



Universidade de Brasília

Departamento de Biologia Celular

Programa de Pós Graduação em Biologia Molecular

**Edição de genes associados à produção de ácidos graxos em soja.**

**Débora Torres Alves Figueirêdo**

Orientador: Prof. Dr. Elíbio Leopoldo Rech Filho

Brasília

**2017**

## **Edição de genes associados à produção de ácidos graxos em soja.**

**Débora Torres Alves Figueirêdo**

Orientador: Prof. Dr. Elíbio Leopoldo Rech Filho

Versão final da Tese de Doutorado apresentada a banca examinadora do Programa de Pós-graduação em Biologia Molecular do Departamento de Biologia Celular como requisito parcial para qualificação no Curso de Doutorado em Biologia Molecular.

Brasília

**2017**

Departamento de Biologia Celular  
Programa de Pós Graduação em Biologia Molecular

**Edição de genes associados a produção de ácidos graxos em soja.**

**Débora Torres Alves Figueirêdo**

Orientador: Prof. Dr. Elíbio Leopoldo Rech Filho

**Banca Examinadora:**

Prof<sup>a</sup>. Dr<sup>a</sup>. Nádia Skorupa Parachin (UnB)  
Examinadora

Prof<sup>a</sup>. Dr<sup>a</sup>. Maria Cristina Mattar da Silva (EMBRAPA)  
Examinadora

Prof<sup>a</sup>. Dr<sup>a</sup>. Angela Mehta dos Reis (EMBRAPA)  
Examinadora

Prof<sup>a</sup>. Dr<sup>a</sup>. Ana Cristina Miranda Brasileiro (EMBRAPA)  
Examinadora

Prof<sup>a</sup>. Dr<sup>a</sup>. Luciana Ramalho de Farias (UniCEUB)  
Suplente

Dedico este trabalho aos grandes amores da minha vida,  
Marcos Junior, Marcos Neto e Maria Alice  
e aos meus pais, Sonia Regina e José Paulo.

## AGRADECIMENTOS

Primeiramente gostaria de agradecer à Deus, por me guiar e nunca me abandonar. Por sempre olhar pelos meus projetos e meus caminhos, e a Nossa Senhora por sempre interceder pelas minhas necessidades

Ao Professor Dr. Elíbio Rech, por me receber como orientanda, por me acolher em seu laboratório e permitir que eu faça parte da sua equipe, pela oportunidade de realizar este trabalho, pela paciência e por todo o conhecimento compartilhado. Obrigada por ser este exemplo de cientista.

Aos membros da banca, Prof<sup>a</sup>. Dr<sup>a</sup>. Ana Cristina Miranda Brasileiro, Prof<sup>a</sup>. Dr<sup>a</sup>. Angela Mehta dos Reis, Prof<sup>a</sup>. Dr<sup>a</sup>. Luciana Ramalho de Farias, Prof<sup>a</sup>. Dr<sup>a</sup>. Maria Cristina Mattar da Silva e a Prof<sup>a</sup>. Dr<sup>a</sup>. Nádia Skorupa Parachin, por aceitarem participar da banca e por todas as contribuições.

Ao meu amado marido Marcos Figueiredo Junior, pelo companheirismo, por nunca desistir de mim, por ser um exemplo, pelos ensinamentos, por sempre me apoiar, por fazer parte da minha vida, da minha história. Muito Obrigada!!

Aos meus filhos Marcos Neto e Maria Alice. Muito obrigada por me ensinarem a ser uma pessoa melhor. Obrigada por me inspirar a lutar por um futuro melhor, a querer mais e ir sempre além. Pelo amor incondicional, pela paciência e pelas palavras simples e sinceras, meu muito obrigado. Amo muito vocês!!!

Aos meus pais, José Paulo e Sonia Regina, pelo amor incondicional. Por me ensinarem tudo o que eu sei e terem feito de mim o que eu sou hoje. Muito obrigada pelos ensinamentos, conselhos, palavras de incentivo, amor, enfim por tudo.

Às minhas avós Maria Katuco, Ailce Brandão. Obrigada por serem exemplos de vida e perseverança.

À minha “avó” de coração, Guiomar Vieira. Obrigada por tantos ensinamentos, pelo exemplo, por ser sempre um refugio para minha pessoa e para os meus filhos. Obrigada por sempre cuidar tão bem de nós, me dando segurança para seguir meus caminhos e meus sonhos.

Àos meus sogros Marcos Antonio e Leniz Genil, por me acompanharem em mais essa jornada. Muito obrigada pelo apoio constante, pela amizade e amor compartilhado durante todo esse período.

Aos meus irmãos, José Paulo Junior, Felipe e Beatriz, por serem figuras tão importantes na minha vida. Obrigada pela paciência, pela amizade, pelos puxões de orelha e pela eterna disponibilidade em ajudar. Sem vocês eu não estaria aqui!!!

Aos meus cunhados Carlos Ragassi, Luciana Neves, Artur, Lucianna Figueiredo e João Paulo Figueiredo e a minha sobrinha Cecilia Alves. Muito obrigada por todo incentivo, amizade, colaboração e carinho durante esse tempo.

A toda a minha família, tios, tias, primos e primas, que sempre torceram por mim.

À Prof<sup>a</sup>. Dr<sup>a</sup>. Izabela Marques Dourado Bastos e ao Prof Dr Octavio Luiz Franco, por me acolherem em seus laboratórios, me ensinarem a caminhar e a crescer cientificamente, e a todos que de alguma forma me ajudaram durante minha vida acadêmica.

Gostaria de agradecer também aos meus amigos do Laboratório de Biologia Sintética da EMBRAPA Recursos Genéticos e Biotecnologia: Nicolau Brito, Cristiano Lacorte, Gabriel Vasconcelos, Débora Sant'Ana, André Murad, Marly Felipe, Otávio Bravim, Thaina Lima, Marina Silva, Andressa Rezende, Jessica Fernandes, Juliana Viana.

Agradeço especialmente ao Prof. Dr. Giovanni Vianna e a Prof Dra. Cintia Coelho, pelo apoio científico e técnico, pela contribuição com idéias, ensinamentos, análises e críticas para o desenvolvimento do trabalho.

Agradeço ao amigo Gustavo José, que muito contribuiu para o desenvolvimento desta tese.

Gostaria de agradecer especialmente as minhas amigas Jessica Carrijo e Valquíria Lacerda, que sempre me incentivaram, apoiaram e me ajudaram em tudo o que eu precisei. Muito Obrigada pela amizade e por nunca deixarem eu desistir.

Agradeço a EMBRAPA, CAPES, e ao Programa de Pós Graduação em Biologia Molecular, pelo apoio financeiro e material deste trabalho.

Sometimes the world seems against you  
The journey may leave a scar  
But scars can heal and reveal just  
Where you are  
The people you love will change you  
The things you have learned will guide you  
And nothing on Earth can silence  
The quiet voice still inside you  
And when that voice starts to whisper  
\_\_\_\_\_, you've come so far  
\_\_\_\_\_, listen  
Do you know who you are?

Música: I Am Moana  
Do filme "Moana" - Disney  
*Musica e letra por Lin-Manuel Miranda e Mark Mancina*

## LISTA DE FIGURAS

Figura 1: **Estrutura química dos ácidos graxos da soja.**

Figura 2: **Representação esquemática das reações envolvidas na biossíntese dos ácidos graxos.**

Figura 3: **Estrutura dos Zinc Finger Nucleases.** A, domínio C2H2, onde o íon de zinco coordena os resíduos de Cisteína e Histidina. B, cada ZF interage com 3 nucleotídeos do DNA. C, representação esquemática do ZFN interagindo com o DNA, e a dimerização da FokI para a clivagem do DNA (Adaptado de Fichtner et al., 2014).

Figura 4: **Representação do Sistema CRISPR-Cas.** Fases do Sistema CRISPR-Cas: Aquisição, Expressão e Interferência. 1- Fase da Aquisição onde o DNA invasor é integrado ao *locus* CRISPR. 2 -Fase de Expressão, transcrição do *loci* CRISPR, e processamento do Pre-crRNA para liberação dos crRNA. 3- Fase de Interferência, clivagem do DNA invasor. Adaptado de Bhaya et al., 2011.

Figura 5: **Sistema CRISPR-Cas9.** A, Sistema Tipo II CRISPR-Cas9 original. B, CRISPR-Cas9 quimera. Adaptado de Barrangou 2012.

Figura 6: **Estrutura do TALEN.** A representação esquemática de um par de TALEN, onde cada TALE é composto de 33-35 aminoácidos, e os aminoácidos nas posições 12 e 13, referidos como RVDs (em vermelho), são responsáveis pelo reconhecimento da base nucleica. B, Estrutura em forma de V de cada TALE, onde o segundo resíduo RVD faz contato com o DNA enquanto o primeiro estabiliza a estrutura. Adaptado de Kim & Kim, 2014.

Figura 7: **Desenho representativo dos vetores pBI\_TAL3\_5 e pBI\_TAL\_3\_6.** Os vetores TALEN pTAL\_3\_5 e pTAL\_3\_6 foram digeridos com as enzimas *Bgl* II e *Sac* I e clonados no mesmo sítio do vetor pBI426, gerando os vetores pBI\_TAL3\_5 e pBI\_TAL\_3\_6.

Figura 8: **Alinhamento das sequências dos genes *FAD2-1A (Glyma 10)* e *FAD2-1B (Glyma20)*.** As sequências dos genes foram alinhadas e a região alvo foi determinada com o auxílio do programa TALE-NT. Glyma10 representa a sequência do gene *Glyma10g42470* e Glyma20 representa a sequência do gene *Glyma20g24530*. Em amarelo observamos as sequências alvo. Em verde está destacado a sequência espaçadora, necessária para a enzima FokI.

Figura 9: **Representação esquemática do par de TALEN.** Em amarelo observamos a sequência-alvo e seus respectivos RVDs. Cada RVD reconhece uma única base do DNA. Os RVDs estão representados em cores onde: azul NN, reconhece a base adenina; laranja HD, reconhece citosina; verde NG reconhece timina, e roxo NN reconhece guanina.

Figura 10: **Análise de restrição dos vetores pBITAL 3.5 (A) e pBITAL 3.6 (B) com a enzima *PvuII* em gel de agarose.** Os vetores pBITAL 3.5 e pBITAL 3.6 foram digeridos para confirmação do padrão de tamanho do vetor, onde foram observados 4 fragmentos com tamanhos aproximados 620 pb, 1.000 pb, 2.350 pb, 3.800 pb. MM: marcador molecular Gene Ruler™ 1Kb ladder (Invitrogen®).

Figura 11: **Amplificação dos fragmentos do vetor pBITAL3\_5 e pBITAL3\_6.** O DNA das plantas transformadas foi extraído e amplificado para confirmação da introdução dos vetores, utilizando o par de iniciadores TAL 353 e TAL 354. 1, planta 137; 2, planta 138; 3, controle pBITAL 3\_5; 4, controle pBITAL 3\_6. 5, controle negativo. 6, branco. MM, marcador molecular 1Kb Plus DNA (Invitrogen).

Figura 12: **Perfil de ácidos graxos das sementes T1 de soja.** As porcentagens de ácidos graxos foram determinadas a partir de reação de esterificação e analisadas em GC-MS. As porcentagens dos ácidos graxos palmítico, esteárico, oleico, linoleico e linolênico foram obtidas para as sementes de soja T1 das plantas 57, 119 e 137 e para as plantas BR16 não transformadas (WT). As linhas verticais representam desvio padrão encontrados nas 5 sementes analisadas. Asteriscos representam diferença sigficativo quando comparado a BR16 (Teste T-Student; \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ ).

Figura 13: **Análise da amplificação do gene *Glyma10* em gel de eletroforese capilar.** Utilizando os primers específicos Glyma10F e Glyma10R, o gene *Glyma10* foi amplificado e pequenas diferenças de tamanho foram observadas. A1, planta 137-T0; A2, planta 119-9; A3, planta 119-22; A4, planta 119-24; A5, planta 119-37; A6, planta 119-38; A7, água; A8, planta 137-32; A9, controle; A10, planta 119-T0;

Figure 14: **Amplificação do gene *Glyma20* em gel de eletroforese capilar.** A, Utilizando os primers específicos Glyma20F e Glyma20R, o gene *Glyma20* foi amplificado e pequenas diferenças de tamanho foram observadas. B1, planta 119-37; B2, planta 119-38; B3, planta 137-32; B4, controle; B5, água; B6, planta 119-T0; B7, planta 137-T0; B8, planta 119-9; B9, planta 119-22; B10, planta 119-24; B, Eletroferograma das plantas 119-38, 119-22 e 119-24, mostrando os picos de fluorescência e o tamanho em pares de base.

Figura 15: **Mutações induzidas por TALENs no gene *Glyma10 (FAD2-1A)* nas plantas transformadas.** Representação das mutações observadas na planta 119-24. O sítio alvo do par de TALEN está indicado em amarelo e em verde está indicado o espaço para atividade da enzima FokI. Deleções estão indicadas por traços.

Figura 16: **Mutações induzidas por TALENs no gene *Glyma20 (FAD2-1B)* nas plantas transformadas.** Representação de todas as mutações observadas nas diferentes plantas. O sítio alvo do par de TALEN está indicado em vermelho e em rosa está indicado o espaço para atividade da enzima Fok I. Deleções estão indicadas por traços, e estrelas marcam sequências conservadas.

Figura 17: **Alinhamento dos genes das família FAD2-1 e FAD2-2.** Os genes da família FAD2-1, correspondentes aos genes *Glma10* e *Glyma20*, foram alinhados com os genes da família FAD2-2 (*Glyma09*, *Glyma03* e *Glyma19*). Pares de base destacados na cor amarela representam as regiões direita e esquerda escolhidas como alvo do TALEN neste trabalho. Pares de base destacados em verdes representam a região espaçodara necessária para a atividade da enzima Fok I.

Figura 18: **Amplificação do gene *Glyma03* em gel de eletroforese capilar.** Utilizando os primers específicos Glyma03F e Glyma03R, o gene *Glyma03* foi amplificado e pequenas diferenças de tamanho foram observadas. C1, marcador molecular; C2, planta 119-T0; C3, planta 137-T0; C4, planta 119-9; C5, planta 119-22; C6, planta 119-24; C7, água; C8, planta 119-37; C9, planta 119-38; C10, planta 137-32; C11, controle;

Figura 19: **Amplificação do gene *Glyma19* em gel de eletroforese capilar.** Utilizando os primers específicos Glyma19F e Glyma19R, o gene *Glyma19* foi amplificado e pequenas diferenças de tamanho foram observadas. D1, planta 119-22; D2, planta 119-24; D3, planta 119-37; D4, planta 119-38; D5, planta 137-32; D6, controle; D7, água; D8, planta 119-T0; D9, planta 137-T0; D10, planta 119-9;

Figura 20: **Mutações induzidas por TALENs no gene *Glyma03* nas plantas transformadas.** Representação das mutações observadas nas diferentes plantas transformadas. O possível sítio alvo do par de TALEN está indicado em amarelo e em azul está indicado o espaço para atividade da enzima Fok I. Deleções estão indicadas por traços e estrelas indicam regiões conservadas.

Figura 21: **Mutações induzidas por TALENs no gene *Glyma19* nas plantas transformadas.** Representação das mutações observadas nas diferentes plantas transformadas. O possível sítio alvo do par de TALEN está indicado em vermelho e em roxo está indicado o espaço para atividade da enzima Fok I. Deleções estão indicadas por traços e estrelas indicam regiões conservadas.

Figura 22: **Alinhamento das sequencias alvo e espaçadora dos TALENs utilizados neste trabalho e os utilizados no trabalho de Haun e colaboradores (2014).** Glyma 10 e Glyma 20 corresponde a região alvo do TALENs utilizados neste trabalho para o gene *Glyma10* e *Glyma 20*, respectivamente. HaunFAD2\_T02Glyma10 e HaunFAD2\_T02Glyma20 correspondem ao par de TALEN denominado FAD2-T02 no trabalho de Haun e colaboradores (2014). Em amarelo estão destacado os pares de base alvos do TALEN, e em verde está destacado a sequencia espaçadora.

Figura 23: **Alinhamento das sequências dos genes *Glyma10* e *Glyma20*, e comparação das regiões alvos selecionadas neste trabalho e no trabalho de Haun e colaboradores (2014).** Em amarelo e em verde estão destacados as sequências alvo do TALEN utilizado neste trabalho e a sequência espaçadora. Em azul e roxo está destacado a sequência alvo e a sequência espaçadora utilizada no trabalho de Haun.

## LISTA DE TABELAS

Tabela 1: Produção de grão de soja nas safras de 2014/15 e 2015/16 (milhões de toneladas) e Exportação Mundial do grão de soja em 2015/16 (milhões de toneladas).

Tabela 2. Sequência dos iniciadores utilizados para amplificação dos genes por meio de PCR.

Tabela 3. Média da porcentagem das concentrações dos ácidos graxos em sementes de sojas nas linhagens transformadas.

Tabela 4. Resultado da amplificação do gene *Glyma10*.

Tabela 5. Resultado da amplificação do gene *Glyma20*.

Tabela 6. Análise das mutações encontradas para o gene *Glyma10* nas diferentes linhagens e o perfil de oleico e linoleico encontrado em cada linhagem.

Tabela 7. Análise das mutações encontradas para o gene *Glyma20* nas diferentes linhagens e o perfil de oleico e linoleico encontrado em cada linhagem.

Tabela 8. Resumo dos diferentes tipos de mutações encontrados nos genes *Glyma10* e *Glyma20* e a frequência de cada tipo de mutação.

Tabela 9. Resultado da amplificação do gene *Glyma03*.

Tabela 10. Resultado da amplificação do gene *Glyma19*.

Tabela 11: Análise das mutações encontradas nos genes *Glyma03* nas diferentes linhagens obtidas.

Tabela 12: Análise das mutações encontradas nos genes *Glyma19* nas diferentes linhagens obtidas.

Tabela 13: Resumo das diferentes mutações encontradas nos genes *Glyma03* e *Glyma19*.

**LISTA DE ABREVIATURAS**

Da	Dalton
DNA	Ácido desoxirribonucléico
RNA	Ácido ribonucléico
mRNA	RNA mensageiro
dNTP	Desoxirribonucleotídeos fosfatados
EDTA	Etileno bis(oxi-etilenonitrilo) do ácido tetraacético
IPTG	Isopropil- $\beta$ -D-tiogalactopiranosídeo
kb	Kilobase
pb	Pares de bases
PBS	Tampão fosfato 50 mM, NaCl 0,15 M pH 7,2
LB	Meio de cultura Luria-Bertani
PCR	reação de polimerização em cadeia
X-Gal	5-bromo-4-cloro-3-indolil-beta-D-galactopiranosídeo
GEEN	Edição de denoma com nucleases modificadas
RNAi	RNA de interferência
miRNA	microRNA
siRNA	Pequeno RNA de interferência
RdRP	Enzimas RNA dependente de RNA polimerase
DsRNA	RNA de fita dupla
RISC	Complexo de indução do Silenciamento do RNA
PTGS	Silenciamento gênico pós-transcricional
ZFN	Nuclease de dedo de Zinco
ZF	Dedo de Zinco
ZFP-TF	Proteínas Zinc-Finger ligas a um domínio de ativação transcricional
TLP	<i>Trait landing pads</i>
TALEN	Nucleases com efetores do tipo ativador transcricional
TALE	Efetores do tipo ativador transcricional

CRISPR	Repetições Palindrômicas Curtas Agrupadas e Regularmente Interespaçadas
PAM	Motivo Adjacente ao Protoespaçador
DBS	Quebra da dupla fita
HR	Recombinação Homóloga
NHEJ	Junção de extremidades não homólogas
<i>E. coli</i>	<i>Escherichia coli</i>
AK	Aspartato quinase
Pre-crRNA	Precursor CRISPR RNA
crRNA	CRISPR RNA
sgRNA	único RNA guia
tracrRNA	RNA transativador
Indel	Inserções e Deleções
FAD	Ácido Graxo dessaturase

## SUMÁRIO

<b>INTRODUÇÃO .....</b>	<b>1</b>
<b>REVISÃO BIBLIOGRÁFICA .....</b>	<b>3</b>
1. A cultura da Soja .....	4
2. Engenharia Metabólica.....	7
3. Biosíntese de ácidos graxos .....	9
4. Família FAD2.....	11
5. O Genoma da planta de Soja .....	13
6. Estratégias genômicas atuais para a aplicação na melhoria de características alvo.....	14
6.1. RNA de Interferência .....	15
6.2. Metodologias de Edição de Genoma:.....	17
<b>HIPÓTESE .....</b>	<b>33</b>
<b>OBJETIVO .....</b>	<b>35</b>
<b>METODOLOGIA .....</b>	<b>37</b>
1. Construção dos vetores TALEN.....	38
2. Vetor de recombinação pREC56 .....	39
3. Transformação das plantas de soja por biobalística .....	39
4. Extração de DNA genômico.....	42
5. Oligonucleotídeos Iniciadores.....	43
6. Análise por PCR das plantas transgênicas.....	44
7. Eletroforese em gel de agarose.....	45
8. PCR-Eletroforese Capilar .....	45
9. Clonagem dos fragmentos de PCR.....	45
10. Transformação em bactérias competentes .....	46
11. Extração de DNA Plasmideal .....	46
12. Esterificação dos ácidos graxos. ....	47
13. Análise por GC-MS.....	48
<b>RESULTADOS.....</b>	<b>49</b>
<b>RESULTADOS .....</b>	<b>50</b>

<b>1. Desenho do par de TALEN .....</b>	<b>50</b>
<b>2. Obtenção de plantas de soja transgênica expressando os vetores TALEN.....</b>	<b>53</b>
<b>3. Análise do perfil de óleo das sementes de soja das plantas transgênicas.....</b>	<b>54</b>
<b>4. Sequenciamento das plantas com perfil de óleo alterado .....</b>	<b>56</b>
<b>5. Busca por potenciais genes <i>Não alvos</i> .....</b>	<b>65</b>
<b>DISCUSSÃO .....</b>	<b>74</b>
<b>CONCLUSÃO .....</b>	<b>86</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>89</b>

## RESUMO

A soja (*Glycine max*) é uma das principais comódites agrícola do mundo. Por conter um alto teor de óleo, a semente de soja é de grande interesse tanto para os programas de produção de biodiesel, como para a indústria alimentícia. Sementes de soja apresentam um perfil de ácidos graxos composto principalmente de 13% de ácido palmítico (16:0), 4% de ácido esteárico (18:0), 18% de ácido oleico (18:1), 55% de ácido linoleico (18:2) e 10% de ácido linolênico (18:3). Assim, o óleo da semente de soja é rico em ácidos graxos poli-insaturados, o que contribui para a sua instabilidade oxidativa. Uma alternativa para a redução de teores de ácidos graxos poli-insaturados em semente de soja é a utilização da engenharia metabólica para diminuir a expressão dos genes *FAD2*. Durante a síntese dos ácidos graxos sabe-se que os genes *FAD2-1A* e *FAD2-1B* são importantes na conversão do óleo monoinsaturado (oleico) em óleo poli-insaturado (linoleico). Neste trabalho *Transcription activator-like effector nucleases* (TALENs) foram desenhados a fim de obter mutações nos genes *FAD2-1A* e *FAD2-1B* utilizando um sítio-alvo conservado em ambos os genes. Três linhagens transgênicas de soja (57, 119 e 137) transformadas com os vetores TALEN, foram analisadas quanto ao seu perfil de óleos, mostrando aumento na concentração do óleo oleico (variação de 31-66%, contra 18% normalmente produzido pelo tipo selvagem). A região-alvo do TALEN foi sequenciada e deleções, inserções e mutações foram observadas, sendo também transmitidas para a próxima geração. Os genes da família *FAD2* foram analisados como possíveis genes não alvos, já que os genes *FAD2-1*, *FAD2-2*, e *FAD2-3* compartilham alta similaridade de sequências. Os candidatos não alvos apresentaram significativos índices de atividade do TALEN. Este trabalho foi importante para a consolidação da metodologia TALEN em plantas, e para avaliar os níveis de não alvo deste sistema.

## ABSTRACT

Soybean is one of the most important agricultural commodities in the world. Because soybean seeds contain a high content of oil, they are of great interest as for biodiesel production programs as for the food industry. Soybean seeds have a fatty acid profile composed of 13% palmitic acid (16: 0), 4% stearic acid (18: 0), 18% oleic acid (18: 1), 55% linoleic acid (18 : 2) and 10% linolenic acid (18: 3). Thus, the soybean oil is rich in polyunsaturated fatty acids, which contributes to its oxidative instability. An alternative to reduce the levels of polyunsaturated fatty acids in soybean seeds is the use of metabolic engineering to reduce the *FAD2* gene expression. During the synthesis of fatty acids it is known that *FAD2-1A* and *FAD2-1B* genes are important in the conversion of monounsaturated fats in polyunsaturated fats. In this work Transcription activator-like effector nucleases (TALENs) were designed in order to obtain mutations in *FAD2-1A* and *FAD2-1B* genes using a conserved target site in both genes. Three soybean lines (57, 119 and 137) transformed with TALEN vectors were analyzed, and the oil profile showed increases in oleic concentration (ranging from 31-66%, against 19% normally produced by the wild type). The target regions of TALEN were sequenced and deletions, insertions and mutations were observed, and mutations, which were also transmitted to the next generation. *FAD2* gene family have been examined as possible off targets, since the genes *FAD2-1*, *FAD2-2* and *FAD2-3* sequences shares high similarity. The off-targets candidates were PCR amplified and their products were sequenced, showing significant off-target rates. This work was important for the consolidation of this system in plants, and to reevaluate its off-target range.

## INTRODUÇÃO

## INTRODUÇÃO

O biodiesel vem se destacando como uma importante fonte de energia renovável e sustentável. Se comparado ao petróleo, a emissão de poluentes na atmosfera pela queima (combustão) do biodiesel é significativamente menor, causando menor impacto ao meio ambiente. A demanda crescente por energia de forma sustentável é um constante desafio e terá cada vez mais importância. Uma das grandes preocupações mundiais está na busca por alternativas energéticas que possam substituir os combustíveis fósseis.

Entre as matérias-primas usadas para a produção do biodiesel, a soja (*Glycine max* (L. Merrill)) se destaca pelo alto teor e qualidade do óleo de suas sementes e pelo domínio do sistema de produção em larga escala, com baixo custos quando comparado a qualquer outro óleo (Ribeiro, 2006). A soja por conter um alto teor de óleo, é considerada uma planta de grande interesse para os programas de produção de biodiesel, desde que seja efetuado o desenvolvimento de linhagens de plantas com teores reduzido de ácido linoleico e palmítico. Além de ser matéria prima para produção de biodiesel, o óleo da soja também é uma importante matéria-prima para a indústria alimentícia.

Uma alternativa para a melhoria do conteúdo de ácidos graxos em semente de soja é a utilização da engenharia metabólica. Diversos estudos tiveram como foco diminuir a expressão dos genes da família FAD2. Os genes *FAD2-1*, são responsáveis pela conversão de ácido oleico em ácido linolênico, aumentando a concentração do óleo oleico que além de melhorar seu valor nutricional, também aumenta a performance deste óleo como um biodiesel. Estudos manipulando as isoformas FAD2-1A e FAD2-1B demonstraram que, utilizando a engenharia genética metabólica com o intuito de diminuir a expressão desses genes, foi possível obter sementes de soja com conteúdo de óleo oleico elevado (Murad et al., 2014).

Com o avanço das técnicas de edição de genoma ficou, cada vez mais prático e rápido obter plantas expressando características desejáveis. O desenvolvimento de ferramentas moleculares, tais como dedos de zinco (ZFN) TALENs e CRISPR/Cas9 vem promovendo avanços na pesquisa de plantas, assim como na melhoria de plantações. Estas metodologias são capazes de realizar uma quebra na dupla fita de DNA, que será naturalmente reparado por uma das vias do sistema de reparo da célula, que poderá realizar modificações tais como inserção ou deleção no local do reparo.

Dentre estas metodologias, TALEN foi apontado como sendo mais específico que ZFN e CRISPR/Cas9, uma vez que pode ter como alvo sequências de DNA pequenas, como microRNAs e potenciadores, que não poderiam ser alvos de ZFN ou CRISPR (Neff et al., 2013). Além disso, TALENs atuam como dímeros para a clivagem do DNA, o que influencia para a alta especificidade desta metodologia, uma vez que dificilmente ambos os TALENs se ligaram ao mesmo local fora do alvo (Koo et al., 2015). TALEN tem sido utilizado com sucesso em uma variedade de plantas, incluindo *Arabidopsis thaliana*, cevada, *Brachypodium spp.*, milho, batata, arroz, soja, tabaco, tomate e trigo (Lee et al., 2016). Apesar do significativo avanço desta tecnologia, muitos aspectos centrais ainda permanecem obscuros. Uma importante preocupação é a atividade fora do alvo desta ferramenta, que ainda necessita ser melhor elucidada.

Diante deste contexto, este trabalho visou produzir plantas de soja contendo um perfil de óleo favorável à indústria de biodiesel, assim como para a indústria alimentícia, aumentando a concentração do óleo oleico nas sementes de soja, utilizando a tecnologia de TALEN. Tendo conhecimento de que esta tecnologia pode gerar mutações fora do alvo, genes que são relativamente próximos à sequência-alvo, os genes da família FAD2 foram analisados como potenciais genes afetados.

## **REVISÃO BIBLIOGÁFICA**

## REVISÃO BIBLIOGRÁFICA

### 1. A cultura da Soja

A soja (*Glycine max* L.) é uma das principais comódites agrícola do mundo, sendo utilizada para diversas finalidades como fonte de alimento e extração de óleo, e devido a esta versatilidade, o complexo soja (grão, farelo e óleo) apresenta alto valor econômico a nível nacional e internacional (SILVA et al., 2006).

O Brasil é atualmente o maior exportador global do grão de soja e o segundo maior produtor do grão soja no mundo, produzindo na safra de 2015/16 cerca de 96,5 milhões de toneladas, dos quais 54,4 milhões de toneladas foram utilizadas para a exportação (Tabela 1). O complexo soja (grão, farelo e óleo) representou 13% das exportações totais do Brasil em 2015, superando o minério de ferro e liderando as exportações da balança comercial, sendo o principal produto de exportação do Brasil (Secretaria de Comércio Exterior- SECEX).

Tabela 1: Produção de grão de soja nas safras de 2014/15 e 2015/16 (milhões de toneladas) e Exportação Mundial do grão de soja em 2015/16 (milhões de toneladas).

Países	Safras		Exportação
	2014/15	2015/16	2015/16
<b>EUA</b>	106,88	106,86	52,70
<b>Brasil</b>	97,20	96,50	54,40
<b>Argentina</b>	61,40	56,80	9,90
<b>China</b>	12,15	11,79	0,11
<b>Demais</b>	42,15	41,02	15,19
<b>Mundo</b>	319,78	312,97	132,30

Fonte: USDA

As sementes de soja fornecem cerca de 27% da oferta mundial de óleos e 44% da oferta mundial de proteínas (Li et al., 2015). A qualidade das sementes é determinada pela sua composição, incluindo proteína, ácidos graxos, açúcares e minerais (Bellaloui et. al., 2015). As sementes de soja possuem uma composição nutricional de em média 40% de proteínas, 19% de lipídeos, 23% de carboidratos e 15% de fibras alimentares.

O óleo da semente de soja é composto predominantemente por cinco tipos de ácidos graxos: o ácido palmítico (16:0), o ácido esteárico (18:0), o ácido oleico (18:1), o ácido linoleico (18:2) e o ácido linolênico(18:3) (Figura 1), e a porcentagem destes cinco ácidos graxos no óleo de soja corresponde em média a 13%, 4%, 18%, 55% e 10%, respectivamente (Clemente & Cahoon, 2009). Os ácidos oleico, linoleico e linolênicos são considerados insaturados, pois apresentam uma ou mais dupla ligação ao longo da cadeia de carbono, e a posição desta dupla ligação pode ser indicada pelo símbolo  $\Delta$ , quando a primeira dupla ligação é contada a partir do grupo funcional ou  $\omega$  (n), quando a primeira dupla ligação é contada a partir do radical metil, como no caso do , onde a primeira dupla ligação está presente no carbo 9, contado a partir do grupo funcional (Rose & Connolly, 1999).

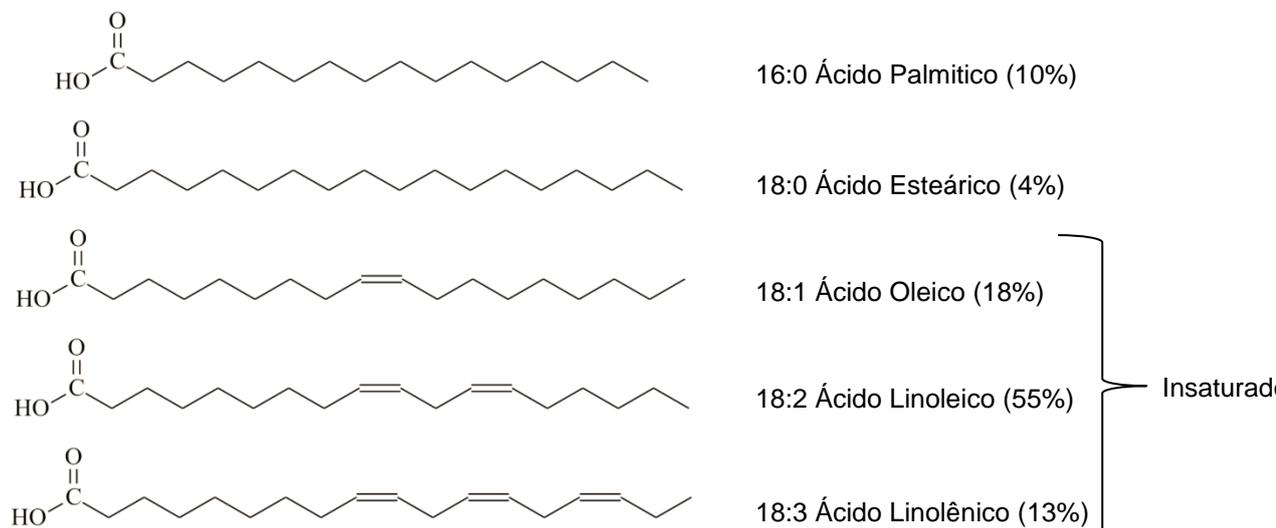


Figura 1: **Estrutura química dos ácidos graxos da soja.**

A qualidade do óleo da semente de soja depende da composição dos ácidos graxos e influência nas propriedades físico-químicas e biológicas do óleo. O perfil de ácidos graxos da soja resulta em baixa estabilidade oxidativa, tendo em vista que os ácidos graxos poli-insaturados são mais suscetíveis às hidroperoxidações, que resultam em cetonas e aldeídos voláteis que alteram as propriedades físico-químicas do óleo, o que limita o uso do óleo de soja em produtos alimentícios e sua aplicação na indústria (Cardinal et al., 2011). A oxidação do óleo de soja propicia a produção de peróxidos, que podem se decompor em produtos secundários como aldeídos, álcoois, ácidos, que são responsáveis pela deterioração do óleo, resultando em sabores indesejados e rancidez; e quando utilizado como biodiesel a oxidação do óleo de soja pode influenciar no acúmulo de substâncias viscosas que entopem os filtros de óleo (Canakci et al., 1999). Os ácidos graxos poli-insaturados (linoleico e linolênico) estão diretamente relacionados com processos oxidativos, visto que possuem um maior número de duplas ligações, que são suscetíveis à oxidação.

Para aumentar a estabilidade oxidativa do óleo da semente de soja algumas estratégias podem ser adotadas como a adição de antioxidantes sintéticos, a hidrogenação e a diminuição do conteúdo dos ácidos poli-insaturados. Os antioxidantes sintéticos mais utilizados são o butil-hidroxianisol (BHA), o butil-hidroxitolueno (BHT) e o t-butil-hidroquinona (TBHQ), mas apesar de prevenir a oxidação do óleo de soja, possuem uso restrito, pois apresentam efeitos carcinogênicos e tóxicos em humanos (Nascimento et al., 2014).

A hidrogenação pode ser utilizada para aumentar a estabilidade do óleo, mas gera quantidades significativas de gorduras *trans* que estão ligadas a doenças cardiovasculares (Yadav, 1996). Além disso, a hidrogenação do óleo de soja, entre outras coisas, reduz a lubricidade e aumenta a viscosidade, limitando assim seu valor como biodiesel (Moser et al., 2007).

Estudos demonstram que altas concentrações de ácido oleico podem aumentar a estabilidade oxidativa do óleo, a sua utilidade em altas

temperaturas, e o uso do óleo de soja com fins farmacêuticos, cosméticos e em produtos industriais como o biodiesel (Lee et al., 2009).

## **2. Engenharia Metabólica**

A engenharia de rotas metabólicas consiste na manipulação das rotas metabólicas existentes ou inclusão de novas rotas, a fim de melhorar processos genéticos e regulatórios para favorecer a produção de produtos ou propriedades celulares, com minimização do gasto energético celular (Bayley et al., 1991). No entanto, devido à complexidade das redes metabólicas, tanto em termos de estrutura como em termos de regulação, é difícil prever os efeitos das modificações genéticas sobre um determinado produto, já que muitos produtos participam de diversas reações químicas, e alterando um desses produtos alteram-se todas as demais reações (Patil, et al., 2005).

E dentre as diversas áreas de atuação, a engenharia de rotas metabólicas pode atuar modificando microrganismos especificamente para produzir produtos que respondam as necessidades do homem como: atuar na produção de medicamentos, vacinas e agentes de diagnóstico de doenças; atuar na detecção de poluentes e sua degradação ou remoção do ambiente e produção de compostos químicos; produzir alimentos mais nutritivos, que cause menos reações adversas, ou que aumente ou diminua o rendimento de produtos de interesse biotecnológico em plantas (Howard et al., 2013; Avalos & Fink, 2013; Kondo et al., 2012; O’Keefe et al., 2015; Cunha et al., 2011; Cunha et al., 2010).

No contexto da soja, melhorar a qualidade das sementes é cada vez mais necessário, visto que as sementes com um teor maior de proteínas são de interesse pecuário, devido a adição no farelo oferecido ao gado. As sementes de soja promovem uma rica fonte de proteína, já que 40% da sua composição é de proteína. Devido a esta quantidade, acredita-se ser possível expressar grandes quantidades de proteínas recombinantes em sementes de soja (Hudson et al., 2014). Por outro lado, sementes de soja com um alto teor de óleo oleico e baixo teor de óleo linolênico é interessante, pois contribuiria

para a estabilidade oxidativa do óleo, aumentando a durabilidade do produto nas prateleiras. Acrescenta-se às características do sistema de produção de soja as vantagens relacionadas ao seu cultivo. A soja possui um sistema de produção relativamente mais econômico para a produção em larga escala, e práticas agronômicas e infraestrutura necessárias para a colheita, armazenamento e processamento, já serem bem estabelecidos quando comparado a outras oleogênicas. Mono e dissacarídeos são também já foram considerados alvos para o aprimoramento do gosto da soja. Já o aumento da estaquiose (um tetrassacarídeo) nas sementes é indesejável, devido a sua contribuição para a flatulência ou diarreia em não ruminantes, como frango e porco (Bellaloui et al., 2015).

O melhoramento em rotas metabólicas de soja também já foi realizado, no intuito de aumentar a quantidade de treonina, um dos aminoácidos essenciais na indústria de alimentação animal. A quantidade desse aminoácido nas rações pode afetar a produção de fontes de carnes, como suínos e aves. A superexpressão da enzima *feedback-resistant aspartate kinase* (AK) aumentou os níveis de treonina livre em 100 vezes, e apresentou também um aumento substancial nos níveis de outros aminoácidos livres, resultando em um aumento de até 3.5 vezes o teor total de aminoácidos livres (Qi et al., 2011). Além disso, outros aminoácidos também já foram relatados, por terem seus níveis aumentados, utilizando a engenharia de rotas metabólicas. A cisteína é um outro exemplo de aminoácido cuja quantidade em sementes de soja foi aumentada significativamente por meio da expressão de uma enzima chave da assimilação de enxofre (OASS), resultando em 58-74% de aumento nos níveis de cisteína ligados a proteínas e 22-32% de aumento nos níveis de cisteína livre, além de um aumento na acumulação de inibidor de protease *Bowman-Birk*, uma proteína rica em cisteína (Kim et al., 2012).

Produtos alimentares com baixos teores de gordura trans tem sido foco do mercado mundial, tendo em vista que a preocupação com este tipo de gordura aumentou, ainda mais, após o mandato da *Food and Drug Administration* (FDA) em 2006, que obriga a informação em todos os produtos, da porcentagem de gorduras trans em seus rótulos. Óleos com elevados teores

de ácido oleico exigem menos processamento oferecendo uma rota para diminuir ainda mais os níveis de gordura trans nos produtos alimentares. Muitos trabalhos têm sido realizados utilizando como alvo o gene da enzima  $\Delta 12$  ácido graxo dessaturase (*GmFAD2*), uma vez que esta enzima é responsável pela conversão do oleico em linoleico. Uma mutação no gene *GmFAD2-1A* levou a interrupção da via de biossíntese do linoleico a partir do ácido oleico, resultando em um aumento de 46% no conteúdo de oleico em sementes de soja Bay, gerando a linhagem M23 (Alt et al., 2005).

Assim a engenharia de rotas metabólicas se apresenta como um importante estratégia para a descoberta e manipulação de enzimas chaves em rotas metabólicas pelo homem.

### 3. Biosíntese de ácidos graxos

A biossíntese dos ácidos graxos se inicia com a formação de acetil-CoA a partir do piruvato. Este acetil-CoA é ativado para malonil-CoA, pela enzima acetil-CoA carboxilase, e o malonil-CoA será então transferido para uma proteína carreadora de grupo acil (ACP), gerando malonil-ACP, o substrato primário da síntese de ácidos graxos (Browse & Somerville, 1991). Para a síntese destes ácidos, o grupo acetado do acetil-CoA é condensado com malonil-CoA, resultando na liberação do CO<sub>2</sub> e na formação de uma ligação carbono-carbono, gerando acetoacil-ACP. Esta reação é catalisada pela 3-cetoacil-ACP sintase III (KASIII). A cadeia continua crescendo com um novo ciclo de alongamento, onde o produto do último ciclo é condensado com malonil-ACP, reduzido, desidratado e novamente reduzido gerando um novo produto. Este produto será condensado com malonil-ACP iniciando um novo ciclo. Assim, a cadeia em crescimento vai aumentando em duas unidades de carbonos, a cada ciclo, até formarem palmitoil-ACP e/ou esteroil-ACP (Yadav, 1996). Estas outras reações de condensação são catalizadas pelas isoformas KAS I/ KASB, e as isoformas KASII e KASA, promovem a condensação do palmitoil-ACP (16 carbonos) a esteroil-ACP (18 carbonos) (Figura 2).

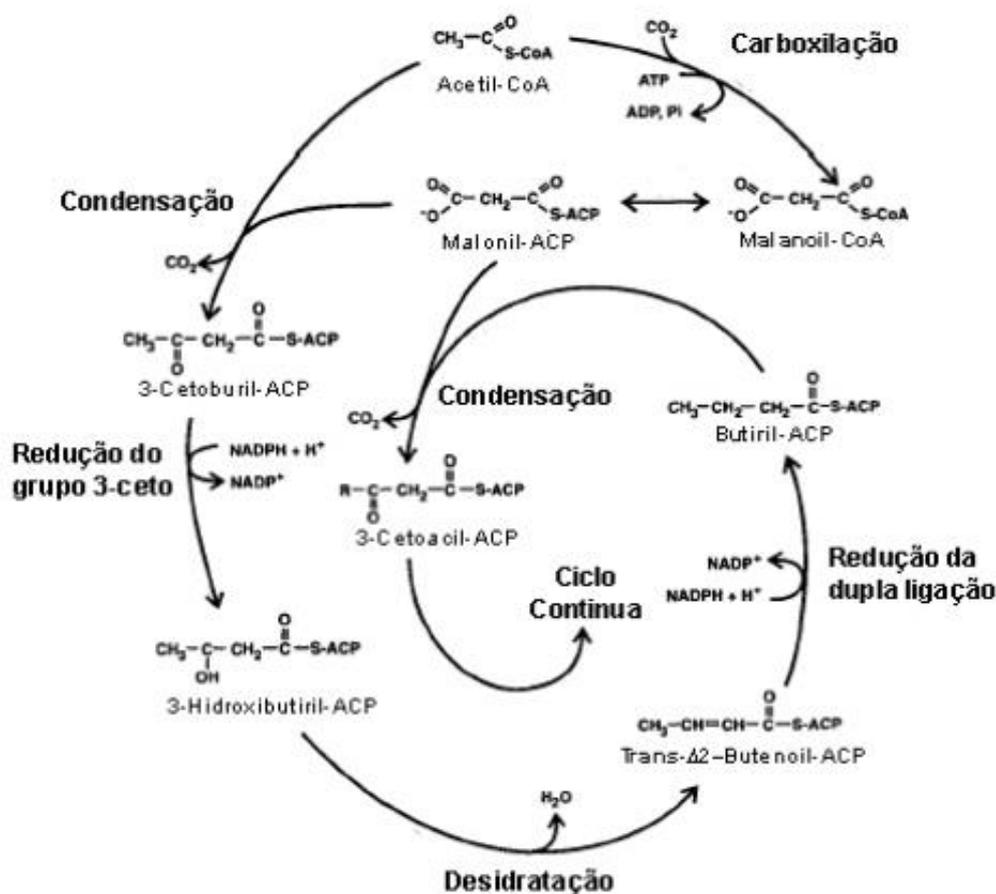


Figura 2: Representação esquemática das reações envolvidas na biossíntese dos ácidos graxos.

Estes ácidos graxos podem então ser incorporados em duas vias para a produção de ácidos graxos insaturados, a via plastidial e ou a via do retículo endoplasmático. Ambas ocorrem nos tecidos foliares com troca significativa entre seus intermediários, mas a via do retículo endoplasmático predomina durante a biossíntese de óleo no desenvolvimento de sementes e em outros tecidos não verdes (YADAV, 1996).

Em ambas as vias, a dessaturação dos ácidos graxos é obtida por meio de uma série de enzimas chamadas de ácido graxos dessaturases (*fatty acid desaturase* - FAD), que são responsáveis por introduzir uma dupla ligação na cadeia de hidrocarboneto, produzindo os ácidos graxos insaturados. Estas enzimas são codificadas por genes nucleares e diferenciam-se na

especificidade de seus substratos e em sua localização subcelular. A maior parte destas enzimas foi inferida partindo de uma biblioteca de mutantes de *Arabidopsis*, defeituosas em insaturação de ácidos graxos (Wallis & Browse, 2002).

Em soja, as enzimas FAD2 e FAD3 estão localizadas no retículo endoplasmático enquanto que as FAB2, FAD5 FAD6 FAD7 e FAD8 estão localizadas nos plastídeos (Wallis & Browse, 2002). FAB2 é uma dessaturase solúvel e catalisa a dessaturação do ácido esteárico em ácido oleico na forma ligada a proteína carreadora de grupos acil (ACP) (Murphy & Piffanelli, 1998). FAD2 e FAD6 são  $\omega$ 6 dessaturases que sintetizam ácido linoleico a partir do oleico no retículo endoplasmático e no plastídeo, respectivamente. As enzimas FAD3, FAD7 e FAD8 são  $\omega$ 3 dessaturases que sintetizam ácido linolênico a partir de linoleico, no retículo endoplasmático (FAD3) e nos plastídeos (FAD7 e FAD8) sendo a FAD8 induzida pelo frio (Berberich, et al., 1998).

#### 4. Família FAD2

A maioria dos ácidos graxos poli-insaturados são sintetizados por dessaturases no retículo endoplasmático. A FAD2 (1-acyl-2oleoyl-sn-glycero-3-phosphocholine  $\Delta$ -12dessaturase ou  $\omega$ -6 dessaturase) é a enzima chave responsável pela produção de ácido linoleico em plantas (Okuley et al., 1994; Szychalla et al., 1997; Cao, et al., 2013).

A FAD2 foi originalmente caracterizada em *Arabidopsis thaliana* e associada com a conversão de ácido oleico em linoleico pela inserção de uma dupla ligação no 12º carbono na cadeia de ácido graxo (Okuley et al., 1994).

Em soja, três genes para a enzima  $\omega$ -6 dessaturase foram identificados, *FAD2-1*, *FAD2-2* e *FAD2-3* (Schlueter et al., 2007). Estudos mostraram que *FAD2-1* é expressa significativamente no desenvolvimento das sementes, enquanto *FAD2-2* é expresso em tecidos vegetativos e no desenvolvimento das sementes, embora seu maior nível de expressão seja em tecidos vegetativos, e provavelmente, codifica a  $\omega$ -6 dessaturase responsável pela produção de

ácidos graxos de membrana (Tang et al., 2005; Schlueter et al., 2007; Heppard et al., 1996). Adicionalmente, o *FAD2-3* é um gene de expressão constitutiva (Li et al., 2007).

Foram encontradas duas isoformas para o gene *FAD2-1* denominadas *FAD2-1A* (Glyma10g42470) localizado no cromossomo 10 e anteriormente descrito como *FAD2-1*, e *FAD2-1B* (Glyma20g24530) localizado no cromossomo 20 (Tang et al., 2005; Schlueter et al., 2007). Ambas as sequências gênicas possuem 93.8% de nucleotídeos idênticos, diferindo em apenas 24 resíduos de aminoácidos na proteína final, que não alteram o modelo topológico predito destas enzimas que são expressas durante o desenvolvimento das sementes (Tang et al., 2005).

A dessaturase *FAD2-2* possui três isoformas, os genes *FAD2-2A* (Glyma19g32930), *FAD2-2B* (Glyma19g32940), e *FAD2-2C* (Glyma03g30070). Os genes para *FAD2-2A* e *FAD2-2B*, enquanto a enzima codificada pelo gene *FAD2-3* se apresenta como uma típica dessaturase de membrana, responsável pela conversão de ácido oleico em ácido linolênico.

As modificações genéticas com o objetivo de regular a expressão da enzima *FAD2* foram realizadas para produzir um óleo com maior teor de oleico em soja (Oliva et al., 2006; Alt et al., 2005). Por meio da supressão *antisense* de *FAD2* em soja, uma nova linhagem foi obtida, produzindo sementes com um conteúdo de óleo que possui uma concentração de oleico duas vezes superior em comparação com a variedade selvagem (Buhr et al., 2002). Similarmente, a supressão de *FAD2-1* e *FatB*, utilizando a metodologia de RNA de interferência, foi capaz de produzir soja com uma concentração significativamente alta de oleico (94%) e uma redução de ácido palmítico (<3%) (Murad et al., 2014), enquanto sementes de soja cultivar Plenish, com alto teor de oleico estão comercialmente disponíveis pela DuPont, desde 2012, sendo desenvolvidas também por silenciamento gênico (Wilson, 2012). Estes resultados demonstram que os membros da família *GmFAD2* desempenham um papel muito importante como alvos para a engenharia metabólica de óleo em sementes de plantas de soja.

## 5. O Genoma da planta de Soja

Em 2010 Schmutz e colaboradores publicaram o sequenciamento shotgun em larga escala de *Glycine max* variedade Williams 82. Aproximadamente 950 megabases foram capturados em 20 cromossomos, e uma pequena quantidade está presente em *scaffolds* não mapeados que consiste em DNA repetitivo. Assim como outras leguminosas, o genoma da soja foi submetido a uma duplicação total (WGD) 59 milhões de anos atrás, seguido pela duplicação de uma glicina específica há aproximadamente 13 milhões de anos atrás. Como resultado aproximadamente 75% dos genes da soja estão presente em mais de uma cópia no seu genoma (Lin et al., 2010; Roulin et al., 2012). Estas duas duplicações foram seguidas de rearranjos do cromossomo, diversificação e perda de genes, o que evidência a complexidade do genoma da soja (Schmutz et al., 2010).

Em 2014, uma nova montagem do genoma da soja tornou-se disponível a partir do Phytozome (<http://phytozome.jgi.doe.gov/>). Esta nova versão substituiu a primeira montagem e foi construída utilizando o montador ARACHNE. A sintonia com *Phaseolus spp* e os mapas genéticos disponíveis para soja foram utilizados para identificar falsos grupos (joins) dentro do genoma anterior. A nova versão do genoma de *Glycine max* compreende 955,4 Mb, reunidos em 20 cromossomos e 1170 *scaffolds* não mapeados. A comparação com o conjunto de etiquetas de sequências expressas (ETS) de soja (soybean EST), sugere que cerca de 98% dos genes codificadores de proteínas conhecidos estão representados nesta montagem.

A disponibilidade do genoma facilita os estudos de genômica funcional comparativa, que ajuda responder questionamentos sobre a biologia da planta e promove um meio eficaz da descoberta de genes envolvidos no metabolismo da planta, que futuramente poderão ser alvos de estudo. Além disso, o sequenciamento da soja nos permite realizar melhorias nas plantas necessárias para o aumento da produção, demandada pelo consumo humano e consumo animal, para a produção de energia limpa e para um balanço na

agricultura em todo o mundo. Com a disponibilidade do genoma ficou acessível a aplicação de várias técnicas visando a melhoria de característica alvo.

## **6. Estratégias gênomicas atuais para a aplicação na melhoria de características alvo**

Tendo em vista que as modificações naturais ou introduzidas alteram o fenótipo do indivíduo, cientistas vem desenvolvendo novos cultivares, por meio de plantas geneticamente modificadas, a fim de melhorar a qualidade e o desempenho das lavouras, visando posterior aplicação agrícola e industrial (Petolino, 2015).

Historicamente, mutações naturais têm contribuído para o desenvolvimento de variações genéticas em programas de melhoramento de plantas, com o objetivo de aumentar a qualidade e produção da planta. Por muito tempo a utilização de mutações induzidas visava buscar fenótipos específicos em plantas, cujas mutações eram introduzidas aleatoriamente no genoma. Entre as técnicas utilizadas, a radiação frequentemente era utilizada, e mutações físicas e químicas eram obtidas utilizando *methanesulfonate* (EMS) e/ou raios de radiação gama e feixes de íons, induziam os sistemas de reparo do DNA. Estes sistemas já estão bem estabelecidos como meio de geração de novos fenótipos por meio da introdução aleatória de mutações no genoma (Osakabe et al., 2015). Para obter plantas superiores utilizando estas estratégias é necessário uma grande triagem nas populações a fim de identificar indivíduos que possuam os fenótipos desejados, sendo um dos grandes problemas dessas metodologias.

Nos últimos anos, genomas inteiros tem sido elucidados e anotados (Michael & Jackson, 2013), e bancos de dados com informações sobre uma infinidade de genes e sua expressão tem sido montados (Wingender et al., 2000). Além disso, tendo em vista a elucidação da dinâmica estrutural e função do genoma, a biotecnologia moderna tem disponibilizado ferramentas que permitem o controle de alterações específicas em sequências do DNA, dentro dos genomas de plantas. Com a aplicação da engenharia genética foi possível

modificar cultivares de plantas para melhorar práticas agrícolas e acrescentar valor nutricional aos produtos. Exemplificando, o arroz dourado que foi projetado para produzir beta-caroteno e aumentar o valor nutricional do grão (Gartland et al., 2013), a soja resistente a herbicida (García-Villalba et al., 2008), o trigo modificado para melhorar a qualidade do processamento (Baker et al., 2006), e a batata com metabolismo de carboidrato alterado (Shepherd et al., 2006).

### 6.1. RNA de Interferência

O RNA de interferência (RNAi) é um mecanismo natural presente em vários organismos incluindo protozoários, invertebrados, vertebrados, fungo, algas e plantas, podendo receber nomes diferentes como silenciamento gênico pós-transcricional (PTGS) (em plantas e *quelling* em fungos). Em plantas, o RNAi foi primeiramente descrito em 1990, quando um grupo de pesquisa resolveu introduzir várias cópias do gene da chalcona sintase (*chs*) responsável pela cor púrpura em flores de petúnia, a fim de obter flores com cores mais intensa. No entanto, a superexpressão do gene de alguma forma bloqueou a síntese do pigmento, gerando 42% de plantas com flores de cor branca e/ou com partes brancas. Também observaram que o nível de RNA mensageiro (mRNA), produzido pelo gene introduzido, era 50 vezes menor nas plantas geneticamente modificadas, quando comparadas com as plantas selvagens (Napoli et al., 1990).

Anos mais tarde esse fenômeno foi esclarecido com o trabalho de Fire e colaboradores (1998), com *Caenorhabditis elegans*, onde verificaram que o silenciamento era mais eficiente quando se utilizava um RNA de fita dupla (dsRNA), ao invés de utilizar cada fita separadamente. Isso sugere que o dsRNA poderia ser responsável por desencadear um silenciamento do gene contendo a sequência homóloga ao dsRNA e nomearam este fenômeno como RNA de interferência. Em 2006, Andrew Z Fire e Crag C Mello receberam o prêmio Nobel em fisiologia por esta descoberta.

Os mecanismos de silenciamento de RNA controlam a expressão dos genes a nível transcricional e pós-transcricional, podendo operar nos

compartimentos nuclear ou citoplasmático, e estão envolvidos em vários processos de regulação de genes no genoma. Esta regulação pode ocorrer à níveis de função do genoma, incluindo a estrutura da cromatina, a segregação dos cromossomos, a transcrição e/ou processamento de RNA, a estabilidade do RNA e a tradução, sendo que os efeitos dos pequenos RNA na expressão e controle de genes são geralmente de inibição da transcrição (Carthew & Sontheimer, 2009).

Em plantas, duas classes de dsRNAs envolvidos no mecanismo de RNAi podem ser destacados, devido a sua origem e função: os microRNA (miRNA) e os *short interfering RNAs* (siRNA). Os miRNAs são todos os pequenos RNAs originados a partir de um transcrito endógeno cuja principal função é atuar como silenciadores pós-transcricionais, inibindo a tradução do RNAm em proteína. Os miRNA são gerados a partir de um precursor de fita única, que forma uma estrutura de grampo auto complementar. Os siRNAs são caracterizados por serem moléculas perfeitamente complementares, introduzidos diretamente no citoplasma das células ou absorvidos do ambiente (Mello & Conte, 2004). Por muito tempo foram caracterizado por serem originados apenas por fontes exógenas. Porém alguns animais podem empregar enzimas RNA dependente de RNA polimerase (RdRP) para gerar siRNA e seus precursores, e em muitos casos, os siRNA dependentes de RdRP correspondem a *loci* endógeno (Golden et al., 2008).

Em ambas as classes o dsRNA é processado em siRNAs pela endonuclease do tipo RNase III, específica para dsRNA, denominada *Dicer*, que apresenta dois domínios: um com atividade catalítica (RNaseIII) e outro domínio de ligação ao dsRNA (dsRBD), gerando moléculas menores com cerca de 21-24 pares de base (Meister & Tusch, 2004). Essas moléculas são então associadas ao complexo denominado *RNA-Induced Silencing Complex* (RISC). Todo complexo RISC possui um membro da família de proteínas chamada *ArgonAUTA* (Hammond et al., 2001). Neste complexo uma das fitas do dsRNA é eliminado e a fita remanescente “guia” o reconhecimento para a degradação ou inibição do mRNA-alvo, por complementaridade de bases (Hutvagner & Zamore, 2002).

Apesar de ser uma metodologia muito utilizada em plantas e animais, o silenciamento gênico por meio de RNAi ainda precisa ser muito estudado, visto que alguns experimentos resultaram em um efeito inibitório não específico, proveniente do tamanho das sequências escolhidas como alvo. Além disso, transgenes expressando construções de RNAi são sujeitos a variação na expressão do transgene, o que requer o rastreamento de um grande número de eventos transgênicos para identificar possíveis candidatos que estávelmente expressam a característica ao longo de gerações.

## 6.2. Metodologias de Edição de Genoma:

A edição de genoma com nucleases modificadas (GEEN) tem sido desenvolvida como um método de engenharia genética eficaz, que utiliza nucleases modificadas artificialmente quebrar o DNA no local desejado no genoma. Neste contexto, uma quebra na dupla fita de DNA (DSB) induzida pela nuclease em um local específico, é reparada por meio de recombinação homóloga (HR) ou *non-homologous end joining* (NHEJ). Em HR, uma sequência altamente similar ao genoma é utilizada como fita-molde para reparo e, dependendo da similaridade da sequência utilizada, as mudanças no sítio reparado podem ser introduzidas. No sistema de SOS celular, durante a NHEJ, a extremidade final do DNA quebrado se religa, geralmente com imprecisão, introduzindo assim mudanças na sequência de nucleotídeos no local da quebra. Essas duas vias conservadas podem ser utilizadas para mudar a sequência de uma região-alvo, inserindo ou retirando bases às sequências no local da clivagem, resultando na interrupção da sequência e/ou integração de sequências exógenas por meio de recombinação homóloga. Atualmente, três tipos de engenharia de nucleases tem sido utilizadas para a edição de genoma em várias espécies de plantas: dedos de zinco (*Zinc-Finger Nucleases* - ZFN), *Transcription Activator-like Effector Nuclease* (TALEN), e o sistema CRISPR/Cas9 (*Clustered Regularly Interspaced Palindromic Repeats*-CRISPR).

A capacidade de modificar a informação genética de uma maneira precisa e específica e, recuperar plantas modificadas não só permite estudos de função do gene e de mecanismos biológicos, mas também, potencialmente,

o desenvolvimento de novos fenótipos. Como a demanda por produção agrícola cresce com a expansão da população humana, a necessidade de modificar mais características complexas em culturas, como aumentar a produtividade e a tolerância a estresse, exigirá abordagens mais sofisticadas. A edição de genoma representa uma das ferramentas que irá permitir melhorias específicas em plantações.

### 6.2.1. Zinc Fingers Nucleases

Os *Zinc finger* (ZF) são pequenos motivos estruturais de proteínas, onde um ou mais íons de zinco são responsáveis por estabilizar sua estrutura. Os ZF foram primeiramente reconhecidos como repetidos motivos com ligação com zinco (*repeated zinc-binding motif*), contendo uma cisteína e uma histidina ligantes conservadas, descrita no fator de transcrição de *Xenopus laevis* IIIA (TFIIIA). Desde sua descoberta, inúmeros outros motivos de ligação com o zinco (*zinc binding motifs*) foram identificados e designados como ZF. Estes variam amplamente em estrutura assim e função, podendo exercer funções desde a ligação com o DNA ou RNA, à interação proteína-proteína e associação com membrana.

Proteínas contendo ZF podem ser classificadas em diferentes grupos, de acordo com sua estrutura, sendo mais comuns os grupos *treble clef*, *zinc ribbon* e *Cys2-His2-like* (Krishna et al., 2003). O grupo *Cys2-His2-like* é o melhor caracterizado até o momento e extremamente comum entre os organismos. Assim que a especificidade dos motivos *Cys2His2* foi entendida, estes foram comumente utilizados para reconhecimento de ácidos nucleicos.

As *Zinc Finger Nucleases* (ZFN) são proteínas híbridas, compostas de um domínio de clivagem que consiste na enzima de restrição do Tipo IIS, FokI, que é responsável por catalisar a clivagem do DNA, e um domínio de ligação ao DNA composto de dedos de zinco (*zinc fingers- ZF*) do tipo *Cys2His2* em *tandem*. Elas foram construídas inicialmente como endonucleases com especificidade alterada, sendo logo em seguida desenvolvidas como reagentes

de clivagem *in vivo* capazes de estimular as mutações e/ou as recombinações homólogas localizadas.

O domínio Cys<sub>2</sub>His<sub>2</sub> é composto de aproximadamente 30 aminoácidos, contendo duas folhas  $\beta$  antiparalelas e uma  $\alpha$ -hélice, e é caracterizado pelo íon de zinco coordenado por dois resíduos de cisteína nas folhas  $\beta$  e dois resíduos de histidina na  $\alpha$ -hélice, por isso o nome Cys<sub>2</sub>His<sub>2</sub> (Figura 3A). Em proteínas naturais, o domínio C<sub>2</sub>H<sub>2</sub> está principalmente envolvido na ligação ao DNA. Assim como muitas proteínas de ligação ao DNA, ZF se liga ao DNA por meio da inserção de uma  $\alpha$ -hélice no sulco maior (*major groove*) da dupla hélice. Cada ZF reconhece 3 pares de bases consecutivos do DNA por meio de interações de um aminoácido da cadeia lateral com um par de base, por isso mais de um ZF é necessário para dar especificidade na ligação com a sequência alvo (Pavletich & Pablo, 1991) (Figura 3B). O número de ZF em cada matriz de ZFN pode variar bastante. O número mínimo para se atingir especificidade é 3, e as combinações com até 6 já foram realizadas e testadas em algum contexto.

O domínio de clivagem não específico do Tipo IIS, FokI é normalmente utilizado como domínio de clivagem em ZFN (Kim et al., 1996). Uma característica crítica deste domínio é que é necessário que ele dimerize para que este domínio seja capaz de clivar o DNA. Assim, um par de ZFNs é essencial para agir no DNA (Urnov et al., 2010). Com a finalidade de permitir que os dois domínios de clivagem formem o dímero e clive o DNA, os dois ZFN devem se ligar em fitas opostas com uma distância de aproximadamente 6 nucleotídeos (Bibikova et al., 2001)(Figura 3C).

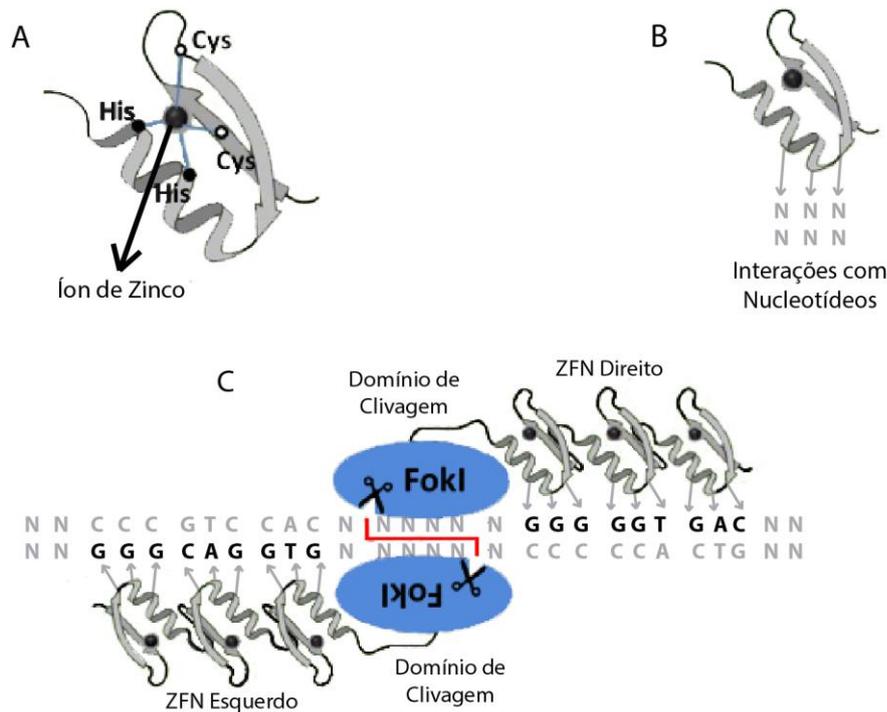


Figura 3: **Estrutura dos Zinc Finger Nucleases.** A, domínio C2H2, onde o íon de zinco coordena os resíduos de Cisteína e Histidina. B, cada ZF interage com 3 nucleotídeos do DNA. C, representação esquemática do ZFN interagindo com o DNA, e a dimerização da FokI para a clivagem do DNA (Adaptado de Fichtner et al., 2014).

Em plantas, vários trabalhos já foram reportados utilizando ZFN para edição de genomas em *Arabidopsis*, soja, tabaco, canola e milho. O primeiro trabalho foi realizado em *Arabidopsis* onde foi demonstrada a expressão de um ZFN sob controle de um promotor de uma proteína de choque térmico, o qual seria responsável pela perda de um sítio de restrição da enzima *EcoRI* por meio de uma correta mutação. Foi confirmado não só a interrupção da sequência alvo, mas também a herança de alguns genes mutados para a geração T1 (Lloyd et al., 2005). Trabalhos similares foram publicados utilizando ZFN em *Arabidopsis* para obter modificações em sequências de um gene reporter pré-integrado (de Pater et al., 2009; Tovkach et al., 2009).

Em soja dois genes DICER-LIKE (*DCL4a* e *DCL4b*) foram alvos de ZFN, gerando plantas com mutações nos genes alvos, onde plantas com mutações no gene *DCL4a* pareciam normais produzindo progênies com mutações segregando 1:2:1 como esperado, enquanto as mutações no gene *DCL4b* exibiram anormalidades fenotípicas, incluindo sementes abortadas, evidenciando que a expressão de genes codificando ZFN podem gerar mutações hereditárias em genes endógenos (Curtin et al., 2011). Também utilizando soja como modelo, Curtin e colaboradores (2013) desenvolveram um método para utilização de ZFN, com a finalidade de criar mutações alvo no genoma paleopoliploide da soja.

A tecnologia ZFN, embora muito elegante e promissora, enfrenta alguns obstáculos técnicos e biológicos. Os níveis de atividade fora do alvo, a imperfeita modularidade de alguns domínios, certas combinações de domínios ZF sendo menos específicos ou menos eficiente em se ligar ao DNA, dificuldade de construir as proteínas ZF, entre outros, são problemas enfrentados frequentemente por pesquisadores que utilizam esta técnica. Além disto, estudos recentes de citotoxicidade mostraram a letalidade potencial em linhagens de células superexpressando ZFNs (Porteus & Baltimore, 2003, Kim et al., 2009), embora melhoramentos na arquitetura e desenhos baseados em bioinformática tem amenizado a atividade fora do alvo (Cradick et al., 2011; Fine et al., 2014).

A montagem sequência específica de matrizes ZF é trabalhosa e demorada (Kim et al., 2009; Gupta et al., 2012) e também associada com altas taxas de falha (Ramirez et al., 2008). Isto tem levado os cientistas a procurar outras tecnologias, que permitam o desenho, a montagem e a utilização de forma prática e comum, na maioria dos laboratórios de biologia molecular.

### 6.2.2. Sistema CRISPR/Cas9

A tecnologia de edição de genoma denominada CRISPR/Cas 9 provém de um sistema de defesa em bactérias e arqueas, originalmente encontrada no genoma de *E. coli* (Ishino et al., 1987), sendo utilizado para protegê-las de

ácidos nucleicos exógenos, como de vírus e plasmídeos. Estima-se que cerca de 40% das bactérias e 90% das arqueas sequenciadas possuam este sistema de defesa (Mojica et al., 2000). Apesar de descrita em 1987, CRISPR começou a realmente a ser investigado após o ano de 2000, e utilizado como tecnologia de edição de genomas, a partir de 2012 (Ishino et al., 1987; Jinek et al., 2012; Gasiunas et al., 2012).

Em bactérias e arqueas, o *loci* CRISPR, que significa repetições curtas palindrômicas agrupadas e regularmente intercaladas (*clustered regularly interspaced palindromic repeats*), é constituído por várias sequências palindrômicas conservadas repetidas e intercaladas por trechos de sequências variáveis chamados espaçadores (spacers) que basicamente correspondem a sequências derivadas de vírus e/ou plasmídeo (região CRISPR); uma sequência líder e um conjunto de genes conservados codificadores de proteínas associados a CRISPR, denominado genes *Cas* (*CRISPR-associated*).

O sistema CRISPR faz uso de RNA antisense como memória de invasões anteriores e é composto basicamente por três fases: aquisição, expressão e interferência. O primeiro estágio é a aquisição, que ocorre logo após a entrada de um DNA invasor, onde novas sequências de fagos e plasmídeos são obtidas e integradas ao *locus* CRISPR. Na fase de expressão, o *locus* CRISPR é transcrito, produzindo um precursor CRISPR RNA (pre-crRNA), que será processado liberando pequenos fragmentos CRISPRs RNAs (crRNAs). Os crRNA maduros levarão a clivagem do DNA exógeno, em uma próxima tentativa de invasão, por meio da formação de complexos com proteínas *Cas* durante a fase de interferência (Figura 4) (Bhaya et al., 2011).

Baseado em classificações que integram filogenia, conservação de gene, organização e conteúdo do *loci* CRISPR, o sistema CRISPR-Cas é classificado em três diferentes tipos, Tipo I, Tipo II e Tipo III (Bhaya et al., 2011; Wiedenheft et al., 2012; Makarova et al., 2015). Cada tipo possui uma proteína *Cas* “de assinatura”, sendo que o Tipo I possui a Cas3 helicase-nuclease, o sistema Tipo II possui Cas9 nuclease, e o Tipo III possui a Cas10.

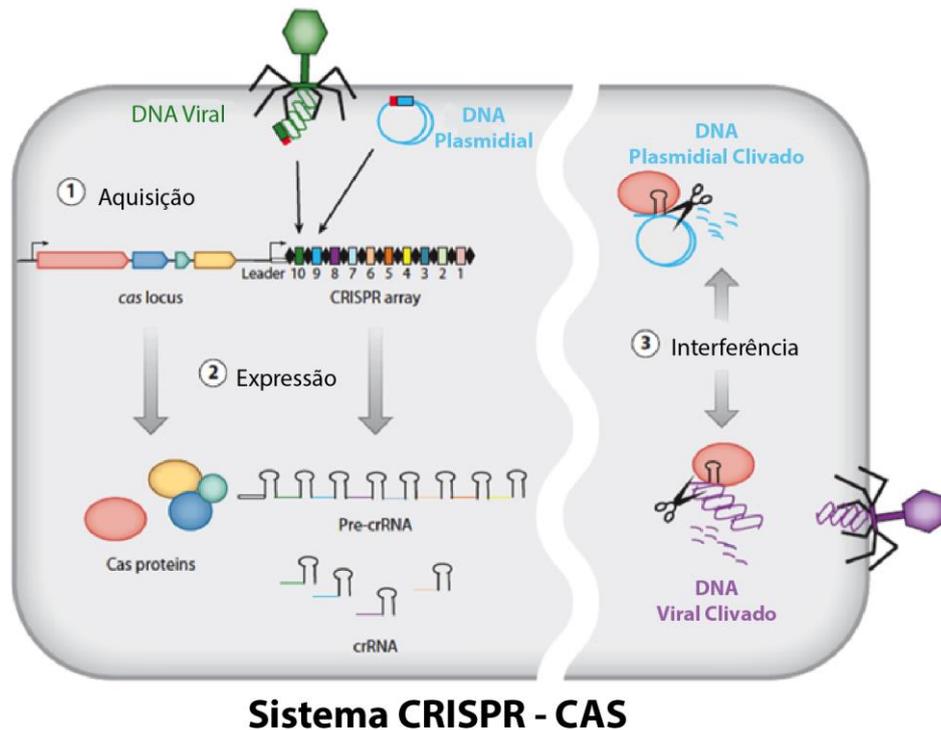


Figura 4: **Representação do Sistema CRISPR-Cas.** Fases do Sistema CRISPR-Cas: Aquisição, Expressão e Interferência. 1- Fase da Aquisição onde o DNA invasor é integrado ao *locus* CRISPR. 2 -Fase de Expressão, transcrição do *loci* CRISPR, e processamento do Pre-crRNA para liberação dos crRNA. 3- Fase de Interferência, clivagem do DNA invasor. Adaptado de Bhaya et al., 2011.

O sistema utilizado para ao desenvolvimento da ferramenta de edição de genoma é o sistema Tipo II, que é o mais estudado e mais simples entre os três sistemas, com apenas 4 genes *Cas*, sendo um deles sempre *Cas9* (Ran et al., 2013). No sistema de defesa do Tipo II, durante a fase de aquisição as sequências precursoras do espaçador dependem do reconhecimento da sequência PAM para serem adquiridas (Mojica et al., 2009). A sequência PAM também é responsável por direcionar a orientação dos espaçadores no *locus* CRISPR, que normalmente são introduzidos na região 5' do *locus*, logo após a sequência líder (Deveau et al., 2008).

Para o processamento do crRNA no sistema do Tipo II, uma outra região do *loci* CRISPR é transcrita gerando os *transactivating crRNA* (tracrRNA), que

são necessários para a maturação dos crRNAs (Deltcheva et al., 2011). Os tracrRNAs contêm uma região que é complementar a região repetida no pre-crRNA, que permite tracrRNA se ligar a sequência complementar no pre-crRNA, formando uma estrutura duplex que é reconhecida e clivada pela ribonuclease RNase III, gerando o crRNA maduro (Van der Oost et al., 2015). Estudos sugerem que a Cas 9 é necessária para o processamento do crRNA, por provavelmente ligar e posicionar as moléculas de RNA para serem clivadas pela RNase III (Fonfara et al., 2014). Após o processamento pela RNase III tracrRNA-crRNA se mantém firmemente associados com a Cas9 formando o complexo Cas9 (Jinek et al., 2012; Gasiunas et al., 2012).

O sistema CRISPR/Cas9 para edição de genoma provém de uma versão melhorada do sistema Tipo II de *Streptococcus pyrogenes* (Jinek et al., 2012), onde uma molécula única RNA guia (sgRNA- single guide RNA) é construída a partir da fusão de crRNA contendo uma sequência guia e o tracrRNA (Figura 5). O sgRNA é capaz de direcionar a atividade da Cas9 para qualquer sítio de interesse adjacente a sequência PAM. Ele possui uma sequência de 20 nucleotídeos complementares ao sítio alvo na porção crRNA do sgRNA, que confere ao sgRNA a habilidade de complementaridade. Uma vez clivado pela Cas9, o alvo pode ser reparado tanto por NHEJ como HR.

Esse sistema se mostrou eficaz como uma ferramenta de edição de genoma pela primeira vez em 2013 com células humanas (Cong et al., 2013; Mali et al., 2013) . Desde então tem sido utilizado em diversos organismos incluindo plantas. Diversos trabalhos já foram publicados utilizando edição de genoma em uma grande variedade de plantas incluindo *Arabidopsis*, tabaco, milho, sorgum, trigo, soja, batata, tomate, laranja e arroz.

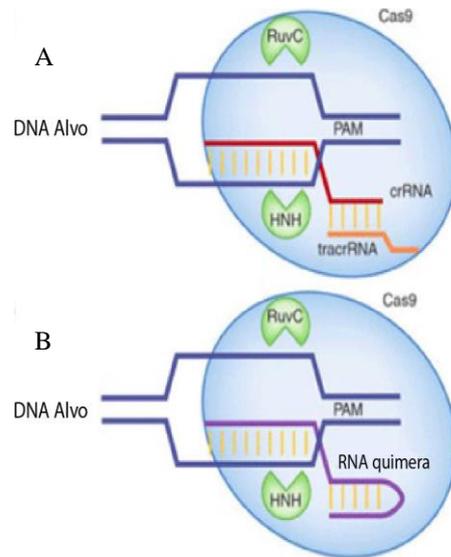


Figura 5: **Sistema CRISPR-Cas9.** A, Sistema Tipo II CRISPR-Cas9 original. B, CRISPR-Cas9 quimera. Adaptado de Barrangou 2012.

Os primeiros trabalhos publicados de aplicações de CRISPR-Cas9 em plantas demonstraram uma imensa versatilidade da tecnologia utilizando espécies modelos como *Arabidopsis thaliana* e *Nicotiana benthamiana*, assim como plantações como de arroz, explorando uma vasta plataforma de transformação (*transfection protoplast*, agroinfiltração e geração de plantas transgênicas estáveis) sendo utilizado tanto genes endógenos como transgenes e explorando tanto NHEJ como HR para gerar pequenas deleções, inserções em genes alvos, e modificações multiplex de genoma (Feng et al., 2013; Li et al., 2013; Nekrasov et al., 2013; Shan et al.; 2013; Xie & Yang, 2013).

Trabalhos subsequentes com sorgo, trigo e milho forneceram os primeiros dados comparativos relativos a aspectos como eficiência de mutação, especificidade de clivagem, resolução da estrutura local (*resolution of locus structure*) e o potencial de criar grandes deleções cromossômicas e também demonstraram que sgRNAs podem ser expressos sob o controle de diferentes promotores, incluindo os reconhecidos por RNA polimerase II e III (Fauser et al., 2014; Feng et al., 2014; Gao et al., 2014; Jiang et al., 2013b; Mao et al., 2013; Miao et al., 2013; Sugano et al., 2014; Upadhyay et al., 2013; Zhang et al., 2014; Zhou et al., 2014; Wang et al., 2014; Liang et al., 2014). Estes

estudos também confirmaram que uma quimera sgRNA é mais eficiente do que os componentes crRNA e tracrRNA separados, assim como ocorre em outros eucariotos (Miao et al., 2013; Zhou et al., 2014).

Devido aos genes de sgRNA serem relativamente pequenos e poderem facilmente serem colocados em uma única construção plasmidial múltiplos alvos também foram testados em *Arabidopsis* (Ma et al., 2015), arroz (Ma et al., 2015; Xi et al., 2015; Wang et al., 2015) e raízes de soja (Cai et al., 2015). Além de novos métodos de construções contendo múltiplos sgRNAs, algumas abordagens na montagem dos genes Cas9 tem sido publicadas como a comparação de eficiência de diferente tipos de nucleases CRISPR/Cas9 produzidas em vários códons otimizados e 12 diferentes *de novo* genes sgRNA sintetizados (Johnson et al., 2015). Outro exemplo é o estudo de Cas9/sgRNA providas de diferentes fontes, como por exemplo de *Staphylococcus aureus* (Ran et al., 2015).

Apesar de ser uma ferramenta recente, robusta e de fácil acesso, possui alta frequência de mutações não alvo como demonstrado em estudos anteriores (Cong et al., 2013, Fu et al., 2013, Hsu et al., 2013, Jiang et al., 2013a, Mali et al., 2013; Pattanayak et al., 2013). Embora a sequência de 20 nucleotídeos do sgRNA fosse inicialmente considerada necessária para determinar a especificidade, foi posteriormente demonstrado que apenas 8-12 nucleotídeos na extremidade 3' (sequência *seed*) são necessários para o reconhecimento do sítio alvo e clivagem (Cong et al., 2013, Jiang et al., 2013a; Jinek et al., 2012), enquanto múltiplos *mismatches* na região distal de PAM podem ser tolerados, dependendo do número total e arranjo (Fu et al., 2013, Hsu et al., 2013 and Pattanayak et al., 2013). A baixa especificidade do complexo CRISPR/Cas9 nas posições não *seed* no crRNA parece ser uma propriedade intrínseca que reduz a probabilidade de evasão do sistema imune por vírus com mutações pontuais (Semenova et al., 2011). Além disso, outro fator que negativo desta tecnologia é a limitação que a sequência PAM impõe ao alvo, visto que esta região é necessária para o reconhecimento e clivagem do alvo.

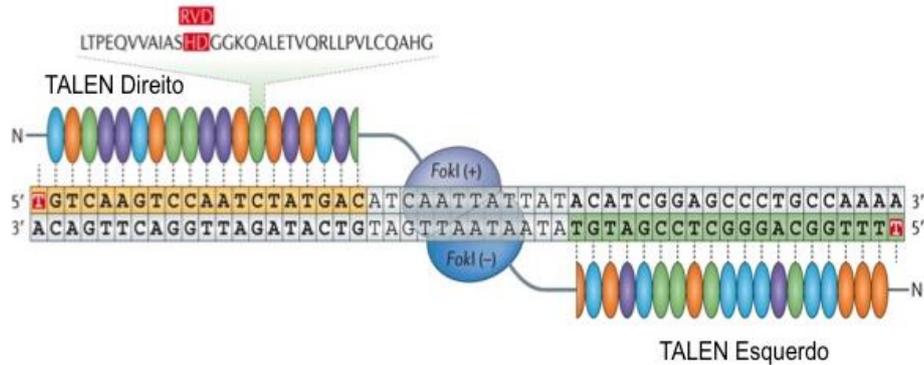
### 6.2.3. TALEN

Efetores de ativação de transcrição *like* (TALEs) são uma classe de proteínas que se ligam ao DNA que podem ser encontradas nas espécies da bactéria patogênica de plantas *Xanthomonas*. Assim como ZFN, TALEN são proteínas customizadas compostas de um domínio de ligação ao DNA, neste caso pelo TALEs, e um domínio de clivagem, constituído pela FokI.

TALE consiste em um fator de virulência chave secretado pelo sistema de secreção do Tipo III de bactéria do gênero *Xanthomonas*, sendo primeiramente isolado de *X. Campestris* (Bonas et al., 1989). Estas proteínas são translocadas para dentro do citoplasma da célula vegetal pelo sistema de secreção do Tipo III, e possuem a habilidade de reprogramarem as células hospedeiras por mimetizar fatores de transcrição eucarióticos, se ligando a promotores na planta hospedeira e ativando a expressão de genes que facilitam a infecção bacteriana (Schornack et al., 2013; Boch et al., 2009; Cermak et al., 2011).

Na natureza, TALE é composto por 12-27 unidades repetidas quase idênticas e ordenadas em *tandem*. Essas unidades são formadas por uma série de 33-35 de aminoácidos, e cada unidade reconhece um nucleotídeo (Figura 6A)(Kay et al., 2007; Boch & Bonas 2010). O número destas repetições pode variar, mas a singularidade de cada TALE, vem de um polimorfismo nos resíduos 12 e 13, referidos como di-resíduos variáveis repetidos (repeat-variable di-residues-RVD), que determinam a especificidade do TALE. Embora diferentes RVDs possam ocorrer entre TALEs, quatro deles, HD, NG, NI e NN, responsáveis por 75% do total dos RVDs e são respectivamente associadas com uma das quatro bases do DNA C, T, A e G/A. O simples código que governa a atividade de TALE fez possível customizar estas proteínas para conseguir modificações em sequencias de DNA de interesse.

A



B

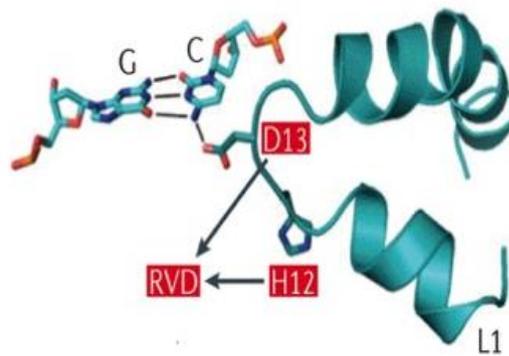


Figura 6: **Estrutura do TALEN.** A representação esquemática de um par de TALEN, onde cada TALE é composto de 33-35 aminoácidos, e os aminoácidos nas posições 12 e 13, referidos como RVDs (em vermelho), são responsáveis pelo reconhecimento da base nucleica. B, Estrutura em forma de V de cada TALE, onde o segundo resíduo RVD faz contato com o DNA enquanto o primeiro estabiliza a estrutura. Adaptado de Kim & Kim, 2014.

A estrutura cristalizada de um par de TALEN ligado ao DNA indica que cada repetição compreende duas alfas hélices ligadas para produzirem uma estrutura em forma de V, onde os RVDs ocorrem no centro do looping conectando as hélices, no qual o segundo resíduo do RVD faz contato com o DNA, enquanto o primeiro estabiliza o domínio da estrutura (Figura 6B).

Os TALENs, assim como os ZFN, possuem um domínio de ligação ao DNA, composto pelas proteínas DNA ligantes TALE, e um domínio de clivagem composto pela enzima endonuclease inespecífica Fok I. O resultado é uma ferramenta de modificação genética capaz de induzir quebras em fita duplas de

DNA (DSB) *in vivo*. É importante ressaltar novamente que a FokI cliva como um dímero, por isso dois TALENs opostos são necessários para criar a DSB, sendo que um espaço entre eles deve ser deixado para que os dois domínios da FokI possam agir (Cermak et al., 2011; Miller et al., 2011). Dois processos principais podem estar envolvidos no reparo de DSB: *non-homologous end joining* (NHEJ), que resultará em pequenas inserções ou deleções (indels) e recombinação homóloga (HR), utilizada para modificações de sequências providas por uma fita molde doadora (Cade et al., 2012).

A montagem dos TALENs é um processo no qual todos os RVDs, que conferem especificidade de ligação a um TALEN, são colocados juntos em um vetor previamente construído que inclui, entre outros, o domínio catalítico FokI. Após o trabalho de Cermak e colaboradores (2011) a montagem dos TALENs se tornou mais acessível e rápida, visto o desenvolvimento da metodologia de clonagem *Golden Gate* que é comumente utilizada na montagem dos TALENs. Por meio da clonagem *Golden Gate*, alguns plasmídeos separados podem eficientemente ser clonados em um vetor em uma única reação em um tubo. Utilizando endonucleases de restrição do tipo IIS, *BsaI* e *Esp3I* que clivam fora dos sítios de reconhecimento, e a enzima de ligação, a reação é submetida a múltiplos passos de digestão/ligação, de acordo com a temperatura ótima para cada enzima utilizada. Visto que a correta montagem elimina o sítio de reconhecimento da enzima de restrição, este processo possui alta eficiência e grande parte das colônias de bactéria, após a transformação, possuem a construção desejada (Engler et al., 2009; Cermak et al., 2011).

A rápida e fácil montagem, disponibilidade de recursos, além de uma alta taxa de sucesso tornaram a tecnologia TALEN uma escolha fácil para aplicações de edição de genoma. A tecnologia TALEN tem sido amplamente utilizada em uma gama de organismos incluindo leveduras, plantas, algas, protozoários, nematoides, peixes, insetos, mamíferos e células humanas.

Apesar de o primeiro sucesso do uso de TALEN para edição de genoma ter sido em levedura (Christian et al., 2010; Li et al., 2011a) o desenvolvimento e a facilidade de novas metodologias de montagem de TALENs (Cemark et al.,

2015; Li et al., 2011b; Reyon et al., 2012, 2013; Schmid-Burgk et al., 2015) rapidamente levou a aplicações dessa ferramenta em plantas (Cermak et al., 2011; Christian et al., 2013; Li et al., 2012; Mahfouz et al., 2011; Zhang et al., 2015), incluindo demonstrações de substituição de genes por recombinação homóloga (Zhang et al., 2013).

Um dos primeiros trabalhos utilizando TALEN como uma ferramenta de edição de genoma foi realizado utilizando ensaios com leveduras e um plasmídeo do gene repórter *lacZ*, para comprovar a funcionalidade dos TALEN (Christian et al., 2010). Neste mesmo trabalho o gene *ADH1* de *Arabidopsis* também foi alvo da atividade dos TALENs (Christian et al., 2010). O mesmo gene foi utilizado com alvo em outro estudo em protoplastos de *Arabidopsis*, onde a metodologia *Golden Gate* foi utilizada para a construção dos TALENs (Cermak et al., 2011). *Arabidopsis* foi ainda utilizada para demonstrar que TALENs são reagentes úteis para alcançar mutagêneses dirigidas nesta importante planta modelo (Christian et al., 2013).

Em soja a utilização da tecnologia TALEN incapacitou a atividade de duas enzimas ácido graxo desaturase 2 (FAD2) envolvidas na biossíntese de ácidos graxos em soja, gerando plantas com sementes possuindo alta concentração de ácido oleico (~80%), e baixa concentração do ácido poli-insaturado, o ácido linoleico (4%) (Haun et al., 2014). Recentemente Demorest e colaboradores (2016) utilizaram as plantas de soja com o teor de oleico alterado geradas no trabalho de Haun e colaboradores (2014), e introduziram mutações no gene da enzima ácido graxo desaturase 3 A (FAD3A), com o intuito de diminuir os níveis de ácido linolênico, novamente utilizando TALENs, conseguindo diminuir significativamente os níveis de ácido linolênico inicialmente 4.7% para 2.5%.

Uma clara desvantagem do TALEN, quando comparado a outras tecnologias, é seu significativo tamanho (aproximadamente 3kb por par) que pode limitar sua ampla aplicação, especialmente nos casos quando não se tem uma eficiente entrega do gene TALEN. O grande tamanho do TALEN faz com que seja difícil entregar e expressar um par de TALENs nas células, tornando-

os menos atrativos para aplicações terapêuticas, onde eles devem ser entregues em vetores virais com limite de tamanho ou como moléculas de RNA. O desenvolvimento de estratégias para entrega eficiente dos genes TALEN nas células poderiam permitir uma edição de genoma mediada por TALEN em mais organismos e células (Gupta & Musunuru, 2014). Outra desvantagem desta tecnologia é a falta de dados relacionados a efeitos *não alvo*. Assim como em ZFNs e CRISPRs, os efeitos *não alvo* do TALEN ainda precisam ser melhor estudados.

#### 6.2.4. **Atividade *não alvo* das metodologias de Edição de Genoma:**

Uma importante qualidade em qualquer metodologia de edição de genoma é a especificidade. Uma boa metodologia deve inserir mutações apenas no alvo, deixando o resto do genoma intacto. TALEN e CRISPR são amplamente utilizados em edição de genomas, no entanto nenhuma delas possui especificidade de reconhecimento do DNA perfeita, sendo capazes de se ligar a outras sequências que não ao DNA alvo, gerando DSB em locais não desejadas no genoma (Carroll, 2014). Estes efeitos não alvo podem introduzir mudanças indesejadas em sequências do genoma com consequências imprevisíveis para as células, órgãos, organismos e até mesmo para o ambiente (Stella e Montoya, 2016).

Com o objetivo de economizar tempo e trabalho, faz-se necessário minimizar as chances de que ocorra uma atividade fora do alvo. Várias medidas podem ser consideradas a fim de evitar atividade *não alvo*, quando se deseja utilizar uma das metodologias de edição de genoma, entre elas a escolha da nucleasse, a escolha do alvo e o método de entrega da nucleasse (Hendel et al., 2015).

Ao escolher a nucleasse, deve-se ter em mente as diferenças entre os sistemas, uma vez que ZFN e TALEN são mais específicos mas relativamente mais complexos de se produzir, enquanto CRISPR, são considerados simples, mas possuem um alto índice de atividade não alvo quando comparado com TALEN e ZFN (Hendel et al., 2015) .

Algoritmos com base em dados coletados em estudos *in vitro* e *in vivo* podem ser utilizados para auxiliar a escolha do alvo e prever possíveis atividades fora do alvo, sem a necessidade de realizar experimentos mais complexos. Exemplos de alguns algoritmos incluem PROGNOS, que podem ser utilizados para predição de *não alvos* para ZFNs e TALENs; *CRISPR Design Tool*, que foi especificamente desenvolvido para prever e pontuar não alvos para o sistema CRISPR/Cas; e o CHOPCHOP, um algoritmo adequado para predição de *não alvo* em CRISPR/Cas e TALEN (Lee et al., 2016).

O método como a nucleasse é entregue nas células já demonstrou ter influência na quantidade de atividade fora do alvo. Quando comparado com a transformação de plasmídeos de expressão, a liberação de nucleases manipuladas com RNAm ou proteína purificada reduz as possibilidades de efeitos fora do alvo, uma vez que diminui a concentração da proteína e o tempo de expressão, sendo a escolha do promotor também um fator importante (Liu & Shiu, 2016).

Apesar das dificuldades, as tecnologias RNAi, ZFNs, CRISPRs e TALENs são poderosas ferramentas de manipulação de genomas em plantas e outros organismos. É provável que o refinamento destas ferramentas continue e que elas se adaptem aumentando significativamente novas descobertas na biologia de plantas e animais.

A tecnologia TALEN tem sido amplamente utilizado em plantas de diversas espécies, e se mostrado funcional como mutagênico, porém apenas 2 estudos foram realizados com soja, uma planta com significativa importância econômica. Neste trabalho foram criadas mutagêneses nos genes *FAD2-1A* e *FAD2-1B* por meio da tecnologia de TALEN, utilizando um sítio-alvo conservado entre os dois genes. Com esta estratégia foi possível gerar três linhagens transgênicas de soja com a porcentagem de ácido oleico alterada, mostrando 31%, 39% e 66% de concentração de oleico, contra 20% normalmente produzido pelo selvagem. Além disso, verificou-se possíveis *não alvos* confirmando que apenas o gene de interesse foi alvo do TALEN.

## **HIPÓTESE**

**HIPÓTESE:**

É possível obter plantas de soja da variedade BR16, contendo um perfil de óleo nas sementes alterado por meio da metodologia TALEN, sendo esta específica para alvos, não atingindo outros locais do genoma?

## **OBJETIVO**

**OBJETIVO:**

O objetivo principal foi alterar o perfil de óleo em sementes de soja por meio de mutações nos genes *FAD2-1A* e *FAD2-1B* utilizando a tecnologia TALEN, e identificar potenciais genes não alvos.

**OBJETIVOS ESPECÍFICOS:**

Este trabalho teve como objetivos específicos selecionar uma região alvo para os TALENs que fosse comum aos genes *FAD2-1A* e *FAD2-1B* e produzir plantas de soja BR16 transgênicas por meio de biobalística, que expressassem os vetores TALENs e apresentassem mutações e deleções nos genes alvos do TALEN, com o intuito de gerar plantas de soja com o conteúdo de óleo oleico nas sementes aumentado. Este trabalho também buscou indentificar possíveis ações fora do alvo provinda da metodologia TALEN em genes da família FAD2.

## **METODOLOGIA**

## METODOLOGIA

### 1. Construção dos vetores TALEN

Os desenhos dos TALENs foram construídos utilizando o programa TALE-NT, disponível online em <http://boglabx.plp.iastate.edu/TALENT/> onde um par de TALEN (contendo 28 RVDs) e uma região espaçadora foram complementares a uma região conservada nos genes FAD2-1A e FAD2-1B. Os vetores TALEN foram obtidos por meio da metodologia *Golden Gate* de Cermak e colaboradores (2011), utilizando o *Addgene Golden Gate TALEN kit*, que utiliza reações de digestão e ligação na mesma reação. Os RVDs completos foram inseridos no plasmídeo pTAL\_3 *backbone* gerando os vetores pTAL3\_5 para a borda esquerda e o vetor pTAL3\_6 para a borda direita. Os plasmídeos foram inseridos em *E. coli* DH5 alfa, transformadas pelo protocolo de choque térmico, e o DNA foi amplificado e extraído conforme protocolo de extração de DNA plasmidial. Para a transformação em plantas de soja, os vetores pTAL3\_5 e pTAL3\_6 foram digeridos com as enzimas de restrição *Bgl* II e *Sac* I e inseridos no plasmídeo pBI426, gerando os vetores pBI\_TAL3\_5 e pBI\_TAL3\_6 (Figura 7). O vetor pBI426 possui o promotor constitutivo 35S, o AMV *translational enhancer* e o terminador NOS.

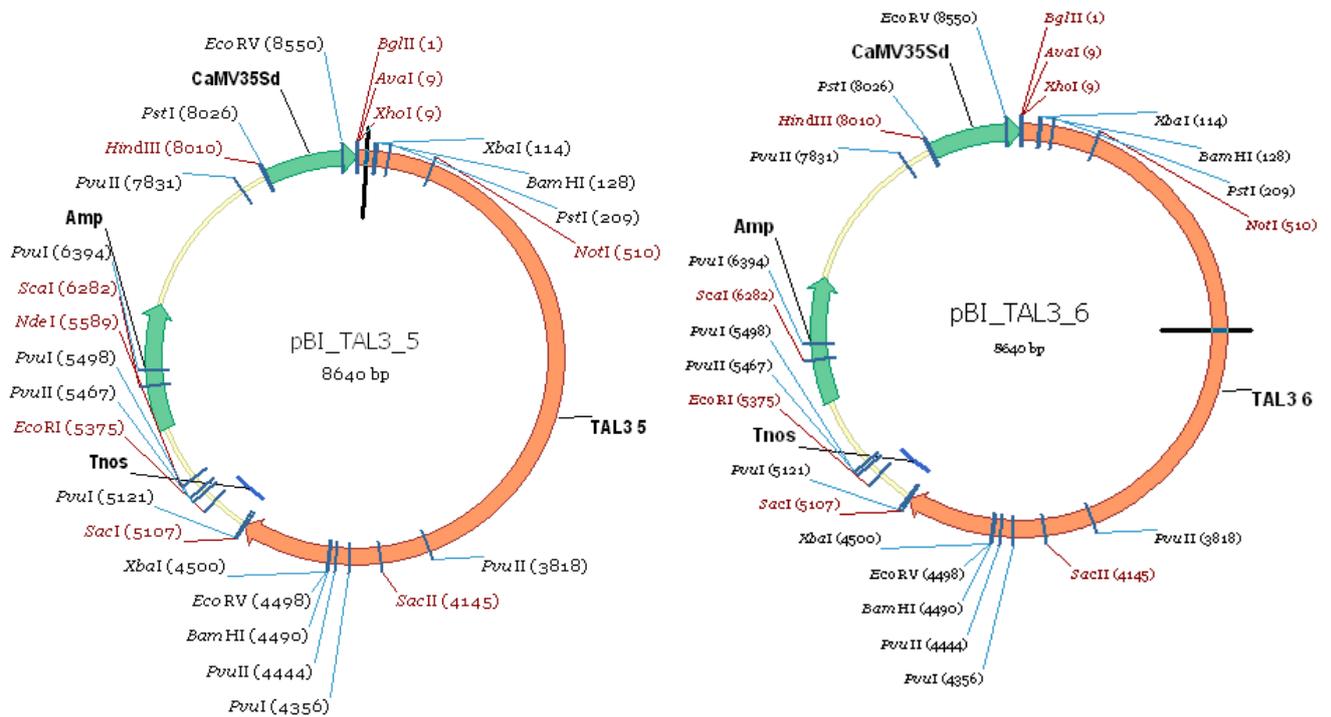


Figura 7: **Desenho representativo dos vetores pBI\_TAL3\_5 e pBI\_TAL3\_6.** Os vetores TALEN pTAL\_3\_5 e pTAL\_3\_6 foram digeridos com as enzimas *Bgl* II e *Sac* I e clonados no mesmo sítio do vetor pBI426, gerando os vetores pBI\_TAL3\_5 e pBI\_TAL3\_6.

## 2. Vetor de recombinação pREC56

O vetor pREC56 foi construído com o intuito de recombinar o gene de seleção *AHAS*, que confere resistência ao herbicida Imazapyr<sup>®</sup> (Arsenal), no local dos genes *FAD2-1A* e *FAD2-1B*. Para isso, um fragmento de 500 pb da região 5'e 3' UTR da região alvo do TALEN foi utilizado para construir o vetor. Entre as regiões 5' e 3' UTR foi inserido o gene *AHAS*, gerando o cassete 56. Este cassete foi então inserido no vetor pREC, gerando o vetor pREC56.

## 3. Transformação das plantas de soja por biobalística

### **6.2.5. Preparo das micropartículas utilizadas para o bombardeamento**

Para o preparo das micropartículas separou-se 60 mg de micropartículas de tungstênio, que foram colocadas em um microtubo de 1,5 mL, onde foi adicionado 1 mL de etanol 70 % para esterilização. Homogeneizou-se no vortex por 15 min. e centrifugou-se por 5 min. a 3.000  $\times g$ . Com o auxílio de uma micropipeta de 1 mL, o sobrenadante foi descartado. Após esterilização, as micropartículas foram lavadas com 1 mL de água destilada estéril, misturada vigorosamente por 15 min. em um vortex e, em seguida, centrifugadas por 5 min. a 3.000  $\times g$ . As micropartículas foram lavadas 4 vezes, e após a última lavagem foram então solubilizadas em glicerol 50% (vol/vol), sonicadas por 5 min, agitadas no vórtex por 30 s e distribuídas em alíquotas de 50  $\mu L$ .

A uma alíquota das micropartículas foram adicionados 8  $\mu L$  do plasmídeo solubilizado a ser bombardeado e, rapidamente, homogeneizou-se em um vortex durante 2s. Então, foram adicionados 50  $\mu L$  de  $CaCl_2$ , 20  $\mu L$  de espermidine 0,1 M, e a solução foi homogeneizada e incubada a temperatura ambiente, sob baixa agitação, em um vórtex, por 10 min. Em seguida, a amostra foi centrifugada por 10 s a 3.000  $\times g$  e o sobrenadante descartado com o auxílio de uma micropipeta de 100  $\mu L$ . Ao precipitado foram adicionados 150  $\mu L$  de etanol absoluto, homogeneizado e centrifugado a 3.000  $\times g$  por 10 s. O sobrenadante foi descartado com o auxílio de uma pipeta de 100  $\mu L$  e, novamente lavado com etanol absoluto. Ao final da segunda lavagem, foram adicionados 24  $\mu L$  de etanol absoluto ao precipitado, e vigorosamente homogeneizado. A amostra foi então submetida à sonicação por 2-3 s e distribuída em alíquotas de 3,2  $\mu L$  na região ventral de cada membrana, previamente posicionada no suporte de membranas, conforme metodologia de transformação genética, descritas em Aragão e colaboradores (2000) e Rech e colaboradores (2008).

### **6.2.6. Preparo do meristema apical de soja para bombardeamento**

Aproximadamente 30 sementes de soja do cultivar BR16 foram esterilizadas em 30 mL de etanol 70 % por 1 min., depois submergidas em 30 mL de hipoclorito de sódio 1% por 20-30 min., e lavadas cinco vezes com água destilada. As sementes descontaminadas foram incubadas por 16-18 horas em água estéril à temperatura ambiente. Aproximadamente 6 a 8 sementes descontaminadas foram selecionadas e transferidas para uma placa estéril contendo papel Whatman estéril. Cada semente teve seus cotilédones abertos, as folhas primárias cortadas, e a região do meristema apical exposta.

### **6.2.7. Bombardeamento dos meristemas**

Os eixos embrionários foram posicionados (10-15 axes por placa), em uma placa de 5 cm de diâmetro, contendo 12 mL de meio apropriado para bombardeamento (meio basal de sais MS (Murashige & Skoog, 1962), sacarose 3% e fitagel 0.8%, pH 5,7 (Sigma). Os eixos embrionários foram dispostos em um círculo no centro da placa, com a região do meristema apical direcionada para cima. O bombardeamento dos eixos embrionários foi realizado de acordo com Rech e colaboradores (2008).

### **6.2.8. Cultura e seleção das plantas transgênicas**

Imediatamente após o bombardeamento, os eixos embrionários foram transferidos para uma placa de vidro com 10 cm de diâmetro e as múltiplas brotações foram induzidas pela completa imersão em 15 mL de meio de indução (Meio basal de sais MS, suplementado com benzilaminopurina 10 mg/L, sacarose 3 % e ágar 0,6 % pH 5.7 (Sigma) por 16 h no escuro, a 26 °C.

Após este período, os embriões foram transferidos para um frasco “baby food” contendo 20 mL de meio de seleção (Meio basal MS suplementado com sucrose 3 %, Imazapyr® (Arsenal) 500 nM, ágar 0,6%, pH 5.7 (Sigma) e foram cultivados a 28 °C, com o fotoperíodo de 16 horas. Após aproximadamente 4 semanas em meio de cultura, os meristemas apicais bombardeados que produziram alongamento dos brotos e das raízes foram transferidos individualmente, para um copo plástico contendo 0,2 dm<sup>3</sup> de solo fertilizado:vermiculita (1:1). Os copos foram cobertos com um saco plástico, selado com uma liga de borracha e mantidos na casa de vegetação a 25 °C . Após uma semana, a liga foi removida e na semana seguinte, o saco plástico foi removido. Assim que as plantas aclimatadas alcançaram aproximadamente 10 cm de comprimento foram transferidas para um vaso contendo 5 dm<sup>3</sup> de solo fertilizado, para crescerem e produzirem sementes.

#### **4. Extração de DNA genômico**

O DNA das plantas T0 de soja selecionadas com Imazapyr® foi isolado utilizando 3 g de folha fresca das plantas. Elas foram maceradas em cadinhos e pistilo de cerâmica, com o auxílio de nitrogênio líquido e colocadas em 2 tubos falcons de 50 mL. Foram adicionados 15 mL do tampão CTAB 2x (brometo de cetil-trimetil-amônio) pré-aquecido à 65 °C, adicionados 30 µL de beta-mercaptoetanol, e colocado em um banho-maria a 65 °C por 30 min. Então, foram adicionados 15 mL de clorofil (24 partes de clorofórmio para 1 parte de álcool isoamílico) e agitado lentamente por 10 min. Os tubos foram então centrifugados por 10 min à 5.000 xg e a fase aquosa foi transferida para um novo tubo de 50 mL, e a extração com clorofil foi repetida. À nova fase aquosa foram adicionados 1,5 µL de RNase A 100mg/mL e incubado à 37 °C, por 30 min. À solução foi adicionado 0,6 x volume de isopropanol gelado, centrifugou por 20 min à 10.000 xg, o sobrenadante foi descartado e o tubo contendo o material precipitado foi colocado na posição invertida em cima de um papel toalha para secar. Após seco, o precipitado foi solubilizado em 300 µL de água destilada autoclavada, quantificado e armazenado a 20° C.

## 5. Oligonucleotídeos Iniciadores

Os iniciadores utilizados foram desenhados utilizando como referência a as sequencias dos genes, disponíveis no banco de dados acessado em [www.soybase.org](http://www.soybase.org) :Glyma 03g30070., Glyma 09g17170, Glyma 10g42470, Glyma 19g32940 e Glyma 20g24530. Os iniciadores utilizados para o reconhecimento do vetor nas plantas de interesse foram desenhados a partir da sequencia do vetor pBITAL3-5. Os iniciadores foram sintetizados pela empresa INTEGRATED DNA TECHNOLOGIES (IDT) e diluídos com água milliQ estéril, para a concentração de 1 M, sendo subestocados nas concentrações 100  $\mu$ M e 10  $\mu$ M. A descrição de todos os iniciadores utilizados estão apresentados na Tabela 2.

Tabela 2. Sequência dos iniciadores utilizados para amplificação dos genes por meio de PCR.

Nome do iniciador	Sequência (5'→3')
<b>Glyma03F</b>	GTGCCATTTGAAAAACCTCCA
<b>Glyma03R</b>	GATATATATATTTGAAGTCGTTACGA
<b>Glyma09F</b>	CGAAGCTCAGCTACTCTCAAGC
<b>Glyma09R</b>	GCCCTTTTGCCAAGACAGCCTT
<b>Glyma10F</b>	GTGGCCAAAGTGGAAGTTCAAGG
<b>Glyma10R</b>	GAGGGTTGTTTAAGTACTTGAAA
<b>Glyma19F</b>	GGGTACCATTTGAAAAACCTCAG
<b>Glyma19R</b>	GATATATATATTTGAAGTCGTTCTCGG

<b>Glyma20F</b>	GTGGCCAAAGTTGAAATTCAGCA
<b>Glyma20R</b>	GAGGGTTGTTTCAGGTACTIONTGGTGT
<b>AHAS P124 senso</b>	ACTAGAGATTCCAGCGTCAC
<b>AHAS 500 antisenso</b>	GTGGCTATACAGATACCTGG
<b>TAL353</b>	GATCGGATCGTGCTGTCACC
<b>TAL354</b>	CCGGCAACAGGATTCAATCT

## 6. Análise por PCR das plantas transgênicas.

Para a análise de reação em cadeia da polimerase (PCR) das plantas transformadas (T<sub>0</sub>) e das progênes (T<sub>1</sub>), o DNA das plantas selecionadas com o herbicida Imazapyr® (Arsenal) foi extraído conforme previamente descrito. A PCR foi realizada com Tris-HCL 10 mM (pH 8.4), KCl 50mM, MgCl<sub>2</sub> 2 mM, 160 µM de cada dNTP, 200 nM de cada iniciador, 2U de *Taq* polimerase (Invitrogen) e cerca de 20 ng de DNA genômico para um volume final de 25 µL. Para amplificação do gene de seleção *AHAS* a reação de PCR ocorreu sob as seguintes condições: 95 °C por 5 min para a desnaturação inicial do DNA, seguidos de 35 ciclos à 95 °C por 1 min, 55 °C por 1 min para anelamento dos iniciadores e 72 °C por 2 min. para extensão, seguidos da extensão final por 7 min a 72 °C. Para a confirmação da inserção do vetor nas plantas selecionadas, a PCR foi realizada como descrito anteriormente exceto pela temperatura de anelamento dos iniciadores, que foi 57 °C por 1 min. Para amplificação dos genes de interesse das plantas selecionadas, e das progênes, a reação de PCR foi realizada da seguinte forma: 95 °C por 5 min para a desnaturação inicial do DNA, seguidos de 35 ciclos à 95 °C por 1 min,

57 °C por 1 min para anelamento dos iniciadores e 68 °C por 2 min para extensão, seguidos da extensão final por 7 min a 68 °C para os genes Glyma 03g30070, Glyma 10g42470 e Glyma 19g32940. Para os genes Glyma 09g17170 e Glyma 20g24530 a única mudança foi na temperatura de anelamento, que foi de 60 °C por 1 min.

## **7. Eletroforese em gel de agarose**

Os produtos amplicados por PCR foram analisados em um gel de agarose 1 % (g/v) onde foram aplicados 10 µL do produto da PCR e coradas com brometo de etídeo na concentração de 0,5 µg/mL e visualizadas sob luz ultravioleta (UV). A amostra foi dissolvida em tampão de amostra 1× (azul de bromofenol 25%, xileno cianol 25%, Tris 50 mM pH7.6 e glicerol 6%) e corrida em um sistema de eletroforese horizontal com o tampão Tris-acetato 1× (Tris 40 mM, EDTA 1mM pH 8.0 e ácido acético 0.1%) sob 90 V constantes. Para analisar o tamanho dos amplicons de DNA, o marcador 1 kb plus DNA ladder Invitrogen foi utilizado como padrão, e o marcador High DNA Mass Ladder Invitrogen foi utilizado para estimativa de massa.

## **8. PCR-Eletroforese Capilar**

Os fragmentos de PCR foram amplificados e analisados em eletroforese capilar *QIAxcel* (*QIAGEN*, Alemanha) com o kit *FAST DNA* utilizando os primers e ciclos citados na análise por PCR das plantas e analisadas utilizando o *QIAXCEL ScreenGel Software* (Qiagen). O Marcador Molecular *QX Alignment Marker* 15pb/1000 pb foi utilizado para estimativa de massa e concentração das bandas.

## **9. Clonagem dos fragmentos de PCR**

Os produtos obtidos nas amplificações por PCR dos genes Glyma 03g30070, Glyma 09g17170, Glyma 10g42470 Glyma 19g32940 e Glyma 20g24530 foram ligados a um plasmídeo de clonagem pGEM® T-Easy (Promega), que possui extremidades de timina livres para complementar à adenina inserida à extremidade do amplicon durante o ciclo de extensão da PCR, além de um gene de resistência à ampicilina como antibiótico de seleção, e um sítio múltiplo de clonagem no gene que codifica a enzima  $\beta$ -Galactosidase. Para a inserção do amplicon no vetor de clonagem foram utilizados 6  $\mu$ L de solução tampão T4 DNA Ligase 2x, 3 U de T4 DNA Ligase, 50 ng de , e 4  $\mu$ L do produto de PCR. As reações de ligação ocorreram por 16 horas a 4 °C.

## 10. Transformação em bactérias competentes

Uma alíquota contendo as bactérias competentes de *Escherichia coli* DH5-alpha (100  $\mu$ L) foi incubada no gelo por 45 min., após receberem 5  $\mu$ L do produto de ligação em seguida, as células foram submetidas a choque térmico sendo colocadas por 1 min. a 42 °C, e depois incubadas 2 min no gelo. Em seguida, foi adicionado 1 mL de meio LB (Luria-Bertani) à solução de suspensão celular de bactérias após a transformação e, elas foram matidas sob agitação constante por 45 min. a 37 °C. Após esse período, toda a amostra foi espalhada em placas seletivas contendo meio LB ágar, ampicilina 100  $\mu$ g/mL, X-Gal 80  $\mu$ g/mL e IPTG 0,5 mM, e incubadas por 16 horas a 37 °C, e as colônias brancas crescidas foram selecionadas.

## 11. Extração de DNA Plasmideal

As colônias brancas foram então colocadas para crescer em 5 mL de meio LB líquido contendo ampicilina 100  $\mu$ g/mL em tubos de 15 mL, por 16 horas a 37 °C, e então foram centrifugadas por 10 min. a 4.000 RPM. O sobrenadante foi descartado e ao precipitado recebeu 340  $\mu$ L de TE gelado

(Tris 25mM pH 8.0 e EDTA 10 mM pH 8.0, e RNase 10 µg/mL) e incubado a temperatura ambiente por 5 min, e adicionado 340 µL da solução de lise alcalina (NaOH 0.2 N e SDS 1%), e homogeneizada gentilmente e incubada a temperatura ambiente por 5 min. Em seguida, foram adicionados 500 µL de acetato de potássio 5 M pH 4.8 gelado, homogeneizado e as amostras foram incubadas no gelo por 5 min. As amostras foram então centrifugada por 5 min. a 4 °C e 12.000 RPM, e o sobrenadante foi transferido para um novo tubo de polipropileno de 1,5 mL, onde foi adicionado 0,7x o volume de isopropanol, homogeneizado e centrifugado por 10 min. a 12.000 RPM à 4 °C, e o sobrenadante foi descartado. Ao precipitado foram adicionados 200 µL de etanol 70 % e centrifugado por 5 min a 12.000 RPM e 4 °C. O sobrenadante foi então descartado, e o material precipitado foi deixado para secar em exposição ao ar, e após a secagem foi solubilizado em 30 µL de água MilliQ com RNase 10 µg/mL, sendo incubado a 37 °C por 30 min. O DNA plasmidial foi então armazenado a -20 °C até o uso.

## **12. Esterificação dos ácidos graxos.**

Neste trabalho foi avaliada a concentração dos diferentes ácidos presentes nas sementes de soja. As reações foram realizadas em tubos de vidro âmbar de 3 mL. Foram cortadas 10 mg do cotilédone da semente de soja e colocadas nos tubos âmbar, e secos em um bloco digestor a 90 °C por toda a noite. Os frascos foram então retirados do bloco digestor e as sementes foram maceradas até virarem um fino pó. Ao pó, foram adicionado 500 µL da solução de KOH 0,5 N em metanol, e foi recolocado no bloco digestor por 1 h a 80 °C. Após esse período, os frascos foram retirados do bloco digestor e esfriados até alcançarem a temperatura ambiente, e então foram adicionado 2 mL da solução de H<sub>2</sub>SO<sub>4</sub> (6 mL de H<sub>2</sub>SO<sub>4</sub> em metanol), e as amostras foram agitadas no vórtex por 3 min., e aquecidas em um bloco digestor por 1 h a 80 °C. As amostras foram então retiradas do bloco digestor e resfriadas a temperatura ambiente por 10 min, e foram adicionado 2 mL de hexano, e agitadas no vórtex por 3 min. As amostras foram então colocadas na bancada para que possa se

separar em duas fases e 1 mL do sobrenadante foi coletado e colocado em um frasco para análise em GC-MS.

### **13. Análise por GC-MS**

A cromatografia gasosa acoplada a espectrometria foi conduzida utilizando modelo GC-2010 (SHIMADZU, Brasil) e um modelo GCMS-QP2010 (SHIMADZU, Brasil). Para isso 2  $\mu$ L da fase sobrenadante da reação de esterificação foi aplicada em uma coluna capilar Omega Wax 250 (30 m x 0,25 mm x 0,25  $\mu$ m) da Supelco (Sigma-24136). As amostras foram submetidas às seguintes condições durante a cromatografia gasosa: temperatura inicial à 50 °C e a temperatura final 240 °C, com aumento de 5 °C/min até atingir a 240°C, onde se mantém por 10 min., sendo um total de tempo de corrida de 58 min. A temperatura de injeção foi 250 °C. Um fluxo linear de 2 mL/min com o gás hélio sob pressão de 117,6 kPa foi utilizado, com um fluxo total de 25 mL/min. A velocidade linear foi configurada para 51,3 cm/s com um purgue de 3 mL/min e uma taxa de divisão de 10. O espectrômetro de massa foi operado em uma temperatura de 250 °C com uma velocidade de scan de 833-Hz de 100-500m/z por 58 min. Os dados dos picos da cromatografia e os fragmentos de MS foram pesquisados no banco de dados de fragmentos FAME.

## **RESULTADOS**

## RESULTADOS

### 1. Desenho do par de TALEN

Duas isoformas do gene *FAD2-1* estão presentes no genoma da soja, e correspondem aos genes *FAD2-1A* e *FAD2-1B*, que apesar de estarem localizados em cromossomos diferentes, possuem uma alta conservação das sequências de nucleotídeos. As sequências dos genes *FAD2-1A* (Glyma10g42470) e *FAD2-1B* (Glyma20g24530) disponíveis na página [www.soybase.org](http://www.soybase.org) foram alinhadas, e com o auxílio do programa *TAL Effector-Nucleotide Targeter*, TALE\_NT, disponível online em <http://boglabx.plp.iastate.edu/TALENT/>, e uma região conservada em ambos os genes foi selecionada para ser utilizada como alvo do par de TALEN. Esta região alvo é composta de uma sequência conservada de 28 pb no lado esquerdo do DNA e 28 pb no lado direito do DNA, e uma região espaçadora de 19 pb que não precisa ser conservada tendo em vista que esta região será alvo da enzima Fok I que não é específica de uma sequência de DNA. A região-alvo foi selecionada, sendo conservada em ambas as sequências, e dentro de 300 pb do códon de início dos genes *FAD2-1A* (Glyma10g42470) e *FAD2-1B* (Glyma20g24530) (Figura 8; Figura 9).

Após a seleção do alvo, os TALENs propostos foram montados utilizando a metodologia *Golden Gate*, que corresponde a uma metodologia de montagem de múltiplos fragmentos de DNA em uma determinada ordem numa única reação, resultando nos vetores pTAL3\_5 e pTAL3\_6. Os vetores pTAL3\_5 e pTAL3\_6 foram então digeridos com as enzimas de restrição *Bgl* II e *Sac* I e inseridos no vetor pBI426, gerando os vetores pBITAL3\_5, para o monômero direito, e pBITAL3\_6, para o monômero esquerdo. Estes novos vetores foram digeridos com a enzima de restrição *Pvu* II para confirmação das construções, gerando fragmentos de aproximadamente 620 pb, 1.000 pb, 2.350 pb e 3.800 pb como esperados (Figura 10).

Glyma10	1	ATGGGTCTAGCAAAGGAAACAACAATGGGAGGTAGAGGTCGTGGCCAA	50
Glyma20	1	ATGGGTCTAGCAAAGGAAACAATAATGGGAGGTGGAGGCCGTGGCCAA	50
Glyma10	51	AGTGAAGTTCAAGGAAGAAGCCTCTCTCAAGGGTTCCAAACACAAAGC	100
Glyma20	51	AGTTGAAATTCAGCAGAAGAAGCCTCTCTCAAGGGTTCCAAACACAAAGC	100
Glyma10	101	CACCATTCACTGTTGGCCAACCTCAAGAAAGCAATCCACCACACTGCTTT	150
Glyma20	101	CACCATTCACTGTTGGCCAACCTCAAGAAAGCCATCCACCACACTGCTTT	150
Glyma10	151	CAGCGCTCCCTCCTCACTTCATTCTCCTATGTTGTTTATGACCTTTCATT	200
Glyma20	151	CAGCGTTCCTCCTCACTTCATTGTCCTATGTTGTTTATGACCTTTCATT	200
Glyma10	201	TGCCTTCATTTTCTACATTGCCACCACCTACTTCCACCTCCTTCCTCAAC	250
Glyma20	201	GGCTTTCATTTTCTACATTGCCACCACCTACTTCCACCTCCTCCCTCACC	250
Glyma10	251	CCTTTCCCTCATTGCATGGCCAATCTATTGGGTTCTCCAAGGTTGCCTT	300
Glyma20	251	CCTTTCCCTCATTGCATGGCCAATCTATTGGGTTCTCCAAGGTTGCATT	300
Glyma10	301	CTCACTGGTGTGTGGGTGATTGCTCACGAGTGTGGTCACCATGCCTCAG	350
Glyma20	301	CTTACTGGCGTGTGGGTGATTGCTCACGAGTGTGGTCACCATGCCTCAG	350
Glyma10	351	CAAGTACCAATGGGTTGATGATGTTGTGGGTTTGACCCTTCACTCAACAC	400
Glyma20	351	CAAGTACCCATGGGTTGATGATGTTATGGGTTTGACCCTTCACTCAGCAC	400
Glyma10	401	TTTLAGTCCCTTATTTCTCATGGAAAATAAGCCATCGCCGCCACTCC	450
Glyma20	401	TTTLAGTCCCTTATTTCTCATGGAAAATAAGCCATCGCCGCCACTCC	450
Glyma10	451	AACACAGGTTCCCTTGACCGTGATGAAGTGTGTTGTCCAAAACCAAATC	500
Glyma20	451	AACACGGGTTCCCTTGACCGTGATGAAGTGTGTTGTCCAAAACCAAATC	500
Glyma10	501	CAAAGTTGCATGGTTTTCCAAGTACTTAAACAACCCTCTAGGAAGGGCTG	550
Glyma20	501	CAAAGTTGCATGGTACACCAAGTACCTGAACAACCCTCTAGGAAGGGCTG	550

Figura 8: Alinhamento das sequências dos genes *FAD2-1A (Glyma 10)* e *FAD2-1B (Glyma20)*. As sequências dos genes foram alinhadas e a região alvo foi determinada com o auxílio do programa TALE-NT. Glyma10 representa a sequência do gene *Glyma10g42470* e Glyma20 representa a sequência do gene *Glyma20g24530*. Em amarelo observamos as sequências alvo. Em verde está destacado a sequência espaçadora, necessária para a enzima FokI.

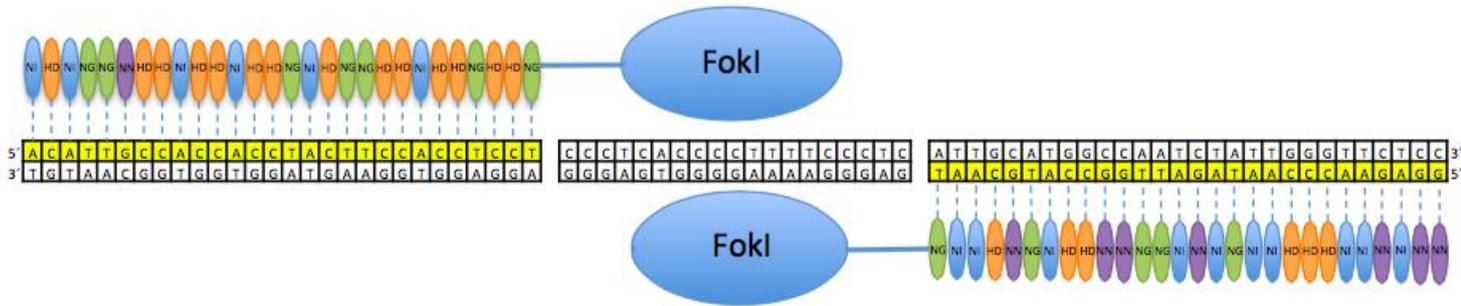


Figura 9: **Representação esquemática do par de TALEN.** Em amarelo observamos a sequência-alvo e seus respectivos RVDs. Cada RVD reconhece uma única base do DNA. Os RVDs estão representados em cores onde: azul NN, reconhece a base adenina; laranja HD, reconhece citosina; verde NG reconhece timina, e roxo NN reconhece guanina.

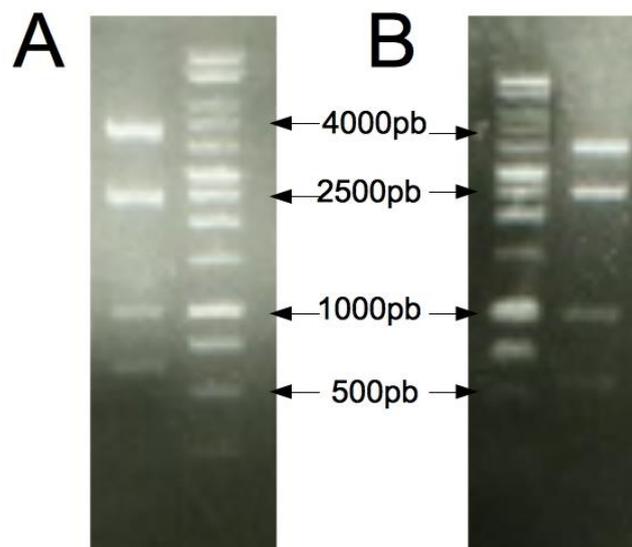


Figura 10: **Análise de restrição dos vetores pBITAL 3.5 (A) e pBITAL 3.6 (B) com a enzima *PvuII* em gel de agarose.** Os vetores pBITAL 3.5 e pBITAL 3.6 foram digeridos para confirmação do padrão de tamanho do vetor, onde foram observados 4 fragmentos com tamanhos aproximados 620 pb, 1.000 pb, 2.350 pb, 3.800 pb. MM: marcador molecular Gene Ruler™ 1Kb ladder (Invitrogen®).

## 2. Obtenção de plantas de soja transgênica expressando os vetores TALEN.

Os vetores pBITAL3\_5 e pBITAL3\_6 foram sequenciados para confirmação da sequência da construção, e, após confirmação, foram submetidos a midi-preparação dos vetores e utilizados para a transformação por biobalística em embriões de soja. Após o bombardeamento de 230 eixos embrionários, com os vetores pBITAL3\_5, pBITAL3\_6, e pREC56, os explantes foram cultivados *in vitro* por 6 semanas em meio de seleção contendo herbicida Imazapyr®, e cerca de 130 linhagens foram selecionadas transferidas para vasos, contendo solo na casa de vegetação. O DNA genômico destas plantas foi extraído a partir de discos foliares, e a presença dos vetores foi confirmada por amplificação por PCR, que identificou fragmentos internos do vetor. Quatro linhagens transgênicas T<sub>0</sub> denominadas 57, 119, 137 e 138 apresentaram o fragmento de 818 pb como esperado (Figura 11), porém em nenhuma das plantas observou-se a presença do vetor pREC56 (dados não mostrados).

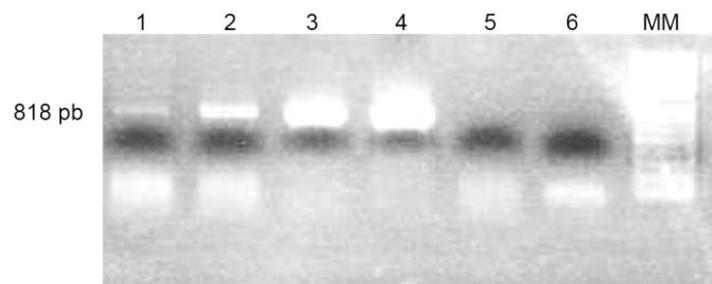


Figura 11: **Amplificação dos fragmentos do vetor pBITAL3\_5 e pBITAL3\_6.** O DNA das plantas transformadas foi extraído e amplificado para confirmação da introdução dos vetores, utilizando o par de iniciadores TAL 353 e TAL 354. 1, planta 137; 2, planta 138; 3, controle pBITAL 3\_5; 4, controle pBITAL 3\_6. 5, controle negativo. 6, branco. MM, marcador molecular 1Kb Plus DNA (Invitrogen).

### **3. Análise do perfil de óleo das sementes de soja das plantas transgênicas.**

Buscando observar mudanças fenotípicas na composição dos óleos das sementes de soja transgênicas, as plantas de soja que apresentaram o resultado positivo para a presença dos vetores por meio de PCR, tiveram todas suas sementes T1 coletadas, armazenadas e analisadas quanto a sua composição de ácidos graxos. As porcentagens de ácidos graxos das sementes de soja 57, 119, 137 e 138 foram obtidas a partir de uma reação de esterificação e análise dos ácidos em GC-MS. Para a montagem do gráfico apresentado na Figura 12, 5 sementes de soja foram selecionadas e utilizadas para obtenção da média de porcentagem de óleo e do desvio padrão. Como demonstrado na Tabela 3, as sementes da planta 57 obtiveram um perfil de óleo alterado, quando comparado ao tipo selvagem, apresentando sementes com o perfil contendo em média 13,67% de palmítico, 0,87% de esteárico, 31,48% de oleico, 51,15% de linoleico e 2,82% de linolênico. As sementes da planta 119 apresentaram o melhor perfil de óleo, dentre todas as plantas transformadas apresentando um perfil contendo 8,62% de palmítico, 1,24% de esteárico, 56,25% de oleico, 31,70% de linoleico e 2,17% de linolênico. Por fim, as sementes de planta 137 apresentaram um perfil de óleos composto de 12,61% de palmítico, 0,26% de esteárico, 29,99% de oleico, 53,07% de linoleico e 3,82% de linolênico, resultados também diferentes do encontrados em plantas selvagens. Em contraste, as plantas da linhagem BR 16 não transformadas apresentaram um perfil de óleo de 11,96% de palmítico, 2,22% de esteárico, 19% de oleico, 61,20% de linoleico e 5,60% de linolênico. As sementes das plantas 138 não apresentaram nenhuma modificação na composição dos ácidos graxos apresentando um perfil de óleo de 12,32% de palmítico, 3% de esteárico, 19,53% de oleico, 60,78% de linoleico e 5,23% de linolênico (Figura 12).

Tabela 3. Média da porcentagem das concentrações dos ácidos graxos em sementes de sojas nas linhagens transformadas.

	Palmítico	Estearíco	Oleico	Linoleico	Linolênico
57	13,67	0,87	31,48	51,15	2,82
119	8,62	1,24	56,25	31,7	2,17
137	12,61	0,26	29,99	53,07	3,82
138	12,32	3	19,53	60,78	5,23
BR16	11,96	2,22	19	61,2	5,6

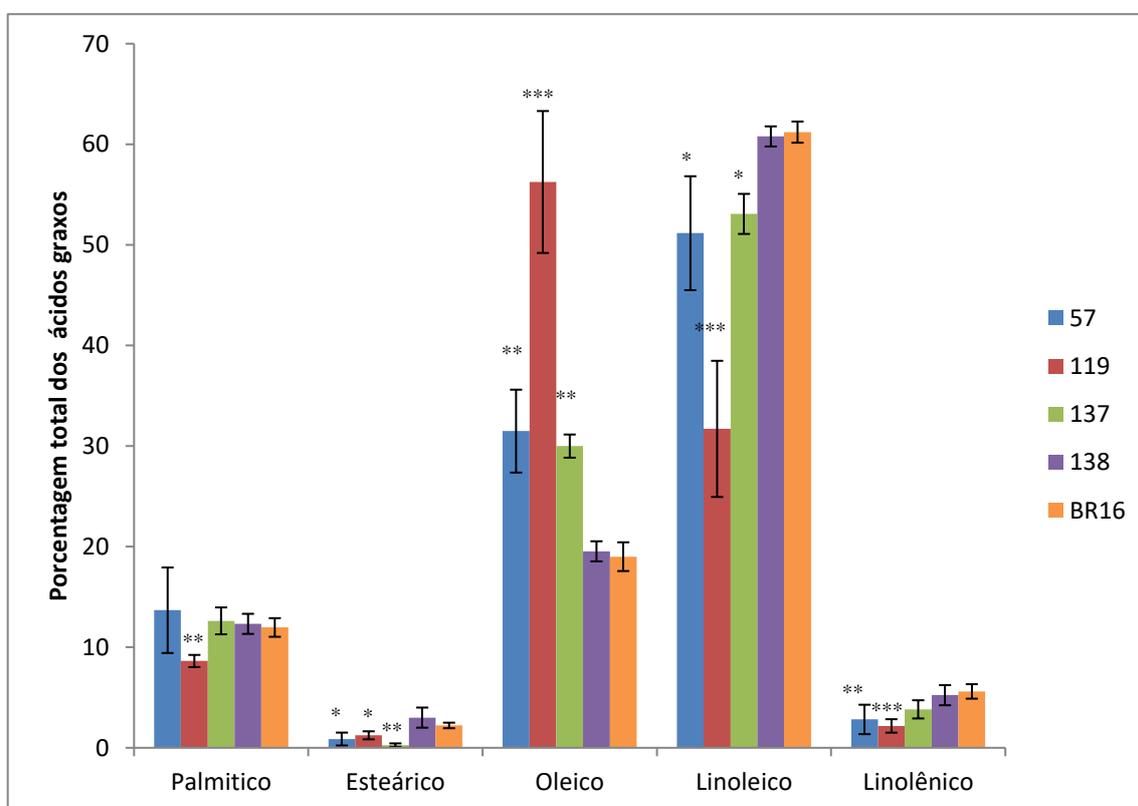


Figura 12: **Perfil de ácidos graxos das sementes T1 de soja.** As porcentagens de ácidos graxos foram determinadas a partir de reação de esterificação e analisadas em GC-MS. As porcentagens dos ácidos graxos palmítico, estearíco, oleico, linoleico e linolênico foram obtidas para as sementes de soja T1 das plantas 57, 119 e 137 e para as plantas BR16 não transformadas (WT). As linhas verticais representam desvio padrão encontrados nas 5 sementes analisadas.

Asteriscos representam diferença sigficativo quando comparado a BR16 (Teste T-Student; \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ ).

#### 4. Sequenciamento das plantas com perfil de óleo alterado

Após as análises dos perfis de óleos das sementes T1 das plantas 57, 119 e 137, 10 sementes com o melhor conteúdo de óleo oleico de cada uma das linhagem foram selecionadas para serem cultivadas na casa de vegetação e submetidas a futuras análises de sequenciamento. Após a germinação das sementes em copos plásticos, as plântulas foram transferidas para vasos na casa de vegetação, as quais posteriormente foram submetidas a extração de DNA genômico de discos foliares, que foram utilizados para amplificação por PCR dos genes *FAD2-1A (Glyma10)* e *FAD2-1B (Glyma20)*. No mesmo período, plantas BR16 não transformadas também foram germinadas e aclimatadas na casa de vegetação, e da mesma forma, o seu DNA genômico foi extraído de discos foliares para amplificação dos mesmos genes. A comparação entre os fragmentos amplificados das plantas BR16 e das plantas T0 e T1, apresentou diferenças de alguns pares de bases no tamanho dos fragmentos (Figura 13 e Figura 14).

A Figura 13 mostra os diferentes tamanhos de amplicons encontrados para o gene *Glyma10* em um gel de eletroforese capilar. Na Tabela 4 podemos observar os diferentes tamanhos em pares de bases e as diferentes concentrações detectadas no gel. As plantas de soja T1, 119-9, 119-37 e 119-38 não tiveram o tamanho dos seus amplicons alterados, já as plantas de soja T0 137 e 119 apresentaram um aumento de 9 pb e 4 pb no número dos pares de bases, respectivamente. As plantas de soja T1, 119-22, e 119-24 tiveram, respectivamente, 3 e 1 pares de bases aumentados enquanto a planta de soja T1 137-32 apresentou um tamanho de 493 pb, 3 pares de base menor do que o esperado.

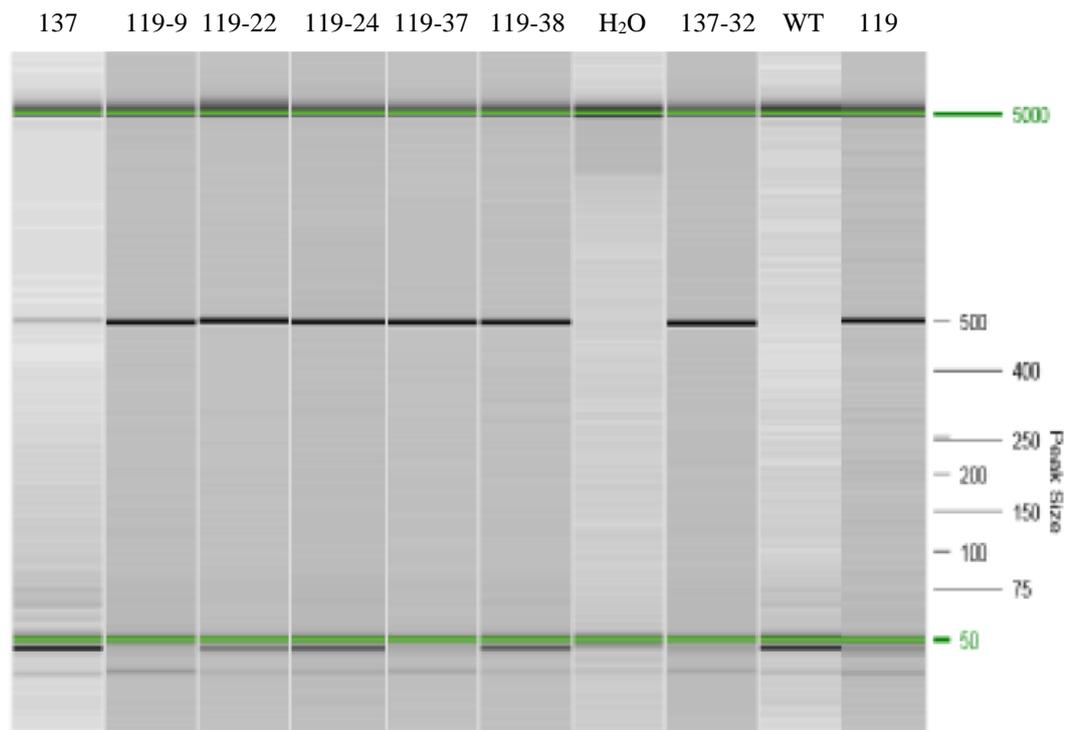
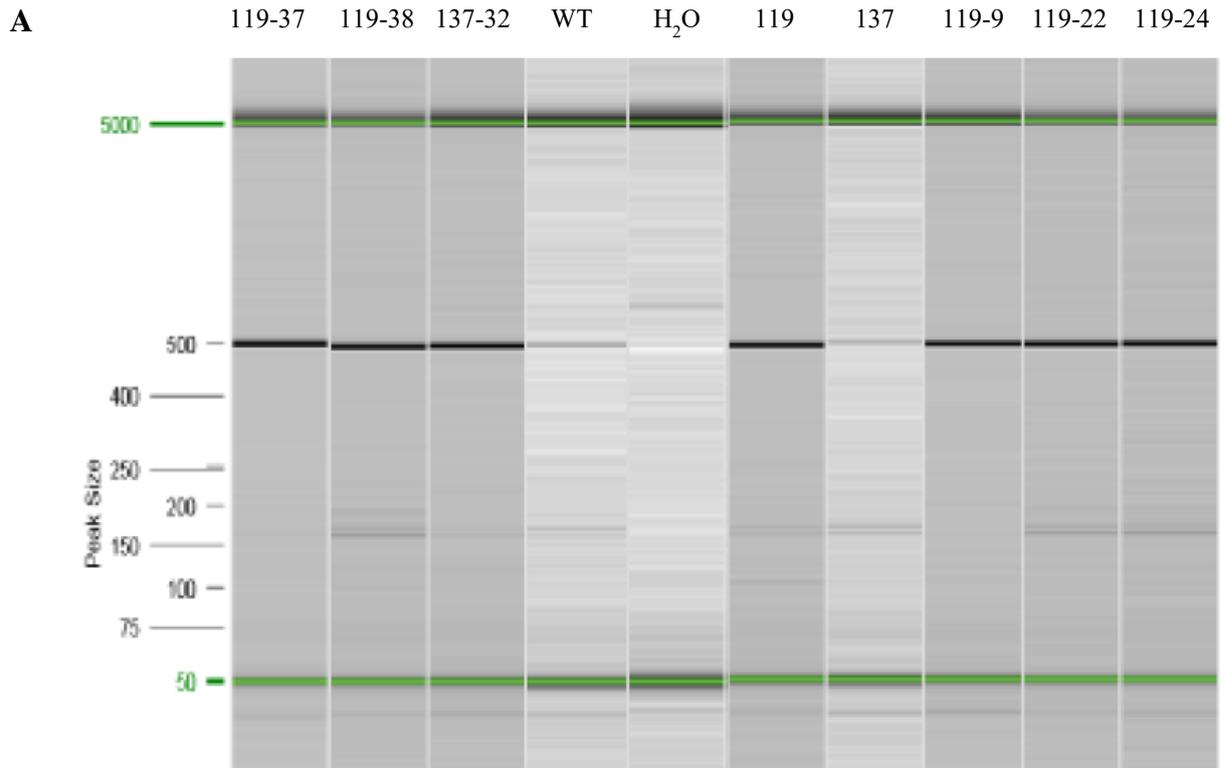


Figura 13: **Análise da amplificação do gene *Glyma10* em gel de eletroforese capilar.** Utilizando os primers específicos Glyma10F e Glyma10R, o gene *Glyma10* foi amplificado e pequenas diferenças de tamanho foram observadas. A1, planta 137-T0; A2, planta 119-9; A3, planta 119-22; A4, planta 119-24; A5, planta 119-37; A6, planta 119-38; A7, água; A8, planta 137-32; A9, controle; A10, planta 119-T0;

Tabela 4. Resultado da amplificação do gene *Glyma10*.

Gene	Planta	Tamanho Observado (pb)	Concentração (ng/μl)
<i>Glyma10</i>	137-T0	507	0.02
<i>Glyma10</i>	119-9	496	0.35
<i>Glyma10</i>	119-22	499	0.49
<i>Glyma10</i>	119-24	497	0.34
<i>Glyma10</i>	119-37	496	0.42
<i>Glyma10</i>	119-38	496	0.35
<i>Glyma10</i>	137-32	493	0.41
<i>Glyma10</i>	WT	496	0.00
<i>Glyma10</i>	119-T0	500	0.27

A Figura 14A apresenta os diferentes tamanhos de amplicons encontrados para o gene *Glyma20* em um gel de eletroforese capilar, e na Tabela 5 podemos observar os diferentes tamanhos em pares de bases e as diferentes concentrações detectadas no gel. As plantas de soja T1, 119-22, 119-24 e 119-38, apresentaram dois picos de amplicons com tamanhos diferentes, um próximo ao tamanho esperado, de 495 pb, 495 pb e 493 pb, respectivamente e um de 162 pb, 162 pb e 163 pb, respectivamente (Figura 14B). Nenhuma planta de soja transgênica apresentou o tamanho esperado de 496 pb. No entanto, as plantas de soja 119-T0, 137-32-T1 e 119-9-T1 demonstraram um tamanho inferior 4, 2 e 1 pb respectivamente, enquanto a planta de soja T1, 119-37 apresentou um aumento de 2 pb. A planta 137-T0 não teve nenhuma banda detectada.



**B**

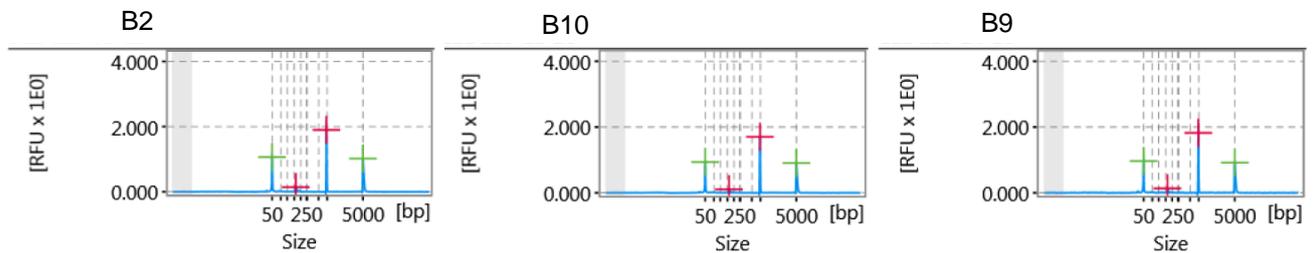


Figure 14: **Amplificação do gene *Glyma20* em gel de eletroforese capilar.** A, Utilizando os primers específicos *Glyma20F* e *Glyma20R*, o gene *Glyma20* foi amplificado e pequenas diferenças de tamanho foram observadas. B1, planta 119-37; B2, planta 119-38; B3, planta 137-32; B4, controle; B5, água; B6, planta 119-T0; B7, planta 137-T0; B8, planta 119-9; B9, planta 119-22; B10, planta 119-24; B, Eletroferograma das plantas 119-38, 119-22 e 119-24, mostrando os picos de fluorescência e o tamanho em pares de base.

Tabela 5. Resultado da amplificação do gene *Glyma20*.

Gene	Planta	Tamanho Observado (pb)	Concentração (ng/μl)
<i>Glyma20</i>	119-37	498	0.57
<i>Glyma20</i>	119-38	162	0.05
<i>Glyma20</i>	119-38	493	0.46
<i>Glyma20</i>	137-32	494	0.30
<i>Glyma20</i>	WT	496	-
<i>Glyma20</i>	119-T0	492	0.29
<i>Glyma20</i>	137-T0	-	-
<i>Glyma20</i>	119-9	495	0.25
<i>Glyma20</i>	119-22	163	0.04
<i>Glyma20</i>	119-22	495	0.41
<i>Glyma20</i>	119-24	163	0.05
<i>Glyma20</i>	119-24	495	0.43

Para confirmar a introdução de mutações provindas da utilização da metodologia TALEN nas plantas de soja transformadas, os fragmentos amplificados com os iniciadores específicos para os genes *Glyma10* e *Glyma20*, foram clonados e sequenciados, inclusive os fragmentos amplificados provindo das plantas de soja transformadas, que de acordo com o gel de capilaridade, não apresentaram diferença no tamanho do fragmento, uma vez que os TALENs podem realizar pequenas inserções e/ou deleções que poderiam ao final resultar em um fragmento com o tamanho esperado. Da transformação, 10 colônias brancas foram selecionadas, crescidas em meio LB, e utilizadas para realização de midipreparação de plasmídeo. Os DNAs extraídos destas bactérias foram enviados para sequenciamento, e o resultado apresentou mutações nos genes *FAD2-1A* (Figura 15) e *FAD2-1B* (Figura 16). A Tabela 6 e 7 apresenta um resumo das mutações encontradas nas diferentes plantas analisadas para os genes *Glyma10* e *Glyma20*.

Para o gene *Glyma 10*, constatou-se por meio de sequenciamento na planta 119-24 uma deleção de 447 pb e uma única mutação em 1 pb, que foi denominada mutação do tipo *Glyma10a*, enquanto que para o gene *Glyma20* diferentes tipos de mutações foram observadas, como inserções, deleções e mutações nas sequências, denominadas mutações do tipo *Glyma20a*, *Glyma20b*, *Glyma20c*, *Glyma20d*, *Glyma20e*, *Glyma20f*, *Glyma20g*, *Glyma20h*, *Glyma20i*, *Glyma20j*,

Glyma20k e por fim a mutação do tipo Glyma20l. A Tabela 6 apresenta um resumo de todos os tipos de mutações encontrados para o gene *Glyma10* e a tabela 7 apresenta dos tipos de mutações encontrados para o gene *Glyma20*, sendo que apenas a mutação do tipo Glyma10a foi observada para o gene *Glyma10*, e a mutação do tipo Glyma20j foi a mais frequente (28.58%) no gene *Glyma20*, sendo encontrada nas plantas 119-T0, 119-24, 119-38 e 137-32. Também foi possível observar por meio da eletroforese capilar do gene *Glyma20* uma banda de 162/163 pb, com intensidade suficiente para ser detectada nas plantas 119-22, 119-24 e 119-38, mas nenhum resultado de sequenciamento apresentou a mesma quantidade de número de bases. As mutações do tipo Glyma20b e Glyma20c possuem deleções de 271pb sendo possível encontrar indels e mutações nos pares de base, sendo que todas as mutações observadas apresentaram apenas indels ou indels e mutações.

A Tabela 8 apresenta os diferentes tipos de mutações sequenciados nas diferentes linhagens, o número de linhagens que apresentaram a mutação e a frequência que a mutação ocorreu dentro do gene, sendo que para o gene *Glyma10* apenas um tipo de mutação foi observada (mutação do tipo Glyma10a), em uma linhagem. Já o gene *Glyma20* apresentou diferentes tipos de mutações (mutações do tipo Glyma20a até Glymal), sendo que a mutação mais frequente foi a Glyma20j, sendo que 4 diferentes linhagens apresentaram este tipo de mutação.



```

Glyma20      GTGGCCAAAGTTGAAATTCAGCAGAGAAGGCTCTCTCAAGGGTTCCAAACACAAAGCCACCATTCACTGTTGGCCAACTCAAGAAAGCCATTCCACCGCACTGCTTTGAGGGTTCCCTC
Glyma20a     -----GAAATTCAGCAAAATGTA AACCCACCTTTGGTGG-----ATTGCCAAACACAGGAGC-----
Glyma20b     GTGGCCAAAGTTGAAATTCAGCATTCCACTAATTTTCCACTCAAGT-----
Glyma20c     GTGGCCAAAGTTGAAATTCAGCATTCCACTAATTTTCCACTCAAGT-----
Glyma20d     GTGGCCAAAGTTGAAATTCAGCACAAATGGTTAGTTGAGATAAACCT-----GAATAAACTCAATTGCATGCA-----
Glyma20e     GTGGCCAAAGTTGAAATTCAGCTTTGGT-----
Glyma20f     GTGGCCAAAGTTGAAATTCAGCATATACTGC-----
Glyma20g     GTGGCCAAAGTTGAAATTCAGCACCCCGAGAAAATCTTACACTTCCAAGT-----ATTGAGCTTGTTCOA
Glyma20h     GTGGCCAAAGTTGAAATTCAGC-----
Glyma20i     GTGGCCAAAGTTGAAATTCAGCACCCAGAAATCGGAGACAAGAGACTTA-----
Glyma20j     GTGGCCAAAGTTGAAATTCAGCA-----
Glyma20k     GTGGCCAAAGTTGAAATTCAGCAGA-----
Glyma20l     GTGGCCAAAGTTGAAATTC-----
                *****

Glyma20      CATTTGGCTTTCAATTTCTAATTGCCACCACTACTTCCACCTCCCTCAGCCCTTTTCCCTCATTTGATGGCCAAATCTATTGGGTTCTCAGAGTTGCATTCAAGGTTGCATTCTTA
Glyma20a     -----TTCAAGTAAAGAACACATGTA AAAAAACAGCATTTC AAAAAAATTTTC-----
Glyma20b     -----TTTGGTTAAAGAACACATGGA--AAACAGCATTTC AAAAAAATTTTC-----
Glyma20c     -----TTTGGTTAAAGAACACATGGA--AAACAGCATTTC AAAAAAATTTTC-----ATCAAAC-----AACTGTTC
Glyma20d     -----
Glyma20e     -----
Glyma20f     -----
Glyma20g     TATTCCGTTCCAATGCTAGAGGGTA--A-----
Glyma20h     -----
Glyma20i     -----
Glyma20j     -----
Glyma20k     -----
Glyma20l     -----

Glyma20      TCATGGAAAATAAGCCATCGCCGCCACCCTCCAAACAGGGTTCCCTTGACCGTGATGAAGTGTGTTGCCAAAACCCAAAATCCAAAGTTGCATGGTACACCAAGTACCTGAACAACCCCTC
Glyma20a     -----CTAACACTTTTGGCTTAAGTTCA CACCAAGTACCTGAACAACCC-----
Glyma20b     TCATGTAATAA-----CAGACAATCAGAGCAAG-----G-----AAGAACGGTGAATAACTGAACAAGTACCTGAACAACCCCTC-----
Glyma20c     CCATGTACCAA-----CAGACAATCAGGGCAAG-----G-----AAGAACGGTGAATAACTGAACAAGTACCTGAACAACCCCTC-----
Glyma20d     CCATTGCCAATAAGCTAAAGACAA-----TTTGATAAACT--CATGCATAGTAATTAT-----CCAAATCATCACCGAGGACACCAAGTACCTGAACAACCC-----
Glyma20e     -----C-----CACACCAAGTACCTGAACAACCC-----
Glyma20f     -----TCAAACGGTGAAGTGC CCCCACCAAGTACCTGAACAACCCCTC-----
Glyma20g     -----CGACCATTC-----GCATGCTTTTCTCTACACCAAGTACCTGAACAACCCCTC-----
Glyma20h     -----TTTGACACCAAGTACCTGAACAACCCCTC-----
Glyma20i     -----G--GTATAGAAAGATCTACACCAAGTACCTGAACAACCCCTC-----
Glyma20j     -----CACCAAGTACCTGAACAACCCCTC-----
Glyma20k     -----CACCAAGTACCTGAACAACCCCTC-----
Glyma20l     -----ACACCAAGTACCTGAACAACCCCT-----
                *****
    
```

**Figura 16: Mutações induzidas por TALENs no gene *Glyma20 (FAD2-1B)* nas plantas transformadas.** Representação de todas as mutações observadas nas diferentes plantas. O sítio alvo do par de TALEN está indicado em vermelho e em rosa está indicado o espaço para atividade da enzima Fok I. Deleções estão indicadas por traços, e estrelas marcam sequências conservadas.

Tabela 6. Análise das mutações encontradas para o gene *Glyma10* nas diferentes linhagens e o perfil de oleico e linoleico encontrado em cada linhagem.

Linhagem	Geração	Gene	Fenótipo(%oleic/%linoleic)	Indel	Indel e mutações	alvos ok	Total de Colonias
119	T0	Glyma10		0.00	0.00	1.00	4
119-38	T1	Glyma10	63.02/26.77	0.00	0.00	1.00	3
119-37	T1	Glyma10	44.31/40.83	0.00	0.00	1.00	2
119-24	T1	Glyma10	66.35/21.11	0.50	0.00	0.50	2
119-22	T1	Glyma10	46.60/40.36	0.00	0.00	1.00	5
119-9	T1	Glyma10	52.63/34.11	0.00	0.00	1.00	1
137	T0	Glyma10		0.00	0.00	1.00	3
137-32	T1	Glyma10	29.80/49.96	0.00	0.00	0.00	0

Tabela 7. Análise das mutações encontradas para o gene *Glyma20* nas diferentes linhagens e o perfil de oleico e linoleico encontrado em cada linhagem.

Linhagem	Geração	Gene	Fenótipo(%oleic/%linoleic)	Indel	Indel e mutações	alvos ok	Total de Colonias
119	T0	Glyma20		0.07	0.33	0.6	15
119-38	T1	Glyma20	63.02/26.77	0.50	0.17	0.33	6
119-37	T1	Glyma20	44.31/40.83	0.00	0.14	0.86	7
119-24	T1	Glyma20	66.35/21.11	0.27	0.45	0.28	11
119-22	T1	Glyma20	46.60/40.36	0.14	0.14	0.72	7
119-9	T1	Glyma20	52.63/34.11	0.14	0.14	0.72	7
137	T0	Glyma20		0.00	0.00	0.00	0
137-32	T1	Glyma20	29.80/49.96	0.50	0.00	0.50	10

Tabela 8. Resumo dos diferentes tipos de mutações encontrados nos genes *Glyma10* e *Glyma20* e a frequência de cada tipo de mutação.

Gene	Tipo de Mutação	Numero de diferentes plantas de soja com Mutação	Frequencia deste tipo de Mutação
------	-----------------	--	----------------------------------

<b><i>Glyma20</i></b>	Glyma20a	1	7.15%
<b><i>Glyma20</i></b>	Glyma20b	1	10.71%
<b><i>Glyma20</i></b>	Glyma20c	3	14.28%
<b><i>Glyma20</i></b>	Glyma20d	1	3.57%
<b><i>Glyma20</i></b>	Glyma20e	1	3.57%
<b><i>Glyma20</i></b>	Glyma20f	1	3.57%
<b><i>Glyma20</i></b>	Glyma20g	1	3.57%
<b><i>Glyma20</i></b>	Glyma20h	1	3.57%
<b><i>Glyma20</i></b>	Glyma20i	2	7.15%
<b><i>Glyma20</i></b>	Glyma20j	4	28.58%
<b><i>Glyma20</i></b>	Glyma20k	1	3.57%
<b><i>Glyma20</i></b>	Glyma20l	2	10.71%
<b><i>Glyma10</i></b>	Glyma10a	1	100%

## 5. Busca por potenciais genes *Não alvos*

Estudos recentes sugerem que TALENs podem criar mutações fora do local desejado, em sequências similares à sequência alvo. Na família dos genes FAD2, existem os membros FAA2-1 e FAD2-2 que compartilham alta similaridades entre si. Para verificar a similaridade entre estes membros, realizou-se o alinhamento das sequências dos membros da família FAD2-1, os genes *Glyma10* (Glyma10g42470) e *Glyma20* (Glyma20g24530), e as sequências dos membros da família FAD2-2, *Glyma03* (Glyma03g30070), *Glyma09* (Glyma09g17170) e *Glyma19* (Glyma19g32940) (Figura 17).

```

Glyma10      CTCAAGAAAGCAATTCCACCACACTGCTTTCAGCGCTCCCTCCTCACTTCATTCTCCTAT
Glyma20      CTCAAGAAAGCCATTCCACCGCACTGCTTTCAGCGTTCCTCCTCACTTCATTGTCCCTAT
Glyma09      CTGAAGAAGGCAATTTCCACCACATTGCTTCCACCGTTCAACATTCCGTTCACTTCTCCTAC
Glyma03      ATCAAGAAGGTCATTCCACCTCACTGTTTCCAGCGTTCTGTTTTCCGCTCATTCTCCTAT
Glyma19      ATTAAGAAGGCCATTCCACCTCACTGTTTCCAGCGCTCTGTTTCTCCGCTCATTCTCCTAT
              *  * * * * *  *   * * *  * * *  * * *  * * *  * *   * *   * * * * * * * *

Glyma10      GTTGTATTATGACCTTTCATTGCTTCATTTTCTACATTGCCACCACC---TACTTCCAC
Glyma20      GTTGTATTATGACCTTTCATTGGCTTTCATTTTCTACATTGCCACCACC---TACTTCCAC
Glyma09      GTCCTCTATGACCTAACCATAGCCTCATGCCTCTTCTATGCCGAGTAAATTACATCCCT
Glyma03      GTTTACGACCTCACCATAGCCTTCTGCCTCTATTATGTTGCCACCACAT---TACTTCCAC
Glyma19      GTTTATGACCTCACCATAGCCTTCTGCCTCTATTATGTTGCCACCACAT---TACTTCCAC
              **  * * *  * * * * * *   *   * * * *   * * * * * *   * * * * * *

Glyma10      C---TCCTTCCTCAACCCTTTTCCCTCATTGCATGGCCAATCTATTGGGTTCTCCAAGGT
Glyma20      C---TCCTTCCTCACCCCTTTTCCCTCATTGCATGGCCAATCTATTGGGTTCTCCAAGGT
Glyma09      ACCCTTCCCCATGAAAACCTTCCCTCCTAGCATGGCCTCTCTATTGGTTTATCCAAGGT
Glyma03      C---TCCTTCCAGCCCTCTCTCTTTCTTGGCATGGCCAATCTACTGGGCTGTCCAAGGT
Glyma19      C---TCCTTCCCGTCCCTCTCTCTTTCTTGGCATGGCCAATCTATTGGGCTGTCCAAGGT
              *  *   * * * * *   * * * * * * * * * * * * * * * * * * * * * *

Glyma10      TGCCTTCTCACTGGTGTGTGGGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAG
Glyma20      TGCATTCTTACTGGCGTGTGGGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAGCAAG
Glyma09      TCCATCCTAACCAGGGTTTGGGTCATCGCACACGAATGCGGCCACCACGCCTTTAGCGAT
Glyma03      TGCATCCTTACTGGAGTTTGGGTCATTGCCCATGAGTGTGGCCACCATGCATTAGTGAC
Glyma19      TGCATCCTTACTGGTGTGGGTCATTGCCCATGAGTGTGGTCACCATGCATTAGTGAC
              *  * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Glyma10      TACCAATGGGTTGATGATGTTGTGGGTTTACCCTTCACTCAACACTTTTAGTCCCTTAT
Glyma20      TACCCATGGGTTGATGATGTTATGGGTTTACCCTTCACTCAGCACTTTTAGTCCCTTAT
Glyma09      CACCAATGGCTCGATGACCTTGTGGCCTAATCCTCCACTCACTTCTCCTAGTGCCCTAT
Glyma03      TACCAGTTGCTTGTGATGATATTGTTGGCCTTGTCTCCACTCCGGTCTCCTAGTCCCATAC
Glyma19      TACCAGCTGCTTGTGATGATATTGTTGGCCTTATCCTCCACTCCGCTCTCCTAGTCCCCTAC
              * * *   * * * * * * * * * * *   * * * * * * * *   * * * * * * * *
  
```

Figura 17: Alinhamento dos genes das família FAD2-1 e FAD2-2. Os genes da família FAD2-1, correspondentes aos genes *Glyma10* e *Glyma20*, foram alinhados com os genes da família FAD2-2 (*Glyma09*, *Glyma03* e *Glyma19*). Pares de base destacados na cor amarela representam as regiões direita e esquerda escolhidas como alvo do TALEN neste trabalho. Pares de base destacados em verdes representam a região espaçadora necessária para a atividade da enzima Fok I.

A análise comparativa das sequências dos genes da família FAD2-1 e FAD2-2, revelam alta similaridade entre os genes *Glyma10* e *Glyma20*, e os genes *Glyma03* e *Glyma19*. Para a região alvo selecionada neste trabalho, o gene *Glyma03*, possui 20 pb diferentes na região esquerda, sendo que destes, 6 pb são únicos deste gene, e 8pb diferentes na região alvo direita. A região espaçadora não necessita ser idêntica a dos genes alvos, visto que esta região é necessária para

atuação da enzima Fok I que não é específica de uma determinada sequência de DNA. Os genes *Glyma03* e *Glyma19*, possuem uma sequência contendo 5 pb diferentes da sequência alvo esquerda escolhida neste trabalho, e o gene *Glyma03* possui 5 pb diferentes na região direita, enquanto o gene *Glyma19* possui 4 pb diferentes da região alvo direita.

Tendo isto em vista resolveu-se estudar os genes da família FAD2-2, *Glyma03g30070* e *Glyma19g32940* como potenciais genes não alvo dos nossos TALEN. Tendo em vista a alta similariedade das sequencias e com objetivo de amplificar cada gene separadamente, um par de iniciadores foi desenhado específico para cada gene. Os possíveis genes não alvo foram amplificados a partir de DNA genômico extraído de discos foliares da geração T0 e T1 das plantas transformadas. Foram amplificados fragmentos com os tamanhos de 1391 pb e 1393 pb que correspondendo aos genes *Glyma03g30070*, *Glyma19g32940*, respectivamente.

A Figura 18 mostra o padrão de bandas encontrado para a amplificação do gene *Glyma03*. Nenhuma planta apresentou a banda com o tamanho esperado, sugerindo que todas as plantas tiveram o gene *Glyma03* alterado pelo TALEN. Apesar de não serem evidentes as diferenças no tamanho das bandas, a Tabela 9 apresenta o tamanho observado e a concentração das bandas, mostrando os diferentes tamanhos encontrados. A planta 137-T0 apresentou uma banda se 1425, sendo a única planta a apresentar uma banda maior do que o tamanho esperado de 1391. As plantas 119-22, 119-37, 119-38 e 137-32 apresentaram mais de 53 pb de diferença, sendo a planta 119-37 a que apresentou maior diferença quando comparado ao WT (60 pb), enquanto as plantas 119-T0, 119-9, 119-24 apresentaram mais de 28 pb a menos do que o esperado.

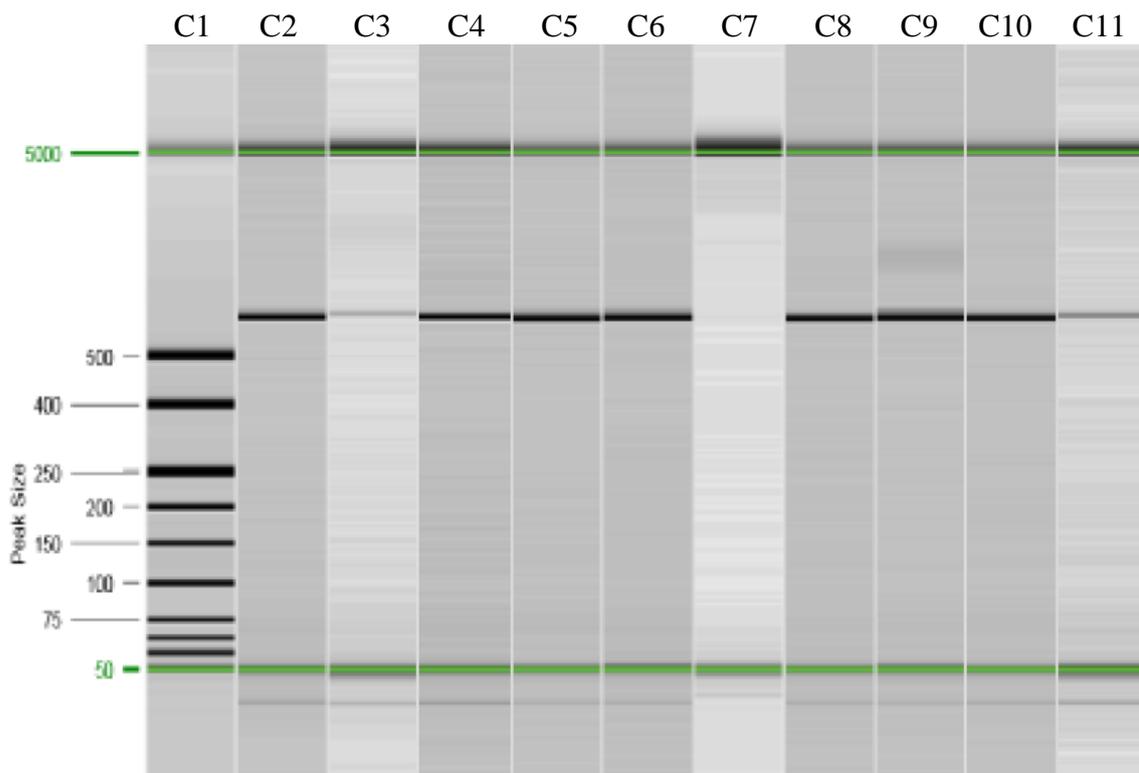


Figura 18: **Amplificação do gene *Glyma03* em gel de eletroforese capilar.** Utilizando os primers específicos Glyma03F e Glyma03R, o gene *Glyma03* foi amplificado e pequenas diferenças de tamanho foram observadas. C1, marcador molecular; C2, planta 119-T0; C3, planta 137-T0; C4, planta 119-9; C5, planta 119-22; C6, planta 119-24; C7, água; C8, planta 119-37; C9, planta 119-38; C10, planta 137-32; C11, controle;

Tabela 9. Resultado da amplificação do gene *Glyma03*.

Gene	Planta	Tamanho Observado (pb)	Concentração (ng/μl)
<b>Glyma03</b>	119-T0	1355	0.23
<b>Glyma03</b>	137-T0	1425	0.02
<b>Glyma03</b>	119-9	1363	0.18
<b>Glyma03</b>	119-22	1334	0.39
<b>Glyma03</b>	119-24	1347	0.37
<b>Glyma03</b>	119-37	1331	0.33
<b>Glyma03</b>	119-38	1338	0.35
<b>Glyma03</b>	137-32	1333	0.30
<b>Glyma03</b>	WT	1391	0.03

A Figura 19 apresenta o padrão de bandas encontrado quando amplificado o gene *Glyma19*. A planta 137-T0 foi a única que apresentou a banda no tamanho esperado de 1393 pb. As outras plantas apresentaram o padrão de bandas com tamanhos diferentes do esperado para o gene. A Tabela 10 expõe um resumo do tamanho das bandas observadas e suas concentrações. Nenhuma planta apresentou o tamanho maior do que esperado, sendo que todas as outras plantas, apresentaram uma banda com o tamanho menor do que o esperado, sugerindo que o gene *Glyma19* foi modificado pelo TALEN. A planta 119-24 foi a que apresentou a maior diferença no tamanho sendo observado uma banda de 1301 pb, 92 pb menor do que o esperado.

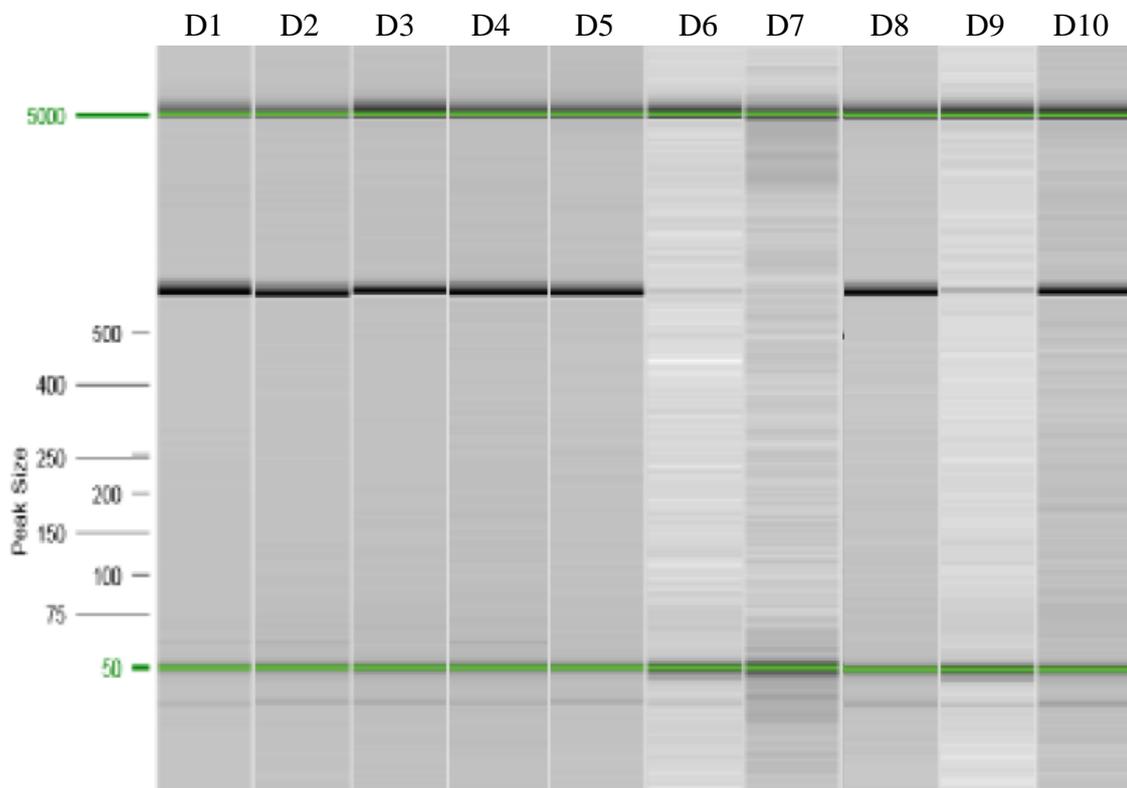


Figura 19: **Amplificação do gene *Glyma19* em gel de eletroforese capilar.** Utilizando os primers específicos Glyma19F e Glyma19R, o gene *Glyma19* foi amplificado e pequenas diferenças de tamanho foram observadas. D1, planta 119-

22; D2, planta 119-24; D3, planta 119-37; D4, planta 119-38; D5, planta 137-32; D6, controle; D7, água; D8, planta 119-T0; D9, planta 137-T0; D10, planta 119-9;

Tabela 10. Resultado da amplificação do gene *Glyma19*.

Gene	Planta	Tamanho Observado (pb)	Concentração (ng/μl)
<b>Glyma19</b>	119-22	1333	0.81
<b>Glyma19</b>	119-24	1301	0.52
<b>Glyma19</b>	119-37	1355	0.30
<b>Glyma19</b>	119-38	1331	0.37
<b>Glyma19</b>	137-32	1325	0.35
<b>Glyma19</b>	WT	1393	-
<b>Glyma19</b>	119-T0	1341	0.26
<b>Glyma19</b>	137-T0	1393	0.01
<b>Glyma19</b>	119-9	1363	0.20

Uma vez que os fragmentos amplificados apresentaram um tamanho diferente do esperado, sugerindo uma possível ação do TALEN, decidiu-se sequenciar os fragmentos amplificados buscando possíveis indels para confirmar a ação dos TALEN nestes não alvos. Os fragmentos amplificados foram então inseridos no vetor de clonagem pGEM-T Easy, e utilizados para transformação de bactérias competentes DH5 alfa segundo o protocolo de transformação por meio de choque térmico. Desta forma, várias colônias brancas foram obtidas, das quais 10 foram selecionadas para realização de minipreparação do vetor e enviadas para sequenciamento. O resultado do sequenciamento foi comparado com as sequências esperadas para cada gene, e mostrou que o par de TALEN foi capaz de realizar quebra na dupla fita nos genes *não alvo* gerando grandes indels (Figura 20 e Figura 21).

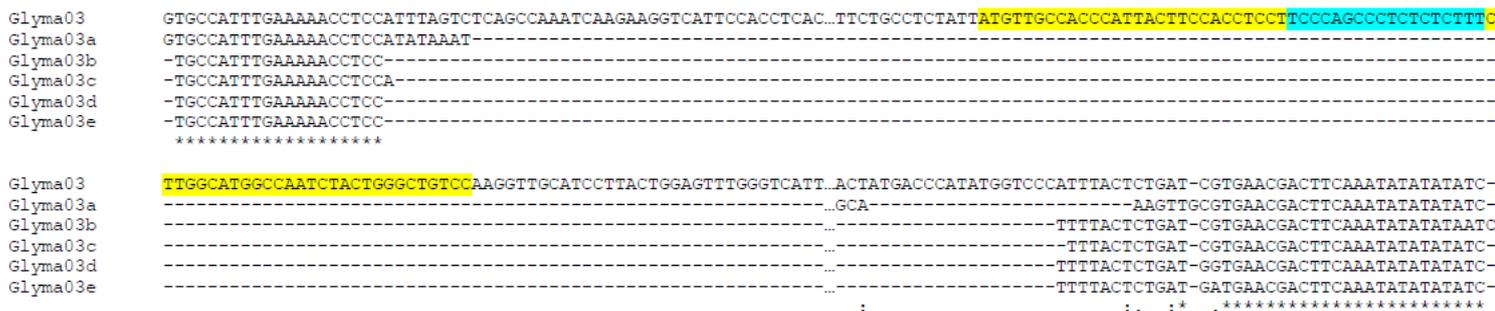


Figura 20: **Mutações induzidas por TALENs no gene *Glyma03* nas plantas transformadas.** Representação das mutações observadas nas diferentes plantas transformadas. O possível sítio alvo do par de TALEN está indicado em amarelo e em azul está indicado o espaço para atividade da enzima Fok I. Deleções estão indicadas por traços e estrelas indicam regiões conservadas.

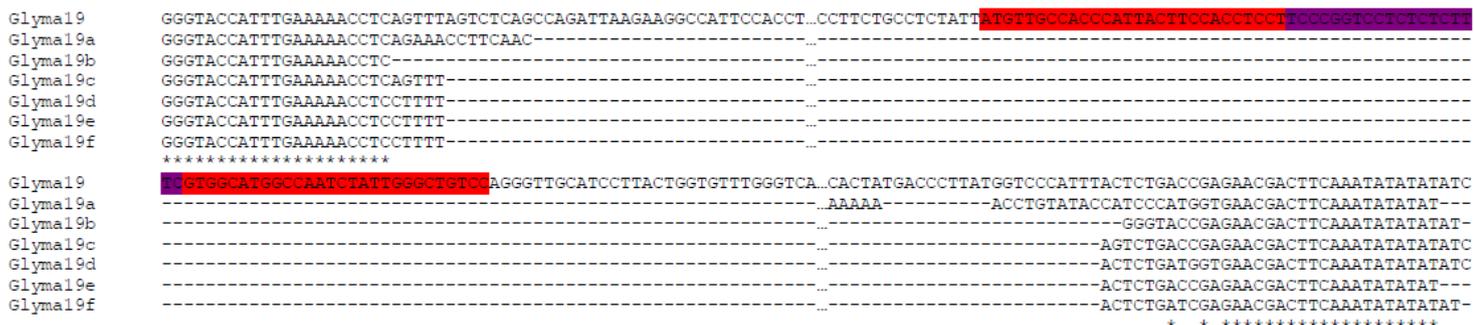


Figura 21: **Mutações induzidas por TALENs no gene *Glyma19* nas plantas transformadas.** Representação das mutações observadas nas diferentes plantas transformadas. O possível sítio alvo do par de TALEN está indicado em vermelho e em roxo está indicado o espaço para atividade da enzima Fok I. Deleções estão indicadas por traços e estrelas indicam regiões conservadas.

Na Tabela 11 e 12 estão apresentadas os diferentes tipos de mutações encontradas em cada planta nos genes *Glyma03* e *Glyma19*. Já na Tabela 13 podemos observar os diferentes tipos de mutações observadas e a frequência de cada mutação no gene. Para o gene *Glyma03* o tipo de mutação mais frequente foi

o *Glyma3b* (48.39%), sendo que 6 diferentes plantas transformadas apresentaram este tipo de mutação, e a planta 119-9 não apresentou esta mutação, enquanto que para o gene *Glyma19* o tipo de mutação mais frequente foi *Glyma19c* (36.36%), que estava presente 3 plantas de sojas transgênicas, sendo 137-T0, 119-24 e 119-22.

Tabela 11: Análise das mutações encontradas nos genes *Glyma03* nas diferentes linhagens obtidas.

Linagem	Geração	Gene	Fenótipo(%oleic/%linoleic)	Indel	Indel e mutações	alvos ok	Total de Colonias
119	T0	<i>Glyma03</i>		0.25	0.25	0.50	8
119-38	T1	<i>Glyma03</i>	63.02/26.77	0.40	0.20	0.40	5
119-37	T1	<i>Glyma03</i>	44.31/40.83	0.16	0.00	0.84	6
119-24	T1	<i>Glyma03</i>	66.35/21.11	0.70	0.10	0.20	10
119-22*	T1	<i>Glyma03</i>	46.60/40.36	0.43	0.14	0.43	7
119-9	T1	<i>Glyma03</i>	52.63/34.11	0.75	0.00	0.25	8
137	T0	<i>Glyma03</i>		0.83	0.00	0.17	6
137-32	T1	<i>Glyma03</i>	29.80/49.96	0.00	0.00	0.00	0

Tabela 12: Análise das mutações encontradas nos genes *Glyma19* nas diferentes linhagens obtidas.

Linagem	Geração	Gene	Fenótipo(%oleic/%linoleic)	Indel	Indel e mutações	alvos ok	Total de Colonias
119	T0	19		0.00	0.00	1.00	1
119-38	T1	19	63.02/26.77	0.00	0.00	1.00	8
119-37	T1	19	44.31/40.83	0.00	0.00	1.00	3
119-24	T1	19	66.35/21.11	0.20	0.40	0.40	10
119-22*	T1	19	46.60/40.36	0.31	0.46	0.23	13
119-9	T1	19	52.63/34.11	0.00	0.14	0.86	7
137	T0	19		1.00	0.00	0.00	2
137-32	T1	19	29.80/49.96	0.00	0.42	0.58	7

Tabela 13: Resumo das diferentes mutações encontradas nos genes *Glyma03* e *Glyma19*.

<b>Gene</b>	<b>Tipo de Mutação</b>	<b>Número de plantas com Mutação</b>	<b>Frequência da ocorrência da Mutação no gene</b>
<b><i>Glyma3</i></b>	Glyma3a	4	16.13%
<b><i>Glyma3</i></b>	Glyma3b	6	48.39%
<b><i>Glyma3</i></b>	Glyma3c	1	9.68%
<b><i>Glyma3</i></b>	Glyma3d	4	22.58%
<b><i>Glyma3</i></b>	Glyma3e	1	3.22%
<b><i>Glyma19</i></b>	Glyma19a	2	9.10%
<b><i>Glyma19</i></b>	Glyma19b	1	4.55%
<b><i>Glyma19</i></b>	Glyma19c	3	36.36%
<b><i>Glyma19</i></b>	Glyma19d	2	18.18%
<b><i>Glyma19</i></b>	Glyma19e	1	18.18%
<b><i>Glyma19</i></b>	Glyma19f	2	13.63%

## **DISCUSSÃO**

## DISCUSSÃO

A redução da expressão dos genes da família FAD2 já se mostrou uma poderosa ferramenta na busca de aumentar as concentrações de ácidos graxos monoinsaturados em soja (Clemente & Cahoon, 2009; Wagner et al., 2011; Murad et al., 2014). Dentre eles, os genes *Glyma10* (FAD2-1A) e *Glyma20* (FAD2-1B) são responsáveis pela conversão do ácido oleico (monoinsaturado) em linoleico (poli-insaturado). Estratégias de inibição da expressão destes genes já resultaram em acúmulo de ácido oleico em sementes de soja. Nos últimos anos o perfil do óleo em sementes de soja foi modificado com sucesso por meio da engenharia de vias metabólicas (Graef et al., 2009; Qu et al., 2012; Qian et al., 2010). Modificando vias específicas na rota de produção de óleo em sementes de soja é possível manipular sua composição, de forma que alcance as qualidades desejadas pelas indústrias. Porém ao alterarmos uma via metabólica, alteramos o conteúdo total das sementes, inclusive de alguns subprodutos. O aumento da concentração do óleo na semente leva a uma diminuição da produção de proteínas da sementes, o que faz com que a semente de soja tenha pouco valor como grão e farelo, porém sendo de grande interesse para a produção de óleo. Isto indica uma tendência na diversificação da produção de semente de soja, de acordo com o produto desejado. Por exemplo, uma semente rica em proteínas, seria de grande interesse para a produção de farelo, mas seria de pouco interesse para a produção de óleo, enquanto uma semente com alta quantidade de óleo seria de pouco interesse para a comercialização de farelo, mas de grande interesse para a produção de óleo.

Neste trabalho utilizou-se a metodologia *Golden Gate* para rapidamente e eficazmente montar-se os vetores TALENs, onde os desenhos dos vetores TALENs foram realizados utilizando o programa TALE-NT. Desde o começo do desenvolvimento dos TALENs, sabe-se que é recomendado que o alvo contenha um nucleotídeo timina na posição 0 (Moscou et al., 2009; Boch et al., 2009). Neste trabalho foi possível mostrar que os TALENs também podem se ligar a sequências que possuam outros nucleotídeos na posição 0, uma vez que os TALENs utilizado neste trabalho possui uma adenina e uma guanina na posição 0.

Haun e colaboradores (2014) também utilizaram a metodologia TALEN em plantas de sojas e obtiveram sementes de soja com o perfil de óleo correspondente a 80% de óleo oleico e menos de 4% de óleo linoleico. Haun e colaboradores (2014) sintetizaram diferentes 4 pares de TALEN, que tinham como alvo diferentes sequências dos genes FAD2-1A e FAD2-1B. Antes de transformarem as plantas de soja, fizeram uma avaliação da atividade destes TALEN em leveduras e em raízes de soja. Um dos pares de TALEN utilizados por eles possuía a mesma região como alvo do TALEN utilizado neste trabalho (Figura 22), porém o TALEN desenhado pelo grupo de Hauns continha 3 nucleotídeos a menos na região direita, quando comparado com o alvo deste trabalho, e o TALEN esquerdo possuía 8 nucleotídeos a menos no tamanho (Figura 22). Esse par de TALEN estudado, no trabalho de Haun, foi denominado FAD2\_T02 e teve sua atividade testada em um ensaio com leveduras, apresentando a melhor atividade dentre os demais pares testados. Porém, quando foi testado em ensaio contra raiz de soja, não apresentou nenhum clone contendo mutações, sendo sugerido pelo autor como tóxico para a planta, por talvez se ligar a locais não alvos, sendo assim descartado pelos autores. Neste trabalho, conseguiu-se demonstrar que aumentando em apenas alguns pares de bases o tamanho do TALEN, pôde-se aumentar a especificidade do TALEN, e diminuir, em partes, as ligações em locais não alvo, reduzindo sua toxicidade. O par de TALEN utilizado neste trabalho foi capaz de clivar a região alvo conseguindo produzir plantas de soja viáveis com mutações nos genes alvos.

```

Glyma10          ACATTGCCACCACCTACTTCCACCTCCTTCCTCAACCCTTTCCCTCATTGCATGGCCAATCTATTGGGTTCTCC
HaunFAD2_T02Glyma10 ACATTGCCACCACCTACTTCCACCTCCTTCCTCAACCCTTTCCCTCATTGCATGGCCAATCT-----
Glyma20          ACATTGCCACCACCTACTTCCACCTCCTTCCTCACCCTTTCCCTCATTGCATGGCCAATCTATTGGGTTCTCC
HaunFAD2_T02Glyma20 ACATTGCCACCACCTACTTCCACCTCCTTCCTCACCCTTTCCCTCATTGCATGGCCAATCT-----
*****

```

Figura 22: **Alinhamento das sequencias alvo e espaçadora dos TALENs utilizados neste trabalho e os utilizados no trabalho de Haun e colaboradores (2014).** Glyma 10 e Glyma 20 corresponde a região alvo do TALENs utilizados neste trabalho para o gene *Glyma10* e *Glyma 20*, respectivamente. HaunFAD2\_T02Glyma10 e HaunFAD2\_T02Glyma20 correspondem ao par de TALEN denominado FAD2-T02 no trabalho de Haun e colaboradores (2014). Em

amarelo estão destacado os pares de base alvos do TALEN, e em verde está destacado a sequencia espaçadora.

No trabalho de Haun e colaboradores (2014) o par de TALEN utilizado por eles para a produção de plantas transgênicas tinha como alvo uma região localizada a 24 pares de base de distancia da região escolhida como alvo neste trabalho (Figura 23). Esta região foi escolhida pelo grupo de Hauns devido a ensaios de teste de atividade em leveduras e em raízes de soja, onde o par de TALEN FAD2\_T04 apresentou um bom desempenho em ambos os testes. Este par de TALEN foi capaz de produzir plantas de soja contendo um perfil de óleo corespondente a 80% de óleo oleico e menos de 4% de óleo linoleico, enquanto que neste trabalho conseguiu-se obter plantas de soja contendo um perfil de óleo nas sementes de 56,25% de oleico e 31,7 de linoleico. Estas variações nos níveis de ácidos graxos pode ser resultado de diferenças na região alvo escolhida, ou podem ser provinientes da genética do cultivar de soja selecionado para o trabalho, tendo em vista que o genoma utilizado para a escolha do alvo foi proviniente do cultivar Williams 82, e o cultivar utilizado neste trabalho foi BR16. Outro fator importante que também pode ter influenciado na diferença de concentração de ácido graxo observado foram as condições de crescimento diferentes.

Por fim, o grande diferencial deste trabalho, quando comparado com o trabalho de Haun e colaboradores (2014), é a busca pela atividade fora do alvo dos TALENs. O par de TALEN utilizado por Hauns e colaboradores apresentava 4 e 5 nucleotídeos diferentes nas regiões esquerda e direita de reconhecimento do TALEN no gene *FAD2-2*, e quando analisados quanto a presença de atividade do TALEN, nenhuma mutação foi observada, diferentemente deste trabalho onde a análise dos genes da família *FAD2-2* revelou também grandes deleções e mutações provinientes da atividade dos TALENs.

Glyma10	1	ATGGGTCTAGCAAAGGAAACAACAATGGGAGGTAGAGGTCGTGGCCAA	50
Glyma20	1	ATGGGTCTAGCAAAGGAAACAATAATGGGAGGTGGAGGCCGTGGCCAA	50
Glyma10	51	AGTGGAAAGTTCAAGGGAAGAAGCCTCTCTCAAGGGTTCCAAACACAAAGC	100
Glyma20	51	AGTTGAAATTCAGCAGAAGAAGCCTCTCTCAAGGGTTCCAAACACAAAGC	100
Glyma10	101	CACCATTCACTGTTGGCCAACCTCAAGAAAGCAATCCACCACACTGCTTT	150
Glyma20	101	CACCATTCACTGTTGGCCAACCTCAAGAAAGCCATTCCACCACACTGCTTT	150
Glyma10	151	CAGCGCTCCCTCCTCACTTCAATCTCCTATGTTGTTTATGACCTTTCATT	200
Glyma20	151	CAGCGTCCCTCCTCACTTCAATGTCTATGTTGTTTATGACCTTTCATT	200
Glyma10	201	TGCCTTCATTTTCTACATTGCCACCACCTACTTCCACCTCCTTCCTCAAC	250
Glyma20	201	GGCTTTTCATTTTCTACATTGCCACCACCTACTTCCACCTCCTCCTCACC	250
Glyma10	251	CCTTTTCCTCATTGCATGGCCAATCTATTGGGTTCTCCAAGGTTGCCTT	300
Glyma20	251	CCTTTTCCTCATTGCATGGCCAATCTATTGGGTTCTCCAAGGTTGCATT	300
Glyma10	301	CTCACTGGTGTGTGGGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAG	350
Glyma20	301	CTTACTGGCGTGTGGGTGATTGCTCACGAGTGTGGTCACCATGCCTTCAG	350
Glyma10	351	CAAGTACCAATGGGTTGATGATGTTGTGGGTTTGACCCTTCACTCAACAC	400
Glyma20	351	CAAGTACCCATGGGTTGATGATGTTATGGGTTTGACCCTTCACTCAGCAC	400
Glyma10	401	TTTTAGTCCCTTATTTCTCATGGAAAATAAGCCATCGCCGCCATCACTCC	450
Glyma20	401	TTTTAGTCCCTTATTTCTCATGGAAAATAAGCCATCGCCGCCACCCTCC	450
Glyma10	451	AACACAGGTTCCCTTGACCGTGATGAAGTGTGTTGTCCAAAACCAAATC	500
Glyma20	451	AACACGGGTTCCCTTGACCGTGATGAAGTGTGTTGTCCAAAACCAAATC	500
Glyma10	501	CAAAGTTGCATGGTTTTCCAAGTACTTAAACAACCCTCTAGGAAGGGCTG	550
Glyma20	501	CAAAGTTGCATGGTACCCAAGTACCTGAACAACCCTCTAGGAAGGGCTG	550
Glyma10	551	TTTCTCTCTCGTCACTCACAATAGGGTGGCCTATGTATTTAGCCTTC	600
Glyma20	551	CTTCTCTCTCATCACTCACAATAGGGTGGCCTTGTATTTAGCCTTC	600
Glyma10	601	AATGTCTCTGGTAGACCCTATGATAGTTTGTCAAGCCACTACCACCCTTA	650
Glyma20	601	AATGTCTCTGGCAGACCCTATGATGGTTTGTCTAGCCACTACCACCCTTA	650

Figura 23: Alinhamento das sequências dos genes *Glyma10* e *Glyma20*, e comparação das regiões alvos selecionadas neste trabalho e no trabalho de Haun e colaboradores (2014). Em amarelo e em verde estão destacados as sequências alvo do TALEN utilizado neste trabalho e a sequência espaçadora. Em azul e roxo está destacado a sequência alvo e a sequência espaçadora utilizada no trabalho de Haun.

A maioria das metodologias utilizadas para a detecção de mutações em sequências é baseada em reação em PCR e, portanto, tendem a subestimar a frequência da atividade no alvo, já que grandes deleções que se estendem além dos limites dos produtos de PCR não são detectadas, e grandes inserções são amplificadas com menor eficiência que menores mutações, sendo assim menos prováveis de serem detectadas (Zischewski et al., 2016). Neste trabalho foi possível, através de PCR obter ampliações que apresentasse grandes deleções. Isto se deve ao amplo espectro amplificado, para obtenção de fragmentos que apresentasse todos os possíveis tipos de mutações sendo pequenas, ou grandes.

Apesar de vários trabalhos já terem publicado grandes deleções utilizando a ferramenta TALEN, em plantas, grandes deleções foram observadas em número reduzido. Gupta e colaboradores (2013) utilizando a metodologia TALEN, geraram deleções de até 69 kb e retiraram o gene *lincRNA* por completo em linhagem germinativa de zebrafish. Semelhantemente, Shan e colaboradores (2013) também observaram a deleção de genes, porém apenas quando 2 pares de TALENs foram utilizados com alvos no mesmo cromossomo. Nishizawa-Yokoi e colaboradores (2015) obtiveram grandes deleções em arroz, porém utilizaram plantas com defeitos na DNA ligase 4. Utilizando a ferramenta TALEN, foi observado em tabaco 4-28 pb deleções (Zhang et al., 2013), e em *Arabidopsis* foi observado deleções variando entre 4-50 pb (Cermak et al., 2011; Christian et al., 2013). Já em batata, deleções entre 1-25 pb foram observadas (Nicolia et al., 2015; Clasen et al., 2016), enquanto que em soja 1-63 pb deleções já foram observadas (Haun et al., 2014; Du et al., 2016). Em cevada foram observadas deleções variando de 1-36 pb (Wendt et al., 2013), e em milho pequenas deleções de 2-17pb foram observados (Liang et al., 2014; Char et al., 2015). Pequenas deleções variando entre 2-32 pb também foram observadas em trigo (Wang et al. 2014) e deleções um pouco maiores variando entre 1-72 pb foram observadas em arroz (Shan et al., 2013; Shan et al., 2015; Wang et al., 2015; Li et al., 2012). As maiores deleções foram obtidas em tomate, variando de 2-168pb (Lor et al., 2014). Neste trabalho observamos grandes deleções nos genes FAD2-1A e FAD2-1B variando de 240-562 pb, decorrentes da ação dos TALEN nos locais alvos. Um perfil semelhante de deleção e inserção foi observado por Haun e colaboradores (2014) em mutações

induzidas em raízes de soja utilizando o par de TALEN FAD2\_4, porém com um tamanho menor. Esse perfil pode ser resultado de uma tentativa de recombinação homóloga após a quebra da dupla fita do DNA.

Um vez que grandes deleções foram observadas nas plantas de soja transformadas, estas precisam ser melhor estudadas, já que plantas M-23 que possuem uma larga deleção no cromossomo 10, apresentam redução na produção de grão de soja, o que parece estar relacionado ao tamanho da deleção (Taylor et al., 2002). Neste sentido, as plantas desenvolvidas neste trabalho devem ter sua produtividade no campo estudada.

Diferentes metodologias já foram descritas na literatura para a identificação de mutações em eventos mediados por nucleases específica para sequências alvo. Dentre elas podemos destacar os ensaios de detecção de mutação baseado em gel, onde o protocolo mais comum utiliza a nucleasse CEL-I ou T7 endonuclease I (T7E1), ensaios de PCR fluorescente, análise de DNA derretido, e análise de polimorfismo no comprimento de fragmentos de restrição (RFLP). Eventos de Mutações também podem ser identificados através de sequenciamento do DNA, que são baseados em sequenciamento de amplificações de PCR por Sanger ou pela técnica de sequenciamento de nova geração. Neste trabalho foi possível identificar mutações utilizando Eletroforese Capilar, uma técnica rápida e econômica, e através de sequenciamento de amplificações de PCR. Apesar do gel de eletroforese capilar apresentar possíveis mutações nos genes, as mesmas não foram confirmadas por meio de sequenciamento.

Apesar de um maior número de sequências mutadas neste trabalho terem sido observadas no sequenciamento do gene FAD2-1B (*Glyma20*), acredita-se que o fenótipo analisado provém de mutações em ambos os genes, uma vez que estudos recentes sugerem que FAD2-1A e FAD2-1B são redundantes e contribuem de forma igual para a conversão de ácido oleico em linoleico, já que plantas com mutações apenas nos genes *FAD2-1A* ou *FAD2-1B* não tiveram seus níveis de oleico alterado (Haun et al., 2014).

Tendo em vista os eventos de duplicação do genoma da soja a 13 e 59 milhões de ano, estima-se que cerca de 75% dos genes preditos em soja estão

presentes em múltiplas cópias(Schmutz et al., 2010). Neste sentido é importante que as novas metodologias de edição de genoma sejam capazes de editar mais de um gene sem que seja necessário a Este trabalho também mostrou que é possível obter mutações em múltiplos genes usando apenas um par de TALEN, o que é uma característica importante para o seu uso em plantas, visto que diversas plantas possuem genes redundantes espalhados pelo genoma, mostrando ser possível usando apenas um par de TALEN, modificar mais de um gene.

A diminuição da atividade dos genes FAD2-1A e FAD2-1B já foi foco de diversos trabalhos utilizando a tecnologia de RNAi. Murad e colaboradores (2014) produziram plantas de soja contendo os genes FAD2-1A (*Glyma10*), FAD2-1B (*Glyma20*) e *FatB*, regulados para baixo devido a utilização da tecnologia de RNAi, sendo capazes de obter plantas possuindo até 94.58% de óleo oleico e menos de 3% de óleo palmitico, enquanto plantas não transgênicas possuíam em média 25% de oleico e 13% de palmítico. Este trabalho foi capaz de obter plantas de soja contendo 56,25% de óleo oleico e 8.62% de palmitico. Neste trabalho o gene *FatB* não foi alvo do par de TALEN, mas mesmo assim foi possível em alguns transgenes observar uma diferença significativa na concentração de palmítico, provavelmente devido a uma forma de compensação metabólica. Apesar de possuir um maior nível de oleico, plantas de soja expressando construções RNAi são sujeitas a variações na expressão do transgene durante as gerações, o que requer uma triagem da expressão por varias gerações.

Outro grande problema enfrentado por variedades de soja derivadas dessa metodologia é que elas são necessariamente transgênicas e necessitam ser aprovadas para comercialização, um processo que pode ser longo e caro. Por outro lado, plantas provinientes de metodologias de edição de genoma, como TALEN e CRISPR não necessariamente são consideradas transgênicas uma vez que, após triagem, podem se obter plantas que apenas apresentem a mutação no local desejado, sem a presença do vetor de transformação, e sem a adição de nenhum gene ou fragmento de DNA provindo de outro organismo. Neste sentido, a USDA já está aprovando o cultivo de plantas provinda destas metodologias, sendo um processo muito mais ágil e com custos significativamente reduzidos (Waltz, 2016).

A estratégia de edição de genoma pode desencadear alterações no DNA por meio de recombinação homóloga ou NHEJ providas da tentativa de reparo da quebra da dupla fita de DNA. Quando comparadas com as metodologias de genética clássica, as técnicas de edição de genoma apresentam uma grande vantagem já que possuem um tempo reduzido para obtenção de eventos, e possui um direcionamento para o alvo a ser modificado, o que em tese diminui o risco da ocorrência de mutações em diversos locais do genoma. Várias ferramentas vem utilizando a estratégia de com sucesso, entre elas ZFN, TALEN e CRISPR. ZFN foi uma das primeiras ferramenta desenvolvida (Urnov et al., 2010), porém é uma ferramenta laborosa e possui várias limitações de desenho e implementação, o que dificulta sua utilização por parte de diversos pesquisadores. Por outro lado, o TALEN e o CRISPR tiveram uma maior aceitação por parte dos pesquisadores, por serem mais práticos e por produzirem resultados em menor tempo.

Du e colaboradores (2016) recentemente realizaram uma análise comparativa entre as tecnologias de TALEN e de CRISPR em soja. Eles utilizaram como alvo os genes *GMPDS11* e *GmPDS18*, envolvidos na via de biossíntese de carotenoides, e desenharam um par de TALEN, e dois CRISPR contendo promotores diferentes. Quando comparado com CRISPR, a ferramenta TALEN obteve maior frequência de mutagêneses (17,5 - 21,1 %) do que o CRISPR contendo o promotor *AtU6-26* (11,7 - 18,1%), porém, ao utilizar o promotor *GMU6-16g-1* (43,4 - 48,1 %) houve um aumento significativo na frequência de mutagêneses.

Outros trabalhos publicados com CRISPR em soja mostraram uma média de frequência de mutagêneses baixa quando comparado ao TALEN. Sun e colaboradores (2015) estudando os genes *Glyma06g14180*, *Glyma08g02290* e *Glyma12g37050*, utilizaram os promotores U6-10 de soja e *AtU6-26* de *Arabidopsis thaliana*, e conseguiram uma média de frequência de mutagêneses de 14,7-20,2%, utilizando o promotor de soja, e 6,6-9,7% utilizando o promotor *AtU6-26* em raízes transgênicas. Cai e colaboradores (2015) também utilizaram a ferramenta CRISPR para obter mutações em genes endógenos e exógenos de soja, e obtiveram uma média de frequência de mutagêneses semelhante a observada utilizando TALEN.

Utilizando um gene exógeno como alvo eles conseguiram uma média de 36,7%, enquanto utilizando genes endógenos GmFEI2 e GmSHR conseguiram uma média variando entre 10-93.3% para o gene GmFEI2 e 35,7-50% para o gene GmSHR. Sun e colaboradores ressaltam que um aumento nas concentrações do complexo Cas9-SgRNA gerado pelo promotor utilizado para expressão pode levar a um aumento da atividade não alvo do CRISPR.

Além disso Cermák e colaboradores (2015) também observaram que TALEN e CRISPR tiveram eficiências similares em atingir o alvo em tomate, e Liang e colaboradores (2014) também observaram que TALEN e CRISPR tiveram eficiências de mutação similares em milho.

Vários critérios devem ser considerados na hora da escolha de uma metodologia de edição de genoma. Variáveis como a nucleasse específica, o stio alvo, o tipo de célula, o método de entrega, entre outros podem contribuir para determinar qual metodologia é melhor para um propósito particular. Um dos principais critérios para a seleção da metodologia utilizada é o quanto essa metodologia é específica para o alvo. Metodologias de edição de genoma são capazes de realizar uma quebra na dupla fita de DNA em lugares específicos, mas também são capazes de clivar em lugares não desejados, em locais que possuem alta homologia com as sequencias alvo. Vários estudos já mostraram que mutações fora do alvo é comum quando utilizando a metodologia CRISPR (Wright et al., 2014), e embora ocorra em menor taxa quando comparados a outras ferramentas de edição de genoma, a possibilidade de ação fora do local alvo utilizando a metodologia TALEN deve ser melhor estudada, uma vez que vários estudos foram realizados no sentido de mostrar a atividade fora do alvo em CRISPR e poucos dos estudos realizados com TALEN estudaram possíveis interações com não alvos.

Neste sentido este trabalho também teve como objetivo buscar como potenciais genes não alvos os genes pretencentes a família FAD2, uma vez que possuem maior similaridade com os genes alvos, sendo possível demonstrar a presença de grandes deleções causadas pela ação do TALEN nestes locais. Mesmo sendo possuindo difereças nas sequências, quando comparadas com a

sequência alvo, as sequências dos genes *Glyma03* e *Glyma19* foram alvo dos TALENs. Haun e colaboradores quando estudaram a possível utilização de um par semelhante ao utilizado neste trabalho, o par de TALEN denominado FAD2\_T02, durante o teste de atividade em raiz de soja, observaram uma potencial atividade fora do local alvo, que poderia estar levando a uma toxicidade na planta fazendo com que poucas ou nenhuma raiz fossem recuperadas. Neste trabalho observou-se que mesmo aumentando o número de bases na sequência alvo, não foi possível evitar uma atividade fora do alvo, mas foi possível recuperar plantas viáveis, porém em pequena quantidade.

Uma possível justificativa para a pouca quantidade de plantas recuperadas, uma vez que aproximadamente 230 eixos foram transformados, 130 foram selecionados e apenas 4 apresentaram os vetores TALEN, poderia ser a atividade fora do local alvo, que poderia ter aumentado a toxicidade das plantas, inviabilizando as plantas. Outra possibilidade poderia ser a grande quantidade de vetores que foram utilizados para a transformação, já que para a planta ser selecionada, ela deveria apresentar o vetor pREC, responsável por conferir resistência ao meio de seleção, e apresentar os dois vetores TALEN, um para cada lado alvo do DNA.

Outro diferencial neste trabalho foi a utilização de biobalística para a obtenção de plantas transgênicas. O trabalho de Haun e colaboradores (2014) utilizou a técnica de transformação por meio de agrobactéria para obtenção das plantas de soja contendo mutações nos genes *FAD2-1A* e *FAD2-1B*, enquanto neste trabalho obteve-se plantas de soja com mutações nos mesmos genes por meio de biobalística. Jacobs e colaboradores (2015), realizaram uma análise comparativa da eficiência de transformação utilizando as tecnologias de biobalística e agrobactéria, para os vetores de CRISPR, e observaram que o gene completo do Cas9 foi incorporado em apenas 12.5% dos eventos transformados com biobalística, enquanto 95% dos eventos transformados com agrobactéria apresentaram o gene completo. Jacobs e colaboradores (2015) sugerem que esta grande diferença observada entre o número de eventos transformados com biobalística e agrobactéria pode ser resultado de uma possível incompleta incorporação do gene Cas9 quando transformado por meio de biobalística. Esta

também poderia ser uma possível justificativa para o baixo número de eventos encontrado neste trabalho, uma vez que utilizou-se a metodologia de biobalística, e obteve-se uma planta que apresentou a amplificação do vetor TALEN em PCR, mas não obteve-se mutações nos genes.

Assim este trabalho foi capaz de corroborar a atividade de TALENs em plantas de soja, sendo capaz de produzir plantas contendo o conteúdo de óleo alterado como o esperado. Porém esta metodologia também se mostrou inespecífica sendo capaz de atingir genes não alvo. A determinação da atividade não alvo é essencial para o sucesso das metodologias de edição de genoma, e uma vez que a metodologia TALEN vem sendo amplamente utilizada faz-se necessário um estudo mais profundo para verificar e aperfeiçoar esta metodologia.

## **CONCLUSÃO**

## CONCLUSÃO

Os resultados apresentados neste trabalho nos permite concluir que:

- ❖ A ferramenta TALEN pode ser utilizada com sucesso em plantas de soja BR16;
- ❖ A montagem dos vetores TALENs utilizando a metodologia Golden Gate foi eficiente;
- ❖ O protocolo de transformação genética de soja por meio de biobalística mostrou-se eficiente para integrar os vetores TALEN no DNA da planta de soja;
- ❖ Foi possível observar mudanças no perfil de óleo da soja após a transformação com os vetores TALEN;
- ❖ Os genes *FAD2-1A* e *FAD2-1B* sofreram mutações, inserções e deleções decorrentes do uso da ferramenta TALEN;
- ❖ A metodologia de eletroforese capilar permitiu a identificação de diferentes tipos de mutação e a concentração de cada amplificação, mostrando-se uma ferramenta eficaz na identificação de possíveis mutantes;
- ❖ Foi observado atividade *off target* provinda da utilização da ferramenta TALEN nos genes *Glyma03* e *Glyma19*;
- ❖ É necessário realizar maior análise de *não alvo* quando utilizando metodologias de edição de genomas;

## **REFERÊNCIAS BIBLIOGRÁFICAS**

## REFERÊNCIAS BIBLIOGRÁFICAS

ALT, Jessie L. et al. Phenotypic and molecular analysis of oleate content in the mutant soybean line M23. **Crop Science**, v. 45, n. 5, p. 1997-2000, 2005.

ARAGÃO, F. J. L. et al. Selection of transgenic meristematic cells utilizing a herbicidal molecule results in the recovery of fertile transgenic soybean [*Glycine max* (L.) Merrill] plants at a high frequency. **TAG Theoretical and Applied Genetics**, v. 101, n. 1, p. 1-6, 2000.

AVALOS, José L.; FINK, Gerald R.; STEPHANOPOULOS, Gregory. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. **Nature Biotechnology**, v. 31, n. 4, p. 335-341, 2013.

BAKER, John M. et al. A metabolomic study of substantial equivalence of field-grown genetically modified wheat. **Plant Biotechnology Journal**, v. 4, n. 4, p. 381-392, 2006.

BAILEY, James E. Toward a science of metabolic engineering. **Science**, v. 252, n. 5013, p. 1668-1676, 1991.

BARRANGOU, Rodolphe. RNA-mediated programmable DNA cleavage. **Nature biotechnology**, v. 30, n. 9, p. 836, 2012.

BELLALOU, Nacer et al. Agricultural practices altered soybean seed protein, oil, fatty acids, sugars, and minerals in the Midsouth USA. **Frontiers in plant science**, v. 6, 2015.

BERBERICH, Thomas et al. Two maize genes encoding  $\omega$ -3 fatty acid desaturase and their differential expression to temperature. **Plant molecular biology**, v. 36, n. 2, p. 297-306, 1998.

BHAYA, Devaki; DAVISON, Michelle; BARRANGOU, Rodolphe. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. **Annual review of genetics**, v. 45, p. 273-297, 2011.

BIBIKOVA, Marina et al. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. **Molecular and cellular biology**, v. 21, n. 1, p. 289-297, 2001.

BOCH, Jens et al. Breaking the code of DNA binding specificity of TAL-type III effectors. **Science**, v. 326, n. 5959, p. 1509-1512, 2009.

BOCH, Jens; BONAS, Ulla. Xanthomonas AvrBs3 family-type III effectors: discovery and function. **Phytopathology**, v. 48, n. 1, p. 419, 2010.

BONAS, Ulla; STALL, Robert E.; STASKAWICZ, Brian. Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. **Molecular and General Genetics MGG**, v. 218, n. 1, p. 127-136, 1989.

BROWSE, J.; SOMERVILLE, C. Glycerolipid synthesis: biochemistry and regulation. **Annual review of plant biology**, v. 42, n. 1, p. 467-506, 1991.

BUHR, Tony et al. Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. **The Plant Journal**, v. 30, n. 2, p. 155-163, 2002.

CAI, Yupeng et al. CRISPR/Cas9-Mediated Genome Editing in Soybean Hairy Roots. **PLoS one**, v. 10, n. 8, p. e0136064, 2015.

CADY, Kyle C. et al. The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. **Journal of bacteriology**, v. 194, n. 21, p. 5728-5738, 2012.

CANAKCI, M.; MONYEM, A.; VAN GERPEN, J. Accelerated oxidation processes in biodiesel. **Transactions of the ASAE**, v. 42, n. 6, p. 1565, 1999.

CAO, Shijiang et al. A large and functionally diverse family of *Fad2* genes in safflower (*Carthamus tinctorius* L.). **BMC plant biology**, v. 13, n. 1, p. 1, 2013.

CARDINAL, Andrea J. et al. Molecular Analysis of in Two Soybean Populations Segregating for the, and Loci. **Crop science**, v. 51, n. 5, p. 2104-2112, 2011.

CARROLL, Dana. Genome engineering with targetable nucleases. **Annual review of biochemistry**, v. 83, p. 409-439, 2014.

CARTHEW, Richard W.; SONTHEIMER, Erik J. Origins and mechanisms of miRNAs and siRNAs. **Cell**, v. 136, n. 4, p. 642-655, 2009.

CERMAK, Tomas et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. **Nucleic acids research**, p. gkr218, 2011.

CERMAK, Tomas; STARKER, Colby G.; VOYTAS, Daniel F. Efficient design and assembly of custom TALENs using the Golden Gate platform. **Chromosomal Mutagenesis**, p. 133-159, 2015.

CHAR, Si Nian et al. Heritable site-specific mutagenesis using TALENs in maize. **Plant biotechnology journal**, v. 13, n. 7, p. 1002-1010, 2015.

CHRISTIAN, Michelle et al. Targeting DNA double-strand breaks with TAL effector nucleases. **Genetics**, v. 186, n. 2, p. 757-761, 2010.

CHRISTIAN, Michelle et al. Targeted mutagenesis of Arabidopsis thaliana using engineered TAL effector nucleases. **G3: Genes| Genomes| Genetics**, v. 3, n. 10, p. 1697-1705, 2013.

CLASEN, Benjamin M. et al. Improving cold storage and processing traits in potato through targeted gene knockout. **Plant biotechnology journal**, v. 14, n. 1, p. 169-176, 2016.

CLEMENTE, Tom E.; CAHOON, Edgar B. Soybean oil: genetic approaches for modification of functionality and total content. **Plant physiology**, v. 151, n. 3, p. 1030-1040, 2009.

CONG, Le et al. Multiplex genome engineering using CRISPR/Cas systems. **Science**, v. 339, n. 6121, p. 819-823, 2013.

CRADICK, Thomas J. et al. ZFN-site searches genomes for zinc finger nuclease target sites and off-target sites. **BMC bioinformatics**, v. 12, n. 1, p. 152, 2011.

CUNHA, N. B. et al. Correct targeting of proinsulin in protein storage vacuoles of transgenic soybean seeds. **Genet Mol Res**, v. 9, n. 2, p. 1163-1170, 2010.

CUNHA, Nicolau B. et al. Expression of functional recombinant human growth hormone in transgenic soybean seeds. **Transgenic research**, v. 20, n. 4, p. 811-826, 2011a.

CURTIN, Shaun J. et al. Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. **Plant physiology**, v. 156, n. 2, p. 466-473, 2011.

CURTIN, Shaun J. et al. Targeted mutagenesis for functional analysis of gene duplication in legumes. **Legume Genomics: Methods and Protocols**, p. 25-42, 2013.

DE PATER, Sylvia et al. ZFN-induced mutagenesis and gene-targeting in Arabidopsis through Agrobacterium-mediated floral dip transformation. **Plant biotechnology journal**, v. 7, n. 8, p. 821-835, 2009.

DE PATER, Sylvia et al. ZFN-mediated gene targeting of the Arabidopsis protoporphyrinogen oxidase gene through Agrobacterium-mediated floral dip transformation. **Plant biotechnology journal**, v. 11, n. 4, p. 510-515, 2013.

DELTCHEVA, Elitza et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. **Nature**, v. 471, n. 7340, p. 602-607, 2011.

DEMAREST, Zachary L. et al. Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. **BMC plant biology**, v. 16, n. 1, p. 225, 2016.

DEVEAU, Hélène et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. **Journal of bacteriology**, v. 190, n. 4, p. 1390-1400, 2008.

DU, Hongyang et al. Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. **Journal of biotechnology**, v. 217, p. 90-97, 2016.

ENGLER, Carola et al. Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. **PloS one**, v. 4, n. 5, p. e5553, 2009.

FAUSER, Friedrich; SCHIML, Simon; PUCHTA, Holger. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. **The Plant Journal**, v. 79, n. 2, p. 348-359, 2014.

FENG, Zhengyan et al. Efficient genome editing in plants using a CRISPR/Cas system. **Cell research**, v. 23, n. 10, p. 1229, 2013.

FENG, Zhengyan et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. **Proceedings of the National Academy of Sciences**, v. 111, n. 12, p. 4632-4637, 2014.

FICHTNER, Franziska; CASTELLANOS, Reynel Urrea; ÜLKER, Bekir. Precision genetic modifications: a new era in molecular biology and crop improvement. **Planta**, v. 239, n. 4, p. 921-939, 2014.

FINE, Eli J. et al. An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. **Nucleic acids research**, v. 42, n. 6, p. e42-e42, 2014.

FIRE, Andrew et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. **nature**, v. 391, n. 6669, p. 806-811, 1998.

FONFARA, Ines et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. **Nucleic acids research**, v. 42, n. 4, p. 2577-2590, 2014.

FU, Yanfang et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. **Nature biotechnology**, v. 31, n. 9, p. 822-826, 2013.

GAO, Han et al. TALE activation of endogenous genes in *Chlamydomonas reinhardtii*. **Algal Research**, v. 5, p. 52-60, 2014.

GARCÍA-VILLALBA, Rocio et al. Comparative metabolomic study of transgenic versus conventional soybean using capillary electrophoresis–time-of-flight mass spectrometry. **Journal of Chromatography A**, v. 1195, n. 1, p. 164-173, 2008.

GARTLAND, K. M. A. et al. Progress towards the ‘Golden Age’ of biotechnology. **Current opinion in biotechnology**, v. 24, p. S6-S13, 2013.

GASIUNAS, Giedrius et al. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. **Proceedings of the National Academy of Sciences**, v. 109, n. 39, p. E2579-E2586, 2012.

GOLDEN, Daniel E.; GERBASI, Vincent R.; SONTHEIMER, Erik J. An inside job for siRNAs. **Molecular cell**, v. 31, n. 3, p. 309-312, 2008.

GRAEF, George et al. A high-oleic-acid and low-palmitic-acid soybean: agronomic performance and evaluation as a feedstock for biodiesel. **Plant biotechnology journal**, v. 7, n. 5, p. 411-421, 2009.

GUPTA, Manju et al. Transcriptional activation of *Brassica napus*  $\beta$ -ketoacyl-ACP synthase II with an engineered zinc finger protein transcription factor. **Plant biotechnology journal**, v. 10, n. 7, p. 783-791, 2012.

GUPTA, Ankit et al. Targeted chromosomal deletions and inversions in zebrafish. **Genome research**, v. 23, n. 6, p. 1008-1017, 2013.

GUPTA, Rajat M.; MUSUNURU, Kiran. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. **The Journal of clinical investigation**, v. 124, n. 10, p. 4154-4161, 2014.

HAMMOND, Scott M.; CAUDY, Amy A.; HANNON, Gregory J. Post-transcriptional gene silencing by double-stranded RNA. **Nature Reviews Genetics**, v. 2, n. 2, p. 110-119, 2001.

HAUN, William et al. Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. **Plant biotechnology journal**, v. 12, n. 7, p. 934-940, 2014.

HENDEL, Ayal et al. Quantifying on-and off-target genome editing. **Trends in biotechnology**, v. 33, n. 2, p. 132-140, 2015.

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

HOWARD, Thomas P. et al. Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in Escherichia coli. **Proceedings of the National Academy of Sciences**, v. 110, n. 19, p. 7636-7641, 2013.

HSU, Patrick D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. **Nature biotechnology**, v. 31, n. 9, p. 827-832, 2013.

HUDSON, Laura C. et al. Soybean seeds: a practical host for the production of functional subunit vaccines. **BioMed research international**, v. 2014, 2014.

HUTVÁGNER, György; ZAMORE, Phillip D. A microRNA in a multiple-turnover RNAi enzyme complex. **Science**, v. 297, n. 5589, p. 2056-2060, 2002.

ISHINO, Yoshizumi et al. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. **Journal of bacteriology**, v. 169, n. 12, p. 5429-5433, 1987.

JIANG, Wenyan et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. **Nature biotechnology**, v. 31, n. 3, p. 233-239, 2013a.

JIANG, Wenzhi et al. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. **Nucleic acids research**, p. gkt780, 2013b.

JINEK, Martin et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. **Science**, v. 337, n. 6096, p. 816-821, 2012.

JOHNSON, Ross A. et al. Comparative assessments of CRISPR-Cas nucleases' cleavage efficiency in planta. **Plant molecular biology**, v. 87, n. 1-2, p. 143-156, 2015.

KAY, Sabine et al. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. **Science**, v. 318, n. 5850, p. 648-651, 2007.

KIM, Hye Joo et al. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. **Genome research**, v. 19, n. 7, p. 1279-1288, 2009.

KIM, Hyongbum; KIM, Jin-Soo. A guide to genome engineering with programmable nucleases. **Nature Reviews Genetics**, v. 15, n. 5, p. 321-334, 2014.

KIM, Won-Seok et al. Transgenic soybean plants overexpressing O-acetylserine sulfhydrylase accumulate enhanced levels of cysteine and Bowman-Birk protease inhibitor in seeds. **Planta**, v. 235, n. 1, p. 13-23, 2012.

KIM, Yang-Gyun; CHA, Jooyeun; CHANDRASEGARAN, Srinivasan. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. **Proceedings of the National Academy of Sciences**, v. 93, n. 3, p. 1156-1160, 1996.

KONDO, Takashi et al. Genetic engineering to enhance the Ehrlich pathway and alter carbon flux for increased isobutanol production from glucose by *Saccharomyces cerevisiae*. **Journal of biotechnology**, v. 159, n. 1, p. 32-37, 2012.

KOO, Taeyoung; LEE, Jungjoon; KIM, Jin-Soo. Measuring and reducing off-target activities of programmable nucleases including CRISPR-Cas9. **Molecules and cells**, v. 38, n. 6, p. 475-481, 2015.

KRISHNA, S. Sri; MAJUMDAR, Indraneel; GRISHIN, Nick V. Structural classification of zinc fingers SURVEY AND SUMMARY. **Nucleic acids research**, v. 31, n. 2, p. 532-550, 2003.

LEE, Hyung Joo; KIM, Eunji; KIM, Jin-Soo. Targeted chromosomal deletions in human cells using zinc finger nucleases. **Genome research**, v. 20, n. 1, p. 81-89, 2010.

LEE, Jeong-Dong et al. Environmental effects on oleic acid in soybean seed oil of plant introductions with elevated oleic concentration. **Crop Science**, v. 49, n. 5, p. 1762-1768, 2009.

LEE, Junwon et al. Designed nucleases for targeted genome editing. **Plant biotechnology journal**, v. 14, n. 2, p. 448-462, 2016.

LI, Jian-Feng et al. Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. **Nature biotechnology**, v. 31, n. 8, p. 688-691, 2013.

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

LI, Ling et al. A systems biology approach toward understanding seed composition in soybean. **BMC genomics**, v. 16, n. Suppl 3, p. S9, 2015.

LI, Ting et al. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. **Nucleic acids research**, v. 39, n. 1, p. 359-372, 2011a.

LI, Ting et al. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. **Nucleic acids research**, p. gkr188, 2011b.

LI, Ting et al. High-efficiency TALEN-based gene editing produces disease-resistant rice. **Nature biotechnology**, v. 30, n. 5, p. 390-392, 2012.

LIANG, Zhen et al. Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. **Journal of Genetics and Genomics**, v. 41, n. 2, p. 63-68, 2014.

LIN, Jer-Young et al. Structural and functional divergence of a 1-Mb duplicated region in the soybean (*Glycine max*) genome and comparison to an orthologous region from *Phaseolus vulgaris*. **The Plant Cell**, v. 22, n. 8, p. 2545-2561, 2010.

LIU, Jia; SHUI, Sai-lan. Delivery methods for site-specific nucleases: Achieving the full potential of therapeutic gene editing. **Journal of Controlled Release**, 2016.

LLOYD, Alan et al. Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 6, p. 2232-2237, 2005.

LOR, Vai S. et al. Targeted mutagenesis of the tomato PROCERA gene using transcription activator-like effector nucleases. **Plant physiology**, v. 166, n. 3, p. 1288-1291, 2014.

MA, Xingliang et al. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. **Molecular plant**, v. 8, n. 8, p. 1274-1284, 2015.

MAHFOUZ, Magdy M. et al. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. **Proceedings of the National Academy of Sciences**, v. 108, n. 6, p. 2623-2628, 2011.

MAKAROVA, Kira S. et al. An updated evolutionary classification of CRISPR-Cas systems. **Nature Reviews Microbiology**, 2015.

MALI, Prashant et al. RNA-guided human genome engineering via Cas9. **Science**, v. 339, n. 6121, p. 823-826, 2013.

MAO, Yanfei et al. Application of the CRISPR–Cas system for efficient genome engineering in plants. **Molecular plant**, v. 6, n. 6, p. 2008, 2013.

MEISTER, Gunter; TUSCHL, Thomas. Mechanisms of gene silencing by double-stranded RNA. **Nature**, v. 431, n. 7006, p. 343-349, 2004.

MELLO, Craig C.; CONTE, Darryl. Revealing the world of RNA interference. **Nature**, v. 431, n. 7006, p. 338-342, 2004.

MIAO, Jin et al. Targeted mutagenesis in rice using CRISPR-Cas system. **Cell research**, v. 23, n. 10, p. 1233, 2013.

MICHAEL, Todd P.; JACKSON, Scott. The first 50 plant genomes. **The Plant Genome**, v. 6, n. 2, 2013.

MILLER, Jeffrey C. et al. A TALE nuclease architecture for efficient genome editing. **Nature biotechnology**, v. 29, n. 2, p. 143-148, 2011.

MOJICA, Francisco JM et al. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. **Molecular microbiology**, v. 36, n. 1, p. 244-246, 2000.

MOJICA, F. J. M. et al. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. **Microbiology**, v. 155, n. 3, p. 733-740, 2009.

MOSCOU, Matthew J.; BOGDANOVE, Adam J. A simple cipher governs DNA recognition by TAL effectors. **Science**, v. 326, n. 5959, p. 1501-1501, 2009.

MOSER, Bryan R. et al. Evaluation of partially hydrogenated methyl esters of soybean oil as biodiesel. **European Journal of Lipid Science and Technology**, v. 109, n. 1, p. 17-24, 2007.

MURAD, André M. et al. Mass spectrometry characterisation of fatty acids from metabolically engineered soybean seeds. **Analytical and bioanalytical chemistry**, v. 406, n. 12, p. 2873-2883, 2014.

MURASHIGE, Toshio; SKOOG, Folke. A revised medium for rapid growth and bio assays with tobacco tissue cultures. **Physiologia plantarum**, v. 15, n. 3, p. 473-497, 1962.

MURPHY, D. J.; PIFFANELLI, P. Fatty acid desaturases: structure, mechanism and regulation.

**Plant lipid biosynthesis. Cambridge University Press, Cambridge**, p. 95-130, 1998.

NAPOLI, Carolyn; LEMIEUX, Christine; JORGENSEN, Richard. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. **The plant cell**, v. 2, n. 4, p. 279-289, 1990.

NASCIMENTO, Jaqueline A. et al. Ethanolics extracts of Moringa. **Journal of Thermal Analysis and Calorimetry**, v. 117, n. 1, p. 265-268, 2014.

NEFF, Kevin L. et al. Mojo Hand, a TALEN design tool for genome editing applications. **BMC bioinformatics**, v. 14, n. 1, p. 1, 2013.

NEKRASOV, Vladimir et al. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. **Nature biotechnology**, v. 31, n. 8, p. 691-693, 2013.

NICOLIA, Alessandro et al. Targeted gene mutation in tetraploid potato through transient TALEN expression in protoplasts. **Journal of biotechnology**, v. 204, p. 17-24, 2015.

NISHIZAWA-YOKOI, Ayako et al. A defect in DNA ligase 4 enhances the frequency of TALEN-mediated targeted mutagenesis in rice. **Plant physiology**, p. pp. 01542.2015, 2015.

O'KEEFE, Barry R. et al. Engineering soya bean seeds as a scalable platform to produce cyanovirin-N, a non-ARV microbicide against HIV. **Plant biotechnology journal**, v. 13, n. 7, p. 884-892, 2015.

OKULEY, John et al. Arabidopsis FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. **The Plant Cell**, v. 6, n. 1, p. 147-158, 1994.

OLIVA, M. L. et al. Stability of fatty acid profile in soybean genotypes with modified seed oil composition. **Crop Science**, v. 46, n. 5, p. 2069-2075, 2006.

OSAKABE, Yuriko; OSAKABE, Keishi. Genome editing in higher plants. In: **Targeted Genome Editing Using Site-Specific Nucleases**. Springer Japan, p. 197-205, 2015.

PATIL, Kiran et al. Evolutionary programming as a platform for in silico metabolic engineering. **BMC bioinformatics**, v. 6, n. 1, p. 1, 2005.

PATTANAYAK, Vikram et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. **Nature biotechnology**, v. 31, n. 9, p. 839-843, 2013.

PAVLETICH, Nikola P.; PABO, Carl O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. **Science**, v. 252, n. 5007, p. 809-817, 1991.

PETOLINO, Joseph F. Genome editing in plants via designed zinc finger nucleases. **In Vitro Cellular & Developmental Biology-Plant**, v. 51, n. 1, p. 1-8, 2015.

PORTEUS, M.H. AND BALTIMORE, D. Chimeric nucleases stimulate gene targeting in human cells. **Science**, v. 300, p.763, 2003.

QI, Qungang et al. Metabolically engineered soybean seed with enhanced threonine levels: biochemical characterization and seed-specific expression of lysine-insensitive variants of aspartate kinases from the enteric bacterium *Xenorhabdus bovienii*. **Plant biotechnology journal**, v. 9, n. 2, p. 193-204, 2011.

QIAN, Junfeng; SHI, Haixian; YUN, Zhi. Preparation of biodiesel from *Jatropha curcas* L. oil produced by two-phase solvent extraction. **Bioresource technology**, v. 101, n. 18, p. 7025-7031, 2010.

QU, Jing et al. Development of marker-free transgenic *Jatropha* plants with increased levels of seed oleic acid. **Biotechnology for biofuels**, v. 5, n. 1, p. 10, 2012.

RAMIREZ, Cherie L. et al. Unexpected failure rates for modular assembly of engineered zinc fingers. **Nature methods**, v. 5, n. 5, p. 374-375, 2008.

RAN, F. Ann et al. Genome engineering using the CRISPR-Cas9 system. **Nature protocols**, v. 8, n. 11, p. 2281-2308, 2013.

RAN, F. Ann et al. In vivo genome editing using *Staphylococcus aureus* Cas9. **Nature**, v. 520, n. 7546, p. 186-191, 2015.

REYON, Deepak et al. FLASH assembly of TALENs for high-throughput genome editing. **Nature** 101

**biotechnology**, v. 30, n. 5, p. 460-465, 2012.

REYON, Deepak et al. Engineering Customized TALE Nucleases (TALENs) and TALE Transcription Factors by Fast Ligation-Based Automatable Solid-Phase High-Throughput (FLASH) Assembly. **Current protocols in molecular biology**, p. 12.16. 1-12.16. 18, 2013.

ROSE, David P.; CONNOLLY, Jeanne M. Omega-3 fatty acids as cancer chemopreventive agents. **Pharmacology & therapeutics**, v. 83, n. 3, p. 217-244, 1999.

ROULIN, Anne et al. The fate of duplicated genes in a polyploid plant genome. **The Plant Journal**, v. 73, n. 1, p. 143-153, 2013.

SCHLUETER, Jessica A. et al. The FAD2 gene family of soybean. **Crop Science**, v. 47, n. Supplement\_1, p. S-14-S-26, 2007.

SCHMID-BURGK, Jonathan L.; SCHMIDT, Tobias; HORNUNG, Veit. Ligation-independent cloning (LIC) assembly of TALEN genes. **Chromosomal Mutagenesis**, p. 161-169, 2015.

SCHMUTZ, Jeremy et al. Genome sequence of the palaeopolyploid soybean. **Nature**, v. 463, n. 7278, p. 178-183, 2010.

SCHORNACK, Sebastian et al. Engineering plant disease resistance based on TAL effectors. **Annual review of phytopathology**, v. 51, p. 383-406, 2013.

SCHUSTER, Ivan et al. Soybean gene flow in the Western Region of Paraná. **Pesquisa Agropecuária Brasileira**, v. 42, n. 4, p. 515-520, 2007.

SEMENOVA, Ekaterina et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. **Proceedings of the National Academy of Sciences**, v. 108, n. 25, p. 10098-10103, 2011.

SHAN, Qiwei et al. Targeted genome modification of crop plants using a CRISPR-Cas system. **Nature biotechnology**, v. 31, n. 8, p. 686-688, 2013.

SHAN, Qiwei et al. Creation of fragrant rice by targeted knockout of the OsBADH2 gene using TALEN technology. **Plant biotechnology journal**, v. 13, n. 6, p. 791-800, 2015.

SHEPHERD, Louise VT et al. Assessing the potential for unintended effects in genetically modified potatoes perturbed in metabolic and developmental processes. Targeted analysis of key nutrients and anti-nutrients. **Transgenic Research**, v. 15, n. 4, p. 409-425, 2006.

SILVA, Maria Sebastiana et al. Composição química e valor protéico do resíduo de soja em relação ao grão de soja. **Ciência e Tecnologia de Alimentos**, v. 26, n. 3, p. 571-576, 2006.

SPYCHALLA, James P. et al. Identification of an animal  $\omega$ -3 fatty acid desaturase by heterologous expression in Arabidopsis. **Proceedings of the National Academy of Sciences**, v. 94, n. 4, p. 1142-1147, 1997.

STELLA, Stefano; MONTOYA, Guillermo. The genome editing revolution: A CRISPR-Cas TALE off-target story. **Bioessays**, v. 38, n. S1, p. S4-S13, 2016.

SUGANO, Shigeo S. et al. CRISPR/Cas9 Mediated Targeted Mutagenesis in the Liverwort *Marchantia polymorpha* L. **Plant and Cell Physiology**, p. pcu014, 2014.

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

TAYLOR, David C. et al. Field testing of transgenic rapeseed cv. Hero transformed with a yeast sn-2 acyltransferase results in increased oil content, erucic acid content and seed yield. **Molecular Breeding**, v. 8, n. 4, p. 317-322, 2002.

TOVKACH, Andriy; ZEEVI, Vardit; TZFIRA, Tzvi. A toolbox and procedural notes for characterizing novel zinc finger nucleases for genome editing in plant cells. **The Plant Journal**, v. 57, n. 4, p. 747-757, 2009.

URNOV, Fyodor D. et al. Genome editing with engineered zinc finger nucleases. **Nature Reviews Genetics**, v. 11, n. 9, p. 636-646, 2010.

UPADHYAY, Santosh Kumar et al. RNA-guided genome editing for target gene mutations in wheat. **G3: Genes| Genomes| Genetics**, v. 3, n. 12, p. 2233-2238, 2013.

VAN DER OOST, John; BROUNS, Stan JJ. CRISPR sabotage. **Genome biology**, v. 16, n. 1, p. 1-3, 2015.

WAGNER, Nicholas et al. RNAi trigger fragment truncation attenuates soybean FAD2-1 transcript suppression and yields intermediate oil phenotypes. **Plant biotechnology journal**, v. 9, n. 7, p. 723-728, 2011.

WALLIS, James G.; BROWSE, John. Mutants of Arabidopsis reveal many roles for membrane lipids. **Progress in lipid research**, v. 41, n. 3, p. 254-278, 2002.

WALTZ, Emily. CRISPR-edited crops free to enter market, skip regulation. **Nature Biotechnology**, v.34, p.582, 2016.

WANG, Mugui et al. Gene editing by co-transformation of TALEN and chimeric RNA/DNA oligonucleotides on the rice OsEPSPS Gene and the inheritance of mutations. **PloS one**, v. 10, n. 4, p. e0122755, 2015.

WANG, Yanpeng et al. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. **Nature Biotechnology**, v. 32, n. 9, p. 947-951, 2014.

WENDT, Toni et al. TAL effector nucleases induce mutations at a pre-selected location in the genome of primary barley transformants. **Plant molecular biology**, v. 83, n. 3, p. 279-285, 2013.

WIEDENHEFT, Blake; STERNBERG, Samuel H.; DOUDNA, Jennifer A. RNA-guided genetic silencing systems in bacteria and archaea. **Nature**, v. 482, n. 7385, p. 331-338, 2012.

WILSON, Richard F. The role of genomics and biotechnology in achieving global food security for high-oleic vegetable oil. **Journal of oleo science**, v. 61, n. 7, p. 357-367, 2012.

WINGENDER, Edgar et al. TRANSFAC: an integrated system for gene expression regulation.

**Nucleic acids research**, v. 28, n. 1, p. 316-319, 2000.

XIE, Kabin; YANG, Yinong. RNA-guided genome editing in plants using a CRISPR–Cas system. **Molecular plant**, v. 6, n. 6, p. 1975-1983, 2013.

XIE, Kabin; MINKENBERG, Bastian; YANG, Yinong. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. **Proceedings of the National Academy of Sciences**, v. 112, n. 11, p. 3570-3575, 2015.

YADAV, N. S. Genetic modification of soybean oil quality. **Soybean: genetics, molecular biology and biotechnology**, 1996.

ZHANG, Yong et al. Transcription activator-like effector nucleases enable efficient plant genome engineering. **Plant physiology**, v. 161, n. 1, p. 20-27, 2013.

ZHANG, Hui et al. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. **Plant biotechnology journal**, v. 12, n. 6, p. 797-807, 2014.

ZHANG, Hui et al. TALEN-mediated targeted mutagenesis produces a large variety of heritable mutations in rice. **Plant biotechnology journal**, v. 14, n. 1, p. 186-194, 2015.

ZHOU, Huanbin et al. Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. **Nucleic acids research**, v. 42, n. 17, p. 10903-10914, 2014.

ZISCHEWSKI, Julia; FISCHER, Rainer; BORTESI, Luisa. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. **Biotechnology Advances**, 2016.