

RAFAEL DELMOND BUENO

**Identificação de marcadores SNP em genes candidatos
associados ao conteúdo e qualidade do óleo de soja**

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

VIÇOSA
MINAS GERAIS – BRASIL
2012

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

B928i
2012

Bueno, Rafael Delmond, 1981-
Identificação de marcadores SNP em genes candidatos
associados ao conteúdo e qualidade do óleo de soja /
Rafael Delmond Bueno. – Viçosa, MG, 2012.
xii, 102f. : il. (algumas color.) ; 29cm.

Orientador: Maurílio Alves Moreira
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Soja - Melhoramento genético. 2. Ácidos graxos.
3. Marcadores genéticos. I. Universidade Federal de Viçosa.
Departamento de Bioquímica e Biologia Molecular. Programa
de Pós-Graduação em Bioquímica Agrícola. II. Título.

CDD 22. ed. 633.342

RAFAEL DELMOND BUENO

Identificação de marcadores SNP em genes candidatos associados ao conteúdo e qualidade do óleo de soja

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

APROVADA: 05 de novembro de 2012

Dr. Newton Deniz Piovesan

Prof. Arlindo Inês Teixeira

Prof. Everaldo Gonçalves de Barros
(Coorientador)

Prof. Pedro Ivo Vieira Good God
(Coorientador)

Prof. Maurílio Alves Moreira
(Orientador)

DEDICATÓRIA

A Deus, por sempre iluminar meu caminho
Aos meus pais Amilton e Evanir, pelo exemplo de vida e esforço para
garantir educação à mim e meu irmão
As meu irmão George, pelo companheirismo e exemplo de dedicação
A minha avó Carolina pelo carinho e por tudo mais.

Por tudo que significam para mim

AGRADECIMENTOS

Ao senhor bom Deus, sobretudo. A estrela sempre presente que guia e ilumina as trilhas do meu viver. Obrigado Senhor!

Ao sistema público de ensino (municipal, estadual e federal) por possibilitar minha formação intelectual desde a alfabetização até a pós-graduação.

À Universidade Federal de Viçosa e ao Instituto de Biotecnologia Aplicado à Agropecuária – BIOAGRO – pela oportunidade de realizar este trabalho.

Ao programa de Pós-Graduação em Bioquímica e Biologia Molecular, pela oportunidade oferecida para realização desse curso, ampliando os meus conhecimentos.

A todos os professores da UFV, especialmente aos do Departamento de Bioquímica e Biologia Molecular, que contribuíram muito para a minha formação acadêmica e pessoal.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudo.

Aos meus familiares, principalmente aos meus pais e meu irmão, que me apoiaram durante essa longa caminhada.

Ao meu professor e orientador Maurílio Alves Moreira, eterno exemplo a seguir, por ter aberto as portas da pesquisa científica, pelos ensinamentos, pela orientação, amizade, paciência, compreensão e confiança, durante todo o Doutorado.

Ao meu co-orientador, Everaldo Gonçalves de Barros, pela amizade, pelos ensinamentos, sugestões, apoio e confiança.

Ao meu co-orientador Pedro Ivo Good-God e ao pesquisador Dr. Newton Deniz Piovesan, pela amizade, orientação e acompanhamento nos experimentos.

Aos professores Arlindo Inês Teixeira, Fábio Demolinari de Miranda e ao pesquisador Dr. Klever Márcio Antunes Arruda, pela amizade, ensinamentos e ajuda na condução dos experimentos.

À Dr. Maria Fernanda Antunes Cruz, pela ajuda na condução dos experimentos.

Aos estudantes de iniciação científica Isadora Oliveira Prata e Pedro Henrique Scarpelli Pereira, pela dedicação e contribuição para a realização desse trabalho.

Ao secretário Eduardo pelo carinho, pela dedicação e boa vontade indispensáveis em todos os momentos.

Aos amigos dos laboratórios Proteína, Biomol, Sequenciamento e de Análises Bioquímicas por toda ajuda e agradável convivência.

Aos funcionários, Eduardo, Reginaldo e José Carlos, pela amizade e colaboração na condução dos experimentos.

Aos amigos de curso de pós-graduação, e amigos da UFV, pelos ensinamentos, convívio e amizade.

À minha namorada Larissa, pelo companheirismo.

À todos meus amigos de República, pela amizade e companheirismo.

A todos os meus familiares e amigos que torcem pelo meu sucesso.

BIOGRAFIA

RAFAEL DELMOND BUENO, filho de Amilton de Souza Bueno e Evanir Delmond Bueno, nasceu em 08 de junho de 1981, São Paulo, Estado de São Paulo, Brasil.

Em Dezembro de ano de 2006, formou se em Ciências Biológicas – Licenciado-pela Universidade Federal do Mato Grosso do Sul, em Três Lagoas, Mato Grosso do Sul.

Em março do ano de 2007, ingressou no Curso de Mestrado em Bioquímica Agrícola da Universidade Federal de Viçosa, em Viçosa, Minas Gerais, submetendo se à defesa de tese em fevereiro de 2009.

Em março de 2008, ingressou no Curso de Doutorado em Bioquímica Agrícola da Universidade Federal de Viçosa – UFV – em Viçosa, Minas Gerais, submetendo se à defesa de tese em novembro de 2012.

SUMÁRIO

RESUMO	ix
ABSTRACT	xi
1. Introdução geral	1
2. Objetivo geral	2
2.1. Objetivos Específicos.....	2
3. Revisão de literatura	3
3.1. Produção e importância econômica do óleo de soja	3
3.2. Composição e qualidade do óleo de soja.....	5
3.3. Biossíntese e armazenamento de triacilgliceróis (TGA) na semente	6
3.3.1. Biossíntese de ácidos graxos	8
3.3.2. Ativação do programa de maturação das sementes.....	12
3.4. Genes candidatos	15
3.5. Marcadores SNP	16
3.6. Metodologias de genotipagem	19
4. Referências bibliográficas.....	23
Capítulo I – Identificação de marcadores SNP (<i>Single Nucleotide Polymorphism</i>) em regiões gênicas regulatórias associadas ao conteúdo de óleo em sementes de soja.....	33
1. Resumo	34
2. Introdução.....	35
3. Materiais e Métodos	36
3.1. Escolha do material genético	36
3.2. Seleção dos genes candidatos e desenho de <i>primers</i>	37
3.3. Reações de amplificação e sequenciamento	41
3.4. Análise de expressão <i>in silico</i>	41
3.5. Populações	42
3.6. Extração do DNA.....	43
3.7. Determinação dos teores de óleo e proteínas	43
3.8. Análise estatística dos dados.....	43
3.8.1. Teste de normalidade.....	43

3.8.1.1. Teste de assimetria.....	43
3.8.1.2. Teste de curtose.....	44
3.8.1.3. Teste de Lilliefors.....	45
3.8.2. Análise de variância	46
3.8.3. Análise associativa	47
3.8.4. Recursos computacionais.....	48
3.9. Desenhos dos <i>primers</i> de genotipagem TSPCR.....	48
3.10. Genotipagem de SNP utilizando a metodologia TSPCR.....	49
4. Resultados e discussão	49
4.1. Amplificação de fragmentos dos genes candidatos por PCR	49
4.2. Sequenciamento e identificação de polimorfismos	51
4.3. Análise de expressão gênica em Northern Eletrônico	54
4.4. Avaliação genética e estatística das populações de mapeamento	57
4.4.1. Testes de normalidade	57
4.4.2. Análise de variância	57
4.5. Teste com os <i>primers</i> TSPCR.....	59
4.6. Genotipagem das populações de RILs e teste de segregação.....	64
4.7. Análise de associação.....	66
5. Conclusões.....	68
6. Referências bibliográficas.....	70
Capítulo II – Validação de marcadores SNP e utilização da metodologia	
de genotipagem TSPCR na discriminação de alelos de genes associados ao	
conteúdo de ácidos graxos em óleo de soja	
1. Resumo	75
2. Introdução.....	75
3. Materiais e métodos	77
3.1. Material genético	77
3.2. Seleção dos genes candidatos e desenho de <i>primers</i>	78
3.3. Reações de amplificação e sequenciamento	81
3.4. Populações de soja utilizadas.....	81
3.5. Extração do DNA	82
3.6. Determinação da composição de ácidos graxos	82

3.7. Análise estatística dos dados	82
3.8. Desenhos dos <i>primers</i> de genotipagem TSPCR	82
3.9. Genotipagem de SNP utilizando a metodologia TSPCR.....	83
4. Resultados e discussão	83
4.1. Amplificação de fragmentos dos genes candidatos por PCR	83
4.2. Sequenciamento e identificação de polimorfismos	84
4.3. Avaliação genética e estatística das populações de mapeamento	85
4.3.1. Valores médios dos teores de ácidos graxos.....	85
4.3.2. Teste de normalidade	86
4.4. Teste com os <i>primers</i> TSPCR.....	89
4.5. Genotipagem das populações e teste de segregação	92
4.6. Análise de associação.....	96
5. Conclusões.....	99
6. Referências bibliográficas.....	100

RESUMO

BUENO, Rafael Delmond, D.Sc., Universidade Federal de Viçosa, novembro de 2012. **Identificação de marcadores SNP em genes candidatos associados ao conteúdo e qualidade do óleo de soja.** Orientador: Maurílio Alves Moreira. Coorientadores: Everaldo Gonçalves de Barros e Pedro Ivo Vieira Good God.

O desenvolvimento de cultivares produtivas com incremento significativo no teor de proteína e/ou de óleo, bem como melhorar a composição de ácidos graxos na fração óleo, está entre os principais objetivos dos programas de melhoramento genético de soja. A identificação e validação de marcadores SNP (*Single Nucleotide Polymorphism*) em regiões gênicas associadas a essas características e o seu emprego na seleção assistida por marcadores moleculares (SAM), representa uma poderosa ferramenta para o melhoramento genético da soja, no sentido de aumentar a eficiência e proporcionar maiores ganhos genéticos. O objetivo deste trabalho foi identificar SNP em regiões gênicas associadas ao conteúdo e a qualidade do óleo em genótipos contrastantes do banco de germoplasma do Programa de Melhoramento da Qualidade da Soja da UFV e validar os polimorfismos como marcadores moleculares em populações segregantes e RILs. Foram selecionados 22 genes candidatos que estão potencialmente relacionados com o conteúdo e a qualidade do óleo em soja. As sequências parciais dos 22 genes candidatos ou de genes homólogos, foram obtidas a partir do GenBank. As sequências gênicas completas foram obtidas por meio do algoritmo BLASTn no banco de dados PHYTOZOME. Os *primers* desenhados para amplificar regiões específicas de cada gene candidato em genótipos contrastantes para conteúdo e qualidade do óleo de soja, produziram 33,91 kpb de sequências com boa qualidade, com valor de Phred superior ou igual a 30, sendo identificados 88 SNP e 26 INDELS. Com o intuito de obter informações sobre a expressão dos transcritos dos genes candidatos foi feito um alinhamento por meio do algoritmo BLASTn, contra bibliotecas de ESTs derivadas de sete tecidos de soja, depositados no banco de dados do NCBI. Os resultados mostraram que todos os transcritos selecionados dos genes ADPRI e G3PDH foram expressos constitutivamente, enquanto que os transcritos dos genes ABI3, ACC, AMIP, ARAF, ASIL, CLPPR, CPT, DGAT, DOF4, FUS3, LEC1A, LEC1B, LPCAT, MDH e WRI, apresentaram expressão órgão-específica. *Primers* de genotipagem TSPCR (*Temperature-Swicht Polymerase Chain Reaction*) foram utilizados para discriminar os

alelos de SNP em cinco populações (RIL e F₂). Os polimorfismos identificados por meio do sequenciamento foram inicialmente testados nos respectivos progenitores, sendo para isso, desenhados 20 conjuntos de *primers* TSPCR. Somente os *primers* TSPCR desenhados no polimorfismo dos genes AMIP-5, AMIP-3, ABI3, LEC1B, ARAF, LPCAT, PDAT, FAD3B e FAD3C mostraram-se polimórficos, sem bandas inespecíficas e/ou marcação de bandas fracas, nos respectivos progenitores. Os *primers* de genotipagem TSPCR desenhados no polimorfismo dos genes AMIP-5, AMIP-3, ABI3 e LEC1B foram amplificados em 176 progênies F₆ do cruzamento PI371611 x CD222 (contrastantes para teor de óleo). Os *primers* TSPCR desenhados no polimorfismo dos genes ARAF, LPCAT, e LEC1B foram genotipados em 236 progênies, na geração F₆, derivadas do cruzamento CD01RR8311 x SUPREMA (contrastantes para teor de óleo). Os *primers* TSPCR, desenhados para detectar o polimorfismo presente nos genes ABI3 e LEC1B foram amplificados em 170 progênies F₆, derivadas do cruzamento entre A7002 x CD219 (contrastantes para teor de óleo). Os *primers* TSPCR desenhados para discriminar os alelos de SNP identificados nos genes candidatos ARAF e PDAT, foram amplificados em 259 progênies F₆ do cruzamento FA22 x CD219 (contrastantes para teor de ácido oleico). Os *primers* TSPCR desenhados no polimorfismo dos genes FAD3B, FAD3C e ABI3 foram amplificados em 185 progênies na geração F₂ do cruzamento A29 x Tucunaré (contrastantes para teor de ácido linolênico). O polimorfismo do gene ABI3, apresentou associações significativas com teor de proteína na população F₆ (A7002 x CD219) e com a característica teor de ácido palmítico, na população F_{2,3} do cruzamento entre A29 e Tucunaré. Nessa mesma população foram detectadas associações significativas do marcador SNP presente no gene FAD3B com os teores dos ácidos graxos esteárico, oléico e linolênico e do marcador SNP do gene FAD3C com os teores de ácidos linoléico e linolênico. A metodologia de genotipagem TSPCR mostrou ser eficaz na discriminação dos alelos de SNP nas cinco populações estudadas. Associação dos marcadores SNP presentes nos genes ABI3, FAD3B e FA3C, que estão relacionados com a biossíntese dos ácidos graxos, possibilitará o emprego da SAM, com intuito de selecionar indivíduos que apresentem melhor perfil de ácidos graxos, otimizando o desenvolvimento de linhagens de soja especiais para a agroindústria.

ABSTRACT

BUENO, Rafael Delmond, D.Sc., Universidade Federal de Viçosa, November, 2012. **Identification of SNP markers in candidate genes associated with the content and quality of soybean oil.** Adviser: Maurílio Alves Moreira. Co-advisers: Everaldo Gonçalves de Barros and Pedro Ivo Vieira Good God.

The development of productive cultivars with significant increases in protein and/or oil, as well as improving the fatty acid composition of the oil fraction, are among the main objectives of soybean breeding programs. Identification and validation of SNPs (Single Nucleotide Polymorphism) in gene regions associated with these characteristics and their use in marker-assisted selection (MAS) represent a powerful tool for genetic improvement of soybean, when seeking to increase efficiency and provide greater genetic gains. The objective of this study was to identify SNPs in gene regions associated with the content and quality of oil for contrasting genotypes of the germplasm bank of the Soybean Quality Improvement Program (*Programa de Melhoramento da Qualidade da Soja*) of UFV, and validate polymorphisms as molecular markers for segregating populations and RILs. Twenty-two candidate genes were selected that are potentially related to the content and quality of soybean oil. The partial sequences of the 22 candidate genes or homologous genes were obtained from the GenBank. Complete gene sequences were obtained from the BLASTn algorithm in the PHYTOZOME database. The primers designed to amplify specific regions of each candidate gene in contrasting genotypes with respect to content and quality of soybean oil, yielded 33.91 kbp of high quality sequences, with Phred value greater than or equal to 30, where 88 were identified as SNPs and 26 as INDELS. In order to obtain information on expression of candidate gene transcripts, an alignment was performed using the BLASTn algorithm against ESTs libraries derived from seven soybean tissues deposited in the NCBI database. The results showed that all transcripts selected of the genes ADPRI and G3PDH were expressed constitutively, while transcripts of the genes ABI3, ACC, AMIP, ARAF, ASIL, CLPPR, CPT, DGAT, DOF4, FUS3, LEC1A, LEC1B, LPCAT, MDH and WRI showed organ-specific expression. Primers for TSPCR genotyping (Temperature-Switch Polymerase Chain Reaction) were used to discriminate the SNP alleles in five populations (RIL and F₂). The polymorphisms identified by sequencing were first tested in their respective progenitors, and for this purpose 20 TSPCR primer sets were designed. Only the TSPCR primers designed in

polymorphism of the genes AMIP-5, AMIP-3, ABI3, LEC1B, ARAF, LPCAT, PDAT, FAD3B and FAD3C showed polymorphism, without nonspecific bands and/or marking of weak bands in their respective progenitors. The TSPCR genotyping primers designed in the polymorphism of genes AMIP-5, AMIP-3, ABI3 and LEC1B were amplified in 176 F₆ progenies of the cross PI371611 x CD222 (contrasting oil contents). TSPCR primers designed in polymorphism of genes ARAF, LPCAT, and LEC1B were genotyped in 236 progenies in the F₆ generation, derived from the cross CD01RR8311 x SUPREMA (contrasting oil contents). The TSPCR primers designed to detect polymorphism present in genes ABI3 and LEC1B were amplified in 170 F₆ progenies derived from a cross between A7002 x CD219 (contrasting oil contents). The TSPCR primers designed to discriminate the SNP alleles identified in the candidate genes ARAF and PDAT were amplified in 259 F₆ progenies of the cross FA22 x CD219 (contrasting oleic acid content). TSPCR primers designed in the polymorphism of genes FAD3B, FAD3C and ABI3 were amplified in 185 progeny in generation F₂ of the cross A29 x Tucunaré (contrasting linolenic acid content). Polymorphism of the gene ABI3 presented significant associations with protein concentration in the population F₆ (A7002 x CD219) and with the characteristic palmitic acid content in population F_{2:3} of the cross between A29 and Tucunaré. In this same population significant associations were detected for the SNP marker present in gene FAD3B with the levels of stearic, oleic and linolenic fatty acids, and SNP marker of gene FAD3C with levels of linoleic and linolenic acids. The TSPCR genotyping methodology proved effective in discriminating SNP alleles in the five populations studied. Association of SNP markers present in genes ABI3, FAD3B and FA3C, which are related to the biosynthesis of fatty acids, allow the use of SAM in order to select individuals with the best fatty acid profile, optimizing the development of special soybean lines for agribusiness.

1. Introdução geral

A soja (*Glycine max* (L.) Merrill) é a mais importante leguminosa cultivada no mundo, devido aos seus altos conteúdos de proteína (40%) e óleo (20%), além da alta produtividade (2.823 kg ha⁻¹) e baixo custo de produção (R\$ 1.100,00 a 1.400,00 ha⁻¹). Essas excelentes características, aliadas às suas múltiplas utilidades, geram uma demanda mundial elevada e crescente, superior a 254 milhões de toneladas no ano de 2012. No Brasil, o complexo soja é um dos que mais se destacam, tanto pela sua expressiva participação na pauta de exportações sob a forma de farelo, óleo e grãos, como também para o suprimento do mercado interno de óleos comestíveis e derivados protéicos. O Brasil ocupa a posição de segundo maior produtor mundial, sendo que na safra 2011/2012, a produção brasileira de soja atingiu a marca de 65 milhões de toneladas, correspondentes a 28% da produção mundial (USDA 2012).

A soja é uma das mais importantes culturas produtoras de óleo do mundo, contribuindo com 28% da produção global de óleo no mercado internacional, os quais são destinados principalmente para o consumo humano e atualmente tem sido utilizado também para a produção de biodiesel. O óleo de soja é composto de triglicerídeos dos ácidos graxos palmítico (13%), esteárico (4%), oléico (20%), linoléico (55%) e linolênico (8%). Os elevados teores de ácido linoléico e linolênico são os principais responsáveis pela alta instabilidade oxidativa do óleo. Desta forma, a composição de ácidos graxos determina o uso final do óleo, o qual pode ser utilizado na alimentação, na produção de biodiesel ou para outros fins.

Atualmente, os programas de melhoramento da qualidade do óleo de soja têm concentrado esforços para aumentar o conteúdo de óleo em cultivares de soja, bem como, aumentar sua estabilidade oxidativa, aumentando o teor de ácido oléico e reduzindo teor dos ácidos linolênico e linoléico. Essas modificações na composição dos ácidos graxos insaturados são benéficas tanto para o uso do óleo de soja para o consumo humano quanto para o biodiesel.

A manipulação de características tais como, conteúdo e qualidade de óleo em soja, é uma tarefa bastante complexa. O controle genético envolve geralmente muitos locos e apresenta significativa interação com o ambiente. Portanto, com o intuito de tornar mais eficiente à condução dos trabalhos de melhoramento, as metodologias de uso de marcadores de DNA têm sido utilizadas como ferramentas para seleção de genótipos adequados. A seleção assistida por marcadores moleculares (SAM) pode

ajudar na obtenção de variedades com uma ou mais características em um menor espaço de tempo. Neste contexto, a identificação de QTLs (*Quantitative Trait Loci*) que influenciam tais caracteres e o seu emprego na SAM, representa uma poderosa ferramenta para o melhoramento genético da soja, no sentido de aumentar a eficiência e proporcionar maiores ganhos genéticos.

O sequenciamento do DNA permite a detecção de alterações de bases únicas no DNA, possibilitando o desenvolvimento de marcadores baseados em polimorfismo em um único par de bases (*Single Nucleotide Polymorphism* – SNP). Os marcadores SNP por contribuírem, às vezes, diretamente na expressão da característica estudada, por terem taxa de mutação relativamente baixa, serem abundantes no genoma e a genotipagem ser facilmente automatizada, podem ser empregados em um programa de melhoramento genético, na SAM, em estudos de associação genética, desequilíbrio de ligação, na clonagem posicional e na integração de mapas físicos e genético.

Considerando o grande potencial dos marcadores SNP, múltiplos esforços vêm sendo aplicados para o desenvolvimento de protocolos e procedimentos que facilitem a identificação de tais marcadores.

A identificação de novos marcadores SNP, em regiões genômicas de genes envolvidos tanto na regulação ou diretamente na biossíntese e montagem de ácidos graxos em soja, possibilitará identificar passos limitantes dessa via, bem como utilizar esses novos polimorfismos na SAM.

2. Objetivo geral

Identificar SNP (*Single Nucleotide Polymorphism*) em regiões gênicas potencialmente associadas ao conteúdo e a qualidade do óleo em genótipos contrastantes do banco de germoplasma do Programa de Melhoramento da Qualidade da Soja (PMQS-UFV). Validar os polimorfismos como marcadores moleculares em populações segregantes.

2.1. Objetivos Específicos

- Selecionar genes candidatos possivelmente associados ao conteúdo e a qualidade do óleo em soja;

- Identificar SNP nas regiões gênicas pré-selecionadas, associadas ao conteúdo e a qualidade do óleo em genótipos contrastantes do PMQS-UFV;
- Validar os marcadores SNP em indivíduos de diversas populações segregantes de soja, utilizando a metodologias de genotipagem TSPCR;
- Fenotipar as populações segregantes para as características; teores de óleo, proteína e composição de ácidos graxos;
- Associar marcadores SNP aos possíveis QTLs para conteúdo e qualidade de óleo;

3. Revisão de literatura

3.1. Produção e importância econômica do óleo de soja

O óleo de soja é o segundo no ranking mundial entre os óleos vegetais, em termos de produção e consumo doméstico, ficando atrás apenas do óleo de palma, sendo que cerca de 28% do óleo vegetal consumido no mundo, advém da soja (USDA, 2012).

A produção brasileira de óleo de soja no ano agrícola 2011/2012 foi de 7,5 milhões de toneladas, sendo que aproximadamente 5,5 milhões de toneladas foram destinadas ao consumo interno e 1,8 milhões de toneladas foram exportadas como óleo bruto (ABIOVE, 2012). Atualmente, cerca de 90% da produção de óleo vegetal é oriunda da soja. A indústria esmagadora é altamente dependente dessa cultura e, apesar do incremento de produção de outras oleaginosas, esse cenário não deverá se alterar nos próximos 15 anos (UBRABIO e FGV, 2010).

O óleo de soja é processado em três produtos: óleo refinado comestível, óleo refinado para fins não alimentares e lecitina. O óleo refinado comestível é a matéria-prima para a produção de margarinas, gordura vegetal, óleo de cozinha, maionese e tempero de salada. Além de ser utilizado na fabricação de produtos não comestíveis, como: esparadrapo, papel carbono, alguns medicamentos e explosivos. O óleo refinado para fins não alimentares é destinado à produção de velas, desinfetantes, sabões, vernizes, tintas, plásticos e lubrificantes, enquanto a lecitina é muito utilizada em confeitos, sorvetes, produtos químicos, cosméticos e têxteis (COROL, 2008). Além desses produtos básicos, recentemente o óleo de soja vem sendo utilizado para a produção de biodiesel.

A atual matriz energética mundial é composta de fontes não renováveis de carbono fóssil, como petróleo (35%), carvão (23%) e gás natural (21%) (PERES et al., 2005). Contudo, existe interesse mundial de substituir os combustíveis fósseis (não renováveis) por combustíveis renováveis. Esse interesse está diretamente ligado ao possível esgotamento das reservas mundiais de carbono fóssil nas próximas décadas e, principalmente, pelo fato da queima desses combustíveis fósseis liberarem uma grande quantidade de gases poluentes, que estão associados principalmente às mudanças climáticas. Levando em conta esses dois fatores, juntamente com as projeções de crescimento da população mundial, há estimativas de que o mercado mundial de óleo de soja expandirá 43% no período de 2006 a 2017 (FAPRI, 2007).

No ano de 2005, a lei nº 11.097/05 estabeleceu a obrigatoriedade da adição de um percentual mínimo de biodiesel ao óleo diesel comercializado, sendo 2% a partir de 2008 e 5% a partir de 2013 em qualquer parte do território nacional. Essa meta já foi antecipada e hoje já se utiliza 5% de biodiesel adicionado ao óleo diesel. Com isso, estima-se em 800 milhões de litros de biodiesel/ano a demanda no Brasil, com a mistura de 2% de biodiesel ao diesel de petróleo. Em 2008 a produção de biodiesel alcançou o volume de 1,1 bilhões de litros/ano, superando a meta para a utilização do B2. A meta para o ano de 2013 é de 2,40 bilhões de litros/ano (UBRABIO e FGV, 2010).

Nesse contexto, o Brasil se destaca pela sua grande diversidade e produtividade de grãos que podem ser utilizados na fabricação de óleos vegetais, apresentando neste sentido, uma grande abertura para uma nova alternativa energética, substituindo o diesel por biodiesel. Dentre as cinco culturas oleaginosas com maior potencial para a produção de biodiesel estão: dendê, canola, mamona, amendoim e soja. A curto prazo, a soja é a cultura que apresenta a melhor perspectiva dentre as demais oleaginosas para produção de biodiesel, devido: i) a existência de cultivares adaptadas as diferentes condições edafoclimáticas; ii) apresenta um rendimento médio de óleo bruto em torno de 0,2 – 0,6 t ha⁻¹ (num prazo de quatro a cinco meses de cultivo); além de possuir uma logística e rede industrial bem definida, voltada para o seu escoamento e processamento (BRIEU e PARENTE, 2009).

A produção de biodiesel brasileira evidencia claramente a utilização do óleo de soja como a principal matéria-prima utilizada pelas indústrias produtoras de biodiesel. Visto que 78%, dos cerca de 2,4 bilhões de litros de biodiesel produzidos em 2010 pelas usinas instaladas no Brasil, utilizaram o óleo de soja como matéria-prima. Os 20%

restantes correspondem à gordura animal (16%), óleo de algodão (3%) e a outras oleaginosas (3%) (UBRABIO e FGV, 2010).

3.2. Composição e qualidade do óleo de soja

A composição do óleo de soja varia em função do genótipo da planta e do ambiente em que ela é cultivada. Sendo o óleo de soja composto essencialmente de triacilgliceróis, os quais são constituídos por uma molécula de glicerol esterificada a três ácidos graxos, variando no tamanho e no grau de insaturação, constituindo a principal forma de armazenamento de lipídios no grão. Os principais ácidos graxos encontrados no óleo de soja são: ácido palmítico (16:0) - 13%; ácido esteárico (18:0) - 4%; ácido oléico (18:1 Δ^9) - 18%; ácido linoléico (18:2 $\Delta^9,12$) - 55% e ácido linolênico (18:3 $\Delta^9,12,15$) - 10% (GRAEF et al., 2009).

A composição e a distribuição de ácidos graxos na molécula de triacilglicerol determinam a qualidade do óleo, influenciando no seu valor nutricional, sabor, odor e propriedades físico-químicas, como estabilidade oxidativa e ponto de fusão. Determinando assim, o uso final do óleo de soja, o qual pode ser utilizado na alimentação, na produção de biodiesel ou para outros fins. Muitas vezes, a composição de ácidos graxos não é adequada para determinados fins, por isso existe considerável interesse na sua modificação por melhoramento convencional ou pelo uso de técnicas de biologia molecular.

Os elevados teores de ácido linoléico e linolênico são os principais responsáveis pela alta instabilidade oxidativa e pelo sabor e odor do óleo de soja. Isso ocorre porque os ácidos graxos polinsaturados são mais suscetíveis às hidroperoxidações, resultando em cetonas e aldeídos voláteis que alteram o sabor e odor do produto (DUNTON et al., 1951; CARDINAL et al., 2011).

Com objetivo de aumentar a estabilidade oxidativa do óleo de soja, as indústrias alimentícias tem historicamente utilizado a hidrogenação química, com objetivo de reduzir os teores dos ácidos, linoléico (18:2) e linolênico (18:3). Este processo gera quantidades significativas de ácidos graxos saturados e de isômeros trans de ácidos graxos. O consumo desses compostos parcialmente hidrogenado está diretamente relacionado com a incidência de certas doenças cardíacas, altos níveis de colesterol e riscos de desenvolvimento de diabetes tipo 2 (YADAV, 1996; HU et al., 1997; MOZAFFARIAN et al., 2006).

A hidrogenação parcial do óleo de soja, com intuito de aumentar a estabilidade oxidativa do óleo, também traz impactos negativos nas propriedades físico-químicas do biodiesel, principalmente por diminuir a fluidez do óleo no frio (GRAEF et al., 2009). DUFFIELD et al. (1998) sugerem o desenvolvimento de plantas oleaginosas que apresentem alto teor de ácido oléico (C18:1) e baixos teores de ácidos graxos saturados (palmítico-16:0 e esteárico-18:0) na fração óleo, melhorando simultaneamente a estabilidade oxidativa e aumentando a fluidez do óleo no frio.

Outra alternativa de aumentar efetivamente a estabilidade oxidativa, sem a necessidade de hidrogenação, seria aumentar geneticamente os teores de ácido oléico em detrimento dos ácidos linolênico e linoléico na semente de soja. O ácido oléico é monoinsaturado e está no centro da via de biossíntese de ácidos graxos, sendo precedido pelos ácidos graxos saturados, palmítico (16:0) e esteárico (18:0) e seguido pelos ácidos graxos poli-insaturados, linoléico (18:2) e linolênico (18:3). O aumento do teor de ácido oléico através de seleção provoca modificações em toda a cadeia. Visto que a herança para alto teor de ácido oléico (18:1) é simples e recessiva (RAHMAN et al., 1996) e tem correlações negativas e altas ($r = - 0,80$) com os teores de linoléico (18:2) e linolênico (18:3).

Com isso, o conhecimento e a manipulação dos genes envolvidos na via de biossíntese de ácidos graxos, bem como os genes que regulam essa via, oferecem inúmeras possibilidades de modificar a composição dos ácidos graxos presentes no óleo de soja.

3.3. Biossíntese e armazenamento de triacilgliceróis (TAG) na semente

A biossíntese de lipídios em vegetais superiores envolve a biossíntese de ácidos graxos, a produção dos lipídios estruturais das membranas plastidiais, a síntese de triacilgliceróis (TAG) e a produção dos lipídios estruturais das membranas extraplastidiais (MEKHEDOV et al., 2000).

A formação das sementes requer o crescimento coordenado dos tecidos de diferentes origens e, compreende duas fases principais: morfogênese e maturação. Durante a morfogênese, o embrião se desenvolve através de uma série de divisões celulares controladas, para assim, adquirir progressivamente o plano básico da planta. A fase de maturação é caracterizada pelo acúmulo dos componentes de reservas nos tecidos das sementes, aquisição de dormência e tolerância à dessecação pelo embrião

que se torna quiescente (GOLDBERG et al., 1994).

Componentes de reserva das sementes, geralmente são formados por; i) proteínas de armazenamento de sementes (SSPs), carboidratos (principalmente amido), e/ou lipídios de reserva (ceras ou TGA). Esses componentes são remobilizados para nutrir as mudas após a fase germinativa, atuando no estabelecimento dessas mudas até momento em que a fotossíntese torna-se eficiente. A proporção relativa e a localização desses componentes nos tecidos variam muito, dependendo da espécie considerada.

Os ácidos graxos presentes na molécula de TAG, não são esterificados aleatoriamente aos grupos hidroxila a cadeia de glicerol. Análises estereoquímicas dos ácidos graxos presentes nas moléculas de TAG, demonstraram que as espécies saturadas são normalmente limitadas às posições sn^{-1} e sn^{-3} do glicerol, enquanto que as espécies insaturadas ocorrem principalmente na posição sn^{-2} (MIQUEL e BROWSE, 1995).

A via de biossíntese de TAG tem sido amplamente estudada em sementes de oleaginosas (por ex; canola e soja,). Apesar de um entendimento detalhado dos percursos metabólicos e dos genes que codificam as enzimas envolvidas na biossíntese de TAG, pouco se conhece sobre o controle metabólico de acúmulo de TAG (BAUD e LEPINIEC, 2010). Esta falta de conhecimento sobre os gargalos ou pontos de controle, conduziu a uma modesta, ou muitas vezes nenhuma alteração nos conteúdos de TAG das linhagens transgênicas, projetadas para modular a expressão das chamadas enzimas reguladoras (OHLROGGE e JAWORSKI, 1997; RAMLI et al., 2002).

Estudos recentes, demonstraram que o controle da biossíntese ácidos graxos é exercida tanto pelo grupo de reações que ocorre nos plastídeos (Bloco A), como pelo grupo de montagem dos TAG, reações que ocorrem no retículo endoplasmático (RE) (Bloco B) (WESELAKE et al., 2008; RAMLI et al., 2009) (Figura 1). Dentro de cada bloco, o controle metabólico é compartilhado entre as várias etapas enzimáticas. Também é essencial considerar o metabolismo central, que fornece precursores para a produção e armazenamento de TAG (BAUD e LEPINIEC, 2010).

O metabolismo central é descrito como uma rede metabólica, ligando nutrientes fornecidos pela planta progenitora (açúcares e aminoácidos) aos precursores da síntese ácidos graxos e TAG, ou seja, o glicerol-3-fosfato (G3P) e acetil-CoA. A sacarose representa a principal forma em que carbono é assimilado fotossinteticamente, sendo transportado para dissipadores, tais como o óleo acumulado nas sementes. A hexose fosfato, gerada pela clivagem da sacarose pode ser metabolizada através das vias glicolítica e oxidativa das pentoses fosfato (OPPP) (BAUD e LEPINIEC, 2010).

Dois intermediários glicolíticos são utilizados como precursores na síntese de TAG. Primeiro, a conversão de diidroxiacetona fosfato (DHAP) para glicerol-3-fosfato (G3P), catalisada pela glicerol-3-fosfato desidrogenase (G3PDH), fornecendo a cadeia de glicerol necessária para a síntese de moléculas TAG. Segundo, a descarboxilação oxidativa do piruvato plastidial, pelo complexo piruvato desidrogenase (PDC), produz CO₂ e acetil-CoA, sendo este último usado como um precursor na síntese *de novo* ácido graxo nos plastídios (JOHNSTON et al., 1997; LIN et al., 2003).

Foi proposto por PERRY et al. (1999) que a disponibilidade de G3P restringe a taxa global de acúmulo de TAG. Em concordância com esse estudo, anos mais tarde, através de uma abordagem transgênica, baseando-se na expressão semente-específica de um gene de levedura que codifica G3PDH em canola, foi demonstrado que o nível de G3P aumentou de três a quatro vezes em sementes em desenvolvimento, resultando em um aumento de 40% no conteúdo final de óleo nas sementes (VIGEOLAS et al., 2007).

3.3.1. Biossíntese de ácidos graxos

A biossíntese de ácidos graxos em soja, inicia-se com a formação dos ácidos palmítico (C16:0), esteárico (C18:0), oléico (C18:1) e uma vez sintetizados nos plastídios, eles podem ser exportados para o citosol associados à proteína carreadora de grupos acil (ACP) e serem utilizados na produção de TGAs e lipídios estruturais de membranas extracloroplastídicas (YADAV, 1996; LOS e MURATA, 1998, BAUD e LIPINIEC, 2010).

Acredita-se que as enzimas da rota são mantidas juntas em um complexo conhecido como ácido graxo sintetase (FAS). Este complexo provavelmente permite que a série de reações ocorra com maior eficiência e catalisam a formação de ácidos graxos a partir de acetil-CoA e malonil-CoA (YADAV, 1996).

A síntese *de novo* ácidos graxos nos plastídios inicia com a formação de acetil Co-A nos plastídios a partir do piruvato (Figura 1). O acetil Co-A é ativado para malonil-CoA pela enzima acetil-CoA carboxilase. A forte regulação da acetil-CoA carboxilase parece controlar a taxa global de síntese de ácidos graxos (OHLROGGE e JAWORSKI, 1997; NIKOLAU et al., 2003). O malonil-CoA, reage com a proteína carregadora de acil (ACP, do inglês *acyl carrier protein*) gerando malonil-ACP, o substrato primário da síntese de ácidos graxos (BONAVENTURE e OHLROGGE, 2002).

Após transferência ao ACP, o malonil-ACP participa de uma reação de condensação com o acetil-CoA, resultando na formação de uma ligação carbono-carbono e liberação do CO₂, tornando a reação essencialmente irreversível (BAUD e LEPINIEC, 2010).

A primeira condensação de acetil-CoA e malonil-ACP para formar um produto de quatro carbonos é catalisada pela 3-cetoacil-ACP sintase III (KASIII). As condensações seguintes (até 16-ACP) são catalisadas pelas isoformas KASI/KASB e as isoformas KASII/KASA, promovem a condensação do palmitoil-ACP (16C) a esteroil-ACP (18C) (PIDKOWITCH et al., 2007).

Três reações adicionais ocorrem após cada condensação para a formação de um ácido graxo saturado. Essas reações são catalisadas pelas enzimas, cetoacil-ACP redutase (KAR), 3-hidroxiacil-ACP desidratase e enoil-ACP redutase (ENR) (MOU et al., 2000). A ação combinada destas quatro reações promove o alongamento do precursor de ácido graxo em dois carbonos por ciclo, formando no final os ácidos graxos C16:0, C18:0 e C18:1. Estes por sua vez, podem permanecer no plastídio para síntese dos lipídeos estruturais de membranas cloroplastidiais, como também podem ser exportados para o citosol para produção de óleo e lipídeos estruturais de membranas extracoloroplastídicas.

Após a síntese ácidos graxos (C16:0, C18:0 e C18:1) no plastídio, os grupos acil são hidrolisados pela acil-ACP thioesterases (FATA ou FATB), que liberam ácidos graxos livres (SALAS e OHLROGGE, 2002; BONAVENTURE et al., 2003). Os ácidos graxos então são ativados a ésteres CoA na membrana externa do cloroplasto, através da sintetase de acil-CoA de cadeia longa (LACS) antes da sua exportação para o ER (Figura 1).

Durante a biossíntese de (TGA) no retículo endoplasmático, a glicerol-3-fosfato aciltransferase (GPAT) transfere o grupo acetil do acetil-Coa para a posição *sn-1* do glicerol-3-fosfato produzindo ácido lisofosfatídico (LPA). Em seguida, a ácido lisofosfatídico aciltransferase (LPAAT) transfere outro grupo acil do acil-Coa para a posição *sn-2* do ácido lisofosfatídico, formando o ácido fosfatídico (PA). A defosforilação do ácido fosfatídico pela ácido fosfatídico fosfohidrolase (PAP) leva à formação do diacilglicerol (DAG). Assim, os diacilgliceróis representam o ponto de ramificação para a biossíntese de lipídios de membrana ou de reserva (BAUD e LIPINIEC, 2010).

Para formar lipídios de membrana, o DAG é reversivelmente convertido em fosfatidilcolina pela colinafosfotransferase (CPT). Então, a fosfatidilcolina (PC) atua como substrato para a biossíntese de ácidos graxos polinsaturados, os quais podem retornar ao *pool* de acil-CoA citoplasmático pela ação da lisofosfatidilcolina aciltransferase (LPCAT). Além das dessaturases microsomais, as enzimas CPT e LPCAT também atuam na biossíntese de lipídios no retículo endoplasmático. A CPT tem a função de converter reversivelmente DAG em fosfatidilcolina para a formação de lipídios de membrana e a LPCAT tem a função de repor o estoque de ácidos graxos polinsaturados do *pool* de acil-CoA no citoplasma (YADAV, 1996) (Figura 1).

Atualmente são descritos dois genes candidatos que codificam as ω -6-dessaturases microsomais em soja, denominados FAD2-1 e FAD2-2. Enquanto para a ω -3-dessaturase foram identificados três genes candidatos, denominados *GmFAD3A*, *GmFAD3B* e *GmFAD3C*. Já a ω -9-dessaturase é uma enzima solúvel presente no plastídio responsável pela conversão do estearoil-ACP em oleil-ACP. Na cultivar de soja Dare, foi descrita a presença de dois genes candidatos para a ω -9-dessaturase, designados de *SACPD-A* e *SACPD-B*. BYFIELD (2006) avaliou o nível de expressão desses genes por meio de qPCR em 51 genótipos de soja e verificou que todos apresentavam ambos os genes, *SACPD-A* e *SACPD-B*.

Por fim, o DAG pode sofrer acilação final na posição 3 da molécula do glicerol e ser convertido a triacilglicerol (TAG), sendo esta reação catalisada pela diacilglicerol aciltransferase (DAGAT) (YADAV, 1996). Os TAG uma vez sintetizados são armazenados em estruturas subcelulares denominadas de corpos lipídicos. Dependendo da espécie considerada, a abundância e localização destes corpos lipídicos em tecidos de sementes podem variar muito. Em soja e girassol, grandes quantidades de óleo são armazenados predominantemente em tecidos embrionários (por exemplo, cotilédones),

ocupando a maior parte do espaço disponível entre os tegumentos da semente. Estudos com *Arabidopsis thaliana*, demonstraram que os cotilédones contêm cerca de 60% dos ácidos graxos totais; radícula e hipocótilo 30% e os 10% restantes, estão localizados no endosperma (PENFIELD et al., 2004; LI et al., 2006).

3.3.2. Ativação do programa de maturação das sementes

O acúmulo de óleo é parte do processo de maturação das sementes, um programa de desenvolvimento altamente controlado, o qual é definido nos tecidos de embriões, uma vez que a morfogênese é alcançada. Assim, estudos realizados em sementes em desenvolvimento e/ou embriões têm demonstrado que as vias de biossíntese dos ácidos graxos e TAGs, são fortemente reguladas a nível transcricional (FAWCETT et al., 1994; O'HARA et al., 2002; BAUD e GRAHAM, 2006).

A regulação coordenada dos genes que codificam enzimas da via biossintética de ácidos graxos, observados no início da fase de maturação das sementes, sugere, que a via pode ser submetida a um sistema de regulação transcricional global (Figura 2), (BAUD e LEPINIEC, 2010).

Estudos genéticos em milho e *A. thaliana* demonstraram uma rede complexa de fatores de transcrição, que controlam os programas de expressão gênica, os quais são essenciais para que ocorra a maturação das sementes. Um vasto conjunto de ativadores e repressores da transcrição regulam a transição, entre as fases de desenvolvimento vegetativo e maturação das sementes (SANTOS et al., 2008).

Eventos de remodelamento da cromatina juntamente com a regulação transcricional, favorecem a atividade dos ativadores do programa de maturação das sementes em desenvolvimento, enquanto que a atividade dos repressores torna-se predominante no final do processo de maturação (SUZUKI E MCCARTY, 2008; ZHANG e OGAS, 2009).

Ativação dos genes chaves em *A. thaliana*, também chamados de reguladores mestres do processo de maturação, inclui membros da superfamília do domínio B3, de proteínas ligadoras do DNA planta-específica, ou seja, ABI3, LEC2, e FUS3 (Figura 2). Os membros dessa rede AFL (ABI3/FUS3/LEC2), atuam em conjunto com LEC1, um fator de transcrição do tipo CAAT *binding factor* (CBF), e exibem padrões de expressão parcialmente sobrepostos, que incluem regulação cruzada e feedback complexo

(LOTAN et al., 1998).

Ao analisar a expressão de ABI3 e FUS3 em vários mutantes, foi demonstrado, que a principal função de LEC2 é “*up-regulated*” a expressão de FUS3 e ABI3. Uma mutação em LEC2 foi responsável por uma diminuição drástica na expressão dos fenótipos ABI3 e FUS3, porém, LEC2 pode ser resgatado pela expressão constitutiva de ABI3 ou FUS3 (TO et al., 2006). Um conjunto de técnicas complementares permitiu isolar alguns dos alvos diretos desses reguladores em *A. thaliana*. Por exemplo, foi demonstrado que LEC2 liga-se aos promotores dos genes que codificam oleosina e HSD1(*hidroxiesteróide desidrogenase*) (BRAYBROOK et al., 2006; BAUD et al., 2009). A expressão de HSD1 é fortemente e especificamente induzida em tecidos que acumulam óleo (BAUD et al., 2009a).

Outro elemento importante na regulação de acúmulo de óleo em semente de *A. thaliana* é o fator de transcrição WRINKLED1 (WRI1) (BAUD et al., 2009). WRI1 codifica um fator de transcrição da família AP2-EREBP (APETALA2-ethylene responsive element-binding protein) (CERNAC e BENNING, 2004), especificando a ação regulatória de LEC2 e, possivelmente de LEC1 para a rede de biossíntese de ácido graxo (BAUD et al., 2010). Foi demonstrado recentemente, que WRI1 interage com as sequencias promotoras dos genes PKP-β1, BCCP2 e KAS1 *in vitro*, genes que codificam enzimas da via glicolítica e de biossíntese de ácidos graxos (MAEO et al., 2009; BAUD et al., 2009b).

Por outro lado, um conjunto de três fatores intimamente relacionado ao domínio B3 (HIS, VP1 e VAL), define um “clade irmão” para os fatores do domínio B3 AFL, que desliga o programa de maturação da semente, antes da germinação, reprimindo os genes AFL e/ou genes alvo da rede AFL, através de ligação competitiva nos motivos RY (TSUKAGOSHI et al., 2007; SUZUKI et al., 2008; GUERRIERO et al., 2009). Outro fator transcricional que contribui para desligar o programa de maturação da semente é o repressor transcricional trihelix ASIL1, que compete com os ativadores da rede AFL, pela ligação aos elementos GT (GTGATT), estando estreitamente próximo ao motivo repetido RY do seu promotor alvo (GAO et al., 2009).

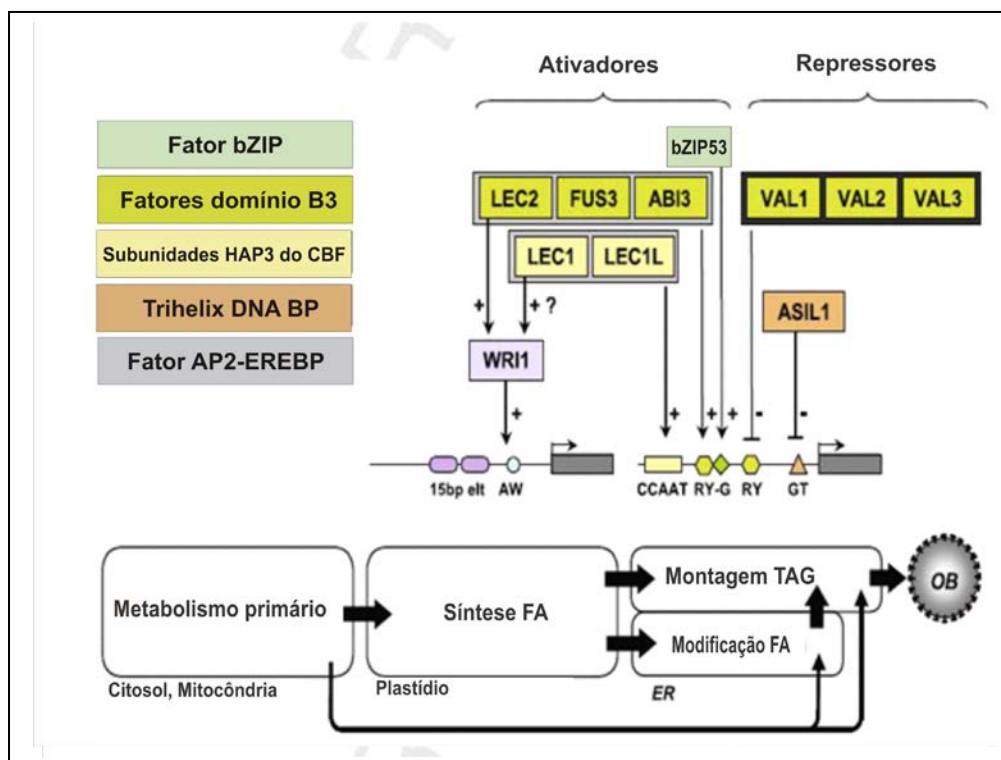


Figura 2. Modelo para a regulação transcricional da biossíntese de triacilglicerol durante a fase de maturação das sementes de *Arabidopsis thaliana*. Fatores que podem induzir ou reprimir a lipogênese estão nas caixas abertas. Genes alvos codificando os atores da rede biossintéticas de ácido graxos e triacilgliceróis são apresentados abaixo dos reguladores transcisionais, juntamente com os elementos cis-regulatórios, reconhecidos por esses reguladores. ASIL1, ARABIDOPSIS 6B-INTERACTING PROTEIN 1-LIKE1; AP2-ERE BP, APETALA2-ethylene responsive element-binding protein; CBF, fator de ligação CAAT-box; ER, retículo endoplasmático, FA, ácidos graxos, OB, corpos lipídicos; FUS3, FUSCA3; LEC1, 2, LEAFY COTYLEDON1, 2; LEC1L, LEAFY COTYLEDON1-LIKE; TAG, triacilglicerol; Trihelix DNA BP, trihelix DNA binding protein; VAL1, 2, 3, VP1/ABSCISIC ACID INSENSITIVE3-LIKE1, 2, 3; WRI1, WRINKLED1. Adaptado de BAUD e LIPINIEC (2010).

Os conhecimentos adquiridos, nos últimos anos, sobre a regulação dos genes envolvidos na biossíntese de ácidos graxos em plantas serão úteis, pelo fato de até agora, as contribuições para melhorar o conteúdo de óleo em sementes terem se concentrado principalmente no melhoramento convencional (WESELAKE et al., 2009).

É importante mencionar, que aumentar o teor e/ou melhorar a qualidade do óleo em sementes de soja é tarefa bastante complicada, visto que são características quantitativas, onde inúmeros fatores contribuem para essas características, de modo que estratégias complexas para adquirir o fenótipo desejável, podem ser necessárias. Naturalmente, essas estratégias vão explorar os conhecimentos adquiridos até agora sobre as redes metabólicas envolvidas e dos vários fatores que regulam biossíntese de óleo em sementes de soja (BAUD e LEPINIEC, 2010). Uma estratégia que depende diretamente desses conhecimentos, é a abordagem baseada em genes candidatos.

3.4. Genes candidatos

Genes candidatos são genes de ação biológica conhecida e que estão envolvidos com o desenvolvimento ou a fisiologia de uma característica de interesse econômico. Estes genes podem ser estruturais, regulatórios ou que participam de vias metabólicas, afetando a expressão do caráter (BRYNE e MCMULLEN, 1996).

Os estudos baseados em genes candidatos buscam indentificar polimorfismos diretamente nas regiões codificantes de um dado gene alvo ou em elementos regulatórios flanqueadores a essas regiões (FITZSIMMONS et al., 1998; LIEFERS et al., 2002; ALMEIDA, 2003). Esses estudos envolvem uma série de etapas: i) escolha do gene candidato, ii) desenho de *primers* para amplificação da região de interesse do gene candidato, iii) descoberta de polimorfismos, iv) escolha e estabelecimento da técnica de genotipagem dos sítios polimórficos, v) formação de uma população referência para genotipagem e, vi) realização dos testes de associação com características fenotípicas da população.

Dentre as vantagens da análise de genes candidatos estão: i) a obtenção de marcadores no próprio gene de interesse; ii) genes candidatos podem gerar marcadores eficientes e facilitar a obtenção de variedades em um menor espaço de tempo; iii) clonagem de genes de importância agrônômica; iv) o maior poder estatístico obtido com menor número de famílias e indivíduos; v) o baixo custo dos procedimentos de genotipagem em larga escala; vi) a simplicidade operacional, pois se pode trabalhar com um único gene e não com um genoma por inteiro; vii) e finalmente, o polimorfismo do gene candidato se validado, pode ser empregado na seleção assistida por marcadores (SAM).

As limitações dessa abordagem estão; no pequeno número de genes conhecidos que controlam características de interesse, o efeito pleiotrópico e de epistasia de outros genes sobre o gene candidato, o alto custo da etapa inicial de detecção de polimorfismos e a dificuldade no estabelecimento definitivo do efeito do gene candidato (ROTHSCHILD e SOLLER, 1999). Além destes, o efeito de interação entre genótipo e ambiente também exerce grande influência. Até que se estabeleça a variante causal do gene responsável pelo seu efeito quantitativo, sempre haverá a possibilidade de que o gene estudado não seja o gene que realmente esteja causando a diferença na característica, mas esteja somente ligado ao QTLs (ROTHSCHILD e SOLLER, 1999). Isso pode levar a resultados contraditórios, dependendo da população estudada, pois

associações significativas em uma população podem não ser verificadas em outra. Por exemplo, em estudos realizados por LINVILLE et al. (2001) não houve associação significativa dos marcadores dos receptores do estrogênio (ESR) e da prolactina (PRLR), com o tamanho da leitegada em suas populações de suínos, enquanto que associação significativa foi observada anteriormente nas populações utilizadas por SHORT et al. (1997) e VINCENT et al. (1998).

3.5. Marcadores SNP

O desenvolvimento de marcadores moleculares tem possibilitado utilizar a variabilidade genética para identificar genes responsáveis pelo controle de características de importância econômica. Os marcadores moleculares são categorizados em três tipos; i) os morfológicos (SAX, 1923), ii) os bioquímicos e/ou isoenzimáticos (MARKET e MOLLER, 1959) e, iii) com a descoberta da reação em cadeia da polimerase (*Polymerase Chain Reaction* – PCR) por MULLIS e FALONA (1987), foi possível o surgimento de uma nova geração de marcadores moleculares, os baseados em sequências de DNA.

Os marcadores baseados em sequências de DNA são diferenciados em dois tipos: marcadores que utilizam o método de hibridização, por exemplo; RFLP (*Restriction Fragment Length Polymorphism*) (BOTSTEIN et al., 1980), e aqueles que utilizam a técnica de PCR, como os marcadores; RAPD - *Random Amplified Polymorphic DNA*, (WILLIAMS et al., 1990; WELSH e MACLELLAND, 1990), microssatélites ou SSR – *Simple Sequence Repeats* – (LITT e LUTY, 1989; WEBER e MAY, 1989) e AFLP – *Amplified Fragment Length Polymorphism* (VOS et al., 1995).

A disponibilidade de sequências gênicas tem aumentado muito nos últimos anos, principalmente com o término do projeto genoma de diversas espécies. Com o fim desses projetos, foi possível a detecção de um grande número de alterações de bases únicas no DNA. Essas alterações possibilitou o desenvolvimento de uma nova classe de marcadores moleculares, os baseados em polimorfismo em um único par de bases (SNP - *Single Nucleotide Polymorphism*).

Os SNP, são variações de nucleotídeos ou pequenas inserções/deleções (INDELs) nas sequências de bases em fragmentos homólogos de DNA (KWOK, 1996; KRUGLYAK, 1997; COLLINS, 1998), e são consideradas um dos mais valiosos marcadores para o mapeamento genético, geração de mapas de alta resolução e estudos

de associação (ZHU e SALMERON, 2007). Para que uma dada variação seja considerada um SNP é necessário que o seu alelo tenha uma frequência mínima de 1% na população testada (NEI e LI, 1979; BROOKES, 1999; VIGNAL et al., 2002).

Basicamente, os SNP compreendem duas categorias: transições e transversões. Nas transições, uma base purina é substituída por outra base purina ($A \leftrightarrow G$) ou uma base pirimidina é substituída por outra base pirimidina ($C \leftrightarrow T$). As transversões consistem na mudança de bases purina para pirimidina e vice-versa ($A \leftrightarrow C$, $A \leftrightarrow T$, $G \leftrightarrow T$ e $G \leftrightarrow C$).

Os SNP se tornam muito mais atrativos ao se considerar que muitos deles podem estar localizados no interior de sequências gênicas expressas (éxons) e este fato pode significar a redução de tempo e custos na obtenção de genes de interesse, se comparados aos marcadores atualmente disponíveis. Porém, existe uma maior tendência para que haja mais SNP nas regiões de íntrons, quando comparados com as regiões de éxons, onde tendem a serem mais conservadas (LIAO e LEE, 2010). De acordo com EMARA e KIM (2003), os SNP podem ser categorizados em dois tipos quando localizados em sequências gênicas expressas: não sinônima, ou seja, aquelas que resultam na mudança de aminoácidos (Figura 4), ou sinônimas que não resultam em mudança de aminoácido (Figura 3).

As mudanças não sinônimas podem determinar o fenótipo alternativo e mostrará 100% de associação com a característica e dessa forma será de grande importância para a seleção assistida por marcadores (MAS) ou para o isolamento do gene. Conforme CHEN et al. (2007), do total de SNP encontrados, 1 a 2% dos SNP estão localizados em regiões codificadoras de modo que os SNP não sinônimos podem ser considerados como candidatos para mudanças de funções gênicas.

Ao contrário, os SNP sinônimos são aqueles que têm efeito pequeno na expressão gênica, entretanto tanto, os SNP sinônimos quanto os não sinônimos são de grande importância como marcadores genéticos nos estudos de mapeamento. Mesmo que a identificação não seja no próprio gene de interesse e sim próxima a ele, ainda assim pode ser usado em análises de associação.



Indivíduo 1	GCA AGA GAT AAT TGT ... Ala Arg Asp Asn Cys	Proteína  Ativa
Indivíduo 2 Mudança no códon, sem alterar aminoácido	GCG AGA GAT AAT TGT ... Ala Arg Asp Asn Cys	Proteína  Ativa

Figura 3. SNP do tipo sinônima, a troca de uma adenina por guanina não acarretou na mudança de aminoácido. Este tipo de mutação também é conhecida por mutação silenciosa. Neste exemplo, a troca nucleotídica é do tipo transição, ou seja, uma base púrica é substituída por outra base púrica (A↔G).



Indivíduo 1	GCA AGA GAT AAT TGT ... Ala Arg Asp Asn Cys	Proteína  Ativa
Indivíduo 2 Mudança no códon e Mudança no aminoácido	GCG ACA GAT AAT TGT ... Ala Thr Asp Asn Cys	Proteína  Inativa

Figura 4. SNP do tipo não-sinônima, ou seja, a troca de uma guanina por citosina acarretou na mudança de aminoácido. Neste exemplo, a troca nucleotídica é do tipo transversões, ou seja, uma base púrica é substituída por uma base pirimídica (G↔C).

Em soja, já foram identificados mais de 30 mil SNP (HYTEN et al., 2010a,b; WU et al., 2010). Segundo CHOI et al. (2007), dois importantes fatores a respeito do genoma de uma espécie devem ser considerados para a identificação de SNP, a magnitude da variabilidade genética e a presença de regiões duplicadas no genoma. Para a cultura da soja, considerando o processo de seleção ao longo do seu melhoramento e a estreita base genética das cultivares utilizadas para a obtenção das primeiras variedades comerciais, uma baixa frequência de sequências variantes é esperada.

Os SNP são muito mais abundantes do que outros marcadores moleculares, como os microssatélites nos genomas de humanos, animais e também de plantas. Enquanto que no genoma da soja foram encontrados pouco mais de 1000 microssatélites, estima-se que existam de 4 a 5 milhões de SNP, baseando-se na taxa de 280 SNP observados nos 76,3 Kpb estudados (ZHU et al. 2003) e na frequência habitual de SNP relatados para genomas de plantas, que é de cerca de 1 SNP a cada 100-300 pb (GUPTA et al., 2001).

Na prática, SNP são marcadores bialélicos e codominantes, de forma que o conteúdo informativo em um único SNP é limitado em comparação com os marcadores SSR que são multialélicos (GRIFFITHS e SMITH, 2000; GUPTA et al., 2001; SCHLOTTERER et al., 2004; ORAGUZIE et al., 2007). Entretanto, um mapa genético de alta densidade com SNP bem espaçados, embora bialélico ofereça várias vantagens sobre mapas com baixa densidade de microssatélites, incluindo poder superior para detectar ligação (GODDARD e WIJSMAN, 2002; MATISE et al., 2003; MIDDLETON et al., 2004; EVANS e CARDON, 2004). Isto se deve ao fato de que a densidade de marcadores SNP pode ser mensurada na escala de kilobase, enquanto que os mapas atuais dispõem de marcadores em escala de megabase para a sua determinação. Além dos SNP serem a mais abundante forma de variação genética encontrada em genomas (ZHU et al., 2003), os SNP possui outra vantagem quando comparado com outros marcadores, eles possuem facilidade de automatização, viabilizada com extensivas tecnologias de genotipagem que tem sido desenvolvidas atualmente (FAN et al., 2006).

Por essas características, os SNP são uma rica fonte de polimorfismo em soja e podem ser utilizados; para saturar mapas genéticos, na detecção de QTLs, na estimação da diversidade genética, em análises de associação, na clonagem posicional, na seleção assistida por marcadores (MAS) e também são potencialmente úteis para a identificação de genes candidatos.

3.6. Metodologias de genotipagem

A maioria dos métodos de genotipagem de SNP, com maior potencial para realizar multiplex, é baseada em um dos quatro mecanismos moleculares: hibridização alelo-específico, extensão de *primer*, ligação de oligonucleotídeos e clivagem enzimática (KIM e MISRA, 2007).

O método mais básico para genotipagem de SNP em regiões específicas do genoma é baseado na técnica de PCR-RFLP (*Restriction Fragment Length Polymorphism*) (MAEDA et al. 1989). O fragmento contendo o SNP de interesse é amplificado com iniciadores específicos, purificado e submetido a um tratamento com uma enzima de restrição que reconheça apenas um dos alelos. Posteriormente, os fragmentos são separados por eletroforese em gel de poliacrilamida ou agarose para diferenciação dos alelos por tamanho. Além de ser extremamente trabalhoso, para o emprego desse método, é necessário que haja uma enzima de restrição que diferencie os

dois alelos, o que nem sempre é possível. Essa metodologia não permite a paralelização de ensaios – geralmente, só é possível genotipar um loco por ensaio. Por outro lado, é uma técnica que não requer equipamentos avançados: apenas um termociclador e equipamentos de eletroforese (CAETANO, 2009).

Posteriormente, surgiram novas técnicas de identificação sem a necessidade do sequenciamento de todo o genoma. Entre as tecnologias atualmente disponíveis para genotipar SNP estão; a TaqMan (LIVAK, 1999), ARMS-PCR (Amplification Refractory Mutation Systems) (YE et al., 2001), HRM (High Resolution Melting) (ERALI e WITTWER, 2010) e entre outras.

Dentre outras metodologias, a extensão de *primer* associada à espectrometria de massa (MALDI-TOF), é uma técnica adequada para estudos com grandes grupos de amostras, em termos de automação e custos. A interpretação dos dados é automatizada e vem como parte do *software* da maioria dos equipamentos (SEMAGN et al., 2006). Essa técnica consiste na extensão do *primer* com uma única base utilizando didesoxinucleotídeos para terminação da cadeia a partir de um *primer* com terminal 3' que se anela imediatamente upstream ao sítio polimórfico (minisequenciamento). A vantagem da análise por MALDI-TOF é a exatidão das massas das reações de genotipagem e o potencial multiplex.

O primeiro ensaio descrito para genotipar SNP por MALDI-TOF, Pinpoint, baseia-se na extensão do *primer* na presença dos quatro ddNTPs e a base no sítio polimórfico é identificada pela massa adicionada ao *primer*. As desvantagens desse método é que ele se baseia na adição de didesoxinucleotídeos, as diferenças de massa são pequenas, variam de 9 Da (T e A) a 40 Da (C e G) e os passos de dessalinização e concentração podem interferir na análise de amostras com pequenas diferenças de massa. Por essas razões o ensaio pontual foi substituído por ensaio “não pontual” que resulta em produtos com diferenças maiores de massa. A técnica modificada Pinpoint utiliza ddNTPs modificados com um tag de massa para extensão do *primer*, o que aumenta a diferença de massa entre produtos dos alelos SNP e a acurácia da genotipagem (BLONDAL et al., 2003).

BLONDAL et al, (2003) propôs uma outra metodologia de detecção de SNP por MALDI-TOF, (*Nucleotide depletion genotyping assay*), a qual utiliza uma única DNA polimerase e não requer uso de dNTPs ou ddNTP modificados. Na reação de minisequenciamento utiliza-se uma mistura de dNTPs com um dos nucleotídeos do sítio polimórfico ausente, fazendo com que a DNA polimerase se dissocia da fita de DNA na

posição em que é requerido o nucleotídeo depletado. São desenhados dois *primers*, sendo o primer informativo desenhado com o terminal 3' uma base upstream do sítio SNP para permitir a diferenciação entre o *primer* não estendido que se refere ao produto de extensão menor. Entre as principais vantagens dessa metodologia estão a acurácia, a possibilidade de se avaliar múltiplas amostras simultaneamente e a redução dos custos das análises.

Um método recentemente relatado para o desenvolvimento de marcadores de PCR alelo-específico para rápida genotipagem de SNP codominantes, é a metodologia TSPCR (*Temperature-Switch Polymerase Chain Reaction*), foi descrita por HAYDEN et al. (2009), num trabalho de genotipagem de SNP em trigo. TSPCR é um método novo para genotipagem rápida e que necessita de baixa tecnologia. Ensaios de marcador TSPCR são realizadas sob condições de reações padronizadas e os produtos da genotipagem podem ser detectados por eletroforese em gel, tornando-os fáceis de implementação. TSPCR pode fornecer uma fonte cada vez maior de SNP informativos para as culturas, podendo ser utilizados pelos programas de melhoramento genético na SAM. Cada marcador TSP compreende um conjunto de três *primers*: um par de *primers* locus específico (R/F), com temperatura de anelamento (T_a) entre 60 e 65 °C, visando amplificar fragmentos de 400 pb, e um *primer* informativo, com seu terminal 3' complementar a um dos alelos presente no locus, com T_a entre 43 e 48 °C, ao qual é adicionado, no terminal 5', uma cauda de 2 a 3 nucleotídeos aleatórios e não complementares à sequência do DNA molde, para elevar o seu valor de T_a para 53 °C. Cada primer informativo deve formar produtos de amplificação com o tamanho de aproximadamente 100 pb com um dos dois *primers* locus específico (Figura 5). Os produtos de PCR são posteriormente separados eletroforese em gel de agarose 2%. Neste sentido, em um ensaio de genotipagem TSPCR utilizando uma população segregante, os indivíduos podem apresentar uma banda de 400 pb se possuírem um alelo alternativo, ou uma banda de 100 pb se possuírem o outro alelo (alelo específico), enquanto que os indivíduos heterozigotos apresentaram as 2 bandas (Figura 6).

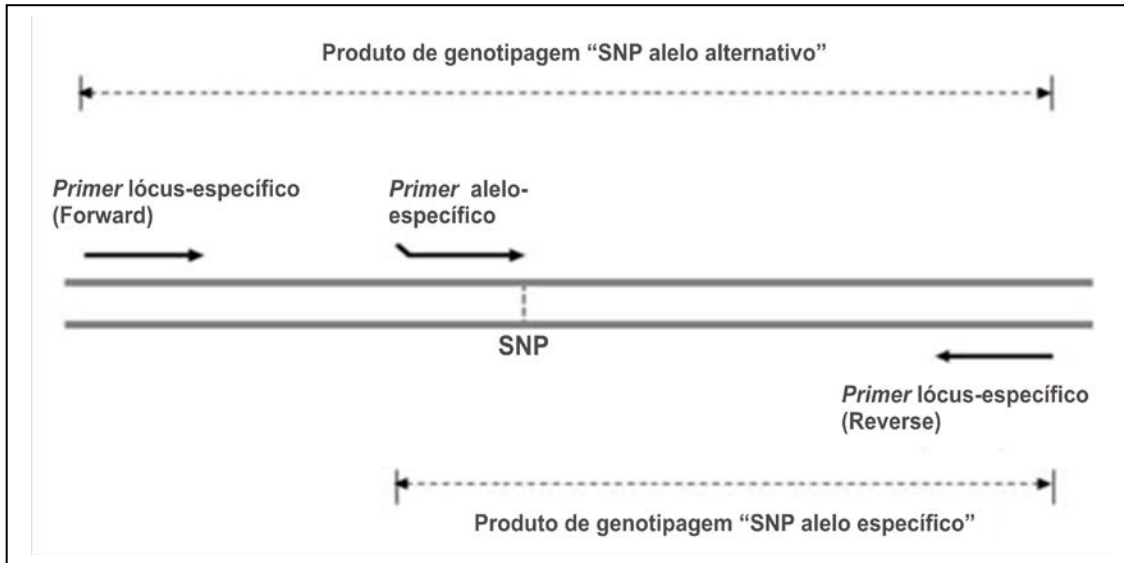


Figura 5. Ilustrando a posição e orientação dos primers de PCR do ensaio de marcador TSPCR alelo-específico, em relação ao DNA. Adaptado de HAYDEN et al. (2009).

As metodologias de genotipagem somente citadas e as descritas acima podem ser implementadas pelos programas de melhoramento genético. Essas metodologias devem ser escolhidas de acordo com custo, precisão, agilidade, disponibilidade de equipamentos, dificuldade de desenvolvimento dos ensaios e potencial para detecções de ensaios em *multiplex* (RAFALSKI, 2002).

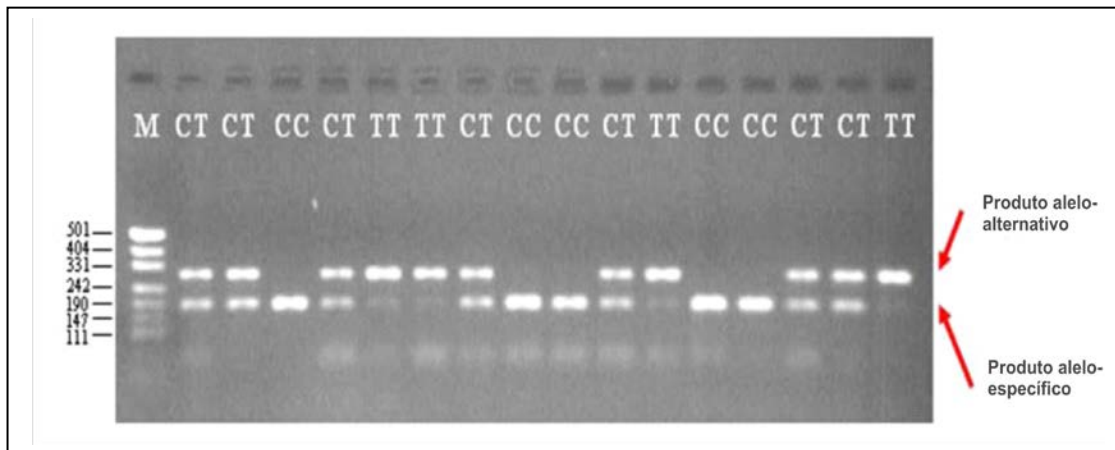


Figura 6. Gel de agarose 2%, mostrando o produto de amplificação da metodologia TSPCR. M (marcador de pares de base); CC (produto alelo-específico); TT (produto alelo alternativo) e CT (heterozigoto Adaptado de HAYDEN et al. (2009).

4. Referências Bibliográficas

ABIOVE - Associação Brasileira das Indústrias de Óleos Vegetais: Disponível em <<http://www.abiove.com.br/>>. **Acessado em**: setembro de 2012.

ALMEIDA, S. E. M. Marcadores moleculares em sintenia com os genes da leptina e de seu receptor em bovinos: associação com parâmetros produtivos. **Tese de doutorado**. Genética e biologia molecular. Universidade Federal do Rio Grande do Sul, 2003.

BAUD S.; LEPINIEC L. Physiological and developmental regulation of seed oil production. **Progress in Lipid Research**, 2010.

BAUD, S.; GRAHAM, I.A. A spatiotemporal analysis of enzymatic activities associated with carbon metabolism in wild-type and mutant embryos of *Arabidopsis* using in situ histochemistry. **The Plant Journal**; v.46, p.155–69, 2006.

BAUD, S.; REINHARD, D.N.; KELEMEN, Z.; D'ANDRÉA, S.; TO, A.; BERGER, N. Regulation of HSD1 in seeds of *Arabidopsis thaliana*. **Plant Cell Physiol**; v.50, p.1463–78, 2009a.

BAUD, S.; WUILLÈME, S.; TO, A.; ROCHAT, C.; LEPINIEC, L. Role of WRINKLED1 in the transcriptional regulation of glycolytic and fatty acid biosynthetic genes in *Arabidopsis*. **The Plant Journal**; v.60, p.933–47, 2009b.

BLONDAL, T.; WAAGE, B.G.; SMARASON, S.V.; JONSSON, F.; FJALLDAL, S.B.; STEFANSSON, K.; GULCHER, J.; SMITH, A.V. A novel MALDI-TOF based methodology for genotyping single nucleotide polymorphisms. **Nucleic Acids Research**; v.31, n.24, p.1-7, 2003.

BONAVENTURE, G.; OHLROGGE, J.B. Differential regulation of mRNA levels of acyl carrier protein isoforms in *Arabidopsis*. **Plant Physiology**; v.128, p.223–35, 2002.

BONAVENTURE, G.; SALAS, J.J.; POLLARD, M.R.; OHLROGGE, J.B. Disruption of the FATB gene in *Arabidopsis* demonstrates an essential role of saturated fatty acids in plant growth. **The Plant Cell**; v.15, p.1020–33, 2003.

BOTSTEIN, D.; WHITE, R.L.; SKOLNICK, M. Construction of genetic linkage map in man using restriction fragment length polymorphisms. **American Journal of Human Genetics**; v.32 (1), p.314-331, 1980.

BRAYBROOK, S.A.; STONE, S.L.; PARK, S.; BUI, A.Q.; LE, B.H.; FISCHER, R.L. Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. **Proceedings of the National Academy Sciences USA**; v.103, p.3468–73, 2006.

BRIEU, T. P.; PARENTE, V. Programa nacional de produção de biodiesel: um balanço da primeira fase até 2008. **Revista Biodiesel-BR**, Curitiba, v. 2, n. 12, p. 62 – 64, ago./set. 2009.

BROOKES A.J. The essence of SNPs. **Gene**; v.234, p.177–186, 1999.

BRYNE, P.F.; MCMULLEN, M.D. Defining genes for agricultural traits: QTL analysis and the candidate gene approach. **Probe**, v.7, p.24-27, 1996.

BYFIELD, G. E.; UPCHURCH, R. G. Effect of temperature on delta-9 stearyl-acyl and 51 microsomal omega-6 desaturase gene expression and fatty acid content in developing soybean seeds. **Crop Science**, v. 47, p. 1698-1704, 2007.

CAETANO, A.R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**; v.38, p.64-71, 2009.

CARDINAL, A.; BURTON, J. W.; CAMACHO-ROGER, A. M.; WHETTEN, R.; CHAPPELL, A. S.; BILYEU, K. D.; AUCLAIR, J.; DEWEY, R.E. Molecular analysis of *GmFAD3A* in two soybean populations segregating for the *fan*, *fap1*, and *fapnc* loci. **Crop Science**, v.51, 2011.

CHEN, K.; BAXTER, T.; MUIR, W.M. Genetic Resources, Genome Mapping and evolutionary Genomics of the Pig (*Sus scrofa*). **International Journal of Biological Sciences**. Sci. 3; 2007.

CHOI, I.Y.; HYTEN, D.L.; MATUKUMALLI, L.K.; SONG, Q.; CHAKY, J.M.; QUIGLEY, C.V.; CHASE, K.; LARK, K.G.; REITER, R.S.; YOON, M.S.; HWANG, E.Y.; YI, S.I.; YOUNG, N.D.; SHOEMAKER, R.C.; VANTASSELL, C.P.; SPECHT, J.E.; CREGAN, P.B. A soybean transcript map: gene distribution, haplotype and SNP analysis. **Genetics**. v.176, p.685–696, 2007.

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

COROL – Cooperativa Agroindustrial. Disponível em: <<http://www.corol.com.br/curi2.htm>>. Acesso em: fevereiro de 2008.

DUFFIELD, J.; SHAPOURI, H.; GRABOSKI, M.; MCCORMICK, R.; Wilson, R. US Biodiesel Development: New Markets for Conventional and Genetically Modified Agricultural Products. Washington, DC, USA: **Economic Research Service/United States Department of Agriculture (USDA)**. 1998.

DUNTON, H. J.; LANCASTER, C. R.; EVANS, C. D.; COWAN, J. C. The flavor problem of soybean oil VII. Linolenic acid. **Journal of the American Oil Chemists Society**; v.28, p.115-118, 1951.

- EMARA, M.G.; KIM, H. Genetic Markers and their Application in Poultry Breeding. **Poultry Science**; v.82, p.952–957, 2003.
- ERALI, M.; WITTEWER, C.T. High resolution melting analysis for gene scanning. **Methods**; v.50, p. 250–261, 2010.
- EVANS, D.M.; CARDON, L.R. Guidelines for Genotyping in Genome wide Linkage Studies: Single Nucleotide Polymorphism Maps Versus Microsatellite Maps. **American Journal of Human. Genetics**; v.75, p.687–692, 2004.
- FAN, J.B.; CHEE, M.S.; GUNDERSON, K.L. Highly parallel genomic assays. **Nature Reviews, Genetics**; v.7, p.632–644, 2006.
- FAPRI - Food and Agricultural Policy Research Institute. World Agricultural Outlook. Disponível em: <[http:// www.fapri.iastate.edu](http://www.fapri.iastate.edu)>. Acesso em: abril 2007.
- FAWCETT, T.; SIMON, W.J.; SWINHOE, R.; SHANKLIN, J.; NISHIDA, I.; CHRISTIE, W.W. Expression of mRNA and steady-state levels of protein isoforms of enoyl-ACP reductases from *Brassica napus*. **Plant Molecular Biology**; v.26, p.155–63, 1994.
- FITZSIMMONS, C. J.; SCHMUTZ, S. M.; BERGEN, R. D.; MCKINNON, J. J. A potential association between the BM 1500 microsatellite and fat deposition in beef cattle. **Mammalian Genome**; v.9, p.432- 434, 1998.
- GAO, M.J.; LYDIATE, D.J.; LI, X.; LUI, H.; GJETVAJ, B.; HEGEDIUS, D.D. Repression of seed maturation genes by a trihelix transcriptional repressor in *Arabidopsis* seedlings. **Plant Cell**; v.21, p.54–71, 2009.
- GODDARD, K.A.B.; WIJSMAN E.M. Characteristics of genetic markers and maps for cost-effective genome screens using diallelic markers. **Genetic Epidemiology**; v.22, p.205–220, 2002.
- GOLDBERG, R.B.; DE PAIVA, G.; YADEGARI, R. Plant embryogenesis: zygote to seed. **Science**; v.14, p. 266:605, 1994.
- GRAEF, G.; LAVALLEE, B. J.; TENOPIR, P.; TAT, M.; SCHWEIGER, B.; KINNEY, A.J.; VAN GERPEN, J.H.; CLEMENTE, T.E. A high-oleic-acid and low-palmitic-acid soybean: agronomic performance and evaluation as a feedstock for biodiesel. **Plant Biotechnology Journal**; v. 7, p. 411–421, 2009.
- GRIFFITHS, A. J; MILLER, J. H.; SUZUKI, D. T.; LEWONTIN, R. C.; GELBART, W. M. An introduction to genetic analysis. **W. H. Freeman and Company**; New York; p.1-103, 2000.
- GUERRIERO, G.; MARTIN, N.; GOLOVKO, A.; SUNDSTRÖM, J.F.; RASK, L.; EZCURRA, I. The RY/Sph element mediates transcriptional repression of maturation

genes from late maturation to early seedling growth. **New Phytologist**; 2009;184:552–65.

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.4524-535, 2001.

HAYDEN, M.J.; TABONE, T.; MATHE, D.E. Development and assessment of simple PCR markers for SNP genotyping in barley. **Theoretical and Applied Genetics**; v.119, p.939–951, 2009.

HU, F.B.; SAMPFER, M.J.; MANSON, J.E.; RIMM, E.; COLDITZ, G.A.; ROSNER, B.A.; HENNEKENS, C.H.; WILLETT, W.C. Dietary fat intake and the risk of coronary heart disease in women. **The New England Journal of Medicine**; v.337, p.1491–1499, 1997.

HYTEN, D. L.; CANNON, S. B.; SONG, Q.; WEEKS, N.; FICKUS, E. W.; SHOEMAKER, R.; SPECHT, J.; FARMER, A. D.; MAY, G. D.; CREGAN, P. High-throughput snp discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. **BMC Genomics**; v.11, p.38, 2010b.

HYTEN, D.L.; CHOI, I.-Y.; SONG, Q.; SPECHT, J.E.; CARTER, T.E.; SHOEMAKER, R.C.; HWANG, E.-Y.; MATUKUMALLI, L.K.; CREGAN, P.B. A high density integrated genetic linkage map of soybean and the development of a 1536 Universal Soy Linkage Panel for QTL mapping. **Crop Science**; v.50, 2010a.

JOHNSTON, M.L.; LUETHY, M.H.; MIERNYK, J.A.; RANDALL, D.D. Cloning and molecular analyses of the *Arabidopsis thaliana* plastid pyruvate dehydrogenase subunits. **Biochimica Et Biophysica Acta-bioenergetics**; v.1321, p.200–6, 1997.

KIM, S.; MISRA, A. SNP genotyping: technologies and biomedical applications. **Annual Review of Biomedical Engineering**; v. 9, p. 289-320, 2007.

KRUGLYAK, L. The use of a genetic map of biallelic markers in linkage studies. **Natural Genetics**; v.17, p.21–24, 1997.

KWOK, P.Y.; DENG, Q.; ZAKERI, H.; NICKERSON, D. A. Increasing the information content of STS-based genome maps: Identifying polymorphisms in mapped STSs. **Genomics**; v.31, p.123–126, 1996.

LI, Y.; BEISSON, F.; POLLARD, M.; OHLROGGE, J. Oil content of *Arabidopsis* seeds: The influence of seed anatomy, light and plant-to-plant variation. **Phytochemistry**; v.67, p.904–15, 2006.

LIAO, M P.Y.; LEE, K.H. From SNPs to functional polymorphism: The insight into biotechnology applications. **Biochemical Engineering Journal**; v.49, p.149-158, 2010.

- LIEFERS, S. C.; te PAS, M. F. W.; VEERKAMP, R. F.; van der LENDE, T. Associations between leptin gene polymorphisms and production, live weight, energy balance, feed intake, and fertility in *Holstein heifers*. **Journal of Dairy Science**; v.85, p.1633-1638, 2002.
- LIN, M.; BEHAL, R.; OLIVER, D.J. Disruption of *plE2*, the gene for the E2 subunit of the plastid pyruvate dehydrogenase complex, in *Arabidopsis* causes an early embryo lethal phenotype. **Plant Molecular Biology**; v.52, p.865-72, 2003.
- LINVILLE, R.C.; POMP, D.; JOHNSON, R.K.; ROTHSCHILD, M.F. Candidate gene analysis for loci affecting litter size and ovulation rate in swine. **Journal of Animal Science**; v.79, p.60-67, 2001.
- LITT, M.; LUTTY, J.A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American Journal of Human Genetics**; v.44, p.398-401, 1989.
- LIVAK, K.J. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genetic Analytical – **Biomolecular Engineering**; v.14, p. 143-149, 1999.
- LOS, D. A.; MURATA, N. Structure and expression of fatty acid desaturases. **Biochimica et Biophysica Acta**; v. 1394, p. 3-15. 1998.
- LOTAN, T.; OHTO, M.; YEE, K.M.; WEST, M.A.; LO, R.; KWONG, R.W. *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. **Cell**; v.93, p.1195-205, 1998.
- MAEDA, M.; MURAYAMA, N.; ISHII, H.; URYU, N.; OTA, M.; TSUJI, K.; INOKO, H. A simple and rapid method for HLA-DQA1 genotyping by digestion of PCR amplified DNA with allele specific restriction endonucleases. **Tissue Antigens**; v.34, n.5, p.290-298, 1989.
- MARKET, C.L.; MOLLER, F. Multiple forms of enzymes tissue, ontogenetic and species specific patterns. **Proceedings of the National Academy of Sciences USA**; v.45, p.753-762, 1959.
- MATISE, T.C.; SACHIDANANDAM, R.; CLARK, A.G.; KRUGLYAK, L.; WIJSMAN, E.; KAKOL, J.; BUYSKE, S.; CHUI, B.; COHEN, P.; TOMA, C.; EHM, M.; GLANOW, S.; HE, C.; HEIL, J.; MARKIANOS, K.; MCMULLEN, I.; PERICAK-VANCE, M.A.; SILBERGLEIT, A.; STEIN, L.; WAGNER, M.; WILSON, A.F. ; WINICK, J.D.; WINN-DEEN, E.S.; YAMASHIRO, C.T.; CANN, H.M.; LAI, E.; HOLDEN, A.L. A 3.9-centimorgan resolution human single-nucleotide polymorphism linkage map and screening set. **The American Journal of Human Genetics**; v.73, p.271-284, 2003.

MEKHEDOV, S.; ILÁRDUYA, M.; OHLROGGE, J. Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis. **Plant Physiology**; v.122, p.389-402, 2000.

MIDDLETON, F.A.; PATO, M.T.; GENTILE, K.L.; MORLEY, C.P.; ZHAO, X.; EISENER, A.F.; BROWN, A.; PETRYSHEN, T.L.; KIRBY, A.N.; MEDEIROS, H.; CARVALHO, C.; MACEDO, A.; DOURADO, A.; COELHO, I.; VALENTE, J.; SOARES, M.J.; FERREIRA, C.P.; LEI, M.; AZEVEDO, M.H.; KENNEDY, J.L.; DALY, M.J.; SKLAR, P.; PATO, C.N. Genomewide linkage analysis of bipolar disorder by use of a high-density single nucleotide polymorphism (SNP) genotyping assay: a comparison with microsatellite marker assays and finding of significant linkage to chromosome 6q22. **American Journal Of Human Genetics**; v.74, p.886–897, 2004.

MIQUEL, M.; BROWSE, J. Lipid biosynthesis in developing seeds. In: Kigel J, Galili G, editors. Seed development and germination. **New York: Marcel Dekker**; p. 169–93, 1995.

MOU, Z.; HE, Y.; DAI, Y.; LIU, X.; LI, J. Deficiency in fatty acid synthase leads to premature cell death and dramatic alterations in plant morphology. **Plant Cell**; v.12, p.405–17, 2000.

MOZAFFARIAN, D.; KATAN, M.B.; ASCHERIO, A.; STAMPFER, M.J.; WILLETT, W.C. Trans fatty acids and cardiovascular disease. **The New England Journal of Medicine**; v.354, p.1601–1613, 2006.

MULLIS, K.B.; FALOONA, F.A. Specific synthesis of DNA in-vitro via a polymerase catalyzed chain reaction. **Methods Enzymol**; v.155, p.335-350, 1987.

NEI, M.; LI, W.H. Mathematical model for studying genetic for variation in terms of restriction endonucleases. **Proceedings of the National Academy Sciences USA**; v.76, p.5269–5273, 1979.

NIKOLAU, B. J.; OHLROGGE, J. B.; WURTELE, E. S. Plant biotin- containing carboxylases. **Archives of Biochemistry and Bio-Physics**; v.414, n.2, p.211–222, 2003.

O'HARA, P.; SLABAS, A.R.; FAWCETT, T. Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. **Plant Physiology**; v.129, p.310–20, 2002.

OHLROGGE, J.; JAWORSKI, J. G. Regulation of fatty acid synthesis. **Plant Physiology. Plant Molecular Biology**; v.48, p.109-136, 1997.

ORAGUZIE, N. C. Association mapping in plants. NEW YORK; USA: **Springer**; v.ix, p.277, 2007.

PENFIELD, S.; RYLOTT, EL.; GILDAY, A.D.; GRAHAM, S.; LARSON, T.R.; GRAHAM, I.A. Reserve mobilization in the *Arabidopsis* endosperm fuels hypocotyl

elongation in the dark, is dependent of abscisic acid, and requires PHOSPHOENOLPYRUVATE CARBOXYKINASE1. **Plant Cell**; v.16, p.2705–18, 2004.

PERES, J. R. R.; FREITAS JUNIOR, E. de; GAZZONI, D. L. Biocombustíveis: uma oportunidade para o agronegócio brasileiro. **Revista de Política Agrícola**, Brasília, v. 14, n.1, p. 31 - 41, 2005.

PIDKOWITCH, M.S.; NGUYEN, H.T.; HEILMANN, I.; ISCHEBECK, T.; SHANKLIN, J. Modulating seed β -ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil. **Proceedings of the National Academy Sciences USA**; v.104, p.4742–7, 2007.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**; v.5, p.94-100, 2002.

RAMLI, U.S.; BAKER, D.S.; QUANT, P.A.; HARWOOD, J.L. Control analysis of lipid biosynthesis in tissue cultures from oil crops shows that flux control is shared between fatty acid synthesis and lipid assembly. **Biochemical Journal**; v.364, p.393–401, 2002.

RAMLI, U.S.; SALAS, J.J.; QUANT, P.A.; HARWOOD, J.J. Use of metabolic control analysis to give quantitative information on control of lipid biosynthesis in the important oil crop, *Elaeis guineensis* (oilpalm). **New Phytologist**; v.184, p.330–9, 2009.

ROTHSCHILD, M.F.; SOLLER, M. Candidate gene analysis to detect genes controlling of economic importance in domestic livestock. In: **Simpósio internacional de genética e melhoramento animal**, Viçosa - MG. Anais. Viçosa: UFV, 1999. p.219-242.

SALAS, J.J.; OHLROGGE, J.B. Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. **Archives of Biochemistry and Biophysics**; v.403, p.25–34, 2002.

SANTOS, M.M.; DUBREUCQ, B.; BAUD, S.; PARCY, F.; CABOCHE, M.; LEPINIEC, L. Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. **The Plant journal**; v.54, p.608–20, 2008

SAX K. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. **Genetics**; v.8, p.522-560, 1923.

SCHLOTTERER, C. The evolution of molecular markers - just a matter of fashion? **Nature Reviews Genetics**; v.5, n.1, p.63-69, 2004.

SEMAGN, K.; BJRNSTAD, A.B.; AND NDJIONDJOP, M.N. An overview of molecular marker methods for plants. **African Journal of Biotechnology**; v.5, n.25, p. 2540-2568, 2006.

SHORT, T.H.; ROTHSCHILD, M.F.; SOUTHWOOD, O.I.; MCLAREN, D.G.; DE VRIES, A.; VAN DER STEEN, H.; ECKARDT, G.R.; TUGGLE, C.K.; HELM, J.; VASKE, D.A.; MILEHAM, A.J.; PLASTOW, G.S. Effect of the estrogen receptor locus on reproduction and production traits in four commercial pig lines. **Journal of Animal Science**; v.75, n.12, p.3138-3142, 1997.

SMITH, E.J.; SHI, L.; DRUMMOND, P. Expressed sequence tags for the chicken genome from a normalized 10-day-old White Leghorn whole embryo cDNA library: 1. DNA sequence characterization and linkage analysis, **Journal of Heredity**; v.92, p.1-8, 2001.

SUZUKI M.; MCCARTY D.R. Functional symmetry of the B3 network controlling seed development. **Current Opinion in Plant Biology**; v.11, p.548–53, 2008.

TABONE, T.; MATHER, D.E.; HAYDEN, M.J. Temperature Switch PCR (TSP): Robust assay design for reliable amplification and genotyping of SNPs. **BMC Genomics**; v.10, 580, 2009.

TO, A.; VALON, C.; SAVINO, G.; GUILLEMINOT, J.; DEVIC, M.; GIRAUDAT, J. A network of local and redundant gene regulation governs *Arabidopsis* seed maturation. **Plant Cell**; v.18, p.1642–51, 2006.

TSUKAGOSHI, H.; MORIKAMI, A.; NAKAMURA, K. Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in *Arabidopsis* seedlings. **Proceedings of the National Academy Sciences USA**; v.104, p.2543–7, 2007.

UBRABIO - União Brasileira do Biodiesel. O biodiesel e sua contribuição ao desenvolvimento brasileiro. Disponível em: <<http://www.ubrablo.com.br/sites/1700/1729/00000169.pdf>> Acesso em: outubro de 2010.

USDA - Departamento de Agricultura dos Estados Unidos. Disponível em: <http://www.fas.usda.gov/oilseeds/circular/2012/April/oilseeds_full04-12.pdf>. Acesso em: agosto de 2012.

USDA - Departamento de Agricultura dos Estados Unidos. Disponível em: <<http://www.sistemafeap.org.br/noticia.aspx?id=1345>>. Acesso em: maio de 2012.

VIGEOLAS, H.; WALDECK, P.; ZANK, T.; GEIGENBERGER, P. Increasing seed oil content in oil-seed rape (*Brassica napus L.*) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. **Plant Biotechnology Journal**; v.5, p.431–41, 2007.

VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M. A review on SNP and other types of molecular markers and their use in animal genetics. **Genetics, Selection, Evolution**; v.34, p.275-305, 2002.

- VINCENT, A.L.; EVANS, G.; SHORT, T.H.; SOUTHWOOD, O.I.; PLASTOW, G.S.; TOGGLE, C.K.; ROTHSCHILD, M.F. The prolactin receptor gene is associated with increased litter size in pigs. In: **WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK. PRODUCTION**, 6., Armidale, 1998. **Proceedings. Armidale: NSW**; v.27, p.15-18, 1998.
- VOS, P.; HOGERS, R.; BLEEKER, M. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Research**; v.23(5), p.4407-4414, 1995.
- WEBER, J.L.; MAY, P.E. Abundant class of human DNA polymorphism which can be typed using the Polymerase Chain Reaction. **American Journal of Human Genetics**; v.44, p.388-396, 1989.
- WELSH, J.; MCCLELLAND, M. Fingerprinting genomes using PCR with arbitrary primers **Nucleic Acids Research**; v.18(7), p.7213-7218, 1990.
- WESELAKE, R.J.; TAYLOR, D.C.; RAHMAN, M.H.; SHAH, S.; LAROCHE, A.; MCVETTY, P.B.E. Increasing the flow of carbon into seed oil. **Biotechnology Advances**; v.27, p.866–78, 2009.
- WESELAKE, R.J.; SHAH, S.; TANG, M.; QUANT, PA.; SNYDER, C.; FURUKAWA-STOFFER, T.L. Metabolic control analysis is helpful for informed genetic manipulation of oilseed rape (*Brassica napus*) to increase seed oil content. **Journal of Experimental Botany**; v.59, p.3543–9, 2008.
- WILLIAMS, J.G.K.; KUBELICK, A.R.; LIVAK, K.J. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Research**; v.18(6), p.6531-6535, 1990.
- WU, E.T AL. SNP discovery by high-throughput sequencing in soybean. **BMC Genomics**, v.11, p.469, 2010.
- YADAV, N. S. Genetic modification of soybean oil quality. In: Verma, D. P. S.; SHOEMAKER, R. C. (Eds.). Soybean genetics, molecular biology and biotechnology. **USA: CAB INTERNATIONAL, WALLINGFORD, CT**, p. 127-188. 1996.
- YE, S.; DHILLON, S.; XIAYI, K.; ANDREW, R.C.; DAY, N.M. An efficient procedure for genotyping single nucleotide polymorphisms. **Nucleic Acids Research**; v.29, No. 17 e 88, 2001.
- ZHANG, H.; OGAS, J. An epigenetic perspective on developmental regulation of seed genes. **Molecular Plant**; v.2, p.610–27, 2009.
- ZHU, T.; SALMERON, J. High-definition genome profiling for genetic marker discovery. **Trends Plant Science**; v.12, p.196–202, 2007.

ZHU, Y. L.; SONG, Q. J.; HYTEN, D. L.; VAN TASSELL, C. P.; MATUKUMALLI, L. K. Single-nucleotide polymorphisms in soybean. **Genetics**; v.163; p.1123-1134, 2003.

CAPÍTULO 1

Identificação de marcadores SNP (*Single Nucleotide Polymorphism*) em regiões gênicas regulatórias potencialmente associadas ao conteúdo de óleo e proteína em sementes de soja

1. Resumo

O objetivo deste trabalho foi identificar e validar SNP em regiões gênicas associadas ao conteúdo de óleo em genótipos divergentes para essas características. Para isso, foram desenhados “*primers*” para amplificar regiões específicas dos transcritos de 19 genes candidatos, em seis genótipos contrastantes para os caracteres teor de óleo. O sequenciamento dessas regiões gênicas, possibilitou a identificação de 81 SNP e 26 INDELs em 31,2 Kpb de sequencias, sendo a média de um SNP para cada 292 pb. Com o intuito de obter informações sobre a expressão dos transcritos dos 19 genes candidatos selecionados em soja, foi feito um alinhamento por meio do algoritmo BLASTn, contra bibliotecas de ESTs derivadas de sete tecidos de soja, depositados no banco de dados do NCBI. Os resultados mostraram que todos os transcritos selecionados dos genes ADPRI e G3PDH foram expressos constitutivamente, enquanto que os transcritos dos genes ABI3, ACC, AMIP, ARAF, ASIL, CLPPR, CPT, DGAT, DOF4, FUS3, LEC1A, LEC1B, LPCAT, MDH e WRI, apresentaram expressão órgão-específica nas sete bibliotecas testadas. *Primers* de genotipagem TSPCR (*Temperature-Switch Polymerase Chain Reaction*) foram utilizados para discriminar os alelos de SNP em três populações de RILs. Os polimorfismos identificados por meio do sequenciamento foram inicialmente testados nos respectivos genitores, sendo para isso, desenhados 17 conjuntos de *primers* de genotipagem TSPCR (*Temperature-Switch Polymerase Chain Reaction*). Somente os *primers* TSPCR desenhados no polimorfismo dos genes AMIP-5, AMIP-3, ABI3, LEC1B, ARAF, LPCAT, mostraram-se polimórficos sem bandas inespecíficas e/ou marcação de bandas fracas nos respectivos progenitores. Os *primers* de genotipagem TSPCR desenhados no polimorfismo dos genes AMIP-5, AMIP-3, ABI3 e LEC1B amplificados em 176 progênies F₆ da população PI371611 x CD222. Os *primers* TSPCR desenhados no polimorfismo dos genes ARAF, LPCAT, e LEC1B foram genotipados em 236 progênies na geração F₆, derivadas do cruzamento CD01RR8311 x SUPREMA. Enquanto que os *primers* TSPCR, desenhados para detectar o polimorfismo presente nos genes ABI3 e LEC1B foram amplificados em 170 progênies F₆, derivadas do cruzamento entre A7002 x CD219. Apenas o polimorfismo (A/C) do gene ABI3, apresentou associação significativa para teor de proteína na população de mapeamento A7002 e CD219, explicando 4,7 % da variação fenotípica.

2. Introdução

Com uma produção estimada em 72 milhões de toneladas de grãos para a safra 2011/2012, o Brasil é o segundo maior produtor mundial de soja (*Glycine max* (L.) Merrill) e deverá tornar-se o maior exportador desta leguminosa ainda neste ano agrícola (USDA, 2012). O sucesso do cultivo da soja é atribuído à combinação dos seus altos teores de proteína (cerca de 40%) e óleo (cerca de 20%), juntamente com níveis satisfatórios de produtividade de grãos em diversas condições edafoclimáticas.

A importância econômica dessa cultura está pautada na possibilidade de extrair do grão um óleo de qualidade desejável tanto para a alimentação humana quanto para a produção de biodiesel, e do farelo (resultante do esmagamento do grão para a produção do óleo de soja), o qual apresenta elevado teor protéico, podem ser elaboradas rações animais de excelente valor nutricional.

O desenvolvimento de cultivares produtivas com incremento significativo no teor de proteína e/ou de óleo está entre os principais objetivos dos programas de melhoramento genético de soja (MAHMOUD et al., 2006). O conteúdo de proteína e óleo dos grãos de soja tem controle genético quantitativo, sendo influenciado pelo ambiente (BONATO, 2000). A existência de correlação negativa entre teores de proteína e óleo nos grãos de soja tem impedido o aumento simultâneo dessas características nas linhagens comerciais (RODRIGUES et al., 2010).

O acúmulo de óleo e proteína em sementes de soja faz parte do processo de maturação das sementes, um programa de desenvolvimento altamente controlado a nível transcricional, o qual é definido nos tecidos de embriões, uma vez que a morfogênese é alcançada (FAWCETT et al., 1994; O'HARA et al., 2002; BAUD e GRAHAM, 2006).

Ativação dos genes chaves em *Arabidopsis thaliana*, também chamados de reguladores mestres do processo de maturação, inclui membros da superfamília do domínio B3, ou seja, membros da rede AFL (ABI3, LEC2, e FUS3) (BAUD e LEPINIEC, 2010). Os membros da rede AFL, atuam em conjunto com LEC1, um fator de transcrição do tipo CAAT *binding factor* (CBF). Foi proposto por BAUD e LEPINIEC (2010) um modelo da regulação transcricional da biossíntese de triacilglicerol (TGA), onde os membros da rede AFL, juntamente com LEC1 ligam-se a domínios específicos da molécula de DNA ativando genes envolvidos; i) no metabolismo primário; ii) na síntese de ácidos graxos; iii) na montagem de TGA e; iv) na modificação dos ácidos graxos.

Possivelmente, a identificação de novos reguladores e dos principais precursores metabólicos relacionados à biossíntese de óleo e de proteínas em sementes de soja irá facilitar a obtenção de genótipos de soja com características superiores.

Dois precursores metabólicos envolvidos na biossíntese de TGA são; i) o glicerol-3-fosfato (G3P) que fornece a cadeia de glicerol necessária para a síntese de moléculas TAG, ii) a descarboxilação oxidativa do piruvato plastidial, pelo complexo piruvato desidrogenase (PDC), produz CO₂ e acetil-CoA, sendo este último usado como um precursor na síntese *de novo* ácido graxo nos plastídios (JOHNSTON et al., 1997; LIN et al., 2003).

A identificação de novos marcadores SNP (*Single Nucleotide Polymorphism*), a partir de genes candidatos, envolvidos tanto na regulação ou diretamente nas vias biossintéticas de acúmulo de óleo e/ou proteína em sementes de soja, pode acelerar a obtenção de variedades com maior teor de óleo e/ou proteína, em um menor espaço de tempo.

Neste contexto, o presente trabalho teve como objetivo identificar e validar SNP em regiões gênicas associadas ao conteúdo de óleo e/ou proteína, em cultivares divergentes do Programa de Melhoramento da Qualidade da Soja (PMQS-UFV).

3. Materiais e Métodos

3.1. Escolha do material genético

Foram utilizados os genótipos de soja CD222, SUPREMA, CD219, CD01RR83841, A7002, PI371611, contrastantes para o teor de óleo e/ou proteína na semente, oriundos do banco de germoplasma do Programa de Melhoramento da Qualidade da Soja, em desenvolvimento no BIOAGRO/Universidade Federal de Viçosa (PMQS/BIOAGRO/UFV). Estes materiais foram plantados em casa de vegetação e as folhas foram coletadas aos 15 dias após o plantio para a extração do DNA, usando a metodologia proposta por DOYLE & DOYLE (1990), com modificações. A concentração e qualidade do DNA foi determinada por meio do espectrofotômetro NanoDrop (NanoDrop Technologies, Wilmington, DE, EUA). O DNA obtido foi também submetido à eletroforese em gel de agarose 0,8%, para avaliar a qualidade. Com base na concentração estimada, procedeu-se à diluição das amostras para uma concentração final de 15 ng/μL.

3.2. Seleção dos genes candidatos e desenho de *primers*

Com base em informações disponíveis na literatura, foram selecionados 19 genes candidatos, que de acordo com a literatura estão diretamente relacionados com acúmulo de óleo e/ou proteína no grão de soja.

Os genes candidatos selecionados, ou codificam enzimas da via de biossíntese de ácido graxos, ou são fatores de transcrição envolvidos na regulação de genes envolvidos no acúmulo de óleo e/ou proteína. Foram também selecionados genes envolvidos no direcionamento e transporte de moléculas em sementes de soja ou em outras espécies. Na Tabela 1 estão listados os genes selecionados e suas respectivas funções, além das referências bibliográficas.

As sequências parciais dos genes listados na Tabela 1 ou de genes homólogos, foram obtidas a partir do GenBank (<http://www.ncbi.nlm.nih.gov/Genbank>). As sequências gênicas completas (regiões promotoras, regulatórias, terminadoras, introns e exons) foram obtidas realizando um BLASTn (*Basic Local Alignment Search Tool*) no banco de dados PHYTOZOME (<http://www.phytozome.net/soybean>), onde está depositado o genoma da soja sequenciado (SCHMUTZ et al., 2010).

A seleção de cada gene candidato seguiu os seguintes critérios: i) ser copia única ou no máximo 2 cópias; ii) apresentar valor de *e-value* igual ou próximo a zero e iii) apresentar anotação funcional no PHYTOZOME.

Visando a amplificação das regiões selecionadas para cada gene candidato, foram desenhados oligonucleotídeos iniciadores (*primers*) usando o programa “*Primer3 Input Program*” (<http://frodo.wi.mit.edu/primer3>). Para 14 dos genes candidatos, *primers* foram desenhados nas regiões regulatórias (5’UTR e 3’UTR). Para cinco dos genes candidatos (Dof4, Dof11, MDH, LEC1A e LEC1B), por apresentarem sequências genômicas curtas, foram desenhados de três a quatro pares de *primers* para cada gene candidato, com o objetivo de amplificar e sequenciar todo o gene.

Os *primers* foram desenhados seguindo os parâmetros: i) o fragmento a ser amplificado variando de 600 a 1000 pb (pares de bases); ii) tamanho do primer entre 19 e 25 pb; iii) temperatura de anelamento entre 59 e 61°C; e iv) conteúdo de CG% entre 40 e 60%. Cada *primer* desenhado foi testado quanto a sua estabilidade termodinâmica usando o programa *Primer Select 3.1* (DNASTAR, Madison, WI, EUA), enquanto que a especificidade foi testada pelo programa *PCR virtual* (iPCR)

(http://www.ch.embnet.org/software/iPCR_form.html) e pelo banco de dados PHYTOZOME. Após serem testados, os *primers* foram sintetizados pela empresa BIONEER (Belo Horizonte, Minas Gerais, Brasil).

Tabela 1. Genes candidatos selecionados com os códigos e as funções das proteínas por eles codificadas em soja ou em plantas relacionadas. Bem como as referências bibliográficas e os transcritos selecionados do banco de dados phytozome

Gene candidato	Código	Função	Transcritos Selecionados	Referências
Fator de transcrição Dof4	Dof4	Regulação de genes envolvidos na biossíntese de ácidos graxos.	Glyma17g08950.1	WANG et al., 2007
Fator de transcrição Dof11	Dof11	Regulação de genes envolvidos na biossíntese de ácidos graxos.	Glyma13g40420.1	WANG et al., 2007
Fator de transcrição LEC1-A	LEC1-A	Regulação da síntese de proteínas de reserva.	Glyma07g39820.1	VERDIER et al. 2008
Fator de transcrição LEC1-B	LEC1-B	Regulação da síntese de proteínas de reserva.	Glyma17g00950.1	VERDIER et al. 2008
Fator de transcrição Insensível ao ABA 3	ABI3	Regulação do desenvolvimento da semente.	Glyma08g47240.1 Glyma18g38490.1	WESELAKE et al., 2009
Fator de ribosilação ADP-dependente	ADPRI	Direcionamento e transporte de moléculas.	Glyma19g00200.1 Glyma05g08700.1	WEI et al., 2008
Fator de transcrição FUS3	FUS3	Regulação do desenvolvimento da semente.	Glyma19g27340.1 Glyma16g05480.1	WESELAKE et al., 2009
Fator de transcrição WRINKLED1	WRI 1	Regulação do acúmulo de óleo durante o desenvolvimento da semente.	Glyma15g34770.1 Glyma08g24420.1	BAUD et al., 2007
Fator de transcrição ASIL	ASIL	Regulação do desenvolvimento da semente.	Glyma07g04790.1 Glyma16g01370.1	BAUD et al., 2010
Malato Desidrogenase Nodule-Enhanced	MDH	Assimilação de nitrogênio.	Glyma17g10880.1	IMSANDE et al. 2001
Alfa-L-Arabinofuranosidase	ARAF	Hidrólise de hemicelulose.	Glyma19g30450.1 Glyma03g27460.2	WEI et al., 2008
Permease de aminoácido	AMIP	Aumento do teor de proteínas de reserva.	Glyma12g03580.1 Glyma11g11440.1	WEIGELT et al., 2008
Subunidade proteolítica da protease Clp ATP-dependente	CLPPRO	Captação de aminoácidos para síntese de proteínas de reserva.	Glyma10g01550.1 Glyma02g01510.1	WEI et al., 2008
Gliceraldeído-3-fosfato desidrogenase	G3PDH	Via biossintética dos triacilgliceróis.	Glyma18g01330.1 Glyma11g37360.1	VIGEOLAS et al., 2007

Diacilglicerol aciltransferase	DGAT	Via biossintética dos triacilgliceróis	Glyma09g07520.1 Glyma13g16560.1	BAUD et al., 2010
Colinafosfotransferase	CPT	Via biossintética dos triacilgliceróis	Glyma12g08720.1 Glyma02g14210.1	BAUD et al., 2010
Lisofosfatidilcolina aciltransferase	LPCAT	Via biossintética dos triacilgliceróis	Glyma05g03510.1 Glyma17g14070.1	BAUD et al., 2010
Fosfolípídeos diacilglicerol-aciltransferase	PDAT	Via biossintética dos triacilgliceróis	Glyma13g16790.1 Glyma17g05910.1	BAUD et al., 2010
Acetil-CoA-carboxilase	ACC	Via biossintética dos triacilgliceróis	Glyma06g11120.1 Glyma04g11550.1	BAUD et al., 2010

3.3. Reações de amplificação e sequenciamento

As reações de amplificação foram feitas para um volume final de 30 µL, usando-se 1U de Platinum®Taq DNA polimerase; Tris-HCl 20 mM pH 8,4; KCl 50 mM; MgCl₂ 1,5 mM; 0,2 mM de cada dNTP; 0,2 µM de cada primer e 30 ng de DNA molde. As reações de amplificação foram conduzidas em termociclador Eppendorff (AG-22331), inicialmente a 94 °C por 2 minutos, seguida de 35 ciclos de 94 °C por 30 segundos, 55 °C por 30 segundos e 72 °C por 60 segundos. Realizadas as reações de amplificação, os produtos gerados foram submetidos à eletroforese em gel de agarose 1,7%, contendo brometo de etídeo, e fotodigitalizados utilizando o sistema L-PIX (Loccus Biotecnologia, São Paulo, Brasil).

Quando apenas o único produto de amplificação esperado foi gerado, estes foram purificados usando-se o kit ExoSAP IT (USB Corporation, Cleveland, Ohio, EUA) na proporção de 6 µL de ExoSap e 18 µL de produto de PCR. Posteriormente, as amostras purificadas foram sequenciadas pelo método didesoxi, utilizando o sequenciador ABI3730xl (Applied Biosystems). As reações de sequenciamento foram feitas pela empresa Macrogen (Gasandong, Geumchun-gu, Seoul, Korea). As sequências obtidas foram editadas usando-se o programa SEQUENCER 4.1.4 (Genes Code Corporation) e alinhadas pelo programa ClustalW (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). As mesmas sequências foram também editadas e alinhadas pelo programa CodonCode Aligner (<http://www.codoncode.com/aligner/>).

3.4. Análise de expressão *in silico*

Objetivando-se obter informações sobre a expressão dos genes candidatos selecionados, foi feito um alinhamento por meio do algoritmo BLASTn, das sequências codificadoras dos transcritos codificados pelos 19 genes candidatos, bem como dos transcritos alternativos, contra bibliotecas de ESTs (*Expressed Sequence Tag*) de soja, depositados no banco de dados do NCBI. As bibliotecas comparadas foram construídas a partir de embriões somáticos, cotilédones, flores, folhas, vargem/sementes, planta inteira e meristema.

O ponto de corte utilizado foi um E-value menor ou igual a $1e^{-100}$, respeitando um ponto de corte mínimo de 90 % de identidade e 100 pares de base (pb) de cobertura, respectivamente.

3.5. Populações

Para a validação dos SNP relacionados às características em estudo, foram desenvolvidas três populações de linhagens endogâmicas recombinantes (*Recombinant Inbred Lines - RILs*), originadas dos cruzamentos entre genótipos contrastantes (A7002 X CD219, PI371611 X CD222 e SUPREMA X CD01RR8384). As variedades A7002 e SUPREMA foram obtidas originalmente pela empresa AVENTIS, atualmente adquirida pela empresa NIDERA sementes, tendo como característica alto teor de óleo. Os genótipos CD219, CD01RR8384 (resistentes ao glifosato) e CD222 (alto teor de óleo) são variedades comerciais da empresa COODETEC (Cooperativa Central de Pesquisa Agrícola). Já o genótipo PI371611, é um acesso originalmente classificado para alto teor de óleo, também fornecido pela COODETEC.

Os cruzamentos foram iniciados no ano agrícola de 2006/2007 na Universidade Federal de Viçosa (UFV). Inicialmente plantas F_1 foram plantadas em casa de vegetação, posteriormente as sementes F_2 foram plantadas e avançadas até a geração $F_{2:8}$, usando o método descendente de uma única semente (*Single Seed Descent - SSD*) como reportado por BRIM (1966).

As sementes F_2 foram plantadas em janeiro de 2008 na estação experimental da COOPADAP (Cooperativa Agropecuária do Alto Paranaíba) no município de São Gotardo, Minas Gerais, Brasil. Em outubro de 2009 sementes F_3 foram plantadas no campo experimental da UFV em Rio Branco, Minas Gerais, Brasil. Em junho de 2010, sementes F_4 também foram plantadas no campo experimental da UFV em Rio Branco. As sementes F_5 foram plantadas e avançadas até a geração F_6 em casa de vegetação da UFV, no ano de 2011. No ano agrícola 2011/2012 foi montando um experimento na estação experimental da COOPADAP, São Gotardo – MG, para avaliar o comportamento das três populações de mapeamento. Para a população CD01RR8384 x SUPREMA foi montado um ensaio com duas repetições, contendo 236 linhagens e os dois genitores; para a população CD219 x A7002 foram avaliadas 170 linhagens e os respectivos genitores em um ensaio com três repetições; para a população PI371611 x CD222 foram avaliadas 176 linhagens e os dois genitores em um ensaio com duas

repetições. Os três ensaios foram montados no delineamento de blocos casualizados, com parcelas de 2 metros, espaçadas em 0,5 metro, com densidade de semeadura de 13 a 15 sementes por metro. O número de repetições variável de uma população para outra, deu-se em função da limitada quantidade de sementes disponível.

3.6. Extração do DNA

Para a genotipagem das populações RILs, foram coletadas folhas F₆ das progênies 15 dias após o plantio para a extração e análise do DNA, usando a metodologia proposta por DOYLE & DOYLE (1990), com modificações. A concentração e qualidade do DNA foi determinada por meio do espectrofotômetro NanoDrop (NanoDrop Technologies, Wilmington, DE). O DNA obtido foi também submetido à eletroforese em gel de agarose 0,8%, para avaliar a qualidade. Com base na concentração estimada, procedeu-se à diluição das amostras para uma concentração final de 15 ng/μL.

3.7. Determinação dos teores de óleo e proteínas

As determinações dos teores de proteína e óleo das populações RILs foram feitas no Laboratório de Bioquímica Molecular de Plantas/BIOAGRO/UFV. Sendo coletados 10 g de sementes de cada progênie. As sementes foram trituradas em um moinho de forma a gerar granulometria adequada para leitura em espectrômetro infravermelho próximo (FT-NIR, equipamento *Thermo Scientific*, modelo Antaris II).

3.8. Análise estatística dos dados

3.8.1. Teste de normalidade

3.8.1.1. Teste de assimetria

O coeficiente de assimetria fornece o grau de desvio, ou afastamento da simetria, de uma distribuição de dados. Em uma distribuição simétrica tem-se igualdade dos valores da média, mediana e moda. A estimativa do coeficiente de assimetria (s) de uma variável X, numa amostra de tamanho n, é dada por:

$$\hat{s} = \frac{n}{(n-1)(n-2)} \cdot \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s(X)} \right)^3$$

Se o resultado for zero, a distribuição é simétrica, se o resultado for negativo, a distribuição é assimétrica negativa (inclinada para a esquerda) e se o resultado for positivo, a distribuição é assimétrica positiva (inclinada para a direita).

As hipóteses $H_0: s = 0$ e $H_a: s \neq 0$ serão testadas pelo teste t:

$$t = \frac{\hat{s}}{\sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}}$$

O valor de t calculado será comparado ao de t tabelado, a nível α de probabilidade, com infinitos graus de liberdade. Se $|t_{cal}| \geq t_{tab}$, rejeita-se H_0 . (SCHUSTER e CUZ, 2004).

3.8.1.2. Teste de curtose

Curtose é o grau de achatamento de uma distribuição, considerado usualmente em relação à distribuição normal. Segundo SCHUSTER e CRUZ (2004), para se estimar o grau de curtose (k) de uma variável X, utiliza-se a seguinte fórmula:

$$\hat{k} = \frac{n \cdot (n+1)}{(n-1)(n-2)(n-3)} \cdot \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s(X)} \right)^4 - \frac{3 \cdot (n-1)^2}{(n-2)(n-3)}$$

Se o resultado for zero, a distribuição de freqüências é a própria distribuição normal, sendo chamada de mesocúrtica ($k = 0$), se o resultado for negativo ($k < 0$), a distribuição é achatada (alta variabilidade) e chamada de planicúrtica e se o resultado for positivo ($k > 0$), a distribuição é concentrada em torno da média (alta homogeneidade) e chamada de leptocúrtica.

Para testar a hipótese $H_0: k = 0$ vs $H_a: k \neq 0$ tem-se o teste t:

$$t = \frac{\hat{k}}{\sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}}$$

Este valor de t calculado será comparado ao valor de t tabelado, a nível α de probabilidade, com infinitos graus de liberdade. Se $|t_{\text{cal}}| \geq t_{\text{tab}}$, rejeita-se H_0 .

3.8.1.3. Teste de Lilliefors

O teste de Lilliefors apresenta o mesmo princípio que o teste de qui-quadrado (χ^2), porém considera os dados individualmente, ou seja, dada uma amostra de tamanho n, agrupam-se os dados em k classes, podendo-se formar uma distribuição de frequência com intervalos de classes iguais a 0,25, 0,50, 0,75, 1,0 do desvio padrão. As frequências esperadas (FE_i) de cada classe são encontradas de acordo com a distribuição normal reduzida z, multiplicando-se a área de cada classe pelo número total de observações (n) e as frequências observadas (FO_i) são os números de observação de cada classe (n_i). Para isso, obtêm-se os valores de z com a introdução de uma modificação, ampliando o seu uso para os casos em que a média e a variância não são conhecidas, mas sim estimadas através dos dados da amostra, como segue:

$$z_i = \frac{X_i - \bar{X}}{s},$$

em que:

z_i = valor da variável normal padronizada da classe i;

X_i = valor máximo da classe i;

\bar{X} = média geral da amostra de dados;

s = desvio padrão da amostra de dados.

Na maioria das vezes, o teste de Lilliefors é mais poderoso que o χ^2 .

O teste de Lilliefors pressupõe o cálculo de todos os z_i , os quais devem ser ordenados para as seguintes considerações:

$F(z_i) = FE_i \leq z_i$ = valor da tabela de distribuição normal reduzida (área);

$S(z_i) = FO_i \leq z_i = n_i/n$, em que:

n_i = número de valores observados $\leq z_i$;

n = número total de observações da amostra.

D = Máximo $|F(z_i) - S(z_i)|$.

O teste é bilateral, como segue:

H_0 : é razoável estudar os dados através da distribuição normal;

H_a : não é razoável estudar os dados através da distribuição normal.

Rejeita-se a hipótese de nulidade quando $D_{\text{cal}} \geq D_{\text{tab}}$, em um nível α de probabilidade com n observações, caso contrário não se rejeita H_0

Deve-se lembrar, porém, que a não rejeição de H_0 indica apenas que esta é uma razoável aproximação da distribuição conhecida.

3.8.2. Análise de variância

A análise de variância dos caracteres teor de óleo e proteína quantificados nas três populações de mapeamento foi feita utilizando o modelo indicado abaixo, realizada segundo o delineamento em blocos completos casualizados.

$$Y_{ij} = \mu + T_i + B_j + \varepsilon_{ij},$$

em que Y_{ij} é o valor observado da característica no i -ésimo tratamento (família ou testemunha) no j -ésimo bloco; μ é a média geral; T_i é o efeito do i -ésimo tratamento. Para ($i = 1, 2, \dots, g$) tem-se o efeito aleatório ou fixo para os genótipos e ($i = g + 1, g + 2, \dots, g + t$) tem-se o efeito fixo para as testemunhas; B_j é o efeito do j -ésimo bloco ($j = 1, 2, \dots, r$); ε_{ij} é o erro aleatório, $\varepsilon_{ij} \sim NID(0, \sigma^2)$.

O esquema da análise de variância, com base neste modelo, está apresentado na Tabela 2.

Tabela 2. Esquema da análise de variância do modelo em blocos completos casualizados

FV	GL	QM	E(QM)	F
Blocos	r-1	QMB	$\sigma^2 + T\sigma_b^2$	
Tratamentos	T-1	QMT	$\sigma^2 + r\phi_T$	QM _t /QM _r
Resíduo	(r-1)(T-1)	QMR	σ^2	
Total	rT-1			

Sendo T = g + t o número total de tratamentos

As médias; somas dos quadrados e quadrados médios foram obtidos por meio das seguintes expressões:

$$\text{Média geral} = M = \frac{Y_{..}}{Tr} = \frac{\sum_{i=1}^T \sum_{j=1}^r Y_{ij}}{Tr}; C = \frac{Y_{..}^2}{Tr} = C_g + C_t$$

$$\text{Média dos genótipos} = M_g = \frac{Y_g}{gr} = \frac{\sum_{i=1}^g \sum_{j=1}^r Y_{ij}}{gr}; C_g = \frac{Y_g^2}{gr}$$

Obtiveram-se as estimativas dos parâmetros da seguinte forma:

$$\text{Variabilidade genética: } \hat{\sigma}_g^2 = \frac{QMG - QMR}{r}$$

$$\text{Coeficiente de variação genético: } CV_g \% = \frac{(100\sqrt{\hat{\sigma}_g^2})}{M_g}$$

3.8.3. Análise associativa

A análise de associação marcador-característica foi determinada pela análise de marca simples, por meio de regressão linear. Desta forma, o seguinte modelo foi utilizado:

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

em que:

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

X_j = é o código do marcador (MM = 2, Mm = 1 e mm = 0);

β_0 = intercepto da regressão (média da característica);

β_1 = inclinação da reta para efeito aditivo;

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

Assim, na avaliação de N (N = 1.000) indivíduos em cada geração os dados foram codificados conforme a Tabela 3.

Tabela 3. Código para os valores fenotípicos da característica de acordo com os genótipos dos marcadores

Genótipo do marcador	Número de indivíduos	Valores da característica	Código de X
MM	n_1	$Y_{11}, Y_{21}, \dots, Y_{n1,1}$	2
Mm	n_2	$Y_{12}, Y_{22}, \dots, Y_{n2,2}$	1
mm	n_3	$Y_{13}, Y_{23}, \dots, Y_{n3,3}$	0

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

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

3.8.4. Recursos computacionais

Todas as análises estatísticas foram realizadas empregando os aplicativos computacionais GENES (CRUZ, 2006) e GQMOL (CRUZ e SCHUSTER, 2004).

3.9. Desenhos dos *primers* de genotipagem TSPCR

Para investigar os polimorfismos identificados nos genes candidatos, foram desenhados *primers* de genotipagem TSPCR (*Temperature-Switch Polymerase Chain Reaction*) (HAYDEN et al., 2009). Os *primers* TSPCR foram desenhados com auxílio dos programas *Primer 3* e *Primer Select*.

Foram desenhados para cada um dos polimorfismos selecionados, um par de *primers* loco específico-LE (R/F), com temperatura de anelamento (Ta) entre 60 e 65 °C, visando amplificar fragmentos de 400 pb, e um *primer* informativo (*primer* alelo específico-ALE), com seu terminal 3' complementar a um dos alelos presente no loco, com Ta entre 43 e 48 °C, ao qual é adicionado, no terminal 5', uma cauda de 2 a 3 nucleotídeos aleatórios e não complementares à sequência do DNA molde, para elevar o seu valor de Ta para 53 °C. Cada *primer* informativo forma produtos de amplificação com o tamanho de aproximadamente 100 pb com um dos dois *primers* loco específico.

3.10. Genotipagem de SNP utilizando a metodologia TSPCR

As reações de genotipagem dos polimorfismos selecionados nas populações segregantes RILs, foram realizadas em um volume total de 15 µL, contendo 1U de Platinum®Taq DNA polimerase; Tris-HCl 20 mM pH 8,4; KCl 50 mM; MgCl₂ 1,5 mM; 0,1 µg/µL de soroalbumina bovina; 0,01% tween-20; 1,0% de formamida; 0,2 mM de cada dNTP (dATP, dTTP, dGTP e dCTP); 0,083 µM do *primer* loco específico, 0,83 µM do *primer* informativo e 30 ng de DNA molde. As reações de amplificação foram conduzidas em termociclador Eppendorff (modelo AG-22331), inicialmente a 94 °C por 2 min, seguida de: 10 ciclos de 94 °C por 30 s, 58 °C por 30 s e 72 °C por 60 s; 5 ciclos de 94 °C por 10 s e 45 °C por 30 s; 15 ciclos de 94 °C por 30 s, 53 °C por 30 s e 72 °C por 5 s. Realizadas as reações de amplificação, os produtos gerados foram submetidos à eletroforese em gel de agarose a 2%, contendo brometo de etídeo (0,2 µg/mL), e fotodigitalizados, utilizando o sistema L-PIX (Loccus Biotecnologia).

4. Resultados e Discussão

4.1. Amplificação de fragmentos dos genes candidatos por PCR

Para a obtenção dos fragmentos de interesse de cada gene candidato, foram feitas reações de amplificação por PCR, utilizando DNA total de folhas de soja e os *primers* específicos correspondentes para cada gene candidato. Realizadas as reações de amplificação, os produtos gerados foram submetidos à eletroforese em gel de agarose. Quando mais de um produto de amplificação foram gerados, a reação foi repetida em

temperaturas de anelamento maiores, até se conseguir uma única banda ou descartar definitivamente este par de *primers*.

Foram desenhados três conjuntos de *primers* (R/F) para o gene Dof4, enquanto que para os genes Dof11, MDH, LEC1A e LEC1B foram desenhados quatro conjuntos de *primers*, com objetivo de amplificar e sequenciar todo o gene. Para os genes Dof4 (Glyma17g08950.1) e LEC1A (Glyma07g39820.1) somente um conjunto de *primer* de cada gene, produziram amplicons sem bandas inespecíficas de aproximadamente 500 pb (com aproximadamente 400 pb correspondente a região codificadora do gene). Para os genes Dof11 (Glyma13g40420.1) e LEC1B (Glyma17g00950.1), três conjuntos de *primers* desenhados para cada gene, geraram amplicons sem bandas inespecíficas (cada amplicon variando de 400 a 500 pb). No entanto, todos os conjuntos de *primers* desenhados para o gene MDH (Glyma17g10880.1), produziram amplicons sem amplificação inespecífica.

Para os 15 genes candidatos restantes listados na Tabela 1, as regiões regulatórias (5'UTR e 3'UTR) foram selecionadas para o desenho dos *primers* (R/F). Os *primers* desenhados na região 5'UTR dos transcritos dos genes ABI (Glyma08g47240.1), ARAF (Glyma19g30450.1), CPT (Glyma02g14210.1), G3PDH (Glyma18g01330.1), LPCAT (Glyma05g03510.1), PDAT (Glyma13g16790.1), WRI1 (Glyma15g34770.1), DGAT (Glyma09g07520.1), bem como os *primers* desenhados na região 3'UTR dos transcritos dos genes ADPRI (Glyma05g08700.1), CPT (Glyma12g08720.1), FUS3 (Glyma19g27340.1), PDAT (Glyma17g05910.1), WRI1 (Glyma15g34770.1) e DGAT (Glyma13g16560.1), geraram produtos de amplificação inespecíficos ou, em alguns casos, não geraram amplicons, sendo os *primers* descartados (resultados não mostrados).

Todos os *primers* desenhados nas regiões regulatórias dos transcritos dos genes AMIP (Glyma12g03580.1 e Glyma11g11440.1), CLPPRO (Glyma10g01550.1 e Glyma02g01510.1), ACC (Glyma06g11120.1 e Glyma04g11550.1) e ASIL (Glyma07g04790.1 e Glyma16g01370.1), produziram amplicons com banda única.

Quando apenas o único produto de amplificação esperado foi gerado para cada progenitor, estes foram posteriormente purificados com ExoSap e sequenciados. A figura 1 ilustra a escolha dos produtos de amplificação por PCR em gel de agarose 1,7%. Os *primers* desenhados para o gene CPT geraram duas bandas e foram descartados, enquanto que os *primers* desenhados para os genes ABI3, FUS3 e ADPRI

produziram único produto de amplificação, sendo os produtos de PCR purificados e sequenciados.

Todos os *primers* desenhados para a obtenção dos fragmentos de interesse em cada gene candidato, tiveram a especificidade testada pelo programa *PCR virtual* (iPCR) e pelo banco de dados PHYTOZOME. No entanto, a presença de várias cópias do mesmo gene no genoma da soja, provavelmente possibilitou a geração de produtos inespecíficos para alguns genes candidatos, pois os oligos específicos desenhados para cada gene candidato pode ter anelado em regiões inespecíficas do genoma. De acordo com SCHMUTZ et al. (2010), o genoma da soja passou por dois eventos de duplicações, que ocorreram a cerca de 59 e 13 milhões de anos atrás, resultando em um genoma altamente repetido, com cerca de 75% dos genes presentes em múltiplas cópias. Estes dois eventos de duplicação foram seguidos por numerosos rearranjos cromossômicos.

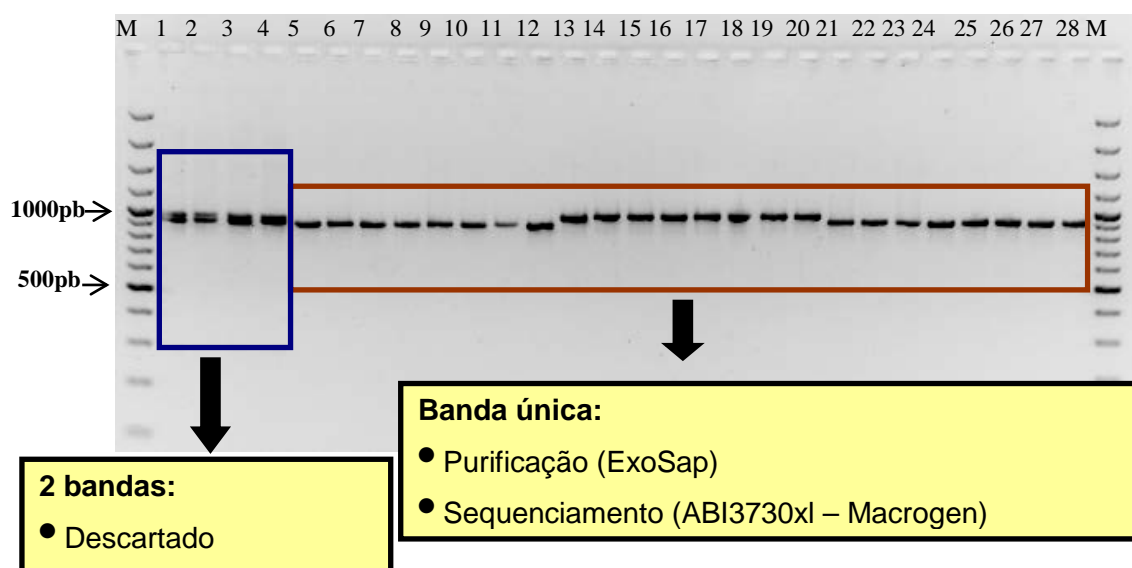


Figura 1. Análise eletroforética em gel de agarose (1,7%), mostrando a escolha dos produtos de amplificação por PCR, que foram purificados e sequenciados. Caneletas de 1 a 4 contêm os fragmentos amplificados a partir dos pares de *primers* desenhados para o gene CPT; canaletas de 5-12, 13-20 e 21-28, fragmentos amplificados a partir dos pares de *primers* desenhados para os genes ABI3, FUS3 e ADPRI. Caneletas M – Marcador de tamanho (*GeneRuler 100 bp DNA Ladder Plus* – Fermentas).

4.2. Sequenciamento e identificação de polimorfismos

Os *primers* desenhados para amplificar regiões específicas de cada transcrito dos 19 genes candidatos, e que produziram apenas o único produto de amplificação para seis

progenitores contrastantes do PMQS, foram purificados e sequenciados utilizando o sequenciador ABI3730xl. As sequências obtidas foram editadas usando os programas SEQUENCER 4.1.4 e CodonCode Aligner, e alinhadas pelo programa ClustalW para identificação dos polimorfismos. Na Figura 2, mostra o alinhamento e a identificação de três SNP e uma INDEL na região 3'UTR do transcrito do gene ADPRI (Glyma19g27340.1), utilizando o programa ClustalW.

A identificação de SNP através de genes candidatos, consiste em sequenciamento direto de indivíduos contrastantes para característica estudada, subsequente alinhamento utilizando ferramentas de bioinformática, para posterior comparação das sequências na procura de SNP, esta metodologia também foi utilizada com sucesso em diversos tipos de estudos, como em humanos (LAI et al., 1998; LEVY et al., 2009), bovinos (CALVO et al., 2006; ABATEPAULO et al., 2008), plantas (VAN et al., 2004; NOVAES et al., 2008), além de outros organismos (LIAO e LEE, 2010).

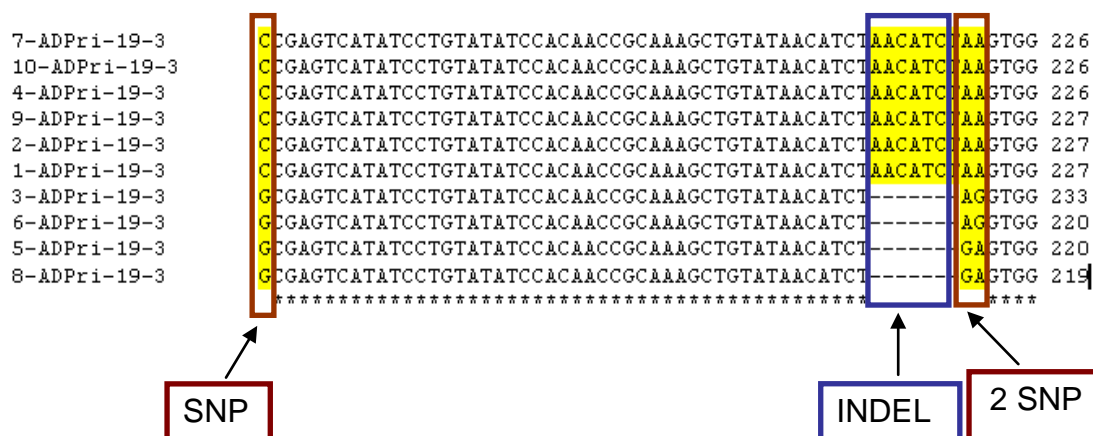


Figura 2. Alinhamento das sequências utilizando programa ClustalW. Mostrando a identificação de três SNP e uma INDEL na região 3'UTR do gene ADPRI (Glyma19g27340.1).

Segundo BROOKES (1999), para que um dado polimorfismo seja considerado SNP, assume-se para um dado loco, que o alelo menos frequente na população tenha uma abundância acima de 1%. No presente trabalho, foram obtidos aproximadamente 31,2 kpb de sequências com boa qualidade, com valor de Phred superior ou igual a 30, sendo identificados 81 SNP e 26 INDELs, sendo a média de um SNP para cada 292 pb (Tabela 4). A frequência habitual de SNP relatados para genomas de plantas é de cerca de 1 SNP a cada 100-300 pb (GUPTA et al., 2001).

Tabela 4. Regiões gênicas sequenciadas de cada transcrito e o respectivo número de polimorfismos encontrados

Genes	Transcritos	Região dos Genes Total de pb compreendidos			Número de polimorfismo	
		5' UTR	Codificante	3' UTR	SNP	INDELS
ABI3	Glyma08g47240.1	1640	-	822	5	0
	Glyma18g38490.1					
ACC	Glyma06g11120.1	780	-	1450	2	1
	Glyma04g11550.1					
ADPRI	Glyma19g27340.1	800	-	790	25	10
AMIP	Glyma12g03580.1	1640	-	1680	14	1
	Glyma11g11440.1					
ARAF	Glyma19g30450.1	849	-	1630	8	3
	Glyma03g27460.2					
ASIL	Glyma07g04790.1	1125	-	1520	0	2
CLPPR	Glyma10g01550.1	830	-	1230	0	0
	Glyma02g01510.1					
CPT	Glyma12g08720.1	735	-	840	1	1
DGAT	Glyma09g07520.1	810	-	820	0	0
Dof4	Glyma17g08950.1	-	490	-	3	1
Dof11	Glyma13g40420.1	310	540	200	1	0
FUS3	Glyma19g27340.1	-	-	1630	3	2
	Glyma16g05480.1					
G3PDH	Glyma18g01330.1			1010	4	1
	Glyma11g37360.1					
LEC1A	Glyma07g39820.1	150	350	-	0	0
LEC1B	Glyma17g00950.1	120	1266	-	2	2
LPCAT	Glyma05g03510.1	780	-	1520	7	2
	Glyma17g14070.1					
MDH	Glyma17g10880.1	110	1206	105	1	0
PDAT	Glyma17g05910.1	810	-	-	5	0
WRI	Glyma08g24420.1	-	-	610	0	0
Total		11489	3852	15857	81	26

Em soja, a frequência de SNP é relativamente baixa em comparação com outras espécies agrícolas cultivadas (ZHU et al., 2003; HYTEN et al., 2006). A baixa frequência de sequências variantes nas cultivares modernas é esperada, devido à baixa variabilidade genética do ancestral selvagem *G. soja* e principalmente, pela perda de 50 % da diversidade genética durante o processo de domesticação da soja (HYTEN et al., 2006).

No trabalho realizado por SCALLOM et al. (1987) com genótipos de soja divergentes, 3543 pb do gene *Gy4* que codifica uma das proteínas de reserva no grão, foram analisados, possibilitando a identificação de três SNP. ZAKHAROVA et al. (1989) compararam 789 pb de sequências de cDNAs do gene que codifica para a subunidade A₃B₄ da proteína glicinina em três cultivares e encontraram dois SNP. ZHU et al. (2003) realizou um estudo com as principais variedades de soja cultivadas nos Estados Unidos, neste trabalho foram sequenciados 66,6 Kpb, sendo 28,7 kpb de sequências codificantes e 37,9 kpb de sequências não-codificantes, sendo identificado 250 SNP, com uma frequência de 1 SNP a cada 270 pb.

4.3. Análise de expressão gênica em Northern Eletrônico

Com o intuito de obter informações sobre a expressão dos genes candidatos selecionados em soja, foi feito um alinhamento por meio do algoritmo BLASTn, das sequências codificadoras dos transcritos codificados pelos 19 genes candidatos selecionados, contra bibliotecas de ESTs derivadas de sete tecidos de soja, depositados no banco de dados do NCBI. A Figura 3 mostra o número de vezes que cada *read* formador de um EST apareceu expresso em cada biblioteca.

Os resultados demonstraram que os transcritos dos genes ADPRI (Glyma19g00200.1, Glyma19g00200.2, Glyma19g00200.3, Glyma19g00200.4 e Glyma05g08700.1) e G3PDH (Glyma18g01330.1, Glyma18g01330.2 e Glyma11g37360.1) foram expressos constitutivamente nas sete bibliotecas testadas. Apesar dos transcritos apresentarem expressão em todas as bibliotecas, o maior número de *hits*, foi detectado nas bibliotecas de folhas, embriões somáticos, cotilédones e planta inteira.

A expressão constitutiva dos genes ADPRI e G3PDH nas sete bibliotecas testadas demonstrando que estes genes podem estar desempenhando um papel fundamental no acúmulo de óleo em sementes soja. WEI et al. (2008) analisaram a expressão diferencial de genes, utilizando sementes de soja com diferentes estágios de maturação de duas cultivares contrastantes para teor de óleo. Neste estudo, os autores verificaram uma alta expressão do gene ADPRI, em todos os estágios de maturação avaliados na cultivar que apresentava alto teor de óleo.

Foi proposto por PERRY et al. (1999) que a disponibilidade de G3P restringe a taxa global de acúmulo de TAG. Em concordância com esse estudo, anos mais tarde

através de uma abordagem transgênica, baseando-se na expressão semente-específica de um gene de levedura que codifica G3PDH em canola, foi demonstrado que o nível de G3P aumentou de três a quatro vezes em sementes em desenvolvimento, resultando em um aumento de 40% no conteúdo final de óleo nas sementes (VIGEOLAS et al., 2007).

Os transcritos dos genes DOF11 (Glyma13g40420.1 e Glyma15g04980.1) e PDAT (Glyma13g16790.1 e Glyma17g05910.1), não foram detectados em nenhuma das bibliotecas testadas. Transcritos dos genes ABI3, ACC, AMIP, ARAF, ASIL, CLPPR, CPT, DGAT, DOF4, FUS3, LEC1A, LEC1B, LPCAT, MDH e WRI, apresentaram expressão somente em algumas bibliotecas.

4.4. Avaliação genética e estatística das populações de mapeamento

4.4.1. Testes de normalidade

As características teor de óleo e proteína nas três populações de RIL's apresentou distribuição aproximadamente normal ($P < 0,01$), como pode ser observado na Tabela 5 e confirmado pelas estimativas não significativas de assimetria, curtose e valor D do teste de Lilliefors. Este fato indica a existência de variação contínua e a possibilidade de que vários genes estejam envolvidos no controle destas características nestas populações, possibilitando o emprego de métodos de estimação de QTLs e testes estatísticos de significância que pressupõe uma distribuição normal dos dados.

4.4.2. Análise de variância

Os resultados das análises de variância dos caracteres teor de óleo e proteína quantificados nas três populações de mapeamento estão apresentadas na Tabela 6. O coeficiente de variação ambiental mostrou-se inferior a 5% para os dois caracteres avaliados nas três populações, indicando uma ótima precisão experimental e que grande parte da variação total observada é consequência da variância genética. Estes valores estão próximos aos encontrados por CHUNG et. al. (2003) para a característica teor de proteína e por RODRIGUES et al. (2010) para as características teor de óleo e proteína.

Nas três populações foi detectada variabilidade genética entre os tratamentos para os caracteres teor de óleo e proteína, condição esta essencial para a detecção de QTLs.

Tabela 5. Estimativas dos valores de curtose, simetria e D para a distribuição dos caracteres teor de óleo e proteína das três populações de RILs avaliadas na safra 2011/2012, em São Gotardo - MG

Populações RILs	A7002 X CD219			SUPREMA X CD01RR8384			PI371611 X CD222		
Teor de proteína									
	Curtose	Simetria	D ⁽¹⁾	Curtose	Simetria	D ⁽¹⁾	Curtose	Simetria	D ⁽¹⁾
Estimativa	3,02	0,35	0,11*	3,813	0,1812	0,0596 ^{ns}	2,74	-0,031	0,054 ^{ns}
Variância	0,035	0,035	-	0,025	0,0251	-	0,034	0,034	-
Valor de t	0,054	1,85	-	2,577	1,144	-	-0,719	-0,171	-
Probabilidade (%)	95,58	6,06	-	0,983	25,18	>1,00% <5,00%	52,073	85,86	-
Teor de Óleo									
Estimativa	3,20	-0,0845 ^{ns}	0,034 ^{ns}	3,44*	-0,386	0,576 ^{ns}	2,69	0,0134	0,0435 ^{ns}
Variância	0,035	0,0347	-	0,025	0,025	-	0,034	0,0335	-
Valor de t	0,53	-0,4536	-	1,41	-2,44	-	-0,87	0,0733	-
Probabilidade (%)	60,03	65,49	-	15,49	1,44	>1,00% <5,00%	61,05	93,963	-

(1) Estatística D para o Teste de Lilliefors.

(*) Significativo a 1% de probabilidade pelo teste t; (ns) – Não significativo pelo teste t.

Tabela 6. Resumo da análise de variância dos caracteres teor de óleo e proteína das três populações de RILs derivada dos cruzamentos A7002 x CD219, SUPREMA x CD01RR8384 e PI371611 x CD222, avaliadas na safra 2011/2012, em São Gotardo - MG

FV	Populações RILs								
	A7002 X CD219			SUPREMA X CD01RR8384			PI371611 X CD222		
	GL	Óleo	Proteína	GL	Óleo	Proteína	GL	Óleo	Proteína
Blocos	2	2,94	2,78	1	0,368	1,20	1	6,96	0,027
Tratamentos	171	1,69**	4,55**	237	1,42**	2,94**	176	1,63**	3,38**
Resíduo	336	0,34	0,65	216	0,34	0,63	172	0,38	1,10
Média	-	21,38	36,95	-	22,58	36,12	-	20,42	38,15
CV(%)	-	2,72	2,19	-	2,59	2,20	-	3,00	2,74
Limite Inferior	-	18,22	33,42	-	19,80	31,87	-	17,59	33,99
Limite Superior	-	24,05	41,57	-	25,66	40,39	-	22,71	41,96

** significativo a 1 % de probabilidade

4.5. Teste com os *primers* TSPCR

A genotipagem dos polimorfismos nas três populações RILs, foram feitas por meio da metodologia TSPCR (HAYDEN et al., 2009). Na Tabela 7 estão listadas as sequencias dos 17 conjuntos de *primers* TSPCR desenhados para cada polimorfismo testado. Com exceção do gene ADPRI (Glyma19g27340.1), onde foram selecionados três polimorfismos para o desenho dos *primers* TSPCR, para os demais genes candidatos foi selecionado um único polimorfismo. Após serem sintetizados, os *primers* TSPCR foram submetidos a reações de PCR em condições reacionais padronizadas, nos 2 progenitores onde se identificou o respectivo polimorfismo por sequenciamento.

Inicialmente, os SNP e INDELS identificados nos genes ADPRI (Glyma19g27340.1), ARAF (Glyma03g27460.2), ABI3 (Glyma18g38490.1), LEC1B (Glyma17g00950.1), PDAT (Glyma17g05910.1), G3PDH (Glyma18g01330.1), FUS3 (Glyma16g05480.1), CPT (Glyma12g08720.1) e DOF4 (Glyma17g08950.1), foram testados nos progenitores (A7002 e CD219), com objetivo de confirmar esses polimorfismos identificados pelo sequenciamento, antes de testá-los na população. Os *primers* de genotipagem TSPCR, desenhados para os polimorfismos presentes nos genes ADPRI (Glyma19g27340.1), ARAF (Glyma03g27460.2) e DOF4 (Glyma17g08950.1), mostraram-se monomórficos nos progenitores. Enquanto que os *primers* TSPCR desenhados para os polimorfismos dos genes CPT (Glyma12g08720.1),

PDAT (Glyma17g05910.1) e G3PDH (Glyma18g01330.1) apresentaram marcação de bandas fracas e/ou bandas inespecíficas. Contudo, os *primers* TSPCR desenhados para testar os polimorfismos dos genes ABI3 (Glyma18g38490.1) e LEC1B (Glyma17g00950.1) mostraram-se polimórficos, sendo esses *primers* posteriormente utilizados na genotipagem da população (Figura 4).

Nos progenitores PI371611 e CD222, foram testados *primers* TSPCR desenhados para os polimorfismos identificados nos genes ADPRI (Glyma19g27340.1), AMIP-5 (Glyma12g03580.1, região 5'UTR), AMIP-3 (Glyma12g03580.1, região 3'UTR), ABI3 (Glyma18g38490.1), DOF4 (Glyma17g08950.1), G3PDH (Glyma18g01330.1) e LEC1B (Glyma17g00950.1). Os *primers* TSPCR desenhados para testar os polimorfismos dos genes ADPRI (Glyma19g27340.1) e DOF4 (Glyma17g08950.1) mostraram-se monomórficos nos progenitores. Para o polimorfismo presente no gene G3PDH (Glyma18g01330.1), os *primers* TSPCR mostraram-se polimórficos, porem produziram bandas inespecíficas. Por outro lado, os *primers* TSPCR desenhados para discriminar os alelos dos genes AMIP-5 (Glyma12g03580.1), AMIP-3 (Glyma12g03580.1), ABI3 (Glyma18g38490.1) e LEC1B (Glyma17g00950.1), foram polimórficos nos progenitores (Figura 4).

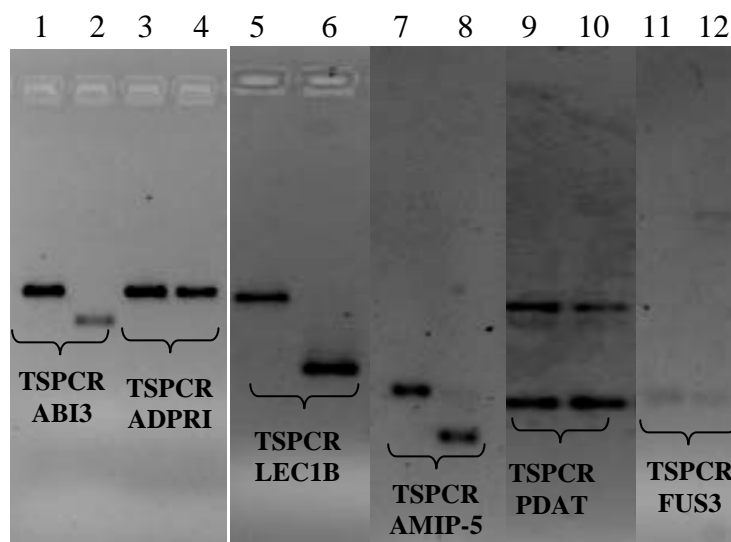


Figura 4. Teste com alguns dos *primers* TSPCR em gel de agarose 2%, com os 2 progenitores onde se identificou o respectivo polimorfismo. Mostrando *primers* TSPCR polimórficos (canaletas 1-2, 5-6 e 7-8), monomórfico (canaletas 3-4), bandas inespecíficas (canaletas 9-10) e bandas fracas (canaletas 11-12). Canaletas 1, 3, 5, 9 e 11 (A7002), canaletas 2, 4, 6, 10 e 12 (CD219); canaletas 7 (PI371611), canaletas 8 (CD222).

Os polimorfismos dos genes ARAF (Glyma03g27460.2), LPCAT (Glyma05g03510.1), LEC1B (Glyma17g00950.1), FUS3 (Glyma16g05480.1) e CPT (Glyma12g08720.1), foram testados com *primers* TSPCR nos progenitores SUPREMA e CD01RR8384. Para os *primers* TSPCR desenhados no polimorfismo do gene CPT (Glyma12g08720.1), não amplificaram, enquanto para o polimorfismo do gene FUS3 (Glyma16g05480.1), os *primers* TSPCR foram polimórficos, no entanto produziram amplicons fracos (Figura 4). Os *primers* TSPCR desenhados para testar os polimorfismos dos genes ARAF (Glyma03g27460.2), LPCAT (Glyma05g03510.1) e LEC1B (Glyma17g00950.1), mostraram-se polimórficos nos progenitores.

Os *primers* TSPCR que se mostraram monomórficos foram redesenhados em outras regiões polimórficas do mesmo gene candidato, no entanto os *primers* se mostraram novamente monomórficos, sendo assim descartados. Enquanto que para os *primers* TSPCR que produziram bandas inespecíficas e/ou marcação de bandas fracas, as condições reacionais foram alteradas com intuito de otimizar a produção de bandas específicas e/ou aumentar a resolução das bandas. Mas essas alterações muitas vezes aumentaram o número de bandas inespecíficas e/ou os *primers* não geraram produtos de amplificação (resultados não mostrados).

Tabela 7. Gene candidato (GC), cromossomos, regiões gênicas (Exon, 5'UTR e 3'UTR), alelos de SNP e sequências de *primers* TSPCR para 17 SNP investigados

GC	Cromossomos	Regiões gênicas	Alelos SNP		Primers LE	Primers ALE ^{a,b}	Alelo de SNP investigado
DOF	17	CDS	T/A	Fwd	GAAGAAGTACGTTGTTGCAGGG	<u>GG</u> ATGTTCACTCAACAACAT	T
				Rev	CGCCGCCTTTAGTCCAATAG		
DOF	13	CDS	C/-	Fwd	CCGCCACAAGAAAAATCAAA	<u>GCG</u> AATACAAGTAAAGAAAGTG	C
				Rev	TGAGGAGCTAATTAAAAAACGGAG		
LEC1B	17	CDS	T/C	Fwd	GCCTGAACGTTTTTGACAGG	<u>CG</u> ATGATCCTAATCACGTTA	T
				Rev	TTTGCATCGTCCGAGATTTT		
ABI3	18	5'UTR	A/C	Fwd	ATCCTATGAAGGGATGCAACA	<u>CGG</u> TATGATTATCATTGTGTATAT	T
				Rev	GCGTTTTGATCATCTTCGAGT		
ADPRI	19	3'UTR	A/G	Fwd	GGGATGACACATCAGTCGGTTA	<u>CG</u> TTATCAATGTTTGGAAAA	A
				Rev	TCCATGCGTTTTGTTCTTGAAA		
ADPRI	19	3'UTR	C/G	Fwd	TTGGCTTGAAAAGTAACATCCTG	<u>CG</u> TTGATTGATACCTTTATTC	C
				Rev	TATACAGCTTTGCGGTTGTGG		
ADPRI	19	3'UTR	A/G	Fwd	CCGAACATTAACCCTATTTTTATGG	<u>G</u> CTTATCAATGTTTGGAAAA	A
				Rev	CATGCGTTTTGTTCTTGAAAAGT		
FUS3	16	3'UTR	A/C	Fwd	TTCCATTACAAGACAAAACCTGAA	<u>CG</u> AATTTTAGAGTACTATCATGGT	A
				Rev	GGGGAATTATTTGTATGTTTGGGA		
ARAF	3	3'UTR	T/G	Fwd	TGACACCTGGTCTACTTTGCC	<u>GGG</u> GATAAATGTTATTTACAGCA	T
				Rev	CCCACGTTGTTCTAAGTCAAAAA		
ARAF	3	5'UTR	A/T	Fwd	CCCCCTTCCTTAACTCCTTATT	<u>GG</u> ACTCTGATTCTGTTACTGTTA	A
				Rev	ACCCACCCACACCATGTTAC		
AMIP	12	3'UTR	T/C	Fwd	AGTTAATCCAGATGTTTCGGGAAA	<u>CCG</u> TATTTACCGAATTTGT	T

				Rev	CAGCGAGAAAGAACTTGGCTAT		
AMIP	12	5'UTR	A/T	Fwd	CATCAGATCAATAACGTGATCGAA	<u>GGGCTATAGGAAAAGATTGA</u>	A
				Rev	TTGCCCATTAACCTTTGAAAAT		
G3PDH	18	3'UTR	T/C	Fwd	TCCTCTCGTAAGGGACGATTG	<u>AGCCTTTTATAATCAAATAGTTT</u>	T
				Rev	TTCTTCAACGAACGTAAGAGGAAA		
CPT	12	5'UTR	T/-	Fwd	TGGGATGAAAAGTGAAAGAAAAGA	<u>CGGTGAGATTTTTTTTTTTTT</u>	T
				Rev	TGGAGAATATCTTGAGTTTGCTCAT		
CPT	12	3'UTR	T/C	Fwd	TCTGCTAATGCACTCCCTGTTAG	<u>GGCTATTAATTTGTAAAGGAATT</u>	T
				Rev	TTAAGAGTGTTGATGGGATAGTTGA		
LPCAT	5	3'UTR	A/G	Fwd	TGGGCGCTTTCATACTCTTG	<u>GCCAGATACTCTGATGTATGA</u>	A
				Rev	AAGCACTGTCCCAGACATTTCA		
PDAT	17	5'UTR	C/T	Fwd	GCTTGTCAGACATGAAACGTCC	<u>CGATCAGATAGATCATTTAAGACT</u>	T
				Rev	GGCAATTAAGCGGGAATGAA		

^a Um dos alelos de SNP complementar ao primer alelo específico (ALE) produz o alelo de SNP referência, produto de PCR menor, enquanto que o alelo sem complementaridade produz o alelo de SNP alternativo "produto de PCR maior. Amostras heterozigotas produzem ambos os produtos de genotipagem TSPCR

^b A sequência 5'-arbitrária de nucleotídeos, indicado em negrito e sublinhado, não é complementar ao DNA alvo.

ALE- primer alelo específico

LE- primer locus específico

4.6. Genotipagem das populações de RILs e teste de segregação

Os *primers* TSPCR que se mostraram polimórficos nos respectivos progenitores foram utilizados na genotipagem das populações segregantes (RILs) do PMQS, desenvolvidas para mapear QTLs para alto teor de óleo e/ou proteína no grão.

Assim, os *primers* TSPCR desenhados nos genes AMIP-5 (Glyma12g03580.1), AMIP-3 (Glyma12g03580.1), ABI3 (Glyma18g38490.1) e LEC1B (Glyma17g00950.1), foram amplificados em 176 progênies F₆ da população PI371611 x CD222. Enquanto que para população CD01RR8311 x SUPREMA, foram genotipadas 236 progênies na geração F₆, com os *primers* TSPCR desenhados nos genes ARAF (Glyma03g27460.2), LPCAT (Glyma05g03510.1) e LEC1B (Glyma17g00950.1).

Dois conjuntos de *primers* TSPCR que se mostraram polimórficos nos genitores foram selecionados para genotipar as progênies da população de RILs A7002 x CD219. Esses *primers* TSPCR desenhados para detectar o polimorfismo presente nos genes ABI3 (Glyma18g38490.1) e LEC1B (Glyma17g00950.1) foram amplificados em 170 progênies e nos genitores dessa população. Os perfis polimórficos para o gene ABI3 nos genótipos A7002, CD219 e em 24 progênies F₆ são apresentados na Figura 5.

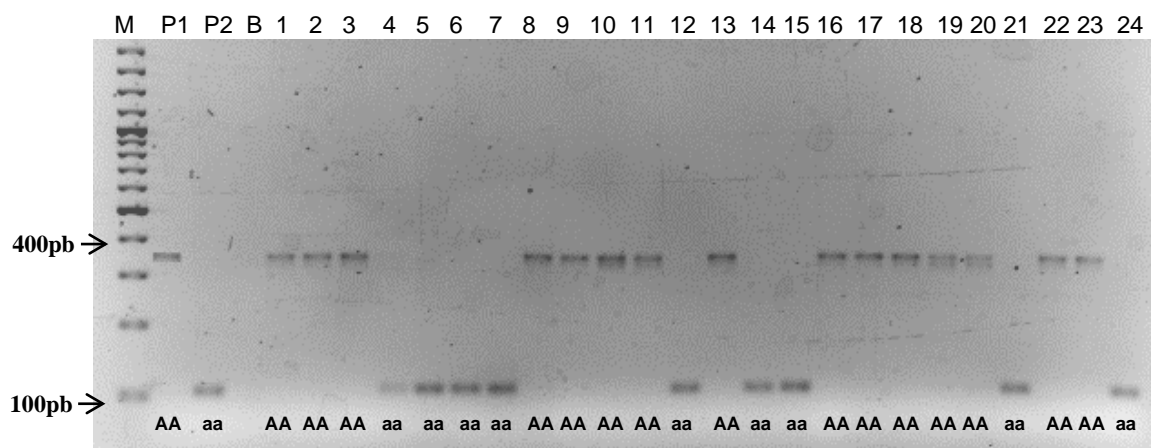


Figura 5. Genotipagem dos indivíduos utilizando *primers* TSPCR desenhado no polimorfismo (A/C) do gene candidato ABI3, em 24 RILs F₆ (A7002 x CD219). P1(A7002), P2 (CD219), B (branco), Canaletas de 1 a 24 (progênies) e canaleta M – Marcador de tamanho (*GeneRuler 100 bp DNA Ladder Plus* – Fermentas). Note que o marcador ABI produziu dois amplicons, o primeiro amplicom de 350 pb (produto de amplificação alelo-alternativo - AA) e um segundo amplicom de aproximadamente 100 pb (produto de amplificação alelo-específico - aa).

A segregação de cada marcador foi testada por meio do teste qui-quadrado ($P < 0,05$) utilizando o critério de Proteção Bonferroni. Todos os marcadores genotipados nas três populações de mapeamento seguiram o padrão esperado de segregação 1:1 (Tabela 8).

Tabela 8. Teste de segregação dos marcadores moleculares genotipados nas populações RILs. Classes genóticas, hipótese, valor do teste de Qui-quadrado e probabilidade para os marcadores

Populações RILs	A7002 X CD219						SUPREMA X CD01RR8384						PI371611 X CD222					
	Classes			Hipotese	χ^2	Probabilidade	Classes			Hipotese	χ^2	Probabilidade	Classes			Hipotese	χ^2	Probabilidade
Marcadores	0	1	2				0	1	2				0	1	2			
ABI3	86		78	1:1	0,39	53,22 ^{ns}	-	-	-	-	-	-	86	8	82	1:1	0,095	75,76 ^{ns}
LEC1B	80		88	1:1	0,38	53,71 ^{ns}	-	-	-	-	-	-	70	3	94	1:1	4,53	10,36 ^{ns}
AMIP-3	-		-	-	-	-	-	-	-	-	-	-	87	3	86	1:1	0,006	93,94 ^{ns}
AMIP-5	-		-	-	-	-	-	-	-	-	-	-	85	1	90	1:1	0,143	70,55 ^{ns}
ARAF	-		-	-	-	-	117	-	113	1:1	0,07	79,20 ^{ns}	-	-	-	-	-	-
LPCAT	-		-	-	-	-	99	3	116	1:1	1,47	47,88 ^{ns}	-	-	-	-	-	-

Nível crítico: 10%

Número total de testes: 20

Critério: Proteção Bonferroni

Número de distorções: 0

4.7. Análise de associação

A análise de associação marcador-característica foi determinada pela análise de marca simples, por meio de regressão linear (Tabela 9 e 10). Nenhum dos marcadores SNP genotipados nas populações de RILs (CD01RR8384 x SUPREMA e PI371611 x CD222) apresentaram associações significativas com os teores de óleo e proteína. Apenas o marcador do gene ABI3 (Glyma18g38490.1), apresentou associação significativa para teor de proteína na população de mapeamento A7002 e CD219. Esse marcador explica 4,7 % da variação do fenótipo determinado pelo R².

Tabela 9. Análise associativa dos marcadores SNP identificados em genes que possivelmente controlam teor protéico, avaliadas em três populações de RIL's pela Regressão Linear

Populações RILs	Proteína Total								
	A7002 X CD219			SUPREMA X CD01RR8384			PI371611 X CD222		
Marcador	Prob F	Média	R ² _%	Prob F	Média	R ² _%	Prob F	Média	R ² _%
ABI3	0,0051 **	36,95	4,74	-	-	-	1	38,14	0,009
LEC1B	1	36,95	0,030	-	-	-	1	38,15	0,0524
AMIP-3	-	-	-	-	-	-	1	38,14	0,0001
AMIP-5	-	-	-	-	-	-	1	38,15	0,1714
ARAF	-	-	-	0,0684	36,10	1,44	-	-	-
LPCAT	-	-	-	1	36,07	0,09	-	-	-

** significativo a 1 % de probabilidade

Tabela 10. Análise associativa dos marcadores SNP identificados em genes que possivelmente controlam teor de óleo, avaliadas em três populações de RIL's pela Regressão Linear

Populações RILs	Óleo								
	A7002 X CD219			SUPREMA X CD01RR8384			PI371611 X CD222		
Marcador	Prob F	Média	R ² _%	Prob F	Média	R ² _%	Prob F	Média	R ² _%
ABI3	1	21,38	0,3025	-	-	-	1	20,42	0,0005
LEC1B	1	21,37	0,2038	-	-	-	1	20,43	0,084
AMIP-3	-	-	-	-	-	-	1	20,42	0,0084
AMIP-5	-	-	-	-	-	-	1	20,42	0,4891
ARAF	-	-	-	1	22,58	0,169	-	-	-
LPCAT	-	-	-	1	22,59	0,0503	-	-	-

** significativo a 1 % de probabilidade

O polimorfismo (A/C) do gene candidato ABI3 (Glyma18g38490.1) que apresentou associação com teor de proteína, foi identificado na região 5'UTR do transcrito Glyma18g38490.1. O gene ABI3 faz parte da rede de reguladores mestres do processo de maturação da semente, denominada AFL, que engloba os fatores

transcricionais ABI3, FUS3 e LEC2 (BAUD e LEPINIEC, 2010). Os membros da rede AFL, atuam em conjunto com LEC1, e exibem padrões de expressão parcialmente sobrepostos, que incluem regulação cruzada e *feedback* complexo (LOTAN et al., 1998). Foi demonstrado que a função de LEC2 é “*up-regulated*” a expressão de FUS3 e ABI3. Uma mutação em LEC2 foi responsável por uma diminuição drástica na expressão dos fenótipos ABI3 e FUS3, porém, LEC2 pode ser resgatado pela expressão constitutiva de ABI3 ou FUS3 (TO et al., 2006).

Segundo WANG et al. (2007), a principal função do fator de transcrição FUS3 é de regular a expressão de genes envolvidos na biossíntese de ácidos graxos, enquanto que LEC1, LEC2 e principalmente ABI3 estão envolvidos na regulação de genes relacionados com o armazenamento de proteína.

HA & BOERMA (2008) em estudo do gene *rca3* que confere às plantas de soja resistência ao fungo *Cercospora sojina* Hara, identificaram um SNP (A/G) posicionado no nucleotídeo 209 do gene que está 100% correlacionado com o fenótipo. A associação marcador-característica foi feita utilizando uma população de 96 plantas F₂; utilizando a metodologia de genotipagem “*Melting Curve Assay*”, a qual utiliza sondas fluorescentes alelo-específicas para discriminação dos alelos. SKONECZKA et al. (2009) identificou uma deleção de 3 pb no gene galactosiltransferase do genótipo PI200508 que apresenta alto conteúdo de sacarose e conteúdo reduzido de estaquiose. A associação do marcador SNP com os fenótipos foi feita utilizando 60 e 82 progênies F₂ derivadas dos cruzamentos PI87013 × PI200508 e PI243545 × PI200508. Essa mutação associou significativamente com as características estudadas, explicando 88 a 94% da variação fenotípica para o conteúdo de estaquiose e 76% para o teor de sacarose nas sementes.

Apesar do presente trabalho, ter encontrado apenas um marcador associado com teor de proteína, é importante enfatizar, que o conteúdo de proteína e óleo em grãos de soja, apresenta controle genético quantitativo, o qual envolve muitos locos, onde normalmente apenas genes de pequeno efeito influenciam essas características. O que dificulta localizar as regiões cromossômicas (QTLs) responsáveis pela variação fenotípica.

5. Conclusões

- Os *primers* desenhados em regiões específicas dos transcritos dos 19 genes candidatos selecionados, que de acordo com a literatura estão diretamente relacionados com acúmulo de óleo e/ou proteína no grão de soja, possibilitou a identificação de 81 SNP e 26 INDELS em 31,2 Kpb de sequências, sendo a média de um SNP para cada 292 pb.
- A análise por northern eletrônico permitiu avaliar a expressão das sequências codificadoras dos transcritos codificados pelos 19 genes candidatos, em sete tecidos de soja e identificar padrões importantes do ponto de vista biotecnológico. Como a expressão constitutiva de todos os transcritos selecionados dos genes ADPRI e G3PDH nas sete bibliotecas testadas, bem como a expressão órgão-específica dos transcritos dos genes ABI3, ACC, AMIP, ARAF, ASIL, CLPPR, CPT, DGAT, DOF4, FUS3, LEC1A, LEC1B, LPCAT, MDH e WRI.
- Os caracteres teor de óleo e proteína avaliados nas três populações de RILs apresentaram distribuição normal, bem como ótima precisão experimental com coeficiente de variação ambiental inferior a 5%. Análise de variância mostrou variabilidade genética entre os tratamentos, condição esta essencial para a detecção de QTLs.
- Dos 17 conjuntos de *primers* TSPCR desenhados para genotipar os polimorfismos nas três populações RILs, somente os *primers* TSPCR desenhados no polimorfismo dos genes AMIP-5 (Glyma12g03580.1), AMIP-3 (Glyma12g03580.1), ABI3 (Glyma18g38490.1), LEC1B (Glyma17g00950.1), ARAF (Glyma03g27460.2), LPCAT (Glyma05g03510.1), mostraram-se polimórficos sem bandas inespecíficas e/ou marcação de bandas fracas nos respectivos progenitores testados.
- Os marcadores genotipados nas três populações de mapeamento seguiram o padrão esperado de segregação. Análise de associação revelou que apenas o marcador do gene ABI3 (Glyma18g38490.1), apresentou associação significativa para teor de proteína na população de mapeamento A7002 e CD219, explicando 4,7 % da variação do fenótipo.
- O polimorfismo dos genes AMIP-5, AMIP-3, LEC1B, ARAF e LPCAT não apresentaram associação com nenhuma das características em estudo. No

entanto, esses polimorfismos poderão ser testados em outras populações do PMQS, com “*background*” genético diferente das utilizadas no presente trabalho, o que possivelmente possibilitará encontrar novas associações.

6. Referências Bibliográficas

- ABATEPAULO, A.R.; CAETANO, A.R.; MENDES, J.R. Detection of SNPs in bovine immune-response genes that may mediate resistance to the cattle tick *Rhipicephalus* (Boophilus) microplus. **Animal Genetics**; v.39, p.328-329, 2008.
- BAUD, S.; GRAHAM, I.A. A spatiotemporal analysis of enzymatic activities associated with carbon metabolism in wild-type and mutant embryos of *Arabidopsis* using in situ histochemistry. **The Plant Journal**; v.46, p.155-69, 2006.
- BAUD, S.; SANTOS, M.M.; TO, A.; HARSOËT LEPINIEC, L.; DUBREUCQ, B. WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in *Arabidopsis*. **The Plant Journal**; v.50, p.825-838. 2007.
- BAUD, S.; LEPINIEC, L. Physiological and developmental regulation of seed oil production. **Progress in lipid research**; v. 49, p. 235-249. 2010.
- BONATO, E.R.; BERTAGNOLLI, P.F.; LANGE, C.E.; RUBIN, S.A.L. Teor de óleo e de proteína em genótipos de soja desenvolvidos após 1990. **Pesquisa Agropecuária Brasileira**; v.35, p.2391-2398, 2000.
- BRIM, C.A. A modified pedigree method of selection in soybeans. **Crop Science**; v.6, p.220, 1966.
- BROOKES, A.J. The essence of SNPs. **Gene**; v.234, p.177-86, 1999.
- CALVO JH, MARTÍNEZ-ROYO A, SILVERI L, et al, Isolation, mapping and identification of SNPs for four genes (ACP6, CGN, ANXA9, SLC27A3) from a bovine QTL region on BTA3. **Cytogenetics and Genome Research**; v.114, n.1, p.39-43, 2006.
- CHUNG, J.; BABKA, H.L.; GRAEF, G.L; STASWICK, P.E.; LEE, D.J.; CREGAN, P.B; SHOEMAKER, R.C.; SPEECH, J.E. The seed protein, oil, and yield QTL on soybean linkage group I. **Crop Science**; v.43, p.1053-1067, 2003.
- CRUZ, C.D.; SCHUSTER, I. GQMOL: aplicativo computacional para análise de dados moleculares e de suas associações com caracteres quantitativos. Versão 2.1. Viçosa: UFV, 2004. Disponível em: <<http://www.ufv.br/dbg/gqmol/gqmol.htm>>. Acesso em: 13 jan. 2005.
- DOYLE, J.J.; DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**; v.12, p.13-15, 1990.
- FAWCETT, T.; SIMON, W.J.; SWINHOE, R.; SHANKLIN, J.; NISHIDA, I.; CHRISTIE, W.W. Expression of mRNA and steady-state levels of protein isoforms of

enoyl-ACP reductases from *Brassica napus*. **Plant Molecular Biology**; v.26, p.155–63, 1994.

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, p.25, 2001.

HA, B.K. & BOERMA, H.R. High-throughput SNP genotyping by melting Curve analysis for resistance to southern root-knot nematode and Frogeye leaf spot in soybean. **Journal Crop Science Biotechnology**; v.11, n.2, p.91-100, 2008.

HAYDEN, M.J.; TABONE, T.; MATHE, D.E. Development and assessment of simple PCR markers for SNP genotyping in barley. **Theoretical and Applied Genetics**; v.119, p.939–951, 2009.

HYTEN, L.D.; SONG, Q.; ZHU, Y.; CHOI, I.-Y.; NELSON, L.R.; COSTA, M.J.; SPECHT, E.J.; SHOEMAKER, R.C.; CREGAN, B.P. Impact of genetic bottlenecks on soybean genome diversity. **Proceedings of the National Academy of Sciences USA**; v.103, p.16666–16671, 2006.

IMSANDE, J.; BERKEMEYER, M.; SCHEIBE, R.; SCHUMANN, U.; GIETL, C.; PALMER, R. G. A. Soybean Plastid-Targeted NADH-Malate Dehydrogenase: Cloning and Expression Analyses. **American Journal of Botany**; v.88, p.2136–2142, 2001.

JOHNSTON, M.L.; LUETHY, M.H.; MIERNYK, J.A.; RANDALL, D.D. Cloning and molecular analyses of the *Arabidopsis thaliana* plastid pyruvate dehydrogenase subunits. **Biochim Biophys Acta**; v.1321, p.200–6, 1997.

LAI, E.; RILEY, J.; PURVIS, I. A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. **Genomics**; v.54, p. 31-38, 1998.

LEVY, D.E.; GEORG, B.; RICE, K.; VERWOERT, G.C.; LAUNER, L.J.; DEGHAN, A.; GLAZER, N.L.; MORRISON, A.C.; JOHNSON, A.D.; ASPELUND, T.; AULCHENKO, Y.; LUMLEY, T.; KÖTTGEN, A.; VASAN, R.S.; RIVADENEIRA, F.; EIRIKSDOTTIR, G.; GUO, X.; ARKING, D.E.; MITCHELL, G.F.; MATTACE-RASO, F.U.S.; SMITH, A.V.; TAYLOR, K.; SCHARPF, R.B.; HWANG, S.; SIJBRANDS, E.J.G.; BIS, J.; HARRIS, T.B.; GANESH, S.K.; O'DONNELL, C.J.; HOFMAN, ALBERT ; ROTTER, JEROME I ; CORESH, JOSEF ; BENJAMIN, EMELIA J ; UITTERLINDEN, ANDRÉ G.; HEISS, G.; FOX, C.S.; WITTEMAN, J.C.M ; BOERWINKLE, E.; WANG, T.J.; GUDNASON, V.; LARSON, M.G.; CHAKRAVARTI, A.; PSATY, B.M.; VAN DUIJN, C.M. Genome-wide association study of blood pressure and hypertension. **Nature Genetics**; v.41, p. 677-687, 2009.

LIAO, P.; LEE, K.H. From SNPs to functional polymorphism: The insight into biotechnology applications. **Biochemical Engineering Journal**; p.149-158, 2010.

LIN, M.; BEHAL, R.; OLIVER, D.J. Disruption of *pLE2*, the gene for the E2 subunit of the plastid pyruvate dehydrogenase complex, in *Arabidopsis* causes an early embryo lethal phenotype. **Plant Molecular Biology**; v.52, p.865–72, 2003.

LOTAN, T.; OHTO, M.; YEE, K.M.; WEST, M.A.; LO, R.; KWONG, R.W. *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. **Cell**; v.93, p.1195–205, 1998.

MAHMOUD, A.A.; NATARAJAN, S.S.; BENNETT, J.O.; MAWHINNEY, T.P.; WIEBOLD, W.J.; KRISHNAN, H.B. Effect of six decades of selective breeding on soybean protein composition and quality: A biochemical and molecular analysis. **Journal of Agricultural and Food Chemistry**; v.54, p. 3916-3922, 2006.

NOVAES, E.; DROST, D.R.; FARMERIE, W.G, et al, High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. **BMC Genomics**; v.9, p.320, 2008.

O'HARA, P.; SLABAS, A.R.; FAWCETT, T. Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. **Plant Physiology**; v.129, p. 310–20, 2002.

PERRY, H.J.; BLIGNY, R.; GOUT E, HARWOOD, J.L. Changes in Kennedy pathway intermediates associated with increased triacylglycerol synthesis in oil-seed rape. **Phytochemistry**; v.52, p.799–804, 1999.

RODRIGUES, J.I.S.; MIRANDA, F.D.; FERREIRA, A.; BORGES, L.L.; FERREIRA, M.F.S.; GOOD-GOD, P.I.V.; PIOVESAN, N.D.; BARROS, E.G.; CRUZ, C.D.; MOREIRA, M.A. Mapeamento de QTL para conteúdos de proteína e óleo em soja. **Pesquisa Agropecuária Brasileira**; v.45, p. 472-480, 2010.

SCALLON, B. J.; DICKINSON, C. D.; NIELSEN, N. C. Characterization of a null-allele for the *Gy4* glycinin gene from soybean. **Molecular Gene and Genetic**; v.208; p.107–113, 1987.

SCHMUTZ, J.; CANNON, S.B.; SCHLUETER, J.; MA, J.; MITROS, T.; NELSON, W.; HYTEN, D.L.; SONG, Q.; THELEN, J.J.; CHENG, J. Genome sequence of the paleopolyploid soybean. **Nature**; v.463, p. 178-183. 2010.

SKONECZKA, J.A.; SAGHAI MAROOF ,M.A.; SHANG, C.; BUSS, G.R. Identification of Candidate Gene Mutation Associated With Low Stachyose Phenotype in Soybean Line PI200508. **Crop Science**; v.49, p. 247–255, 2009.

TO, A.; VALON, C.; SAVINO, G.; GUILLEMINOT, J.; DEVIC, M.; GIRAUDAT, J. A network of local and redundant gene regulation governs *Arabidopsis* seed maturation. **Plant Cell**; v.18, p.1642–51, 2006.

USDA. United States Department of Agriculture. Disponível em <<http://www.fas.usda.gov/oilseeds/current/default.asp>>. Acesso em, 21 Fev. 2012.

- VAN, K, et al, Discovery of single nucleotide polymorphisms in soybean using primers designed from ESTs. **Euphytica**; v.139, p.147-155, 2004.
- VERDIER, J.; THOMPSON, R.D. Transcriptional Regulation of Storage Protein Synthesis During Dicotyledon Seed Filling. **Plant and Cell Physiology**; v.49, p. 1263–1271, 2008.
- VIGEOLAS, H.; WALDECK, P.; ZANK, T.; GEIGENBERGER, P. Increasing seed oil content in oil-seed rape (*Brassica napus L.*) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. **Plant Biotechnology Journal**; v.5, p. 431–41, 2007.
- WANG, H. W.; ZHANG, B.; HAO, Y. J.; HUANG, J.; TIAN, A. G.; LIAO, Y.; ZHANG, J. S.; CHEN, S. Y. The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic *Arabidopsis* plants. **The Plant Journal**; v.52, p.716-129, 2007.
- WEI, W.H.; CHEN, B.; YAN, X.H.; WANG, L.J.; ZHANG, H.F.; CHENG, J.P.; ZHOU, X.A.; AI-HUA SHA, A.H.; SHEN, H. Identification of differentially expressed genes in soybean seeds differing in oil content. **Plant Science**; v.175: p. 663–673, 2008.
- WEIGELT, K.; KÜSTER, H.; RADCHUK, R.; MÜLLER, M.; WEICHERT, H.; FAIT, A.; FERNIE, A.R.; SAALBACH, I.; WEBER, H. Increasing amino acid supply in pea embryos reveals specific interactions of N and C metabolism and highlights the importance of mitochondrial metabolism. **The Plant Journal**; v.55, p. 909-26, 2008.
- WESELAKE, R.J.; TAYLOR, D.C.; RAHMAN, M.H.; SHAH, S.; LAROCHE, A.; MCVETTY, P.B.E.; HARWOOD, J.O.H.N. L. Increasing the flow of carbon into seed oil. **Biotechnology Advances**; v. 27, p. 866-78, 2009.
- ZAKHAROVA, E. S.; EPISHIN, S. M.; VINETSKI, Y. P. An attempt to elucidate the origin of cultivated soybean via comparison of nucleotide sequences encoding glycinin B4 polypeptide of cultivated soybean, *Glycine max*, and its presumed wild progenitor, *Glycine soja*. **Theoretical and Applied Genetics**; v.78; p.852–856, 1989.
- ZHU, Y. L.; SONG, Q. J.; HYTEN, D. L.; VAN TASSELL, C. P.; MATUKUMALLI, L. K. Single-nucleotide polymorphisms in soybean. **Genetics**; v.163; p.1123-1134, 2003.
- ZHU, Y.L.; SONG, Q.J.; HYTEN, D.L.; VANTASSELL, C.P.; MATUKUMALLI, L.K.; GRIMM, D.R.; HYATT, S.M.; FICKUS, E.W.; YOUNG, N.D.; CREGAN, P.B.. Single-nucleotide polymorphisms in soybean. **Genetics**; v.163, p.1123 -1134, 2003.

CAPÍTULO 2

Validação de marcadores SNP e utilização da metodologia de genotipagem TSPCR na discriminação de alelos de genes associados ao conteúdo de ácidos graxos em óleo de soja

1. Resumo

Várias metodologias para a genotipagem de SNP estão disponíveis na literatura, entretanto, a maioria delas requer equipamentos e/ou reagentes de custo elevado, fora da realidade da maioria dos grupos de pesquisa e das empresas de melhoramento genético. A metodologia de genotipagem TSPCR (*Temperature-Switch Polymerase Chain Reaction*) utiliza a amplificação por PCR baseada na hibridização alelo-específica e a discriminação dos alelos é feita em gel de agarose, o que a torna de fácil implementação e de baixo custo. O objetivo do trabalho foi identificar e validar SNP em regiões gênicas associadas ao conteúdo de ácidos graxos em genótipos de soja e utilizar a metodologia de genotipagem TSPCR para discriminar alelos dos genes candidatos. Foram selecionados quatro genes candidatos e os genes FAD3B e FAD3C que codificam ω -3-dessaturases em sementes de soja. O sequenciamento de regiões dos seis genes permitiram identificar 25 SNP e 3 INDELS. Os *primers* TSPCR desenhados para discriminar os alelos de SNP nos genes candidatos ARAF e PDAT foram amplificados em 259 RILs do cruzamento FA22 x CD219 (contrastantes para teor de ácido oleico). Os *primers* TSPCR desenhados no polimorfismo dos genes FAD3B, FAD3C e ABI3 foram amplificados em 185 progênies F₂ do cruzamento A29 x Tucunaré (contrastantes para teor de ácido linolênico). O polimorfismo dos genes ABI3, FAD3B e FAD3C associaram significativamente com o teor de um ou mais ácido graxo. A metodologia de genotipagem TSPCR mostrou ser eficaz na discriminação dos alelos de SNP nas duas populações. Os marcadores SNP dos genes ABI3, FAD3B e FAD3C poderão ser usados para selecionar linhagens de soja com baixíssimos teores de ácido linolênico e/ou altos teores de ácido oleico.

2. Introdução

O óleo de soja é o segundo no ranking mundial entre os óleos vegetais, em termos de produção e consumo doméstico, ficando atrás apenas do óleo de palma, sendo que cerca de 28% do óleo vegetal consumido no mundo advém da soja (USDA, 2012).

A composição do óleo de soja varia em função do genótipo da planta e do ambiente em que ela é cultivada. Os triacilgliceróis (TAG), moléculas compostas do glicerol esterificado a três ácidos graxos, representam a classe de lipídeos mais abundante nas sementes dessa leguminosa. Os principais ácidos graxos constituintes dos

TAG são: ácido palmítico (16:0) - 13%; ácido esteárico (18:0) - 4%; ácido oléico (18:1 Δ^9) - 18%; ácido linoléico (18:2 $\Delta^{9,12}$) - 55% e ácido linolênico (18:3 $\Delta^{9,12,15}$) - 10% (GRAEF et al., 2009).

A composição e a distribuição de ácidos graxos na molécula de TGA estão diretamente relacionadas com a qualidade do óleo, valor nutricional, sabor e odor, além das propriedades físico-químicas, como estabilidade oxidativa e ponto de fusão. Os elevados teores de ácido linoléico e linolênico são responsáveis pela alta instabilidade oxidativa e pelo sabor e odor do óleo de soja. Isso ocorre porque os ácidos graxos polinsaturados são mais suscetíveis às hidroxilações, resultando em cetonas e aldeídos voláteis que alteram o sabor e odor do produto (AKER, 1983; DUNTON et al., 1951; CARDINAL et al., 2011). A utilização de óleo de soja com alto teor de ácidos graxos insaturados para a produção de biodiesel leva à produção de um combustível com qualidade inferior, com baixo índice de cetano e maior instabilidade oxidativa (KNOTHE, 2007; RAMOS, 2009).

O conteúdo de ácidos graxos monoinsaturados e poliinsaturados no óleo de soja apresenta herança quantitativa, sendo condicionado por vários genes e possui interação significativa com o ambiente. Com o intuito de tornar mais eficiente os trabalhos de melhoramento, as metodologias de marcadores de DNA têm sido utilizadas na seleção de genótipos adequados. A identificação de marcadores SNP (*Single Nucleotide Polymorphism*) em genes candidatos, envolvidos tanto na regulação ou na via de biossíntese de ácidos graxos, pode acelerar a obtenção de variedades de soja com a composição de ácidos graxos modificada na fração óleo.

Os principais genes responsáveis pela biossíntese de ácidos graxos polinsaturados de reserva são o GmFAD3A, que codifica uma ω -3-dessaturase (BILYEU et al., 2003), e o FAD2-1, que codifica uma ω -6-dessaturase (HEPPARD et al., 1996). No entanto, o número exato de genes que controla a via de biossíntese de ácidos graxos insaturados ainda não é conhecido e os seus efeitos individuais e as possíveis interações não estão bem caracterizados.

Os genes FAD3A (Glyma14g37350), FAD3B (Glyma02g39230) e FAD3C (Glyma18g06950), que codificam ω -3-dessaturases, afetam diretamente o acúmulo de ácido linolênico em sementes de soja. O uso de marcadores moleculares associados a esses genes é indicado na seleção de genótipos com teores reduzidos desse ácido graxo (BILYEU et al., 2005 ; BILYEU et al., 2006).

Uma metodologia de genotipagem de SNP de fácil implementação pelos

programas de melhoramento genético e de baixo custo, é a metodologia TSPCR (*Temperature-Switch Polymerase Chain Reaction*) (HAYDEN et al., 2009). Essa metodologia utiliza a amplificação por PCR baseada na hibridização alelo-específica e a discriminação dos alelos é feita em gel de agarose 2%.

O presente trabalho teve como objetivo, identificar e validar SNP em regiões gênicas associadas ao conteúdo de ácidos graxos, em cultivares divergentes do Programa de Melhoramento da Qualidade da Soja da UFV (PMQS-UFV), bem como, utilizar a metodologia de genotipagem TSPCR na discriminação dos alelos desses genes.

3. Materiais e Métodos

3.1. Material genético

Foram utilizados quatro genótipos de soja: A29, FA22, CD219RR e TUCUNARÉ. O genótipo A29 (1% de ácido linolênico) foi desenvolvido por meio de mutagênese por pesquisadores da Universidade Estadual de Iowa (EUA). O genótipo FA22 foi selecionado em uma população de melhoramento e possui teor reduzido de ácido linoléico e elevado conteúdo de ácido oleico (aproximadamente 50%). A variedade CD219RR possui teores normais de ácido linolênico (~8 %) e ácido oléico (~19 %), foi desenvolvida pela empresa COODETEC (Cooperativa Central de Pesquisa Agrícola). A variedade Tucunaré, desenvolvida pela Fundação Matogrosso, apresenta um elevado teor de óleo.

Os materiais genéticos foram plantados em casa de vegetação e as folhas foram coletadas aos 15 dias após o plantio para a extração do DNA, usando a metodologia proposta por DOYLE & DOYLE (1990), com modificações. A concentração e qualidade do DNA foi determinada por meio do espectrofotômetro NanoDrop (NanoDrop Technologies, Wilmington, DE, EUA). O DNA obtido foi também submetido à eletroforese em gel de agarose 0,8%, para avaliar a qualidade. Com base na concentração estimada, procedeu-se à diluição das amostras para uma concentração final de 15 ng/ μ L.

3.2. Seleção dos genes candidatos e desenho de *primers*

Com base em informações disponíveis na literatura, foram selecionados os genes candidatos ABI3, FAD2-1, ARAF, PDAT que possivelmente estão relacionados com os conteúdos de ácidos graxos em sementes de soja, e os genes FAD3B e FAD3C, que codificam ω -3-dessaturases, afetando diretamente o acúmulo de ácido linolênico em sementes de soja. O gene candidato ABI3 é fator de transcrição envolvido no processo de maturação das sementes (BAUD e LEPINIEC, 2010). O gene FAD2-1, codifica uma ω -6-dessaturase que converte ácido oléico em linoléico. Para os genes FAD3B e FAD3C, já foram desenvolvidos marcadores moleculares do tipo SNP (BILYEU *et al.*, 2006). No entanto, o presente trabalho está propondo utilizar a metodologia de genotipagem TSPCR para discriminar os alelos desses dois genes em populações segregantes, visando selecionar genótipos com teores reduzidos de ácido linolênico. Na Tabela 1 estão listados os genes selecionados, com os códigos e as funções das proteínas por eles codificadas, bem como suas respectivas funções, além das referências bibliográficas.

As sequências parciais dos genes candidatos ABI3, ARAF, PDAT e FAD2-1, ou de genes homólogos, foram obtidas a partir do GenBank (<http://www.ncbi.nlm.nih.gov/Genbank>). As sequências gênicas completas (regiões promotoras, regulatórias, terminadoras, introns e exons) inclusive dos genes FAD3B (Glyma02g39230) e FAD3C (Glyma18g06950) foram obtidas realizando um BLASTn (*Basic Local Alignment Search Tool*) no banco de dados PHYTOZOME (<http://www.phytozome.net/soybean>), onde está depositado o genoma da soja sequenciado (SCHMUTZ *et al.*, 2010).

A seleção dos genes candidatos ABI3, ARAF, PDAT e FAD2-1 seguiu os seguintes critérios: i) ser copia única ou no máximo 2 cópias; ii) apresentar valor de *e-value* igual ou próximo a zero e iii) apresentar anotação funcional no PHYTOZOME.

Visando a amplificação das regiões selecionadas para cada gene candidato, foram desenhados oligonucleotídeos iniciadores (*primers*) usando o programa “*Primer3 Input Program*” (<http://frodo.wi.mit.edu/primer3>). Para o gene candidato ABI3 (Glyma08g47240.1 e Glyma18g38490), ARAF (Glyma19g30450.1 e Glyma03g27460.2) e PDAT (Glyma13g16790.1 e Glyma17g05910.1), os *primers* foram desenhados nas regiões regulatórias (5’UTR e 3’UTR) dos seus respectivos transcritos selecionados. Para o transcrito (Glyma20g24530.1) do gene candidato

FAD2-1, por apresentar sequência genômica curta, foi desenhado cinco conjuntos de *primers*, com o objetivo de amplificar e sequenciar todo o gene. Enquanto que para os genes FAD3B (Glyma02g39230) e FAD3C (Glyma18g06950), foram desenhados *primers* na região codificante desses genes, visando amplificar e sequenciar apenas a região onde foi identificado o SNP (BILYEU et al., 2006).

Os *primers* foram desenhados seguindo os parâmetros: i) o fragmento a ser amplificado variando de 400 a 900pb (pares de bases); ii) tamanho do primer entre 19 e 25 pb; iii) temperatura de anelamento entre 59 e 61°C; e iv) conteúdo de CG% entre 40 e 60%. Cada *primer* desenhado foi testado quanto a sua estabilidade termodinâmica usando o programa *Primer Select 3.1* (DNASTAR, Madison, WI, EUA), enquanto que a especificidade foi testada pelo programa *PCR virtual* (iPCR) (http://www.ch.embnet.org/software/iPCR_form.html) e pelo banco de dados PHYTOZOME. Após serem testados, os *primers* foram sintetizados pela empresa BIONEER (Belo Horizonte, Minas Gerais, Brasil).

Tabela 1. Genes selecionados com os códigos e as funções das proteínas por eles codificadas em soja ou em plantas relacionadas. Bem como as referências bibliográficas e os transcritos selecionados do banco de dados phytozome

Gene candidato	Código	Função	Transcritos Selecionados	Referências
Fator de transcrição Insensível ao ABA 3	ABI3	Regulação do desenvolvimento da semente	Glyma08g47240.1 Glyma18g38490.1	WESELAKE et al., 2009
Fosfolipídeos diacilglicerol-aciltransferase	PDAT	Via biossintética dos triacilgliceróis	Glyma13g16790.1 Glyma17g05910.1	BAUD et al., 2010
Alfa-L-Arabinofuranosidase	ARAF	Hidrólise de hemicelulose	Glyma19g30450.1 Glyma03g27460.2	WEI et al., 2008
ω -6-dessaturase	FAD2-1	Síntese de ácidos graxos poliinsaturados	Glyma20g24530.1	HEPPARD et al. 1996
ω -3-dessaturase	FAD3B	Síntese de ácidos graxos poliinsaturados	Glyma02g39230	BILYEU et al., 2006
ω -3-dessaturase	FAD3C	Síntese de ácidos graxos poliinsaturados	Glyma18g06950	BILYEU et al., 2006

3.3. Reações de amplificação e sequenciamento

As reações de amplificação foram feitas para um volume final de 30 µL, usando-se 1U de Platinum®Taq DNA polimerase; Tris-HCl 20 mM pH 8,4; KCl 50 mM; MgCl₂ 1,5 mM; 0,2 mM de cada dNTP; 0,2 µM de cada primer e 30 ng de DNA molde. As reações de amplificação foram conduzidas em termociclador Eppendorff (AG-22331), inicialmente a 94 °C por 2 minutos, seguida de 35 ciclos de 94 °C por 30 segundos, 55 °C por 30 segundos e 72 °C por 60 segundos. Realizadas as reações de amplificação, os produtos gerados foram submetidos à eletroforese em gel de agarose 1,7%, contendo brometo de etídeo, e fotodigitalizados utilizando o sistema L-PIX (Loccus Biotecnologia, São Paulo, Brasil).

Quando apenas o único produto de amplificação esperado foi gerado, estes foram purificados usando-se o kit ExoSAP IT (USB Corporation, Cleveland, Ohio, EUA) na proporção de 6 µL de ExoSap e 18 µL de produto de PCR. Posteriormente, as amostras purificadas foram sequenciadas pelo método didesoxi, utilizando o sequenciador ABI3730xl (Applied Biosystems). As reações de sequenciamento foram feitas pela empresa Macrogen (Gasandong, Geumchun-gu, Seoul, Korea). As sequências obtidas foram editadas usando-se o programa SEQUENCER 4.1.4 (Genes Code Corporation) e alinhadas pelo programa ClustalW (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). As mesmas sequências foram também editadas e alinhadas pelo programa CodonCode Aligner (<http://www.codoncode.com/aligner/>).

3.4. Populações de soja utilizadas

Para a validação dos SNP foram desenvolvidas duas populações. Uma de linhagens endogâmicas recombinantes (*Recombinant Inbred Lines - RILs*), originada do cruzamento entre genótipos contrastantes para teor de ácido oléico (FA22 x CD219) e uma população F_{2,3} contrastante para teor de ácido linolênico (A29 x Tucunaré).

Os cruzamentos para obtenção das RILs foram iniciados no ano agrícola de 2006/2007 na Universidade Federal de Viçosa (UFV). Sementes F₁ foram plantadas em casa de vegetação, sementes F₂ foram plantadas e avançadas até a geração F_{2,8}, usando o método descendente de uma única semente (*Single Seed Descent - SSD*) como reportado por Brim (1966). Para obtenção da população segregante (A29 x Tucunaré)

foram efetuados cruzamentos entre os progenitores e as sementes híbridas foram autofecundadas e obtidas 185 progênies F₂. Esse material foi novamente autofecundado obtendo-se a população F_{2:3}.

3.5. Extração do DNA

Para a genotipagem das duas populações de mapeamento, foram coletadas folhas F₂ (A29 x Tucunaré) e F₆ (FA22 x CD219) das progênies 15 dias após o plantio para a extração e análise do DNA, usando a metodologia proposta por DOYLE & DOYLE (1990), com modificações. A concentração e qualidade do DNA foi determinada por meio do espectrofotômetro NanoDrop (NanoDrop Technologies, Wilmington, DE). O DNA obtido foi também submetido à eletroforese em gel de agarose 0,8%, para avaliar a qualidade. Com base na concentração estimada, procedeu-se à diluição das amostras para uma concentração final de 15 ng/μL.

3.6. Determinação da composição de ácidos graxos

A determinação da composição de ácidos graxos na fração óleo foi feita por cromatografia gasosa. As etapas de extração e transesterificação dos ácidos graxos em metil-ésteres derivados foram realizadas segundo a metodologia de BUBECK et al. (1989). Os metil-ésteres derivados foram fracionados em coluna Carbowax de 30 m de comprimento e 0,32 mm de diâmetro interno. As temperaturas da coluna e do injetor foram mantidas a 240 °C e a do detector a 280 °C

3.7. Análise estatística dos dados

Todas as análises estatísticas foram realizadas empregando os aplicativos computacionais GENES (CRUZ, 2006) e GQMOL (CRUZ e SCHUSTER, 2004).

3.8. Desenhos dos *primers* de genotipagem TSPCR

Para investigar os polimorfismos identificados nos genes ABI3, ARAF, FAD2-1, FAD3B e FAD3C, foram desenhados *primers* de genotipagem TSPCR (*Temperature-*

Switch Polymerase Chain Reaction) (HAYDEN et al., 2009). Os *primers* TSPCR foram desenhados com auxílio dos programas *Primer 3* e *Primer Select*.

Foram desenhados para cada um dos polimorfismos selecionados, um par de *primers* loco específico-LE (R/F), com temperatura de anelamento (T_a) entre 60 e 65 °C, visando amplificar fragmentos de 400 pb, e um *primer* informativo (*primer* alelo específico-ALE), com seu terminal 3' complementar a um dos alelos presente no loco, com T_a entre 43 e 48 °C, ao qual é adicionado, no terminal 5', uma cauda de 2 a 3 nucleotídeos aleatórios e não complementares à sequência do DNA molde, para elevar o seu valor de T_a para 53 °C. Cada *primer* informativo forma produtos de amplificação com o tamanho de aproximadamente 100 pb com um dos dois *primers* loco específico.

3.9. Genotipagem de SNP utilizando a metodologia TSPCR

As reações de genotipagem dos polimorfismos selecionados nas populações segregantes RILs, foram realizadas em um volume total de 15 µL, contendo 1U de Platinum®Taq DNA polimerase; Tris-HCl 20 mM pH 8,4; KCl 50 mM; MgCl₂ 1,5 mM; 0,1 µg/µL de soroalbumina bovina; 0,01% tween-20; 1,0% de formamida; 0,2 mM de cada dNTP (dATP, dTTP, dGTP e dCTP); 0,083 µM do *primer* loco específico, 0,83 µM do *primer* informativo e 30 ng de DNA molde. As reações de amplificação foram conduzidas em termociclador Eppendorff (modelo AG-22331), inicialmente a 94 °C por 2 min, seguida de: 10 ciclos de 94 °C por 30 s, 58 °C por 30 s e 72 °C por 60 s; 5 ciclos de 94 °C por 10 s e 45 °C por 30 s; 15 ciclos de 94 °C por 30 s, 53 °C por 30 s e 72 °C por 5 s. Realizadas as reações de amplificação, os produtos gerados foram submetidos à eletroforese em gel de agarose a 2%, contendo brometo de etídeo (0,2 µg/mL), e fotodigitalizados, utilizando o sistema L-PIX (Loccus Biotecnologia).

4. Resultados e Discussão

4.1. Amplificação de fragmentos dos genes candidatos por PCR

Para a obtenção dos fragmentos de interesse de cada gene candidato, foram feitas reações de amplificação por PCR, utilizando DNA total de folhas de soja e os *primers* específicos correspondentes para cada gene candidato. Realizadas as reações de amplificação, os produtos gerados foram submetidos à eletroforese em gel de agarose

1,7 %. Quando mais de um produto de amplificação foram gerados, a reação foi repetida em temperaturas de anelamento maiores, até se conseguir uma única banda ou descartar definitivamente este par de *primers*.

Foram desenhados cinco conjuntos de *primers* (R/F) para o gene FAD2-1 (Glyma20g24530.1) com objetivo de amplificar e sequenciar todo o gene. Quatro conjuntos de *primers* produziram amplicons sem bandas inespecíficas (cada amplicon variando de 500 a 600 pb).

Os *primers* desenhados na região 5'UTR dos transcritos dos genes ABI (Glyma08g47240.1), PDAT (Glyma13g16790.1) e ARAF (Glyma19g30450.1), bem como os *primers* desenhados na região 3'UTR do transcrito do gene PDAT (Glyma17g05910.1) produziram produtos de amplificação inespecífica nos progenitores testados, sendo descartados. Os *primers* desenhados para amplificar as regiões gênicas que abriga SNP (BILYEU *et al.*, 2006) nos genes FAD3B (Glyma02g39230) e FAD3C (Glyma18g06950), produziram amplicons sem amplificação inespecífica. Quando apenas um único produto de amplificação esperado foi gerado para cada progenitor, estes foram posteriormente purificados com ExoSap e sequenciados.

4.2. Sequenciamento e identificação de polimorfismos

Os *primers* desenhados para amplificar regiões específicas dos transcritos dos genes ARAF (Glyma19g30450.1 e Glyma03g27460.2), PDAT (Glyma13g16790.1 e Glyma17g05910.1), FAD2-1 (Glyma20g24530.1), ABI3 (Glyma08g47240.1 e Glyma18g38490), FAD3B (Glyma02g39230) e FAD3C (Glyma18g06950) que produziram apenas o único produto de amplificação nos respectivos genitores testados, foram purificados e sequenciados utilizando o sequenciador ABI3730xl. As sequências obtidas foram editadas usando os programas SEQUENCER 4.1.4 e CodonCode Aligner, e alinhadas pelo programa ClustalW para identificação dos polimorfismos.

Foram obtidos aproximadamente 8,461 kpb de sequências com boa qualidade, com valor de Phred superior ou igual a 30, sendo identificados 25 SNP e 3 INDELS, média de um SNP para cada 302 pb (Tabela 2). A frequência habitual de SNP relatados para genomas de plantas é de cerca de 1 SNP a cada 100-300 pb (GUPTA *et al.*, 2001). Em milho, esta frequência aumenta para cerca de 1 polimorfismo a cada 60,8 pb (CHING, 2002) e, em trigo, a densidade de polimorfismos é de 1 a cada 370 pb (PROCUNIER *et al.*, 2003).

Tabela 2. Regiões gênicas sequenciada dos transcritos de cada gene candidato e o número de polimorfismos encontrado

Genes	Transcritos	Região dos Genes			Número de polimorfismo	
		Total de pb compreendidos			SNP	INDELS
		5' UTR	Codificante	3' UTR		
ABI3	Glyma08g47240.1 Glyma18g38490.1	1640	-	822	5	0
ARAF	Glyma19g30450.1 Glyma03g27460.2	849	-	1630	8	3
PDAT	Glyma17g05910.1	810	-	-	5	0
FAD2-1	Glyma20g24530.1	400	1250	450	5	0
FAD3B	Glyma02g39230	-	300	-	1	0
FAD3C	Glyma18g06950	-	310	-	1	0
Total		3699	1860	2902	25	3

A frequência de SNP depende da região gênica sequenciada, bem como da espécie. Na cultura da soja a frequência de SNP é relativamente baixa em comparação com outras espécies agrícolas cultivadas (HYTEN et al, 2006;. ZHU et al, 2003). A baixa frequência de sequências variantes é esperada, devido ao processo de seleção da soja ao longo dos anos, reduzindo a variação em 50%, e a estreita base genética dos genótipos utilizados para a obtenção das primeiras cultivares, além da baixa variação do ancestral selvagem *G. soja* (HYTEN et al., 2006). Entretanto, ZHU et al. (2003) realizou um estudo com as principais variedades de soja cultivadas nos Estados Unidos, neste trabalho foram sequenciados 66,6 Kpb, sendo 28,7 kpb de sequências codificantes e 37,9 kpb de sequencias não-codificantes, sendo identificado 250 SNP, com uma frequência de 1 SNP a cada 270 pb.

4.3. Avaliação genética e estatística das populações de mapeamento

4.3.1. Valores médios dos ácidos graxos

Os valores médios dos teores de ácidos graxos nos genótipos FA22, CD219, A29 e Tucunaré, das populações segregantes $F_{2:3}$ (A29 x Tucunaré) e RILs (FA22 x CD219), são apresentados na Tabela 2. Observa-se que os progenitores A29 e Tucunaré são contrastantes para os teores de ácido linolênico. A população $F_{2:3}$ derivada do

cruzamento entre A29 e Tucunaré, apresenta o maior valor máximo e mínimo para teores de ácido linolênico, entre 0,97 - 12,90 % respectivamente. Os progenitores FA22 e CD219 são divergentes para teor de ácido oléico. A população de RILs, derivada desse cruzamento, apresentou valor máximo e mínimo para essa característica entre 12,81 e 46,38.

4.3.2. Teste de normalidade

Os ácidos graxos (palmítico, esteárico, oléico, linoléico e linolênico) nas duas populações apresentaram distribuição aproximadamente normal ($P < 0,01$), como pode ser observado na Tabela 4, sendo confirmado pelas estimativas não significativas de assimetria, curtose e valor D do teste de Lilliefors. Este fato indica a existência de variação contínua e a possibilidade de que vários genes estejam envolvidos no controle dessas características.

Tabela 3. Valores de média (X), desvio padrão (DP), mínimo (MIN), máximo (MAX), para teores de ácidos graxos dos progenitores FA22, CD219, A29 e Tucunaré e das populações segregantes F_{2:3} do cruzamento A29 x Tucunaré e RILs FA22 x CD219

Progenitor		TEORES DE ÁCIDOS GRAXOS				
População		Palmítico	Estearico	Oléico	Linoléico	Linolênico
FA22	\bar{X}	9,81	2,70	38,57	40,68	8,24
	DP	0,30	0,30	4,04	3,39	0,70
	MIN	9,24	2,09	30,97	35,59	7,11
	MAX	10,24	3,16	44,65	46,88	9,36
CD219	\bar{X}	10,46	3,97	19,27	48,34	8,42
	DP	0,60	0,37	1,29	1,72	0,62
	MIN	10,67	3,66	18,91	51,06	8,50
	MAX	12,33	5,05	22,91	56,34	10,57
A29	\bar{X}	11,009	4,072	21,715	62,025	1,191
	DP	0,470	0,452	3,222	3,412	0,175
	MIN	9,663	3,046	15,217	54,668	0,904
	MAX	11,882	5,112	28,329	69,854	1,846
Tucunaré	\bar{X}	11,414	3,109	19,138	58,175	7,995
	DP	0,357	0,262	1,193	0,828	0,728
	MIN	10,911	2,713	17,075	56,566	5,999
	MAX	12,173	3,670	21,957	59,364	9,069
FA22 x CD219	\bar{X}	8,80	2,70	35,73	46,87	5,90
	DP	0,85	0,66	4,06	2,78	1,02
	MIN	8,59	1,87	12,81	36,46	6,39
	MAX	15,16	5,87	46,38	60,75	12,35
A29 x Tucunaré	\bar{X}	10,33	3,67	24,74	57,04	4,21
	DP	1,45	1,28	4,37	4,02	1,59
	MIN	0,67	2,27	9,66	37,06	0,97
	MAX	14,23	19,97	47,89	65,92	12,90

Tabela 4. Estimativas dos valores de curtose, simetria e D, para a distribuição dos teores de ácido palmítico, esteárico, oléico, linoléico e linolênico, nas populações A29 x Tucunaré (F_{2:3}) e FA22 x CD219 (RILs).

População	FA22 x CD219														
	Ácido Palmítico			Ácido Esteárico			Ácido Oléico			Ácido Linoléico			Ácido Linolênico		
	Curtose	Simetria	D ⁽¹⁾	Curtose	Simetria	D ⁽¹⁾	Curtose	Simetria	D ⁽¹⁾	Curtose	Simetria	D ⁽¹⁾	Curtose	Simetria	D ⁽¹⁾
Estimativa	2,82	-0,04	0,03 ^{ns}	3,87	0,55	0,05 ^{ns}	2,63	0,45	0,065 ^{ns}	2,65	-0,50	0,067 ^{ns}	2,65	0,20	0,04 ^{ns}
Variância	0,09	0,02		0,09	0,02		0,09	0,02		0,09	0,02		0,09	0,02	
Valor de t	-0,60	-0,28		2,87	3,64		-1,24	2,98		-1,14	-3,3207		-1,15	1,29	
Probabilidade (%)	55,78	77,33		0,43	0,04		21,40	0,31	>1,00% <5,00%	25,12	0,11	>1,00% <5,00%	24,87	19,27	
População	A29 X Tucunaré														
Estimativa	2,76	0,14	0,05 ^{ns}	2,98	-0,11	0,05 ^{ns}	3,44	1,84	.0,06 ^{ns}	2,53	-0,17	0,045 ^{ns}	2,29	0,28	0,067 ^{ns}
Variância	0,13	0,03		0,13	0,03		0,13	0,03		0,13	0,032		0,13	0,032	
Valor de t	-0,66	0,794		-0,04	-0,62		3,94	3,30		-1,33	-0,95		-1,98	1,59	
Probabilidade (%)	51,33	56,69		96,69	54,49		20,94	10,30	>1,00% <5,00%	18,14	65,73		12,49	22,70	>1,00% <5,00%

(1) Estatística D para o Teste de Lilliefors.

(*) Significativo a 1% de probabilidade pelo teste t; (ns) – Não significativo pelo teste t.

4.4. Teste com os *primers* TSPCR

A genotipagem dos polimorfismos nas duas populações de mapeamento foi feita por meio da metodologia TSPCR (HAYDEN et al., 2009).

Na Tabela 5 estão listadas as sequências dos 6 conjuntos de *primers* TSPCR desenhados para cada polimorfismo testado. Para seis genes selecionados, foi selecionado um único polimorfismo para o desenho dos *primers* TSPCR. Após serem sintetizados, os *primers* TSPCR foram submetidos a reações de PCR em condições reacionais padronizadas, nos 2 progenitores onde se identificou o respectivo polimorfismo por sequenciamento.

Os SNP e INDELS identificados no transcrito dos genes, ARAF (Glyma03g27460.2), PDAT (Glyma17g05910.1) e FAD2-1 (Glyma20g24530.1) foram testados nos progenitores (FA22 e CD219), com objetivo de confirmar esses polimorfismos identificados pelo sequenciamento, antes de testá-los na população segregante. Os *primers* de genotipagem TSPCR desenhados para o polimorfismo presente no gene FAD2-1 (Glyma20g24530.1) apresentaram ampliações inespecíficas. Os *primers* TSPCR desenhados para os polimorfismos dos genes ARAF (Glyma03g27460.2), PDAT (Glyma17g05910.1) mostraram-se polimórficos (Figura 1). Os polimorfismos dos genes FAD3B (Glyma02g39230), FAD3C (Glyma18g06950) e ABI3 (Glyma18g38490.1) foram testados com *primers* TSPCR nos progenitores A29 e Tucunaré. Todos os conjuntos de *primers* mostraram-se polimórficos, sem marcação de bandas fracas e/ou bandas inespecíficas (Figura 1).

Todos os *primers* desenhados para a obtenção dos fragmentos de interesse em cada gene candidato, tiveram à especificidade testada pelo programa PCR virtual (iPCR) e pelo banco de dados PHYTOZOME. No entanto, a presença de várias cópias do mesmo gene no genoma da soja, provavelmente possibilitou a geração de produtos inespecíficos para o gene candidato FAD2-1. Pois os *primers* TSPCR desenhados para discriminar os alelos de SNP nesse gene candidato, pode ter anelado em regiões inespecíficas do genoma. O gene candidato FAD2-1, apresenta quatro transcritos alternativos (Glyma20g24530.2, Glyma20g24530.3, Glyma20g24530.4 e Glyma20g24530.5) com alta similaridade nas sequências, Além disso, esse gene apresenta outras três cópias altamente similares nos cromossomos 10, 19 e 3, o que provavelmente possibilitou a geração de produtos inespecíficos para esse polimorfismo. O genoma da soja passou por dois eventos de duplicações, um mais antigo a 59 milhões de anos atrás e segundo mais recente a 13

milhões de anos atrás, resultando em um genoma altamente repetido, com cerca de 75 % dos genes presentes em múltiplas cópias (SCHMUTZ et al., 2010).

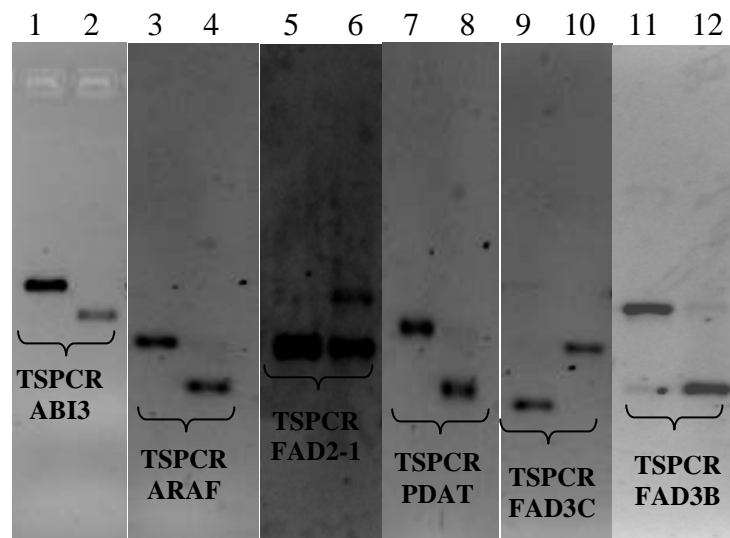


Figura 1. Teste com alguns dos *primers* TSPCR em gel de agarose 2%, com os 2 progenitores onde se identificou o respectivo polimorfismo. Mostrando *primers* TSPCR polimórficos (canaletas 1-2, 3-4, 7-8, 9-10 e 11-12) e bandas inespecíficas (canaletas 5-6). Canaletas 1, 10 e 11 (Tucunaré), canaletas 2, 9 e 12 (A29); canaletas 3, 6 e 8 (CD219), canaletas 4, 5 e 7 (FA22).

Tabela 5. Gene candidato (GC), cromossomos, regiões gênicas (Exon, 5'UTR e 3'UTR), alelos de SNP e sequências de *primers* TSPCR para seis SNP investigados

GC	Cromossomos	Regiões gênicas	Alelos SNP		Primers LE	Primers ALE ^{a,b}	Alelo de SNP investigado
ABI3	18	5'UTR	A/C	Fwd	ATCCTATGAAGGGATGCAACA	<u>CGGTATGATTATCATTGTGTATAT</u>	T
				Rev	GCGTTTTGATCATCTTCGAGT		
ARAF	3	5'UTR	A/T	Fwd	CCCCCTTCCTTTAACTCCTTATT	<u>GGACTCTGATTCTGTTACTGTTA</u>	A
				Rev	ACCCACCCACACCATGTTAC		
FAD2-1	20	3'UTR	C/T	Fwd	AGGCAACCAATGCAATGAAG	<u>GGTACAAGCTGTTTCACAATA</u>	T
				Rev	TGGAGTGTGCATTTAAGGAAGAA		
FAD3B	2	Exon	G/A	Fwd	GGATGAGTCATGGGTTCCAG	<u>CCGTTTATGTGTCATTATCTTAG</u>	G
				Rev	CCGGGGCTTCTTGAAAATA		
FAD3C	18	Exon	C/T	Fwd	GCAGTTTTGCCATCCTTCAA	<u>ACAATTTGAGGGAAAAGGTA</u>	T
				Rev	TTGACTTAGATCACATAGGCAGACA		
PDAT	17	5'UTR	C/T	Fwd	GCTTGTCAGACATGAAACGTCC	<u>CGATCAGATAGATCATTTAAGACT</u>	T
				Rev	GGCAATTAAGCGGGAATGAA		

^a Um dos alelos de SNP complementar ao primer alelo específico (ALE) produz o alelo de SNP referência, produto de PCR menor, enquanto que o alelo sem complementaridade produz o alelo de SNP alternativo produto de PCR maior. Amostras heterozigotas produzem ambos os produtos de genotipagem TSPCR

^b A sequência 5'-arbitrária de nucleotídeos, indicado em negrito e sublinhado, não é complementar ao DNA alvo.

ALE- primer alelo específico

LE- primer locus específico

4.5. Genotipagem das populações e teste de segregação

Os *primers* TSPCR polimórficos nos respectivos progenitores foram utilizados na genotipagem das duas populações do PMQS, desenvolvidas para mapear QTLs para alto teor de ácido oléico (FA22 x CD219-RILs) e para baixo teor de ácido linolênico (A29 x Tucunaré-F_{2:3}).

Os *primers* TSPCR desenhados para discriminar os alelos de SNP identificados nos genes *ARAF* (Glyma03g27460.2) e *PDAT* (Glyma17g05910.1), foram amplificados em 259 progênies F₆ e nos genitores da população FA22 x CD219 (Figura 2). Enquanto que para população A29 x Tucunaré, foram genotipadas 185 progênies na geração F₂, com os *primers* TSPCR desenhados no polimorfismo dos genes *FAD3B* (Glyma02g39230), *FAD3C* (Glyma18g06950) e *ABI3* (Glyma18g38490.1) (Figura 3).

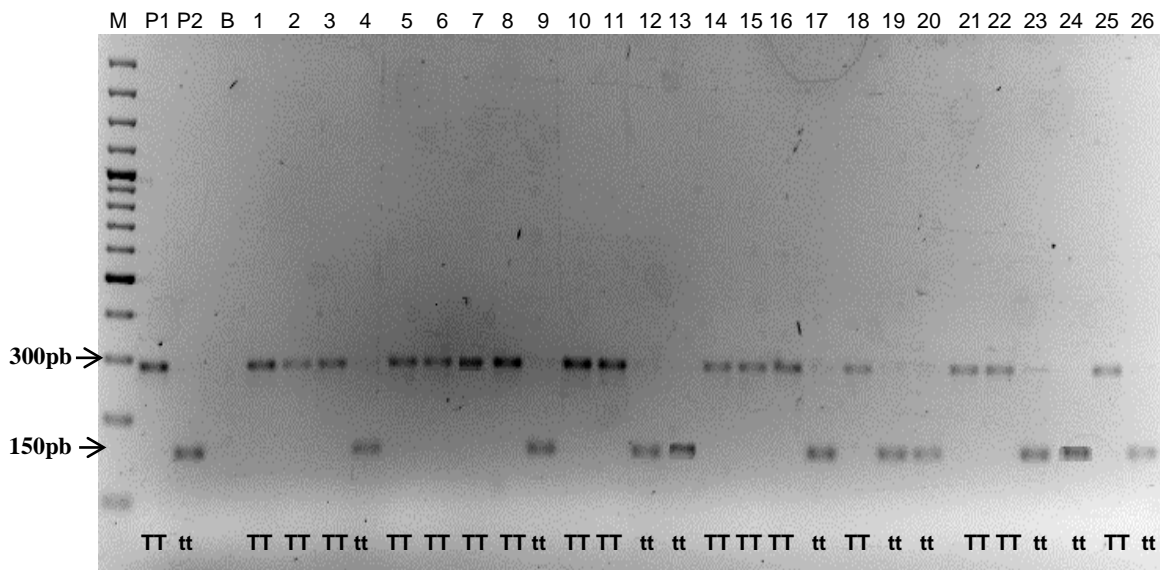


Figura 2. Genotipagem dos indivíduos utilizando *primers* TSPCR desenhados no polimorfismo (C/T) do gene candidato *PDAT*, em 26 RILs F₆ (FA22 x CD219). P1(FA22), P2 (CD219), B (branco), Canaletas de 1 a 26 (progênies) e canaleta M – Marcador de tamanho (*GeneRuler 100 bp DNA Ladder Plus* – Fermentas). Note que o marcador *PDAT* produziu dois amplicons, o primeiro amplicom de 300 pb (produto de amplificação alelo-alternativo - TT) e um segundo amplicom de aproximadamente 150 pb (produto de amplificação alelo-específico - tt).

A segregação de cada marcador foi testada por meio do teste qui-quadrado ($P < 0,05$) utilizando o critério de Proteção Bonferroni. A segregação dos marcadores moleculares dos genes *ABI3*, *FAD3B* e *FAD3C* na população F₂ do cruzamento entre A29 e Tucunaré seguiu o padrão esperado (1:2:1). Os dois os marcadores genotipados na população de RILs, derivada do cruzamento entre os genótipos FA22 e CD219, seguiram o padrão esperado de 1:1 (Tabela 6).

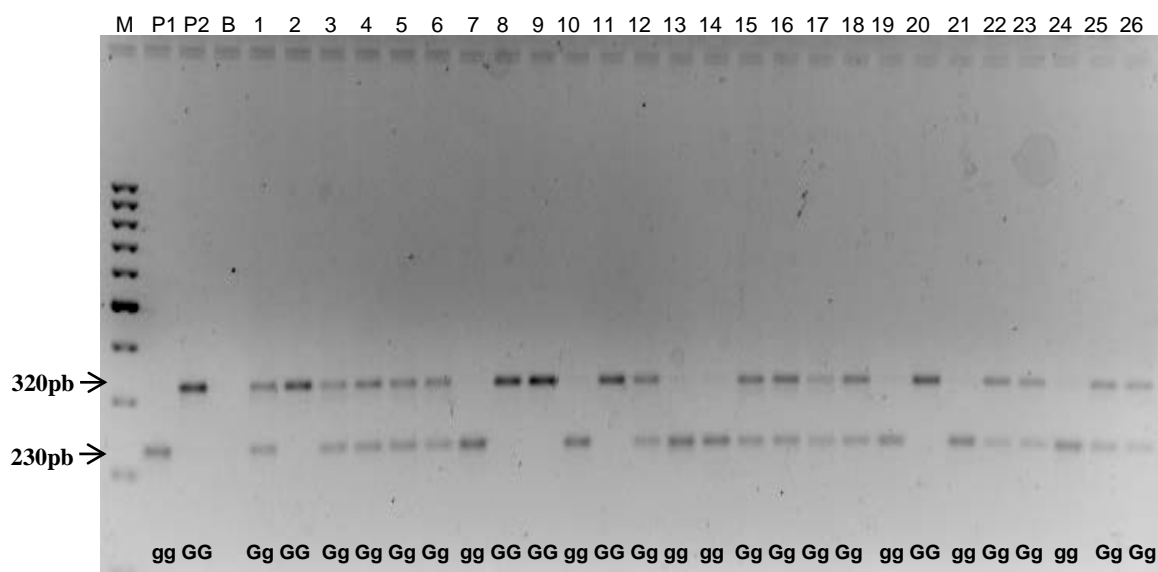


Figura 3. Genotipagem dos indivíduos utilizando *primers* TSPCR desenhados no polimorfismo (G/A) do gene candidato FAD3C, em 26 progenies F₂ (A29 x Tucunaré). P1(A29), P2 (Tucunaré), B (branco), Canaletas de 1 a 26 (progênes) e canaleta M – Marcador de tamanho (*GeneRuler 100 bp DNA Ladder Plus* – Fermentas). Note que o marcador FAD3C produziu dois amplicons, o primeiro amplicom de 320 pb (produto de amplificação alelo-alternativo - GG) e um segundo amplicom de aproximadamente 230 pb (produto de amplificação alelo-específico - gg).

A metodologia de genotipagem TSPCR mostrou-se eficaz na discriminação dos alelos de SNP nas duas populações estudadas. Na população F₂ A29 X Tucunaré, a metodologia TSPCR foi eficaz na discriminação das classes genotípicas (GG, Gg e gg) dos marcadores genotipados nessa população (Figura 3). Da mesma forma, para população de RILs FA22 x CD219, na discriminação das duas classes genotípicas (TT e tt) dos marcadores genotipados nessa população (Figura 4).

Os SNP nos genes FAD3B e FAD3C do progenitor A29 foi caracterizado por BILYEU et al. (2006), permitindo o desenvolvimento de marcadores moleculares do tipo PCR-RFLP específicos para os alelos de A29. A técnica de PCR-RFLP (*Restriction Fragment Length Polymorphism*) consiste na utilização de enzimas de restrição, também conhecidas como endonucleases, cujo papel biológico é clivar moléculas de DNA exógeno e são encontradas em uma ampla variedade de organismos procariontes. O polimorfismo é observado por hibridização dos fragmentos de DNA digeridos com sequências homólogas (sondas) marcadas com radioatividade ou compostos que desencadeiem uma reação de luminescência. Este processo, no entanto, é caro e demanda muito tempo para sua execução. Além disso, devido à limitada existência de endonucleases, se torna muito difícil detectar toda forma de SNP presente no genoma.

Atualmente, outras metodologias para a genotipagem de SNP estão disponíveis na literatura, entretanto, a maioria delas requer equipamentos e/ou reagentes de custo elevado, fora da realidade da maioria dos grupos de pesquisa e das empresas de melhoramento genético do Brasil. A metodologia de genotipagem TSPCR foi descrita pela primeira vez por HAYDEN et al. (2009), num trabalho de genotipagem de SNP em trigo. Ela se baseia na hibridização alelo-específica, onde o oligonucleotídeo iniciador é desenhado com seu terminal 3' complementar ao SNP interrogado. Somente um dos dois possíveis alelos vai gerar um produto de amplificação do tamanho desejado, e a discriminação dos alelos é feita em gel de agarose, o que a torna de baixo custo.

Tabela 6. Teste de segregação com os marcadores moleculares genotipados nas populações FA22 X CD219 (RILs) e A29 x Tucunaré (F2). Classes genotípicas, hipótese, valor do teste de Qui-quadrado e probabilidade para os marcadores para os marcadores

Populações	FA22 x CD219						A29 x Tucunaré					
	Classe genotípica			Hipotese	χ^2	Probabilidade	Classe genotípica			Hipotese	χ^2	Probabilidade
Marcadores	AA	Aa	aa				BB	Bb	bb			
ARAF	126	2	131	1:1	0,10	75,5 ^{ns}	-	-	-	-	-	-
PDAT	115	5	139	1:1	2,27	13,2 ^{ns}	-	-	-	-	-	-
ABI3	-	-	-	-	-	-	38	66	38	1:2:1	0,70	70,32 ^{ns}
FAD3B	-	-	-	-	-	-	38	100	48	1:2:1	2,13	34,49 ^{ns}
FAD3C	-	-	-	-	-	-	39	90	54	1:2:1	2,508	28,53 ^{ns}

Nível crítico: 10%
 Total de testes: 20
 Critério: Proteção Bonferroni
 Número de distorções: 0

4.6. Análise de associação

A associação marcador-característica foi determinada pela análise de marca simples, por meio de regressão linear (Tabela 7 e 8). Nenhum dos marcadores SNP genotipados na população de RILs (FA22 x CD219) apresentaram associações significativas com os ácidos graxos estudados.

No entanto, para população segregante $F_{2:3}$, derivada do cruzamento entre A29 e Tucunaré, os três marcadores apresentaram associações significativas com um ou mais constituinte dos ácidos graxos avaliados. O marcador SNP do gene ABI3 (Glyma18g38490.1), apresentou associação significativa com a característica teor de ácido palmítico, explicando 5,42 % da variação do fenótipo (12,67 – 9,65 %) determinado pelo R^2 . O marcador SNP do gene FAD3B (Glyma02g39230), associou significativamente com os teores de ácidos graxos esteárico, explicando 3,38 % da variação fenotípica (5,59 – 2,40 %); oléico, explicando 6,22 % da variação da característica (27,82 – 14,14); e linolênico, explicando 2,79 % da variação do fenótipo (11,95 - 1,24). O marcador SNP do gene FAD3C (Glyma18g06950), apresentou associação significativa com os teores de ácidos graxos linoléico, explicando 9,21 % da variação fenotípica (68 – 55 %) e linolênico, explicando 9,51 % da variação dessa característica.

Apesar do marcador SNP do gene ABI3 ter explicado somente 5,42 % da variação do teor de ácido palmítico, esse marcador pode ser utilizado pelo programa de melhoramento genético, em populações com “*background*” genéticos diferentes da utilizada no presente trabalho, visto que a população A29 x Tucunaré apresenta pouca variabilidade genética para essa característica. Atualmente, existe um grande interesse em reduzir o teor de ácido palmítico no óleo de soja que é destinado para produção de biodiesel, pois esse ácido graxo diminui a fluidez do biodiesel no frio (GRAEF et al., 2009).

Na população F_2 , derivada do cruzamento entre A29 e Tucunaré, as progênies com os menores valores de ácidos linolênico na geração $F_{2:3}$, apresentaram as formas recessivas para os genes FAD3B e FAD3C. Resultados semelhantes foram observados por BILYEU et al. (2006), avaliando uma população contrastante para teor de ácido linolênico, derivada do cruzamento entre os genótipos SS97-6946 e IA3017. Nesse estudo, foram avaliados 63 indivíduos e encontrado apenas um genótipo que apresentava as formas recessivas para os genes FAD3B e FAD3C e FAD3A, esse

indivíduo apresentou o menor conteúdo de ácido linolênico. PINTO (2012) realizou análise associativa do conteúdo de ácido linolênico com progênies F_2 e $F_{2:3}$, derivadas do cruzamento entre os genitores A29 e Tucunaré, observando que os indivíduos com menor conteúdo de ácido linolênico apresentam predominância de alelos recessivos (FAD3A, FAD3B e FAD3C), demonstrando que a herança para baixo teor de ácido linolênico é recessiva.

A progênie que apresentou o maior teor de ácido oléico (25 %) na população $F_{2:3}$ (A29 x Tucunaré) foi recessiva para o gene FAD3B. Esse resultado indica a possibilidade da seleção de indivíduos com alto teor de ácido oléico, tendo em vista que a herança para teor de ácido oléico (18:1) é recessiva (RAHMAN et al., 1996). No entanto, esta efetividade deve ser testada em outras populações com “*backgrounds*” genéticos diferentes.

A efetividade dos marcadores FAD3B e FAD3C na seleção de baixo teor de ácido linolênico já foi testada em outras populações com “*backgrounds*” genéticos diferentes da utilizada no presente trabalho (BILYEU et al., 2005; BILYEU et al., 2006).

Tabela 7. Análise associativa dos marcadores SNP identificados nos genes ABI3, FAD3B e FADC com a composição de ácidos graxos, avaliado na população derivada do cruzamento entre A29 x Tucunaré em F_{2:3}

A29 x Tucunaré															
Ácidos Graxos	Ácido Palmítico			Ácido Esteárico			Ácido Oléico			Ácido Linoléico			Ácido Linolênico		
	Marcador	Prob, F	Média	R ² _%	Prob, F	Média	R ² _%	Prob, F	Média	R ² _%	Prob, F	Média	R ² _%	Prob, F	Média
ABI3	0,0055**	11,18	5,42	1	4,14	0,015	0,145	17,33	1,52	1	60,99	0,02	1	6,35	0,34
FAD3B	1	11,14	0,37	0,01*	3,90	3,38	0,0006**	17,84	6,22	1	61,80	0,04	0,02*	5,32	2,79
FAD3C	0,1505	11,16	1,13	1	3,99	0,01	1	17,54	0,0146	0**	61,17	9,21	0**	6,14	9,51

** e *: Significativo a 1 e 5% de probabilidade

Tabela 8. Análise associativa dos marcadores SNP identificados nos genes ARAF e PDAT com a composição de ácidos graxos, avaliado na população derivada do cruzamento entre FA22 x CD219 em RILs

FA22 x CD219															
Ácidos Graxos	Ácido Palmítico			Ácido Esteárico			Ácido Oléico			Ácido Linoléico			Ácido Linolênico		
	Marcador	Prob, F	Média	R ² _%	Prob, F	Média	R ² _%	Prob, F	Média	R ² _%	Prob, F	Média	R ² _%	Prob, F	Média
ARAF	1	8,78	0,15	1	2,69	0,0001	1	35,89	0,03	1	46,75	0,01	1	5,90	0,12
PDAT	1	8,78	0,006	1	2,69	0,1038	1	35,92	0,18	1	46,70	0,32	1	5,90	0,25

** e *: Significativo a 1 e 5% de probabilidade

5. Conclusões

- Os *primers* desenhados em regiões específicas dos transcritos dos quatro genes candidatos selecionados, diretamente relacionados com composição de ácidos graxos, bem como os genes FAD3B e FAD3C que codificam ω -3-dessaturases, afetando diretamente o acúmulo de ácido linolênico em sementes de soja, possibilitou a identificação 25 SNP e 3 INDELS em 8,461 Kpb de sequências, sendo a média de um SNP para cada 302 pb.
- A metodologia TSPCR discriminou as três classes genóticas dos três marcadores genotipados na população F₂ (A29 x Tucunaré), bem como as duas classes genóticas dos dois marcadores genotipados na população de RILs (FA22 x CD219).
- Nenhum dos marcadores SNP genotipados na população de RILs (FA22 x CD219) apresentou associação significativa com os teores de ácidos graxos estudados. No entanto, para população segregante F_{2:3}, derivada do cruzamento entre A29 e Tucunaré, os três marcadores apresentaram associações significativas com uma ou mais característica.
- O marcador SNP do gene ABI3 associou significativamente com teor de ácido palmítico. O marcador SNP do gene FAD3B, associou significativamente com os teores de ácidos graxos esteárico, oléico e linolênico. Enquanto que o marcador SNP do gene FAD3C apresentou associação significativa com os teores de ácidos graxos; linoléico e linolênico.
- Os marcadores SNP dos genes ABI3, FAD3B e FAD3C poderão ser usados para selecionar linhagens de soja com baixíssimos teores de ácido linolênico e/ou altos teores de ácido oléico.

6. Referências Bibliográficas

BAUD S.; LEPINIEC L. Physiological and developmental regulation of seed oil production. **Progress in Lipid Research**, 2010.

BILYEYU, K. D.; PALAVALLI, L.; SLEPER, D. A.; BEUSELINCK, P. R. Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels **Crop Science**; v. 43, p.1833–1838, 2003.

BILYEYU, K. D.; PALAVALLI, L.; SLEPER, D. A.; BEUSELINCK, P. R. Molecular genetic resources for development of 1% linolenic acid soybeans **Crop Science**; v. 46, p.1913–1918, 2006.

BILYEYU, K. D.; PALAVALLI, L.; SLEPER, D. A.; BEUSELINCK, P. R. Mutations in Soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. **Crop Science**; v. 45, p.1830-1836, 2005.

BRIM, C.A. A modified pedigree method of selection in soybeans. **Crop science**; v.6, p.22, 1966.

BUBECK, D.M.; FEHR, W.R.; HAMMOND, E.G. Inheritance of palmitic and stearic acid mutants of soybean. **Crop Science**, Madison; v.29, p.652-56, 1989.

CARDINAL, A.; BURTON, J. W.; CAMACHO-ROGER, A. M.; WHETTEN, R.; CHAPPELL, A. S.; BILYEYU, K. D.; AUCLAIR, J.; DEWEY, R.E. Molecular analysis of GmFAD3A in two soybean populations segregating for the fan, fap1, and fapnc loci. **Crop Science**; v.51, 2011.

CHING, A.; CALDWELL, K. S.; JUNG, M.; DOLAN, M.; SMITH, O. S. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. **BMC Genetic**; v.3, p.19, 2002.

CRUZ, C.D.; SCHUSTER, I. GQMOL: aplicativo computacional para análise de dados moleculares e de suas associações com caracteres quantitativos. Versão 2.1. Viçosa: UFV, 2004. Disponível em: <<http://www.ufv.br/dbg/gqmol/gqmol.htm>>. Acesso em: 13 jan. 2005.

CRUZ, C.D.; SCHUSTER, I. GQMOL: aplicativo computacional para análise de dados moleculares e de suas associações com caracteres quantitativos. Versão 2.1. Viçosa: UFV, 2004. Disponível em: <<http://www.ufv.br/dbg/gqmol/>>.

DUNTON, H. J.; LANCASTER, C. R.; EVANS, C. D.; COWAN, J. C. The flavor problem of soybean oil VII. Linolenic acid. **Journal of the American Oil Chemists Society**; v.28, p.115-118, 1951.

GRAEF, G.; LAVALLEE, B. J.; TENOPIR, P.; TAT, M.; SCHWEIGER, B.; KINNEY, A.J.; VAN GERPEN, J.H.; CLEMENTE, T.E. A high-oleic-acid and low-palmitic-acid soybean: agronomic performance and evaluation as a feedstock for biodiesel. **Plant Biotechnology Journal**; v. 7, p. 411–421, 2009.

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.4524-535, 2001.

HAYDEN, M.J.; TABONE, T.; MATHE, D.E. Development and assessment of simple PCR markers for SNP genotyping in barley. **Theoretical and Applied Genetics**; v.119, p.939–951, 2009.

HEPPARD, E. P.; KINNEY, A. J.; STECCA, K. L.; MIAO, G. H. Developmental and growth temperature regulation of two different microsomal ω -6-desaturase genes in soybeans. **Journal Plant Physiology**; v.110, p.311-319, 1996.

HYTEN, L.D.; SONG, Q.; ZHU, Y.; CHOI, I.-Y.; NELSON, L.R.; COSTA, M.J.; SPECHT, E.J.; SHOEMAKER, R.C.; CREGAN, B.P. Impact of genetic bottlenecks on soybean genome diversity. **Proceedings of the National Academy Sciences USA**; v.103, p.16666–16671, 2006.

KNOTHE, G. Some aspects of biodiesel oxidative stability. **Fuel Processing Technology**; v. 88, p. 669-677, 2007.

KONIECZNY, A.; AUSUBEL, F.M. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific per- based markers. **The Plant Journal**; v.4, p.403–410, 1993.

PINTO, M.O. **Expressão e substituição alélica dos genes de dessaturases de ácidos graxos em soja**. 2012. 80p. Tese (Doutorado) - Universidade Federal de Viçosa, Viçosa

PROCUNIER, J.D.; GRAY, M.; LIAKAT, A.M. Single nucleotide polymorphisms (SNPs) in hexaploid wheat and high Throughput SNP Detection by Invader Operation System. Proc. **Plant and Animal Genomes**. 11th conference., San-Diego. p.251, 2003.

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

RAMOS, M. J.; FERNÁNDEZ, C. M.; CASAS, A.; RODRÍGUEZ, L.; PÉREZ, A. Influence of fatty acid composition of raw materials on biodiesel properties. **Bioresource Technology**; v.100, p.261–268, 2009.

SCHMUTZ, J.; CANNON, S.B.; SCHLUETER, J.; MA, J.; MITROS, T.; NELSON, W.; HYTEN, D.L.; SONG, Q.; THELEN, J.J.; CHENG, J.; XU, D.; HELLSTEN, U.; MAY, G.D.; YU, Y.; SAKURAI, T.; UMEZAWA, T.; BHATTACHARYYA, M.K.; SANDHU, D.; VALLIYODAN, B.; LINDQUIST, E.; PETO, M.; GRANT, D.; SHU, S.; GOODSTEIN, D.; BARRY, K.; FUTRELL-GRIGGS, M.; ABERNATHY, B.; DU, J.; TIAN, Z.; ZHU, L.; GILL, N.; JOSHI, T.; LIBAULT, M.; SETHURAMAN, A.; ZHANG X.C.; SHINOZAKI, K.; NGUYEN, H.T.; WING, R.A.; CREGAN, P.; SPECHT, J.; GRIMWOOD, J.; ROKHSAR, D.; STACEY, G.; SHOEMAKER, R.C.; JACKSON, S.A. Genome sequence of the paleopolyploid soybean. **Nature**; v.463, p. 178-183. 2010.

USDA - Departamento de Agricultura dos Estados Unidos. Disponível em: <http://www.fas.usda.gov/oilseeds/circular/2012/April/oilseeds_full04-12.pdf>. **Acesso em:** agosto de 2012.

WEI, W.H.; CHEN, B.; YAN, X.H.; WANG, L.J.; ZHANG, H.F.; CHENG, J.P.; ZHOU, X.A.; AI-HUA SHA, A.H.; SHEN, H. Identification of differentially expressed genes in soybean seeds differing in oil content. *Plant Science*; v.175: p. 663–673, 2008.

WESELAKE, R.J.; TAYLOR, D.C.; RAHMAN, M.H.; SHAH, S.; LAROCHE, A.; MCVETTY, P.B.E.; HARWOOD, J.O.H.N. L. Increasing the flow of carbon into seed oil. **Biotechnology Advances**; v. 27, p. 866-78, 2009.

ZHU, Y.L.; SONG, Q.J.; HYTEN, D.L.; VANTASSELL, C.P.; MATUKUMALLI, L.K.; GRIMM, D.R.; HYATT, S.M.; FICKUS, E.W.; YOUNG, N.D.; CREGAN, P.B.. Single-nucleotide polymorphisms in soybean. **Genetics**; v.163, p.1123-1134, 2003.