LUIZ CLÁUDIO COSTA SILVA

CARACTERIZAÇÃO MOLECULAR DE ALELOS ASSOCIADOS À QUALIDADE DO GRÃO DE SOJA

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

VIÇOSA
MINAS GERAIS – BRASIL
2017

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

T
S586c
2017

Silva, Luiz Cláudio Costa, 1987-
Caracterização molecular de alelos associados à qualidade do grão de soja / Luiz Cláudio Costa Silva. – Viçosa, MG, 2017.
xiii, 117f. : il. (algumas color.) ; 29 cm.

Inclui apêndice.
Orientador: Elizabeth Pacheco Batista Fontes.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Soja - Melhoramento genético. 2. Soja - Produtos. 3. Proteínas de soja. 4. Soja como alimento. I. Universidade Federal de Viçosa. Departamento de Bioquímica e Biologia Molecular. Programa de Pós-graduação em Genética e Melhoramento. II. Título.

CDD 22 ed. 633.343

LUIZ CLÁUDIO COSTA SILVA

CARACTERIZAÇÃO MOLECULAR DE ALELOS ASSOCIADOS À QUALIDADE DO GRÃO DE SOJA

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

APROVADA: 08 de março de 2017.

Eveline Teixeira Caixeta	Ivan Schuster
Pedro Augusto Braga dos Reis	Maximiller Dal-Bianco Lamas Costa (Coorientador)

Elizabeth Pacheco Batista Fontes
(Orientadora)

DEDICATÓRIA

Ao meu pai Francisco, pelo amor e exemplo no pouco tempo juntos

À minha mãe D. Gininha, por todo amor, carinho e incentivo irrestrito

Aos meus irmãos Tunico, Zé Roberto, Rubinho, Roberta, Flavinho e Ana Paula, pelo exemplo, carinho e incentivo

À minha amada Cândida, por todo amor, companheirismo e força

Dedico.

AGRADECIMENTOS

A Deus, por ter me dado uma família abençoada, por ter estado à minha frente me guiando por todas as dificuldades, e por ter colocado em meu caminho pessoas tão importantes.

À minha mãe, D. Gininha, pelo exemplo, pela confiança, pela educação, e pelo amor incondicional que formou e manteve nossa família sempre unida.

Ao meu pai, Francisco (*in memoriam*), pelo amor recebido, e pelo exemplo como pai e homem, mesmo no pouco tempo em que convivemos.

Aos meus irmãos Tunico, Zé Roberto, Rubinho, Roberta, Flavinho e Ana Paula, pelo amor, companheirismo, paciência e apoio na minha jornada acadêmica.

À minha amada Cândida, por todo o amor, força, companheirismo, apoio, confiança, risadas e tantas outras coisas, por todos os momentos que dividiu comigo, e tantos outros que ainda virão.

Aos meus cunhados Silvinho, Tatiana, Kelly e Sandra, pelo convívio, amizade, boas risadas, e por terem, junto com meus irmãos, me dado uma das grandes felicidades da minha vida, que são meus sobrinhos.

Aos meus sobrinhos, Mariana, Luiz Felipe, Humberto, Jamille, Marcus, Henrique, Lucas, Kaíque e Júlia, por toda a alegria que têm depositado em minha vida.

À toda minha família, em especial às tias Rosa e Catarina (*in memoriam*) e ao tio Mário (*in memorian*), por todo amor, carinho, apoio e ensinamentos.

À professora Elizabeth Pacheco Batista Fontes, pela oportunidade de orientação, pelo aprendizado e exemplo como professora e pesquisadora, pela cordialidade, e pelas ótimas confraternizações.

Ao professor Maximiller Dal-Bianco Lamas Costa, pela orientação, ensinamentos, conselhos e exemplo como professor e pesquisador, pela paciência e disponibilidade, e pela amizade.

Aos membros da banca de defesa de tese, doutor Ivan Schuster, doutora Eveline Caixeita e professor Pedro Reis, pela disponibilidade, conselhos, correções e sugestões para a melhoria desta tese, e para pesquisas futuras.

Ao doutor Newton Deniz Piovesan, pela orientação, ensinamentos e exemplo como profissional, pela amizade e pelos conselhos.

Ao pós-doutorando Rafael Delmond Bueno, pela contribuição para a realização deste trabalho, pelos conselhos e ideias, pela amizade e pelas boas prosas.

Ao professor Maurilio Alves Moreira (*in memoriam*), pela orientação no mestrado e início do doutorado, pelo exemplo de profissionalismo e força de vontade, e pelos ensinamentos e amizade que ficaram.

A todos os professores da Universidade Federal de Viçosa, que contribuíram com substancialmente para minha formação profissional e pessoal.

À Universidade Federal de Viçosa, pela formação acadêmica, profissional e pessoal, e pelos bons momentos vividos nestes 11 anos.

Ao Programa de Pós-Graduação em Genética e Melhoramento, pela formação acadêmica e pessoal, e pela oportunidade de realização do curso de doutorado.

Ao CNPq e à CAPES, pela concessão de bolsa.

Ao CNPq, FAPEMIG e Caramuru Alimentos, pelo auxílio financeiro para a execução deste trabalho.

Aos estagiários Danyelle, Guilherme, Júlia, Larissa, Letícia e Raíssa, pela contribuição na realização deste projeto, pela amizade e boa convivência.

Aos técnicos Eduardo, José Carlos, Marina, Naldo e Seu Paulo, pela contribuição e facilitação na realização deste trabalho, pela amizade e descontração.

A todos os ex-companheiros de laboratório, que contribuíram para minha formação, e pelos bons momentos vividos, em especial à Loreta, Pedro e Isadora, que contribuíram diretamente para este trabalho.

Aos companheiros dos laboratórios Proteína, Análises Bioquímicas, Enzimologia, Biologia de Populações de Patógenos e NCT BIOAGRO, pela ajuda e disponibilidade de utilização de equipamentos e reagentes.

A todos companheiros da república jogatina, em especial Cris, Daniel, Edison, Gilson, Jonatas, Leandro e Leone, por compartilhar estes quatro anos comigo, com muita amizade, cerveja e bom humor.

A todos os amigos que fiz em Viçosa, e que continuam sendo parte importante da minha vida.

E a todos aqueles que de alguma forma contribuíram para a realização deste trabalho, o meu sincero obrigado.

BIOGRAFIA

Luiz Cláudio Costa Silva, filho de Francisco Guerra Silva e Maria Eugênia da Costa Silva, nasceu em Manhuaçu, Minas Gerais, em 19 de julho de 1987. Residente em Manhuaçu, concluiu o ensino médio no Centro Educacional de Manhuaçu, em 2004.

Em maio de 2006, iniciou o curso de graduação em Agronomia na Universidade Federal de Viçosa (UFV). Graduou-se Engenheiro Agrônomo em janeiro de 2011.

Em fevereiro de 2011, iniciou o curso de mestrado em Genética e Melhoramento pela Universidade Federal de Viçosa, sob orientação do professor Maurilio Alves Moreira, submetendo-se à defesa da dissertação em 15 de março de 2013.

Em abril de 2013, iniciou o curso de doutorado em Genética e Melhoramento pela Universidade Federal de Viçosa, sob orientação do professor Maurílio Alves Moreira, e posteriormente sob orientação da professora Elizabeth Pacheco Batista Fontes, e coorientação do professor Maximiller Dal-Bianco Lamas Costa e do doutor Newton Deniz Piovesan, submetendo-se à defesa da tese em março de 2017.

SUMÁRIO

RESUMO x
ABSTRACT xii
1. INTRODUÇÃO GERAL 1
2. OBJETIVOS 5
2.1. Objetivo geral 5
2.2. Objetivos específicos 5
3. REVISÃO DE LITERATURA 5
3.1. Importância econômica da soja e do óleo de soja 5
3.2. Composição do óleo de soja 7
3.3. Biossíntese de ácidos graxos em soja 8
3.4. Avaliação da atividade enzimática de dessaturases por meio da expressão heteróloga 11
3.5. Melhoramento genético da soja para composição do óleo 14
3.6. Oligossacarídeos presentes no grão de soja 16
3.7. Melhoramento genético da soja para teor de oligossacarídeos 17
3.8. Metodologias para detecção de SNPs 18

CAPÍTULO I – CARACTERIZAÇÃO DE UM NOVO ALELO PARA O GENE *GmFAD3A* ASSOCIADO COM REDUÇÃO NO TEOR DE ÁCIDO LINOLÊNICO NO GRÃO DE SOJA 22
1. INTRODUÇÃO 23
2. MATERIAL E MÉTODOS 26
2.1. Material vegetal 26
2.2. Desenvolvimento da população segregante 26
2.3. Extração de DNA, sequenciamento e identificação de mutações 27
2.4. Genotipagem 27
2.5. RT-PCR 30
2.6. Análises fenotípicas 31
2.7. Análises estatísticas 31
2.8. Análise de atividade enzimática 32
2.9. Análise de expressão gênica 34
3. RESULTADOS 37

3.1. Caracterização dos genes *GmFAD2-1A* e *GmFAD2-1B* da linhagem FA22 37
3.2. Caracterização dos genes *GmFAD3* da variedade CS303TNKCA 37
3.3. Análises genotípicas 41
3.4. Análises fenotípicas 41
3.5. Análise de atividade enzimática 46
3.6. Análises de expressão gênica 46
4. DISCUSSÃO 49
5. CONCLUSÕES 53

CAPÍTULO II – EFEITO DE UMA MUTAÇÃO NO GENE RAFINOSE SINTASE 2 (*GmRS2*) EM CARACTERÍSTICAS DE QUALIDADE DO GRÃO DE SOJA 55
1. INTRODUÇÃO 56
2. MATERIAL E MÉTODOS 58
2.1. Material vegetal e desenvolvimento da população 58
2.2. Análises genotípicas 58
2.3. Análises fenotípicas 59
2.4. Análises estatísticas 60
3. RESULTADOS E DISCUSSÃO 60
3.1. Desenvolvimento da população e metodologia de genotipagem 60
3.2. Estatística descritiva 63
3.3. Regressão linear e correlações entre características de qualidade do grão 65
4. CONCLUSÕES 69
CONCLUSÕES GERAIS 70
REFERÊNCIAS BIBLIOGRÁFICAS 71

APÊNDICE I – MELHORAMENTO GENÉTICO PARA O DESENVOLVIMENTO DE SOJAS ESPECIAIS ATRAVÉS DE RETOCRUZAMENTOS ASSISTIDOS 88
1. INTRODUÇÃO 89
2. MATERIAL E MÉTODOS 92
2.1. Material genético 92

2.2. Extração de DNA, desenho dos conjuntos de primers e condições de HRM 93
2.3. Desenvolvimento de soja com mais de 80% de oleico 97
2.4. Programa de retrocruzamentos assistidos por marcadores 97
3. RESULTADOS E DISCUSSÃO 101
3.1. Marcadores HRM 101
3.2. Desenvolvimento de soja com mais de 80% de ácido oleico 101
3.3. Definição dos programas de retrocruzamentos assistidos por marcadores 105
3.4. Condução dos retrocruzamentos 108
4. CONCLUSÕES 112
5. PERSPECTIVAS 112
REFERÊNCIAS BIBLIOGRÁFICAS 113

RESUMO

SILVA, Luiz Cláudio Costa, D.Sc., Universidade Federal de Viçosa, março de 2017. **Caracterização molecular de alelos associados à qualidade do grão de soja**. Orientador: Elizabeth Pacheco Batista Fontes. Coorientadores: Maximiller Dal-Bianco Lamas Costa e Newton Deniz Piovesan.

A soja (*Glycine max* (L.) Merrill) é um dos mais importantes produtos do agronegócio brasileiro, apresentando como principais atrativos seus altos teores de óleo e proteína no grão. A proteína de soja apresenta grande importância histórica na pecuária brasileira, sendo ingrediente básico da maioria das rações de suínos, aves e peixes. Atualmente, a soja também é importante fonte de proteína na alimentação humana, sendo consumida na forma de diferentes produtos, alguns dos quais têm na sacarose um importante componente. O óleo de soja, por sua vez, apresenta grande importância tanto na indústria alimentícia quanto na produção de biodiesel, uma forma de combustível biodegradável cuja utilização vem aumentando consideravelmente no país devido a leis de incentivo do Governo Federal. O Programa de Melhoramento da Qualidade da Soja (PMQS/BIOAGRO/UFV) tem como objetivo o desenvolvimento de cultivares de soja especiais, visando melhor atender ao mercado de soja para alimentação humana e para produção de óleo de qualidade. O presente trabalho insere-se dentro deste contexto e foi dividido em dois capítulos. No Capítulo I foi caracterizada uma nova mutação no gene *GmFAD3A* da variedade CS303TNKCA, desenvolvida pelo PMQS e que apresenta baixo teor de ácido linolênico. Esta mutação (delA) foi avaliada em uma população F_2 segregante para teor de ácido linolênico contendo 187 plantas, e em populações $F_{2:3}$ e $F_{2:4}$, em diferentes ambientes, por meio de um marcador molecular baseado em *High Resolution Melt* (HRM). DelA explicou entre 48,81-65,97% da variação da característica, reduzindo os teores de ácido linolênico para 3,4-4,4%, se mostrando uma boa fonte de variabilidade genética para o desenvolvimento de cultivares de soja com óleo de maior qualidade. No Capítulo II, foi avaliado o efeito de uma mutação anteriormente caracterizada por outros pesquisadores no gene rafinose sintase 2 (*GmRS2)* da linhagem PI200508 sobre características de qualidade do grão de soja, pela da avaliação de uma população F_2 contendo 168 indivíduos. A mutação explicou 69,61%, 51,81% e 31,96% dos teores de estaquiose, rafinose e sacarose, respectivamente, e foi capaz de produzir soja

com 0,18% de estaquiose no grão, em média. Os baixos valores de coeficiente de determinação encontrados para teor de proteína e óleo indicam que a mutação pode ser utilizada para aumentar o teor de sacarose e reduzir os teores de rafinose e estaquiose sem mudanças significativas nos teores de óleo e proteína.

ABSTRACT

SILVA, Luiz Cláudio Costa, D.Sc., Universidade Federal de Viçosa, March, 2017. **Molecular characterization of alleles associated with soybean grain quality**. Adviser: Elizabeth Pacheco Batista Fontes. Co-advisers: Maximiller Dal-Bianco Lamas Costa and Newton Deniz Piovesan.

Soybean (Glycine max (L.) Merrill) is one of the most important products in Brazilian agribusiness, presenting high oil and protein contents as major attractives. Soy protein has a historical importance in Brazilian livestock, as the basic ingredient in pig, poultry and fish feed. Currently, soy is also an important source of protein for human consumption, in the form of different products, some of which depend on the amount of sucrose in the grain. Soybean oil, in turn, has great importance both in food industry and in the biodiesel production, a form of biodegradable fuel whose use has been increasing considerably in the country due to Federal Government's incentive laws. The purpose of the Soybean Quality Improvement Program (PMQS / BIOAGRO / UFV) is to develop special soybean cultivars, for better attend the soybean market for human consumption and for the production of high quality oil. The present work falls within this context and was divided into two chapters. In Chapter I, a new mutation in *GmFAD3A* gene of CS303TNKCA variety was characterized. This variety was developed by the PMQS and shows low linolenic acid content. This mutation (delA) was evaluated in a F_2 population containing 187 plants segregating for linolenic acid content, and in $F_{2:3}$ and $F_{2:4}$ populations, in different environments, using a molecular marker based on High Resolution Melt (HRM). DelA explained between 48.81-65.97% of the phenotype variation, reducing the linolenic acid levels to 3.4-4.4%, and was considered a good source of genetic variability for the development of soybean cultivars with a better quality soybean oil. In Chapter II, the effect of a previously characterized mutation in the raffinose synthase 2 gene (GmRS2), from lineage PI200508, on soybean quality characteristics was evaluated by an F_2 population containing 168 individuals. The mutation explained 69.61%, 51.81% and 31.96% of the stachyose, raffinose and sucrose contents, respectively, and we were able to produce soybean with 0.18% of stachyose in the grain, on average. The low coefficient of determination value found for protein and oil content indicates that

the mutation can be used to increase sucrose and reduce raffinose and stachyose contents without significant changes in oil and protein.

1. INTRODUÇÃO GERAL

A soja (*Glycine max* (L.) Merrill) é uma das principais culturas brasileiras, sendo responsável por cerca da metade da produção de grãos do país. O Brasil é o segundo maior produtor de soja no mundo, correspondendo por aproximadamente um terço de toda a produção mundial, sendo o estado do Mato Grosso o seu principal produtor (CONAB 2017). Os principais atrativos para a produção de soja são seus altos teores de óleo e proteína. Além disso, a soja é uma boa fonte de minerais como ferro, potássio, magnésio, zinco, cobre, fósforo, manganês e vitaminas do complexo B, e contém ainda isoflavonas, saponinas, fitatos, fitosteróis, peptídeos com baixo peso molecular, sacarose e ácidos graxos poli-insaturados (Carrão-Panizzi et al. 1998).

Apesar de ser conhecida e consumida pela civilização oriental por milhares de anos, a soja só foi introduzida na Europa no fim do século XV, mas apenas em jardins botânicos. Até o final do século XIX, a produção de soja era restrita à China. Somente no começo do século XX os altos teores de óleo e proteína da soja começaram a despertar o interesse das indústrias mundiais (EMBRAPA – https://www.embrapa.br/web/portal/soja/cultivos/soja1/historia).

No Brasil, cultivos experimentais de soja se iniciaram na Bahia em 1882, segundo apontam alguns registros históricos. Entretanto, o ano de 1901 é considerado o marco da introdução da soja no país, quando começaram cultivos na Estação Agropecuária de Campinas, e a distribuição de sementes para produtores paulistas. Em 1914, a soja foi introduzida oficialmente no Rio Grande do Sul, estado que apresenta condições climáticas similares às regiões produtoras do grão nos Estados Unidos. Até a década de 70, as cultivares brasileiras eram importados deste país (APROSOJA – http://www.aprosoja.com.br/soja-e-milho/a-historia-da-soja).

A partir da década de 70, o cultivo da soja começou a se expandir no Brasil, período onde a indústria do óleo iniciou sua ampliação no país, acompanhado por um cenário mundial de aumento na demanda pelo grão, com consequente aumento em seu preço. Neste período, a soja se tornou o principal produto agrícola nacional, passando de uma produção de 1,5 milhões de toneladas em 1970 para 15 milhões em 1979. Este aumento foi alcançado em parte pelo aumento na área plantada, mas

também pelo aumento da produtividade média (de 1,14 t/ha para 1,74 t/ha; APROSOJA – http://www.aprosoja.com.br/soja-e-milho/a-historia-da-soja).

Os investimentos em pesquisas visando ao desenvolvimento de novas tecnologias adaptadas às condições brasileiras levaram à "tropicalização" da soja, permitindo expandir seu plantio para regiões de baixas latitudes, incluindo o cerrado brasileiro. Atualmente, a produtividade média nacional está próxima de 3,0 t/ha, com uma produção na ordem de 95 milhões de toneladas na safra 2015/2016 (CONAB 2017). Neste sentido, o melhoramento genético teve parcela relevante, por meio do desenvolvimento de cultivares adaptadas ao clima e ao solo brasileiros, aumentando seu potencial produtivo.

O grão de soja é composto, em média, por 20% de lipídeos, 40% de proteínas, 34% de carboidratos e 5% de minerais (EMBRAPA – http://www.cnpso.embrapa.br/soja_alimentacao/index.php?pagina=23), sendo proteína e óleo seus componentes mais importantes (Kim et al. 2006, Yang et al. 2014). A soja é uma das principais fontes proteicas de origem vegetal, apresentando composição de aminoácidos razoavelmente balanceada, sendo deficiente apenas em aminoácidos sulfurados, como metionina e cisteína. Dentre os subprodutos da soja, o farelo é o que tem sido mais utilizado para a alimentação animal, devido ao seu alto teor proteico, boa aceitação pelos animais e facilidade de obtenção no mercado. Com relação à alimentação humana, diversos alimentos são produzidos a partir da soja, como a farinha de soja, proteína texturizada, isolado proteico, lecitina, leite, iogurte, tofu, natto, miso, além de ser consumida também *in natura* em saladas e edamame ("soja-vagem"; Sá et al. 2006).

O uso da soja na alimentação dos brasileiros vem aumentando devido à maior divulgação dos benefícios da soja para a saúde humana e do crescimento da oferta de produtos à base de soja de melhor qualidade. Além de boa fonte de proteína, seu consumo ajuda a melhorar os níveis sanguíneos de colesterol, e se mostra uma boa alternativa para pessoas com intolerância à lactose (Carrão-Panizzi & Silva 2011). Entretanto, o valor nutricional da soja é limitado pela presença de diversos fatores antinutricionais, como inibidores das proteases tripsina e quimiotripsina, compostos fenólicos (taninos), lectinas e oligossacarídeos (Stech et al. 2011, Yang et al. 2014).

Dentre os oligossacarídeos presentes no grão de soja, apenas a sacarose é nutricionalmente útil para a digestão de humanos e animais monogástricos, e seu teor no grão é fator crítico para a produção de vários derivados de soja, como leite de soja, tofu, miso e natto (Sato et al. 2014, Wang et al. 2014). Em contrapartida, os oligossacarídeos de rafinose (rafinose e estaquiose) são açúcares não completamente digeríveis por humanos e animais monogástricos, e seu consumo pode causar disfunções como diarreia, náuseas e flatulência (Liener 1994, Reddy et al. 2016). Deste modo, a presença de estaquiose e rafinose é um fator restritivo para o consumo de soja como fonte proteica, e vários processos têm sido descritos na literatura visando remover ou metabolizar estes compostos para reduzir seus efeitos.

Outro componente importante do grão de soja, o óleo de soja é o segundo óleo vegetal mais consumido no mundo, representando quase um terço do mercado mundial (USDA 2016a). No Brasil, sua importância é ainda maior, correspondendo a mais de 80% do mercado de óleos vegetais na indústria alimentícia. Neste setor, o óleo de soja é processado em três produtos básicos: óleo refinado comestível (matéria-prima de margarinas, óleo de cozinha, maionese e temperos, gordura vegetal e produtos farmacêuticos, como os repositores hormonais); óleo refinado para fins não alimentícios (matéria-prima para produção de biodiesel, além de ingrediente de velas, sabões, tintas, plásticos, lubrificantes, desinfetantes e inseticidas); e lecitina (usada em produtos químicos, cosméticos e têxteis, alguns alimentos e sorvetes) (AGROLINK – https://www.agrolink.com.br/culturas/soja/comercializacao.aspx). Dentre estes produtos, o biodiesel, um combustível biodegradável e renovável capaz de reduzir a poluição ambiental, tem apresentado importância crescente no mercado brasileiro. Segundo a Lei nº 13.263 de 23 de março de 2016, que altera a Lei nº 13.033 de 24 de setembro de 2014, todo o óleo diesel comercializado no Brasil deve apresentar 8% de biodiesel a partir de 23 de março de 2017, e este valor deve chegar a 10% em 2019 (BRASIL 2016). Deste modo, é esperado que a demanda por óleo de soja continue crescendo nos próximos anos.

A qualidade do óleo e suas propriedades físicas e químicas, tais como ponto de fusão e estabilidade oxidativa, são determinados pelos ácidos graxos que o compõe (Fehr 2007). O óleo de soja possui em média 13% de ácido palmítico, 4% de ácido esteárico, 20% de ácido de oleico, 55% de ácido linoleico e 8% de ácido

linolênico (Pham et al. 2010, Goettel et al. 2014). Ácidos graxos poli-insaturados (ácidos linoleico e linolênico) apresentam baixa estabilidade oxidativa, o que pode levar ao processo de rancificação do óleo. A rancificação é responsável por alterações no sabor e odor do óleo, diminuindo sua aceitação e tempo de armazenamento (Warner & Fehr 2008). Para reduzir os teores de ácidos graxos poli-insaturados e aumentar o teor de ácido oleico, o óleo de soja passa por um processo de hidrogenação química. Entretanto, este processo leva à formação de 10-40% de gorduras *trans*, moléculas cujo consumo tem sido associado ao aumento do risco de diabetes e doenças coronárias (Mozaffarian et al. 2006, Zaloga et al. 2006). Recentemente, o FDA (*Food and Drug Administration*), agência responsável pela regulamentação de alimentos e remédios nos EUA, emitiu uma determinação de que óleos parcialmente hidrogenados não são reconhecidos como seguros (FDA 2015), e estabeleceu o prazo de três anos para que os produtores de alimentos se encaixem nesta medida ou façam uma requisição para o uso destes óleos (FDA 2016). Esta medida provavelmente deve se estender para outros países em breve, incluindo o Brasil.

A soja é considerada um alimento funcional, devido aos vários compostos benéficos encontrados em seu grão. Entretanto, a presença de certos compostos, como os oligossacarídeos de rafinose, ácidos graxos poli-insaturados e inibidores de proteases, restringem e interferem em sua utilização como alimento e como matéria prima para a indústria, demandando esforços e recursos para inativá-los, extraí-los ou convertê-los em formas menos danosas. Neste sentido, o melhoramento genético tem se mostrado uma ferramenta importante para melhorar a genética das cultivares e torná-las mais bem adaptadas às necessidades do mercado. O desenvolvimento de variedades de soja com maior qualidade do grão, como composição do óleo mais adequada, altos teores de proteína e sacarose, e baixos teores de oligossacarídeos de rafinose, é de grande importância para atender os vários nichos do mercado da soja, como a indústria alimentícia, produção animal e indústria do biodiesel, além de diminuir tempo e custos demandados para o beneficiamento do grão e derivados.

Neste trabalho, foram caracterizadas mutações em genes controlando características de qualidade do grão de soja. Estas mutações se mostram

promissoras para o desenvolvimento de variedades de soja que melhor atendam ao mercado de soja na alimentação e à indústria de biocombustíveis.

2. OBJETIVOS

2.1. Objetivo geral

O objetivo deste trabalho foi a caracterização de alelos associados com características especiais do grão de soja para a indústria e para o consumo humano.

2.2. Objetivos específicos

- Caracterizar os genes *GmFAD2* e *GmFAD3* da variedade CS303TNKCA e da linhagem FA22;
- Quantificar a associação de uma nova mutação no gene *GmFAD3A* da variedade CS303TNKCA com baixo teor de ácido linolênico;
- Analisar a expressão dos genes *GmFAD2-1* e *GmFAD3* em sementes em desenvolvimento;
- Analisar a atividade enzimática da dessaturase codificada pelo gene *GmFAD3A* da variedade CS303TNKCA;
- Desenvolver diagnóstico molecular para uma mutação no gene *GmRS2*, associada com redução dos teores de rafinose e estaquiose
- Avaliar o efeito da mutação no gene *GmRS2* em características de qualidade do grão;
- Iniciar um programa de retrocruzamentos assistidos por marcadores visando o desenvolvimento de variedades de soja com características especiais.

3. REVISÃO DE LITERATURA

3.1. Importância econômica da soja e do óleo de soja

A soja (*Glycine max* (L.) Merrill) é uma leguminosa amplamente cultivada no mundo. Apresenta alta produtividade, custo de produção relativamente baixo, e altos teores de óleo e proteína (Bilyeu & Wiebold 2016). Na safra 2015/2016 foram cultivados no mundo cerca de 120,17 milhões de hectares de soja, com uma produção de cerca de 315,86 milhões de toneladas (USDA 2016b). O Brasil é o

segundo maior produtor mundial de soja, atrás apenas dos EUA, e produziu cerca de 95,43 milhões de toneladas nesta safra, com produtividade média de 2,87 t/ha, em uma área de 33,25 milhões de hectares (57,02% da área cultivada em grãos no país). A queda na produtividade, em relação à safra 2014/2015 se deveu principalmente ao clima, que provocou atrasos no plantio e comprometeu lavouras em diversos estados. O estado do Mato Grosso é o maior produtor nacional, com 26,03 milhões de toneladas (CONAB 2017).

Os principais subprodutos da soja são o óleo e o farelo de soja. O farelo de soja apresenta alto teor proteico, sendo utilizado principalmente na alimentação animal. Em 2015, foram produzidos cerca de 30,76 milhões de toneladas de farelo de soja no Brasil. Deste montante, cerca de 52% foram consumidos internamente, e 48% foram exportados (ABIOVE 2017a). Para 2016, a estimativa é de uma produção de 29,60 milhões de toneladas, fruto da queda na produção da safra 2015/2016 (CONAB 2017).

O óleo de soja ocupa a segunda posição no ranking mundial entre os óleos vegetais em termos de produção e consumo doméstico. Na safra 2015/2016 foram produzidos no mundo 52,01 milhões de toneladas de óleo de soja, correspondendo a 29% do óleo vegetal consumido no mundo, ficando atrás apenas do óleo de palma (USDA 2016a). Em 2015, o Brasil produziu cerca de 8,07 milhões de toneladas de óleo de soja, com cerca de 6,52 milhões de toneladas sendo consumidas internamente, e 1,67 milhões de toneladas exportadas. Para 2016, as estimativas são de uma produção de 7,75 milhões de toneladas de óleo de soja (ABIOVE 2017a, CONAB 2017). O óleo de soja contabiliza mais de 80% do mercado de óleos vegetais para a indústria alimentícia no Brasil, e é processado em três produtos básicos: óleo refinado comestível (matéria-prima de margarinas, óleo de cozinha, maionese e temperos, gordura vegetal e produtos farmacêuticos, como os repositores hormonais); óleo refinado para fins não alimentares (matéria-prima para produção de biodiesel, além de ingrediente de velas, sabões, tintas, plásticos, lubrificantes, desinfetantes e inseticidas); lecitina (usada em produtos químicos, cosméticos e têxteis, alguns alimentos e sorvetes) (AGROLINK – https://www.agrolink.com.br/culturas/soja/comercializacao.aspx).

Dentre os muitos usos do óleo de soja, tem se destacado a produção de biodiesel, um combustível renovável, biodegradável e ambientalmente correto, obtido a partir da transesterificação de triacilglicerídeos, e que pode substituir o óleo diesel mineral (Parente 2003). Segundo a Lei nº 13.263 de 23 de março de 2016, o óleo diesel comercializado no Brasil deve conter 8% de biodiesel a partir de março de 2017, valor que deve chegar a 10% em 2019 (BRASIL 2016). Em 2016, foram produzidos 3,81 milhões de metros cúbicos de biodiesel no Brasil, acumulando um crescimento de 942% desde o ano de 2007 (ABIOVE 2017b). Desta maneira, a demanda por biodiesel e consequentemente de óleo de soja deve continuar aumentando nos próximos anos.

3.2. Composição do óleo de soja

O óleo de soja é composto basicamente por triglicerídeos dos seguintes ácidos graxos: 13% de ácido palmítico (16:0); 4% de ácido esteárico (18:0); 20% de ácido de oleico (18:1 Δ^{9}); 55% de ácido linoleico (18:2 $\Delta^{9,12}$); 8% de ácido linolênico (18:3 $\Delta^{9,12,15}$; Pham et al. 2010, Goettel et al. 2014). A qualidade do óleo e suas propriedades físicas e químicas, tais como ponto de fusão e estabilidade oxidativa, são determinados pelo balanço destes ácidos graxos (Fehr 2007). Elevados teores de ácidos graxos poli-insaturados (ácidos linoleico e linolênico) são os principais responsáveis pela alta instabilidade oxidativa do óleo de soja, o que pode levar a rancificação do óleo. Este processo é responsável por alterar sabor e odor, e diminuir seu tempo de armazenamento (Warner & Fehr, 2008).

O processo de hidrogenação química é utilizado para reduzir os teores de ácidos graxos poli-insaturados e aumentar o teor de ácido oleico no óleo de soja. Entretanto, este processo leva à formação de 10-40% de ácidos graxos *trans*, moléculas cujo consumo tem sido associado ao aumento do risco de obesidade e doenças coronárias, dentre outras patologias (Mozaffarian et al. 2006; Zaloga et al. 2006). Recentemente, o FDA (*Food and Drug Administration*), agência que regula alimentos e remédios nos EUA, estabeleceu o prazo de três anos para a remoção de produtos alimentícios que contenham gordura *trans* do mercado norte americano (FDA 2015). E esta medida provavelmente deve se estender para outros países em breve, incluindo o Brasil.

A presença de ácidos graxos poli-insaturados e sua maior susceptibilidade à degradação oxidativa é também um entrave para produção de biodiesel. A auto oxidação do biodiesel ocorre pela reação em cadeia de radicais livres entre o biodiesel e o oxigênio da atmosfera, produzindo peróxidos, hidroperóxidos e ácidos carboxílicos, dentre outros compostos (Vieira & Pasquini 2014). A degradação oxidativa do biodiesel leva a problemas como aumento da viscosidade, elevação da acidez e formação de gomas. Além disso, os compostos formados podem ser corrosivos para as câmaras dos motores e podem levar ao entupimento de bombas de injeção e filtros (Tang et al. 2008, Dantas et al. 2011, Teleken et al. 2015). Malafaia e Souza (2014) avaliaram a estabilidade oxidativa do biodiesel de óleo de soja bruto, e encontraram um período de indução de 5,10 h, por meio do método de Rancimat. O período de indução é o tempo gasto para a detecção de ácidos orgânicos voláteis produzidos a partir do processo de oxidação do biodiesel. Este valor abaixo do período mínimo aceitável de 6 horas (Lôbo & Ferreira 2009) está relacionado com a baixa estabilidade oxidativa do óleo de soja bruto.

Apesar de ser considerado um óleo de qualidade, o óleo de soja necessita ser beneficiado para melhor atender ao mercado alimentício e de biocombustíveis. Neste sentido, o melhoramento genético visando alteração na composição de ácidos graxos apresenta grande potencial para a produção de óleo com maior qualidade, reduzindo gastos e efeitos indesejados gerados pelo beneficiamento.

3.3. Biossíntese de ácidos graxos em soja

A biossíntese de ácidos graxos poli-insaturados em sementes de soja envolve as enzimas ácido graxo dessaturases (no inglês *fatty acid desaturases* – FAD), principalmente as do retículo endoplasmático (microssoma). Estas enzimas introduzem uma ligação dupla entre dois carbonos de um ácido graxo, removendo dois átomos de hidrogênio, numa reação que requer oxigênio molecular, NAD(P)H, um sistema de transporte de elétrons (ferredoxina NADPH redutase e ferredoxina, ou citocromo *b5* redutase e citocromo b_5) e um terminal de dessaturase (Nakamura & Nara 2004). Os principais substratos para estas enzimas são os resíduos de ácidos graxos ligados à fosfatidil colina (PC) ou a coenzima A (CoA). A ω6-dessaturase

converte oleoil em linoleoil, enquanto a ω3-dessaturase é responsável pela conversão de linoleoil em linolenil (Zhang et al. 2012).

Análises das sequências de aminoácidos indicam que as dessaturases ligadas à membrana contém dois longos domínios hidrofóbicos com potencial de atravessar a dupla camada lipídica por duas vezes. A comparação de sequências também revelou a existência de três motivos conservados de histidina (H-box) que contém oito resíduos de histidina: $HX_{3-4}H$, $HX_{2-3}HH$, e $H/QX_{2-3}HH$. Estes H-boxs são possíveis ligantes para átomos de ferro, atuando no sítio catalítico das enzimas (Shanklin et al. 1994). Alterações em apenas um único resíduo nestes boxes de histidina podem reduzir a atividade da enzima (Bilyeu et al. 2006, Radovanovic et al. 2014, Reinprecht & Pauls 2016). A Figura 1 apresenta o esquema de uma dessaturase ligada à membrana do fungo *Claviceps purpurea*, adaptado de Meesapyodsuk et al. (2007).

Heppard et al. (1996) descreveram a existência de dois tipos de genes codificando para ω6-dessaturases do retículo endoplasmático em soja, *GmFAD2-1* e *GmFAD2-2*. *GmFAD2-1A* (Glyma.10G278000) e *GmFAD2-1B* (Glyma.20G111000) são expressos principalmente em sementes em desenvolvimento, codificando as ω6-dessaturases que controlam os níveis de ácido oleico no óleo (Figura 2). Apesar de estas enzimas diferirem entre si por apenas 24 resíduos, a enzima *GmFAD2-1B* é mais estável que *GmFAD2-1A*, que é mais susceptível à degradação pelo proteossomo, principalmente em altas temperaturas (Tang et al. 2005). Já os genes *GmFAD2-2A* (Glyma.19g147300), *GmFAD2-2B* (Glyma.19g147400) e *GmFAD2-2C* (Glyma.03g144500) são expressos constitutivamente em todos os tecidos e órgãos vegetais (Schlueter et al. 2007). Li et al. (2007) caracterizaram uma nova isoforma de ω6-dessaturase, denominada *GmFAD2-3*. Análises *in silico* e de RT-PCR mostraram que este gene é expresso tanto em sementes em desenvolvimento quanto em tecidos vegetativos.

Em grãos de soja, três genes *GmFAD3* que contribuem para o conteúdo de ácido linolênico foram caracterizados (Figura 1; Bilyeu et al. (2003). Os genes *GmFAD3A* (Glyma.14G194300) e *GmFAD3B* (Glyma.02G227200) possuem 94% de identidade na sequência das regiões codificadoras. No entanto, os níveis de transcritos são distintos nos diferentes estádios de desenvolvimento e tecidos de

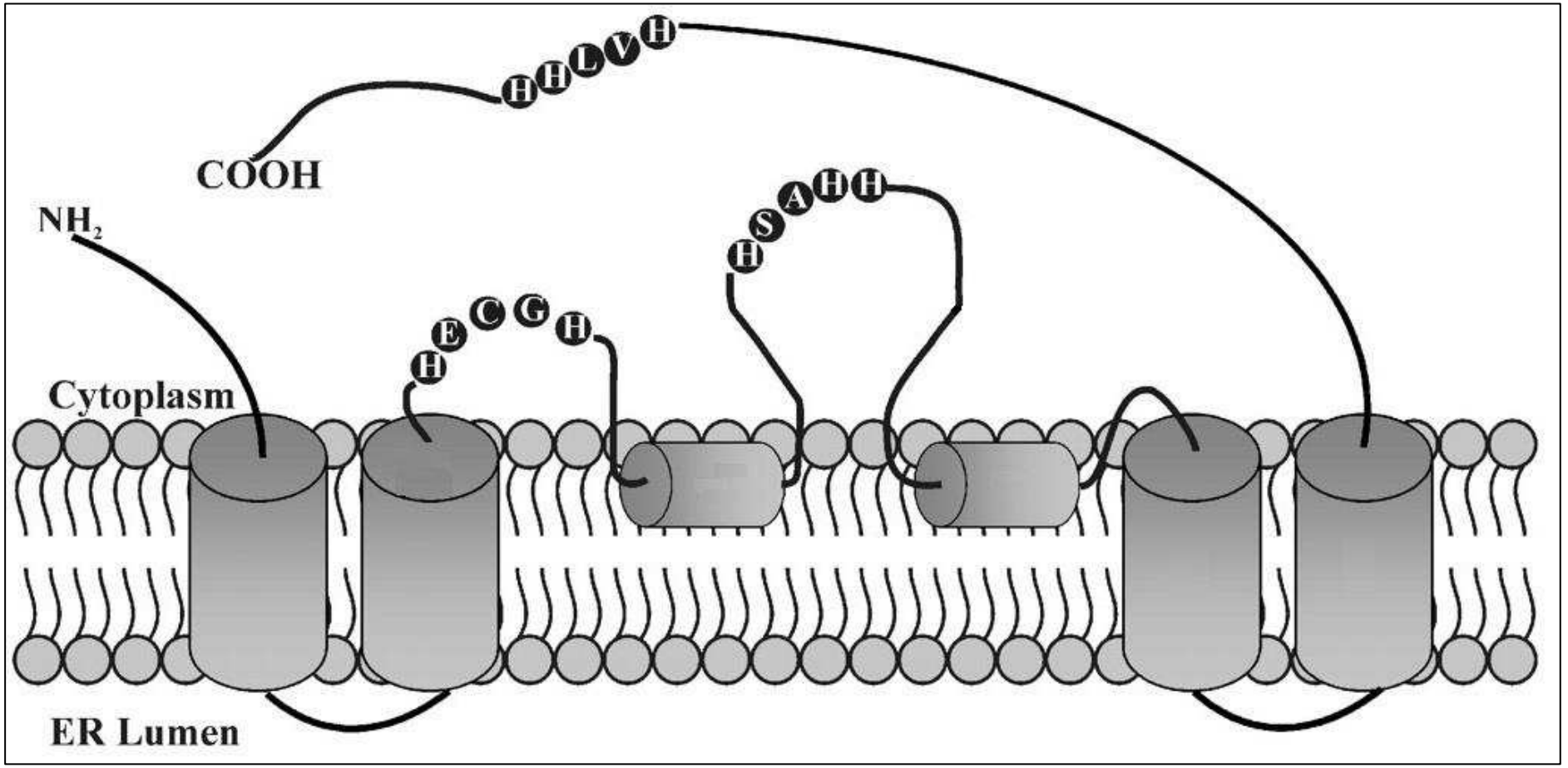

Figura 1 – Modelo de topologia de membrana da dessaturase *CpDesX* de *Claviceps purpúrea*. NH_2 e COOH indicam as extremidades amino-terminal e carboxi-terminal, respectivamente. Os três boxes de histidina são mostrados no lado citosólico da membrana. Adaptado de Meesapyodsuk et al. (2007)

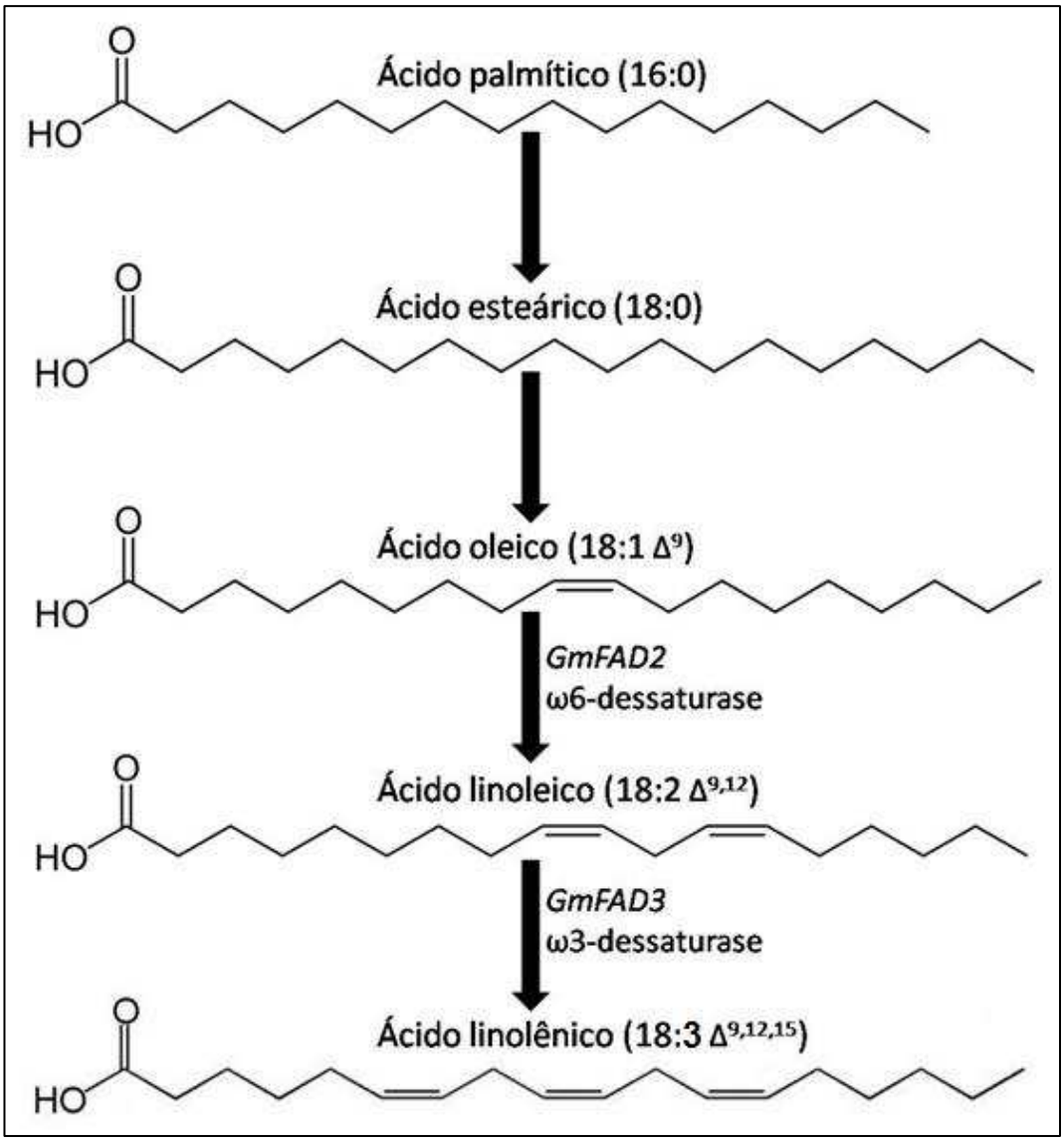

Figura 2 – Biossíntese de ácidos graxos poli-insaturados.

soja. O *GmFAD3A* é altamente expresso nos grãos, e apresenta o maior efeito sobre o teor de ácido linolênico (Bilyeu et al. 2005, Pinto et al. 2013), enquanto *GmFAD3B* e *GmFAD3C* (Glyma.18G062000) apresentaram menores níveis de expressão neste órgão.

3.4. Avaliação da atividade enzimática de dessaturases por meio da expressão heteróloga

Dessaturases desempenham papel importante na composição do óleo de espécies oleaginosas, tais como a soja, a canola e o dendê. A atividade de enzimas

como a ω6-dessaturase e a ω3-dessaturase estão diretamente ligadas com o acúmulo de ácidos graxos poli-insaturados, que conferem uma menor estabilidade para o óleo vegetal. A oxidação destes ácidos graxos resulta no acúmulo de odores e sabores indesejáveis, reduzindo a palatabilidade e qualidade alimentar (Reinprecht & Pauls, 2016). Por outro lado, ácidos graxos poli-insaturados, como o ácido linolênico, são elementos essenciais na dieta humana, e uma maior ingestão deste ácido reduz significantemente o risco de depressão e câncer de mama (Lucas et al. 2011). Além disso, o ácido linolênico é um precursor de ácidos graxos poli-insaturados de cadeia longa, como os ácidos eicosapentaenoico e docosahexaenoico, cujo consumo está associado com redução do risco de doenças cardiovasculares, além de ter efeitos benéficos no desenvolvimento inicial do sistema visual e do cérebro durante a gravidez (Ha et al. 2014). Assim, o entendimento da função e atividade das dessaturases é alvo de estudos de vários grupos de pesquisa no mundo, que têm empregado diferentes metodologias, como expressão heteróloga, mutação induzida e análise de atividade enzimática *in vitro*.

A expressão heteróloga tem sido uma das principais metodologias para estudo da atividade de dessaturases. Consiste na expressão de um gene ou parte de um gene em um organismo hospedeiro, que geralmente não os possui. A inserção do gene no hospedeiro geralmente é realizada por meio de transformação genética utilizando DNA recombinante, e a atividade da enzima pode ser avaliada por meio de mudanças fenotípicas nos organismos transformados. A levedura *Saccharomyces cerevisiae* tem sido amplamente utilizada para expressão heteróloga de dessaturases visando à avaliação de atividade enzimática. Estas leveduras não produzem ácidos graxos poli-insaturados em condições normais (Tang et al. 2005, Zhang et al. 2009, Shi et al. 2012, Sun et al. 2016).

No intuito de descobrir as diferenças entre as isoformas das ω3-dessaturases microssomais *GmFAD3A-2* e *GmFAD3C*, que contribuem para o conteúdo de ácido linolênico em sementes em desenvolvimento de soja, Zhang et al. (2009) clonaram estes genes de uma cultivar de soja e os inseriram em *S. cerevisiae* cepa K601. As leveduras transformadas foram crescidas e tiveram seu perfil de ácidos graxos analisado por cromatografia gasosa. Leveduras expressando *GmFAD3A-2* e *GmFAD3C* produziram 0,77 e 4,13% de ácido linolênico, respectivamente, ácido

não detectado nas leveduras sem os genes inseridos, implicando *GmFAD3C* como uma enzima de maior atividade ou estabilidade em comparação à *GmFAD3A-2*.

Heppard et al. (1996) descreveram a existência de dois tipos de genes codificando para ω6-dessaturases do retículo endoplasmático em soja, *GmFAD2-1* e *GmFAD2-2*. Ambos os genes ω6-dessaturases são capazes de produzir ácido linoleico quando expressos em *S. cerevisiae* (Tang et al. 2005). Neste estudo, os autores avaliaram a influência da temperatura nas duas isoformas, e constataram que colônias contendo a enzima FAD2-1B apresentaram maior produção de ácido linoleico nas temperaturas de 20 e 30 ºC. Com o aumento da temperatura de 20 para 30 ºC, foi verificado que a queda na atividade das enzimas foi maior para colônias contendo a isoforma FAD2-1A (de 18% para 4,5%) e menor para aquelas contendo FAD2-1B (de 30% para 18%). Assim, os autores concluíram que a isoforma FAD2-1A é uma enzima mais instável que FAD2-1B, principalmente em altas temperaturas.

Um nova isoforma de ω6-dessaturase, denominada *GmFAD2-3*, foi identificada em soja e caracterizada por meio de expressão em leveduras (Li et al. 2007). Colônias de leveduras *S. cerevisiae* cepa INVSc1 transformadas com plasmídeo contendo o gene *GmFAD2-3* acumularam uma quantidade considerável de ácido linoleico, sugerindo que o gene codifica uma enzima funcional.

Iskandarov et al. (2010) estudaram a alga verde *Parietochloris incisa*, um tipo de alga que acumula grandes quantidades de ácido araquidônico (um ácido graxo poli-insaturado de cadeia longa), e isolaram 3 cDNAs (*PiDesD12*, *PiDesD6*, *PiDesD5*). Por meio da expressão heteróloga em *S. cerevisiae* cepa W303, os autores confirmaram que *PiDesD6* e *PiDesD5* codificam para dessaturases de membrana com atividade Δ6 e Δ5, respectivamente.

A expressão heteróloga é uma técnica que tem sido utilizada com êxito para estudo de diversas enzimas, incluindo as dessaturases de membrana. Apresenta como principais vantagens: utiliza técnicas relativamente simples, como clonagem molecular, transformação genética e cultivo de células; permite analisar diferentes isoformas de uma mesma enzima; não necessita de purificação da enzima ou obtenção de homogenatos; e não utiliza substratos marcados radioativamente, como as análises *in vitro*.

3.5. Melhoramento genético da soja para composição do óleo

A principal meta do melhoramento da qualidade do óleo de soja é o aumento da sua estabilidade oxidativa, por meio da redução do teor de ácido linolênico e aumento do teor de ácido oleico. A herança do conteúdo de ácidos graxos poli-insaturados e monoinsaturados em cruzamentos entre parentais não mutantes normalmente é de natureza quantitativa (Burton et al. 1983). Entretanto, pelo efeito de genes maiores, herança simples já foi constatada em cruzamentos envolvendo linhagens mutantes, com teores modificados de ácidos graxos (Rahmam et al. 1996).

ROSS et al. (2000) descreveram o mais baixo teor de ácido linolênico encontrado em grão de soja para a linhagem A29 (1%). Esta linhagem combina as mutações *fan1* (acesso A5), *fan2* (acesso A23) e *fan3* (acesso gerado por mutação induzida da linhagem A89-144003), localizadas nos genes *GmFAD3A*, *GmFAD3B* e *GmFAD3C*, respectivamente. Pham et al. (2014) caracterizaram a mutação *fan1* como sendo uma deleção de 6.372 pb, que coincide em 4.002 pb com o gene *GmFAD3A*, restando apenas 179 pb da região 3' UTR do gene. Bilyeu et al. (2006) identificaram um SNP (G→A) na extremidade 3' do sítio de *splicing* que precede o éxon 4 do gene FAD3B do acesso M23. Esta mutação acarreta em *splicing* impreciso no mRNA desta região, levando a alteração no quadro de leitura e terminação prematura da síntese da enzima. Ainda neste trabalho, os autores caracterizaram um SNP no gene *GmFAD3C* de *fan3* (C910T), que resultou em uma troca de histidina por tirosina no resíduo 304. Esta posição coincide com a região rica em histidina II da sequência da ϖ-3 dessaturase, que foi caracterizada anteriormente como sendo essencial para a atividade da enzima (Shanklin et al. 1994).

Outros autores descreveram mutações no gene *GmFAD3A* associadas com baixo teor de ácido linolênico. Reinprecht et al. (2009) caracterizaram uma troca (G>A) na posição 798 da linhagem RG10, levando a um stop códon prematuro, e uma enzima faltando 111 resíduos, incluindo o terceiro box de histidina. Uma mutação encontrada no primeiro par de bases do íntron 6 da linhagem CX1512-44 foi identificada por Bilyeu et al. (2005), levando a alteração no sítio de *splicing*.

Kim et al. (2015) identificaram uma troca (G>A) no segundo sítio de *splicing* da linhagem PE1690, levando a um stop códon prematuro.

Quanto ao aumento no teor de ácido oleico, Pham et al. (2010) produziram linhagens de soja com 80% de ácido oleico por meio da contribuição de alelos mutantes nos genes *GmFAD2-1A* (linhagens 17D e M23) e *GmFAD2-1B* (linhagem PI283327). A deleção do gene *GmFAD2-1A* da linhagem M23 foi descrita por Sandhu et al. (2007), por meio de análises por PCR e DNA blot. Pham et al. (2010) caracterizaram uma mutação missense na linhagem PI283327 (C>G) na posição 410 da sequência codificante do gene *GmFAD2-1B*, resultando na troca de um resíduo de prolina (apolar) por outro de arginina (básico) em uma posição perfeitamente conservada. Esta troca afeta a estrutura da proteína e/ou sua função, de acordo com o *software* de predição de efeito de SNPs PolyPhen. Posteriormente, Pham et al. (2011) utilizaram o cruzamento entre as linhagens PI603452 e PI283327, para obter resultados similares. Neste trabalho, os autores identificaram a deleção de uma adenosina na posição 544/545 em uma região de éxon do gene *GmFAD2-1A* da PI603452, resultando em alteração no quadro de leitura e terminação prematura do peptídeo, após 191 aminoácidos.

A combinação de mutações nos genes *GmFAD2-1* e *GmFAD3* tem sido estudada, visando produzir sementes com alto teor de ácido oleico e baixo teor de ácido linolênico. A inserção dos genes mutantes *GmFAD2-1A* (obtido da variedade M23), *GmFAD3-A* (obtido da variedade M24) e *GmFAD3-B* (obtido da variedade M5) levou `a produção de variedades com aproximadamente 55% de ácido oleico e 4,2% de ácido linolênico (Rahman et al. 2001). O silenciamento transcricional dos genes *GmFAD2-1A* e *GmFAD2-1B* e mutações nos genes *GmFAD3-A* (da variedade C1640) e *GmFAD3-B* (da variedade RG10) resultaram em 77-79 % de ácido oleico e 2,1-2,8% de ácido linolênico (Brace et al. 2011). Pham et al. (2012) foram capazes de obter soja com uma média de 10,5% de ácidos graxos saturados, 84,9% de ácido oleico, 2,9% de ácido linoleico, e 1,7% de ácido linolênico, combinando mutações nos genes *GmFAD2-1A*, *GmFAD2-1B*, *GmFAD3A* e *GmFAD3C*. É importante salientar que pelos estudos de Pham et al. (2012) ficou claramente evidenciado que, apesar de mutações nos genes *GmFAD3A* e *GmFAD3C*, o conteúdo de ácido linolênico variou conforme o meio ambiente. Estes resultados sugerem que serão necessários estudos da variabilidade ambiental para

esta característica, assim como direcionar futuramente uma pesquisa básica para entender porque este gene depende mais de interações com o meio ambiente quando comparado com o *GmFAD2-1*.

O desenvolvimento de soja com teor de ácidos graxos modificado pode beneficiar as aplicações industriais, além de fornecer uma alternativa mais barata e igualmente funcional ao azeite de oliva no setor alimentício (Pham et al. 2012). Atualmente, grandes empresas também têm empregado esforços neste sentido. A Monsanto desenvolveu a cultivar Vistive Gold (MON 87705), uma soja geneticamente modificada com alto teor de oleico e baixo teor de linolênico, por meio do silenciamento de genes *GmFatB* e *GmFAD2* por RNA de interferência. Já a Dupont Pioneer possui a cultivar Plenish, também desenvolvida por modificação genética, com alto oleico e cerca de 3% de ácido linolênico. O desenvolvimento de cultivares de soja com altos teores de ácido oleico e baixos teores de ácido linolênico, adaptados às condições de plantio do território brasileiro é de grande importância para a indústria alimentícia e de biodiesel, de modo que será importante produzir óleo de melhor qualidade nutricional, mais estável e com menor custo de beneficiamento.

3.6. Oligossacarídeos presentes no grão de soja

O grão de soja apresenta cerca de 34% de carboidratos, a maioria deles sendo polissacarídeos não-amídicos (celulose, hemicelulose e pectina; Kanchana et al. 2015). Sacarose, rafinose e estaquiose são os principais carboidratos solúveis presentes no grão de soja, e são fatores responsáveis pela qualidade nutricional do grão, sendo considerados como potenciais contribuintes para aumentar sua energia metabolizável (Bilyeu & Wiebold 2016). A sacarose é o único oligossacarídeo nutricionalmente útil para digestão de monogástricos (Yang et al. 2014), e seu teor no grão geralmente varia entre 2,9 e 5,6% em linhagens e cultivares de soja nos EUA (Sato et al. 2012). O teor de sacarose é fator crítico na elaboração de produtos à base de soja, tais como leite de soja, tofu, miso e natto (Geater et al. 2000, Kumar et al. 2010, Sato et al. 2014, Wang et al. 2014). Por outro lado, a presença de oligossacarídeos de rafinose (ROs), que inclui os α-galactosídeos de rafinose e estaquiose, é considerada um fator antinutricional. A ingestão destes

oligossacarídeos por humanos e animais monogástricos pode causar disfunções orgânicas, tais como diarreia, náuseas e flatulência (Liener 1994, Wang et al. 2014, Reddy et al. 2016), devido à ausência de atividade α-galactosidase nestes organismos (Hagely et al. 2013; Yang et al. 2014). A estaquiose é o segundo principal açúcar solúvel no grão de soja, comumente variando entre 1,4-4,1% (Hymowitz et al. 1972; Zeng et al. 2015), e assume papel restritivo quanto ao consumo do grão de soja como fonte proteica.

Vários procedimentos visando reduzir ou eliminar os teores de estaquiose em grãos de soja e produtos derivados estão descritos na literatura, tais como: embebição e germinação (Viana et al. 2005); processos de fermentação (Mital & Steinkraus, 1975); extração de oligossacarídeos em água (Ku et al. 1976); ultrafiltração do extrato hidrossolúvel de soja (Omosaiye et al.1978); extração dos oligossacarídeos com etanol a partir da farinha de soja (Leske et al. 1991); utilização de α-galactosidase de origem vegetal e microbiana (Guimarães et al. 2001). A obtenção de variedades de soja com baixo teor de estaquiose e rafinose, associado com altos teores de proteína e/ou sacarose é uma alternativa interessante para melhorar a qualidade nutricional do grão (Sato et al. 2014), uma vez que estas variedades irão reduzir o tempo e custos demandados para o beneficiamento do grão e seus derivados.

3.7. Melhoramento genético da soja para teor de oligossacarídeos

A demanda atual por alimentos mais saudáveis tem ajudado a descobrir o valor nutricional e benefícios para a saúde da soja. Do mesmo modo, a demanda por sojas especiais destinadas ao consumo humano vem aumentando, seja para o consumo do grão diretamente, ou para a produção de derivados, como proteína texturizada, tofu, leite de soja, natto e edamame (Wang et al. 2014, O'Keefe et al. 2015). Assim, o desenvolvimento de cultivares de soja com maior qualidade nutricional é considerado imperativo para os melhoristas, e um novo nicho de mercado para produtores. Tamanho de sementes, alto teor de proteína, alto teor de sacarose e baixos teores de oligossacarídeos de rafinose são objetivos comuns de programas de melhoramento visando ao desenvolvimento de sojas especiais (Jaureguy et al. 2011).

A biossíntese de rafinose e estaquiose envolve as enzimas rafinose sintase, da família das hidrolases, que realizam a transferência do galactinol para a sacarose (Dierking & Bilyeu 2008; Qiu et al. 2015; Bilyeu & Wiebold 2016). A Figura 3 mostra um esquema da biossíntese de ROs, adaptado de Gangl et al. (2015). Skoneczka et al. (2009) caracterizaram uma deleção de três pares de bases na sequência do gene rafinose sintase 2 (*GmRS2* – Glyma.06G179200) do acesso PI200508, que apresenta baixo teor de estaquiose. A deleção encontrada explicou entre 88 e 94% da variação observada no teor de estaquiose e 76% do teor de sacarose do grão de duas populações F_2 derivadas da PI200508. Estudos anteriores mostraram que a PI200508 é um bom doador de alelos para baixo teor de oligossacarídeos de rafinose. O vigor das sementes não é afetado em linhagens com baixo teor de ROs geradas por retrocruzamentos e derivadas da PI200508, de modo que não há diferença significativa em características de qualidade da semente, tais como emergência no campo, rendimento de sementes, maturidade, altura e teor de ácidos graxos (Neus et al. 2005). Estas características são importantes para melhorar a qualidade nutricional da soja sem afetar o desempenho agronômico das cultivares (Yang et al. 2004).

O desenvolvimento de variedades de soja com baixo teor de estaquiose, associado com altos teores de sacarose e de proteína tem se mostrado um desafio pela presença de correlação negativa entre proteína e sacarose (Jaureguy et al. 2011, Sato et al. 2014). Estudos de correlação entre caracteres de qualidade do grão em populações segregantes para baixo teor de estaquiose têm se mostrado uma importante ferramenta para guiar melhoristas buscando desenvolver sojas com características desejáveis para o consumo humano. Considerando que as sínteses de sacarose, rafinose e estaquiose compartilham precursores na mesma via metabólica (Dierking & Bilyeu 2008, Bilyeu & Wiebold 2016), é esperado que o aumento do teor de sacarose pela inibição da síntese de ROs possa ser alcançado sem diminuição no teor de proteína, conforme sugerido por Sato et al. (2014).

3.8. Metodologias para detecção de SNPs

Com o advento das técnicas de biologia molecular, a manipulação de DNA se tornou possível, levando ao surgimento dos marcadores moleculares na década

de 1980 (Guimarães et al. 2009). Marcadores moleculares são variantes na estrutura do DNA em determinada posição do genoma, que podem ser utilizados para marcar alelos de genes de interesse. Algumas das grandes vantagens sobre os marcadores morfológicos são o número quase ilimitado de polimorfismos distribuídos aleatoriamente ao longo de todo o genoma, a velocidade de diagnóstico e a ausência de interferência do meio (Guimarães et al. 2009).

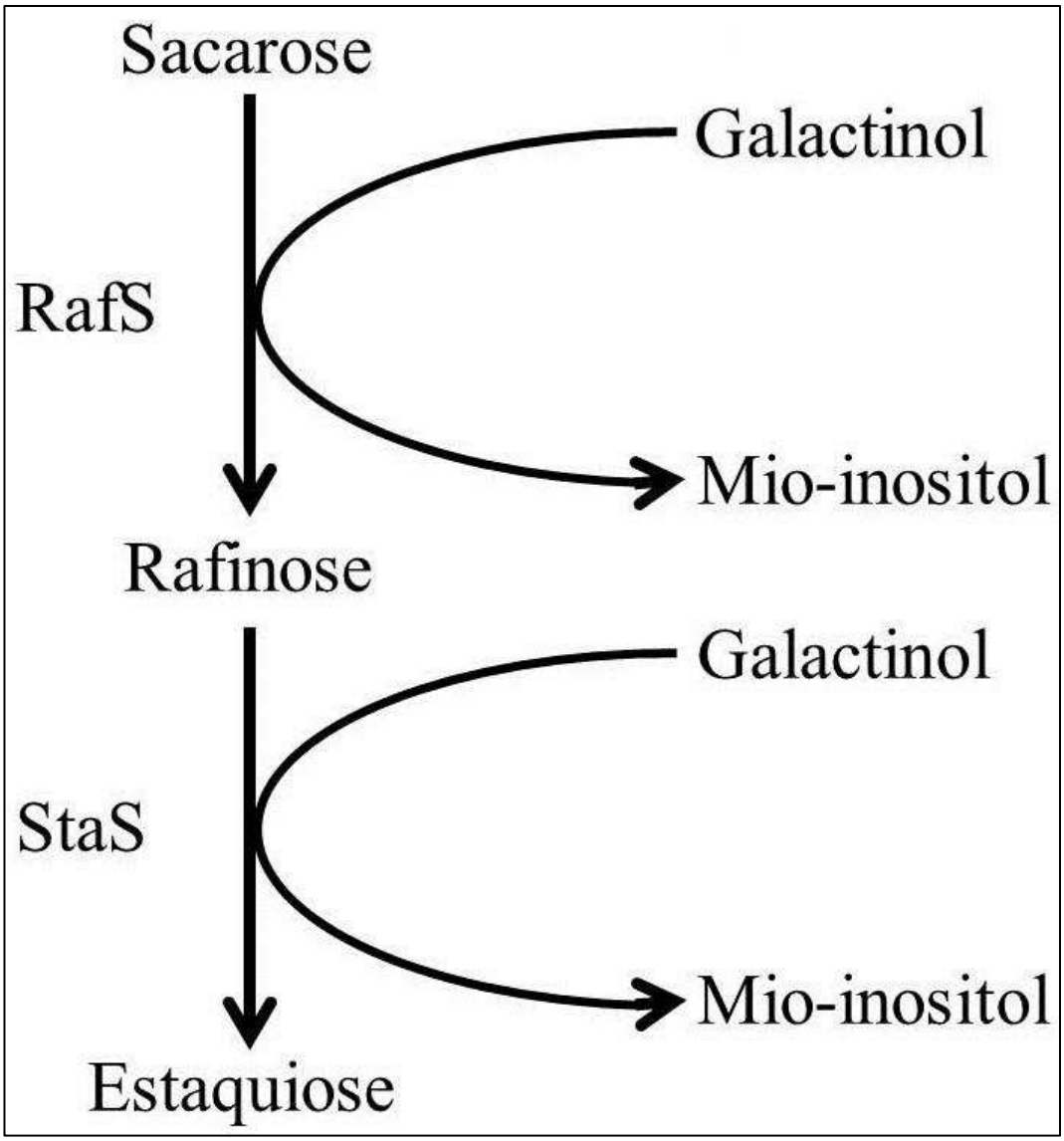

Figura 3 – Biossíntese de rafinose e estaquiose em plantas. A síntese de rafinose é mediada por enzimas rafinose sintases (RafS), que promove a transferência de um resíduo de galactosil da molécula de galactinol para a sacarose. A enzima estaquiose sintase (StaS) realiza a transferência de outro resíduo de galactosil para a rafinose, produzindo estaquiose.

SNPs (*Simple Nucleotide Polymorphism*) são variações de nucleotídeos ou pequenas inserções/deleções (indels) em sequências de DNA homólogas (Collins et al. 1998). Os marcadores baseados em SNPs ganharam interesse da comunidade científica por serem a forma de polimorfismo mais abundante em genomas (Zhu et al. 2003), por sua facilidade de automação, viabilizada por tecnologias que têm sido constantemente desenvolvidas (Fan et al. 2006). Para ser considerado um SNP, um polimorfismo precisa ser verificado em pelo menos 1% da população analisada. A ocorrência de SNP em genomas de plantas é de cerca de 1 a cada 100-300 pb (Gupta et al. 2001; Caixeta et al. 2009).

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 (Hyten et al. 2006). Ainda assim, Song et al. (2013) encontraram 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 da Williams 82.

Várias metodologias para a detecção de SNP têm sido desenvolvidas e utilizadas pela comunidade científica, a maioria delas utilizando fluoróforos intercalantes. *High Resolution Melt* (HRM) é uma metodologia pós-PCR que compara diferenças nas curvas de desnaturação de fragmentos contendo SNP e pequenos *indels* (Smith et al. 2009). O produto de amplificação é submetido a um gradiente de temperatura, com leitura de fluorescência em tempo real, gerando uma curva de desnaturação. Pequenas variações na sequência dos fragmentos são suficientes para gerar alterações nas curvas de desnaturação e permitir a genotipagem dos indivíduos.

O sistema Taqman® (ThermoFisher Scientific) para discriminação alélica utiliza um sistema de sondas marcadas com diferentes fluoróforos, e a atividade 5’ nuclease da Taq polimerase (Shi et al. 1999). São desenvolvidas duas sondas complementares à região alvo, cada uma específica para um alelo do SNP, e contendo uma molécula repórter na extremidade 5’ e um *quencher* na extremidade 3’. Este *quencher* é uma molécula que interfere na intensidade da fluorescência

emitida pelo repórter. Quando as sondas estão em solução, a proximidade entre o *quencher* e o repórter impossibilita a leitura de fluorescência do repórter. Durante a PCR, a sonda hibridiza especificamente com seu alelo alvo, e durante a fase de extensão, o repórter presente na extremidade 5' é clivado pela atividade 5' nuclease da Taq polimerase. A separação do repórter e do *quencher* resulta em aumento da liberação de fluorescência pelo repórter. Por meio da medição da intensidade de fluorescência de ambos os repórteres (específicos de cada alelo) após a realização da PCR, o genótipo é determinado (Shi et al. 1999, Alves & Novais 2004).

Outra técnica bastante utilizada para genotipagem de SNPs é KASP (Kompetitive Allele Specific PCR). A técnica utiliza dois *primers forward* alelo-específicos contendo uma cauda específica para cada um deles, um *primer reverse* comum, e sondas específicas para anelar na sequência complementar à cauda dos *primers*, contendo diferentes fluorófilos (geralmente FAM e HEX). Após a síntese das fitas, são geradas regiões complementares às caudas dos *primers* alelo-específicos. As sondas, que normalmente estão ligadas a um *quencher*, se ligam a estas regiões complementares, possibilitando que emitam fluorescência (LGC Group).

O desenvolvimento de marcadores SNPs baseados em mutações associadas com características de qualidade do grão apresenta grande potencial para o melhoramento genético da soja, visando ao desenvolvimento de cultivares de soja com características especiais.

CAPÍTULO I

CARACTERIZAÇÃO DE UM NOVO ALELO PARA O GENE *GmFAD3A* ASSOCIADO COM REDUÇÃO NO TEOR DE ÁCIDO LINOLÊNICO NO GRÃO DE SOJA

1. INTRODUÇÃO

A soja é a segunda principal fonte de óleo vegetal no mundo, correspondendo por cerca de um terço do mercado internacional, em termos de produção e consumo (USDA 2016a). No Brasil, o óleo de soja corresponde por mais de 80% do mercado de óleos vegetais para a indústria alimentícia. Estima-se que foram produzidos no país cerca de 7,8 milhões de toneladas de óleo de soja em 2016, uma queda de 4,0% em relação ao ano de 2015 (ABIOVE 2017a). Apesar desta queda, a produção de óleo de soja vem crescendo nos últimos anos, e para 2017, as previsões são de uma produção de 8,1 milhões de toneladas de óleo, um crescimento de 4,51%. A partir desta previsão para 2017, estima-se um crescimento de 34% nos últimos 10 anos (ABIOVE 2017a).

Uma das explicações para os incrementos na produção de óleo de soja nos últimos anos é o crescimento na produção de biodiesel, cuja principal matéria prima é o óleo de soja (ANP 2016). A produção de biodiesel no país passou de 69,01 mil metros cúbicos em 2006, para de 3,94 milhões de metros cúbicos em 2015, um crescimento de 5.700%. Neste contexto, a soja foi responsável por 77,7% da produção total (ANP 2016). Segundo a Lei nº 13.263 de março de 2016, que altera a Lei nº 13.033, de 24 de setembro de 2014, o percentual de biodiesel misturado ao óleo diesel comercializado no país deve passar para 8% a partir de 23 de março de 2017, e este percentual deve chegar a 10% em 2019 (BRASIL 2016).

A soja acumula, em média, 20% de óleo no grão, composto por cerca de 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). A composição de ácidos graxos determina a qualidade do óleo, e influencia suas propriedades físicas e químicas, tais como ponto de fusão e estabilidade oxidativa (Yadav 1996, Fehr 2007). A alta concentração de ácidos graxos poli-insaturados (ácidos linoleico e linolênico) no óleo de soja é a principal causa de baixa estabilidade oxidativa, resultando em rancificação, rápido declínio do sabor ideal e redução no tempo de armazenamento de produtos manufaturados (Warner & Fehr 2008).

Com o intuito de aumentar a estabilidade oxidativa do óleo de soja, a indústria alimentícia tem utilizado o processo de hidrogenação química para reduzir

o teor de ácidos graxos poli-insaturados e aumentar o teor de ácido oleico; entretanto, este processo leva a produção de ácidos graxos *trans*. O consumo de componentes parcialmente hidrogenados está diretamente relacionado à incidência de doenças coronárias, altos níveis de colesterol e desenvolvimento de diabetes tipo 2 (Hu et al. 1997, Mozaffarian et al. 2006, Pham et al. 2014). A relevância deste fato é tamanha que o FDA (US *Food and Drug Administration*) emitiu uma determinação de que óleos parcialmente hidrogenados (OPHs) não são mais reconhecidos como seguros (FDA 2015), e estabeleceu um período de três anos para que os fabricantes de alimentos se adequem a esta determinação ou façam um requerimento para o uso de OPHs (FDA 2016).

Melhoramento convencional e engenharia genética têm sido utilizados para desenvolver variedades de soja com baixo teor de ácidos graxos poli-insaturados, aumentando a estabilidade oxidativa sem a necessidade de hidrogenação química (Fehr 2007, Flores et al. 2008, Pham et al. 2014). O primeiro programa de melhoramento de soja buscando alterar a composição de ácidos graxos foi iniciado na década de 1950 (Dutton et al. 1951). Ao longo dos anos, foi caracterizado que a biossíntese de ácidos graxos em soja depende principalmente das enzimas dessaturases do retículo endoplasmático (microssomo) (Baud & Lepiniec 2010). O substrato primário para estas enzimas são resíduos de ácidos graxos ligados a fosfatidil colina ou coenzima-A (CoA). As ω6-dessaturases convertem oleoil (18:1) em linoleoil (18:2), enquanto que as ω3-dessaturases são responsáveis pela conversão de linoleoil em linolenil (18:3) (Baud & Lepiniec 2010).

O teor de ácido oleico no grão de soja é controlado pelas enzimas ω6-dessaturases, codificadas pelos genes *GmFAD2-1A* e *GmFAD2-1B* (Heppard et al. 1996). Mutações nestes genes estão associadas com aumento no teor de ácido oleico e redução no teor de ácido linoleico. A combinação de mutações nestes dois genes possibilita o desenvolvimento de soja com mais de 80% de ácido oleico no grão (Pham et al. 2010, Pham et al. 2011, Pham et al. 2012). Outras isoformas destes genes são responsáveis pela produção de ácido linoleico em outros tecidos da planta (Heppard et al. 1996, Li et al. 2007).

Os genes *GmFAD3* codificam para enzimas ω3-dessaturases. A redução do conteúdo de ácido linolênico em genótipos de soja é controlada por mutações em

alelos nos locos *fan1*, *fan2*, *fan3* e *fanx* relacionados com a expressão de enzimas ω3-dessaturases microssomais (Bilyeu et al. 2005, Anai et al. 2005, Bilyeu et al. 2006, Chappell & Bilyeu 2006). O gene *GmFAD3A/fan1* (Glyma.14g194300) está localizado no grupo de ligação (GL) B2/cromossomo 14; o gene *GmFAD3B*/*fan3* (Glyma.02g227200) está localizado no GL D1b/cromossomo 2; o gene *GmFAD3C/fan2* (Glyma.18g062000) está localizado no GL-G/cromossomo 18 (Bilyeu et al. 2011, Cardinal et al. 2011, Goodstein et al. 2012).

Seleção recorrente utilizando materiais com baixos níveis de ácido linolênico como doador de alelos tem tornado possível reduzir o conteúdo de ácido linolênico para cerca de 4,2% (Wilson et al. 1981). Mais recentemente, o uso de alterações genéticas induzidas por mutagênese de irradiação física e química produziu mutantes com baixos teores de ácido linolênico: A5, C1640, RG10, M5, IL8, A23 e A26 (Hammond & Fehr 1983, Wilcox et al. 1984, Bubeck et al. 1989, Rahman et al. 1996, Stijšin et al. 1998, Fehr & Hammond 2000). O sequenciamento dos genes *GmFAD3* nas linhagens de soja A5, A26 e A23, aliado a análises de associação de marcadores, possibilitou a identificação de três genes *GmFAD3* mutantes, que contribuem diretamente para o baixo teor de ácido linolênico nestas linhagens de soja (Byrum et al. 1997, Bilyeu et al. 2003, Bilyeu et al. 2006, Pham et al. 2014). A combinação dos alelos *fan1* (A5), *fan2* (A23) e *fan3* (A26) produziu a linhagem A29, com apenas 1% de ácido linolênico (Fehr & Hammond 2000, Ross et al. 2000, Pham et al. 2014).

A linhagem FA22 apresenta teor médio de ácido oleico (cerca de 50%), e foi desenvolvida através de mutagênese pela Iowa State University. Apesar de ser um potencial doador de alelos para aumento do teor de ácido oleico, as bases genéticas desta característica ainda não são conhecidas para esta linhagem.

Por meio de um programa de melhoramento, nosso grupo desenvolveu a variedade CS303TNKCA, com baixo teor de ácido linolênico (3,5-4,0%), ausência das três lipoxigenases (*lox1*, *lox2* e *lox3*) e do inibidor de protease Kunitz. Esta variedade é uma isolinha obtida por retrocruzamento da linhagem BARC-12 (Leffel 1994), doadora para baixo teor de ácido linolênico, com a variedade Monarca como recorrente (Brune et al. 2010). Neste sentido, o objetivo deste estudo foi caracterizar

o baixo teor de ácido linolênico no grão presente na variedade CS303TNKCA e o médio teor de ácido oleico presente na linhagem FA22.

2. MATERIAL E MÉTODOS

2.1. Material vegetal

Foram utilizadas duas linhagens de soja contrastantes para teor de ácido linolênico. A variedade CS303TNKCA foi desenvolvida pelo Programa de Melhoramento da Qualidade da Soja (PMQS/BIOAGRO/UFV), e apresenta baixo teor de ácido linolênico (3,5-4,0%), teor normal de ácido oleico, médio teor de óleo (20-23%), ausência das três lipoxigenases (*lox1*, *lox2* e *lox 3*) e do inibidor de protease Kunitz. Esta variedade é uma isolinha obtida por retrocruzamentos entre a linhagem BARC-12 (Leffel 1994) e a variedade Monarca (COOPADAP) como genitor recorrente. A linhagem FA22 foi desenvolvida por mutagênese na Iowa State University e contém teor médio de ácido oleico (50%) e conteúdo normal de ácido linolênico (5-6%; Alt et al. 2005).

2.2. Desenvolvimento da população segregante

A variedade CS303TNKCA e a linhagem FA22 foram utilizadas para desenvolver a população do estudo. O cruzamento entre elas foi realizado em casa de vegetação (CV) na Universidade Federal de Viçosa, em Viçosa-MG, para obter a população F_2. Famílias $F_{2:3}$ foram cultivadas em campo experimental da Universidade Federal de Viçosa em Visconde do Rio Branco-MG (VRB). O experimento foi conduzido em blocos ao acaso com duas repetições. Cada parcela consistiu de uma linha de 1,5 metros plantada com 25 sementes por fileira, com 0,5 metros entre linhas. Famílias $F_{2:4}$ foram cultivadas em Viçosa (VIC) e em São Gotardo-MG, em dois experimentos (SG1 e SG2), sob condições similares às anteriores.

2.3. Extração de DNA, sequenciamento e identificação de mutações

Amostras foliares dos genitores CS303TNKCA e FA22, e das plantas F_1 e F_2 foram coletadas no estágio de desenvolvimento V2 (Fehr & Caviness 1977), congeladas em nitrogênio líquido e liofilizadas. DNA genômico foi extraído conforme metodologia descrita por Doyle & Doyle (1990). A concentração do DNA foi determinada utilizando um espectrofotômetro NanoDrop (NanoDrop Technologies, Wilmington, DE), e sua qualidade foi verificada por eletroforese em gel de agarose 0,8%.

As sequências completas dos genes *GmFAD2-1A* (Glyma.10G278000), *GmFAD2-1B* (Glyma.20G111000), *GmFAD3A* (Glyma.14G194300), *GmFAD3B* (Glyma.02G227200) e *GmFAD3C* (Glyma.18G062000) foram obtidas no banco de dados Phytozome (Goodstein et al. 2012). O *software* Primer3 Input v. 0.4.0 (Untergrasser et al. 2012) foi utilizado para desenhar conjuntos de *primers* para cada gene selecionado de modo a amplificar toda a região do gene (Tabela 1). Cada conjunto de *primers* foi testado para produzir um único produto de amplificação, que foi purificado utilizando o kit ExoSAP (USB Corporation, Cleveland, Ohio, USA) e então sequenciado na empresa Macrogen (Gasan-dong, Geumchun-gu, Seoul, Korea). As sequências geradas foram editadas utilizando o *software* Sequencher 4.1.4 (Gene Codes Corporation) e alinhadas no *software* ClustalW (Thompson et al. 2002). Os polimorfismos encontrados foram analisados nas plataformas ExPASy (Artimo et al. 2012) e NetPlantGene Server (Hebsgaard et al. 1996). O genoma de referência da linhagem Williams 82 foi utilizado para identificar e comparar mutações.

2.4. Genotipagem

A genotipagem da população F_2 contrastante para teor de ácido linolênico foi realizada utilizando um par de *primers* capaz de diagnosticar a mutação encontrada no gene *GmFAD3A* da variedade CS303TNKCA por meio da metodologia de *High Resolution Melt* (HRM; Liew et al. 2004): Forward: TCCCACCCAGTGAGAGAAAG; Reverse:

Tabela 1 – *Primers* utilizados para sequenciar os genes *GmFAD2-1A*, *GmFAD2-1B*, *GmFAD3A*, *GmFAD3B* e *GmFAD3C* na variedade de soja CS303TNKCA e na linhagem FA22. Temperatura de anelamento (Tm), tamanho do *amplicon* e tamanho do *amplicon* após edição são mostrados.

Gene	**Marcador**	**Região**	**Sequência dos *Primers***	**Tm (ºC)**	***Amplicon* (pb)**	***Amplicon* editado (pb)**
GmFAD2-1A	FAD2-1A1	Região Upstream-Íntron 1	Fw: TCCGTGTCTCCAAGTCCTCT Rv: GCACCTAAGCCGACAAAAAG	54	788	695
GmFAD2-1A	FAD2-1A2	Íntron 1-Éxon2	Fw: TCATGCATTTGGTTCTGTGTC Rv: TTTCTCCTTCATTGGGCATC	54	971	708
GmFAD2-1A	FAD2-1A3	Éxon 2-Região downstream	Fw: ATGGGGATATGGTCAACGTC Rv: CGTGACCCCTTACCAACCTA	54	810	745
GmFAD2-1B	FAD2-1B1	Região Upstream-5' UTR	Fw: TGATCACAGGTGGAAACCAA Rv: AAAGCGAATCAACCTTGTGG	54	876	797
GmFAD2-1B	FAD2-1B16	Região 5' UTR	Fw: GGCAATATTGTCTTCTTTTTCCTC Rv: ACAAAAAGGCACAAAACAGG	60	877	470
GmFAD2-1B	FAD2-1B2	5' UTR-Íntron 1	Fw: ATGTTTCAGCGTTGGCTTTC Rv: TGCACGATTCAATTTGGAGA	54	871	839
GmFAD2-1B	FAD2-1B17	5' UTR-Éxon2	Fw: CCCGCTGTCCCTTTTAAACT Rv: TACTTGCTGAAGGCATGGTG	58	869	771
GmFAD2-1B	FAD2-1B3	Íntron 1-Éxon2	Fw: TGTCACTTCCCTCCATTCATT Rv: ACATAGCAGCCAAACCAACC	54	904	731
GmFAD2-1B	FAD2-1B4	Éxon2-3' UTR	Fw: GGGCTGCTTCTCTTCTCAT Rv: TGGAGTGTGCATTTAAGGAA	54	820	794
GmFAD3A	FAD3A8	Íntron 6-íntron 7	Fw: CCCCTTAACCTTACCTCAACAA Rv: GGGACCAGAAGTATCAAAGCA	55	756	724
GmFAD3A	FAD3A17	5' UTR-éxon 5	Fw: TATTACGCACCACCCACCAC Rv: AAGCCGGGATGAACTAACCA	55	637	607
GmFAD3A	FAD3A18	Éxon 6-íntron 6	Fw: AATTGTTTGAACACGTGCAGAT Rv: GTAGCTTTCCATTGAATCGTGA	55	704	581

GmFAD3A	FAD3A19	Éxon 5-íntron 6	Fw: AGCAATCTGTTCCCACCCAG Rv: CCAAACACTACGTAGACAACAG	55	780	673
GmFAD3A	FAD3A20	Íntron 2-íntron 6	Fw: ATTGCGTGTTTGGAATCCC Rv: CTTGAGCAATAGAACTGGACTAG	55	724	603
GmFAD3A	FAD3A21	Íntron 7-região downstream	Fw: GGAACTTGGTGGTGAAAGGG Rv: CGGTGGAACTCTGTTTCTGA	55	848	808
GmFAD3B	FAD3B1	5' UTR-íntron 2	Fw: TTTTTCGAAGATTTCCTTCTTCC Rv: AAGAACTTGCAACGACAGATGA	56	864	766
GmFAD3B	FAD3B2	Íntron 2-íntron 6	Fw: AACCACATTTTGGATTTGGAAC Rv: CGCATTTAAGTGTTCACCCACT	56	899	829
GmFAD3B	FAD3B3	Íntron 6-íntron 7	Fw: AGCCCCGTTTTTCATTTATAGC Rv: AAATGAAGTGGGTGTGAAAACA	56	496	472
GmFAD3B	FAD3B4	Íntron 7-região downstream	Fw: CTATACTGGGAGCTTGGTGGTG Rv: ATGCTCTGTGCAAGTGGTAGAA	56	630	603
GmFAD3C	FAD3C1	5' UTR-íntron 3	Fw: TCCAATAGAAGAAAAGGGGGTA Rv: ATAGTGGACGGTTTGGACTTGT	56	933	866
GmFAD3C	FAD3C21	Éxon 3-íntron 5	Fw: TGTTGAGAAGGATGAATCATGG Rv: TGGACATGTGTTAACCCCATTA	56	706	704
GmFAD3C	FAD3C22	Éxon 5-íntron 6	Fw: AGAAGCCCTGGAAAAGAAGG Rv: TTTGAGGGAAAAGGTGATGG	56	882	847
GmFAD3C	FAD3C3	Íntron 6-região downstream	Fw: ATTTATGTGCAGTTTTGCCATC Rv: TGCAATTCTGACAAAGCACATA	56	924	898

GAAAACATGGTAACCCAACACA. O *software* uMelt (Dwight et al. 2011) foi utilizado para estimar a eficiência de diferenciação para o marcador desenvolvido.

O experimento foi realizado em um aparelho RotorGene-Q (QIAGEN), em reação contendo 30 ng de DNA, 0,7 µM de cada *primer*, e 5 µL de Type-it HRM 2X PCR kit (QIAGEN), em um volume final de 10 µL. PCR foi executada sob as seguintes condições: 94 ºC por 10 minutos; 40 ciclos iniciados a 94 ºC por 20 segundos, 54 ºC por 20 segundos, e 72 ºC por 30 segundos. Por fim, os produtos de amplificação foram submetidos a um gradiente de temperatura variando entre 60 e 90 ºC, com aumento de 0,1 ºC por passo. A fluorescência foi medida em tempo real a cada incremento de temperatura para determinar a curva de *melt*. Cada amostra foi executada em duplicata, e foram determinados indivíduos padrões para curva de genótipos mutantes, selvagens e heterozigotos.

2.5. RT-PCR

Para verificar alteração nos processos de *splicing*, foi realizado uma análise através de RT-PCR. Amostras de RNA foram extraídas de sementes em estágio de desenvolvimento R5 utilizando Trizol (LifeTechnology), seguindo recomendações do fabricante, e quantificadas usando Picogreen (Thermo Scientific) em um aparelho SpectraMax M5 (Molecular Devices). A qualidade do RNA foi verificada por eletroforese em gel de agarose 1,5%. Um µg de RNA foi tratado com DNAse I Amplification Grade (Thermo Scientific) e então submetido à transcrição reversa utilizando a enzima Superscript III (Invitrogen). PCR foi realizado em volume final de 15 µL, contendo 2 µL do produto da reação de transcrição reversa, 0,33 µM de cada *primer*, 0,5 U de Platinum® *Taq* DNA Polymerase (Invitrogen), 200 µM de cada deoxirribonucleotídeo, 1,5 µL de PCR Buffer (Invitrogen), e 1,5 mM de $MgCl_2$, sob as seguintes condições: 94 ºC por 2 minutos; 38 ciclos iniciados a 94 ºC por 30 segundos, 55 ºC por 30 segundos, e 72 ºC por 2 minutos; e uma etapa final a 72 ºC por 7 minutos. Os produtos de PCR foram visualizados por eletroforese em gel de agarose 1,2%. A Tabela 2 apresenta os conjuntos de *primers* utilizados para amplificar cada gene *GmFAD3*.

Tabela 2 – *Primers* utilizados para amplificação dos genes *GmFAD3* de soja através de RT-PCR. *Primers forward* anelam no início do primeiro éxon, e *primer reverse* anelam no fim do último éxon.

Gene	Primer Forward	Primer Reverse
GmFAD3A	ATGGTTAAAGACACAAAGCCTT	TCACTCGCACCGAGACTGA
GmFAD3B	ATGGTTAAAGACACAAAGCCTT	CCACTCGCAACGAGACTGA
GmFAD3C	AAGGCACAATGTTTTGGGCAC	CTTGGACCCAGTCCAACTAA

2.6. Análises fenotípicas

Lipídeos de sementes das linhas avaliadas nos experimentos CV, VRB, VIC, SG1 e SG2, e dos genitores da população CS303TNKCA e FA22 foram extraídos conforme metodologia adaptada de Gesteira et al. (2003). Quinze miligramas de sementes trituradas foram colocadas em tubo de vidro de tampa rosqueável, e adicionadas com 1 mL de hexano. Em seguida, as amostras foram borbulhadas com nitrogênio, agitadas em vórtex e acondicionadas a 4 ºC por 16 horas. Posteriormente, a solução de hexano foi transferida para um novo tubo, evaporada com nitrogênio líquido, adicionada com 0,4 mL de metóxido de sódio, agitada em vórtex e levada para banho-maria a 30 ºC por 1h. Logo depois, foram adicionados a cada tubo 1 mL de água milliQ e 1 mL de hexano, agitados em vórtex e acondicionados à temperatura ambiente por 1h. Por fim, a fase orgânica (sobrenadante) foi coletada para novo tubo, adicionada com sulfato de sódio anidro, e transferida para tubos de injeção. A composição de ácidos graxos foi realizada por cromatografia gasosa em um aparelho Shimadzu GC-2010 Plus, conforme descrito por Burkey et al. (2007). O conteúdo de ácidos graxos nas amostras foi determinado como uma proporção do total de ácidos graxos, representado em % de óleo. O teor de óleo das sementes foi determinado por espectrometria de ressonância magnética nuclear (NMR – Resonance Instruments, Witney, Oxfordshite, UK). As análises fenotípicas foram realizadas por Matta (2012).

2.7. Análises estatísticas

Teste de χ^2 foi realizado para verificar a segregação da mutação encontrada no gene *GmFAD3A* na população F_2. Foram calculados médias, desvios padrões, máximos e mínimos para conteúdo dos ácidos palmítico, esteárico, oleico, linoleico

e linolênico para cada experimento, utilizando o *software* GENES (Cruz 2013). Regressão linear foi realizada no *software* GQMOL (Cruz & Schuster 2004) para verificar a associação entre a mutação e o conteúdo de ácido linolênico, bem como avaliar efeitos aditivos e dominantes. Foram gerados gráficos de distribuição de frequência, separando indivíduos por genótipo e agrupando os valores de ácido linolênico em intervalos de desvio padrão. Os gráficos foram gerados com auxílio dos *softwares* Microsoft Excel e Microsoft Power Point (Microsoft Corporation).

2.8. Análise de atividade enzimática

Para avaliar a atividade da ω3-dessaturase codificada pelo gene *GmFAD3A* da variedade CS303TNKCA, foi realizada expressão heteróloga na levedura *Saccharomyces cerevisiae*. Produtos da reação de transcrição reversa (ver item 2.5) foram amplificados com *primers* desenhados especificamente para a região codificante do gene *GmFAD3A* selvagem e mutante (FAD3A/Fw: CCCGGATCCATGGTTAAAGACACAAAGCCTT; FAD3A/Rv: GGGGCTCGAGTCAGTCTCGGTGCGAGTGA; e FAD3-CS303/Rv: GGGGCTCGAGCTATCCCTTTCTCTCACTGG). Estes *primers* possuem em sua sequência os sítios de restrição das enzimas BamHI (FAD3A/Fw) e XhoI (FAD3A/Rv e FAD3A-CS303/Rv), para auxiliar nos processos de clonagem. PCR foi realizada em volume final de 30 µL, contendo 2,0 µL de cDNA, 3,0 µL de 10X PCR Rxn Buffer (Invitrogen), 1,5 mM de $MgSO_4$, 200 µM de cada deoxinucleotídeo, 0,33 µM de cada *primer*, e 0,5 U de Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen), sob as seguintes condições: uma etapa inicial a 94 ºC/2 minutos, seguida de 40 ciclos iniciados a 94 ºC/15 segundos, 55 ºC/30 segundos, e 68 ºC/2 minutos, finalizando a 68 ºC/10 minutos. Os produtos de amplificação gerados foram visualizados por eletroforese em gel de agarose 1,2%, e purificados utilizando o PureLink Quick Gel Extraction Kit (Thermo Fisher Scientific).

Posteriormente, os produtos de amplificação foram digeridos com as enzimas de restrição BamHI e XhoI (New England Biolabs), em reação contendo 100 ng do produto purificado, 10 U de cada enzima, 2,0 µL de NEBuffer 3.1 (New England Biolabs), em volume final de 20 µL. A reação foi realizada a 37ºC por 2

horas, e as enzimas foram inativadas a 85 ºC por 10 minutos. Os fragmentos digeridos foram clonados no vetor pYES2 Yeast Expression Vector (Thermo Fisher Scientific), previamente digerido com as mesmas enzimas. A reação de clonagem foi realizada em volume final de 20 µL, contendo 1,0 µL da enzima T4 DNA Ligase (New England Biolabs), 90 fmol do fragmento digerido, 30 fmol do vetor e 2,0 µL do T4 DNA Ligase Reaction Buffer (New England Biolabs).

Os plasmídeos construídos foram transformados em bactérias da cepa DH5α de *Escherichia coli* para validação e multiplicação dos vetores. Estes vetores foram utilizados para transformar células de leveduras da cepa W303 de *S. cerevisiae*, conforme metodologia adaptada de Gietz & Schiestl (2007). Células da levedura foram crescidas em 7 mL de meio YPD a 28 ºC/180 giros até atingirem OD entre 0,5 e 1,5. Então, as colônias foram centrifugadas a 5.000 giros/4 minutos, o sobrenadante foi descartado, e as células foram ressuspendidas em água estéril e centrifugadas novamente. Após o descarte do sobrenadante, as células foram ressuspendidas em 125 µL acetato de lítio 100 mM (AcLi) estéril, e transferidas para microtubos de 2,0 µL. Após nova centrifugação a 10.000 giros/30 segundos, o sobrenadante foi descartado, e as células ressuspendidas em 50 µL de AcLi 100 mM. Seguiu-se uma nova centrifugação a 10.000 giros/30 segundos e descarte do sobrenadante. Em seguida, foram adicionados nesta ordem: 240 µL de PEG 50% (polietilenoglicol), 36 µL de AcLi 1M, 25 µL de DNA de esperma de salmão 2 µg/µL e 50 µL de solução do vetor dissolvido em água (entre 2-5 µg). As amostras foram ressuspendidas em vórtex com baixa vibração, e incubadas a 30 ºC por 30 minutos, e depois a 42 ºC por 20-25 minutos. Após centrifugação a 8.000 giros/15 minutos, o sobrenadante foi descartado, e foi adicionado 1 mL de água estéril. Seguindo a uma nova centrifugação e descarte do sobrenadante, o pellet foi ressuspendido em 200 µL de água estéril, plaqueado em meio SD-ura (meio contendo YNB, glicose, e todos os aminoácidos essenciais, à exceção de uracila), e incubado em estufa por cerca de três dias a 28 ºC/180 giros.

Para análise da atividade enzimática, colônias transformadas foram crescidas em 7 mL de meio SD contendo galactose ao invés de glicose, para estimular a expressão do gene clonado, a 28 ºC por 12-16 horas e a 26 ºC por cerca de 48 horas. A extração de lipídeos e quantificação dos ácidos graxos foi realizada de modo semelhante ao item 2.6. Para a análise do perfil de ácidos graxos, foram

utilizadas colônias transformadas com o vetor não clonado, e colônias contendo os genes *GmFAD3A* selvagem e mutante.

2.9. Análise de expressão gênica

Para avaliar a expressão dos genes *GmFAD2-1* e *GmFAD3*, foram desenhados conjuntos de *primers* específicos para cada gene (Tabela 3). A Tabela 4 apresenta informações sobre os indivíduos analisados neste estudo.

A análise foi realizada em um aparelho 7500 Real Time PCR Systems (AppliedBiosystems), em reação contendo 1,0 µL de cDNA diluído 1:10, 5 µL de solução contendo os *primers forward* e *reverse* na concentração de 1,5 µM, e 6 µL de SYBR Green PCR Master Mix (Thermo Fisher Scientific). As reações de PCR em tempo real foram realizadas nas seguintes condições: 95ºC por 10 min e 40 ciclos de: desnaturação 95 ºC por 15 seg, seguido de anelamento e extensão a 60 ºC por 1 min. A especificidade dos produtos amplificados foi analisada pelas curvas de dissociação geradas pelo equipamento. Foram utilizados controles negativos sem a adição de cDNA sintetizado e/ou *primers* para confirmar a ausência de qualquer tipo de contaminante no meio reacional (No Template Control, NTC). Foram utilizadas três réplicas biológicas e duas réplicas de pipetagem. Para a obtenção de controles endógenos utilizados no experimento, foi feita uma busca na literatura para os controles endógenos já utilizados e validados em soja (Van de Mortel et al. 2007, Libault et al. 2008, Hu et al. 2009, Le et al. 2012, Li, Q. et al. 2012, Miranda et al. 2013). Estes controles foram utilizados para testes em nossas amostras e verificação de sua expressão estável. A estatística foi conduzida utilizando os *softwares* GeNorm (Vandesompele et al. 2002), BestKeeper (Pfaffl et al. 2004) e NormFinder (Andersen et al. 2004). Foram selecionados com base nas análises os genes CONS7 e UKN2, por se mostrarem os mais estáveis (Dados não mostrados).

Tabela 3 – *Primers* utilizados para a análise de expressão gênica.

Gene	***Primers***	**Tm (ºC)**	**Referência**
GmFAD2-1A	Fw: ACTCATGTGGCTCACCATCTCTT	59	
	Rv: TTGGTTGCCTCCATTGCAT		
GmFAD2-1B	Fw: CATGTGGCTCACCATCTTTTCTC	59	
	Rv: TGGTTGCCTCCGTTGCA		
GmFAD3A	Fw: TCTCACAACTGTGGATCGTGACT	58	
	Rv: TGGGTGCCAATGTCATGGT		
GmFAD3B	Fw: AAGAAGCCCCGGAAAGGA	58	
	Rv: CACTGGGTGGAAACAGATTGC		
GmFAD3C	Fw: AAAGCAGCAAAGGCAGTGCTA	59	
	Rv: TGGCAATGGCCCAGATTT		
CYP2	Fw: CGGGACCAGTGTGCTTCTTCA	55	Miranda et
	Rv: CCCCTCCACTACAAAGGCTCG		al. 2013
UBC4	Fw: GAGCGAGCAGTTTCAGAC	55	Miranda et
	Rv: CATAGGAGGGACGATACG		al. 2013
ELF1A	Fw: GACCTTCTTCGTTTCTCGCA	55	Miranda et
	Rv: CGAACCTCTCAATCACACGC		al. 2013
ELF1B	Fw: GTTGAAAAGCCAGGGGACA	55	Miranda et
	Rv: TCTTACCCCTTGAGCGTGG		al. 2013
60s	Fw: TCAGGACATTCTCCGCAAGATTCC	60	Le et al.
	Rv: GATCAGCAATTATGCACAACG		2012
ABC	Fw: GATCAGCAATTATGCACAACG	60	Le et al.
	Rv: CCGCCACCATTCAGATTATGT		2012
Fbox	Fw: AGATAGGGAAATTGTGCAGGT	60	Le et al.
	Rv: CTAATGGCAATTGCAGCTCTC		2012
SKIP16	Fw: GAGCCCAAGACATTGCGAGAG	60	Hu et al.
	Rv: CGGAAGCGGAAGAACTGAACC		2009
UKN1	Fw: TGGTGCTGCCGCTATTTACTG	60	Hu et al.
	Rv: GGTGGAAGGAACTGCTAACAATC		2009
UKN2	Fw: GCCTCTGGATACCTGCTCAAG	60	Hu et al.
	Rv: ACCTCCTCCTCAAACTCCTCTG		2009
ACT11	Fw: ATCTTGACTGAGCGTGGTTATTCC	60	Hu et al.
	Rv: GCTGGTCCTGGCTGTCTCC		2009
TUA5	Fw: TGCCACCATCAAGACTAAGAGG	60	Li, Q. et al.
	Rv: ACCACCAGGAACAACAGAAGG		2012

Glyma05g37470	Fw: TTCCACCTCGCAACCATGAT	60	Li, Q. et al.
	Rv: CGAAGCAAACTCCCTCTTGG		2012
Glyma08g28550	Fw: CTGGGCAAAAAGAACTGAAT	60	Li, Q. et al.
	Rv: AACTGAACTGACTAACAATTCGG		2012
Glyma18g04130	Fw: ATCATGAGTGAACAAAAAGGTTC	60	Li, Q. et al.
	Rv: AGTCTGAAACAAACTCTCGCC		2012
CYP	Fw: ACGACGAAGACGGAGTGG	60	Li, Q. et al.
	Rv: CGACGACGACAGGCTTGG		2012
Cons6	Fw: AGATAGGGAAATGGTGCAGGT	60	Libault et
	Rv: CTAATGGCAATTGCAGCTCTC		al. 2008
Cons7	Fw: ATGAATGACGGTTCCCATGTA	60	Libault et
	Rv: GGCATTAAGGCAGCTCACTCT		al. 2008
UBI3	Fw: GTGTAATGTTGGATGTGTTCCC	60	Van de
	Rv: ACACAATTGAGTTCAACACAAACCG		Mortel et
			al. 2007

Tabela 4 – Plantas de soja analisadas para expressão dos genes *GmFAD2-1* e *GmFAD3*.

Planta	**Característica**	**Mutações conhecidas**
A29	Baixo linolênico	*GmFAD3A*, *GmFAD3B* e *GmFAD3C*
CS303TNKCA	Baixo linolênico	Não caracterizada
PI603452	Médio oleico	*GmFAD2-1A*
PI283327	Médio oleico	*GmFAD-1B*
FA22	Médio oleico	Não caracterizada
PL04*	Alto oleico	*GmFAD2-1A* e *GmFAD2-1B*
PL21*	Alto oleico	*GmFAD2-1A* e *GmFAD2-1B*
PL126*	Alto oleico	*GmFAD2-1A* e *GmFAD2-1B*
NT12	Alta proteína	-
BARC8	Alta proteína	-
BR80	Alta proteína	-
CD01 RR	Alto óleo	-
SUPREMA	Alto óleo	-

* Ver item 3.2 do Apêndice I

3. RESULTADOS

3.1. Caracterização dos genes *GmFAD2-1A* e *GmFAD2-1B* da linhagem FA22

O sequenciamento do gene *GmFAD2-1A* (Glyma.10G278000) cobriu um total de 2.114 pb na linhagem FA22, dos quais 564 pb estavam *upstream* e 219 pb estavam *downstream* à ORF. Já o sequenciamento do *GmFAD2-1B* (Glyma.20G111000) cobriu um total de 3.525 pb na linhagem FA22, incluindo 1767 pb *upstream* e 188 pb *downstream* à ORF. A comparação entre as sequências geradas e as do genoma padrão Williams 82 não mostrou mutações nos genes *GmFAD2-1A* e *GmFAD2-1B* para a linhagem FA22. Ambos os genes foram completamente sequenciados para regiões de éxons, íntrons e regiões UTR.

3.2. Caracterização dos genes *GmFAD3* da variedade CS303TNKCA

O sequenciamento do gene *GmFAD3A* (Glyma.14G194300) cobriu um total de 3.721 pb na variedade CS303TNKCA, incluindo aproximadamente 56 pb *upstream* e 408 pb *downstream* à ORF. A comparação entre a sequência gerada para CS303TNKCA e o genoma de referência da variedade Williams 82 revelou a presença de oito mutações no gene *GmFAD3A*: duas mutações em regiões de éxon, cinco mutações em regiões de íntrons e uma mutação na região 3'UTR (Tabela 5). Foi observada uma troca (A>G) na posição 34 do éxon 2. O íntron 3 apresentou uma inserção de uma timina na posição 94, enquanto que o éxon 5 mostrou uma deleção de uma adenina na posição 79. Análises na sequência do íntron 6 revelaram uma inserção de uma citosina na posição 66 e uma troca (G>T) na posição 383. Foram detectadas uma troca (T>C) na posição 184 e uma inserção de uma adenina na posição 252 do íntron 7. Por fim, foi identificada uma troca (A>G) na posição 140 da região 3'UTR. Segundo a plataforma NetPlantGene Server (Hebsgaard et al. 1996), nenhuma destas mutações causa alterações nos sítios de *splicing*. A troca (A>G) na posição 34 do éxon 2 não resultou em troca de aminoácidos. No entanto, a deleção de A na posição 79 no éxon 5 (delA) resultou na formação de um stop códon prematuro, levando à formação de uma proteína truncada com apenas 207 resíduos, comparada com os 376 resíduos da proteína codificada pelo genótipo padrão Williams 82 (Figura 1). O fragmento de 169 resíduos faltantes na

CS303TNKCA abriga a região rica em histidina III (H-box3). Estas regiões são consideradas essenciais para a função catalítica das enzimas dessaturases (Shanklin et al. 1994).

Tabela 5 – Polimorfismos encontrados entre sequências de DNA do gene *GmFAD3A* das linhagens de soja CS303TNKCA e Williams 82, destacando a deleção de 1 pb que leva à alteração no quadro de leitura.

Linhagem de soja	**Éxon 2**	**Íntron 3**	**Éxon 5**	**Íntron 6**		**Íntron 7**		**3' UTR**
Posição	34	94	79	66	383	184	252	140
Williams 82	A	-	A	-	G	T	-	A
CS303TNKCA	G	T	-	C	T	C	A	G

O sequenciamento do gene *GmFAD3B* (Glyma.02G227200) cobriu um total de 2.670 pb na CS303TNKCA, dos quais 62 pb foram *upstream* e 203 pb foram *downstream* à ORF. O sequenciamento do gene *GmFAD3C* (Glyma.18G062000) cobriu um total de 2.803 pb, incluindo 44 pb *upstream* e 225 pb *downstream* à ORF. A comparação das sequências entre os locos correspondentes entre CS303TNKCA e Williams 82 não relevaram polimorfismos nas regiões sequenciadas dos genes *GmFAD3B* e *GmFAD3C*. Os éxons foram completamente sequenciados para os três genes, enquanto que as regiões intrônicas foram parcialmente sequenciadas para os genes *GmFAD3A* e *GmFAD3B*, conforme Figura 2.

Em vista de que os genes *GmFAD3* são altamente expressos em sementes (Bilyeu et al. 2005), foram realizadas análises de RT-PCR para verificar se existe alguma mutação em região de íntron não identificada que poderia resultar em diferentes transcritos devido a variação no processo de *splicing* (Figura 3). Os resultados indicam que sementes de CS303TNKCA não apresentam formas alternativas dos genes *GmFAD3A*, *GmFAD3B* e *GmFAD3C*.

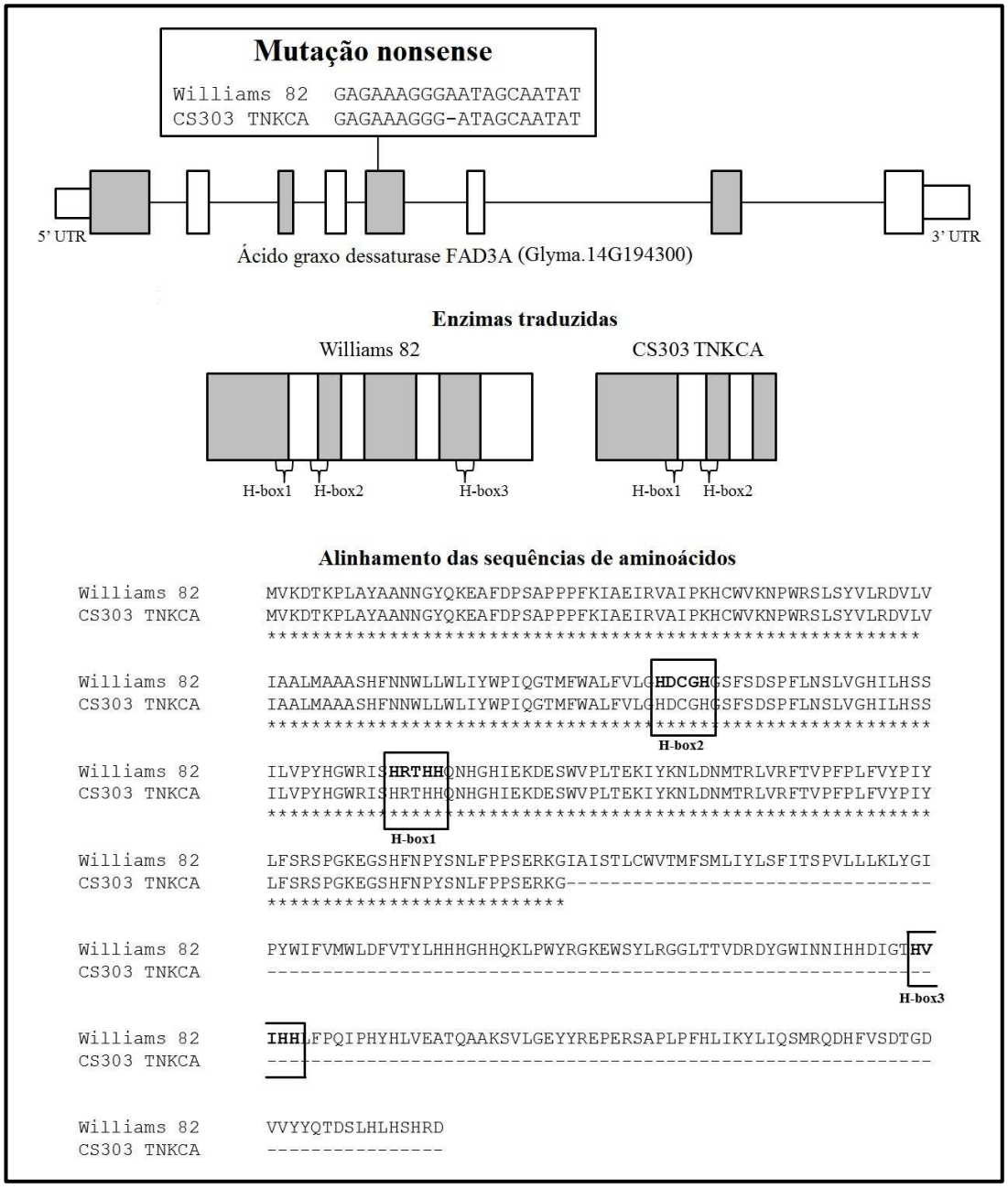

Figura 1 – Deleção de 1 pb no éxon 5 do gene *GmFAD3A* da variedade de soja CS303TNKCA (delA), levando a um stop códon prematuro e a perda do terceiro box de histidina da enzima. Na estrutura do gene, os blocos menores representam regiões UTR, blocos maiores representam éxons, e as linhas representam regiões de íntron.

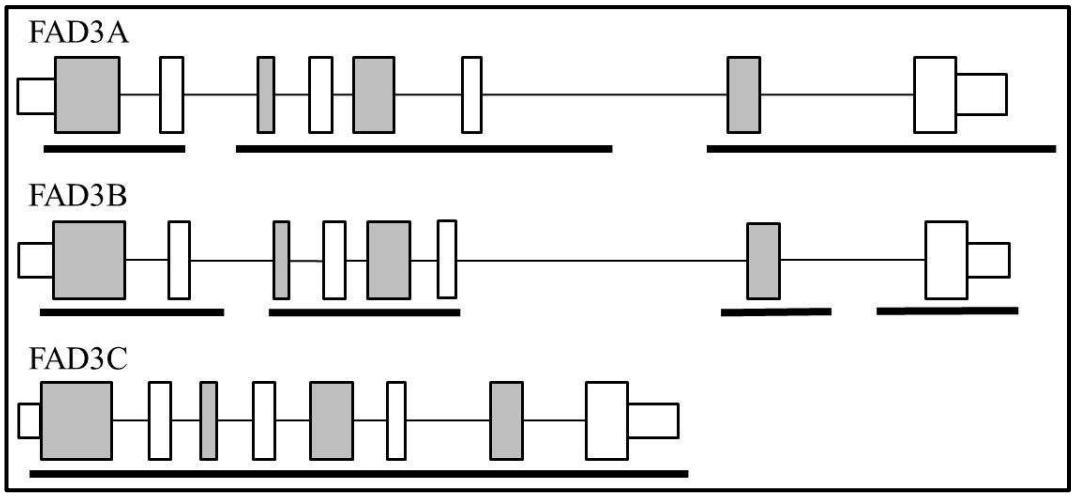

Figura 2 – Estrutura dos genes *GmFAD3A*, *GmFAD3B* e *GmFAD3C*. Os blocos menores nas extremidades indicam regiões UTR, os blocos maiores indicam éxons, e as linhas indicam íntrons. As barras abaixo de cada gene indicam a porção sequenciada de cada gene.

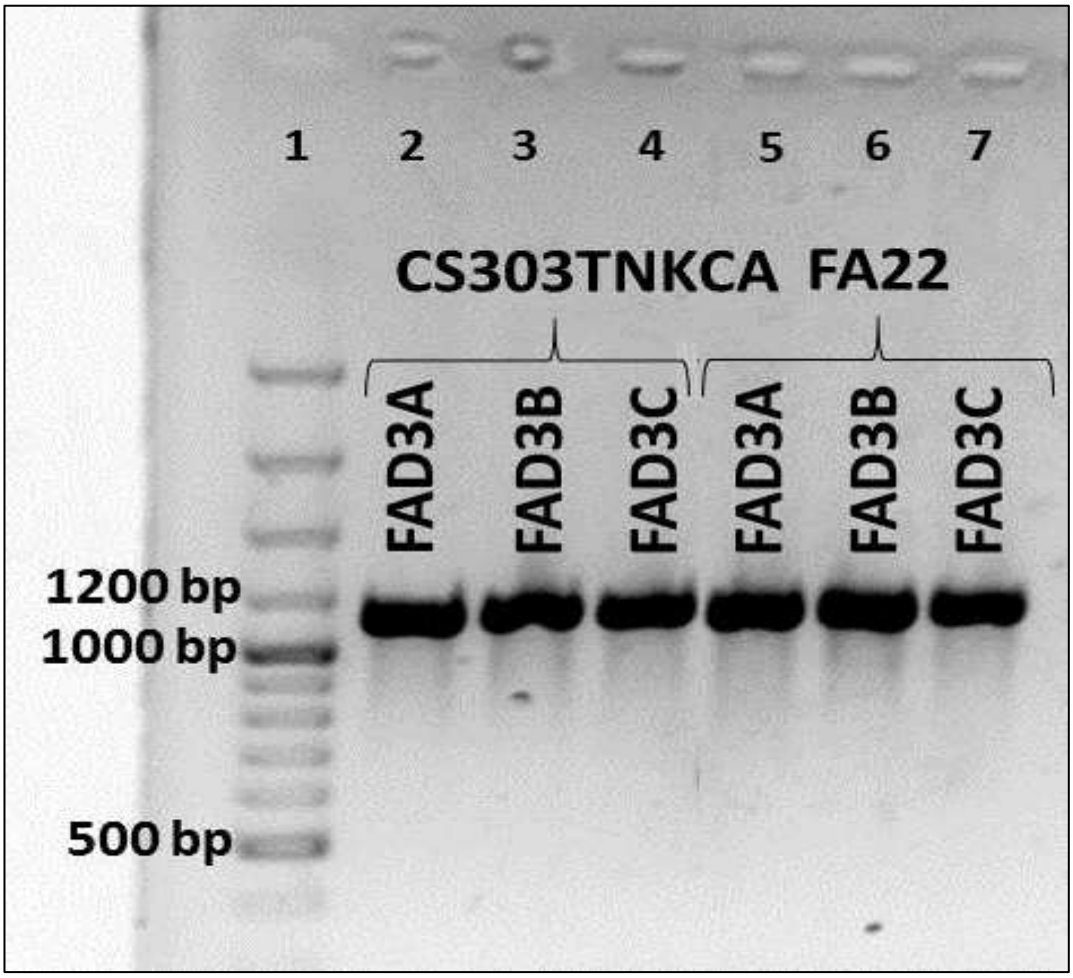

Figura 3 – RT-PCR dos genes *GmFAD3A*, *GmFAD3B* e *GmFAD3C* na variedade de baixo teor de ácido linolênico CS303TNKCA e na linhagem de teor normal de ácido linolênico FA22.

3.3. Análises genotípicas

Foi desenhado um conjunto de *primers* flanqueando a mutação delA para gerar um marcador que diferencie os indivíduos contrastantes para teor de ácido linolênico através da metodologia de HRM. Comparações entre CS303TNKCA, FA22 e indivíduos F_1 obtidos do cruzamento entre estas linhagens mostraram que o diagnóstico desenvolvido foi eficiente em diferenciar a mutação (Figura 4). A partir deste marcador baseado em HRM, 187 plantas F_2 foram genotipadas. Foram encontrados 41 indivíduos mutantes, 41 selvagens e 105 heterozigotos, obtendo um χ^2 = 2,83 ns (p = 24,3%), indicando que o marcador se comporta como esperado para segregação mendeliana.

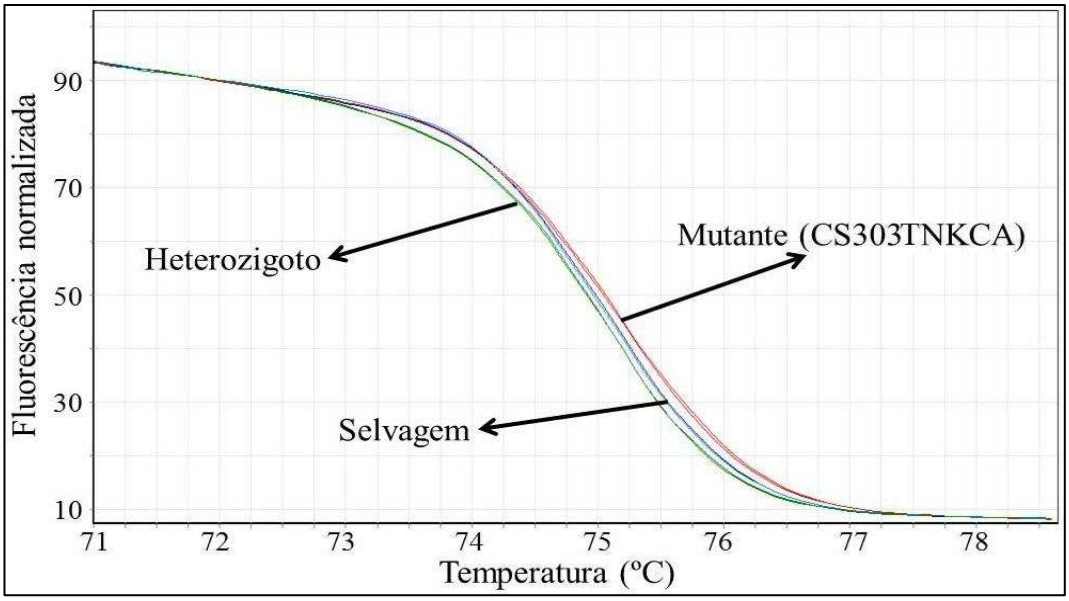

Figura 4 – Marcador HRM baseado em uma deleção de 1-pb encontrada no gene *GmFAD3A* da variedade CS303TNKCA e associada com baixo teor de ácido linolênico. As curvas normalizadas de indivíduos mutante (CS303TNKCA), selvagem e heterozigoto são mostradas.

3.4. Análises fenotípicas

Os teores dos ácidos palmítico, esteárico, oleico, linoleico e linolênico para 187 plantas F_2 derivadas do cruzamento entre CS303TNKCA e FA22, cultivadas em casa de vegetação (CV), e para 187 famílias derivadas das plantas F_2, cultivadas em Visconde do Rio Branco (VRB-$F_{2:3}$), Viçosa (VIC-$F_{2:4}$) e São Gotardo (SG1 e SG2-$F_{2:4}$) foram analisados. A Tabela 6 apresenta mínimos, máximos e médias para cada característica avaliada em cada experimento. Em todos os experimentos, o teor

de ácido linolênico variou entre 2,67% e 9,45%. A média do teor de ácido linolênico em indivíduos mutantes variou entre 3,35 % e 4,38%.

Análises de regressão linear mostraram que o efeito da mutação delA sobre o teor de ácido linolênico é predominantemente de origem aditiva, e nenhum efeito significativo de dominância foi encontrado (Tabela 7). A Figura 5 apresenta a média e a frequência dos teores de ácido linolênico para cada genótipo entre as 187 linhagens avaliados nos cinco experimentos, bem como o coeficiente de determinação da regressão "delA x teor de ácido linolênico". A mutação delA explicou entre 48,81% e 65,97% da variação no conteúdo de ácido linolênico.

Tabela 6 – Valores médios do perfil de ácidos graxos (%) de uma população segregante de soja derivada do cruzamento entre CS303TNKCA e FA22 avaliada em diferentes locais, e de plantas de genótipo mutante, selvagem e heterozigoto para a mutação delA do gene *GmFAD3A*.

	Palmítico	**Esteárico**	**Oleico**	**Linoleico**	**Linolênico**
			CV		
Mutante	10,52	3,57	21,59	60,07	4,25
Heterozigoto	10,28	3,45	21,46	58,78	6,04
Selvagem	10,14	3,33	21,24	57,64	7,66
Média	10,30±1,01	3,45±0,45	21,44±4,55	58,81±4,27	6,00±1,44
Mínimo	8,53	2,53	15,72	39,04	2,67
Máximo	19,06	6,36	42,61	65,18	9,45
			VRB		
Mutante	10,31	4,29	28,48	52,54	4,38
Heterozigoto	10,27	4,12	28,25	51,62	5,74
Selvagem	10,21	3,90	27,30	51,84	6,75
Média	10,26±0,45	4,11±0,31	28,09±3,53	51,87±3,17	5,66±0,97
Mínimo	9,06	3,31	21,47	40,52	3,86
Máximo	11,59	5,14	41,16	59,07	8,02
			VIC		
Mutante	10,10	3,25	42,02	41,28	3,35
Heterozigoto	10,02	3,16	42,44	40,33	4,06
Selvagem	10,05	3,09	40,48	41,59	4,80
Média	10,04±0,54	3,16±0,30	41,92±6,22	40,81±5,37	4,07±0,68
Mínimo	8,79	2,43	25,69	27,25	2,71
Máximo	11,62	4,01	57,74	54,03	6,35
			SG1		
Mutante	10,44	3,57	31,78	49,87	4,35
Heterozigoto	10,30	3,40	31,85	49,09	5,37
Selvagem	10,27	3,28	30,94	49,15	6,37
Média	10,32±0,54	3,41±0,28	31,63±6,14	49,27±5,33	5,36±0,93
Mínimo	8,95	2,62	18,89	35,29	3,72
Máximo	11,82	4,28	47,51	60,99	8,23
			SG2		
Mutante	10,46	3,54	31,70	50,01	4,30
Heterozigoto	10,31	3,40	31,92	49,00	5,38
Selvagem	10,28	3,27	30,87	49,18	6,40
Média	10,33±0,59	3,40±0,30	31,64±6,68	49,26±5,84	5,37±1,00
Mínimo	8,67	2,68	19,25	30,83	3,66
Máximo	12,11	4,38	52,37	60,97	8,41

CV = Casa de vegetação (Viçosa-MG); **VRB** = Visconde do Rio Branco-MG; **VIC** = Viçosa-MG; **SG** = São Gotardo-MG

Tabela 7 – Variância dos componentes aditivos e dominantes do teor de ácido linolênico explicados pela mutação delA da variedade CS303TNKCA, calculados através de regressão linear.

Fonte de Variação	F	Prob (%)	R^2 (%)
	GH		
Completo	147,71	0	61,62
Aditivo	296,09	0	61,55
Dominante	0,14	100 ns	0,07
	VRB		
Completo	184,77	0	66,76
Aditivo	358,67	0	65,97
Dominante	1,47	0,23 ns	0,79
	VIC		
Completo	91,61	0	49,89
Aditivo	184,09	0	49,88
Dominante	0,03	100 ns	0,02
	SG1		
Completo	101,05	0	52,34
Aditivo	203,18	0	52,34
Dominante	0,00	100 ns	0
	SG2		
Completo	87,83	0	48,84
Aditivo	176,43	0	48,81
Dominante	0,05	100 ns	0,03

Completo = soma dos modelos aditivo e dominante.

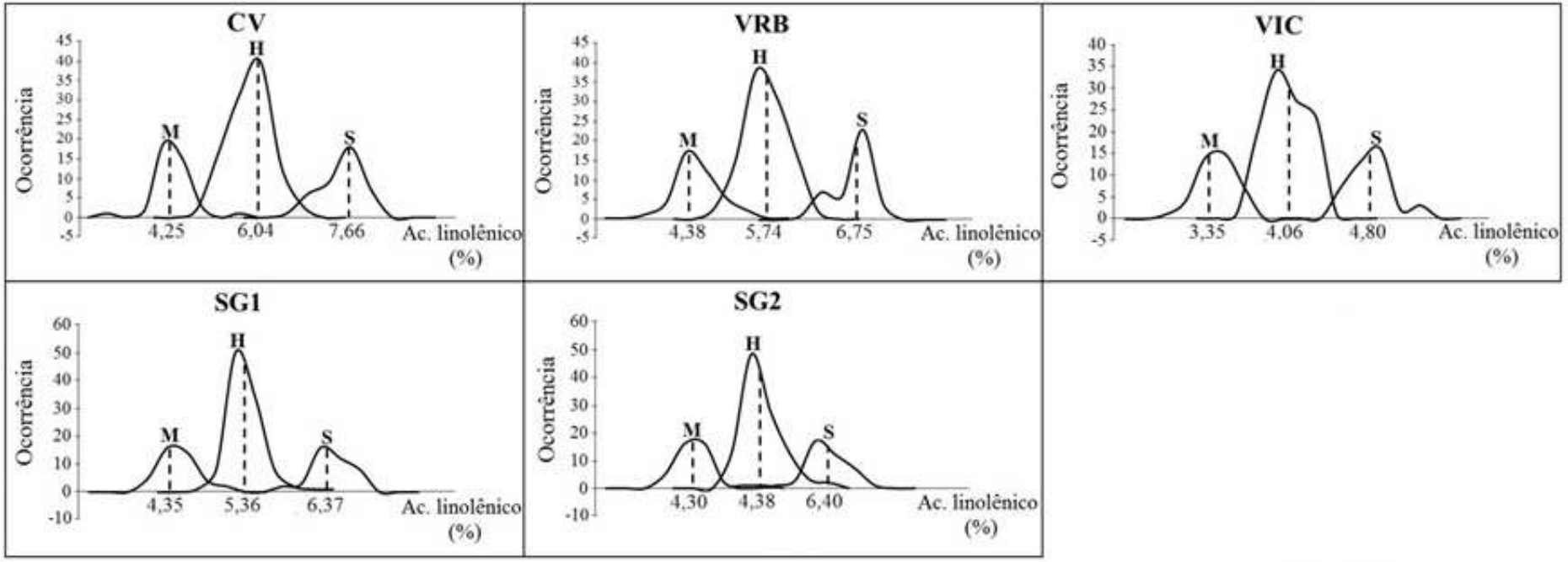

Figura 5 – Gráficos de distribuição de frequência do conteúdo de ácido linolênico de 187 linhas de soja de uma população segregante, separados pelos genótipos para o gene *GmFAD3A*, avaliados em diferentes locais. As médias do conteúdo de ácido linolênico para cada genótipo são mostrados. As letras acima das curvas identificam os genótipos: mutante (M), heterozigoto (H) e selvagem (S). Os locais avaliados foram: CV (casa de vegetação – Viçosa-MG), VRB (Visconde do Rio Branco-MG), VIC (Viçosa-MG), SG (São Gotardo-MG).

3.5. Análise de atividade enzimática

Por meio de expressão heteróloga em *S. cerevisiae* cepa W303, foi avaliada a atividade de ω3-dessaturases codificadas pelo alelo mutante *GmFAD3A* da variedade CS303TNKCA, e pelo alelo selvagem da planta PL04, desenvolvida através do cruzamento entre os acessos PI283327 e PI603452 (ver item 3.2 do Apêndice I). O perfil de ácidos graxos das colônias de leveduras transformadas está apresentado na Tabela 8. Colônias transformadas com o alelo *GmFAD3A* selvagem foram capazes de produzir 9,86% de ácido linoleico, e 4,21% de ácido linolênico. Por outro lado, leveduras transformadas com o alelo *GmFAD3A* contendo delA produziram em média 2,42% de ácido linoleico e não produziram ácido linolênico em níveis detectáveis, semelhante ao encontrado para leveduras selvagens (sem inserção do gene), produziram em média 2,30% de ácido linoleico e não produziram ácido linolênico em níveis detectáveis.

Tabela 8 – Perfil médio de ácidos graxos de leveduras selvagens e transformadas com o gene *GmFAD3A*.

		Ácidos Graxos (%)					
Colônia	**Doador do alelo**	**Palm**	**Pole**	**Este**	**Olei**	**Lole**	**Lico**
Selvagem	-	14,66	36,27	7,73	39,04	2,30	ND
GmFAD3A	PL04	16,92	22,30	13,71	34,04	8,48	4,55
GmFAD3A delA	CS303TNKCA	15,53	36,35	10,57	35,14	2,42	ND

Palm = ácido palmítco; **Pole** = ácido palmitoleico; **Este** = ácido esteárico; **Olei** = ácido oleico; **Lole** = ácido linoleico; **Lico** = ácido linolênico; **ND** = não detectável.
Valores calculados a partir da média de três análises, cada uma utilizando três colônias.

3.6. Análises de expressão gênica

A partir da análise de expressão gênica, verificou-se que o gene *GmFAD2-1A* foi mais expresso em plantas com teores normais de ácido oleico, em relação à isoforma *GmFAD2-1B* (Figura 6A). A linhagem FA22 apresentou níveis de expressão similares ao de outras linhagens que apresentam teores normais de ácido oleico. Por outro lado, a PI603452 e plantas derivadas apresentaram baixos níveis de expressão do gene *GmFAD2-1A*. Não foi verificado um padrão de expressão para o gene *GmFAD2-1B* entre indivíduos com teores normais e médios de ácido

oleico. Linhagens com alto teor de proteína (BARC-8, NT12 e BR80) e baixo teor de óleo apresentaram níveis aparentemente mais baixos de expressão dos genes *GmFAD2-1*, principalmente a linhagem NT12.

A análise da expressão dos genes *GmFAD3* (Figura 6B) mostrou padrões relativamente próximos de expressão entre linhagens de baixo teor de ácido linolênico (CS303TNKCA e A29) em relação a outras com teores normais desta característica (NT12, BARC8, BR80, CD01 RR, FA22, PI283327 e PI603452). Entretanto, foi observada uma maior expressão do gene *GmFAD3A* em relação aos genes *GmFAD3B* e *GmFAD3C*, em plantas com alto teor de ácido oleico, derivadas do cruzamento entre as linhagens PI603452 e PI283327 (PL04, PL21 e PL126). A cultivar Suprema, que apresenta alto teor de óleo, apresentou maiores níveis de expressão para os genes *GmFAD3A* e *GmFAD3C*.

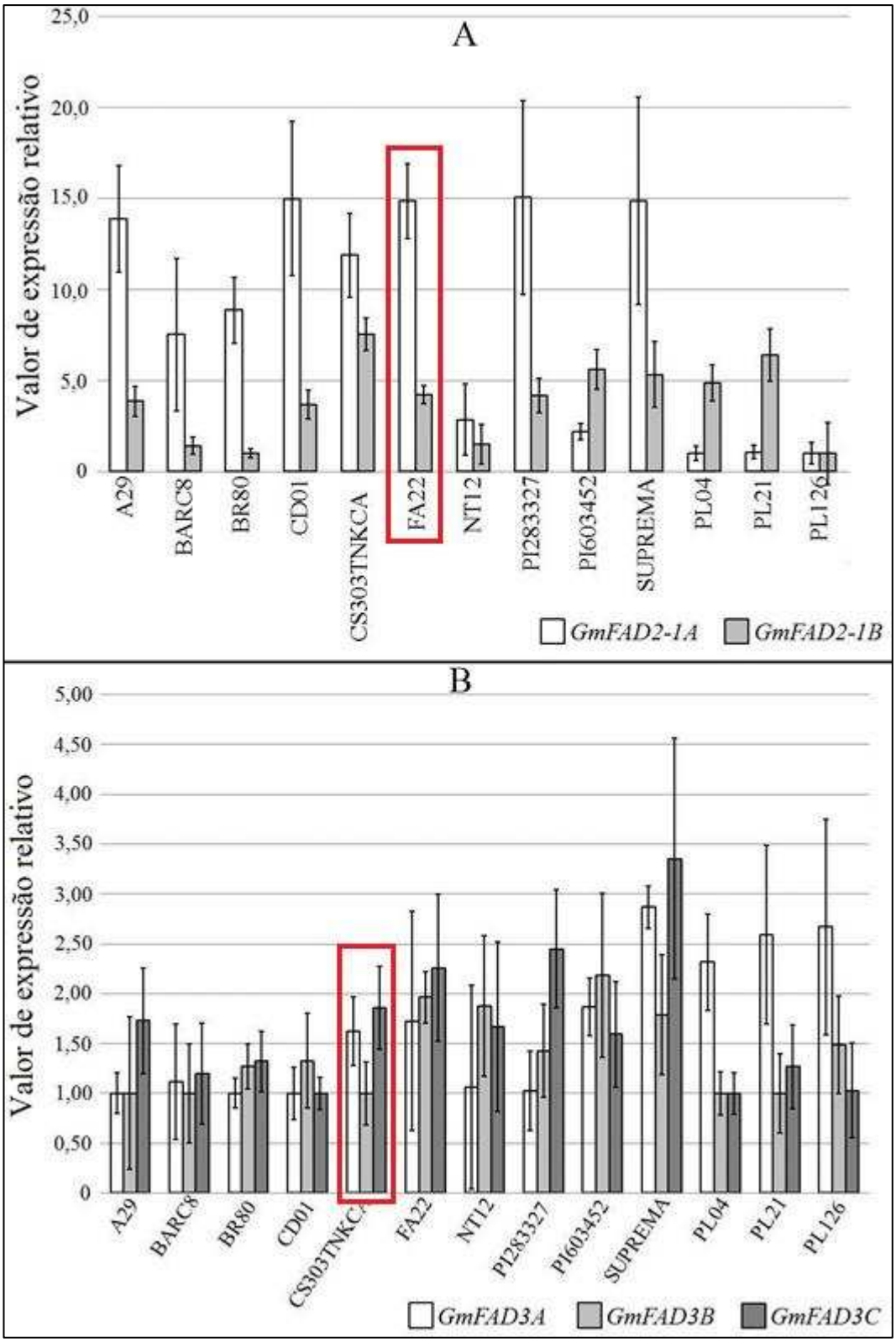

Figura 6 – Padrão de expressão gênica de sementes em desenvolvimento para os genes *GmFAD2-1* (**A**) e *GmFAD3* (**B**) de diferentes acessos de soja, utilizando os genes *Cons7* e *UKN2* como controles endógenos, com destaque para a linhagem FA22 (**A**) e para a variedade CS303TNKCA (**B**)

4. DISCUSSÃO

O BIOAGRO/UFV possui um programa de melhoramento (PMQS) que busca o desenvolvimento de soja com reduzido teor de ácido linolênico e elevado teor de ácido oleico, características desejáveis para alimentação e indústria de biodiesel no Brasil. A variedade CS303TNKCA apresenta baixo teor de ácido linolênico (cerca de 3,5-4,0%), e foi desenvolvida pelo PMQS, derivada da cultivar Monarca (COOPADAP) e da linhagem BARC-12 (Leffel 1994). CS303TNKCA também apresenta ausência das três lipoxigenases (*lox1*, *lox2* e *lox3*) e do inibidor de protease Kunitz, além de médio teor de óleo (cerca de 20-23%) e rendimento por volta de 3.600 kg de grãos por hectare.

Neste estudo, procuramos elucidar as bases genéticas do baixo teor de ácido linolênico da variedade CS303TNKCA, e do médio teor de ácido oleico (cerca de 50%) da linhagem FA22, desenvolvida por mutagênese na Iowa State University, e que apresenta conteúdo normal de ácido linolênico (5-6%; Alt et al. 2005).

Consistente com trabalhos prévios que demonstraram que o conteúdo médio de ácido oleico presente na linhagem FA22 não é condicionado pela deleção do gene *GmFAD2-1A* presente na linhagem M23 (Alt et al. 2005), na presente investigação, não foram detectadas mutações no gene *GmFAD2-1A* da linhagem FA22. Entretanto, o sequenciamento completo dos genes *GmFAD2-1A* e *GmFAD2-1B*, também não permitiu identificar alguma mutação que pudesse explicar o conteúdo médio de ácido oleico apresentado pela FA22, o que implica na necessidade de novos estudos para explicar a característica desta linhagem.

Análises de expressão gênica de sementes em estágio de desenvolvimento R5 mostraram padrões de expressão dos genes *GmFAD2-1A* e *GmFAD2-1B* da linhagem FA22 similares aos encontrados para outros acessos com teores normais de ácido oleico (Figura 6A). Além disso, o gene *GmFAD2-1A* foi o mais expresso em todos materiais analisados que apresentam teores normais de ácido oleico. Tem sido demonstrado que a isoforma GmFAD2-1A é menos estável que *GmFAD2-1B* (Tang et al. 2005), sendo provável então que apresente maiores níveis de expressão para compensar sua maior instabilidade. Aparentemente, a mutação caracterizada por Pham et al. (2011) no gene *GmFAD2-1A* da linhagem PI603452, que leva a um stop códon prematuro e terminação da tradução após 191 resíduos, é capaz de alterar

os níveis de expressão do gene, haja vista que a PI603452 e plantas derivadas do cruzamento desta com a linhagem PI283327 (PL04, PL21 e PL126 – ver item 3.2 do Apêndice I) apresentaram menores níveis de expressão para este gene.

A expressão dos genes *GmFAD3* (Figura 6A) em linhagens com baixo teor de ácido linolênico (CS303TNKCA e A29) foi similar ao encontrado em outras linhagens de teores normais desta característica. Foi observado maior nível de expressão do gene *GmFAD3A* em relação aos genes *GmFAD3B* e *GmFAD3C* em plantas com alto teor de ácido oleico, derivadas do cruzamento entre PI603452 e PI283327; entretanto, não foi identificada a causa deste fenômeno. Apesar de a mutação caracterizada no gene *GmFAD3A* da variedade CS303TNKCA ser semelhante àquela caracterizada por Pham et al. (2011) na linhagem PI603452 (produção de uma enzima com 191 resíduos faltantes, incluindo o H-box3, na CS303TNKCA), delA parece não alterar a expressão do gene, diferentemente da mutação na linhagem PI603452 e plantas derivadas, nas quais nota-se uma menor expressão relativa do gene *GmFAD2-1A*.

Os genes *GmFAD2-1* codificam para ω6-dessaturases expressas em sementes em desenvolvimento, enzimas que convertem ácido oleico em ácido linoleico. Por meio do sequenciamento destes genes e de análises de expressão, não foi possível identificar as bases genéticas para o médio teor de ácido oleico (cerca de 50%) na linhagem FA22. É possível que este fenótipo seja explicado por outros mecanismos, como regulação pós-tranducional. Durante o período de formação da semente, foi demonstrado que ω6-dessaturases codificadas pelos genes *GmFAD2-1A* e *GmFAD2-1B* de soja podem ser fosforiladas no sítio Ser182, um mecanismo comum de regulação da atividade enzimática (Tang et al. 2005). Até o momento, as bases genéticas para o teor médio de ácido oleico da linhagem FA22 permanecem desconhecidas, e outras abordagens devem ser consideradas visando elucidar esta questão.

Por outro lado, foi identificada uma deleção de 1 pb (delA) no gene *GmFAD3A*, que codifica para a ω3-dessaturase na variedade CS303TNKCA (Figura 1). A mutação delA resulta em uma proteína truncada faltando 169 resíduos, um fragmento que compreende a região do terceiro box de histidina (H-box3) presente na enzima. Resultados similares foram encontrados por Reinprecht et al.

(2009), que caracterizaram uma troca (G>A) na posição 798 da região codificante do mesmo gene da linhagem RG10. Esta mutação leva a um stop códon prematuro e uma proteína truncada faltando 111 resíduos, incluindo o H-box3, e é responsável por reduzir o teor de ácido linolênico a cerca de 3,4%. As ω3-dessaturases são responsáveis pela síntese de ácido linolênico a partir do ácido linoleico, e pertencem a uma família de enzimas caracterizadas pela presença de um cofator *di-iron* que interage com as três regiões de motivos conservados de histidina (H-box) na proteína (Shanklin et al. 1994). Estes boxes de histidina das dessaturases de membrana são essenciais para coordenar o cofator *di-iron* requerido para a atividade catalítica (Shanklin et al. 1994, McCartney et al. 2004), e a mudança de um único resíduo dentro destes boxes de histidina pode reduzir a atividade da enzima (McCartney et al. 2004, Bilyeu et al. 2006, Radovanovic et al. 2014, Reinprecht & Pauls 2016).

Análises de atividade enzimática a partir da expressão heteróloga do gene *GmFAD3A* em *S. cerevisiae* mostraram que, enquanto colônias transformadas com o gene selvagem clonado da PL04 (ver item 3.2 do Apêndice I) foram capazes de produzir ácido linoleico (8,48%) e ácido linolênico (4,55% em média), colônias transformadas com o gene mutante da variedade CS303TNKCA produziram apenas 2,42% de ácido linoleico e não foram capazes de produzir ácido linolênico em níveis detectáveis no estudo (Tabela 8). Estratégia similar foi utilizada anteriormente por Singer et al. (2014) para avaliar a atividade de uma enzima mutante de *Brassica oleracea*, codificada pelo gene *BoFAD3-2*. A comparação da expressão deste gene mutante e de um gene *BoFAD3-2* selvagem mostrou que a enzima mutante apresenta um declínio drástico na atividade Δ15-dessaturase, levando à baixíssima produção de ácido linolênico.

Análises de expressão gênica (Figura 6B) não mostraram alteração no padrão de expressão dos genes *GmFAD3* da variedade CS303TNKCA, em relação a outras linhagens com teores normais de ácido linolênico, ou em relação à A29, linhagem com baixíssimo teor de ácido linolênico (cerca de 1%). Além disso, análises de RT-PCR não indicaram alteração no tamanho dos transcritos produzidos pelos genes *GmFAD3* em sementes em desenvolvimento da variedade CS303TNKCA, a partir de *primers* amplificando toda a região codificante (do códon inicial até o final; Figura 3), de modo que não há evidências de alteração no

processo de *splicing*. Os resultados encontrados na análise de atividade enzimática, expressão gênica e RT-PCR corroboram a hipótese de que delA é a responsável pelo baixo teor de ácido linolênico apresentado pela variedade CS303TNKCA.

Foi desenvolvido um diagnóstico para detectar a mutação delA no gene *GmFAD3A* da variedade CS303TNKCA, através da metodologia de HRM (Figura 4), uma técnica que tem sido utilizada com sucesso no mapeamento e genotipagem de várias espécies (Liew et al. 2004, Wu et al. 2008, Cruz et al. 2013, Simko et al. 2016). A genotipagem das plantas F_2 derivadas do cruzamento entre CS303TNKCA e FA22 foi utilizada para associar delA com o conteúdo de ácido linolênico em três gerações. A mutação levou à redução do teor de ácido linolênico para 3,4-4,4% na população segregante avaliada em diferentes locais (Figura 5). Estudos disponíveis na literatura têm reportado resultados similares: uma troca (G>A) no primeiro nucleotídeo do íntron 6 da linhagem CX1512-44 leva a uma alteração no processo de *splicing* (Bilyeu et al. 2005); uma substituição (G>A) na posição 798 da sequência codificante da linhagem RG10 leva a um stop códon prematuro (Reinprecht et al. 2009); uma deleção de 6,4 kpb no gene *GmFAD3A* foi observado na linhagem A5 (Pham et al. 2014); uma substituição (G>A) no segundo sítio de *splicing* da linhagem PE1690 leva a um stop códon prematuro (Kim et al. 2015). Neste estudo, a mutação delA foi responsável por reduzir entre 30,11 e 44,48% do teor de ácido linolênico em sementes de soja, explicando 48,81-65,75% da variação na característica.

Cultivares de soja comumente apresentam cerca de 20% de óleo, que contém em média 10% de ácido palmítico (16:0), 4% de ácido esteárico (18:0), 20 a 25% de ácido oleico (18:1 Δ^9), 55-60% de ácido linoleico (18:2 $\Delta^{9,12}$) e 7 a 10% ácido linolênico (18:3 $\Delta^{9,12,15}$). Óleos de fritura estáveis geralmente necessitam de baixos teores de ácidos linolênico (<3%) e linoleico (<50%) e alto teor de ácido oleico (>50%) (Warner & Fehr 2008). Os genes *GmFAD2-1A* e *GmFAD2-1B* codificam para as ω6-dessaturases, que são responsáveis pela conversão de ácido oleico em linoleico em sementes em desenvolvimento (Heppard et al. 1996, Schlueter et al. 2007), enquanto os genes *GmFAD3A*, *GmFAD3B* e *GmFAD3C* codificam para as ω3-dessaturases, que convertem ácido linoleico em linolênico em sementes em desenvolvimento (Bilyeu et al. 2003). A combinação de mutações nos genes *GmFAD2-1A* e *GmFAD2-1B* eleva o teor de ácido oleico para cerca de 80-

85%, e reduz o teor de ácido linolênico para cerca de 2,5-4,9% (Pham et al. 2011, Pham et al. 2012). Linhagens mutantes contendo os genes *GmFAD2-1A*, *GmFAD2-1B* e *GmFAD3A* não funcionais produziram grãos com mais de 80% de ácido oleico, 1,9-7,0% de ácido linoleico e 1,8-3,5% de ácido linolênico (Pham et al. 2012). Neste caso particular, o uso de somente o gene *GmFAD3A* é suficiente para apresentar um grande efeito sobre o teor de ácido linolênico (Pinto et al. 2003), e deve-se considerar que a queda no fluxo metabólico da produção de ácido linoleico por dessaturases *GmFAD2-1A* e *GmFAD2-1B* mutantes irá afetar a quantidade de substrato disponível para a ω3-ácido graxo dessaturase (*GmFAD3*).

O marcador HRM para baixo teor de ácido linolênico baseado na mutação delA se provou eficiente para seleção assistida por marcadores em programas de melhoramento, reduzindo o conteúdo de ácido linolênico para 3,4-4,4%. Atualmente, o PMQS está desenvolvendo um programa de retrocruzamentos assistidos buscando desenvolver variedades de soja combinando a mutação delA da variedade CS303TNKCA com as mutações detectadas nos genes *GmFAD2-1A* e *GmFAD2-1B* de uma planta com cerca de 85% de ácido oleico derivada do cruzamento das linhagens PI283327 e PI603452 (Pham et al. 2011 – ver Apêndice I). Espera-se desta forma produzir cultivares de soja com alto teor de ácido oleico e baixo teor dos ácidos linoleico e linolênico, gerando óleo com boa estabilidade oxidativa e aumentando os benefícios para a saúde humana e a indústria brasileira de biodiesel.

5. CONCLUSÕES

- Não foi possível caracterizar as bases genéticas do médio conteúdo de ácido oleico na linhagem FA22;
- Foi identificada uma deleção de uma adenina (delA) na sequência codificante do gene *GmFAD3A* da variedade CS303TNKCA, que está associada com redução no teor de ácido linolênico para 3,4-4,4%, explicando entre 48,81% e 65,97% da variação na característica;
- O marcador HRM baseado em delA se mostrou eficiente para realizar a discriminação de alelos do gene *GmFAD3A*;

- A mutação delA se mostra promissora para o desenvolvimento de cultivares de soja com óleo de maior estabilidade e qualidade nutricional.

CAPÍTULO II

EFEITO DE UMA MUTAÇÃO NO GENE RAFINOSE SINTASE 2 (*GmRS2*) EM CARACTERÍSTICAS DE QUALIDADE DO GRÃO DE SOJA

1. INTRODUÇÃO

A soja (*Glycine max*) é considerada como a fonte de proteína vegetal com maior impacto na alimentação de animais e humanos, devido ao seu alto conteúdo proteico, e por sua composição de aminoácidos razoavelmente balanceada (Gu et al. 2010, Ari 2014). Apesar de seu potencial para uso na alimentação ser conhecido há milênios pelos povos da China, somente mais recentemente a soja passou a ser considerada no ocidente como alimento funcional (alimento que, além das funções nutricionais básicas, produz efeitos benéficos para a saúde; EMBRAPA; https://www.embrapa.br/web/portal/soja/cultivos/soja1/soja-na-alimentacao).
Assim, com o passar dos anos, a demanda por sojas especiais tem aumentado devido ao uso do grão diretamente, ou para produção de derivados, como proteína texturizada, tofu, leite de soja, natto e edamame (Wang et al. 2014, O'Keefe et al. 2015).

Além do alto teor de proteína, outros fatores também são responsáveis pela qualidade do grão de soja, como o conteúdo de oligossacarídeos. Sacarose é o único oligossacarídeo presente no grão de soja considerado nutricionalmente útil para digestão de humanos e animais monogástricos (Yang et al. 2014), e seu conteúdo no grão é um fator crítico para produzir derivados de soja (Kumar et al. 2011, Sato et al. 2014, Wang et al. 2014). Além disso, sacarose é provavelmente o principal fator que influencia o sabor de *vegetable soybean* (soja vagem; Li, Y. et al. 2012b).

Por outro lado, a presença de oligossacarídeos de rafinose (ROs) em soja, que inclui os α-galactosídeos rafinose e estaquiose, é considerada um fator antinutricional, já que animais monogástricos não possuem α-galactosidase, uma enzima responsável pela hidrolização de ROs (Hagely et al. 2013, Yang et al. 2014). A ingestão destes ROs na ausência de α-galactosidase pode causar disfunções orgânicas como diarreia, náuseas e flatulência (Liener 1994, Reddy et al. 2016). Estaquiose é o segundo principal açúcar presente em soja, usualmente variando entre 1,4-4,1% do grão (Hymowitz et al. 1972, Zeng et al. 2015), e sua presença no grão é considerada restritiva ao consumo de soja como fonte de proteína. A redução do teor de estaquiose e rafinose aumenta a energia metabolizável e promove diminuição em seus efeitos indesejáveis (Suarez et al. 1999, Parsons et al. 2000, Dierking & Bilyeu 2008).

Vários procedimentos para reduzir ou eliminar o conteúdo de estaquiose no grão de soja e produtos derivados estão descritos na literatura, tais como embebição e germinação (Kim et al. 1973); processos de fermentação (Mital & Steinkraus 1975); extração de oligossacarídeos em água (Ku et al. 1976); ultrafiltração do extrato hidrossolúvel de soja (Omosaiye et al. 1978); extração dos oligossacarídeos com etanol a partir da farinha de soja (Leske et al. 1991); utilização de α-galactosidase de origem vegetal e microbiana (Guimarães et al. 2001), dentre outros. Entretanto, a produção de variedades de soja com baixo teor de estaquiose, associado com altos teores de proteína e sacarose, é uma alternativa interessante para aumentar a qualidade nutricional da soja no mercado (Sato et al. 2014), uma vez que estas variedades reduziriam o tempo e custos para processar o grão e seus derivados.

A biossíntese de rafinose e estaquiose é mediada por enzimas rafinose sintase, que realizam a transferência de resíduos de galactosil do galactinol para a sacarose (Dierking & Bilyeu 2008, Qiu et al. 2015, Bilyeu & Wiebold 2016). Skoneczka et al. (2009) caracterizaram uma deleção de três pares de bases no gene rafinose sintase 2 (*GmRS2* – Glyma.06G179200) do acesso PI200508, que possui alto teor de sacarose e baixo teor de estaquiose. Esta mutação foi analisada em duas populações F_2 derivadas da PI200508, e explicou 88-94% e 76% da variação nos teores de estaquiose e sacarose, respectivamente. O baixo teor de ROs em linhagens derivadas da PI200508 não afeta o vigor da semente e tampouco as características de qualidade do grão, tais como emergência no campo, rendimento de sementes, maturidade, peso e conteúdo de ácidos graxos. Estas características são importantes para a qualidade nutricional da soja (Yang et al. 2004), tornando a PI200508 um bom doador de alelos para baixo conteúdo de ROs.

Correlações entre características de qualidade do grão têm sido investigadas por muitos pesquisadores (Hymowitz et al. 1992, Hartwig et al. 1997, Wilcox & Shibles 2001, Jaureguy et al. 2011, Sato et al. 2014, Yu et al. 2016). Como as sínteses de sacarose, rafinose e estaquiose estão na mesma via metabólica (Dierking & Bilyeu 2008, Bilyeu & Wiebold 2016), é esperado alcançar um aumento no teor de sacarose pela inibição da síntese de ROs, sem um decréscimo no teor de proteína (Sato et al. 2014). Deste modo, o objetivo deste trabalho foi validar um novo

marcador molecular baseado na mutação encontrada no acesso PI200508, e estudar seu efeito em características de qualidade do grão de soja.

2. MATERIAL E MÉTODOS

2.1. Material vegetal e desenvolvimento da população

Foi desenvolvida uma população F_2 com 168 indivíduos, a partir de cruzamentos diferentes acessos. PI283327 foi cruzada com PI603452, e a F_1 gerada foi cruzada com PI200508. A nova F_1 foi cruzada com a cultivar NA5909. As plantas geradas foram genotipadas para o gene rafinose sintase 2, e aquelas heterozigotas foram selecionadas para formar a população F_2. PI200508 apresenta uma mutação associada com redução no teor de estaquiose e aumento no teor de sacarose (Skoneczka et al. 2009). Os acessos PI603452 e PI283327 possuem mutações nos genes *GmFAD2-1A* e *GmFAD2-1B,* respectivamente, que quando combinadas podem produzir soja com cerca de 85% de ácido oleico (Pham et al. 2011, Pham et al. 2012). A cultivar NA5909 (Nidera Sementes) é uma variedade de alto desempenho agronômico. Todos os cruzamentos e o desenvolvimento da população F_2 foram realizados em casa de vegetação na Universidade Federal de Viçosa, em Viçosa-MG. Sementes F_2 foram plantadas em junho de 2015 e colhidas em outubro do mesmo ano. Os genitores da população e as variedades UFVTN105AP (alta proteína – Programa de Melhoramento da Qualidade da Soja – PMQS/BIOAGRO/UFV), CS303TNKCA (baixo teor de ácido linolênico – PMQS/BIOAGRO/UFV) e Tucunaré (boas características agronômicas – Fundação Mato Grosso) foram utilizados como testemunhas adicionais. Sementes das plantas F_2, dos genitores e das testemunhas adicionais foram colhidas e utilizadas para análises fenotípicas.

2.2. Análises genotípicas

Amostras foliares dos genitores PI603452, PI283327, PI200508, NA5909, plantas F_1 e de todas as plantas F_2 foram coletadas em estágio de desenvolvimento V2 (Fehr & Caviness 1977), congeladas em nitrogênio líquido e armazenadas em freezer a -80 ºC. DNA genômico foi extraído conforme metodologia proposta por

Doyle & Doyle (1990). A concentração do DNA foi determinada utilizando um espectrofotômetro NanoDrop (NanoDrop Technologies, Wilmington, DE), e sua qualidade foi conferida por eletroforese em gel de agarose 0,8%.

O conjunto de *primers* desenvolvido por Skoneczka et al. (2009) (Forward 5'-GGACTTGAAGGAACAGTTTAGG-3'; e Reverse 5'-CGTTACTGACGATCTTATCCAC-3') foi avaliado para verificar sua adequação à metodologia de genotipagem *High Resolution Melt* (HRM) (Liew et al. 2008, Simko 2016). Posteriormente, o mesmo conjunto de *primers* foi utilizado para sequenciar a região da mutação do gene *GmRS2* nos quatro genitores utilizados no desenvolvimento da população. As sequências geradas foram editadas e alinhadas nos *softwares* Sequencher v. 4.1.4 (Gene Codes Corporation) e ClustalW (Thompson et al. 2002). A partir destas sequências, foi gerado um novo conjunto de *primers* HRM (Forward 5'-GTGGAGCAGGTGTATGTG-3', e Reverse 5'-GTCTGACCCCACCCCAATAC-3). As reações de genotipagem da população F_2 foram realizadas conforme item 2.4 do Capítulo I.

2.3. Análises fenotípicas

Os teores de umidade, óleo e proteína foram determinados utilizando espectroscopia de infravermelho próximo (NIR – FT-NIR analyzer, Thermo Scientific, model Antaris II), utilizando 20-30 sementes moídas. A extração dos oligossacarídeos sacarose, rafinose e estaquiose foi realizada conforme metodologia adaptada de Teixeira et al. (2012). Na primeira etapa, 50 mg de grãos de soja moídos foram colocados em tubos de centrifugação rosqueáveis, adicionados com 1,0 mL de éter de petróleo, agitadas em vortex e mantidos em banho-maria a 42 ºC sob agitação constante. Após 5 minutos, as amostras foram centrifugadas a 14.000 giros por 5 minutos, e o sobrenadante foi descartado. Esta etapa foi realizada três vezes. Na segunda etapa, foi adicionado 1 mL de etanol 80%, as amostras foram agitadas em vortex e levadas para banho-maria em ebulição sob agitação constante por 5 minutos. Após serem resfriadas a temperatura ambiente, as amostras foram agitadas em vortex novamente, centrifugadas a 14.000 giros por 5 minutos, e a fase líquida foi coletada em béquer de 10 mL. Após realizar a segunda etapa três vezes, os béqueres foram levados para secar em estufa a 50 ºC

por cerca de 14 horas. Por fim, os oligossacarídeos presentes no fundo do béquer foram ressuspendidos em 1 mL de etanol 80 %. O extrato foi analisado por cromatografia líquida de alta performance (HPLC) em um cromatógrafo Shimadzu Prominence, utilizando acetonitrila como fase móvel. A extração de lipídeos foi realizada conforme metodologia adaptada de Gesteira et al. (2003) (ver Capítulo I), e o perfil de ácidos graxos foi determinado por cromatografia gasosa em um cromatógrafo Shimadzu GC-2010 Plus.

2.4. Análises estatísticas

Para verificar a segregação da mutação do gene *GmRS2* na população F_2, foi utilizado o teste χ^2. Estatísticas descritivas (média, máximo, mínimo, amplitude, variância e desvio padrão) foram calculadas para cada genótipo e para cada característica. O Coeficiente de Correlação de Pearson foi calculado para cada par de características, e sua significância foi verificada por teste t. Regressão linear foi realizada para estimar o efeito da mutação no gene *GmRS2* em todas características avaliadas. A significância do coeficiente da regressão (β_1) foi verificada por teste t, e o ajuste do modelo linear para cada regressão foi quantificado por meio do coeficiente de determinação (R^2). As análises estatísticas foram realizadas no *software* GENES (Cruz 2013), com exceção das análises de regressão, que foram realizadas no *software* GQMOL (Cruz & Schuster 2004). Foram construídos gráficos de distribuição de frequência dos teores de sacarose, rafinose e estaquiose para indivíduos mutantes, selvagens e heterozigotos da população F_2, evidenciando suas médias. Estes gráficos foram construídos utilizando os *softwares* Microsoft Excel e Microsoft Power Point (Microsoft Corporation).

3. RESULTADOS E DISCUSSÃO

3.1. Desenvolvimento da população e metodologia de genotipagem

A população utilizada no estudo foi desenvolvida pelo cruzamento entre quatro genitores, como descrito na metodologia. O conjunto de *primers* descrito por Skoneczka et al. (2009) foi utilizado para genotipar e selecionar indivíduos da população F_2 contendo a mutação no gene *GmRS2* (Glyma.06G179200), através da

metodologia de HRM. No entanto, o marcador não mostrou segregação mendeliana na população avaliada, provavelmente devido a alguma mutação não identificada na região de amplificação dos genitores da população, que poderia levar alterações na curva de *melt*. Para avaliar esta hipótese, os fragmentos amplificados pelos *primers* desenvolvidos por Skoneczka et al. (2009) foram sequenciados para todos os genitores da população (PI200508, PI283327, PI603452 e NA5909), gerando fragmentos variando entre 182 e 185 pares de bases.

A partir da análise das sequências geradas, a deleção de três pares de bases no acesso PI200508 foi reidentificada. Entretanto, foi encontrada uma mutação silenciosa (C>T) na posição 25 da PI603452 (Figura 1), acesso utilizado para o desenvolvimento de soja com mais de 80% de ácido oleico (Pham et al. 2011, Pham et al. 2012). Uma vez que *High Resolution Melt* é sensível a variações de um único nucleotídeo (Wu et al. 2008, Simko 2016), esta mutação pode alterar o padrão de curva de *melt* e consequentemente a correta identificação de genótipos derivados da PI603452. Assim, foi desenhado um novo diagnóstico molecular para a metodologia de HRM, amplificando um fragmento de 52 a 55 pares de bases que inclui a mutação da PI200508 no gene *GmRS2*, mas não inclui a região da mutação na PI603452. Este novo marcador HRM (Figura 2) se comportou conforme esperado para segregação mendeliana. Um total de 168 plantas F_2 foram genotipadas, sendo encontradas 42 plantas mutantes, 78 heterozigotas e 48 selvagens, obtendo um $\chi^2 = 1,29$ (p = 0,53 ns).

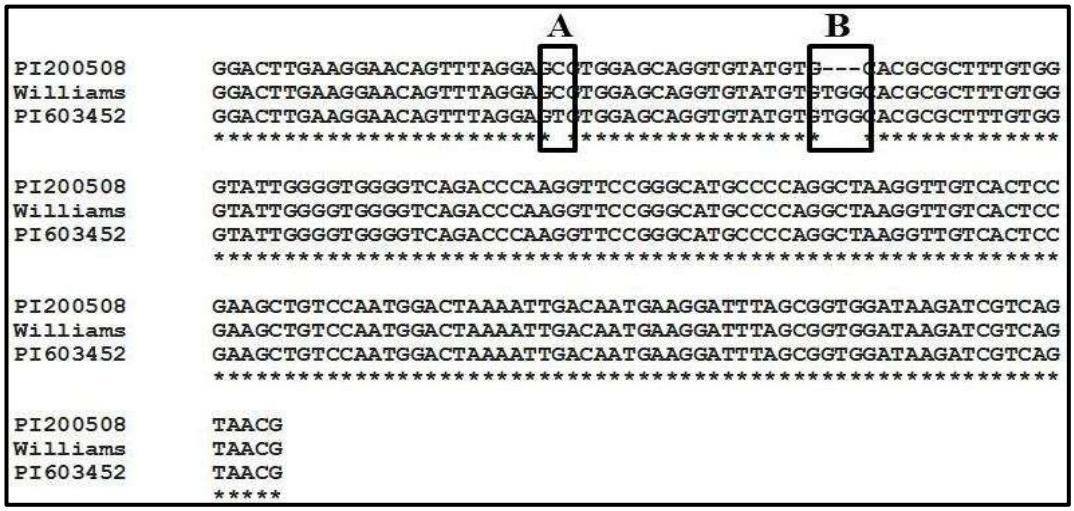

Figura 1 – Alinhamento das sequências geradas pelos *primers* descritos por Skoneczka et al. (2009) com o genoma de referência Williams 82, mostrando a troca (C>G) encontrada no acesso PI603452 (A) e a deleção de três pares de bases do acesso PI200508 (B).

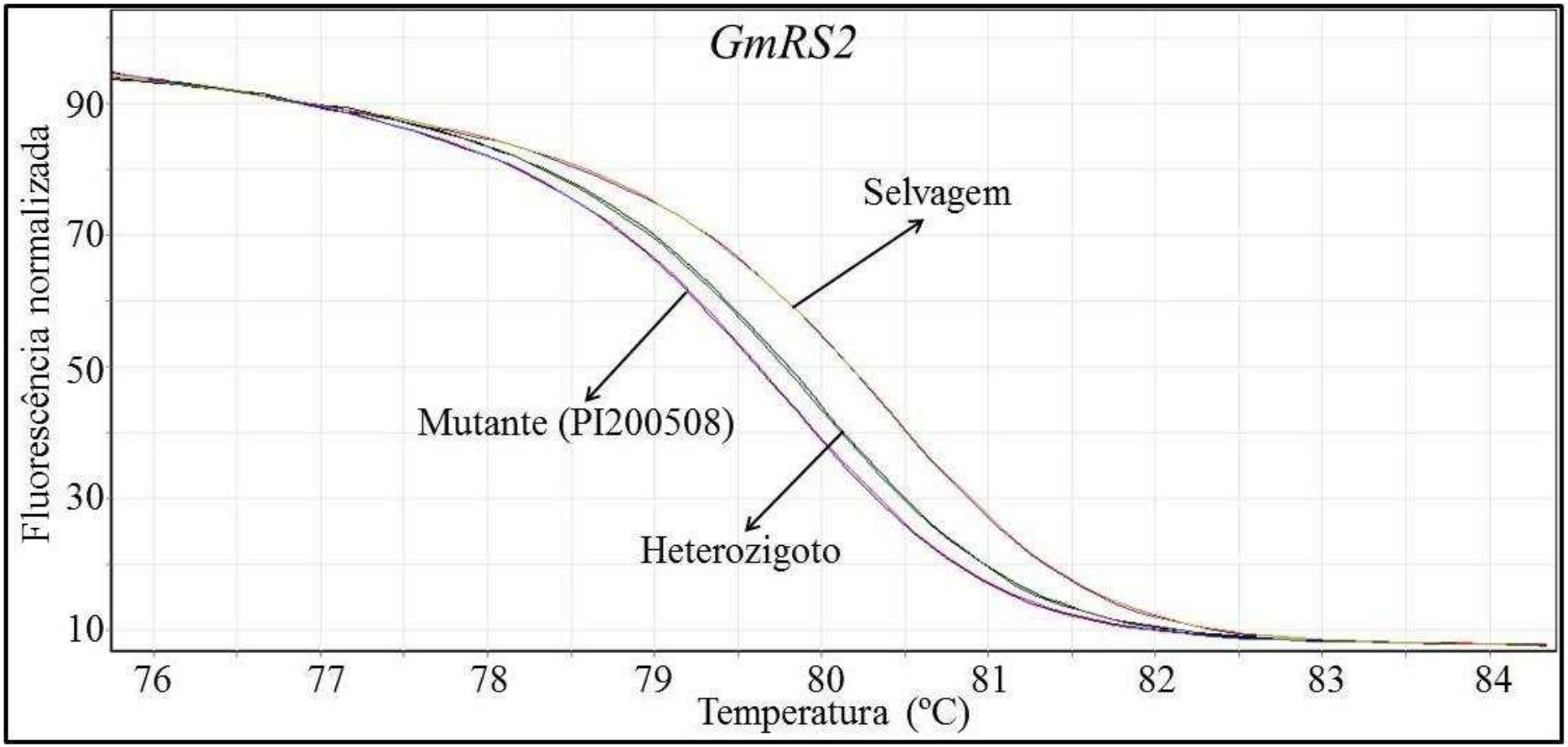

Figura 2 – Marcador desenvolvido para a metodologia de HRM, baseado em uma deleção de três pares de bases encontrada no gene *GmRS2* do acesso de soja PI200508. As curvas de *melt* normalizadas para indivíduos mutante, heterozigoto e selvagem são mostradas.

3.2. Estatística descritiva

Sementes $F_{2:3}$ foram analisadas para várias características de qualidade do grão (teores de proteína, óleo, estaquiose, rafinose, sacarose, oligossacarídeos totais, e ácidos palmítico, esteárico, oleico, linoleico e linolênico), e foi encontrada variabilidade para todas as características analisadas. A Tabela 1 apresenta os valores das características avaliadas para a população F_2, para cada genótipo *GmRS2*, e para os genitores e testemunhas adicionais.

O conteúdo de estaquiose variou entre 0,04 e 0,55% para indivíduos mutantes e 1,11-2,96% para indivíduos do tipo selvagem, com média de 0,18% e 2,08%, respectivamente. Skoneczka et al. (2009) encontraram teores de estaquiose variando entre 0,56-1,47% para indivíduos mutantes, e 2,58-5,63% para indivíduos selvagens, provenientes de uma população F_2 derivada do cruzamento entre o acesso PI200508 com os acessos PI87013 ou PI243545.

Os valores máximos e mínimos da população F_2 excederam os valores dos genitores para várias características, indicando a ocorrência de segregação transgressiva. Este fenômeno tem sido observado em outros estudos envolvendo populações de segregantes de soja para teor de proteína (Hyten et al. 2004), rendimento e peso de grãos (Mansur et al. 1993), teor de sacarose e oligossacarídeos (Kim et al. 2006), e conteúdo de ácidos oleico e linoleico (Pham et al. 2011). É esperado que progênies derivadas de cruzamentos entre muitos genitores mostrem um amplo espectro de variabilidade genética, fornecendo uma maior possibilidade de isolamento de indivíduos segregantes transgressivos (Tyagi & Khan 2010, Mughal et al. 2015).

Tabela 1 – Estatísticas descritivas de características de qualidade do grão de uma população F_2 de soja gerada por cruzamentos entre PI603452, PI283327, PI200508 e NA5909, mostrando valores para cada genótipo identificado para o gene *GmRS2*, genitores e testemunhas adicionais.

		Proteína	**Óleo**	**Sacarose**	**Rafinose**	**Estaquiose**	**Oligossacarídeos**	**Palmítico**	**Esteárico**	**Oleico**	**Linoleico**	**Linolênico**
Genitores e Testemunhas												
	NA5909	37,53	24,42	1,89	0,51	1,38	3,42	11,91	4,65	21,02	55,77	6,65
	PI200508	41,98	16,95	4,84	0,11	0,30	5,25	9,79	2,69	41,91	39,02	6,60
	PI283327	42,85	17,46	3,02	0,27	1,76	5,05	10,11	5,09	28,14	45,91	10,74
	PI603452	41,35	18,29	3,18	0,19	1,69	5,06	10,85	3,61	33,86	43,16	8,53
	UFVTN105AP	48,09	17,62	1,81	0,43	1,61	3,85	11,25	3,81	27,10	52,76	5,08
	CS303TNKCA	38,01	25,63	1,58	0,77	2,18	4,53	10,67	3,79	28,34	54,26	2,94
	Tucunaré	38,42	20,52	3,17	1,00	1,74	5,91	10,85	3,76	21,11	56,41	7,87
População F_2												
Média	População F_2	40,66	19,74	3,84	0,27	1,34	5,45	10,64	3,61	29,66	47,99	8,15
	Mutante	39,21	20,80	5,07	0,11	0,18	5,36	10,69	3,63	30,22	47,63	7,98
	Heterozigoto	41,22	19,28	3,62	0,25	1,51	5,38	10,73	3,61	29,25	48,11	8,30
	Selvagem	41,02	19,55	3,11	0,45	2,08	5,64	10,43	3,57	29,84	48,12	8,06
Mínimo	População F_2	33,00	12,75	1,33	0,02	0,04	2,46	6,67	2,38	17,11	4,01	3,70
	Mutante	33,00	15,45	2,49	0,02	0,04	2,66	6,73	2,38	19,36	4,32	3,94
	Heterozigoto	34,70	12,75	1,52	0,02	0,77	2,46	6,67	2,58	17,24	4,01	4,71
	Selvagem	36,06	14,35	1,33	0,23	1,11	2,79	7,39	2,74	17,11	4,22	3,70
Máximo	População F_2	47,82	25,51	7,07	0,84	3,11	10,16	17,32	5,45	82,17	60,99	10,96
	Mutante	45,84	25,37	7,07	0,22	0,55	7,30	13,88	5,26	82,17	60,99	10,08
	Heterozigoto	47,57	24,33	6,21	0,61	3,11	8,41	17,32	5,45	81,42	60,17	10,96
	Selvagem	47,82	25,51	7,01	0,84	2,96	10,16	14,57	4,76	81,78	57,78	10,34
Amplitude	População F_2	14,82	12,76	5,74	0,82	3,07	7,70	10,65	3,07	65,06	56,98	7,26
	Mutante	12,84	9,92	4,58	0,20	0,51	4,64	7,15	2,88	62,81	56,67	6,14
	Heterozigoto	12,87	11,58	4,69	0,59	2,34	5,95	10,65	2,87	64,18	56,16	6,25
	Selvagem	11,76	11,16	5,68	0,61	1,85	7,37	7,18	2,02	64,67	53,56	6,64
Desvio Padrão	População F_2	3,16	2,73	1,25	0,18	0,82	1,22	1,32	0,54	10,54	9,46	1,09
	Mutante	2,81	2,57	1,05	0,05	0,11	1,07	1,37	0,57	11,63	10,17	1,08
	Heterozigoto	3,16	2,74	0,92	0,13	0,43	1,14	1,34	0,56	9,53	8,94	0,95
	Selvagem	3,10	2,65	1,13	0,15	0,53	1,46	1,24	0,48	11,28	9,83	1,29

3.3. Regressão linear e correlações entre características de qualidade do grão

Análises de regressão linear permitiram observar uma associação significativa entre a mutação no gene *GmRS2* e os teores de sacarose, rafinose, estaquiose, proteína e óleo (Tabela 2). Foram gerados gráficos utilizando a distribuição do conteúdo de sacarose, rafinose e estaquiose, para cada genótipo identificado para a mutação (Figura 3). Esta mutação explicou 69,61% do teor de estaquiose, 51,81% do teor de rafinose e 31,96% do teor de sacarose. Skoneczka et al. (2009) avaliaram a mesma mutação em duas populações F_2 derivadas da população PI200508, e encontraram valores de coeficiente de determinação de 88-94% para teor de estaquiose e 76% para teor de sacarose. Esta diferença pode ser explicada em parte pela diferença no ambiente experimental onde as populações foram avaliadas, bem como a formação da população utilizada. Neste estudo, foi utilizada uma população derivada do cruzamento de quatro genitores divergentes, o que pode levar a um aumento na segregação das características, gerando maior variabilidade genética (Hanson 1959, Alliprandini & Vello 2004).

Tabela 2 – Regressão linear de "características de qualidade de soja x genótipos *GmRS2*", mostrando os valores de teste F, constante da regressão (β_0), coeficiente da regressão (β_1) e coeficiente de determinação (R^2)

Característica	F	β_0	β_1	R^2 (%)
Proteína	7,09**	39,76	0,87**	4,10
Óleo	4,37*	20,36	-0,60*	2,56
Sacarose	77,97**	4,84	-0,97**	31,96
Rafinose	178,5**	0,09	0,17**	51,81
Estaquiose	380,2**	0,37	0,93**	69,61
Oligossacarídeos	1,19	5,31	0,14	0,71
Ac. Palmítico	0,99	10,78	-0,14	0,59
Ac. Esteárico	0,28	3,64	-0,03	0,17
Ac. Oleico	0,02	29,84	-0,17	0,01
Ac. Linoleico	0,06	47,74	0,24	0,03
Ac. Linolênico	0,07	8,12	0,03	0,04

* $p \leq 0,05$; ** $p \leq 0,01$; Oligossacarídeos = sacarose + rafinose + estaquiose.

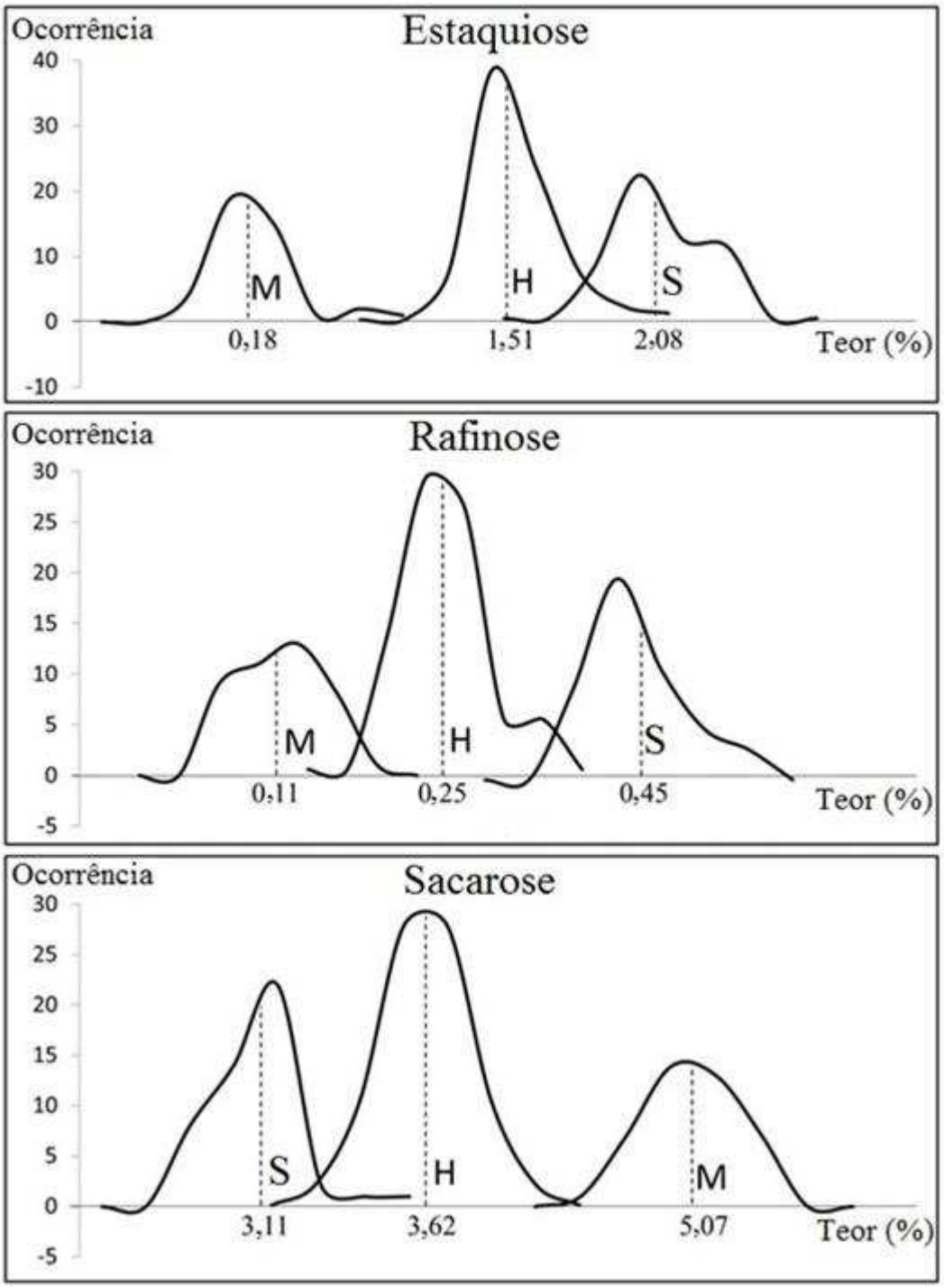

Figura 3 – Gráficos de distribuição de frequência dos teores de estaquiose, rafinose e sacarose de 168 indivíduos de uma população F_2, separados pelos genótipos *GmRS2*. As médias de cada característica para cada genótipo são mostradas. As letras representam os genótipos: mutante (M), heterozigoto (H) e selvagem (S).

Apesar de ter sido encontrada associação significativa entre a mutação do gene *GmRS2* e os teores de proteína e óleo (Tabela 2), os coeficientes de determinação encontrados foram baixos (4,1% para teor de proteína e 2,6% para teor de óleo), indicando que a mutação pode ser utilizada para reduzir os teores de rafinose e estaquiose e aumentar o teor de sacarose, sem mudanças significativas nos teores de óleo e proteína, conforme sugerido por Sato et al. (2014). Além disso, nenhuma associação foi verificada entre a mutação e os teores de oligossacarídeos totais (soma dos teores de sacarose, rafinose e estaquiose), indicando que a variação no teor dos oligossacarídeos foi devido à redução na conversão de sacarose em rafinose (Skoneczka et al. 2009, Yang et al. 2014). Com relação à composição do óleo, nenhuma associação foi encontrada entre a mutação e o conteúdo dos ácidos palmítico, esteárico, oleico, linoleico e linolênico, concordando com Neus et al. (2005).

Correlações entre as características analisadas estão apresentadas na Tabela 3. Foi encontrada correlação negativa entre os teores de proteína e sacarose (-0,40), valor inferior ao que tem sido encontrado em outros trabalhos (Hartwig et al. 1997, Wilcox and Shibles 2001, Jaureguy et al. 2011, Sato et al. 2014). Esta correlação tem se mostrado um desafio para o desenvolvimento de soja com altos teores de ambas as características (Jaureguy et al. 2011, Sato et al. 2014). Segundo Li, Y. et al. (2012), sacarose está diretamente envolvida na biossíntese de proteína, e a ocorrência desta correlação negativa entre as características se explica provavelmente pela competição de ATP e requerimento de cadeia de carbono para cada via (Paul & Foyer 2001). O acúmulo de carboidratos é um importante fator envolvido na produção de proteína durante o período de enchimento de grãos, e por esta razão, os metabolismos de carbono e nitrogênio não são completamente independentes (Li, Y. et al. 2012). Conforme discutido anteriormente, a menor correlação entre sacarose e proteína encontrada em nosso estudo pode ser explicada pela segregação de outros genes com efeito em ambas as características.

É importante mencionar que outras correlações encontradas neste estudo mostraram concordância com aquelas relatadas em outros estudos envolvendo características de qualidade do grão, como correlação negativa entre óleo e proteína (Hartwig et al. 1997, Wilcox and Shibles 2001, Sato et al. 2014), e

Tabela 3 – Coeficientes de Correlação de Pearson entre características de qualidade do grão de soja em uma população F_2 segregante gerada por cruzamentos entre PI283327, PI603452, PI200508 e NA5909.

	Proteína	**Óleo**	**Sacarose**	**Rafinose**	**Estaquiose**	**Oligo**	**Palmítico**	**Esteárico**	**Oleico**	**Linoleico**	**Linolênico**
Proteína	1	-0,8483**	-0,4019**	-0,0442	0,0099	-0,4120**	0,1007	0,0320	0,0391	-0,0940	0,2183**
Óleo		1	0,3125**	0,0529	0,0240	0,3448**	-0,2858**	-0,0804	0,0905	-0,0043	-0,4010**
Sacarose			1	-0,2704**	-0,4147**	0,7091**	0,0073	-0,0023	0,0393	-0,0036	-0,2320**
Rafinose				1	0,6465**	0,3015**	-0,1041	-0,0387	0,0245	-0,0106	-0,0541
Estaquiose					1	0,3397**	-0,1093	-0,0164	-0,0048	0,0205	-0,0399
Oligo						1	-0,0816	-0,0193	0,0412	0,0082	-0,2734**
Palmítico							1	0,3718**	-0,4213**	0,2926**	0,1916*
Esteárico								1	-0,1531*	0,0457	0,0960
Oleico									1	-0,9811**	-0,5806**
Linoleico										1	0,5019**
Linolênico											1

* $p \leq 0.05$; ** $p \leq 0.01$ (teste t); **Oligo** = oligossacarídeos (sacarose + rafinose + estaquiose).

correlação positiva entre óleo e sacarose (Hartwig et al. 1997, Sato et al. 2014). Com relação aos teores de açúcares, correlações significativas foram encontradas entre sacarose e rafinose (-0,27), sacarose e estaquiose (-0,41) e rafinose e estaquiose (0,65). Para estas características, diferentes correlações têm sido reportadas (Hartwig et al. 1997, Skoneczka et al. 2009, Wang et al. 2014, Qiu et al. 2015).

Programas de melhoramento buscando o desenvolvimento de sojas especiais têm focado em características como tamanho adequado de grãos, altos teores de proteína e sacarose, e baixos teores dos oligossacarídeos rafinose e estaquiose, dentre outros (Chen 2004, Jaureguy et al. 2011). Neste estudo, foi desenvolvida uma metodologia de genotipagem para a mutação no gene *GmRS2* (Skoneczka et al. 2009), utilizando *High Resolution Melt*, e a utilizamos para avaliar seu efeito em várias características de qualidade do grão. Os resultados encontrados mostram ser possível aumentar o teor de sacarose e reduzir estaquiose e rafinose, sem grandes alterações em outras características de qualidade do grão, como teor de óleo e proteína, através da interrupção da conversão de sacarose em rafinose. Esta informação é importante para auxiliar melhoristas no desenvolvimento de variedades que melhor se adequem ao mercado de sojas especiais.

4. CONCLUSÕES

- O diagnóstico desenvolvido pela metodologia de HRM foi eficiente para promover a identificação dos genótipos do gene *GmRS2*;
- A mutação encontrada no gene *GmRS2* está associada com os teores de sacarose, rafinose e estaquiose;
- Esta mutação possibilita reduzir os teores de rafinose e estaquiose e aumentar o teor de sacarose sem grandes efeitos em outras características de qualidade do grão, como teores de óleo e proteína.

CONCLUSÕES GERAIS

- Os marcadores HRM desenvolvidos foram eficientes para realizar a discriminação alélica e estão associados com características de qualidade do grão de soja.
- A caracterização de novos alelos relacionados à qualidade do grão de soja é de grande importância para o desenvolvimento de cultivares com características especiais, visando melhor atender à indústria alimentícia e de biocombustíveis.

REFERÊNCIAS BIBLIOGRÁFICAS

ABIOVE – ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE ÓLEOS VEGETAIS. **Estatística Mensal do Complexo Soja**. Fevereiro, 2017a. Disponível em: <http://www.abiove.org.br/site/index.php?page=estatistica&area=NC0yLTE=> Acessado em 28/02/2017.

ABIOVE - ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE ÓLEOS VEGETAIS. **Biodiesel: entrega e produção**. Fevereiro, 2017b. Disponível em: <http://www.abiove.org.br/site/index.php?page=estatistica&area=NC0yLTE=> Acessado em 28/02/2017.

AGROLINK – Soja: Formas de consumo. Disponível em: <https://www.agrolink.com.br/culturas/soja/comercializacao.aspx> Acessado em 05/02/2017.

ALLIPRANDINI, L. F.; VELLO, N. A. Heritability and correlations among traits in four-way soybean crosses. **Euphytica**, v. 136, n. 1, p. 81-91, 2004.

ALT, J. L.; FEHR, W. R.; WELKE, G. A.; SANDHU, D. Phenotypic and molecular analysis of oleate content in the mutant soybean line M23. **Crop Science**, v. 45, n. 5, p. 1997-2000, 2005.

ALVES, M. P.; NOVAIS, C. M. PCR em tempo real: Uma inovação tecnológica da reação em cadeia da polimerase (PCR). **Biotecnologia Ciência e Desenvolvimento, Brasília**, ed. 33, p. 10-13, 2004.

ANAI, T. et al. Identification of corresponding genes for three low-α-linolenic acid mutants and elucidation of their contribution to fatty acid biosynthesis in soybean seed. **Plant Science**, v. 168, n. 6, p. 1615-1623, 2005.

ANDERSEN, C. L.; JENSEN, J. L.; ØRNTOFT, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. **Cancer Research**, v. 64, n. 15, p. 5245-5250, 2004.

ANP – AGÊNCIA NACIONAL DE PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS. **Anuário Estatístico Brasileiro do Petróleo, Gás Natural e Biocombustíveis**. Rio de Janeiro-RJ, 2016.

APROSOJA – ASSOCIAÇÃO DOS PRODUTORES DE SOJA DO MATO GROSSO. Disponível em: <http://www.aprosoja.com.br/soja-e-milho/a-historia-da-soja> Acessado em 07/02/2017.

ARI, M. M. Nutrient Composition, Amino Acids Profile and Anti Nutritional Factors of Nixtamalized Soya Bean (*Glycine max*) using different alkali treatment. **Production Agriculture and Technology**, v. 10, n. 1, p. 28-37, 2014.

ARTIMO, P. et al. ExPASy: SIB bioinformatics resource portal. **Nucleic acids research**, v. 40, n. 1, p. 597-603, 2012.

BAUD, S.; LEPINIEC, L. Physiological and developmental regulation of seed oil production. **Progress in Lipid Research**, v. 49, n. 3, p. 235-249, 2010.

BILYEU, K. D. et al. Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels. **Crop Science**, v. 43, n. 5, p. 1833-1838, 2003.

BILYEU K, D. et al. Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. **Crop Science:** v. 45, n. 5, p. 1830-1836, 2005.

BILYEU, K. D.; et al. Molecular genetic resources for development of 1% linolenic acid soybeans. **Crop Science**, v. 46, n. 5, p. 1913-1918, 2006.

BILYEU, K. D. et al. Novel mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. **Crop Science**, v. 51, n. 1, p. 259-264, 2011.

BILYEU, K. D.; WIEBOLD, W. J. Environmental stability of seed carbohydrate profiles in soybeans containing different alleles of the raffinose synthase 2 (RS2) gene. **Journal of Agricultural and Food Chemistry**, v. 64, n. 5, p. 1071-1078, 2016.

BRACE, R. C.; FEHR, W. R.; SCHNEBLY, S. R. Agronomic and seed traits of soybean lines with high oleate concentration. **Crop Science**, v. 51, p. 534–541, 2011.

BRASIL. Lei nº 13.263, de 23 de março de 2016.

BRUNE, M. F. S. S. et al. Avaliação bioquímico-nutricional de uma linhagem de soja livre do inibidor de tripsina Kunitz e de lectinas. **Ciência e Tecnologia de Alimentos**, v. 30, n. 3, p. 657-663, 2010.

BUBECK, D. M.; FEHR, W. R.; HAMMOND, E. G. Inheritance of palmitic and stearic acid mutants of soybean. **Crop Science**, v. 29, n. 3, p. 652-656, 1989.

BURKEY, K. O. et al. Elevated carbon dioxide and ozone effects on peanut: II. Seed yield and quality. **Crop Science**, v. 47, n. 4, p. 1488-1497, 2007.

BURTON, J. W.; WILSON, R. F.; BRIM, C. A. Recurrent selection in soybeans. IV. Selection for increased oleic acid percentage in seed oil. **Crop Science**, v. 23, n. 4, p. 744-747, 1983.

BYRUM, J. R. et al. Alteration of the omega-3 fatty acid desaturase gene is associated with reduced linolenic acid in the A5 soybean genotype. **Theoretical and Applied Genetics**, v. 94, n. 3-4, p. 356-359, 1997.

CAIXETA, E. T. et al. Tipos de marcadores moleculares. In: BORÉM, A.; CAIXETA, E. T. **Marcadores Moleculares**. 2ª edição, Universidade Federal de Viçosa. 532p. Viçosa-MG, 2009.

CARDINAL, A. J. et al. Molecular Analysis of GmFAD3A in Two Soybean Populations Segregating for the fan, fap1, and fap(nc) and Loci. **Crop Science**, v. 51, n. 5, p. 2104-2112, 2011.

CARRÃO-PANIZZI, M. C. et al. Influence of Growth Locations on Isoflavone Contents in Brazilian Soybean Cultivars. **Breeding Science**, v. 48, p. 409-413, 1998.

CARRÃO-PANIZZI, M. C.; SILVA, J. C. Soja na alimentação humana: qualidade na produção de grãos com valor agregado. In: **Quinto Congresso de la Soja del Mercosur**. 2011.

CHAPPELL, A. S.; BILYEU, K. D. A. *GmFAD3A* mutation in the low linolenic acid soybean mutant C1640. **Plant Breeding**, v. 125, n. 5, p. 535-536, 2006.

CHEN, P. Developing high quality identity-preserved soybean for the specialty soyfood market. **Production and marketing of identity-preserved Soybean. American Soybean Association**, p. 23-31, 2004.

COLLINS, F. S.; BROOKS, L. D.; CHAKRAVARTI, A. A DNA polymorphism discovery resource for research on human genetic variation. **Genome research**, v. 8, n. 12, p. 1229-1231, 1998.

CONAB – COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da Safra Brasileira de Grãos**. Quarto levantamento, v. 4, n.4, safra 2016/2017, 2017.

CRUZ, C. D. Genes: a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, v. 35, n. 3, p. 271-276, 2013.

CRUZ, C. D.; SCHUSTER, I. GQMOL: aplicativo computacional para análise de dados moleculares e de suas associações com caracteres quantitativos. 2004.

CRUZ, M. F. A. et al. Identificação de SNPs para conteúdo de ácidos graxos em soja pela técnica HRM. **Pesquisa Agropecuária Brasileira**, v. 48, n. 12, p. 1596-1600, 2013.

DANTAS, M. B. et al. Evaluation of the oxidative stability of corn biodiesel. **Fuel**, v. 90, n. 2, p. 773-778, 2011.

DIERKING, E. C.; BILYEU, K. D. Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. **The Plant Genome**, v. 1, n. 2, p. 135-145, 2008.

DOYLE, J. J.; DOYLE, J. L. Isolation of plant DNA from fresh tissue. **Focus**, v. 12, p. 13-15, 1990.

DUTTON, H. J. et al. The flavor problem of soybean oil. VIII. Linolenic acid. **Journal of the American Oil Chemists' Society**, v. 28, n. 3, p. 115-118, 1951.

DWIGHT, Z.; PALAIS, R.; WITTWER, C. T. uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. **Bioinformatics**, v. 27, n. 7, p. 1019-1020, 2011.

EMBRAPA – EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Disponível em: <https://www.embrapa.br/web/portal/soja/cultivos/soja1/historia> Acessado em 07/02/2017.

EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Disponível em: <http://www.cnpso.embrapa.br/soja_alimentacao/index.php?pagina=23> Acessado em 07/02/2017.

EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Disponível em: <https://www.embrapa.br/web/portal/soja/cultivos/soja1/soja-na-alimentacao> Acessado em: 08/02/2007.

FAN, J. B.; CHEE, M. S.; GUNDERSON, K. L. Highly parallel genomic assays. **Nature Reviews Genetics**, v. 7, n. 8, p. 632-644, 2006.

FDA – US FOOD AND DRUG ADMINISTRATION. **Final Determination Regarding Partially Hydrogenated Oils (Removing Trans Fat)**, 2015. Disponível em: <http://www.fda.gov/Food/IngredientsPackagingLabeling/FoodAdditivesIngredients/ucm449162.htm> Acessado em 06/02/2017.

FDA – US FOOD AND DRUG ADMINISTRATION. **The FDA takes step to remove artificial trans fats in processed foods**, 2016. Disponível em: <http://www.fda.gov/Food/NewsEvents/ConstituentUpdates/ucm449145.htm> Acessado em 06/02/2017.

FEHR, W. R. Breeding for modified fatty acid composition in soybean. **Crop Science**, v. 47, n. 3, p. 72-87, 2007.

FEHR, W. R.; CAVINESS, C. E. Stages of soybean development. Special Report: **Agriculture and Home Economics Experiment Station**. Iowa State University, 1977.

FEHR, W. R.; HAMMOND, E. G. **Reduced linolenic acid production in soybeans**. U.S. Patent n. 6,133,509, 17 out. 2000.

FLORES, T. et al. Silencing of GmFAD3 gene by siRNA leads to low α-linolenic acids (18: 3) of fad3-mutant phenotype in soybean [Glycine max (Merr.)]. **Transgenic Research**, v. 17, n. 5, p. 839-850, 2008.

GANGL, R.; BEHMÜLLER, R.; TENHAKEN, R. Molecular cloning of AtRS4, a seed specific multifunctional RFO synthase/galactosylhydrolase in *Arabidopsis thaliana*. **Frontiers in Plant Science**, v. 6, p. 789, 2015.

GEATER, C. W. Genetic and environmental effects on natto quality characteristics of soybean. Tese (Doutorado em Melhoramento de Plantas) – Iowa State University. 117p. Ames, Iowa, EUA, 2000.

GESTEIRA, A. S. et al. Biometrical analyses of linolenic acid content of soybean seeds. **Genetics and Bolecular Biology**, v. 26, n. 1, p. 65-68, 2003.

GIETZ, R. D.; SCHIESTL, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. **Nature Protocols**, v. 2, n. 1, p. 31-34, 2007.

GOODSTEIN, D. M. et al. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, n. D1, p. D1178-D1186, 2012.

GOETTEL, W. et al. Identification and characterization of transcript polymorphisms in soybean lines varying in oil composition and content. **BMC Genomics**, v. 15, n. 1, p. 299, 2014.

GU, C. et al. Effect of soybean variety on anti-nutritional factors content, and growth performance and nutrients metabolism in rat. **International Journal of Molecular Sciences**, v. 11, n. 3, p. 1048-1056, 2010.

GUIMARÃES, C. T. et al. Marcadores moleculares e suas aplicações no melhoramento genético. **Informe Agropecuário,** Belo Horizonte, v. 30, n. 253, p. 86-95, 2009.

GUIMARÃES, V. M. et al. Characterization of α-galactosidases from germinating soybean seed and their use for hydrolysis of oligosaccharides. **Phytochemistry**, v. 58, n. 1, p. 67-73, 2001.

GUPTA, P. K.; ROY, J. K.; PRASAD, M. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. **Current Science**, v. 80, n. 4, p. 524-535, 2001.

HA, B. K. et al. Identification of quantitative trait loci controlling linolenic acid concentration in PI483463 (*Glycine soja*). **Theoretical and Applied Genetics**, v. 127, n. 7, p. 1501-1512, 2014.

HAGELY, K. B.; PALMQUIST, D.; BILYEU, K. D. Classification of distinct seed carbohydrate profiles in soybean. **Journal of Agricultural and Food Chemistry**, v. 61, n. 5, p. 1105-1111, 2013.

HAMMOND, E.G.; FEHR, W. R. Registration of A5 germplasm line of soybean (Reg. No. GP44). Crop Science, v. 23, p.192-193, 1983.

HANSON, W. D. The breakup of initial linkage blocks under selected mating systems. **Genetics**, v. 44, n. 5, p. 857, 1959.

HARTWIG, E. E.; KUO, T. M.; KENTY, M. M. Seed protein and its relationship to soluble sugars in soybean. **Crop Science**, v. 37, n. 3, p. 770-773, 1997.

HEBSGAARD, S. M. et al. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. **Nucleic Acids Research**, v. 24, n. 17, p. 3439-3452, 1996.

HEPPARD, E. P. et al. Developmental and growth temperature regulation of two different microsomal [omega]-6 desaturase genes in soybeans. **Plant Physiology**, v. 110, n. 1, p. 311-319, 1996.

HU, F. B. et al. Dietary fat intake and the risk of coronary heart disease in women. **New England Journal of Medicine**, v. 337, n. 21, p. 1491-1499, 1997.

HU, R. et al. Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. **BMC Molecular Biology**, v. 10, n. 1, p. 93, 2009.

HYMOWITZ, T. et al. Relationship between the content of oil, protein, and sugar in soybean seed. **Agronomy Journal**, v. 64, n. 5, p. 613-616, 1972.

HYTEN, D. L. et al. Seed quality QTL in a prominent soybean population. **Theoretical and Applied Genetics**, v. 109, n. 3, p. 552-561, 2004.

HYTEN, D. L. et al. Impacts of genetic bottlenecks on soybean genome diversity. **Proceedings of the National Academy of Sciences**, v. 103, n. 45, p. 16666-16671, 2006.

ISKANDAROV, U.; KHOZIN-GOLDBERG, I.; COHEN, Z. Identification and characterization of Δ12, Δ6, and Δ5 desaturases from the green microalga *Parietochloris incisa*. **Lipids**, v. 45, n. 6, p. 519-530, 2010.

JAUREGUY, L. M.; CHEN, P.; SCABOO, A. M. Heritability and correlations among food-grade traits in soybean. **Plant Breeding**, v. 130, n. 6, p. 647-652, 2011.

KANCHANA, P.; SANTHA, M. L.; RAJA, K. D. A review on *Glycine max* (L.) Merr.(soybean). **World Journal of Pharmacy and Pharmaceutical Sciences**, v. 5, n. 1, p. 356-371, 2015.

KIM, H. K.; KANG, S. T.; OH, K. W. Mapping of putative quantitative trait loci controlling the total oligosaccharide and sucrose content of *Glycine max* seeds. **Journal of Plant Research**, v. 119, n. 5, p. 533-538, 2006.

KIM, M. et al. A new low linolenic acid allele of GmFAD3A gene in soybean PE1690. **Molecular Breeding**, v. 35, n. 8, p. 155, 2015.

KIM, W. J.; SMIT, C. J. B.; NAKAYAMA, T. O. M. Removal of oligosaccharides from soybeans. **LWT Lebensmitt Wissensch Technol**, 1973.

KU, S. et al. Extraction of oligosaccharides during cooking of whole soybeans. **Journal of Food Science**, v. 41, n. 2, p. 361-364, 1976.

KUMAR, V. et al. Evaluation of vegetable-type soybean for sucrose, taste-related amino acids, and isoflavones contents. **International Journal of Food Properties**, v. 14, n. 5, p. 1142-1151, 2011.

LE, D. T. et al. Evaluation of candidate reference genes for normalization of quantitative RT-PCR in soybean tissues under various abiotic stress conditions. **PLoS One**, v. 7, n. 9, p. e46487, 2012.

LEFFEL, R. C. Registration of BARC-12, a low linolenic acid soybean germplasm line. **Crop Science**, v. 34, n. 5, p. 1426-1427, 1994.

LESKE, K. L. et al. Effect of ethanol extract on nitrogen-corrected true metabolizable energy for soybean meal with broilers and roosters. **Poultry Science**, v. 70, n. 4, p. 892-895, 1991.

LI, L. et al. Molecular cloning and characterization of a novel microsomal oleate desaturase gene from soybean. **Journal of Plant Physiology**, v. 164, n. 11, p. 1516-1526, 2007.

LI, Q. et al. Validation of reference genes for real-time quantitative PCR normalization in soybean developmental and germinating seeds. **Plant Cell Reports**, v. 31, n. 10, p. 1789-1798, 2012.

LI, Y. S. et al. Greater differences exist in seed protein, oil, total soluble sugar and sucrose content of vegetable soybean genotypes ['*Glycine max*'(L.) Merrill] in Northeast China. **Australian Journal of Crop Science**, v. 6, n. 12, p. 1681, 2012.

LIBAULT, M. et al. Identification of four soybean reference genes for gene expression normalization. **The Plant Genome**, v. 1, n. 1, p. 44-54, 2008.

LIENER, I. E. Implications of antinutritional components in soybean foods. **Critical Reviews in Food Science & Nutrition**, v. 34, n. 1, p. 31-67, 1994.

LIEW, M. et al. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. **Clinical Chemistry**, v. 50, n. 7, p. 1156-1164, 2004.

LÔBO, I. P.; FERREIRA, S. L. C.; CRUZ, R. S. D. Biodiesel: parâmetros de qualidade e métodos analíticos. **Química Nova**, v. 32, n. 6, p. 1596-1608, 2009.

LUCAS, M. et al. Dietary intake of n−3 and n−6 fatty acids and the risk of clinical depression in women: a 10-y prospective follow-up study. **The American Journal of Clinical Nutrition**, v. 93, n. 6, p. 1337-1343, 2011.

MALAFAIA, F. G.; SOUZA, L. G. **Estudo comparativo de especificações do biodiesel proveniente da soja e da mamona**. Monografia (Graduação em Engenharia de Energia) – Universidade de Brasília. 53p. Brasília-DF, 2014.

MANSUR, L. M. et al. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). **Theoretical and Applied Genetics**, v. 86, n. 8, p. 907-913, 1993.

MATTA, L. B. Mapeamento de QTL para conteúdo de ácido oleico em soja utilizando marcadores SNP. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa. 69p. Viçosa-MG, 2012.

MCCARTNEY, A. W. et al. Membrane-bound fatty acid desaturases are inserted co-translationally into the ER and contain different ER retrieval motifs at their carboxy termini. **The Plant Journal**, v. 37, n. 2, p. 156-173, 2004.

MEESAPYODSUK, D. et al. Primary structure, regioselectivity, and evolution of the membrane-bound fatty acid desaturases of *Claviceps purpurea*. **Journal of Biological Chemistry**, v. 282, n. 28, p. 20191-20199, 2007.

MIRANDA, V. J. et al. Validation of reference genes aiming accurate normalization of qPCR data in soybean upon nematode parasitism and insect attack. **BMC Research Notes**, v. 6, n. 1, p. 196, 2013.

MITAL, B. K.; STEINKRAUS, K. H. Utilization of oligosaccharides by lactic acid bacteria during fermentation of soy milk. **Journal of Food Science**, v. 40, n. 1, p. 114-118, 1975.

MOZAFFARIAN, D. et al. Trans fatty acids and cardiovascular disease. **New England Journal of Medicine**, v. 354, n. 15, p. 1601-1613, 2006.

MUGHAL, A. H. et al. Oil content variability and genetic divergence in half-sib families of *Prunus armeniaca* L. in Kashmir Valley India. **African Journal of Agricultural Research**, v. 10, n. 14, p. 1693-1701, 2015.

NAKAMURA, M. T.; NARA, T. Y. Structure, function, and dietary regulation of Δ6, Δ5, and Δ9 desaturases. **Annual Review Nutrition**, v. 24, p. 345-376, 2004.

NEUS, J. D.; FEHR, W. R.; SCHNEBLY, S. R. Agronomic and seed characteristics of soybean with reduced raffinose and stachyose. **Crop Science**, v. 45, n. 2, p. 589-592, 2005.

O'KEEFE, S.; BIANCHI, L.; SHARMAN, J. Soybean Nutrition. **SM Journal of Nutrition and Metabolism**, v. 1, n. 2, p. 1006, 2015.

OMOSAIYE, O.; CHERYAN, M.; MATTHEWS, M. E. Removal of oligosaccharides from soybean water extracts by ultrafiltration. **Journal of Food Science**, v. 43, n. 2, p. 354-360, 1978.

PARENTE, E. J. S.; **Biodiesel: uma aventura tecnológica num país engraçado**. Unigráfica. 60p. Fortaleza-CE, 2003.

PARSONS, C. M.; ZHANG, Y.; ARABA, M. Nutritional evaluation of soybean meals varying in oligosaccharide content. **Poultry Science**, v. 79, n. 8, p. 1127-1131, 2000.

PAUL, M. J.; FOYER, C. H. Sink regulation of photosynthesis. **Journal of Experimental Botany**, v. 52, n. 360, p. 1383-1400, 2001.

PFAFFL, M. W. et al. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. **Biotechnology Letters**, v. 26, n. 6, p. 509-515, 2004.

PHAM, A. T. et al. Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. **BMC Plant Biology**, v. 10, n. 1, p. 195, 2010.

PHAM, A. T. et al. A novel FAD2-1 A allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. **Theoretical and Applied Genetics**, v. 123, n. 5, p. 793-802, 2011.

PHAM, A. T.; SHANNON, J. G.; BILYEU, K. D. Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. **Theoretical and Applied Genetics**, v. 125, n. 3, p. 503-515, 2012.

PHAM, A. T. et al. Characterization of the fan1 locus in soybean line A5 and development of molecular assays for high-throughput genotyping of FAD3 genes. **Molecular breeding**, v. 33, n. 4, p. 895-907, 2014.

PINTO, M. O. et al. Associação de marcadores moleculares SNP com o conteúdo de ácido linolênico em sementes de soja. **Pesquisa Agropecuária Brasileira**, v. 48, n. 3, p. 263-269, 2013.

QIU, D. et al. Identification and characterization of a stachyose synthase gene controlling reduced stachyose content in soybean. **Theoretical and Applied Genetics**, v. 128, n. 11, p. 2167-2176, 2015.

RADOVANOVIC, N. et al. Functional characterization of flax fatty acid desaturase FAD2 and FAD3 isoforms expressed in yeast reveals a broad diversity in activity. **Molecular Biotechnology**, v. 56, n. 7, p. 609-620, 2014.

RAHMAN, S. M.; TAKAGI, Y.; KINOSHITA, T. Genetic control of high oleic acid content in the seed oil of two soybean mutants. **Crop science**, v. 36, n. 5, p. 1125-1128, 1996.

RAHMAN, S. M. et al. High stearic acid soybean mutant induced by x-ray irradiation. **Bioscience, Biotechnology, and Biochemistry**, v. 59, p. 922–923, 2001.

REDDY, K. R. et al. Ultraviolet-B Radiation Alters Soybean Growth and Seed Quality. **Food and Nutrition Sciences**, v. 7, n. 01, p. 55, 2016.

REINPRECHT, Y. et al. Molecular basis of the low linolenic acid trait in soybean EMS mutant line RG10. **Plant Breeding**, v. 128, n. 3, p. 253-258, 2009.

REINPRECHT, Y.; PAULS, K. P. Microsomal omega-3 fatty acid desaturase genes in low linolenic acid soybean line RG10 and validation of major linolenic acid QTL. **Frontiers in Genetics**, v. 7, 2016.

ROSS, A. J. et al. Agronomic and seed traits of 1%-linolenate soybean genotypes. **Crop Science**, v. 40, n. 2, p. 383-386, 2000.

SÁ, M. E. L.; CIABOTTI, S.; BARCELOS, M. F. P. Formas de consumo da soja. **Informe Agropecuário**, v. 27, n. 230, jan/fev, 2006.

SANDHU, D. et al. Enhanced oleic acid content in the soybean mutant M23 is associated with the deletion in the Fad2-1a gene encoding a fatty acid

desaturase. **Journal of the American Oil Chemists' Society**, v. 84, n. 3, p. 229-235, 2007.

SATO, T. et al. Near-infrared reflectance calibrations for determining sucrose content in soybean breeding using artificial reference samples. **Plant Breeding**, v. 131, n. 4, p. 531-534, 2012.

SATO, T. et al. Effects of divergent selection for seed protein content in high-protein vs. food-grade populations of early maturity soybean. **Plant Breeding**, v. 133, n. 1, p. 74-79, 2014.

SCHLUETER, J. A. et al. The FAD2 gene family of soybean. **Crop Science**, v. 47, n. Supplement_1, p. 14-26, 2007.

SHANKLIN, J.; WHITTLE, E.; FOX, B. G. Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. **Biochemistry**, v. 33, n. 43, p. 12787-12794, 1994.

SHI, J. et al. A rice microsomal delta-12 fatty acid desaturase can enhance resistance to cold stress in yeast and *Oryza sativa*. **Molecular breeding**, v. 29, n. 3, p. 743-757, 2012.

SHI, M. M. et al. High throughput genotyping for the detection of a single nucleotide polymorphism in NAD (P) H quinone oxidoreductase (DT diaphorase) using TaqMan probes. **Molecular Pathology**, v. 52, n. 5, p. 295, 1999.

SIMKO, I. High-resolution DNA melting analysis in plant research. **Trends in Plant Science**, v. 21, n. 6, p. 528-537, 2016.

SINGER, S. D.; WESELAKE, R. J.; RAHMAN, H.. Development and characterization of low α-linolenic acid *Brassica oleracea* lines bearing a novel mutation in a 'class a'FATTY ACID DESATURASE 3 gene. **BMC genetics**, v. 15, n. 1, p. 94, 2014.

SKONECZKA, J. A. et al. Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI200508. **Crop Science**, v. 49, n. 1, p. 247-255, 2009.

SMITH, B. L.; LU, C. P.; BREMER, A. J. R. High-resolution melting analysis (HRMA): a highly sensitive inexpensive genotyping alternative for population studies. **Molecular Ecology Resources**, v. 10, n. 1, p. 193-196, 2009.

SONG, Q. et al. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. **PLoS One**, v. 8, n. 1, p. e54985, 2013.

STECH, M. R.; CARNEIRO, D. J.; CARVALHO, M. R. B. Fatores antinutricionais e coeficientes de digestibilidade aparente da proteína de produtos de soja para o pacu (*Piaractus mesopotamicus*). **Acta Scientiarum. Animal Sciences**, v. 32, n. 3, p. 255-262, 2010.

STIJŠIN, D. et al. Inheritance of low linolenic acid level in the soybean line RG10. **Crop Science**, v. 38, n. 6, p. 1441-1444, 1998.

SUAREZ, F. L. et al. Gas production in humans ingesting a soybean flour derived from beans naturally low in oligosaccharides. **The American Journal of Clinical Nutrition**, v. 69, n. 1, p. 135-139, 1999.

SUN, R. et al. Identification of a Δ12 fatty acid desaturase from oil palm (*Elaeis guineensis* Jacq.) involved in the biosynthesis of linoleic acid by heterologous expression in Saccharomyces cerevisiae. **Gene**, v. 591, n. 1, p. 21-26, 2016.

TANG, G. et al. Oleate desaturase enzymes of soybean: evidence of regulation through differential stability and phosphorylation. **The Plant Journal**, v. 44, n. 3, p. 433-446, 2005.

TANG, H. et al. The effect of natural and synthetic antioxidants on the oxidative stability of biodiesel. **Journal of the American Oil Chemists' Society**, v. 85, n. 4, p. 373-382, 2008.

TEIXEIRA, A. I. et al. Development of a method to quantify sucrose in soybean grains. **Food Chemistry**, v. 130, n. 4, p. 1134-1136, 2012.

TELEKEN, J. G. et al.Período de indução de blendas de biodiesel de diferentes fontes graxas. **Revista Gestão & Sustentabilidade Ambiental**, v. 4, p. 532-541, 2015.

THOMPSON, J. D.; GIBSON, T. J.; HIGGINS, D. G. Multiple sequence alignment using ClustalW and ClustalX. **Current Protocols in Bioinformatics**, p. 2.3. 1-2.3. 22, 2002.

TYAGI, S. D.; KHAN, M. H. Genetic divergence in lentil. **African Crop Science Journal**, v. 18, n. 2, 2010.

UNTERGASSER, A. et al. Primer3—new capabilities and interfaces. **Nucleic Acids Research**, v. 40, n. 15, e115, 2012.

USDAa – US DEPARTMENT OF AGRICULTURE. Table 47. **World vegetable oils supply and distribution**. Disponível em: <https://www.ers.usda.gov/data-products/oil-crops-yearbook.aspx> Acessado em 07/02/2017.

USDAb – US DEPARTMENT OF AGRICULTURE. Table 45. **Supply and use: Soybeans, soybean meal, and soybean oil**. Disponível em: <https://www.ers.usda.gov/data-products/oil-crops-yearbook.aspx> Acessado em 07/02/2017.

VAN DE MORTEL, M. et al. Distinct biphasic mRNA changes in response to Asian soybean rust infection. **Molecular Plant-Microbe Interactions**, v. 20, n. 8, p. 887-899, 2007.

VANDESOMPELE, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. **Genome Biology**, v. 3, n. 7, p. research0034.1-12, 2002.

VIANA, P.A**. Caracterização cinético-bioquímica e aplicações biotecnológicas de alfa-galactosidades de *Debaryomices hansenii* UFV-1**. Dissertação (Mestrado em Bioquímica Agrícola) – Universidade Federal de Viçosa. 174p. Viçosa-MG, 2005.

VIEIRA, F. S.; PASQUINI, C. Determination of the oxidative stability of biodiesel using near infrared emission spectroscopy. **Fuel**, v. 117, p. 1004-1009, 2014.

WANG, Y.; CHEN, P.; ZHANG, B. Quantitative trait loci analysis of soluble sugar contents in soybean. **Plant Breeding**, v. 133, n. 4, p. 493-498, 2014.

WARNER, K.; FEHR, W. Mid-oleic/ultra low linolenic acid soybean oil: a healthful new alternative to hydrogenated oil for frying. **Journal of the American Oil Chemists' Society**, v. 85, n. 10, p. 945, 2008.

WILCOX, J. R.; CAVINS, J. F.; NIELSEN, N. C. Genetic alteration of soybean oil composition by a chemical mutagen. **Journal of the American Oil Chemists' Society**, v. 61, n. 1, p. 97-100, 1984.

WILCOX, J. R.; SHIBLES, R. M. Interrelationships among seed quality attributes in soybean. **Crop Science**, v. 41, n. 1, p. 11-14, 2001.

WILSON, R. F.; BURTON, J. W.; BRIM, C. A. Progress in the selection for altered fatty acid composition in soybeans. **Crop Science**, v. 21, n. 5, p. 788-791, 1981.

WU, Shu-Biao et al. High resolution melting analysis of almond SNPs derived from ESTs. **Theoretical and Applied Genetics**, v. 118, n. 1, p. 1-14, 2008.

YADAV, N. S. Soybean genetics, molecular biology and biotechnology. Biotechnology in Agriculture Series. CAB International, 1996.

YANG, K. et al. Development of molecular markers for low raffinose and stachyose in korean soybean cultivars. **Plant Breeding and Biotechnology**, v. 2, n. 2, p. 151-157, 2014.

YU, X. et al. Profiling and relationship of water-soluble sugar and protein compositions in soybean seeds. **Food chemistry**, v. 196, p. 776-782, 2016.

ZALOGA, G. P. et al. Trans fatty acids and coronary heart disease. **Nutrition in Clinical Practice**, v. 21, n. 5, p. 505-512, 2006.

ZENG, A. et al. Identification and confirmation of quantitative trait loci for stachyose content in soybean seed. **Plant Breeding**, v. 134, n. 2, p. 178-185, 2015.

ZHANG, H. T. et al. Heterologous expression of two *Glycine max* ω-3 fatty acid desaturases in *Saccharomyces cerevisiae*. **Russian Journal of Plant Physiology**, v. 56, n. 4, p. 569-574, 2009.

ZHANG, J. et al. Arabidopsis fatty acid desaturase FAD2 is required for salt tolerance during seed germination and early seedling growth. **PLoS One**, v. 7, n. 1, p. e30355, 2012.

ZHU, Y. L. et al. Single-nucleotide polymorphisms in soybean. **Genetics**, v. 163, n. 3, p. 1123-1134, 2003.

APÊNDICE I

MELHORAMENTO GENÉTICO PARA O DESENVOLVIMENTO DE SOJAS ESPECIAIS ATRAVÉS DE RETOCRUZAMENTOS ASSISTIDOS

1. INTRODUÇÃO

A soja (*Glycine max* (L.) Merril) é uma leguminosa de grande importância como fonte de óleo e proteína vegetal (Wang et al. 2014). Os diferentes usos da soja podem ser agrupados em grãos, farelo, óleo, farinha de soja desengordurada e proteína isolada, possuindo aplicações tanto na alimentação humana e animal quanto para a indústria. A demanda social por alimentos mais saudáveis tem auxiliado na descoberta do valor nutricional e os benefícios da soja para a saúde (Jaureguy et al. 2011), de modo que a busca por sojas de maior qualidade nutricional tem aumentado no mercado mundial (Kathmale et al. 2013). Portanto, o desenvolvimento de cultivares de soja com características especiais é imperativo para os programas de melhoramento, e um nicho de mercado promissor para produtores (Jaureguy et al. 2011).

Um dos principais componentes do grão de soja, o óleo de soja geralmente apresenta grandes quantidades de ácidos graxos poli-insaturados (ácidos linoleico e linolênico), que apresentam baixa estabilidade oxidativa, levando a produção de sabor e odor desagradáveis, reduzindo sua aceitação e diminuindo seu tempo de armazenamento (Warner & Fehr 2008). Para aumentar sua estabilidade, o óleo de soja é submetido à hidrogenação química, que visa quebrar ligações duplas de ácidos graxos poli-insaturados, reduzindo seu teor e aumentando o teor de ácido oleico. Entretanto, este processo leva à produção de 10-40% de ácidos graxos *trans*, moléculas cujo consumo está associado com males como obesidade e doenças coronárias (Mozaffarian et al. 2006, Zaloga et al. 2006). Assim, muitos programas de melhoramento visam o desenvolvimento de cultivares de soja com alto teor de ácido oleico e baixo teor de ácidos graxos poli-insaturados (Bilyeu et al. 2006, Pham et al. 2012, Kim et al. 2015).

Por outro lado, maior tamanho de sementes, alto teor de proteína, alto teor de sacarose e baixos teores de oligossacarídeos de rafinose (rafinose e estaquiose – ROs) são objetivos comuns de programas de melhoramento visando o desenvolvimento de sojas especiais destinadas ao consumo humano (Jaureguy et al. 2011).

Características como teores de proteína, sacarose e oligossacarídeos de rafinose (ROs), e composição do óleo apresentam herança de natureza quantitativa.

Entretanto, pelo efeito de genes maiores, herança simples tem sido constatada em cruzamentos envolvendo linhagens mutantes (Bilyeu et al. 2006, Skoneczka et al. 2009, Pham et al. 2011, Bandillo et al. 2015). Mutações encontradas em genes específicos têm sido relatadas e associadas com fenótipos de interesse, sendo importantes ferramentas para o desenvolvimento de sojas com características especiais.

Pham et al. (2010) caracterizaram uma troca (C>G) PI283327 na posição 410 da sequência codificante do gene *GmFAD2-1B* (Glyma.20G111000), resultando na troca de um resíduo de prolina por outro de arginina em uma posição perfeitamente conservada. Posteriormente, Pham et al. (2011) identificaram uma deleção de uma adenosina na posição 544/545 da região codificante do gene *GmFAD2-1A* (Glyma.10G278000) do acesso PI603452, levando a alteração no quadro de leitura e um stop códon prematuro, após 191 aminoácidos. Estas duas mutações foram associadas com aumento no teor de ácido oleico, e sua combinação produziu grãos com teor de ácido oleico variando entre 83,9% e 86,4% em média, em experimento avaliado em dois ambientes (Pham et al. 2011). Ainda neste trabalho, os autores avaliaram o potencial da linhagem M23, que apresenta uma deleção genômica de 160 kb, incluindo o gene *GmFAD2-1A*, para a obtenção de soja com alto teor de ácido oleico. A combinação desta mutação com a mutação no gene *GmFAD2-1B* da PI283327 produziu grãos com 82,7% em média, em experimento avaliado nas mesmas condições.

A respeito do teor de ácido linolênico, Bilyeu et al. (2006) caracterizaram mutações nos genes FAD3 (*GmFAD3A*, *GmFAD3B* e *GmFAD3C*), associando-as com baixo teor de ácido linolênico. Destes, o gene *GmFAD3A* (Glyma.14G194300) apresenta o maior efeito sobre a característica (Bilyeu et al. 2005, Pinto et al. 2013). A combinação de mutações nos genes *GmFAD3A*, *GmFAD3B* (Glyma.02G227200) e *GmFAD3C* (Glyma.18G062000) possibilitou o desenvolvimento da linhagem A29, que possui cerca de 1% de ácido linolênico (Bilyeu et al. 2006).

Pham et al. (2012) combinaram mutações nos genes *GmFAD2-1A*, *GmFAD2-1B*, *GmFAD3A* e *GmFAD3C* para produzir óleo de soja com cerca de 85% de ácido oleico, 3,0% de ácido linoleico e 1,5% de ácido linolênico, uma

composição semelhante ao azeite de oliva, considerado o óleo vegetal natural mais saudável existente (White 2007).

Com relação à utilização da soja na alimentação, seu valor nutricional é limitado pela presença de fatores antinutricionais, como inibidores das proteases tripsina e quimiotripsina, compostos fenólicos (taninos), lectinas e ROs (Stech et al. 2011, Yang et al. 2014). A presença destes ROs é considerada um fator restritivo para o consumo de soja como fonte de proteica, uma vez que estes compostos não são completamente digeríveis por humanos e animais monogástricos, e seu consumo pode levar a disfunções como diarreia, náuseas e flatulência (Liener 1994, Reddy et al. 2016). Por outro lado, a presença de sacarose, o único oligossacarídeo nutricionalmente útil presente no grão de soja, apresenta grande importância na produção de vários derivados de soja, como leite de soja, tofu, miso e natto (Sato et al. 2014, Wang et al. 2014). Entretanto, desenvolvimento de sojas especiais combinando altos teores de proteína e sacarose tem sido dificultado pela correlação negativa existente entre estas características (Jaureguy et al. 2011, Sato et al. 2014).

Skoneczka et al. (2009) identificaram uma mutação no gene rafinose sintase 2 (*GmRS2* - Glyma.06G179200) do acesso PI200508, associada com baixos teores ROs e aumento no teor de sacarose. Considerando que sacarose, rafinose e estaquiose se encontram em uma mesma via metabólica (Dierking & Bilyeu 2008, Bilyeu & Wiebold 2016), e que a mutação encontrada na PI200508 apresenta pequeno efeito sobre o teor de proteína (Capítulo II), esta mutação se mostra de grande interesse para o melhoramento genético visando o desenvolvimento de cultivares de soja com altos teores de proteína e sacarose, e baixos teores de rafinose e estaquiose.

O método dos retrocruzamentos é um procedimento que permite a transferência de um ou poucos genes de um dos genitores (doador) para o genitor recorrente, que geralmente é um cultivar ou material elite de grande interesse. Por meio dele, é possível desenvolver uma nova planta, semelhante ao genitor recorrente na maioria das características, mas portando o fenótipo melhorado para a característica alvo (Montalván & Destro 1999). Com o intuito de facilitar a seleção dos indivíduos nos ciclos de seleção e diminuir o tempo gasto entre os ciclos de seleção, principalmente para características expressas somente no período

reprodutivo, como aquelas relacionadas aos grãos, a utilização de marcadores moleculares associados às características a serem introgredidas apresenta-se como estratégia ideal (Guimarães et al. 2009). Neste contexto, retrocruzamentos assistidos por marcadores (RAM) têm sido utilizados com sucesso no melhoramento de soja para introgressão de características como baixo teor de fitato (Landau-Ellis & Pantalone 2009, Boehm Jr 2014), ausência da α-subunidade de β-conglicinina (Song et al. 2014) e ausência do inibidor de tripsina Kunitz (Maranna et al. 2016).

A utilização de RAM para a introgressão de mutações em genes controlando características de qualidade do grão para cultivares elite é uma estratégia promissora para produzir novas cultivares de soja que possam atender a diferentes nichos do mercado. Neste sentido, o objetivo deste trabalho foi iniciar um programa de retrocruzamentos assistidos por marcadores para o desenvolvimento de cultivares de soja com características especiais.

2. MATERIAL E MÉTODOS

2.1. Material genético

A linhagem 4218, derivada do acesso A29 (baixíssimo teor de ácido linolênico), os acessos PI283327 e PI6033452 (alto teor de ácido oleico) e o acesso PI200508 (baixo teor de ROs) foram selecionados como doadores para os programas de retrocruzamentos. Estes acessos foram caracterizados anteriormente para as características mencionadas, e as mutações causais de seus fenótipos foram identificadas (Bilyeu et al. 2006, Pham et al. 2010, Pham et al. 2011, Skoneczka et al. 2009, Pham et al. 2014) (Tabela 1).

As variedades CS303TNKCA, M7110 IPRO, NA5909, TMG 4182 foram escolhidas como genitores recorrentes para o programa de melhoramento para qualidade do óleo. A cultivar UFVTN105 AP foi escolhida como genitor recorrente para o melhoramento visando à redução do teor de ROs e aumento no teor de sacarose. A variedade CS303TNKCA foi desenvolvida pelo Programa de Melhoramento da Qualidade da Soja (PMQS/BIOAGRO/UFV) a partir da linhagem BARC-12 (Leffel 1994) e da cultivar Monarca, e apresenta médio a alto

teor de óleo (20-23%), baixo teor de ácido linolênico (3,5-4,0%), ausência das três lipoxigenases (*lox1*, *lox2* e *lox3*) e do inibidor de protease Kunitz. A cultivar M7110 IPRO (Monsoy) é precoce, resistente a lagartas (Bt), tolerante ao glifosato (RR2), resistente ao acamamento, resistente à mancha olho-de-rã e de ampla distribuição geográfica. NA5909 (Nidera Sementes) é uma cultivar resistente ao glifosato, semiprecoce, resistente ao acamamento, resistente ao cancro da haste, mancha olho-de-rã e crestamento bacteriano, e adaptada ao sul do Brasil. A cultivar TMG 4182 (Tropical Melhoramento e Genética) é de ciclo precoce, moderadamente resistente ao acamamento, resistente ao cancro da haste, mancha olho-de-rã, pústula bacteriana e várias raças do nematoide do cisto e adaptada a vários estados. A variedade UFVTN105 AP foi desenvolvida pelo PMQS/BIOAGRO/UFV, e apresenta alto teor de proteína e ausência das três lipoxigenases (*lox1*, *lox2* e *lox3*).

Tabela 1 – Acessos doadores de alelos no programa de retrocruzamentos.

Acesso	**Característica**	**Mutação**	**Gene**	**Referência**
A29	Baixo linolênico	Deleção do gene	*GmFAD3A*	Pham et al. 2014
		G>A	*GmFAD3B*	Bilyeu et al. 2006
		C>T	*GmFAD3C*	Bilyeu et al. 2006
PI603452	Alto oleico	Deleção A	*GmFAD2-1A*	Pham et al. 2011
PI283327	Alto oleico	C>G	*GmFAD2-1B*	Pham et al. 2010
PI200508	Baixo ROs	Deleção TGG	*GmRS2*	Skoneczka et al. 2009

2.2. Extração de DNA, desenho dos conjuntos de *primers* e condições de HRM

Amostras foliares dos acessos constantes na Tabela 1 foram utilizadas para purificação de DNA genômico, conforme Doyle & Doyle (1990). O DNA purificado foi quantificado em espectrofotômetro NanoDrop (NanoDrop Technologies, Wilmington, DE), e sua qualidade foi verificada pro eletroforese em gel de agarose 0,8%.

As sequências dos genes constantes na Tabela 1 foram obtidas do banco de dados Phytozome (Goodstein et al. 2012) e utilizadas para desenhar conjuntos de *primers* flanqueando as mutações previamente identificadas nestes genes, com

auxílio do *software* Primer3 Input v. 0.4.0 (Untergrasser et al. 2012; Tabela 2), a exceção do gene *GmRS2*, para o qual foram utilizados os *primers* descritos no Capítulo II. Os produtos de amplificação foram enviados para sequenciamento (Macrogen Inc), e as sequências geradas foram editadas no *software* Sequencher v. 4.1.4 (Gene Codes Corporation), e alinhadas com o genoma de referência Williams 82. A partir da análise das sequências, foram desenhados novos conjuntos de *primers*, amplificando fragmentos variando entre 55 e 117 pb, contendo a região da mutação causal descrita para cada gene, com exceção do gene *GmFAD3A*, cuja mutação é a deleção completa do gene. Neste caso, os *primers* foram construídos flanqueando uma inserção de 10 pares de bases do acesso A29 a 1115 pb *downstream* ao gene (Tabela 3). O conjunto de *primers* desenhados para a mutação do gene *GmFAD3C* já havia sido desenhado anteriormente por nosso programa. As sequências dos *primers* foram avaliadas no banco de dados Phytozome (Goodstein et al. 2012), para verificar sua especificidade com a sequência alvo. O *software* uMelt (Dwight et al. 2011), uma ferramenta para predição virtual de curvas de *melt*, foi utilizado para estimar a eficiência de discriminação alélica de cada marcador.

Para avaliar a eficiência da distinção dos genótipos através da técnica de HRM, foram realizadas para cada par de *primers* uma reação HRM contendo um indivíduo mutante, um selvagem e um indivíduo F_1 entre os genótipos mutante e selvagem. As reações de HRM foram realizadas conforme descrito no item 2.4 do Capítulo I.

Tabela 2 – *Primers* utilizados para sequenciamento da região das mutações caracterizadas nos genes descritos na tabela 1.

Gene	***Primers* Para Sequenciamento**	***Amplicon***
GmFAD2-1A	Fw: TCATGCATTTGGTTCTGTGTC	971 pb
	Rv: TTTCTCCTTCATTGGGCATC	
GmFAD2-1B	Fw: TGTCACTTCCCTCCATTCATT	904 pb
	Rv: ACATAGCAGCCAAACCAACC	
GmFAD3A	Fw: TGTTGGTATGCTCACCCTGA	617 pb
	Rv: GTGTCCACTCTATCCCAGTTGTC	
GmFAD3B	Fw: AACCACATTTTGGATTTGGAAC	899 pb
	Rv: CGCATTTAAGTGTTCACCCACT	
GmFAD3C	Fw: ATTTATGTGCAGTTTTGCCATC	924 pb
	Rv: TGCAATTCTGACAAAGCACATA	
GmRS2	Fw: GGACTTGAAGGAACAGTTTAGG	185 pb
	Rv: CGTTACTGACGATCTTATCCAC	

Tabela 3 – *Primers* desenvolvidos para genotipagem das mutações pela metodologia de HRM, mostrando tamanho do fragmento e alelos selvagens e mutantes.

Gene	***Primers* HRM**	***Amplicon***	**Alelo Selvagem**	**Alelo Mutante**
GmFAD2-1A	Fw: AGTTGCATGGTTTTCCAAGT	70 pb	A	-
	Rv: TGTGAGTGTGACGAGAAGAGAAA			
GmFAD2-1B	Fw: TGGGTTTGACCGTTCACTCAG	65 pb	G	C
	Rv: GCGGCGATGGCTTATTTTC			
GmFAD3A	Fw: TCATTGGAGGGAGCAAAGGT	73 pb	-	AAGTTAAAGC
	Rv: TTCAATATTTTAGCTTGATCG			
GmFAD3B	Fw: TGCTTCCTGAGGCTGTTCTT	117 pb	G	A
	Rv: TGGCACAGTGAATCTAATGAGTC			
GmFAD3C	Fw: TGGCACCCATGTTATCCAT	73 pb	C	T
	Rv: GAGAATTAATACCGCTTCGATTAAA			
GmRS2	Fw: GTGGAGCAGGTGTATGTG	55 pb	-	TGG
	Rv: GTCTGACCCCACCCCAATAC			

2.3. Desenvolvimento de soja com mais de 80% de oleico

Sementes dos acessos PI603452 e PI283327 foram plantadas de modo escalonado, para permitir a coincidência de plantas em antese com plantas ainda não fecundadas. Os cruzamentos foram realizados utilizando o acesso PI603452 como doador de pólen e o acesso PI283327 como receptor. As sementes F_1 foram plantadas em vasos e produziram 163 sementes F_2, que foram plantadas em julho de 2014. Amostras foliares das plantas F_2 foram coletadas em estágio V2 (Fehr & Caviness 1977), congeladas em nitrogênio líquido, e armazenadas em freezer a -80 ^{o}C.

As 163 plantas foram genotipadas para os genes FAD2 (*GmFAD2-1A* e *GmFAD2-1B*). Aquelas que se mostraram homozigotas para as mutações dos dois genes foram colhidas e tiveram seu perfil de ácidos graxos analisado. A extração de lipídeos foi realizada conforme metodologia adaptada de Gesteira et al. (2003) (ver Capítulo I). A análise do perfil de ácidos graxos foi realizada em um cromatógrafo gasoso Shimadzu GC-2010 Plus, através de quantificação relativa de cada ácido graxo em relação ao total de ácidos graxos.

Sementes $F_{2:3}$ das plantas F_2 selecionadas foram plantadas e também tiverem se perfil de ácidos graxos analisado, conforme descrito acima. A planta F_2 que mostrou o perfil de ácidos graxos mais adequado e estável nas duas avaliações foi selecionada como doadora de alelos para o programa de retrocruzamentos.

2.4. Programa de retrocruzamentos assistidos por marcadores

Todos os cruzamentos iniciais e retrocruzamentos foram realizados em casa de vegetação com controle de temperatura. Os plantios foram realizados em vasos com solo adubado, contendo duas sementes para genótipos doadores de pólen e três sementes para receptores de pólen (genitores recorrentes). Foram desenvolvidos cinco programas de retrocruzamentos, um para cada variedade a ser melhorada (CS303TNKCA, M7110 IPRO, NA5909, TMG 4182 e UFVTN105 AP), conforme Tabela 4.

Tabela 4 – Programas de retrocruzamentos desenvolvidos.

Variedade recorrente	**Doadores**	**Característica a ser transferida**	**Genes**
CS303TNKCA	Planta AO*	Alto oleico	*GmFAD2*
M7110 IPRO	Planta AO*	Alto oleico	*GmFAD2*
	Linhagem 4218**	Baixo linolênico	*GmFAD3*
NA5909	Planta AO*	Alto oleico	*GmFAD2*
	Linhagem 4218**	Baixo linolênico	*GmFAD3*
TMG 4182	Planta AO*	Alto oleico	*GmFAD2*
	Linhagem 4218**	Baixo linolênico	*GmFAD3*
UFVTN105 AP	PI200508	Baixo estaquiose	*GmRS2*

* Planta derivada do cruzamento entre os acessos PI603452 e PI283327, conforme item 2.3.
** Linhagem derivada do acesso A29.

Para fazer uma estimativa do número de cruzamentos necessários a serem realizados para cada RAM, foi utilizada a seguinte fórmula:

$$c = 1 - (1 - p)\text{^}n$$

onde *c* é a certeza de se encontrar ao menos uma planta contendo o genótipo desejado, dado o número de plantas avaliado; *p* é a probabilidade de uma planta conter o genótipo desejado; e *n* é o número de plantas avaliado. Fixando-se um valor de certeza (*c*), é possível, aplicando logaritmo, determinar o número mínimo de plantas a serem avaliadas para encontrar este valor de certeza, dada a probabilidade de se encontrar o genótipo desejado em um cruzamento.

Para as cultivares M7110 IPRO, NA5909 e TMG 4182 foram conduzidos dois subprogramas de retrocruzamentos, um com os genes *GmFAD2-1* e outro com os genes *GmFAD3*. Ao final dos ciclos de retrocruzamentos dos dois subprogramas, as plantas geradas serão cruzadas, para combinar os alelos dos cinco genes.

Para iniciar os programas de retrocruzamentos, foram realizados plantios escalonados das variedades recorrentes. A cada dois dias, foram plantados dois vasos contendo três sementes para cada variedade, num período mínimo de duas semanas. Após uma semana do primeiro plantio, os genitores doadores foram

plantados em vasos contendo duas sementes cada. Os cruzamentos foram realizados conforme descrito por Borém et al. (2009). As flores polinizadas foram etiquetadas para marcação dos cruzamentos. Posteriormente, as vagens oriundas de cruzamentos foram colhidas separadamente daquelas geradas por autofecundação.

Para a realização da primeira geração de retrocruzamentos, o esquema de plantio foi realizado conforme descrito acima. Após a identificação das plantas F_1, foi realizada a genotipagem através da técnica de HRM, conforme descrito no item 2.2, utilizando os marcadores descritos na Tabela 3, visando à confirmação dos cruzamentos e presença dos alelos favoráveis. As plantas F_1 que se mostraram heterozigotas para todos os genes indicados na Tabela 4 foram selecionadas para a realização da primeira geração de retrocruzamentos (RC1). As demais plantas foram descartadas. O procedimento foi repetido para todas as outras gerações de retrocruzamentos.

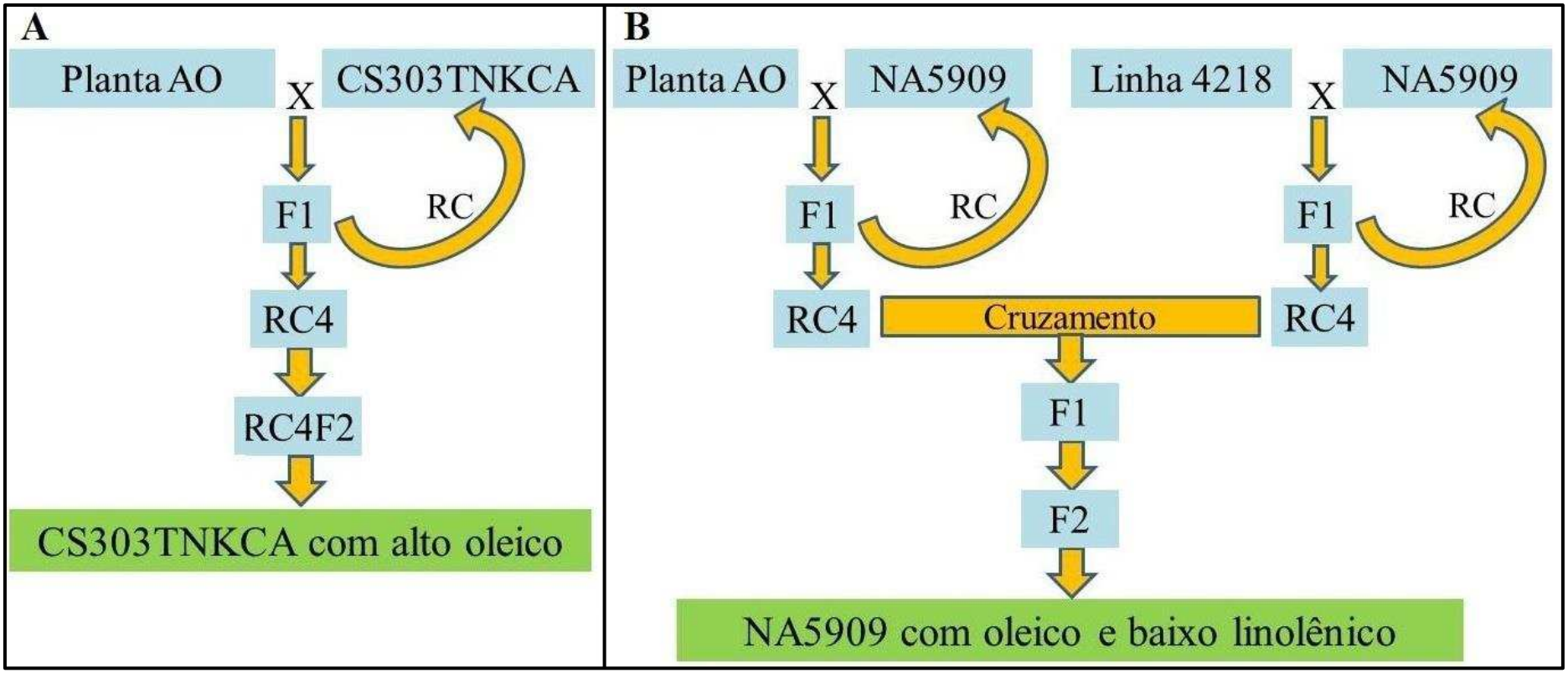

Figura 1 – (**A**) Esquema do programa de retrocruzamentos assistidos (variedade CS303TNKCA). (**B**) Esquema do programa de retrocruzamentos assistidos divididos em dois subprogramas (variedade NA5909). Planta AO = planta doadora de alelos para alto oleico (derivada dos acessos PI283327 e PI603452); Linha 4218 = planta doadora de alelos para baixo linolênico (derivada do acesso A29).

3. RESULTADOS E DISCUSSÃO

3.1. Marcadores HRM

Todos os marcadores desenvolvidos foram capazes de discriminar indivíduos homozigotos mutantes, homozigotos selvagens e heterozigotos. A Figura 2 apresenta as curvas de *melt* geradas para cada marcador HRM.

A metodologia de HRM foi desenvolvida para detectar variações genéticas em sequências de DNA, através da determinação da curva de *melt* de uma sequência particular. A forma da curva de *melt* depende do tamanho do *amplicon*, da sequência, conteúdo de GC e complementariedade das fitas de DNA (Simko 2016). Estes fatores são influenciados pelo tipo de mutação presente no *amplicon* do marcador HRM, definindo a eficiência da técnica em discriminar alelos. Assim, é notável que houve maior eficiência de discriminação alélica para os marcadores baseados nas mutações do genes *GmFAD3A* (inserção de 10 pb), *GmFAD3B* (G>A), *GmFAD3C* (C>T) e *GmRS2* (deleção de 3 pb). Por outro lado, os marcadores baseados nas mutações dos genes *GmFAD2-1A* (deleção de uma adenina) e *GmFAD2-1B* (G>C) apresentaram curvas de *melt* mais próximas, tornando a discriminação alélica mais dificultada.

3.2. Desenvolvimento de soja com mais de 80% de ácido oleico

Dentre as 163 plantas F_2 geradas a partir do cruzamento entre os acessos PI283327 e PI603452, foram identificadas oito plantas homozigotas mutantes para os dois genes, através dos marcadores HRM desenvolvidos para as mutações nos genes *GmFAD2-1A* e *GmFAD2-1B* (χ^2 = 0,501 ns). A partir da análise do perfil de ácidos graxos de sementes $F_{2:3}$ e $F_{2:4}$ destas oito plantas (Tabela 5), a planta PL04 apresentou o perfil de ácidos mais estável e adequado, sendo selecionada como doadora de alelos para o programa de retrocruzamentos.

Foi verificada uma grande variação no teor de ácido oleico das outras plantas avaliadas (Tabela 5). De modo que o teor de ácido oleico no grão de soja é uma característica quantitativa (Bachlava et al. 2009), a influência do ambiente e de outros genes envolvidos no controle da característica podem explicar em parte esta variação, assim como a possibilidade de erro na identificação dos genótipos.

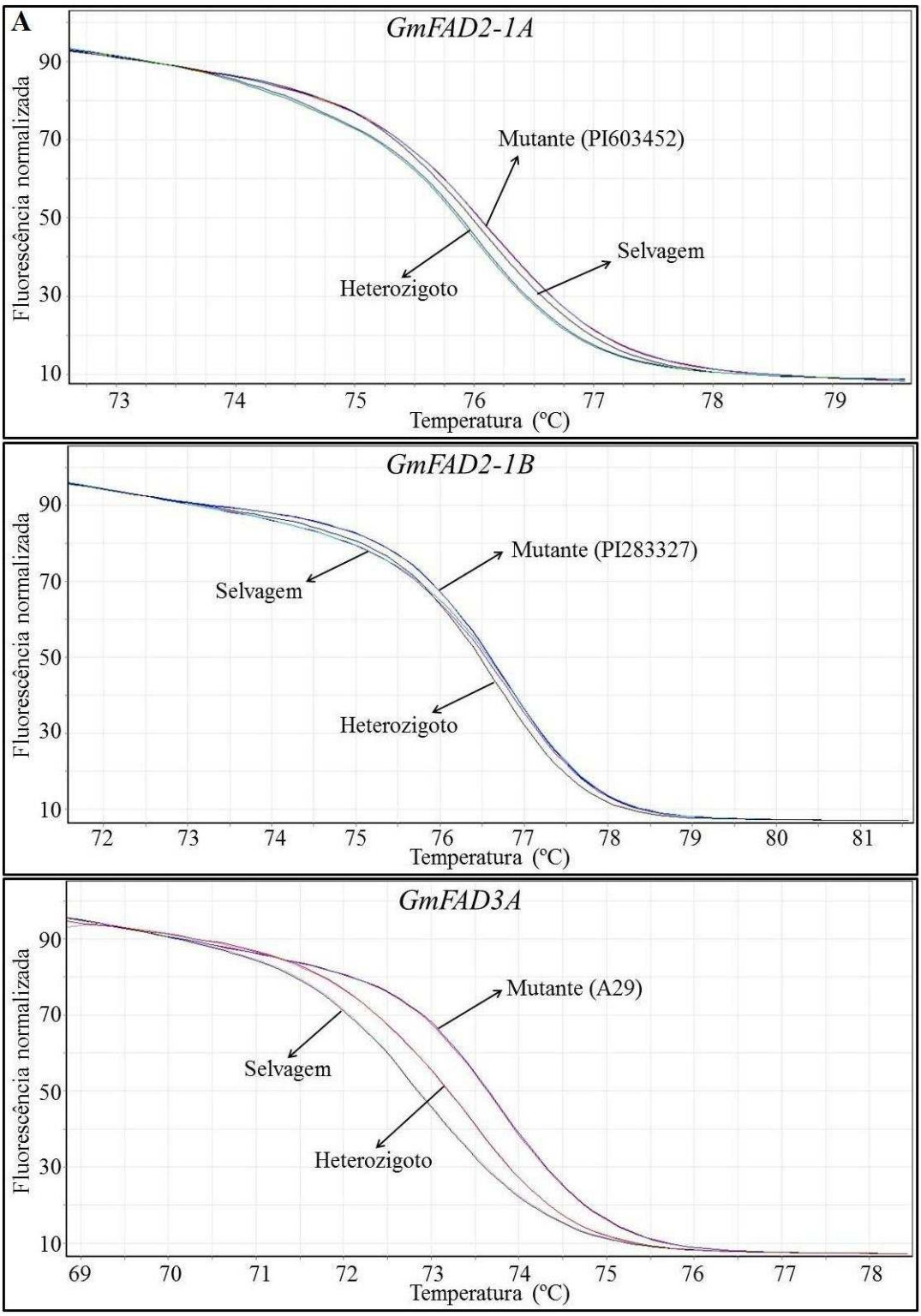

Figura 2A – Curvas de *melt* normalizadas geradas por marcadores HRM baseados em mutações nos genes *GmFAD2-1A*, *GmFAD2-1B* e *GmFAD3A*, mostrando indivíduos mutantes, selvagens e heterozigotos.

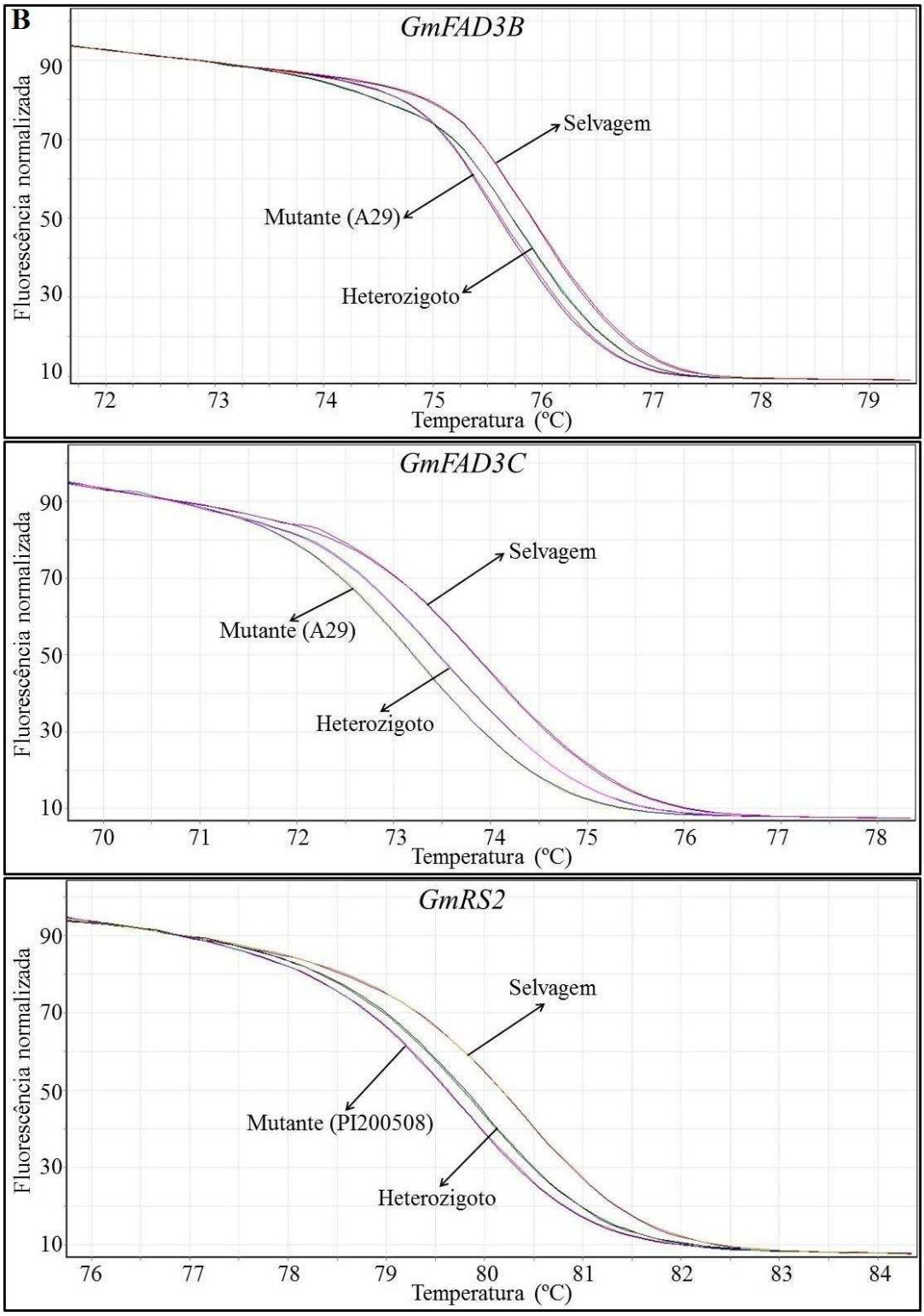

Figura 2B – Curvas de *melt* normalizadas geradas por marcadores HRM baseados em mutações nos genes *GmFAD3B*, *GmFAD3C* e *GmRS2*, mostrando indivíduos mutantes, selvagens e heterozigotos.

Tabela 5 – Perfil de ácidos graxos das oito plantas de uma população F_2 derivada do cruzamento entre os acessos PI603452 e PI283327, identificadas através dos marcadores HRM baseados nas mutações dos genes *GmFAD2-1A* e *GmFAD2-1B*, evidenciando o teor de ácido oleico da planta PL04.

Plantas	**Palmítico**		**Esteárico**		**Oleico**		**Linoleico**		**Linolênico**	
	$F_{2:3}$	**$F_{2:4}$**	**$F_{2:3}$**	**$F_{2:4}$**	**$F_{2:3}$**	**$F_{2:4}$**	**$F_{2:3}$**	**$F_{2:4}$**	**$F_{2:3}$**	**$F_{2:4}$**
PL04	6,1711	6,3362	3,0269	3,4717	85,6467	84,2320	1,6530	2,2236	3,5188	3,7365
PL21	8,7650	6,3719	3,3407	3,5061	55,6846	85,6508	25,8523	1,5504	6,3573	2,9209
PL61	10,4592	10,6648	2,8921	3,2826	35,5134	32,5494	43,5392	45,6788	7,5921	7,8244
PL79	8,8400	9,7408	3,2633	3,7669	57,0033	47,3697	24,0000	31,4872	6,8900	7,6354
PL84	9,0333	8,9670	2,8233	3,3901	42,4100	50,4928	38,1333	29,9097	7,5900	7,2404
PL88	8,2867	9,8941	3,1566	3,1932	60,7400	39,5852	22,4133	40,5210	5,3967	6,8065
PL106	9,2250	8,7709	3,3400	3,2485	53,3650	67,2709	26,6550	15,5692	7,4100	5,1406
PL126	8,2867	7,5811	3,0733	3,3104	60,2066	70,6443	21,4733	12,9539	6,9367	5,5102

A PL04 apresentou 84,94% de ácido oleico, na média das avaliações de sementes $F_{2:3}$ e $F_{2:4}$. Este valor é semelhante aos encontrados por Pham et al. (2011), que obtiveram plantas contendo entre 83,9% e 86,4% de ácido oleico em média, avaliadas em dois ambientes. A combinação das mutações nos genes *GmFAD2-1A* e *GmFAD2-1B* dos acessos PI603452 e PI283327, respectivamente, possibilita a obtenção dos mais altos teores de ácido oleico encontrados para soja.

Concomitantemente ao alto teor de ácido oleico, a PL04 apresenta reduzidos teores médios de ácido linoleico (1,94%) e linolênico (3,63%). Correlações negativamente significativas entre ácido oleico e os ácidos linoleico e linolênico têm sido relatada para várias espécies, incluindo a soja (Gross & Stefansson 1966, Yermanos et al. 1972, Hammond & Fehr 1975, Léon et al. 2004, Van der Merwe et al. 2012). Como estes ácidos se encontram numa mesma via metabólica, onde as ω6-dessaturases convertem ácido oleico em linoleico, e as ω3-dessaturases convertem ácido linoleico em linolênico (Zhang et al. 2012), é esperado que o aumento no teor de ácido oleico seja acompanhado por queda no teor de ácidos graxos poli-insaturados. É importante ressaltar que a combinação das mutações em *GmFAD2-1A* e *GmFAD2-1B*, sem nenhuma mutação nos genes *GmFAD3* possibilitou a obtenção de soja com 3,63% de ácido linolênico em média, valor semelhante ao encontrado para linhagens portando mutações no gene *GmFAD3A*, como a linhagem PI361088B, que possui cerca de 3,8% (Chappell & Bilyeu 2007), a linhagem RG10 (3,4%; Reinprecht et al. 2009), a linhagem A5 (3,5-4,1%; Pham et al. 2014), a linhagem PE1690 (5,1%; Kim et al. 2015), e a variedade CS303TNKCA (Capítulo I).

3.3. Definição dos programas de retrocruzamentos assistidos por marcadores

Foram iniciados cinco programas de RAM. A PL04, que apresentou cerca de 85% de ácido oleico (item 2.3), foi selecionada como doadora de alelos para as variedades CS303TNKCA, M7110 IPRO, NA5909 e TMG 4182. A linhagem 4218, derivada da A29, genótipo com cerca de 1% de ácido linolênico (Bilyeu et al. 2006), foi utilizada como doadora de alelos para as variedades M7110 IPRO, NA5909 e TMG 4182. Por fim, a PI200508, que apresenta baixo teor de estaquiose

(Skoneczka et al. 2009), foi utilizada para introgredir esta característica na variedade UFVTN105 AP.

CS303TNKCA é uma variedade de soja desenvolvida pelo PMQS/BIOAGRO/UFV, e as bases para seu baixo teor de ácido linolênico (cerca de 3,5-4,0%) foram caracterizadas no Capítulo I. Conforme Pham et al. (2012), a combinação entre mutações nos genes *GmFAD2-1A*, *GmFAD3B* e *GmFAD3A* é capaz de produzir soja com cerca de 85% de ácido oleico e 1,8-2,6% de ácido linolênico. Valores abaixo de 3% são o padrão para óleo de soja com baixo teor de ácido linolênico para a indústria (Pham et al. 2012). Assim, a combinação das mutações nos genes *GmFAD2-1A* e *GmFAD2-1B* com a mutação no gene *GmFAD3A* da CS303TNKCA, aliada a outras características desta variedade, como teor de óleo, ausência das três lipoxigenases e produtividade média, podem torná-la uma boa variedade produtora de óleo vegetal de qualidade.

As cultivares M7110 IPRO, NA5909 e TMG 4182 apresentam bom desempenho agronômico e estão adaptadas a importantes regiões produtoras de soja no país. O melhoramento destas cultivares visando o aumento do teor de ácido oleico e redução dos teores de ácidos graxos poli-insaturados possibilitará a oferta de óleo de soja de maior qualidade no mercado, apresentando maior estabilidade e com menor custo de beneficiamento para atender a indústria alimentícia e de biocombustíveis. Como estas cultivares apresentam teores normais de ácidos graxos, optou-se pela transferência dos alelos presentes nos genes *GmFAD2-1A*, *GmFAD2-1B*, *GmFAD3A*, *GmFAD3B* e *GmFAD3C*.

A cultivar UFVTN105 AP foi desenvolvida pelo PMQS/BIOAGRO/UFV para uso na alimentação humana, e apresenta alto teor de proteína e ausência das três lipoxigenases. Conforme discutido no Capítulo II, o acesso PI200508 apresenta uma mutação no gene *GmRS2*, que está associada com baixo teor de estaquiose e aumento no teor de sacarose (Skoneczka et al. 2009), mas que apresenta pequeno efeito sobre o teor de proteína. Deste modo, a PI200508 foi selecionada como doadora de alelos para baixo teor de estaquiose e alto teor de sacarose, sem que ocorra diminuição do teor de proteína, possibilitando desenvolver uma variedade de soja mais adequada ao consumo humano.

Durante o planejamento de um programa de retrocruzamentos, é importante avaliar o número de plantas geradas a cada ciclo, de modo que seja possível conduzir o maior número de programas ao mesmo tempo, mas que seja possível avaliar um número seguro de plantas para cada programa. Para tal, deve se levar em conta, dentre outros fatores, o número de genes e a probabilidade de se encontrar os genótipos desejados (heterozigotos) para cada programa; a possibilidade de perda de material genético por problemas com germinação, pragas, doenças e outros; a destreza e experiência da pessoa que executará os cruzamentos; tempo e custos demandados para a análise das progênies, que no caso de um programa de RAM está associado com a genotipagem. Além disso, ao se trabalhar com vários programas ao mesmo, o espaço necessário para a condução dos programas também deve ser levado em conta. Neste sentido, utilizou-se uma fórmula para determinar o número necessário de plantas geradas em um cruzamento, para se ter 90% ou 95% de certeza de ser encontrada ao menos uma planta contendo alelos favoráveis para todos genes envolvidos no programa de retrocruzamentos. Os valores encontrados para cada variedade estão descritos na Tabela 6.

O programa de retrocruzamento da variedade UFVTN105 AP envolve apenas um gene (*GmRS2*). Deste modo, a cada geração são necessárias apenas quatro plantas por cruzamentos para se ter 90% de certeza de que pelo menos uma delas seja heterozigota, e cinco plantas para se obter 95% de certeza. Já para a variedade CS303TNKCA, o programa de retrocruzamentos envolve dois genes (*GmFAD2-1A* e *GmFAD2-1B*). Neste caso são necessárias 9 plantas para alcançar 90% de certeza, e 11 plantas para alcançar 95% de certeza.

Entretanto, para os programas de RAM das cultivares M7110 IPRO, NA5909 e TMG 4182, que envolvem cinco genes (*GmFAD2-1A*, *GmFAD2-1B*, *GmFAD3A*, *GmFAD3B* e *GmFAD3C*), seria necessário avaliar 73 ou 95 plantas para cada retrocruzamento, para se ter respectivamente 90% ou 95% de certeza de se encontrar ao menos uma planta heterozigota para os cinco genes. Levando-se em consideração a condução de todos os programas ao mesmo tempo, e que trabalhar com apenas uma planta selecionada por geração pode aumentar a chance de perda do material genético, optamos por dividir os programas de RAM destas três cultivares em dois subprogramas em paralelo, um para a transferência das mutações dos genes *GmFAD2* e outro para a transferência dos genes *GmFAD3*. Desta forma,

serão necessárias 9 ou 11 plantas para se alcançar 90 e 95% de certeza para os genes *GmFAD2*, e 18 ou 23 plantas para se alcançar 90 e 95% de certeza para os genes *GmFAD3*. A redução no número de plantas necessárias a serem geradas e avaliadas possibilitou aumentar a viabilidade da condução dos programas, tanto do ponto de vista prático (espaço em casa de vegetação, rendimento da pessoa responsável pelos cruzamentos e esforços gastos para a realização da genotipagem), quanto do ponto de vista econômico (número de plantas a serem genotipadas).

Tabela 6 – Número de plantas a serem geradas em um retrocruzamento para se encontrar ao menos uma planta contendo alelos favoráveis para todos os genes a serem melhorados para cada variedade, com 90% ou 95% de certeza.

Variedade	**Genes envolvidos**	***p***	***c***	***n***
CS303TNKCA	2	1/4	90%	9
			95%	11
M7110 IPRO	5	1/32	90%	73
			95%	95
NA5909	5	1/32	90%	73
			95%	95
TMG 4182	5	1/32	90%	73
			95%	95
UFVTN105 AP	1	1/2	90%	4
			95%	5
-	3	1/8	90%	18
			95%	23

p = probabilidade de encontrar um genótipo heterozigoto para todos os genes envolvidos;

c = percentual de certeza de se encontrar ao menos uma planta contendo o genótipo desejado;
n = número de plantas a serem avaliadas

3.4. Condução dos retrocruzamentos

Os programas de retrocruzamentos assistidos por marcadores foram iniciados em julho de 2014. Devido à quantidade de cruzamentos a serem realizados, optou-se por escalonar a realização dos cruzamentos, de acordo com a

disponibilidade de espaço na casa de vegetação e do rendimento do responsável pela realização dos cruzamentos. A Tabela 7 apresenta um resumo do andamento dos programas de RAM.

Em um cruzamento, o indivíduo F_1 apresenta 50% do genoma de cada genitor. A cada geração de retrocruzamento, em média, o genoma do genitor doador reduz pela metade. Assim, após a geração RC1, espera-se que uma planta possua 25% de genoma doador, e consequentemente 75% do genoma recorrente. Seguindo-se este raciocínio, espera-se 87,5% do genoma recorrente em plantas da geração RC2, 93,75% na geração RC3 e 96,875% na geração RC4, e assim por diante. Para avaliar a porcentagem de progênies segregantes similares aos genitores comerciais em termos agronômicos em cada geração de retrocruzamentos, Carpenter & Fehr (1986) introgrediram germoplasma de *Glycine soja*, o parente mais próximo da espécie *G. max*, em dois cultivares comerciais de soja (Amsoy-71 e Century), realizando cinco gerações de retrocruzamentos com estas cultivares. Os valores encontrados foram de zero para RC0 (F_1) e RC1, 2% para RC2, 22% para RC3, 51% para RC4 e 65% para RC5. Estes resultados mostram que são necessárias pelo menos duas gerações de retrocruzamentos para recuperar linhagens segregantes agronomicamente aceitáveis, quando ao menos 50 plantas são retrocruzadas em cada geração, e que a geração RC4 apresentou metade das linhagens agronomicamente aceitáveis. Martins et al. (2002) realizaram cruzamentos múltiplos entre a variedade de soja FT-Cristalina e linhagens de soja alelos para ausência das três lipoxigenases. Em seguida, utilizaram quatro gerações de retrocruzamentos para desenvolver linhagens triplo-nulas (TN) e triplo-positivas para as três lipoxigenases, e agronomicamente superiores à variedade FT-Cristalina original. Landau-Ellis & Pantalone (2009) utilizaram quatro gerações de retrocruzamentos para transferir a característica de baixo teor de fitato do acesso CX1834-1-2 para a cultivar de soja 5601T. Dentre 15 plantas heterozigotas $RC4F_1$, duas mostraram perfeita identidade genética com o genoma do genitor recorrente, através de análise com marcadores SSR. Deste modo, optamos pela realização de quatro gerações de retrocruzamentos para a condução de nossos programas de RAM.

Tabela 7 – Resumo do andamento dos programas de retrocruzamentos assistidos por marcadores. O número de sementes geradas em cada geração de cruzamentos, e o número de plantas selecionadas a partir das sementes geradas por cruzamentos são mostrados.

Geração		**CS303TNKCA X PL04**	**M7110 IPRO X PL04**	**M7110 IPRO X 4218**	**NA5909 X PL04**	**NA5909 X 4218**	**TMG 4182 X PL04**	**TMG 4182 X 4218**	**UFVTN105 AP X PI200508**
F_1	Geradas	1	15	16	47	18	4	13	2
	Selecionadas	1	6	2	4		4	3	2
RC1	Geradas	16	23	6	13		21	8	11
	Selecionadas	1	5	1	4		4		3
RC2	Geradas	11	97	4	81		102		18
	Selecionadas	1	5		6		6		8
RC3	Geradas	62							75
	Selecionadas	4							11

PL04 = linha com cerca de 85% de ácido oleico; 4218 = linhagem derivada do acesso A29, com cerca de 1% de ácido linolênico.

Até o momento, os programas de RAM para as variedades CS303TNKCA e UFVTN105 AP foram avançados até a geração RC3. A realização dos cruzamentos para a geração RC4 foi iniciada em janeiro de 2017, com previsão de colheita para abril e maio de 2017.

Para a cultivar M7110 IPRO, o subprograma de retrocruzamentos para alto teor de ácido oleico foi avançado até a geração RC2. A realização dos cruzamentos para RC3 foi iniciada em janeiro de 2007, com previsão de colheita para abril e maio de 2017. O subprograma para baixo teor de ácido linolênico também se encontra na geração RC2.

Com relação à cultivar NA5909, foram realizadas duas gerações de retrocruzamentos para alto teor de oleico (RC2), e a realização dos cruzamentos para RC3 foi iniciada em janeiro de 2007, conforme a cultivar M7110 IPRO. Por outro lado, o subprograma para baixo teor de ácido linolênico se encontra na geração F_1.

Foram realizadas duas gerações de retrocruzamentos para alto teor de oleico para a cultivar TMG 4182. A realização da geração RC3 foi iniciada em janeiro de 2017, conforme às demais cultivares. Já para o subprograma de retrocruzamentos para baixo teor de ácido linolênico, foi realizada uma geração de retrocruzamentos (RC1).

Retrocruzamentos assistidos por marcadores é uma estratégia interessante para a introgressão de um ou pouco alelos em cultivares agronomicamente superiores, e tem sido utilizada com sucesso no melhoramento de soja (Martins et al. 2002, Landau-Ellis & Pantalone 2009, Boehm Jr 2014, Song et al. 2014, Maranna et al. 2016). Neste trabalho, foram utilizados retrocruzamentos assistidos por marcadores para incorporar características de qualidade do grão em diferentes cultivares de soja. O desenvolvimento de cultivares de soja com características especiais é de grande importância para diversificar o mercado, fornecendo alimento de maior qualidade para a alimentação humana e animal, óleo de maior qualidade para a indústria alimentícia e de biocombustíveis, além de reduzir custos e esforços em seu beneficiamento.

4. CONCLUSÕES

- Os marcadores HRM desenvolvidos para as mutações nos genes *GmFAD2-1A*, *GmFAD2-1B*, *GmFAD3A*, *GmFAD3B*, *GmFAD3C* e *GmRS2* foram eficientes para a identificação dos genótipos;
- A PL04 se mostrou uma boa doadora para alto teor de ácido oleico;
- O programa de retrocruzamentos assistidos por marcadores é um método interessante para a incorporação de novos alelos em cultivares de soja, visando à produção de sojas especiais.

5. PERSPECTIVAS

- Realizar novas gerações de retrocruzamentos para as cultivares M7110 IPRO, NA5909 e TMG 4182, visando alcançar a geração RC4;
- Realizar o cruzamento entre plantas $RC4F_1$ dos subprogramas de RAM para alto oleico e baixo linolênico das cultivares M7110 IPRO, NA5909 e TMG 4182;
- Produzir as gerações $RC4F_2$ para todos os programas de RAM e identificar as linhagens homozigotas para as mutações introgredidas;
- Avaliar as linhagens em ensaios de campo para identificar aquelas mais promissoras para serem lançadas novas cultivares.

REFERÊNCIAS BIBLIOGRÁFICAS

BACHLAVA, E. et al. Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. **Crop Science**, v. 49, n. 2, p. 433-442, 2009.

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, 2015.

BILYEU K, D. et al. Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. **Crop Science:** v. 45, n. 5, p. 1830-1836, 2005.

BILYEU, K. D.; WIEBOLD, W. J. Environmental stability of seed carbohydrate profiles in soybeans containing different alleles of the raffinose synthase 2 (RS2) gene. **Journal of Agricultural and Food Chemistry**, v. 64, n. 5, p. 1071-1078, 2016.

BOEHM JR, J. D. **Molecular Marker Assisted Backcross Development and Evaluation of an Environmentally Friendly, Commercially Acceptable Low Seed Phytate Soybean**. Dissertação (Mestrado) – University of Tennessee. 135p. Knoxville, Tennesse, EUA, 2014.

BORÉM, A.; ALMEIDA, L. A.; KIIHL, R. A. Hibridação em soja. In: BORÉM, A. Hibridação artificial de plantas. 2ª Edição, Universidade Federal de Viçosa. 625p. Viçosa-MG, 2009.

CARPENTER, J. A.; FEHR, W. R. Genetic variability for desirable agronomic traits in populations containing *Glycine soja* germplasm. **Crop Science**, v. 26, n. 4, p. 681-686, 1986.

CHAPPELL, A. S.; BILYEU, K. D. The low linolenic acid soybean line PI 361088B contains a novel mutation. **Crop Science**, v. 47, n. 4, p. 1705-1710, 2007.

DIERKING, E. C.; BILYEU, K. D. Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. **The Plant Genome**, v. 1, n. 2, p. 135-145, 2008.

DOYLE, J. J.; DOYLE, J. L. Isolation of plant DNA from fresh tissue. **Focus**, v. 12, p. 13-15, 1990.

DWIGHT, Z.; PALAIS, R.; WITTWER, C. T. uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. **Bioinformatics**, v. 27, n. 7, p. 1019-1020, 2011.

FEHR, W. R.; CAVINESS, C. E. Stages of soybean development. Special Report: **Agriculture and Home Economics Experiment Station**. Iowa State University, 1977.

GESTEIRA, A. S. et al. Biometrical analyses of linolenic acid content of soybean seeds. **Genetics and Bolecular Biology**, v. 26, n. 1, p. 65-68, 2003.

GOODSTEIN, D. M. et al. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, n. D1, p. D1178-D1186, 2012.

GROSS, A. T. H.; STEFANSSON, B. R. Effect of planting date on protein, oil, and fatty acid content of rape seed and turnip rape. **Canadian Journal of Plant Science**, v. 46, n. 4, p. 389-395, 1966.

GUIMARÃES, C. T. et al. Marcadores Moleculares no Melhoramento. In: BORÉM, A.; CAIXETA, E. T. **Marcadores Moleculares**. 2ª Edição, Universidade Federal de Viçosa. 532p. Viçosa-MG, 2009.

HAMMOND, E. G.; FEHR, W. R. Oil Quality Improvement in Soybeans-*Glycine max* (L.) Merr. **Fette, Seifen, Anstrichmittel**, v. 77, n. 3, p. 97-101, 1975.

JAUREGUY, L. M.; CHEN, P.; SCABOO, A. M. Heritability and correlations among food-grade traits in soybean. **Plant Breeding**, v. 130, n. 6, p. 647-652, 2011.

KATHMALE, D. K.; ANDHALE, A. U.; DESHMUKH, M. P. Growth and Yield of Soybean Genotypes as Influenced by Sowing Time at Different Locations under Climate Change Situation in Maharashtra, India. **International Journal of Bio-resource and Stress Management**, v. 4, n. 4, p. 492-495, 2013.

KIM, M. et al. A new low linolenic acid allele of GmFAD3A gene in soybean PE1690. **Molecular Breeding**, v. 35, n. 8, p. 155, 2015.

LANDAU-ELLIS, D.; PANTALONE, V. R. Marker-assisted backcrossing to incorporate two low phytate alleles into the Tennessee soybean cultivar 5601T. **Accounting Principles Fifth Canadian Edition**, v. 1, p. 316-318, 2009.

LEFFEL, R. C. Registration of BARC-12, a low linolenic acid soybean germplasm line. **Crop Science**, v. 34, n. 5, p. 1426-1427, 1994.

LEÓN, L.; MARTÍN, L. M.; RALLO, L. Phenotypic correlations among agronomic traits in olive progenies. **Journal of the American Society for Horticultural Science**, v. 129, n. 2, p. 271-276, 2004.

LIENER, I. E. Implications of antinutritional components in soybean foods. **Critical Reviews in Food Science & Nutrition**, v. 34, n. 1, p. 31-67, 1994.

MARANNA, S. et al. Introgression of null allele of Kunitz trypsin inhibitor through marker-assisted backcross breeding in soybean (*Glycine max* L. Merr.). **BMC Genetics**, v. 17, n. 1, p. 106, 2016.

MARTINS, C. A. O. et al. Efeito da eliminação genética das lipoxigenases das sementes sobre as características agronômicas da soja. **Pesquisa Agropecuária Brasileira**, v. 37, n. 10, p. 1389-1398, 2002.

MONTALVÁN, R.; DESTRO, D. Método dos retrocruzamentos. In: DESTRO, D.; MONTALVÁN, R. **Melhoramento genético de plantas**. Universidade Estadual de Londrina. 749p. Londrina-PR, 1999.

MOZAFFARIAN, D. et al. Trans fatty acids and cardiovascular disease. **New England Journal of Medicine**, v. 354, n. 15, p. 1601-1613, 2006.

PHAM, A. T. et al. Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. **BMC Plant Biology**, v. 10, n. 1, p. 195, 2010.

PHAM, A. T. et al. A novel FAD2-1 A allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. **Theoretical and Applied Genetics**, v. 123, n. 5, p. 793-802, 2011.

PHAM, A. T.; SHANNON, J. G.; BILYEU, K. D. Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. **Theoretical and Applied Genetics**, v. 125, n. 3, p. 503-515, 2012.

PHAM, A. T. et al. Characterization of the fan1 locus in soybean line A5 and development of molecular assays for high-throughput genotyping of FAD3 genes. **Molecular breeding**, v. 33, n. 4, p. 895-907, 2014.

PINTO, M. O. et al. Associação de marcadores moleculares SNP com o conteúdo de ácido linolênico em sementes de soja. **Pesquisa Agropecuária Brasileira**, v. 48, n. 3, p. 263-269, 2013.

REDDY, K. R. et al. Ultraviolet-B Radiation Alters Soybean Growth and Seed Quality. **Food and Nutrition Sciences**, v. 7, n. 01, p. 55, 2016.

REINPRECHT, Y. et al. Molecular basis of the low linolenic acid trait in soybean EMS mutant line RG10. **Plant Breeding**, v. 128, n. 3, p. 253-258, 2009.

SATO, T. et al. Near-infrared reflectance calibrations for determining sucrose content in soybean breeding using artificial reference samples. **Plant Breeding**, v. 131, n. 4, p. 531-534, 2012.

SIMKO, I. High-resolution DNA melting analysis in plant research. **Trends in Plant Science**, v. 21, n. 6, p. 528-537, 2016.

SKONECZKA, J. A. et al. Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI200508. **Crop Science**, v. 49, n. 1, p. 247-255, 2009.

SONG, B. et al. Marker-assisted backcrossing of a null allele of the α-subunit of Soybean (*Glycine max*) β-conglycinin with a Chinese soybean cultivar (a). The development of improved lines. **Plant Breeding**, v. 133, n. 5, p. 638-648, 2014.

STECH, M. R.; CARNEIRO, D. J.; CARVALHO, M. R. B. Fatores antinutricionais e coeficientes de digestibilidade aparente da proteína de produtos de soja para o pacu (*Piaractus mesopotamicus*). **Acta Scientiarum. Animal Sciences**, v. 32, n. 3, p. 255-262, 2010.

TAI, G. et al. The Results of Using Simple Backcross Breeding Method on Resistant to Frogeye Leaf Spot of Soybean. **Chinese Agricultural Science Bulletin**, v. 5, p. 002, 2002.

UNTERGASSER, A. et al. Primer3—new capabilities and interfaces. **Nucleic Acids Research**, v. 40, n. 15, e115, 2012.

VAN DER MERWE, R. et al. Physicochemical and oxidative stability characteristics of high-and mid-oleic sunflower seed oil. In: **Proceedings of the 18th Intern. Sunf. Conf. Mar del Plata**. 2012. p. 955-960.

WANG, Y.; CHEN, P.; ZHANG, B. Quantitative trait loci analysis of soluble sugar contents in soybean. **Plant Breeding**, v. 133, n. 4, p. 493-498, 2014.

WARNER, K.; FEHR, W. Mid-oleic/ultra low linolenic acid soybean oil: a healthful new alternative to hydrogenated oil for frying. **Journal of the American Oil Chemists' Society**, v. 85, n. 10, p. 945, 2008.

WHITE, P. J. Fatty acid in oil seeds (vegetable oils). In: CHOW, C. K. **Fatty acids in foods and their health implications**, CRC press, New York, p 210–263, 2007.

YANG, K. et al. Development of molecular markers for low raffinose and stachyose in korean soybean cultivars. **Plant Breeding and Biotechnology**, v. 2, n. 2, p. 151-157, 2014.

YERMANOS, D. M. et al. Oil content and composition of the seed in the world collection of sesame introductions. **Journal of the American Oil Chemists' Society**, v. 49, n. 1, p. 20-23, 1972.

ZALOGA, G. P. et al. Trans fatty acids and coronary heart disease. **Nutrition in Clinical Practice**, v. 21, n. 5, p. 505-512, 2006.

ZHANG, J. et al. Arabidopsis fatty acid desaturase FAD2 is required for salt tolerance during seed germination and early seedling growth. **PLoS One**, v. 7, n. 1, p. e30355, 2012.