



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

JOÃO MATHEUS KA FER

**PERFIL TRANSCRICIONAL DE SEMENTES DE SOJA COM  
COLORAÇÃO DE TEGUMENTO CONTRASTANTE**

---

Londrina  
2022

JOÃO MATHEUS KAFER

**PERFIL TRANSCRICIONAL DE SEMENTES DE SOJA COM  
COLORAÇÃO DE TEGUMENTO CONTRASTANTE**

Dissertação apresentada ao Programa de Pós-graduação em Biotecnologia da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Mestre.

Orientadora: Dra. Liliane Marcia Mertz Henning

Co-orientadora: Dra. Mayla Daiane Correa Molinari

Londrina  
2022

**Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL**

K11      Kafer, João Matheus.  
            Perfil transcricional de sementes de soja com coloração de tegumento contrastante / João Matheus Kafer. - Londrina, 2022.  
            78 f. : il.

            Orientador: Liliane Marcia Mertz Henning.  
            Coorientador: Mayla Molinari.  
            Dissertação (Mestrado em Biotecnologia) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Biotecnologia, 2022.  
            Inclui bibliografia.

            1. Glycine max - Tese. 2. Coloração de tegumento - Tese. 3. Armazenamento - Tese. 4. RNA-seq - Tese. I. Mertz Henning, Liliane Marcia . II. Molinari, Mayla . III. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Biotecnologia. IV. Título.

CDU 66

JOÃO MATHEUS KAFER

**PERFIL TRANSCRICIONAL DE SEMENTES DE SOJA COM  
COLORAÇÃO DE TEGUMENTO CONTRASTANTE**

Dissertação apresentada ao Programa de Pós-graduação em Biotecnologia da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Mestre.

**BANCA EXAMINADORA**

---

Orientadora: Dra. Liliane Marcia Mertz Henning  
Universidade Estadual de Londrina – UEL

---

Dra. Mayla Daiane Correa Molinari  
Universidade Estadual de Londrina – UEL

---

Dra. Julia Abati  
Universidade Estadual de Londrina – UEL

---

Dr. Américo Viana  
Embrapa Soja

Londrina, 25 de fevereiro de 2022

## AGRADECIMENTOS

Em primeiro lugar, agradeço a minha família por todo amor e carinho que me permitiram chegar até aqui;

Agradeço ao meu namorado Julio por ter compartilhado todos os momentos do mestrado, nos dias ruins e bons;

Agradeço a todos os amigos que estiveram presentes em toda a jornada do mestrado e a tornaram mais leve;

Agradeço a minha orientadora a Dra. Liliane M. Mertz Henning por todo apoio durante a pesquisa, especialmente, durante o período da pandemia;

Agradeço também, ao Dr. Fernando Henning pela oportunidade de participar do projeto e por toda ajuda, especialmente na obtenção do material vegetal;

Agradeço a minha Co-orientadora, Dra. Mayla Molinari por toda a dedicação no ensino das análises de bioinformática;

Agradeço a Dra. Silvana Rockenbach Marin por todo auxílio na análise de RT-qPCR;

Agradeço também a equipe do Laboratório de Biotecnologia Vegetal da Embrapa Soja, por toda a ajuda no projeto;

Agradeço a Universidade Estadual de Londrina (UEL) pela oportunidade do Mestrado;

Ao Programa de Pós-Graduação em Biotecnologia da UEL, estendendo o agradecimento a todos os Professores;

A Embrapa por disponibilizar a estrutura física e o apoio técnico para desenvolvimento deste trabalho;

Agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudo e apoio a pesquisa.

KAFER, João Matheus. **Perfil transcricional de sementes de soja com coloração de tegumento contrastante**. 2022. 78 f. Dissertação (Mestrado em Biotecnologia) - Universidade Estadual de Londrina, Londrina, 2022.

## RESUMO

A soja é a principal fonte de proteína vegetal que é utilizada para vários fins, mas seu principal uso é como ração para proteína animal. A soja obteve muito sucesso devido a sua adaptação a várias regiões do País, resultado do melhoramento genético da cultura. No entanto, ao longo dos anos, o foco principal dos programas de melhoramento foi voltado para o aumento em produtividade, o que acarretou redução da diversidade genética e perda de outras características de interesse. Entretanto, é possível encontrar diversidade genética na espécie, inclusive para a coloração do tegumento dos grãos de soja, que podem variar entre o amarelo, preto, marrom, verde e bicolor. Cultivares de tegumento preto foram avaliadas como benéficas tanto quando utilizada como semente, quanto quando utilizada como grão. Entender os mecanismos genéticos envolvidos no controle dessa característica, pode ser útil para o desenho de estratégias biotecnológicas visando identificar genes relacionados a melhoria da qualidade de sementes e grãos para inúmeras aplicações. Dessa forma, este trabalho visa identificar genes associados a sementes de soja com coloração de tegumento contrastante, com foco na identificação de genes relacionados a qualidade de sementes. Os resultados do transcriptoma foram obtidos de bibliotecas de RNAseq de sementes das cultivares BRSMG 715A (tegumento preto), BRS 413 RR e DM 6563 IPRO (tegumento amarelo), a partir de duas condições: sementes recém colhidas e sementes armazenadas em câmara fria por seis meses. Foram encontrados 318 genes presentes em todas as cultivares. Foram identificados vias superexpressas na cultivar de tegumento preto e nas cultivares de tegumento amarelo. A partir das análises *in silico* dos transcriptomas, foram selecionados sete genes para validação por meio de PCR em tempo real: *ACSI*, *ACSF3*, *CYP90A1*, *CYP710A1*, *HCT*, *CBL*, *SAHH*. A cultivar de tegumento preto BRSMG 715A foi submetida ao cálculo de expressão relativa e foi possível observar que os genes apresentaram padrão semelhante tanto para a análise de RNA-Seq quanto para análise de RT-qPCR. O gene *CBL* apresentou maior expressão na cultivar de tegumento preto quando comparada as outras duas cultivares de tegumento amarelo. Além disso, os genes *CYP90A1*, *CYP710A1* e *HCT* foram superexpressos na cultivar de tegumento preto em relação a cultivar de tegumento amarelo BRS 413 RR. Esses genes estão relacionados a maior teor de lignina na semente, indução de crescimento, biossíntese de aminoácidos e estabilidade de membrana e podem contribuir para maior concentração de lignina e maior qualidade de sementes observada na cultivar BRSMG 715A.

**Palavras-chave:** *Glycine max*; lignina; armazenamento; RNA-Seq, qualidade de sementes.

KAFER, João Matheus. **Transcriptional profile of soybean seeds with contrasting seed coat color**. 2022. 78 p. Dissertation (Master Degree – Biotechnology) - Universidade Estadual de Londrina, Londrina, 2022.

## ABSTRACT

Soybean is the main source of vegetable protein which is used for various purposes, but its main use is as animal protein feed. Soybean was very successful due to its adaptation to various regions due to the action of genetic improvement. However, this resulted in reduced genetic diversity and loss of other traits. However, it is possible to find diversity in the color of the seed coat that can vary from black, brown, green and bicolor. Black seed coat cultivars were evaluated as having benefits both when used as seed and as grain. Understanding the genetic mechanisms involved in the control of this trait can be useful for the design of biotechnological strategies that aim to identify genes related to improvement in seed and grain quality. Thus, this work aims to identify genes that are associated with soybean seeds with contrasting seed coat color. The transcriptome results were validated by real-time PCR, in the cultivars BRS 715A (black seed coat), BRS 413 RR and DM 6563 IPRO (yellow seed coat). We found 318 genes present in all cultivars and in two treatments: freshly harvested and cold chamber for 6 months. Overexpressed pathways were identified in the seed coat cultivar and in the yellow seed coat cultivars. From the *in silico* analysis of the transcriptomes, the following genes were selected: *ACS1*, *ACSF3*, *CYP90A1*, *CYP710A1*, *HCT*, *CBL*, *SAHH* validated by RT-qPCR in the freshly harvested. Regarding the calibration, the cultivar BRSMG 715A was used, it is possible to observe the genes presented a similar pattern for both the RNA-Seq analysis and RT-qPCR. The *CBL* gene was induced in the BRSMG715A cultivar when compared to the two yellow seed coat cultivars. In addition, *CYP90A1*, *CYP710A1* and *HCT* genes were overexpressed in the black tegument cultivar in relation to the BRS 413 RR cultivar. The genes are related to higher lignin content in the seed, growth induction, amino acid biosynthesis and membrane stability and they may contribute to higher lignin concentration and higher seed quality observed in the BRSMG 715A cultivar.

**Keywords:** *Glycine max*; seed coat color; lignin; RNA-Seq; storage.

## LISTA DE FIGURAS

- Figura 1** – Diversidade genética da soja verificada nas características morfológicas e de coloração do tegumento do grão. Foto: Antônio Neto (Arquivo Embrapa Soja). ..... 19
- Figura 2** – Silenciamento pós-transcricional via RNA de interferência (siRNA) do gene CHS no tegumento da soja, devido a ação das repetições invertidas presentes nos genes CHS1, 3 e 4 que serão clusterizadas (Hernandez-Garcia, 2019). ..... 27
- Figura 3** – Fenótipos dos tegumentos governados pelos quatro alelos do locus I em soja. O alelo recessivo pode levar a expressão da coloração marrom ou preta (Hernandez-Garcia, 2019). ..... 28
- Figura 4** – Via resumida dos fenilpropanóides e origem dos produtos lignina, flavonóides e antocianina em plantas (Zhang et al., 2021). ..... 30
- Figura 5** – Via de biossíntese de molignóis em angiospermas (Weng & Chapple, 2010). ..... 32
- Figura 6** – Polimerização oxidativa de lignina pelo acoplamento dos três álcoois hidroxicinamil (Weng & Chapple, 2010). ..... 33
- Figura 7** – Rota geral de formação dos isoflavonóides (Wang, 2010). ..... 34
- Figura 8** – Estrutura básica das antocianinas com representantes mais presentes a esquerda e outros de menor importância a direita(a). Seguimento da rota dos fenilpropanóides que dá origem ao grupo das antocianinas(b). ..... 36
- Figura 9** – Diagramas de Venn considerando genes com expressão relativa (A) e absoluta (B) significativas. Em A, Genes com expressão relativa significativa na cultivar de tegumento preto BRSMG 715A em comparação com as cultivares amarelas BRS 413 RR e DM 6563 IPRO por tratamento (T0 - recém-colhidas; CF câmara-fria). Em B, 318 genes com expressão absoluta significativa comuns a T0 e CF (TPM  $\geq 3$ ). Somente genes com expressão relativa significativa (Log<sub>2</sub>FC  $\leq -1$  e  $\geq 1$ ) comuns a T0 e CF (318) em (A) foram avaliados quanto sua expressão absoluta (TPM  $\geq 3$ ) (B). ..... 44
- Figura 10** – Vias metabólicas enriquecidas na cultivar BRSMG 715A em T0 e CF (A); e nas cultivares BRS 413 RR e DM 6563 IPRO em T0 e CF (B). ..... 49

- Figura 11** – Heatmap contendo valores da expressão absoluta em TPM. Barras representam a expressão relativa em Log2FC dos sete genes alvo. Os genes selecionados são todos regulados positivamente (up-regulados) na cultivar de tegumento preto BRSMG 715A. Os quadros em verde mostram os valores de TPM utilizados para calcular o Log2FC de cada gene (barra azul)..... 50
- Figura 12** – Expressão relativa dos genes HCT, CBL SAHH, ACSF3, CYP90A1, CYP710A1 e ACS1 em sementes de soja inteiras recém-colhidas. Valores positivos representam expressão na cultivar de tegumento preto BRSMG 715A quando comparada com as cultivares de tegumento amarelo BRS 413 RR (A) e DM 6563 IPRO (B). Para RT-qPCR ( $2^{\Delta\Delta Ct}$ ) a análise estatística foi realizada por meio do teste T ( $p \leq 0.1$ ). Para a análise de RNA-Seq (Log2FC)  $FDR \leq 0.05$ . Asteriscos(\*) representam valores com diferença significativa..... 51
- Figura 13** – Esquematização dos genes induzidos na cultivar de tegumento preto BRSMG 715A preto em relação as cultivares de tegumento amarelo BRS 413 RR e DM 6563 IPRO com suas respectivas contribuições para aumento na qualidade e longevidade de sementes. .... 57
- Figura 14** – Teor de lignina na semente de soja da cultivar de tegumento preto BRSMG 715A e das cultivares amarelas BRS 413 RR e DM 6563 IPRO, letras distintas demonstram diferença significativa pelo teste de Tukey ( $p \leq 0.05$ ). .... 62

## LISTA DE TABELAS

<b>Tabela 1</b> – Lista de primers utilizados na quantificação da expressão relativa dos alvos via RT-qPCR .....	42
<b>Tabela 2</b> – Genes que apresentam expressão apenas na cultivar de tegumento preto BRSMG 715A .....	46
<b>Tabela 3</b> – Avaliação do número de cópias, ortólogos e domínio proteico entre soja e Arabidopsis. ....	60

## LISTA DE ABREVIATURAS E SIGLAS

<i>ACSI</i>	1-aminocyclopropane-1-carboxylate synthase
<i>ACSF3</i>	Malonyl-CoA/methylmalonyl-CoA synthetase
<i>AGO5</i>	Proteína Argonaute, complexo RISC (silenciamento gênico)
<i>CBL</i>	Cystathionine beta-lyase
<i>CHS</i>	Chalcone Synthase
<i>CYP710A</i>	c-3 Oxidase
<i>CYP90A1</i>	Sterol 22-desaturase
<i>F3'H</i>	Flavonoid hidroxylase
<i>HCT</i>	Shikimate O-hydroxycinnamoyltransferase
<i>MYB R2R3</i>	Fator de transcrição MYB
ORFs	Open Reading Frame
<i>SAHH</i>	Adenosylhomocysteinase

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	13
<b>2</b>	<b>OBJETIVOS</b> .....	16
2.1	OBJETIVO GERAL.....	16
2.2	OBJETIVOS ESPECÍFICOS .....	16
<b>3</b>	<b>REVISÃO DE LITERATURA</b> .....	17
3.1	ASPECTOS GERAIS DA CULTURA DA SOJA .....	17
3.2	FATORES QUE INFLUENCIAM NA QUALIDADE DAS SEMENTES E GRÃOS DE SOJA. ....	19
3.2.1	Qualidade Fisiológica De Sementes De Soja .....	20
3.2.2	Qualidade De Grãos De Soja.....	22
3.3	INOVAÇÃO E DESENVOLVIMENTO DE NOVOS PRODUTOS À BASE DE SOJA.....	24
3.4	HISTÓRICO BASE DO MELHORAMENTO GENÉTICO DA SOJA NO BRASIL .....	24
3.5	VARIAÇÕES NA COLORAÇÃO DO TEGUMENTO DA SEMENTE DE SOJA .....	26
3.5.1	Qualidade de Sementes e sua Relação com a Coloração do Tegumento.....	28
3.6	ROTA DOS FENILPROPANÓIDES .....	29
3.6.1	Lignina.....	30
3.6.2	Isoflavonas.....	33
3.6.3	Antocianinas .....	35
3.7	CARACTERÍSTICAS TECNOLÓGICAS DA CULTIVAR BRSMG 715A .....	37
3.8	Rna-seq COMO FERRAMENTA PARA IDENTIFICAÇÃO DE GENES .....	37
<b>4</b>	<b>MATERIAL E MÉTODOS</b> .....	39
4.1	MATERIAL BIOLÓGICO E EXPERIMENTO EM CASA DE VEGETAÇÃO PARA COLETA DE AMOSTRA PARA RNA-Seq e RT-qPCR .....	39
4.2	ANÁLISE DE RNA-Seq .....	40
4.2.1	Análise De Bioinformática .....	40
4.3	ANÁLISE DE RT-qPCR .....	41
4.3.1	Primers Dos Genes Alvo Para Validação Da Expressão Relativa Por RT-qPCR.....	42
4.4	ANÁLISE BIOQUÍMICA.....	43

4.4.1	Análise De Teor De Lignina.....	43
4.4.2	Análise Estatística.....	43
<b>5</b>	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>44</b>
5.1	EXPRESSÃO DIFERENCIAL.....	44
5.2	VIAS METABÓLICAS ENRIQUECIDAS.....	46
5.3	EXPRESSÃO ABSOLUTA E RELATIVA .....	50
5.3.1	Genes Envolvidos Na Biossíntese De Aminoácidos .....	52
5.3.2	Genes Envolvidos Na Biossíntese De Fitorreguladores Vegetais .....	54
5.3.3	Genes Envolvidos Na Biossíntese De Ácidos Graxos.....	56
5.3.4	Genes Envolvidos Na Biossíntese De Lignina .....	56
5.4	AVALIAÇÃO DO NÚMERO DE CÓPIAS E ORTÓLOGOS EM ARABDOPSIS.....	58
5.5	AVALIAÇÃO DO TEOR DE LIGNINA NA SEMENTE .....	61
<b>6</b>	<b>CONCLUSÕES.....</b>	<b>63</b>
<b>7</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>64</b>
	<b>REFERÊNCIAS.....</b>	<b>65</b>

## 1. INTRODUÇÃO

A soja (*Glycine max* Merrill.) é uma das principais commodities agrícolas em âmbito mundial. É uma cultura de grande importância econômica com produção crescente. Os principais países produtores são o Brasil, Estados Unidos, Argentina, China e Índia (USDA, 2019, FAO, 2020). A soja apresenta em sua composição uma das maiores fontes de proteína, (aproximadamente 40%) e no quesito teor de óleo, encontra-se em segundo lugar (aproximadamente 20%), ficando atrás apenas da cultura do amendoim (Medic; Atkinson; Hurburgh, 2014, Sharma et al., 2014, Pagano; Misanrari, 2016).

No Brasil, a soja é utilizada principalmente na fabricação de rações animais, mas podem ser empregadas na alimentação humana, na indústria química para fabricação de cosméticos, fibras, na produção de biodiesel, entre outros. Devido a essa versatilidade em sua aplicação, a produção brasileira de soja cresce a cada ano. Na safra 2020/2021, houve um crescimento de 2,7% na área plantada de soja brasileira em relação à safra anterior, passando de aproximadamente 36,95 milhões hectares (ha) para 38,92 milhões de ha. Na safra 2021/2022 a produção brasileira de soja deve ultrapassar 142 milhões de toneladas, 3,4% superior em relação à safra anterior (CONAB, 2021).

Um dos principais fatores que contribuíram para o sucesso da cultura da soja no Brasil foi o desenvolvimento de cultivares adaptadas às diversas regiões de cultivo no país. No entanto, estudos evidenciam que apesar do grande número de cultivares de soja disponíveis, há uma baixa variabilidade genética entre os genótipos, em razão principalmente da estreita base genética do germoplasma brasileiro (Gwinner et al., 2017). Apesar disso, é possível encontrar variabilidade na espécie para diversas características, incluindo a coloração do tegumento da semente, que pode variar entre amarelo, verde, preto, marrom e bicolor (Song et al., 2016).

Apesar da maioria das cultivares comerciais de soja apresentarem tegumento de coloração amarela, a soja de tegumento preta apresenta propriedades que a tornam atrativa, tanto para utilização como semente, quanto para grão. Foi identificado que a concentração de compostos fenólicos como as antocianinas, carotenoides e ligninas são mais elevados em sementes de soja de tegumento preto quando comparadas à soja de tegumento amarelo (Xu & Chang, 2008, Bahry et al., 2015).

Assim, quando utilizada como grãos, a soja de tegumento preto torna-se uma alternativa interessante para a alimentação humana, sendo considerada fonte de vários nutrientes essenciais (Medic; Atkinson; Hurburgh, 2014, Keffe; Bianchi; Sharman,

2015). Além disso, apresenta uma grande variedade de substâncias fitoquímicas, destacando os isoflavonóides, os ácidos fenólicos, saponinas e triterpenóides que beneficiam a saúde humana (Cho et al., 2012). Sementes de soja com tegumento preto podem apresentar elevados níveis de antocianina (antioxidantes) que podem apresentar muitos benefícios a saúde como na prevenção de doenças como o câncer (Astadi; Paice, 2011, Wu et al., 2017). Além de atuarem como antioxidantes, apresentam características anti-inflamatórias, nefroprotetivas, antidiabéticas, anti-infertilidade, anti-obesidade, anti-ártrico, neuroprotetores, anti-hiperlipidêmico, anti-catarata e cicatrizante (Ganesan; Xu, 2017).

Adicionalmente, no que diz respeito à qualidade das sementes, já foi observado que genótipos com níveis elevados de lignina no tegumento conferem maior espessura e, conseqüentemente, maior resistência do tecido aos possíveis danos físicos em função da umidade, danos mecânicos e/ou presença de patógenos (Menezes et al., 2009, Carvalho et al., 2015, Mertz-Henning et al., 2015, Abati et al., 2021, Abati et al., 2022, Kuchlan et al., 2018). Outro fator relevante, é que geralmente, sementes de soja com tegumento de coloração preta apresentam melhor qualidade fisiológica em relação a soja de tegumento amarelo. Além disso, sementes com tegumento preto estão associadas a melhor longevidade durante o armazenamento (Kuchlan; Dadlani; Samuel, 2010, Liu et al., 2017, Adsul; Chimote; Deshmukh, 2018, Abati et al., 2021).

A pigmentação do tegumento da soja é uma característica complexa sendo governada por cinco loci (*I*, *R*, *T*, *WI* e *O*) que interagem entre si (Song et al., 2016). Apesar destes cinco loci serem os principais e mais relevantes para coloração do tegumento, estudos mostraram que a cor verde do tegumento envolve diferentes padrões de segregação de outros loci como o *G1*, *G2* e *G3* (Reese; Boerma, 1989). Mecanismos de silenciamento gênico pós-transcricional mediado por RNA interferente (siRNA - *small interfering RNA*) também estão envolvidos no controle da característica (Tuteja et al., 2009).

É estimado que muitas outras moléculas benéficas estejam presentes e associadas com o tegumento preto da cultivar de soja, entretanto, precisam ser exploradas em larga escala, a fim de identificar genes promissores para o melhoramento da cultura e/ou ainda identificar oportunidades no desenvolvimento de novos produtos e agregar valor ao produto.

Diante do exposto, o estudo do transcriptoma de sementes de soja com tegumento de coloração contrastante contribuirá para o avanço do conhecimento acerca

dos mecanismos genético-moleculares que conferem características superiores as sementes e grãos de coloração preta.

## **2. OBJETIVOS**

### 2.1 OBJETIVO GERAL

O objetivo deste trabalho foi identificar diferenças no transcriptoma de sementes de soja com tegumento de coloração contrastante.

### 2.2 OBJETIVOS ESPECÍFICOS

Identificar por meio de RNA-Seq diferenças entre o transcriptoma de cultivares de soja de tegumento de coloração preta e amarela.

Prospectar genes que se diferenciam na cultivar preta que possam estar associados a melhor qualidade de sementes e grãos de soja.

Caracterizar as sementes de soja preta e amarela dos genótipos estudados quanto ao teor e lignina e correlacionar estes resultados com os dados transcricionais de cada cultivar.

Validar os genes mais promissores por meio de PCR quantitativo em tempo real (RT-qPCR).

### 3. REVISÃO DE LITERATURA

#### 3.1 ASPECTOS GERAIS DA CULTURA DA SOJA

A soja (*Glycine max* Merrill.) é uma leguminosa importante que representa uma das principais fontes de proteína e óleo para produção animal, bem como para o consumo humano. A maioria dos pesquisadores concordam que a origem do cultivo da soja é na China, onde a soja selvagem anual (*Glycine soja*), o parente ancestral da soja cultivada (*G. max*), é amplamente distribuído (Kim et al., 2012, Sedivy; Wu; Hanzawa, 2017).

O processo de domesticação se deu por cruzamento natural e seleção artificial pelos chineses, que a selecionavam para o aumento no tamanho de grãos e legumes, altura de planta reduzida e espessamento da haste (Kim et al., 2012). Embora o local exato de origem da soja seja desconhecido, vários locais na China foram atribuídos como origem por meio de descobertas arqueológicas e de evidências moleculares. Entre eles destacam-se o nordeste da China, a região em torno do rio Huanghe e o vale Yangtze (Kim et al., 2012, Sedivy; Wu; Hanzawa, 2017).

O grão possuía grande importância para a segurança alimentar na China, sendo considerado sagrado durante os anos de 2883 e 2838 AC em conjunto com o arroz, trigo, milho e cevada (Hymowitz, 2008, Kim et al., 2012). Era e ainda é matéria prima essencial para produção de vários produtos, como o tofu (Sedivy; Wu; Hanzawa, 2017). O cultivo de soja espalhou-se na Ásia lentamente tendo sido primeiramente introduzida na Coreia e Japão (Jeoung et al., 2018). No ocidente, o grão foi introduzido através das grandes navegações entre o século XV e XVI (Specht et al., 2014, Sedivy; Wu; Hanzawa, 2017).

Sua adoção no ocidente como planta de importância para a agricultura começou no início do século XX, nos Estados Unidos, principalmente como planta forrageira, e posteriormente para o cultivo do grão, que superou a produção de forragem, tornando-se o principal uso da soja até o presente (Hymowitz, 2008, Anderson et al., 2019). No Brasil, a soja foi introduzida na Bahia em 1882, estudada como planta forrageira na época (Oliveira, 2016). Entretanto, o principal marco ocorreu em 1901 com a introdução da soja pelos cultivos na Estação Agropecuária de Campinas e a distribuição de sementes a agricultores paulistas. A grande expansão da soja no Brasil se deu nos anos de 1970, devido à grande demanda internacional pelo grão, e ao país possuir a vantagem do

escoamento da safra em período diferenciado a outros países, permitindo a venda do produto em maior cotação (Martinelli et al., 2017).

A partir deste momento o país investiu em melhoramento genético para adaptar a soja às condições climáticas brasileiras, e dessa forma, permitiu que a cultura fosse cultivada em regiões de baixas latitudes, entre o trópico de capricórnio e a linha do equador (Martinelli et al., 2017). Em virtude do investimento na adaptação e melhoramento da soja, atualmente a cultura é cultivada praticamente em todo território nacional, desde as altas latitudes presentes no Sul e até as áreas tropicais no Norte do país (Specht et al., 2014, Todeschini et al., 2019).

Por se tratar de uma das principais culturas produzidas mundialmente, a soja é responsável pelo fornecimento de mais de um quarto da proteína total para produção de diversos alimentos e rações animais (Hartman; West; Herman, 2011, Liu et al., 2020). Segundo a FAO (2020), os países que apresentam maior média de produção são Brasil, Estados Unidos da América e Argentina, juntos representam mais de 80% da produção mundial do grão.

No Brasil, na safra de 2020/2021 houve um crescimento de 2,7% na área plantada de soja brasileira em relação à safra anterior, passando de aproximadamente 36,95 milhões hectares (ha) para 38,92 milhões de ha. Na safra 2021/2022 a produção brasileira de soja deve ultrapassar 142 milhões de toneladas representando aumento de 3,4% em relação ao período anterior (CONAB, 2021). No país, as principais regiões produtoras são a centro-oeste e a sul. Na região centro-oeste, em 2021, ocorreu um incremento na área cultivada, de 4,3% em relação ao verificado no plantio passado, atingindo um 18.337,6 mil ha. Na região Sul, um incremento de 1,6% em relação à safra anterior foi observado (CONAB, 2021).

Com relação a produtividade da soja brasileira, ela foi influenciada positivamente pelo melhoramento genético e pelo aprimoramento das práticas agrícolas. Comparativamente, entre as décadas de 1980 e 1990 a produtividade da soja brasileira se mantinha abaixo de 2000 kg/ha, nas décadas de 2000 a 2010, a produtividade da soja continuou crescente, até o ano safra 2020/2021 onde a produtividade da soja brasileira foi de 3527,49 kg/ha, maior valor registrado até então. A safra 2021/2022 sofreu redução de produtividade devido a fatores climáticos que causaram a redução da produtividade para 3091 kg/ha (CONAB, 2022).

Visto a importância mundial do grão é imperativo identificar características de interesse em cultivares e, assim, aprimorar o melhoramento genético da soja. No entanto,

a base genética das variedades de soja cultivadas atualmente é estreita, devido a seleção de atributos relacionados a alta produtividade (Gwinner et al., 2017). Consequentemente, outras características interessantes como qualidade de semente, compostos fenólicos, teor de proteína e óleo acabaram sendo negligenciadas. Todavia, é possível explorar a diversidade genética da soja prospectando genes em cultivares selvagens e espécies próximas. A diversidade observada na morfologia e coloração dos grãos de soja mostra que é possível ainda identificar variabilidade genética bastante interessante ou ampla na espécie (Figura 1).

**Figura 1** - Diversidade genética da soja verificada nas características morfológicas e de coloração do tegumento do grão. Foto: Antônio Neto (Arquivo Embrapa Soja).



### 3.2 FATORES QUE INFLUENCIAM NA QUALIDADE DAS SEMENTES E GRÃOS DE SOJA

A soja possui grande importância econômica e, de modo geral, é utilizada como grãos, para processamento e consumo direto, já como sementes são destinadas para a semeadura da próxima safra. Ambos, sementes e grãos são influenciados pelas etapas de produção, colheita, beneficiamento, recepção e secagem e armazenamento da soja de formas distintas. Além disso, as diferentes características intrínsecas dos genótipos influenciam a qualidade de sementes e grãos de soja (Fraça-Neto 2018, Mertz-Henning et al., 2018, Sarkar et al., 2021). O conceito de qualidade de sementes pode ser definido como o conjunto atuante de fatores fisiológicos, genéticos, sanitários e físicos (Fraça-Neto, 2009). A qualidade fisiológica de sementes é um atributo imprescindível para o estabelecimento do estande total de plantas vigorosas, que chegarão ao estágio

reprodutivo com alta produtividade, devendo possuir germinação e vigor elevados (Han & Yang, 2015).

A qualidade fisiológica de sementes é afetada pelo tempo e condições de armazenamento, portanto ambos podem reduzir a germinação e vigor das sementes. Embora a deterioração da semente seja um processo natural, aumento na temperatura e umidade podem acentuá-la. Assim como a presença de patógenos, como fungos e insetos que reduzem significativamente a viabilidade e vigor da semente durante o armazenamento (Krzyzowski; França-Neto; Henning, 2018). Apesar de todos os atributos da qualidade apresentarem importância equivalente e serem fundamentais para o adequado desempenho agrônomo das culturas, nesse trabalho deu-se o foco para o atributo fisiológico.

### 3.2.1 Qualidade fisiológica de sementes de soja

A manutenção da qualidade de semente de soja é imprescindível para a formação de um estande saudável e com plântulas vigorosas que por fim serão essenciais para maior produtividade da cultura (França-Neto et al., 2016).

Durante a etapa de produção da semente, pode ocorrer o processo conhecido como deterioração da semente por umidade. Isso ocorre antes e após de se atingir a maturidade fisiológica antes da colheita. A exposição da semente a ciclo alternados de umidade alta e baixa, associado a elevadas temperaturas, resultará na deterioração da semente (Huth et al., 2016). Esse problema, constitui um dos mais relevantes para a manutenção da qualidade fisiológica da semente (França-Neto et al., 2018). Vários trabalhos evidenciaram redução da qualidade fisiológica de semente por danos oriundos deste processo (Forti et al., 2010, Moreano et al., 2011, Forti et al., 2013). Além disso, a infecção por fungos de campo como *Cercospora kikuchii*, *Phomopsis* spp e *Fusarium* spp podem agravar a deterioração, contribuindo para diminuição do vigor e germinação (França-Neto et al., 2018, Turner et al., 2020, Chang et al., 2020). Esse processo de deterioração da semente pode ser agravado durante a secagem, beneficiamento e armazenamento. Dessa forma, diminuindo ainda mais a qualidade fisiológica inicial da semente (França-Neto et al., 2018).

Ainda durante a etapa de produção, a soja está suscetível a várias intempéries climáticas como a ocorrência de veranicos associados a elevadas temperaturas durante a fase de enchimento do grão, causando enrugamento e diminuindo a qualidade da semente.

Ademais, a soja pode sofrer estresses ambientais (bióticos e abióticos), estes ocasionam a morte prematura da planta, maturação forçada, redução da produtividade e alteração no tegumento da semente como a coloração esverdeada (França-Neto et al., 2018, Sarkar et al., 2021). As sementes esverdeadas possuem vigor e germinação inferiores quando comparadas às sementes normais. Além disso, outra causa de deterioração pode ser identificada após o ataque de percevejos. Quando se alimentam da semente, eles a inoculam com a levedura *Nematospora coryli* Peglion, quando infectada a semente sofre necroses, que podem resultar na perda de germinação e vigor. Em campos de produção de sementes esses insetos devem ser monitorados e, em caso positivo, controlados imediatamente (França-Neto et al., 2018).

Já durante a colheita, predominam danos mecânicos, que podem ser formas caracterizados pelo conteúdo de água na semente durante o impacto: Em sementes secas (teor de umidade abaixo de 12%), apresentaram danos mecânicos imediatos, como fissuras, rachaduras e quebras. Por outro lado, sementes úmidas (teor de umidade acima de 14%) apresentam danos latentes, como amassamentos e abrasões (França-Neto, 2018, Moreano et al., 2011). Para ambos os casos, é essencial um bom manejo da colhedora, de forma que os mecanismos de trilha estejam ajustados corretamente.

Após a colheita, a semente passa pela recepção, que consiste na entrada na Unidade de beneficiamento de sementes (UBS) pelas moegas e em seguida pela máquina de pré-limpeza, essas etapas são importantes para remoção das impurezas grosseiras e as menores que a semente. Posteriormente, caso as sementes cheguem a UBS com mais de 12,5% de umidade, é realizado o processo de secagem, tal operação é necessário para reduzir a umidade inicial de forma adequada, para as etapas seguintes. Durante esse procedimento, a temperatura deve ser inferior a 40 °C e a umidade relativa deve superior que 35%, excedentes a estes requisitos acarretam dano térmico na semente. Ainda, a semente pode estar suscetível a danos mecânicos durante a movimentação da semente no secador, caso as condições não sejam adequadas (França-Neto et al., 2016).

Após a recepção e secagem, a semente passa pelo beneficiamento que visa remover contaminantes como materiais estranhos (vagens, ramos, torrões e insetos), sementes de outras culturas e de erva-daninhas. Ademais, o processo é necessário para: classificar a semente por tamanho; melhorar a qualidade do lote pela remoção de semente danificada e deteriorada; aplicar fungicidas e inseticidas à semente, quando necessários; e para embalar adequadamente a semente para a sua comercialização (França-Neto et al., 2016).

Após o beneficiamento, a semente de soja é armazenada, até a utilização para a próxima safra. Durante esse período é essencial que as condições de armazenamento sejam ideais para manter a qualidade fisiológica que foi obtida durante o processo de produção. Para isso, os fatores mais importantes para manter a qualidade da semente de soja são a temperatura e umidade do ar do ambiente e qualidade inicial do material de armazenamento. Geralmente, durante o armazenamento nos armazéns de sementes, as sementes são mantidas sobre condições ideais de temperatura (10-15 °C) e umidade relativa do ar (60-70%), em big bags semipermeáveis que podem ser de diferentes materiais (Mbofung et al., 2013, Smaniotto et al., 2014, Ferreira et al., 2019). Entretanto, durante o processo de transporte e semeadura, as sementes podem ser expostas a condições não adequadas, sem controle de temperatura e umidade por períodos prolongados até o momento da semeadura (Coradi et al., 2020).

A qualidade da semente pode ainda estar relacionada aos fatores genéticos da cultivar. Maior longevidade foi associada a características como maior teor de lignina e menor permeabilidade do tegumento (Adsul; Chimote; Deshmuk, 2018). Ademais, para a manutenção da qualidade fisiológica é necessária a atividade de várias enzimas antioxidantes como superóxido dismutase e peroxidase responsáveis por reduzir o dano causado por espécies reativas de oxigênio, sendo que a atividade dessas enzimas, geralmente é diferente entre os genótipos (Carvalho et al., 2011). Inúmeros trabalhos avaliaram genótipos com coloração de tegumento preto e o caracterizaram com maior qualidade fisiológica durante o armazenamento em função da presença de diferentes compostos como antocianinas e lignina (Bahry et al., 2016, Bahry et al., 2018, Abati, 2021). Segundo Huth et al. (2016) e Castro et al. (2019) danos causados por estresse oxidativo em sementes de soja, em decorrência da chuva e umidade, é menor em sementes com maiores teores de lignina no tegumento, pois também apresentam maior atividade de enzimas antioxidantes como catalase e a álcool desidrogenase.

### 3.2.2 Qualidade de grãos de soja

Assim como na semente, os grãos são influenciados pelas etapas de produção a campo, colheita, secagem, beneficiamento e armazenamento. O grão de soja é avaliado de acordo com as suas características tecnológicas, entre elas, destacam-se a umidade, óleo e proteína. Os dois últimos são componentes de alto valor comercial que serão utilizados para compor os produtos derivados da soja (Mandarino et al., 2018). Além

disso, destaca-se que os grãos são classificados comercialmente conforme a IN número 11 de 2007 (MAPA), em que se verifica os defeitos dos grãos de soja, e estabelecem o limite de comercialização e os descontos de cada item excedente da normalidade. Dessa forma, é possível identificar grãos de qualidade exigida pelo mercado (Lorini et al., 2016).

Com relação aos danos no grão, ele pode ser deteriorado pela umidade, assim como na semente, durante o processo de pré-colheita, o dano causado por chuvas resulta em grãos com maior umidade durante o armazenamento, que promove a proliferação dos fungos como *Aspergillus* spp e/ou *Penicillium* spp e acarreta a degradação do grão (Mandarino et al., 2018). Para evitar esse tipo de danos nos grãos, o processo de secagem deve começar o quanto antes para reduzir a umidade inicial dos grãos.

Além disso, podem sofrer danos após infestação de percevejos, esse dano é caracterizado por ocasionar grãos chochos e redução de produtividade, reduzindo o valor econômico atribuído ao produto. Os danos causados por percevejos podem ser controlados pelo uso de Manejo Integrado de Pragas (MIP). Ademais, durante a etapa de produção, a morte prematura das plantas resulta em grãos esverdeados que possuem qualidade inferior, assim como nas sementes (França-Neto et al., 2016).

Os danos causados durante as etapas de produção vão ser determinantes para as qualidades tecnológicas do grão de soja e de outras leguminosas (Sarkar et al., 2021). O teor de proteína do grão é o mais alto entre as culturas agrícolas, possuindo concentrações elevadas de aminoácidos essenciais, com exceção de metionina e cisteína (Medic, Artkinson, Hurburgh, 2014, Gorissen et al., 2018). A proteína da soja é composta por dois tipos de proteína de armazenamento: a glicinina que representa cerca de 60% das proteínas totais da soja; e a  $\beta$ -conglucina que representa os 40% restante do total das proteínas de armazenamento no grão (Wijewardana et al., 2019).

A concentração de óleo no grão é a segunda maior entre as culturas agrícolas, e sua qualidade é avaliada em função do índice de acidez e peróxidos, que indicam rancidez hidrolítica e oxidativa, respectivamente (Mandarino et al., 2018). A acidez é indicativa da qualidade da matéria prima e é influenciada pela maturação, armazenagem, ação enzimática, qualidade dos grãos e processo de extração, considerada ideal entre 0,3 e 0,5% (Hwang et al., 2014). Ambas as características são influenciadas por diversos fatores como: local e condições de cultivo, cultivar, armazenagem e danos nos grãos (Hu; Wiatrak, 2012, Mandarino et al., 2018, Ziegler et al., 2018, Mertz-Henning et al., 2018, Taher et al., 2019). O processo de acúmulo de proteína e óleo nos grãos de soja ocorre entre 10-12 e 15-25 dias após o florescimento, respectivamente. Esse processo é contínuo

até 70 dias após o florescimento, o que determina as etapas mais críticas para a concentração destes compostos na soja (Poeta et al., 2014, Wijewardana et al., 2019, Tamagno et al., 2022).

### 3.3 INOVAÇÃO E DESENVOLVIMENTO DE NOVOS PRODUTOS À BASE DE SOJA

A soja é utilizada para o desenvolvimento de vários produtos, inicialmente, voltados para a indústria alimentícia como óleo de soja, farinhas, molho de soja, leite de soja e tofu (Dukariya; Singh, 2020). Atualmente, a soja recebeu mais atenção como fonte potencial para o desenvolvimento de novos produtos, devido as suas características como teores elevados de compostos fenólicos, proteínas e ácidos graxos saudáveis que são importantes para a saúde humana (Gandhi, 2009, Sanjukta; Rai, 2016, Qin et al., 2022). Ademais, o consumo de proteína vegetal vem crescendo como substituto à proteína animal, devido aos benefícios a saúde, em especial a carne vermelha e produtos processados, que estão associados com ácidos graxos saturados e maior risco a doenças cardiovasculares (Alamu et al., 2017, Qin et al., 2022). Além disso, fatores ambientais também estão envolvidos como o aumento do efeito estufa oriundo de animais de rebanho (Qin et al., 2022).

Tendo em vista estes aspectos, a soja apresenta amplo potencial para ser uma das principais culturas no que diz respeito ao desenvolvimento de novos produtos. Sua diversidade genética permite explorar diferentes cultivares que podem ser melhoradas para vários fins. Vários trabalhos associaram cultivares de soja com tegumento preto como sendo superior a cultivares de tegumento amarelo com relação ao desenvolvimento de produtos alimentícios, principalmente, pelo maior teor de compostos fenólicos que estão presentes nestas cultivares (Khosravi; Razavi, 2018, Hou et al., 2020, Heish et al., 2020, Lee, 2021).

### 3.4 HISTÓRICO BASE DO MELHORAMENTO GENÉTICO DA SOJA NO BRASIL

Após ser disseminada do oriente para o ocidente, a soja sofreu adaptação para diversos ambientes devido suas características climáticas e geográficas como temperatura e latitude, respectivamente (Song et al., 2013). Assim, as espécies crioulas (*landraces*) mais adaptadas para cada região se tornaram predominantes, estreitando a base genética regional ao longo dos anos (Song et al., 2013). Este estreitamento histórico na diversidade

da cultura é intensificado ainda mais pela seleção de linhagens altamente produtivas nos programas de melhoramento genético atuais (Mikel et al., 2010, Zhou et al., 2015, Hegstad et al., 2019). Além disso, características cleistogâmicas da soja (fecundação ocorre antes da abertura floral) podem ter reduzido a homogeneidade e variação genômica, característica que a tornou mais sensível ao estreitamento genético durante o processo de domesticação (Valliyodan et al., 2016, Hegstad et al., 2019).

Segundo Smith et al. (2015), a redução de diversidade genética diminui o potencial de ganhos genéticos, que pode representar um aumento na susceptibilidade a estresses bióticos e na redução da adaptabilidade a mudanças ambientais. Portanto, um dos principais objetivos dos programas de melhoramento modernos deve ser também voltado para a preservação e enriquecimento da diversidade genética e não somente a produtividade da cultura (Gwinner et al., 2017). De acordo com Mikel et al. (2010), a diversidade genética na soja pode ser melhorada pelo aumento da mistura de germoplasmas exóticos e distintos nos programas de melhoramento enquanto a produtividade é mantida pela reciclagem de cultivares elite.

Atualmente, vários países são considerados centros de diversidade por meio de grandes bancos de germoplasma, com cultivares adaptadas às diferentes condições edafoclimáticas. As coleções de germoplasma incluem cultivares primitivas, crioulas (*landraces*), espécies selvagens próximas a planta cultivada, estoques genéticos, linhagens recombinantes e híbridos (Kim et al., 2012). De acordo com Genesys, plataforma que agrupa os diferentes acessos de plantas pelo mundo e permite aos pesquisadores encontrar material nos bancos de germoplasma registrados, estão registrados 61,688 acessos de soja e outras espécies selvagens do gênero *Glycine* (Genesys, 2021).

O banco de germoplasma brasileiro mantido pela Embrapa possui, aproximadamente, 55 mil acessos do gênero *Glycine*, que consistem em introduções de plantas, linhagens nacionais melhoradas e cultivares adaptadas a áreas tropicais e subtropicais (Landgraf, 2020).

Cultivares de soja com coloração de tegumento preto, representam 150 acessos na base Genesys (<https://www.genesys-pgr.org/a/v2IZDVB1WIW>). Destes, 41 são mantidos na coleção de germoplasma brasileiro da Embrapa. O germoplasma mantido no Brasil possui grande redundância com a coleção preservada pelo USDA, e recentemente houve aumento da diversidade de acessos no germoplasma brasileiro (Smýkal et al., 2014; Landgraf, 2020).

### 3.5 VARIAÇÕES NA COLORAÇÃO DO TEGUMENTO DA SEMENTE DE SOJA

A pigmentação do tegumento é uma característica moderadamente complexa controlada por 5 *loci* principais, identificados pela genética clássica (*I*, *R*, *T*, *WI* e *O*). Entre eles os *loci* *I*, *R* e *T* estão envolvidos na via de pigmentação baseada em flavonoides, enquanto os *loci* *WI* e *O*, apenas influenciam a pigmentação sobre um *background* de alelos recessivos *i* e *r* ou *i* e *t*, respectivamente (Song et al., 2016). Além desses *loci*, a pigmentação da semente pode sofrer influência dos *loci* *G1*, *G2* e *G3* que causa a coloração verde no tegumento (Reese; Boerma, 1989, Wang et al., 2018). Adicionalmente, verificou-se o que o *locus* *KI* também pode resultar em alterações espaciais na coloração da semente (Cho; Jones; Vodkin, 2017).

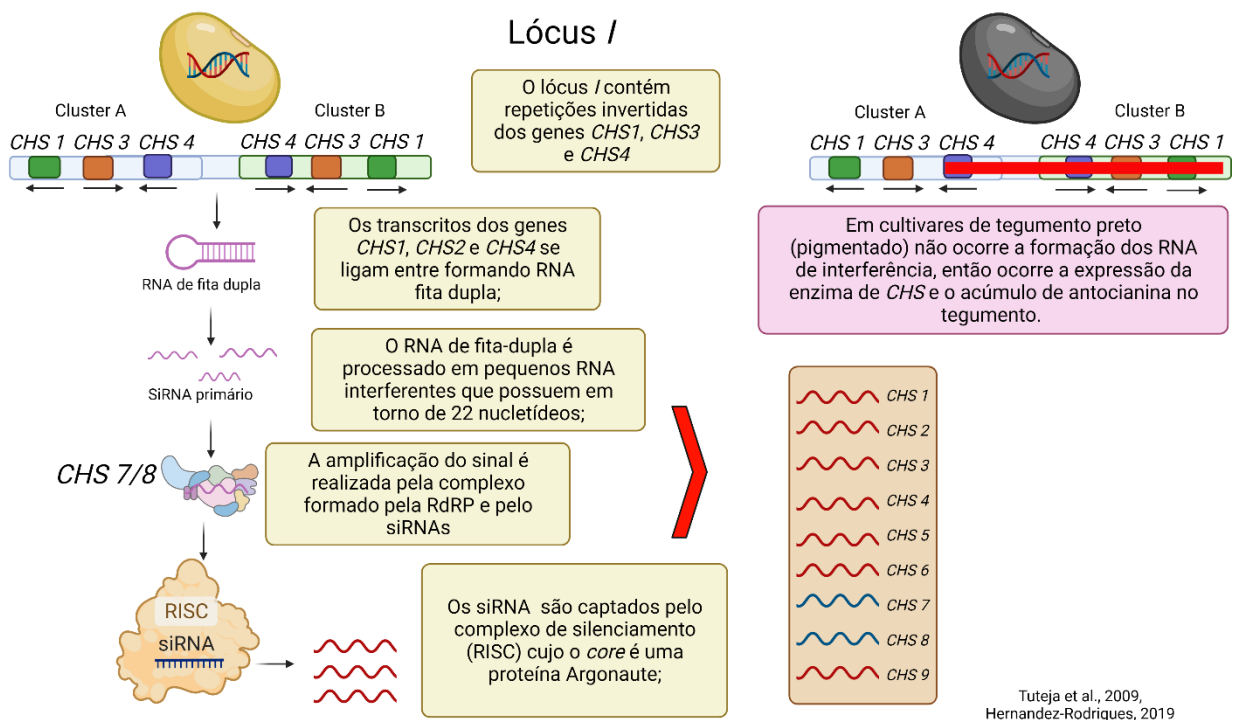
Na soja, há quatro alelos (*I*, *i<sup>i</sup>*, *i<sup>k</sup>*, *i*) do *locus* Inibidor (*I*) que governa a pigmentação do tegumento da semente, os alelos *I*, *i<sup>i</sup>* e *i<sup>k</sup>*, dão origem a coloração amarelo claro, hilo pigmentado e regiões pigmentadas em formato de “sela”, respectivamente. Já o alelo recessivo *i* torna o tegumento pigmentado de preto ou marrom (Tuteja et al., 2009; Senda et al., 2012). O mapeamento do *locus* *I* revelou que ele possui regiões que codificam a enzima *Chalcona Sintases* (*CHS*) (Todd; Vodkin, 1996; Tuteja et al., 2004). Esta enzima é a primeira atuante na via dos fenilpropanóides que leva a uma grande variedade de compostos do metabolismo secundário, incluindo as antocianinas que estão presentes em vários tecidos e conferem a coloração preta ao tegumento da semente de soja (Hernandez-Garcia, 2019).

Em soja, a *CHS* pertence a uma família multigênica, de 9 membros (*CHS1-CHS9*) que apresentam padrões de expressão variáveis de acordo com o tecido (Senda et al., 2012, Hernandez-Garcia, 2019). Através de análises filogenéticas de suas *ORFs*, (*Open Reading Frames*), os membros da família foram classificados em duas subfamílias, que compreende os genes *CHS7/CHS8* e os membros restantes em outra (*CHS1-CHS6*, *CHS9*) (Tuteja et al., 2008, Kurauchi et al., 2009).

A expressão dos genes *CHS* pode ser influenciada por regulação pós-transcricional via RNA de interferência (*siRNAs*). Entretanto sua expressão é cultivar dependente. Cultivares com tegumento amarelo claro (alelo *I*) produzem *siRNAs* de *CHS* no tegumento, enquanto sementes de tegumento pigmentado (alelo *i*) não produzem (Tuteja et al., 2009, Hernandez-Garcia, 2019). Evidências sugerem que os genes *CHS 1,3* e *4* contém repetições de sequencias invertidas que expressam *siRNA* (Hernandez-Garcia, 2019). Estes *siRNAs* transcritos são processados por um complexo composto por uma

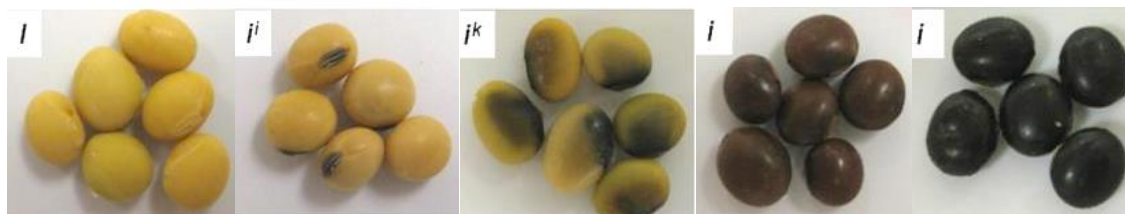
enzima *DICER*, um *RNA* maduro *CHS7-8* (*AGO5*) e uma *RNA polimerase* (*RdRP*) (Figura 2). Este complexo uma vez formado origina o *RNA-induced silencing complex* (*RISC*), que irá produzir *siRNAs* secundários que são capazes de amplificar o sinal dos *siRNAs* e silenciar os *mRNA* de todos os nove genes da família *CHS* (Tuteja et al., 2009, Kurauchi et al., 2011, Senda et al., 2017, Figura 2). Já os *loci R* e *T* controlam o tipo e a abundância dos pigmentos no tegumento da soja a partir de mecanismos transcricionais, que resultam em coloração preta (*i, R, T*), preta imperfeita (*i,R,t*), marrom (*i,r,T*) e camurça (*i, r, t*). De acordo com estudos de clonagem molecular realizados por Zabala e Vodkin (2003) e Nagamatsu et al. (2007), o *locus T* codifica uma enzima flavonóide 3' hidroxilase (*F3'H*) que é responsável pela síntese de antocianinas e proantocianidinas. Enquanto o *locus R*, codifica um fator de transcrição *MYB R2R3* que regula a expressão de cascatas gênicas envolvidas na coloração do tegumento (Gilman et al., 2011).

**Figura 2** - Silenciamento pós-transcricional via RNA de interferência (*siRNA*) do gene *CHS* no tegumento da soja, devido a ação das repetições invertidas presentes nos genes *CHS1*, 3 e 4 que serão clusterizadas (Hernandez-Garcia, 2019).



De acordo com estudos de mapeamento genético desenvolvido por Yang et al. (2010), o *locus O* codifica uma enzima redutase de antocianinas. O *locus W1* por sua vez, codifica para uma *Flavonoide 3'5' hidroxilase*, enzima que está relacionada com a coloração do tegumento e ao florescimento (Zabala; Vodkin, 2007).

**Figura 3-** Fenótipos dos tegumentos governados pelos quatro alelos do locus *I* em soja. O alelo recessivo pode levar a expressão da coloração marrom ou preta (Hernandez-Garcia, 2019).



A coloração do tegumento também é afetada por interações epistáticas como demonstrado por Cho, Jones e Vodkin (2017). Os autores observaram que a mutação no locus *K1*, resulta em diferenças espaciais entre a coloração do tegumento ao interagir com o locus *I* (Figura 3). O locus *K1* codifica uma enzima argonauta (AGO5) que é envolvida no carregamento de miRNAs e siRNAs de ação *trans*. Uma vez mutado produz uma enzima AGO5 deficiente, prejudicando a ação dos pequenos RNAs produzidos no locus *I*, seja na forma dominante ou recessiva (Cho; Jones; Vodkin, 2017).

### 3.5.1 Qualidade de Sementes e sua Relação com a Coloração do Tegumento

O tegumento da semente de soja é a camada externa da semente, originado a partir dos integumentos do óvulo. A sua estrutura é composta pela epiderme da testa, formada pelas cutículas pelas células paliçádicas ou macrosclerídeos, a hipoderme da testa, formada pelas células em ampulheta ou fosteosclerídeos e as células parenquimatosas (Mertz-Henning, 2009, Miller et al., 2010, Park et al., 2012).

As principais funções do tegumento são proporcionar a proteção do embrião contra danificações mecânicas e ataque de microrganismos, a regulação das trocas gasosas entre o embrião e o ambiente externo, do intercâmbio de água e a proteção durante a embebição, evitando a ruptura celular (Park et al., 2012). Desta forma, percebe-se que o tegumento é condicionante/está vinculado da qualidade de sementes

Neste contexto, vários estudos têm demonstrado uma correlação positiva entre a coloração preto do tegumento e a melhoria na qualidade da semente, especialmente com relação a resistência ao armazenamento e longevidade. Adsul, Chimote e Deshmukh

(2018) verificaram que a longevidade de semente está relacionada positivamente não só com a coloração do tegumento, mas também com a densidade da semente, proporção do tegumento e conteúdo de lignina no tegumento. Enquanto a permeabilidade de tegumento foi relacionada negativamente com a longevidade da semente.

Cultivares com tegumento preto apresentaram valores elevados para teor de lignina e outros compostos da via dos fenilpropanóides, como flavonoides e antocianina que também podem estar relacionados a longevidade da semente (Menezes et al., 2009, Abati et al., 2021). A camada paliçádica e as células em formato de ampulheta são mais espessas em genótipos de tegumento preto, e essa camada mais rígida na soja preta confere maior qualidade fisiológica e menor permeabilidade (Mertz et al., 2009, Kuchlan; Daniel; Samuel, 2010, Bahry et al., 2015). Por este motivo, em sementes de tegumento preto o processo de embebição da semente ocorre mais lentamente atenuando o processo de deterioração (Bahry et al., 2017).

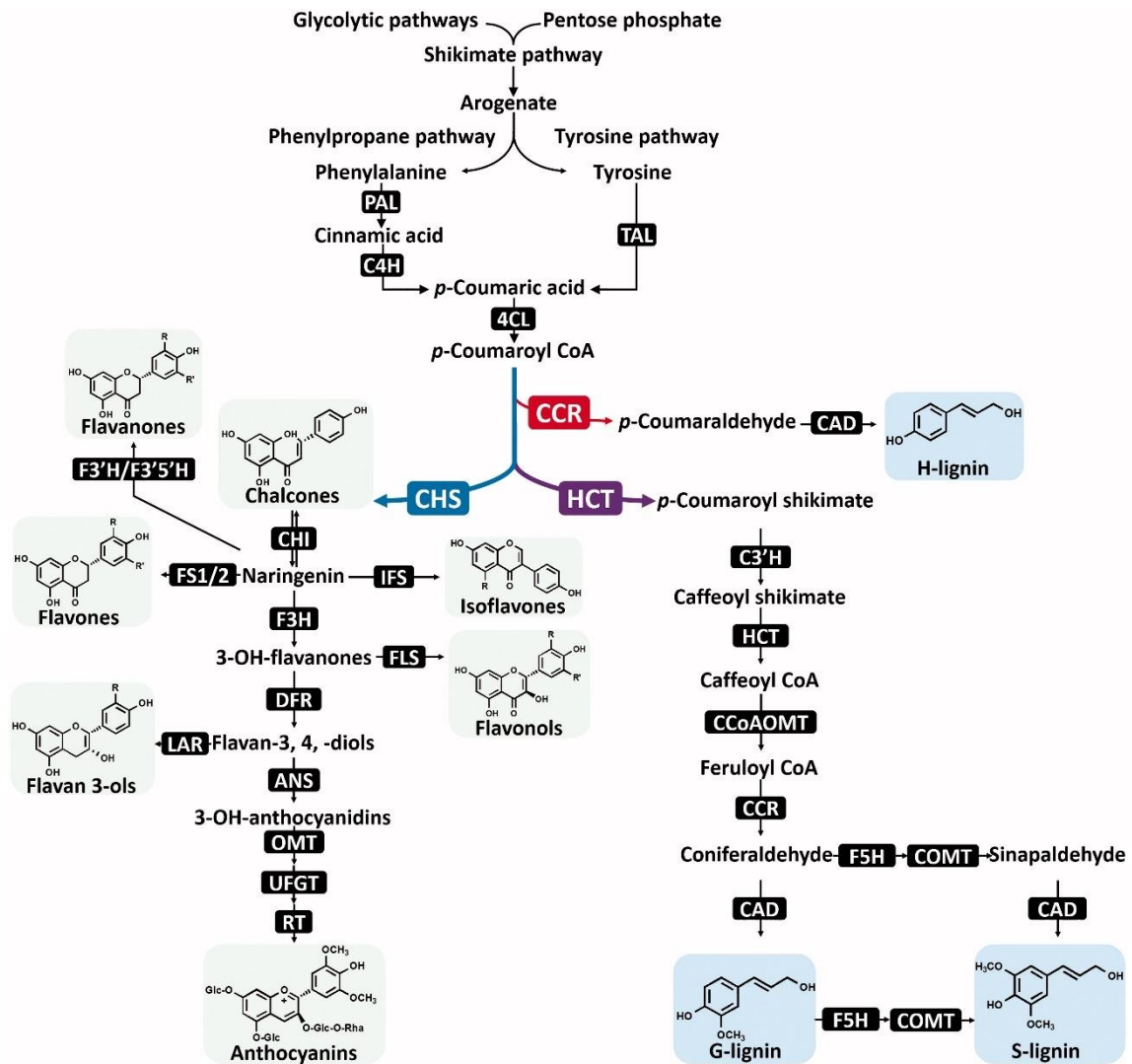
### 3.6 ROTA DOS FENILPROPANÓIDES

A rota dos fenilpropanóides produz uma ampla gama de compostos derivados do esqueleto da fenilalanina que estão relacionados na defesa, suporte estrutural e sobrevivência das plantas (Vogt, 2010, Fraser & Chapple, 2011, Dastmalchi & Dhauchadel, 2015). Entre os principais compostos produzidos pela rota dos fenilpropanóides destacam-se a lignina como o segundo maior biopolímero produzido na Terra e vários compostos que demonstram coloração como flavonoides, isoflavonoides, antocianinas e antocianidinas (Zhang et al., 2021).

A rota dos fenilpropanóides tem início com a desaminação da fenilalanina que produz ácido transcinâmico e amônia, essa reação é catalisada pela fenilalanina amônia liase (PAL) (Figura 4). Em seguida, o ácido transcinâmico sofre reação de hidroxilação gerando ácido cumárico pela cinamato 4-hidroxilase (*CH4*). Posteriormente, a curamato CoA ligase (*4CL*) catalisa a formação dependente de ATP do p-cumaril-CoA. A partir desta etapa, a via se divide em diversos ramos que irão formar vários compostos de acordo com a ação de diferentes enzimas (Vogt, 2010, Liu; Osborn; Ma, 2015, Rock, 2017).

Muitos compostos da rota dos fenilpropanóides estão relacionados a longevidade e melhor desempenho fisiológico da semente (Liu et al., 2017, Chanda et al., 2022). Entre eles destacam-se a lignina, flavonoides e antocianinas.

**Figura 4** - Via resumida dos fenilpropanóides e origem dos produtos lignina, flavonóides e antocianina em plantas (Zhang et al., 2021).



### 3.6.1 Lignina

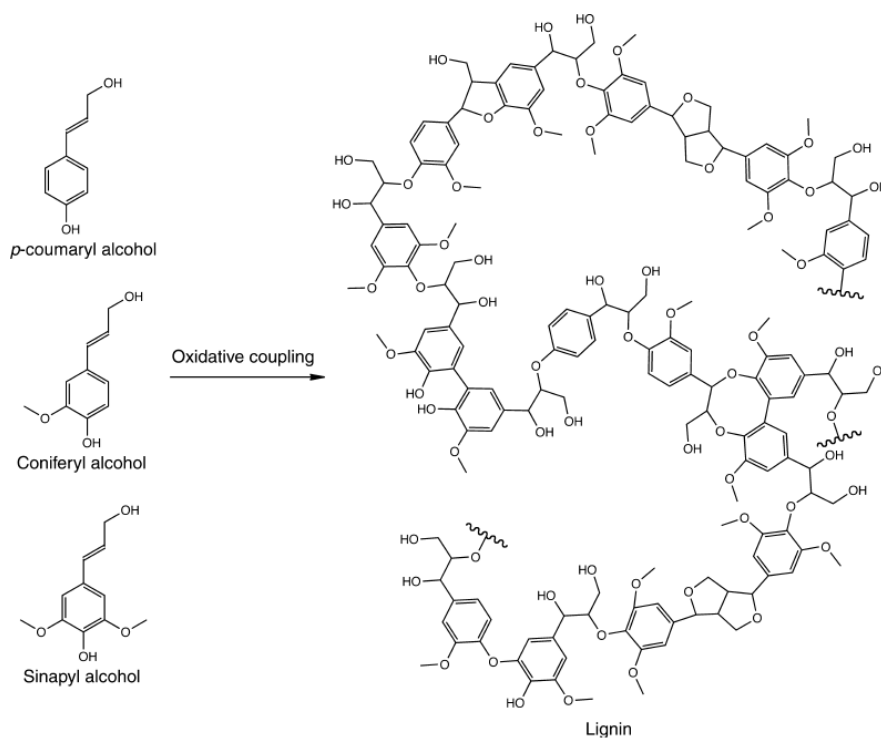
A lignina é formada por monômeros denominados de monolignóis, eles são álcoois não metoxilados p-cumaril, e álcoois metoxilados coniferil e sinaptil, esses compostos formam as unidades do polímero de lignina H-(hidroxifenil), G-(guaiacil) e S-(siringil), respectivamente (Figura 4) (Weng & Chapple, 2010; Barros et al., 2015, Vanholme et al., 2019).

A via de biossintese de monolignóis é formada por 8 enzimas principais e está exemplificada na Figura 5. Primeiramente ocorre com a conversão de fenilalanina para ácido transcinâmico pela Fenilalanina amônia-liase (PAL). Em seguida, o ácido transcinâmico sofre a ação das enzimas Cinnamato 4-hidroxilase (C4H) que catalisa a primeira reação de hidroxilação aromática da via, formando os produtos ácido p-

cumárico. Após isso, 4-Hidroxicinamoil-CoA ligase (4CL) transforma o substrato p ácido p-cumárico em p-cumaroil CoA, esse produto pode ser utilizado diretamente para a produção de lignina H, ou ainda pode ser usado para os outros passos da via. Posteriormente, hidroxicinamoil-CoA:shiquimato hidroxicinamoil transferase (*HCT*), transforma p-cumaroil CoA em p-cumaroil-shiquimato, ela ainda atua posteriormente na via transformando Cafeoil shiquimato em cafeoil-CoA. Seguindo a reação do produto p-cumaroil-shiquimato, ele é transformado em Cafeoil chiquimato pela enzima 4-Hidroxicinamoil-CoA ligase (4CL). Seguindo a etapa mais avançada da via, Cafeoil-CoA O-metiltransferase (CCoAOMT), catalisa a primeira trans-metilação da via, utilizando como o substrato S-adenosil metionina (SAM) como doador do grupamento metil. As duas últimas reações da via de biossíntese dos molignóis direcionam a produção de lignina H-(hidroxifenil), G-(guaiacil) e S-(siringil) e dependem do substrato da reação. Primeiramente, a enzima (hydroxy) cinamoil-CoA redutase (CCR) catalisa individualmente a reação de redução entre os substratos p-coumaroyl-CoA, cafeoil-CoA e feruloil-CoA nos produtos p-cumaraldeído, Cafealdeído e coniferaldeído, respectivamente. Por fim, a (hydroxy) álcool cinamílico desidrogenase (CAD) utiliza os produtos obtidos da oxirredução e catalisa a reação de redução para álcool p-cumaroil, álcool coniferílico e álcool sinapílico (Figura 6). Esses álcoois são metoxilados formando os diferentes tipos de lignina: lignina H-(hidroxifenil), G-(guaiacil) e S-(siringil) (Weng & Chapple, 2010; Dixon & Barros, 2018, Vanholme et al., 2019, Zhang et al., 2021).



**Figura 6-** Polimerização oxidativa de lignina pelo acoplamento dos três álcoois hidroxicinamil (Weng & Chapple, 2010).



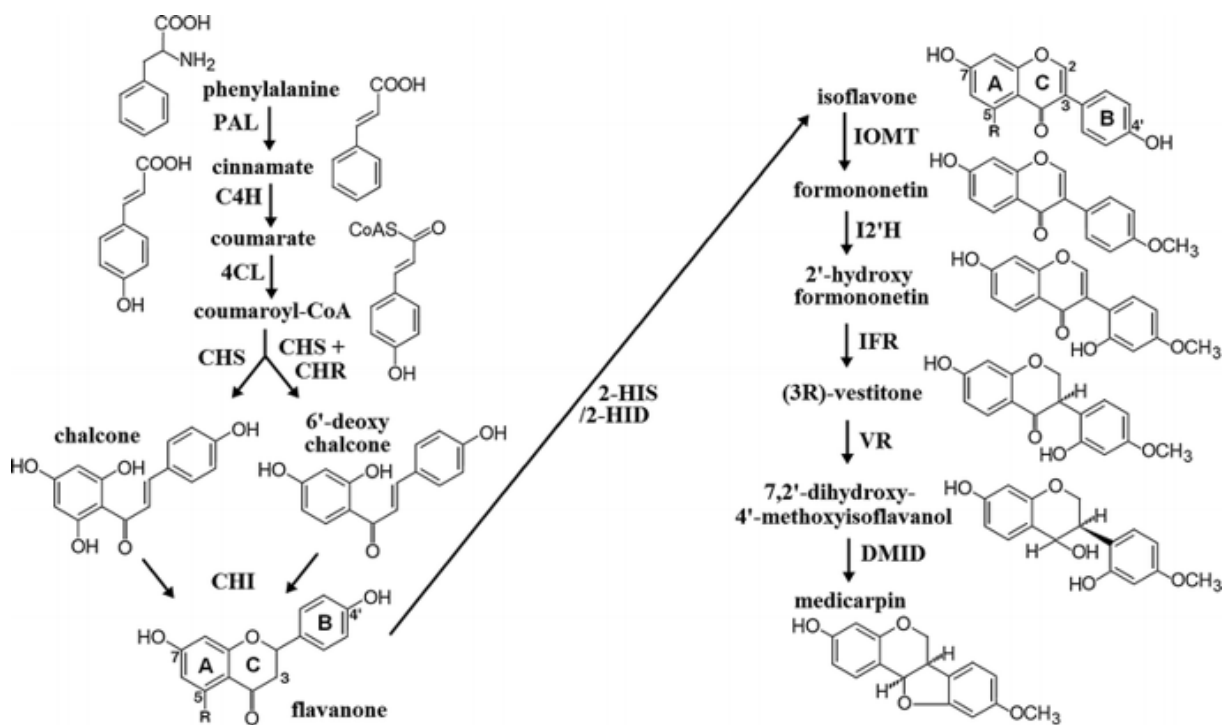
### 3.6.2 Isoflavonas

As isoflavonas são um grande grupo de produtos gerados pelas plantas que possuem funções importantes na defesa vegetal e na sinalização para interações simbióticas (Pei et al., 2018, Deng et al., 2019)

As isoflavonas são um dos produtos oriundos da rota dos fenilpropanóides que já foi descrita anteriormente. A partir do momento de formação de *p*-cumaroil-CoA, a enzima *CHS* (*chalcone sinase*) catalisa a primeira etapa para a biossíntese dos flavonoides, a condensação da molécula *p*-cumaroil-CoA em três moléculas de malonil-CoA que se tornam um andaime de chalcona. Em seguida, a enzima *CHR* (*chalcone reductase*) age como um intermediário na reação de *CHS*, produzindo deoxichalcona da reação catalítica agrupada da *CHS*. Posteriormente, a enzima *CHI* (*chalcone isomerase*) catalisa a ciclização das chalconas ou desoxichalconas para formar o centro estrutural da flavanona. A partir da ação da ação de *2-HIS* (*2-hydroxyisoflavanone*) a flavanona é convertida em 2-hidroxisoflavanona pela migração de um anel-aril para transferir o anel-B aromático para a posição C-2 para C-3 e hidroxilação no C-2. Ocorre a desidratação da 2-hidroxisoflavanona de forma espontânea ou pela ação da enzima 2-hidroxisoflavanona desidratase para geração das isoflavonas daidzeína e genisteína (Figura 7). A partir deste

ponto, vias específicas são direcionadas, em que vários derivados são sintetizados, como as fitoalexinas, medicarpina, pisatina e maackiaina (Wang, 2010, Veitch, 2013, Dastmalchi; Bernards; Dhaudabel, 2016).

**Figura 7** – Rota geral de formação dos isoflavonóides (Wang, 2010).



Na soja foram identificadas doze isômeros de isoflavonas que incluem agliconas (daidzeína, genisteína e gliciteína), seus glicosídeos conjugados (daidzína, genistina e glicitina), os glicosídeos em conjunto com o malonil (6''-O-malonil daidzína, 6''-O-manonil genistina e 6''-O-malonil glicitina), e os glicosídeos acetil (6''-O-acetil daidzína, 6''-O-acetil genistina e 6''-O-acetil glicitina) (Dastmalchi & Dhaudabel, 2014). A soja é considerada uma fonte rica em isoflavonas (3,2-4,2 mg/g massa seca de sementes) (Mujic et al., 2011). Sendo que o conteúdo individual de isoflavonas varia conforme o genótipo, condições de cultivo e do ano (Medic; Atkinson; Hurburgh, 2014, Júnior & Ida, 2015).

O acúmulo de isoflavona ocorre nas sementes de soja entre 35 a 60 dias depois do florescimento (Dastmalchi & Dhaubadel, 2015). A distribuição de isoflavonas nas sementes não é uniforme, com a maior concentração nos hipocótilos, seguidos dos cotilédones e do tegumento (Medic; Atkinson; Hurburgh, 2014).

O teor elevado de isoflavonas nas sementes foi relacionado a maior tolerância ao armazenamento e a uma qualidade fisiológica das sementes superior pois, esses

compostos apresentam função antioxidante e possibilitariam uma maior defesa contra espécies reativas de oxigênio (Abati et al., 2021).

Adicionalmente, isoflavonas possuem grande importância para saúde e prevenção de doenças crônicas em humanos. Os autores Zaheer e Humayoun Akhtar (2017) identificaram em revisão de literatura vários benefícios que o consumo de isoflavonas pode proporcionar, entre eles destacam-se: alívio em sintomas da menopausa, câncer de mama, câncer de próstata, na incidência de doenças cardiovasculares, osteoporose, diabetes e obesidade, melhoria nas funções cognitivas e até mesmo na prevenção de infecções virais.

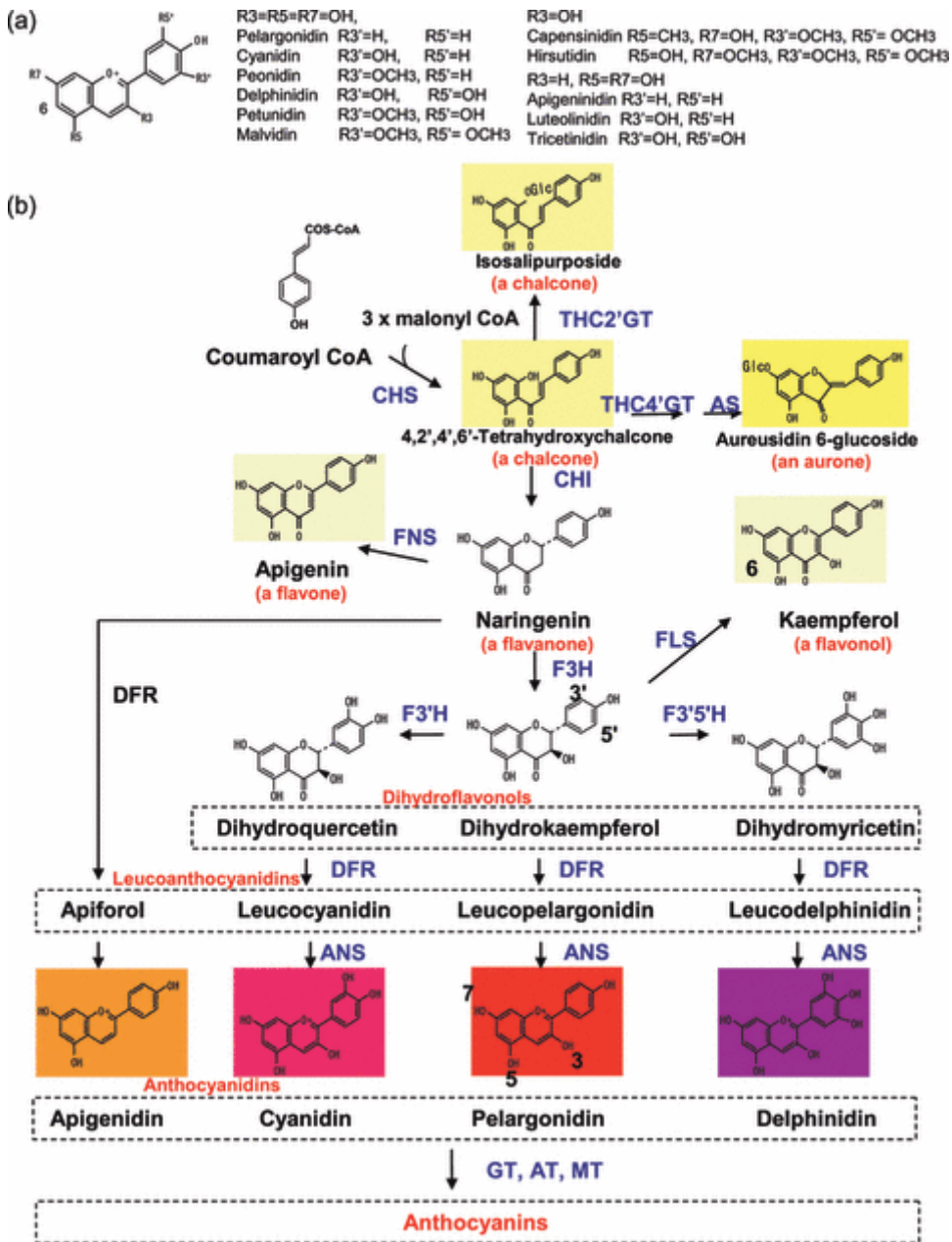
### 3.6.3 Antocianinas

As antocianinas são amplamente distribuídas no reino vegetal, e são responsáveis por diferentes colorações que variam do laranja ao azul em vários tecidos como flores, folhas, frutos e sementes. Esses compostos são tipicamente solúveis em água e armazenados nos vacúolos da célula vegetal (Tanaka; Sasaki; Ohmiya, 2008, Oancea & Oprean, 2011). Além de proporcionar pigmentação do tecido vegetal as antocianinas estão relacionadas a defesa contra estresses e patógenos (Zhang; Butelli; Martin, 2014). Ainda, possuem função antioxidante protegendo o tecido vegetal contra danos oxidativos causados por irradiação elevada quando a capacidade de fixar carbono é pequena (Silva et al., 2016).

As principais antocianinas conhecidas são pelargonidina, cianidina, peonidina, delphinidina, petunidina e malvidina (Figura 8a) (Landi; Tattini; Gould, 2015). A coloração é dependente do número de grupos hidroxila nos anéis B, quanto maior o número de grupos, mais azul a cor. A metilação de antocianinas tem um leve efeito avermelhado (Tanaka; Sasaki; Ohmiya, 2008).

As antocianinas são produzidas na rota dos fenilpropanóides até a formação de antocianidinas, a partir desta etapa, a molécula é modificada com os grupamentos glicosil, acil ou metil, reações que são catalisadas por glucosiltransferases, aciltransferases e metiltransferases, respectivamente (Landi; Tattini; Gould, 2015).

**Figura 8** - Estrutura básica das antocianinas com representantes mais presentes a esquerda e outros de menor importância a direita(a). Seguimento da rota dos fenilpropanóides que dá origem ao grupo das antocianinas(b).



As antocianinas conferem a coloração preta ao tegumento de soja, pelo acúmulo desse componente na parede paliçádica da epiderme no tegumento (Todd; Vodkin, 1993). Em sementes de soja preta podem ser encontrados um perfil de 9 antocianinas, que são afetadas por condições genéticas e de ambiente, entretanto, a cianidina-3-O-B-D-glicosídeo é a forma mais presente (Koh; Yun; Kim, 2014, Carrão-Panizzi et al., 2019). Já foram apontados nove tipos de antocianinas, sendo que as formas mais presentes são: cianidina-3-O-glicosídeo, delphinidina-3-O-glicosídeo e pelargonidina-3-O-glicosídeo (Lee;

Cho, 2012, Cho et al., 2013). Muitos estudos foram realizados associando a soja preta a benefícios à saúde humana, em especial ao teor de antocianina elevado encontrado no tegumento. Ganesan e Xu (2017) em revisão de literatura identificaram vários atributos que o consumo de soja preta e o teor elevado de antocianina conferem a saúde humana, entre eles incluem: propriedades antioxidantes, anti-inflamatórias, nefroprotetoras, antidiabéticas, anti-câncer, antifertilidade, antiobesidade, antiartríticas, neuroprotetoras, anti-hiperlipidêmicas, anticataratas e cicatrizantes de feridas.

### 3.7 CARACTERÍSTICAS TECNOLÓGICAS DA CULTIVAR BRS MG715A

A cultivar de soja BRSMG 715A apresenta tegumento preto e foi desenvolvida pela Empresa Brasileira de Agropecuária (Embrapa) em parceria com a Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig). Possui teor de proteína elevado para alimentação humana e pode ser utilizado como um substituto do feijão. Por possuir sabor suave, característica relevante para a aceitação do consumidor, já que a maioria das cultivares de soja podem ser associadas a um sabor característico e forte. Além disso, possui elevada atividade antioxidante, aproximadamente, 1.8x mais que cultivares de soja de tegumento amarelo. Ademais, esta cultivar é de fácil cozimento, uma característica tecnológica importante para cultivares de alimentação humana. (Landgraf, 2019).

No tocante a atributos agronômicos a cultivar apresenta produtividade semelhante a outras cultivares convencionais disponíveis no mercado, além disso, possui resistência aos nematoides do cisto e da galha (Landgraf, 2019). Sobre a qualidade da semente, Abati et al. (2022) demonstraram que as sementes da cultivar apresenta maior qualidade fisiológica em relação a cultivares de tegumento amarelo durante o armazenamento, sobretudo quando armazenadas em ambientes sem controle de temperatura e umidade relativa do ar. Adicionalmente, a cultivar possuía teor elevado de lignina e isoflavona que são relacionadas ao maior potencial de armazenamento da semente (Abati et al., 2021).

### 3.8 RNA-SEQ COMO FERRAMENTA PARA IDENTIFICAÇÃO DE GENES

Sequenciamento de RNA (RNA-Seq) é uma poderosa ferramenta para quantificar a expressão de genes, detectar variantes de permuta, variações de nucleotídeos

únicos (SNPs) e explorar novos transcritos codificantes e não codificantes (Kratz; Carninci, 2014; Molinari et al., 2021a).

O RNA-Seq permite identificar e quantificar moléculas de RNA em diferentes tipos de amostras biológicas. Para se obter respostas biológicas precisas a partir do RNA-Seq devem ser considerados alguns aspectos, como: o desenho experimental, o tipo de biblioteca sintetizada (*single-end* ou *paired-end*), o tamanho dos fragmentos gerados (pares de bases - pb), o número de repetições biológicas, a profundidade (milhões de reads) e cobertura do sequenciamento (número de vezes que cobriu o genoma referência), a disponibilidade do genoma da espécie e a escolha dos softwares adequados para executar as análises computacionais (Conesa et al., 2016; Molinari et al., 2021a).

Basicamente, um experimento de RNA-Seq segue as seguintes etapas: 1) Delineamento experimental adequado ao objetivo da pesquisa e escolha da plataforma de sequenciamento mais adequada; 2) Extração do RNA, que pode ser total ou fracionado/tipo específico; 3) Construção das bibliotecas de cDNA, 4) Sequenciamento; 5) Análises de bioinformática (Fang et al., 2012, Hrdlickova et al., 2016, Conesa et al., 2016; Molinari et al., 2021a).

As análises de bioinformática seguem um pipeline baseado em: verificar a qualidade dos fragmentos saídos do sequenciador, limpeza dos fragmentos, montagem dos transcritos no genoma referência ou montagem *de novo* e expressão diferencial considerando o algoritmo adequado de acordo com o número de repetições biológicas do experimento. Outras etapas intermediárias de tratamento dos dados podem ser adotadas (Conesa et al., 2016; Molinari et al., 2021a). A última etapa trata-se da biologia de sistema que visa anotar e identificar as funções biológicas dos genes a partir de diferentes softwares (Garg; Jain, 2013, Costa-Silva et al., 2018).

Em plantas, a utilização da plataforma de RNA-Seq permitiu explorar a dinâmica transcricional em vários aspectos do crescimento e desenvolvimento vegetal (Martin et al., 2013). Especificamente em soja, a ferramenta tem sido utilizada para identificar com robustez e assertividade de genes diferencialmente expressos em diferentes tecidos, estádios de desenvolvimento e tratamentos (Severin et al., 2010, Jones; Vodkin, 2012; Marcolino-Gomes et al., 2014; Wang et al., 2014; Liu et al., 2017; Leisner; Yendrek; Ainsworth, 2017; Lima et al., 2017; Reis et al., 2020; Molinari et al., 2021b).

## 4. MATERIAL E MÉTODOS

### 4.1 MATERIAL BIOLÓGICO E EXPERIMENTO EM CASA DE VEGETAÇÃO PARA COLETA DE AMOSTRA PARA RNA-SEQ E RT-QPCR

Para coleta dos dados, dois experimentos idênticos foram realizados em casa de vegetação em datas distintas. O primeiro foi utilizado para coleta do material biológico para análise de RNA-Seq e o segundo, para coleta do material biológico utilizado nas análises de RT-qPCRs. As cultivares escolhidas neste experimento foram previamente caracterizadas quanto aos seus teores de lignina, compostos fenólicos e qualidade fisiológica (Abati et al, 2021, Abati et al., 2022). As três cultivares foram selecionadas em função da resposta ao comportamento no armazenamento: BRSMG 715A - tegumento preto como resistente ao armazenamento, BRS 413 RR - tegumento amarelo como moderadamente resistente ao armazenamento e DM 6563 IPRO - tegumento amarelo como suscetível ao armazenamento.

As sementes foram germinadas em papel para germinação em períodos alternados de luz e escuro em câmara de germinação, por 4 dias, a  $28 \pm 1^\circ\text{C}$  e 100% de umidade relativa (UR). Após a germinação, foram transferidas duas plântulas com raiz primária uniforme por vaso (8L). O experimento totalizou em 100 vasos. Cada vaso continha uma mistura de substrato 1:1 (solo fertilizado e areia lavada). O experimento foi conduzido em casa de vegetação em condição de dias curtos (10h de luz/ 14h de escuro) a  $28 \pm 2^\circ\text{C}$  e condições ótimas de irrigação.

As sementes foram colhidas quando as cultivares atingiram a maturação plena. As sementes foram debulhadas manualmente e na sequência homogeneizadas para compor uma amostra de trabalho de 500g. Para cada cultivar, a amostra foi subdividida em dezoito partes iguais, as quais foram utilizadas para compor os diferentes tratamentos T0 - recém-colhidas contendo três repetições; CF - armazenamento em câmara fria por 6 meses ( $10^\circ\text{C}$  e umidade relativa 50%), sendo cada tratamento composto por 3 repetições. As amostras das sementes maduras inteiras foram utilizadas para extração do RNA total.

## 4.2 ANÁLISE DE RNA-SEQ

As amostras de sementes coletadas nos dois tratamentos foram congeladas em Nitrogênio (N<sub>2</sub>) líquido e armazenadas a -80 °C até o momento da extração do RNA. As amostras foram maceradas em N<sub>2</sub> e o RNA total extraído utilizando o kit Concert Plant RNA reagente (Invitrogen, Califórnia, EUA) de acordo com as especificações do fabricante. A quantificação do RNA foi realizada em espectrofotômetro NanoDrop, de acordo com os seguintes parâmetros de qualidade e pureza: concentração > 600 ng µL, relação 260/280 variando entre 1,8 e 2,0 e relação 260/230 ≥ 2,0.

O RNA foi tratado com o kit turbo DNase-free para remoção de DNA genômico remanescente (Invitrogen, Califórnia, EUA). A integridade do RNA foi avaliada por eletroforese em gel de agarose 1% (p/v) corado com brometo de etídio (1 µg/mL) (Sambrook et al., 1989). As amostras de RNA total com alta qualidade foram enviadas para sequenciamento na Universidade da Georgia (Georgia Genomics Facility – GGF, EUA). Antes das amostras serem sequenciadas, foram avaliadas no equipamento Agilent Bioanalyzer 2100 (Agilent Technologies, Inc.) e apenas as amostras com número de integridade de RNA (RIN) ≥ 7,00 foram utilizadas para sintetizar as bibliotecas de mRNA-Seq. As bibliotecas foram sintetizadas usando o kit Illumina TruSeq™ SBS v5 poli-A. em plataforma de sequenciamento Illumina NextSeq 500 1.9 de 75 bp de extremidade pareada (Illumina, San Diego, CA, EUA), com pelo menos 1X de cobertura do genoma. Cada amostra foi sequenciada em triplicata biológica, totalizando dezoito bibliotecas de mRNA (3 repetições BRSMG 715A - T0; 3 repetições BRSMG 715A - CF, 3 repetições BRS 413 RR - T0; 3 repetições BRS 413 RR - CF, 3 repetições DM 6563 IPRO - T0; 3 repetições DM 6563 IPRO - CF).

### 4.2.1 Análise de bioinformática

A qualidade dos fragmentos brutos (FASTQ) foi avaliada usando o software FastQC v.0.11.5 antes e após a remoção de adaptadores e de sequências de baixa qualidade (Andrews 2010; Patel & Jain 2012). A limpeza dos fragmentos foi realizada por meio do software Trimmomatic versão 0.36 (Bolger et al., 2014) padronizando cortes a cada quatro nucleotídeos nas extremidades das sequências que apresentavam score de qualidade inferior a 30 (Phred Quality Score, Q ≥ 30).

O alinhamento dos fragmentos foi realizado pelo software HISAT2 v.2.1.0 utilizando o genoma da soja Wm82.a2.v1 como referência (Goodstein et al., 2011, Kim et al., 2015). Em seguida, artefatos da PCR provenientes do sequenciamento Illumina foram removidos usando o software Samtools v.1.5 (Li et al. 2009). A montagem dos transcritos foi realizada usando o software Stringtie v.1.3.3 (Pertea et al., 2015) e as expressões relativas foram obtidas a partir do software EdgeR v.3.22.3 (Robinson et al., 2010) via RStudio v.3.5.1 (Racine, 2012).

Para cada cultivar, os genes diferencialmente expressos foram obtidos pela comparação dentro de cada tratamento (T0; CF) entre a cultivar de tegumento preto (BRSMG 715A) e as cultivares de tegumento amarelo (BRS 413 RR; DM 6563 IPRO). Foram considerados diferenciais os genes com valores de Log<sub>2</sub> Fold Change (Log<sub>2</sub>FC)  $\leq -1$  e  $\geq +1$  com taxa de falso positivo (FDR)  $\leq 0,05$  e logCPM positivo (Molinari et al., 2021a). A fim de aumentar a confiabilidade da análise, foram selecionados para o estudo apenas os genes diferencialmente expressos entre as cultivares preta e amarela que se mostraram conservados nos tratamentos avaliados.

A anotação dos genes foi realizada a partir da ferramenta Phytomine (Goodstein et al., 2011). A análise de enriquecimento gênico e ontologia gênica foi realizada pelo software ShinyGO baseado em informações do banco de dados KEGG pathways (Kanehisa & Goto, 2000).

#### 4.3 ANÁLISE DE RT-QPCR

O RNA total foi extraído de amostras de sementes da soja usando o reagente Trizol<sup>®</sup> e em seguida as amostras foram tratadas com o kit DNase I (Invitrogen, Carlsbad, CA) para remover possível contaminação com DNA remanescente. Depois disso, o cDNA foi sintetizado usando Super Script<sup>®</sup> III First-Strand Synthesis System (Invitrogen, Carlsbad, CA) de acordo com as instruções do fabricante. As reações de RT-qPCR foram compostas de cDNAs, primers F e R de 0,2  $\mu$ M e tampão de reação 1x Platinum<sup>®</sup> SYBR Green<sup>®</sup> qPCR SuperMix UDG (Invitrogen, Carlsbad, CA).

A calibração da expressão gênica foi realizada usando primers da  $\beta$ -actina (Glyma.15G050200 – iniciador F 5'GAGCTATGAATTGCCTGATGG3'/ iniciador R 5'CGTTTCATGAATTCCAGTAGC3') e  $\beta$ -Elf (Glyma.13G114700 – iniciador F 5'TTCTGTCTTCTGCAAGTGGTG3'/ primer R 5'GATCCCTCATCCATACATTTTCAG3') (Gilman et al., 2011).

A expressão relativa foi realizada em triplicata biológica e triplicata técnica (n = 9). As reações de PCR quantitativo foram rodadas em termociclador modelo 7900HT (Applied Biosystems). As condições de ciclagem utilizadas foram desnaturação a 95 °C por 20 segundos (s), seguido por 40 ciclos de 95 °C por 3s, 60 °C por 26s e 1 ciclo para a curva de melting a 95 °C por 15s, 60 °C por 1 min e 95 °C por 15s. Apenas primers com eficiência  $\geq 85\%$  foram considerados a partir da fórmula  $E=[10^{(-1/\text{slope})-1}]$  (Rutledge & Cote, 2003). O nível de expressão foi determinado a partir da fórmula  $2^{-\Delta\Delta Ct}$  adaptada de acordo com a eficiência dos primers (Livak & Schmittgen 2001). A análise estatística foi realizada usando o teste t ( $p \leq 0,05$ ) via software Sasm-Agri (Canteri et al., 2001). A correlação de Pearson entre RT-qPCR e expressão de RNA-Seq log2FC foi realizada via RStudio v.3.5.1 (Racine, 2012).

#### 4.3.1 Primers dos genes alvo para validação da expressão relativa por RT-qPCR

Sete genes anotados como os mais expressos em sementes da cultivar de soja BRSMG 715A com tegumento preto foram selecionados (Tabela 1). Os genes foram escolhidos utilizando a expressão absoluta  $TPM \geq 3$  para cada cultivar (BRSMG 715A - tegumento preto; DM 6563 IPRO e BRS 413 RR- tegumento amarelo) e tratamento (T0, recém-colhidas; CF, armazenamento em câmara-fria durante 6 meses) indicando expressão. As sequências dos CDS dos genes alvo foram obtidas a partir do banco de dados Phytozome, e os primers específicos foram desenhados usando o software Primer3Plus (<https://primer3plus.com/cgi-bin/dev/primer3plus.cgi>). Homo e heterodímeros foram eliminados usando Software Multiple Primer Analyze (<https://www.thermofisher.com/br/multiple-primer-analyzer.html>).

**Tabela 1-** Lista de primers utilizados na quantificação da expressão relativa dos alvos via RT-qPCR.

Gene ID	Anotação	Primer 5' - 3'	Tm (°C)
Glyma.11G228900	<i>CYP90A1</i>	Primer F: CCTGCTCGGAAAACACTCTC	60.0
		Primer R: GAGTTGGCGAAGCTCATTGT	60.4
Glyma.13G217400	<i>CYP710A</i>	Primer F: ATGGCTCAATCAGTCCCAAG	60.1
		Primer R: TCTGGGATGTCTGGAGGTTC	60.0
Glyma.08G204300	<i>ACSF3</i>	Primer F: ATCTCGGTGGAGCTCGAGTA	60.0
		Primer R: CCACTAAGCCAAATCCCAAG	59.6
Glyma.03G129700	<i>CBL</i>	Primer F: AGTTGTGTGCTGAGGGGTTC	60.2
		Primer R: GGAAGCATCATCCACCAAAG	60.5
Glyma.02G254600	<i>HCT</i>	Primer F: TTCCCAGACCAAAGTCCAC	59.9
		Primer R: AAGAGGCATCCCTTTTGGAT	59.9
Glyma.16G032200	<i>ACS1</i>	Primer F: ATGAGTGGAGGGGCAACA	60.0
		Primer R: GGGTGGGAATCATAAAAGCA	60.6

Glyma.08G108800	<i>SAHH</i>	Primer F: CTATGAGAAGACCGGCGAAC	59.8
		Primer R: GAGACGCTCCTTCATCTGC	60.1

**Legenda:** **CYP90A1:** cytochrome P450, family 90, subfamily A, polypeptide 1; **CYP710A:** K09832 - cytochrome P450, family 710, subfamily A; **ACSF3:** malonyl-CoA/methylmalonyl-CoA synthetase, EC.6.2.1.3; **CBL:** cystathionine beta-lyase, (EC.4.4.1.8) -; **HCT:** Shikimate O-hydroxycinnamoyltransferase (EC.2.3.1.133); **ACS1:** S-adenosyl-L-methionine methylthioadenosine-lyase (EC.4.4.1.14); **SAHH:** Adenosylhomocysteinase (EC.3.3.1.1).

#### 4.4 ANÁLISE BIOQUÍMICA

##### 4.4.1 Análise de teor de Lignina

O teor de lignina foi determinado utilizando a metodologia proposta por Moreira-Vilar et al. (2014), que consiste na preparação da amostra para excluir proteínas e outros materiais que absorvem luz UV. Posteriormente, por meio do método do Brometo de acetila, em que 20mg de parede celular ausente de proteínas é colocado em uma centrífuga contendo 0,5mL de brometo de acetil 25% (v/v em ácido acético glacial) e incubado a 70 °C por 30 minutos.

Depois da digestão completa, a amostra é resfriada rapidamente em banho de gelo, e então misturado com 0,9 mL de 2M NaOH, 0.1 de 5M HCl-hidroxilamino, e volume de ácido acético glacial para completar a solução. Após a centrifugação (1400xg, 5 minutos), a absorbância do sobrenadante foi aferida a 280 nm. Uma curva padrão foi gerada com lignina alcalina e a absorbilidade 22.9 g<sup>-1</sup> L cm<sup>-1</sup>. Os resultados foram expressos em miligramas (mg) de lignina.g<sup>-1</sup> de parede celular.

##### 4.4.2 Análise Estatística

Os dados obtidos a partir da análise bioquímica foram submetidos a análise de variância (ANOVA), utilizando o software RStudio. Quando diferenças significativas foram encontradas através do teste F (p<0,05), as médias foram comparadas com teste de separação de médias Tukey (p<0,05).

## 5. RESULTADOS E DISCUSSÃO

### 5.1 EXPRESSÃO DIFERENCIAL

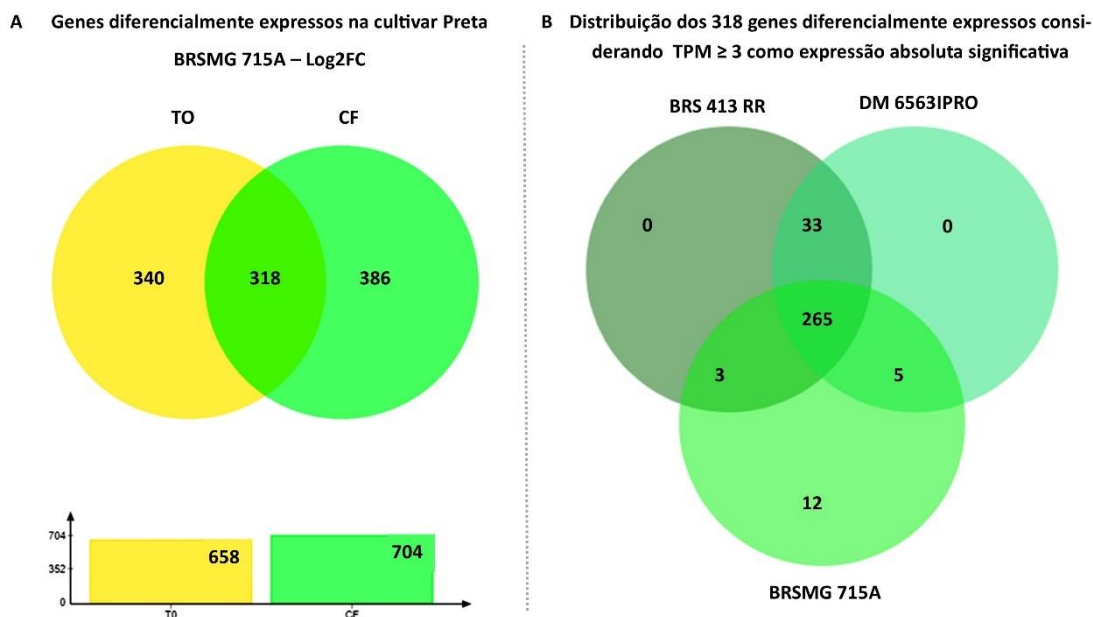
Em T0, foram identificados 658 genes diferencialmente expressos na cultivar de tegumento preto BRSMG 715A quando comparado as cultivares de tegumento amarelo (BRS 413 RR e DM 6563 IPRO). Destes, 340 são expressos exclusivamente neste tratamento. Em CF, foram identificados 704 genes diferenciais, sendo 386 expressos exclusivamente neste tratamento (Figura 9A). Ambos os tratamentos compartilharam 318 genes (Figura 9A; material complementar 2). Com o objetivo de identificar genes que pudessem estar envolvidos apenas nas diferenças entre cultivares o grupamento gênico de 318 genes foi utilizado para seleção de genes via RT-qPCR (Figura 9A).

Os genes diferencialmente expressos foram divididos em dois subgrupos a partir de seus valores de expressão absoluta em transcritos por milhão (TPM). O primeiro grupo refere-se a genes com expressão absoluta significativa ( $TPM \geq 3$ ) e o segundo grupo refere-se a genes com expressão absoluta não significativa ou ausente ( $TPM < 3$ ) (Arquivo adicional 2).

A figura 9B mostra um diagrama de *venn* subdividindo o grupo de genes com expressão absoluta significativa ( $TPM \geq 3$ ). Dentro deste grupo, foram encontrados 265 genes com TPM significativo em todas as cultivares, 33 genes com TPM significativo apenas nas cultivares amarelas, 3 genes com TPM significativo comuns entre cultivar amarela BRS 413 RR e a preta, 5 genes com TPM significativo comuns entre a cultivar amarela DM 6563 IPRO e a cultivar preta e 12 genes com TPM significativo exclusivos da cultivar de tegumento preto (BRSMG 715A).

**Figura 9** - Diagramas de Venn considerando genes com expressão relativa (A) e absoluta (B) significativas. Em A, sobreposição de genes com expressão relativa significativa entre a cultivar de tegumento preto BRSMG 715A e as cultivares amarelas BRS 413 RR e DM 6563 iPRO por tratamento (T0 - recém-colhidas; CF câmara-fria). Em B, sobreposição entre a cultivar BRSMG 715A e as cultivares BRS 413 RR e DM 6563 iPRO para os 318 genes com expressão absoluta significativa comuns a T0 e CF ( $TPM \geq 3$ ). Somente genes

com expressão relativa significativa ( $\text{Log}_2\text{FC} \leq -1$  e  $\geq 1$ ) comuns a T0 e CF (318) em (A) foram avaliados quanto sua expressão absoluta ( $\text{TPM} \geq 3$ ) (B).



Os genes exclusivos da cultivar de tegumento preto estão apresentados na tabela 2, o gene Glyma.06G213600 codifica para uma enzima isoamilase, que realiza desramificação do amido, transformando o amido em maltodextrina (Aguirre et al., 2018, Tayade et al., 2019). Já o Glyma.06G043900 codifica para ciclina A, que são proteínas importantes para o processo de diferenciação do embrião (Zhang et al., 2019). O gene Glyma.01G016700 inositol-pentaquisfosfato 2-quinase, essa enzima foi associada ao ácido fítico (considerado um fator antinutricional quelante de minerais), e como um dos candidatos para redução da substância através de nocaute e silenciamento, entretanto, o gene também é necessário para o crescimento de plântula e homeostase de fosfato (Kuo et al., 2014, Basak et al., 2020). O gene Glyma.02G064300 codifica para a enzima ribonuclease t2, e está envolvido nos processos de nodulação na soja (Azizkhani et al., 2021). Além disso, quatro genes estão relacionados a resistência a doenças, entre esses, três são receptores *leucine rich repeats* (LRR) e estão envolvidos no reconhecimento de sinais de patógenos, fitorreguladores vegetais e sinais de desenvolvimento (Belkhadir et al., 2014). O gene Glyma.20G138000 codifica cysteine rich repeat, associado a resistência a doenças (Miyakama et al., 2014). Já o gene Glyma.20G138000 codifica para fosfatidilinositol glicano, classe s (pigs) e está envolvido na glicolisação de proteínas, ação importante para vários processos biológicos como controle de qualidade de proteínas

no retículo endoplasmático, estabilidade, interações entre proteínas (Strasser, 2016). Esses genes podem ser utilizados posteriormente em trabalhos para superexpressão e/ou silenciamento para validar as suas funções na qualidade de sementes e a sua relação com tegumento

**Tabela 2-** Genes que apresentam expressão apenas na cultivar de tegumento preto BRSMG 715A

<b>GENE ID</b>	<b>Anotação Phytomine</b>	<b>KEGG</b>
Glyma.06G213600	isoamylase / debranching enzyme	sugar metabolism
Glyma.06G043900	cyclin a (ccna)	seed development
Glyma.01G016700	inositol-pentakisphosphate 2-kinase / ip5 2-kinase	phytic acid biosynthetic pathway
Glyma.02G064300	ribonuclease t2 [ec:3.1.27.1] (e3.1.27.1)	nodule development
Glyma.02G293700	N/A	uncharacterized protein
Glyma.06G267400	leucine-rich repeat-containing protein	disease resistance
Glyma.09G075800	N/A	uncharacterized protein
Glyma.13G194800	leucine-rich repeat-containing protein	disease resistance
Glyma.13G194900	leucine-rich repeat-containing protein	disease resistance
Glyma.16G028300	N/A	uncharacterized protein
Glyma.18G166600	phosphatidylinositol glycan, class s (pigs)	protein glycosilation
Glyma.20G138000	cysteine-rich repeat secretory protein	disease resistance

A expressão relativa dos genes diferencialmente expressos citados na Figura 9A, estão descritos no material complementar 1. Esse arquivo mostra um total de 1044 genes, sendo 658 expressos no tratamento T0 e 704 expressos no tratamento CF (Figura 9A). Dos 658 genes expressos em T0, 341 são down-regulados e 317 up-regulados na cultivar preta. Dos 704 genes expressos em CF, 351 são down-regulados e 353 up-regulados na cultivar preta (Arquivo adicional 1). Todos os 1044 genes avaliados estão presentes nas duas versões do genoma referência Willians 82 (Wm82.a2; Wm82.a4).

## 5.2 VIAS METABÓLICAS ENRIQUECIDAS

O grupo de genes com expressão absoluta significativa ( $TPM \geq 3$ ) foi utilizado para identificar as vias metabólicas enriquecidas nas cultivares de tegumento preto e tegumento amarelo. Apenas vias enriquecidas com pelo menos três genes foram consideradas (Figura 10).

A cultivar de tegumento preto BRSMG 715A apresentou um total de 22 vias enriquecidas em suas sementes, enquanto as cultivares de tegumento amarelo apresentaram apenas 14 vias (Figura 10). Para melhor organização, as 22 vias

enriquecidas foram organizadas em dois grupos de genes: Grupo A) número de genes acima de 4% e grupo B) abaixo de 4%.

Grupo A: 18,57% dos genes estão envolvidos em respostas de defesa; 13,57% são proteínas não caracterizadas; 5,71% participam da regulação da transcrição; 5,71% estão associados ao teor de proteína na semente; 5% estão relacionados a biossíntese de aminoácidos; 5% são transportadores; 5% associados a biossíntese de hormônios vegetais; 4,29% são genes responsáveis pela resistência a doenças. No grupo B, foram identificados 3,57% pertencentes ao grupo que regula as enzimas responsáveis pela nodulação de raízes; 3,57% envolvidos no metabolismo do açúcar; 3,57% participam da biossíntese de flavonoides; 3,57% são responsáveis pelo dobramento de proteínas; 2,86% são transportadores de aminoácidos; 2,86% fazem parte do processo de ubiquitilação de proteínas; 2,14% fazem parte da classe de transportadores ABC; 2,14% são compostos antinutricionais; 2,14% estão relacionados ao metabolismo do ácido linolênico; 2,14% fazem parte da iniciação da transcrição; 2,14% participam da biossíntese de ácidos graxos; 2,14% estão relacionados a biossíntese de terpenóides; 2,14% pertencem a proteínas de manutenção de água na célula e 2,14% estão envolvidos no reparo do DNA (Figura 10A).

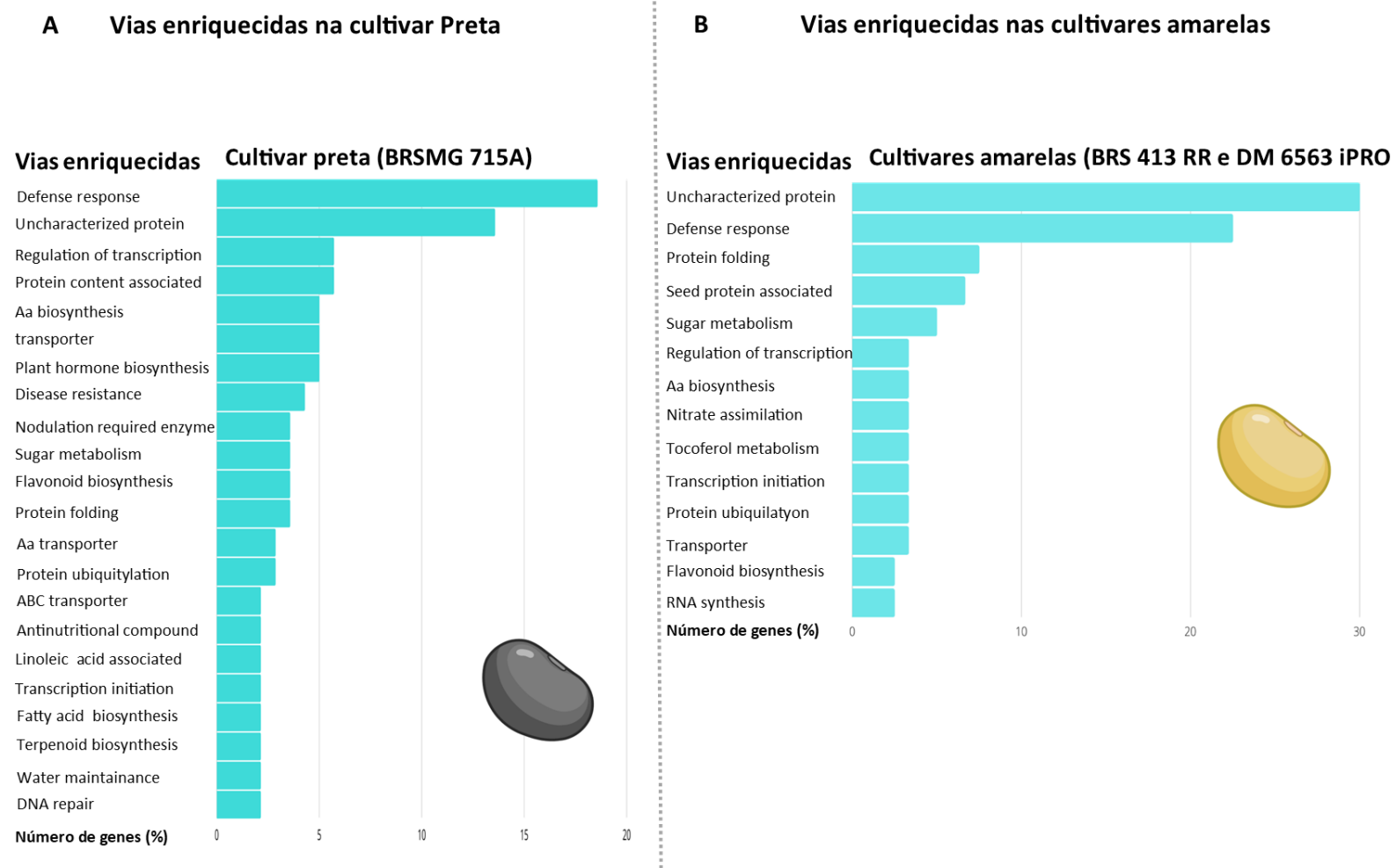
Entre as vias enriquecidas comuns nas cultivares amarelas BRS 413 RR e DM 6563 IPRO (Figura 10B), 30% são proteínas não caracterizadas; 22,50% estão associados a resposta a defesa; 7,5% estão envolvidas no processo de dobramento de proteínas; 6,67% são associados ao teor de proteína na semente; 5,00% participam do metabolismo do açúcar. Ainda, 3,33% estão relacionados à regulação da transcrição; 3,33% estão associados à biossíntese de aminoácidos; 3,33% fazem parte do processo de assimilação de nitrato; 3,33% estão associados a biossíntese de tocoferol; 3,33% estão associados a iniciação da transcrição; 3,33% fazem parte do processo de ubiquitilação de proteínas; 3,33% participam do transporte de solutos na célula; 2,5% são associados a biossíntese de flavonoides e 2,5% estão relacionados a biossíntese de RNA (Figura 10B).

Os resultados obtidos nestas condições identificaram, 11 vias enriquecidas exclusivas da cultivar de tegumento preto BRSMG 715A, sendo elas: biossíntese de hormônios vegetais, resistência a doenças, enzimas necessárias à nodulação, transporte de aminoácidos, transportadores ABC, compostos antinutricionais, metabolismo do ácido linoleico, biossíntese de ácidos graxos, biossíntese de terpenoides, manutenção do teor de água e reparo de DNA, essas vias representam cerca de 30,7% das vias enriquecidas. Já nas cultivares amarelas (BRS 413 RR e DM 6563 IPRO), foram

identificadas apenas três vias exclusivas: assimilação de nitrato, metabolismo do tocoferol e síntese de RNA. Nas cultivares de tegumento amarela, as vias enriquecidas representaram cerca de 9,16%.

Além disso, 11 vias enriquecidas foram comuns para as três cultivares avaliadas, apresentando diferenças entre o número de genes encontrados. A cultivar de tegumento preto BRSMG 715A possui 2,38% mais genes associados a regulação da transcrição, 1,67% mais genes relacionados a biossíntese de aminoácidos, 1,67% mais genes que fazem parte de transportadores e 1,07% mais genes associados a biossíntese de flavonoides. Já as cultivares DM 6563 IPRO e BRS 413 RR de tegumento amarelo, apresentaram 16,43% mais genes que codificam proteínas não caracterizadas, 3,93% mais genes associados a resposta a defesa, 3,93% mais genes relacionados ao dobramento de proteínas, 1,43% mais genes que fazem parte do metabolismo do açúcar, 1,19% que são associados ao processo de iniciação da transcrição, 0,96% mais genes relacionados ao teor de proteína na semente e 0,47% mais genes que fazem parte do processo de ubiquitinação de proteínas.

**Figura 10** - Vias metabólicas enriquecidas na cultivar BRSMG 715A em T0 e CF (A); e nas cultivares BRS 413 RR e DM 6563 IPRO em T0 e CF (B).

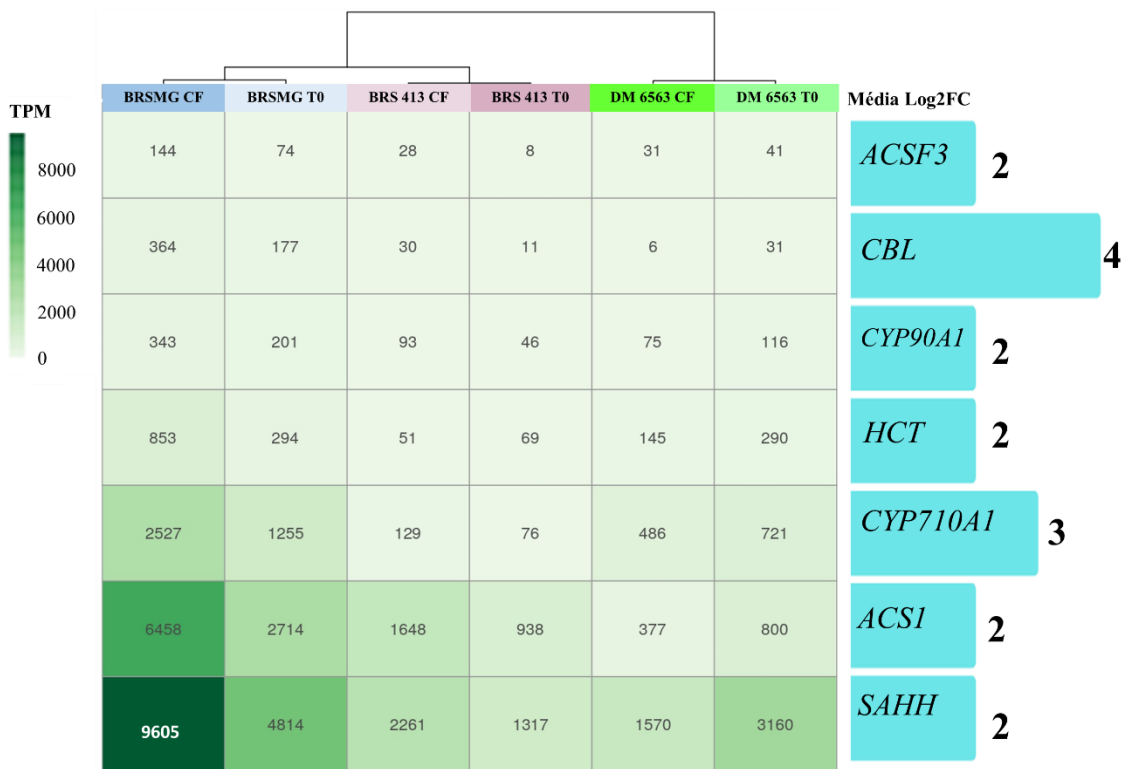


### 5.3 EXPRESSÃO ABSOLUTA E RELATIVA

Para validação por meio de RT-qPCR foram selecionados sete genes com expressão =  $TPM \geq 3$ : s(a) Genes envolvidos na biossíntese de aminoácidos (CBL; SAHH); (b) Genes envolvidos na biossíntese de hormônios vegetais (ACS1; CYP90A1; CYP710A1); (c) Genes envolvidos na biossíntese de ácidos graxos (ACSF3); (d) Genes envolvidos na síntese de lignina (HCT). Os genes foram comuns nos tratamentos T0 e CF e comuns para as três cultivares, além disso, como os genes são mais expressos na cultivar de tegumento preto BRSMG 715A, foi enfatizado explorar diferenças transcricionais entre os genótipos além das já relacionadas com a diversidade de pigmentação do tegumento (Figura 11). Os genes identificados apresentaram funções biológicas interessantes que foram associadas às características atribuídas às cultivares pretas, como maior qualidade fisiológica de semente, maior qualidade de proteína e compostos benéficos a saúde (Krishnan; Jez, 2018, Raksha et al., 2016, Mertz-Henning et al., 2015). Considerando o conjunto de expressão absoluta (TPM) dos sete genes analisados, foi observado maior similaridade entre os tratamentos (T0; CF) dentro de cada cultivar (Figura 3). Em função desta resposta, três clusters (azul, roxo, verde) foram gerados. Esse dado evidencia que a expressão dos genes alvos é semelhante em ambos os tratamentos.

Além disso, o cluster (azul) gerado para a cultivar BRSMG 715A apresenta maior similaridade com o cluster (roxo) da cultivar BRS 413 RR. Estes dados indicam que a cultivar com tegumento preto BRSMG 715A apresenta expressão absoluta mais próxima a cultivar de tegumento amarelo BRS 413 RR. Independente da variação na expressão absoluta (TPM), todos os genes foram em média 4X mais expressos ( $\text{Log}_2\text{FC} = 2$ ) na cultivar de tegumento preto quando comparado as cultivares de tegumento amarelo.

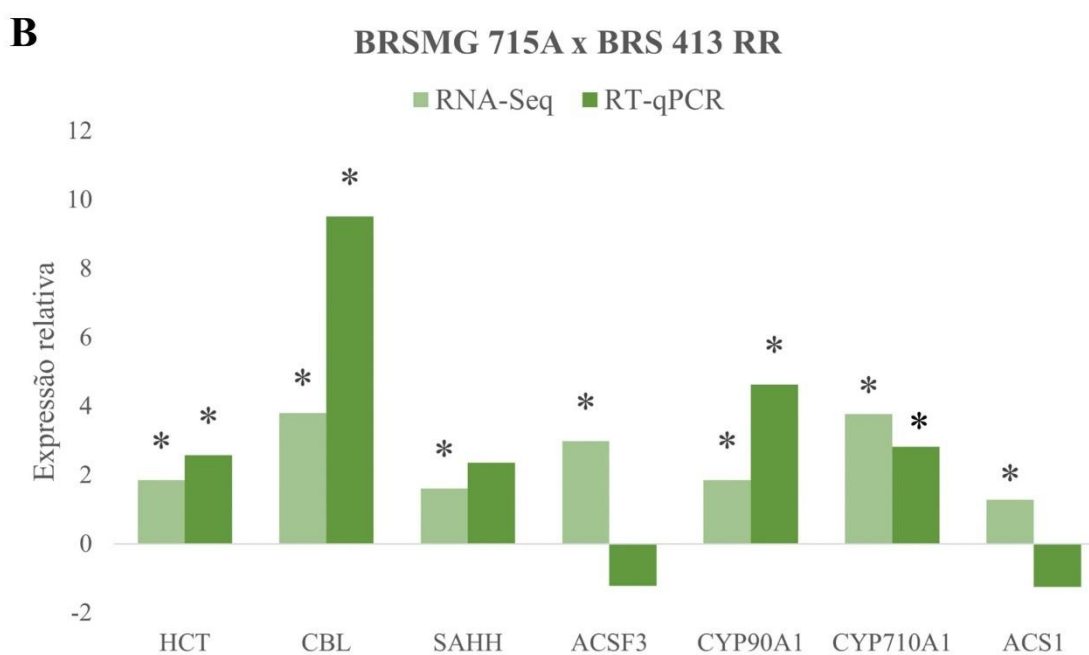
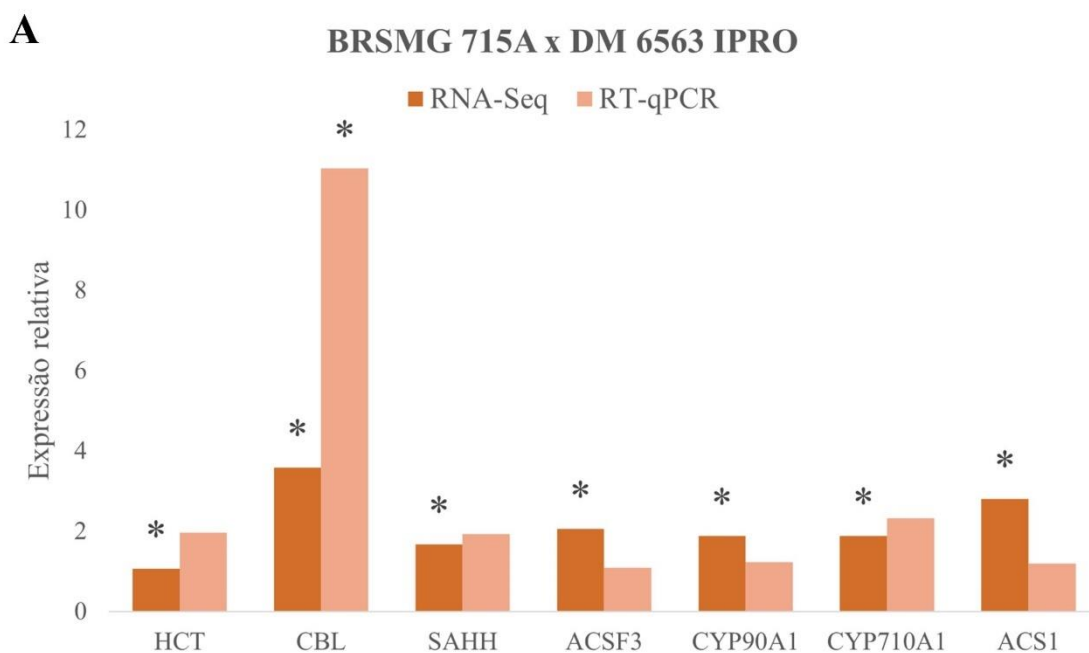
**Figura 11** - Heatmap contendo valores da expressão absoluta em TPM. Barras representam a expressão relativa em  $\text{Log}_2\text{FC}$  dos sete genes alvo. Os genes selecionados são todos regulados positivamente (up-regulados) na cultivar de tegumento preto BRSMG 715A. Os quadros em verde mostram os valores de TPM utilizados para calcular o  $\text{Log}_2\text{FC}$  de cada gene (barra azul).



Estes sete genes foram divididos em quatro grupos de acordo com a sua função biológica: (a) Genes envolvidos na biossíntese de aminoácidos (*CBL*; *SAHH*); (b) Genes envolvidos na biossíntese de hormônios vegetais (*ACSI*; *CYP90A1*; *CYP710A1*); (c) Genes envolvidos na biossíntese de ácidos graxos (*ACSF3*); (d) Genes envolvidos na síntese de lignina (*HCT*).

**Figura 12-** Expressão relativa dos genes *HCT*, *CBL*, *SAHH*, *ACSF3*, *CYP90A1*, *CYP710A1* e *ACSI* em sementes de soja inteiras recém-colhidas. Valores positivos representam expressão na cultivar de tegumento preto BRSMG 715A quando comparada com as cultivares de tegumento amarelo BRS 413 RR (A) e DM 6563 IPRO (B). Para RT-qPCR ( $2^{-\Delta\Delta C_t}$ ) a análise estatística foi realizada por meio do teste T ( $p \leq 0.1$ ). Para a

análise de RNA-Seq (Log2FC)  $FDR \leq 0.05$ . Asteriscos(\*) representam valores com diferença significativa.



### 5.3.1 Genes envolvidos na Biossíntese de aminoácidos

Foram selecionados os genes *CBL* e *SAHH* que fazem parte da biossíntese dos aminoácidos sulfurados metionina e cisteína. A metionina e a cisteína são considerados aminoácidos essenciais, pois não são produzidos e, portanto, devem ser adquiridos a partir da alimentação. Seu consumo é fundamental para a biossíntese de muitas moléculas

importantes para a atuação correta de diversos organismos, como creatina, carnitina, poliaminas, epinefrina, colina e melatonina (Singer et al., 2019, Malle et al., 2020). Além disso, na biossíntese de aminoácidos sulfurados o produto, S-adenosylmethionine é precursor para a metilação de molignóis, processo essencial para a formação de lignina (Bai et al., 2018).

O gene *CBL*, expresso 8X mais na cultivar de tegumento preto (Log2FC 4) pela análise de RNA-Seq, apresentou mesmo padrão de regulação pela análise de RT-qPCR (Figura 12, sendo 11 e 9X mais expressos na cultivar de tegumento preto BRSMG 715A em relação as cultivares de tegumento amarelo DM 6563 IPRO e BRS 413 RR, respectivamente. O gene codifica para a enzima cistationina beta-liase, que realiza as interconversões entre L-homocisteína e L-cisteína. Outros estudos já foram realizados e sugerem que o gene *CBL* é um candidato promissor para o aumento de cisteína e metionina na soja (Krishnan; Jez, 2018, Ma et al., 2019). Foi demonstrado, por meio de RNAi, que a repressão do gene causa redução na biossíntese de metionina, que resulta em menor crescimento e desenvolvimento em batata (Maimann et al., 2001). Além disso, por meio do silenciamento gênico foi verificado que este gene é crucial para o desenvolvimento do embrião e de raízes em *Arabidopsis*, afetando a regulação da expressão gênica e a biossíntese de metionina (Liu et al., 2019).

Já o gene *SAHH*, codifica para a enzima adenosilhomocisteínase, que transforma o produto S-adenosil-L-homocisteína em L-homocysteína e adenosina, possui importância para o crescimento celular e regulação da expressão gênica por meio de metilação do DNA (Ding et al., 2016). O gene apresentou expressão relativa 4X maior na cultivar de tegumento preto BRSMG 715A (Log2FC = 2) do que nas cultivares de tegumento amarelo. Na expressão relativa observada pela análise de RT-qPCR, os genes não apresentaram diferença significativa quando comparada a BRSMG 715A de tegumento preto em comparação com as cultivares de tegumento amarelo DM 6563 IPRO e BRS 413 RR, no entanto, seguiram o mesmo padrão de expressão que foi observado pela análise de RNA-Seq. Além disso, a enzima interage com a formação de lignina, transformando o substrato S-adenosilhomocisteína (inibidor da metilação de molignóis) em L-homocisteína, dessa forma, sua expressão é essencial para a biossíntese de lignina (Bai et al., 2018). A enzima também está envolvida na recuperação de metionina através da formação do produto L-homocisteína pela síntese *de novo* da metionina (Watanabe et al., 2021). A deficiência em um ou mais aminoácidos é suficiente para impactar negativamente no crescimento e desenvolvimento de diferentes organismos (Singer et al.,

2019). Portanto, um dos focos do melhoramento genético têm sido enriquecer a composição dos aminoácidos presentes na soja e não apenas no teor de proteína bruto em si.

### 5.3.2 Genes envolvidos na Biossíntese de Fitorreguladores Vegetais

Com relação aos hormônios vegetais, foram identificados genes que correspondem a biossíntese de etileno e brassinoesteróides.

O gene *ACSI* codifica para a enzima 1-aminociclopropano-1-carboxilato sintase, uma das principais enzimas responsáveis pela biossíntese de etileno. O gene apresentou expressão 4X maior na cultivar de tegumento preto BRSMG 715A (Log<sub>2</sub>FC 2) para a análise de RNA-Seq. Já na análise de expressão relativa por RT-qPCR, os genes não apresentaram diferença significativa quando comparados na cultivar de tegumento preto a BRSMG 715A com as cultivares de tegumento amarelo DM 6563 IPRO e BRS 413 RR (Figura 12AB), no entanto, este gene apresentou expressão superior na cultivar BRS 413 RR (Figura 12B). O etileno é um dos hormônios vegetais que influenciam no crescimento vegetal, desenvolvimento e resposta a estresses bióticos e abióticos. *ACSI* age convertendo o substrato D-adenosil-metionina em ácido 1-aminociclopropano-1-carboxílico (ACC), que posteriormente, é transformado em etileno pela enzima aminociclopropano-1-carboxilato oxidase (ACO) (Pattyn et al., 2021). Etileno é um dos hormônios responsáveis pelo processo de germinação, aumentando a sua concentração e induzindo o crescimento da radícula até um ponto limítrofe, posteriormente, reduz o crescimento da plântula (Ishibashi et al., 2013). Além disso, o etileno participa na formação, manutenção e alongação do gancho plumular. Ele também participa da alongação do hipocótilo (Ahammed et al., 2020).

O gene *CPD* que codifica uma c-3 oxidase (CYP90A1) é um dos genes atuantes na biossíntese de brassinoesteróides (BRs) catalisando a reação entre (22S)-22-hidroxicampesterol e (22R,23R)-22,23-dihidroxicampesterol, bem como a reação entre o substrato 6-desoxocasterona e o produto 6-desoxoteasterona. O gene CYP90A1 apresentou expressão de relativa 4X superior na cultivar BRSMG 715A de tegumento preto em relação às cultivares de tegumento amarelo (Log<sub>2</sub>FC 2). Na análise de expressão relativa por RT-qPCR, o gene apresentou diferença significativa quando foi comparado a cultivar de tegumento preto BRSMG 715A com a cultivar de tegumento amarelo BRS 413 RR. No entanto, quando a cultivar de tegumento preto BRSMG 715A foi comparada

com a cultivar de tegumento amarelo DM6563 IPRO, não apresentou diferença significativa, mas manteve a regulação observada no RNA-Seq (Figura 12AB).

Em genótipos mutantes do gene *cpd* a síntese de BRs é interrompida na etapa anterior a sua atividade (Ohnishi et al., 2012). Na semente, os BRs promovem a germinação e agem em contraste ao ácido abscísico que inibem o processo. Em soja, a forma mais ativa de BR promove a elongação de epicótilo e hipocótilo. Além disso, o hormônio atrasou a senescência foliar e induziu genes que modificam a parede celular e genes relacionados ao AIA, enquanto suprimiu fatores de transcrição WRKY envolvidos na senescência e estresses (Yin et al., 2018). O gene *CPD* é essencial para o desenvolvimento vegetal, a ausência do gene ocasionou nanismo e a recuperação do gene retornam ao gene com desenvolvimento normal (Wang et al., 2015). Ainda, foi verificado que a atuação dos brassinosteróides, pode causar o aumento de lignina bem como alterar a expressão de genes envolvidos na biossíntese de fenilpropanóides (Li et al., 2016, Fan et al., 2020).

Já o gene *CYP710A1* que consiste na esterol 22-dessaturase, uma das principais enzimas na biossíntese de esteróis vegetais, é precursor dos brassinosteróides. Atua no final da via dos esteróis vegetais dando origem aos produtos brassicasterol e stigmasterol. O gene apresentou expressão relativa 6X maior na cultivar BRSMG 715A de tegumento preto (Log2FC 3). Na análise de expressão relativa por RT-qPCR, o gene apresentou diferença significativa quando foi comparado a BRSMG 715A e a cultivar BRS 413 RR, quando foi comparado com a cultivar DM6563 IPRO não apresentou diferença significativa, mas manteve a regulação observada no RNA-Seq (Figura 12AB). A enzima introduz uma ligação dupla no C22 do  $\beta$ -sitosterol que o transforma em stigmasterol (Valitova; Sulkarnayeva; Minibayeva, 2016, Aboobucker; Suza, 2019). A interconversões entre  $\beta$ -sitosterol e stigmasterol, estão associadas a tolerância a estresses bióticos e abióticos por modificarem a permeabilidade da membrana plasmática das células vegetais (Griebel; Zeier, 2010, Wang et al., 2012, Cabianca et al., 2021). Na semente de algodão, o gene induzido durante o desenvolvimento, em conjunto com outros genes que participam da via dos esteróis vegetais (Suo et al., 2021).

Ainda, o stigmasterol possui diversos benefícios para a saúde humana, a superexpressão do gene *CYP710A* gerou um aumento de 2.6X do composto no fruto de tomate (Raksha et al., 2016). Além disso, o gene foi identificado em QTL como candidato ao aumento de número de vagens por planta, assim como outros genes da via dos BR que

influenciam o crescimento e desenvolvimento de vagens de soja (Song et al., 2020, Ariyoshi et al., 2016).

### 5.3.3 Genes envolvidos na Biossíntese de ácidos graxos

Com relação a *ACSF3*, o gene codifica para enzima malonil-CoA/metilmalonil-CoA sintetase catalisando a reação entre o substrato malonato e o produto malonil-CoA, o produto é um dos principais precursores para a síntese e alongação de ácidos graxos, ainda é necessário para a formação de outros compostos como fitoalexinas, flavonóides e antocianinas. O gene apresentou expressão relativa 4X maior na cultivar de tegumento preto BRSMG 715A (Log<sub>2</sub>FC 2). Na expressão relativa observada na análise de RT-qPCR, os genes não apresentaram diferença significativa quando comparada com as cultivares de tegumento amarelo DM6563 IPRO e BRS 413 RR, entretanto, o gene apresentou maior expressão relativa na cultivar BRS 413 RR (Figura 12B). O gene é considerado essencial para desenvolvimento e crescimento vegetal, em *Arabidopsis*, mutantes deficientes da enzima apresentaram crescimento retardado e acumularam ácido malônico e succinico (Chen et al., 2011). O malonil-CoA, é um dos precursores necessários para enzima *CHS* (primeira enzima na via de formação de flavonoides e isoflavonoides), que condensa três moléculas de malonil-CoA em conjunto com uma molécula de *p*-coumaroil-CoA, que origina o restante dos flavonóides na semente de soja (Dastmalchi & Dhaubhadel, 2015). Entre eles os produtos oriundos da via dos flavonóides estão as antocianinas, que confere a pigmentação preta na cultivar BRSMG 715A.

### 5.3.4 Genes envolvidos na Biossíntese de lignina

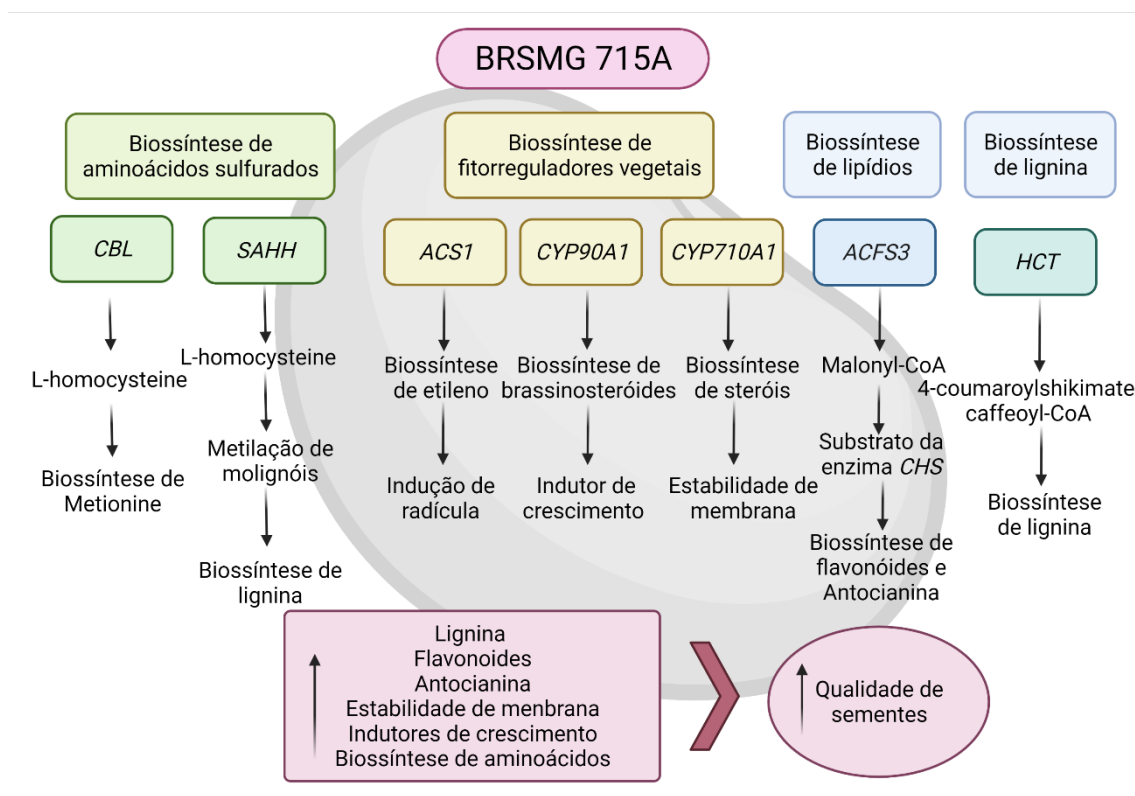
Com relação a biossíntese de lignina, foi selecionado o gene *HCT* que codifica para a enzima shiquimato O-hidroxicinamoiltransferase, sendo uma das principais enzimas responsáveis pela biossíntese de lignina, catalisando duas reações entre 4-coumaroil-CoA para 4-coumaroilshikimato, e a reação entre o substrato cafeoil shikimate no produto cafeoil-CoA. O gene foi 4X mais expresso na cultivar BRSMG 715A de tegumento preto (Log<sub>2</sub>FC 2). Na análise de expressão relativa por RT-qPCR, o gene apresentou diferença significativa quando comparado a BRSMG 715A e a cultivar BRS 413 RR (Figura 12B). Quando foi comparado com a cultivar DM6563 IPRO não

apresentou diferença significativa, mas manteve a regulação observada no RNA-Seq (Figura 12A). Segundo estudo realizado por Mertz-Henning et al., (2015) o gene foi reprimido no tegumento da semente, em diferentes estádios de desenvolvimento da semente. Vários estudos demonstraram que o silenciamento por meio de RNAi e nocaute do gene *HCT* resulta em menor teor de lignina e redução do desenvolvimento em plantas (Shadle et al., 2007, Besseau et al., 2007, Bhattarai et al., 2018, Ha et al., 2019, Serrani-Yarce et al., 2021). Maiores teores de lignina na semente estão associados a maior qualidade fisiológica, redução de danos oriundos de umidade, menor dano causado por fungos de armazenamento e menor danos mecânicos (Huth et al., 2016, Bellaloui et al., 2012, Kuchlan et al., 2018). Ainda, variedades de soja com tegumento preto, possuem melhor armazenabilidade devido a possuírem poros menores na membrana e maior teor de lignina (Kuchlan; Dadlani; Samuel, 2010). A cultivar BRSMG 715A possui teor elevado de lignina, considerado um dos fatores determinantes para maior qualidade de semente (Abati et al., 2021).

Em resumo, os genes induzidos na cultivar preta possuem importância para a qualidade de sementes (Figura 13). As enzimas *CBL* e *SAHH* catalisam reações importantes para a biossíntese de aminoácidos sulfurados em especial a metionina, o nocaute e silenciamento do gene *CBL* acarretou menor desenvolvimento do embrião (Maimann et al. 2001, Liu et al., 2019). Ainda, o gene *SAHH* possui importância para a biossíntese de lignina, sendo necessário para a metilação dos molignóis (Bai et al., 2018). Já os genes *ACSI* e *CYP90A1* fazem parte do grupo de biossíntese de fitorreguladores vegetais, e possuem comportamento indutor na germinação, já o gene *CYP710A1* faz parte dos esteróis vegetais, que são responsáveis pela estabilidade da membrana (Ishibashi et al., 2013, Wang et al., 2015, Suo et al., 2021). O gene *ACSF3* faz parte do metabolismo de lipídios e possui grande importância para a síntese de flavonoides e antocianina, como substrato enzima *CHS* (Chen et al., 2011, Dastmalchi & Dhaubhadel, 2015). Com relação ao gene *HCT*, ele é responsável pela polimerização de 4-coumaroilshikimato e cafeoil-CoA, ambos necessários para síntese dos diferentes tipos de lignina, o silenciamento do gene causa menor teor de lignina e redução do desenvolvimento (Bhattarai et al., 2018, Ha et al., 2019, Serrani-Yarce et al., 2021).

**Figura 13** - Esquematização dos genes induzidos na cultivar de tegumento preto BRSMG 715A preto em relação as cultivares de tegumento amarelo BRS 413 RR e DM 6563 IPRO

com suas respectivas contribuições para aumento na qualidade e longevidade de sementes.



#### 5.4 AVALIAÇÃO DO NÚMERO DE CÓPIAS E ORTÓLOGOS EM ARABDOPSIS

Os genes também foram avaliados conforme seus ortólogos, cópias encontradas no genoma da soja e domínio proteico. Os genes selecionados possuem duas cópias, com exceção do gene *SAHH* (Glyma.08G108800) que possui três cópias. Além disso, apresentam pelo menos 90% de similaridade com o gene alvo (Tabela 3). Essa similaridade entre os genes está relacionada com a duplicação do genoma da soja, que possui cerca de 75% dos genes com múltiplas cópias (Schmutz et al., 2010). A duplicação é um processo evolutivo que pode causar subfuncionalização gênica, gerando acúmulo de cópias de genes não funcionais (pseudogenes).

Com relação aos ortólogos encontrados em *Arabidopsis*, apenas os genes *CYP710A1* (Glyma.13G217400) e *HCT* (Glyma.02G254600) apresentaram ortólogos por sintenia (derivam de um mesmo ancestral comum por especiação) diretos. O *CYP710A1* foi descrito em *Arabidopsis* como responsável pela conversão entre  $\beta$ -sitosterol e stigmasterol (Arnqvist et al., 2006, Cabianca et al., 2016), é provável que os genes

apresentam a mesma função. Já o gene HCT, foi descrito em *Arabidopsis* catalisando entre 4-coumaroyl-CoA para 4-coumaroylshikimate, e a reação entre o substrato caffeoyl shikimate no produto caffeoyl-CoA (Hoffman et al., 2004, Eudes et al., 2016). Ambos os genes avaliados apresentam 70% de similaridade com o ortólogo em *Arabidopsis*. Enquanto os outros apresentaram o *Best-hit-arabi-name* que corresponde a sequência com maior similaridade físico-química em *Arabidopsis*, com pelo menos 60% de similaridade. Além disso, genes apresentaram domínios proteicos conservados entre soja e *Arabidopsis* o que representa que os genes cumprem funções semelhantes entre essas espécies.

**Tabela 3** -Avaliação do número de cópias, ortólogos e domínio proteico entre soja e Arabdopsis.

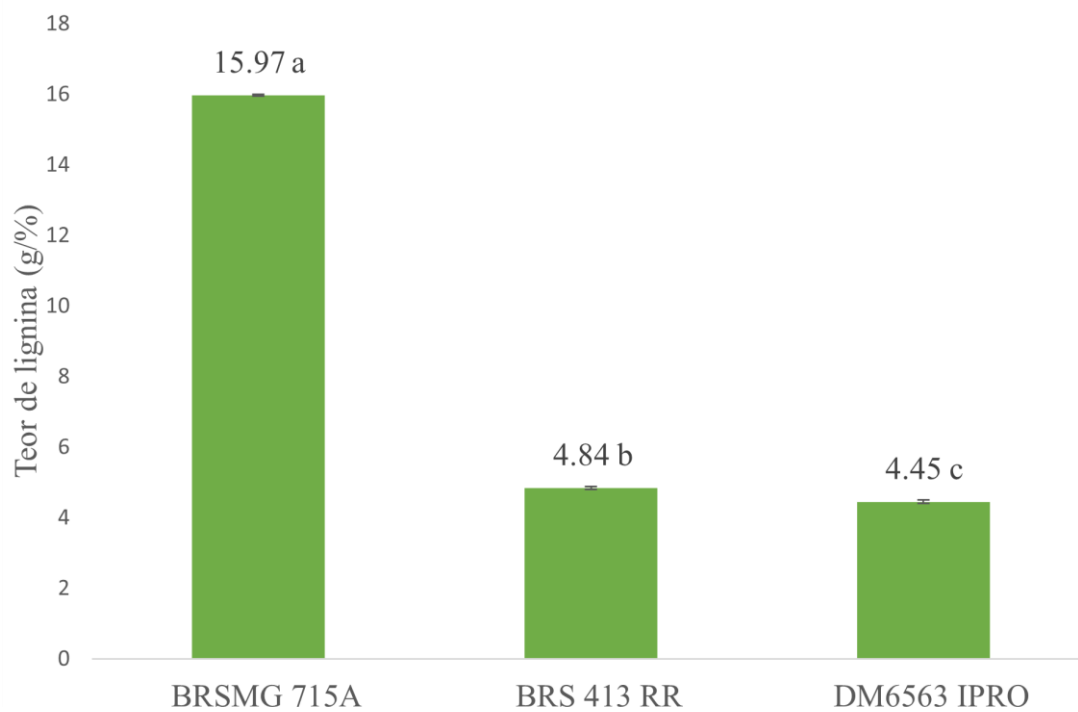
<b>GeneID</b>	<b>NCS</b>	<b>Copia</b>	<b>%S</b>	<b>Ortólogo</b>	<b>%S</b>	<b>Best-hit-arabi</b>	<b>S%</b>	<b>Dominio proteico</b>
Glyma.11G228900	2	Glyma.18G028300.1	92	N/A	N/A	AT5G05690.1	80	p450
Glyma.13G217400	2	Glyma.15G095000.1	90	AT2G34500.1	70	N/A	N/A	p450
Glyma.08G204300	2	Glyma.07G019100.1	94	N/A	N/A	AT3G16170.1	61	AMP-binding enzyme
Glyma.03G129700	2	Glyma.19G132000.1	92	N/A	N/A	AT3G57050.1	80	Cys/Met metabolism PLP-dependent enzyme
Glyma.02G254600	2	Glyma.14G061800.1	92	AT2G39980.1	70	N/A	N/A	Transferase family
Glyma.16G032200	2	Glyma.07G065700	96	N/A	N/A	AT3G61510.1	63	Aminotransferase class I and II
Glyma.08G108800	3	Glyma.05G152000.1/G lyma.11G254700.1	99/ 93	N/A	N/A	AT4G13940.1	93	S-adenosyl-L-homocysteine hydrolase

**Legenda:** NCS- Número de cópias na soja; S%-similaridade; N/A- Não está disponível.

## 5.5 AVALIAÇÃO DO TEOR DE LIGNINA NA SEMENTE

As sementes foram avaliadas conforme o teor de lignina, a cultivar de tegumento preto BRSMG 715A apresentou 15.97 % de lignina obtendo o maior teor entre as cultivares de tegumento amarelo BRS 413 RR e DM 6563 IPRO (Figura 14). Da mesma forma a cultivar BRS 413 RR diferiu da DM 6563 com teor de lignina de 4.84 g/% em relação a 4.45 g/% (Figura 14). Esses resultados corroboram o que foi observado em outros estudos para a cultivar BRSMG 715A e outros genótipos de tegumento preto (Bahry et al., 2015, Adsul; Chimote; Deshmuk, 2018, Abati et al., 2021). Ainda, cultivares de tegumento preto destacam-se por possuírem maior concentração de compostos fenólicos como antocianina e flavonóides (Ciabotti et al., 2019). Ademais, com elevado teor de lignina possuem menor permeabilidade no tegumento e absorvem menos água, a absorção de água durante ao armazenamento está diretamente relacionada com a deterioração da semente, especialmente em ambiente não controlado (Abati et al., 2022). Dessa forma, cultivares de tegumento preto foram associadas a melhor armazenabilidade e longevidade de sementes, tanto pelos potenciais antioxidantes quanto pelo teor de lignina (Liu et al., 2017, Adsul; Chimote; Deshmuk, 2018, Chandra et al., 2022).

**Figura 14** - Teor de lignina na semente de soja da cultivar de tegumento preto BRSMG 715A e das cultivares amarelas BRS 413 RR e DM 6563 IPRO, letras distintas demonstram diferença significativa pelo teste de Tukey ( $p \leq 0.05$ ).



## 6. CONCLUSÕES

Foram encontrados 265 genes expressos em todas as cultivares, sendo 12 genes exclusivos da cultivar de tegumento preto (BRS MG715A);

Os genes foram prospectados com relação a sua função e ontologia gênica, dessa forma, verificou-se vias induzidas na cultivar BRSMG 715A que podem estar relacionadas a maior qualidade de sementes e desenvolvimento de plântula;

O gene *CBL* foi induzido na cultivar BRSMG 715A de tegumento preto em relação as cultivares de tegumento amarelo, enquanto os genes *CYP710A1*, *CYP90A1* e *HCT* foram superexpressos na cultivar de tegumento preto em relação a cultivar BRS 413 RR;

A cultivar BRSMG 715A possui maior teor de lignina na semente comparado as outras cultivares de tegumento amarelo.

## **7. CONSIDERAÇÕES FINAIS**

Os genes encontrados estão associados a maior qualidade de sementes de plântulas, além disso, podem estar associados a biossíntese de aminoácidos sulfurados, importantes para a qualidade de proteína da soja. Ainda, maiores diferenças foram encontradas entre a cultivar BRSMG 715A e a cultivar DM 6563 IPRO, que possuem comportamento distinto durante o armazenamento e teores de lignina contrastantes, comparados com relação a expressão absoluta em RNA-Seq, a cultivar BRS 413 RR possui comportamento intermediário, foi mais próxima a cultivar BRSMG 715A. Portanto, mais estudos devem ser desenvolvidos a fim de revelar mais sobre o comportamento desses genes durante o período de armazenamento.

## REFERÊNCIAS

- ABATI, J. et al. Metabolites of the phenylpropanoid pathway and physiological quality of soybean seeds in storage. **Journal of Seed Science**, v. 43, 8 nov. 2021a.
- ABATI, J. et al. Water absorption and storage tolerance of soybean seeds with contrasting seed coat characteristics. **Acta Scientiarum. Agronomy**, v. 44, n. 1, p. e53096, 21 dez. 2021b.
- ADSUL, A. T.; CHIMOTE, V. P.; DESHMUKH, M. P. Inheritance of Seed Longevity and Its Association with Other Seed-Related Traits in Soybean (*Glycine max*). **Agricultural Research**, v. 7, n. 2, p. 105–111, 2018.
- AHAMMED, G. J. et al. Role of ethylene crosstalk in seed germination and early seedling development: A review. **Plant Physiology and Biochemistry**, v. 151, p. 124–131, 1 jun. 2020.
- ALAMU, E. O. et al. Assessment of nutritional characteristics of products developed using soybean (*Glycine max* (L.) Merr.) pipeline and improved varieties. <http://www.editorialmanager.com/cogentagri>, v. 3, n. 1, p. 1398042, 2017.
- ANDERSON, E. J. et al. Soybean [*Glycine max* (L.) Merr.] Breeding: History, Improvement, Production and Future Opportunities. **Advances in Plant Breeding Strategies: Legumes**, v. 7, p. 431–516, 8 out. 2019.
- ARIYOSHI, Y. et al. Regulation of brassinosteroid on pod growth through cell hypertrophy in soybean (*Glycine max* (L.) Merr.). **Plant Growth Regulation**, v. 80, n. 3, p. 391–395, 2016.
- AZIZKHANI, N.; MIRZAEI, S.; TORKZADEH-MAHANI, M. Genome-wide identification and characterization of legume T2 Ribonuclease gene family and analysis of GmaRNS9, a soybean T2 Ribonuclease gene, function in nodulation. **3 Biotech 2021 11:12**, v. 11, n. 12, p. 1–21, 15 nov. 2021.
- BAHRY, C. A. et al. Chemical composition and structural characterization of contrasting colors of soybean seed coats. **Semina: Ciências Agrárias**, v. 36, n. 3, p. 1913–1926, 2015.
- BAHRY, C. A. et al. Physiological quality and imbibitions of soybean seeds with contrasting coats. **Revista Ciencia Agronomica**, v. 48, n. 1, p. 125–133, 2017.
- BAI, Z. et al. Alteration of S-adenosylhomocysteine levels affects lignin biosynthesis in switchgrass. **Plant Biotechnology Journal**, v. 16, n. 12, p. 2016–2026, 1 dez. 2018.
- BARROS, J. et al. The cell biology of lignification in higher plants. **Annals of Botany**,

v. 115, n. 7, p. 1053–1074, 1 jun. 2015.

BELKHADIR, Y. et al. The growth–defense pivot: crisis management in plants mediated by LRR-RK surface receptors. **Trends in Biochemical Sciences**, v. 39, n. 10, p. 447–456, 1 out. 2014.

BELLALOU, N. et al. Soybean Seed Composition Constituents as Affected by Drought and Phomopsis in Phomopsis Susceptible and Resistant Genotypes. <http://dx.doi.org/10.1080/15427528.2011.651774>, v. 26, n. 3, p. 428–453, maio 2012.

BESSEAU, S. et al. Flavonoid Accumulation in Arabidopsis Repressed in Lignin Synthesis Affects Auxin Transport and Plant Growth. **The Plant Cell**, v. 19, n. 1, p. 148–162, 27 fev. 2007.

BHATTARAI, K. et al. Agronomic Performance and Lignin Content of HCT Down-Regulated Alfalfa (*Medicago sativa* L.). **Bioenergy Research**, v. 11, n. 3, p. 505–515, 1 set. 2018.

CABIANCA, A. et al. Changes in the plant  $\beta$ -sitosterol/stigmasterol ratio caused by the plant parasitic nematode *meloidogyne incognita*. **Plants**, v. 10, n. 2, p. 1–15, 2021.

CANTERI, M. G. et al. SASM-Agri - Sistema para análise e separação de médias em experimentos agrícolas pelos métodos Scott - Knott, Tukey e Duncan. 2018.

CARVALHO, E. R. et al. Mn foliar sobre a qualidade sanitária e lignina de sementes de soja convencional e resistente ao glifosato 1 Foliar manganese in the health and lignin quality of conventional and glyphosate-resistant soybean seeds. v. 46, n. 1, p. 135–143, [s.d.].

CASTRO, E. DE M. et al. Lignin and Activity Enzymatic in Susceptibility to Weathering Damage on Soybean Seeds. **Journal of Agricultural Science**, v. 11, n. 11, p. 203, 2019.

CHANDRA, S. et al. Seed longevity studies in wild type, cultivated and inter-specific recombinant inbred lines (RILs) of soybean [*Glycine max* (L.) Merr.]. **Genetic Resources and Crop Evolution**, v. 69, n. 1, p. 399–409, 1 jan. 2022.

CHANG, X. et al. Diversity of the Seedborne Fungi and Pathogenicity of *Fusarium* Species Associated with Intercropped Soybean. **Pathogens 2020, Vol. 9, Page 531**, v. 9, n. 7, p. 531, 1 jul. 2020.

CHEN, H. et al. Malonyl-CoA synthetase, encoded by *ACYL ACTIVATING ENZYME13*, is essential for growth and development of *Arabidopsis*. **Plant Cell**, v. 23, n. 6, p. 2247–2262, 2011.

CHO, K. M. et al. Soluble phenolics and antioxidant properties of soybean (*Glycine max* L.) cultivars with varying seed coat colours. **Journal of Functional Foods**, v. 5, n. 3, p.

1065–1076, 1 jul. 2013.

CHO, Y. B.; JONES, S. I.; VODKIN, L. O. Mutations in Argonaute5 illuminate epistatic interactions of the K1 and I loci leading to saddle seed color patterns in glycine max. **Plant Cell**, v. 29, n. 4, p. 708–725, 2017.

CIABOTTI, S. et al. Chemical composition, protein profile, and isoflavones content in soybean genotypes with different seed coat colors. **International Food Research Journal**, v. 23, n. 2, p. 621–629, 2016.

**Conab - Safra Brasileira de Grãos**. Disponível em: <<https://www.conab.gov.br/info-agro/safras/graos>>. Acesso em: 4 nov. 2021.

CONESA, A. et al. A survey of best practices for RNA-seq data analysis. **Genome Biology** 2016 **17:1**, v. 17, n. 1, p. 1–19, 26 jan. 2016.

CORADI, P. C. et al. Soybean seed storage: Packaging technologies and conditions of storage environments. **Journal of Stored Products Research**, v. 89, p. 101709, 1 dez. 2020.

COSTA-SILVA, J.; DOMINGUES, D.; LOPES, F. M. RNA-Seq differential expression analysis: An extended review and a software tool. **PLOS ONE**, v. 12, n. 12, p. e0190152, 1 dez. 2017.

DASTMALCHI, M.; BERNARDS, M. A.; DHAUBHADEL, S. Twin anchors of the soybean isoflavonoid metabolon: evidence for tethering of the complex to the endoplasmic reticulum by IFS and C4H. **The Plant journal : for cell and molecular biology**, v. 85, n. 6, p. 689–706, 1 mar. 2016.

DASTMALCHI, M.; DHAUBHADEL, S. Proteomic insights into synthesis of isoflavonoids in soybean seeds. **PROTEOMICS**, v. 15, n. 10, p. 1646–1657, 1 maio 2015.

DENG, J. et al. Seed quality deterioration dynamics for isoflavones biosynthesis in soybean (*Glycine max* L. Merr.) seeds against field mildew stress. **Acta Physiologiae Plantarum**, v. 41, n. 5, p. 1–9, 1 maio 2019.

DING, Y. et al. Identification and functional characterization of the sulfate transporter gene GmSULTR1;2b in soybean. **BMC Genomics**, v. 17, n. 1, p. 1–19, 20 maio 2016.

DIXON, R. A.; BARROS, J. Lignin biosynthesis: old roads revisited and new roads explored. **Open Biology**, v. 9, n. 12, 1 dez. 2019.

DUKARIYA, G. et al. Soybean and Its Products: Nutritional and Health Benefits Cite this paper Related papers Pulses Facing the New Age: Functional Compounds on Gene Expression and Health Connection Dave Oomah Soy, Soy Foods and Their Role in

Vegetarian Diets Soybean and Its Products: Nutritional and Health Benefits. [s.d.].

EUDES, A. et al. Exploiting the Substrate Promiscuity of Hydroxycinnamoyl-CoA:Shikimate Hydroxycinnamoyl Transferase to Reduce Lignin. **Plant and Cell Physiology**, v. 57, n. 3, p. 568–579, 1 mar. 2016.

FANG, Z.; MARTIN, J.; WANG, Z. Statistical methods for identifying differentially expressed genes in RNA-Seq experiments. **Cell and Bioscience**, v. 2, n. 1, p. 1–8, 31 jul. 2012.

**FAOSTAT**. Disponível em: <[https://www.fao.org/faostat/en/#rankings/countries\\_by\\_commodity](https://www.fao.org/faostat/en/#rankings/countries_by_commodity)>. Acesso em: 11 jan. 2022.

FERREIRA, C. D. et al. Changes in Phenolic Acid and Isoflavone Contents during Soybean Drying and Storage. **Journal of Agricultural and Food Chemistry**, v. 67, n. 4, p. 1146–1155, 30 jan. 2019.

FRANÇA-NETO, J. DE B. et al. Documentos 380 Empresa Brasileira de Pesquisa Agropecuária Embrapa Soja Ministério da Agricultura, Pecuária e Abastecimento. 2016.

GANESAN, K.; XU, B. A critical review on polyphenols and health benefits of black soybeans. **Nutrients**, v. 9, n. 5, p. 1–17, 2017.

GARG, R.; JAIN, M. Legume Genomics. **Methods in molecular biology (Clifton, N.J.)**, v. 1069, p. 43–58, 2013.

GILLMAN, J. D. et al. Loss-of-function mutations affecting a specific Glycine max R2R3 MYB transcription factor result in brown hilum and brown seed coats. **BMC Plant Biology**, v. 11, 9 nov. 2011.

GOODSTEIN, D. M. et al. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, n. D1, p. D1178–D1186, 1 jan. 2012.

GORISSEN, S. H. M. et al. Protein content and amino acid composition of commercially available plant-based protein isolates. **Amino Acids**, v. 50, n. 12, p. 1685–1695, 1 dez. 2018.

GRIEBEL, T.; ZEIER, J. A role for  $\beta$ -sitosterol to stigmasterol conversion in plant-pathogen interactions. **Plant Journal**, v. 63, n. 2, p. 254–268, 2010.

GWINNER, R. et al. Genetic diversity in Brazilian soybean germplasm. **Crop Breeding and Applied Biotechnology**, v. 17, n. 4, p. 373–381, 2017.

HA, C. M. et al. Ectopic Defense Gene Expression Is Associated with Growth Defects in *Medicago truncatula* Lignin Pathway Mutants. **Plant Physiology**, v. 181, n. 1, p. 63–84, 29 ago. 2019.

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.

HEGSTAD, J. M. et al. Introgression of novel genetic diversity to improve soybean yield. **Theoretical and Applied Genetics** 2019 **132:9**, v. 132, n. 9, p. 2541–2552, 17 jun. 2019.

HRDLICKOVA, R.; TOLOUE, M.; TIAN, B. RNA-Seq methods for transcriptome analysis. **Wiley Interdisciplinary Reviews: RNA**, v. 8, n. 1, p. e1364, 1 jan. 2017.

HSIEH, S. L. et al. By-Products of the Black Soybean Sauce Manufacturing Process as Potential Antioxidant and Anti-Inflammatory Materials for Use as Functional Foods. **Plants** 2021, **Vol. 10, Page 2579**, v. 10, n. 12, p. 2579, 25 nov. 2021.

HU, M.; WIATRAK, P. Effect of Planting Date on Soybean Growth, Yield, and Grain Quality: Review. **Agronomy Journal**, v. 104, n. 3, p. 785–790, 1 maio 2012.

HUTH, C. et al. Susceptibility to weathering damage and oxidative stress on soybean seeds with different lignin contents in the seed coat 1. **Journal of Seed Science**, p. 296–304, 2016.

HWANG, E. Y. et al. A genome-wide association study of seed protein and oil content in soybean. **BMC Genomics**, v. 15, n. 1, p. 1–12, 2 jan. 2014.

ISHIBASHI, Y. et al. Regulation of soybean seed germination through ethylene production in response to reactive oxygen species. **Annals of Botany**, v. 111, n. 1, p. 95–102, 1 jan. 2013.

JONES, S. I.; VODKIN, L. O. Using RNA-Seq to Profile Soybean Seed Development from Fertilization to Maturity. **PLoS ONE**, v. 8, n. 3, 2013.

KANEHISA, M.; GOTO, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, v. 28, n. 1, p. 27–30, 1 jan. 2000.

KHOSRAVI, A.; RAZAVI, S. H. Therapeutic effects of polyphenols in fermented soybean and black soybean products. **Journal of Functional Foods**, v. 81, p. 104467, 1 jun. 2021.

KIM, D.; LANGMEAD, B.; SALZBERG, S. L. HISAT: a fast spliced aligner with low memory requirements. **Nature methods**, v. 12, n. 4, p. 357, 31 mar. 2015.

KIM, M. Y. et al. Tracing soybean domestication history: From nucleotide to genome. **Breeding Science**, v. 61, n. 5, p. 445–452, 2012.

KOH, K.; YOUN, J. E.; KIM, H. S. Identification of anthocyanins in black soybean (*Glycine max* (L.) Merr.) varieties. **Journal of Food Science and Technology**, v. 51, n. 2, p. 377–381, 1 fev. 2014.

- KRATZ, A.; CARNINCI, P. The devil in the details of RNA-seq. **Nature Biotechnology**, v. 32, n. 9, p. 882–884, 2014.
- KRISHNAN, H. B.; JEZ, J. M. Review: The promise and limits for enhancing sulfur-containing amino acid content of soybean seed. **Plant Science**, v. 272, p. 14–21, 1 jul. 2018.
- KRZYZANOWSKI, F. C.; FRANÇA-NETO, J. D. B. Looking at Lignin its physiological performance. n. April, p. 1999–2001, 2018.
- KRZYZANOWSKI, F. C.; FRANÇA-NETO, J. DE B.; HENNING, A. A. A alta qualidade da semente de soja: fator importante para a produção da cultura. **Circular Técnica** **136**, v. 1, n. Londrina PR, p. 24, 2018.
- KUCHLAN, M. K.; DADLANI, M.; SAMUEL, D. V. K. Seed coat properties and longevity of soybean seeds. **Journal of New Seeds**, v. 11, n. 3, p. 239–249, 2010.
- KURAUCHI, T. et al. Endogenous RNA interference of chalcone synthase genes in soybean: Formation of double-stranded RNA of GmIRCHS transcripts and structure of the 5' and 3' ends of short interfering RNAs. **Journal of Plant Physiology**, v. 168, n. 11, p. 1264–1270, 2011.
- LANDGRAF. **SEEDNews**. Disponível em: <<https://seednews.com.br/artigos/3364-bancos-de-germoplasma-seguranca-e-desenvolvimento-edicao-setembro-2020>>. Acesso em: 13 jan. 2022a.
- LANDGRAF. **New black soybeans have more protein than beans and approved taste - Portal Embrapa**. Disponível em: <<https://www.embrapa.br/en/busca-de-noticias/-/noticia/47445298/nova-soja-preta-tem-mais-proteina-que-feijao-e-sabor-aprovado>>. Acesso em: 26 nov. 2020b.
- LANDI, M.; TATTINI, M.; GOULD, K. S. Multiple functional roles of anthocyanins in plant-environment interactions. **Environmental and Experimental Botany**, v. 119, p. 4–17, 1 nov. 2015.
- LEE, J. H.; CHO, K. M. Changes occurring in compositional components of black soybeans maintained at room temperature for different storage periods. **Food Chemistry**, v. 131, n. 1, p. 161–169, 2012.
- LEE, K. J. et al. Phytochemicals and Antioxidant Activity of Korean Black Soybean (*Glycine max* L.) Landraces. **Antioxidants** **2020**, Vol. **9**, Page **213**, v. 9, n. 3, p. 213, 5 mar. 2020.
- LEISNER, C. P.; YENDREK, C. R.; AINSWORTH, E. A. Physiological and transcriptomic responses in the seed coat of field-grown soybean (*Glycine max* L. Merr.)

to abiotic stress. **BMC Plant Biology**, v. 17, n. 1, p. 1–11, 12 dez. 2017.

LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics (Oxford, England)**, v. 25, n. 16, p. 2078–2079, ago. 2009.

LI, H. et al. BpMADS12 gene role in lignin biosynthesis of *Betula platyphylla* Suk by transcriptome analysis. **Journal of Forestry Research**, v. 27, n. 5, p. 1111–1120, 1 out. 2016.

LIMA, J. J. P. et al. Molecular characterization of the acquisition of longevity during seed maturation in soybean. **PLOS ONE**, v. 12, n. 7, p. e0180282, 1 jul. 2017.

LIU, G. et al. Cystathionine beta-lyase is crucial for embryo patterning and the maintenance of root stem cell niche in *Arabidopsis*. 2019.

LIU, J. et al. Metabolism variation and better storability of dark- versus light-coloured soybean (*Glycine max* L. Merr.) seeds. **Food Chemistry**, v. 223, p. 104–113, 2017.

LIU, J.; OSBOURN, A.; MA, P. MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. **Molecular plant**, v. 8, n. 5, p. 689–708, 4 maio 2015.

LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. **Methods**, v. 25, n. 4, p. 402–408, 1 dez. 2001.

MA, Y. et al. QTL Mapping for Protein and Sulfur-Containing Amino Acid Contents Using a High-Density Bin-Map in Soybean (*Glycine max* L. Merr.). **Journal of Agricultural and Food Chemistry**, v. 67, n. 44, p. 12313–12321, 2019.

MAIMANN, S. et al. Transgenic potato plants reveal the indispensable role of cystathionine  $\beta$ -lyase in plant growth and development. **The Plant Journal**, v. 23, n. 6, p. 747–758, 1 set. 2000.

MALLE, S. et al. Genome-wide association identifies several QTLs controlling cysteine and methionine content in soybean seed including some promising candidate genes. **Scientific Reports 2020 10:1**, v. 10, n. 1, p. 1–14, 11 dez. 2020.

MANDARINO, JOSÉ MARCOS GONTIJO OLIVEIRA, MARCELO ALVARES DE A BENASSI, V. DE T.; LEITE, R. S. Qualidade de sementes e grãos comerciais de soja no Brasil - safra 2014/15. 2016.

MARCOLINO-GOMES, J. et al. Diurnal oscillations of soybean circadian clock and drought responsive genes. **PLoS ONE**, v. 9, n. 1, 2014.

MARTIN, L. B. B. et al. **Catalyzing plant science research with RNA-seq** **Frontiers in Plant Science** Frontiers Research Foundation, , 1 abr. 2013. Disponível em:

<www.frontiersin.org>. Acesso em: 17 ago. 2020

MBOFUNG, G. C. Y. et al. Effects of storage temperature and relative humidity on viability and vigor of treated soybean seeds. **Crop Science**, v. 53, n. 3, p. 1086–1095, 2013.

MEDIC, J.; ATKINSON, C.; HURBURGH, C. R. Current knowledge in soybean composition. **JAOCs, Journal of the American Oil Chemists' Society**, v. 91, n. 3, p. 363–384, 2014.

MENEZES, M. DE et al. Aspectos químicos e estruturais da qualidade fisiológica de sementes de soja. **Pesquisa Agropecuária Brasileira**, v. 44, n. 12, p. 1716–1723, 2009.

MERTZ-HENNING, L. M. et al. Relative quantification of gene expression levels associated with lignin biosynthesis in soybean seed coat. **Seed Science and Technology**, v. 43, n. 3, p. 445–455, 2015.

MERTZ-HENNING, L. M. et al. Effect of water deficit-induced at vegetative and reproductive stages on protein and oil content in soybean grains. **Agronomy**, v. 8, n. 1, 2018.

MERTZ, L. M. et al. Structural differences between soybean seeds coat with contrasting permeability. **Revista Brasileira de Sementes**, v. 31, n. 1, p. 23–29, 2009.

MERTZ, L. M. et al. CDNA-AFLP analyses between black and yellow soybean seed coats. **Seed Science and Technology**, v. 38, n. 1, p. 88–95, 2010.

MIKEL, M. A. et al. Genetic Diversity and Agronomic Improvement of North American Soybean Germplasm. **Crop Science**, v. 50, n. 4, p. 1219–1229, 1 jul. 2010.

MILLER, S. S. et al. Hourglass cell development in the soybean seed coat. **Annals of Botany**, v. 106, n. 2, p. 235–242, 1 ago. 2010.

MOLINARI, M. et al. Transcriptome analysis using RNA-Seq from experiments with and without biological replicates : a review. **Revista de Ciências Agrárias**, v. 64, p. 1–13, 2021.

MOREANO, T. B. et al. Changes in the effects of weathering and mechanical damage on soybean seed during storage. **Seed Science and Technology**, v. 39, n. 3, p. 604–611, 2011.

MOREIRA-VILAR, F. C. et al. The Acetyl Bromide Method Is Faster, Simpler and Presents Best Recovery of Lignin in Different Herbaceous Tissues than Klason and Thioglycolic Acid Methods. **PLOS ONE**, v. 9, n. 10, p. e110000, 16 out. 2014.

NAGAMATSU, A. et al. Functional analysis of soybean genes involved in flavonoid biosynthesis by virus-induced gene silencing. **Plant Biotechnology Journal**, v. 5, n. 6,

p. 778–790, nov. 2007.

OANCEA, S.; OPREAN, L.; LUCIAN, ". ANTHOCYANINS, FROM BIOSYNTHESIS IN PLANTS TO HUMAN HEALTH BENEFITS-review. **Acta Universitatis Cibiniensis Series E: FOOD TECHNOLOGY**, v. XV, n. 1, 2011.

OLIVEIRA, G. DE L. T. The geopolitics of Brazilian soybeans. <https://doi.org/10.1080/03066150.2014.992337>, v. 43, n. 2, p. 348–372, 3 mar. 2015.

PAGANO, M. C.; MIRANSARI, M. **Production Worldwide**. [s.l.] Elsevier Inc., 2016.

PATTYN, J.; VAUGHAN-HIRSCH, J.; VAN DE POEL, B. The regulation of ethylene biosynthesis: a complex multilevel control circuitry. **New Phytologist**, v. 229, n. 2, p. 770–782, 1 jan. 2021.

PEI, R. et al. Identification of novel QTL associated with soybean isoflavone content. **The Crop Journal**, v. 6, n. 3, p. 244–252, 1 jun. 2018.

PERTEA, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. **Nature Biotechnology** **2015 33:3**, v. 33, n. 3, p. 290–295, 18 fev. 2015.

POETA, F. B. et al. Seed Water Concentration and Accumulation of Protein and Oil in Soybean Seeds. **Crop Science**, v. 54, n. 6, p. 2752–2759, 1 nov. 2014.

QIN, P.; WANG, T.; LUO, Y. A review on plant-based proteins from soybean: Health benefits and soy product development. **Journal of Agriculture and Food Research**, v. 7, p. 100265, 1 mar. 2022.

RACINE, J. S. RStudio: A Platform-Independent IDE for R and Sweave. **Journal of Applied Econometrics**, v. 27, n. 1, p. 167–172, 1 jan. 2012.

RADIX ASTADI, I.; PAICE, A. G. Black Soybean (*Glycine max* L: Merrill) Seeds' Antioxidant Capacity. In: **Nuts and Seeds in Health and Disease Prevention**. [s.l.] Elsevier Inc., 2011. p. 229–236.

RAKSHA, B. R. et al. Spatio-varietal differences in stigmasterol biosynthesis in tomato and overexpression of a sterol desaturase gene for enhanced stigmasterol production. **In Vitro Cellular & Developmental Biology - Plant** **2016 52:6**, v. 52, n. 6, p. 571–579, 1 dez. 2016.

REESE, P. F.; BOERMA, H. R. Additional genes for green seed coat in soybean. **Journal of Heredity**, v. 80, n. 1, p. 86–88, 1989.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139–140, 1 jan. 2010.

ROCK, C. D. Phenylpropanoid Metabolism. **eLS**, p. 1–18, 17 abr. 2017.

RUTLEDGE, R. G.; CÔTÉ, C. Mathematics of quantitative kinetic PCR and the application of standard curves. **Nucleic Acids Research**, v. 31, n. 16, p. e93–e93, 15 ago. 2003.

SANJUKTA, S.; RAI, A. K. Production of bioactive peptides during soybean fermentation and their potential health benefits. **Trends in Food Science & Technology**, v. 50, p. 1–10, 1 abr. 2016.

SARKAR, S. et al. Abiotic Stresses: Alteration of Composition and Grain Quality in Food Legumes. **Agronomy** **2021**, Vol. 11, Page 2238, v. 11, n. 11, p. 2238, 4 nov. 2021.

SCHMUTZ, J. et al. Genome sequence of the palaeopolyploid soybean. **Nature** **2010** **463:7278**, v. 463, n. 7278, p. 178–183, 14 jan. 2010.

SEDIVY, E. J.; WU, F.; HANZAWA, Y. Soybean domestication: the origin, genetic architecture and molecular bases. **New Phytologist**, v. 214, n. 2, p. 539–553, 1 abr. 2017.

SENDA, M. et al. Accumulation of proanthocyanidins and/or lignin deposition in buff-pigmented soybean seed coats may lead to frequent defective cracking. **Planta**, v. 245, n. 3, p. 659–670, 1 mar. 2017.

SERRANI-YARCE, J. C. et al. Targeting hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase for lignin modification in *Brachypodium distachyon*. **Biotechnology for Biofuels**, v. 14, n. 1, p. 1–17, 1 dez. 2021.

SEVERIN, A. J. et al. RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome. **BMC Plant Biology**, v. 10, n. 1, p. 1–16, 5 ago. 2010.

SHADLE, G. et al. Down-regulation of hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and forage quality. **Phytochemistry**, v. 68, n. 11, p. 1521–1529, 2007.

SHARMA, S. et al. Positional effects on soybean seed composition during storage. **Journal of Food Science and Technology**, v. 50, n. 2, p. 353–359, 1 abr. 2013.

SHARMA, S. et al. Physical characteristics and nutritional composition of some new soybean (*Glycine max* (L.) Merrill) genotypes. **Journal of Food Science and Technology**, v. 51, n. 3, p. 551–557, 2014.

SILVA, V. O. et al. Chemistry and photochemistry of natural plant pigments: the anthocyanins. **Journal of Physical Organic Chemistry**, v. 29, n. 11, p. 594–599, 1 nov. 2016.

SINGER, W. M. et al. Soybean Amino Acids in Health, Genetics, and Evaluation. **Soybean for Human Consumption and Animal Feed**, 7 out. 2019.

SM Journal of Nutrition and Metabolism Gr up SM. [s.d.].

SMANIOTTO, T. A. DE S. et al. Qualidade fisiológica das sementes de soja armazenadas em diferentes condições. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 18, n. 4, p. 446–453, 2014.

SMITH, S. et al. Genetic Diversity and Modern Plant Breeding. p. 55–88, 2015.

SMÝKAL, P. et al. Legume Crops Phylogeny and Genetic Diversity for Science and Breeding. <https://doi.org/10.1080/07352689.2014.897904>, v. 34, p. 43–104, 7 jun. 2014.

SONG, J. et al. Identification and validation of loci governing seed coat color by combining association mapping and bulk segregation analysis in soybean. **PLoS ONE**, v. 11, n. 7, p. 1–19, 2016.

SONG, J. et al. Identification of QTL and genes for pod number in soybean by linkage analysis and genome-wide association studies. **Molecular Breeding**, v. 40, n. 6, 2020.

SONG, J. Y. et al. Evaluation of Genetic Diversity and Comparison of Biochemical Traits of Soybean (*Glycine max* L.) Germplasm Collections. **Plant Breeding and Biotechnology**, v. 1, n. 4, p. 374–384, 30 dez. 2013.

SONG, Q. et al. Fingerprinting soybean germplasm and its utility in genomic research. **G3: Genes, Genomes, Genetics**, v. 5, n. 10, p. 1999–2006, 2015.

STRASSER, R. Plant protein glycosylation. **Glycobiology**, v. 26, n. 9, p. 926–939, 1 set. 2016.

STRICKLER, S. R.; BOMBARELY, A.; MUELLER, L. A. **Designing a transcriptome next-generation sequencing project for a nonmodel plant species** *American Journal of Botany* John Wiley & Sons, Ltd, , 1 fev. 2012. Disponível em: <<https://bsapubs.onlinelibrary.wiley.com/doi/full/10.3732/ajb.1100292>>. Acesso em: 18 ago. 2020

SUO, X. et al. Functions of phytosterols in seed development of upland cotton (*Gossypium hirsutum* L.). **Industrial Crops and Products**, v. 170, p. 113802, 15 out. 2021.

TAHER, H. I. et al. Predicting soybean losses using carbon dioxide monitoring during storage in silo bags. **Journal of Stored Products Research**, v. 82, p. 1–8, 1 jun. 2019.

TAMAGNO, S. et al. Selection for yield shifted the proportion of oil and protein in favor of low-energy seed fractions in soybean. **Field Crops Research**, v. 279, p. 108446, 1 abr. 2022.

TANAKA, Y.; SASAKI, N.; OHMIYA, A. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. **The Plant Journal**, v. 54, n. 4, p. 733–749, 1 maio 2008.

- TODD, J. J.; VODKIN, L. O. Pigmented Soybean. **Plant physiology**, v. 102, p. 663–670, 1993.
- TODD, J. J.; VODKIN, L. O. Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. **Plant Cell**, v. 8, n. 4, p. 687–699, 1996.
- TODESCHINI, M. H. et al. Soybean genetic progress in South Brazil: physiological, phenological and agronomic traits. **Euphytica**, v. 215, n. 7, p. 1–12, 1 jul. 2019.
- TURNER, R. E. et al. Effects of Purple Seed Stain on Seed Quality and Composition in Soybean. **Plants 2020, Vol. 9, Page 993**, v. 9, n. 8, p. 993, 5 ago. 2020.
- TUTEJA, J. H. et al. Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in Glycine max. **Plant Cell**, v. 16, n. 4, p. 819–835, 2004.
- TUTEJA, J. H. et al. Endogenous, tissue-specific short interfering RNAs silence the chalcone synthase gene family in glycine max seed coats WOA. **Plant Cell**, v. 21, n. 10, p. 3063–3077, 1 out. 2009.
- TUTEJA, J. H.; VODKIN, L. O. Structural features of the endogenous CHS silencing and target loci in the soybean genome. **Crop Science**, v. 48, n. SUPPL. 1, 2008.
- VALLIYODAN, B. et al. Genetic diversity and genomic strategies for improving drought and waterlogging tolerance in soybeans. **Journal of Experimental Botany**, v. 68, n. 8, p. 1835–1849, 1 abr. 2017.
- VANHOLME, R. et al. Lignin biosynthesis and its integration into metabolism. **Current Opinion in Biotechnology**, v. 56, p. 230–239, 1 abr. 2019.
- VEITCH, N. C. Isoflavonoids of the leguminosae. **Natural product reports**, v. 30, n. 7, p. 988–1027, 2013.
- VOGT, T. Phenylpropanoid Biosynthesis. **Molecular Plant**, v. 3, n. 1, p. 2–20, 1 jan. 2010.
- WANG, K. et al. Phytosterols play a key role in plant innate immunity against bacterial pathogens by regulating nutrient efflux into the apoplast. **Plant Physiology**, v. 158, n. 4, p. 1789–1802, 2012.
- WANG, L. et al. RNA-seq analyses of multiple meristems of soybean: Novel and alternative transcripts, evolutionary and functional implications. **BMC Plant Biology**, v. 14, n. 1, p. 1–19, 17 jun. 2014a.
- WANG, L. et al. RNA-seq analyses of multiple meristems of soybean: Novel and alternative transcripts, evolutionary and functional implications. **BMC Plant Biology**, v. 14, n. 1, p. 1–19, 17 jun. 2014b.

- WANG, M. et al. Functional analysis of GmCPDs and investigation of their roles in flowering. **PLoS ONE**, v. 10, n. 3, p. 1–25, 2015.
- WANG, M. et al. Parallel selection on a dormancy gene during domestication of crops from multiple families. **Nature Genetics**, v. 50, n. 10, p. 1435–1441, 2018.
- WANG, X. Structure, function, and engineering of enzymes in isoflavonoid biosynthesis. **Functional and Integrative Genomics**, v. 11, n. 1, p. 13–22, 30 mar. 2011.
- WANG, X. et al. Quantitative trait locus analysis of seed sulfur-containing amino acids in two recombinant inbred line populations of soybean. **Euphytica** **2014** **201:2**, v. 201, n. 2, p. 293–305, 5 ago. 2014c.
- WATANABE, M.; CHIBA, Y.; HIRAI, M. Y. Metabolism and Regulatory Functions of O-Acetylserine, S-Adenosylmethionine, Homocysteine, and Serine in Plant Development and Environmental Responses. **Frontiers in Plant Science**, v. 12, p. 593, 7 maio 2021.
- Weathering Damage in Soybean Seeds: Assessment, Seed Anatomy and Seed Physiological Potential on JSTOR.** Disponível em: <[https://www.jstor.org/stable/24642271?casa\\_token=nqpYmmsKtqEAAAAA%3AQc3snLAJ3e8ZoFx4jTt0iqOf\\_HkI311HeFho6dOuVmvyq6UtzzwNWPnRP84hfTw0SkkK UyvIkOn1VAw3NkLJaTdHd-HISSwXqZIn7XJdOT8IAJHbQuxf&seq=1#metadata\\_info\\_tab\\_contents](https://www.jstor.org/stable/24642271?casa_token=nqpYmmsKtqEAAAAA%3AQc3snLAJ3e8ZoFx4jTt0iqOf_HkI311HeFho6dOuVmvyq6UtzzwNWPnRP84hfTw0SkkK UyvIkOn1VAw3NkLJaTdHd-HISSwXqZIn7XJdOT8IAJHbQuxf&seq=1#metadata_info_tab_contents)>. Acesso em: 2 fev. 2022.
- WENG, J. K.; CHAPPLE, C. The origin and evolution of lignin biosynthesis. **New Phytologist**, v. 187, n. 2, p. 273–285, 1 jul. 2010.
- WIJEWARDANA, C.; REDDY, K. R.; BELLALLOUI, N. Soybean seed physiology, quality, and chemical composition under soil moisture stress. **Food Chemistry**, v. 278, p. 92–100, 25 abr. 2019.
- WU, T. et al. Anthocyanins in black rice, soybean and purple corn increase fecal butyric acid and prevent liver inflammation in high fat diet-induced obese mice. **Food & Function**, v. 8, p. 3178, 2017.
- XU, B.; CHANG, S. K. C. Antioxidant capacity of seed coat, dehulled bean, and whole black soybeans in relation to their distributions of total phenolics, phenolic acids, anthocyanins, and isoflavones. **Journal of agricultural and food chemistry**, v. 56, n. 18, p. 8365–8373, 24 set. 2008.
- YAN, F. et al. Loss-of-Function Mutation of Soybean R2R3 MYB Transcription Factor Dilutes Tawny Pubescence Color. **Frontiers in plant science**, v. 10, p. 1809, 31 jan. 2019.

YANG, K. et al. Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean. **Journal of Heredity**, v. 101, n. 6, p. 757–768, 2010.

YIN, W. et al. Brassinosteroid-regulated plant growth and development and gene expression in soybean. **Crop Journal**, v. 7, n. 3, p. 411–418, 2019.

ZABALA, G.; VODKIN, L. Cloning of the pleiotropic T locus in soybean and two recessive alleles that differentially affect structure and expression of the encoded flavonoid 3' hydroxylase. **Genetics**, v. 163, n. 1, p. 295–309, 2003.

ZAHEER, K.; HUMAYOUN AKHTAR, M. An updated review of dietary isoflavones: Nutrition, processing, bioavailability and impacts on human health. **Critical reviews in food science and nutrition**, v. 57, n. 6, p. 1280–1293, 13 abr. 2017.

ZHANG, J. et al. Genome-wide Scan for Seed Composition Provides Insights into Soybean Quality Improvement and the Impacts of Domestication and Breeding. **Molecular Plant**, v. 11, n. 3, p. 460–472, 5 mar. 2018.

ZHANG, S. et al. Cooperative Regulation of Flavonoid and Lignin Biosynthesis in Plants. <https://doi.org/10.1080/07352689.2021.1898083>, v. 40, n. 2, p. 109–126, 2021.

ZHANG, X. et al. Transcriptome sequencing analysis of maize embryonic callus during early redifferentiation. **BMC Genomics**, v. 20, n. 1, p. 1–22, 27 fev. 2019.

ZHOU, Z. et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. **Nature Biotechnology** 2015 33:4, v. 33, n. 4, p. 408–414, 2 fev. 2015.

ZIEGLER, V. et al. Effects of moisture and temperature during grain storage on the functional properties and isoflavone profile of soy protein concentrate. **Food Chemistry**, v. 242, p. 37–44, 1 mar. 2018.