



UnB

Pós-Graduação

Biologia Molecular



Embrapa

*Recursos Genéticos e
Biotecnologia*

**REDUÇÃO DO TEOR DE ÁCIDO FÍTICO EM SEMENTES DE SOJA MEDIADO
POR CRISPR E RNA INTERFERENTE**

Aluna: **Jéssica Carrijo de Souza**

Orientador: **Prof. Dr. Francisco JL Aragão**

Coorientador: **Dr. Giovanni R Vianna**

Brasília-DF
2021

JÉSSICA CARRIJO DE SOUZA

**REDUÇÃO DO TEOR DE ÁCIDO FÍTICO EM SEMENTES DE SOJA MEDIADO
POR CRISPR E RNA INTERFERENTE**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Biologia Molecular do Instituto de Ciências Biológicas da Universidade de Brasília, como requisito parcial à obtenção do título de Doutor em Biologia molecular.

Orientador: **Prof. Dr. Francisco José Lima Aragão**

Coorientador: **Dr. Giovanni Rodrigues Vianna**

Brasília-DF
2021

Tese de autoria de Jéssica Carrijo de Souza, intitulada “REDUÇÃO DO TEOR DE ÁCIDO FÍTICO EM SEMENTES DE SOJA MEDIADO POR CRISPR E RNA INTERFERENTE”, apresentada como requisito parcial para obtenção do grau de Doutor em Biologia Molecular da Universidade de Brasília, em 14 de maio de 2021, defendida e aprovada pela banca examinadora abaixo assinada:

Prof. Dr. Francisco José Lima Aragão
Orientador
(Embrapa Recursos Genéticos e Biotecnologia - Cenargen)

Dr. Giovanni Rodrigues Vianna
Coorientador
(Embrapa Recursos Genéticos e Biotecnologia - Cenargen)

Prof. Dr. Robert Miller
Membro interno
(Programa de Pós Graduação em Biologia Molecular- UnB)

Dr. Josias Correa de Faria
Membro externo
(Embrapa Arroz e Feijão - CNPAF)

Dra. Rosana Pereira Vianello
Membro externo
(Embrapa Arroz e Feijão - CNPAF)

Prof. Dr. Tatsuya Nagata
Suplente
(Programa de Pós Graduação em Biologia Molecular- UnB)

Brasília
2021

Dedico este trabalho aos meus grandes
exemplos, papai e mamãe! ♥

AGRADECIMENTOS

Nessa jornada intensa e emocionante, não podia deixar de agradecer aos que de alguma forma esteve comigo nessa caminhada.

Ao meu *Bom Deus* por abençoar essa jornada e por me ajudar a ter discernimento e tranquilidade para concluir cada passo. Muitas noites recorridas a Ele, buscando o conforto para o meu coração, com a esperança de que tudo daria certo.

Aos meus pais *Adilmar* e *Sônia*, e meus irmãos *Viviane* e *Júnior* que mesmo sem entender muito das alegrias e frustrações do dia a dia de um laboratório, estiveram sempre ao meu lado, dando palavras de conforto e me acolhendo com os abraços. Além também da minha *Dinda*, por ouvir tantos desabafos e também comemorar com as minhas pequenas conquistas no dia a dia.

Ao meu melhor amigo e coorientador *Giovanni Vianna* pela companhia diária, pela ajuda nos experimentos e por ser meu melhor amigo nessa longa caminhada. Obrigada por todas as risadas e broncas que me permitiu dar os passos certos no doutorado como também ajudou a me tornar um ser humano melhor. Minha eterna admiração e gratidão.

Ao Dr. *Francisco Aragão* por aceitar ser meu orientador e esteve sempre disponível para ajudar.

Aos meus queridos: *Julio Carlyle*, *Nicolau Brito* e *Pedro Berbert* que direta e indiretamente sempre me deram forças para continuar, me incentivando e mostrando que sou capaz.

As minhas amigas, que mesmo longe sempre tiveram uma palavra de incentivo: *Andressa Rezende*, *Debora Torres*, *Jéssica Sousa*, *Marina Silva*, *Shara Vasconcelos* e *Stephany Brito*.

Aos colegas de laboratório em especial a *Kenny Bonfim*, *Elsa Nogueira*, *Natalia Sousa*, *Lídia Queiroz*, *Elinea* e *Carolzinha*. Sem deixar de falar da minha adorável estagiária *Nathalia Aline* que apesar das raivas que me passava, foi minha companheira em boa parte do doutorado.

Aos amigos que fiz durante o período do “doutorado sanduíche” na Universidade da Georgia, *Mónica Prias*, *Peter LaFayette*, *Soraya Bertioli*, *Tim Chappell*, *Gary Orr*, *Sueme Ueno* e *Renan Souza*. Minha estada foi tranquila por conta de vocês. E ao Dr. *Wayne* por me receber no seu lab.

Agradeço a disponibilidade dos membros da banca: Prof. *Robert Miller*, Dr. *Josias Farias*, Dra. *Rosana Vianello* e Prof. *Tatsuya Nagata* que aceitaram avaliar e contribuir com o meu trabalho.

A Todos que de alguma forma contribuíram para o desenvolvimento deste trabalho, incluído as agencias financiadoras **CAPES**, **CNPq**, **Embrapa** e **UnB**.

Meus sinceros agradecimentos!!

Nós somos o que fazemos repetidas vezes. A excelência, então, não é um ato, mas um hábito.

Will Durant

RESUMO

CARRIJO, Jéssica. **Redução do teor de ácido fítico em sementes de soja mediado por CRISPR e RNA interferente**. 2021. 98 folhas. Doutorado em Biologia Molecular- Universidade de Brasília, 2021.

A maior parte do fósforo encontrado nas sementes de soja está acumulada na forma de ácido fítico, representando mais de 85% do fósforo total. Sua biossíntese ocorre através da via do inositol-fosfato. Quando esta molécula está na forma de sal, fitato, é considerado um fator antinutricional, pois limita a disponibilidade de alguns nutrientes essenciais, diminuindo o valor nutricional das sementes. Além disso, o ácido fítico não é digerível pelos monogástricos, sendo amplamente excretado em vez de absorvido, podendo resultar no acúmulo de fósforo nos solos, causando também poluição das águas. De modo que o objetivo deste trabalho foi desenvolver uma linhagem de soja com reduzido teor de ácido fítico, por meio da engenharia genética. Para isso, foram utilizadas duas metodologias, CRISPR e RNA interferente. Inicialmente duas configurações do sistema CRISPR foram avaliadas por *hairy-root*. (1) Um cassete simplificado (STU) onde tanto a Cas9 quanto os RNA guias (sgRNA) são controlados por um mesmo promotor, *versus* (2) um cassete tradicional (TCTU), mais comumente utilizado, onde a Cas9 está sob a regulação de um promotor e cada sgRNA por outro promotor (TCTU). O sistema STU aqui utilizado, permitiu uma redução de mais de 1.000pb no tamanho dos vetores de transformação. Foram escolhidos dois genes que atuam ao final da via de biossíntese do ácido fítico, *GmIPK1* e *GmIPK2*. Os resultados indicaram que o sistema tradicional se apresentou mais eficiente em gerar edições no genoma das raízes de soja, embora os dois sistemas tenham demonstrado funcionalidade. O cassete TCTU foi então escolhido para gerar os transformantes definitivos de soja via sistema biobalístico de transformação. Plantas com edição em ambos os genes (*GmIPK1* e *GmIPK2*) foram geradas individualmente, e o conteúdo de ácido fítico foram avaliados, destacando assim 3 eventos: IPK1-5D2 (11,56%) IPK2-24B2 (14,11%) IPK2-24C4 (25,35%) que apresentaram uma aparente redução de ácido fítico sem alteração nas características agrônômicas. Também foram geradas plantas de soja transformadas utilizando a técnica de RNA interferente para redução da expressão do gene *GmIPK2*. Análises da quantidade de ácido fítico foram avaliadas nas plantas geradas, destacando 3 eventos IPK2-RNAi_1A1 (23,96%), IPK2-RNAi_1D2 (41,67%) e IPK2-RNAi_7C2 (13,71%) com aparente redução. Foi observado alterações no fenótipo da linhagem 1D2 com retardo no desenvolvimento e baixa produtividade. No entanto a linhagem 1A1 se mostrou um potencial candidato, visto ter apresentado redução de ácido fítico sem alterações fenotípicas. A utilização de ambas as metodologias (CRISPR e RNAi), foram eficazes na geração de plantas de soja com redução no conteúdo de ácido fítico em sementes. No entanto novos experimentos ainda são necessários para compreender a natureza das mutações para os eventos CRISPRs além de avaliar a biodisponibilidade de nutrientes nas sementes após a redução do ácido fítico em todos os eventos em destaque gerados nesse trabalho.

Palavras-chave: Edição de genes; Soja transgênica; *Glycine max*; Engenharia genética de plantas; *GmIPK1* e *GmIPK2*.

ABSTRACT

CARRIJO, Jessica. **Reduction of phytic acid content in soybean seeds by CRISPR and RNA interference**. 2021. 98 pages. PhD in Molecular Biology - University of Brasília, 2021.

Most of the phosphorus found in soybean seeds is accumulated in the form of phytic acid, representing more than 85% of the total phosphorus. Its biosynthesis occurs by the inositol-phosphate pathway. When this molecule is in the salt form, phytate, it is considered an anti-nutritional factor, as it limits the availability of some essential nutrients and decreases the nutritional value of the seeds. In addition, phytic acid is not digestible by monogastrics, being largely excreted rather than absorbed, which can result in the accumulation of phosphorus in soils, also causing water pollution. So, the objective of this work was to develop a soybean event with reduced phytic acid content, through genetic engineering. CRISPR and RNA interference were used. Initially, two configurations of the CRISPR system were evaluated by hairy-root. (1) A simplified cassette (STU) where both Cas9 and guide RNA (sgRNA) are controlled by the same promoter *versus* (2) a traditional cassette (TCTU) traditional cassette, most commonly used, where Cas9 is under the regulation of a promoter and each sgRNA by another promoter. The STU system allowed a reduction of more than 1,000bp in the size of the transformation vectors. Two target genes that act at the end of the phytic acid biosynthesis pathway, namely *GmIPK1* and *GmIPK2* were chosen. The results indicated that the traditional system was more efficient in generating genome edits in soybean roots, although both systems demonstrated functionality. The TCTU cassette was then chosen to generate the definitive soybean events using a biolistic transformation system. Plants edited in both genes (*GmIPK1* and *GmIPK2*) were generated individually, and the phytic acid content was evaluated, thus prominent três events: IPK1-5D2 (11.56%), IPK2-24B2 (14.11%) and IPK2- 24C4 (25.35%) that showed an apparent reduction of phytic acid without change in agronomic characteristics. Transformed soybean plants were also generated using the RNA interference technique to reduce the expression of the *GmIPK2* gene. Analyzes of the amount of phytic acid were evaluated in the generated plants prominent três events IPK2-RNAi_1A1 (23.96%) IPK2-RNAi_1D2 (41.67%) and IPK2-RNAi_7C2 (13.71%) with apparent reduction. Changes in the phenotype of the 1D2 strain were observed with developmental delay and low productivity. However, the 1A1 strain proved to be a potential candidate, as it showed a reduction in phytic acid without phenotypic changes. The use of both methodologies (CRISPR and RNAi) was effective in generating soybean plants with reduced phytic acid content in seeds. However, new experiments are still needed to understand the nature of mutations for CRISPRs events, in addition to evaluating the bioavailability of nutrients in seeds after phytic acid reduction in all prominent events generated in this work.

Keywords: Gene editing; Transgenic soybeans; *Glycine max*; Genetic engineering of plants; *GmIPK1* and *GmIPK2*.

LISTA DE ILUSTRAÇÕES

Figura 1: Representação esquemática da via de biossíntese de ácido fítico, indicando o ponto de atuação das enzimas IPK1 e IPK2 ao final da via, (adaptado de Sashidhar <i>et al.</i> 2020) (A) e Estrutura química da molécula do ácido fítico (B)	15
Figura 2: Esquematização da montagem dos vetores.....	26
Figura 3: Esquematização do protocolo de <i>hairy-root</i>	27
Figura 4: Alinhamento das três cópias do gene <i>GmIPK1</i> presentes no genoma da soja Willians 82.....	30
Figura 5: Esquema dos dois sistemas de expressão CRISPR/Cas9: Sistema TCTU e Sistema STU.....	30
Figura 6: Detecção dos fragmentos dos genes <i>GmIPK1</i> e <i>GmIPK2</i> por PCR em gel de agarose sob luz UV.....	31
Figura 7: Alinhamento das sequências obtidas pelo sequenciamento Sanger (Clustal) da região alvo dos genes <i>GmIPK1</i> e <i>GmIPK2</i> editadas	33
Figura 8: Comparação dos sistemas de edição STU (single transcriptional unit) e TCTU (two-component transcriptional unit) do gene <i>GmIPK1</i> em raízes de soja por CRIPSR/Cas9	35
Figura 9: Comparação dos sistemas de edição STU (single transcriptional unit) e TCTU (two-component transcriptional unit) do gene <i>GmIPK2</i> em raízes de soja por CRIPSR/Cas9.....	37
Figura 10: Esquema do vetor CRISPR/Cas9 para biobalística, sistema TCTU.....	50
Figura 11: Representação gráfica dos resultados obtidos após análise dos sequenciamentos das amostras de DNA-genômico das plantas transformadas com vetores CRISPR-Cas9-TCTU.....	51

Figura 12: Esquema do vetor pAHAS-GUS-RNAI_IPK2.....	68
Figura 13: Ensaio histoquímico para detecção do gene <i>GUS</i> nas linhagens T ₀ plantas transformadas com o vetor pAHAS-GUS-RNAI-IPK2.	68
Figura 14: Fenótipos alterados observados nas linhagens de soja IPK2-RNAi #1 progênies A (1, 2 e 3) e D (2 e 3).....	70
Figura 15: Expressão normalizada do gene <i>GmIPK2</i> em sementes e folhas de sojas geneticamente modificadas mediada por qRT-PCR.....	73

LISTA DE TABELAS

Tabela 1. Características dos sgRNAs selecionados para GmIPK1 e GmIPK2.....	25
Tabela 2: <i>Primers</i> usados para detecção das edições mediadas por CRISPR/Cas9.....	28
Tabela 3: Trabalhos da literatura que visaram à obtenção de plantas lpa	45
Tabela 4: Conteúdo de ácido fítico nas sementes das plantas IPK1 e IPK2 CRISPR T ₂	52
Tabela 5: Conteúdo de ácido fítico nas sementes das plantas IPK1-CRISPR T ₃	54
Tabela 6: Conteúdo de ácido fítico nas sementes das plantas IPK2-CRISPR T ₃	55
Tabela 7: Conteúdo de ácido fítico nas sementes das plantas destacadas IPK2-CRISPR T ₃	56
Tabela 8: Conteúdo de ácido fítico nas sementes das plantas IPK2-RNAi T ₂	69
Tabela 9: Conteúdo de ácido fítico nas sementes das plantas IPK2-RNAi T ₃	71
Tabela 10: Conteúdo de ácido fítico nas sementes destacadas IPK2-RNAi T ₃	72

SUMÁRIO

INTRODUÇÃO – Uma visão geral	13
CAPÍTULO I: Dois sistemas eficientes de CRISPR/Cas9 para edição de genes em soja.....	19
1.1 INTRODUÇÃO - CRISPR/Cas como uma ferramenta biotecnológica	20
1.2 OBJETIVO GERAL	24
Objetivos específicos	24
1.3 METODOLOGIA	25
Escolha das regiões alvo.....	25
Obtenção de raízes transgênicas pelo sistema <i>hairy-root</i>	27
Detecção de edição de genes mediada por CRISPR / Cas9	28
1.4 RESULTADOS	29
Design dos alvos e construção dos vetores CRISPR/Cas	29
Obtenção de raízes de soja transgênicas	31
Análises das sequências editadas via sequenciamento tipo Sanger.....	31
Análises das edições e avaliação da eficiência via sequenciamento tipo Illumina	34
1.5 DISCUSSÃO	38
1.6 CONSIDERAÇÕES	41
CAPÍTULO II: Plantas de soja editadas via CRISPR/SpCas9.....	42
2.1 INTRODUÇÃO - Um gene nocauteado por CRISPR	43
2.2 OBJETIVO GERAL	47
Objetivos específicos	47
2.3 METODOLOGIA	48
Construção do vetor para transformação de soja por biobalística.....	48
Confirmação das plantas transgênicas e nocaute dos genes alvos	48
Determinação dos teores de ácido fítico nas sementes de soja	49
2.4 RESULTADOS	50
Construção do vetor para biobalística e transformação genética.....	50

Análises do sequenciamento das plantas T ₀ obtidas.....	50
Análise do conteúdo de ácido fítico.....	51
2.5 DISCUSSÃO.....	57
2.5 CONSIDERAÇÕES.....	61
CAPÍTULO III: Silenciamento do gene <i>GmIPK2</i> em soja mediado por RNA interferente	62
3.1 INTRODUÇÃO – Regulação gênica mediada por RNAi.....	63
3.2 OBJETIVO GERAL.....	65
Objetivos específicos	65
3.3 METODOLOGIA	66
Construção do vetor de RNAi para silenciamento do gene <i>IPK2</i>	66
Transformação de plantas de soja via biobalística.....	66
Detecção das plantas transgênicas.....	66
Determinação dos teores de ácido fítico nas sementes de soja	67
Quantificação da transcrição do gene por qRT-PCR.....	67
3.4 RESULTADOS.....	68
Construção do vetor pAHAS-GUS-RNAi_IPK2 para biobalística.....	68
Obtenção das plantas transgênicas.....	68
Análise do conteúdo de ácido fítico.....	69
Análise quantitativa da transcrição do gene <i>GmIPK2</i>	72
3.5 DISCUSSÃO.....	74
CONCLUSÃO FINAL E PERSPECTIVAS.....	77
REFERÊNCIAS BIBLIOGRÁFICAS	78
ANEXO I: <i>Two efficient CRISPR/Cas9 systems for gene editing in soybean</i>.....	86

INTRODUÇÃO – Uma visão geral

A soja (*Glycine max* (L.) Merrill.) é considerada uma das principais *commodities* do mundo, com produção mundial de 337,298 milhões de toneladas na safra 2019/2020 (FAOSTAT, 2019). Neste mesmo período o Brasil alcançou o *status* de maior produtor mundial, com uma produção superior a 130 milhões de toneladas. Desse total o Brasil exportou aproximadamente 74 milhões de toneladas, se tornando o maior produtor e exportador de soja do mundo (CONAB, 2020).

O aumento considerável da utilização dessa cultura nas últimas décadas está relacionado estritamente a ser uma fonte riquíssima de nutrientes, devido a sua composição, com alto teor de proteínas (36%) e óleos (18%). Um percentual de 87% da produção total mundial é destinado à alimentação, produção de farelos e óleos, enquanto que os 13% restante está relacionando a produção direta (ex.: leite e tofu) e alimentação humana, que ainda está restrita a alguns países (países Orientais) e culturas alimentares (vegetarianos) (Thrane *et al.*, 2017).

Do total de farelo produzido, 98% está destinado a alimentação animal. O uso da soja para essa finalidade é devido sua composição química única que a faz ser o grão com o maior conteúdo de proteínas, além de diversos outros componentes valiosos como: polissacarídeos, vitaminas, minerais entre outros (Thrane *et al.*, 2017). Apesar das suas propriedades funcionais, o grão de soja apresenta uma grande quantidade de componentes antinutricionais, incluindo as saponinas, isoflavonas, lecitinas, inibidores de proteases e também o ácido fítico (Punjabi *et al.*, 2018A).

Esse último, o ácido fítico (InsP6 ou 1,2,3,4,5,6-hexaquis[dihidrogenofosfato]mio- inositol- Figura 1B) também conhecido como fitato, quando na forma de sal, é uma das principais formas de acúmulo (estoque) de fósforo nas plantas, representando mais de 85% do fósforo total nas sementes de soja (Sparvoli & Cominelli, 2015). Devido a sua estrutura química (altamente negativo em pH fisiológico) o ácido fítico é facilmente precipitado em sal de fitato (ou apenas fitato), esse por sua vez se liga a importantes minerais cátions divalentes, dentre os principais estão: ferro (Fe^{2+}), zinco (Zn^{2+}), cálcio (Ca^{2+}), magnésio (Mg^{2+}), também são capazes de interagir com proteínas (Goßner *et al.*, 2019) tornando esses

nutrientes/proteínas indisponível para serem absorvidos pelo trato gastrointestinal dos animais que se alimentam do farelo. Essa característica faz com que o fitato seja considerado um fator antinutricional, o que contribui para a diminuição do valor nutricional da semente (Vincent *et al.*, 2015).

Além disso, a molécula do fitato não é processada pelos animais monogástricos tais como suínos, equinos, aves, peixes e humanos, uma vez que esses seres vivos não são capazes de digerir por completo essa molécula. Essa então é excretada nas fezes em vez de absorvida. Esta questão pode ser agravada, uma vez que, para atender as exigências nutricionais na alimentação de não ruminantes, há a suplementação da ração com fósforo (P), o que aumenta a concentração desse elemento no esterco produzido, levando ao acúmulo de fósforos nos solos e ao conseqüente risco de poluição da água (eutrofização) pela alta concentração de P liberado sempre nas mesmas áreas (Kumar *et al.*, 2019).

O fósforo é um dos principais macronutrientes das plantas, uma vez que é responsável pelo desenvolvimento e crescimento, sendo também parte essencial de diversos processos, tais como geração e transferência de energia, metabolismo de carbono, síntese de membrana, ativação de enzima, fixação de nitrogênio, constituinte de ácidos nucléicos, fosfolipídios e adenosina trifosfato (ATP), transdução e regulação de sinais, reparo e recombinação de DNA entre outros (Malhotra *et al.*, 2018). Fazendo o um componente vital, sendo esse em sua maioria estocado na forma de ácido fítico.

A biossíntese do ácido fítico (Figura 1) ocorre em duas rotas na via do inositol-fosfato, uma conhecida como lipídeo-dependente, que opera em todos os tecidos das plantas, e outra, lipídeo-independente, que é predominante em sementes. A síntese se inicia com a produção de *myo-inositol* (referido como Ins) através da reação enzimática da *D-myoinositol 3-phosphate synthase (MIPS)* que converte D-glucose-6 em *myo-inositol 3-phosphate (Ins(3)P1)*. A partir dessa fase, diversos InsPs (InsP2, InsP3, InsP4, InsP5 e InsP6) são formados através da adição de fósforo por enzimas da via.

O que distingue a rota lipídeo-dependente da rota lipídeo-independente é a forma como o InsP3 é gerado. Na rota lipídeo-dependente, como o próprio nome sugere, é necessário primeiro a síntese de lipídeos de inositol, para depois ser gerado o Ins (1,4,5)P3 que então pela ação da enzima *Inositol multiphosphate kinase 2 (IPK2)* gera InsP4, seguido

pelo InsP5. Enquanto na via lipídeo-independente esse processo é direto e regulado pela enzima *ITPK* (*Inositol 1, 3, 4 trisphosphate 5/6 kinase*). Ao final do processo em ambas as rotas é gerado a molécula Ins (1,3,4,5,6)P5 que pela ação da *Inositol pentakisphosphate 2-kinase* (*IPK1*), o ácido fítico (Ins(1,2,3,4,5,6)P6) é sintetizado. Uma vez sintetizado as moléculas são transportadas para dentro dos vacúolos por transportadores *MRP* (*Multidrug-resistance-associated protein*) pertencente à família ABC (*ATP-binding cassette*) e são armazenadas dentro de globoides que estão nos corpos proteicos em sementes maduras (Sparvoli & Cominelli, 2015).

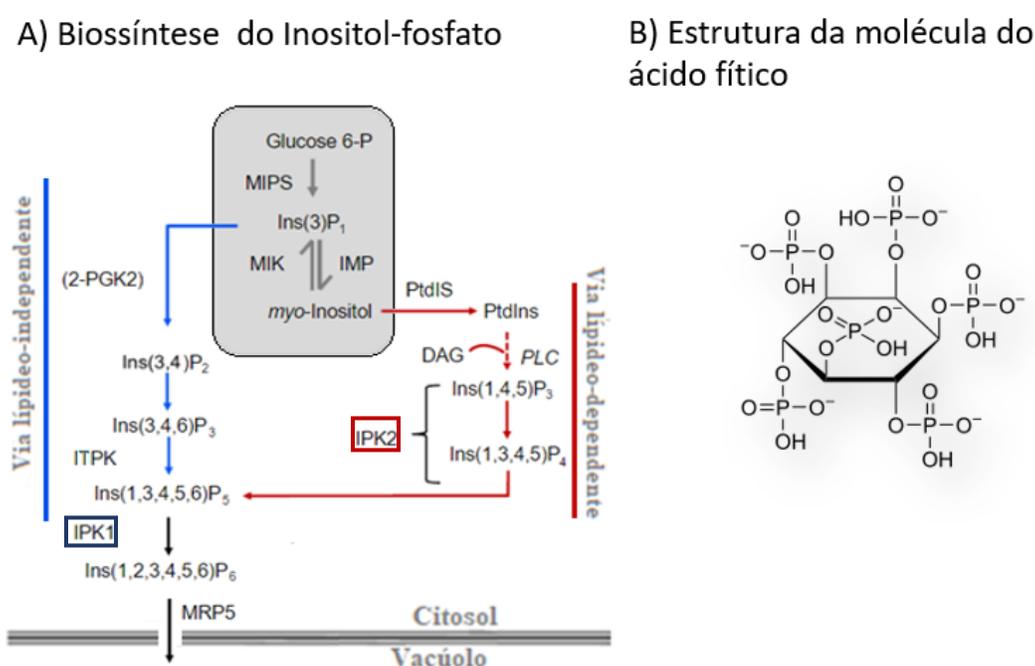


Figura 1: Representação esquemática da via de biossíntese de ácido fítico, indicando o ponto de atuação das enzimas IPK1 e IPK2 no final da via (Adaptado de Sashidhar *et al.* 2020) (A) e Estrutura química da molécula do ácido fítico (B). São apresentados a via de manutenção para outras vias dependentes do Myo-Inositol (box cinza) com as enzimas MIPS: *myo-inositol phosphate synthase*, MIK: *myo-inositol kinase*, IMP: *myo-inositol monophosphatase*, 2-PGK2: *2-phosphoglyceric acid kinase*. A via lipídeo independente (setas azuis) sendo regulado pela enzima ITPK: *inositol tetrakisphosphate kinase*. Na via lipídeo dependente (setas vermelhas) ocorre inicialmente a síntese de lipídeos de Inositol com a atuação das enzimas PtdIS: *phosphatidyl inositol synthase*, DAG: *diacyl glycerol*, PLC: *phospholipase C.*, seguida ao final pela quinase IPK2: *Inositol multiphosphate kinase*. O produto resultante de ambas as vias (InsP5) sofre a ação da enzima IPK1: *inositol pentakisphosphate 2-kinase* o que resulta na geração do InsP6 ou ácido fítico. O Ácido fítico é enviado para dentro do vacúolo através do transportador MRP5: *multidrug resistance protein*.

A via do inositol-fosfato é responsável por diversos processos biológicos, pois os InsPs são moléculas sinalizadoras em processos celulares como: homeostase de fósforo, resposta hormonal, resposta a estresses abióticos e à patógenos (Adepoju *et al.*, 2017). No

entanto, devido à presença em excesso desse fator antinutricional, que reduz o valor nutricional dos grãos, há um grande interesse no desenvolvimento de estratégias para redução do conteúdo de ácido fítico, sem afetar o desenvolvimento da planta como um todo. Acredita-se que uma redução de pelo menos 4% no nível de fitato poderia melhorar o valor econômico e nutricional dessa cultura (Marathe *et al.*, 2018).

Já foram realizadas tentativas de obtenção de soja com redução nos teores de ácido fítico pelo melhoramento clássico (Wilcox *et al.*, 2000; Hitz *et al.*, 2002). No entanto, esse caminho mostrou-se limitado, devido à falta de variação natural nos níveis de fitato no germoplasma de soja (Marathe *et al.*, 2018). Nesse sentido, o desenvolvimento de novas ferramentas biotecnológicas tais como RNA interferente, Zinc-finger (ZNFs) e CRISPR surgiram como uma possibilidade de obtenção de um organismo geneticamente modificado com as características desejadas sem alterar as características agrônômicas importantes, visto que essas técnicas possibilitam modificações genéticas precisas e rápidas (Limerá *et al.*, 2017), e já foram usadas de forma eficiente para geração de plantas com alta produtividade agrícola e resistência a estresses bióticos e abióticos (Mohanta *et al.*, 2017).

Devido à grande importância que a via do inositol-fosfato representa a germinação e desenvolvimento inicial das plantas, é interessante que se encontre um equilíbrio entre mutantes que apresentem baixo teor de ácido fítico (*low phytic acid - lpa*), mas que tenha efeitos pleiotrópicos reduzidos (Sparvoli & Cominelli, 2015). Assim faz-se necessário o estudo dos genes componentes dessa via, de modo a identificar aqueles que podem ter sua expressão alterada, e gerar uma variedade *lpa* sem prejuízo da planta. Como apenas as moléculas InsP5 e InsP6 têm efeitos negativos na biodisponibilidade de minerais, e que o nocaute ou silenciamento de um gene que atue ao final da via, pode diminuir os níveis de ácido fítico sem redução dos outros intermediários (Basak *et al.*, 2017), dois genes foram selecionados como objetos de estudo desse projeto. São eles os genes *IPK1* e *IPK2*.

O gene *IPK1* codifica para a enzima *Inositol-pentakisphosphate-2-kinase*, enzima que catalisa a fosforilação do *myo-inositol 1,3,4,5,6 pentacisfosfato* em ácido fítico, reação final da biossíntese do ácido fítico (Martynov *et al.*, 2014). Possui três genes ortólogos no genoma da soja, nos cromossomos 4, 6 e 14. O gene presente no cromossomo 14 tem uma alta expressão em sementes, 20 dias após a floração, enquanto as cópias dos cromossomos

4 e 6, apresentam expressão apenas parcial (Yuan *et al.*, 2012). As três cópias apresentam diferenças em nível de DNA, nas sequências não codantes (introns), porém geram transcritos similares e codifica peptídeos de similar tamanho (Yuan *et al.*, 2012). De acordo com Basak *et al.* (2017) há expressão do gene *GmIPK1* em todos os tecidos, no entanto em sementes o nível de expressão pode chegar até oito vezes mais que em raízes, que é o tecido que apresenta menor expressão, sendo o pico de expressão ainda nas sementes imaturas quando estas alcançam tamanho entre 6-8mm.

Por outro lado, o gene *IPK2* (*Inositol 1,4,5-tris-phosphatekinase*) possui apenas uma cópia presente no cromossomo 12, codifica para uma das quinases responsáveis por converter InsP3 em InsP4, e InsP4 em InsP5 na via lipídio dependente da biossíntese do ácido fítico. Embora não seja a rota principal de produção do ácido fítico, apresenta uma expressão temporal no final do ciclo de enchimento das vagens (Punjabi *et al.*, 2018A). Punjabi *et al.* (2018B) relataram que o nível de expressão do gene é maior quando o estágio de desenvolvimento da semente está entre 8 e 10mm (tamanho da semente), A expressão do gene *GmIPK2* é naturalmente baixa em todos os tecidos, mas quando são comparados os tecidos, a expressão de *GmIPK2* é maior em sementes. Esse gene também está envolvido em outros processos fisiológicos, como desenvolvimento da planta e resposta a estresses (Li *et al.*, 2017).

Alta redução no teor de ácido fítico pode gerar fenótipos como alteração do crescimento das plantas, no desenvolvimento e germinação das sementes, e na resposta a estresses abiótico e biótico (Raboy *et al.*, 2009; Kumar *et al.*, 2019). Mas de todo modo é esperado que a redução na expressão dos genes *IPK1* e *IPK2* causem uma diminuição do acúmulo de ácido fítico acompanhado por um pequeno aumento no teor de fósforo livre e aumento do conteúdo de InsPs que possuam até cinco resíduos de fósforo (Xu *et al.*, 2005; Sparvolli & Cominelli, 2015).

Essa tese tem como objetivo final gerar um evento de soja com reduzido teor de ácido fítico, a partir de uma cultivar comercial BRS7080, com o auxílio de duas ferramentas biotecnológicas, CRISPR e RNA interferente. Para uma melhor organização do manuscrito, este foi dividido em capítulos. O capítulo I refere-se à validação da melhor configuração vetorial do sistema CRISPR/Cas9. Seguido pelo capítulo II que apresenta o

estudo com as plantas de soja editadas para gerar uma linhagem com redução de ácido fítico. Finalizando com o capítulo III onde utilizando a técnica de RNA interferente o gene *GmIPK2* foi silenciado, resultando em plantas de soja com reduzido teor de ácido fítico.

CAPÍTULO I:

Dois sistemas eficientes de CRISPR/Cas9 para edição de genes em soja

1.1 INTRODUÇÃO - CRISPR/Cas como uma ferramenta biotecnológica

Ao longo de muitos anos, várias estratégias foram desenvolvidas e se mostraram de grande importância para o avanço na área de edição de genoma em plantas, como o uso das meganucleases (Adli, 2018) Zinc-finger nucleases (ZFNs) (Davies *et al.*, 2017) e TALENs (*Transcription Activator-Like Effector Nucleases*) (Du *et al.*, 2016). No entanto, algumas limitações foram apresentadas para o uso dessas tecnologias, tanto relacionado à baixa eficiência de edição; quanto à necessidade de vários passos para montagens dos vetores, bem como os efeitos adversos das mutações em locais inespecíficos ou *off-targets* (Gaj *et al.*, 2013; Adli, 2018). Desse modo, a tecnologia mais recentemente de edição de genomas, CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*), apresentou-se como a mais promissora, por ser mais simples quando comparada as demais técnicas e com especificidade maior no reconhecimento do local a ser editado. (Pellagatti *et al.*, 2015; Wolter *et al.*, 2019).

O sistema CRISPR é originalmente parte do sistema imune dos procarionotos, protegendo-os através da clivagem de DNAs dos organismos invasores guiados por um RNA (Wiedenheft *et al.*, 2012). O sistema CRISPR/Cas se vale de proteínas denominadas Cas (*CRISPR associated*), normalmente com dois domínios catalíticos responsáveis pela clivagem da dupla fita, o HNH (cliva a fita-alvo, complementar ao guia) e RuvC (cliva a outra fita); além de um complexo sgRNA (*Single Guide RNA*), que é a fusão de um crRNA maduro (*CRISPR-derived RNA*) e um tracrRNA (*trans-activating RNA*), resultando em uma estrutura simplificada e funcional para a ativação e direcionamento do processo (Shan *et al.*, 2013; Barrangou, 2015).

Uma característica primordial do mecanismo de CRISPR/Cas é a exigência de uma sequência, normalmente de 2 a 6 nucleotídeos, chamada *Protospacer-Adjacent Motifs* (PAM), localizada de três a quatro nucleotídeos de distância da região de clivagem do alvo no sentido 5'-3' (Ghogare *et al.*, 2020). Anterior ao PAM está a sequência alvo, normalmente escolhida com tamanho aproximado de 20 pares de bases, que determina o local de ancoragem do sistema e da clivagem da fita dupla (DSB *Double-Strand Break*) gerando em sua maioria extremidades *blunt-end* nas duas fitas (Jinek *et al.*, 2012).

Diversas enzimas Cas foram descritas ao longo dos anos a partir de diferentes microrganismos (Nakade *et al.*, 2017; Perez-Rojo *et al.*, 2018), e a classificação é baseada na estrutura da endonuclease (Makarova *et al.*, 2011; Ghogare *et al.*, 2020). Atualmente a endonuclease do tipo II conhecida como Cas9, identificada de *Streptococcus pyogenes*, é mais amplamente utilizada, por possuir uma região PAM simples, identificada como NGG (Jinek *et al.*, 2012).

A utilização do sistema CRISPR/Cas assim como das demais tecnologias de edição de genoma, só foi possível devido ao maior entendimento do sistema de reparo natural de DNA dos organismos, uma vez que, as endonucleases fazem somente a quebra da dupla fita. O reparo é feito pela maquinaria celular do próprio organismo, podendo acontecer de duas maneiras, por NHEJ (*Non-Homologous End-Joining*), ocorrendo em maior frequência em plantas, e geralmente verificam-se modificações *Indels* (inserção ou deleção) no DNA, o que pode alterar a *frameshift* (alteração da janela de leitura) ou romper domínios importantes para a atividade do gene (Zhou *et al.*, 2014). Ou por HDR (*Homology-Directed Repair*), nesse tipo de reparo é necessária a presença de um fragmento homólogo de DNA, natural ou doado (Gratz *et al.*, 2014).

Apesar dos avanços apresentados no melhoramento genético convencional de plantas, a dinâmica de desenvolvimento de novas cultivares não tem ocorrido na velocidade demandada pela agricultura moderna (Gao, 2018). Dessa forma, visto as vantagens do sistema CRISPR, este surge como uma ferramenta auxiliar importante aos programas de melhoramento. Principalmente para algumas culturas, como é o caso da soja que possui um genoma complexo. Isso porque no processo de evolução a soja que é diploide (2n) houve uma duplicação, tornando-se 4n, seguido por novas pressões evolutivas, retornando a seu estado diploide. Sendo assim considerado um poliploide diploizado (Schmutz *et al.*, 2010).

O sistema CRISPR/Cas já tem sido utilizado para edição de genes na cultura de soja (Chilcoat *et al.*, 2017; Cai *et al.*, 2018; Bao *et al.*, 2019; Cheng *et al.*, 2019). Entretanto, devido ao tempo relativamente longo gasto desde a idealização e a construção dos vetores, até a obtenção de um evento transformado, é pertinente a validação do sistema antes da realização do protocolo definitivo para obtenção desse evento. Isto porque, embora se desconheça o motivo biológico, nem todos os alvos escolhidos têm a mesma eficiência de

ação (Doench *et al.*, 2014). Ainda que existam várias ferramentas de predição *in-silico* para o desenho/escolha de sgRNA, estudos experimentais *in vivo* prévios e de forma rápida e que possam mostrar a eficiência de clivagem de DNA, tornam-se fundamentais na escolha dos melhores alvos para a modificação genética desejada (Zhang *et al.*, 2016; Lee, 2017). A tecnologia de CRISPR/Cas associada à metodologia conhecida como *hairy-root*, que induz raízes transgênicas expressando o cassete alvo no organismo de interesse, pode ser adotada como um modelo para investigação de função de genes, bem como avaliar de maneira prévia a eficiência do sgRNAs na edição gênica (Cai *et al.*, 2015; Morriss *et al.*, 2017; Li *et al.*, 2019).

Tradicionalmente o sistema CRISPR/Cas9 o RNA guia está sob o controle de promotores polimerase tipo III, assim como U6 e U3 (U-snRNA *plant promoters*), isso pode ser uma desvantagem, uma vez que para muitas espécies não modelo, esta classe de promotores ainda não é bem caracterizada (Sun *et al.*, 2015). Essa condição fica ainda mais desafiadora em casos de sistemas multiplex, onde é necessário um promotor para controlar cada sgRNA, já que, repetições de um mesmo promotor, podem causar alterações dos níveis de expressão dos sgRNA ou silenciamento do transgene (Ma *et al.*, 2015). Ainda considerando que a Cas9 é codificada por um gene de tamanho aproximado de 4,2kb, o uso do multiplex pode aumentar significativamente o tamanho dos vetores a serem utilizados nos sistemas de transformação genética de plantas.

A avaliação da capacidade de edição de um sistema simplificado onde tanto a Cas9 como o sgRNA sejam controlados por um único promotor polimerase tipo II, configuração denominada de *Single Transcriptional Unit* (STU) (Tang *et al.*, 2016; Xu *et al.*, 2019), comparado ao sistema convencional onde a Cas9 é controlada por um promotor tipo II e o sgRNA controlado por um promotor do tipo III, configuração essa chamada de *Two-Component Transcriptional Unit* (TCTU), pode ser uma interessante estratégia para utilização do sistema CRISPR de uma forma compacta como no STU. Ainda há a possibilidade de utilização de um promotor do tipo II controlando sgRNA, com regulação espaço-temporal, permitindo uma regulação transcricional mais precisa com promotores induzidos (Mikami *et al.*, 2017).

Considerando as dificuldades no processo de transformação genética da cultura, tanto relacionado ao tamanho dos cassetes gênicos inseridos bem como a necessidade de uma avaliação prévia da eficiência de edição dos diferentes sgRNAs possíveis, este capítulo teve por objetivo avaliar duas configurações do sistema CRISPR/Cas9, o sistema *Two-Component Transcriptional* (TCTU) versus *Single Transcriptional Unit* (STU), via transformação de raízes de soja (*hairy-root*), para garantir a escolha do melhor cassete CRISPR/Cas9 para edição de genes em soja. Os genes escolhidos estão envolvidos no final da biossíntese de ácido fítico os genes *IPK1* e *IPK2* (já descritos anteriormente). De modo a alcançar o objetivo final desta tese, que é a obtenção de um evento de soja com redução no conteúdo de ácido fítico.

Este primeiro capítulo deu origem ao artigo publicado na revista *Transgenic Research* (<https://link.springer.com/article/10.1007/s11248-021-00246-x>), anexado ao final do trabalho.

1.2 OBJETIVO GERAL

Avaliar duas configurações vetoriais para CRISPR/Cas9 na edição de genes em soja.

Objetivos específicos

- Avaliar e escolher os *targets* adequados para edição dos genes alvos (*GmIPK1* e *GmIPK2*).
- Construir vetores para as configurações STU e TCTU do cassete CRISPR/*SpCas9* para edição de cada um dos genes *IPK1* e *IPK2* em raízes de soja.
- Obter raízes transgênicas (*hairy-root*) de soja expressando ambos os cassetes STU e TCTU de CRISPR/*SpCas9*, de maneira individual, para edição dos genes *GmIPK1* e *GmIPK2*.
- Avaliar as edições geradas no genoma da soja pelos sistemas STU e TCTU, para determinar o sistema apropriado para transformação estável de soja.

1.3 METODOLOGIA

Escolha das regiões alvo

Sequências genômicas de dois genes da biossíntese do ácido fítico Glyma.14G072200 e Glyma.12G240900 (*GmIPK1* e *GmIPK2*, respectivamente) foram obtidos na plataforma Phytozome (<https://phytozome-next.jgi.doe.gov>, Goodstein *et al.*, 2012), baseando-se no genoma montado Wm82.a2.v1 (Schmutz *et al.*, 2010). As sequências foram analisadas pelo programa *Geneious Prime* 2019.2.3 (<https://www.geneious.com>) para *Streptococcus pyogenes*-Cas9 (*SpCas9*) e os potenciais efeitos *off-target* foram calculados de acordo com o método desenvolvido por Hsu *et al.* (2013). A atividade *on-target* foi avaliada com o software online *GPP sgRNA Designer* (<https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design>) usando Azimuth 2.0 (Doench *et al.*, 2016, Sanson *et al.*, 2018). Três regiões alvo foram selecionadas para cada gene, com base em suas localizações genômicas *off-target score* e predições da atividade *on-target* (Tabela 1).

Tabela 1. Características dos sgRNAs selecionados para *GmIPK1* e *GmIPK2*.

Nome do guia	Sequência do guia ¹	PAM	GC (%)	Posição no genoma (Wm82.a2.v1)	Direção ²	Mismatch ³	<i>Off-target score</i> ⁴	<i>On-target efficacy score</i> ⁵
GmIPK1-Target1	GCAGCTGACTGGGTTTACAG	AGG	55	Gm14:6063363..6063382	<i>sense</i>	4	3.64%	0.6765
GmIPK1-Target2	GAAAGTGGTCCGCATACGTA	AGG	50	Gm14:6063188..6063207	<i>sense</i>	4	1.87%	0.5864
GmIPK1-Target3	GTGGCGAGTTGAAGCTGCCA	GGG	61	Gm14:6062617..6062636	<i>sense</i>	3	7.11%	0.4512
GmIPK2-Target1	GGTGTAAGGAGAGTTCCGG	TGG	50	Gm12:39951279..39951298	<i>antisense</i>	3	4.57%	0.6646
GmIPK2-Target2	GGTCTCCAGGACCAGGTGA	GGG	65	Gm12:39951158..39951177	<i>antisense</i>	4	3.27%	0.5795
GmIPK2-Target3	GCACTTGCAGATGTAGTCCT	CGG	50	Gm12:39951063..39951063	<i>antisense</i>	3	7.96%	0.6359

¹ sgRNAs identificado pelo programa Geneious Prime 2019.2.3

² Direções baseada no genoma modelo Wm82.a2.v1.

³ Números mínimo de *mismatches* entre as sequências sgRNA e a possibilidade de *off-target*.

⁴ Identificação *off-target score*. Baseado em Hsu *et al.* 2013.

⁵ CRISPRko (Doench *et al.* 2016 e Sanson *et al.*, 2018). Escala 0 a 1

Construção de vetores de CRISPR/Cas para obtenção de raízes transformadas

O vetor CRISPR/Cas9 p201-EGFP-C9 foi montado para avaliar os sistemas STU e TCTU (Figura 2). Esse vetor contém o gene repórter EGFP (*Enhanced Green Fluorescent*

Protein) (Chiu *et al.*, 1996), sob o controle do promotor CsVMV (Cassava vein mosaic virus) (Verdaguer *et al.*, 1996) e o terminador do gene *nos* de *A. tumefaciens* (An *et al.*, 1995). A sequência que codifica para *SpCas9* (Mali *et al.*, 2013) foi obtida do plasmídeo Addgene #41815, e está sob o controle do promotor *GmUbi3 Glycine max* (Chiera *et al.*, 2007), e terminador *PsRbcs (Pisum sativum rubisco small subunit-* Schardl *et al.*, 1987). Cada sgRNA das configurações de TCTU está sob o controle do promotor *MtU6* de *Medicago truncatula* (Kim *et al.*, 2013). No sistema STU (Wang *et al.*, 2018), os sgRNAs, (*target+Scaffold-* Dang *et al.*, 2015) são separados por uma sequência única (UNS), que são *linkers* para facilitar a clonagem (Torella *et al.*, 2014).

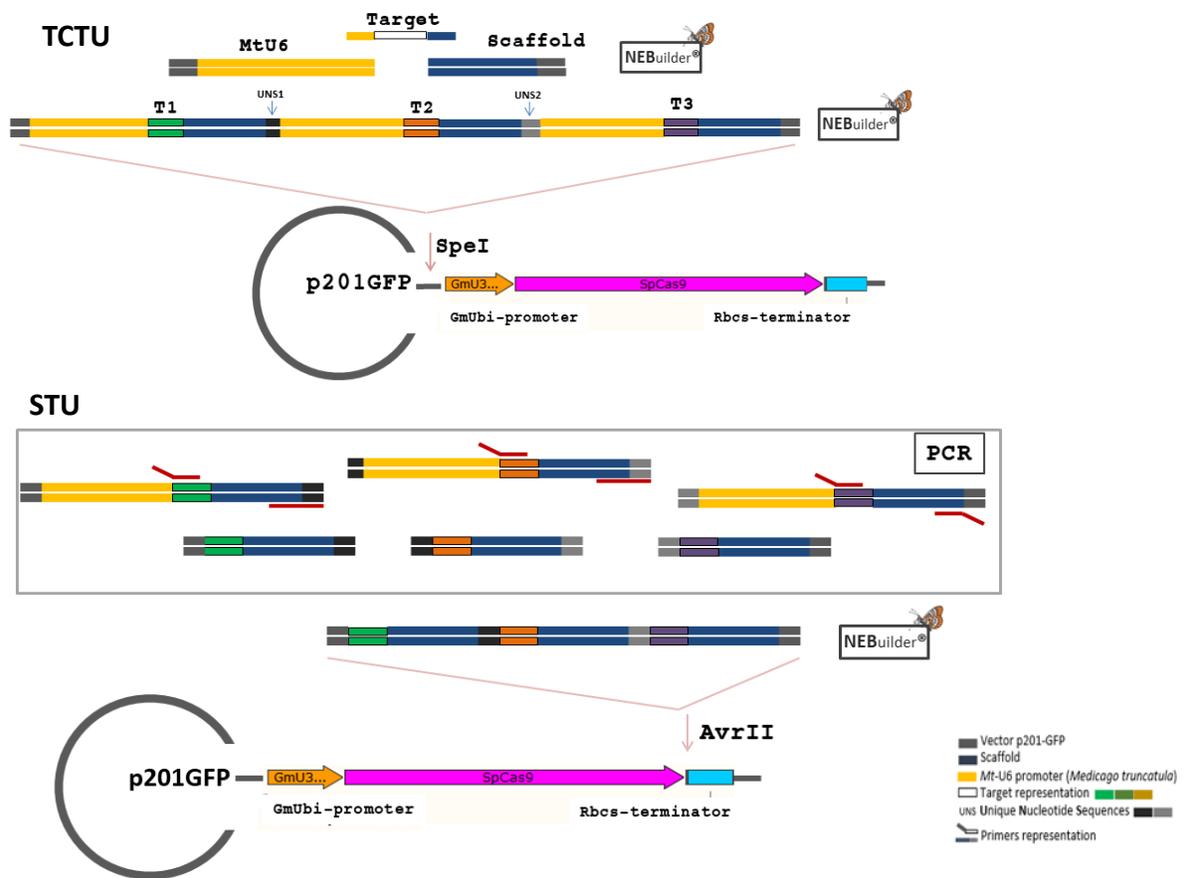


Figura 2: Esquemática da montagem dos vetores. Sistema TCTU (Two-Component Transcriptional Unit): 1: Modelo dos alvos desenhados; 2: **Cassete primário**, resultado da reação NEBuilder, do promotor *MtU6*+*Target*+*Scaffold*; 3: **Cassete intermediário TCTU**, apresentando a configuração *MtU6*-*Target1*-*Scaffold*-*UNS1*-*MtU6*-*Target2*-*Scaffold*-*UNS2*-*MtU6*-*Target3*-*Scaffold*; 4: Clonagem do cassete no vetor p201GFP-*SpCas9* na sítio de restrição *SpeI*, gerando assim o vetor final do sistema TCTU. **Sistema STU (Single Transcriptional Unit):** 5: Fragmentos gerados após PCR usando os cassetes primários apresentados em 2 e *primers* específicos; 6: **Cassete intermediário STU** gerado após reação de NEBuilder apresentando a configuração *Target1*-*Scaffold*-*UNS1*-*Target2*-*Scaffold*-*UNS2*-*Target3*-*Scaffold*; 7: Clonagem do cassete no vetor p201GFP-*SpCas9* na sítio de restrição *AvrII*, gerando assim o vetor final do sistema STU.

Os cassetes da configuração CRISPR/Cas9-TCTU foram montados pelo sistema NEBuilder (NEBuilder[®] HiFi DNA Assembly Cloning Kit- Biolabs[®]) e clonados no vetor p201-EGFP-C9 digerido com a enzima de restrição *SpeI*, gerando o vetor final de 15.454pb (*MtU6-sgRNA(3X) -GmUbi-SpCas9-RbcsT*) (Figura 2-A). Para a montagem dos vetores CRISPR/Cas9-STU, os cassetes foram clonados em p201-EGFP-C9 digerido com a enzima de restrição *AvrII*, gerando os vetores finais de 14.396pb (*GmUbi-SpCas9-sgRNA(3X)-RbcsT*) (Figura 2-B).

Obtenção de raízes transgênicas pelo sistema *hairy-root*

Foram utilizadas sementes de soja, cultivar *Jack*, variedade com melhor desempenho na embriogênese somática e regeneração de plantas (Raza *et al.* 2019). O protocolo de *hairy-root* foi realizado segundo Jacobs *et al.* (2015) sem seleção. Os vetores foram transformados na cepa K599 de *Agrobacterium rhizogenes*. A fluorescência de EGFP foi usada para detectar raízes transgênicas 20 dias após o processo de transformação. Cada raiz positiva foi considerada um único evento e usada para análise posterior. Além disso, reação de PCR foi feita para resgate dos fragmentos referentes aos genes alvos, para posterior detecção das alterações, os *amplicons* referem-se aos três alvos juntos. O *amplicon* de 1322 pb é esperado para o gene *GmIPK1* utilizando os primers IPK1-97F: ACACAATTCCTTTCCACCA e IPK1-1399R: AGCAGAGGCTAGATCCTTGA, e para o gene *GmIPK2*, é esperado um fragmento de 1053pb com os primers IPK2-80F: TTGCATTGCTTTGTGTAAGG e IPK2-1113R: CTGCGACACTAATTCAAGCA.

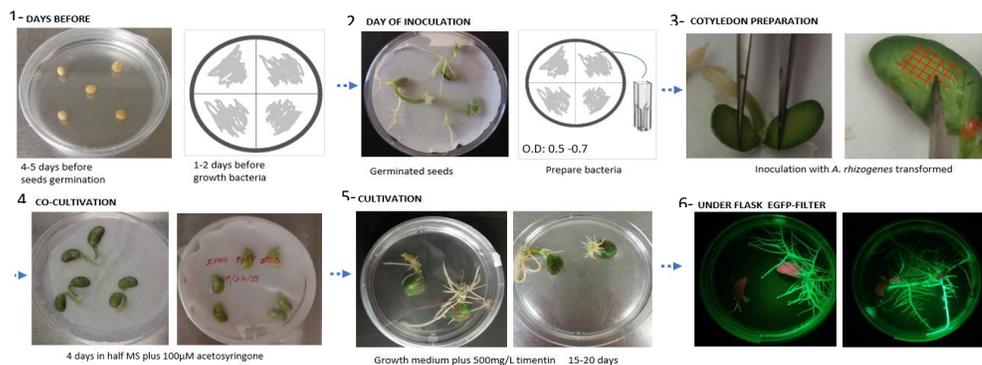


Figura 3: Esquemática do protocolo de *hairy-root*. 1: Germinação das sementes em meio MS-sólido 4 dias antes do experimento, crescimento bacteriano 1 dia antes do experimento; 2: Sementes germinadas e preparação da bactéria (cepa K599) transformada crescida para um densidade óptica entre 0,5-0,7 diluída em tampão PB; 3: Preparação dos cotilédones, apresentando como deve ser separado os cotilédones e o corte para a inoculação (traços em vermelho); 4: Co cultivo em papel filtro umedecido com MS+100 µM acetosiringona por 4 dias; 5: Cotilédones em MS + timetin 500mg/L⁻¹, crescimento das raízes; 6: Raízes crescidas avaliadas sob filtro para EGFP.

Detecção de edição de genes mediada por CRISPR / Cas9

DNA genômico foi isolado das raízes positivas para presença da EGFP (coletado aproximadamente 2 mm da extremidade da raiz), usando o kit *Wizard Genomic DNA Purification* (Promega, Madison, WI, EUA). Os padrões de edição de genes foram determinados inicialmente pelo sequenciamento tipo Sanger. Os fragmentos obtidos pela PCR foram sequenciados e analisados pelo programa SYNTHEGO (www.synthego.com) (arquivo *.abi*).

Em seguida um novo sequenciamento utilizando *primers* específicos com sequências adaptadoras para TruSeq Illumina (Tabela 2) foi realizado para determinação da eficiência dos sistemas. Os *amplicons* foram agrupados e codificados individualmente com primers universais indexados iTru/iNext com códigos de barras. As amostras foram agrupadas em quantidades equimolares e sequenciadas em um sequenciador do tipo Illumina MiSeq, usando o modo de extremidade emparelhada de 150-pb (2x150 PE). Esse experimento foi feito pelo *Georgia Genomics and Bioinformatics Core, University of Georgia*. Os dados de sequenciamento dos *amplicons* foram analisados usando o software AGEseq (Xue & Tsai, 2015).

Tabela 2: Primers usados para detecção das edições mediadas por CRISPR/Cas9

Gene (Gene modelo)	Target	Primer Forward 5' to 3' †	Primer Reverse 5' to 3' †	Amplicon (pb)
<i>GmIPK1</i> (Glyma.14G072200)	IPK1-Target 1	<u>CCTACACGACGCTCTTCCGATCT</u> AGCGTAAGCGAGCACAAGAT	<u>G TTCAGACGTGTGCTCTTCCGATC</u> GCTGCATTCCATTTTCTCC	238
	IPK1-Target 2	<u>CCTACACGACGCTCTTCCGATCT</u> CGTTCATGTGGAGTCAAAGC	<u>G TTCAGACGTGTGCTCTTCCGATC</u> GTTGTAGCTGAATTGTGTGCTTAT	211
	IPK1-Target 3	<u>CCTACACGACGCTCTTCCGATCT</u> TCTGACATGAGAAGGCCAAA	<u>G TTCAGACGTGTGCTCTTCCGATC</u> TTGACCTATGCCTCCTTGGT	204
<i>GmIPK2</i> (Glyma.12G240900)	IPK2-Target 1	<u>CCTACACGACGCTCTTCCGATCT</u> GAAGAAGGAGCGGATGGAGT	<u>G TTCAGACGTGTGCTCTTCCGATC</u> AAGGCCAAGGACGGAATC	209
	IPK2-Target 2	<u>CCTACACGACGCTCTTCCGATCT</u> AGGTTCTGGAGCCGATCTTT	<u>G TTCAGACGTGTGCTCTTCCGATC</u> ACTCCATCCGCTCCTTCTTC	200
	IPK2-Target 3	<u>CCTACACGACGCTCTTCCGATCT</u> GGCGGATAGACACTGGAGAG	<u>G TTCAGACGTGTGCTCTTCCGATC</u> CTACTCCAAACCCTCCGTCA	246

† Sublinhado indica sequencia universal 5' TruSeqHT

1.4 RESULTADOS

Design dos alvos e construção dos vetores CRISPR/Cas

Para avaliação do sistema convencional e simplificado, foram usados dois genes que fazem parte da via do inositol-fosfato (*GmIPK1* e *GmIPK2*). O gene *IPK1* (Glyma.14G072200), renomeado como IPPK na plataforma do Phytozome, apresenta dois ortólogos Glyma.04G030000 e Glyma.06G03010. Embora o gene do cromossomo 14 possua similaridade nos níveis de transcrição quando comparado com as outras duas cópias (85,9% com a cópia do cromossomo 4 e 85% de similaridade com a do cromossomo 6), tem diferença com as sequências a nível de DNA, similaridade de 75% com a cópia do cromossomo 4 (tamanho 7.535pb) e 78% com a cópia do cromossomo 6 (tamanho do 6.400pb). A cópia do cromossomo 14 do gene *GmIPK1*, apresenta níveis maiores de expressão em sementes de soja imaturas (Yuan *et al.*, 2012). De modo que essa cópia foi usada como modelo para o desenho dos alvos, no entanto as sequências alvos escolhidas apresentaram similaridade com as outras cópias. Um alinhamento da região alvo das três cópias do gene *IPK1* estão apresentados na figura 4.

As sequências genômicas de ambos os genes foram obtidas do genoma de *G. max* (Schmutz *et al.*, 2010). O gene *GmIPK1* tem 4.158 nucleotídeos distribuídos em sete exons. Um total de 46 sequências-alvo para *SpCas9* foram identificados no *GmIPK1*, 22 delas na sequência CDS, dos quais três alvos foram escolhidos. Os alvos estão localizados no início do exon 1, exon 2 e exon 3 respectivamente e foram nomeados GmIPK1-target1, GmIPK1-target2 e GmIPK1-target3. Seu *off-target score* variou de 1,87% a 7,11%, com a atividade *on-target score* variando de 45,12% a 67,65%. Todos os três alvos estão na orientação *sense*.

Glyma.12G240900 (*GmIPK2*) é um gene com um único exon (1.630 nucleotídeos). Das 65 sequências-alvo identificadas no gene, 47 tinham como alvo o CDS. Os três guias selecionados para este gene foram todos escolhidos na orientação *anti-sense* e separados por pelo menos 75 pb. Seu *off-target score* variou de 3,27% a 7,96%, e *on-target score* entre 57,96% e 66,46%. Igualmente, os alvos selecionados para *GmIPK2* foram nomeados GmIPK2-target1, GmIPK2-target2 e GmIPK2-target3.

Target 1

```

14 gaggaagacgcagctgactgggtttacagaggggaagg
6  gagaaggacgcaaatgagtgggtttaccgagggcgaagg
4  gagaaggacgcagctgagtgggtttaccgagggcgaagg
   *** * *****   *** *****   *** *****

```

Target 2

```

14 gaaagtggtcgcatacgttaaggctccaaggaatggat
6  gaaagtgatgcgcatcgttaaggctccaaggaatggcg
4  gaaagtgatgcgcatcgttaaggctccaaggaatggcg
   ***** * ***** ***** *****

```

Target 3

```

14 cctgcgtggcgagttgaagctgccagggttgatgcacat
6  cctgcttggagagttgataatgctcgggtcgacatgcat
4  cctgcttggagagttgataatgctcaggctcgacatgcat
   ***** *** *****   ***   *** **   ***

```

Figura 4: Alinhamento das três cópias do gene *GmIPK1* presentes no genoma da soja Willians 82: São apresentados o alinhamento da cópia do cromossomo 14 (alvo desses estudos) e as cópias dos cromossomos 4 e 6. Também são apresentados a região referente aos alvos 1 (destacados em amarelo), alvo 2 (destacado em cinza) e alvo 3 (destacados em rosa). Contendo também a região PAM (destacado em verde).

Os sgRNAs foram desenhados de acordo com os sistemas: *single transcriptional unit* (STU) e *two-component transcriptional unit* (TCTU). No sistema STU, os sgRNAs foram expressos sob regulação de um único promotor *GmUbi3*. No sistema TCTU, cada sgRNA está sob regulação de um promotor *MtU6*. Após a montagem dos quatro vetores pelo sistema *NEBuilder assembly*, os vetores finais (Figura 5), foram nomeados de p201-EGFP-C9, IPK1-TCTU, IPK1-STU, IPK2-TCTU e IPK2-STU, sendo usados para a transformação das raízes de soja.

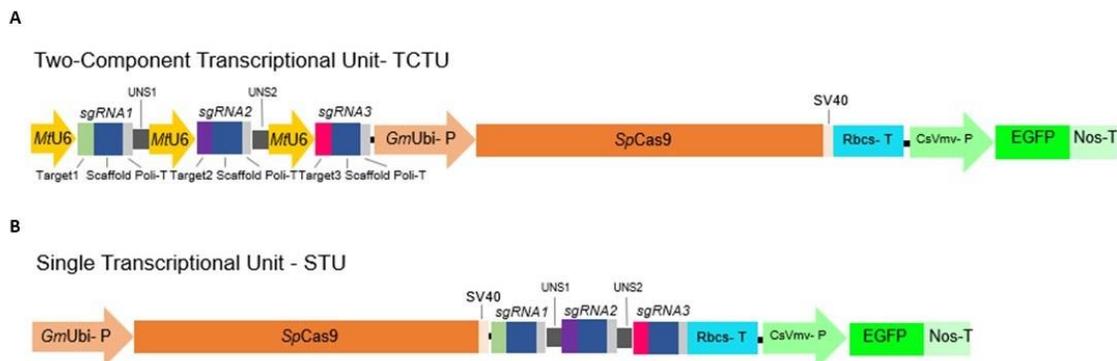


Figura 5: Esquema dos dois sistemas de expressão CRISPR/Cas9: Sistema TCTU e Sistema STU. Vetores de *hairy-root* CRISPR/Cas9 para editar os genes *GmIPK1* e *GmIPK2* com os dois sistemas TCTU (A) e STU (B). A endonuclease Cas 9 de *Streptococcus pyogenes* (*SpCas9*) está sob o controle do promotor Ubiquitin-3 de *Glycine max* (laranja). O cassete do gene repórter EGFP está sob o controle do promotor CsVMV (Cassava vein mosaic virus) (verde claro) e do terminador T-nos (Nopalina Sintase). No sistema TCTU, cada sgRNA é conduzido pelo promotor *MtU6* separado por uma sequência UNS. No sistema STU, os sgRNAs são separados apenas pela sequência UNS. SV40: Peptídeo de sinalização nuclear.

Obtenção de raízes de soja transgênicas

Para cada construção, 10 cotilédones de sementes de soja em germinação foram transformados. As raízes transgênicas geradas, foram confirmadas como positivas por avaliação sob luz fluorescência de EGFP. Um total de 25 raízes positivas para EGFP individuais foram coletadas para cada transformação, o DNA genômico foi isolado e uma biblioteca de sequenciamento de *amplicons* foram preparados para determinar a edição dos genes mediada por CRISPR/Cas9. Após a reação de PCR foi observado para algumas amostras fragmentos de tamanho menores (*amplicon* esperado de 1322 pb (*GmIPK1*) e 1053 pb (*GmIPK2*) (Figura 6).

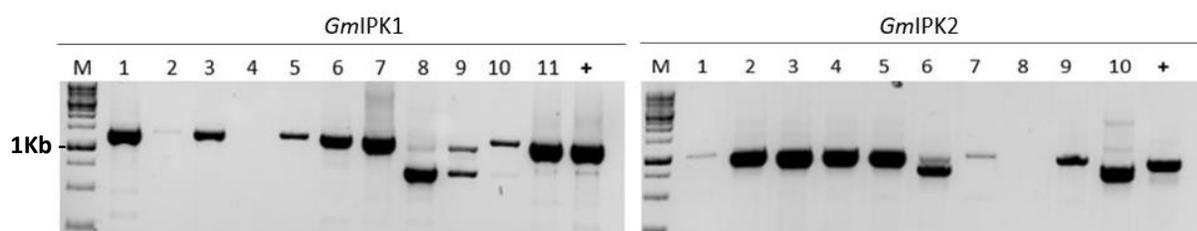


Figura 6: Detecção dos fragmentos dos genes *GmIPK1* e *GmIPK2* por PCR em gel de agarose sob luz UV. Fragmentos esperado de 1233pb para IPK1 e 1053pb para o gene IPK2. *GmIPK1*: Linhas 1- 6 IPK1-STU, linhas 7-11 IPK1-TCTU +: Raiz controle (NT); *GmIPK2*. Linhas 1-5 IPK1-STU, linhas 7-10 IPK1-TCTU +: Raiz controle (NT), M: 1Kb ladder LIBPBio

Inicialmente foi realizado sequenciamento tipo Sanger, onde foi possível verificar que os sistemas foram capazes de gerar edições nos genes de interesse e ainda obter uma porcentagem de *Indels* presentes nos *amplicons* analisados. Em seguida um sequenciamento tipo Illumina MiSeq foi realizado para determinar a real eficiência dos sistemas.

Análises das sequências editadas via sequenciamento tipo Sanger

Com as análises preliminares das sequências obtidas após sequenciamento tipo Sanger foi possível observar se os cassetes CRISPR/Cas9 foram capazes de gerar edições, e com as análises das sequências no programa SYNTHGO foi possível obter a porcentagem de *Indels* presentes nas amostras avaliadas. As amostras potencialmente editadas foram comparadas a uma amostra controle. A análise individual das 23 amostras para o sistema IPK1-TCTU mostrou 15 amostras com percentual de edição (*Indels*) entre 50 a 80%, e para as sete restantes a porcentagem de edição foi inferior a 45%. Da mesma forma, para o sistema IPK1-STU, 13 amostras apresentaram taxas de edição entre 50 e 65%, e 10

amostras apresentaram uma taxa de edição entre 10 e 40%. De modo que a média de edições observadas pela análise no SYNTHEGO nos sistemas TCTU e STU foram de 48% e 43% respectivamente para o gene *GmIPK1*.

As sequências obtidas com o sequenciamento das amostras que foram transformadas com o vetor da configuração IPK1-TCTU continham mutações pontuais nas regiões de cada alvo, e também apresentou uma deleção de 572pb na região localizada entre o segundo e o terceiro alvos para algumas amostras, indicando que o fragmento entre dois sítios de clivagem tinha sido removido (Figura 7-A). Nas sequências das amostras IPK1-STU além das mutações pontuais detectadas na região alvo, para algumas sequências, as deleções foram observadas entre o primeiro e o terceiro alvos (750pb) (Figura 7-B).

Já para as análises das sequências do gene *GmIPK2*, das 22 amostras analisadas para IPK2-TCTU, 13 amostras apresentaram taxas de edição entre 55 e 99%, nove apresentaram taxas de edição entre 10 e 40% ou nenhuma edição. O alinhamento das sequências IPK2-TCTU revelou uma deleção correspondente à região entre o primeiro e o terceiro alvo (215pb) na maioria das amostras (Figura 7-C). No entanto, no sistema IPK2-STU, de 22 amostras sequenciadas, cinco amostras apresentaram edições acima de 70%, 10 amostras apresentaram taxas de até 40% e sete não foram editadas. No alinhamento das

sequências obtidas apenas três das 22 sequências analisadas mostraram a deleção abrangendo do primeiro ao terceiro alvo (Figura 7-D). A média de eficiência para os dois sistemas no gene *IPK2* avaliado pelo programa SYNTHEGO foi de 55% no sistema TCTU e 21% no sistema STU.

Análises das edições e avaliação da eficiência via sequenciamento tipo Illumina

Após a avaliação da porcentagem de *Indels* presentes nas amostras analisadas pelo sequenciamento tipo Sanger, fez-se necessário uma análise mais acurada nas sequências, uma vez que a heterozigotidade é um fator importante para determinar a eficiência de cada sistema. Uma biblioteca de *amplicons* dos alvos foi preparada para determinar a edição dos genes mediada por CRISPR/Cas9, utilizando tecnologia de sequenciamento Illumina MiSeq *paired-end*. Após tratamento um total de 376.180 *paired-end* foram gerados nas bibliotecas de *amplicon-seq* de *GmIPK1* e *GmIPK2*. A análise subsequente mostrou que 180.971 (48,11%) e 195.209 (51,89%) *reads* foram obtidas de *GmIPK1* e *GmIPK2*, respectivamente. Os padrões de modificação genética geradas por CRISPR/Cas9 foram analisados usando o software AGEseq.

A análise mostrou que o *GmIPK1*-STU apresentou baixa taxa de edição (Figura 7), onde 99,1% dos *reads* eram iguais ao controle NT, 6 das 15 amostras não apresentaram edições, enquanto menos de 2% apresentaram edições em dois dos três alvos. Das sequências que apresentaram edições a amostra *GmIPK1*-STU- #04 apresentou deleção de 1-base em todos os três sítios (<0,3%). A sequência entre o alvo 1 e o alvo 3 (751pb) foi deletada em 156 dos *reads* (3,99% e 5,40%). *GmIPK1*-STU-#06 exibiu apenas modificações para o alvo 3, com deleção de 5 pb detectada em 19% das *reads*, já para *GmIPK1*-STU-#08 edições foram apresentadas para todos os alvos. Uma deleção de 2pb foi detectada no alvo um com uma frequência maior que 87%. Simultaneamente, uma deleção de 4pb foi observada em 18,52% das *reads* para o alvo três. Para o alvo dois uma porcentagem reduzida de *reads* apresentou alguma modificação, 4% das *reads* mostraram uma deleção de três bases e outros 3% mostrando uma inserção de base única. Para *GmIPK1*-STU-#09, foi observada uma deleção de 749pb compreendendo a sequência entre os alvos 1 e 3.

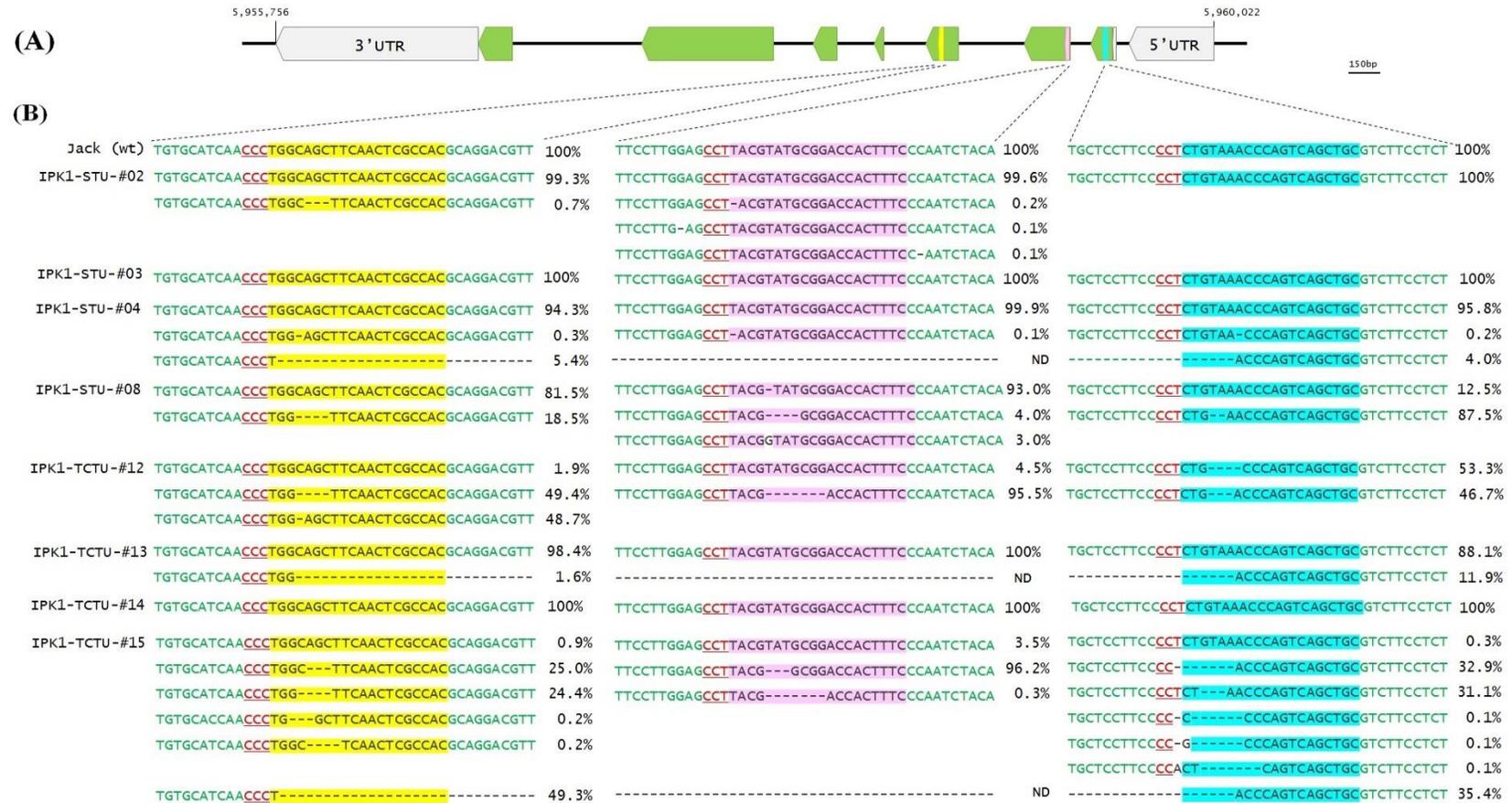


Figura 8: Comparação dos sistemas de edição STU (single transcriptional unit) e TCTU (two-component transcriptional unit) do gene *GmIPK1* em raízes de soja por CRISPR/Cas9. (A) Ilustração esquemática do gene *GmIPK1* (Glyma.14G072200). Todos os exons são representados por setas verdes e os íntrons são representados por linhas. Ambas as regiões não traduzidas (UTR) 5' e 3' são representadas como caixas cinza. *GmIPK1*-Alvo 1, *GmIPK1*-Alvo 2 e *GmIPK1*-Alvo 3 são representados por setas azuis, rosa e amarelas, respectivamente. (B) Mutações induzidas por CRISPR/Cas9 pelos sistemas STU e TCTU. As sequências *wild-type* estão em verde, as deleções são mostradas como traços e os SNPs são mostrados em preto. As sequências-alvo são destacadas em azul, rosa e amarelo, e o PAM é sublinhado em vermelho. As porcentagens próximas às sequências indicam o número de leituras mapeadas sobre o número total de leituras sequenciadas para um determinado alvo em cada amostra. ND: Não detectado.

Enquanto isso para o sistema *GmIPK1-TCTU* (Figura 8) uma alta frequência de edição foi apresentada, onde 73,20% dos *reads* apresentavam mutações. A natureza das mutações foi desde deleções de 1 a 12pb ou inserções de uma base, até grandes deleções que removeram sequências entre dois dos sgRNA. Apenas dois dos 15 eventos analisados não apresentaram mutações. Deleções de 751pb (compreendendo a sequência entre alvo um e alvo três) foram observadas em 8 das 15 amostras para *GmIPK1-TCTU*. No entanto, deleções de fragmentos menores, como entre o alvo um e alvo dois (*GmIPK1-TCTU-#04* e *GmIPK1-TCTU-#06*) e entre o alvo dois e o alvo três (*GmIPK1-TCTU-#05*) também foram detectadas. Alguns eventos únicos, como *GmIPK1-TCTU- #09*, *GmIPK1-TCTU- #12* e *GmIPK1-TCTU- #15*, mostraram níveis de edição superiores a 95% em todos os três sítios. As frequências médias de edição de 89,30%, 69,13% e 79,39% foram estimadas para o alvo um, alvo dois e alvo três, respectivamente. A frequência mais baixa de deleção no alvo dois em relação aos outros, pode ser devido a possibilidade de remoção do fragmento entre o primeiro e o terceiro alvo, resultando na não detecção do fragmento referente ao alvo dois.

Na análise do sistema *GmIPK2-STU* (Figura 9) um total de 96,38% das *reads* eram similares à linhagem selvagem. Os alvos 1 e 2 apresentaram taxas de edição próximas, 12,71% e 13,76% respectivamente, enquanto para o alvo 2 a taxa de edição foi tecnicamente indetectável (0,04%). A mutação observada em maior frequência foi a deleção de um fragmento de 215pb correspondente entre os alvos 1 e 3, estando presente em 54,08% dos eventos analisados no sistema *GmIPK2-STU*. Todos os eventos apresentaram alguma edição mediada por CRISPR/Cas9, mesmo que menores com deleções de 1 a 2 pb.

O sistema *GmIPK2-TCTU* (Figura 8) promoveu um número maior de edições, quando comparado ao sistema *STU*, com 79,33% de edições. Todos os eventos apresentaram algum tipo de edição, sendo a deleção entre o alvo 1 e alvo 3 as mais frequentes, além disso tivemos para o evento *IPK2-TCTU-#05* deleções entre o alvo 1 e alvo 2. Já os eventos *IPK2-TCTU-#08* e *IPK2-TCTU-#04* apresentaram os maiores índices de edição, variando entre 60,85% e 41,57%, respectivamente. Apenas *IPK2-TCTU- #04* apresentou edição no alvo 2, de modo que a eficiência do alvo 2 foi a menor taxa (0,67%) (dos 15 eventos 10 não apresentaram edições para o alvo 2).

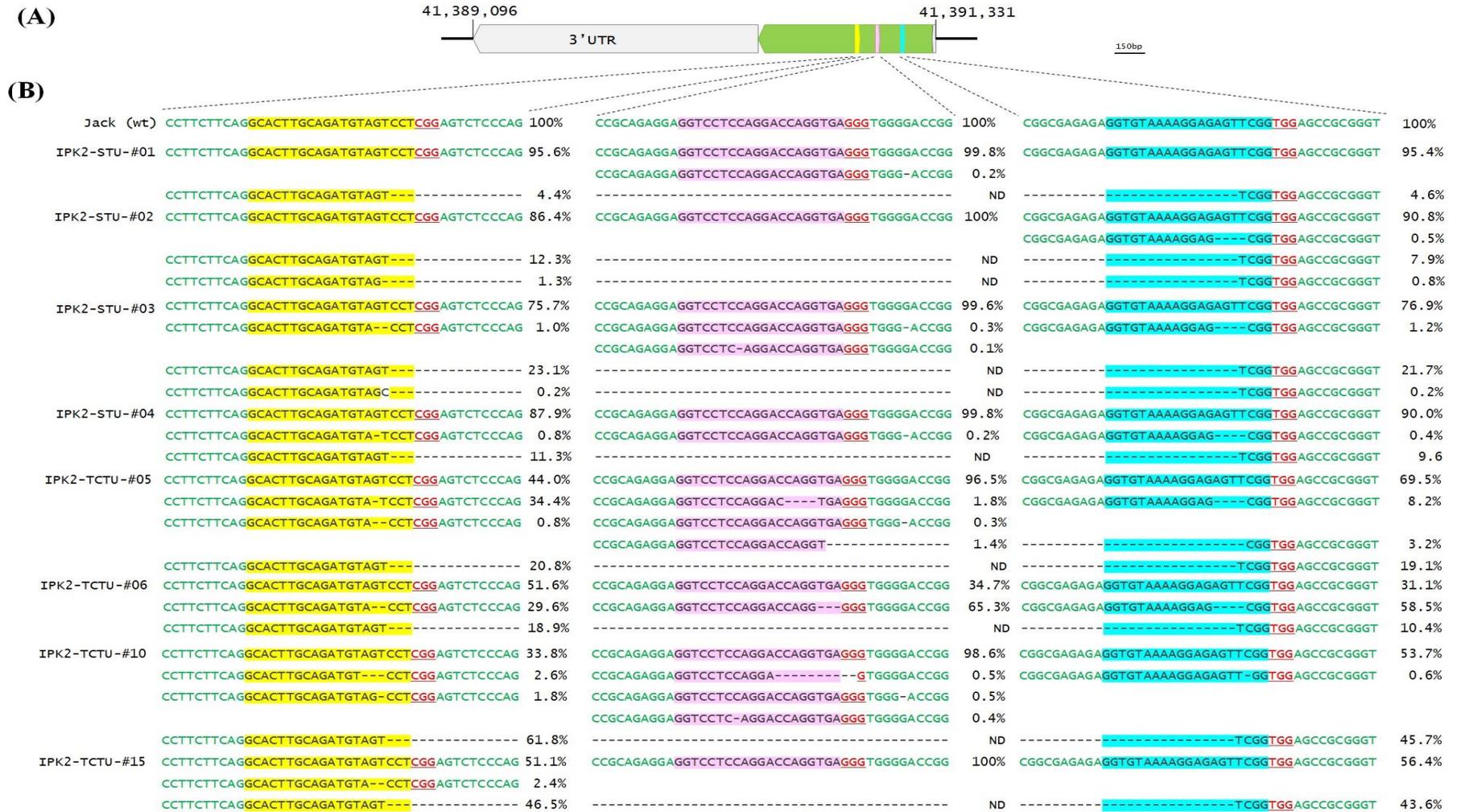


Figura 9: Comparação dos sistemas de edição STU (single transcriptional unit) e TCTU (two-component transcriptional unit) do gene *GmIPK2* em raízes de soja por CRISPR/Cas9. (A) Representação esquemática do gene *GmIPK2* (Glyma.12G240900). Todos os exons são representados por setas verdes e os íntrons são mostrados como linhas. As regiões não traduzidas 5' e 3' (UTR) são mostradas como caixas cinza. *GmIPK2*-Alvo 1, *GmIPK2*-Alvo 2 e *GmIPK2*-Alvo 3 são representados por setas azuis, rosas e amarelas, respectivamente. (B) Mutações induzidas por CRISPR/Cas9 nos sistemas STU e TCTU. As sequências *Wild-type* estão em verde, as deleções são mostradas como traços e os SNPs são mostrados em preto. As sequências alvo são destacadas em azul, rosa e amarelo e o PAM é destacado e sublinhado em vermelho. As porcentagens ao lado das sequências indicam o número de leituras mapeadas sobre o número total de leituras sequenciadas para um determinado alvo em cada amostra. ND: Não detectado.

1.5 DISCUSSÃO

Embora a tecnologia CRISPR tenha revolucionado a edição de genes desde sua descoberta, novos avanços vêm sendo feitos para tornar esse sistema cada vez mais eficiente. Tradicionalmente, o sistema CRISPR/Cas9 usa uma combinação de dois tipos de promotores, um promotor polimerase de classe II que regula a expressão da endonuclease Cas e outro da classe III que regula o sgRNA (Tang *et al.*, 2019). Algumas limitações de uso tornaram-se aparentes após seu desenvolvimento, isso por conta da baixa caracterização de promotores da classe pol III em alguns organismos (Sun *et al.*, 2015). Além de uma expressão descoordenada da Cas9 e dos sgRNA, uma vez que estão sendo direcionados por diferentes promotores (Tang *et al.*, al. 2019), ou extrapolando para uma possível variação dos níveis de expressão ou silenciamento do transgene, com o uso repetitivo de sequências promotoras (Ma *et al.*, 2015).

Estudos têm demonstrado que o sistema CRISPR/Cas9 para edição gênica pode sim ser mais simplificado com a possibilidade de utilização de cassetes sem comprometer sua eficiência. Isso poderá facilitar o caminho para a edição de genes aplicada à melhoria de características comerciais de culturas de importância econômica, especialmente no que diz respeito à quantidade de material genético que é suportado para ser introduzido em uma célula usando quaisquer sistemas de transformação. Isso é um ponto crítico para o sistema biobalístico, onde o uso de grandes vetores pode resultar em uma interrupção em uma região não específica do genoma (Baltes *et al.*, 2017). O sistema CRISPR/Cas9 mais compacto como o STU, permitiria a utilização de vetores mais curtos, com maior número de alvos simultâneos sem a necessidade de repetição de promotores.

A possibilidade de uso de vetores menores e mais simplificados gera bastante interesse. Esse trabalho avaliou a eficiência de edição em genes de soja, pela expressão do sistema CRISPR/Cas9 em duas configurações vetoriais diferentes, o primeiro conhecido por TCTU, é o sistema tradicionalmente usado *versus* um sistema simplificado (STU). Sendo possível reduzir o cassete de expressão CRISPR/Cas9 de 7050pb no sistema TCTU para 5990pb no sistema STU, uma redução de mais de 1000pb. Dado que um sgRNA tem 120pb, em 1000pb disponíveis para utilização, seria possível adicionar aproximadamente oito novos sgRNAs em um design simplificado, o que significa oito novos alvos para

outros genes em um sistema multiplex. Isso se torna importante quando se deseja causar ruído numa família inteira de genes, ou em rota completa de síntese.

Um sistema vetorial simplificado similar foi usado para editar genes em plantas de arroz, apresentando uma eficiência entre 29 a 38%, com mutações pontuais observadas no local de clivagem ou no máximo duas bases à frente. Quando comparado ao sistema tradicional a taxa de eficiência manteve-se a mesma (Tang *et al.*, 2016). Com algumas modificações para aperfeiçoar o sistema, usando sequências de clivagem, a taxa de mutação foi próxima a 50% com seis alvos. (Tang *et al.*, 2019). Diferentes endonucleases foram avaliadas com a configuração STU (*FnCpf1*, *LbCpf1* ou *SpCas9*) usando um sistema multiplex, e a eficiência de edição foi superior a 50% (Wang *et al.*, 2018).

No presente trabalho foi utilizado um sistema multiplex com três alvos para cada gene, de forma a garantir um *knockout* total do gene. Foram observados resultados gene-específico tanto para TCTU quanto para STU. Ambos foram capazes de criar *Indels* a partir de DSBs, incluindo deleções que variaram de 1 a 10pb para *IPK2* e de 1 a 12pb para o gene *IPK1*. Inserções de 1 base também foram detectadas para ambos os genes. Além disso, foram detectadas deleções de 752pb em *GmIPK1*, fragmento referente entre o alvo um e alvo três na sequência de *GmIPK1*. Para *GmIPK2*, a maior deleção foi de 217pb, distância entre alvo um e alvo três. Essas grandes deleções foram possíveis provavelmente pela ação simultânea do sistema CRISPR/Cas9 entre o primeiro e o terceiro alvo (Zhang *et al.*, 2019).

O sistema TCTU apresentou frequências de edição mais altas para ambos os genes. O sistema STU empregado aqui foi baseado nos relatos de Wang *et al.* (2018) e Mikami *et al.* (2017) para a cultura do arroz, onde não foi necessário a utilização de sistema de clivagem para individualizar os sgRNA após a transcrição, uma vez que, a própria maquinaria da planta pode fazer esse processamento (Mikami *et al.*, 2017). Mikami *et al.* (2017) observaram uma taxa de edição de até 100% utilizando um sistema simplificado sem a presença de qualquer estrutura de clivagem. Já outros trabalhos usaram sistemas de clivagem, tais como: RNA-Cys4 (Čermák *et al.*, 2017), ribozimas exógenas (Gao *et al.*, 2015, Tang *et al.*, 2016), ou sistema de gene tRNA-gRNA policistrônico (Xie *et al.*, 2015) para liberar os sgRNAs. Essas abordagens podem contornar as frequências de

deleções/inserções reduzidas em relação ao sistema tradicional, obtidas nesse trabalho com o sistema STU.

Notavelmente, o fragmento inteiro entre os alvos um e três foi removido com mais frequência para *GmIPK2*, já que todos os eventos de STU e TCTU tiveram essa deleção. Para *GmIPK1*, todo o fragmento entre os alvos um e três (aproximadamente 715 pb) foi eliminado em 4 de 15 eventos com o sistema STU e 8 de 15 eventos para o sistema TCTU. As edições para o alvo 2 do *GmIPK2* foram raras. Este último é o gene menor, e a proximidade dos locais alvo (101 pb entre os alvos 1 e 2; 76 pb entre o alvo 3 e 2) podem ter levado à interferência com as endonucleases Cas. O alvo dois para ambos os genes apresentaram uma baixa eficiência em induzir mutações, uma explicação possível e que foi observada tanto para o gene *IPK1* quanto *IPK2* em ambos os sistemas, é a possibilidade de remoção do fragmento entre o primeiro e o terceiro alvo, como observado no alinhamento das sequências (Figura 6). A ausência da região referente ao segundo alvo pode ter subestimado a porcentagem de edição.

Em células diploides, três zigosidades são possíveis, monoalélico, heterozigoto dialélico (mais comumente referido como bialélico) e homozigoto dialélico (Luttgeharm *et al.*, 2017). Conhecer a natureza das mutações ajuda a entender a eficiência de cada sistema. Em nossos resultados, a edição bialélica foi rara para o sistema STU, embora o evento #8 para *GmIPK1* tivesse uma frequência de edição bialélica de 87,5% para o alvo 1, consistindo em uma deleção de 2 pb. Para *GmIPK1* com o sistema TCTU, todos os três alvos eram bialélicos para os eventos #4 e # 9, enquanto os eventos #6 e #8 eram monoalélicos para os alvos 1 e 3. O evento #10 também era bialélico para os alvos 1 e 3, mas monoalélico para o alvo 2. Edições dialélicas eram predominantemente heterozigotos. Nenhum dos alvos no *GmIPK2* resultou em edições dialélicas para nenhum dos sistemas.

Em resumo, os sistemas TCTU e STU foram capazes de gerar edições em genes de soja que codificam para enzimas da via de síntese de ácido fítico, embora as frequências de edição mais altas obtidas fossem quando utilizamos o sistema TCTU. Diante dessa informação prévia, utilizamos essa configuração de vetor como base para a produção de plantas transgênicas com baixo teor de ácido fítico.

1.6 CONSIDERAÇÕES

O tempo necessário para obtenção de um evento transgênico é muito longo, e ter um indicativo prévio de funcionamento do sistema CRISPR é uma garantia. A possibilidade de utilização de um sistema onde se combina a tecnologia CRISPR com a metodologia de *hairy-root*, foi de grande ajuda na seleção prévia de um sistema eficiente de edição gênica, considerando tanto a estruturação vetorial quanto os alvos escolhidos para os genes escolhidos para esse estudo. Os resultados aqui obtidos indicaram que as duas configurações vetoriais utilizadas foram capazes de gerar edições nos genes *GmIPK1* e *GmIPK2*, sendo a resposta considerada gene-específica. Após as análises concluímos que a configuração tradicional (TCTU) deveria ser a escolhida para a obtenção das plantas transgênicas. Entretanto, se o objetivo desse trabalho fosse ter como alvo vários genes simultaneamente, a configuração STU também poderia ser utilizada para trabalhar-se com vetores menores e mais compactos. Concluiu-se também que todas as seis regiões alvo indicadas pelo programa Geneious foram adequadas para a edição dos genes alvos em soja.

CAPÍTULO II:

Plantas de soja editadas via CRISPR/*SpCas9*

2.1 INTRODUÇÃO - Um gene nocauteado por CRISPR

Nos últimos anos foi notável o aumento das aplicações da tecnologia CRISPR na edição de genes, principalmente em plantas modelos, em grande parte devido à sua maior simplicidade, se comparado com os outros métodos de edição existentes, e acessibilidade dessa tecnologia, que também oferece uma nova abordagem para os estudos de função gênica (Wang *et al.*, 2017).

O setor produtivo de soja tem inúmeras demandas, e essas têm sido, de certa forma, bem atendidas pelos programas de melhoramento. Entretanto outras características demandadas não têm sido obtidas nesses programas. Uma dessas abordagens é a obtenção de soja com reduzido teor de ácido fítico nas sementes. Plantas que apresentam redução de ácido fítico são conhecidas como *lpa* (*low phytic acid*). As primeiras gerações de culturas *lpa* foram obtidas pelo melhoramento clássico, entretanto, sabe-se que a introgressão de uma característica em uma cultivar comercial não é trivial (Goßner *et al.*, 2019). Desse modo, vislumbra-se o uso da engenharia genética no auxílio desse objetivo, incluindo o sistema CRISPR.

Na literatura apenas três trabalhos foram publicados usando CRISPR/Cas com a finalidade de gerar plantas com fenótipo *lpa*. O gene *ITPK1* teve sua expressão interrompida em *Oryza sativa*, o que resultou em redução de até 32% no conteúdo de ácido fítico, e um aumento de aproximadamente cinco vezes na concentração de fósforo inorgânico (Pi), porém essa alteração teve impactos no crescimento e reprodução das plantas de arroz obtidas (Jiang *et al.*, 2019). Esse mesmo gene (*ITPK1*) foi alvo de estudo em *Brassica napus*. A interrupção de sua expressão resultou em plantas com redução de ácido fítico e aumento na presença de fósforo livre com crescimento normal sem nenhuma alteração perceptível de fenótipo (Sashidhan *et al.*, 2020). Vlčko & Ohnoutková (2020) avaliaram o gene *HvITPK1* em cevada, e obtiveram mutantes com níveis alterados de fósforo nos grãos maduros, variando de 65% a 174% o aumento do conteúdo quando comparada com a linhagem selvagem.

Trabalhos prévios demonstraram que a redução na expressão dos genes *IPK1* ou *IPK2*, seja em mono ou dicotiledôneas, pode permitir a obtenção de plantas *lpa*, sem causar

alterações significativas no desenvolvimento dessas plantas (Stevenson-Paulik *et al.*, 2005; Ali *et al.*, 2013; Raboy *et al.*, 2014; Punjabi *et al.*, 2018B).

Alguns trabalhos descritos na literatura objetivando eventos *lpa* são listados na tabela 3. Apenas três desses trabalhos tiveram como alvo a cultura da soja, utilizando a metodologia de RNA interferente (estes serão discutidos no capítulo 3), os outros trabalhos em soja foram desenvolvendo linhagens *lpa* por mutações induzidas.

Dentre esses, Yuan *et al.* (2007) caracterizaram duas mutações induzidas por raios gama (γ -irradiation), em *G. max* usando as variedades *Taiwan 75 (Gm-lpa Tw-1)* e *Zhechun No3 (Zc-3) (Gm-lpa Zc-2)* para gerar um evento *lpa*, o que resultou em uma redução de aproximadamente 50% de ácido fítico e aumento de fósforo inorgânico. Essas duas linhagens foram utilizadas em outros trabalhos, como o desenvolvido por Vicent *et al.*, (2015) que utilizando a linhagem *Gm-lpa Zc-2* determinaram o efeito de fósforo nas sementes e induziu uma nova mutação, selecionando o evento que apresentou mutação no cromossomo 6. Após novas análises foram observados níveis ainda mais baixos de ácido fítico, e acúmulo moderado de fósforo inorgânico, contudo as sementes apresentaram perda na capacidade de germinação. Já Goßner *et al.* (2019) fizeram avaliações comparativas através do cruzamento (*cross-breeding*) das duas linhagens *Gm-lpa Tw-1* e *Gm-lpa Zc-2* para investigar alterações na redução do ácido fítico, quando as mutações estivessem presentes em uma mesma linhagem. Naquele momento já se sabia que as mutações estavam nos genes *MIPS* e *IPK1* respectivamente. Quando notada a dupla mutação, as sementes das plantas apresentavam uma drástica redução do conteúdo de ácido fítico quando comparada com linhagens com mutação simples. A redução foi de até 88% de ácido fítico e acúmulo moderado de InsP3, InsP4 e InsP5.

Fenótipos *lpa* foram descritos em outras espécies como em trigo (*Triticum aestivum* L.), onde mutações não letais foram identificadas, apresentando sementes com uma redução do nível de ácido fítico para 48,2%, de modo a diminuir a concentração de ácido fítico no farelo em 43% e aumentou a concentração de P inorgânico no farelo em quase quatro vezes (Guttieri *et al.*, 2004). Em milho, foi feito um silenciamento embrião específico, tendo como alvo o transportador MRP, isso gerou uma redução de 93% de ácido fítico, sem redução do peso dos grãos e sem problemas na germinação (Shi *et al.*, 2007).

Tabela 3: Trabalhos da literatura que visaram à obtenção de plantas *lpa*

Referência	Gene alvo	Estratégia	Planta	Observação
Yuan <i>et al.</i> , 2007	<i>IPK1</i>	Mutações induzidas	Soja	50% de redução de PA e aumento de Pi
Yuan <i>et al.</i> , 2009	<i>IPK1</i> <i>MIPS</i>	Mutações induzidas	Soja	46,3% de redução de PA 66,6% de redução de PA
Goßner <i>et al.</i> , 2019	<i>MIPS x IPK1</i>	<i>Cross-breeding</i>	Soja	79-88% redução de PA e acúmulo de Ins menores
Nunes <i>et al.</i> , 2006	<i>GmMIPS1</i>	RNAi	Soja	Aumento no fósforo disponível, redução de até 94,5% de PA, impactos negativos na germinação.
Ali <i>et al.</i> , 2013	<i>IPK1</i>	RNAi	Arroz	Promotor semente-específico Redução de PA e aumento de Pi
Aggarwall <i>et al.</i> , 2018	<i>IPK1</i>	RNAi	Trigo	Redução de 28% de PA, aumento de fósforo livre, redução no número de sementes por espiguetas.
Punjabi <i>et al.</i> , 2018B	<i>IPK2</i>	RNAi	Soja	Promotor específico de semente, 42 a 45% de redução de PA (variedade Pusa 16), aumento no conteúdo de minerais essenciais.
Kumar <i>et al.</i> , 2019	<i>GmMIPS1</i>	RNAi	Soja	Promotor semente-específico, 38,75 a 41,34% redução de fitato, sem impactos negativos
Jiang <i>et al.</i> , 2019	<i>ITPK</i>	CRISPR	Arroz	Redução de até 32% nos níveis de ácido fítico
Sashidhan <i>et al.</i> , 2020	<i>ITPK</i>	CRISPR	Canola	Redução de PA e aumento da concentração de fósforo livre
Vlčko & Ohnoutková 2020	<i>ITPK</i>	CRISPR	Cevada	Aumento de 65% a 174% do conteúdo fósforo

Já em feijão (*Phaseolus vulgaris*) uma linhagem mutada *lpa*, mostrou uma redução de 90% de ácido fítico quando comparada à linhagem selvagem, além de um aumento de cátions de ferro e uma melhor resposta na germinação (Campion *et al.*, 2009).

Mutações químicas em ervilha (*Pisum sativum* L) levaram a redução de 60% de fitato com um aumento de fósforo livre, porém as plantas apresentaram um tempo maior para floração e maturação, peso das sementes ligeiramente menor e consequente um menor rendimento (Warkentin *et al.*, 2012). Essas ervilhas *lpa* foram usadas em um estudo com frangos, as aves alimentadas com as ervilhas *lpa* apresentaram menor ganho de peso, no entanto a estrutura óssea aumentou. Além disso, foi observado que, uma vez que as ervilhas *lpa* apresentaram uma maior disponibilidade de fósforo, houve a redução da suplementação nutricional com fósforo, para atender as exigências nutricionais (Thacker, *et al.*, 2013).

Alguns trabalhos relataram como consequência, a redução da viabilidade das e morfologia indesejada nas sementes e maior susceptibilidade a patógenos como efeitos negativos gerados pela interrupção de algum gene da via do inositol-fosfato. Assim a possibilidade de utilização de uma técnica precisa e específica, como CRISPR, surge como uma excelente alternativa, para gerar um evento de soja que apresente fenótipo de redução de ácido fítico sem efeitos pleiotrópicos.

Desse modo, o trabalho descrito nesse capítulo teve como objetivo gerar plantas de soja com redução de ácido fítico através da deleção dos genes *GmIPK1* e *GmIPK2* via CRISPR/Cas9, avaliar o padrão de produção de ácido fítico nas sementes e observar o desenvolvimento das plantas, gerando uma linhagem *lpa* para ser inserida no programa de melhoramento. Utilizamos uma cultivar comercial da Embrapa (BRS 7980) o que pode ser considerado um passo à frente no caminho para a liberação de um produto. Além disso, a progênie selecionada não terá a presença do cassete transgênico, o que facilitará no processo de desregulamentação.

2.2 OBJETIVO GERAL

Desenvolver plantas de soja com edição dos genes *GmIPK1* e *GmIPK2* individualmente via sistema CRISPR/Cas9.

Objetivos específicos

- Obter plantas de soja com o gene *GmIPK1* editado via CRISPR/Cas9.
- Obter plantas de soja com o gene *GmIPK2* editado via CRISPR/Cas9.
- Avaliar o perfil de acúmulo de ácido fítico nas sementes das plantas editadas, e qualificar aquelas com o fenótipo desejado.

2.3 METODOLOGIA

Construção do vetor para transformação de soja por biobalística

Os cassetes CRISPR/Cas-TCTU para os dois genes, validados no capítulo I, contendo a sequência que codifica para a enzima Cas9 de *Streptococcus pyogenes* (SpCas9) sob o controle do promotor *GmUbi3*, e cada sgRNA (três no total), regulados pelo promotor *MtU6*, foram transferidos para o vetor pAHAS, adequado para transformação de soja por biobalística. O vetor possui o gene *ahas* (*Acetolactate synthase*) mutado de *A. thaliana* como gene marcador de seleção.

Obtenção das plantas transgênicas

As plantas transgênicas foram obtidas por bombardeamento de genes para as células do meristema apical de sementes maduras de soja, de acordo com o protocolo estabelecido por Rech *et al.* (2008). Após o bombardeamento os embriões ficaram 12 horas em meio de indução com BAP (*6-benzilaminopurina*) na concentração de 5µg/µL, em seguida foram transferidas para o meio de seleção com o agente seletivo *imazapyr* 500nM, para regeneração dos embriões. Após um mês as plantas desenvolvidas foram transferidas para meio de enraizamento com o hormônio IBA (*Indole-3-butyric acid*), e as plantas permaneceram neste meio até desenvolverem raízes sendo então aclimatadas e levadas para casa de vegetação, onde permaneceram até a produção das sementes.

Confirmação das plantas transgênicas e nocaute dos genes alvos

Plantas T₀ foram inicialmente avaliadas como positivas por PCR para a presença do gene codificante para Cas9 (SpCas9-F3: GGACAAGAAGTACTCCATTGG e SpCas9-R583: GCTGATTGTAAGTCTGAACC). Seguido pela avaliação da edição nos genes por sequenciamento da região dos alvos para cada gene (o fragmento gerado compreende os três alvos). Uma reação de PCR foi realizada com os *primers* desenhados, para o gene *GmIPK1*, IPK1-97F: ACACAATTCCTTTCCCACCA e IPK1-1399R: AGCAGAGGCTAGATCCTTGA (*amplicon* esperado de 1.322-pb), e para o gene *GmIPK2*, e para IPK2-80F: TTGCATTGCTTTGTGTAAGG e IPK2-1113R: CTGCGACTAATTCAAGCA (*amplicoon* esperado de 1.053-pb). Os fragmentos de PCR (*pool*) foram purificados e enviados para sequenciamento (tipo Sanger).

Os dados obtidos no sequenciamento Sanger foram avaliados pelo programa SYNTHEGO, que apresentou a probabilidade de inserções/deleções (*Indels*) nas amostras analisadas. Foram consideradas com potenciais candidatos amostras que apresentaram a probabilidade de *Indels* superior a 10% de edição.

Determinação dos teores de ácido fítico nas sementes de soja

As sementes geradas das plantas T_0 foram poucas e eram pequenas e não seria possível fazer as análises de quantificação de ácido fítico, já que o processo é destrutivo, assim algumas sementes foram germinadas para obtenção de mais sementes de modo a ser possível realizar as análises. As análises foram realizadas com a progênie T_2 e T_3 .

Para avaliação do teor de ácido fítico nas sementes, foi utilizado o kit *Phytic Acid Assay* (Megazyme[®], Bray, Irlanda), de acordo com as instruções do fabricante (https://www.megazyme.com/documents/Booklet/K-PHYT_DATA.pdf). Para esse ensaio colorimétrico, apenas foi adaptado a quantidade de material vegetal inicial, sendo reduzida proporcionalmente para 0,2g (T_2) ou 0,1g (T_3). As sementes de soja foram lixadas para obtenção do pó necessário para a realização do procedimento.

Na análise da T_3 duas condições foram avaliadas para cada amostras: Sementes com aspecto normal (N) e aspecto anormal/abortado (A) (deformidade no grão ou mal desenvolvimento do grão).

Após as análises foram obtidos dados de quantidade de fósforo total e de ácido fítico, esses então foram comparados com uma amostra controle (semente de soja BRS7980 não transformada). Para o cálculo dos valores, foi necessário estabelecer uma curva padrão de calibração, feita utilizando uma solução fósforo padrão para linearizar os resultados do ensaio colorimétrico. A leitura da absorbância (nm) foi realizada no comprimento de onda de 655nm.

Com os resultados obtidos foram seguidos 4 critérios para selecionar potenciais eventos: (1) Linhagens que apresentaram menor concentração de ácido fítico comparado ao controle; (2) Indivíduos dentro de uma mesma linhagem que apresentaram valores inferior à média da linhagem; (3) Alteração do valor de absorbância de fósforo livre ou (4) alteração na quantidade final de fósforo total.

2.4 RESULTADOS

Construção do vetor para biobalística e transformação genética

A configuração TCTU para CRISPR foi escolhida para dar continuidade ao trabalho de obtenção das plantas editadas, uma vez que apresentou uma maior eficiência de gerar edições tanto para os genes *GmIPK1* e *GmIPK2*. Os cassetes do CRISPR-TCTU (sgRNAs/Cas9) para *IPK1* e *IPK2* foram removidos dos respectivos vetores (p201-EGFP-C9) e clonados individualmente no vetor pAHAS gerando assim os vetores pAHAS-Cas9IPK1-TCTU e pAHAS-Cas9IPK2-TCTU (Figura 10).

Após confirmação da integridade dos vetores, seguiram-se os experimentos de biobalística. Foram realizados três experimentos com um total de 864 embriões bombardeados para o vetor pAHAS-Cas9IPK1 e dois experimentos com um total de 756 embriões bombardeados com o vetor pAHAS-Cas9IPK2. Desse total, 129 plantas crescidas em meio seletivo foram aclimatadas, sendo 59 vindas dos experimentos onde se utilizou o vetor para *IPK1* e 70 do vetor para *IPK2*.

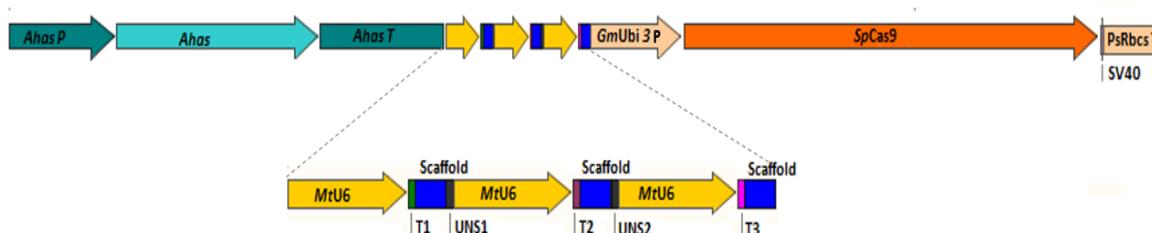


Figura 10: Esquema do vetor CRISPR/Cas9 para biobalística, sistema TCTU. Estrutura base do vetor pAHAS-Cas9-TCTU para nocaute dos genes *GmIPK1* e *GmIPK2*. O cassete do gene de seleção é composto pelo gene *ahas* mutado de *A. thaliana* (azul claro) sendo regulado pelo promotor e terminador *ahas* (azul escuro). A endonuclease Cas 9 de *Streptococcus pyogenes* (*SpCas9*) (laranja) está sob o controle do promotor Ubiquitin-3 de *Glycine max* e terminador *Rbcs* de *Pisum sativum* (laranja claro). Na parte amplificada é apresentado os sgRNA, que são regulados pelo promotor U6 de *Medicago truncatula* (amarelo). Os alvos (verde T1, roxo T2, e rosa T3) são seguidos pela sequência *Scaffold* (azul). Os sgRNA são separados por uma sequência UNS (preto) O cassete CRISPR/sgRNAs possui 11.347bp e o cassete *ahas* 4.296bp.

Análises do sequenciamento das plantas T₀ obtidas

Embora algumas plantas não tenham se desenvolvido após serem transferidas para casa de vegetação, as sobreviventes foram avaliadas pela presença da Cas9 por PCR (dados não apresentados) e em seguida os *amplicons* foram enviadas para sequenciamento (Sanger). As sequências dos genes obtidas após o sequenciamento dos fragmentos de PCR

das amostras analisadas não detectaram as edições ocorridas. No entanto, os dados do sequenciamento foram avaliados no programa SYNTHEGO e das 28 amostras analisadas para o gene *GmIPK1*, 10 indicaram presença de *Indels*. Enquanto para o *GmIPK2* das 35 amostras analisadas, nove apresentaram alguma porcentagem de *Indels*. Como pode ser observado na figura 11-A para o *IPK1*, obtivemos eficiência da edição (presença de *Indels*) entre 2% a 60%, enquanto para o *IPK2*, a variação na eficiência de edição foi de 5% a 77% (Figura 11-B). O programa SYNTHEGO indica a probabilidade de edições presentes na amostra analisada.

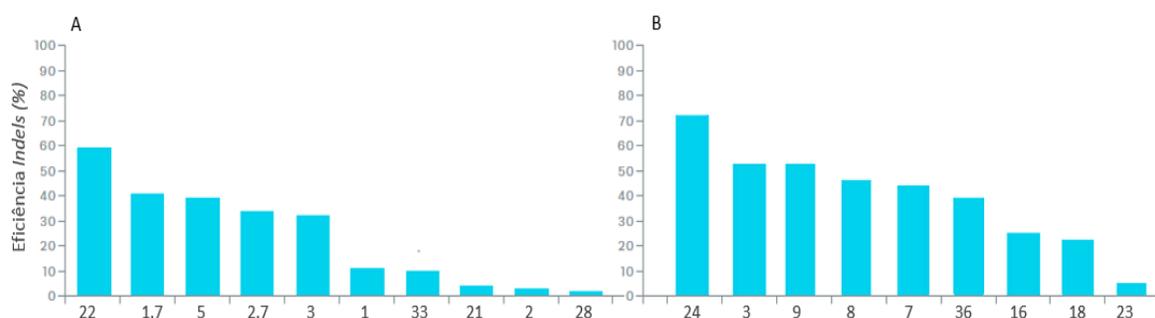


Figura 11: Representação gráfica dos resultados obtidos após análise dos sequenciamentos das amostras de DNA-genômico das plantas transformadas com vetores CRISPR-Cas9-TCTU. O eixo X apresenta as linhagens avaliadas, e o eixo Y a porcentagem de edição/Indels (Inserção e deleção) presente na amostra, para os genes *IPK1* (A) e *IPK2* (B).

Análise do conteúdo de ácido fítico

Uma vez que a maioria das plantas T₀ produziram poucas sementes e/ou com tamanho pequeno (efeito normal para plantas oriundas de cultura *in vitro*), não foi viável a realização das análises nessa fase. Assim foram escolhidos cinco eventos para avançar à próxima geração (T₂). A escolha das linhagens foram determinadas pela porcentagem de *Indel* apresentadas na análise do programa SYNTHEGO, sendo escolhidas linhagens que apresentaram as maiores porcentagem e as menores porcentagens (Figura 11-A e 11-B). As quantidades de sementes germinadas variaram conforme a disponibilidade para cada evento. Foram escolhidas as linhagens #1, #2, #5, #1.7, #2.7 para o *IPK1* e as linhagens #7, #9, #16, #23 e #24 para *IPK2*.

Com o auxílio do kit *Phytic Acid Assay* (Megazyme[®]) para quantificação do nível de ácido fítico, foram realizadas análises com sementes da geração T₂ das linhagens anteriormente selecionadas. Foram necessárias de três a quatro sementes para obter 0,2g de

sementes em pó necessárias para realização das avaliações. Os resultados podem ser visualizados na tabela 4.

Tabela 4: Conteúdo de ácido fítico nas sementes das plantas IPK1 e IPK2 CRISPR T₂.

Amostras	Fósforo livre (nm)	Fósforo total (nm)	Fósforo (g/100 g)	Ácido fítico (g/100 g)	Média
IPK1					
1A	0,169	1,491	0,523	1,853	
1B	0,169	1,779	0,636	2,257	
1C	0,145	1,835	0,668	2,369	
1D	0,157	1,548	0,550	1,950	
1E	0,165	1,598	0,566	2,008	2,087
2A	0,124	1,603	0,584	2,072	
2B	0,123	1,671	0,612	2,169	
2C	0,103	1,202	0,434	1,541	
2D	0,137	1,634	0,591	2,097	
2E	0,262	1,352	0,431	1,528	
2F	0,115	1,426	0,518	1,838	1,874
5A	0,102	1,434	0,527	1,867	
5B	0,160	1,951	0,708	2,510	
5C	0,130	1,757	0,643	2,280	
5D	0,269	1,696	0,564	2,000	2,164
1.7A	0,120	1,365	0,492	1,744	1,744
2.7A	0,137	1,497	0,537	1,905	1,905
IPK2					
9A	0,139	0,816	0,669	2,373	
9B	0,133	0,722	0,582	2,062	2,218
16A	0,050	0,725	0,667	2,365	
16B	0,056	0,741	0,677	2,401	2,383
23A	0,115	0,740	0,617	2,190	
23B	0,127	0,816	0,680	2,413	
23C	0,129	0,748	0,612	2,171	
23D	0,123	0,792	0,661	2,345	
23E	0,145	0,775	0,622	2,207	
23F	0,136	0,852	0,707	2,507	2,305
24A	0,044	0,725	0,673	2,386	
24B	0,041	0,687	0,639	2,264	
24C	0,044	0,663	0,612	2,169	
24D	0,042	0,722	0,672	2,384	
24E	0,059	0,718	0,652	2,311	
24F	0,049	0,738	0,680	2,412	2,321
BRS7980	0,029	0,632	0,596	2,113	
BRS7980	0,105	0,614	0,503	1,783	1,948

Descrição: Valor de ácido fítico inferior ao controle BRS7980 (verde); indivíduo com valor inferior à média da linhagem (rosa); Alteração da absorbância de fósforo livre (variação de azul) alteração na quantidade final de fósforo (laranja)

É possível observar que em sua maioria as amostras analisadas apresentaram valores numericamente elevado de ácido fítico, maior até que a planta BRS7980 controle não transgênica. Sendo as linhagens 2C e 2E de IPK1 a exceção, apresentando concentrações de ácido fítico menor quando comparado ao controle. O resultado geral foi confuso, de modo que se seguiu os quatro critérios para selecionar as linhagens para uma nova análise de quantificação do ácido fítico. (1) Linhagens que apresentaram menor concentração de ácido fítico comparado ao controle. Visto que se trata da progênie T₂ e ainda há segregação gênica, (2) foi feita uma média entre os indivíduos de uma mesma linhagem, e destacar aqueles que apresentaram valores menores que a média (destacados em rosa). Quando se obtém uma característica *lpa*, é esperado um aumento na concentração de fósforo livre, por conta da redução da molécula do ácido fítico. Sendo assim algumas linhagens se destacaram (3) pelo aumento na absorvância de fósforo livre (destacados em azul) ou (4) alteração na quantidade final de fósforo (destacado em laranja).

As linhagens que se destacaram por alguma característica discrepante em relação aos critérios pré estabelecidos foram selecionadas e avançadas à próxima geração. Sendo assim as linhagens #1A, #1C, #2C, #2E, #5A, #5C, #5D e #1.7A para IPK1, e as linhagens #9A (foi usado a linhagem A, pois a B não tinha mais sementes), #23A, #23E, #24B, #24C, #24E para o IPK2, foram escolhidas para obtenção da progênie T₃. Quatro sementes de cada linhagem foram germinadas.

Nessa nova geração as linhagens produziram sementes em maior proporção com aspecto normal para um grão de soja, sementes com desenvolvimento anormal (pequenas e deformadas) ou que sofreram aborto (desenvolvimento interrompido com a semente ainda no estágio inicial). Com isso foram feitas duas análises na T₃, uma representando sementes abortadas ou com desenvolvimento anormal (A) e outra representando sementes normais (N). Para esta análise foi utilizado 0,1g de extrato de pó de sementes, o que representou entre uma ou duas sementes normais.

As sementes T₃ para IPK1 e IPK2 foram analisadas, e os resultados estão apresentados nas tabelas 5 e 6 respectivamente, apenas para as amostras identificadas como normal (N). As linhagens analisadas que não apresenta a nomenclatura N, não foram analisadas na condição abortada/desenvolvimento anormal (A).

Como pode ser observado, os resultados obtidos para o gene *IPK1* não foram expressivos, poucas linhagens apresentaram valores menores no conteúdo de ácido fítico em comparação com a média de todas as outras amostras (1C-4, 5A-3, 5D-1, 5D-2, 7A-3, 7A-4). A linhagem 1C-3 chama a atenção, uma vez que apresentou a absorvância de fósforo livre elevada e que resultou em valores numericamente alterados na quantidade de ácido fítico.

Tabela 5: Conteúdo de ácido fítico nas sementes das plantas IPK1-CRISPR T₃

Amostras	Fósforo livre (nm)	Fósforo total (nm)	Fósforo (g/100 g)	Ácido fítico (g/100 g)
1A1 N	0,159	1,143	0,647	2,294
1A2 N	0,127	1,189	0,698	2,477
1A3 N	0,140	1,152	0,665	2,359
1C1 N	0,132	1,135	0,660	2,340
1C3 N	0,180	1,686	0,990	3,510
1C4 N	0,136	1,069	0,613	2,174
1F1	0,166	1,210	0,687	2,436
2C1 N	0,137	1,166	0,677	2,400
2C2 N	0,142	1,238	0,720	2,554
2C3 N	0,152	1,312	0,763	2,704
2E1 N	0,142	1,130	0,649	2,302
5A1 N	0,137	1,177	0,684	2,425
5A2 N	0,138	1,119	0,645	2,288
5A3	0,183	1,118	0,614	2,179
5C1	0,144	1,215	0,704	2,498
5C2 N	0,151	1,148	0,655	2,323
5C3	0,131	1,075	0,620	2,200
5D1 N	0,132	1,062	0,611	2,167
5D2 N	0,157	1,096	0,618	2,191
7A2	0,139	1,135	0,655	2,321
7A3	0,117	1,020	0,594	2,106
7A4	0,123	1,076	0,626	2,221
BRS7980	0,1182	1,0078	0,5848	2,0739

Descrição: Indivíduo com valor inferior à média da linhagem (rosa); Alteração da absorvância de fósforo livre (variação de azul); Resultado discrepante (amarelo).

Com as plantas com o gene *GmIPK2* editado (Tabela 6), os indivíduos 24B-1, 24B-2, 24C-4 e 24C-4 apresentaram redução no conteúdo de ácido fítico em comparação ao controle BRS7980. Algumas outras amostras se destacaram quanto ao aumento na absorvância de fósforo livre (destacadas em azul).

Para todas as amostras analisadas na condição de abortada ou desenvolvimento anormal (A) na análise da T₃ para as linhagens IPK1 e IPK2, foram obtidos valores muito

abaixo do observado no controle BRS7980, de modo a não ser foi possível correlacionar a redução do conteúdo de ácido fítico com a deleção do gene (dados não apresentados).

Tabela 6: Conteúdo de ácido fítico nas sementes das plantas IPK2-CRISPR T₃

Amostras	Fósforo livre (nm)	Fósforo total (nm)	Fósforo (g/100 g)	Ácido fítico (g/100 g)
9A1 N	0,134	1,115	0,645	2,286
9A2 N	0,161	1,417	0,825	2,927
23A1 N	0,171	1,167	0,654	2,321
23A2 N	0,175	1,367	0,784	2,779
23A3 N	0,135	1,223	0,715	2,537
23C1 N	0,134	1,115	0,645	2,286
23C2 N	0,161	1,417	0,825	2,926
23E1 N	0,171	1,167	0,654	2,321
23E2 N	0,175	1,367	0,784	2,779
24C2 N	0,141	1,347	0,777	2,754
24C3 N	0,112	0,979	0,558	1,980
24C4 N	0,114	0,971	0,552	1,959
24B1 N	0,129	1,019	0,573	2,032
24B2 N	0,141	1,030	0,572	2,029
24B3 N	0,145	1,125	0,632	2,239
24B4 N	0,137	1,177	0,670	2,375
24E1 N	0,155	1,110	0,615	2,180
24E2 N	0,151	1,128	0,629	2,231
24E3 N	0,164	1,349	0,763	2,707
24E4 N	0,136	1,150	0,653	2,316
BRS7980	0,1246	1,0562	0,6073	2,1534

Descrição: Valor de ácido fítico inferior ao controle BRS7980 (verde); indivíduo com valor inferior à média da linhagem (rosa); Alteração da absorvância de fósforo livre (variação de azul).

Devido à grande quantidade de amostras que foram analisadas na geração T₃, onde foram analisadas todas as linhagens (IPK1 e IPK2) nas duas condições: abortada (A) e normal (N) e levando em conta que o ensaio colorimétrico pode sofrer alterações na coloração dependendo de tempo. Decidiu-se por refazer o experimento apenas com as linhagens que se destacaram em algum dos critérios estabelecidos anteriormente. Com essa análise (Tabela 7) concluir-se que para o gene *GmIPK1*, duas linhagens são eventos em potenciais, a linhagem 5D-2 que apresentou uma redução de 11,56% no conteúdo de ácido fítico em comparação ao controle, e a linhagem 1C-3 que apresentou um aumento na absorvância de fósforo livre e uma elevada quantidade de ácido fítico. Para o gene *GmIPK2*, duas linhagens 24B-2 e 24C-4 apresentaram redução de 14,11% e 25,35% do conteúdo de ácido fítico quando comparadas ao controle BRS7980. Enquanto que a linhagem 23E-2 apresentou uma absorvância elevada de fósforo livre.

Quanto à alteração no fenótipo para as plantas analisadas na tabela 7 para o gene *IPK1* todas apresentaram desenvolvimento visualmente, no entanto houve redução no tamanho/peso dos grãos para quase todas as linhagens. Já para o gene *IPK2*, as linhagens 23C-2, 23E-1 e 23E-2 apresentaram redução significativa do tamanho/peso dos grãos. Além disso, a linhagem 23E-2 (Abs. do fósforo livre elevada) apresentou deiscência precoce de folhas e com isso as vagens não completaram seu desenvolvimento, de modo que produziu apenas 24 sementes sendo 9 delas com aspecto de abortamento e o tamanho/peso bem reduzido. As demais linhagens apresentaram desenvolvimento normal. A linhagem 24C-4 que apresentou a maior porcentagem de redução de ácido fítico (25,35%) não apresentou redução no tamanho/massa (peso dos grãos), seguido pela linhagem 24B-2 (14,11%) que houve uma pequena diminuição.

Tabela 7: Conteúdo de ácido fítico nas sementes das plantas destacadas IPK2-CRISPR T₃

Amostras	Fósforo livre (nm)	Fósforo total (nm)	Fósforo (g/100 g)	Ácido fítico (g/100 g)	% redução	Massa grãos (100 sementes)
IPK1						
1 C3	0,184	1,866	1,064	3,774	-55,27	12,46g
1 C4	0,128	1,114	0,624	2,214	8,91	12,84g
5 A3	0,154	1,158	0,635	2,253	7,28	9g*
5 D1	0,125	1,130	0,637	2,257	7,13	13,2g
5 D2	0,116	1,073	0,606	2,149	11,56	10,92g
7 A3	0,128	1,150	0,647	2,295	5,56	20,0g
IPK2						
23 C2	0,096	1,110	0,642	2,276	6,36	9,38g*
23 E1	0,137	1,211	0,680	2,410	0,85	8,20g*
23 E2	0,164	1,170	0,637	2,258	7,10	7,20g*
24 B1	0,123	1,167	0,661	2,342	3,62	21,02g
24 B2	0,138	1,068	0,589	2,087	14,11	14,66g
24 C3	0,097	1,139	0,659	2,337	3,83	20,54g
24 C4	0,104	0,912	0,512	1,814	25,35	19,8g
BRS7980	0,1347	1,2175	0,6854	2,4304	-	18,4g

Descrição: Valor de ácido fítico inferior ao controle BRS7980 (verde); Alteração da absorvância de fósforo livre (variação de azul) (*Não apresentava a quantidade de sementes suficientes, foi feita uma estimativa).

2.5 DISCUSSÃO

O aperfeiçoamento da tecnologia CRISPR nos últimos anos vem tornando disponível a possibilidade do melhoramento genético em áreas de grande importância, tais como a produção de alimentos agronomicamente melhorados ou com aplicação direta na saúde humana. Embora muitos trabalhos ainda estejam em fase inicial, é evidente a diversidade de estudos que vêm sendo realizados.

Considerando a importância da cultura de soja para o Brasil e para o mundo, vê-se como uma possibilidade muito promissora o desenvolvimento de cultivares de soja modificada, com a finalidade de ressaltar características de interesses agrônômicos, tais como: aumento de proteínas, produção de óleos específicos, aumento do valor nutricional, ou tolerância aos estresses bióticos e abióticos. O desenvolvimento de potenciais produtos utilizando essa ferramenta além de facilitar a tramitação nos processos de biossegurança para desregulamentação do produto, ainda pode permitir uma melhor percepção pública quanto aos produtos geneticamente modificados.

No Brasil até o presente momento, o único produto analisado pela CTNBIO gerado pelo sistema CRISPR foi o milho ceroso, que apresenta maior teor de amilopectina (Qi *et al.*, 2020), em contrapartida a agência de regulamentação norte-americana USDA, dentre os diversos produtos já analisados, há duas variedades de soja, para tolerância a seca e com alteração do tamanho do pecíolo. Isso abre precedentes para que no futuro mais produtos sejam analisados e liberados. Dentre as diversas vertentes para utilização de CRISPR, um dos focos, é o melhoramento da qualidade de culturas destinadas à alimentação (Sedeek *et al.*, 2019). A redução de fatores antinutricionais, como o fitato, pode ser considerada um dos principais alvos para produção das chamadas *lpa crops*.

Nesse capítulo objetivamos o desenvolvimento de um evento de soja que apresentasse redução do teor de ácido fítico, através do nocaute de dois genes, que desempenham papéis importantes na via do inositol-fosfato. O gene *IPK2* embora não tenha sua rota principal em sementes, apresenta uma expressão elevada neste tecido, enquanto o gene *IPK1* é a enzima final responsável pela produção do ácido fítico. Ambos os genes estão envolvidos nas rotas de biossíntese de ácido fítico, e a interrupção de sua expressão resulta em eventos com reduzidos teores de ácido fítico (Tabela 3). Neste estudo

foi possível observar uma aparente redução de até 26% do conteúdo de ácido fítico após o nocaute individual dos genes estudados, com destaque para as linhagens IPK1-5D2 (redução de 11,6%), IPK2-24B2 (14,1%) e IPK2-24C4 (25,3%). As duas linhagens referentes ao gene *IPK2* não apresentaram alterações visíveis durante o desenvolvimento da planta, redução de produtividade ou no tamanho dos grãos. Marathe *et al.* (2018) acreditam que uma redução de 4% no teor de ácido fítico em sementes de soja seria suficiente para o aumento do valor nutricional das sementes, visto que muitas leguminosas como a soja acumulam mais fósforo do que o necessário para as atividades celulares.

O gene *IPK1*, foi alvo de estudo em duas espécies, arroz (Ali *et al.*, 2013) e trigo (Aggarwall *et al.*, 2018), utilizando a metodologia de RNA interferente, e em soja com mutações induzidas (Yuang *et al.*, 2007 e Yuang *et al.*, 2009). Os trabalhos com RNAi apresentaram resultados divergentes, onde em arroz foi um utilizado promotor semente-específico, isso resultou na redução dos níveis de ácido fítico e aumento de fósforo inorgânico e acúmulo de ferro no endosperma sem efeitos negativos para a cultura, enquanto que em trigo, utilizando o promotor constitutivo CaMV35S, houve redução entre 28 a 56% do conteúdo de ácido fítico, porém as linhagens apresentaram redução do número de sementes por espiguetas. O gene *IPK2* foi alvo de um único estudo usado RNAi em cultura de soja (Punjabi *et al.*, 2018B), por isso os resultados serão melhores discutidos no capítulo 3 (RNAi é o foco do capítulo 3). Mas é válido ressaltar que o silenciamento desse gene resultou em sementes de soja com redução do teor de ácido fítico, sem alteração fenotípica nas sementes.

Usando CRISPR, os genes alvos deste estudo não foram avaliados anteriormente. Outro gene da via do inositol-fosfato, o *ITPK6*, foi nocauteado em arroz, o que gerou uma redução de 32% do conteúdo de ácido fítico, com níveis elevados de fósforo inorgânico quando comparados a uma linhagem controle. Entretanto essa alteração prejudicou significativamente o crescimento e reprodução das plantas (Jiang *et al.*, 2019). Em contrapartida, Sashidhan *et al.* (2020) mutagenisaram três parálogos do gene *ITPK* de *Brassica napus*, obtendo redução do conteúdo de ácido fítico e aumento da concentração de fósforo livre, sem apresentar efeitos negativos para as plantas.

Jiang *et al.* (2019) sugerem que seja possível a obtenção de um evento *lpa* sem apresentar efeitos pleiotrópicos, com a otimização do design do vetor ou através da identificação e análises de novas mutações. Vlčko & Ohnoutková (2020) interromperam a expressão do gene *OsITPK* por CRISPR, o padrão de edição apresentou inserção ou deleção de bases. Aquelas linhagens que apresentaram inserções, apresentaram maior tolerância a estresse salino quando comparado às linhagens que apresentaram deleções. Dando assim indícios que a natureza da mutação pode ser um fator a ser considerado quando correlacionar a mutação com o fenótipo obtido.

Alterações que causem perturbações no final da via tem como resposta a diminuição do conteúdo de ácido fítico, seguido por aumento do fósforo livre e aumento do conteúdo de InsPs menores (InsP3 até InsP5) (Sparvoli & Cominelli, 2015). E segundo Raboy (2009) o conteúdo de fósforo total não altera, mesmo quando há um evento *lpa*, o que muda é a forma como o fósforo está estocado (P celular, P inorgânico ou ácido fítico). A linhagem IPK1-1C3 obtida nesse trabalho precisa ser mais bem estudada, pois apresentou o valor da absorvância muito elevada para fósforo livre e fósforo total, e isso resultou em uma quantidade elevada no conteúdo de ácido fítico. A linhagem IPK2-23E2 também apresentou o valor da absorvância de fósforo livre elevada, no entanto não houve alteração na absorvância de fósforo total, de modo a não resultar em um elevado teor de ácido fítico, como observado na IPK1-1C3. Novos estudos devem ser feitos para avaliar esses resultados contrastantes.

A análise do conteúdo de ácido fítico pelo kit da Megazyme nas sementes das gerações T₂ e T₃ apresentaram variações nos valores obtidos entre as plantas editadas e o controle BRS7980, o que dificultou a avaliação direta das amostras. Na análise das linhagens tendo como alvo o gene *GmIPK1* apresentada na tabela 5, em quase todas as amostras a quantidade de ácido fítico estava entre 2,1 e 3,5g/100mL, enquanto o controle apresentou 2,0g/100mL. Esse mesmo resultado foi observado para as amostras das linhagens tendo como alvo o gene *GmIPK2* (Tabela 6), nessa análise, com exceção de quatro amostras, as demais apresentaram uma variação do conteúdo de ácido fítico entre 2,1 e 2,9 g/mL, e em contra partida o controle BRS7980 apresentava apenas 2,1 g/100mL de ácido fítico. Essa variação poderia ser por conta da heterozigosidade ainda existentes

nestas gerações. Uma vez que para a obtenção do pó necessário para realizar o protocolo de quantificação de ácido fítico, foram usadas duas ou mais sementes, e por conta da segregação poderiam ter sementes positivas e negativas, essa condição talvez justifique o fato dos valores de ácido fítico obtidos estarem iguais ou superiores ao controle. A mesma variação foi observada nos resultados dos indivíduos de uma mesma linhagem, como exemplo a linhagem 24C que apresentou variações entre 1,95 e 2.7 g/100mL para as amostras analisadas (Tabela 6). Novas análises poderão ser feitas para uma maior acuraria dos dados.

Análises para melhor compreensão da natureza das mutações causadas pela ação do CRISPR nas plantas de soja ainda serão feitas, no entanto, é possível sugerir que ambos os genes quando nocauteados foram capazes de reduzir o conteúdo de ácido fítico. Com as análises realizadas até agora, podemos dizer que deleções no gene *GmIPK2* pode ter gerado uma linhagem (24-C4) com redução do teor de ácido fítico sem alterações visíveis no desenvolvimento da planta e no tamanho dos grãos, já para *GmIPK1* foi evidente que a maioria das plantas analisadas apresentaram redução no tamanho/peso das sementes, no entanto ainda não podemos afirmar a razão para esse fenótipo, seja ele a redução do conteúdo de ácido fítico ou a natureza da mutação.

2.5 CONSIDERAÇÕES

O estudo apresentado neste capítulo foi de grande relevância, por confirmar que os genes *GmIPK1* e *GmIPK2* estão mesmo envolvidos na via de síntese do ácido fítico em soja, e que a interrupção de sua expressão permite gerar plantas com fenótipo *lpa* sem aparente alteração das características agronômicas da cultura (germinação, desenvolvimento e produtividade).

A metodologia CRISPR anteriormente avaliada apenas em raízes de soja, se mostrou eficiente quando usada na geração de um evento de soja transformado, reforçando ainda mais o potencial dessa técnica. A resolução normativa N°16 aprovada em 2018 pelo CTNBio, afirma que os produtos gerados por quaisquer técnicas inovadoras de melhoramento por precisão serão avaliados caso-a-caso se serão considerados convencional (não OGM) ou um organismo transgênico (OGM). Sendo aqueles com ausência de DNA/RNA recombinante na progênie sejam considerados convencional e podem ser registrados sem a necessidade do processo de desregulamentação (Entine *et al.*, 2021). Isso reforça a utilidade do uso da metodologia CRISPR para obtenção de eventos transformados.

Ainda serão necessários estudos para compreender a natureza das mutações e ainda avaliar a disponibilidade de nutrientes que provavelmente se tornaram disponível depois da redução do fator antinutricional, fitato. De modo a aumentar o valor nutricional das sementes de soja *lpa*.

Uma nova perspectiva para as plantas geradas neste estudo, foi apresentada por Jiang *et al.* (2020), que após a deleção do gene *IPK1* em arroz por CRISPR, essas plantas mostraram um aumento na tolerância a estresse salino e déficit hídrico. O papel da via do inositol-fosfato ainda é pouco conhecido quanto à resposta aos estresses abióticos, entretanto, literaturas recentes de Jiang *et al.* (2020) e Vlčko & Ohnoutková (2020) apresentaram trabalhos onde a deleção de um gene relacionado a via do inositol-fosfato resultou em resposta positiva à tolerância aos estresses abióticos. Surgindo assim novas perspectivas para os eventos gerados neste estudo.

CAPÍTULO III:

Silenciamento do gene *GmIPK2* em soja mediado por RNA interferente

3.1 INTRODUÇÃO – Regulação gênica mediada por RNAi

Embora a tecnologia de RNA interferente tenha sido apresentada como uma ferramenta biotecnológica há bastante tempo, ainda continua sendo muito utilizada (Punjabi *et al.*, 2018B; Gao *et al.*, 2019; Kumar *et al.*, 2019; Luan *et al.*, 2020). A interferência mediada por RNA foi primeiramente identificada em plantas e nomeada *post-transcriptional gene silencing* (PTGS) (Lindbo, 2012). Essa foi identificada como um mecanismo natural para regular a expressão gênica, atuando como uma resposta de defesa contra invasores. De modo que a ativação desse sistema ocorre pela presença de RNAs dupla-fita (dsRNA) normalmente de organismos invasores (Jagtap *et al.*, 2011).

Na natureza dsRNAs longos (vindo dos invasores) são processados por uma enzima da família *RNAse* III conhecida como *Dicer*, que corta a cadeia dos longos RNAs em pequenos RNAs (20-24 nucleotídeos) denominados *small interfering* RNAs ou siRNA, que por sua vez são reconhecidos e incorporados dentro de um complexo de proteínas *RNA-induced silencing complex* (RISC). Dentro desse complexo ocorre a degradação de uma das fitas do dsRNA, mantendo a outra como molde para o reconhecimento das sequências complementares nos invasores. Após reconhecimento ocorre a interrupção da tradução do RNA mensageiro (mRNA), o que resulta no silenciamento do gene (Han, 2018).

Com o intuito do uso desse sistema na engenharia genética, notou-se que a expressão de pequenos RNAs em forma de *short hairpin* RNA (shRNA) entregues à célula por meio de vetores, poderia induzir a ativação do sistema RNAi (Jagtap *et al.*, 2011). Assim surge o RNAi, como uma ferramenta biotecnológica e com importantes aplicações práticas para melhoramento genético de plantas através de um silenciamento gene-específico.

Como observado na tabela 3, plantas *lpa* obtidas utilizando a metodologia de RNAi são predominantes. Um dos trabalhos pioneiros em soja com o objetivo de gerar planta *lpa* com uso de ferramentas biotecnológicas, foi o de Nunes *et al.* (2006) que geraram uma linhagem de soja com redução do teor de fitato, a partir da redução da expressão do gene *GmMIPS* (*myo-inositol-1-phosphate synthase*). Embora o teor de fitato tenha sido reduzido em até 94,5%, as plantas obtidas apresentaram alteração no desenvolvimento das sementes, não sendo capazes de se desenvolverem. Foi proposto que esse gene podia não ser um bom

candidato para gerar o fenótipo *lpa* em soja. Entretanto Kumar *et al.* (2019) obtiveram plantas de soja *lpa* usando um promotor específico para sementes, tendo como alvo o mesmo gene, *GmMIPS*, o que resultou em uma redução de até 41,34% do ácido fítico, além do aumento do conteúdo de Fe^{2+} (21,7%), Zn^{2+} (11,15%) e Ca^{2+} (15,35%), sem efeitos negativos para as plantas. Outro trabalho foi apresentado por Punjabi *et al.* (2018B) que silenciaram o gene *GmIPK2*, utilizando promotor semente-específico e obtiveram plantas de soja, com redução nos níveis de ácido fítico de até 45%, acumulação moderada de fósforos inorgânicos e um elevado conteúdo de minerais essenciais.

Em outras culturas também já foram obtidos o fenótipo *lpa* por intermédio do RNAi, como em arroz, onde a utilização de um promotor específico de semente (Oleoin18) promoveu o silenciamento do gene *OsIPK1*. As análises indicaram uma regulação do gene 3,8 vezes menor (*down-regulation*) quando comparada ao *wild-type*, isso foi correlacionado com a diminuição de ácido fítico e um aumento concomitante de fósforo inorgânico. Os grãos de arroz apresentaram um aumento de 1,8 vezes de ferro no endosperma, sem efeito negativo observado tanto na germinação quanto em qualquer outro desempenho agrônômico (Ali *et al.*, 2013). Esse mesmo gene foi *down-regulated* em trigo, *TaIPK1*, e as plantas apresentaram uma redução de 28% de ácido fítico, com um aumento de fósforo livre nos grãos maduros e aumento significativo na disponibilidade dos cátions ferro e zinco, porém houve uma redução no número de sementes por espiga (Aggarwall *et al.*, 2018).

Alguns trabalhos obtiveram plantas *lpa* utilizando RNA interferente, sem alteração negativa das características da planta. Uma possível justificativa seja que o uso do RNA interferente pode apresentar diferentes níveis de expressão do gene alvo. E entendendo a importância da via do inositol-fosfato, este capítulo tem por objetivo discutir o silenciamento do gene *GmIPK2* via RNAi, como meio para obtenção de uma planta geneticamente modificada que apresente uma redução de ácido fítico em sementes.

3.2 OBJETIVO GERAL

Desenvolver plantas de soja transgênicas com redução da expressão do gene *GmIPK2* mediado por RNA interferente.

Objetivos específicos

- Construir um vetor para regulação negativa do gene *IPK2* utilizando a metodologia RNAi.
- Obter plantas transgênicas de soja via biobalística utilizando o vetor anteriormente gerado pAHAS-GUS-RNAi_IPK2.
- Avaliar o perfil de acúmulo de ácido fítico nas sementes das plantas transgênicas obtidas.
- Quantificar o nível de expressão do gene *GmIPK2* em sementes e folhas das plantas transformadas por qRT-PCR.
- Avaliar o desenvolvimento das progênies e selecionar na progênie segregante aquelas que apresentaram redução de ácido fítico e seu efeito sobre características agronômicas da planta.

3.3 METODOLOGIA

Construção do vetor de RNAi para silenciamento do gene *IPK2*

O RNA total a partir 100mg de extrato foliar de soja (linhagem BRS7980) foi isolado com o auxílio do Kit *RNAeasy* (Qiagen®). Após a integridade do RNA ser confirmada em gel de agarose foi realizada a reação com *transcriptase reversa* para obtenção do cDNA (utilizando *primers* oligo dT), seguida por uma reação de PCR. Um fragmento de 326 pb referente à parte do gene *GmIPK2* (acesso: Glyma.12G240900) foi obtido, utilizando os primers *IPK2_RNAi-F*: ctctagactcgagGACTACTCCATCCGCTCCTT e *IPK2_RNAi-R*: ggatccggtaccGCGGATAGACTGGAGAGA contendo na extremidade dos *primers* (apresentados em minúsculo), sequências referentes às enzimas *XbaI* e *XhoI* (*forward*) e *BamHI* e *KpnI* (*reverse*) que foram usados para a clonagem do fragmento no vetor pKANNIBAL (Wesley et al., 2001).

Anteriormente os fragmentos *GmIPK2* foram clonados do vetor pGEM-T-*easy* e em seguida digerido com as enzimas *XbaI-BamHI* e *XhoI-KpnI*, criando assim as extremidades necessárias para clonagem do fragmento nos sentidos *sense* e *anti-sense*. O cassete do RNAi está sob regulação do promotor CaMV35S (*Cauliflower mosaic virus*) e terminador do gene *ocs* (*Octopine synthase*) de *A. tumefaciens*, além de possuir o íntron do gene *PDK* (*Pyruvate dehydrogenase kinase*), que facilitará a formação da estrutura no formato *hairpin*. Após clonagem dos fragmentos, o cassete de RNAi completo foi transferido para o vetor pAHAS-GUS, que além do gene de seleção *ahas* mutado de *A. thaliana* que confere resistência ao herbicida sistêmico *Imazapyr*, possui o cassete do gene repórter GUS regulado pelo promotor 35S e o enhancer AMV (*alfalfa mosaic virus*). O vetor gerado foi nomeado pAHAS-GUS-RNAi_IPK2 e contém 13.355pb.

Transformação de plantas de soja via biobalística

As plantas transgênicas de soja foram obtidas utilizando o sistema biobalístico de acordo com metodologia desenvolvida por Rech *et al.* (2008), por bombardeamento do vetor pAHAS-GUS-RNAi_IPK2 anteriormente obtido.

Detecção das plantas transgênicas

Foi realizado um ensaio histoquímico com a solução X-GLUC (*5-bromo-4-cloro-3-indolil glucuronida*) com as plantas T₀ potencialmente transformadas que foram

aclimatadas, aquelas linhagens que apresentaram a coloração azul, foram consideradas como um evento transgênico.

Posteriormente, DNA genômico foi isolado de acordo com Romano & Leal-Bertioli (2015). As análises de *Southern-blot* e a hibridização foram conduzidas como descrito por Romano & Vianna (2015), utilizando sondas específicas para o íntron PDK, obtida a partir de uma reação de PCR com primers específicos e usando o vetor pAHAS-GUS-RNAi_IPK2 como *template*. A digestão do DNA-genômico quantificado foi realizada com a enzima *SphI*, para avaliação do número de cópias do transgene.

Determinação dos teores de ácido fítico nas sementes de soja

Essa análise foi realizada como descrito no capítulo anterior com o kit da Megazyme para quantificação de ácido fítico e de fósforo total, para as sementes das linhagens obtidas. Utilizando 200mg (T₂) e 100mg (T₃) de material inicial (semente lixadas, aspecto de pó).

Quantificação da transcrição do gene por qRT-PCR

Foi extraído RNA de sementes (6 e 8 mm) e folhas (com 30 dias após germinação). das linhagens de IPK2-RNAi T₃ selecionadas, com o auxílio do kit RNAeasy *plant* (Qiagen®), conforme descrição do fabricante. O cDNA foi obtido a partir de 2µg de RNA total utilizando primers oligo dT com a reação da enzima *transcriptase reserva* (Invitrogen®). Com os *primers* endógenos *GmChib* (Matthews *et al.*, 2014), foram avaliados a integridade do cDNA e também a presença de contaminantes de DNA-genômico. O cDNA foi diluído 50x para a realização da qRT-PCR, utilizando o kit Master mix Invitrogen®. Os *primers* GmIPK2QF (AGGGAAAGGGAAAGCTACCA) e GmIPK2QR (GCCACCCAAGAAGTTGTGAT) foram desenhados especificamente para a amplificação do gene *IPK2*, que permitia gerar um fragmento de 96 pb. Como controle endógeno foi utilizado *primers* correspondentes ao gene *GmELF1* (Miranda *et al.*, 2013)

3.4 RESULTADOS

Construção do vetor pAHAS-GUS-RNAi_IPK2 para biobalística

Os fragmentos do gene *IPK2* obtidos, após a PCR com o cDNA, continham os sítios de restrições nas extremidades 5' para as enzimas *Xba*I e *Xho*I (*forward*) e *Bam*HI e *Kpn*I (*reverse*) para posterior clonagem do gene *GmIPK2* nos sentidos *sense* (*Xho*I e *Bam*HI) e *antisense* (*Xba*I e *Kpn*I) no vetor intermediário pKANNIBAL. Após a clonagem do fragmento nos dois sentidos, o cassette completo do RNAi foi removido com a enzima de restrição *Not*I e clonado no mesmo sítio *Not*I (único) do vetor pAHAS-GUS (Figura 12).

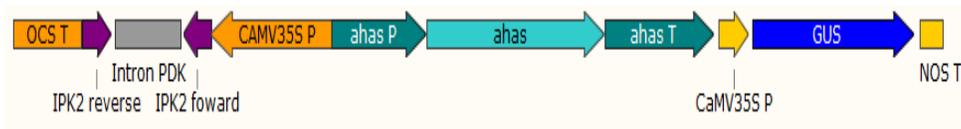


Figura 12: Esquema do vetor pAHAS-GUS-RNAi-IPK2. O cassette do RNAi está sob regulação do promotor duplicado de CAMV35s e terminador OCS (laranja). Os fragmentos do gene *IPK2* no sentido *sense* (representado como *foward*) e *antisense* (representado como *reverse*) (roxo) estão separados por um intron PDK (cinza). Além de conter também o cassette do gene de seleção *ahas* mutado de *A. thaliana* (azul claro), sendo regulado pelo promotor e terminador *ahas* (azul escuro). Possui o gene marcador *Gus* que está sob o controle do promotor CaMV35s e terminador *nos* (amarelo).

Obtenção das plantas transgênicas

Foram realizados dois experimentos de bombardeamento com um total de 648 embriões bombardeados, dos quais 50 plantas se desenvolveram em meio seletivo contendo *Imazapyr*, e foram aclimatadas, sendo quatro positivas para o gene repórter *GUS* na geração T₀ (Figura 13), que foram denominadas como linhagens #1, #7, #17 e #46. Dessas, apenas a progênie da linhagem #1 apresentou uma expressão forte do gene *GUS* na geração T₁ e essa mesma linhagem mostrou resultados positivos no *Southern-blot* (Dados não apresentados). As demais progênies, algumas apresentaram o nível expressão menor do gene *GUS*, baseando-se na intensidade da cor azul obtida.

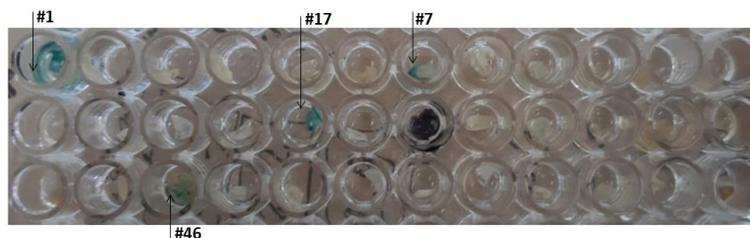


Figura 13: Ensaio histoquímico para detecção do gene *GUS* nas linhagens T₀ plantas transformadas com o vetor pAHAS-GUS-RNAi-IPK2. As linhagens #1, #7, #17 e #46 apresentaram-se positivas para a presença do gene *GUS*.

Análise do conteúdo de ácido fítico

Um ensaio preliminar foi realizado com sementes da geração T₂ de todas as linhagens obtidas. Foram necessários 0,2g de extrato, como as sementes apresentavam um tamanho pequeno, foram usadas entre três a quatro sementes de cada evento para obtenção do extrato a ser utilizado no experimento. O resultado obtido para essa análise é apresentado na tabela 7. É possível observar que duas linhagens, 7C e 17C (destacadas em verde na tabela) apresentaram redução do conteúdo de ácido fítico quando comparadas ao controle. Os mesmos critérios descritos para as plantas CRISPR foram utilizados nesta análise, onde uma média foi feita entre os indivíduos da mesma linhagem, destacando aqueles que apresentaram valor inferior à média (destacado em rosa). Algumas linhagens se destacaram pela alteração no valor da absorvância de fósforo livre (destacado em azul) ou aumento na concentração de fósforo (destacado em laranja).

Tabela 8: Conteúdo de ácido fítico nas sementes das plantas IPK2-RNAi T₂

Amostras	Fósforo livre (nm)	Fósforo total (nm)	Fósforo (g/100 g)	Ácido fítico (g/100 g)	Média
1A	0,057	1,010	0,540	1,914	
1B	0,113	1,250	0,644	2,284	
1C	0,185	1,370	0,671	2,380	
1D	0,197	1,260	0,602	2,135	1,999
7A	0,049	1,190	0,646	2,292	
7B	0,039	1,020	0,556	1,970	
7C	0,07	0,795	0,411	1,456	1,765
17A	0,095	1,290	0,677	2,400	
17B	0,049	1,090	0,590	2,091	
17C	0,088	0,934	0,479	1,699	
17D	0,055	0,991	0,530	1,880	1,882
26A	0,049	1,130	0,612	2,171	2,171
46A	0,028	0,927	0,509	1,806	
46B	0,067	0,997	0,527	1,868	
46C	0,087	1,300	0,687	2,437	
46D	0,059	1,200	0,646	2,292	1,941
BRS7980	0,070	1,090	0,578	2,049	

Descrição: Valor de ácido fítico inferior ao controle BRS7980 (verde); indivíduo com valor inferior à média da linhagem (rosa); Alteração da absorvância de fósforo livre (variação de azul) e alteração na quantidade final de fósforo (laranja).

Com isso as linhagens denominadas 1A, 1D, 7C, 17C, 17D, 46A e 46B foram colocadas para avanço da geração. Quatro sementes de cada uma delas foram germinadas, algumas não sobreviveram e outras apresentaram fenótipos diferentes das demais. Das quatro sementes da linhagem 1A, três se desenvolveram, porém apresentaram baixa

estatura (1A-1), retardo no crescimento (1A-2 e 1A-3) e as folhas senesceram antes mesmo das sementes se desenvolverem completamente (1A-3 principalmente, mas também foi observado na 1A-2). Já para a linhagem 1D, apenas duas plantas sobreviveram das 4 sementes inicialmente colocadas para germinação, as duas apresentaram retardo no desenvolvimento, as sementes demoraram mais de 4 meses para amadurecer (Figura 14). Apesar das alterações fenotípicas para a maioria das progênies da linhagem mãe #1, a linhagem 1A-1 apresentou um desenvolvimento mais próximo do normal, apenas com crescimento reduzido (baixa estatura), porém não alterou a produção.

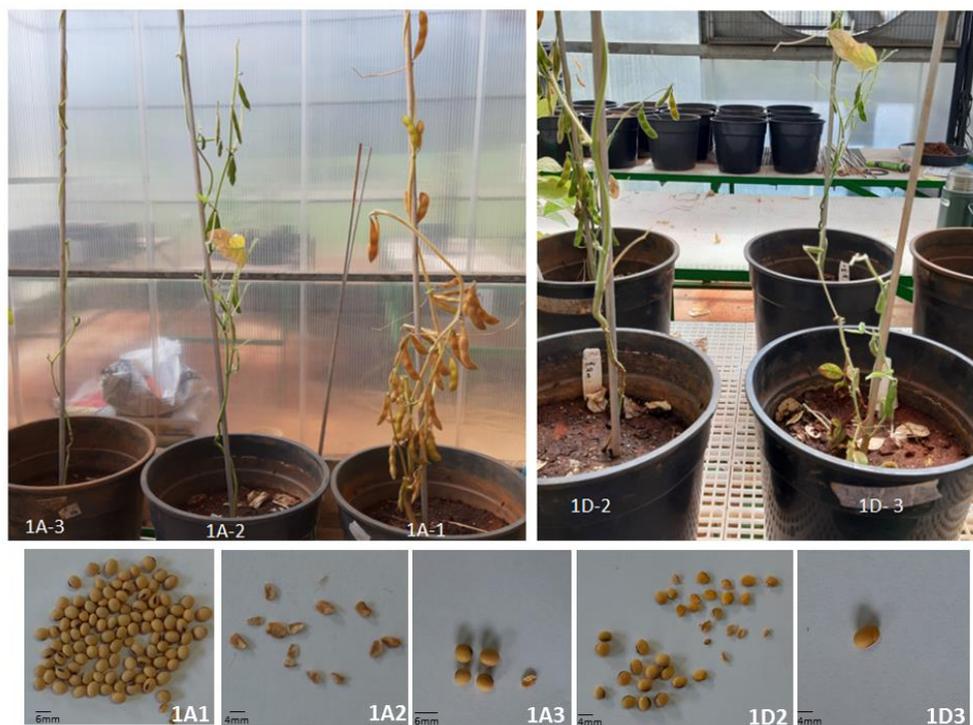


Figura 14: Fenótipos alterados observados nas linhagens de soja IPK2-RNAi #1 progênies A (1, 2 e 3) e D (2 e 3). Na parte inferior são apresentadas as sementes obtidas para as linhagens 1A1, 1A2, 1A3 1D2 e 1D3. As linhagens 1A3 e 1D3 apresentaram retardo no desenvolvimento.

Uma segunda análise do conteúdo de ácido fítico com a progênie T_3 foi feita utilizando 0,1g de extrato, sendo necessário para algumas amostras utilizar até três sementes. Foram analisadas duas amostras para cada linhagem, uma com a morfologia normal (N) e outra abortada/desenvolvimento anormal (A).

O padrão observado anteriormente, se repetiu na análise com as progênies T₃ para a grande maioria das amostras, com a quantidade de ácido fítico superior ao do controle BRS7980 (Tabela 8). Com exceção das linhagens 1A-1, 1D-2, 7C-2 e 17D2 que apresentaram redução de ácido fítico quando comparadas ao controle. As sementes com um padrão anormal/abortada apresentaram quantidade de ácido fítico menor para todas as amostras quando comparado ao controle BRS7980 (dados não apresentados). De modo que não foi possível fazer uma correlação entre o fenótipo (abortada/anormal) com a redução do conteúdo de ácido fítico.

Tabela 9: Conteúdo de ácido fítico nas sementes das plantas IPK2-RNAi T₃

Amostras	Fósforo livre (nm)	Fósforo total (nm)	Fósforo (g/100 g)	Ácido fítico (g/100 g)
1A1 N	0,111	0,690	0,372	1,320
1D2 N	0,545	1,260	0,459	1,628
7C1 N	0,150	1,074	0,594	2,105
7C2 N	0,138	1,006	0,558	1,978
7C3 N	0,146	1,234	0,699	2,479
17C2 N	0,135	1,126	0,637	2,258
17C3 N	0,160	1,363	0,773	2,740
17D1 N	0,159	1,108	0,610	2,163
17D2 N	0,158	1,064	0,582	2,064
46A1 N	0,157	1,191	0,664	2,355
46B1 N	0,101	1,063	0,619	2,194
BRS7980	0,112	1,025	0,586	2,080

Descrição: Valor de ácido fítico inferior ao controle BRS7980 (verde); indivíduo com valor inferior à média da linhagem (rosa); Alteração da absorvância de fósforo livre (variação de azul) e alteração na quantidade final de fósforo (laranja).

As amostras que se destacaram na análise apresentada na tabela 9, foram usadas em um novo experimento para confirmar o resultado anteriormente obtido. Assim a tabela 10 apresenta os dados finais, sendo possível detectar uma redução do teor de ácido fítico de 41,67% na linhagem 1D-2, 23,96% para linhagem 1A-1 e 13,71% na linhagem 7C-2. Quanto aos fenótipos dessas linhagens, o desenvolvimento das plantas 1A-1 e 1D-2 foram apresentadas na figura 13; as linhagens 7C-1, 7C-2 e 17D-2 apresentaram desenvolvimento normal.

Quanto à produtividade, a linhagem 1A-1 produziu aproximadamente 100 sementes normais com apenas 3 com desenvolvimento anormal sem redução no tamanho/massa das sementes (Tabela 10), a linhagem 1D-2 produziu apenas 14 sementes normais pequenas e

21 entre abortadas e com desenvolvimento anormal (Figura 13). Já as linhagens 7C-2 e 17D-2 produziram 150 e 132 sementes respectivamente. A linhagem 1D-2 apresentou uma redução numericamente significativa do tamanho/massa, enquanto que 7C-1, 7C-2 e 17D-2 apresentou a massa reduzido, talvez dentro do esperado (Tabela 10).

Tabela 10: Conteúdo de ácido fítico nas sementes destacadas IPK2-RNAi T₃

Amostras	Fósforo livre (nm)	Fósforo total (nm)	Fósforo (g/100 g)	Ácido fítico (g/100 g)	% de PA reduzido	Massa grãos (100 sementes)
1 A1	0,106	0,929	0,521	1,847	23,96	17,6g
1 D2	0,101	0,733	0,399	1,417	41,67	4,0g*
7 C1	0,142	1,233	0,690	2,448	0	10,66g
7 C2	0,118	1,052	0,591	2,097	13,71	15,22g
17 D2	0,129	1,151	0,646	2,293	5,62	12,70g
BRS7980	0,153	1,301	0,726	2,430	-	17,0g

Descrição: Valor de ácido fítico inferior ao controle BRS7980 (verde); indivíduo com valor inferior à média da linhagem (rosa); Alteração da absorvância de fósforo livre (variação de azul). (*Não apresentava a quantidade de sementes suficientes, foi feita uma estimativa). PA Ácido fítico.

Análise quantitativa da transcrição do gene *GmIPK2*

RNA total das linhagens de soja T₃ de IPK2-RNAi apresentadas na tabela 10 foram extraídos a partir de sementes com tamanho aproximado entre 6 e 8 mm e de folhas (com 30 dias após germinação). As reações foram feitas para o gene *GmIPK2*, tendo como controle o gene endógeno, fator de alongação *GmELF1*. Nos resultados apresentados na figura 15, as duas linhagens que apresentaram redução mais significativa do conteúdo de ácido fítico, 1A1 e 1D2, também apresentaram redução na transcrição do gene *GmIPK2*, sendo essa redução na expressão mais evidente em sementes. Para as amostras 7C1 e 7C2 houve uma redução da expressão em sementes, porém enquanto a linhagem 7C1 apresentou uma grande redução na expressão do gene alvo tanto em sementes como em folhas linhagem, a 7C2 apresentou uma expressão elevada de *GmIPK2* em folhas, quando comparada ao observado no controle BRS7980. Um mesmo padrão controverso foi observado para a linhagem 17D2 que apresentou uma transcrição elevada em sementes.

Embora apenas as linhagens 1A1, 1D2 e 7C2 apresentaram redução no conteúdo de ácido fítico, alguns fenótipos apresentados em alguns dos indivíduos das linhagens estudadas podem justificar que a redução na transcrição do gene *GmIPK2* pode ter

influenciado em outros processos. Durante o desenvolvimento das plantas, antes do completo enchimento das vagens, alguns indivíduos apresentaram amarelamento das folhas e aspecto de senescência foliar precoce, de modo que as folhas caíram antes mesmo do desenvolvimento completo das vagens. E como consequência indireta essas linhagens apresentaram redução na quantidade de sementes produzidas.

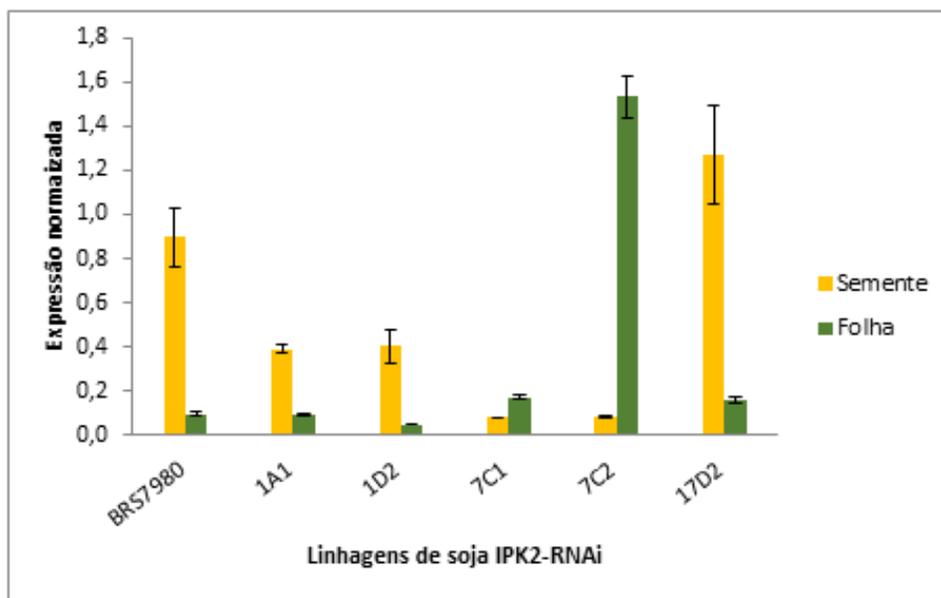


Figura 15: Expressão normalizada do gene *GmIPK2* em sementes e folhas de sojas geneticamente modificadas mediada por qRT-PCR. Indicando níveis variados de silenciamento da expressão do gene *GmIPK2* em sementes (Amarelo) e folhas (verde) para as linhagens que apresentaram algum nível de redução de ácido fítico.

3.5 DISCUSSÃO

A escolha da ferramenta biotecnologia apropriada é dependente do objetivo que se deseja alcançar. Embora já se tenha a variante dead/Cas9 que é capaz de modular a expressão gênica (Brezgin *et al.*, 2019), o sistema CRISPR/Cas9 em sua forma tradicional, atua no nível de DNA causando um *knockout*, ou seja, um silenciamento completo e permanente do gene. Não sendo em alguns casos a melhor estratégia a ser usada. Já o RNA interferente reduz a expressão gênica a nível transcricional (mRNA) causando um *knockdown*, de modo que podem existir variações nos níveis de silenciamento do gene em diferentes eventos (Prabhune, 2020).

Considerando que os produtos da via do inositol-fosfato apresentam papéis importantes na germinação e desenvolvimento da planta, talvez um silenciamento completo de um dos genes possa tornar a planta inviável, como foi observado por Nunes *et al.*, (2006), que tiveram como objetivo a regulação da expressão do gene *GmMIPS* de *G. max*, e observaram que quando houve um silenciamento completo do gene, as sementes tornaram-se inviáveis devido a terem o desenvolvimento paralisado precocemente. A hipótese para os resultados obtido por Nunes *et al.* (2006) estaria na escolha do gene *MIPS*, uma vez que este é o primeiro gene da via, o que interrompeu a rota do inositol-fosfato completamente, ocorrendo uma redução total no nível de ácido fítico. No entanto Kumar *et al.*, (2019) silenciaram o mesmo gene *GmMIPS* e conseguiram obter redução de até 41,34% dos níveis de ácido fítico em soja, sem comprometer a germinação. A diferença entre os dois trabalhos foi a utilização do promotor específico de semente (*promoter vicilin*), promotor esse de proteína de reserva.

Neste trabalho foi possível reduzir o nível de ácido fítico com a repressão da expressão do gene *GmIPK2* em uma cultivar comercial da Embrapa, BRS7980, com redução de 41,63% (IPK2-RNAi_1D2), 23,96% (IPK2-RNAi_1A1) e 13,71% (IPK2-RNAi_7C2). Porém, foram observados impactos negativos no desenvolvimento da linhagem 1D2, tanto no desenvolvimento da planta quanto na produtividade, gerando apenas 14 sementes normais e 21 sementes com aspecto anormal ou abortado. Contudo, as outras duas linhagens (1A1 e 7C2) apresentaram desenvolvimento normal. Com objetivo semelhante Punjabi *et al.* (2018B) utilizando um promotor específico de semente (*promoter vicilin*), conseguiram uma redução entre 42 e 45% de ácido fítico, além de acúmulo de

fósforo inorgânico e elevado conteúdo de minerais em linhagens de soja com a regulação do gene *GmIPK2*. Esses indícios mostram que o sucesso ou não na obtenção de um evento *lpa* sem efeito deletério para a planta parece estar correlacionado com a utilização de promotores específicos de sementes.

Embora a rota lipídeo-dependente, rota de atuação do gene *GmIPK2*, não atue de forma principal em sementes, a sua expressão é maior em sementes do que em outros tecidos (Punjabi *et al.*, 2018A). Os resultados obtidos na análise da expressão gênica mostraram que para as linhagens 1D2 e 1A1 que apresentaram as maiores reduções do teor de ácido fítico apresentaram também redução do transcrito referente ao gene *GmIPK2*, porém não tão significativo quanto as linhagens 7C1 e 7C2, porém nessas últimas linhagens quando analisado o conteúdo de ácido fítico não houve redução significativa.

Algumas hipóteses podem ser consideradas para entendimento dessas questões, dentre elas o fato de que a rota principal da produção de ácido fítico em sementes (lipídeo-independente) pode ter recompensado a produção da molécula InsP5, molécula que antecede o InsP6 (ácido fítico), de modo a não ser possível detectar a redução do conteúdo de ácido fítico, mas ter observado a redução da expressão do gene. Outra opção está relacionada à segregação, onde o material vegetal usados para os dois experimentos (quantificação de ácido fítico e qRT-PCR) foram diferentes, de modo a não ter como confirmar se todas as sementes utilizadas eram positivas.

Quando comparado os dois tecidos, folhas e sementes, a redução dos transcritos em folhas não foi perceptível, em sementes essa redução foi mais evidente, mas não houve uma correlação entre, apenas as plantas com maior redução de ácido fítico apresentar maior redução da expressão do transcrito. Um resultado discrepante foi observado para as linhagens 7C2 e 17D2 que apresentaram uma expressão elevada de *GmIPK2* em folhas e sementes respectivamente quando comparada com o controle BRS7980.

Novas análises ainda serão realizadas para melhor compreensão dos efeitos causados pela redução do conteúdo de ácido fítico nas linhagens que se destacaram. E ainda novos estudos serão feitos para uma melhor correlação da expressão gênica com a redução dos níveis de ácido fítico.

CONSIDERAÇÕES

A relação existente entre fazer apenas o *Knockdown* (redução da expressão gênica) de um gene que pertence a uma via de extrema importância pode estar correlacionada ao sucesso da obtenção de uma planta transformada, que apresente as características desejadas. Conseguimos a redução do conteúdo de ácido fítico entre 13,71 e 41,34% em três linhagens utilizando RNAi, sendo que o evento com maior redução na produção de ácido fítico, também apresentou efeitos negativos na produtividade.

Ainda serão necessários novos experimentos para o melhor entendimento da correlação da expressão gênica do gene *GmIPK2* e a redução na quantidade de ácido fítico detectada. Também será necessário realizar experimentos para verificar se houve aumento na biodisponibilidade de nutrientes após a redução do conteúdo de ácido fítico, de modo a ter uma planta de soja com um maior valor nutricional.

Além disso, as plantas aqui geradas podem ser usadas para avaliar se a alteração na expressão do gene *IPK2*, está relacionada com alteração nas respostas aos estresses abióticos conforme observado por Jiang *et al.*, (2020).

CONCLUSÃO FINAL E PERSPECTIVAS

Até este ponto do projeto, foi possível concluir que a utilização prévia da técnica de geração de raízes transgênicas (*hairy-root*), associada ao sistema CRISPR é recomendada para avaliar se os alvos escolhidos, bem como a configuração do sistema de edição, são eficientes. Uma vez que devido ao longo tempo gasto no processo de obtenção de uma planta transgênica, essa avaliação permite economia de tempo e garantia da utilização de um cassete CRISPR viável. Além disso, ambas as configurações CRISPR seja a tradicional (TCTU) ou a simplificada (STU) se mostraram viáveis em gerar edições no genoma da soja. Embora a configuração tradicional tenha se mostrado mais eficiente, em gerar edições nas três regiões alvos.

Os genes *IPK1* e *IPK2* que já foram descritos na literatura como genes essenciais na via do Inositol-fosfato, quando nocauteado ou silenciado foram capazes de reduzir o conteúdo de ácido fítico. De modo que uma redução drástica na expressão dos genes *IPK2* mediada por RNAi condicionou a obtenção de plantas com redução do teor de ácido fítico nas sementes, com alterações no desenvolvimento e produtividade. Enquanto que uma redução moderada no conteúdo de ácido fítico para o mesmo gene mediado por CRISPR ou RNAi pode ter resultado em uma planta sem alterações no desenvolvimento, produtividade e tamanho dos grãos.

Todas as plantas obtidas neste trabalho que apresentaram redução do conteúdo de ácido fítico serão avaliadas quanto ao aumento de nutrientes que provavelmente se tornaram disponíveis após a redução do ácido fítico. Além de ensaios para avaliar a tolerância dessas linhagens a estresses abióticos.

Utilizando ambas as ferramentas biotecnologias CRISPR ou RNAi, foi possível gerar plantas de soja com conteúdo reduzido de ácido fítico. Sendo possível afirmar que as metodologias avaliadas neste estudo poderão ser utilizadas para gerar novas plantas com características desejadas, podendo também extrapolado para outras culturas, como feijão e algodão.

REFERÊNCIAS BIBLIOGRÁFICAS

- Adepoju, O. A., Williams, S. P. & Gillaspay, G. (2017) Inositol Phosphates and Energy Signaling in Plants. *The FASEB Journal*, 31, 628.612-628.612
- Adli, M. (2018) The CRISPR tool kit for genome editing and beyond. *Nature Communications*, 9, 1911
- Aggarwal, S., Kumar, A., Bhati, K. K., Kaur, G., Shukla, V., Tiwari, S., *et al.* (2018) RNAi-Mediated Downregulation of Inositol Pentakisphosphate Kinase (*IPKI*) in Wheat Grains Decreases Phytic Acid Levels and Increases Fe and Zn Accumulation. *Frontiers in Plant Science*, 9
- Ali, N., Paul, S., Gayen, D., Sarkar, S. N., Datta, K. & Datta, S. K. (2013) Development of low phytate rice by RNAi mediated seed-specific silencing of inositol 1,3,4,5,6-pentakisphosphate 2-kinase gene (*IPKI*). *PLoS One*, 8
- An, G., Watson, B. D., Stachel, S., Gordon, M. P. & Nester, E. W. (1985) New cloning vehicles for transformation of higher plants. *Embo J*, 4, 277-284
- Baltes, N. J., Gil-Humanes, J. & Voytas, D. F. (2017) Chapter One - Genome Engineering and Agriculture: Opportunities and Challenges. In *Progress in Molecular Biology and Translational Science* (ed. by D. P. Weeks & B. Yang), pp. 1-26. Academic Press.
- Bao, A., Chen, H., Chen, L., Chen, S., Hao, Q., Guo, W., *et al.* (2019) CRISPR/Cas9-mediated targeted mutagenesis of *GmSPL9* genes alters plant architecture in soybean. *BMC Plant Biol*, 19, 019-1746
- Barrangou, R. (2015) Diversity of CRISPR-Cas immune systems and molecular machines. *Genome Biol*, 16, 015-0816
- Basak, N., Krishnan, V., Pandey, V., Punjabi, M., Hada, A., Marathe, A., *et al.* (2017) Molecular characterization of inositol pentakisphosphate 2-kinase (*GmlPki*) from soybean and its expression pattern in the developing seeds. *Indian Journal of Genetics and Plant Breeding (The)*, 77, 371
- Brezgin, S., Kostyusheva, A., Kostyushev, D. & Chulanov, V. (2019) Dead Cas Systems: Types, Principles, and Applications. *International journal of molecular sciences*, 20, 6041
- Cai, Y., Chen, L., Sun, S., Wu, C., Yao, W., Jiang, B., *et al.* (2018) CRISPR/Cas9-Mediated Deletion of Large Genomic Fragments in Soybean. *Int J Mol Sci*, 19
- Campion, B., Sparvoli, F., Doria, E., Tagliabue, G., Galasso, I., Fileppi, M., *et al.* (2009) Isolation and characterisation of a lpa (low phytic acid) mutant in common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet*, 118, 1211-1221
- Čermák, T., Curtin, S. J., Gil-Humanes, J., Čegan, R., Kono, T. J. Y., Konečná, E., *et al.* (2017) A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. *Plant Cell*, 29, 1196-1217
- Cheng, Q., Dong, L., Su, T., Li, T., Gan, Z., Nan, H., *et al.* (2019) CRISPR/Cas9-mediated targeted mutagenesis of *GmLHY* genes alters plant height and internode length in soybean. *BMC Plant Biol*, 19, 019-2145

- Chiera, J. M., Bouchard, R. A., Dorsey, S. L., Park, E., Buenrostro-Nava, M. T., Ling, P. P., *et al.* (2007) Isolation of two highly active soybean (*Glycine max* (L.) Merr.) promoters and their characterization using a new automated image collection and analysis system. *Plant Cell Rep*, 26, 1501-1509
- Chilcoat, D., Liu, Z. B. & Sander, J. (2017) Use of CRISPR/Cas9 for Crop Improvement in Maize and Soybean. *Prog Mol Biol Transl Sci*, 149, 27-46
- Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H. & Sheen, J. (1996) Engineered GFP as a vital reporter in plants. *Curr Biol*, 6, 325-330
- CONAB, 2020 link acessado em dezembro, 2020 <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>
- Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., *et al.* (2015) Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol*, 16, 015-0846
- Davies, J. P., Kumar, S. & Sastry-Dent, L. (2017) Use of Zinc-Finger Nucleases for Crop Improvement. *Prog Mol Biol Transl Sci*, 149, 47-63
- Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*, 34, 184-191
- Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., *et al.* (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol*, 32, 1262-1267
- Du, H., Zeng, X., Zhao, M., Cui, X., Wang, Q., Yang, H., *et al.* (2016) Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *J Biotechnol*, 217, 90-97
- Entine, J., Felipe, M.S.S., Groenewald, JH., Kershen, D.L, Lema, M., McHughen, A., *et al.* (2021) Regulatory approaches for genome edited agricultural plants in select countries and jurisdictions around the world. *Transgenic Res*.
- FAOSTAT, 2019, Link acessado em janeiro de 2021 <http://www.fao.org/faostat/en/#data/QC>
- Gaj, T., Gersbach, C. A. & Barbas, C. F., 3rd. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*, 31, 397-405
- Gao, C. (2018) The future of CRISPR technologies in agriculture. *Nat Rev Mol Cell Biol*, 19, 275-276
- Gao, Y., Zhang, Y., Zhang, D., Dai, X., Estelle, M. & Zhao, Y. (2015) Auxin binding protein 1 (ABP1) is not required for either auxin signaling or Arabidopsis development. *Proc Natl Acad Sci U S A*, 112, 2275-2280
- Ghogare, R., Williamson-Benavides, B., Ramírez-Torres, F. & Dhingra, A. (2020) CRISPR-associated nucleases: the Dawn of a new age of efficient crop improvement. *Transgenic Res*, 29, 1-35
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*, 40, 22

- Goßner, S., Yuan, F., Zhou, C., Tan, Y., Shu, Q. & Engel, K. H. (2019) Impact of Cross-Breeding of Low Phytic Acid MIPS1 and IPK1 Soybean (*Glycine max* L. Merr.) Mutants on Their Contents of Inositol Phosphate Isomers. *J Agric Food Chem*, 67, 247-257
- Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M., *et al.* (2014) Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics*, 196, 961-971
- Guttieri, M., Bowen, D., Dorsch, J., Raboy, V. & Souza, E. (2004) Identification and Characterization of a Low Phytic Acid Wheat. *Crop Science*, 44
- Han, H. (2018) RNA Interference to Knock Down Gene Expression. *Methods Mol Biol*, 7471-7479_7416
- Hitz, W. D., Carlson, T. J., Kerr, P. S. & Sebastian, S. A. (2002) Biochemical and molecular characterization of a mutation that confers a decreased raffinose and phytic acid phenotype on soybean seeds. *Plant Physiol*, 128, 650-660
- Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 31, 827-832
- Jacobs, T. B., LaFayette, P. R., Schmitz, R. J. & Parrott, W. A. (2015) Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol*, 15, 015-0131
- Jagtap, U. B., Gurav, R. G. & Bapat, V. A. (2011) Role of RNA interference in plant improvement. *Naturwissenschaften*, 98, 473-492
- Jiang, M., Liu, Y., Li, R., Li, S., Tan, Y., Huang, J., *et al.* (2020) An Inositol 1,3,4,5,6-Pentakisphosphate 2-Kinase 1 Mutant with a 33-nt Deletion Showed Enhanced Tolerance to Salt and Drought Stress in Rice. *Plants*, 10
- Jiang, M., Liu, Y., Tan, Y., Huang, J. & Shu, Q. (2019) Mutation of Inositol 1,3,4-trisphosphate 5/6-kinase6 Impairs Plant Growth and Phytic Acid Synthesis in Rice. *Plants*, 8
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-821
- Kim, G.-B. & Nam, Y.-W. (2013) Isolation and Characterization of *Medicago truncatula* U6 Promoters for the Construction of Small Hairpin RNA-Mediated Gene Silencing Vectors. *Plant Molecular Biology Reporter*, 31, 581-593
- Kumar, A., Kumar, V., Krishnan, V., Hada, A., Marathe, A., C, P., *et al.* (2019) Seed targeted RNAi-mediated silencing of *GmMIPS1* limits phytate accumulation and improves mineral bioavailability in soybean. *Sci Rep*, 9, 019-44255
- Lee, C. M., Zhu, H., Davis, T. H., Deshmukh, H. & Bao, G. (2017) Design and Validation of CRISPR/Cas9 Systems for Targeted Gene Modification in Induced Pluripotent Stem Cells. *Methods Mol Biol*, 6472-6477_6471
- Li, C., Nguyen, V., Liu, J., Fu, W., Chen, C., Yu, K., *et al.* (2019) Mutagenesis of seed storage protein genes in Soybean using CRISPR/Cas9. *BMC Res Notes*, 12, 019-4207

- Li, J., Zhang, B., Ma, T., Wang, H., Yu, Q. & Li, M. (2017) Role of the Inositol Polyphosphate Multikinase *IPK2* in Regulation of Hyphal Development, Calcium Signaling and Secretion in *Candida albicans*. *Mycopathologia*, 182, 609-623
- Liang, Z., Zhang, K., Chen, K. & Gao, C. (2014) Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *J Genet Genomics*, 41, 63-68
- Limera, C., Sabbadini, S., Sweet, J. B. & Mezzetti, B. (2017) New Biotechnological Tools for the Genetic Improvement of Major Woody Fruit Species. *Front Plant Sci*, 8
- Lindbo, J. A. (2012) A historical overview of RNAi in plants. *Methods Mol Biol*, 894, 1-16
- Luan, H., Liao, W., Song, Y., Niu, H., Hu, T. & Zhi, H. (2020) Transgenic plant generated by RNAi-mediated knocking down of soybean *Vma12* and soybean mosaic virus resistance evaluation. *AMB Express*, 10, 020-00997
- Luttgeharm, K. D., Wong, K. S. & Siembieda, S. (2017) Heteroduplex cleavage assay for screening of probable zygosity resulting from CRISPR mutations in diploid single cell lines. *Biotechniques*, 62, 268-274
- Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., *et al.* (2015) A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. *Mol Plant*, 8, 1274-1284
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., *et al.* Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*. 2011 Jun;9(6):467-77.
- Malhotra H., Vandana, Sharma S., Pandey R. (2018) Phosphorus Nutrition: Plant Growth in Response to Deficiency and Excess. In: *Plant Nutrients and Abiotic Stress Tolerance*. Springer, Singapore.
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., *et al.* (2013) RNA-guided human genome engineering via Cas9. *Science*, 339, 823-826
- Marathe, A., Krishnan, V., Mahajan, M. M., Thimmegowda, V., Dahuja, A., Jolly, M., *et al.* (2018) Characterization and molecular modeling of Inositol 1,3,4 tris phosphate 5/6 kinase-2 from *Glycine max* (L) Merr.: comprehending its evolutionary conservancy at functional level. 8, 017-1076
- Martynov, V. V. & Dorokhov, D. B. (2014) [Polymorphism of the *IPK1* gene among members of the genus *Glycine*]. *Genetika*, 50, 787-793
- Miranda Vde, J., Coelho, R. R., Viana, A. A., de Oliveira Neto, O. B., Carneiro, R. M., Rocha, T. L., *et al.* (2013) Validation of reference genes aiming accurate normalization of qPCR data in soybean upon nematode parasitism and insect attack. *BMC Res Notes*, 6, 1756-0500
- Mikami, M., Toki, S. & Endo, M. (2017) In Planta Processing of the SpCas9-gRNA Complex. *Plant Cell Physiol*, 58, 1857-1867
- Mohanta, T. K., Bashir, T., Hashem, A. & Abd Allah, E. F. (2017) Systems biology approach in plant abiotic stresses. *Plant Physiol Biochem*, 121, 58-73

- Morriss, S. C., Studham, M. E., Tylka, G. L. & MacIntosh, G. C. (2017) Validation of a hairy roots system to study soybean-soybean aphid interactions. *PLoS One*, 12
- Nakade, S., Yamamoto, T. & Sakuma, T. (2017) Cas9, Cpf1 and C2c1/2/3-What's next? *Bioengineered*, 8, 265-273
- Nunes, A. C., Vianna, G. R., Cuneo, F., Amaya-Farfán, J., de Capdeville, G., Rech, E. L., *et al.* (2006) RNAi-mediated silencing of the myo-inositol-1-phosphate synthase gene (GmMIPS1) in transgenic soybean inhibited seed development and reduced phytate content. *Planta*, 224, 125-132
- Prabhune, M. 2020. RNAi vs. CRISPR: Guide to selecting the best gene silencing method. SYNTHEG'S The Bench Blog, Abril, 2020. <https://www.synthego.com/blog/rnai-vs-crispr-guide>.
- Pellagatti, A., Dolatshad, H., Valletta, S. & Boultonwood, J. (2015) Application of CRISPR/Cas9 genome editing to the study and treatment of disease. *Arch Toxicol*, 89, 1023-1034
- Perez-Rojo, F., Nyman, R. K. M., Johnson, A. A. T., Navarro, M. P., Ryan, M. H., Erskine, W., *et al.* (2018) CRISPR-Cas systems: ushering in the new genome editing era. *Bioengineered*, 9, 214-221
- Punjabi, M., Bharadvaja, N., Sachdev, A. & Krishnan, V. (2018A) Molecular characterization, modeling, and docking analysis of late phytic acid biosynthesis pathway gene, inositol polyphosphate 6-/3-/5-kinase, a potential candidate for developing low phytate crops. 8, 018-1343
- Punjabi, M., Bharadvaja, N., Jolly, M., Dahuja, A. & Sachdev, A. (2018B) Development and Evaluation of Low Phytic Acid Soybean by siRNA Triggered Seed Specific Silencing of Inositol Polyphosphate 6-/3-/5-Kinase Gene. *Front Plant Sci*, 9
- Qi, X., Wu, H., Jiang, H., Zhu, J., Huang, C., Zhang, X., *et al.* (2020) Conversion of a normal maize hybrid into a waxy version using in vivo CRISPR/Cas9 targeted mutation activity. *The Crop Journal*, 8, 440-448.
- Raboy, V., Cichy, K., Peterson, K., Reichman, S., Sompong, U., Srinives, P., *et al.* (2014) Barley (*Hordeum vulgare* L.) low phytic acid 1-1: an endosperm-specific, filial determinant of seed total phosphorus. *J Hered*, 105, 656-665
- Raboy, V. (2009) Approaches and challenges to engineering seed phytate and total phosphorus. *Plant Science*, 177, 281-296
- Raza, G., Singh, M. B. & Bhalla, P. L. (2019) Somatic Embryogenesis and Plant Regeneration from Commercial Soybean Cultivars. *Plants*, 9
- Rech, E. L., Vianna, G. R. & Aragão, F. J. (2008) High-efficiency transformation by biolistics of soybean, common bean and cotton transgenic plants. *Nat Protoc*, 3, 410-418
- Romano E.; Leal-bertioli, S. C. M. Extração e quantificação de dna de tecidos vegetais. In: Brasileiro, A.C.M; Carneiro, V.T.C. (Org.). *Manual de Transformação Genética de Plantas*. 2ed.Brasília: Embrapa, 2015, v. 1, p. 165-180.

- Romano, E.; Vianna, G. R. Análise da integração de transgenes pela técnica de Southern blot. In: Brasileiro, A.C.M; Carneiro, V.T.C. (Org.). Manual de Transformação Genética de Plantas. 2ed. Brasília: Embrapa, 2015, v. 1, p. 215-232
- Sanson, K. R., Hanna, R. E., Hegde, M., Donovan, K. F., Strand, C., Sullender, M. E., *et al.* (2018) Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun*, 9, 018-07901
- Sashidhar, N., Harloff, H. J., Potgieter, L. & Jung, C. (2020) Gene editing of three BnITPK genes in tetraploid oilseed rape leads to significant reduction of phytic acid in seeds. *Plant Biotechnol J*, 18, 2241-2250
- Schardl, C. L., Byrd, A. D., Benzion, G., Altschuler, M. A., Hildebrand, D. F. & Hunt, A. G. (1987) Design and construction of a versatile system for the expression of foreign genes in plants. *Gene*, 61, 1-11
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., *et al.* (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, 463, 178-183
- Sedeek, K. E. M., Mahas, A. & Mahfouz, M. (2019) Plant Genome Engineering for Targeted Improvement of Crop Traits. *Front Plant Sci*, 10
- Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., *et al.* Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol*. 2013 Aug;31(8):686-8.
- Shi, J., Wang, H., Schellin, K., Li, B., Faller, M., Stoop, J. M., *et al.* (2007) Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. *Nat Biotechnol*, 25, 930-937
- Sparvoli, F. & Cominelli, E. (2015) Seed Biofortification and Phytic Acid Reduction: A Conflict of Interest for the Plant? *Plants*, 4, 728-755
- Stevenson-Paulik, J., Bastidas, R. J., Chiou, S.-T., Frye, R. A. & York, J. D. (2005) Generation of phytate-free seeds in *Arabidopsis* through disruption of inositol polyphosphate kinases. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 12612-12617
- Sun, X., Hu, Z., Chen, R., Jiang, Q., Song, G., Zhang, H., *et al.* (2015) Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Sci Rep*, 5
- Tang, X., Ren, Q., Yang, L., Bao, Y., Zhong, Z., He, Y., *et al.* (2019) Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing. *Plant Biotechnol J*, 17, 1431-1445
- Tang, X., Zheng, X., Qi, Y., Zhang, D., Cheng, Y., Tang, A., *et al.* (2016) A Single Transcript CRISPR-Cas9 System for Efficient Genome Editing in Plants. *Mol Plant*. 2016 Jul 6;9(7):1088-91., *Mol Plant*.
- Thacker, P., Deep, A., Petri, D. & Warkentin, T. (2013) Nutritional evaluation of low-phytate peas (*Pisum sativum* L.) for young broiler chicks. *Archives of Animal Nutrition*, 67, 1-14

- Thrane, M., Paulsen, P. V., Orcutt, M. W. & Krieger, T. M. (2017) Chapter 2 - Soy Protein: Impacts, Production, and Applications. In Sustainable Protein Sources (ed. by S. R. Nadathur, J. P. D. Wanasundara & L. Scanlin), pp. 23-45. Academic Press, San Diego.
- Torella, J. P., Lienert, F., Boehm, C. R., Chen, J. H., Way, J. C. & Silver, P. A. (2014) Unique nucleotide sequence-guided assembly of repetitive DNA parts for synthetic biology applications. *Nat Protoc*, 9, 2075-2089
- Verdaguer, B., de Kochko, A., Beachy, R. N. & Fauquet, C. (1996) Isolation and expression in transgenic tobacco and rice plants, of the cassava vein mosaic virus (CVMV) promoter. *Plant Mol Biol*, 31, 1129-1139
- Vlčko, T. & Ohnoutková, L. (2020) Allelic Variants of CRISPR/Cas9 Induced Mutation in an Inositol Trisphosphate 5/6 Kinase Gene Manifest Different Phenotypes in Barley. *Plants*, 9
- Vincent, J. A., Stacey, M., Stacey, G. & Bilyeu, K. D. (2015) Phytic Acid and Inorganic Phosphate Composition in Soybean Lines with Independent IPK1 Mutations. *Plant Genome*, 8
- Wang, L., Zhou, Y. & Duanmu, D. (2017) Use of CRISPR/Cas9 for Symbiotic Nitrogen Fixation Research in Legumes. *Prog Mol Biol Transl Sci*, 149, 187-213
- Wang, M., Mao, Y., Lu, Y., Wang, Z., Tao, X. & Zhu, J. K. (2018) Multiplex gene editing in rice with simplified CRISPR-Cpf1 and CRISPR-Cas9 systems. *J Integr Plant Biol*. 8:626-631. Epub 2018 Jul 7.
- Warkentin, T., Delgerjav, O., Arganosa, G., Rehman, A., Bett, K. E., Anbessa, Y., *et al.* (2012) Development and Characterization of Low-Phytate Pea. *Crop Science*, 52
- Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, 482, 331-338
- Wesley, S. V., Helliwell, C. A., Smith, N. A., Wang, M. B., Rouse, D. T., Liu, Q., *et al.* (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J*, 27, 581-590
- Wolter, F., Schindele, P. & Puchta, H. (2019) Plant breeding at the speed of light: the power of CRISPR/Cas to generate directed genetic diversity at multiple sites. *BMC Plant Biol*, 19, 019-1775
- Xie, K., Minkenberg, B. & Yang, Y. (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci U S A*, 112, 3570-3575
- Xu, J., Brearley, C. A., Lin, W.-H., Wang, Y., Ye, R., Mueller-Roeber, B., *et al.* (2005) A Role of Arabidopsis Inositol Polyphosphate Kinase, AtIPK2 α in Pollen Germination and Root Growth. *Plant Physiology*, 137, 94-103
- Xu, R., Qin, R., Li, H., Li, J., Yang, J. & Wei, P. Enhanced genome editing in rice using single transcript unit CRISPR-*LbCpf1* systems. *Plant Biotechnol J*. 2019 Mar;17(3):553-555. Epub 2018 Nov 19.
- Xue, L. J. & Tsai, C. J. AGEseq: Analysis of Genome Editing by Sequencing. *Mol Plant*. 2015; 8(9):1428-30.

- Yuan, F. J., Zhao, H. J., Ren, X. L., Zhu, S. L., Fu, X. J. & Shu, Q. Y. (2007) Generation and characterization of two novel low phytate mutations in soybean (*Glycine max* L. Merr.). *Theor Appl Genet*, 115, 945-957
- Yuan, F. J., Zhu, D. H., Deng, B., Fu, X. J., Dong, D. K., Zhu, S. L., *et al.* (2009) Effects of two low phytic acid mutations on seed quality and nutritional traits in soybean (*Glycine max* L. Merr.). *J Agric Food Chem*, 57, 3632-3638
- Yuan, F. J., Zhu, D. H., Tan, Y. Y., Dong, D. K., Fu, X. J., Zhu, S. L., *et al.* (2012) Identification and characterization of the soybean IPK1 ortholog of a low phytic acid mutant reveals an exon-excluding splice-site mutation. *Theor Appl Genet*, 125, 1413-1423
- Zhang, Z., Mao, Y., Ha, S., Liu, W., Botella, J. R. & Zhu, J. K. (2016) A multiplex CRISPR/Cas9 platform for fast and efficient editing of multiple genes in Arabidopsis. *Plant Cell Rep*, 35, 1519-1533
- Zhou, H., Liu, B., Weeks, D. P., Spalding, M. H. & Yang, B. (2014) Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Res*, 42, 10903-10914.

ANEXO I

Two efficient CRISPR/Cas9 systems for gene editing in soybean



Two efficient CRISPR/Cas9 systems for gene editing in soybean

Jéssica Carrijo · Eudald Illa-Berenguer · Peter LaFayette ·
Nathalia Torres · Francisco J. L. Aragão · Wayne Parrott ·
Giovanni R. Vianna

Received: 16 December 2020 / Accepted: 17 March 2021 / Published online: 2 April 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract Genome editing using CRISPR/Cas9 has been highlighted as a powerful tool for crop improvement. Nevertheless, its efficiency can be improved, especially for crops with a complex genome, such as soybean. In this work, using the CRISPR/Cas9 technology we evaluated two CRISPR systems, a one-component vs. a two-component strategy. In a simplified system, the single transcriptional unit (STU), *SpCas9* and sgRNA are driven by only one promoter, and in the conventional system, the two-component transcriptional unit (TCTU), *SpCas9*, is under the control of a pol II promoter and the sgRNAs

are under the control of a pol III promoter. A multiplex system with three targets was designed targeting two different genes, *GmIPK1* and *GmIPK2*, coding for enzymes from the phytic acid synthesis pathway. Both systems were tested using the hairy root soybean methodology. Results showed gene-specific edition. For the *GmIPK1* gene, edition was observed in both configurations, with a deletion of 1 to 749 base pairs; however, the TCTU showed higher indel frequencies. For *GmIPK2* major exclusions were observed in both systems, but the editing efficiency was low for STU. Both systems (STU or TCTU) have been shown to be capable of promoting effective gene editing in soybean. The TCTU configuration proved to be preferable, since it was more efficient. The STU system was less efficient, but the size of the CRISPR/Cas cassette was smaller.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11248-021-00246-x>.

J. Carrijo · F. J. L. Aragão · G. R. Vianna (✉)
Embrapa Recursos Genéticos e Biotecnologia, Parque
Estação Biológica, PqEB, Av W5 Norte Final 716,
Brasília, DF 70770-917, Brazil
e-mail: giovanni.vianna@embrapa.br

J. Carrijo · N. Torres · F. J. L. Aragão · G. R. Vianna
Department of Molecular Biology, University of Brasilia,
Brasília, DF 70910-900, Brazil

E. Illa-Berenguer · P. LaFayette · W. Parrott
Center for Applied Genetic Technologies, University of
Georgia, Athens, GA 30602, USA

P. LaFayette · W. Parrott
Department of Crop and Soil Sciences, University of
Georgia, Athens, GA 30602, USA

Keywords Genome editing · Low phytic acid ·
Single transcriptional unit · Two-component
transcriptional unit

Introduction

Over many years, several methodologies have been developed for plant genome editing, such as the use of meganucleases (Adli 2018), zinc-finger nucleases (ZFNs) (Davies et al. 2017) and TALENs

(transcription activator-like effector nucleases) (Du et al. 2016). However, these strategies have shortcomings that limit effective use of these technologies, such as low editing efficiency, complicated vector assembly and off-target mutations (Gaj et al. 2013; Adli 2018). In this context, CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein) has emerged and become prevalent because of its simplicity when recognizing the site to be modified and flexibility of use (Pellagatti et al. 2015; Wolter et al. 2019).

CRISPR/Cas was originally part of the prokaryotic immune system (Wiedenheft et al. 2012). This system is represented by the presence of a Cas endonuclease with two catalytic domains responsible for the cleavage of the double-strand DNA: HNH domain and RuvC nuclease and a single guide RNA complex (sgRNA), which is the fusion of a mature crRNA (CRISPR-derived RNA) and a tracrRNA (trans-activating RNA), creating a functional structure for activation of the endonuclease and recognition of the target sequence (Shan et al. 2013; Barrangou 2015). A Protospacer Adjacent Motifs (PAM) downstream to target sequence determines the anchorage location and the site of the double strand break (DSB) on the DNA strands (Jinek et al. 2012).

Despite major advances in conventional plant breeding, the development of improved plant varieties is proceeding less quickly than necessary, given the increased demand for food caused by the rapid global population growth (Gao 2018; Ahmar et al 2020). Germplasm accesses sometimes have no information available or absence of natural variations to be used for the development of events with desirable characteristics (Marathe et al 2018). For some crops, such as soybean, which have a complex genome (Schmutz et al. 2010), precise and efficient strategies for gene function analysis and crop improvement are of great interest, mainly for breeding programs. The CRISPR/Cas system has already been used to edit genes in soybean (Chilcoat et al. 2017; Cai et al. 2018; Bao et al. 2019; Cheng et al. 2019). However, due to the relatively long time required to obtain a transgenic soybean plant, it is important to have a practical system to validate genome-editing constructs before generating genetically modified plants. Although there are several in silico prediction tools for the design/choice of sgRNA, rapid in vivo experimental studies

might be more reliable in determining the best target sequences (Lee et al. 2017; Zhang et al. 2016).

To find the appropriate system to edit the soybean genome, we evaluated the efficiency of a single transcriptional unit (STU) strategy for the CRISPR/Cas9 system compared to a two-component transcriptional unit (TCTU) using the hairy root-based expression system, evaluating a multiplex sgRNAs. We chose two genes (*GmIPK1* and *GmIPK2*) coding for enzymes from the phytate synthesis pathway (Sparvoli and Cominelli 2015) as a model.

Materials and methods

Selection of CRISPR/Cas target sites

The soybean genomic sequences for two inositol-pentakisphosphate 2-kinase genes (Fig. 1), Glyma.14G072200 and Glyma.12G240900 (*GmIPK1* and *GmIPK2*, respectively), were obtained from Phytozome (<https://phytozome-next.jgi.doe.gov>, Goodstein et al. 2012). The soybean gene models were based on the Wm82.a2.v1 genome assembly (Schmutz et al. 2010). These genes were chosen based on the fact that they are potential candidates for decreased phytic acid content in seeds. *GmIPK1* (Glyma.14G072200) gene is 4158 nucleotides in length distributed in seven exons. Additionally, it has two paralogs, namely Glyma.04G030000 and Glyma.06G03010. Although the gene on chromosome 14 shares similarity at the transcript level with the rest (85.9% and 85% similarity, respectively), they are different at the DNA level. We focused on the *GmIPK1* gene on chromosome 14 because it has the highest expression levels in immature soybean seeds (Yuan et al. 2012), while Glyma.12G240900 (*GmIPK2*) is a much shorter gene (1630 nucleotides) with a single exon.

All possible *Streptococcus pyogenes* Cas9 (SpCas9) target sites within the obtained sequences were identified with Geneious Prime 2019.2.3 (<https://www.geneious.com>). Potential off-target effects were calculated according to the method previously developed by Hsu et al. (2013). Predicted on-target activity was estimated with the online software GPP sgRNA Designer (<https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design>) using Azimuth 2.0 (Doench et al 2016, Sanson et al. 2018). Three target sites were selected for each gene based on their

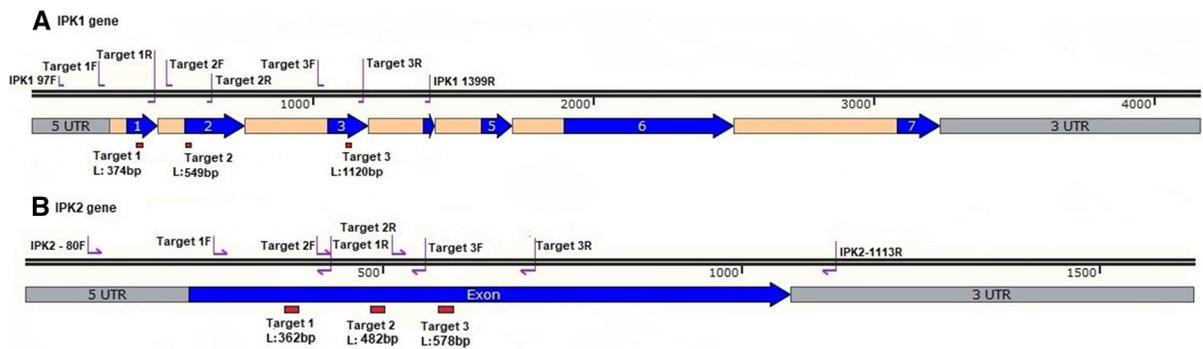


Fig. 1 Scheme showing the targeted sequences for IPK1 gene (a) and IPK2 gene (b) and location of the primers. Blue arrow: exons; light orange oblong: introns; gray oblong: 5' and 3' UTR

(untranslated region). L: Location of target. The scheme shows all primer positions used in this work. The sequence primers are in the supplementary file. (Color figure online)

genomic locations, potential off-target score and predicted on target activity (Table 1).

Vector construction

The binary CRISPR/Cas9 vector p201-EGFP-C9 was assembled to test the STU and TCTU systems (Fig. 2). The vector was constructed using the p201-EGFP-C9, which contains the *EGFP* (Enhanced Green Fluorescent Protein) (Chiu et al. 1996) reporter gene under the control of the CsVMV promoter (from *cassava vein mosaic virus*) (Verdaguer et al. 1996), the *nos* gene terminator (An et al. 1985) and the coding sequence for *SpCas9* (Mali et al. 2013), which was obtained as Addgene plasmid #41815 and is under control of the *GmUbi3* promoter (Chiera et al. 2007) and the *PsRbcs* terminator (Schardl et al. 1987). The sgRNAs of the TCTU configurations are under the control of the *MtU6* promoter from *Medicago truncatula* (Kim et al. 2013). In the STU system (Wang et al. 2018), each sgRNA target sequence was combined with the optimized scaffold (Dang et al. 2015) and separated by a unique sequence (UNS), and these were used as linkers to facilitate cloning (Torella et al. 2014).

The CRISPR/Cas9 TCTU configuration cassettes were assembled and cloned into the p201-EGFP-C9 digested with *SpeI*, generating the final 15,454 bp vector (*MtU6*-sgRNA(thrice)-*GmUbi*-*SpCas9*-*RbcsT*) (Fig. 2a). For assembly of the STU vectors, the cassettes were cloned into the p201-EGFP-C9 digested with *AvrII*, generating the final 14,396 bp vectors (*GmUbi*-*SpCas9*-sgRNA(thrice) -*RbcsT*) (Fig. 2b).

Soybean hairy root transformation

The *Glycine max* cultivar Jack was used in this study. This variety is the best for somatic embryogenesis and plant regeneration (Raza et al. 2020). Soybean hairy root transformation was performed as previously described by Jacobs et al. (2015) with no selection pressure. EGFP fluorescence was used to detect transgenic roots 20 days after transformation. Each positive root was considered as a single event and used for downstream analysis. Additionally, PCR was used to detect changes. A 1322-bp amplicon was expected for the *GmIPK1* gene using the primers IPK1-97F: ACACAATTCCTTTCCCACCA and IPK1-1399R: AGCAGAGGCTAGATCCTTGA. For the *GmIPK2* gene, a 1053-bp amplicon was expected using primers IPK2-80F: TTGCATTGCTTTGTGTAAAGG and IPK2-1113R: CTGCGACACTAATTCAAGCA.

Detection of CRISPR/Cas9-mediated gene editing

Genomic DNA was isolated from GFP-positive roots (approximately 2 mm from the root tip) using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA). Gene editing patterns were determined by amplicon sequencing using specific primers tailed with partial TruSeq Illumina adapter sequences (Supplementary Table 1). PCR products from the same sample were pooled and individually barcoded using indexed universal iTru/iNext primers. Indexed samples were pooled in equimolar amounts and sequenced on an Illumina MiSeq using 150-bp paired-end (2 × 150 PE) mode at the Georgia Genomics and Bioinformatics Core, University of

Table 1 Characteristics of selected sgRNAs for *GmIPK1* and *GmIPK2*

Name guide	Guide sequence ¹	PAM	GC (%)	Genome position (Wm82.a2.v1)	Direction ²	Mismatch number ³	Off-target score ⁴ (%)	On-target efficacy score (CRISPRko) ⁵
GmIPK1-target1	GCAGCTGACTGGGTTTACAG	AGG	55	Gm14:6063363..6063382	Sense	4	3.64	0.6765
GmIPK1-target2	GAAAGTGGTCCGCATACGTA	AGG	50	Gm14:6063188..6063207	Sense	4	1.87	0.5864
GmIPK1-target3	GTGGCGAGTTGAAGCTGCCA	GGG	61	Gm14:6062617..6062636	Sense	3	7.11	0.4512
GmIPK2-target1	GGTGTAAGAGGAGAGTTCGG	TGG	50	Gm12:39951279..39951298	Antisense	3	4.57	0.6646
GmIPK2-target2	GGTCCTCCAGGACCAGGTGA	GGG	65	Gm12:39951158..39951177	Antisense	4	3.27	0.5795
GmIPK2-target3	GCACTTGCAGATGTAGTCTT	CGG	50	Gm12:39951063..39951063	Antisense	3	7.96	0.6359

¹sgRNAs were identified using Geneious Prime 2019.2.3

²Direction based on Wm82.a2.v1 gene model annotation

³Minimum number of mismatches between sgRNA sequences and possible off-target sequences

⁴Top off-target score identified. Off-target score is based on Hsu et al. (2013)

⁵GFP sgRNA Designer (CRISPRko) based on Doench et al. (2016) and Sanson et al. (2018). Scale from 0 to 1

Georgia. Amplicon sequencing data were analyzed using AGESeq software (Xue and Tsai 2015).

Results

A single transcriptional unit (STU) system was compared to the two-component transcriptional (TCTU) system in soybean using hairy root transformation. We evaluated two strategies to express multiple sgRNAs for the genes *GmIPK1* and *GmIPK2* that code for two different phytic acid biosynthetic enzymes: *GmIPK2* codes for an inositol-polyphosphate multikinase, mainly involved in the production of inositol 1,3,4,5,6-pentakisphosphate, while *GmIPK1* codes for an inositol-pentakisphosphate 2-kinase, which phosphorylates inositol 1,3,4,5,6-pentakisphosphate in position 2, releasing phytic acid (Sparvoli and Cominelli, 2015). Genomic sequences of both genes were obtained from the *G. max* genome (assembly Wm82.a2.v1, Schmutz et al. 2010). A total of 46 SpCas9 target sites were identified for *GmIPK1*, 22 of them in the coding sequence (CDS), from which 3 target sites were chosen. Targets were located in exon 1, exon 2 and exon 3, respectively, and were named GmIPK1-target1, GmIPK1-target2 and GmIPK1-target3. Their potential off-target score ranged from 1.87 to 7.11% (all below the 10% threshold) with predicted on-target activity ranging from 45.12 to 67.65%. All three targets are in sense orientation. Of the 65 target sequences identified from the *GmIPK2* gene, 47 targeted the CDS. The three selected guides for this gene were all chosen in the anti-sense orientation and separated by at least 75 bp. Their potential off-target score ranged from 3.27 to 7.96% and predicted on-target activity between 57.96 and 66.46%. Selected targets for *GmIPK2* were named GmIPK2-target1, GmIPK2-target2 and GmIPK2-target3. Details on selected targets are summarized on Table 1.

Four different p201-EGFP-C9 plasmids, IPK1-TCTU, IPK1-STU, IPK2-TCTU and IPK2-STU, were used for soybean hairy root transformation. For each construct, ten cotyledons from germinating seeds were transformed (Fig. 3a), and transgenic hairy roots were confirmed by visual inspection of EGFP fluorescence (Fig. 3b). PCR analysis was used for a primary evaluation to detect changes before Sanger sequencing; 1322- and 1053-bp amplicons were expected for

A Two-Component Transcriptional Unit- TCTU



B Single Transcriptional Unit - STU

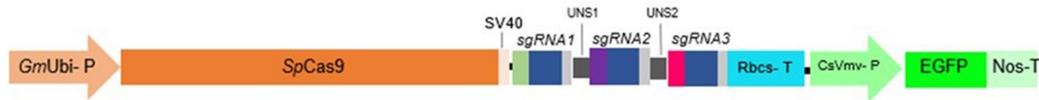


Fig. 2 Schematic diagram of two CRISPR/Cas9 expression systems: dual promoter system (TCTU) and single transcriptional unit (STU) system. Hairy root vector CRISPR/Cas9 is used to edit the *GmIPK1* and *GmIPK2* genes with the two systems: TCTU (a) and STU (b). The Cas 9 endonuclease from *Streptococcus pyogenes* (*SpCas9*) is under the control of the ubiquitin-3 promoter from glycine max (orange). The cassette of

the EGFP reporter gene is under the control of the CsVMV (*cassava vein mosaic virus*) promoter (light green) and T-nos terminator (nopaline synthase). In the TCTU system each sgRNA is driven by the U6 promoter from *Medicago truncatula* separated for a UNS sequence. In the STU system the sgRNAs are only separated for the UNS sequence

the *GmIPK1* and *GmIPK2* genes, respectively (Fig. 3c). Of a total of 15 individual EGFP-positive roots collected from each transformation, genomic DNA was isolated, and targeted amplicon sequencing libraries were prepared to determine CRISPR/Cas9-mediated gene editing.

Illumina MiSeq paired-end sequencing technology yielded a total of 376,180 paired-end clean reads from *GmIPK1* and *GmIPK2* amplicon-seq libraries. Subsequent analysis showed that 180,971 (48.11%) and 195,209 (51.89%) paired-end reads were obtained from *GmIPK1* and *GmIPK2*, respectively. CRISPR/Cas-induced gene modification patterns were analyzed using AGESeq software (Supplementary Table 2 and Supplementary Table 3). The STU system showed very low editing rates for the *GmIPK1* gene, with wild-type sequences detected on 99.1% of the reads. Six of 15 samples had no edits, while 5 only had < 2% editing at 2 of the 3 sites. Editing occurred mainly in four different events. *GmIPK1*-STU-#04 reads with a 1-bp deletion were found in all three sites at a low but detectable level (< 0.3%). The sequence between *GmIPK1*-Target 1 and *GmIPK1*-Target 3 (751 bp) was deleted in 156 of the reads (3.99% and 5.40%) (Fig. 4). *GmIPK1*-STU-#06 only displayed modifications for *GmIPK1*-Target 3, with a 5-bp deletion detected in 19% of the reads. Modifications were identified in all three sites of the *GmIPK1*-STU-#08 event (Fig. 4). A 2-bp deletion was detected in *GmIPK1*-Target 1 with a frequency > 87%. At the same time, a 4-bp deletion sequence was observed in 18.52% of the reads of *GmIPK1*-Target 3. Finally, a

reduced percentage of the reads mapped to *GmIPK1*-Target 2 harbored modifications, with nearly 4% of the reads showing a 3-bp deletion and another 3% showing a single base insertion. For *GmIPK1*-STU-#09, a 749-bp deletion comprising the sequence between *GmIPK1*-Target 1 and *GmIPK1*-Target 3 was observed.

Higher indel frequencies were observed when the TCTU system was tested on the *GmIPK1* gene (73.20% reads carried a mutation) (Fig. 4). Different sgRNA sequences led to a wide collection of mutations, ranging from small deletions (1 to 12 bp) or insertions (only 1 bp) to large deletions that removed the sequences between two sgRNAs. Only two events (*GmIPK1*-TCTU-#02 and *GmIPK1*-TCTU-#14) harbored no modifications. The most commonly found mutations for both STU and TCTU are as follows: for *IPK1*-Target 3–751 bp, –4 bp and –1 bp; for *GmIPK1*-Target 2 –1 bp (STU), –3 and –7 bp only in TCTU; for *IPK1*-Target 1–751 bp, –3 bp and –4 bp (only for the TCTU). As previously detected in *IPK1*-STU samples, deletions of about 751 bp (comprising the sequence between *GmIPK1*-Target 1 and *GmIPK1*-Target 3) are frequently present in *IPK1*-TCTU samples (8/15). Additional gene segment deletions between *GmIPK1*-Target 1 and *GmIPK1*-Target 2 (*GmIPK1*-TCTU-#04 and *GmIPK1*-TCTU-#06), *GmIPK1*-Target 2 and *GmIPK1*-Target 3 (*GmIPK1*-TCTU-#05) were detected as well. Average editing frequencies of 89.30%, 69.13% and 79.39% were estimated for *GmIPK1*-Target 1, *GmIPK1*-Target 2 and *GmIPK1*-Target 3, respectively.

However, some single events such as GmIPK1-TCTU-#09, GmIPK1-TCTU-#12 and GmIPK1-TCTU-#15 show editing levels > 95% in all three sites.

For the *GmIPK2* gene (Fig. 5), the STU system delivered overall 96.38% of wild-type sequencing reads. GmIPK2-Target 1 and GmIPK2-Target 3 showed similar editing rates (12.71% and 13.76%, respectively). In contrast, GmIPK2-Target 2 presented an almost undetectable editing rate of 0.04%. The most identified modification was a deletion of about 215 bp between the GmIPK2-Target 1 and GmIPK2-Target 3 target sites. This modification was present in all 15 events evaluated using the STU system with frequencies up to 54.08%/33.71% (event GmIPK2-STU-#11). No plants were identified without CRISPR/Cas9-mediated edits. Smaller deletions (1 or 2 bp) were also common modifications identified on IPK2-STU events. Again, the TCTU system induced a much higher number of modifications than the STU system (79.33% and 96.38%), but the differences between the STU and TCTU systems were less pronounced for *GmIPK2* than for *GmIPK1*. Despite all events showing some kind of edition, GmIPK2-Target 2 had the lowest editing rate with only a 0.67% rate (10 out of the 15 events showed only wild-type reads). The most frequent edit was the excision of the fragment between GmIPK2-Target 1 and GmIPK2-Target 3. Another large deletion, i.e., between GmIPK2-Target 1 and GmIPK2-Target 2, was only detected in event

GmIPK2-TCTU-#05. Events GmIPK2-TCTU-#08 and GmIPK2-TCTU-#04 presented a higher editing rate, ranging between from 60.85% and 41.57%. Only IPK2-TCTU-#04 presented any editing on GmIPK2-Target 2. GmIPK1-TCTU-#02 library had some issues and was sequenced at low depth coverage, resulting in no sequencing information available on the GmIPK1-target 2.

Discussion

Although CRISPR technology has been revolutionizing gene edition since its discovery, new advances have made this system even more efficient (Fiaz et al 2019). Traditionally, the CRISPR/Cas9 system uses a combination of two types of promoters, one from the class II Pol promoter that regulates Cas endonuclease expression and other from the class III that regulates the sgRNA (Tang et al. 2019). Some limitations have been become apparent after it was developed because of the poor characterization of pol III promoters in some organisms (Sun et al. 2015) and uncoordinated activity of Cas9 and sgRNA expression, given they are being driven by different promoters (Tang et al. 2019), or repetitive use of the sequence that might cause variation of expression levels and transgene silencing (Ma et al. 2015). The CRISPR/Cas9 system can be simplified even further without compromising its efficiency. A more compact system such as the STU

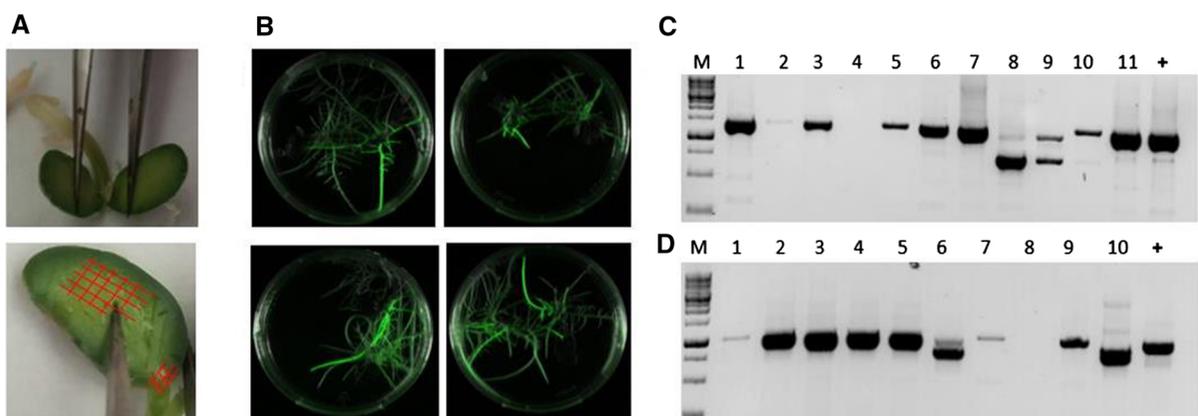


Fig. 3 **a** Hairy-root process and detection of fragments Preparation of the cotyledons, showing the cut for inoculation (outlined in red); **b** roots grown in MS medium under EGFP flashlight; **c** detection of the fragment of the *GmIPK1* gene by PCR on agarose gel under UV light, expected fragment of

1233 bp. Lines 1–6: IPK1-STU, lines 7–11: IPK1-TCTU, +: plant control. **d** Detection of the fragment of the *GmIPK2* gene by PCR on agarose gel under UV light, expected fragment of 1053 bp. Lines 1–5: IPK1-STU, lines 7–10: IPK1-TCTU, +: plant control. **M**: 1-kb ladder LIBPBio

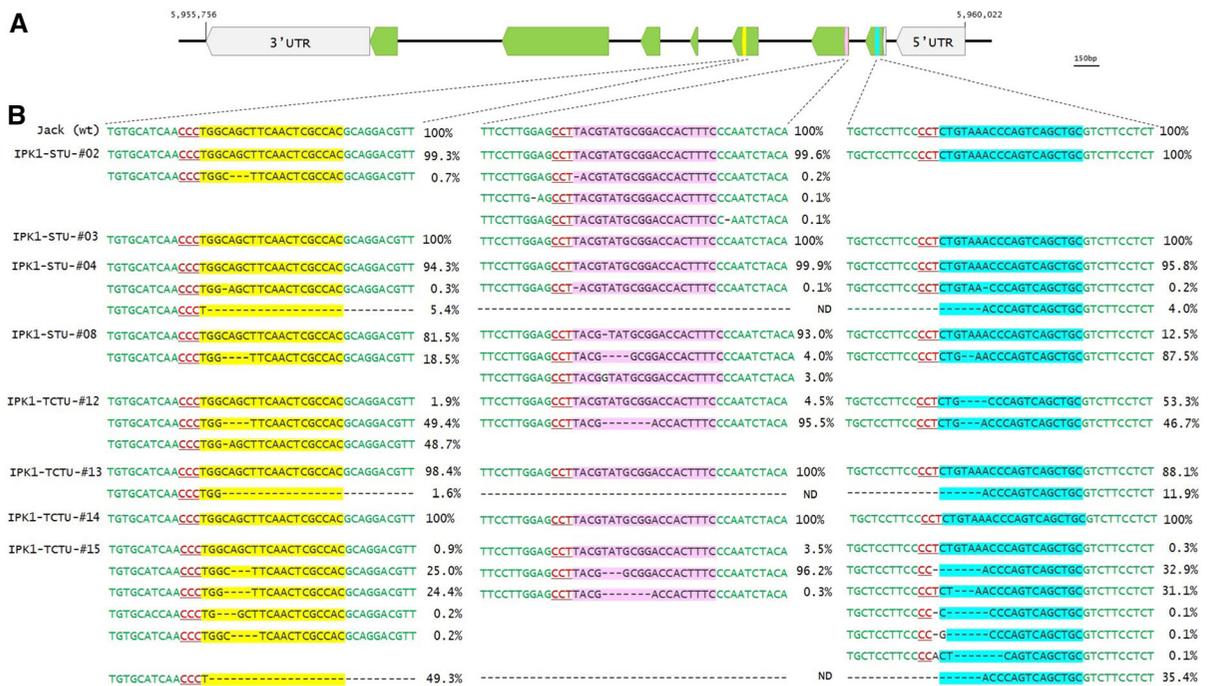


Fig. 4 Comparison of the single transcriptional unit (STU) and two-component transcriptional unit (TCTU) CRISPR/Cas9 gene editing systems in soybean hairy roots for the *GmIPK1* gene. **a** Schematic illustration of the *GmIPK1* gene (Glyma.14G072200). All exons are represented by solid green arrows, and introns are represented as lines. Both the 5' and 3' untranslated regions (UTR) are depicted as gray boxes. *GmIPK1*-Target 1, *GmIPK1*-Target 2 and *GmIPK1*-Target 3 are represented by blue, pink and yellow arrows, respectively.

b CRISPR/Cas9 induced mutations by the STU and TCTU systems. Wild-type sequences are in green, deletions are shown as dashes, and SNPs are shown in black. The targeted sequences are highlighted in blue, pink and yellow, and the PAM is underlined in red. Percentages next to the sequences indicate the number of reads mapped over the total number of reads sequenced for a given target within each sample. ND: Not detected. (Color figure online)

design would enhance flexibility and simplify gRNA construction. In this context, we evaluated the efficiency of editing two soybean genes by using CRISPR/Cas9 with the sgRNA in two different configurations, i.e., the conventional (TCTU) and simplified (STU) strategies. It was possible to reduce the CRISPR/Cas9 expression cassette from 7050 bp in the TCTU system to 5990 bp in the STU system, a reduction of > 1000 bp. Given that each sgRNA has 120 bp, the remaining 1000 bp is equivalent to approximately eight new sgRNAs in a simplified design, which means eight new targets for other genes in a multiplex system.

In this work, when we compare the systems, the TCTU and STU gave gene-specific results, but both were able to create indels from DSBs. These included deletions ranging from 1 to 10 bp for *IPK2*. For *IPK1*, the deletion size varied from 1 to 12 bp. They also

included 1-bp insertions that were detected for both genes. In addition, we obtained a 752-bp deletion in *GmIPK1* gene when the fragment between *GmIPK1*-Target 1 and -Target 3 was deleted. For the *GmIPK2* gene, the largest deletion was 217 bp, which is the distance between *GmIPK2*-Target 1 and -Target 3. Overall, the major difference is that the TCTU system consistently gave higher editing frequencies for both genes. This difference in frequency may simply indicate that the TCTU is the superior system. The STU system employed here was based on reports by Wang et al. (2018) and Mikami et al. (2017) wherein Cas9 multiplexed gRNAs were shown to achieve editing efficiencies from 50 to 94%. An advantage of relying on the plant endogenous processing machinery to cleave the RNA is that it simplifies gRNA construct assembly. Other approaches have used the Cys4 RNA cleaving system (Cermak et al. 2017), exogenous

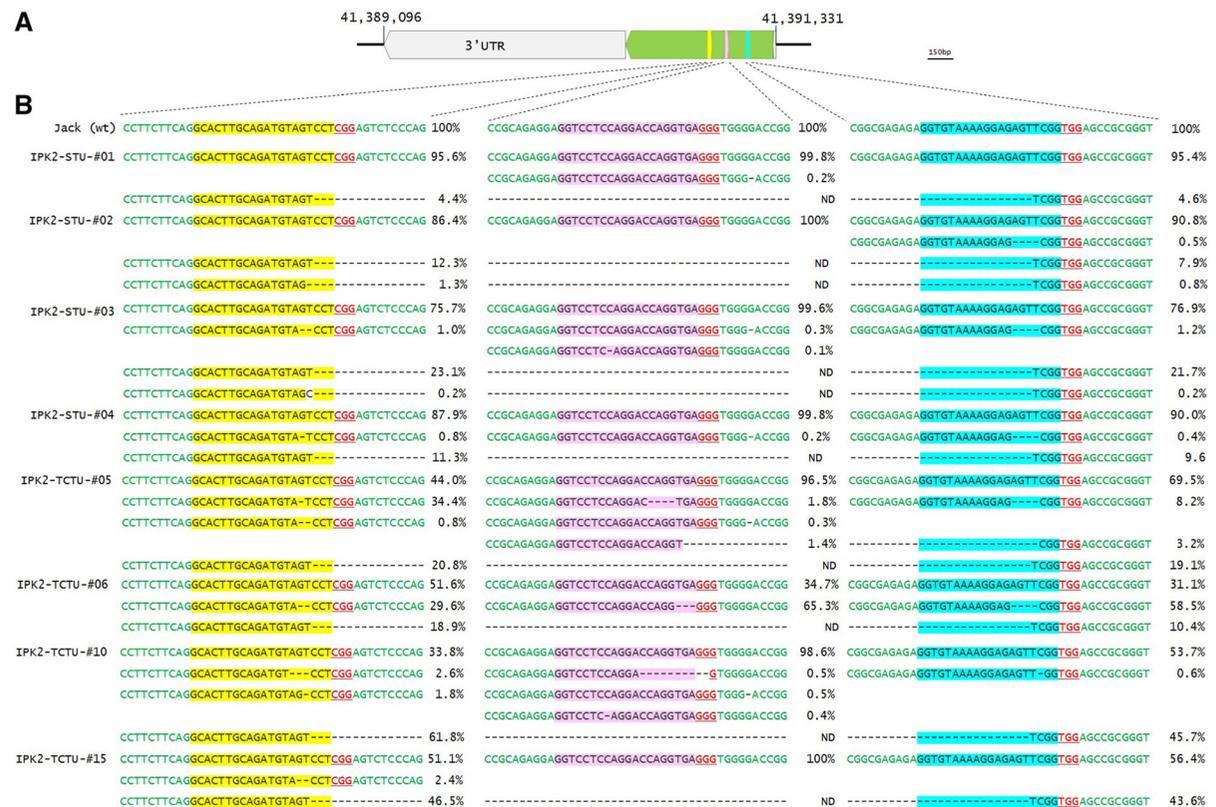


Fig. 5 Comparison of the single transcriptional unit (STU) and two-component transcriptional unit (TCTU) CRISPR/Cas9 gene editing systems in soybean hairy roots for the *GmIPK2* gene (Glyma.12G240900). **a** Schematic representation of the *GmIPK2* gene. All exons are represented by solid green arrows, and introns are shown as lines. The 5' and 3' untranslated regions (UTR) are shown as gray boxes. *GmIPK2*-Target 1, *GmIPK2*-Target 2 and *GmIPK2*-Target 3 are represented by blue, pink and yellow arrows, respectively. **b** CRISPR/Cas9

ribozymes (Gao et al. 2015; Tang et al. 2016) and the polycistronic tRNA-gRNA gene system (Xie et al. 2015) to release individual gRNAs, and these approaches may circumvent the low frequencies obtained here with the STU system. A similar simplified transcriptional unit CRISPR system (STU) was tested in rice and showed editing rate efficiencies between 29 and 38%, and the efficiency remained the same when compared to the traditional system. Point mutations were observed at the cleavage site or at most two bases ahead (Tang et al. 2016). With additional optimization, the mutation rate was close to 50% across six targets (Tang et al. 2019). Different endonucleases were tested with the STU configuration

induced mutations in both STU and TCTU systems. Wild-type sequences are in green, deletions are shown as dashes, and SNPs are shown in black. The targeted sequences are highlighted blue, pink and yellow, and the PAM is highlighted and underlined in red. Percentages next to sequences indicate the number of reads mapped over the total number of reads sequenced for a given target within each sample. ND: Not detected. (Color figure online)

(*Fn*Cpf1, *Lb*Cpf1 or Cas9) using a multiplex system, and the efficiency was > 50% (Wang et al. 2018).

The successful activity from the CRISPR/Cas9 complex activity is associated with target sequence characteristics (Doench et al. 2014), but we are not aware of reports describing preferences based on target position within a multiplex system. Notably, the entire fragment between targets 1 and 3 was more frequently removed for the *GmIPK2* gene (smaller gene) as all events from both STU and TCTU had this deletion in frequencies up to 60.8% for event 4 from the TCTU system. For the *GmIPK1* gene, the entire segment between targets 1 and 3 (approximately 715 bp long, depending on the cut) was lost in 4 of 15

events with the STU system and 8 of 15 events for the TCTU system. Edits for target 2 of the *GmIPK2* gene were rare. The *GmIPK2* is the smaller gene, and the proximity of the target sites (101 bp between target 1 and 2, 76 bp between target 3 and 2) might have led to interference with the Cas endonucleases.

In diploid cells, three zygositys are possible, monoallelic, heterozygous diallelic (more commonly referred to as biallelic) and homozygous diallelic (Luttgeharm et al. 2017). Knowing the nature of the mutations helps understand the efficiency of each system. In our hands, biallelic editing is rare for the STU system, although event 8 for *GmIPK1* had an 87.5% biallelic editing frequency for target 1, consisting of a 2-bp deletion. For *GmIPK1* with the TCTU system, all three targets were biallelic for events 4 and 9, while events 6 and 8 were monoallelic for targets 1 and 3. Event 10 was also biallelic for targets 1 and 3, but monoallelic for target 2. Diallelic edits were predominantly heterozygous. None of the targets in the *GmIPK2* gene resulted in diallelic edits with either system.

In conclusion, the TCTU and STU systems were effective in editing soybean genes coding for enzymes from the phytic acid synthesis pathway, though the higher editing frequencies obtained with the TCTU system make this the preferred technology. This technology will be the foundation for efficiently editing genes in the soybean genome as well as production of low phytic acid genotypes.

Acknowledgements This study was supported in part by Empresa Brasileira de Pesquisa Agropecuária (Embrapa), with funds allocated to the Georgia Agricultural Experiment Stations (Athens, GA, USA) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil). J. Carrijo was supported by a fellowship from Cordenção de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil).

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by JC, EB and NT. The first draft of the manuscript was written by JC and GRV, and all authors commented on and revised previous versions of the manuscript. All authors read and approved the final manuscript.

Availability of data and materials All raw data using in this manuscript will be made available upon request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Consent to participate and publication All authors consented to their participation and the publication of this manuscript.

References

- Adli M (2018) The CRISPR tool kit for genome editing and beyond. *Nat Commun* 9:1911
- Ahmar S, Gill RA, Jung KH, Faheem A, Qasim MU, et al. (2020). Conventional and molecular techniques from simple breeding to speed breeding in crop plants: recent advances and future outlook. *Int J Mol Sci.* 2020 21(7):2590
- An G, Watson BD, Stachel S, Gordon MP, Nester EW (1985) New cloning vehicles for transformation of higher plants. *EMBO J* 4:277–284
- Bao A, Chen H, Chen L, Chen S, Hao Q et al (2019) CRISPR/Cas9-mediated targeted mutagenesis of *GmSPL9* genes alters plant architecture in soybean. *BMC Plant Biol* 19(1):131
- Barrangou R (2015) Diversity of CRISPR-Cas immune systems and molecular machines. *Genome Biol* 16:247
- Cai Y, Chen L, Sun S, Wu C, Yao W, et al. (2018) CRISPR/Cas9-Mediated deletion of large genomic fragments in soybean. *Int J Mol Sci* 19.
- Cermak T, Curtin SJ, Gil-Humanes J, Cegan R, Kono TJY et al (2017) A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell* 29:1196–1217
- Cheng Q, Dong L, Su T, Li T, Gan Z et al (2019) CRISPR/Cas9-mediated targeted mutagenesis of *GmLHY* genes alters plant height and internode length in soybean. *BMC Plant Biol* 19:562
- Chiera JM, Bouchard RA, Dorsey SL, Park EH, Buenrostro-Nava MT, Ling PP, Finer JJ (2007) Isolation of two highly active soybean (*Glycine max* (L.) Merr.) promoters and their characterization using a new automated image collection and analysis system. *Plant Cell Rep* 26:1501–1509
- Chilcoat D, Liu Z-B, Sander J (2017) Use of CRISPR/Cas9 for Crop Improvement in Maize and Soybean. *Prog Mol Biol Transl Sci* 149:27–46
- Chiu W-L, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J (1996) Engineered GFP as a vital reporter in plants. *Curr Biol* 6:325–330
- Dang Y, Jia G, Choi J, Ma H, Anaya E et al (2015) Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol* 16:280
- Davies JP, Kumar S, Sastry-Dent L (2017) Use of Zinc-Finger Nucleases for Crop Improvement. *Progress in Molecular Biology and Transl Sci* 149:47–63
- Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M et al (2014) Rational design of highly active sgRNAs for

- CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* 32:1262–1267
- Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Root DE (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 34(2):184–191
- Du H, Zeng X, Zhao M, Cui X, Wang Q et al (2016) Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *J Biotechnol* 217:90–97
- Fiaz S, Ahmad S, Riaz A, Noor MA, Wang X et al (2019) Applications of the CRISPR/Cas9 System for Rice Grain Quality Improvement: Perspectives and Opportunities. *Int J Mol Sci* 20(4):888
- Gaj T, Gersbach CA, Barbas CF 3rd (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 31:397–405
- Gao Y, Zhang Y, Zhang D, Dai X, Estelle M, Zhao Y (2015) Auxin binding protein 1 (ABP1) is not required for either auxin signaling or Arabidopsis development. *Proc Natl Acad Sci U S A* 112(7):2275–2280
- Gao C (2018) The future of CRISPR technologies in agriculture. *Nat Rev Mol Cell Biol* 19:275–276
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, et al. (2012). *Phytozome: a comparative platform for green plant genomics*. *Nucleic acids research*, 40(Database issue), D1178–D1186.
- Hsu PD, Scott DA, Weinstein J, Ran FA, Konermann S et al (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31(9):827–832
- Jacobs TB, LaFayette PR, Schmitz RJ, Parrott WA (2015) Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol* 15:16
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821
- Kim GB, Nam YW (2013) Isolation and characterization of *Medicago truncatula* U6 promoters for the construction of small hairpin RNA-mediated gene silencing vectors. *Plant Mol Biol Report* 31(3):581–593
- Lee CM, Zhu H, Davis TH, Deshmukh H (2017) Bao G (2017) Design and Validation of CRISPR/Cas9 Systems for Targeted Gene Modification in Induced Pluripotent Stem Cells. *Methods Mol Biol* 1498:3–21
- Luttgeharm KD, Wong K-S, Siembieda S (2017) Heteroduplex cleavage assay for screening of probable zygosity resulting from CRISPR mutations in diploid single cell lines. *Biotechniques* 62:268–274
- Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, Wang B et al (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol Plant* 8:1274–1284
- Mali P, Yang LH, Esvelt KM, Aach J, Guell M, DiCarlo JE et al (2013) RNA-Guided human genome engineering via Cas9. *Science* 339(6121):823–826
- Marathe A, Krishnan V, Mahajan MM, Thimmegowda V, Dahuja A, et al, (2018). A. Characterization and molecular modeling of Inositol 1,3,4 tris phosphate 5/6 kinase-2 from *Glycine max* (L) Merr.: comprehending its evolutionary conservancy at functional level. *3 Biotech*. 8 (1):50.
- McCarty NS, Graham AE, Studená L, Ledesma-Amaro R (2020) Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat Commun* 11(1):1281
- Mikami M, Toki S, Endo M (2017) In Planta Processing of the SpCas9-gRNA Complex. *Plant Cell Physiol* 58:1857–1867
- Pellagatti A, Dolatshad H, Valletta S, Boulwood J (2015) Application of CRISPR/Cas9 genome editing to the study and treatment of disease. *Arch Toxicol* 89:1023–1034
- Punjabi M, Bharadvaja N, Sachdev A, Krishnan V (2018) Molecular characterization, modeling, and docking analysis of late phytic acid biosynthesis pathway gene, inositol polyphosphate 6-/3-/5-kinase, a potential candidate for developing low phytate crops. *8 (8): 344*
- Raza G, Singh MB, Bhalla PL (2020) Somatic Embryogenesis and Plant Regeneration from Commercial Soybean Cultivars. *Plants* 9(1):38
- Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, et al (2018) Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* 9:5416
- Schardl CL, Byrd AD, Benzion G, Altschuler MA, Hildebrand DA et al (1987) Design and construction of a versatile system for the expression of foreign genes in plants. *Gene* 61:1–11
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183
- Shan Q, Wang Y, Li J, Zhang Y, Chen K et al (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* 31(8):686–688
- Sparvoli F, Cominelli E (2015) Seed Biofortification and Phytic Acid Reduction: A Conflict of Interest for the Plant? *Plants* 4:728–755
- Sun X, Hu Z, Chen R, Jiang Q, Song G, et al. (2015) Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Sci Rep* 5.
- Tang X, Zheng X, Qi Y, Zhang D, Cheng Y, Tang A, Voytas DF et al (2016) A single transcript CRISPR-Cas9 system for efficient genome editing in plants. *Mol Plant* 9:1088–1091
- Tang X, Ren Q, Yang L, Bao Y, Zhong Z et al (2019) Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing. *Plant Biotechnol J* 17:1431–1445
- Torella JP, Lienert F, Boehm CR, Chen JH, Way JC et al (2014) Unique nucleotide sequence-guided assembly of repetitive DNA parts for synthetic biology applications. *Nat Protoc* 9:2075–2089
- Verdagner B, de Kochko A, Beachy RN et al (1996) Isolation and expression in transgenic tobacco and rice plants, of the cassava vein mosaic virus (CVMV) promoter. *Plant Mol Biol* 31:1129–1139. <https://doi.org/10.1007/BF00040830>
- Wang M, Mao Y, Lu Y, Wang Z, Tao X, et al. (2018) Multiplex gene editing in rice with simplified CRISPR-Cpf1 and CRISPR-Cas9 systems. *J Integr Plant Biol*. 2018;60(8):626–631.
- Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482:331–338
- Wolter F, Schindele P, Puchta H (2019) Plant breeding at the speed of light: the power of CRISPR/Cas to generate directed genetic diversity at multiple sites. *BMC Plant Biol* 19:176

- Xie K, Mikenberg B, Yang Y (2014) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci USA* 112(11):3570–3575
- Xue LJ, Tsai CJ (2015) AGEseq: Analysis of Genome Editing by Sequencing. *Mol Plant* 8:1428–1430
- Yuan FJ, Zhu DH, Tan YY, Dong DK, Fu XJ et al (2012) Identification and characterization of the soybean IPK1 ortholog of a low phytic acid mutant reveals an exon-excluding splice-site mutation. *Theor Appl Genet* 125(7):1413–1423
- Zhang Z, Mao Y, Ha S, Liu W, Botella JR et al (2016) A multiplex CRISPR/Cas9 platform for fast and efficient editing of multiple genes in Arabidopsis. *Plant Cell Rep* 35:2151–2153

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.