin./2018	451	36,12 ± 1,41	46,95 ± 1,18	42,55 ± 1,86
	Viçosa/2019	403	42,50 ± 0,07	47,86 ± 0,14	44,52 ± 2,32
Óleo (%)	Viçosa/2017	459	-	-	22,62 ± 1,51
	Viçosa/2018	452	17,23 ± 1,70	12,85 ± 1,03	15,23 ± 1,55
	Capin./2018	451	26,62 ± 1,62	20,40 ± 1,24	23,72 ± 1,53
	Viçosa/2019	403	24,62 ± 0,02	20,75 ± 0,13	22,02 ± 1,78

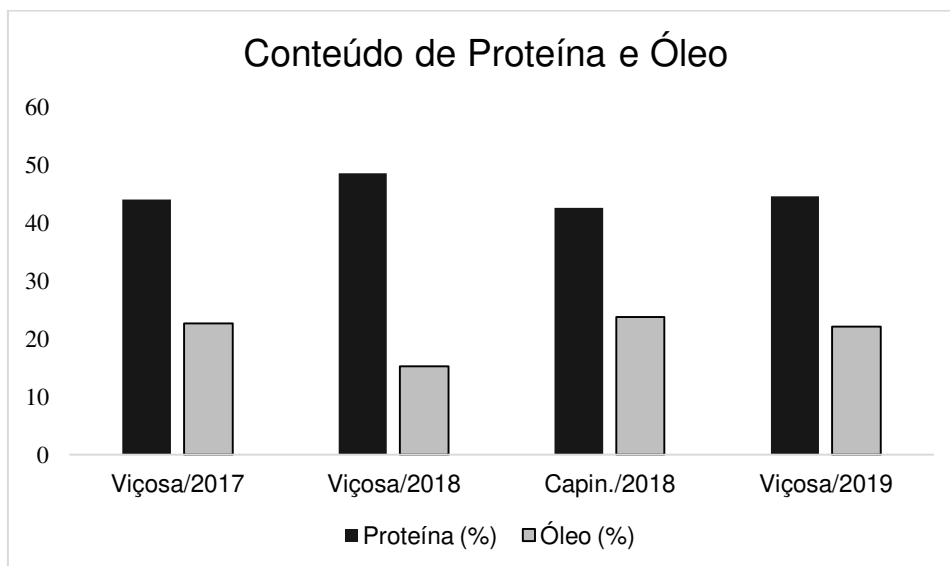


Figura 1. Conteúdo de proteína e óleo da população *RILs* nos quatro ambientes avaliados.

3.3 Associação de marcadores moleculares às características avaliadas

Em trabalho anterior desenvolvido pelo programa de melhoramento (PMQS), 269 marcadores *SNPs* foram selecionados na literatura e utilizados para genotipar 34 acessos de soja (Dados não publicados). Estes genitores foram fenotipados para o conteúdo de proteína e os dados utilizados para a associação nos programas GQMOL e TASSEL. A partir desses dados foram selecionados 24 *SNPs* para serem utilizadas como possíveis marcadores, já disponíveis e em análise em projetos paralelos.

Entre os objetivos deste estudo, em concordância com o PMQS, está a obtenção de novos marcadores associados com o conteúdo de óleo. Com o intuito de atender a este objetivo, os 34 acessos mencionados anteriormente foram fenotipados para o conteúdo de óleo e os dados utilizados para associação utilizando os programas GQMOL e TASSEL. Foram identificados um total de dezoito *SNPs* associadas por marca simples com um p-valor menor do que 5% (47, 61, 73, 115, 129, 134, 153, 164, 184, 185, 187, 190, 220, 221, 223, 245, 250 e 251) e sete *SNPs* identificadas por estudo de associação genômica ampla (99, 115, 134, 146, 161, 190 e 245).

Dentre os marcadores *SNPs* associados para óleo, nove se encontravam disponíveis no laboratório, entretanto apenas sete se apresentaram polimórficos para a população trabalhada (**Tabela 4**). Devido a aproximação dos marcadores ss61 e ss164, ambos localizados no cromossomo 20 da soja, o ss61 foi escolhido em decorrência de melhores dados de associação. Sendo assim, seis marcadores (ss61, ss73, ss115, ss129, ss134 e ss185) foram utilizados para genotipar a população F4 (Viçosa/2017) e esses dados utilizados juntamente com dados de fenotipagem para a realização das análises de associação (**Tabela 5**).

Apenas um marcador apresentou associação significativa para o conteúdo de óleo nos quatro ambientes avaliados, sendo este o ss115, localizado na posição 3,828,587 pb do cromossomo 15 (Gm15). O efeito aditivo deste gene para a característica de óleo variou de 0,4081 a 1,285 nos quatro ambientes. Para os ambientes Viçosa/2017, Capinópolis/2018 e Viçosa/2019 a variação foi de 0,9958 a 1,285. O alelo favorável do *SNP* em questão foi doado pelo genitor PQMS-SUPR.

A análise de dados para o conteúdo de proteína, também revela associação do marcador ss115 para os quatro ambientes avaliados. No entanto, o alelo favorável é doado pelo genitor PMQS-08 e tem efeito aditivo variando de 1,0761 a 1,5276. Além disto, os *SNPs* ss61, ss73, ss134 e ss185 associaram em uma localidade de ensaio cada e o alelo favorável foi doado pelo genitor PMQS-08 (Dados não mostrados).

Tabela 4. Dados genotípicos dos genitores PMQS-08 e PMQS-SUPR para 24 *SNPs* utilizadas no PMQS.

- (1) A primeira linha corresponde ao cromossomo e grupo de ligação.
- (2) A segunda linha o número do marcador. 0: alelo mais abundante. 2: alelo em menor abundância.
- (3) “GQMOL PTN” ou “GQMOL Oil” correspondem a análise da associação como marca simples utilizando p-value menor do que 0,05.
- (4) “TASSEL PTN” ou “TASSEL Oil” correspondem a análise da associação como GWAS utilizando p-value menor do que 0,05 para MLM (Modelo Linear Misto) ou 0,01 para GLM (Modelos Linear Generalizados).

	1/D1a	2/D1b	5/A1	8/A2	8/A2	10/O	11/B1	13/F	14/B2	15/E	15/E	18/G	19/L	20/I	20/I	20/I	20/I	20/I	20/I	20/I	20/I	20/I	20/I	
	46	134	44	10	94	19	129	153	73	115	116	56	100	62	61	164	176	41	185	190	106	108	111	114
PMQS-08	2	2	2	2	2	2	0	2	0	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2
PMQS-SUPR	0	0	2	0	2	0	2	2	2	2	0	0	2	2	0	0	2	2	0	2	2	2	2	2
GQMOL PTN - pvalue	0,03	0,013	0,01		0	0,02	0,046	0,014	0,061	0	0,01	0,05	0,02	0,02	0	0,1			0,04	0		100	100	
GQMOL PTN - R2	-2,01	-4,77	2,36		2,1	2,07	2,077	2,333	1,755	4,33	-2,4	1,83	-2,93	-	-2	-1,9			-2,2	3,5		2,5	2,6	
TASSEL PTN - MLM 0.05		0,03								0,01														
TASSEL PTN - GLM 0.01		0,002								0	0,01									0				
GQMOL Oil - pvalue		0,008					0,007	0,004	0,039	0					0	0			0	0				
GQMOL Oi 1 - R2		3,327					-1,79	-1,8	-1,26	-2,12					1,7	1,6			2,08	-3				
TASSEL Oil - MLM 0.05																								0
TASSEL Oil - GLM 0.01		0,005								0														0

Tabela 5. Associação dos marcadores selecionados para o conteúdo de óleo nos quatro ambientes avaliados, realizada no programa QMOL.

SNP ID	Chr/LG	Prob(F) - PTN	Prob(t) - PTN	R2(%) - PTN	EA Ptn	Genitor Ptn	Prob(F) - Oil	Prob(t) - Oil	R2(%) - Oil	EA Oil	Genitor Oil	
Viçosa/2017	61	20/I	1	0,9229	0,0024			1	0,6066	0,2078		
	73	14/B2	1	0,5676	0,908			1	0,6837	0,0453		
	115	15/E	0,0 **	0,0 **	9	1,1828	PMQS-08	0,0 **	0,0 **	10,7271	0,9958	PMQS-SUPR
	129	11/B1	1	0,9189	0,025			1	0,5283	0,1091		
	134	2/D1b	0,0734	0,0697	0,883			1	0,7173	0,0365		
	185	20/I	1	0,6289	0,0661			1	0,6845	0,461		
Viçosa/2018	61	20/I	0,0582	0,0551	10			0,1611	0,1571	0,5495		
	73	14/B2	1	0,573	0,0883			1	0,6059	0,0739		
	115	15/E	0,0 **	0,0 **	5,4038	1,0761	PMQS-08	0,01**	0,0098**	1,7882	0,4081	PMQS-SUPR
	129	11/B1	1	0,5874	0,081			1	0,6586	0,531		
	134	2/D1b	1	0,7569	0,261			0,1278	0,1234	0,6396		
	185	20/I	0,0417*	0,0394*	1,1282			0,0607	0,575	0,9576		
Capin./2018	61	20/I	0,0434 *	0,041 *	0,1382			1	0,5341	0,1535		
	73	14/B2	1	0,6838	0,0453			1	0,6103	0,0721		
	115	15/E	0,0 **	0,0 **	16,6953	1,5276	PMQS-08	0,0 **	0,0 **	17,3147	1,285	PMQS-SUPR
	129	11/B1	1	0,7902	0,018			1	0,7813	0,0198		
	134	2/D1b	0,0098 **	0,0095**	18			0,2303	0,2281	0,3973		
	185	20/I	0,2919	0,2922	0,3034			1	0,9273	0,0021		
Viçosa/2019	61	20/I	1	0,593	0,1966			1	0,5004	0,1321		
	73	14/B2	0,0169*	0,0162*	15			1	0,6086	0,2		
	115	15/E	0,0 **	0,0 **	7,8343	1,3062	PMQS-08	0,0 **	0,0 **	10,51	1,1378	PMQS-SUPR
	129	11/B1	0,2066	0,2037	0,4216			1	0,6488	0,0566		
	134	2/D1b	0,267	0,2663	0,3402			1	0,6424	0,0615		
	185	20/I	1	0,545	0,1569			1	0,9822	0,0001		

3.4 Genes diferencialmente expressos no cromossomo 15

O mapeamento dos *reads* foram analisados diretamente no cromossomo 15 do genoma de soja (Williams 82 Assembly 2), gerando ao final uma lista de 2.022 genes diferencialmente expressos (DEGs- *Differential Expressed Genes*). Deste total, foram selecionados aqueles localizados em um intervalo de 3,050,845 a 4,829,074 pb no cromossomo 15 da soja.

Ao nível nominal de significância ($p < 0,05$) foram detectados 35 genes, sendo destes 16 *up*-regulados, mais expressos na variedade de alto conteúdo de óleo PMQS-SUPR e 19 *down*-regulados, menos expressos nesta mesma variedade (**Tabela 6**). A partir de uma segunda análise mais rigorosa (FDR $q < 0,05$ e FC > 1.5) apenas nove genes passaram a ser considerados (Glyma.15g050900, Glyma.15g053700, Glyma.15g055700, Glyma.15g057700, Glyma.15g059900, Glyma.15g063000, Glyma.15g064100, Glyma.15g065600, Glyma.15g068700), sendo o primeiro mais próximo da região de interesse. Exigências maiores (FDR $q \leq 0.01$ FC > 1.5) restringe ainda mais o número de genes (Glyma.15g057700, Glyma.15g063000, Glyma.15g064100 e Glyma.15g068700).

A partir dos dados de contagem do mapeamento das variedades, foi gerada uma análise de componentes principais (PCA - *principal component analysis*) para determinar como as variedades se agrupavam quanto a seus perfis gerais de expressão gênica (**Figura 2**). Apesar de ser observado uma maior variação entre as repetições da variedade PMQS-08, foi possível diferenciá-las no componente PCA2, de modo a distingui-las em 2 grupos distintos com explicação de 33% da variação. Foi gerado ainda um gráfico do tipo *heatmap*, que descreve os dados de expressão gênica das replicatas em um esquema de cores. O ideal é que o mesmo grupo de amostras apresente cores semelhantes, conforme observado. Além disso, há um perfil de expressão distinto entre as diferentes condições além de um correto agrupamento entre as repetições biológicas das variedades (**Figura 3**). Também exibimos um gráfico *MAplot* que revela nível de expressão vs. variação de *log Fold Change* (**Figura 4**).

Tabela 6. Genes diferencialmente expressos na região próxima ao *SNP* ss115.

Genes	log2 FC	FC	p value	q value	UP/DOWN	Função
Glyma.15g038400	-1.7	0.3	0.03	0.31	DOWN	NA
Glyma.15g039100	-0.5	0.7	0.00	0.03	DOWN	Metabolismo proteico
Glyma.15g039300	-0.9	0.5	0.00	0.07	DOWN	Regulação da transcrição
Glyma.15g040000	-0.2	0.9	0.04	0.36	DOWN	Biossíntese de carboidrato
Glyma.15g041500	-0.3	0.8	0.00	0.09	DOWN	Metabolismo proteico
Glyma.15g043000	1.9	3.7	0.04	0.35	UP	Sinalização; desenvolvimento
Glyma.15g043400	0.2	1.1	0.01	0.18	UP	Transporte
Glyma.15g048200	-0.3	0.8	0.05	0.42	DOWN	Metabolismo proteico
Glyma.15g049200	-0.5	0.7	0.00	0.00	DOWN	Transporte de carboidrato
Glyma.15g050900	1.6	3.0	0.00	0.01	UP	Transporte
Glyma.15g051000	-0.2	0.9	0.03	0.34	DOWN	Metabolismo proteico; resposta a estresse
Glyma.15g051400	1.6	3.0	0.02	0.29	UP	Biossíntese de carboidrato
Glyma.15g053700	0.8	1.7	0.00	0.01	UP	Sinalização
Glyma.15g054700	-0.4	0.8	0.02	0.22	DOWN	Biossíntese de aminoácidos aromáticos
Glyma.15g055400	1.9	3.8	0.01	0.14	UP	Defesa
Glyma.15g055500	-0.4	0.8	0.02	0.27	DOWN	Metabolismo
Glyma.15g055700	0.8	1.8	0.00	0.01	UP	Biossíntese de riboflavina
Glyma.15g056400	-1.2	0.4	0.03	0.29	DOWN	Metabolismo de carboidratos
Glyma.15g057700	0.9	1.8	0.00	0.01	UP	Metabolismo
Glyma.15g058000	-1.0	0.5	0.04	0.35	DOWN	Regulação da transcrição
Glyma.15g058500	0.7	1.6	0.05	0.42	UP	NA
Glyma.15g059900	0.4	1.3	0.00	0.05	UP	Dobramento proteico
Glyma.15g060000	0.6	1.5	0.01	0.21	UP	Transporte
Glyma.15g060200	-0.4	0.8	0.04	0.38	DOWN	Regulação da transcrição
Glyma.15g061400	-1.3	0.4	0.01	0.17	DOWN	Biossíntese de colesterol
Glyma.15g061600	0.6	1.5	0.02	0.21	UP	Metabolismo de proteínas
Glyma.15g063000	6.4	83.3	0.00	0.01	UP	Regulação da transcrição
Glyma.15g063800	-2.6	0.2	0.01	0.18	DOWN	Sinalização celular
Glyma.15g064000	1.2	2.2	0.01	0.18	UP	Regulação da transcrição
Glyma.15g064100	0.8	1.7	0.00	0.01	UP	Resposta a estresse
Glyma.15g065600	0.9	1.8	0.00	0.03	UP	Biossíntese de terpenóides
Glyma.15g067100	-1.1	0.5	0.01	0.16	DOWN	Transporte
Glyma.15g276500	-1.6	0.3	0.02	0.22	DOWN	Resposta a estresse
Glyma.15g068500	-0.8	0.6	0.03	0.32	DOWN	Tradução
Glyma.15g068700	0.7	1.6	0.00	0.00	UP	Transporte de aminoácidos

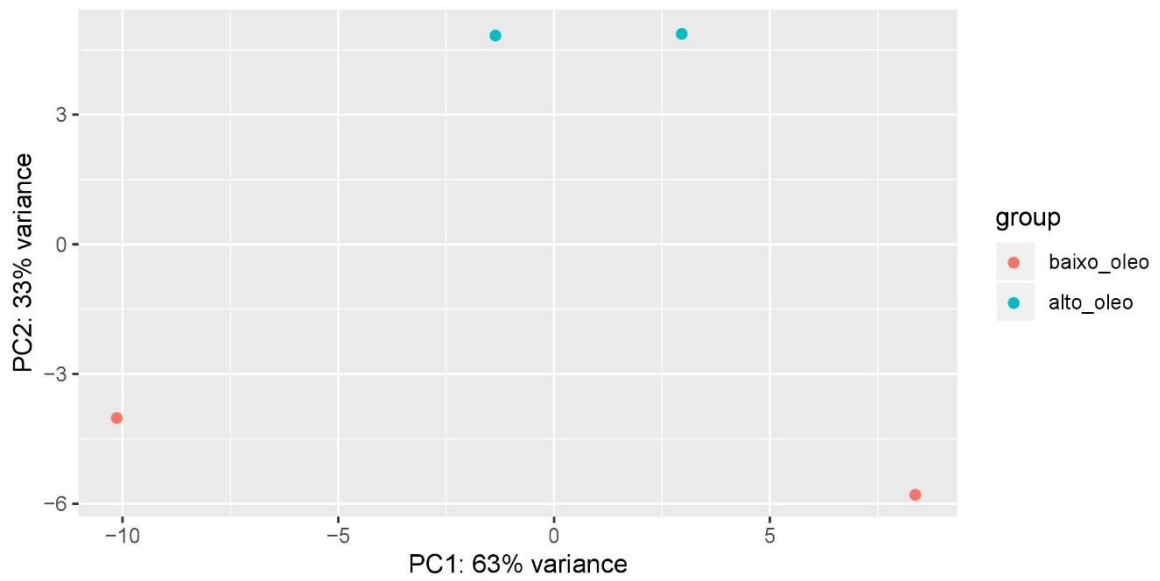


Figura 2. Análise de componentes principais (PCA) dos perfis gerais de expressão das 2 réplicas das variedades selecionadas para o estudo. Variância no PC1 (36%) e PC2 (33%).

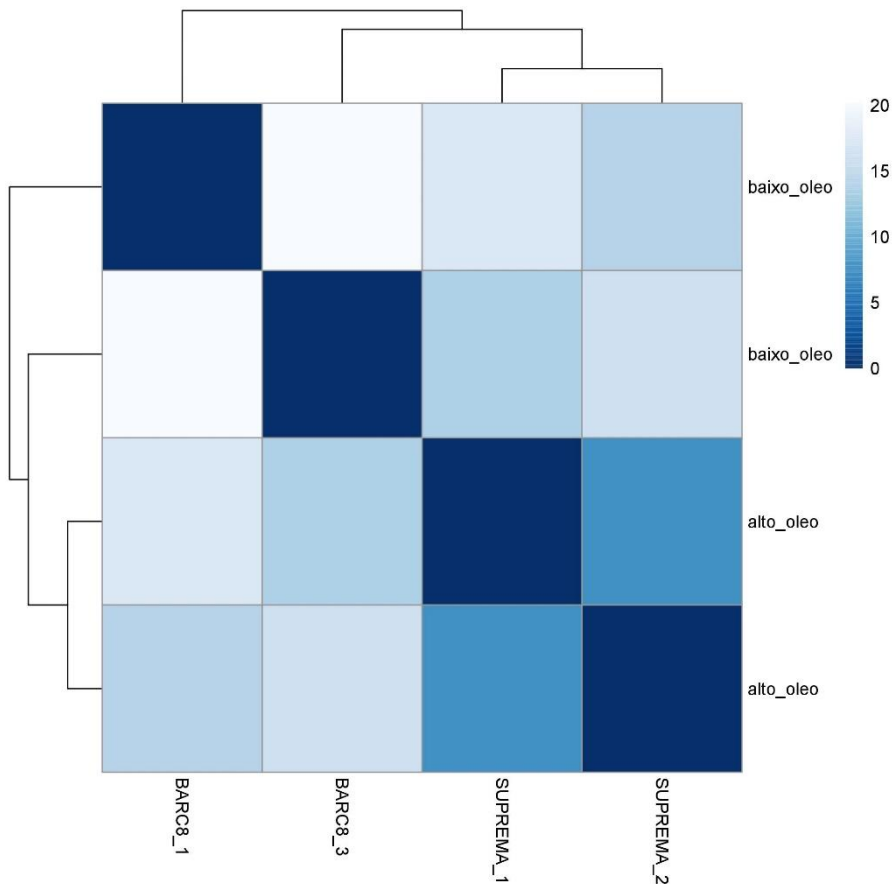


Figura 3. Heatmap mostrando a diferença no perfil das amostras nas condições de alto óleo (PQMS-SUPR) e baixo óleo (PMQS-08). Brilho de cor mais alto indica uma maior magnitude da expressão diferencial e vice-versa.

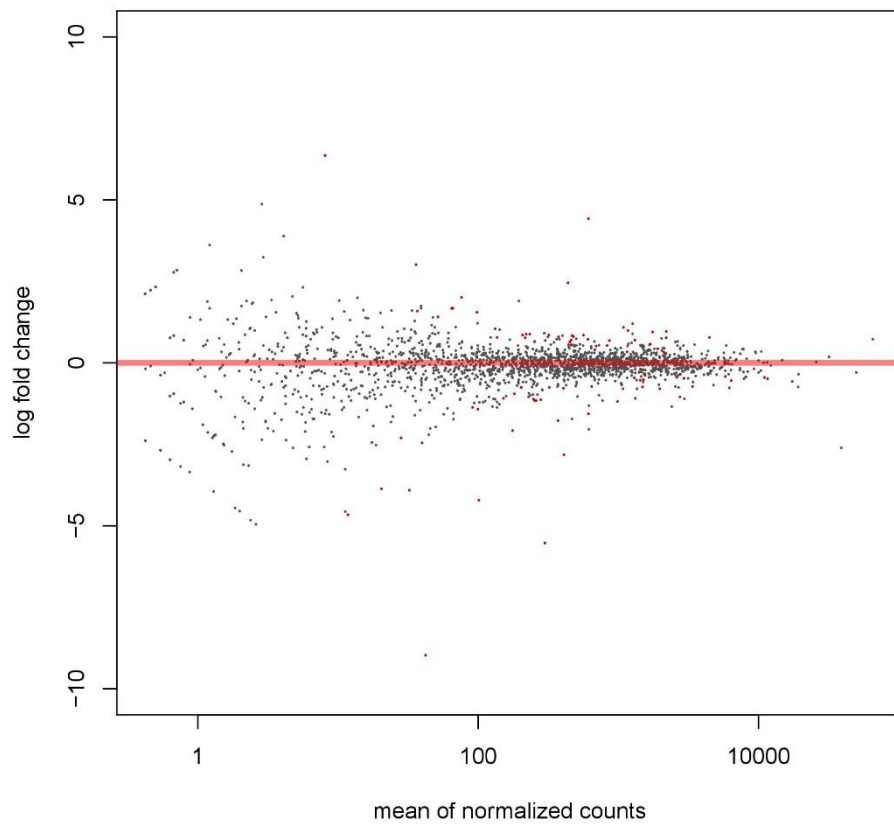


Figura 4. *MAplot*, mostrando nível de expressão vs. variação de *log Fold Change*. Pontos em vermelho são significativos.

4. Discussão

4.1 Parâmetros genéticos e fenotípicos

A obtenção de estimativas de parâmetros genéticos e fenotípicos têm grande importância em programas de melhoramento genético, uma vez que permite conhecer a estrutura genética de populações e auxilia na escolha do método mais apropriado para a seleção e o peso que deve ser atribuído a cada caráter (Vasconcelos et al. 2010; Rossmann, 2001).

O coeficiente de variação (CV) é utilizado para indicar o grau de precisão do experimento e um valor acima de 30% é considerado alto (Espíndola et al. 2012; Nogueira et al. 2012). Nesse estudo foram obtidos valores baixos e isso demonstra boa precisão experimental de acordo com Naoe et al. (2004), estando próximo de valores obtidos por seu trabalho e por Melo Filho (2002). Da mesma forma, o coeficiente de variação genético (CVg) é um parâmetro importante que permite inferir a variabilidade genética presente na população, interferindo diretamente no ganho por seleção (Ferrão et al. 2008). A razão CVg/CVe pode ser empregada como índice indicativo do grau de eficiência seletiva das progênies (Leite et al. 2016). Quando essa razão for igual ou maior que a unidade (>1), indica uma situação favorável à seleção (Gelape Faleiro et al. 2002; Yokomizo et al. 2003). Complementando esses dados, baixos valores da relação $>QMR<QMR$ foram observados e revelam boa precisão e homogeneidade dos dados fenotípicos e das variâncias residuais, assim como no trabalho de (Rodrigues et al. 2014).

O coeficiente de herdabilidade (h^2) é um dos parâmetros genéticos mais úteis para o trabalho dos melhoristas (Ramalho et al. 2012; Falconer, 1996) e é muito importante para o desenvolvimento de estratégias de melhoramento a serem adotadas. Está fortemente relacionado aos ganhos de seleção, sendo que quanto maior a herdabilidade de um caráter, maior a chance de sucesso na seleção. Estes valores de herdabilidade podem variar de zero a um, e estimativas acima de 0,70 são consideradas altas em espécies vegetais (Cruz et al. 2012; Ramalho et al. 2012). Os valores de herdabilidade obtidos nesse estudo, bem como os valores das relações entre os coeficientes de variação genética e ambiental refletem uma situação favorável para um maior ganho de seleção. Valores de herdabilidade semelhantes foram observados nos trabalhos de Zhang et al. (2015) e Hwang et al. (2014) para os conteúdos de óleo e proteína de soja.

Foi encontrada uma forte correlação negativa entre os conteúdos de proteína e óleo, conforme relatado em trabalhos anteriores (Burton, 1987; Li et al. 2018). Essa estimativa é importante pois mostra como a seleção de um caráter pode influenciar a expressão de outros caracteres. Além de visar o aprimoramento de uma característica principal, há forte interesse também em manter ou melhorar a expressão de outros caracteres e essa correlação tem dificultado esse processo (Eskandari et al. 2013). Tal associação pode representar a presença de efeitos pleiotrópicos ou de ligação, dificultando a melhoria em ambas as características simultaneamente (Chung et al. 2003).

O resumo dos dados fenotípicos apresentados neste trabalho evidencia o fato de que o conteúdo de proteína e óleo é muito afetado por condições ambientais. Todas estas informações obtidas podem ser complementadas ao uso de marcadores moleculares para aumentar a eficiência de seleção de progênies.

4.2 Associação de marcadores moleculares às características avaliadas

Dentre os marcadores apontados nesse estudo, os SNPs ss61, ss73, ss134 e ss185 se mostraram associados para proteína em uma localidade. Nestes casos, o alelo favorável exerceu pouco efeito e não se mostrou estável ao longo dos ambientes testados, provavelmente pelo fato de que características de herança quantitativa, como no caso de proteína, estarem envolvidas com múltiplos genes com pequeno efeito e que podem ser sensíveis a mudanças ambientais. A interação entre genótipos e ambientes também pode explicar a baixa consistência dos *locus* identificados, como aponta Contreras-Soto et al. (2017).

Em nosso trabalho ressaltamos a importância do SNP ss115, localizado no cromossomo 15 (Gm15). O marcador em questão se mostrou comum tanto para conteúdo de óleo quanto para conteúdo de proteína na semente e apresenta valores significativos. Essa forte associação indica um possível efeito pleiotrópico dos genes próximos a esta marca ou *QTLs* fortemente ligados. O efeito pleiotrópico é mais provável, pois essas características quase sempre co-segregam e a tentativa de separá-las ainda não obteve êxito (Wang et al. 2014). Outro ponto relevante é a forte estabilidade encontrada ao longo dos quatro ambientes testados, apresentando um valor menor de associação no ambiente de Viçosa/2018, que revelou menor efeito aditivo e R^2 para conteúdo de óleo, por isso a interação tende a ser pequena e conseqüentemente, a associação também. Além disso, estudos anteriores mostram que o efeito dos loci no cromossomo 15 é ambientalmente dependente (Pathan et al. 2013;

Tajuddin et al. 2003). Dessa forma, esse loci representa uma fonte potencial de marcadores para quebrar, pelo menos parcialmente, a correlação negativa existente entre proteína e óleo, fundamental para desenvolvimento de linhagens de soja.

Kim et al. (2016) avaliou uma população de linhagens puras para proteína e óleo e através de mapeamento fino do cromossomo 15 encontrou uma região de 535kb no intervalo entre BARCSOYSSR_15_0161 e BARCSOYSSR_15_0194, que se sobrepõe parcialmente à faixa determinada por ss715622170 e 115. Zhang et al. (2018) encontrou um QTL de grande efeito identificado em 3,8-4,8 Mb no cromossomo 15, sendo liderado por ss715622170 e ss115. Este *locus* foi associado simultaneamente com o conteúdo de proteína, óleo, ácido linolênico e diversos aminoácidos, explicando 22% e 14% da variação genética para teor de proteína e óleo, respectivamente. Lee et al. (2019) aponta para quatro marcadores no cromossomo 15, dentre eles ss115, posicionados dentro de um intervalo de 139 kb. Essas informações revelam o efeito e estabilidade do loco estudado e são extremamente úteis na seleção assistida por marcadores. Todos estes dados juntamente com os obtidos neste projeto corroboram para a aplicabilidade do marcador ss115 para programas de seleção assistida.

4.3 Genes diferencialmente expressos no cromossomo 15

O marcador *SNP* ss115 (Gm15) se mostrou significativo e estável no aumento do conteúdo de óleo nos grãos de soja. Com o objetivo de investigar os genes expressos próximos a essa região de interesse, foi gerada uma lista de DEGs a partir de experimentos de RNASeq para os genitores utilizados neste projeto. Para isto, foi aplicada uma seleção a partir de *p*-value ajustado ou *False Discovery Rate* (FDR) e *Fold Change* (FC) para uma interpretação de dados mais confiável. O valor de $FC \geq 2$ e uma correção de FDR com $q \leq 0,05$ são comuns. No entanto, FC de 1,3 ou 1,5 e FDR com $q \leq 0,1$ e $q \leq 0,2$ podem ser encontrados na literatura (Zhu et al. 2005; Booth et al. 2004; Han et al. 2006; Hartmann et al. 2009). Os critérios para determinar qual nível de FC e FDR deve ser usado como limite dependem da finalidade do experimento e da especificidade desejada (Pawitan et al. 2005; Hartmann et al. 2009).

O número de genes identificados como diferencialmente expressos sem o uso de quaisquer limiares, exceto $p \leq 0,05$, foi de 35 genes, envolvidos em vários aspectos do metabolismo de proteínas, carboidratos e lipídeos, regulação da transcrição, tradução, mecanismos de defesa das plantas, sinalização e transporte celular, incluindo um transportador de aminoácidos.

A aplicação de um limiar de FDR com um nível de significância amplamente utilizado FDR $q \leq 0,05$ e $FC > 1.5$ em nosso conjunto de dados reduziu o número de genes para nove. Dentre eles, Glyma.15g053700 está descrito como pertencente à família de proteínas fosfatases do tipo 2C (PP2C). Estudos em *Arabidopsis thaliana* realizados por Rodriguez (1998) identificaram as PP2C como principais intervenientes nos processos de transdução de sinal das plantas. Fuchs et al. (2003) descreve que esse grupo funciona como um regulador negativo na transdução de sinal do ácido abscísico. Além disso, interrompe a sinalização do ácido abscísico em *Striga*, resultando em alta transpiração e absorção eficiente de nutrientes sob condições de seca (Fujioka et al. 2009).

O gene Glyma.15g050900 corresponde a família de proteínas adaptadoras ENTH/ANTH/VHS. Essas proteínas são responsáveis pelo tráfico de membranas, envolvendo a regulação da localização intracelular de proteínas e lipídios e é necessário para absorção metabólica, crescimento e desenvolvimento celular (De Craene et al. 2012). Glyma.15g055700 está relacionado a síntese de riboflavina, em estreita concordância com o estudo de Waters et al. (2018), que também aponta o gene. Glyma.15g057700 corresponde a atividades de anidrases carbônicas, enzimas que catalisam a hidratação reversível do CO_2 , importante componente da maioria dos tecidos vegetais superiores (DiMario et al. 2017). Glyma.15g059900 pertence a família de chaperonas moleculares, também conhecidas como proteínas de choque térmico, auxiliam proteínas mal dobradas e ajudam outras a alcançarem sua conformação nativa, implicado em múltiplas funções, variando de resposta ao estresse, tráfico intracelular, processamento de antígeno, controle da proliferação celular e diferenciação (Aguilar-Rodríguez et al. 2016). Glyma.15g063000 corresponde ao fator de transcrição *ILR3*. Em *Arabidopsis thaliana*, *ILR3* regula a resposta de deficiência de ferro, a biossíntese de glucosinolato alifático (*GLS*) e a resposta do patógeno (Samira et al. 2018). Glyma.15g064100 corresponde ao gene *DUF538*. Estudos mostram que a expressão do gene *DUF538* leva a degradação das moléculas de clorofila e aumento de compostos antioxidantes para defender as plantas contra condições de estresse (Gholizadeh, 2016). Glyma.15g065600 está envolvido na biossíntese de terpenóide. Glyma.15g068700 é responsável por transporte de aminoácidos, pertence à família *Lysine histidine transporter-like 8*. Transportadores de aminoácidos são os principais mediadores de distribuição de nitrogênio, importantes na alocação de recursos em plantas (Sheng et al. 2014).

Destes genes citados acima, Glyma.15g057700, Glyma.15g063000, Glyma.15g064100 e Glyma.15g068700 são mais significativos ao se utilizar FDR $q < 0,01$, apesar de se encontrarem mais distantes da região de interesse.

O fato dos genes encontrados estarem ligados ao desenvolvimento e crescimento da planta pode ser devido a fase em que foram realizadas as análises, já que o estágio R5 corresponde a fase inicial do enchimento dos grãos e desenvolvimento da planta. Funções relacionadas ao metabolismo e transporte de proteínas também foram encontradas, apontando para uma região favorável a essa característica. Maiores estudos são necessários para investigar a ocorrência destes genes.

5. Conclusão

Este trabalho validou o efeito do marcador SNP ss115 sobre o aumento do conteúdo de óleo e proteína em grãos de soja, apresentando uma alternativa viável para o desafio da correlação negativa existente entre essas características. As informações com base em quatro ambientes distintos revelam o efeito e a estabilidade do marcador, muito útil na seleção assistida por marcadores.

Genes diferencialmente expressos nesta região foram apresentados para maior compreensão dos efeitos observados, revelando funções relacionadas a transporte, sinalização, resposta ao estresse e metabolismo, abrindo caminhos para futuros estudos. O conhecimento sobre esses materiais e genes será útil para futuras pesquisas no programa de melhoramento da soja.

6. Referências

- AGUILAR-RODRÍGUEZ, J. et al. The molecular chaperone DnaK is a source of mutational robustness. **Genome biology and evolution**, v. 8, n. 9, p. 2979-2991, 2016.
- BURTON, J.W. Quantitative genetics: results relevant to soybean breeding. In: R. Shibles (Ed.), Soja: Melhoria, Produção e Usos, 2ª ed.; **Sociedade Americana de Agronomia: Madison**, WI, EUA, 211–241, 1987.
- BRADBURY, Peter J. et al. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**, v. 23, n. 19, p. 2633-2635, 2007.
- AKOND, A.M; RAGIN, B.; BAZZELLE, R.; KANTARTZI, S.K.; MEKSEM, K.; KASSEM, M.A. Quantitative trait loci associated with moisture, protein, and oil content in soybean [*Glycine max* (L.) Merr.] **Journal of Agricultural Science**. v. 4, p. 16-25, 2012.
- APROSOJA. 2019. Aprosoja Brasil. Disponível em: <<http://aprosojabrasil.com.br/>>. Acesso em: 10 de junho. 2019.
- 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.
- BOOTH, S.; BOWMAN C.; BAUMGARTNER, R. Identification of central nervous system genes involved in the host response to the scrapie agent during preclinical and clinical infection. **Journal of General Virology**. v. 85, n. 11, p. 3459-3471, 2004.
- CHOI, I.Y; HYTEN, D.L.; MATUKUMALLI, L.K.; SONG, Q.; CHAKY, J.M.; QUIGLEY, C.V. Soybean Transcript Map: Gene Distribution, Haplotype and Single-Nucleotide Polymorphism Analysis. **Genetics**. v. 176, n. 1, p. 685-696, 2007.
- CHUNG, J. et al. The seed protein, oil, and yield QTL on soybean linkage group I. **Crop science**. v. 43, n. 3, 1053-1067, 2003.
- CLEMENTE, T.E; CAHOON, E.B. Soybean Oil: Genetic Approaches for Modification of Functionality and Total Content. **Plant physiology**. v. 151, n. 3, 1030-1040, 2009.
- COLLINS, F.; BROOKS, L.; CHAKRAVARTI, A. A DNA polymorphism discovery resource for research on human genetic variation. **Genome research**. v. 8, n. 12, 1229-1231, 1998.
- CONAB. 2019. Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira de Grãos. Décimo levantamento. **Safra 2018/2019**. v. 6, n.10.
- CONTRERAS-SOTO, R.I. et al. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-Based haplotype analysis. **PLoS one**. v. 12, n. 2, p.1-22, 2017.

- CRUZ, C.D; Schuster, I. GQMOL: aplicativo computacional para análise de dados moleculares e de suas associações com caracteres quantitativos. Versão 2.1. Viçosa: UFV. 2004. Disponível em: <<http://www.ufv.br/dbg/gqmol/gqmol.htm>>. Acesso em: 10 de abril de 2019.
- CRUZ, C.D; Regazzi, A.J.; Carneiro, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa:UFV. 4ed. v.1, 514p, 2012.
- 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.
- DE CRAENE, J.O.; RIPP, R.; LECOMPTE, O. et al. Evolutionary analysis of the ENTH/ANTH/VHS protein superfamily reveals a coevolution between membrane trafficking and metabolism. **BMC Genomics**. v. 13, n. 1, p. 297, 2012.
- DELLAPORTA, S.L.; WOOD, J.; HICKS, J.B. A plant DNA minipreparation: Version II. **Plant Molecular Biology Reporter**. v.1, n.4, p.19-21, 1983.
- DIERS, B.W.; KEIM, P.; FEHR, W.R.; SHOEMAKER, R.C. RFLP analysis of soybean seed protein and oil content. **Theor Appl Genet**. v.83, p.608–612, 1995.
- DIMARIO, R.; CLAYTON, H.; MUKHERJEE, A. Plant carbonic anhydrases: structures, locations, evolution, and physiological roles. **Elsevier**. v. 10, n. 1, p. 30-46, 2017.
- DOBIN, A. et al. STAR: ultrafast universal RNA-seq aligner. **Bioinformatics**. v. 29, n. 1, p. 15-21, 2013.
- ESKANDARI, M.; COBER, E.R.; RAJCAN, I. Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield. **Theor Appl Genet**. v. 126, n. 6, p. 1677-1687, 2013.
- ESPÍNDOLA, S.; FINHOLDT, R. Avaliação da performance produtiva e agrônômica de genótipos de soja para a região de Jaboticabal, SP. **FAZU em Revista**, n. 08, 2012.
- FALCONER, D.S. et al. Introduction to quantitative genetics. 4 ed. **Trends Genet**. 280p, 1996.
- FAN, J.; CHEE, M. GUNDERSON, K. L. Highly parallel genomic assays. **Nature Reviews Genetics**, v. 7, n. 8, p. 632, 2006.
- FASOULA, V.; HARRIS, D.; BOERMA, H.R. Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. **Crop Science**. v. 44, n. 4, 1218-1225, 2004.
- FERRÃO, R.; SILVA, M.A.; CARNEIRO, P.; CRUZ, C. Parâmetros genéticos em café Conilon. **Pesq. agropec. bras.**, Brasília, v.43, n.1, p.61-69, 2008.
- FILHO, M. Caracterização de populações de soja de alto conteúdo de proteína quanto à

- produtividade e à qualidade fisiológica das sementes. Viçosa. **Tese de Doutorado**. Universidade Federal de Viçosa, 2002.
- FUCHS, S.; GRILL, E.; MESKIENE, I., SCHWEIGHOFER, A. Type 2C protein phosphatases in plants. **The FEBS journal**. v. 280, n. 2, 681-693, 2013.
- FUJIOKA, H.; SAMEJIMA, H.; SUZUKI, H. Aberrant protein phosphatase 2C leads to abscisic acid insensitivity and high transpiration in parasitic *Striga*. Fujioka, H. **Nat. Plants**. v. 5, p. 258-262, 2019.
- GELAPE FALEIRO, F. et al. Comparação de blocos casualizados e testemunhas intercalares na estimação de parâmetros genéticos em feijoeiro. **Pesquisa Agropecuária Brasileira**. v. 37, n. 12, 1675-1680, 2002.
- GHOLIZADEH, A. DUF538 protein superfamily is predicted to be chlorophyll hydrolyzing enzymes in plants. **Physiol Mol Biol Plants**. v. 22, n. 1, p. 77-85, 2016.
- GOETTEL, W.; XIA, E.; UPCHURCH, R.; WANG, M.L.; Chen, P.; Identification and characterization of transcript polymorphisms in soybean lines varying in oil composition and content. **BMC Genomics**. v. 15, n. 1, p. 299, 2014.
- HAN, T.; WANG, J.; TONG, W.; MOORE, M.M.; FUSCOE, J.C.; CHEN, T. Microarray analysis distinguishes differential gene expression patterns from large and small colony Thymidine kinase mutants of L5178Y mouse lymphoma cells. **BMC Bioinformatics**. v. 7, n. 2, p. S9, 2006.
- HARTMANN, A. et al. Effects of threshold choice on the results of gene expression profiling, using microarray analysis, in a model feeding experiment with rats. **Archiv für Tierzucht**. v. 52, n. 1, p. 65-78, 2009.
- HWANG, E.Y. et al. A genome-wide association study of seed protein and oil content in soybean. **BMC Genomics**. v. 15, n. 1, p. 1, 2014.
- HYTEN, D.L. et al. Seed quality QTL in a prominent soybean population. **Theor Appl Genet**. v. 109, n. 3, p. 552-561, 2004.
- HYTEN, D.L.; SONG, Q.; ZHU, Y.; CHOI, I.Y.; NELSON, R.L.; COSTA, J.M.; CREGAN, P. Impacts of genetic bottlenecks on soybean genome diversity. **Proceedings of the National Academy of Sciences**. v. 103, n. 45, p. 16666-16671, 2006.
- JAUREGUY, L.M.; CHEN, P.; SCABOO, A.M. Heritability and correlations among food-grade traits in soybean. **Plant Breed**. v. 130, n. 6, p. 647-652, 2011.
- KIM, M.; SCHULTZ, S.; NELSON, R.; DIERS, B.W. Identification and fine mapping of a soybean seed protein QTL from PI 407788A on chromosome 15. **Crop Science**. v. 56, n. 1, p. 219-225, 2016.

- LEE, S.; VAN, K.; SUNG, M. et al. Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. **Theor Appl Genet.** v. 132, n. 6, p. 1639-1659, 2019.
- LEITE, W.S. et al. Genetic parameters estimation, correlations and selection indexes for six agronomic traits in soybean lines F8. **Comunicata Scientiae.** v. 7, n. 3, p. 302-310, 2016.
- LI, Y.; REIF, J.; HONG, H. et al. Genome-wide association mapping of QTL underlying seed oil and protein contents of a diverse panel of soybean accessions. *Plant Science.* v. 266, p. 95-101, 2018.
- LGC GENOMICS. KASP genotyping chemistry User guide and manual. 2013. Disponível em: <<https://www.lgcgroup.com/>>. Acesso em: 24 de abril. 2019.
- MAHMOUD, A.A.; NATARAJAN, S.S.; BENNETT, J.O. et al. Effect of six decades of selective breeding on soybean protein composition and quality: A biochemical and molecular analysis. **J Agric Food Chem.** v. 54, n. 11, p. 3916-3922, 2006.
- NAOE, L.; SEDIYAMA, C.; PIOVESAN, N. et al. Estimaco de componentes de varincia e parmetros genticos, considerando o parentesco em populao de retrocruzamento de soja. **Revista Cincia Agroambiental.** v. 1, n. 1, p. 21-26, 2006.
- NICHOLS, D.; GLOVER, K.D.; CARLSON, S.R.; SPECHT, J.E.; DIERS, B. Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. **Crop Science,** v. 46, n. 2, p. 834-839, 2006.
- NOGUEIRA, A.; SEDIYAMA, T. et al. Anlise de trilha e correlaes entre caracteres em soja cultivada em duas pocas de semeadura. **Bioscience Journal.** v. 28, n. 6, 2012.
- PANTHEE, D.; PANTALONE, V. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. **Crop Science.** v. 45, n. 5, p. 2015-2022, 2005.
- PATHAN, S.; VUONG, T.; CLARK, K. et al. Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. **Crop Science.** v. 53, n. 3, p. 765-774, 2013.
- PATIL, G. et al. Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. **Scientific Reports.** v. 6, p. 19199, 2016.
- PATIL, G. et al. Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. **Theoretical and Applied Genetics.** v. 130, n. 10, p. 1975-1991, 2017.
- PAWITAN, Y.; MICHIELS, S.; KOSCIELNY, S. et al. False discovery rate, sensitivity and sample size for microarray studies. **Bioinformatics.** v. 21, n. 13, p. 3017-3024, 2005.

- PHAM, A.T.; LEE, J.D.; SHANNON, J.G.; BILYEYU, K.D. Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. **BMC Plant Biol.** v. 10, n. 1, p. 195, 2010.
- PHANSACK, P. et al. Multi-Population Selective genotyping to identify soybean [*Glycine max* (L.) Merr.] seed protein and oil QTLs. **G3: Genes, Genomes, Genetics.** v. 6, n. 6, p. 1635-1648, 2016.
- QI, Z. et al. Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. **Euphytica.** v. 179, n. 3, p. 499-514, 2011.
- RAMALHO, M. A.; SANTOS, J. et al. Aplicações da genética quantitativa no melhoramento de plantas autógamas. **Lavras: Ed. UFLA.** 522 p. 2012.
- RAMALHO, M.; SANTOS, J.; ZIMMERMANN, M. Genética quantitativa em plantas autógamas: aplicações ao melhoramento do feijoeiro. Goiânia: UFG. 271 p.1993.
- RODRIGUES, J.; ARRUDA, K.; CRUZ, C.D.; PIOVESAN, N.D.; BARROS, E.G.D.; MOREIRA, M.A. Biometric analysis of protein and oil contents of soybean genotypes in different environments. **Pesquisa Agropecuária Brasileira.** v. 49, n. 6, p. 475-482, 2014.
- RODRIGUEZ, P.L. Protein phosphatase 2C (PP2C) function in higher plants. **Plant molecular biology.** v. 38, n. 6, p. 919-927, 1998.
- ROSSMANN, H. Estimativas de parâmetros genéticos e fenotípicos de uma população de soja avaliada em quatro anos. **Tese de Doutorado.** Universidade de São Paulo. Universidade de São Paulo, 2001.
- SAMIRA, R. et al (2018) The bHLH transcription factor ILR3 modulates multiple stress responses in Arabidopsis. **Plant Mol Biol.** v. 97, n. 4-5, p. 297-309, 2018.
- SCHUSTER, I.; CRUZ C.D. **Estatística Genômica**, 2. ed. Viçosa, Mg. 2008.
- SEBOLT, A.; SHOEMAKER, R. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. **Crop Science.** v. 40, n. 5, p. 1438-1444, 2000.
- SHENG, L.; DENG, L.; YAN, H. et al. A genome-wide analysis of the AAAP gene family in maize. **Journal of Proteomics and Bioinformatics.** v. 7, n. 1, p. 023-033, 2014.
- SILVA, L.A.C. Caracterização do perfil de expressão em variedades de soja contrastantes para o teor de proteína. **Tese de Doutorado.** Universidade Federal de Viçosa, 2018.
- SILVA, L.C.C.; BUENO, R.D.; DA MATTA, L.B.; PEREIRA, P.H.S.; MAYRINK. D.B.; PIOVESAN, N.D.; DAL-BIANCO, M. Characterization of a new GmFAD3A allele in Brazilian CS303TNKCA soybean cultivar. **Theor Appl Genet.** v. 131, n. 5, p. 1099-1110,

- 2018.
- SONG, Q.; HYTEN, D.L.; JIA, G. et al. Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. **PLoS One**. v. 8, n. 1, p. e54985, 2013.
- SOYSTATS. (2019). Disponível em: <<http://www.soystats.com/>>. Acesso em: 10 de junho. 2019.
- TAJUDDIN, T.; WATANABE, S.; SCIENCE, N.Y.B. Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines. *Breeding Science*. v. 53, n. 2, p. 133-140, 2003.
- TRIPATHI, N. et al. Molecular approaches for genetic improvement of seed quality and characterization of genetic diversity in soybean: a critical review. **Biotechnology letters**. v. 38, n. 10, p. 1645-1654, 2016.
- VAN, K.; MCHALE, L. Meta-analyses of QTLs associated with protein and oil contents and compositions in soybean seed. **International journal of molecular sciences**. v. 18, n. 6, p. 1180, 2017.
- VASCONCELOS, E.S. et al. Estimativas de ganho genético por diferentes critérios de seleção em genótipos de alfafa. *Ceres*. v. 57, n. 2, 2015.
- WANG, X.; JIANG, G.L.; GREEN, M. et al. Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. **Mol Genet Genomics**. v. 289, n. 5, p. 935-949, 2014.
- WATERS, B.M.; AMUNDSEN, K.; GRAEF, G. Gene Expression Profiling of Iron Deficiency Chlorosis Sensitive and Tolerant Soybean Indicates Key Roles for Phenylpropanoids under Alkalinity Stress. *Front Plant Sci*. v. 9, p. 10, 2018.
- WILCOX, J. et al. Interrelationships among seed quality attributes in soybean. *Crop Science*. v. 41, n. 1, p. 11-14, 2011.
- YOKOMIZO, G. et al. Caracterização fenotípica e genotípica de progênies de pupunheira para palmito. **Pesquisa Agropecuária Brasileira**. v. 38, n. 1, p. 67-72, 2003.
- ZAFALON, M. Brasil pode virar líder mundial em soja em 2019. **Vaivém das Commodities**. Disponível em: <<https://www1.folha.uol.com.br/>>. Acesso em: 10 de maio. 2019.
- ZHANG, J.; WANG, X.; LU, Y. 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, 2018.
- ZHANG, Y.H. et al. Marker-assisted breeding for transgressive seed protein content in soybean [*Glycine max* (L.) Merr.]. **Theor Appl Genet**. v. 128, n. 6, p. 1061-1072, 2015.
- ZHU, D.; HERO, A. CHENG, H. et al. Network constrained clustering for gene microarray

data. **Bioinformatics**. v. 21, n. 21, p. 4014-4020, 2005.

ZHU, Y.L.; SONG, Q.J.; HYTEN, D.L. et al. Single-Nucleotide Polymorphisms in Soybean.

Genetics. v. 163, n. 3, p. 1123-1134, 2003.