



UNIVERSIDADE DE BRASÍLIA

Instituto de Ciências Biológicas

Departamento de Biologia Celular

Programa de Pós-graduação em Ciências Biológicas - Biologia Molecular

MATEUS MEIRA DOS SANTOS

**ESTRATÉGIA MOLECULAR PARA EDIÇÃO GÊNICA DO INIBIDOR
BOWMAN-BIRK MAJORITÁRIO EM SEMENTES DE SOJA**

Brasília, 2026



MATEUS MEIRA DOS SANTOS

**ESTRATÉGIA MOLECULAR PARA EDIÇÃO GÊNICA DO INIBIDOR
BOWMAN-BIRK MAJORITÁRIO EM SEMENTES DE SOJA**

Dissertação apresentada ao Programa de Pós-Graduação em Biologia Molecular da Universidade de Brasília, como requisito parcial para a obtenção do título de mestre.

Orientadora: Dra. Maria Fatima Grossi-de-Sa

Coorientadora: Dra. Maria Eugênia Lisei de Sá

Brasília, 2026

MATEUS MEIRA DOS SANTOS

**ESTRATÉGIA MOLECULAR PARA EDIÇÃO GÊNICA DO INIBIDOR
BOWMAN-BIRK MAJORITÁRIO EM SEMENTES DE SOJA**

Dissertação apresentada ao Programa de Pós-Graduação
em Biologia Molecular da Universidade de Brasília, como
requisito parcial para a obtenção do título de Mestre.

Aprovada em: ____/____/____

BANCA EXAMINADORA

Dra. Maria Fatima Grossi-de-Sa
(Orientadora / Presidente)
Universidade de Brasília - UnB

Dr. Marcelo Macedo Brigido
(Membro Interno)
Universidade de Brasília - UnB

Dr. Bruno Paes de Melo
(Membro Externo)
Centro de Tecnologia Canavieira - CTC

Dr. Fabrício Barbosa Monteiro Arraes

(Membro Suplente)

SEMPRE AgTech

Dedico esta dissertação aos meus professores e tutores, que sempre me incentivaram a seguir a carreira acadêmica. Dedico ao Mateus de 5 anos, que amava fazer experiências malucas e que, hoje, se tornou o cientista que sonhou.

AGRADECIMENTOS

À minha família, que deu suporte à minha trajetória, incentivou, sustentou, acompanhou e amou desde o início da minha vida até os dias de hoje. Com vocês, foi possível passar pela vida acadêmica sabendo que eu sempre podia contar com esse apoio. À minha mãe, Adriana, que me ensinou a ser curioso, questionador, honesto e íntegro. Ao meu pai, Mário, por me ensinar que trabalho duro e dedicação superam talento, sorte e privilégios. Aos meus irmãos, por serem um porto seguro, um lugar de riso, alegria, descontração e desconpressão. À minha avó Cipriana, que, pelo exemplo, me ensinou a ser crítico, valorizar a ciência e o conhecimento, me ensinando o que era o método científico mesmo sem saber. À memória de minha avó Rita, que me auxiliou emocional e financeiramente até seus últimos dias, permitindo, em conjunto com toda a minha família, que eu me formasse. Aos meus tios Mauro, Marina e Mara, que sempre estiveram ao meu lado: esta conquista acadêmica, do primeiro da família a se tornar Mestre e futuramente Doutor, também é de vocês. À minha família do coração, os Barbalhos, que me aceitaram, acolheram e apoiam todos os dias: vocês sabem que consegui chegar até aqui, também graças ao apoio de vocês.

À minha orientadora, Dra. Fatima Grossi. Sua história, seu exemplo, sua força e sua intuição me mostram a cada dia por onde seguir como pessoa e como cientista. Agradeço pela oportunidade e pela confiança depositadas em mim para desenvolver este projeto.

À Dra. Maria Eugênia minha Coorientadora, por ter sido a primeira a me dar oportunidade e confiança na Embrapa. Agradeço pelo companheirismo, pelos ensinamentos e por me mostrar, a cada dia, como eu poderia ser um ser humano e cientista melhor. Por me ensinar pelo exemplo, por me ouvir, por toda a ajuda técnica e emocional, por me representar e me defender quando preciso, por cada puxão de orelha e por cada elogio. Sou extremamente grato por cada momento que passei com você.

Às minhas primeiras orientadoras, Dra. Sarah Caldas e Dra. Cintia Coelho, que primeiramente me ensinaram a amar as plantas e a genética, abrindo o caminho e a minha mente, possibilitando que eu chegasse até aqui.

Aos meus amigos de bancada: Ítalo, Jéfyne, Nattany, Sara, Luanna, Thuanne, Mariana, Raíre, Andressa, Gustavo, Júlia, Lorena, Manoely, e tantos outros que passaram pelo laboratório. Agradeço por cada dica, por cada vez que me disseram que eu podia contar com vocês. Pelas risadas, pelos shades, pelas festas do laboratório, por cada café e biscoito após o almoço. Pelos bares e festas depois de uma semana cheia, pelos vôleis de

quarta, pelo fitdance de segunda e pelos treinos com vocês. Por cada aniversário, cada carona. Enfim, por cada dia de companhia ao longo desses anos; sem vocês, eu não teria conseguido.

Aos meus melhores amigos, Jader e Igor, por estarem presentes desde a graduação nos bons e maus momentos, por estarem sempre disponíveis para me ouvir, consolar e incentivar nos momentos de cansaço e desesperança, por me colocarem pra cima e me acompanharem em momentos épicos que nunca vou esquecer.

Aos meus amigos Bruno, Hélio, Nathalia, Hugo, Camila, bem como a todos os amigos que, em diferentes momentos, somaram-se a esse grupo trazendo novas energias e apoio. Por cada dia de descompressão, acolhimento e escuta: vocês me fizeram ver a luz em muitos momentos em que eu não fui capaz.

Aos analistas Dra. Isabela e Dr. Leonardo, e às pesquisadoras Dra. Cristina e Dra. Carolina, por me ensinarem não somente as técnicas e particularidades do laboratório, mas por mostrarem, pelo exemplo, como um cientista trabalha. Por me provarem que a ciência também está fora da universidade e por me ajudarem a ver e entender com clareza o que amo na ciência. Agradeço por cada vez que os interrompi e me socorreram, e por saber que podia e posso contar com vocês sempre.

Às Dras. Maria Eugênia, Luanna e Nayara, em especial, por me ajudarem pessoalmente no desenvolvimento deste projeto, muitas vezes atuando como parte da orientação.

Aos membros da banca, Dr. Bruno Paes, Dr. Marcelo Brígido e Dr. Fabrício por aceitarem contribuir com este trabalho e com a minha formação.

À Universidade de Brasília e ao Instituto de Ciências Biológicas, pela efervescência cultural, pelas oportunidades de vida e pelas vivências que nunca vou esquecer ou superar. Onde me formei biotecnologista e onde agora me formo mestre em Biologia Molecular.

À Embrapa Recursos Genéticos e Biotecnologia e ao Laboratório de Interação Molecular Planta-Praga, pela estrutura de ponta disponibilizada para a execução deste projeto.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

E por fim e mais importante, o amor da minha vida, Filipe Barbalho. Por esse alinhamento milenar que foi encontrar você nesta vida, neste plano e na hora exata. Você foi quem mais me ouviu reclamar, quem me viu perder a cabeça a cada experimento

que dava errado, e quem comemorou cada resultado positivo comigo. Sou grato por sua inteligência e por sua visão científica, que me encantam e me inspiram a me tornar um cientista melhor. Sou imensamente grato por todo o seu tempo, dedicação e trabalho para que eu pudesse concluir esta dissertação.

“Nos campos da observação, o acaso favorece apenas as mentes preparadas.”

Louis Pasteur

RESUMO

A soja (*Glycine max* (L.) Merrill) é uma das principais fontes de proteína vegetal; contudo, seu aproveitamento nutricional é limitado pela presença de fatores antinutricionais, notadamente os Inibidores de Protease do tipo Bowman-Birk (BBI). Estas proteínas, altamente estáveis termicamente devido à sua estrutura rica em pontes dissulfeto, inibem a tripsina e a quimotripsina no trato digestivo de monogástricos, reduzindo a digestibilidade proteica e podendo causar hipertrofia pancreática. O presente estudo teve como objetivo o desenvolvimento de linhagens de soja da cultivar brasileira BRS 537 com teor reduzido de BBI, utilizando a tecnologia de edição genômica CRISPR/Cas9 para o *knockout* do gene *Glyma.16G208900*. A seleção deste gene como alvo prioritário baseou-se em análises *in silico* e validação experimental por RT-qPCR, que indicaram sua expressão predominante durante o enchimento dos grãos na cultivar BRS 537. Análises estruturais via AlphaFold v2 e *docking* molecular atestaram a competência do modelo para a inibição bivalente (“cabeça-dupla”) sobre tripsina e quimotripsina, com laços reativos centrados em Arg16 e Arg43. A estratégia de transformação genética mediada por *Agrobacterium tumefaciens* em eixos embrionários maduros resultou na regeneração de 220 plantas aclimatadas, com uma eficiência de transformação de 56,8%. Dentre os dois sgRNAs avaliados, apenas o sgRNA1 mostrou-se eficiente no direcionamento do complexo CRISPR/Cas9 ao gene alvo. Essa especificidade permitiu a identificação de 10 eventos de edição na geração T_0 (Eventos 23, 32, 85, 88, 34, 21, 18, 84, 31 e 22), os quais exibiram uma diversidade de padrões mutacionais, incluindo deleções de -9 pb e -10 pb, além de inserções de $+1$ pb e $+3$ pb entre outros. Essas alterações geraram tanto mutações do tipo *frameshift*, que levam à formação de códons de parada prematuros, quanto deleções *in-frame*, ambas viabilizando a perda funcional da proteína alvo. A heterogeneidade das frequências alélicas observadas nos eventos T_0 é consistente com o mosaicismo típico dessa geração, indicando a necessidade de avanço para gerações subsequentes (T_1 e posteriores) para a segregação e fixação de alelos editados em homozigose e para a obtenção de linhagens potencialmente livres de transgene. Esses resultados ratificam a eficiência e a viabilidade da edição gênica na cultivar BRS 537, estabelecendo uma base para o desenvolvimento de soja com qualidade nutricional superior.

Palavras-chave: *Glycine max*. Fatores antinutricionais. CRISPR/Cas9. Transformação vegetal. Melhoramento genético. Digestibilidade proteica.

ABSTRACT

Soybean (*Glycine max* (L.) Merrill) is a primary source of plant protein; however, its nutritional utilization is limited by the presence of antinutritional factors, notably the Bowman-Birk protease inhibitors (BBI). These proteins, which are highly thermostable due to their disulfide-rich structure, inhibit trypsin and chymotrypsin in the digestive tract of monogastrics, reducing protein digestibility and potentially causing pancreatic hypertrophy. This study aimed to develop soybean lines from the Brazilian cultivar BRS 537 with reduced BBI content using CRISPR/Cas9 genome editing technology to *knockout* the *Glyma.16G208900* gene. The selection of this gene as a priority target was based on *in silico* analyses and experimental validation by RT-qPCR, which confirmed its predominant expression during grain filling in the BRS 537 cultivar. Structural analyses via AlphaFold v2 and molecular *docking* confirmed the model's competency for bivalent ("double-headed") inhibition of trypsin and chymotrypsin, with reactive loops centered at Arg16 and Arg43. The *Agrobacterium tumefaciens*-mediated genetic transformation strategy in mature embryonic axes resulted in the regeneration of 220 acclimated plants, with a transformation efficiency of 56.8%. Among the two sgRNAs evaluated, only sgRNA1 proved efficient in targeting the CRISPR/Cas9 complex to the target gene. This specificity allowed the identification of 10 editing events in the T_0 generation (Events 23, 32, 85, 88, 34, 21, 18, 84, 31, and 22), which exhibited a diversity of mutational patterns, including deletions of -9 bp and -10 bp, as well as insertions of $+1$ bp and $+3$ bp, among others. These alterations generated both *frameshift* mutations, leading to premature stop codons, and *in-frame* deletions, both enabling the functional loss of the target protein. The heterogeneity of allelic frequencies observed in T_0 events is consistent with the typical mosaicism of this generation, indicating the need to advance to subsequent generations (T_1 and beyond) for the segregation and fixation of edited alleles in homozygosity and to obtain potentially transgene-free lines. These results confirm the efficiency and feasibility of gene editing in the BRS 537 cultivar, establishing a foundation for the development of soybean with superior nutritional quality.

Keywords: *Glycine max*. Antinutritional factors. CRISPR/Cas9. Plant transformation. Genetic improvement. Protein digestibility.

LISTA DE FIGURAS

Figura 1 – Esquema comparativo do fluxo genômico e temporal: introgressão genética via retrocruzamento clássico <i>versus</i> edição gênica direta mediada por CRISPR/Cas9.	18
Figura 2 – Representação tridimensional comparativa dos inibidores Bowman-Birk (A) e Kunitz (B) da soja.	22
Figura 3 – Mecanismo molecular de formação do complexo ternário entre um inibidor da família Bowman-Birk e as serino-proteases tripsina e quimotripsina.	24
Figura 4 – Representação esquemática da edição genômica via sistema CRISPR/Cas9 e vias de reparo de DNA.	29
Figura 5 – Fluxograma da clonagem do sgRNA no vetor de expressão pYPQ141b.	42
Figura 6 – Fluxograma da engenharia do vetor zCas9-BBI por recombinação via Gateway.	43
Figura 7 – Mapa esquemático da região do T-DNA dos vetores zCas9-BBI e localização dos alvos no gene <i>Glyma.16G208900</i>	45
Figura 8 – Análise filogenética e perfil de expressão transcricional da família de genes do Inibidor Bowman-Birk (BBI) em <i>Glycine max</i> (L.) Merrill.	56
Figura 9 – Quantificação relativa dos transcritos de <i>Glyma.16G208900</i> por RT-qPCR durante o desenvolvimento da semente da cultivar BRS 537.	57
Figura 10 – Modelagem estrutural e predição de peptídeo sinal da proteína BBI codificada pelo gene <i>Glyma.16G208900</i>	60
Figura 11 – Predição estrutural do inibidor Bowman-Birk após a remoção do peptídeo sinal.	61
Figura 12 – Validação estrutural e funcional do inibidor Bowman-Birk modelado.	62
Figura 13 – Análise de <i>docking</i> molecular do inibidor BBI em complexos binários.	64
Figura 14 – Modelagem tridimensional do complexo ternário evidenciando a bivalência simultânea do inibidor BBI (<i>Glyma.16G208900</i>).	66
Figura 15 – Mapa de densidade de elementos <i>cis</i> -regulatórios e elementos de acessibilidade.	69
Figura 16 – Validação molecular e mapa esquemático do vetor binário final zCas9-BBI	74
Figura 17 – Validação molecular das plantas T ₀ e estratégia de genotipagem por PCR para o transgene <i>zCas9</i>	76

LISTA DE TABELAS

Tabela 1	– Características estruturais e funcionais dos inibidores de protease Kunitz (KTI) e Bowman–Birk (BBI) da soja.	21
Tabela 2	– Comparação das estratégias utilizadas para a redução de inibidores de protease em soja.	31
Tabela 3	– Alinhamento múltiplo das sequências de inibidores de protease do tipo Bowman–Birk (BBI) com destaque para variações em relação ao consenso estrutural.	38
Tabela 4	– Sequências dos oligonucleotídeos (sgRNAs) desenhados para a clonagem no vetor CRISPR/zCas9.	40
Tabela 5	– Sequência dos primers utilizados para a caracterização molecular das colônias transformadas via PCR de colônia.	44
Tabela 6	– Programa de termociclagem utilizado na PCR para a confirmação dos eventos de transformação.	48
Tabela 7	– Sequência dos primers utilizados para a amplificação e sequenciamento da região alvo no gene <i>Glyma.16G208900</i>	48
Tabela 8	– Categorias dos estádios de desenvolvimento da semente de soja, definidas com base na massa fresca e em características visuais.	50
Tabela 9	– Sequência dos primers utilizados para a amplificação por RT-qPCR do gene alvo BBI e dos genes de referência endógenos.	52
Tabela 10	– Composição da reação para os ensaios de RT-qPCR.	52
Tabela 11	– Condições reacionais e perfil térmico adotados para as análises de RT-qPCR.	53
Tabela 12	– Síntese dos parâmetros termodinâmicos e estruturais obtidos nas simulações de <i>docking</i> molecular do BBI (<i>Glyma.16G208900</i>).	67
Tabela 13	– <i>Heatmap</i> da densidade de elementos <i>cis</i> -regulatórios dos parálogos do BBI alvo.	71
Tabela 14	– Esquema representativo da clonagem direcional dos sgRNAs no vetor pYPQ141b via digestão com <i>Esp3I</i>	73
Tabela 15	– Resumo das eficiências de regeneração, transformação e edição gênica obtidas na cultivar BRS 537.	77
Tabela 16	– Perfil das edições e alinhamento genômico das linhagens de soja BRS 537 com edição no gene <i>Glyma.16G208900</i>	79
Tabela S1	– Lista de loci dos 15 genes selecionados de <i>Glycine max</i> analisados quanto ao perfil de expressão basal.	98
Tabela S2	– Perfil de expressão gênica de 15 transcritos de <i>Glycine max</i> . A intensidade da cor azul representa os níveis de transcrição em $\text{Log}_2(\text{FPKM} + 1)$	98

SUMÁRIO

1	INTRODUÇÃO	16
2	REFERENCIAL TEÓRICO	20
2.1	A SOJA: VALOR NUTRICIONAL E FATORES ANTINUTRI- CIONAIS	20
2.2	INIBIDORES DE PROTEASE EM SOJA: KUNITZ E BOWMAN- BIRK	21
2.3	O INIBIDOR BOWMAN-BIRK (BBI): ESTRUTURA, FAMÍ- LIA GÊNICA E IMPLICAÇÕES	23
2.4	BIOTECNOLOGIA APLICADA AO MELHORAMENTO DA QUALIDADE NUTRICIONAL	26
2.5	EDIÇÃO GENÔMICA COM CRISPR/CAS9 PARA O ME- LHORAMENTO DE PLANTAS	28
3	HIPÓTESE	32
4	OBJETIVOS	33
4.1	OBJETIVO GERAL	33
4.2	OBJETIVOS ESPECÍFICOS	33
5	MATERIAL E MÉTODOS	35
5.1	ANÁLISES <i>IN SILICO</i> E CARACTERIZAÇÃO ESTRUTU- RAL DO ALVO	35
5.1.1	Prospecção Genômica e Seleção do Gene Alvo	35
5.1.2	Análise da Arquitetura Regulatória e Prospecção de Elementos <i>Cis</i> - Atuantes	35
5.1.3	Predição Estrutural e Refinamento	37
5.1.4	Visualização e <i>Docking</i> Molecular	38

5.2	DESENHO DOS SGRNAS E CONSTRUÇÃO DOS VETORES DE EDIÇÃO GENÔMICA	40
5.2.1	Desenho dos RNAs guia e síntese dos oligonucleotídios	40
5.2.2	Preparo dos insertos de sgRNA por fosforilação e anelamento	40
5.2.3	Clonagem dos sgRNAs no vetor pYPQ141b e validação molecular	41
5.2.4	Engenharia do vetor zCas9-BBI por recombinação Gateway	42
5.2.5	Transformação de <i>A. tumefaciens</i>	44
5.3	TRANSFORMAÇÃO GENÉTICA DE SOJA E REGENERAÇÃO DE PLANTAS	46
5.3.1	Preparo do Inóculo de <i>A. tumefaciens</i>	46
5.3.2	Transformação e Regeneração <i>In Vitro</i> de Explantes de Soja	46
5.3.3	Aclimação das Plântulas Regeneradas	47
5.4	CARACTERIZAÇÃO MOLECULAR DOS EVENTOS EDITADOS	47
5.4.1	Extração de DNA Genômico	47
5.4.2	Triagem de Plantas Transformadas por PCR	47
5.4.3	Confirmação das Edições Gênicas por Sequenciamento	48
5.5	EXPRESSÃO TEMPORAL DO GENE BBI (<i>GLYMA.16G208900</i>) DURANTE A MATURAÇÃO DAS SEMENTES	49
5.5.1	Amostragem Vegetal e Estratégia de <i>Pool</i> Biológico	49
5.5.2	Extração de RNA Total	50
5.5.3	Síntese de cDNA	51
5.5.4	Reação de RT-qPCR e Análise de Expressão	51
5.5.5	Análise dos dados de RT-qPCR	53
6	RESULTADOS E DISCUSSÃO	55

6.1	SELEÇÃO DO ALVO E PERFIL DE EXPRESSÃO <i>IN SILICO</i>	55
6.1.1	Validação experimental do perfil de expressão na cultivar de soja BRS 537	56
6.2	VALIDAÇÃO ESTRUTURAL DO ALVO <i>GLYMA.16G208900</i> E PREDIÇÃO DE FUNCIONALIDADE	59
6.2.1	Caracterização Estrutural da Proteína do Gene <i>Glyma.16G208900</i> Madura	59
6.2.2	Obtenção do Modelo Estrutural da Proteína Madura e Mapeamento dos Sítios Ativos	60
6.2.3	Inferência Funcional Baseada em Modelagem Molecular: Análise Comparativa da Interação Enzima-Inibidor	63
6.2.4	Análise de Redundância e Arquitetura Regulatória da Família BBI	67
6.3	CONSTRUÇÃO E VALIDAÇÃO DOS VETORES PARA EDIÇÃO DO GENE (<i>GLYMA.16G208900</i>)	72
6.4	EFICIÊNCIA DE TRANSFORMAÇÃO E CARACTERIZAÇÃO DAS EDIÇÕES	75
7	CONCLUSÕES E PERSPECTIVAS	82
7.1	CONCLUSÕES	82
7.2	PERSPECTIVAS	83
	REFERÊNCIAS	84
	Material Suplementar	98

1 INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill), *Fabaceae* de origem asiática, representa uma das principais culturas agrícolas de importância econômica global, sendo fundamental para a segurança alimentar (HARTMAN et al., 2011). O cultivo e o comércio de seus derivados sustentam um complexo agroindustrial robusto. Nas últimas décadas, esse crescimento consolidou a cultura como fonte estratégica de proteína vegetal e óleo (HIRAKURI; LAZZAROTTO, 2014). Nesse cenário, o Brasil destaca-se como o maior produtor e exportador mundial, com produção estimada em 176,1 milhões de toneladas na safra 2025/2026, evidenciando a relevância da cultura no agronegócio nacional e na economia brasileira (Companhia Nacional de Abastecimento, 2026).

O sucesso agrônômico da soja está diretamente associado ao seu perfil nutricional, caracterizado por elevado teor proteico (aproximadamente 40%) e lipídico (20%), conferindo ampla versatilidade industrial (CARRAO-PANIZZI; MANDARINO, 1998). Além do uso consolidado na formulação de rações animais, a soja tem ampliado sua participação na alimentação humana, impulsionada pela demanda crescente por alimentos funcionais. O grão é rico em fitoquímicos, como isoflavonas e saponinas, cujo consumo está associado à prevenção de doenças crônicas, incluindo doenças cardiovasculares, osteoporose e determinados tipos de câncer, além de representar uma alternativa essencial para dietas baseadas em proteínas vegetais (MESSINA, 2016; RIZZO; BARONI, 2018).

Apesar desse potencial, o aproveitamento biológico da proteína de soja apresenta limitações devido à presença de fatores antinutricionais, com destaque para os inibidores de protease, considerados um dos entraves tecnológicos para a indústria de alimentos (GENOVESE; LAJOLO, 2006). No trato digestivo de animais monogástricos, incluindo os humanos, esses compostos formam complexos estáveis com enzimas pancreáticas essenciais, como a tripsina e a quimotripsina, e comprometem a hidrólise proteica. O consumo crônico desses inibidores pode induzir respostas fisiológicas compensatórias, como hipersecreção pancreática, hipertrofia e hiperplasia do órgão, com consequentes prejuízos ao crescimento e à conversão alimentar (LIENER, 1994).

Na soja, os principais inibidores de protease distribuem-se entre as famílias Kunitz (KTI) e Bowman-Birk (BBI) (KUNITZ, 1947; BIRK, 1985). Ambos apresentam atividade antinutricional; entretanto, exibem propriedades físico-químicas contrastantes. O KTI possui caráter termolábil e sofre inativação eficiente por agregação proteica durante tratamentos térmicos convencionais. Por sua vez, o BBI apresenta elevada estabilidade estrutural em decorrência da extensa rede de pontes dissulfeto, característica associada à

resistência ao processamento industrial e à digestão gastrointestinal (DIPIETRO; LIENER, 1989; LIENER, 1994).

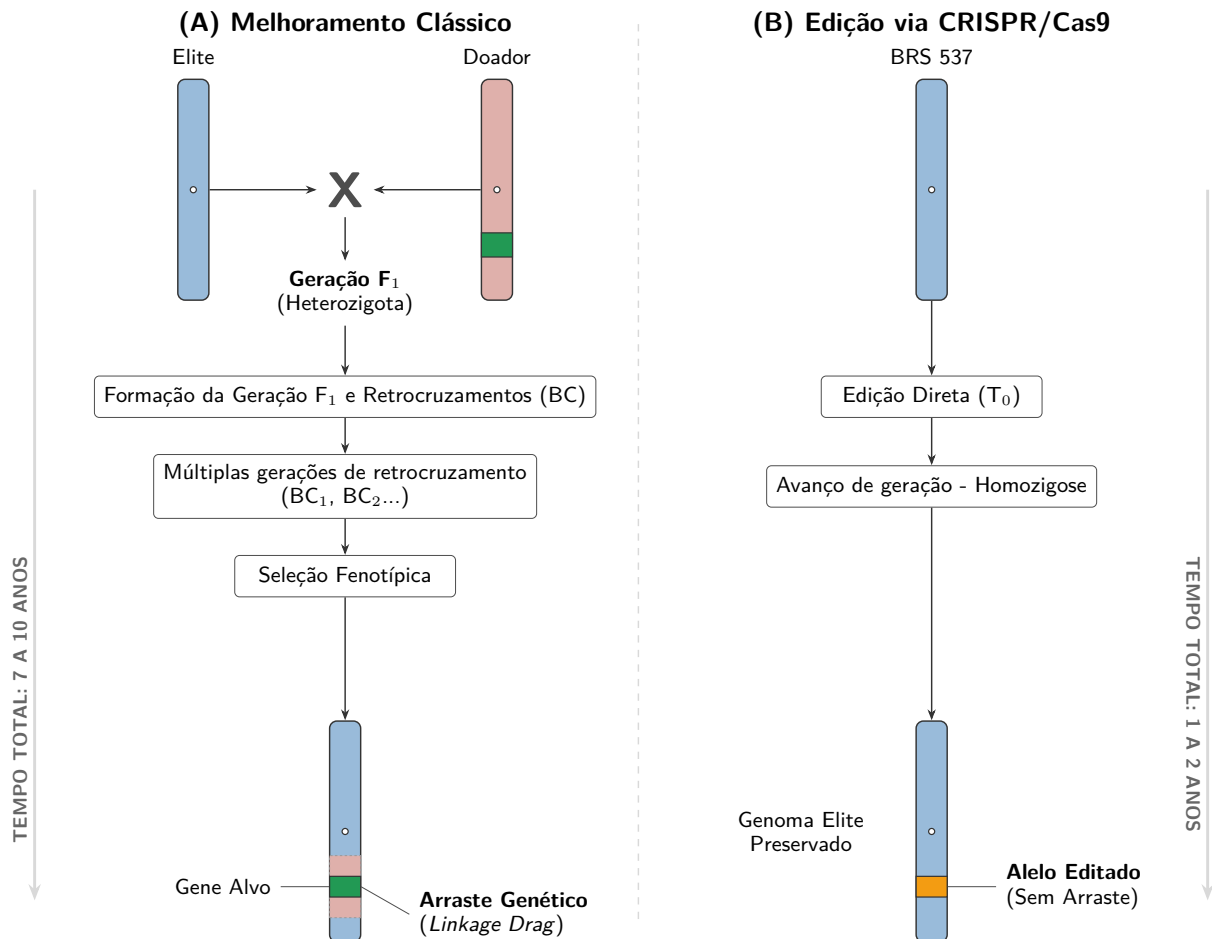
A inativação efetiva do BBI depende de reações químicas lentas, como a β -eliminação das pontes dissulfeto. Esse processo requer tratamentos químicos e térmicos superiores a 100 °C, em pH básico por períodos acima de 20 minutos para a obtenção de uma redução expressiva da atividade inibitória (CHEN et al., 2014; HE et al., 2017). Em razão dessa estabilidade térmica, produtos comerciais derivados de soja podem reter entre 5% e 20% da atividade original. Essa fração residual é atribuída majoritariamente à persistência do BBI (LIENER, 1994). Sua persistência impõe desafios nutricionais e eleva os custos da cadeia produtiva, podendo restringir a aplicação da soja em produtos de alto valor agregado, como fórmulas infantis e rações de alto desempenho, nas quais a digestibilidade máxima constitui um requisito crítico. Portanto, a redução direcionada do teor de BBI das sementes por meio de edição gênica surge como uma estratégia promissora para superar essas limitações tecnológicas e agregar valor à cultura.

A ausência de cultivares comerciais livres do inibidor Bowman-Birk decorre da complexidade inerente ao melhoramento genético para essa característica em uma espécie paleopoliploide como a soja. A presença de famílias multigênicas resultantes de eventos de duplicação do genoma favorece mecanismos de redundância genética. Essa dinâmica explica a dificuldade de eliminação da atividade inibitória, uma vez que o *knockout* de um gene isolado pode sofrer compensação pela expressão de genes parálogos ou pela ativação de vias regulatórias alternativas (EL-BROLOS; STAINIER, 2017; GILLMAN et al., 2015).

O emprego de RNA de interferência (RNAi) para o silenciamento específico em sementes demonstrou eficácia na redução drástica dessas atividades antinutricionais, com consequente melhoria na digestibilidade proteica (KIM et al., 2025). Entretanto, essa abordagem fundamenta-se na transgenia clássica, condição que implica a inserção estável e permanente de DNA exógeno e sujeita o produto final a processos regulatórios onerosos e à menor aceitação comercial, entraves típicos dos Organismos Geneticamente Modificados (OGMs) (KIM et al., 2025). Adicionalmente, a dependência inicial de genótipos não comerciais para a transformação exige extensos e custosos programas de retrocruzamento para a introgressão do alelo de interesse em cultivares elite, como a BRS 537. Conforme esquematizado na Figura 1, o melhoramento clássico requer o cruzamento inicial entre os parentais para a formação de uma geração F₁, seguido por múltiplas etapas de retrocruzamento (BC) e seleção fenotípica. Essa abordagem tradicional é severamente limitada pelo tempo prolongado, estendendo-se frequentemente de 7 a 10 anos, e pela inevitável ocorrência do arraste genético indesejado (*linkage drag*), no qual fragmentos do genoma doador são co-herdados com o gene alvo. Em contrapartida, a edição gênica direta

mediada por CRISPR/Cas9 no *background* genético da cultivar de interesse contorna essas limitações metodológicas, eliminando o arraste genético e reduzindo o tempo de obtenção de linhagens homozigotas estáveis para 1 a 2 anos (CHEN et al., 2019; TANG et al., 2023).

Figura 1 – Esquema comparativo do fluxo genômico e temporal: introgressão genética via retrocruzamento clássico *versus* edição gênica direta mediada por CRISPR/Cas9.



Legenda: Os cromossomos em **azul** representam o genoma da cultivar recorrente elite (BRS 537), enquanto os blocos em **vermelho** indicam o DNA do doador. Em **(A)**, o cruzamento inicial entre Elite e Doador forma a Geração F₁ (par heterozigoto). Em seguida, ocorre a Formação da Geração F₁ e Retrocruzamentos (BC), avançando por múltiplas gerações de retrocruzamento (BC₁, BC₂...) para restaurar o genoma elite. A introgressão do gene alvo (**verde**) neste longo processo é acompanhada por fragmentos genômicos indesejados (*linkage drag*), levando de 7 a 10 anos. Em **(B)**, a edição do alelo (**amarelo**) mediada por CRISPR/Cas9 ocorre diretamente no *background* elite, eliminando o arraste genético e reduzindo o tempo de obtenção para 1 a 2 anos.

Fonte: Elaborada pelo autor (2026).

Para contornar as limitações do melhoramento clássico e os entraves regulatórios da transgenia, o sistema CRISPR/Cas9 consolida-se como uma ferramenta de excelência para o melhoramento genético de precisão. A tecnologia possibilita induzir mutações sítio-dirigidas e obter linhagens livres de DNA exógeno (*transgene-free*), característica responsável por sua distinção conceitual em relação aos organismos transgênicos clássicos (CHEN et al., 2019;

CONG et al., 2013; ARAKI; ISHII, 2015). No Brasil, essa distinção encontra respaldo na Resolução Normativa nº 16/2018 da CTNBio, que estabelece que organismos obtidos por edição gênica podem, a depender da natureza da alteração introduzida, ser enquadrados como produtos de melhoramento convencional (BRASIL. Comissão Técnica Nacional de Biossegurança (CTNBio), 2018).

Dessa forma, a realização deste estudo na cultivar BRS 537, um genótipo elite brasileiro, confere maior relevância translacional aos resultados obtidos. A aplicação da edição gênica em germoplasma comercial representa uma estratégia capaz de encurtar o intervalo entre a descoberta molecular e sua incorporação em programas de melhoramento. Ao contornar as limitações temporais e de arraste genético inerentes ao melhoramento convencional, essa abordagem fortalece o potencial da engenharia genômica para o desenvolvimento de cultivares de soja com características de alto valor agregado e maior proximidade de aplicação no setor produtivo.

Nesse contexto, a edição genômica mediada pelo sistema CRISPR/Cas9 emerge como uma estratégia promissora para a inativação de fatores antinutricionais em soja. O presente estudo teve como objetivo desenvolver um genótipo com níveis reduzidos do inibidor BBI por meio do *knockout* do gene *Glyma.16G208900*, selecionado por corresponder à isoforma predominantemente expressa durante a maturação das sementes, conforme evidenciado por dados de transcriptoma (SEVERIN et al., 2010; SREEDASYAM et al., 2023).

A hipótese central deste estudo parte da premissa de que a inativação direcionada desse gene majoritário teria o potencial de promover uma redução expressiva da atividade antinutricional nas sementes. Entretanto, considerando a organização multigênica da família BBI, torna-se necessário avaliar o potencial de compensação transcricional entre genes parálogos. Para prever se o *knockout* do *locus* principal possui o risco de induzir a ativação compensatória de outros genes, cenário que poderia resgatar o fenótipo inibitório indesejado, o estudo investiga a arquitetura regulatória dessa família por meio da prospecção de motivos conservados (*cis-elements*) em suas regiões promotoras.

Em suma, a integração entre a edição genômica direcionada e a análise da regulação transcricional da família gênica do BBI contribui para elucidar os mecanismos de expressão desses inibidores. Dessa forma, este estudo estabelece bases sólidas para estratégias de melhoramento voltadas à redução de fatores antinutricionais em soja, explorando o principal gene associado a esse fenótipo.

2 REFERENCIAL TEÓRICO

2.1 A SOJA: VALOR NUTRICIONAL E FATORES ANTINUTRICIONAIS

A soja (*Glycine max* (L.) Merrill) é amplamente reconhecida por seu elevado valor nutricional, sendo uma das fontes de proteína vegetal mais completas e economicamente acessíveis disponíveis para a alimentação humana e animal. A qualidade de sua proteína, que corresponde a aproximadamente 80% do valor biológico da proteína do leite de vaca, se aproxima dos padrões de referência estabelecidos pela Organização das Nações Unidas para a Alimentação e a Agricultura (FAO) e pela Organização Mundial da Saúde (OMS) (BURSSENS et al., 2011). Atualmente, a digestibilidade ileal verdadeira substituiu o clássico valor biológico (SCHAAFSMA, 2005) como métrica preferencial, por apresentar maior correlação com o anabolismo proteico real (MOUGHAN; LIM, 2024).

Além do seu perfil proteico, o grão de soja é rico em compostos fitoquímicos com atividade biológica, como isoflavonas, saponinas e ácidos graxos essenciais, os quais têm despertado a atenção da comunidade científica devido aos seus potenciais benefícios à saúde. Diversos estudos científicos corroboram a associação entre o consumo de soja e a prevenção de doenças crônicas, incluindo doenças cardiovasculares, osteoporose e certos tipos de câncer, consolidando sua classificação como um alimento funcional (RIZZO; BARONI, 2018; MESSINA, 2016).

Contudo, o aproveitamento integral desses nutrientes é dificultado pela presença de compostos endógenos conhecidos como fatores antinutricionais. Essas moléculas podem interferir na digestibilidade de macronutrientes, na absorção de minerais ou exercer efeitos fisiológicos adversos. Na soja, os fatores antinutricionais de maior relevância são os inibidores de proteases, que constituem cerca de 6% do total de proteínas do grão, sendo apontados como um dos principais limitantes para a sua ampla utilização na dieta humana (GENOVESE; LAJOLO, 2006; LIENER, 1994).

Os inibidores de protease da soja são classificados em duas famílias principais, com base em suas características estruturais e especificidade: o inibidor de tripsina Kunitz (KTI) e o inibidor de tripsina e quimotripsina Bowman-Birk (BBI) (KUNITZ, 1947; BIRK, 1985). Embora ambos sejam genericamente conhecidos como inibidores de tripsina, eles possuem especificidades distintas e podem inibir outras serinoproteases, como a quimotripsina. A ingestão desses inibidores em sua forma ativa tem sido associada, em estudos com animais monogástricos, à inibição do crescimento e a alterações metabólicas no pâncreas, como hipertrofia e hiperplasia, decorrentes de um aumento compensatório na secreção de enzimas

digestivas (LIENER, 1994).

Portanto, a neutralização desses compostos é uma etapa crucial para garantir a segurança e a qualidade nutricional dos produtos à base de soja (GENOVESE; LAJOLO, 2006). Dentre os fatores antinutricionais proteicos da soja, evidências acumuladas indicam que aqueles estruturalmente mais compactos, como os da família BBI, representam o maior desafio tecnológico e nutricional atual (OKEDIGBA et al., 2023), uma vez que sua extensa rede de pontes dissulfeto permite que persistam ativos mesmo após os protocolos de processamento térmico convencionais (DIPIETRO; LIENER, 1989; HE et al., 2017).

2.2 INIBIDORES DE PROTEASE EM SOJA: KUNITZ E BOWMAN-BIRK

Inibidores de protease presentes na soja, embora agrupados pela sua função biológica, pertencem a famílias proteicas estrutural e funcionalmente distintas. A compreensão de suas características individuais é fundamental para o desenvolvimento de estratégias eficazes de melhoramento nutricional. As duas principais famílias são a do inibidor de tripsina KTI e a do inibidor BBI, que diferem em massa molecular, especificidade de alvo e, em sua estabilidade térmica, conforme resumido na Tabela 1 (KOIDE; IKENAKA, 1973).

Tabela 1 – Características estruturais e funcionais dos inibidores de protease Kunitz (KTI) e Bowman-Birk (BBI) da soja.

Característica	Kunitz (KTI)	Bowman-Birk (BBI)
Massa molecular	~20-25 kDa	~6-10 kDa
Pontes dissulfeto	2	7
Sítios de inibição	1 (Tripsina)	2 (Tripsina e Quimotripsina)
Termoestabilidade	Termolábil	Termoestável
Impacto no processamento	Facilmente inativado	Persiste após processamento

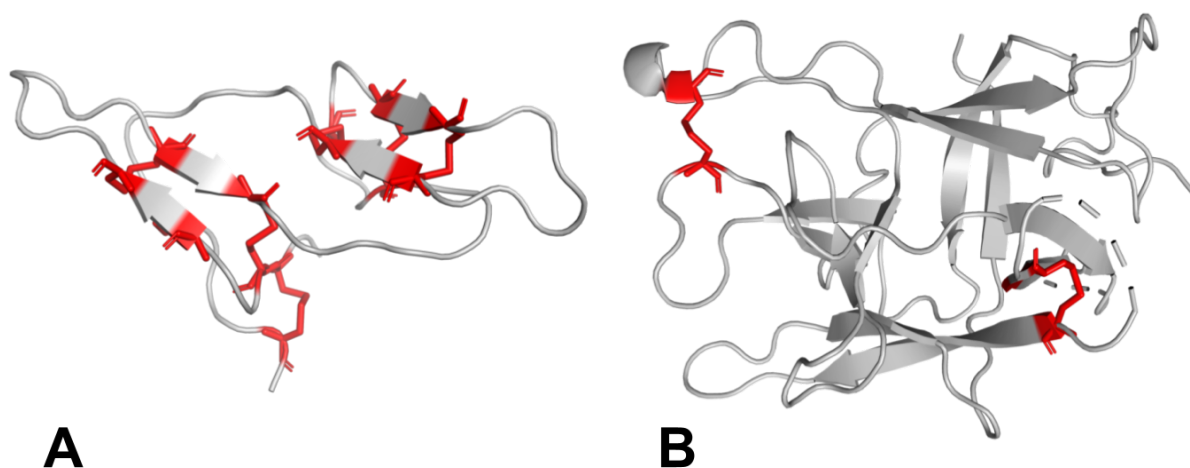
Fonte: Adaptado de Koide e Ikenaka (1973), Liener (1994) e Birk (1985).

O inibidor de tripsina KTI, isolado e cristalizado por Kunitz em 1946, é uma proteína com massa molecular de aproximadamente 21,5 kDa. Sua estrutura polipeptídica, estabilizada por apenas duas pontes dissulfeto, apresenta um único sítio reativo com alta especificidade para a inibição da tripsina (KUNITZ, 1947; LIENER, 1994). A característica mais relevante do KTI para a indústria de alimentos é sua termolabilidade. Estudos clássicos demonstram que o KTI apresenta menor estabilidade térmica frente ao inibidor BBI; assim, a exposição a tratamentos térmicos adequados (como o cozimento) é geralmente suficiente para desnaturar sua estrutura e mitigar sua função antinutricional (DIPIETRO; LIENER, 1989). Por essa razão, embora presente em quantidades significativas no grão cru, o KTI é considerado um problema nutricional de menor impacto em produtos de soja que passam por processamento térmico.

Em contrapartida, o inibidor BBI representa um desafio tecnológico mais significativo devido à sua elevada resiliência estrutural. O BBI é uma proteína de baixo peso molecular, com massa aproximada de 8,8 kDa, conforme determinado por análises de espectrometria de massa (MALDI-TOF-MS) (GU et al., 2014). Sua conformação é altamente compacta e estabilizada por uma rede de sete pontes dissulfeto, característica que confere elevada estabilidade frente a tratamentos térmicos e a variações extremas de pH (BIRK, 1985; OKEDIGBA et al., 2023). O inibidor também se caracteriza por um arranjo estrutural do tipo “cabeça dupla” (*double-headed*), apresentando dois sítios de ligação independentes que permitem a inibição simultânea da tripsina e da quimotripsina (ODANI; IKENAKA, 1973). Além disso, manifesta-se como uma família de isoformas com variações polimórficas em sua sequência de aminoácidos, aspecto que amplia a dificuldade de sua completa inativação durante o processamento industrial de derivados de soja (LAJOLO; GENOVESE, 2002).

Essa disparidade na estabilidade térmica, resumida na Tabela 1, é uma consequência direta de suas arquiteturas tridimensionais distintas. Conforme ilustrado na Figura 2, o BBI apresenta uma conformação globular de intensa coesão estrutural, sendo reticulada por uma rede de sete pontes dissulfeto. Em contrapartida, o KTI possui uma estrutura de maior volume e flexibilidade, estabilizada por apenas duas pontes dissulfeto, o que o torna consideravelmente mais suscetível à desnaturação térmica.

Figura 2 – Representação tridimensional comparativa dos inibidores Bowman-Birk (A) e Kunitz (B) da soja.



Legenda: Modelos moleculares ilustrando a diferença na arquitetura tridimensional das duas famílias de inibidores. Em (A), o inibidor Bowman-Birk (PDB ID: 1BBI) exibe uma conformação globular altamente compacta, estabilizada por uma densa rede de sete pontes dissulfeto (destacadas em **vermelho**). Em (B), o inibidor de tripsina Kunitz (PDB ID: 9CT1) apresenta uma estrutura tridimensional mais volumosa (20-25 kDa) e flexível, contendo apenas duas pontes dissulfeto. O esqueleto polipeptídico (*backbone*) está representado em cinza nas estruturas. Essa expressiva disparidade no padrão de reticulação covalente é o principal fator responsável pela elevada termoestabilidade estrutural do BBI em contraste à termolabilidade do KTI.

Fonte: Elaborada pelo autor (2026).

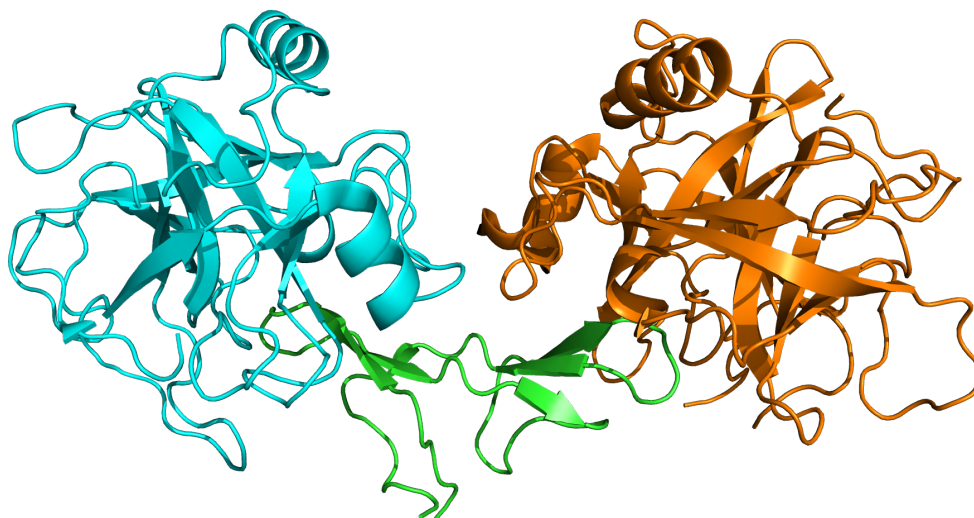
Sob a ótica do processamento industrial, essa divergência torna-se o ponto central que direciona as estratégias de melhoramento nutricional. Enquanto a atividade do KTI é majoritariamente eliminada pelos métodos convencionais de calor, a resiliência intrínseca do BBI permite que ele permaneça funcional nos produtos finais, mantendo seu efeito antinutricional (DIPIETRO; LIENER, 1989). Visando mitigar esse entrave, abordagens biotecnológicas como o silenciamento via RNA de interferência (RNAi) foram recentemente empregadas com sucesso, comprovando que a supressão do BBI eleva substancialmente a digestibilidade proteica (KIM et al., 2025). Entretanto, nas abordagens que envolvem a expressão estável de RNA interferente mediada por transgenia, o desenvolvimento dessas linhagens esbarra em entraves regulatórios e de mercado. Diante desse desafio, o *knockout* gênico via CRISPR/Cas9 surge como a abordagem mais eficaz para mutações de perda de função desses antinutrientes de forma precisa, otimizando a digestibilidade da proteína de soja sem comprometer sua integridade nutricional e com maior viabilidade comercial.

2.3 O INIBIDOR BOWMAN-BIRK (BBI): ESTRUTURA, FAMÍLIA GÊNICA E IMPLICAÇÕES

O inibidor BBI constitui o foco central deste estudo devido à sua arquitetura molecular singular e à sua relevância como fator antinutricional em sementes de soja, cultura de elevada importância agrônômica e econômica no cenário nacional e global. Esta rigidez estrutural não apenas confere a estabilidade térmica típica da família, mas é determinante para sua resistência à proteólise no trato gastrointestinal, permitindo que a molécula alcance seu alvo funcionalmente intacta (BIRK, 1985; GITLIN-DOMAGALSKA et al., 2020).

A funcionalidade inibitória do BBI reside na presença de dois laços reativos independentes expostos em sua superfície. Esta disposição topológica permite que uma única molécula do inibidor forme complexos ternários estáveis ao ligar-se simultaneamente a duas enzimas distintas, a tripsina e a quimotripsina, bloqueando seus sítios ativos e impedindo a hidrólise das proteínas dietéticas (BIRK, 1985; GITLIN-DOMAGALSKA et al., 2020) (Figura 3). Curiosamente, essa notável estabilidade antienzimática também despertou interesse farmacológico recente, com estudos evidenciando o potencial adjuvante do BBI no tratamento de mesotelioma humano (KASHIWAGI et al., 2011), o que ilustra a dualidade funcional da molécula.

Figura 3 – Mecanismo molecular de formação do complexo ternário entre um inibidor da família Bowman-Birk e as serino-proteases tripsina e quimotripsina.



Legenda: Representação tridimensional da estrutura cristalográfica ilustrando a inibição estequiométrica simultânea característica da família Bowman-Birk. O BBI (**verde**) interage e bloqueia de forma independente os sítios ativos da tripsina (**ciano**) e da quimotripsina (**laranja**). Estrutura correspondente ao complexo cristalográfico resolvido em alta resolução sob o PDB ID: [3RU4](#).

Fonte: Elaborada pelo autor (2026), adaptado de [Honda et al. \(2018\)](#).

Do ponto de vista genômico, a complexidade da família BBI em soja está intrinsecamente ligada à história evolutiva de *G. max* (L.) Merrill como espécie paleopoliploide. Esse histórico inclui pelo menos dois eventos de duplicação completa do genoma (Whole Genome Duplication – WGD), ocorridos há aproximadamente 59 e 13 milhões de anos, resultando na retenção de múltiplas cópias para cerca de 75% de seus genes ([SCHMUTZ et al., 2010](#)). Por essa razão, análises bioinformáticas identificaram pelo menos 15 *loci* com potencial para codificar proteínas do tipo BBI, frequentemente organizados em *clusters* genômicos ([BARROS et al., 2012](#); [SCHMUTZ et al., 2010](#)). Os localizadores dos 15 genes estão descritos no suplementar. Essa expansão reflete-se na existência de múltiplas isoformas proteicas que, apesar da alta homologia estrutural, podem apresentar variações cinéticas e de especificidade.

Essa complexidade genômica impõe barreiras biotecnológicas significativas para o melhoramento da soja. Fenômenos de compensação genética, onde a perda funcional de um gene induz a regulação positiva de genes relacionados, são um desafio crítico em famílias genéticas expandidas, isto é, grupos de genes parálogos que retêm funções redundantes sobrepostas ([EL-BROLOSY](#); [STAINIER, 2017](#)). Estudos demonstraram que o *knockout* de isoformas majoritárias pode falhar em eliminar a atividade antinutricional devido à ação compensatória funcional desses outros membros da mesma família ([GILLMAN et al.,](#)

2015).

A conservação evolutiva dessa família em leguminosas reforça seu papel adaptativo na defesa vegetal (GITLIN-DOMAGALSKA *et al.*, 2020; CLEMENTE *et al.*, 2019). Especificamente, os inibidores do tipo BBI atuam na neutralização de proteases digestivas de insetos-praga e patógenos. Contudo, no contexto da nutrição, essa mesma capacidade inibitória reduz drasticamente a digestibilidade proteica, consolidando a inativação do BBI como uma meta prioritária. Apesar dessa prioridade, a eliminação genérica de inibidores de protease levanta preocupações agrônomicas, pois poderia, hipoteticamente, deixar a planta mais suscetível à herbivoria.

No contexto da interação planta-ambiente, os BBIs desempenham uma função primordial no sistema imune inato das leguminosas, atuando como potentes agentes de defesa direta contra a herbivoria e a invasão microbiana. Essa resposta defensiva é frequentemente induzida por estresse biótico, como o dano tecidual causado pela mastigação de insetos, que sinaliza a liberação desses inibidores no trato digestivo do herbívoro. Uma vez no intestino do inseto, os BBIs bloqueiam as proteases digestivas essenciais, como a tripsina e a quimotripsina exógenas, impedindo a absorção de nutrientes vitais, o que acarreta inibição do crescimento e aumento da mortalidade larval (XIE *et al.*, 2021; KUMAR, 2024). Além de sua atividade inseticida, há evidências de que os BBIs sejam induzidos na resposta imune desencadeada por efetores durante o ataque de patógenos fúngicos e bacterianos, onde operam inativando as enzimas proteolíticas secretadas pelos microrganismos para degradar as paredes celulares da planta (XIE *et al.*, 2021; GITLIN-DOMAGALSKA *et al.*, 2020). Portanto, a manutenção dessa arquitetura multigênica em cultivares modernas reflete uma necessidade adaptativa para a sobrevivência em campo, o que reforça o risco agrônomico de estratégias de edição que visem a supressão generalizada de proteases sem considerar a especificidade dos tecidos-alvo (LIU *et al.*, 2026).

Contudo, esse conflito é contornado pela própria biologia e evolução da cultura. A soja possui uma extensa família multigênica de BBIs com forte regulação tecido-específica. O processo evolutivo de duplicação genômica permitiu que a família sofresse uma subfuncionalização regulatória: algumas isoformas especializaram-se na defesa de órgãos vegetativos e reprodutivos externos, enquanto outras se especializaram no acúmulo de reservas exclusivamente na semente madura (FORCE *et al.*, 1999; VORSTER *et al.*, 2023). Essa compartimentalização possibilita que a edição genômica elimine o fator antinutricional da dieta ao focar apenas na isoforma específica da semente, preservando intacta a maquinaria de defesa nas barreiras físicas externas da planta (VORSTER *et al.*, 2023).

Apesar da multiplicidade de cópias, dados transcriptômicos indicam que a con-

tribuição funcional dos membros da família não é equivalente. Durante os estádios de enchimento do grão, a expressão é fortemente dominada pelo gene *Glyma.16G208900*, identificado como a isoforma predominante em sementes (SEVERIN et al., 2010; KIM et al., 2024). Em contrapartida, outros tecidos como raízes (*Glyma.09G158700*) e vagens (*Glyma.18G231500*) apresentam níveis basais de expressão de BBI, dependendo de outros parálogos para a manutenção da defesa contra herbivoria.

Portanto, o *knockout* direcionado ao *Glyma.16G208900* desponta como uma estratégia altamente vantajosa, pois aponta para a possibilidade de redução do fator anti-nutricional no grão com menor probabilidade de comprometimento da proteção agrônômica da planta no campo. Para fundamentar o desenho dessa edição genômica de forma segura, faz-se necessária a validação do alvo em duas frentes complementares: a estrutural e a regulatória. Primeiramente, dada a ausência de uma estrutura cristalográfica experimental para o produto deste gene, exige-se a aplicação de modelagem molecular de alta precisão para atestar sua funcionalidade como inibidor. Em paralelo, a caracterização detalhada da arquitetura do seu promotor, contrastada com a de seus parálogos, é essencial para prever e contornar potenciais mecanismos de compensação funcional genômica na espécie.

2.4 BIOTECNOLOGIA APLICADA AO MELHORAMENTO DA QUALIDADE NUTRICIONAL

O advento da engenharia genética representou um marco na agricultura moderna, culminando no desenvolvimento da primeira geração de culturas geneticamente modificadas. Essas culturas foram projetadas para expressar características agrônômicas de interesse direto para o produtor, como a tolerância a herbicidas e a resistência a insetos-praga, resultando em ganhos de produtividade e redução no uso de defensivos químicos (KLÜMPER; QAIM, 2014). Contudo, nas últimas décadas, a biotecnologia agrícola avançou para uma segunda geração, com foco crescente no melhoramento de características de qualidade nutricional e funcional, visando agregar valor ao produto final e oferecer benefícios diretos à saúde do consumidor (KAMTHAN et al., 2016).

A soja tem se destacado como plataforma central no desenvolvimento de produtos desta segunda geração de biotecnologia agrícola. Historicamente, a mitigação de compostos indesejáveis foi conduzida por meio do melhoramento genético tradicional, cujo marco no Brasil foi o desenvolvimento de cultivares isentas das enzimas lipoxigenases, como a cultivar BRS 257 (CARRÃO-PANIZZI; PÍPOLO, 2009). Essas enzimas são responsáveis pela oxidação de ácidos graxos, gerando compostos voláteis que conferem à soja um sabor adstringente e de difícil aceitação por muitos consumidores. O desenvolvimento desta cultivar representou um grande avanço, fornecendo matéria-prima de qualidade superior para a indústria de alimentos. Mais recentemente, impulsionada pelas ferramentas

biotecnológicas de precisão, essa mesma característica tornou-se um exemplo de sucesso na edição genômica. Wang et al. (2020) demonstraram a obtenção de sementes de soja isentas de lipoxigenases através do *knockout* simultâneo dos genes da família das Lipoxigenases (LOX), especificamente *GmLOX1*, *GmLOX2* e *GmLOX3*, via CRISPR/Cas9 (WANG et al., 2020), evidenciando o potencial da técnica em reproduzir e acelerar o desenvolvimento de características de alto valor agregado.

Em outra vertente do melhoramento nutricional, os esforços concentram-se na modulação do perfil de ácidos graxos do óleo de soja. Historicamente, a obtenção de variedades com elevado teor de ácido oleico, um ácido graxo monoinsaturado associado a benefícios à saúde e elevada estabilidade oxidativa, que favorece aplicações industriais específicas e o uso em *blends* para frituras, dependia, em grande parte, de abordagens transgênicas (WARNER; GUPTA, 2005). Estudos pioneiros, como o de Buhr et al. (2002), demonstraram a viabilidade do silenciamento de genes da via de biossíntese de ácidos graxos, promovendo a acumulação de ácido oleico nas sementes. Tais linhagens apresentavam não apenas elevado teor de ácido oleico, mas também uma redução significativa nos níveis de ácido palmítico, cuja diminuição contribui para o aprimoramento do perfil nutricional e da estabilidade oxidativa do óleo. De forma análoga, a aplicação de ferramentas de edição gênica, como CRISPR/Cas9, para a inativação de inibidores de proteases visa elevar o valor biológico do grão, com foco no aumento da digestibilidade proteica.

No cenário nacional, pesquisadores da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) também obtiveram sucesso ao desenvolver plantas transgênicas de soja via RNA de interferência (RNAi) que apresentavam não apenas elevado teor de ácido oleico, mas também uma redução significativa de ácido palmítico (MURAD et al., 2014). Contudo, assim como observado no silenciamento de fatores antinutricionais, essas abordagens esbarram nas complexidades regulatórias e na menor aceitação pública dos OGMs.

Diante da necessidade contínua de aprimorar a qualidade do grão, a supressão de fatores antinutricionais mediante o emprego de biotecnologias de alta precisão, a exemplo do CRISPR/Cas9, configura-se como uma abordagem promissora no melhoramento da cultura da soja. A aplicação de ferramentas de engenharia genética para o *knockout* do inibidor BBI está em consonância com essa tendência, abordando um importante gargalo nutricional da soja e reforçando seu potencial como fonte proteica de excelência. Esses exemplos evidenciam que a biotecnologia aplicada à soja transpassou o enfoque estritamente agrônomo, consolidando-se como uma estratégia robusta para o aprimoramento direto da qualidade nutricional do grão.

2.5 EDIÇÃO GENÔMICA COM CRISPR/CAS9 PARA O MELHORAMENTO DE PLANTAS

Para viabilizar modificações genéticas de alta precisão, como o *knockout* de *loci* específicos, o sistema CRISPR/Cas9 (Repetições Palindrômicas Curtas Agrupadas e Regularmente Interespaçadas) consolidou-se como a ferramenta de escolha no melhoramento de plantas. Esta plataforma destaca-se por sua elevada eficiência, versatilidade e relação custo-benefício, superando tecnologias de edição anteriores, como as Nucleases de Dedos de Zinco (ZFNs) e as Nucleases Efetoras do Tipo Ativador de Transcrição (TALENs), que exigem complexos processos de engenharia proteica para cada novo alvo gênico (PORTEUS; BALTIMORE, 2003; WOOD et al., 2011).

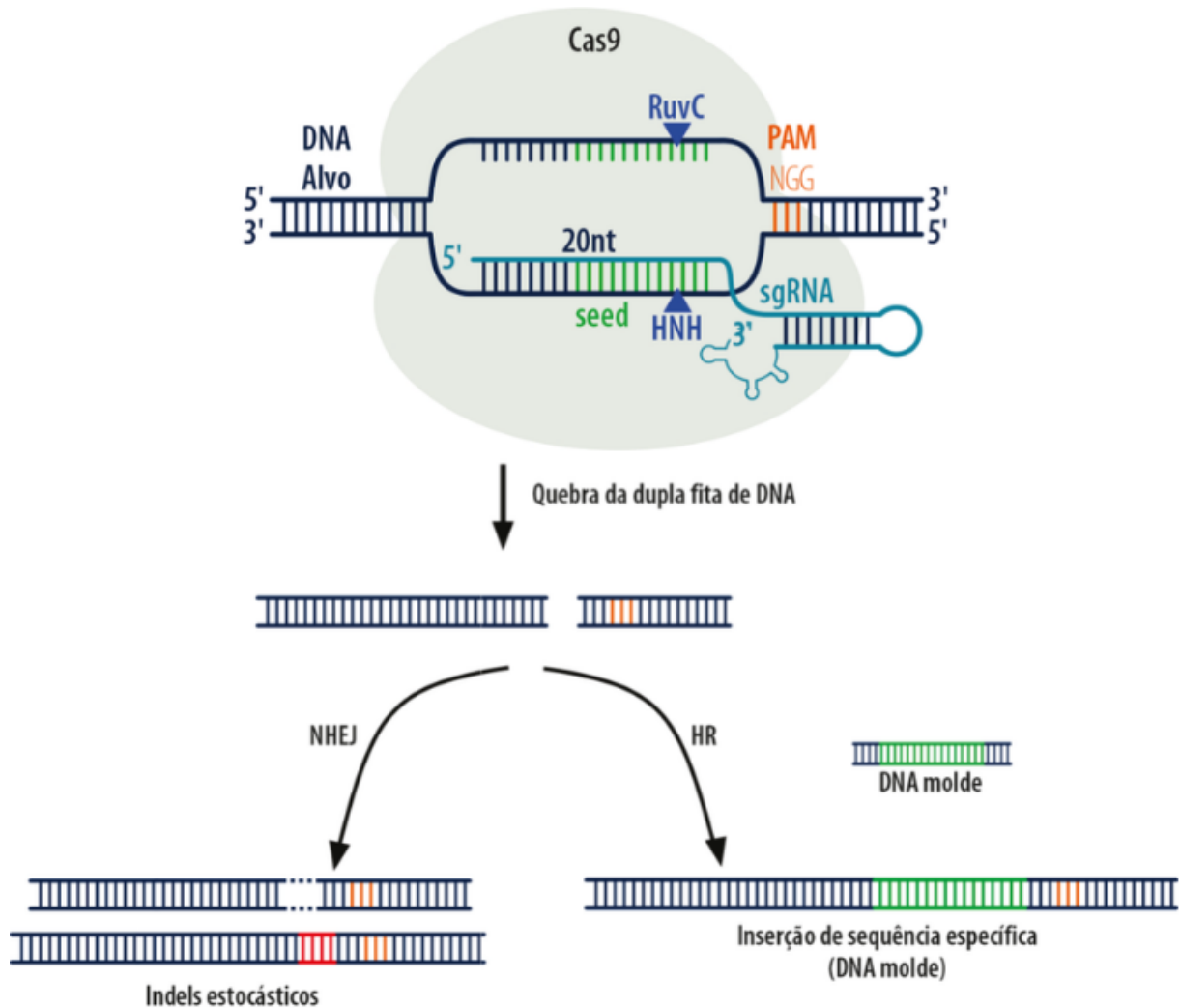
O sistema CRISPR/Cas9 da bactéria *Streptococcus pyogenes* (*SpCas9*) foi adaptado de um mecanismo de defesa imune procariótico, utilizado para clivar DNA de elementos genéticos invasores (GARNEAU et al., 2010). Na sua forma nativa, o sistema utiliza duas moléculas independentes (crRNA e tracrRNA); contudo, para a aplicação biotecnológica em eucariotos, a plataforma foi simplificada pela fusão dessas moléculas em uma única estrutura quimérica, o RNA guia sintético (sgRNA) (JINEK et al., 2012; MOLINARI et al., 2020). O sgRNA, que contém a sequência complementar ao DNA-alvo acoplada às estruturas secundárias para o reconhecimento da enzima, conduz a Cas9 até a sequência específica no genoma. O complexo ancora-se inicialmente a um motivo adjacente ao protoespaçador (PAM), que é reconhecido pelo lobo de reconhecimento da enzima (MOLINARI et al., 2020).

Em seguida, a atividade da helicase promove a abertura da dupla fita de DNA, permitindo o pareamento complementar de 20 a 24 nucleotídeos entre o RNA guia e a sequência alvo. Apesar de sua elevada precisão, a tecnologia CRISPR/Cas9 pode gerar clivagens fora do sítio alvo (*off-targets*), especialmente quando existem sequências genômicas altamente similares ao sgRNA. O sistema possui baixa tolerância a incompatibilidades na região *seed* (10 a 12 nucleotídeos próximos ao sítio PAM), mas pode tolerar pequenos *mismatches* fora dessa região (MOLINARI et al., 2020). Por isso, estratégias de desenho de sgRNAs e análises bioinformáticas são amplamente empregadas para minimizar esse risco (CHEN et al., 2019; LIU et al., 2017).

Após o pareamento completo, os domínios HNH e RuvC, responsáveis pela atividade endonucleásica da Cas9, promovem a clivagem da fita complementar e da não complementar do DNA. Essa clivagem ocorre de forma precisa no terceiro nucleotídeo *upstream* à PAM (CONG et al., 2013; JACOBS et al., 2015; MOLINARI et al., 2020). Essa quebra de fita dupla (*Double Strand Break* - DSB) ativa os mecanismos de reparo da célula vegetal, predominantemente a Junção de Extremidades Não Homólogas (NHEJ). Por ser

propensa a erros, a via NHEJ introduz inserções ou deleções (*indels*) no sítio de clivagem que frequentemente alteram a matriz de leitura da sequência codificante, resultando no *knockout* funcional do gene-alvo (JACOBS et al., 2015; MOLINARI et al., 2020).

Figura 4 – Representação esquemática da edição genômica via sistema CRISPR/Cas9 e vias de reparo de DNA.



Legenda: O diagrama ilustra a dinâmica molecular da edição, iniciando-se com o reconhecimento da sequência genômica alvo pelo sgRNA e a ancoragem do complexo ribonucleoproteico ao motivo PAM. Após a clivagem das fitas de DNA mediada pelos domínios RuvC e HNH no terceiro nucleotídeo *upstream* do PAM, a célula vegetal aciona seus mecanismos de reparo. O esquema contrasta a via de Reparo Dirigido por Homologia (HDR), que depende de um molde para edições precisas, com a Junção de Extremidades Não Homólogas (NHEJ). Destaca-se a rota NHEJ, predominante em plantas, que, devido à sua natureza propensa a erros, introduz inserções ou deleções (*indels*). Estas pequenas mutações provocam o deslocamento da matriz de leitura (*frameshift*), culminando na produção de proteínas truncadas e no *knockout* funcional do gene. **Fonte:** Elaborada pelo autor, adaptado de Molinari et al. (2020).

Essas pequenas mutações *indel* são suficientes para alterar a matriz de leitura do gene, podendo resultar na produção de uma proteína truncada e não funcional, o que

efetivamente causa o *knockout* do gene. A eficiência e a robustez deste sistema foram demonstradas em inúmeras espécies, incluindo *Arabidopsis thaliana*, arroz, milho e soja (FENG et al., 2013; JACOBS et al., 2015). Desta maneira, essa tecnologia viabiliza a otimização de características nutricionais e agronômicas complexas em larga escala.

Uma vantagem técnica determinante do sistema CRISPR/Cas9 é a independência física entre a mutação gerada e o maquinário de edição (transgene). Diferente da transgenia clássica, onde o gene exógeno permanece, a edição via CRISPR permite que, por cruzamentos ou autopolinizações subsequentes, o *locus* contendo o cassete Cas9/sgRNA seja segregado e eliminado, enquanto a mutação de interesse é mantida em homozigose. Essa capacidade de gerar linhagens livres de DNA exógeno (*transgene-free*) é o requisito biológico fundamental que embasa o enquadramento dessas plantas como convencionais (não-OGM) em legislações baseadas na avaliação do produto final, como a vigente no Brasil e em diversos países da América Latina (VIEIRA, 2022; ZARATE et al., 2023; BRASIL. Comissão Técnica Nacional de Biossegurança (CTNBio), 2018). Diante desse cenário, o sistema CRISPR/Cas9 configura-se como a ferramenta de melhoramento genético de precisão mais empregada para o desenvolvimento de culturas de alto valor agronômico, a exemplo da soja (YAO et al., 2023; ABDURAKHMON et al., 2025).

Como síntese das abordagens discutidas ao longo deste capítulo, a Tabela 2 consolida a evolução metodológica das estratégias empregadas para a mitigação de inibidores de protease em soja. A análise dessas abordagens evidencia que a convergência de limitações tecnológicas, agronômicas e regulatórias inerentes ao processamento térmico, ao melhoramento clássico e à transgenia (RNAi) impulsionou a adoção da edição genômica (CRISPR/Cas9) como a ferramenta de precisão metodológica ideal para superar esse entrave nutricional.

Tabela 2 – Comparação das estratégias utilizadas para a redução de inibidores de protease em soja.

Estratégia	Vantagem Principal	Limitação
Processamento térmico	Estabelecido industrialmente e amplamente aplicado.	Não elimina completamente a atividade do inibidor Bowman-Birk (BBI).
Melhoramento convencional	Abordagem não-transgênica.	Lento e suscetível à penalidade agronômica pelo arraste genético (<i>linkage drag</i>).
RNA de interferência (RNAi)	Alta eficiência no silenciamento transcricional do gene alvo.	Requer a integração permanente de DNA exógeno, submetendo a planta a rigorosas restrições regulatórias (OGM).
Edição genômica via CRISPR	Modificação precisa e capacidade de gerar linhagens livres de transgenes (<i>transgene-free</i>).	Risco de <i>off-targets</i> e alto nível de mosaicismo alélico em T ₀ , exigindo triagem complexa e segregação para eliminar o casete Cas9/sgRNA.

Fonte: Elaborada pelo autor (2026).

3 HIPÓTESE

A edição direcionada do *locus* codificante para a isoforma majoritária do Inibidor Bowman-Birk em sementes de soja, via sistema CRISPR/Cas9, promove o *knockout* funcional deste gene, resultando na redução significativa da atividade inibitória de tripsina e quimotripsina na cultivar BRS 537 e no consequente aprimoramento da qualidade proteica do grão.

4 OBJETIVOS

4.1 OBJETIVO GERAL

Obter e caracterizar molecularmente linhagens da cultivar de soja brasileira BRS 537 com *knockout* do gene Bowman-Birk majoritário nos grãos via sistema CRISPR/Cas9, visando à redução da atividade desses inibidores e à melhoria da qualidade proteica das sementes.

4.2 OBJETIVOS ESPECÍFICOS

Para o cumprimento do objetivo geral, o projeto estruturou-se sob três eixos principais:

I. Caracterização *In Silico*

- Consolidar o *locus Glyma.16G208900* como alvo prioritário para edição, fundamentando a escolha em análises comparativas de expressão gênica e filogenia da família BBI;
- Caracterizar a arquitetura *cis*-regulatória dos promotores, comparando o perfil de especialização regulatória do gene alvo com seus parálogos para avaliar a redundância funcional do grupo, visando identificar padrões de divergência funcional e o potencial de compensação genética frente a diferentes estímulos fisiológicos e ambientais;
- Avaliar, por meio de modelagem molecular e *docking*, a competência estrutural e a funcionalidade dos laços reativos da proteína alvo na inibição de tripsina e quimotripsina.

II. Engenharia Genética e Transformação Vegetal

- Projetar os RNAs guias e realizar a montagem dos vetores de expressão gênica via sistemas Golden Gate e Gateway, contendo o cassete da nuclease Cas9 direcionado ao *locus Glyma.16G208900*, confirmar a integridade das construções por mapeamento de restrição e sequenciamento.
- Obter plantas regeneradas *in vitro* da cultivar de soja BRS 537 transformadas pelo método combinado de bombardeamento de partículas de tungstênio e infecção por *Agrobacterium tumefaciens* utilizando como explantes os eixos embrionários.

III. Triagem Molecular e Seleção de Linhagens

- Identificar plantas transformadas da geração T_0 por meio de PCR para confirmar a integração do T-DNA ao genoma;
- Caracterizar a natureza e a eficiência das edições gênicas (mutações do tipo *indel*) por meio de sequenciamento e análise computacional;
- Identificar a perda da função inibitória do Bowman-Birk sobre as enzimas tripsina e quimotripsina nas linhagens editadas, fundamentando-se nas alterações da conformação proteica causadas por mutações de mudança da matriz de leitura (*frameshift*).

5 MATERIAL E MÉTODOS

5.1 ANÁLISES *IN SILICO* E CARACTERIZAÇÃO ESTRUTURAL DO ALVO

5.1.1 Prospecção Genômica e Seleção do Gene Alvo

A identificação dos membros da família multigênica BBI e a obtenção de suas respectivas sequências de DNA e aminoácidos foram conduzidas por meio de buscas no genoma de referência de *G. max* (L.) Merrill (*G. max* Wm82.a4.v1), disponível na base de dados pública Phytozome v14 (GOODSTEIN et al., 2012). O alinhamento das sequências e a inferência da árvore filogenética foram realizados utilizando o *software* MEGA v12.1 (STECHER et al., 2026). A escolha do gene alvo ocorreu por meio de análise comparativa dos perfis de expressão gênica, utilizando dados de transcriptoma (RNA-Seq) referentes aos diferentes estádios de desenvolvimento da semente, obtidos do Gene Atlas v.2 (SREEDASYAM et al., 2023). A integração desses dados permitiu correlacionar a filogenia dos parálogos com seus respectivos níveis de acúmulo de transcrito e identificar o gene *Glyma.16G208900* como a isoforma predominante durante a fase de enchimento de grãos. Para a visualização, a topologia da árvore gerada no MEGA foi exportada e renderizada na ferramenta *Interactive Tree Of Life* iTOL (LETUNIC; BORK, 2007). Concomitantemente, os dados numéricos de FPKM (Fragmentos por Quilobase de Transcrito por Milhão de Leituras Mapeadas) foram submetidos a um agrupamento hierárquico (*hierarchical clustering*) e plotados em formato de *heatmap* utilizando o *software* Morpheus (Broad Institute). A composição final da figura e o alinhamento vetorial dos painéis foram executados no *software* Inkscape v1.4.2.

5.1.2 Análise da Arquitetura Regulatória e Prospecção de Elementos *Cis*-Atuantes

Para a caracterização da arquitetura regulatória e investigação da divergência funcional pós-duplicação genômica, foram analisadas as regiões promotoras do gene alvo BBI (*Glyma.16G208900*) e de seus seis parálogos diretos: *Glyma.09G158700* (Par1A), *Glyma.09G158600* (Par1B), *Glyma.09G158800* (Par2A), *Glyma.14G117600* (Par2B), *Glyma.09G158900* (Par3A) e *Glyma.09G158500* (Par3B). Estes parálogos foram selecionados com base na identidade de suas sequências codificadoras (acima de 70% de similaridade) e organizados em módulos de conservação. As sequências correspondentes a 2000 pb *upstream* do códon de início da tradução (ATG) foram extraídas por meio da plataforma Phytozome v13. A adoção desse intervalo metodológico é validada para o genoma de *G. max* (L.) Merrill, pois abrange a densidade máxima de elementos *cis*-regulatórios proximais

e distais funcionais, garantindo a captura da sinalização genômica, hormonal e de estresse, minimizando o risco de sobreposição com regiões codificadoras adjacentes (CHEN et al., 2023; WANG et al., 2023a).

Para superar as limitações de cobertura e vieses algorítmicos inerentes a bancos de dados individuais, a identificação e o mapeamento dos elementos *cis*-regulatórios foram conduzidos por meio de uma abordagem integrativa, empregando três plataformas bioinformáticas independentes: PlantCARE (LESCOT, 2002), PlantPAN (CHOW et al., 2024) e PlantRegMap (TIAN et al., 2020).

Para evitar a poluição analítica com motivos basais de transcrição (ex.: *TATA-box*) ou elementos inespecíficos (ex.: elementos de resposta à luz), a prospecção inicial foi submetida a um filtro. Foram selecionados apenas os motivos que atingiram os seguintes limiares para cada plataforma: escore de similaridade $\geq 0,95$ no PlantPAN, filtragem funcional por correspondência no PlantCARE, e $p\text{-value} \leq 1 \times 10^{-4}$ no PlantRegMap.

De maneira concomitante à filtragem estatística, foi conduzida uma varredura estrutural *in silico* focada na prospecção de elementos de acessibilidade cromatínica, especificamente os motivos *AT-Hook* e *MARTBOX*. Estes elementos, caracterizados por regiões ricas em adenina e timina (tratos poli-A/T), constituem assinaturas genômicas típicas de Regiões de Fixação à Matriz (*Matrix Attachment Regions* - MARs). A presença e o adensamento destes sítios atuam no recrutamento de proteínas arquiteturais, como as da família HMG (*High Mobility Group*), que se inserem na fenda menor do DNA, induzindo curvaturas locais e a desestabilização de nucleossomos (ARAVIND; LANDSMAN, 1998; ALLEN et al., 1993).

A prospecção destes módulos arquiteturais fundamenta-se na premissa de que a formação de uma “zona de cromatina aberta” é o evento primário que reduz o impedimento estérico. Essa acessibilidade física atua como um pré-requisito cinético para o acoplamento eficiente da maquinaria transcricional. Para consolidar essa rede regulatória, os elementos funcionais triados deveriam se alinhar especificamente a três eixos fisiológicos inerentes à biologia dos inibidores Bowman-Birk (CLEMENTE et al., 2019; VORSTER et al., 2023):

1. **Desenvolvimento da Semente e Metabolismo de Reservas:** Mapeamento do motivo *RY-element* (CATGCA), reconhecido por fatores de transcrição B3 (ABI3/FUS3) em leguminosas, e dos motivos *O2-site* e *GCN4 motif*, associados à síntese de proteínas de reserva e expressão no endosperma.
2. **Defesa Aguda e Sistêmica:** Prospecção de motivos de resposta a ferimento mecânico e herbivoria (*WRE3* e *WUN-motif*), resposta genérica a estresse (*TC-rich*

repeats), além de elementos responsivos a vias hormonais de imunidade, como o ácido salicílico (*TCA-element*) e o ácido jasmônico (sítios de ligação *MYC* e *MYB*). Adicionalmente, investigou-se a presença do *W-box*, alvo específico para fatores de transcrição WRKY associados à senescência e defesa especializada.

3. **Sinalização Hormonal da Maturação:** Foco no complexo responsivo ao ácido abscísico (ABA), englobando variantes do *ABRE* e o *G-box*, que governam a dessecação e a entrada em dormência da semente.

Para a triagem da arquitetura regulatória, em detrimento da contabilização bruta de anotações isoladas obtidas nas três plataformas, foi realizado o rastreamento e o agrupamento dos fatores de transcrição em janelas de 30 pb ao longo dos 2000 pb. A adoção desta métrica baseia-se no impedimento estérico do complexo transcricional, onde motivos localizados dentro dessa distância física atuam de forma cooperativa em um mesmo módulo regulatório (WASSERMAN; SANDELIN, 2004). Essa estratégia permitiu a identificação de domínios funcionais integrando dados de múltiplos *softwares*. A análise comparativa final buscou elucidar os padrões de conservação, perda ou ganho de motivos de regulação (subfuncionalização regulatória) entre o promotor do gene alvo e seus parálogos.

5.1.3 Predição Estrutural e Refinamento

Dada a ausência de estruturas determinadas experimentalmente (cristalografia de raios-X ou RMN) para o *locus Glyma.16G208900* no Protein Data Bank (PDB), a arquitetura tridimensional da proteína foi gerada via modelagem computacional utilizando o algoritmo AlphaFold v2 (JUMPER et al., 2021).

Os inibidores da família BBI são sintetizados como pré-proteínas contendo um peptídeo sinal N-terminal (ODANI; IKENAKA, 1973) para direcionamento à via secretora (CLEMENTE et al., 2019), a identificação e remoção dessa região é imprescindível para a correta delimitação do domínio funcional (RICHARDSON, 1977). Portanto, a sequência de aminoácidos foi processada no servidor SignalP 6.0 (TEUFEL et al., 2022), que indicou a presença de um peptídeo sinal com alta probabilidade de clivagem (> 0,99).

Após a identificação do peptídeo sinal, realizou-se um refinamento da sequência madura por meio de biologia comparativa. A sequência foi alinhada a homólogos com estrutura cristalográfica resolvida (PDB IDs 1K9B, 1D6R, 1BBI, 2BBI, 5J4Q e 5J4S) utilizando o *software* Jalview (WATERHOUSE et al., 2009) (Tabela 3). Observou-se que a sequência deduzida de *Glyma.16G208900* possuía extensões N e C-terminais ausentes nas formas processadas das referências citadas.

Tabela 3 – Alinhamento múltiplo das sequências de inibidores de protease do tipo Bowman–Birk (BBI) com destaque para variações em relação ao consenso estrutural.

Pos.	1	10	20	30	40	50	60	70
1K9B_1	-----	KPCC D	QCACTKSNP P	QCRCSDMRL N	SCHSACKSC I	CALSYPAQC F	CVDITDFCY E	PCK-----
1D6R_2	-----	KPCC D	QCACTKSNP P	QCRCSDMRL N	SCHSACKSC I	CALSYPAQC F	CVDITDFCY E	PCK-----
<i>Glyma.16G208900</i>	DDEYS	KPCC D	LCMCT R SMP P	QCS C ED I RL N	SCHSDCKSC M	CT R SQP G QC R	CLDTNDFCY K	PCK S RDD---
1BBI_1	DDESSKPCC D	QCACTKSNP P	QCRCSDMRL N	SCHSACKSC I	CALSYPAQC F	CVDITDFCY E	PCKPSEDDK E	N
2BBI_1	DDESSKPCC D	QCACTKSNP P	QCRCSDMRL N	SCHSACKSC I	CALSYPAQC F	CVDITDFCY E	PCKPSEDDK E	N
5J4Q_2	DDESSKPCC D	QCACTKSNP P	QCRCSDMRL N	SCHSACKSC I	CALSYPAQC F	CVDITDFCY E	PCKPSEDDK E	N
5J4S_2	DDESSKPCC D	QCACTKSNP P	QCRCS D LRL N	SCHSACKSC I	CT F S I PPQC F	CVDITDFCY E	PCKPSEDDK E	N

Nota: Resíduos idênticos ao consenso estrutural estão indicados em **azul** e **verde** alternados por coluna; divergências estão destacadas em **vermelho**; gaps são representados por traços. Os resíduos em negrito e sublinhados (**R16** e **R43**) na sequência *Glyma.16G208900* indicam os sítios reativos.

Fonte: Elaborado pelo autor a partir de sequências estruturais depositadas no Protein Data Bank e sequência de aminoácidos do gene *Glyma.16G208900* retirada da plataforma Phytozome v13.

Com base nesses parâmetros, a sequência primária foi editada *in silico* para a remoção da região sinalizadora e das extensões terminais que não estão presentes, mimetizando a forma madura da proteína após o processamento. A sequência refinada foi então submetida a uma nova rodada de modelagem no AlphaFold v2. Esta abordagem fundamenta-se na tendência de o AlphaFold v2 tender a atribuir baixos valores de confiança (pLDDT < 50) a regiões intrinsecamente desordenadas ou não estruturadas, que frequentemente englobam o peptídeo sinal. A remoção dessas regiões evita artefatos conformacionais que poderiam comprometer a precisão das simulações subsequentes de *docking* molecular (AKDEL et al., 2022).

Além do processamento das extremidades, o alinhamento evidenciou mutações de carga, polaridade e volume nos aminoácidos da isoforma alvo em comparação aos cristais clássicos (como o 1BBI). Uma das alterações ocorre no segundo laço reativo, onde a substituição de uma Leucina por uma Arginina (**R43**). O alvo também apresenta um mapa eletrostático diferente, com ganho de cargas negativas no primeiro laço e inversão polar na região C-terminal, aliado a um aumento de flexibilidade estérica promovido por resíduos como G47 e Q45. Essa divergência estrutural limita o uso direto de homólogos do PDB, tornando a modelagem computacional (AlphaFold v2) essencial para garantir maior fidelidade nas análises de interação..

5.1.4 Visualização e *Docking* Molecular

Em complemento à validação topológica, realizou-se o mapeamento espacial dos resíduos constituintes dos laços de inibição, visando caracterizar os sítios reativos do modelo predito. Para o locus *Glyma.16G208900*, os resíduos Arg16 e Arg43 (numeração da

proteína refinada) foram identificados como determinantes da especificidade inibitória, com base no alinhamento estrutural e conservação topológica. Em contraste com as estruturas clássicas resolvidas do BBI como (PDB ID: 1BBI) e as demais mostradas na Tabela 3, que possui Lys e Leu (Lisina e Leucina) nos loops reativos (compatíveis com inibição de tripsina e quimotripsina, respectivamente), o alvo deste estudo apresenta duas Argininas. Embora distintos, a sobreposição estrutural indica que a geometria dos *loops* se mantém preservada, sugerindo funcionalidade.

As análises estruturais e representações gráficas foram realizadas no *software* PyMOL v2.5.

Para verificar a funcionalidade do modelo, foram realizadas simulações de atracamento molecular (*molecular docking*) no servidor HDock v1.1 (YAN et al., 2020), que emprega um algoritmo híbrido de reconhecimento de moldes e *docking* proteína-proteína. Para simular o potencial de inibição, foram utilizadas como alvos as estruturas cristalográficas da Tripsina (PDB ID: 1D6R) e α -Quimotripsina (PDB ID: 5J4Q). A metodologia permitiu a avaliação de complexos binários e a viabilidade estérica de complexos ternários. Para a modelagem do complexo ternário, adotou-se a estratégia de *docking* sequencial: o modelo de melhor afinidade obtido no complexo binário inicial (inibidor e a primeira protease) foi empregado como plataforma receptora para a ancoragem independente da segunda enzima. Esta abordagem permite a identificação da interface livre restante, justificando tecnicamente a obtenção de um modelo final onde as três cadeias proteicas coexistem em conformação espacial estável.

A validação das interações baseou-se nos *scores* de energia de *docking* (indicativos de afinidade termodinâmica favorável) e na mensuração da proximidade atômica entre a tríade catalítica das proteases e os resíduos reativos (Arg16 e Arg43) do modelo. Adotou-se um limiar de distância inferior a 4,0 Å para a identificação de contatos fisicamente plausíveis, englobando ligações de hidrogênio e interações de van der Waals. Este critério foi estabelecido não apenas para avaliar a complementaridade de superfície, mas para garantir a viabilidade geométrica necessária à formação do complexo de Michaelis, onde o oxigênio nucleofílico ($O\gamma$) da serina catalítica deve estar adequadamente posicionado em relação ao carbono carbonílico da alça reativa do inibidor. Essa abordagem é consistente com o mecanismo de inibição canônica (mimético de substrato) e de bloqueio estérico do sítio ativo descrito para inibidores de serino-proteases (LASKOWSKI; KATO, 1980; VANGONE; BONVIN, 2015).

5.2 DESENHO DOS SGRNAS E CONSTRUÇÃO DOS VETORES DE EDIÇÃO GENÔMICA

5.2.1 Desenho dos RNAs guia e síntese dos oligonucleotídios

Para a edição genômica direcionada do gene *Glyma.16G208900*, que codifica o alvo, foi utilizado o sistema CRISPR/Cas9. A seleção das sequências-alvo foi realizada *in silico* por meio da plataforma CRISPR-P 2.0 (LIU et al., 2017). Esta ferramenta permitiu a identificação de regiões com motivos PAM (Protospacer Adjacent Motif) do tipo NGG otimizados para a nuclease SpCas9, garantindo alta eficiência de clivagem no genoma da soja. Foram selecionados dois RNAs guia (sgRNAs), designados sgRNA1 e sgRNA6.

Para avaliar a especificidade dos guias e mitigar o risco de clivagens inespecíficas, o potencial de *off-targets* foi rastreado utilizando o algoritmo Cas-OFFinder (BAE et al., 2014). A busca foi parametrizada contra o genoma de referência da soja (*G.max Wm82.a4.v1*), tolerando até quatro *mismatches* e a presença de protuberâncias (*bulges*) de DNA ou RNA na sequência.

Após a validação computacional, os pares de oligonucleotídeos complementares (Forward e Reverse) correspondentes a cada guia foram sintetizados comercialmente (Exxtend, Brasil). Para permitir a clonagem direcional no vetor de expressão CRISPR-zCas9, foram incorporadas extremidades coesivas (*overhangs*) específicas nas extremidades 5' de cada fita: a sequência **ggtc** nos iniciadores *forward* e **aaac** nos iniciadores *reverse* (Tabela 4).

Tabela 4 – Sequências dos oligonucleotídios (sgRNAs) desenhados para a clonagem no vetor CRISPR/zCas9.

Alvo (Guia)	Identificação	Sequência (5' → 3')
sgRNA1	CRISPR-sg1-F	ggtc AGGGGGTTACAGCTGCACGCA
	CRISPR-sg1-R	aaacTGCGTGCAGCTGTAACCCCT
sgRNA6	CRISPR-sg6-F	ggtc ATATCTTCACAGCTGCATTG
	CRISPR-sg6-R	aaacCAATGCAGCTGTGAAGATAT

Nota: Sequências de RNAs guias (sgRNAs) utilizadas para direcionar o sistema CRISPR/Cas9 à região-alvo de edição no genoma da soja. Os sgRNAs foram selecionados utilizando a plataforma CRISPR-P 2.0, e os potenciais off-targets foram identificados por meio do algoritmo Cas-OFFinder. As bases em letras minúsculas representam as extremidades coesivas (*overhangs*) adicionadas para a clonagem via enzima de restrição.

Fonte: Elaborada pelo autor (2026). Oligonucleotídeos sintetizados pela Exxtend (Brasil).

5.2.2 Preparo dos insertos de sgRNA por fosforilação e anelamento

Para gerar os insertos de fita dupla aptos para ligação, foi realizada uma reação de fosforilação e anelamento. Em um volume de reação de 20 µL, foram combinados 1 µL de cada oligonucleotídio (Forward e Reverse a 100 µM), 2 µL de Tampão Tampão T4

PNK 10X com 1 μ L de ATP (10 mM) e 15 μ L de água ultrapura. A reação de fosforilação foi incubada em termociclador a 37 °C por 30 minutos. Em seguida, para o anelamento, a mistura foi submetida a 95 °C por 5 minutos, com resfriamento gradual de -0,1 °C/s até 25 °C. O produto final foi diluído (1:100) em água livre de nucleases para a etapa subsequente.

5.2.3 Clonagem dos sgRNAs no vetor pYPQ141b e validação molecular

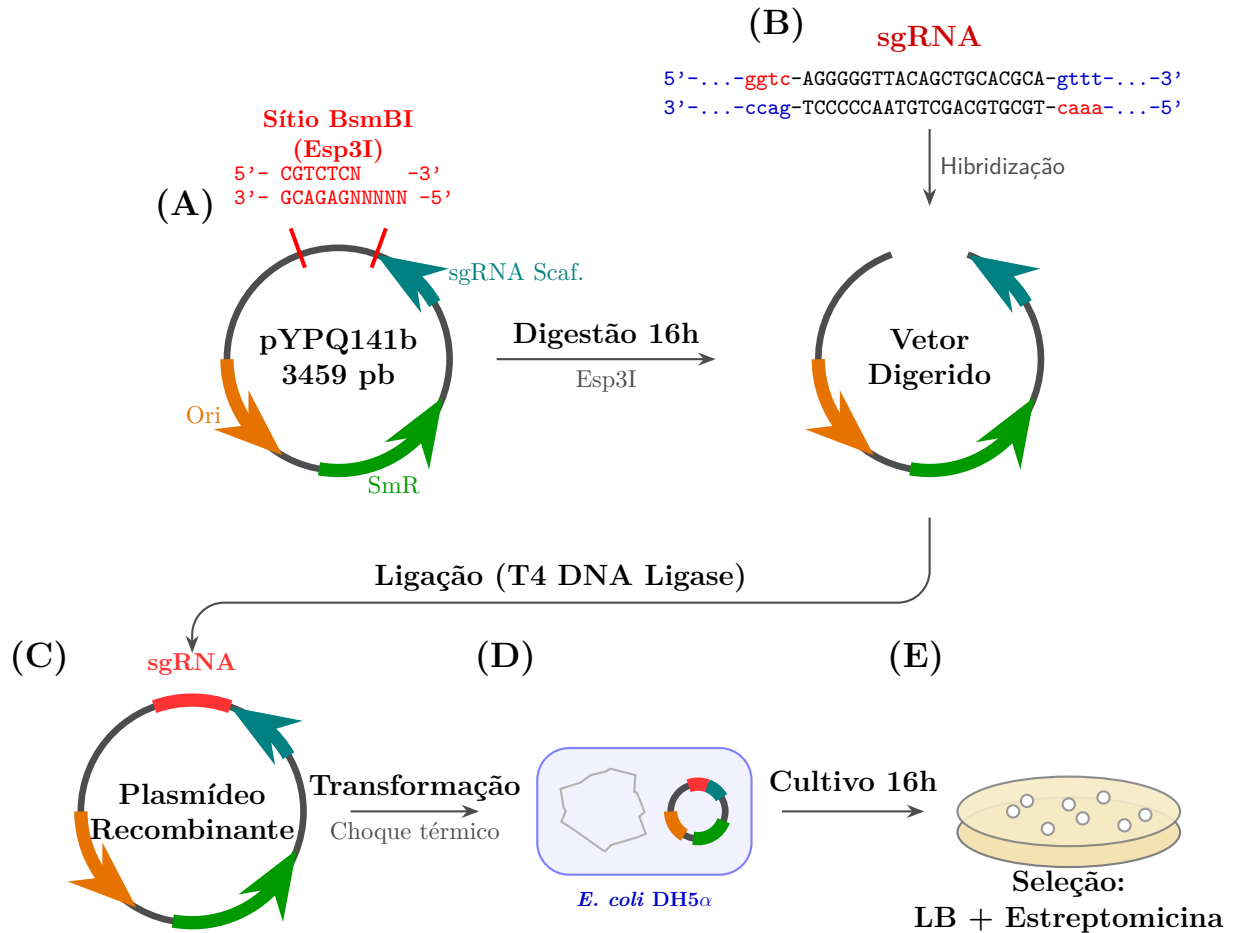
Os sgRNAs foram clonados no vetor de entrada pYPQ141b; como demonstrado na Figura 5, projetado para a inserção de um único guia sob o controle do promotor *U3* de *A. thaliana* (LOWDER et al., 2015). Inicialmente, o vetor foi linearizado por meio de digestão com a enzima de restrição *BsmBI* (Esp3I) (Thermo Fisher Scientific). A reação de clivagem seguiu o protocolo do fabricante, empregando o Tampão Tango 1X suplementado com DTT, com incubação a 37 °C por 16 horas e posterior inativação térmica a 65 °C por 20 minutos.

Após a inativação, o vetor linearizado foi purificado utilizando o QIAEX II Gel Extraction Kit (QIAGEN), seguindo as recomendações do fabricante. A reação de ligação foi preparada em um volume final de 10 μ L, contendo 100 ng do vetor pYPQ141b digerido, 2 μ L do inserto de sgRNA anelado e diluído, 1 μ L de Tampão T4 DNA Ligase 10X e 0,5 μ L da enzima T4 DNA Ligase (Thermo Fisher Scientific). A mistura reacional foi incubada em termociclador a 20 °C por 16 horas.

Na sequência, 5 μ L do produto da ligação foram utilizados para transformar células quimicamente competentes de *Escherichia coli* (linhagem DH5 α) por meio do método de choque térmico. O protocolo consistiu na incubação das células com o plasmídeo em gelo por 10 minutos, seguida de incubação a 42 °C por 45 segundos em banho-maria e imediato retorno ao gelo por 2 minutos. Após a adição de 1 mL de meio Luria-Bertani (LB) líquido (SEZONOV et al., 2007), as células foram recuperadas a 37 °C por 1 hora sob agitação (250 rpm).

As culturas foram então centrifugadas, o sobrenadante parcialmente descartado (restando $\sim$ 100 μ L) e o concentrado celular foi plaqueado em meio LB ágar suplementado com 50 μ g/mL de espectinomicina, com incubação *overnight* a 37 °C. Colônias individuais foram selecionadas para cultivo em meio LB líquido contendo o mesmo antibiótico e incubadas durante a noite a 37 °C com agitação (180 rpm). Para a extração do DNA plasmidial, foi utilizado o QIAprep Spin Miniprep Kit (QIAGEN), seguindo o protocolo do fabricante. A confirmação da inserção correta de cada sgRNA foi realizada por sequenciamento de Sanger (SANGER et al., 1977) (ACTGene Análises Moleculares, Brasil) utilizando o oligonucleotídeo universal M13-F.

Figura 5 – Fluxograma da clonagem do sgRNA no vetor de expressão pYPQ141b.



Legenda: (A) O plasmídeo receptor pYPQ141b (3.459 pb), que possui o gene de resistência à estreptomicina (*SmR*) e a origem de replicação (*Ori*), é submetido à digestão enzimática com a endonuclease de restrição *BsmBI* (*Esp3I*). Este processo gera extremidades coesivas assimétricas de 4 nucleotídeos. (B) Paralelamente, os oligonucleotídeos sintetizados para o *sgRNA* *guide* são submetidos ao anelamento e hibridização para garantir a compatibilidade de ligação. (C) A reação de ligação é mediada pela enzima *T4 DNA Ligase*, promovendo a integração do guia ao vetor e resultando no plasmídeo recombinante circularizado. (D) O produto da ligação é inserido em células competentes de *E. coli* DH5α via choque térmico. (E) Seleção fenotípica de colônias transformantes em meio sólido LB suplementado com espectinomicina. Colônias isoladas foram submetidas à extração de DNA plasmidial e posterior sequenciamento de Sanger para confirmar a correta inserção do sgRNA.

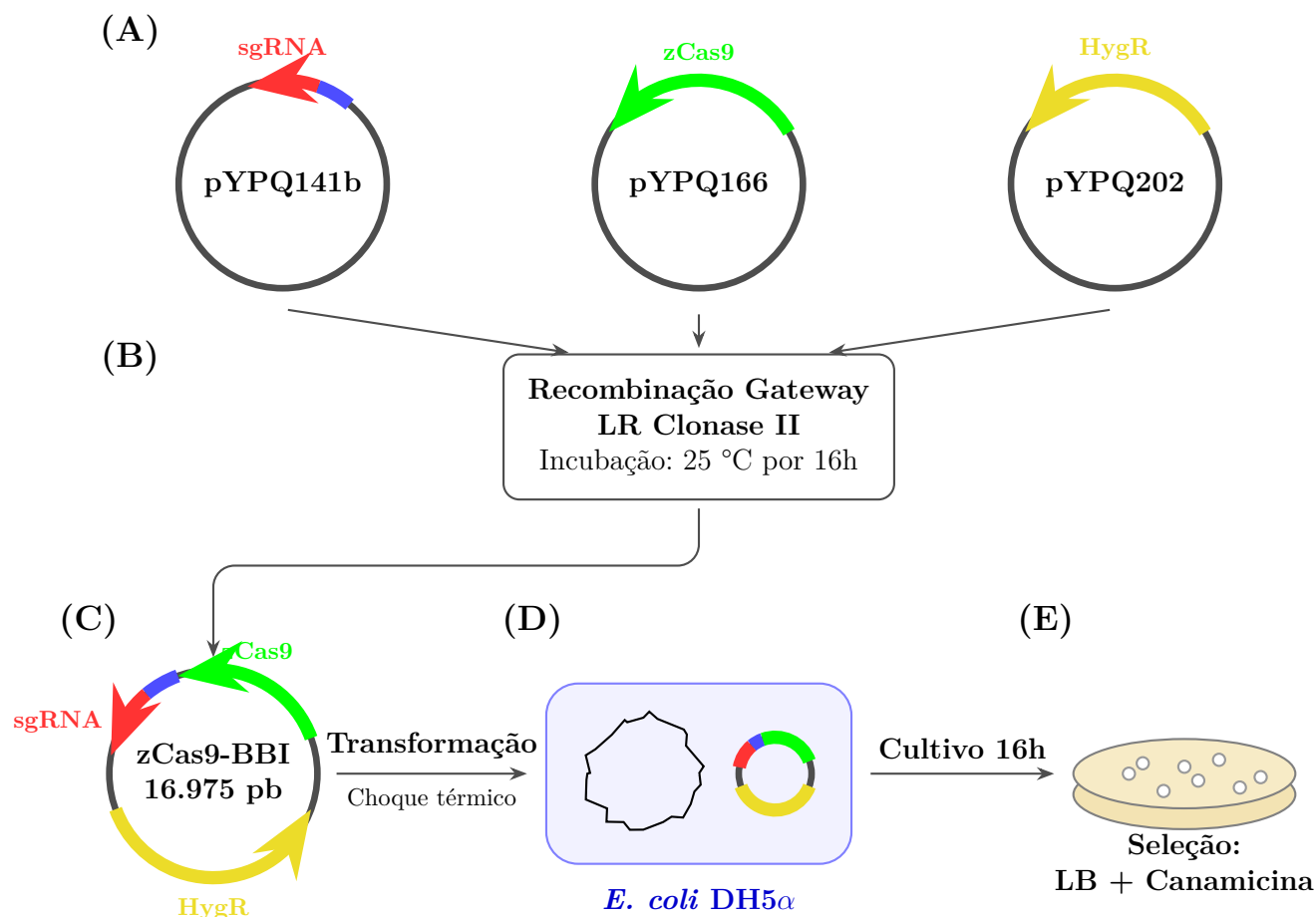
Fonte: Elaborada pelo autor (2026).

5.2.4 Engenharia do vetor zCas9-BBI por recombinação Gateway

Após a validação por sequenciamento e digestão, os vetores de entrada pYPQ141b contendo os sgRNAs validados foram utilizados para a montagem dos vetores de expressão finais, conforme demonstrado na Figura 6 via reação de recombinação Gateway. A reação foi preparada em um volume final de 6 µL, composta por 150 ng do vetor pYPQ141b (portando o sgRNA), 150 ng do vetor pYPQ166 (contendo o gene da nuclease zCas9) e 200 ng do vetor de destino pYPQ202 contendo o gene de resistência à higromicina (LOWDER

et al., 2015). A recombinação foi catalisada pela adição de 0,8 µL da enzima LR Clonase II (Invitrogen, Thermo Fisher Scientific), seguida de incubação a 25 °C por 16 horas.

Figura 6 – Fluxograma da engenharia do vetor zCas9-BBI por recombinação via Gateway.



Legenda: (A) Vetores utilizados na reação de montagem. Os vetores pYPQ141b (contendo o *sgRNA*), pYPQ166 (contendo o gene da *zCas9*) juntamente com o vetor pYPQ202 (contendo o gene *HygR*). (B) A recombinação é mediada pela enzima LR Clonase II, incubada a 25 °C por 16 horas. (C) O produto da reação é o vetor de expressão zCas9-BBI, que une os três cassetes funcionais. (D) O vetor recombinante é introduzido em células competentes de *E. coli* DH5α por choque térmico. (E) As colônias obtidas são cultivadas e o DNA plasmidial foi isolado pelo método de lise alcalina, conforme detalhado na Seção 5.2.3, seguido de validação estrutural via digestão enzimática com *BamHI*.

Fonte: Elaborada pelo autor (2026).

O produto da recombinação, que foi chamado de zCas9-BBI 1 e 6, foi utilizado para a transformação de células termocompetentes de *E. coli* (linhagem DH5α) via choque térmico, conforme descrito na Seção 5.2.3. Os clones obtidos foram selecionados e submetidos à validação estrutural por meio de digestão enzimática com a endonuclease *BamHI* (sistema Anza™ 5, Thermo Fisher Scientific). Previamente à digestão, o DNA plasmidial foi isolado pelo método de lise alcalina, conforme detalhado na Seção 5.2.3. A reação de clivagem foi realizada com o Tampão Anza 1X, incubada a 37 °C por 60 minutos,

com posterior inativação térmica a 80 °C por 5 minutos.

5.2.5 Transformação de *A. tumefaciens*

Para cada vetor zCas9-BBI, 200ng do plasmídeo foram misturados a 50µL de células eletrocompetentes de *A. tumefaciens* GV3101 em uma cubeta de eletroporação (espaçamento de 0,2 cm). Um pulso elétrico foi aplicado utilizando um eletroporador Bio-Rad Gene Pulser® sob as configurações padrão do equipamento. Imediatamente após o pulso, 1 mL de meio LB foi adicionado à cubeta, e a suspensão celular foi transferida para um tubo de 1,5 mL. As células foram recuperadas por incubação a 28°C por 2 horas sob agitação de 100 rpm. Alíquotas da cultura recuperada foram plaqueadas em meio LB ágar suplementado com Rifampicina (50 mg/L) antibiótico usado para seleção da linhagem GV3101 e Canamicina (100 mg/L) para seleção do vetor de transformação. As placas foram incubadas a 28°C por 48 horas, colônias individuais foram selecionadas e a presença das construções de interesse nos clones transformados foi confirmada por PCR de colônia (*Colony PCR*). Para tanto, frações de colônias isoladas foram coletadas com auxílio de ponteiras estéreis e ressuspensas em 50 µL de água ultrapura (Milli-Q) estéril. A lise celular foi promovida por tratamento térmico a 95 °C por 5 minutos, seguida de centrifugação a 14.000 rpm por 15 segundos para a sedimentação dos detritos celulares. Alíquotas de 2 µL do sobrenadante foram utilizadas como DNA molde nas reações de amplificação. As reações foram preparadas utilizando a DNA polimerase *Taq Platus* (Sinapse Biotecnologia), seguindo as condições recomendadas pelo fabricante, utilizando os primers ta Tabela 5.

Tabela 5 – Sequência dos primers utilizados para a caracterização molecular das colônias transformadas via PCR de colônia.

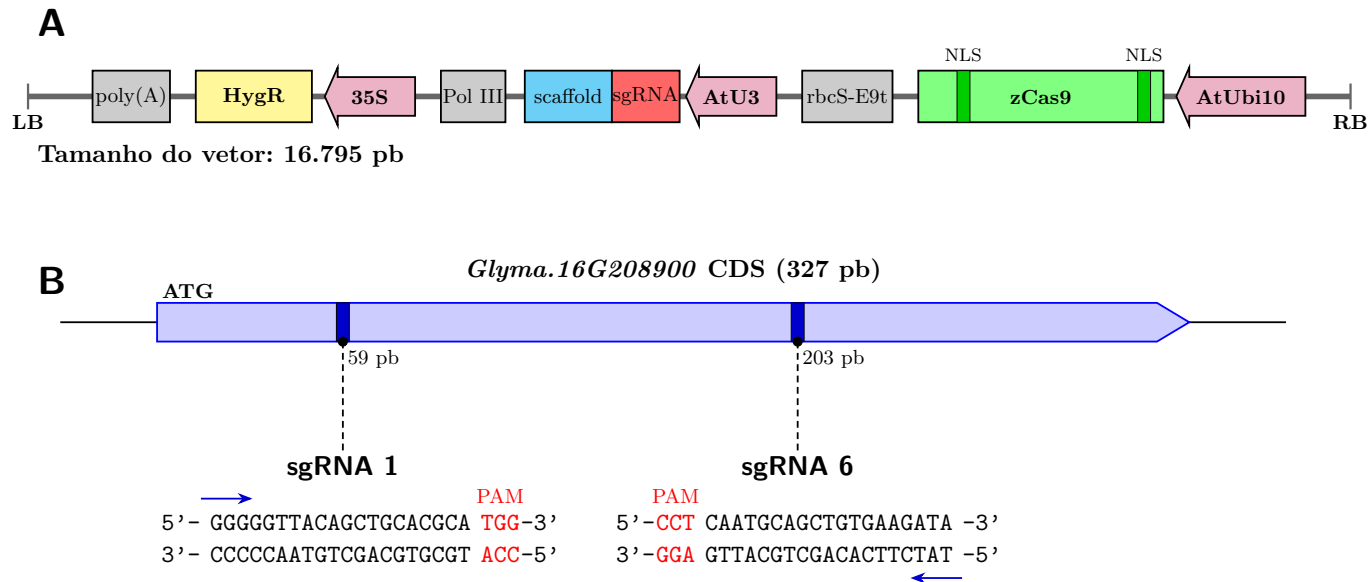
Nome do Primer	Sequência (5' → 3')	Tamanho do amplicon (pb)
zCas9 Fw	CCGAGCTGAGACAGGTTCGAT	1224
zCas9 Rv	TGAAAAAGTGGCACCGAGTCG	

Nota: Primers desenhados para amplificar fragmento interno contendo os resíduos dos vetores pYPQ141b+sgRNAs, pYPQ166 e pYPQ202 (zCas9-BBI) utilizados na recombinação.

Fonte: Elaborada pelo autor (2026).

Estes primers foram desenhados para flanquear regiões que abrangem parte do RNA *scaffold* e parte do gene da nuclease zCas9. Essa estratégia permite verificar a integridade da montagem final, pois o fragmento amplificado atravessa os sítios de junção resultantes da recombinação dos diferentes vetores utilizados. Clones que apresentaram o *amplicon* de tamanho 1224pb foram cultivados e utilizados para a transformação da soja.

Figura 7 – Mapa esquemático da região do T-DNA dos vetores zCas9-BBI e localização dos alvos no gene *Glyma.16G208900*.



Legenda: (A) Representação linear esquemática da região do T-DNA dos vetores zCas9-BBI. A estratégia utiliza dois vetores independentes, cada um carreando um guia específico. Os componentes ilustram a orientação transcricional: **Cassete de seleção:** gene HygR impulsionado pelo promotor 35S; **Cassete Cas9:** nuclease zCas9 impulsionada pelo promotor AtUbi10; **Cassete sgRNA:** sgRNA sob controle do promotor AtU3. (B) Organização estrutural do gene alvo *Glyma.16G208900*. A região codificante (CDS) de 327 pb está representada pela caixa azul, com o códon de iniciação (ATG) demarcado. Os pequenos retângulos indicam a posição de anelamento dos guias na sequência alvo. Abaixo, o detalhamento das sequências-alvo alinhadas à fita do CDS, ladeadas pelos respectivos motivos PAM destacados em vermelho. As setas azuis indicam a fita em que a sequência do sgRNA se liga, sendo a fita senso no caso do sgRNA 1 e a fita antisenso no sgRNA 6.

Fonte: Elaborada pelo autor (2026).

5.3 TRANSFORMAÇÃO GENÉTICA DE SOJA E REGENERAÇÃO DE PLANTAS

5.3.1 Preparo do Inóculo de *A. tumefaciens*

Clones de *A. tumefaciens* da linhagem GV3101, previamente transformados com os vetores binários finais e confirmados por PCR de colônia, foram utilizados para preparar o inóculo. Culturas bacterianas foram iniciadas em 50 mL de meio LB líquido suplementado com Rifampicina (50 mg/L) e Canamicina (100 mg/L) e incubadas a 28°C sob agitação (200 rpm) por 16 a 20 horas, até atingirem a fase logarítmica de crescimento (DO 0.6-0.9). As células bacterianas foram então coletadas por centrifugação (6000 rpm por 5 minutos), o sobrenadante descartado, e o pellet celular ressuspenso em meio de co-cultivo líquido CCM (HOOD et al., 1993) suplementado com acetossiringona (200 µM). A suspensão bacteriana foi incubada a 28°C sob agitação (200 rpm) por 40 minutos para indução dos genes de virulência (*vir*) antes da etapa de infecção dos explantes.

5.3.2 Transformação e Regeneração *In Vitro* de Explantes de Soja

Sementes maduras da cultivar de soja BRS 537 foram submetidas a um processo de desinfestação superficial consistindo em imersão em etanol 70% (v/v) por 5 minutos, seguido de tratamento com solução de hipoclorito de sódio 2,5% (v/v) por 20 minutos e três lavagens consecutivas com água destilada estéril. As sementes desinfestadas foram então embebidas em água destilada estéril e mantidas no escuro a 25°C por $\approx$ 20 horas para hidratação. Os eixos embrionários foram excisados das sementes hidratadas, utilizando a região do meristema apical como explante primário conforme o protocolo de (MELO et al., 2020).

A transformação foi conduzida seguindo o protocolo adaptado de Melo et al. (2020). Inicialmente, os eixos embrionários foram dispostos em meio de cultura e submetidos ao bombardeamento com micropartículas de tungstênio visando a indução de microlesões no tecido do meristema apical. Imediatamente após essa etapa, a suspensão celular de *A. tumefaciens* previamente preparada na Seção 5.2.5, foi gotejada sobre os explantes, assegurando o contato efetivo da bactéria com as regiões lesionadas.

Após a inoculação, os embriões foram incubados no escuro a 28°C por 24 horas. Em sequência, para interromper a fase de infecção, os explantes foram imersos em solução de cefotaxima (500 mg/L) por 40 minutos. O uso deste antibiótico, de natureza bacteriostática, permitiu a inibição do metabolismo bacteriano residual, assegurando condições adequadas para as etapas subsequentes de regeneração *in vitro*. Os explantes lavados foram então transferidos para frascos de cultura contendo meio de regeneração MS (MURASHIGE; SKOOG, 1962), suplementado com 0,1% (m/v) de carvão ativado e higromicina (6 mg/L)

para a seleção química dos transformantes. Os frascos foram mantidos em sala de crescimento a 25°C, sob fotoperíodo de 16 horas, por 3 a 4 semanas de cultivo *in vitro*, até a obtenção de plântulas com parte aérea e sistema radicular desenvolvidos.

5.3.3 Aclimação das Plântulas Regeneradas

Os vasos contendo as plantas de soja regeneradas que apresentavam bom desenvolvimento da parte aérea e radicular foram cuidadosamente removidas do cultivo *in vitro*, sendo transplantadas para vasos contendo substrato BioPlant Plus. Os vasos foram inicialmente cobertos com sacos plásticos transparentes para manter alta umidade relativa e mantidos em casa de vegetação. Após 5 dias, foram feitas perfurações nos sacos plásticos, e a cobertura foi removida ao longo de 7 a 10 dias para permitir a aclimação progressiva das plantas às condições ambientais (geração T_0). As plantas aclimatadas foram cultivadas até a maturidade para produção de sementes (geração T_1). A caracterização molecular inicial para confirmação da transformação foi realizada na geração T_0 , conforme descrito na seção subsequente.

5.4 CARACTERIZAÇÃO MOLECULAR DOS EVENTOS EDITADOS

5.4.1 Extração de DNA Genômico

Para a análise molecular inicial das plantas transformadas (Geração T_0), foi realizada a extração de DNA genômico a partir de tecido foliar jovem. Cerca de 100 mg de tecido vegetal (2-3 discos foliares) foram coletados de cada planta aclimatada assim que os primeiros trifólios se desenvolveram. A extração foi realizada utilizando o método baseado em (DELLAPORTA et al., 1983). Resumidamente, o tecido foliar foi macerado em nitrogênio líquido, seguido pela adição de tampão de extração e incubação a 65°C. Após etapas de deproteinização com clorofórmio: álcool isoamílico (24:1) e precipitação do DNA com isopropanol, o DNA genômico foi lavado com etanol 70%, seco e ressuspenso em água ultrapura. A concentração e a pureza do DNA foram estimadas por espectrofotometria (NanoDrop™, Thermo Fisher Scientific®) e a integridade foi verificada por eletroforese em gel de agarose 1% (m/v).

5.4.2 Triagem de Plantas Transformadas por PCR

A confirmação da presença do T-DNA (contendo o cassete *zCas9*) no genoma das plantas T_0 foi realizada por PCR convencional. A reação foi preparada utilizando 100 ng de DNA genômico como molde e os mesmos primers utilizados para a confirmação da transformação em *A. tumefaciens* (descritos na seção 5.2.5), desenhados para flanquear as regiões do RNA *scaffold* e do gene da nuclease *zCas9*.

A reação de amplificação foi realizada utilizando o kit GoTaq® Master Mix (Promega®), seguindo o programa de termociclagem detalhado na Tabela 6.

Tabela 6 – Programa de termociclagem utilizado na PCR para a confirmação dos eventos de transformação.

Etapa	Temperatura	Tempo	Ciclos
Desnaturação Inicial	95°C	3 min	1
Desnaturação	95°C	30 seg	40
Anelamento	57°C	30 seg	
Extensão	72°C	1,5 min	
Extensão Final	72°C	5 min	1
Manutenção	12°C	∞	-

Fonte: Elaborada pelo autor (2026).

Os produtos da PCR foram analisados por eletroforese em gel de agarose 1% (m/v), corado com brometo de etídio e visualizados em transiluminador UV (Gel Doc EZ Gel Imaging System BioRad). As amostras que apresentaram a banda de tamanho esperado (1224 pb) foram consideradas eventos de transformação positivos. Como controles negativos, foram utilizados DNA genômico de plantas BRS 537 não transformadas (tipo selvagem) e água ultrapura (controle sem molde).

5.4.3 Confirmação das Edições Gênicas por Sequenciamento

Para verificar a ocorrência de mutações do tipo inserção/deleção (*indels*) no gene alvo *Glyma.16G208900*, induzidas pelo sistema CRISPR/Cas9, a região genômica flanqueando os sítios alvo dos sgRNAs foi amplificada por PCR a partir do DNA genômico das plantas T₀ positivas. Foram utilizados os primers listados na Tabela 7.

Tabela 7 – Sequência dos primers utilizados para a amplificação e sequenciamento da região alvo no gene *Glyma.16G208900*.

Nome do Primer	Sequência (5' → 3')	Tamanho do amplicon (pb)
BBI-Seq-F	ATGACATGTAACCTCCACGACAGA	403
BBI-Seq-R	CAACGACACTGTCCTGGCTG	

Nota: O primer Forward (BBI-Seq-F) foi utilizado também para a reação de sequenciamento de Sanger.

Fonte: Elaborada pelo autor (2026).

Estes primers foram desenhados para amplificar um fragmento de 403 pb contendo os sítios de clivagem esperados. A reação de PCR foi realizada conforme descrito na Seção 5.4.2, utilizando o kit GoTaq® Master Mix (Promega®), porém ajustando a temperatura de anelamento para 60°C.

Os produtos de PCR foram purificados por precipitação, ajustando-se o volume da reação (25 μ L) para 200 μ L com água ultrapura, seguido da adição de 400 μ L de isopropanol gelado. A mistura foi incubada a -20°C por 1 hora e, em seguida, centrifugada a $20.000 \times g$ por 30 minutos a 4°C . O sobrenadante foi cuidadosamente descartado e o *pellet* foi lavado com etanol absoluto, centrifugado novamente a $20.000 \times g$ por 5 minutos a 4°C , seco e ressuspenso em 20 μ L de água ultrapura.

O amplicon purificado foi submetido ao sequenciamento pelo método de Sanger (ACTGene Análises Moleculares, Brasil). Para o preparo das amostras, foram utilizados 60 ng do amplicon purificado e 5,0 pmol do Primer Fw descrito na Tabela 7, seguindo as orientações técnicas da empresa para o processamento de produtos de PCR.

As sequências obtidas foram analisadas quanto à presença de *indels* por comparação com a sequência selvagem do gene *Glyma.16G208900*. A análise foi realizada utilizando a ferramenta de bioinformática online ICE (CONANT et al., 2022) (Inference of CRISPR Edits) v3.0 (Synthego/EditCo Bio). A presença de múltiplas sequências sobrepostas no cromatograma a partir do sítio de clivagem esperado e a decomposição dos traços pela ferramenta ICE foram utilizadas para confirmar o sucesso da edição gênica e estimar a frequência e o tipo das mutações induzidas. As plantas editadas foram separadas para completarem o ciclo até a completa maturação das sementes (T_1).

5.5 EXPRESSÃO TEMPORAL DO GENE BBI (*GLYMA.16G208900*) DURANTE A MATURAÇÃO DAS SEMENTES

5.5.1 Amostragem Vegetal e Estratégia de *Pool* Biológico

Para assegurar a representatividade populacional e minimizar a variação individual entre plantas, adotou-se uma estratégia de amostragem composta (*bulk sampling*). Vagens foram colhidas de 20 plantas diferentes da cultivar BRS 537 mantidas em casa de vegetação.

A classificação dos estádios de desenvolvimento (S1 a S9) foi realizada inicialmente por morfometria visual das vagens, seguida pela pesagem individual das sementes frescas para confirmação da faixa de massa (mg) correspondente a cada estágio. Sementes fora do intervalo de peso estabelecido para cada classe foram descartadas.

As sementes selecionadas de cada estágio, variando de aproximadamente 40 unidades para os estádios iniciais, de menor biomassa, a cerca de 20 unidades para os estádios finais, foram agrupadas em um único *pool* biológico por estágio, visando à homogeneização da variabilidade genética e fisiológica intraestádio. Todo o material de cada *pool* foi macerado simultaneamente em nitrogênio líquido até a obtenção de um pó fino homogêneo.

A partir deste *pool* homogeneizado, foram realizadas três extrações de RNA independentes, constituindo três repetições biológicas. Adicionalmente, todas as reações de amplificação por RT-qPCR foram conduzidas em triplicata técnica (três poços por réplica biológica na placa), garantindo a robustez técnica da quantificação.

O estadiamento do desenvolvimento das sementes foi baseado no sistema de nove estádios (S1–S9) adotado pelo projeto de transcriptômica de soja do Joint Genome Institute (JGI), amplamente utilizado em bases públicas como Phytozome v13 e Gene Atlas.

As faixas de massa fresca associadas a cada estágio foram definidas a partir de critérios operacionais empregados na geração dos dados de RNA-seq desse projeto, conforme informação obtida diretamente com membros do consórcio responsável (comunicação pessoal).

Neste estudo, tais critérios foram utilizados como referência para a categorização dos estádios de desenvolvimento, visando à compatibilidade com dados públicos de expressão gênica. As categorias de desenvolvimento definidas para a coleta, correspondendo às respectivas faixas de massa fresca e características visuais, estão detalhadas na Tabela 8.

Tabela 8 – Categorias dos estádios de desenvolvimento da semente de soja, definidas com base na massa fresca e em características visuais.

Estádio	Faixa de Massa Fresca (mg)	Características
S1	< 10	–
S2	27 – 42	–
S3	70 – 95	–
S4	115 – 150	–
S5	200 – 250	–
S6	> 300	Semente Verde
S7	> 300	Semente Amarela
S8	200 – 250	Semente Amarela
S9	< 150	Semente Amarela

Nota: Os estádios S8 e S9 correspondem a fases avançadas de maturação e dessecação e não foram utilizados nas análises de expressão gênica deste estudo, com base nos valores infimos gerados pelo Gene Atlas.

Fonte: Elaborada pelo autor. Informações detalhadas sobre as faixas de massa foram fornecidas por Vikranth K. Chandrasekaran e colaboradores, integrantes do projeto de transcriptômica de soja do JGI, por comunicação pessoal (2022).

5.5.2 Extração de RNA Total

A extração de RNA total foi realizada a partir de aproximadamente 100 mg de tecido de sementes pulverizado em nitrogênio líquido. Devido às alterações drásticas na composição bioquímica da semente ao longo do desenvolvimento, especificamente o acúmulo progressivo de óleos e polissacarídeos (amido) nos estádios finais, adotou-se uma

estratégia diferente de extração para garantir a integridade e pureza do material em todas as fases.

Para os estádios iniciais e intermediários (S1 a S4), caracterizados por menor teor de reservas complexas, utilizou-se o kit Direct-zol™ RNA Miniprep (Zymo Research), seguindo as instruções do fabricante. Este método, baseado em colunas de sílica, proporcionou alta recuperação de RNA com eficiente remoção de contaminantes.

Nos estádios de maturação tardia (S5 a S7), observou-se que a elevada concentração de polissacarídeos (amido) promovia a saturação física das colunas de purificação do kit anterior. Para superar essa limitação técnica, a extração nestas amostras foi conduzida com o reagente Plant RNA Reagent (Invitrogen). Este protocolo baseia-se em um reagente de lise contendo detergentes catiônicos (CTAB) e isotiocianato de guanidina, especificamente formulado para precipitar polissacarídeos e polifenóis antes da precipitação do RNA, evitando a co-precipitação de inibidores de PCR e possíveis contaminações com polissacarídeos.

A concentração e a pureza das amostras de RNA (razões A260/A280 e A260/A230) foram avaliadas em espectrofotômetro NanoDrop™ (Thermo Scientific). A integridade das bandas de rRNA (28S e 18S) foi verificada visualmente por eletroforese em gel de agarose 1,0% (p/v) corado com brometo de etídio.

5.5.3 Síntese de cDNA

Para a síntese de DNA complementar (cDNA), foi utilizada a enzima M-MLV Reverse Transcriptase (M1302, Sigma-Aldrich®) com base nas orientações do fabricante. A reação foi preparada utilizando 1 µg de RNA total como molde e oligo nv(dT)30 como primer. O RNA, o primer oligo nv(dT)30 e dNTPs foram incubados a 65°C por 5 minutos e rapidamente resfriados em gelo. Em seguida, foram adicionados o tampão de reação 10X first strand buffer e a enzima M-MLV. A reação de transcrição reversa foi incubada a 37°C por 50 minutos, seguida pela inativação da enzima a 80°C por 10 minutos. As amostras de cDNA resultantes foram diluídas 1:10 em água ultrapura livre de nucleases e armazenadas a -20°C para uso nas reações de RT-qPCR.

5.5.4 Reação de RT-qPCR e Análise de Expressão

As reações de PCR quantitativo em tempo real (RT-qPCR) foram realizadas para avaliar os níveis de transcritos do gene alvo *BBI* (*Glyma.16G208900*), normalizados por dois genes de referência endógenos. As sequências dos primers utilizados para a amplificação do gene alvo e dos candidatos a controle endógeno estão apresentadas na Tabela 9. Os genes β -conglucina (*Glyma.20G148400.1*) e *GAPDH* (*Glyma.06G172600.1*) foram selecionados

com base na sua validação prévia para análises de expressão em soja (HU et al., 2009; SEVERIN et al., 2010) e por representarem classes funcionais estritamente distintas. O *GAPDH* está envolvido em processos metabólicos constitutivos da via glicolítica, sendo um controle amplamente estável na espécie (HU et al., 2009). Em contrapartida, a β -conglícinina é uma proteína de reserva cujo uso é estabelecido como marcador específico do transcriptoma de sementes (SEVERIN et al., 2010). A adoção de genes pertencentes a vias biológicas independentes segue as diretrizes metodológicas padrão para evitar o viés de co-regulação transcricional no cálculo do fator de normalização (BUSTIN et al., 2009).

Tabela 9 – Sequência dos primers utilizados para a amplificação por RT-qPCR do gene alvo BBI e dos genes de referência endógenos.

Gene	Finalidade	Sentido	Sequência (5' → 3')	Tamanho do amplicon (pb)
<i>BBI</i> (<i>Glyma.16G208900</i>)	Alvo	Fw Rv	TTGACACCAACGACTTCTGC GGCTTCGTCCATTTGAGAGA	83
β -conglícinina (<i>Glyma.20G148400.1</i>)	Referência	Fw Rv	GATGAGGAACAAGATGAACGTCAAT CTTGCTTTCGCTCTTCCTTC	70
<i>GAPDH</i> (<i>Glyma.06G172600.1</i>)	Referência	Fw Rv	GGGTGGTGCAAAGAAGGTTATTA TCGTTGACACCAACAACAACA	72

Fonte: Elaborada pelo autor (2026).

Os ensaios de RT-qPCR foram conduzidos em triplicatas biológicas e técnicas utilizando o equipamento QuantStudio™ 3 Real-Time PCR System (Applied Biosystems™) em placas de 96 poços de 0,2 mL. As reações foram preparadas em um volume final de 10 μ L, e a composição detalhada da mistura reacional por poço está descrita na Tabela 10.

Tabela 10 – Composição da reação para os ensaios de RT-qPCR.

Componente	Volume por reação (μ L)	Concentração Final
GreenMaster® (2X)(Cellco)	5,0	1X
Primer Forward (10 μ M)	0,5	500 nM
Primer Reverse (10 μ M)	0,5	500 nM
cDNA (diluição 1:10)*	2,0	-
Água livre de nucleases	2,0	-
Volume Total	10,0	

Nota: *O cDNA utilizado como molde foi sintetizado a partir de 1000 ng de RNA total.

Fonte: Elaborada pelo autor (2026).

O perfil térmico e os parâmetros da curva de dissociação (*melt curve*) adotados para a amplificação e verificação da especificidade dos amplicons estão sumarizados na

Tabela 11. O ensaio iniciou-se com etapas de pré-aquecimento a 50°C por 2 minutos e ativação da polimerase a 95°C por 10 minutos, seguidas de 40 ciclos de desnaturação a 95°C por 15 segundos e anelamento/extensão a 60°C por 1 minuto. A curva de dissociação foi realizada imediatamente após o término da amplificação, com o objetivo de confirmar a ausência de dímeros de primers e verificar a especificidade da amplificação, consistindo em um patamar de 60°C por 1 minuto, seguido de aquecimento gradual até 95°C (taxa de 0,5°C/s). A temperatura da tampa do termociclador foi mantida a 105°C e não foi utilizada referência passiva.

Tabela 11 – Condições reacionais e perfil térmico adotados para as análises de RT-qPCR.

Perfil Térmico de Amplificação			
Fase	Temperatura	Tempo	Ciclos
Pré-aquecimento	50°C	2 min	1
Ativação da polimerase	95°C	10 min	1
Desnaturação	95°C	15 s	40
Anelamento/Extensão	60°C	1 min	
Curva de Dissociação (<i>Melt Curve</i>)			
Fase	Temperatura	Tempo	Taxa de Aquecimento
Patamar inicial	60°C	1 min	-
Dissociação	60°C → 95°C	Contínuo	0,5°C/s

Fonte: Elaborada pelo autor (2026).

5.5.5 Análise dos dados de RT-qPCR

Os dados de amplificação foram exportados do *software* do equipamento QuantStudio™ 3 Real-Time PCR System (Applied Biosystems™). Para cada reação individual, a eficiência de amplificação (E) e o valor do ciclo de quantificação (Cq) foram estimados utilizando o algoritmo da ferramenta Miner (ZHAO; FERNALD, 2005). Ressalta-se que somente os valores calculados por este método foram utilizados nas análises subsequentes, uma vez que o algoritmo computa esses parâmetros diretamente a partir da modelagem cinética das curvas de fluorescência de cada amostra, não assumindo uma eficiência de amplificação constante (100%) entre as reações.

A análise da expressão relativa foi conduzida com base em um modelo de quantificação generalizado que incorpora as eficiências individuais de amplificação e a normalização por múltiplos genes de referência (HELLEMANS et al., 2007). Inicialmente, calculou-se a média aritmética dos valores de Cq e E das réplicas técnicas correspondentes a cada réplica biológica. Em seguida, as quantidades relativas (RQ) foram calculadas pela fórmula $RQ = (1 + E)^{-Cq}$. Para corrigir possíveis variações na quantidade de RNA inicial e na eficiência da transcrição reversa, calculou-se um Fator de Normalização (NF) para cada amostra, obtido através da média geométrica dos valores de RQ dos genes de referência selecionados (*GAPDH* e *β-conglicina*).

A expressão relativa normalizada (NRQ) do gene alvo BBI (*Glyma.16G208900*) foi determinada dividindo-se o RQ do gene alvo pelo NF da respectiva amostra. Posteriormente, os valores de NRQ foram reescalados pelo método comparativo (SCHMITTGEN; LIVAK, 2008), adotando-se o estágio S1 como amostra calibradora (por apresentar a menor expressão média), ao qual foi atribuído o valor relativo de 1,0.

Para atender à normalidade e homogeneidade de variâncias requeridas pelos testes estatísticos paramétricos, os dados de NRQ reescalados foram submetidos à transformação logarítmica de base 2 ($\log_2$). A análise estatística para identificar diferenças significativas na expressão do gene BBI ao longo dos estágios de desenvolvimento (S1 a S7) foi realizada no ambiente de programação computacional R (versão 4.5.3). As diferenças globais foram avaliadas por meio da Análise de Variância (ANOVA) de via única. Posteriormente, as médias foram comparadas utilizando o teste *post-hoc* de Tukey ($p < 0,05$) com o auxílio do pacote estatístico *agricolae*, permitindo o agrupamento estatístico das médias em classes de significância.

6 RESULTADOS E DISCUSSÃO

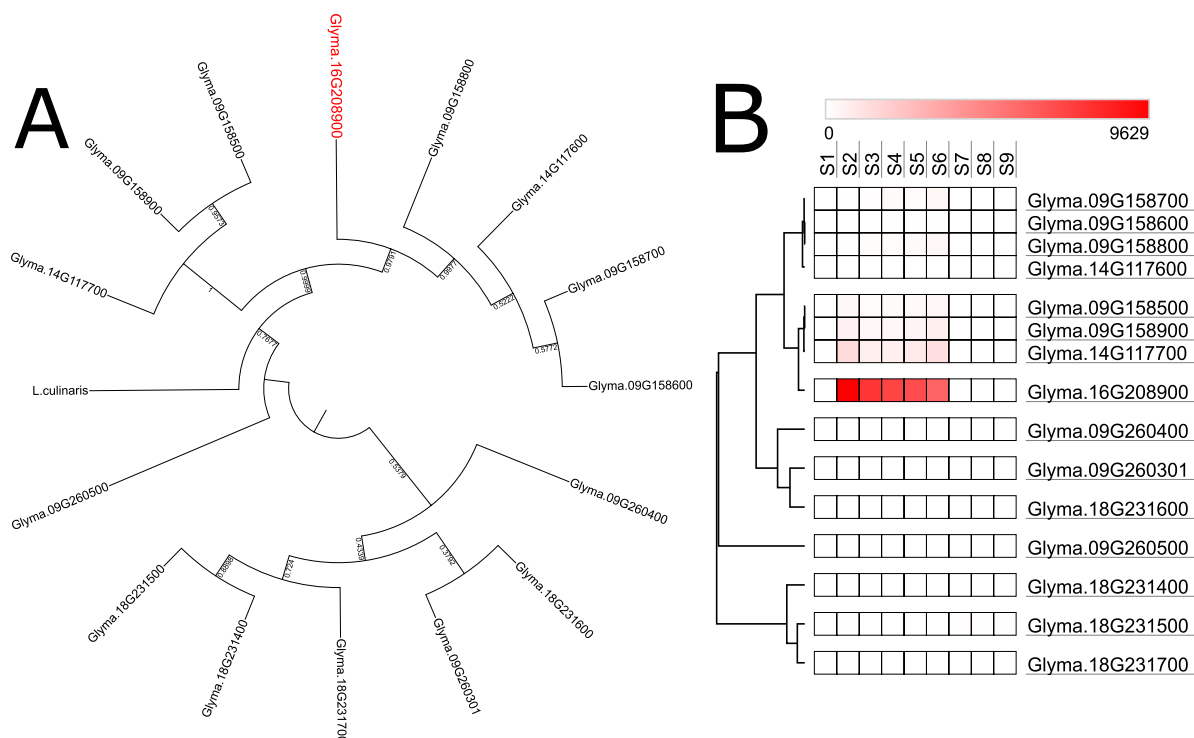
6.1 SELEÇÃO DO ALVO E PERFIL DE EXPRESSÃO *IN SILICO*

A prospecção genômica do gene alvo foi realizada por meio de uma análise comparativa de perfis transcricionais, utilizando dados de RNA-Seq provenientes do Gene Atlas v.2 (SREEDASYAM et al., 2023). A investigação conduzida neste estudo permitiu a identificação de 15 *loci* codificantes para inibidores do tipo Bowman-Birk em *G. max* (L.) Merrill. A partir desse conjunto, foram integradas análises filogenéticas e de expressão gênica nos diferentes estádios de desenvolvimento da semente, visando selecionar o *locus* de maior relevância funcional para a estratégia de edição. De maneira secundária, foi feita a análise de níveis de transcritos dos genes da família BBI, separado pelos tecidos e condições fisiológicas fornecidos pelo PlantGeneAtlas, constando no suplementar.

Conforme ilustrado na Figura 8, a análise revelou forte heterogeneidade nos níveis de transcritos entre os membros da família. Apesar da presença de múltiplos *loci* codificantes no genoma da soja, a contribuição transcricional durante o desenvolvimento da semente é fortemente concentrada em poucos membros. Destaca-se o gene *Glyma.16G208900*, que apresentou níveis de expressão significativamente superiores aos demais, com um pico consistente (atingindo valores superiores a 9.600 FPKM) nos estádios iniciais a intermediários associados ao enchimento do grão.

Esse padrão de dominância indica que este *locus* responde pela maior parcela da atividade transcricional da família Bowman-Birk nesse período crítico, justificando sua escolha como alvo prioritário para a confirmação experimental do perfil de expressão e para as estratégias de edição gênica via CRISPR/Cas9 propostas neste estudo.

Figura 8 – Análise filogenética e perfil de expressão transcricional da família de genes do Inibidor Bowman-Birk (BBI) em *Glycine max* (L.) Merrill.



Legenda: (A) Árvore filogenética ilustrando as relações de ortologia e as distâncias evolutivas entre os 15 *loci* da família BBI identificados na soja. O inibidor de *Lens culinaris* (*L.culinaris*) foi utilizado como grupo externo para o enraizamento da árvore. Os valores nos nós indicam a confiabilidade dos ramos (teste de *bootstrap* com 10.000 réplicas). **(B)** *Heatmap* com agrupamento hierárquico representando os níveis de transcritos ao longo de nove estádios de desenvolvimento da semente (S1 a S9). A escala de cores (variando do branco ao **vermelho**) indica a intensidade de expressão, atingindo um máximo de 9629 (FPKM). O gene alvo selecionado para edição por CRISPR/Cas9, *Glyma.16G208900*, encontra-se destacado em **vermelho** na árvore filogenética, e com os maiores valores de FPKM, evidenciando seu expressivo pico de atividade transcricional nos estádios iniciais a intermediários do enchimento do grão.

Fonte: Elaborada pelo autor (2026), com dados de expressão do *Gene Atlas v.2*.

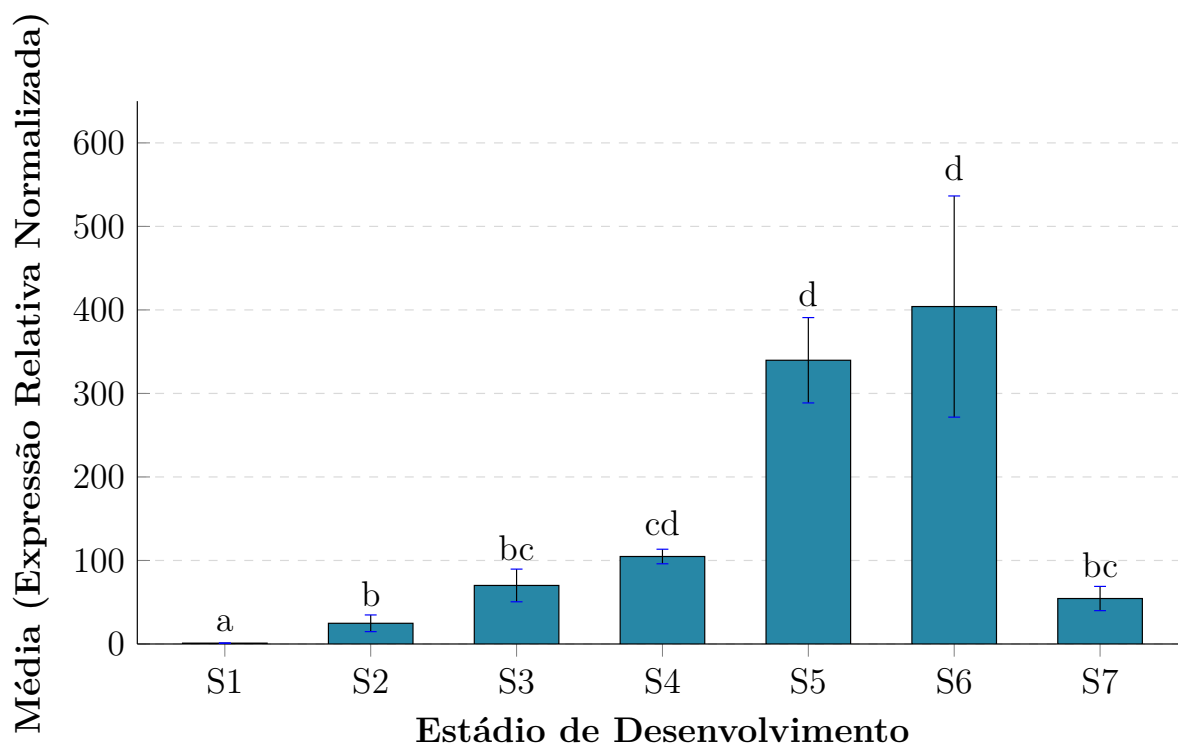
6.1.1 Validação experimental do perfil de expressão na cultivar de soja BRS 537

Embora os dados do genoma de referência (*G. max* (L.) Merrill) cv. Williams 82 demonstrem a predominância transcricional do gene *Glyma.16G208900*, a validação experimental desse perfil é necessária, uma vez que variações genotípicas e ambientais podem influenciar a regulação da expressão gênica. Assim, para confirmar o padrão de expressão na cultivar BRS 537, os níveis de mRNA foram quantificados por RT-qPCR ao longo dos estádios de desenvolvimento da semente selecionados (S1-S7).

Conforme apresentado na Figura 9, a quantificação revelou um padrão dinâmico de transcrição do alvo. Os níveis de expressão relativa foram normalizados pelos genes de referência *GAPDH* e β -conglucina. O estágio inicial de desenvolvimento (S1) foi adotado

como amostra calibradora (valor relativo igual a $1,00 \pm 0,20$) por representar o momento fisiológico de menor atividade transcricional deste gene, servindo como linha de base basal para a embriogênese inicial.

Figura 9 – Quantificação relativa dos transcritos de *Glyma.16G208900* por RT-qPCR durante o desenvolvimento da semente da cultivar BRS 537.



Legenda: Os níveis de expressão foram quantificados nos sete estádios (S1 a S7), representando a progressão temporal da embriogênese até a maturação/dessecação das sementes de soja. Os dados brutos (Cq e Eficiência) foram extraídos via algoritmo Miner e normalizados pela média geométrica da expressão dos genes endógenos de referência *GAPDH* e *β -conglícina*. A expressão relativa foi escalonada adotando o estágio S1 como amostra calibradora (valor = 1). As barras indicam a média da expressão relativa de três repetições biológicas, cada uma analisada em triplicata técnica, e as linhas verticais representam o erro padrão da média (EPM). Letras distintas acima das barras indicam diferença estatística significativa pelo teste *post-hoc* de Tukey ($p < 0,05$). Observa-se uma regulação fortemente positiva (*up-regulation*) atingindo o pico nos estádios S5 e S6, seguida de um declínio acentuado no estágio final (S7).

Fonte: Elaborada pelo autor (2026).

Em relação a essa linha de base, observou-se um nível transcricional basal no estágio S1 ($1,00 \pm 0,20$), seguido por um salto expressivo e estatisticamente significativo ($p < 0,001$) já no estágio S2 ($24,78 \pm 9,99$). Esse acúmulo contínuo culminou em níveis massivos durante a fase crítica de enchimento de grão e transição para a maturação. O patamar de expressão máxima foi detectado de forma ampla nos estádios S5 ($339,73 \pm 51,09$) e S6 ($404,04 \pm 132,42$), não havendo diferença estatística entre estes dois estádios, caracterizando uma janela de intensa atividade transcricional. A análise confirmou que este bloco de

expressão máxima (S5-S6) é significativamente superior aos níveis observados no início do desenvolvimento ($p < 0,001$).

A partir do estágio S6, nota-se uma virada abrupta na biologia da semente. Uma redução drástica na abundância de transcritos ocorre na transição para o estágio S7 ($54,43 \pm 14,51$), configurando um declínio significativo ($p < 0,01$ em relação a S6) em direção aos níveis de dormência.

Curiosamente, a topologia da curva de expressão obtida experimentalmente revela uma divergência temporal marcante em relação aos perfis transcricionais genéricos previstos pela plataforma *in silico*, *Gene Atlas*. Enquanto os dados de RNA-seq de referência (ancorados na cultivar temperada Williams 82) sugerem um pico de expressão de *Glyma.16G208900* nas fases iniciais de desenvolvimento embrionário (a partir da fase S2), a quantificação por RT-qPCR na cultivar BRS 537 demonstrou um deslocamento do pico de expressão máxima para os estádios tardios (S5 e S6), com uma indução superior a 400 vezes.

Essa divergência evidencia a necessidade da validação experimental e pode ser explicada por fatores genotípicos, metodológicos e fisiológicos. Primeiramente, a cultivar BRS 537 possui adaptações ao ambiente tropical que podem alterar a ontogenia e a cronologia da regulação gênica em relação à cultivar de referência. Em segundo lugar, metodologicamente, abordagens de transcriptômica global (*RNA-seq*) expressam valores de abundância relativa que podem sofrer efeito de diluição devido à massiva síntese simultânea de diversos transcritos de reserva durante o enchimento de grãos, enquanto a abordagem de RT-qPCR com múltiplas normalizações e correção de eficiência oferece uma resolução direcionada e altamente sensível.

Do ponto de vista fisiológico e evolutivo, a dinâmica observada via RT-qPCR alinha-se de forma mais consistente com a biologia da planta. Nos estádios iniciais (S1 e S2), a semente encontra-se predominantemente na fase de divisão e expansão celular (JONES; VODKIN, 2013), justificando a atividade moderada de genes ligados à reserva metabólica. Em contraste, a marcante e robusta regulação positiva observada nos estádios posteriores (S4 a S6) coincide perfeitamente com a fase de enchimento de grãos, etapa na qual a semente desloca sua maquinaria para a síntese massiva de proteínas de reserva e defesa.

Fisiologicamente, a superexpressão do gene BBI nesta janela temporal tardia atende a uma demanda de proteção agrônômica e de armazenamento. À medida que o tecido cotiledonar se torna altamente nutritivo, ele eleva o seu risco à herbivoria. Assim, a planta parece responder com a maximização da transcrição de inibidores de protease como

o BBI (*Glyma.16G208900*), sugerindo uma estratégia evolutiva para sincronizar a defesa bioquímica com o acúmulo de nutrientes (CLEMENTE et al., 2019; QI et al., 2005). Em consequência, com a chegada do estágio S7, o declínio nos níveis de mRNA reflete a esperada mudança metabólica, onde a transcrição é suprimida para que a semente sofra dessecação e entre em dormência fisiológica (COLLAKOVA et al., 2013). Em suma, a magnitude de transcrição no momento exato do enchimento de grãos sustenta a adequação biológica desta isoforma como o alvo prioritário para a eliminação do seu traço antinutricional via edição genômica.

6.2 VALIDAÇÃO ESTRUTURAL DO ALVO *GLYMA.16G208900* E PRE-DIÇÃO DE FUNCIONALIDADE

6.2.1 Caracterização Estrutural da Proteína do Gene *Glyma.16G208900* Madura

A validação funcional de genes candidatos via edição gênica requer um conhecimento profundo da estrutura tridimensional da proteína alvo. Uma vez que não há estrutura cristalográfica experimental depositada para o produto do gene *Glyma.16G208900* no Protein Data Bank (PDB), esta etapa do projeto foi dedicada à construção de um modelo molecular que permitisse prever o comportamento do inibidor e validar sua capacidade de interação com proteases.

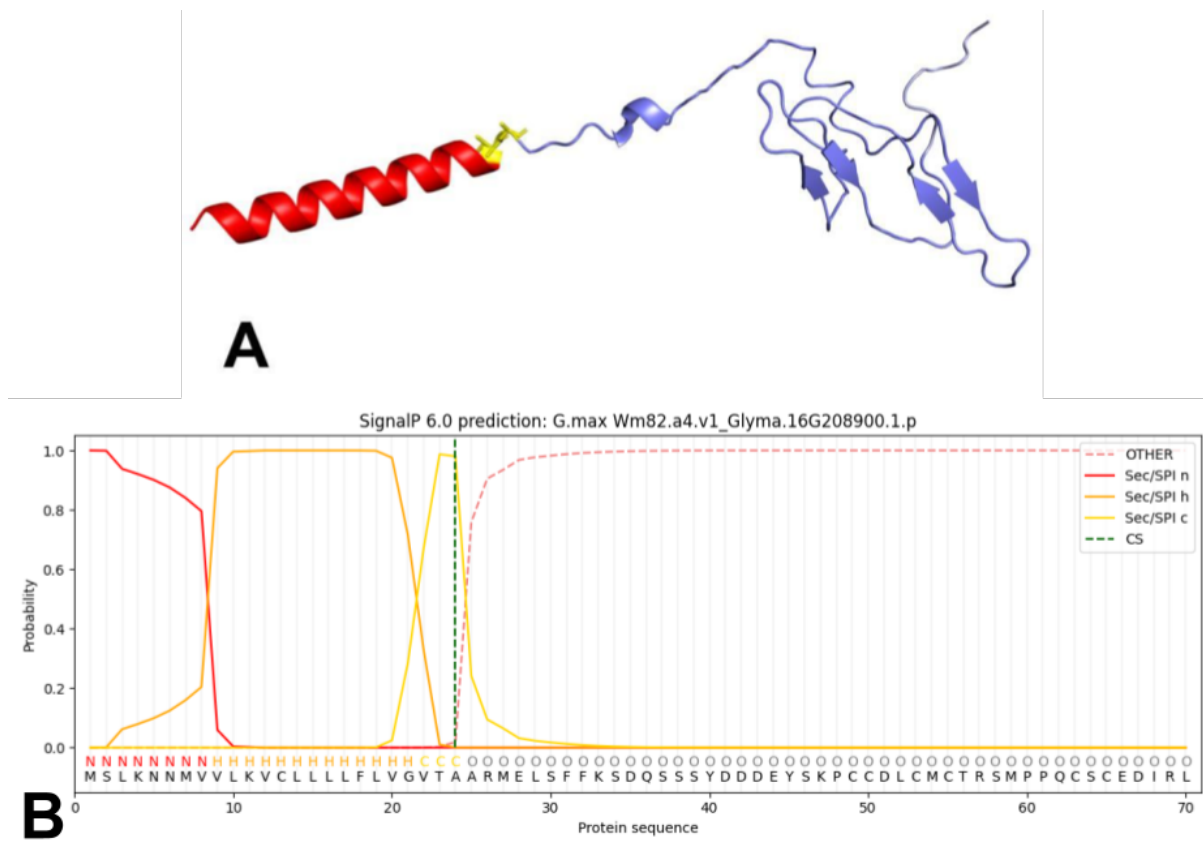
A modelagem inicial da sequência completa de aminoácidos revelou uma arquitetura globular compacta, típica da família BBI (QI et al., 2005). Essa conformação estrutural é caracterizada por um arranjo altamente reticulado e estabilizado por pontes de dissulfeto, o qual confere à proteína sua notável estabilidade térmica e extrema resistência à degradação proteolítica, características estruturais que são fundamentais para a manutenção da sua atividade como inibidor de proteases no ambiente celular.

Contudo, acoplada a esse núcleo estável, observou-se uma extensão hélice N-terminal projetada com índices de confiança muito baixos (pLDDT < 50), sugerindo tratar-se de uma região intrinsecamente flexível ou transitória, conforme ilustrado na Figura 10A.

Para investigar e validar a natureza desta região, a sequência primária foi submetida ao servidor *SignalP* 6.0, que indicou, com probabilidade superior a 99%, a presença de um peptídeo sinal do tipo Sec/SPI (Figura 10B). O sítio de clivagem foi identificado entre os resíduos 24 e 25 (Ala24-Ala25), delimitando a transição exata para a sequência funcional da proteína madura. Com a confirmação desta sequência sinalizadora de endereçamento, procedeu-se à obtenção do modelo refinado da forma processada para as análises funcionais

e de ancoragem (*docking*) subsequentes.

Figura 10 – Modelagem estrutural e predição de peptídeo sinal da proteína BBI codificada pelo gene *Glyma.16G208900*.



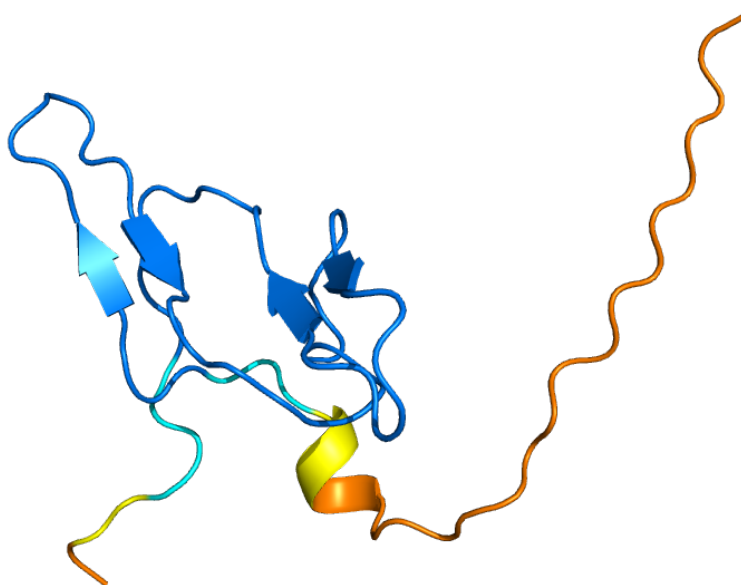
Legenda: (A) Modelo 3D inicial gerado pelo algoritmo AlphaFold2. O gradiente de cores reflete o índice de confiança da predição (pLDDT) estimado por resíduo: a região em azul indica confiança alta a muito alta (pLDDT > 70), correspondendo ao núcleo globular estruturado e maduro da proteína; a região em amarelo representa confiança baixa ($50 < \text{pLDDT} < 70$); e a região em vermelho denota confiança muito baixa (pLDDT < 50), característica de regiões intrinsecamente desordenadas, evidenciando o peptídeo sinal flexível na extremidade N-terminal. (B) Gráfico de probabilidade gerado pelo *SignalP* 6.0, indicando a arquitetura do peptídeo sinal tipo *Sec/SPI*: a região *n* (linha sólida vermelha), a região *h* (linha sólida laranja) e a região *c* (linha sólida amarela). A linha vertical verde tracejada (CS) marca o sítio de clivagem predito entre os resíduos Ala24 e Ala25, a partir do qual a probabilidade de proteína madura (linha tracejada vermelha) atinge o patamar máximo.

Fonte: Elaborada pelo autor (2026).

6.2.2 Obtenção do Modelo Estrutural da Proteína Madura e Mapeamento dos Sítios Ativos

A sequência da proteína, após a remoção do peptídeo sinal predito, foi submetida a uma nova rodada de modelagem estrutural. Conforme ilustrado na Figura 11, foram observadas regiões de desordem estrutural nas extremidades da cadeia polipeptídica. Essa constatação motivou a remoção *in silico* dos segmentos não presentes nos BBIs canônicos mencionados na Seção 5.1.3, a fim de obter a representação exata de sua forma madura.

Figura 11 – Predição estrutural do inibidor Bowman-Birk após a remoção do peptídeo sinal.



Legenda: Modelo tridimensional predito pelo servidor AlphaFold2. A coloração do esqueleto polipeptídico reflete o índice de confiança da predição estimado por resíduo (pLDDT). O núcleo estruturado, que abriga os sítios ativos, apresenta confiança alta a muito alta (azul e ciano). Em contrapartida, as extensões projetadas nas extremidades N-terminal e C-terminal exibem níveis de confiança baixos a muito baixos (amarelo e laranja), evidenciando que se tratam de regiões intrinsecamente flexíveis e desordenadas, que precedem a forma madura e compacta do inibidor.

Fonte: Elaborada pelo autor (2026).

Esta etapa de re-modelagem e corte guiado pelas sequências ativas é crucial, ao permitir que o algoritmo recalcule o enovelamento na ausência das regiões N-terminal e C-terminal desordenadas. A exclusão dessas regiões intrinsecamente flexíveis evita interferências termodinâmicas e espaciais artificiais durante o cálculo da predição, resultando em uma geometria global mais precisa e energeticamente favorável (RUFF; PAPPU, 2021).

A predição estrutural de proteínas de baixa massa molecular e ricas em pontes dissulfeto representa um desafio conhecido. Essas proteínas frequentemente carecem de um núcleo hidrofóbico volumoso que guie o enovelamento termodinâmico, dependendo fortemente do pareamento correto das cisteínas (MCDONALD et al., 2023). Diante desse cenário, a confiabilidade do modelo gerado pelo AlphaFold v2 foi avaliada por meio de superposição estrutural com o inibidor Bowman-Birk de referência de *G. max* (L.) Merrill (PDB ID: 1K9B).

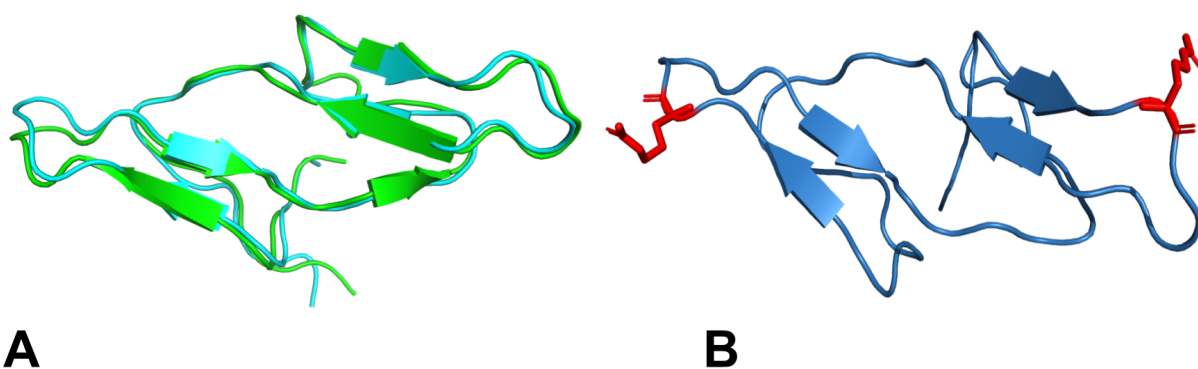
A avaliação topológica global no servidor *TM-align* (ZHANG; SKOLNICK, 2005) resultou em um TM-score (*Template Modeling score*) de 0,885. Como valores estritamente superiores a 0,5 indicam que as estruturas compartilham o mesmo enovelamento, este resultado indica a identidade topológica do modelo predito. Complementarmente, o ali-

nhamento espacial refinado revelou um valor de RMSD (*Root Mean Square Deviation*) de 0,549 Å, calculado sobre as coordenadas de 319 átomos de carbono alfa ($C\alpha$) constituintes do esqueleto polipeptídico (Figura 12A). Este valor em escala sub-angstrom, aliado ao elevado TM-score, evidencia uma coincidência estrutural de altíssima precisão entre os *backbones*. Essas métricas indicam a acurácia do algoritmo na predição conformacional, sugerindo um arranjo adequado das pontes dissulfeto em relação às estruturas depositadas, fator essencial para a estabilidade do enovelamento.

O modelo foi analisado no *software PyMOL v2.5* (SCHRÖDINGER, LLC, 2021), exibindo a conformação canônica de “cabeça dupla” (*double-headed*), típica da família Bowman-Birk (VOSS et al., 1996). A superposição com a estrutura de referência indicou a preservação de dois laços reativos funcionalmente independentes, fundamentais para a atividade inibitória (QI et al., 2005). O Sítio Reativo 1 (laço N-terminal) é caracterizado pela exposição solvente-acessível do resíduo Arg16. O Sítio Reativo 2 (laço C-terminal) apresenta o resíduo Arg43 em conformação análoga, sugerindo a capacidade de interação independente com uma segunda enzima proteolítica conforme demonstrado na Figura 12B.

A inspeção da geometria local desses resíduos chave indicou que ambos se encontram adequadamente posicionados no contexto estrutural, apresentando compatibilidade geométrica com a formação de complexos proteína-proteína, corroborando a utilização do modelo nas simulações subsequentes de *docking* molecular.

Figura 12 – Validação estrutural e funcional do inibidor Bowman-Birk modelado.



Legenda: (A) Superposição estrutural entre o modelo tridimensional predito pelo AlphaFold v2 para a proteína madura (em ciano) e a estrutura cristalográfica experimental de referência do inibidor Bowman-Birk de (*G. max* (L.) Merrill) (PDB ID: 1K9B, em verde). A elevada coincidência dos esqueletos polipeptídicos, evidenciada pelo baixo valor de RMSD (0,549 Å, calculado sobre 319 átomos alinhados), indica a acurácia do enovelamento global e da predição da rede de pontes dissulfeto intracadeia. (B) Detalhe dos dois laços reativos funcionalmente independentes do modelo, destacando a exposição dos resíduos Arg16 e Arg43 em vermelho, configurando a arquitetura estrutural competente para inibição simultânea.

Fonte: Elaborada pelo autor (2026).

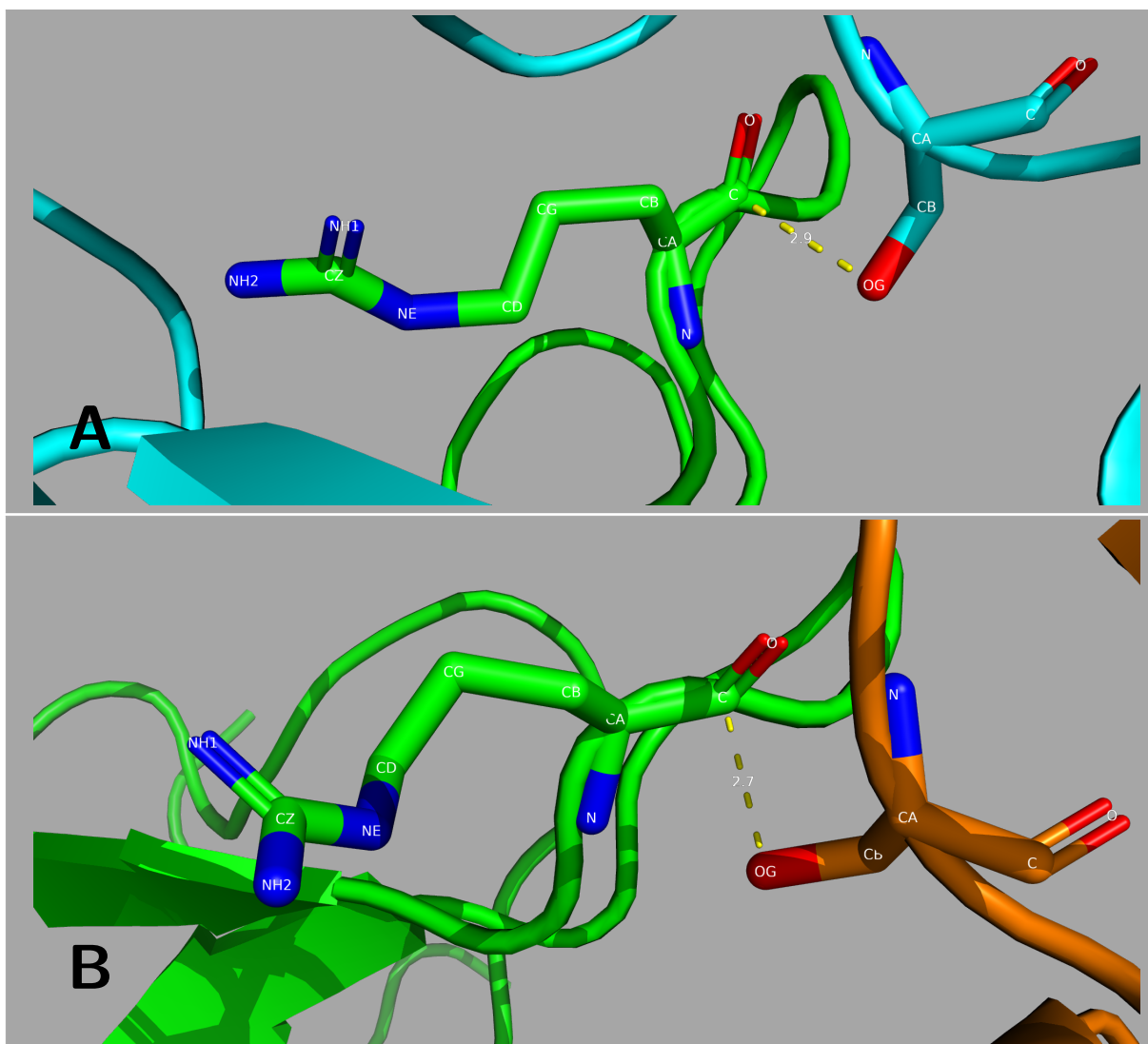
6.2.3 Inferência Funcional Baseada em Modelagem Molecular: Análise Comparativa da Interação Enzima-Inibidor

A competência funcional do modelo *Glyma.16G208900* foi verificada via *docking* molecular frente às proteases Tripsina Bovina e α -Quimotripsina. As análises sugerem que o inibidor possui a plasticidade conformacional necessária para interagir eficientemente com ambas as enzimas, comportando-se como um inibidor “cabeça-dupla” (*double-headed*) clássico. Esse mecanismo de ação é associado ao inibidor Bowman-Birk, cuja arquitetura bifuncional, amplamente validada por cristalografia, permite a neutralização simultânea e independente de duas moléculas de serino-proteases por meio de laços reativos localizados em domínios estruturalmente opostos (VOSS et al., 1996; QI et al., 2005).

No complexo binário com a Tripsina, obteve-se um *Docking Score* altamente favorável de $-355,62$. A confiabilidade estatística do modelo foi atestada pelo elevado índice de confiança (*Confidence Score* de $0,9839$) e por um valor de *Ligand RMSD* (*Root Mean Square Deviation*) residual de $0,23 \text{ \AA}$. O baixo valor de RMSD indica que o laço reativo do inibidor sofreu mínima deformação estrutural para se ancorar na enzima, validando a rigidez predita. A análise estrutural (Figura 13A) confirmou a inserção canônica do resíduo Arg16 na fenda catalítica, com uma distância de $2,9 \text{ \AA}$ entre o seu carbono carbonílico (CO) e o oxigênio gama (OG) da Ser195 catalítica da enzima. Esta proximidade espacial e compatibilidade geométrica são características fundamentais do mecanismo padrão de inibição canônica. O laço reativo do inibidor mimetiza a conformação de um substrato verdadeiro no complexo, posicionando o carbono suscetível de forma termodinamicamente favorável ao ataque nucleofílico inicial; no entanto, a alta rigidez estrutural do inibidor previne o avanço do ciclo catalítico e a subsequente clivagem (KROWARSCH et al., 2003).

Para a α -Quimotripsina, o complexo apresentou um *Score* energético de $-357,95$ e um índice de confiança estatística de $0,9846$, indicando uma interação também altamente favorável e estável (*Ligand RMSD* de $0,37 \text{ \AA}$). A ancoragem foi mediada pelo laço reativo C-terminal do inibidor, destacando-se a inserção do resíduo Arg43 na fenda catalítica, com distância de interação de $2,7 \text{ \AA}$ em relação à Ser195 catalítica da enzima (Figura 13B). Embora a α -Quimotripsina apresente preferência clássica por resíduos hidrofóbicos volumosos no seu sítio principal de clivagem, a formação estável deste complexo com uma arginina corrobora o mecanismo padrão de inibição da família BBI (LASKOWSKI; KATO, 1980; KROWARSCH et al., 2003). Segundo este modelo, a extrema rigidez estrutural do laço reativo, mantida pela extensa rede de pontes dissulfeto, aliada às extensas interações secundárias na interface da proteína, são capazes de ancorar a enzima e estabilizar o complexo de Michaelis, compensando energeticamente a presença de um resíduo subótimo na fenda de clivagem.

Figura 13 – Análise de *docking* molecular do inibidor BBI em complexos binários.



Legenda: (A) Detalhe da interação entre o resíduo reativo Arg16 (**verde**) do inibidor e a Ser195 catalítica da Tripsina Bovina (**ciano**), destacando a distância de 2,9 Å entre os átomos (CO e OG). (B) Interação entre o resíduo reativo Arg43 (**verde**) do inibidor e a Ser195 catalítica da α -Quimotripsina (**laranja**), com distância de 2,7 Å. As representações em varetas seguem a coloração clássica de átomos (Oxigênio em vermelho, Nitrogênio em azul, Carbono na cor da proteína). Alguns resíduos foram suprimidos para melhor visualização das interfaces de ligação. **Fonte:** Elaborada pelo autor (2026).

Conforme descrito na literatura, a especificidade de ligação nestes inibidores é determinada majoritariamente pela conformação rígida do *loop* reativo estabilizado pela rede de pontes dissulfeto e pela superfície de contato com a protease, prevalecendo sobre a simples complementaridade química entre o resíduo reativo do inibidor e a fenda de ligação da enzima (LASKOWSKI; KATO, 1980; BODE; HUBER, 1992; OTLEWSKI et al., 1999). Dessa forma, a inibição ocorre via oclusão estérica do sítio catalítico e estabilização geométrica do complexo, fenômeno que permite superar a seletividade primária da enzima (LASKOWSKI; QASIM, 2000; KROWARSCHE et al., 2003).

A bivalência simultânea foi ratificada pela modelagem do complexo ternário (Tripsina–Inibidor–Quimotripsina), que alcançou um elevado índice de confiança (*Confidence Score* de 0,8089), indicando a probabilidade estrutural do modelo, e um *Score* energético favorável de $-222,13$. Para esta etapa, adotou-se o complexo binário Tripsina-BBI como receptor estático, enquanto a α -Quimotripsina atuou como ligante móvel. Durante esta simulação de *docking* sequencial, observou-se um valor de *Ligand RMSD* de $129,28 \text{ \AA}$. Este parâmetro é esperado e reflete o amplo deslocamento translacional global da α -Quimotripsina desde o seu ponto de partida aleatório até o seu ancoramento estável no sítio reativo C-terminal livre do complexo receptor (ANDRUSIER et al., 2008; MÉNDEZ et al., 2003).

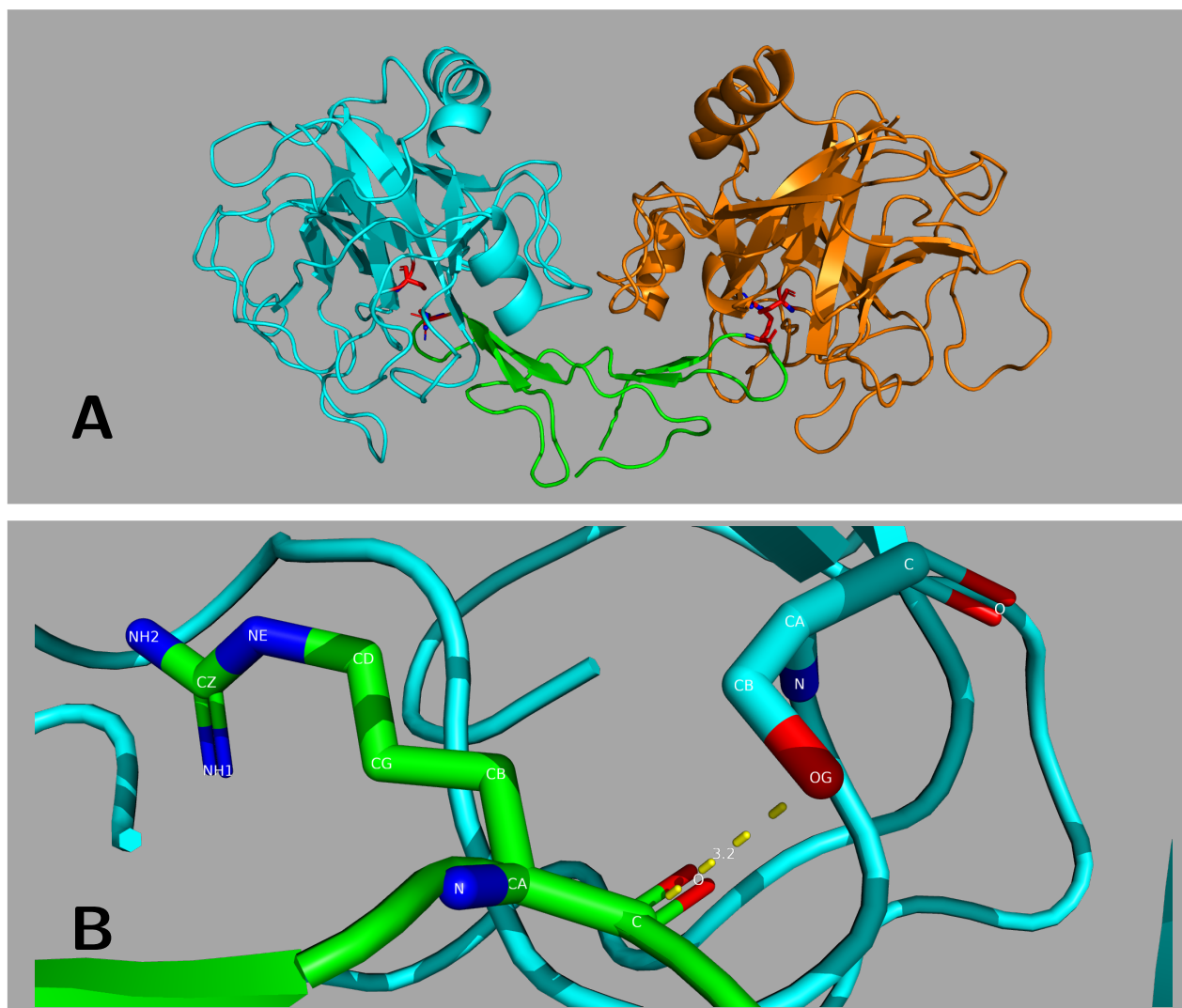
A estrutura tridimensional final (Figura 14A) demonstra a manutenção da competência de ligação em ambos os laços reativos simultaneamente. As distâncias interatômicas nas fendas catalíticas do complexo ternário ($3,2 \text{ \AA}$ para a Tripsina e $3,1 \text{ \AA}$ para a Quimotripsina) sofreram apenas variações marginais em relação aos complexos binários isolados. Além disso, a análise visual do complexo revelou a ausência de impedimentos estéricos severos entre as massas globais da Tripsina e da Quimotripsina, indicando que a arquitetura dicotômica do BBI fornece o espaçamento físico e angular adequado para a acomodação simultânea de ambas as serino-proteases sem repulsão estrutural. Isso corrobora a estabilidade da arquitetura dicotômica do inibidor mesmo sob saturação total dos seus sítios ativos.

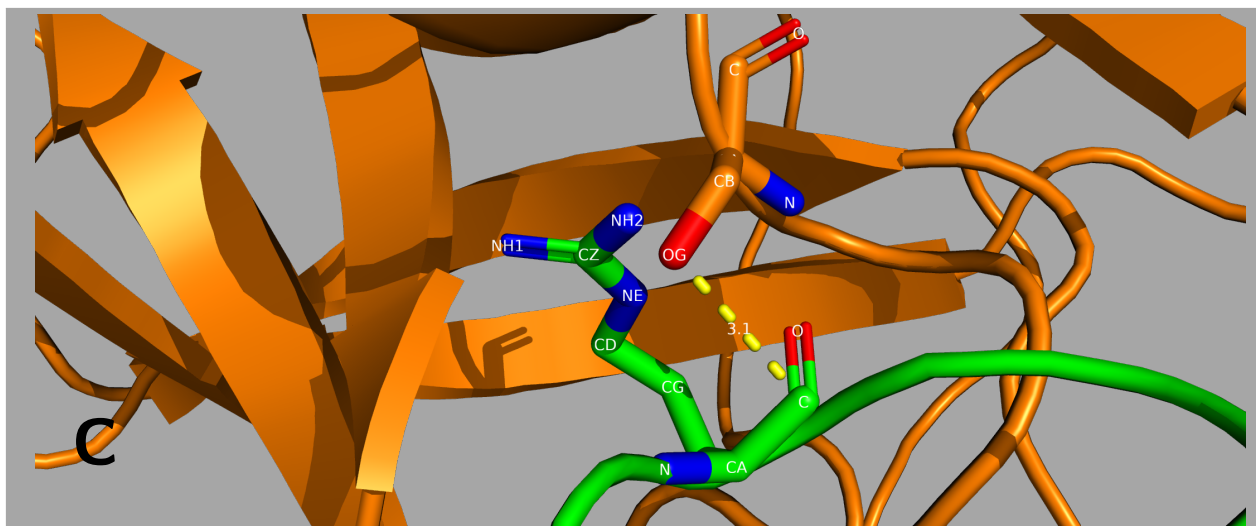
Este resultado possui implicações estruturais e biológicas fundamentais, pois a manutenção das distâncias de interação e a geometria conservada das interfaces indicam a ausência de impedimentos estéricos significativos quando ambas as proteases estão ancoradas. Isso sugere que a molécula do inibidor é estruturalmente robusta e preserva sua conformação global, essencial para a manutenção da função bivalente, uma característica clássica pela independência espacial dos domínios N- e C-terminais da família Bowman-Birk (BIRK, 1985; VOSS et al., 1996).

Em conjunto, as análises de modelagem molecular e *docking* indicam que o BBI majoritário nas sementes apresenta todas as características estruturais e funcionais típicas dos inibidores Bowman-Birk canônicos. O modelo revelou elevada conservação do enovelamento global, estabilizado por uma densa rede de pontes dissulfeto e pela arquitetura de “cabeça dupla”, abrigando dois laços reativos funcionais e independentes. As simulações demonstraram que esses sítios são geometricamente e energeticamente competentes para promover a oclusão estérica e inibição simultânea da tripsina e quimotripsina, conforme os parâmetros termodinâmicos e estruturais sumarizados na Tabela 12. Dessa forma, os resultados indicam que o *locus Glyma.16G208900* codifica uma isoforma íntegra e altamente responsiva do BBI, sendo um determinante primário da atividade antinutricional

associada à inibição de proteases digestivas em sementes de soja. Consequentemente, o *knockout* direcionado deste gene via sistema CRISPR/Cas9 consolida-se como a estratégia promissora para a supressão desta atividade inibitória, visando a obtenção de linhagens *transgene-free* com alta qualidade proteica, sem comprometer a integridade das demais proteínas de reserva do grão.

Figura 14 – Modelagem tridimensional do complexo ternário evidenciando a bivalência simultânea do inibidor BBI (*Glyma.16G208900*.)





Legenda: (A) Visão geral da arquitetura do inibidor *Glyma.16G208900* (verde) mediando a ancoragem simultânea e independente da Tripsina Bovina (ciano, à esquerda) e da α -Quimotripsina (laranja, à direita) em polos estruturalmente opostos. Os resíduos-chave das interfaces de interação (Arg16 do inibidor pareada à Ser195 da Tripsina, e Arg43 do inibidor pareada à Ser195 da Quimotripsina) estão destacados em vermelho. (B) Detalhe estrutural da interface N-terminal, evidenciando a inserção da Arg16 do inibidor na fenda da Tripsina, com distância de interação de 3,2 Å em relação à Ser195 catalítica. (C) Detalhe estrutural da interface C-terminal, apresentando a Arg43 ancorada na fenda da Quimotripsina, com distância de 3,1 Å para a respectiva Ser195 catalítica. Modelos gerados via *docking* sequencial (HDOCK) e analisados no *software* PyMOL.

Fonte: Elaborada pelo autor (2026).

Tabela 12 – Síntese dos parâmetros termodinâmicos e estruturais obtidos nas simulações de *docking* molecular do BBI (*Glyma.16G208900*).

Complexo Simulado	Score (HDOCK)	Confiança	RMSD (Å)	Distância (Å)
Binário: Tripsina–BBI	−355,62	0,9839	0,23	2,9
Binário: Quimotripsina–BBI	−357,95	0,9846	0,37	2,7
Ternário: Trip–BBI–Quimo	−222,13	0,8089	129,28*	3,2 (Trip) 3,1 (Quimo)

Nota: *O elevado valor de RMSD no complexo ternário corresponde ao amplo deslocamento translacional do ligante móvel no espaço tridimensional durante a ancoragem sequencial, não representando deformação estrutural. Fonte: Elaborada pelo autor (2026).

6.2.4 Análise de Redundância e Arquitetura Regulatória da Família BBI

Embora a validação estrutural sugira a competência do gene *Glyma.16G208900* como inibidor de proteases, a eficácia de uma estratégia de *knockout* gênico em soja visando a melhoria da digestibilidade do grão depende fortemente da arquitetura genômica da espécie. Sendo um organismo paleopoliploide, a soja frequentemente retém múltiplas cópias funcionais de genes (parálogos) oriundos de eventos de duplicação completa do genoma

(WGD), o que poderia mascarar o fenótipo de perda de função via compensação genética transcricional (SCHMUTZ et al., 2010).

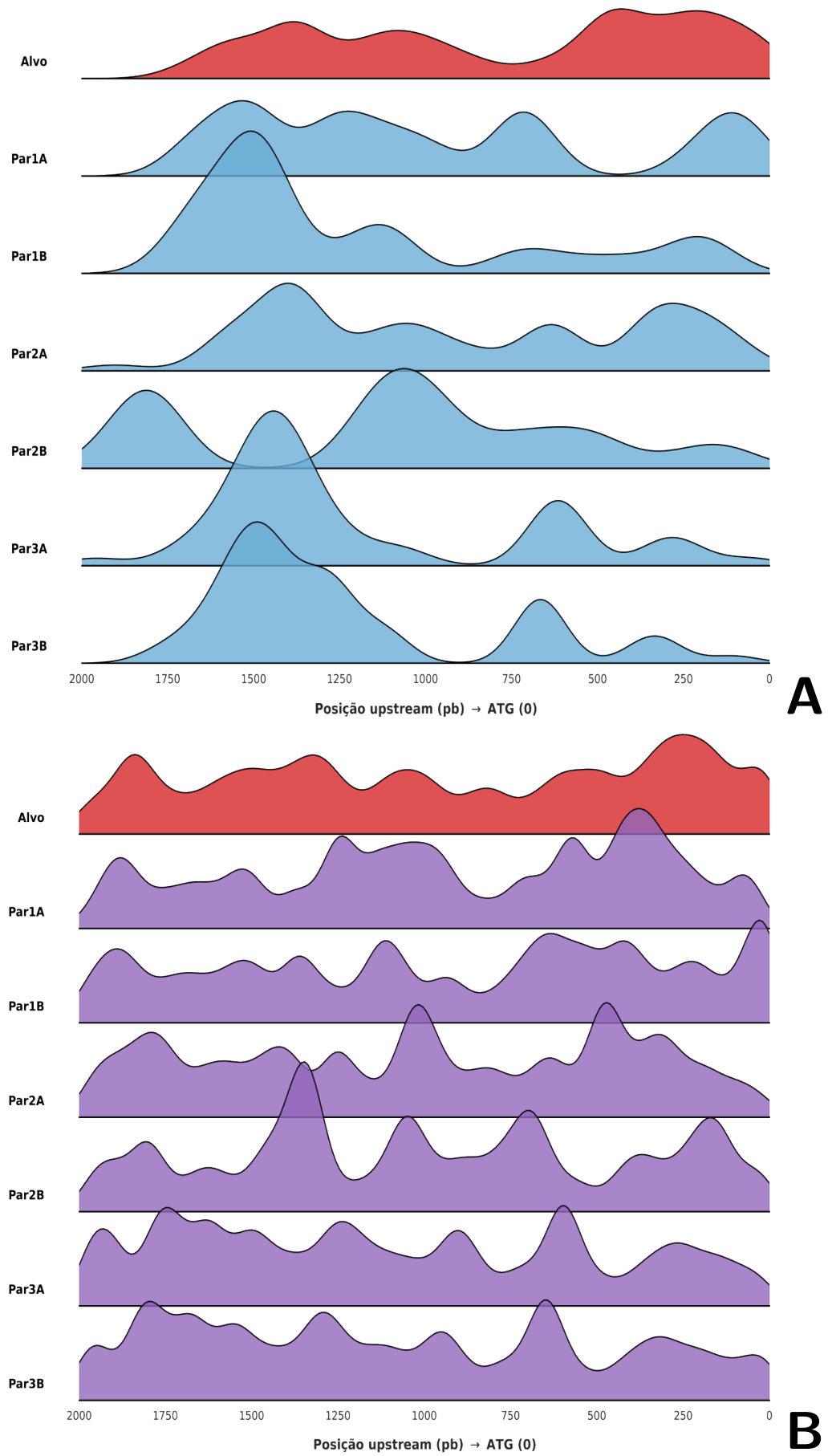
A análise de homologia e identidade da sequência codificadora (CDS) permitiu agrupar os genes da família BBI em módulos de conservação em relação ao gene alvo (*Glyma.16G208900*). Foram identificados três pares de parálogos principais: o **Módulo 1**, composto pelos genes Par1A (*Glyma.09G158700*) e Par1B (*Glyma.09G158600*) (83,3% de identidade); o **Módulo 2**, contendo Par2A (*Glyma.09G158800*) e Par2B (*Glyma.14G117600*) (82,9%); e o **Módulo 3**, formado por Par3A (*Glyma.09G158900*) e Par3B (*Glyma.09G158500*) (70,8% e 70,1% respectivamente).

Notavelmente, a prospecção *in silico* desses parálogos revelou uma sobreposição com os dados de transcriptômica (RNA-Seq) obtidos de forma independente. Os *loci* que satisfizeram o limiar de identidade de sequência (>70%) coincidem com o clado de genes que apresentam expressão basal durante as fases do desenvolvimento da semente (Figura 8). Essa convergência metodológica é um indicativo de que a arquitetura dos elementos *cis*-regulatórios nesses parálogos sofreu pressões evolutivas distintas em relação ao gene alvo. Esse perfil é característico de eventos de subfuncionalização evolutiva pós-duplicação gênica (FORCE et al., 1999), onde a divergência nas regiões promotoras atua como o principal mecanismo para segregação espacial e temporal de tarefas, mesmo quando a identidade da proteína madura é conservada.

A análise integrativa das regiões promotoras (2000 pb *upstream*), cruzando dados de densidade espacial de consenso concomitante por PlantCARE, PlantPAN e FIMO/PlantRegMap (janelas de agrupamento de 30 pb), mapeou a sintaxe regulatória da família. O mapeamento demonstrou que, ao invés de uma compartimentalização estrita, a família BBI opera sob uma regulação combinatória e em gradiente. Todos os membros compartilham uma forte espinha dorsal ancestral baseada em defesas moduladas por ácido jasmônico (*MYC/MYB*) e expressão em sementes (*RY-element*), mas variam na densidade e nas combinações de motivos especializados e motivos ligados a disponibilidade do DNA aos fatores de transcrição, ditando a responsividade de cada *locus* (Tabela 13).

Esta sobreposição espacial e funcional, evidenciada pela modelagem genômica (Figura 15), delineia os perfis de atuação fisiológica entre o alvo e seus parálogos:

Figura 15 – Mapa de densidade de elementos *cis*-regulatórios e elementos de acessibilidade.



Legenda: Modelagem espacial da arquitetura regulatória comparativa da família BBI. **(A)** Distribuição espacial de módulos de motivos *cis*-regulatórios. **(B)** Densidade de elementos de acessibilidade cromatínica (*MARTBOX* e *AT-Hook*). Destaca-se a forte co-localização destes elementos nos primeiros 500 pb do gene alvo (em vermelho).

- **A Assinatura Aguda e Arquitetura Aberta do Gene Alvo:** O gene alvo (*Glyma.16G208900*) não se diferencia por possuir motivos exclusivos, mas por sua sintaxe topológica desbalanceada. Ele mantém a marcação de semente (*RY-element* (TAN et al., 2025)), mas é o único membro de toda a família que se desconectou completamente do metabolismo basal de acúmulo de reservas a longo prazo (zero cópias de *O2/GCN4*). Em compensação, ele concentra a densidade máxima do eixo de resposta aguda a dano mecânico, acumulando seis agrupamentos de ferimento (*WRE3/WUN-motif* (CLEMENTE et al., 2019)). A base mecanística para a sua hiperexpressão (Figura 9) é elucidada pela topologia proximal do seu promotor: diferentemente dos parálogos, o gene alvo concentra uma densidade massiva de elementos de acessibilidade cromatínica estritamente co-localizados com os sítios de ligação de fatores de transcrição ativadores nos primeiros 500 pb *upstream* (Figura 15). Estudos recentes elucidam que as interações guiadas por motivos *AT-Hook* promovem o remodelamento dinâmico da arquitetura da cromatina, mediando a regulação epigenética e atuando como potentes integradores de sinais de estresse biótico (MARTINČOVÁ; SOUKUP, 2025). Essa co-localização na zona proximal cria um ambiente fisicamente desimpedido que viabiliza o recrutamento imediato da maquinaria transcricional. Isso sugere que o alvo pode não atuar apenas como uma proteína de reserva basal, mas apresentar transcrição fortemente induzida em resposta a danos mecânicos pela mastigação de pragas.
- **O Espectro de Maturação e Tolerância Abiótica (Módulos 1 e 3):** Os parálogos evoluíram para sustentar o metabolismo contínuo da planta em oposição à resposta aguda do alvo. Os Módulos 1 e 3 representam um gradiente de regulação ambiental, onde a acessibilidade cromatínica é mais difusa. O **Módulo 3** concentra a maior maquinaria de metabolismo de reservas do endosperma (*O2-site/GCN4* (VERDIER; THOMPSON, 2008)) com quatro cópias por promotor, ancorada à maior densidade da família para vias hormonais sistêmicas de ABA (7 a 9 sítios) e Jasmonato (23 sítios) (LI et al., 2022). O **Módulo 1** atua como um intermediário evolutivo: preserva a expressão basal na semente (pico de sete *RY-elements* no Par1A), equilibra a dessecação pelo ABA (LAMERS et al., 2025) e mantém o alerta biótico secundário via *MYC/MYB*, mas, assim como o Módulo 3, reduziu os motivos de injúria mecânica presentes no gene alvo.
- **Divergência Imunológica Sistêmica (Módulo 2):** Dentro da evolução da família, o gene *Glyma.14G117600* (Par2B) representa o salto evolutivo em relação ao

padrão de semente. Ele perdeu o ativador primário da embriogênese (zero *RY-elements*) e as respostas a mastigação (zero *WRE3/WUN*). Em contrapartida, voltou sua expressão celular para a via imunológica adquirindo densidade para o elemento **W-box** (cinco sítios, reconhecidos por fatores de transcrição sistêmicos WRKY (LESCOT, 2002)). Este rearranjo global indica que o *locus* Par2B escapou do paradigma de acúmulo no grão para atuar na imunidade inata sistêmica (folhas, vagens e raízes) contra patógenos biotróficos e necrotróficos.

Tabela 13 – *Heatmap* da densidade de elementos *cis*-regulatórios dos parálogos do BBI alvo.

Módulo	Gene	Semente & Reservas		Estresse Abiótico	Defesa, Ferimento e Sinalização Hormonal			
		RY-element	O2/GCN4	ABA (ABRE/G-box)	Ferimento (WRE3/WUN)	Imunidade (W-box)	SA (TCA)	JA (MYC/MYB)
Alvo	<i>Glyma.16G208900</i>	5	0	2	6	1	0	16
Par 1	<i>Glyma.09G158700</i> (1A)	7	2	6	3	3	1	20
	<i>Glyma.09G158600</i> (1B)	5	2	2	1	3	1	20
Par 2	<i>Glyma.09G158800</i> (2A)	5	1	4	1	1	1	18
	<i>Glyma.14G117600</i> (2B)	0	1	4	0	5	0	13
Par 3	<i>Glyma.09G158900</i> (3A)	6	4	9	1	2	2	23
	<i>Glyma.09G158500</i> (3B)	6	4	7	2	2	1	23

Nota: Valores indicam a contagem de agrupamentos físicos únicos na região promotora (2000 pb) mapeados por meio do *score* estatístico combinado das ferramentas *in silico* (limiar de janela de 30 pb). A intensidade do sombreado indica a proporção do respectivo motivo (os valores destacados em **negrito** indicam a densidade máxima encontrada na família BBI para cada eixo). O gradiente de cores reflete a abundância de sítios de ligação de consenso, evitando falsos-negativos algorítmicos ao cruzar ocorrências validadas nas plataformas PlantCARE, PlantPAN e FIMO. Todos os promotores conservaram integralmente (uma cópia base) do *TC-rich repeats*.
Fonte: Elaborada pelo autor (2026).

A partir dessa dinâmica transcricional responsiva, a caracterização da região de 2000 pb *upstream* revela um perfil regulatório promissor tanto para a compreensão da fisiologia vegetal quanto para aplicações em biologia sintética. A coocorrência de domínios estruturais como os *AT-hook* e elementos MARTBOX sugere que este promotor coordena a expressão gênica para além do recrutamento clássico de fatores de transcrição, atuando na modulação da arquitetura da cromatina e na sua ancoragem ao arcabouço nuclear (DOLGOVA; DOLGOV, 2019). Do ponto de vista biotecnológico, a conservação e funcionalidade desses elementos conferem a este promotor o potencial de atuar como uma *bio-part* na construção de novos cassetes de expressão. Diferentemente de promotores constitutivos heterólogos, que impõem um custo energético contínuo à planta, o uso de um promotor endógeno com controle espaciotemporal refinado abre perspectivas para a expressão sítio-dirigida de proteínas recombinantes, a superexpressão de genes de defesa

ou o silenciamento gênico focado na semente, minimizando potenciais efeitos pleiotrópicos no desenvolvimento vegetativo (YU et al., 2023).

Esta análise integrativa atende ao arcabouço proposto pela biotecnologia moderna. Estudos recentes que geraram mutantes para inibidores de protease, como linhagens duplo-nulas em ervilha (CLEMENTE et al., 2015) e nocautes da família BBI via CRISPR-Cas9 em soja (KIM et al., 2024), demonstraram sucesso na redução drástica da atividade antinutricional do grão. Contudo, a literatura ressalta que o real impacto agrônômico dessas mutações na defesa contra herbivoria em campo ainda carece de validação (CLEMENTE et al., 2015; KIM et al., 2024). É exatamente nesta lacuna fisiológica que a arquitetura regulatória e de disponibilidade do DNA aqui demonstrada, em conjunto com os dados de expressão, atua: ela fornece o embasamento teórico para prever a viabilidade da planta e contrapor a hipótese de vulnerabilidade sistêmica da cultura a ataques generalizados.

Este *knockout*, tem o potencial de resultar na diminuição expressiva de um fator antinutricional contido na semente, melhorando os índices de conversão alimentar. Simultaneamente, o mapeamento *cis*-regulatório indica que a planta editada não estará metabolicamente indefesa no campo; seus parálogos, indicam a manutenção do alicerce ancestral de sinalização por *MYC/MYB* e ABA, e promotores altamente especializados (como as ilhas de *W-box* no Par2B) asseguram a capacidade de transativar rapidamente a maquinaria de resposta imunológica nos tecidos vegetativos sistêmicos (EL-BROLOS; STAINIER, 2017). Sob esta ótica, o melhoramento nutricional do grão é endossado por uma robusta redundância e resiliência fisiológica global.

6.3 CONSTRUÇÃO E VALIDAÇÃO DOS VETORES PARA EDIÇÃO DO GENE (*GLYMA.16G208900*)

Previamente à etapa de clonagem, a especificidade das sequências-alvo (sgRNA1 e sgRNA6) foram atestadas. A análise preditiva de *off-targets* gerada pelo Cas-OFFinder (BAE et al., 2014) confirmou a elevada especificidade estrutural dos guias. O mapeamento global no genoma da soja identificou potenciais regiões secundárias de ligação; no entanto, todas as ocorrências mais relevantes apresentaram de três a quatro *mismatches* em relação ao guia original, distribuídas em cromossomos distintos como o cromossomo 1, 4 e 6. Como a endonuclease zCas9 possui rigorosa exigência por homologia na região *seed* (proximal ao PAM) para estabilizar a dupla fita e induzir a quebra do DNA, a presença de múltiplos desajustes nestas posições indica que a probabilidade de atividade nucleásica *off-target* é mínima.

Assegurada a viabilidade das sequências para a inibição específica do gene BBI majoritário em sementes, prosseguiu-se para a montagem dos construtos. Os resultados

obtidos na primeira etapa de clonagem demonstraram a construção bem-sucedida dos vetores recombinantes (sgRNA1 e sgRNA6). A avaliação por sequenciamento de Sanger confirmou a integração exata dos oligonucleotídios sintéticos no vetor de entrada *pYPQ141b*, sob a regulação do promotor U3 de *A. thaliana*.

Conforme detalhado no esquema de alinhamento molecular (Tabela 14), a estratégia de clonagem direcional via enzima *Esp3I* preservou a integridade das sequências. A validação estrutural dessa etapa é crítica, uma vez que a ocorrência de quaisquer mutações na região do guia comprometeria o pareamento com o genoma da soja e anularia a capacidade de clivagem pela nuclease zCas9.

Tabela 14 – Esquema representativo da clonagem direcional dos sgRNAs no vetor pYPQ141b via digestão com *Esp3I*.

1. Região do pYPQ141b e Sítios de Clivagem (<i>Esp3I</i>)		
5'-	CACCATT↓GGTCAgagacgctctct↓GTTTTAGAG	-3'
3'-	GTGGTAACCAG↑TctctgcgagagaCAAA↑ATCTC	-5'
2. Insetos Sintéticos (sgRNAs anelados)		
sgRNA1		
5'-	ggtcAGGGGGTTACAGCTGCACGCA	-3'
3'-	TCCCCCAATGTCGACGTGCGTcaaa	-5'
sgRNA6		
5'-	ggtcATATCTTCACAGCTGCATTG	-3'
3'-	TATAGAAGTTCGACGTAACcaaa	-5'
3. Vetores Recombinantes Finais		
pYPQ141b + sgRNA1		
5'-	CACCATTggtcAGGGGGTTACAGCTGCACGCA	GTTTTAGAG -3'
3'-	GTGGTAACCCAGTCCCCCAATGTCGACGTGCGT	caaaATCTC -5'
pYPQ141b + sgRNA6		
5'-	CACCATTggtcATATCTTCACAGCTGCATTG	GTTTTAGAG -3'
3'-	GTGGTAACCCAGTATAGAAGTTCGACGTAAC	caaaATCTC -5'

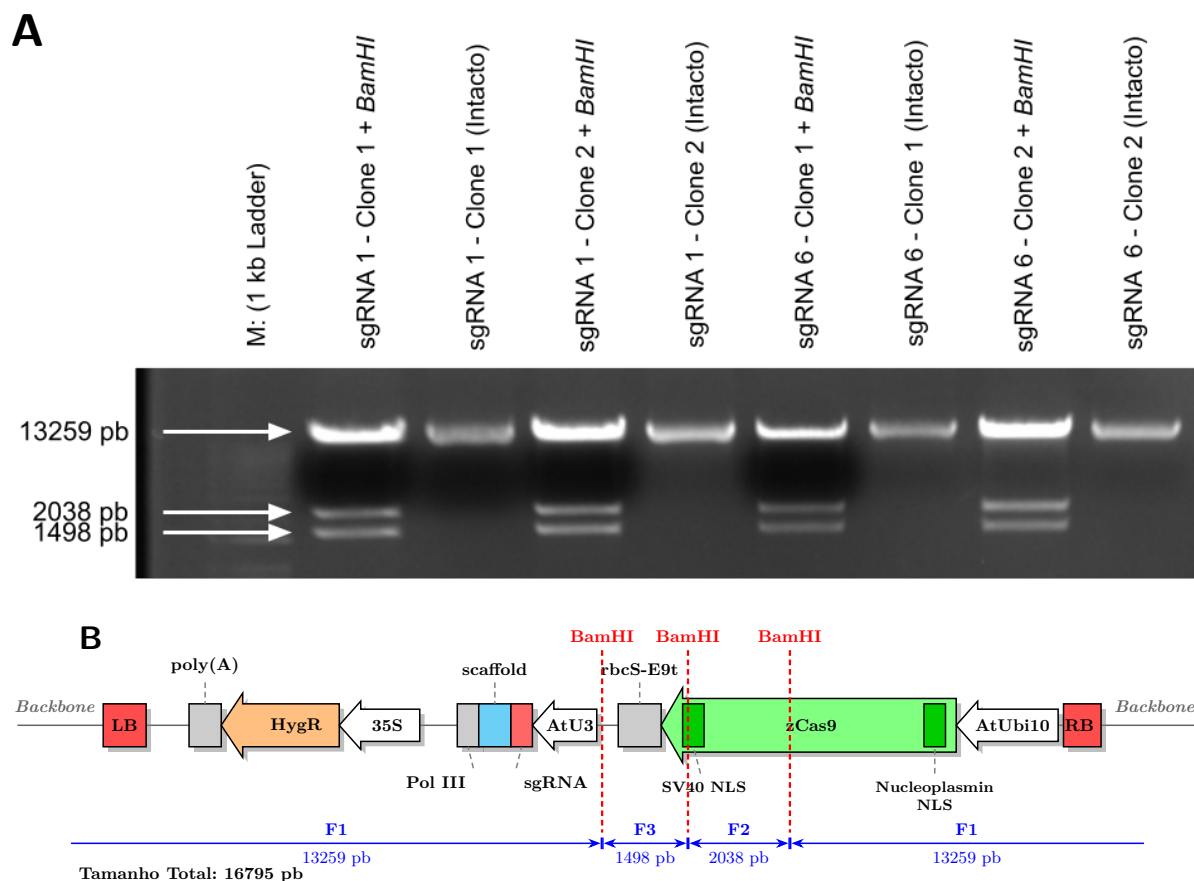
Legenda: O esquema ilustra a estratégia de clonagem direcional. Em (1), os sítios de reconhecimento da enzima de restrição *Esp3I* estão destacados em **vermelho** e os locais de clivagem indicados pelas setas (↓ e ↑), gerando extremidades coesivas assimétricas. Em (2), os oligonucleotídios desenhados contendo os braços de homologia projetados em **azul**. Em (3), a sequência dos plasmídeos após a ligação enzimática, com os guias sublinhados, evidenciando a complementaridade, validado experimentalmente por sequenciamento.

Fonte: Elaborada pelo autor (2026).

A subsequente montagem do vetor final via recombinação Gateway® LR resultou na fusão dos cassetes (*pYPQ141b*) contendo os sgRNAs com a nuclease zCas9 (oriunda do plasmídeo intermediário *pYPQ166*) no vetor de destino *pYPQ202*. Para atestar correta estrutura do vetor, os clones obtidos foram submetidos à análise de restrição enzimática.

A digestão *Bam*HI gerou o padrão de fragmentação condizente com o *in silico* para a montagem correta (Figura 16). A constância do perfil de bandas entre clones independentes corroborou a estabilidade das construções binárias finais.

Figura 16 – Validação molecular e mapa esquemático do vetor binário final zCas9-BBI



Legenda: (A) Análise de restrição enzimática dos vetores binários finais em gel de agarose (1%). Os canais analisam independentemente dois clones (Clone 1 e Clone 2) para os construtos portando o sgRNA 1 e o sgRNA 6. Para cada clone, o plasmídeo digerido com a enzima de restrição *Bam*HI é apresentado ao lado do seu controle circular não clivado (**Intacto**). As setas à esquerda apontam os fragmentos gerados pela digestão. **M:** Marcador de peso molecular (*1 kb Ladder*). (B) Mapa esquemático linearizado do vetor evidenciando a orientação transcricional dos cassetes genéticos. As linhas tracejadas indicam a localização dos três sítios de restrição *Bam*HI. A clivagem simultânea gera o padrão dos três fragmentos observados no gel: **Fragmento 1 (F1):** 13.259 pb, contendo o *backbone* e cassetes periféricos; **Fragmento 2 (F2):** 2.038 pb, correspondente à região interna da nuclease zCas9; **Fragmento 3 (F3):** 1.498 pb, abrangendo o término da zCas9 e o terminador rbcS-E9t, corroborando a integridade arquitetural da montagem. **Fonte:** Elaborada pelo autor (2026).

Após a confirmação estrutural *in vitro* dos vetores zCas9-BBI-sgRNA1 e zCas9-BBI-sgRNA6, a etapa subsequente consistiu na transformação dessas construções para a linhagem GV3101 de *A. tumefaciens*. A eficiência da transformação e a manutenção estável do plasmídeo íntegro no interior da bactéria hospedeira foram testadas via PCR de colônia. Utilizando-se iniciadores (*primers*) específicos desenhados para a região do T-DNA

(Tabela 5), obteve-se a amplificação do fragmento de 1224 pb, o tamanho esperado para o amplicon da região do vetor zCas9-BBI.

6.4 EFICIÊNCIA DE TRANSFORMAÇÃO E CARACTERIZAÇÃO DAS EDIÇÕES

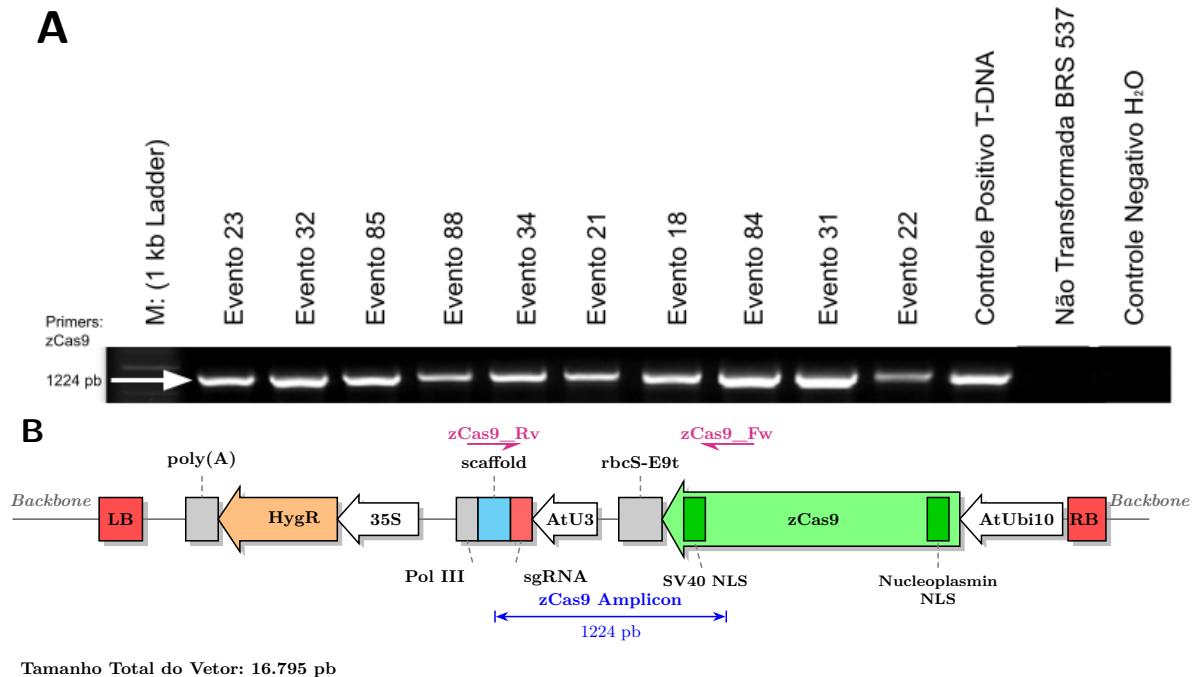
A estratégia de transformação genética do meristema apical mediada por *A. tumefaciens* em eixos embrionários da cultivar de soja BRS 537 resultou na regeneração e aclimatização de 220 plantas (T_0), representando uma taxa de regeneração de aproximadamente 44% a partir de, aproximadamente, 500 explantes iniciais. A triagem molecular primária para a presença do transgene foi conduzida por seleção química celular *in vitro*, utilizando higromicina. Posteriormente, a integração do T-DNA no genoma vegetal foi confirmada via PCR convencional. Essa análise genotípica em larga escala atestou a amplificação do fragmento esperado em 125 indivíduos distintos, indicando uma eficiência de transformação e seleção de 56,8%.

Para atestar a robustez da genotipagem e ilustrar o padrão de integração do T-DNA, a Figura 17A apresenta um perfil eletroforético representativo de dez eventos transgênicos independentes, amostrados aleatoriamente a partir do conjunto total de 125 plantas positivas. O controle genômico negativo (WT), extraído de uma planta BRS 537 não submetida ao processo de co-cultivo, confirmou a ausência de amplificação inespecífica no genoma nativo da soja, enquanto o controle positivo plasmidial (zCas9-BBI) estabeleceu o parâmetro esperado para o inserto.

A estratégia molecular desenhada para esta genotipagem foi pautada na detecção simultânea de múltiplos elementos essenciais à edição, conforme esquematizado no diagrama do vetor linearizado (Figura 17B). Foram utilizados os primers específicos (zCas9 Fw e zCas9 Rv) descritos na Tabela 5. O primer reverso (zCas9 Rv) foi desenhado para anelar na sequência do *scaffold* do sgRNA, enquanto o primer *forward* (zCas9 Fw) ancora-se na nuclease *zCas9*.

A escolha desses sítios de anelamento que geram o fragmento amplificado de 1224pb não detecte apenas a nuclease, mas a integração física de uma fita contínua contendo o *scaffold* do sgRNA, o promotor AtU3 e o terminador rbcS-E9t, adentrando a sequência da *Cas9*. Esta configuração de PCR minimiza o risco de falsos positivos decorrentes de integrações truncadas do T-DNA, garantindo que os 125 eventos selecionados possuam a arquitetura genética mínima e íntegra requerida para a clivagem do gene alvo *Glyma.16G208900*.

Figura 17 – Validação molecular das plantas T₀ e estratégia de genotipagem por PCR para o transgene *zCas9*.



Legenda: Confirmação da integração do T-DNA no genoma da soja cultivar BRS 537. **(A)** Eletroforese em gel de agarose (1%) corado com brometo de etídio evidenciando a amplificação do fragmento esperado de 1224 pb em uma amostragem representativa de dez eventos transgênicos independentes (T₀). **M:** Marcador de massa molecular (1 kb DNA Ladder); **Controle Positivo (PC):** Plasmídeo zCas9BBI; **Controle Negativo Biológico (WT):** DNA genômico de planta não transformada (BRS 537); **Controle Negativo da Reação (NC):** Água (H₂O). **(B)** Representação esquemática da região do T-DNA detalhando a estratégia de genotipagem. As setas em magenta indicam a localização e o sentido de amplificação dos iniciadores zCas9 Fw e zCas9 Rv. O **amplicon de 1224 pb** (barra azul) abrange a região do *scaffold* do sgRNA, perpassa o promotor AtU3 e o terminador rbcS-E9t, adentrando a sequência codificante da nuclease *zCas9*, confirmando a integridade física do cassete de edição.

Fonte: Elaborada pelo autor (2026).

Para a validação da edição no gene alvo *Glyma.16G208900*, um total de 122 plantas transgênicas foram submetidas ao sequenciamento de Sanger. A análise dos cromatogramas pela ferramenta ICE (*Inference of CRISPR Edits*) permitiu a identificação de dez plantas editadas confirmadas (denominadas Eventos 23, 32, 85, 88, 34, 21, 18, 84, 31 e 22), correspondendo a uma eficiência de edição de 8,2% (10/122) na população genotipada (Tabela 15).

Tabela 15 – Resumo das eficiências de regeneração, transformação e edição gênica obtidas na cultivar BRS 537.

Etapa do Processo	N.º de Amostras	Taxa de Eficiência
Explantos Iniciais (Eixos Embrionários)	500	—
Plantas Regeneradas e Aclimatizadas (T ₀)	220	44,0% ^(a)
Plantas Transgênicas Confirmadas (PCR+)	125	56,8% ^(b)
Plantas Genotipadas (Sanger/ICE)	122	—
Plantas Editadas	10	8,2% ^(c)

Notas: (a) Taxa de regeneração calculada sobre o número de explantes iniciais. (b) Eficiência de transformação (presença do T-DNA) calculada sobre o total de plantas regeneradas. (c) Eficiência de edição calculada com base no total de plantas submetidas à genotipagem.

Fonte: Elaborada pelo autor (2026).

A caracterização molecular dos dez eventos está descrita na Tabela 16. A análise das sequências demonstrou uma diversidade no perfil de reparo mediado pela via NHEJ, porém somente o sgRNA 1 demonstrou eficácia na geração de mutações dirigidas nas amostras analisadas. Os Eventos 32 e 85 destacaram-se por apresentarem altas taxas de contribuição de alelos mutados (73% e 69%, respectivamente), caracterizados predominantemente por deleções de 9 pb e 10 pb no sítio de clivagem. Em contrapartida, o Evento 23 exibiu um perfil quimérico mais complexo (29% de edição total), onde o alelo selvagem (37%) coexiste com múltiplas variantes, sendo as mais representativas uma inserção de 1 pb (+1, 19%) e uma inserção de 3 pb (+3, 6%).

Além dos eventos com altas taxas de mutação, a análise identificou um conjunto de sete plantas (88, 34, 21, 18, 84, 31 e 22) exibindo frequências de edição, variando entre 2% e 3%. Embora representem uma contribuição alélica reduzida no tecido amostrado, a manutenção e o relato desses eventos de baixa eficiência são cruciais. Na geração primária (T₀), o mosaico genético implica que a taxa de mutação quantificada nos pequenos discos foliares específicos podem não refletir a composição genômica de toda a planta, especialmente nos tecidos reprodutivos. Dessa forma, essas plantas confirmam a atividade e a funcionalidade da nuclease zCas9 no genoma alvo e preservam o potencial de segregação de alelos editados estáveis nas gerações subsequentes, justificando sua inclusão no painel de genotipagem. Dessa forma, a detecção dessas mutações, mesmo em baixa frequência no tecido foliar analisado, constitui o indicativo biológico necessário para justificar o avanço dessas plantas para a geração T₁. Nesse estágio subsequente, espera-se que a segregação mendeliana permita a fixação dos alelos editados em homozigose e a obtenção de linhagens estáveis.

Esta coexistência de diferentes alelos representa um padrão esperado e frequen-

temente relatado para plantas na geração T_0 , as quais se desenvolvem comumente como quimeras genéticas (ZHENG et al., 2020). Por se tratar do evento primário de regeneração, o tecido vegetal apresenta uma composição em mosaico, fenômeno que decorre não apenas da origem multicelular inerente à organogênese *in vitro*, que pode abrigar tanto células transformadas quanto eventuais escapes selvagens, mas, fundamentalmente, da ação contínua da nuclease Cas9 ao longo do desenvolvimento do explante (MALABARBA et al., 2020). Como a enzima é guiada por um promotor constitutivo, os eventos de clivagem e reparo ocorrem de forma assíncrona durante as sucessivas divisões celulares. Esse prolongamento da atividade enzimática permite que diferentes linhagens celulares estabeleçam perfis de edição distintos em estádios tardios, consolidando o mosaicismo biológico observado no indivíduo regenerado (MALABARBA et al., 2020; IMPENS et al., 2022). Devido à natureza dessas mutações (*indels* de pequena escala), o quimerismo e a presença de células não editadas não podem ser distinguidos visualmente por meio da eletroforese em gel de agarose, exigindo a análise de decomposição de cromatogramas para sua identificação. Assim, as frequências alélicas observadas refletem a atividade contínua e assíncrona da nuclease Cas9 ao longo do desenvolvimento somático, resultando em porcentagens variáveis de edição em diferentes partes da mesma planta.

Tabela 16 – Perfil das edições e alinhamento genômico das linhagens de soja BRS 537 com edição no gene *Glyma.16G208900*

Amostra	Sequência Alvo	Indel Freq. (%)	
Referência	TTCCTTGTGGGGGTTACAGCTGCACGCCATGGAACTGAGCTTCTTCAAAAGT	0	–
Evento 32	TTCCTTGTGGGGGTTACAGCTGCACGCCATGGAACTGAGCTTCTTCAAAAGT	0	27,0
	TTCCTTGTGGGGGTTA-----GCATGGAACTGAGCTTCTTCAAAAGT	-9	73,0
Evento 85	TTCCTTGTGGGGGTTACAGCTGCACGCCATGGAACTGAGCTTCTTCAAAAGT	0	31,0
	TTCCTTGTGGGGGTT-----GCATGGAACTGAGCTTCTTCAAAAGT	-10	69,0
Evento 88	TTCCTTGTGGGGGTTACAGCTGCACGCCATGGAACTGAGCTTCTTCAAAAGT	0	97,0
	TTCCTTGT-----AAAAGT	-37	1,0
	TTCCTTGTG-----TGGAACTGAGCTTCTTCAAAAGT	-19	1,0
	TTCCTTGTGGGGGTTACAGCT-----CTTCTTCAAAAGT	-17	1,0
Evento 34	TTCCTTGTGGGGGTTACAGCTGCACGCCATGGAACTGAGCTTCTTCAAAAGT	0	97,0
	TTCCTTGT-----AAAAGT	-37	1,0
	TTCCTTGTG-----TGGAACTGAGCTTCTTCAAAAGT	-19	1,0
	TTCCTTGTGGGGGTTACAGCT-----CTTCTTCAAAAGT	-17	1,0
Evento 21	TTCCTTGTGGGGGTTACAGCTGCACGCCATGGAACTGAGCTTCTTCAAAAGT	0	97,0
	TTCCTTGT-----AAAAGT	-37	1,0
	TTCCTTGTG-----TGGAACTGAGCTTCTTCAAAAGT	-19	1,0
	TTCCTTGTGGGGGTTACAGCT-----CTTCTTCAAAAGT	-17	1,0
Evento 18	TTCCTTGTGGGGGTTACAGCTGCACGCCATGGAACTGAGCTTCTTCAAAAGT	0	97,0
	TTCCTTGT-----AAAAGT	-37	1,0
	TTCCTTGTG-----TGGAACTGAGCTTCTTCAAAAGT	-19	1,0
	TTCCTTGTGGGGGTTACAGCT-----CTTCTTCAAAAGT	-17	1,0
Evento 84	TTCCTTGTGGGGGTTACAGCTGCACGCCATGGAACTGAGCTTCTTCAAAAGT	0	98,0
	TTCCTT-----CATGGAACTGAGCTTCTTCAAAAGT	-20	1,0
	TTCCTTGTGGGGGTTACAGCTG-----TTCAAAGT	-20	1,0
Evento 31	TTCCTTGTGGGGGTTACAGCTGCACGCCATGGAACTGAGCTTCTTCAAAAGT	0	98,0
	TTCCTT-----CATGGAACTGAGCTTCTTCAAAAGT	-20	1,0
	TTCCTTGTGGGGGTTACAGCTG-----TTCAAAGT	-20	1,0
Evento 22	TTCCTTGTGGGGGTTACAGCTGCACGCCATGGAACTGAGCTTCTTCAAAAGT	0	98,0
	TTCCTT-----CATGGAACTGAGCTTCTTCAAAAGT	-20	1,0
	TTCCTTGTGGGGGTTACAGCTG-----TTCAAAGT	-20	1,0
Evento 23	TTCCTTGTGGGGGTTACAGCTGCACGCCATGGAACTGAGCTTCTTCAAAAGT	0	37,0
	TTCCTTGTGGGGGTTACAGCTGCACNNGCATGGAACTGAGCTTCTTCAAAAG	+1	19,0
	TTCCTTGTGGGGGTTACAGCTGCACNNNGCATGGAACTGAGCTTCTTCAAA	+3	6,0
	TTCCTTGTGGGGGTTACAGCTGCA-GCATGGAACTGAGCTTCTTCAAAAGT	-1	2,0
	TTCCTTGTGGGGGTTACAGCTG---GCATGGAACTGAGCTTCTTCAAAAGT	-3	1,0

Legenda: Alinhamento *in silico* estruturado em grade evidenciando as alterações nucleotídicas. Cada célula possui preenchimento em cor para o nucleotídeo (A em **verde**, T em **vermelho**, C em **amarelo**, G em **azul**). A sequência alvo correspondente ao sgRNA está sublinhada. O motivo PAM distingue-se pela coloração **vermelha** da fonte, mantendo o preenchimento celular correspondente à base. Mutações *indels* (deleções representadas por hifens e inserções por "**N**" **vermelhos**) possuem preenchimento **cinza**. A frequência representa a contribuição relativa de cada variante alélica na amostra quimérica estimada pelo algoritmo do *software* ICE.

Fonte: Elaborada pelo autor (2026).

Do ponto de vista funcional, as mutações geradas nos diferentes eventos apresentam potenciais impactos distintos sobre a integridade da proteína codificada. Mutações cujo tamanho não corresponde a múltiplos de três nucleotídeos, como a deleção de 10 pb (–10) no Evento 85 e a inserção de 1 pb (+1) no Evento 23, promovem o deslocamento da matriz

de leitura (*frameshift*). Esse tipo de alteração geralmente resulta na geração de códons de parada prematuros, o que tende a produzir proteínas truncadas e, conseqüentemente, a formação de alelos de perda de função (JINEK et al., 2012). Por outro lado, mutações múltiplas de três nucleotídeos (*in-frame*), como a deleção de 9 pb (−9) predominante no Evento 32, resultam na remoção específica de três aminoácidos. Embora não causem o deslocamento da matriz de leitura, deleções *in-frame* podem comprometer a funcionalidade da proteína. A remoção de resíduos essenciais, ou a desestabilização da rede de pontes de dissulfeto pode impedir o reconhecimento canônico pelas proteases alvo, inviabilizando a atividade inibitória da proteína codificada e potencialmente levando à perda funcional do inibidor (QI et al., 2005; KROWARSCH et al., 2003; KIM et al., 2024).

Considerando que os inibidores de protease Bowman-Birk dependem criticamente da integridade estrutural desses laços reativos e de sua rede de pontes dissulfeto para manter a interação com serino-proteases, mesmo pequenas alterações na sequência podem comprometer a atividade inibitória característica dessa família de proteínas. Dessa forma, as mutações do tipo *frameshift* observadas sugerem a provável geração de alelos de perda de função com potencial para reduzir ou abolir a atividade inibitória associada ao inibidor Bowman-Birk, contribuindo potencialmente para a mitigação do fator antinutricional em sementes de soja. Entretanto, a confirmação definitiva do impacto funcional dessas alterações requer análises adicionais em nível proteico e bioquímico. Nesse contexto, abordagens como a quantificação da proteína BBI por métodos imunológicos (por exemplo, *Western blot*), aliadas a ensaios de atividade inibitória frente a serino-proteases digestivas, tripsina e quimotripsina, seriam necessárias para verificar experimentalmente a perda da atividade característica desses inibidores nas linhagens editadas.

A obtenção de diferentes perfis mutacionais entre os eventos analisados também evidencia a eficiência da ferramenta de edição no genoma da cultivar BRS 537. Essa diversidade de alelos editados demonstra a capacidade da tecnologia CRISPR/Cas9 de gerar variação genética dirigida em *loci* específicos, ampliando as possibilidades de modulação funcional de genes associados a características agrônômicas de interesse. O melhoramento genético focado na qualidade do grão é um pilar determinante para a segurança e a sustentabilidade alimentar, uma vez que a soja atua como uma das principais e mais ricas fontes de proteína vegetal em escala global.

Nesse contexto, o *knockout* mediado por CRISPR/Cas9 direcionado aos fatores antinutricionais representa uma estratégia biotecnológica de impacto (KAFFER et al.,). O *knockout* racional de genes codificadores de inibidores de proteases, como o inibidor de tripsina Kunitz (WANG et al., 2023b), associado à inativação de outras proteínas deletérias como a aglutinina da soja (*Le1*) (KAFFER et al., 2026) e a alteração nas rotas de biossíntese de saponinas (ASA et al., 2025), viabiliza a geração de plantas com o

metabolismo remodelado a favor do consumo. Dessa forma, a aplicação dessa abordagem não apenas tem o potencial de acelerar o desenvolvimento de cultivares *transgene-free* com perfil nutricional aprimorado, mas também de maximizar a conversão alimentar e a eficiência da utilização proteica nas cadeias de produção animal e humana.

7 CONCLUSÕES E PERSPECTIVAS

7.1 CONCLUSÕES

O presente estudo estabeleceu com sucesso uma plataforma funcional para a edição genômica na cultivar brasileira de soja BRS 537, resultando na obtenção de genótipos mutantes para o inibidor de proteases Bowman-Birk majoritário. Com base nos resultados experimentais e nas validações *in silico*, destacam-se:

- A análise integrada indicou o gene *Glyma.16G208900* como o alvo prioritário para edição gênica, visto que sua expressão superior identificada *in silico* foi corroborada *in vitro*. A prospecção de promotores revelou que, embora a família compartilhe sinalização basal de semente, o gene alvo apresenta especialização para resposta a danos mecânicos, o que sugere que seu *knockout* via CRISPR/Cas9 reduzirá eficientemente o acúmulo reativo de BBI no grão sem comprometer a resiliência da planta, graças à redundância de seus parálogos, a presença dos elementos de acessibilidade do DNA e sua manutenção no restante dos tecidos da planta.
- A modelagem molecular via AlphaFold v2 indica a topologia clássica “cabeça-dupla” do alvo, com sítios reativos (Arg16 e Arg43) geometricamente competentes para a inibição simultânea de tripsina e quimotripsina, ratificando a necessidade de sua inativação.
- O protocolo de transformação genética via organogênese direta em eixos embrionários maduros, mediado por *A. tumefaciens*, mostrou-se robusto para a cultivar BRS 537, alcançando uma taxa de regeneração de 44% e eficiência de transformação (presença de T-DNA) de 56,8%.
- O sistema CRISPR/Cas9 desenhado foi eficaz na clivagem do DNA genômico, resultando na identificação de dez eventos de edição independentes (eficiência de 8,2% em uma população de 122 plantas genotipadas). As mutações geradas abrangeram perfis distintos, incluindo deleções extensas (-37, -20, -19 e -17 pb), deleções curtas (-10 e -9 pb) e inserções (+1 e +3 pb). Alterações que causam *frameshift* (como -10 e +1) geram códons de parada prematuros, enquanto deleções *in-frame* (-9) promovem a exclusão de aminoácidos, ambas as vias convergindo para a desestabilização estrutural e indicando o *knockout* desejado. Adicionalmente, a detecção de eventos com frequências de edição diferenciadas reitera a natureza quimérica das plantas T_0 e a importância de seu avanço geracional para a segregação de alelos estáveis, reiterando que a presença de mutações em plantas quiméricas da geração T_0 é o

indicativo biológico que valida o avanço para a geração T_1 e subsequentes, na qual a segregação mendeliana atuará na fixação dos alelos de interesse.

7.2 PERSPECTIVAS

A obtenção destas dez plantas editadas na geração T_0 abre novas frentes de investigação necessárias para consolidar o desenvolvimento de uma cultivar de soja com maior digestibilidade proteica. Como desdobramentos deste projeto, propõe-se:

1. **Segregação e Obtenção de Linhagens Livres de Transgene:** Conduzir o avanço de gerações para identificar, via Sequenciamento de Sanger e Sequenciamento de Nova Geração (NGS), plantas da geração T_1 e subsequentes que herdem as mutações em homozigose, mas que tenham segregado o cassete CRISPR/Cas9 (T-DNA). A obtenção de plantas *transgene-free* é essencial para facilitar a aprovação regulatória e a aceitação comercial.
2. **Validação Proteômica e Enzimática:** Realizar ensaios de atividade inibitória de tripsina e quimotripsina *in vitro* utilizando extratos proteicos das sementes editadas homozigotas frente a sementes NT. Isso permitirá quantificar a redução real da atividade antinutricional em comparação à cultivar de soja BRS 537, validando o efeito estrutural e funcional das diferentes mutações geradas (deleções extensas vs. *frameshifts*).
3. **Avaliação Agronômica:** Monitorar o desenvolvimento fenotípico das linhagens editadas em casa de vegetação e campo, avaliando parâmetros como altura, número de vagens, peso de sementes e teor total de proteína e óleo, assim como o ataque de pragas, a fim de assegurar que a inativação do gene BBI não acarreta penalidades pleiotrópicas (perda de produtividade ou aumento da suscetibilidade a insetos).

REFERÊNCIAS

- ABDURAKHMOM, N. Y.; IBROKHIM, Y. A.; ZABARDAST, T. B.; LOLA, K. K.; ZIYODULLOXON, H. B.; ANVARJON, A. M.; NURDINJON, S. O.; BEKHZOD, O. M.; AKMAL, M. A.; MIRZAKAMOL, S. A. Recent efforts to increase Soybean (*Glycine max* L.) tolerance to abiotic stresses using CRISPR-Cas technology. **Plant Science Today**, v. 12, n. 4, 2025. ISSN 2348-1900. Disponível em: <<https://horizonepublishing.com/journals/index.php/PST/article/view/8822>>.
- AKDEL, M.; PIRES, D. E. V.; PARDO, E. P.; JÄNES, J.; ZALEVSKY, A. O.; MÉSZÁROS, B.; BRYANT, P.; GOOD, L. L.; LASKOWSKI, R. A.; POZZATI, G.; SHENOY, A.; ZHU, W.; KUNDROTAS, P.; SERRA, V. R.; RODRIGUES, C. H. M.; DUNHAM, A. S.; BURKE, D.; BORKAKOTI, N.; VELANKAR, S.; FROST, A.; BASQUIN, J.; LINDORFF-LARSEN, K.; BATEMAN, A.; KAJAVA, A. V.; VALENCIA, A.; OVCHINNIKOV, S.; DURAIRAJ, J.; ASCHER, D. B.; THORNTON, J. M.; DAVEY, N. E.; STEIN, A.; ELOFSSON, A.; CROLL, T. I.; BELTRAO, P. A structural biology community assessment of AlphaFold2 applications. **Nature Structural & Molecular Biology**, v. 29, n. 11, p. 1056–1067, 2022. ISSN 1545-9993, 1545-9985. Disponível em: <<https://www.nature.com/articles/s41594-022-00849-w>>.
- ALLEN, G. C.; HALL JR, G. E.; CHILDS, L. C.; WEISSINGER, A. K.; SPIKER, S.; THOMPSON, W. F. Scaffold attachment regions increase reporter gene expression in stably transformed plant cells. **The Plant Cell**, v. 5, n. 6, p. 603–613, 1993. ISSN 1040-4651. Disponível em: <<https://doi.org/10.1105/tpc.5.6.603>>.
- ANDRUSIER, N.; MASHIACH, E.; NUSSINOV, R.; WOLFSON, H. J. Principles of flexible protein–protein docking. **Proteins: Structure, Function, and Bioinformatics**, v. 73, n. 2, p. 271–289, 2008. ISSN 1097-0134. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.22170>. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.22170>>.
- ARAKI, M.; ISHII, T. Towards social acceptance of plant breeding by genome editing. **Trends in Plant Science**, v. 20, n. 3, p. 145–149, 2015. ISSN 13601385. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1360138515000291>>.
- ARAVIND, L.; LANDSMAN, D. AT-hook motifs identified in a wide variety of DNA-binding proteins. **Nucleic Acids Research**, v. 26, n. 19, p. 4413–4421, 1998. ISSN 0305-1048. Disponível em: <<https://doi.org/10.1093/nar/26.19.4413>>.
- ASA, H.; KUWABARA, C.; MATSUMOTO, K.; SHIGETA, R.; YAMAMOTO, T.; MASUDA, Y.; YAMADA, T. Simultaneous site-directed mutagenesis for soybean β -amyrin synthase genes via DNA-free CRISPR/Cas9 system using a single gRNA. **Plant Cell Reports**, v. 44, n. 2, p. 40, 2025. ISSN 1432-203X. Disponível em: <<https://doi.org/10.1007/s00299-025-03433-w>>.
- BAE, S.; PARK, J.; KIM, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. **Bioinformatics**, v. 30, n. 10, p. 1473–1475, 2014. ISSN 1367-4803. Disponível em: <<https://doi.org/10.1093/bioinformatics/btu048>>.

BARROS, B. D. A.; SILVA, W. G. D.; MOREIRA, M. A.; BARROS, E. G. D. *In silico* characterization and expression analysis of the multigene family encoding the Bowman–Birk protease inhibitor in soybean. **Molecular Biology Reports**, v. 39, n. 1, p. 327–334, 2012. ISSN 0301-4851, 1573-4978. Disponível em: <<http://link.springer.com/10.1007/s11033-011-0742-1>>.

BIRK, Y. The Bowman-Birk inhibitor. Trypsin- and chymotrypsin-inhibitor from soybeans. **International Journal of Peptide and Protein Research**, v. 25, n. 2, p. 113–131, 1985. ISSN 0367-8377. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1399-3011.1985.tb02155.x>>.

BODE, W.; HUBER, R. Natural protein proteinase inhibitors and their interaction with proteinases. **European Journal of Biochemistry**, v. 204, n. 2, p. 433–451, 1992. ISSN 0014-2956, 1432-1033. Disponível em: <<https://febs.onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1992.tb16654.x>>.

BRASIL. Comissão Técnica Nacional de Biossegurança (CTNBio). **Resolução Normativa nº 16, de 15 de janeiro de 2018**: Dispõe sobre as técnicas inovadoras de melhoramento de precisão (timp). Brasília, 2018. Disponível em: <https://antigo.mctic.gov.br/mctic/opencms/legislacao/outros_atos/resolucoes/Resolucao_Normativa_CTNBio_n_16_de_15012018.html>.

BUHR, T.; SATO, S.; EBRAHIM, F.; XING, A.; ZHOU, Y.; MATHIESEN, M.; SCHWEIGER, B.; KINNEY, A.; STASWICK, P.; CLEMENTE, T. Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. **The Plant Journal**, v. 30, n. 2, p. 155–163, 2002. ISSN 1365-313X. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-313X.2002.01283.x>. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-313X.2002.01283.x>>.

BURSENS, S.; PERTRY, I.; DIASOLUA, D.; KUO, Y.-H.; VAN, M.; LAMBEI, F. Soya, Human Nutrition and Health. In: EL-SHEMY, H. (Ed.). **Soybean and Nutrition**. InTech, 2011. ISBN 978-953-307-536-5. Disponível em: <<http://www.intechopen.com/books/soybean-and-nutrition/soya-human-nutrition-and-health>>.

BUSTIN, S. A.; BENES, V.; GARSON, J. A.; HELLEMANS, J.; HUGGETT, J.; KUBISTA, M.; MUELLER, R.; NOLAN, T.; PFAFFL, M. W.; SHIPLEY, G. L.; VANDESOMPELE, J.; WITTEWER, C. T. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. **Clinical Chemistry**, v. 55, n. 4, p. 611–622, 2009. ISSN 1530-8561. Disponível em: <<https://doi.org/10.1373/clinchem.2008.112797>>.

CARRAO-PANIZZI, M. C.; MANDARINO, J. M. G. Soja: potencial de uso na dieta brasileira. Londrina: EMBRAPA-CNPSO, 1998., 1998. Accepted: 2011-04-10T11:11:11Z. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/handle/doc/460385>>.

CARRÃO-PANIZZI, M. C.; PÍPOLO, A. E. **Cultivares de soja especiais para alimentação humana**. Embrapa Soja, 2009. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/469886>>.

CHEN, K.; WANG, Y.; ZHANG, R.; ZHANG, H.; GAO, C. CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture. **Annual Review of Plant**

Biology, Annual Reviews, v. 70, n. Volume 70, 2019, p. 667–697, 2019. ISSN 1543-5008, 1545-2123. Disponível em: <<https://www.annualreviews.org/content/journals/10.1146/annurev-arplant-050718-100049>>.

CHEN, S.; TIAN, Z.; GUO, Y. Characterization of hexokinase gene family members in *Glycine max* and functional analysis of GmHXK2 under salt stress. **Frontiers in Genetics**, Frontiers, v. 14, 2023. ISSN 1664-8021. Disponível em: <<https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2023.1135290/full>>.

CHEN, Y.; XU, Z.; ZHANG, C.; KONG, X.; HUA, Y. Heat-induced inactivation mechanisms of Kunitz trypsin inhibitor and Bowman-Birk inhibitor in soymilk processing. **Food Chemistry**, v. 154, p. 108–116, 2014. ISSN 0308-8146. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0308814613019729>>.

CHOW, C.-N.; YANG, C.-W.; WU, N.-Y.; WANG, H.-T.; TSENG, K.-C.; CHIU, Y.-H.; LEE, T.-Y.; CHANG, W.-C. PlantPAN 4.0: updated database for identifying conserved non-coding sequences and exploring dynamic transcriptional regulation in plant promoters. **Nucleic Acids Research**, v. 52, n. D1, p. D1569–D1578, 2024. ISSN 0305-1048, 1362-4962. Disponível em: <<https://academic.oup.com/nar/article/52/D1/D1569/7332064>>.

CLEMENTE, A.; ARQUES, M. C.; DALMAIS, M.; SIGNOR, C. L.; CHINOY, C.; OLIAS, R.; RAYNER, T.; ISAAC, P. G.; LAWSON, D. M.; BENDAHDANE, A.; DOMONEY, C. Eliminating Anti-Nutritional Plant Food Proteins: The Case of Seed Protease Inhibitors in Pea. **PLOS ONE**, Public Library of Science, v. 10, n. 8, p. e0134634, 2015. ISSN 1932-6203. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134634>>.

CLEMENTE, M.; CORIGLIANO, M. G.; PARIANI, S. A.; SÁNCHEZ-LÓPEZ, E. F.; SANDER, V. A.; RAMOS-DUARTE, V. A. Plant Serine Protease Inhibitors: Biotechnology Application in Agriculture and Molecular Farming. **International Journal of Molecular Sciences**, v. 20, n. 6, p. 1345, 2019. ISSN 1422-0067. Disponível em: <<https://www.mdpi.com/1422-0067/20/6/1345>>.

COLLAKOVA, E.; AGHAMIRZAIE, D.; FANG, Y.; KLUMAS, C.; TABATABA, F.; KAKUMANU, A.; MYERS, E.; HEATH, L. S.; GRENE, R. Metabolic and Transcriptional Reprogramming in Developing Soybean (*Glycine max*) Embryos. **Metabolites**, Multidisciplinary Digital Publishing Institute, v. 3, n. 2, p. 347–372, 2013. ISSN 2218-1989. Disponível em: <<https://www.mdpi.com/2218-1989/3/2/347>>.

Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de grãos: Safra 2025/26 – 4º Levantamento**. [S.l.], 2026. Disponível em: <<https://www.gov.br/conab/pt-br/atuacao/informacoes-agropecuarias/safras/safra-de-graos/boletim-da-safra-de-graos/4o-levantamento-safra-2025-26/4o-levantamento-safra-2025-26>>.

CONANT, D.; HSIAU, T.; ROSSI, N.; OKI, J.; MAURES, T.; WAITE, K.; YANG, J.; JOSHI, S.; KELSO, R.; HOLDEN, K.; ENZMANN, B. L.; STONER, R. Inference of CRISPR Edits from Sanger Trace Data. **The CRISPR Journal**, v. 5, n. 1, p. 123–130, 2022. ISSN 2573-1599, 2573-1602. Disponível em: <<https://www.liebertpub.com/doi/10.1089/crispr.2021.0113>>.

CONG, L.; RAN, F. A.; COX, D.; LIN, S.; BARRETTO, R.; HABIB, N.; HSU, P. D.; WU, X.; JIANG, W.; MARRAFFINI, L. A.; ZHANG, F. Multiplex Genome Engineering Using CRISPR/Cas Systems. **Science**, v. 339, n. 6121, p. 819–823, 2013. ISSN 0036-8075, 1095-9203. Disponível em: <<https://www.science.org/doi/10.1126/science.1231143>>.

DELLAPORTA, S. L.; WOOD, J.; HICKS, J. B. A plant DNA miniprep: Version II. **Plant Molecular Biology Reporter**, v. 1, n. 4, p. 19–21, 1983. ISSN 0735-9640, 1572-9818. Disponível em: <<http://link.springer.com/10.1007/BF02712670>>.

DIPIETRO, C. M.; LIENER, I. E. Heat inactivation of the kunitz and bowman-birk soybean protease inhibitors. **Journal of Agricultural and Food Chemistry**, v. 37, n. 1, p. 39–44, 1989. Disponível em: <<https://doi.org/10.1021/jf00085a010>>.

DOLGOVA, A. S.; DOLGOV, S. V. Matrix attachment regions as a tool to influence plant transgene expression. **3 Biotech**, v. 9, n. 5, p. 176, 2019. ISSN 2190-5738. Disponível em: <<https://doi.org/10.1007/s13205-019-1709-5>>.

EL-BROLOS, M. A.; STAINIER, D. Y. R. Genetic compensation: A phenomenon in search of mechanisms. **PLOS Genetics**, v. 13, n. 7, p. e1006780, 2017. ISSN 1553-7404. Disponível em: <<https://dx.plos.org/10.1371/journal.pgen.1006780>>.

FENG, Z.; ZHANG, B.; DING, W.; LIU, X.; YANG, D.-L.; WEI, P.; CAO, F.; ZHU, S.; ZHANG, F.; MAO, Y.; ZHU, J.-K. Efficient genome editing in plants using a CRISPR/Cas system. **Cell Research**, Nature Publishing Group, v. 23, n. 10, p. 1229–1232, 2013. Disponível em: <<https://doi.org/10.1038/cr.2013.114>>.

FORCE, A.; LYNCH, M.; PICKETT, F. B.; AMORES, A.; YAN, Y.-L.; POSTLETHWAIT, J. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. **Genetics**, v. 151, n. 4, p. 1531–1545, 1999. ISSN 1943-2631. Disponível em: <<https://doi.org/10.1093/genetics/151.4.1531>>.

GARNEAU, J. E.; DUPUIS, M.; VILLION, M.; ROMERO, D. A.; BARRANGOU, R.; BOYAVAL, P.; FREMAUX, C.; HORVATH, P.; MAGADÁN, A. H.; MOINEAU, S. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. **Nature**, v. 468, n. 7320, p. 67–71, 2010. ISSN 0028-0836, 1476-4687. Disponível em: <<https://www.nature.com/articles/nature09523>>.

GENOVESE, M. I.; LAJOLO, F. M. Fatores antinutricionais da soja. **Informe Agropecuário**, Belo Horizonte, v. 27, n. 230, p. 28–33, 2006. Disponível em: <<https://www.livrariaepamig.com.br/wp-content/uploads/2023/03/IA-230.pdf>>.

GILLMAN, J. D.; KIM, W.-S.; KRISHNAN, H. B. Identification of a New Soybean Kunitz Trypsin Inhibitor Mutation and Its Effect on BowmanBirk Protease Inhibitor Content in Soybean Seed. **Journal of Agricultural and Food Chemistry**, American Chemical Society, v. 63, n. 5, p. 1352–1359, 2015. ISSN 0021-8561. Disponível em: <<https://doi.org/10.1021/jf505220p>>.

GITLIN-DOMAGALSKA, A.; MACIEJEWSKA, A.; DĘBOWSKI, D. Bowman-Birk Inhibitors: Insights into Family of Multifunctional Proteins and Peptides with Potential Therapeutical Applications. **Pharmaceuticals**, Basel, Switzerland, v. 13, n. 12, p. 421, 2020. ISSN 1424-8247. Disponível em: <<https://doi.org/10.3390/ph13120421>>.

GOODSTEIN, D. M.; SHU, S.; HOWSON, R.; NEUPANE, R.; HAYES, R. D.; FAZO, J.; MITROS, T.; DIRKS, W.; HELLSTEN, U.; PUTNAM, N.; ROKHSAR, D. S. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, n. D1, p. D1178–D1186, 2012. ISSN 1362-4962, 0305-1048. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr944>>.

GU, C.; SONG, X.; ZHAO, L.; PAN, S.; QIN, G. Purification and characterization of Bowman-Birk trypsin inhibitor from soybean. **Journal of Food and Nutrition Research**, v. 2, n. 9, p. 505–512, 2014. Disponível em: <<https://doi.org/10.12691/jfnr-2-9-3>>.

HARTMAN, G. L.; WEST, E. D.; HERMAN, T. K. Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. **Food Security**, v. 3, n. 1, p. 5–17, 2011. ISSN 1876-4517, 1876-4525. Disponível em: <<http://link.springer.com/10.1007/s12571-010-0108-x>>.

HE, H.; LI, X.; KONG, X.; HUA, Y.; CHEN, Y. Heat-induced inactivation mechanism of soybean Bowman-Birk inhibitors. **Food Chemistry**, v. 232, p. 712–720, 2017. ISSN 0308-8146. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0308814617306325>>.

HELLEMANS, J.; MORTIER, G.; PAEPE, A. D.; SPELEMAN, F.; VANDESOMPELE, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. **Genome Biology**, v. 8, n. 2, p. R19, 2007. ISSN 1474-760X. Disponível em: <<https://genomebiology.biomedcentral.com/articles/10.1186/gb-2007-8-2-r19>>.

HIRAKURI, M. H.; LAZZAROTTO, J. J. O agronegócio da soja nos contextos mundial e brasileiro. Londrina: Embrapa Soja, 2014., p. 70 p., 2014. ISSN 2176-2937. Accepted: 2014-07-08T11:11:11Z. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/handle/doc/990000>>.

HONDA, D. E.; MARTINS, J. B. L.; VENTURA, M. M.; EYRILMEZ, S. M.; LEPŠÍK, M.; HOBZA, P.; PECINA, A.; FREITAS, S. M. de. Interface Interactions of the Bowman-Birk Inhibitor BTC1 in a Ternary Complex with Trypsin and Chymotrypsin Evaluated by Semiempirical Quantum Mechanical Calculations. **European Journal of Organic Chemistry**, v. 2018, n. 37, p. 5203–5211, 2018. ISSN 1099-0690. eprint: <https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/ejoc.201800754>. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/ejoc.201800754>>.

HOOD, E. E.; GELVIN, S. B.; MELCHERS, L. S.; HOEKEMA, A. New *Agrobacterium* helper plasmids for gene transfer to plants. **Transgenic Research**, v. 2, n. 4, p. 208–218, 1993. ISSN 0962-8819, 1573-9368. Disponível em: <<http://link.springer.com/10.1007/BF01977351>>.

HU, R.; FAN, C.; LI, H.; ZHANG, Q.; FU, Y.-F. Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. **BMC molecular biology**, v. 10, p. 93, 2009. ISSN 1471-2199. Disponível em: <<https://doi.org/10.1186/1471-2199-10-93>>.

IMPENS, L.; JACOBS, T. B.; NELISSEN, H.; INZÉ, D.; PAUWELS, L. Mini-Review: Transgenerational CRISPR/Cas9 Gene Editing in Plants. **Frontiers in Genome**

Editing, *Frontiers*, v. 4, 2022. ISSN 2673-3439. Disponível em: <<https://www.frontiersin.org/journals/genome-editing/articles/10.3389/fgeed.2022.825042/full>>.

JACOBS, T. B.; LAFAYETTE, P. R.; SCHMITZ, R. J.; PARROTT, W. A. Targeted genome modifications in soybean with CRISPR/Cas9. **BMC Biotechnology**, v. 15, n. 1, p. 16, 2015. ISSN 1472-6750. Disponível em: <<https://bmcbiotechnol.biomedcentral.com/articles/10.1186/s12896-015-0131-2>>.

JINEK, M.; CHYLINSKI, K.; FONFARA, I.; HAUER, M.; DOUDNA, J. A.; CHARPENTIER, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. **Science**, v. 337, n. 6096, p. 816–821, 2012. ISSN 0036-8075, 1095-9203. Disponível em: <<https://www.science.org/doi/10.1126/science.1225829>>.

JONES, S. I.; VODKIN, L. O. Using RNA-Seq to Profile Soybean Seed Development from Fertilization to Maturity. **PLOS ONE**, Public Library of Science, v. 8, n. 3, p. e59270, 2013. ISSN 1932-6203. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0059270>>.

JUMPER, J.; EVANS, R.; PRITZEL, A.; GREEN, T.; FIGURNOV, M.; RONNEBERGER, O.; TUNYASUVUNAKOOL, K.; BATES, R.; ŽÍDEK, A.; POTAPENKO, A.; BRIDGLAND, A.; MEYER, C.; KOHL, S. A. A.; BALLARD, A. J.; COWIE, A.; ROMERA-PAREDES, B.; NIKOLOV, S.; JAIN, R.; ADLER, J.; BACK, T.; PETERSEN, S.; REIMAN, D.; CLANCY, E.; ZIELINSKI, M.; STEINEGGER, M.; PACHOLSKA, M.; BERGHAMMER, T.; BODENSTEIN, S.; SILVER, D.; VINYALS, O.; SENIOR, A. W.; KAVUKCUOGLU, K.; KOHLI, P.; HASSABIS, D. Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, n. 7873, p. 583–589, 2021. ISSN 0028-0836, 1476-4687. Disponível em: <<https://www.nature.com/articles/s41586-021-03819-2>>.

KAHER, J. M.; KOLTUN, A.; HOSHINO, R. T.; GIROTTO, L.; SILVEIRA, C. A.; MARIN, S. R. R.; RECH, E. L.; NEPOMUCENO, A. L.; MERTZ-HENNING, L. M. CRISPR/Cas9-mediated inactivation of the soybean agglutinin Le1 gene to improve grain quality. **Frontiers in Plant Science**, *Frontiers*, v. 16, 2026. ISSN 1664-462X. Disponível em: <<https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2025.1737513/full>>.

KAHER, J. M.; SILVA, J. V. da; POLIZELI, S. R. d. A.; HOSHINO, R. T.; ROSA, J. da; FILHO, E. L. R.; NEPOMUCENO, A. L.; MERTZ-HENNING, L. M. Metabolic engineering of soybean for improving grain quality for animal consumption. **Frontiers in Plant Science**, v. 17, p. 1738805. ISSN 1664-462X. Disponível em: <<https://pmc.ncbi.nlm.nih.gov/articles/PMC12957170/>>.

KAMTHAN, A.; CHAUDHURI, A.; KAMTHAN, M.; DATTA, A. Genetically modified (GM) crops: milestones and new advances in crop improvement. **Theoretical and Applied Genetics**, v. 129, n. 9, p. 1639–1655, 2016. ISSN 0040-5752, 1432-2242. Disponível em: <<http://link.springer.com/10.1007/s00122-016-2747-6>>.

KASHIWAGI, K.; VIRGONA, N.; YAMADA, J.; SATO, A.; OTA, M.; YAZAWA, T.; YANO, T. Bowman-Birk protease inhibitor from soybeans enhances cisplatin-induced cytotoxicity in human mesothelioma cells. **Experimental and Therapeutic Medicine**, v. 2, n. 4, p. 719–724, 2011. ISSN 1792-0981, 1792-1015. Disponível em: <<https://www.spandidos-publications.com/10.3892/etm.2011.270>>.

KIM, W.; KIM, S.; KRISHNAN, H. B. Seed-Specific Silencing of Abundantly Expressed Soybean Bowman–Birk Protease Inhibitor Genes by RNAi Lowers Trypsin and Chymotrypsin Inhibitor Activities and Enhances Protein Digestibility. **International Journal of Molecular Sciences**, publisher, v. 26, n. 14, 2025. ISSN 1422-0067. Disponível em: <<https://www.mdpi.com/1422-0067/26/14/6943>>.

KIM, W.-S.; GILLMAN, J. D.; KIM, S.; LIU, J.; JANGA, M. R.; STUPAR, R. M.; KRISHNAN, H. B. Bowman–Birk Inhibitor Mutants of Soybean Generated by CRISPR-Cas9 Reveal Drastic Reductions in Trypsin and Chymotrypsin Inhibitor Activities. **International Journal of Molecular Sciences**, v. 25, n. 11, p. 5578, 2024. ISSN 1422-0067. Disponível em: <<https://www.mdpi.com/1422-0067/25/11/5578>>.

KLÜMPER, W.; QAIM, M. A Meta-Analysis of the Impacts of Genetically Modified Crops. **PLoS ONE**, v. 9, n. 11, p. e111629, 2014. ISSN 1932-6203. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0111629>>.

KOIDE, T.; IKENAKA, T. Studies on Soybean Trypsin Inhibitors: 3. Amino-Acid Sequence of the Carboxyl-Terminal Region and the Complete Amino-Acid Sequence of Soybean Trypsin Inhibitor (Kunitz). **European Journal of Biochemistry**, v. 32, n. 3, p. 417–431, 1973. ISSN 0014-2956, 1432-1033. Disponível em: <<https://febs.onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1973.tb02624.x>>.

KROWARSCHE, D.; CIERPICKI, T.; JELEN, F.; OTLEWSKI, J. Canonical protein inhibitors of serine proteases. **Cellular and Molecular Life Sciences CMLS**, v. 60, n. 11, p. 2427–2444, 2003. ISSN 1420-9071. Disponível em: <<https://doi.org/10.1007/s00018-003-3120-x>>.

KUMAR, D. M. Plant Protease Inhibitors: A Defense Mechanism Against Phytophagous Insects. **Journal for Research in Applied Sciences and Biotechnology**, v. 3, n. 1, p. 70–73, 2024. ISSN 2583-4053. Disponível em: <<https://jrasb.com/index.php/jrasb/article/view/430>>.

KUNITZ, M. CRYSTALLINE SOYBEAN TRYPSIN INHIBITOR. **Journal of General Physiology**, v. 30, n. 4, p. 291–310, 1947. ISSN 1540-7748, 0022-1295. Disponível em: <<https://rupress.org/jgp/article/30/4/291/12180/CRYSTALLINE-SOYBEAN-TRYPSIN-INHIBITOR-II-GENERAL>>.

LAJOLO, F. M.; GENOVESE, M. I. Nutritional significance of lectins and enzyme inhibitors from legumes. **Journal of Agricultural and Food Chemistry**, v. 50, n. 22, p. 6592–6598, 2002. Disponível em: <<https://pubs.acs.org/doi/10.1021/jf020191k>>.

LAMERS, J.; ZHANG, Y.; ZELM, E. van; LEONG, C. K.; MEYER, A. J.; ZEEUW, T. de; VERSTAPPEN, F.; VEEN, M.; DEOLU-AJAYI, A. O.; GOMMERS, C. M. M.; TESTERINK, C. Absciscic acid signaling gates salt-induced responses of plant roots. **Proceedings of the National Academy of Sciences**, Proceedings of the National Academy of Sciences, v. 122, n. 6, p. e2406373122, 2025. Disponível em: <<https://www.pnas.org/doi/10.1073/pnas.2406373122>>.

LASKOWSKI, M.; KATO, I. Protein Inhibitors of Proteinases. **Annual Review of Biochemistry**, v. 49, n. 1, p. 593–626, 1980. ISSN 0066-4154, 1545-4509. Disponível em: <<https://www.annualreviews.org/doi/10.1146/annurev.bi.49.070180.003113>>.

LASKOWSKI, M.; QASIM, M. What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? **Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology**, v. 1477, n. 1-2, p. 324–337, 2000. ISSN 01674838. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0167483899002848>>.

LESCOT, M. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. **Nucleic Acids Research**, v. 30, n. 1, p. 325–327, 2002. ISSN 13624962. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/30.1.325>>.

LETUNIC, I.; BORK, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. **Bioinformatics**, v. 23, n. 1, p. 127–128, 2007. ISSN 1367-4803. Disponível em: <<https://doi.org/10.1093/bioinformatics/btl529>>.

LI, C.; XU, M.; CAI, X.; HAN, Z.; SI, J.; CHEN, D. Jasmonate Signaling Pathway Modulates Plant Defense, Growth, and Their Trade-Offs. **International Journal of Molecular Sciences**, publisher, v. 23, n. 7, 2022. ISSN 1422-0067. Disponível em: <<https://www.mdpi.com/1422-0067/23/7/3945>>.

LIENER, I. E. Implications of antinutritional components in soybean foods. **Critical Reviews in Food Science and Nutrition**, v. 34, n. 1, p. 31–67, 1994. ISSN 1040-8398, 1549-7852. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/10408399409527649>>.

LIU, H.; DING, Y.; ZHOU, Y.; JIN, W.; XIE, K.; CHEN, L.-L. CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. **Molecular Plant**, v. 10, n. 3, p. 530–532, 2017. ISSN 16742052. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1674205217300047>>.

LIU, X.; ZHANG, Y.; FRANCIS, F.; CHENG, Y.; LIU, Y.; CHEN, J.; ZHANG, Y. Induced Bowman-Birk trypsin inhibitor TaBBTI-1 in wheat during poor-host interactions with aphids confers broad-spectrum resistance to insect herbivores. **International Journal of Biological Macromolecules**, v. 340, p. 150204, 2026. ISSN 0141-8130. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0141813026001303>>.

LOWDER, L. G.; ZHANG, D.; BALTES, N. J.; PAUL, J. W.; TANG, X.; ZHENG, X.; VOYTAS, D. F.; HSIEH, T.-F.; ZHANG, Y.; QI, Y. A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation. **Plant Physiology**, v. 169, n. 2, p. 971–985, 2015. ISSN 0032-0889, 1532-2548. Disponível em: <<https://academic.oup.com/plphys/article/169/2/971-985/6114041>>.

MALABARBA, J.; CHEVREAU, E.; DOUSSET, N.; VEILLET, F.; MOIZAN, J.; VERGNE, E. New Strategies to Overcome Present CRISPR/Cas9 Limitations in Apple and Pear: Efficient Dechimerization and Base Editing. **International Journal of Molecular Sciences**, Multidisciplinary Digital Publishing Institute, v. 22, n. 1, 2020. ISSN 1422-0067. Disponível em: <<https://www.mdpi.com/1422-0067/22/1/319>>.

MARTINČOVÁ, M.; SOUKUP, A. Twenty years of *AT-HOOK MOTIF NUCLEAR LOCALIZED (AHL)* gene family research – Their potential in crop improvement. **Current Plant Biology**, v. 42, p. 100460, 2025. ISSN 2214-6628. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2214662825000283>>.

MCDONALD, E. F.; JONES, T.; PLATE, L.; MEILER, J.; GULSEVIN, A. Benchmarking AlphaFold2 on peptide structure prediction. **Structure**, Elsevier, v. 31, n. 1, p. 111–119.e2, 2023. ISSN 0969-2126. Disponível em: <[https://www.cell.com/structure/abstract/S0969-2126\(22\)00479-8](https://www.cell.com/structure/abstract/S0969-2126(22)00479-8)>.

MELO, B. P. D.; LOURENÇO-TESSUTTI, I. T.; MORGANTE, C. V.; SANTOS, N. C.; PINHEIRO, L. B.; LINS, C. B. D. J.; SILVA, M. C. M.; MACEDO, L. L. P.; FONTES, E. P. B.; SA, M. F. Grossi-de. Soybean Embryonic Axis Transformation: Combining Biolistic and *Agrobacterium*-Mediated Protocols to Overcome Typical Complications of *In Vitro* Plant Regeneration. **Frontiers in Plant Science**, v. 11, p. 1228, 2020. ISSN 1664-462X. Disponível em: <<https://www.frontiersin.org/article/10.3389/fpls.2020.01228/full>>.

MESSINA, M. Soy and Health Update: Evaluation of the Clinical and Epidemiologic Literature. **Nutrients**, v. 8, n. 12, p. 754, 2016. ISSN 2072-6643. Disponível em: <<https://www.mdpi.com/2072-6643/8/12/754>>.

MOLINARI, H. B. C.; VIEIRA, L. R.; SILVA, N. V. e.; PRADO, G. S.; FILHO, J. H. L. (Ed.). **Tecnologia CRISPR na edição genômica de plantas: Biotecnologia aplicada à agricultura**. Brasília, DF: Embrapa, 2020. 161 p. Disponível em: <<http://www.alice.cnptia.embrapa.br/alice/handle/doc/1126157>>.

MOUGHAN, P. J.; LIM, W. X. J. Digestible indispensable amino acid score (DIAAS): 10 years on. **Frontiers in Nutrition**, v. 11, p. 1389719, 2024. ISSN 2296-861X. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fnut.2024.1389719/full>>.

MURAD, A. M.; VIANNA, G. R.; MACHADO, A. M.; CUNHA, N. B. da; COELHO, C. M.; LACERDA, V. A. M.; COELHO, M. C.; RECH, E. L. Mass spectrometry characterisation of fatty acids from metabolically engineered soybean seeds. **Analytical and Bioanalytical Chemistry**, v. 406, n. 12, p. 2873–2883, 2014. ISSN 1618-2650. Disponível em: <<https://doi.org/10.1007/s00216-014-7709-8>>.

MURASHIGE, T.; SKOOG, F. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. **Physiologia Plantarum**, v. 15, n. 3, p. 473–497, 1962. ISSN 1399-3054. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1399-3054.1962.tb08052.x>. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1399-3054.1962.tb08052.x>>.

MÉNDEZ, R.; LEPLAE, R.; MARIA, L. D.; WODAK, S. J. Assessment of blind predictions of protein–protein interactions: Current status of docking methods. **Proteins: Structure, Function, and Bioinformatics**, v. 52, n. 1, p. 51–67, 2003. ISSN 1097-0134. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.10393>. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.10393>>.

ODANI, S.; IKENAKA, T. Studies on Soybean Trypsin Inhibitors. **The Journal of Biochemistry**, v. 74, n. 4, p. 697–715, 1973. ISSN 1756-2651, 0021-924X. Disponível em: <<https://academic.oup.com/jb/article-lookup/doi/10.1093/oxfordjournals.jbchem.a130295>>.

OKEDIGBA, A. O.; ROSSO, M. L.; YU, D.; SHANG, C.; HUANG, H.; ZHANG, B.; CAPELLUTO, D. G. S. Comparative Binding Affinity Analysis of Soybean Meal Bowman-Birk and Kunitz Trypsin Inhibitors in Interactions with Animal Serine Proteases.

ACS Food Science & Technology, American Chemical Society, v. 3, n. 8, p. 1344–1352, 2023. Disponível em: <<https://doi.org/10.1021/acsfoodscitech.3c00158>>.

OTLEWSKI, J.; KROWARSCCH, D.; APOSTOLUK, W. Protein inhibitors of serine proteinases. **Acta Biochimica Polonica**, v. 46, n. 3, p. 531–565, 1999. ISSN 0001-527X. Disponível em: <https://doi.org/10.18388/abp.1999_4150>.

PORTEUS, M. H.; BALTIMORE, D. Chimeric Nucleases Stimulate Gene Targeting in Human Cells. **Science**, v. 300, n. 5620, p. 763–763, 2003. ISSN 0036-8075, 1095-9203. Disponível em: <<https://www.science.org/doi/10.1126/science.1078395>>.

QI, R.-F.; SONG, Z.-W.; CHI, C.-W. Structural Features and Molecular Evolution of Bowman-Birk Protease Inhibitors and their Potential Application. **Acta Biochimica et Biophysica Sinica**, Science Press, v. 37, n. 5, p. 283–292, 2005. ISSN 1672-9145. Disponível em: <<https://www.sciengine.com/10.1111/j.1745-7270.2005.00048.x>>.

RICHARDSON, M. The proteinase inhibitors of plants and micro-organisms. **Phytochemistry**, v. 16, n. 2, p. 159–169, 1977. ISSN 00319422. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0031942200867771>>.

RIZZO, G.; BARONI, L. Soy, Soy Foods and Their Role in Vegetarian Diets. **Nutrients**, v. 10, n. 1, p. 43, 2018. ISSN 2072-6643. Disponível em: <<https://www.mdpi.com/2072-6643/10/1/43>>.

RUFF, K. M.; PAPPU, R. V. AlphaFold and Implications for Intrinsically Disordered Proteins. **Journal of Molecular Biology**, v. 433, n. 20, p. 167208, 2021. ISSN 0022-2836. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022283621004411>>.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences**, v. 74, n. 12, p. 5463–5467, 1977. ISSN 0027-8424, 1091-6490. Disponível em: <<https://pnas.org/doi/full/10.1073/pnas.74.12.5463>>.

SCHAAFSMA, G. The Protein Digestibility-Corrected Amino Acid Score (PDCAAS)—a concept for describing protein quality in foods and food ingredients: a critical review. **Journal of AOAC International**, v. 88, n. 3, p. 988–994, 2005. ISSN 1060-3271. Disponível em: <<https://doi.org/10.1093/jaoac/88.3.988>>.

SCHMITTGEN, T. D.; LIVAK, K. J. Analyzing real-time PCR data by the comparative CT method. **Nature Protocols**, v. 3, n. 6, p. 1101–1108, 2008. ISSN 1754-2189, 1750-2799. Disponível em: <<https://www.nature.com/articles/nprot.2008.73>>.

SCHMUTZ, J.; CANNON, S. B.; SCHLUETER, J.; MA, J.; MITROS, T.; NELSON, W.; HYTEN, D. L.; SONG, Q.; THELEN, J. J.; CHENG, J.; XU, D.; HELLSTEN, U.; MAY, G. D.; YU, Y.; SAKURAI, T.; UMEZAWA, T.; BHATTACHARYYA, M. K.; SANDHU, D.; VALLIYODAN, B.; LINDQUIST, E.; PETO, M.; GRANT, D.; SHU, S.; GOODSTEIN, D.; BARRY, K.; FUTRELL-GRIGGS, M.; ABERNATHY, B.; DU, J.; TIAN, Z.; ZHU, L.; GILL, N.; JOSHI, T.; LIBAULT, M.; SETHURAMAN, A.; ZHANG, X.-C.; SHINOZAKI, K.; NGUYEN, H. T.; WING, R. A.; CREGAN, P.; SPECHT, J.; GRIMWOOD, J.; ROKHSAR, D.; STACEY, G.; SHOEMAKER, R. C.; JACKSON, S. A. Genome sequence of the palaeopolyploid soybean. **Nature**, v. 463, n. 7278, p. 178–183, 2010. ISSN 0028-0836, 1476-4687. Disponível em: <<https://www.nature.com/articles/nature08670>>.

SCHRÖDINGER, LLC. **The PyMOL Molecular Graphics System, Version 2.5.** 2021. Disponível em: <<https://www.pymol.org/>>.

SEVERIN, A. J.; WOODY, J. L.; BOLON, Y.-T.; JOSEPH, B.; DIERS, B. W.; FARMER, A. D.; MUEHLBAUER, G. J.; NELSON, R. T.; GRANT, D.; SPECHT, J. E.; GRAHAM, M. A.; CANNON, S. B.; MAY, G. D.; VANCE, C. P.; SHOEMAKER, R. C. RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome. **BMC Plant Biology**, v. 10, n. 1, p. 160, 2010. ISSN 1471-2229. Disponível em: <<http://bmcpplantbiol.biomedcentral.com/articles/10.1186/1471-2229-10-160>>.

SEZONOV, G.; JOSELEAU-PETIT, D.; D'ARI, R. *Escherichia coli* Physiology in Luria-Bertani Broth. **Journal of Bacteriology**, American Society for Microbiology, v. 189, n. 23, p. 8746–8749, 2007. Disponível em: <<https://journals.asm.org/doi/10.1128/jb.01368-07>>.

SREEDASYAM, A.; PLOTT, C.; HOSSAIN, M. S.; LOVELL, J.; GRIMWOOD, J.; JENKINS, J.; DAUM, C.; BARRY, K.; CARLSON, J.; SHU, S.; PHILLIPS, J.; AMIREBRAHIMI, M.; ZANE, M.; WANG, M.; GOODSTEIN, D.; HAAS, F.; HISS, M.; PERROUD, P.-F.; JAWDY, S.; YANG, Y.; HU, R.; JOHNSON, J.; KROPAT, J.; GALLAHER, S.; LIPZEN, A.; SHAKIROV, E.; WENG, X.; TORRES-JEREZ, I.; WEERS, B.; CONDE, D.; PAPPAS, M.; LIU, L.; MUCHLINSKI, A.; JIANG, H.; SHYU, C.; HUANG, P.; SEBASTIAN, J.; LAIBEN, C.; MEDLIN, A.; CAREY, S.; CARRELL, A.; CHEN, J.-G.; PERALES, M.; SWAMINATHAN, K.; ALLONA, I.; GRATAPAGLIA, D.; COOPER, E.; THOLL, D.; VOGEL, J.; WESTON, D. J.; YANG, X.; BRUTNELL, T.; KELLOGG, E.; BAXTER, I.; UDVARDI, M.; TANG, Y.; MOCKLER, T.; JUENGER, T.; MULLET, J.; RENSING, S.; TUSKAN, G.; MERCHANT, S.; STACEY, G.; SCHMUTZ, J. JGI Plant Gene Atlas: an updateable transcriptome resource to improve functional gene descriptions across the plant kingdom. **Nucleic Acids Research**, v. 51, n. 16, p. 8383–8401, 2023. ISSN 0305-1048. Disponível em: <<https://doi.org/10.1093/nar/gkad616>>.

STECHER, G.; SULESKI, M.; TAO, Q.; TAMURA, K.; KUMAR, S. MEGA 12.1: Cross-Platform Release for macOS and Linux Operating Systems. **Journal of Molecular Evolution**, v. 94, n. 1, p. 14–18, 2026. ISSN 1432-1432. Disponível em: <<https://doi.org/10.1007/s00239-025-10287-z>>.

TAN, S.; CAO, J.; LI, S.; LI, Z. Unraveling the Mechanistic Basis for Control of Seed Longevity. **Plants**, publisher, v. 14, n. 5, 2025. ISSN 2223-7747. Disponível em: <<https://www.mdpi.com/2223-7747/14/5/805>>.

TANG, Q.; WANG, X.; JIN, X.; PENG, J.; ZHANG, H.; WANG, Y. CRISPR/Cas Technology Revolutionizes Crop Breeding. **Plants**, Multidisciplinary Digital Publishing Institute, v. 12, n. 17, 2023. ISSN 2223-7747. Disponível em: <<https://www.mdpi.com/2223-7747/12/17/3119>>.

TEUFEL, F.; ARMENTEROS, J. J. A.; JOHANSEN, A. R.; GİSLASON, M. H.; PIHL, S. I.; TSIRIGOS, K. D.; WINTHER, O.; BRUNAK, S.; HEIJNE, G. V.; NIELSEN, H. SignalP 6.0 predicts all five types of signal peptides using protein language models. **Nature Biotechnology**, v. 40, n. 7, p. 1023–1025, 2022. ISSN 1087-0156, 1546-1696. Disponível em: <<https://www.nature.com/articles/s41587-021-01156-3>>.

TIAN, F.; YANG, D.-C.; MENG, Y.-Q.; JIN, J.; GAO, G. PlantRegMap: charting functional regulatory maps in plants. **Nucleic Acids Research**, v. 48, n. D1, p. D1104–D1113, 2020. ISSN 0305-1048. Disponível em: <<https://pmc.ncbi.nlm.nih.gov/articles/PMC7145545/>>.

VANGONE, A.; BONVIN, A. M. Contacts-based prediction of binding affinity in protein–protein complexes. **eLife**, v. 4, p. e07454, 2015. ISSN 2050-084X. Disponível em: <<https://elifesciences.org/articles/07454>>.

VERDIER, J.; THOMPSON, R. D. Transcriptional Regulation of Storage Protein Synthesis During Dicotyledon Seed Filling. **Plant and Cell Physiology**, v. 49, n. 9, p. 1263–1271, 2008. ISSN 0032-0781, 1471-9053. Disponível em: <<https://academic.oup.com/pcp/article-lookup/doi/10.1093/pcp/pcn116>>.

VIEIRA, L. R. Regulatory framework of genome editing in Brazil and worldwide. In: **CRISPR Technology in Plant Genome Editing: Biotechnology Applied to Agriculture**. Brasília, DF: Embrapa, 2022. p. 169–195. Disponível em: <<https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1131958/crispr-technology-in-plant-genome-editing-biotechnology-applied-to-agriculture>>.

VORSTER, J.; WESTHUIZEN, W. v. d.; PLESSIS, G. d.; MARAIS, D.; SPARVOLI, F.; COMINELLI, E.; CAMILLI, E.; FERRARI, M.; DONNE, C. L.; MARCONI, S.; LISCIANI, S.; LOSA, A.; SALA, T.; KUNERT, K. In order to lower the antinutritional activity of serine protease inhibitors, we need to understand their role in seed development. **Frontiers in Plant Science**, Frontiers, v. 14, 2023. ISSN 1664-462X. Disponível em: <<https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1252223/full>>.

VOSS, R.; ERMLER, U.; ESSEN, L.; WENZL, G.; KIM, Y.; FLECKER, P. Crystal Structure of the Bifunctional Soybean Bowman-Birk Inhibitor at 0.28-nm Resolution: Structural Peculiarities in a Folded Protein Conformation. **European Journal of Biochemistry**, v. 242, n. 1, p. 122–131, 1996. ISSN 0014-2956, 1432-1033. Disponível em: <<https://febs.onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1996.0122r.x>>.

WANG, J.; KUANG, H.; ZHANG, Z.; YANG, Y.; YAN, L.; ZHANG, M.; SONG, S.; GUAN, Y. Generation of seed lipoxygenase-free soybean using CRISPR-Cas9. **The Crop Journal**, v. 8, n. 3, p. 432–439, 2020. ISSN 2214-5141. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2214514119301539>>.

WANG, L.; ZHANG, J.; LI, H.; ZHANG, G.; HU, D.; ZHANG, D.; XU, X.; YANG, Y.; HUANG, Z. Genome-Wide Identification of the Phytocyanin Gene Family and Its Potential Function in Salt Stress in Soybean (*Glycine max* (L.) Merr.). **Agronomy**, Multidisciplinary Digital Publishing Institute, v. 13, n. 10, 2023. ISSN 2073-4395. Disponível em: <<https://www.mdpi.com/2073-4395/13/10/2484>>.

WANG, Z.; SHEA, Z.; ROSSO, L.; SHANG, C.; LI, J.; BEWICK, P.; LI, Q.; ZHAO, B.; ZHANG, B. Development of new mutant alleles and markers for KTI1 and KTI3 via CRISPR/Cas9-mediated mutagenesis to reduce trypsin inhibitor content and activity in soybean seeds. **Frontiers in Plant Science**, Frontiers, v. 14, 2023. ISSN 1664-462X. Disponível em: <<https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1111680/full>>.

WARNER, K.; GUPTA, M. Potato chip quality and frying oil stability of high oleic acid soybean oil. **Journal of Food Science**, v. 70, n. 6, p. s395–s400, 2005. Disponível em: <<https://ift.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2621.2005.tb11462.x>>.

WASSERMAN, W. W.; SANDELIN, A. Applied bioinformatics for the identification of regulatory elements. **Nature Reviews. Genetics**, v. 5, n. 4, p. 276–287, 2004. ISSN 1471-0056.

WATERHOUSE, A. M.; PROCTER, J. B.; MARTIN, D. M. A.; CLAMP, M.; BARTON, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. **Bioinformatics**, v. 25, n. 9, p. 1189–1191, 2009. ISSN 1367-4811, 1367-4803. Disponível em: <<https://academic.oup.com/bioinformatics/article/25/9/1189/203460>>.

WOOD, A. J.; LO, T.-W.; ZEITLER, B.; PICKLE, C. S.; RALSTON, E. J.; LEE, A. H.; AMORA, R.; MILLER, J. C.; LEUNG, E.; MENG, X.; ZHANG, L.; REBAR, E. J.; GREGORY, P. D.; URNOV, F. D.; MEYER, B. J. Targeted Genome Editing Across Species Using ZFNs and TALENs. **Science**, v. 333, n. 6040, p. 307–307, 2011. ISSN 0036-8075, 1095-9203. Disponível em: <<https://www.science.org/doi/10.1126/science.1207773>>.

XIE, Y.; RAVET, K.; PEARCE, S. Extensive structural variation in the Bowman-Birk inhibitor family in common wheat (*Triticum aestivum* L.). **BMC Genomics**, v. 22, n. 1, p. 218, 2021. ISSN 1471-2164. Disponível em: <<https://doi.org/10.1186/s12864-021-07475-8>>.

YAN, Y.; TAO, H.; HE, J.; HUANG, S.-Y. The HDock server for integrated protein–protein docking. **Nature Protocols**, v. 15, n. 5, p. 1829–1852, 2020. ISSN 1754-2189, 1750-2799. Disponível em: <<https://www.nature.com/articles/s41596-020-0312-x>>.

YAO, D.; ZHOU, J.; ZHANG, A.; WANG, J.; LIU, Y.; WANG, L.; PI, W.; LI, Z.; YUE, W.; CAI, J.; LIU, H.; HAO, W.; QU, X. Advances in CRISPR/Cas9-based research related to soybean [*Glycine max* (Linn.) Merr.] molecular breeding. **Frontiers in Plant Science**, Frontiers, v. 14, 2023. ISSN 1664-462X. Disponível em: <<https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1247707/full>>.

YU, L.; ZHANG, H.; GUAN, R.; LI, Y.; GUO, Y.; QIU, L. Genome-Wide Tissue-Specific Genes Identification for Novel Tissue-Specific Promoters Discovery in Soybean. **Genes**, Multidisciplinary Digital Publishing Institute, v. 14, n. 6, p. 1150, 2023. ISSN 2073-4425. Disponível em: <<https://www.mdpi.com/2073-4425/14/6/1150>>.

ZARATE, S.; CIMADORI, I.; JONES, M. S.; ROCA, M. M.; BARNHILL-DILLING, S. K. Assessing agricultural gene editing regulation in Latin America: an analysis of how policy windows and policy entrepreneurs shape agricultural gene editing regulatory regimes. **Frontiers in Bioengineering and Biotechnology**, v. 11, p. 1209308, 2023. ISSN 2296-4185. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fbioe.2023.1209308/full>>.

ZHANG, Y.; SKOLNICK, J. TM-align: a protein structure alignment algorithm based on the TM-score. **Nucleic Acids Research**, v. 33, n. 7, p. 2302–2309, 2005. ISSN 0305-1048. Disponível em: <<https://doi.org/10.1093/nar/gki524>>.

ZHAO, S.; FERNALD, R. D. Comprehensive Algorithm for Quantitative Real-Time Polymerase Chain Reaction. **Journal of computational biology : a journal of**

computational molecular cell biology, v. 12, n. 8, p. 1047–1064, 2005. ISSN 1066-5277. Disponível em: <<https://pmc.ncbi.nlm.nih.gov/articles/PMC2716216/>>.

ZHENG, N.; LI, T.; DITTMAN, J. D.; SU, J.; LI, R.; GASSMANN, W.; PENG, D.; WHITHAM, S. A.; LIU, S.; YANG, B. CRISPR/Cas9-Based Gene Editing Using Egg Cell-Specific Promoters in *Arabidopsis* and Soybean. **Frontiers in Plant Science**, Frontiers, v. 11, 2020. ISSN 1664-462X. Disponível em: <<https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2020.00800/full>>.

MATERIAL SUPLEMENTAR

Relação dos Loci Analisados

A Tabela S1 lista os identificadores exatos (Locus) dos 15 transcritos selecionados no genoma de referência de *Glycine max* (Wm82.a4.v1) que compõem a família BBI.

Tabela S1 – Lista de loci dos 15 genes selecionados de *Glycine max* analisados quanto ao perfil de expressão basal.

Número	Locus (<i>GeneID</i>)
1	<i>Glyma.09G158500</i>
2	<i>Glyma.09G158600</i>
3	<i>Glyma.09G158700</i>
4	<i>Glyma.09G158800</i>
5	<i>Glyma.09G158900</i>
6	<i>Glyma.09G260301</i>
7	<i>Glyma.09G260400</i>
8	<i>Glyma.09G260500</i>
9	<i>Glyma.14G117600</i>
10	<i>Glyma.14G117700</i>
11	<i>Glyma.16G208900</i>
12	<i>Glyma.18G231400</i>
13	<i>Glyma.18G231500</i>
14	<i>Glyma.18G231600</i>
15	<i>Glyma.18G231700</i>

Tabela S2 – Perfil de expressão gênica de 15 transcritos de *Glycine max*. A intensidade da cor azul representa os níveis de transcrição em $\text{Log}_2(\text{FPKM} + 1)$.

Amostra / Condição	<i>Glyma.09G158500</i>	<i>Glyma.09G158600</i>	<i>Glyma.09G158700</i>	<i>Glyma.09G158800</i>	<i>Glyma.09G158900</i>	<i>Glyma.09G260301</i>	<i>Glyma.09G260400</i>	<i>Glyma.09G260500</i>	<i>Glyma.14G117600</i>	<i>Glyma.14G117700</i>	<i>Glyma.16G208900</i>	<i>Glyma.18G231400</i>	<i>Glyma.18G231500</i>	<i>Glyma.18G231600</i>	<i>Glyma.18G231700</i>
flower.open	0.0	0.0	0.0	0.0	0.0	3.6	0.0	0.0	0.0	0.0	0.0	1.9	7.8	1.3	0.2
flower.unopen	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	2.6	2.0	1.9	0.4
lateral_root.standard	0.1	0.0	0.0	0.1	0.2	0.0	0.0	0.0	0.0	0.2	0.6	6.8	7.0	0.0	0.0
leaf.1130AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	1.0
leaf.1130PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.9
leaf.130AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.8
leaf.130PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.7
leaf.330AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.9
leaf.330PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	1.0

Continua na próxima página...

Tabela S2 – Continuação da página anterior

Amostra / Condição	<i>Glyma.09G158500</i>	<i>Glyma.09G158600</i>	<i>Glyma.09G158700</i>	<i>Glyma.09G158800</i>	<i>Glyma.09G158900</i>	<i>Glyma.09G260301</i>	<i>Glyma.09G260400</i>	<i>Glyma.09G260500</i>	<i>Glyma.14G117600</i>	<i>Glyma.14G117700</i>	<i>Glyma.16G208900</i>	<i>Glyma.18G231400</i>	<i>Glyma.18G231500</i>	<i>Glyma.18G231600</i>	<i>Glyma.18G231700</i>
leaf.530AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.8
leaf.530PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	1.2
leaf.730AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.4	0.1	1.7
leaf.730PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.5
leaf.930AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	1.0
leaf.930PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.4
leaf.ammonium	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.3	1.1	0.2	1.1
leaf.nitrate	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	1.2	0.8	0.1	0.6
leaf.standard	0.0	0.0	0.0	0.0	0.0	2.1	0.0	0.0	0.0	0.0	0.4	0.1	2.2	1.0	5.4
leaf.symbiotic_condition	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.3	0.6	0.1	2.5
leaf.trifoliolate_stage_1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	6.1	0.0	0.0
leaf.trifoliolate_stage_2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.4	0.0	1.6
leaf.trifoliolate_stage_3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.4	0.0	0.8
leaf.trifoliolate_stage_4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.1	0.0	1.4
leaf.trifoliolate_stage_5	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.3	0.9
leaf.urea	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.5	0.9	0.1	1.4
nodules.1130AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.8	0.0	0.0	0.0	4.8	3.6	0.2	0.0
nodules.1130PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.6	0.0	0.0	0.0	2.9	2.1	0.1	0.0
nodules.130AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.2	0.0	0.0	0.0	3.5	3.0	0.2	0.0
nodules.130PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.9	0.0	0.0	0.0	4.4	3.6	0.2	0.0
nodules.330AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.7	0.0	0.0	0.0	3.6	3.2	0.2	0.0
nodules.330PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.7	0.0	0.0	0.0	4.6	4.3	0.2	0.0
nodules.530AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.1	0.0	0.0	0.0	4.0	3.4	0.2	0.0
nodules.530PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.4	0.0	0.0	0.0	4.5	4.0	0.2	0.0
nodules.730AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.2	0.0	0.0	0.0	4.2	3.4	0.2	0.0
nodules.730PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.6	0.0	0.0	0.0	4.6	3.5	0.3	0.0
nodules.930AM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.7	0.0	0.0	0.0	4.1	3.0	0.1	0.0
nodules.930PM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.3	0.0	0.0	0.0	3.5	3.2	0.2	0.0

Continua na próxima página...

Tabela S2 – Continuação da página anterior

Amostra / Condição	<i>Glyma.09G158500</i>	<i>Glyma.09G158600</i>	<i>Glyma.09G158700</i>	<i>Glyma.09G158800</i>	<i>Glyma.09G158900</i>	<i>Glyma.09G260301</i>	<i>Glyma.09G260400</i>	<i>Glyma.09G260500</i>	<i>Glyma.14G117600</i>	<i>Glyma.14G117700</i>	<i>Glyma.16G208900</i>	<i>Glyma.18G231400</i>	<i>Glyma.18G231500</i>	<i>Glyma.18G231600</i>	<i>Glyma.18G231700</i>
nodules.symbiotic_condition	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.3	0.0	0.0	0.0	2.3	5.7	0.1	0.0
root_tip.standard	0.0	0.0	0.0	0.1	0.3	0.0	0.2	0.0	0.0	0.5	2.5	7.8	7.8	0.0	0.0
root.ammonium	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.6	7.4	0.0	0.0
root.nitrate	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	5.4	6.3	0.0	0.0
root.standard	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.2	1.5	3.9	4.3	0.0	0.0
root.symbiotic_condition	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.3	6.8	0.0	0.0
root.urea	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.7	6.0	0.0	0.0
seed.seed_s1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.9	0.0	0.5	0.7	0.0
seed.seed_s2	8.1	3.3	5.1	5.7	9.1	1.1	0.0	0.0	0.1	10.4	13.2	2.6	4.8	1.6	0.0
seed.seed_s3	7.3	4.0	6.2	6.6	8.2	1.0	0.1	0.0	0.9	9.3	12.9	2.2	1.8	1.6	0.0
seed.seed_s4	7.4	4.7	7.5	7.3	8.3	0.1	0.0	0.0	1.2	9.3	12.8	3.2	2.3	1.6	0.0
seed.seed_s5	7.8	4.9	7.3	7.5	8.7	0.1	0.0	0.0	0.8	9.7	12.7	2.7	1.9	0.9	0.0
seed.seed_s6	8.3	5.7	8.3	8.3	9.1	0.0	0.0	0.0	1.8	10.2	12.5	2.6	2.0	0.4	0.0
seed.seed_s7	0.9	0.4	1.0	1.6	1.3	0.0	0.0	0.0	0.1	2.0	3.0	4.2	6.5	0.1	0.1
seed.seed_s8	0.3	0.1	0.5	0.3	0.3	0.0	0.0	0.0	0.0	0.5	2.1	1.5	5.3	0.0	0.0
seed.seed_s9	0.2	0.2	0.9	0.4	0.5	0.0	0.0	0.0	0.0	0.5	2.3	2.3	5.7	0.0	0.2
shoot_tip.1130AM	0.0	0.0	0.0	0.0	0.0	3.5	0.0	0.0	0.0	0.0	0.0	1.0	0.4	1.6	0.0
shoot_tip.1130PM	0.0	0.0	0.0	0.0	0.0	4.2	0.0	1.4	0.0	0.0	0.0	2.3	0.5	2.3	0.0
shoot_tip.130AM	0.0	0.0	0.0	0.0	0.0	4.1	0.0	0.0	0.0	0.0	0.0	1.8	0.3	2.1	0.0
shoot_tip.130PM	0.0	0.0	0.0	0.0	0.0	3.5	0.0	0.0	0.0	0.0	0.0	1.6	0.3	2.1	0.0
shoot_tip.330AM	0.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0	1.8	0.2	1.7	0.0
shoot_tip.330PM	0.0	0.0	0.0	0.0	0.0	3.2	0.0	0.0	0.0	0.0	0.0	1.6	0.3	1.6	0.0
shoot_tip.530AM	0.0	0.0	0.0	0.0	0.0	4.1	0.0	0.0	0.0	0.0	0.0	1.9	0.2	2.1	0.0
shoot_tip.530PM	0.0	0.0	0.0	0.0	0.0	3.3	0.0	0.0	0.0	0.0	0.0	1.3	0.3	2.0	0.0
shoot_tip.730AM	0.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0	1.9	0.3	2.0	0.0
shoot_tip.730PM	0.0	0.0	0.0	0.0	0.0	3.6	0.0	0.0	0.0	0.0	0.0	0.8	0.3	1.8	0.0
shoot_tip.930AM	0.0	0.0	0.0	0.0	0.0	3.6	0.0	0.0	0.0	0.0	0.0	1.0	0.2	1.7	0.0
shoot_tip.930PM	0.0	0.0	0.0	0.0	0.0	3.9	0.0	0.0	0.0	0.0	0.0	1.3	0.1	1.7	0.0

Continua na próxima página...

Tabela S2 – Continuação da página anterior

Amostra / Condição	<i>Glyma.09G158500</i>	<i>Glyma.09G158600</i>	<i>Glyma.09G158700</i>	<i>Glyma.09G158800</i>	<i>Glyma.09G158900</i>	<i>Glyma.09G260301</i>	<i>Glyma.09G260400</i>	<i>Glyma.09G260500</i>	<i>Glyma.14G117600</i>	<i>Glyma.14G117700</i>	<i>Glyma.16G208900</i>	<i>Glyma.18G231400</i>	<i>Glyma.18G231500</i>	<i>Glyma.18G231600</i>	<i>Glyma.18G231700</i>
shoot_tip.standard	0.0	0.0	0.0	0.0	0.0	4.9	0.0	0.0	0.0	0.0	0.3	2.8	2.3	2.7	0.1
stem.standard	0.0	0.0	0.0	0.0	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.7	0.0