



UNIVERSIDADE
ESTADUAL DE LONDRINA

PAMELA ISAEELY DE LIMA JOAQUIM

**PROSPECÇÃO *IN SILICO* E AVALIAÇÃO DA EXPRESSÃO
DE GENES ASSOCIADOS AO TEOR DE PROTEÍNA EM
GRÃOS DE SOJA**

Londrina
2021

PAMELA ISAEELY DE LIMA JOAQUIM

**PROSPECÇÃO *IN SILICO* E AVALIAÇÃO DA EXPRESSÃO
DE GENES ASSOCIADOS AO TEOR DE PROTEÍNA EM
GRÃOS DE SOJA**

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
2021

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

J62 Joaquim, Pamela Isaely de Lima.
Prospecção in silico e avaliação da expressão de genes associados ao teor de proteína em grãos de soja / Pamela Isaely de Lima Joaquim. - Londrina, 2021. 75 f.: il.

Orientador: Liliane Marcia Mertz Henning.
Coorientador: Mayla Daiane Correa Molinari.
Dissertação (Mestrado em Biotecnologia) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Biotecnologia, 2021. Inclui bibliografia.

1. Glycine max - Tese. 2. teor proteico - Tese. 3. Transportadores de compostos nitrogenados - Tese. 4. RNA-Seq - Tese. I. Henning, Liliane Marcia Mertz. II. Molinari, Mayla Daiane Correa. III. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Biotecnologia. IV. Título.

CDU 66

PAMELA ISAEELY DE LIMA JOAQUIM

**PROSPECÇÃO *IN SILICO* E AVALIAÇÃO DA EXPRESSÃO
DE GENES ASSOCIADOS AO TEOR DE PROTEÍNA EM
GRÃOS DE SOJA**

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. Silvana Regina Rockenbach Marin

Dra. Larissa Giroto

Londrina, 24 de março de 2021

Aos meus pais Márcia Aparecida de Lima e Rodrigo Bonadia Joaquim e a todos que estiveram ligados a essa conquista de alguma forma.

AGRADECIMENTOS

À Deus por todas as conquistas até hoje e pela força para continuar meu caminho.

À minha mãe Márcia Aparecida de Lima, por ser uma mãe incrível, sempre me incentivar, me apoiar, e estar me dando força durante todas as etapas da minha vida e ao meu pai Rodrigo Bonadia Joaquim por todo apoio.

À Capes, pela bolsa concedida.

À minha orientadora Dra. Liliane Marcia Mertz Henning pela oportunidade de realizar o mestrado e por toda dedicação a esse projeto.

À minha co-orientadora Dra. Mayla Daiane Correa Molinari, por todas idéias sugeridas, por toda paciência e compreensão, por todo tempo dedicado ao presente projeto e pela parceria.

Ao programa de Pós-graduação em Biotecnologia pelo apoio, pelos conhecimentos transmitidos por docentes e funcionários do departamento.

À Embrapa, pelo suporte oferecido, pelos materiais, pela dedicação de todos os funcionários da unidade que estiveram envolvidos diretamente ou indiretamente, pela experiência e acompanhamento de outros projetos.

A Silvana Regina Rockenbach Marin, por toda dedicação e auxílio nas atividades realizadas em laboratório.

Ao meu namorado João Kimura pelo apoio e aos meus amigos.

Obrigada a todos que estiveram comigo, mesmo de longe, pois sem o apoio de todos em minha trajetória, não seria possível a realização deste projeto.

JOAQUIM, Pamela Isaely De Lima. **Prospecção *In Silico* e avaliação da expressão de genes associados ao teor de proteína em grãos de soja**. 2021. 74 f. Dissertação (Mestrado em Biotecnologia) - Universidade Estadual De Londrina, Londrina, 2021.

RESUMO

A soja é a principal fonte de proteína vegetal para produção de ração animal. Mas para ser utilizada para tal finalidade, o teor de proteína mínimo do grão deve atingir no mínimo 35%. Do ponto de vista genético, o que se observa é que ao longo dos anos há uma redução no teor de proteína do grão de soja. Assim, em algumas situações, esse teor mínimo não é alcançado. Entender os mecanismos genéticos envolvidos no controle dessa característica pode ser útil para o desenho de estratégias biotecnológicas que visem o aumento do teor de proteína no grão de soja. Dessa forma, este trabalho visa identificar genes que possam estar associados ao teor de proteína dos grãos de soja. O enfoque do trabalho foi voltado a identificar os transportadores de compostos nitrogenados que atuem especialmente em sementes inteiras, tegumento e legume. Os resultados do transcriptoma foram validados por meio de PCR em tempo real, nas cultivares BRS 232 e BRS 284, com alto e baixo teor de proteína, respectivamente. Foram encontrados 76 genes transportadores de compostos nitrogenados em três bibliotecas de RNA-Seq. A partir das análises *in silico* dos transcriptomas, foram selecionados sete genes, validados por RT-qPCR nas fases de 25-100mg, 100-200mg e 400-500mg. Em relação a cultivar de calibração BRS 284, é possível observar que apenas os genes AAP7, AVT3, CAT9 e UMAMIT25 apresentam-se *up*-regulados, em todas as fases de desenvolvimento da semente, na cultivar de maior teor proteico (BRS 232). Em legume, os únicos genes que se apresentaram superexpressos em duas fases, foram o gene CAT9 e o gene UMAMIT25. O gene CAT9 apresentou *up*-regulação nas fases de 25-100mg e 100-200mg. O gene UMAMIT25 apresentou *up*-regulação em legume de 100-200mg e 250-450mg. A superexpressão dos genes UMAMIT25 e CAT9, em futuros estudos, podem acarretar na solução da correlação negativa entre proteína e produtividade.

Palavras-chave: *Glycine max*; teor proteico; transportadores de compostos nitrogenados; RNA-Seq.

JOAQUIM, Pamela Isaely De Lima. ***In silico* prospection and evaluation of the gene expression associated to protein content in soybeans.** 2021. 74 p. Dissertation (Master Degree – Biotechnology) - Universidade Estadual De Londrina, Londrina, 2021.

ABSTRACT

Soybean grain is the main plant source of protein for animal food. But, to be used for that, soybeans need to have at least 35% of protein. Regarding soybean genetic, throughout the years the soybean protein content has been reduced, as the productivity increases. Thus, in some situations, soybean grains have not this minimum protein content. To understand the genetic mechanism involved in grain protein content may be useful to develop biotechnology strategies to increase this characteristic. Therefore, the aim of this work was to identify genes that may be associated to protein content of soybean grain. The work focused in identify carriers of nitrogen compounds that act especially on whole seeds, tegument and pods. The results of the transcriptome were validated by real-time PCR, in cultivars BRS 232 and BRS 284, with high and low protein content, respectively. 76 carrying nitrogenous compounds genes were found in three RNA-Seq libraries. Based on transcriptomes *in silico* analyzes, seven genes were selected, validated by RT-qPCR in the stages of 25-100mg, 100-200mg and 400-500mg. Regarding the BRS 284 calibration cultivar, it is possible to observe that only the AAP7, AVT3, CAT9 and UMAMIT25 genes are up-regulated at all stages of seed development in the cultivar with the highest protein content (BRS 232). In pods, only the genes CAT9 and UMAMIT25 were overexpressed in two phases. The CAT9 gene was up-regulation in the 25-100mg and 100-200mg stages. While, the UMAMIT25 gene exhibited up-regulation in 100-200mg and 250-450mg pods. Overexpression of the UMAMIT25 and CAT9 genes, in further studies, may result in the solution of the negative correlation between protein and yield.

Keywords: *Glycine max*; protein content; nitrogen compounds transporters; RNA-Seq.

LISTA DE FIGURAS

Figura 1	Principais estádios fenológicos da cultura da soja (Fehr; Caviness, 1977; Perdigão et al., 2019)15
Figura 2	Composição média dos grãos de soja (Pípolo, Mandarino, 2016, adaptado pelo autor)18
Figura 3	Representação ilustrativa de compartimentos de um grão de soja (Medic; Atkinson; Hurburgh, 2014, adaptado pelo autor)20
Figura 4	Síntese de Ureídeos (Próprio autor).....26

LISTA DE TABELAS

Tabela 1	Produção de sementes e soja em grão no Brasil na safra de 2008/09 a safra de 2017/18 em mil toneladas (Abrasem, 2019; Conab, 2020, adaptado pelo autor).....	19
Tabela 2	Porcentagem de aminoácidos essenciais e não essenciais na composição das proteínas de soja (Patil <i>et al.</i> , 2017, adaptado pelo autor)	21
Tabela 3	Transportadores de aminoácidos e ureídeos em Arabidopsis e soja (próprio autor)	29
Tabela 4	Genes candidatos a transportadores de aminoácidos e ureídeos com dados de validação funcional. (Próprio autor)	33

LISTA DE ABREVIATURAS E SIGLAS

AAP6	Amino ácido permeasse 6
Aas	Amino Ácidos
Abrasem	Associação Brasileira de Sementes e Mudanças
ANEC	Associação Nacional dos Exportadores de Cereais
Asn	Asparagina
C	Carbono
CONAB	Companhia Nacional de Abastecimento
Cys	Cisteína
Da	Dalton
DAF	Dias após o florescimento
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
EUA	Estados Unidos da América
FBN	Fixação Biológica do Nitrogênio
Gln	Glutamina
GOGAT	Glutamato Sintase
GS	Glutamina Sintetase
K	Potássio
LHT1	Lysine/Histidine transporter (Transportador de Lisina/Histidina)
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
Met	Metionina
N	Nitrogênio
NH ₄ ⁺	Amônio
Nod	Fatores de Nodulação
P	Fósforo
PCR	Polymerase Chain Reaction (Reação em cadeia da polimerase)
PIB	Produto Interno Bruto
PvUPS1	Putative allantoin transporter in Phaseolus vulgaris (Transportador putativo de alantoína em Phaseolus vulgaris)
QTLs	Quantitative trait Locus (locus de característica quantitativa)
R	Reprodutivo
RNA-Seq	Sequenciamento de RNA
RS	Rio Grande do Sul

SP	São Paulo
Trp	Triptofano
V	Vegetativo
VC	Cotilédone
VE	Emergência

SUMÁRIO

1	INTRODUÇÃO	12
2	REVISÃO BIBLIOGRÁFICA	15
2.1	A CULTURA DA SOJA.....	15
2.2	IMPORTÂNCIA ECONÔMICA.....	16
2.3	COMPOSIÇÃO DO GRÃO E DIFERENTES USOS	17
2.4	ACÚMULO DE PROTEÍNA NO GRÃO DE SOJA	19
2.5	FATORES QUE INFLUENCIAM NO TEOR DE PROTEÍNA	22
2.5.1	Fatores Ambientais	22
2.5.2	Fatores Nutricionais	24
2.5.3	Fatores Genéticos	27
2.6	FERRAMENTAS DA BIOTECNOLOGIA PARA AUMENTO DO TEOR DE PROTEÍNA DA SOJA	30
3	OBJETIVOS	34
3.1	OBJETIVO GERAL	34
3.2	OBJETIVOS ESPECÍFICOS	34
	REFERÊNCIAS BIBLIOGRÁFICAS	35
4	ARTIGO 1: PROSPECÇÃO IN SILICO E AVALIAÇÃO DA EXPRESSÃO DE GENES ASSOCIADOS AO TEOR DE PROTEÍNA EM GRÃOS DE SOJA	43

1. INTRODUÇÃO

A soja é uma cultura de fundamental importância sócio-econômica em nível global (Conab, 2020). No Brasil, é a principal *commodity* produzida e exportada, representando 48% da produção total de grãos (Conab, 2019). O País produz em torno de 124,8 milhões de toneladas em uma área aproximada de 37 milhões de hectares, ficando em primeiro lugar no ranking de produção mundial do grão, tendo como principais estados produtores, Mato Grosso, Paraná e Goiás (Conab, 2020).

Por apresentar proteína qualitativamente equivalente à proteína animal, é amplamente utilizada em forma de farelo para indústria de ração animal (Pípolo; Mandarino, 2016; Gazzoni, 2018). A produção de ração animal é estimada em pouco mais de um bilhão de toneladas por ano, movimentando ativamente a economia mundial (IFIF, 2019), porém para ser utilizada para este fim, o farelo deve conter um teor de proteína mínimo de 46%, segundo a ANEC-71 (Pípolo *et al.*, 2015), teor requerido para a alimentação animal (Willis 2003; Patil *et al.*, 2017).

O teor de proteína mínimo exigido pela indústria moageira de soja é de pelo menos 35% nos grãos para atingir o mínimo exigido no farelo (Willis 2003; Patil *et al.*, 2017). O conteúdo de proteína nos grãos de soja geralmente atendia aos padrões industriais, mas ao longo dos anos, em algumas situações devido as alterações genéticas promovida pelo melhoramento genético da cultura, a indústria moageira tem encontrado dificuldade em obter grãos que atendam ao padrão proteico estabelecido (Pathan, 2013). Na tentativa de alcançar o teor de proteína adequado, utiliza-se a estratégia de remover o tegumento da semente. Esta etapa a mais no processamento da soja aumenta consideravelmente os custos do processamento (Pípolo *et al.*, 2015).

Linhas de soja com até 41,28% de proteína são encontradas em cultivares brasileiras (Mandarino *et al.*, 2019), mas essas linhas com alto teor de proteína no grão são caracterizados por um número menor de sementes por planta e consequentemente menor produtividade (Medic; Atkinson; Hurburgh, 2014).

Um dos maiores problemas para o melhoramento genético dessa característica é a correlação negativa entre rendimento de grãos e teor de proteína (Pathan *et al.*, 2013). Na maioria das situações, o melhoramento

genético na cultura da soja busca o desenvolvimento de cultivares mais produtivas, resultando em plantas com menores teores proteicos (Pathan *et al.*, 2013). Estudos realizados em um período de 10 anos, entre 1994 e 2004, mostraram que enquanto havia o aumento da produtividade em 10,9 kg/hectare/ano, o teor proteico decresceu 0,1% neste período (Brumm; Hurburgh 2006; Pathan, 2013). Segundo Pípolo *et al.* (2015), devido ao desempenho no aumento da produtividade da cultura a tendência é aumentar ainda mais a queda dos teores de proteína ao longo dos anos.

Na tentativa de evitar essa redução no teor proteico entender os mecanismos envolvidos na produção, transporte e armazenamento de proteínas no grão pode auxiliar como ferramenta para geração de cultivares elite que apresentem alto teor proteico no grão e que sejam ao mesmo tempo produtivas. Uma possibilidade é prospectar genes envolvidos na síntese de proteína no grão, utilizando ferramentas como RNA-Seq, que vem sendo amplamente utilizadas para estudar a nível transcricional, genes fundamentais a planta para melhorar tais fenótipos (Mardis, 2008; Michalak *et al.*, 2016). Entre os genes que podem estar envolvidos no aumento do teor de proteína no grão encontram-se os transportadores de compostos nitrogenados (TCNs) como aminoácidos e ureídeos (Ohyama, 1984; Ohyama, 2013), uma vez que na soja há uma alta demanda de nitrogênio para obtenção de adequada produtividade e grãos com valor considerável de proteínas (produtividade de 3.500 kg ha⁻¹ com teor de proteína no grão de 40%, necessita de aproximadamente 224 kg ha⁻¹ de N) (Hungria *et al.*, 2012).

O aumento da disponibilidade de N para parte aérea, aumenta o teor de proteína das sementes (Fabre; Planchon, 2012). Trabalhos prévios em planta modelo mostram que a superexpressão de alguns TCNs gera o aumento do transporte do N, aumentando também a produtividade (Carter; Tegeder, 2016; Besnard *et al.* 2018). No trabalho de Carter e Tegeder (2016), a superexpressão do gene UPS1 Ureide permease 1(UPS1) em *Arabidopsis thaliana* gerou um aumento em 36% no rendimento geral das sementes. Besnard *et al.*, (2018), também em *Arabidopsis*, perceberam aumento no número de sementes após superexpressarem o TCN UMAMIT25.

O avanço no conhecimento acerca dos mecanismos moleculares, através dos quais esses transportadores atuam, especialmente durante a formação e enchimento do grão, será útil no desenho de estratégias biotecnológicas que visem a obtenção de cultivares de soja com maior capacidade de síntese de proteínas de reserva resultando em grãos de alto teor de proteína e plantas de alta produtividade.

Diante disso, o objetivo deste trabalho é identificar por meio de estudos de transcriptoma, genes envolvidos no processo de transporte de aminoácidos e ureídeos nas sementes de soja e tecidos adjacentes, a fim de selecionar alvos promissores para o melhoramento genético da cultura através de ferramentas de edição.

2. REVISÃO BIBLIOGRÁFICA

2.1. A CULTURA DA SOJA

A soja (*Glycine max* (L.) Merr.) é uma leguminosa dicotiledônea da família Fabaceae, originária da Ásia (Bellieny-rabelo, 2016), sendo atualmente cultivada em diversas partes do mundo (Nyoki; Ndakidemi; Baltazari *et al.*, 2019). É uma espécie autógama de ciclo anual que apresenta ampla adaptação a climas tropicais e subtropicais (Sediyama, 2009; Borém; Miranda; Fritsche-Neto, 2017). Os principais estádios de desenvolvimento da cultura podem ser divididos em vegetativos (V) e reprodutivos (R) e podem ser observados na **Figura 1**. O primeiro é denominado VE (emergência), o segundo é denominado VC (cotiledonar), este representa o estágio em que os cotilédones se encontram completamente expandidos (Fehr; Caviness, 1977; Wright; Lenssen, 2013). Após o VC, o desenvolvimento vegetativo pode ser dividido de acordo com o número de trifolios expandidos de V1 a VN (Fehr; Caviness, 1977; Wright; Lenssen, 2013). Os estágios reprodutivos são divididos em: R1 (Início da floração); R2 (plena floração); R3 (desenvolvimento das vagens); R4 (vagem completa); R5 (Desenvolvimento de sementes); R6 (Sementes cheias); R7 (início da maturação); R8 (maturação completa) (Wright; Lenssen, 2013; Nyoki; Ndakidemi; Baltazari *et al.*, 2019).

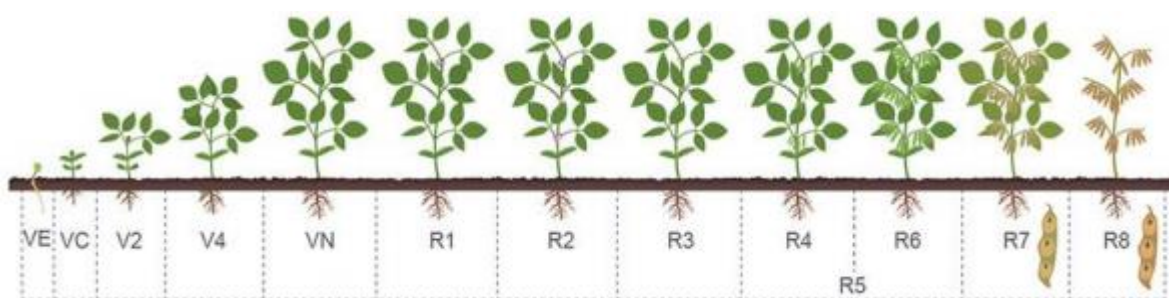


Figura 1- Principais estádios fenológicos da cultura da soja (Perdigão *et al.*, 2019).

Por se tratar de uma leguminosa, uma das características mais importantes da cultura é a capacidade de sua raiz interagir com bactérias simbióticas (rizobia), otimizando a captação de nitrogênio, nutriente muito utilizado pela planta para seu crescimento, desenvolvimento e enchimento do

grão (Oldroyd; Downie, 2008).

2.2. IMPORTÂNCIA ECONÔMICA

A produção de soja está entre as atividades econômicas que apresentam maior crescimento nas últimas décadas, devido a fatores como desenvolvimento tecnológico e comercializações internacionais (Ainsworth *et al.*, 2012). É uma das principais commodities mundiais, com produção em mais de 60 países, de suma importância para a alimentação animal e humana (Conab, 2019). Nos últimos trinta anos o grão passou a ser um dos mais produzidos e consumidos (Conab, 2019). Os principais produtores são Brasil, Estados Unidos e Argentina, que juntos são responsáveis por 80% de sua produção (Ainsworth *et al.*, 2012). A produção mundial gira em torno de 362 milhões de toneladas, destes aproximadamente 124,8 milhões são referentes a produção brasileira, o que deixa o país na primeira posição no ranking de produção mundial do grão (Conab, 2020).

O mercado da soja brasileira atende duas demandas, a exportação do grão inteiro e do grão esmagado, de onde obtém-se dois subprodutos, o óleo e o farelo de soja (Lemos *et al.*, 2017). Em 2019, do total de exportações de soja, 80,76% foram em forma de grãos inteiros, 18,10% de farelo e 1,14% de óleo (Abiove, 2020). O Brasil é o quarto maior produtor mundial de farelo e o principal consumidor, pois internamente consome cerca de 52.8% do que produz (Lemos *et al.*, 2017). Por se tratar da principal fonte de proteína vegetal atende demandas crescentes na produção de rações animais (Hirakuri; Lazzarotto, 2014; Bellieny-Rabelo, 2016;). Como um todo, o mercado da soja gerou ao País ganhos de 34,78 bilhões de dólares com exportação, no ano de 2019 (Conab, 2019). Devido sua forte participação na economia brasileira com foco na indústria alimentícia, o desenvolvimento de cultivares com características melhoradas, como elevados teores proteicos, podem agregar maior valor comercial aos produtos a base de soja e impulsionar ainda mais o mercado internacional (Fukui; Lopes, 2013).

2.3. COMPOSIÇÃO DO GRÃO E DIFERENTES USOS

A soja é considerada a principal fonte de proteína vegetal pois apresenta como componente proteínico ricos em diversos aminoácidos essenciais, fazendo com que ela seja uma das importantes matérias primas para produção de farelo para ração animal destinada a frangos, suínos e bovinos, respondendo por mais de 80% da proteína bruta deste produto (Pípolo; Mandarino, 2016). Além disso, a soja também é rica em óleo e componentes antioxidantes, isoflavonas, fosfolípidios, aminoácidos essenciais, vitaminas e minerais (Hirakuri; Lazzarotto, 2014; Pípolo; Mandarino, 2016).

Embora a variação na composição química dos grãos da soja seja influenciada pela genética das cultivares, esta também sofre forte influência ambiental (Sales *et al.*, 2016). Na **Figura 2** é possível visualizar a composição média do grão de soja, no qual aproximadamente 40,3% é de proteína, 33,9% de carboidratos, 21% de óleo e 4,9% de cinzas (representado por resíduos inorgânicos) (Pípolo; Mandarino, 2016). Ainda, percebe-se uma correlação positiva entre óleo e produtividade, que são duas características que possuem correlação negativa com o teor proteico (Pípolo; Mandarino, 2016). Em um levantamento mais recente realizado pela Embrapa, considerando 863 amostras de grãos coletadas nas diversas regiões produtoras de soja do País, observou-se que o teor médio de proteína foi de 36,88%, com valor mínimo observado de 28,65% e máximo de 41,28% (Mandarino *et al.*, 2019).

O mercado mundial absorve 86% da produção brasileira para processamento industrial, 7% para alimentação humana, e 7% para produção de sementes (De Oliveira Junior *et al.*, 2018). Na soja utilizada como semente o teor proteico é importante para o processo germinativo e desempenho no campo, pois influencia no metabolismo, na quantidade e qualidade de reservas, sendo utilizados no desenvolvimento do embrião e assim no vigor da plântula (Barbieri *et al.*, 2013).

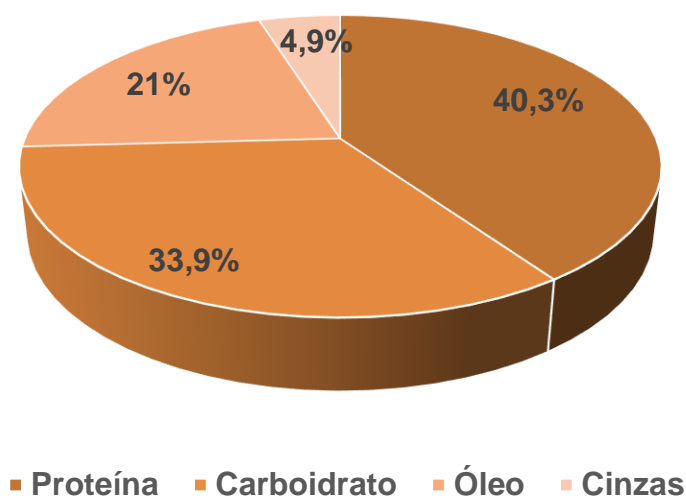


Figura 2 - Composição média dos grãos de soja (Pípolo, Mandarino, 2016, adaptado pelo autor).

Em relação a soja destinada a grãos, 80% é encaminhado as indústrias de alimentação animal, os outros 20% são utilizados na indústria oléica, (Gonçalves; Yada, 2019), de cosméticos, farmacêutica, tintas, biodiesel, emulsificante, alimentação humana (Fontes, 2020). O parâmetro utilizado para classificação de soja grão é sua composição química, principalmente quanto ao teor de proteína e óleo (Gonçalves; Yada, 2019).

Quanto mais alto for o teor de proteínas nos grãos, melhor será para a produção de farelos com teores de proteína mínimos exigidos pela legislação e menor o custo de produção (Pípolo *et al.*, 2015). O teor de proteína mínimo para farelos destinados a alimentação animal é 46%, caso este não seja alcançado a indústria terá que realizar o descasque dos grãos, este é o principal motivo do aumento do custo no processamento do farelo (Willis 2003; Patil *et al.*, 2017; Mandarino *et al.*, 2019).

Segundo dados da Abrasem (2016), o foco principal dos produtores de soja é voltado para produção de grãos. Na **Tabela 1** é possível observar os valores de produção de sementes e grãos no Brasil, em mil toneladas, entre as safras de 2008/09 à 2017/18. De acordo com dados de levantamento realizado pela Embrapa Soja, a partir de amostras representativas de todas as regiões do Brasil, em cultivares brasileiras o teor médio de proteína dos grão de soja gira

em torno de 36,86% e para amostras designadas a sementes 37,6%. Essa diferença observada entre o teor de proteínas de grãos e de sementes de soja pode ser atribuída ao manejo, tendo em vista que lavouras destinadas a produção de sementes preconizam a qualidade, e com isso adotam um manejo mais criterioso. Nos grãos, embora o teor médio observado atenda aos parâmetros mínimos requeridos, é importante destacar que muitas das amostras analisadas apresentam valores bem inferiores, sendo que o mínimo observado foi de 31,59% (Mandarino *et al.*, 2019).

Tabela 1- Produção de sementes e soja em grão no Brasil na safra de 2008/09 a safra de 2017/18 em mil toneladas

ANO	SOJA SEMENTE	SOJA GRÃO
2008/09	1.290	57.165
2009/10	1.560	68.688
2010/11	1.592	75.324
2011/12	2.293	66.383
2012/13	2.149	81.499
2013/14	1.776	86.120
2014/15	1.984	96.228
2015/16	2.561	95.434
2016/17	2.385	114.075
2017/18	3.069	119.281

Fonte: Abrasem, 2019; Conab, 2020, adaptado pelo autor.

2.4. ACÚMULO DE PROTEÍNA NO GRÃO DE SOJA

Os grãos de soja são compostos por três compartimentos: 1. Tegumento: envoltório que protege a semente, este corresponde a 8% do total do grão; 2. Cotilédones: servem como reservas alimentares para a planta em crescimento durante a germinação, e representam 90% da estrutura do grão; 3. Eixo embrionário: composto pelo hipocótilo e epicótilo, responsáveis pelo crescimento e emergência acima da superfície do solo, e pela radícula, que se tornará posteriormente a raiz, representando 2% do total do grão (Guleria *et al.*, 2019).

A concentração de proteínas no cotilédone gira em torno de 43%, enquanto no eixo embrionário e tegumento é de aproximadamente 41% e 9% respectivamente, estes teores proteicos variam de acordo com o genótipo e o

ambiente de cultivo (Medic; Atkinson; Hurburgh, 2014). A **Figura 3** ilustra o tegumento, o cotilédone e o eixo embrionário (epicótilo, hipocótilo e radícula) de um grão de soja.

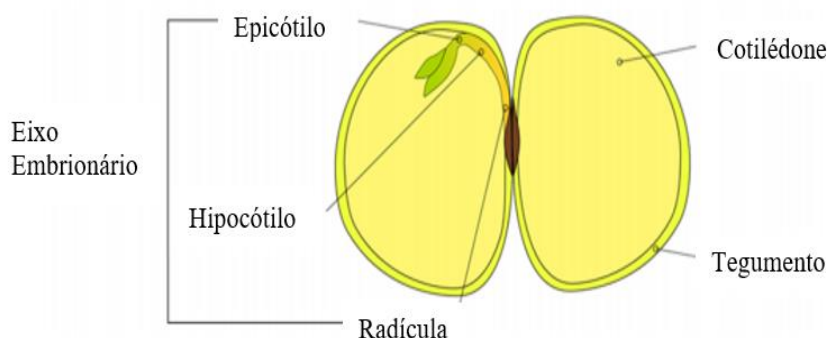


Figura 3 - Representação ilustrativa de compartimentos de um grão de soja (Medic; Atkinson; Hurburgh, 2014, adaptado pelo autor).

A soja apresenta variabilidade genética para o teor de proteína no grão, sendo possível encontrar acessos que apresentam teores muito altos com até 57,9% de proteína (base de peso seco), porém, que geram um baixo número de sementes o que limita a sua utilização (Medic; Atkinson; Hurburgh, 2014). À nível mundial, a média proteica do grão de soja é de 40% em peso seco (Mello Filho *et al.*, 2004), porém, nas cultivares brasileiras utilizadas atualmente, os teores giram em torno de 37%, não sendo raras situações em que teores ficam abaixo do mínimo de 35% preconizado pela indústria moageira (Pipolo *et al.*, 2015; Mandarino *et al.*, 2019).

Proteínas solúveis em água, como exemplo as albuminas, podem ser processadas e utilizadas em alimentos de soja tradicionais (Yu *et al.*, 2016), além de possuírem excelente capacidade de processamento na fabricação de alimentos, como emulsificante, gelificante e capacidade de retenção de água (Chen *et al.*, 2014). Porém, a maior parte das proteínas de soja, denominadas globulinas, são insolúveis em água, mas solúvel em soluções salinas (Damodaran; Parkin, 2018).

Cerca de 90% das proteínas da soja são representadas por globulinas do tipo glicinina e β -conglícinina (11S e 7S) (Murphy, 2008; Medic; Atkinson; Hurburgh, 2014). Estes tipos de proteína são consumidas pela planta rapidamente durante a germinação, sendo reconstituídas posteriormente

durante a captação de nutrientes (Takahashi *et al.*, 2003). As proteínas podem desempenhar papéis fisiológicos como defesa, transporte de água e solutos, divisão celular, defesa durante o estresse hídrico; atuação no metabolismo e síntese de outras proteínas (Zhang *et al.*, 2013; Xu *et al.*, 2015).

Para indústria de alimentação animal, as proteínas tem a finalidade de fornecer quantidades significativas de aminoácidos (AAs) essenciais e não essenciais (AAs) (Patil *et al.*, 2017). AAs essenciais (20% do grão) são aqueles que não podem ser sintetizados em animais e desempenham papel crucial nos processos metabólicos (Panthee *et al.*, 2006). A **Tabela 2**, mostra a composição média de proteínas de soja quanto a aminoácidos essenciais e não essenciais e suas respectivas porcentagens (Patil *et al.*, 2017).

Tabela 2- Porcentagem de aminoácidos essenciais e não essenciais na composição das proteínas de soja.

Aminoácidos essenciais	% em proteína de soja	Aminoácidos não essenciais	% em proteína de soja
Metionina	1	Cistina	3
Triptofano	1	Tirosina	4
Histidina	3	Alanina	4
Treonina	4	Glicina	4
Valina	4	Serina	5
Isoleucina	5	Prolina	5
Fenilalanina	6	Ácido Aspártico	7
Lisina	7	Arginina	8
Leucina	8	Ácido glutâmico	19

Fonte: Patil *et al.*, 2017, adaptado pelo autor.

Pela importância dos AAs no conteúdo final de proteínas, alguns trabalhos relacionados ao seu acúmulo durante o desenvolvimento de grãos vêm sendo realizados. Li *et al.* (2015), combinaram tecnologias de análises metabolômicas, transcriptômicas para revelar as redes metabólicas que determinam a estrutura e composição do grão durante seu desenvolvimento. No período de 25 dias após o florescimento (DAF), os AA mais abundantes são, a asparagina, histidina, glutamina, glutamato, lisina, aspartato e serina. Estes AA apresentam padrão decrescente até o período de 50 DAF. Embora os níveis desses AA tenham diminuído, o glutamato, lisina e aspartato foram os aminoácidos livres mais

abundantes no período de 50 DAF. O acúmulo de alguns aminoácidos (incluindo asparagina, glutamato, serina, isoleucina, tirosina, metionina, prolina, valina e particularmente alanina) atingiu o pico aos 30 DAF, enquanto a lisina, aspartato, fenilalanina, leucina, isoleucina e metionina tiveram um pico pequeno aos 45 DAF.

Os AAs asparagina e glutamina são de extrema importância, pois possuem funções de transporte de nitrogênio na biossíntese de outros aminoácidos, que irão compor as proteínas posteriormente e devido a essa função, o aumento desses aminoácidos tem correlação positiva com o teor de proteína (Pandurangan *et al.*, 2012).

Durante os primeiros dias do desenvolvimento do grão são sintetizadas proteínas que vão atuar em funções metabólicas que decrescem a partir de 22 DAF, após este período inicia-se a síntese de proteínas de armazenamento como glicinina e β -conglucininina (Saldivar *et al.*, 2011). Menos de 5% da proteína total da semente madura são acumuladas em 12 DAF (Saldivar *et al.*, 2011). Em aproximadamente 25 dias após a floração, a composição de proteína nas sementes é apenas 30% da proteína total na semente madura, e então em cerca de 40 a 64 DAF os 70% do total de proteínas são sintetizadas (Saldivar *et al.*, 2011). Para que se possa aumentar a qualidade e capacidade de armazenamento proteico no grão é crucial o entendimento dos fatores que influenciam neste processo, estes podem ser de origem ambiental, nutricional e genético (Schmidt *et al.*, 2011; Assefa *et al.*, 2019).

2.5. FATORES QUE INFLUENCIAM NO TEOR DE PROTEÍNA

2.5.1 Fatores ambientais

A composição proteica da semente, conta com a interação genótipo \times ambiente (Patil *et al.*, 2017; Assefa *et al.*, 2019). Por esse motivo, diferentes cultivares e lotes de uma mesma cultivar podem variar quanto à composição química devido ao cultivo em diferentes condições ambientais, com potenciais reflexos sobre a qualidade fisiológica das sementes e grãos (Delarmelino-Ferraresi; Villela; Aumonde, 2014). Neste contexto, a composição química da

proteína pode ser afetada por fatores bióticos e abióticos, como exemplo, a data de semeadura; temperaturas; condições hídricas e estação de cultivo e armazenagem (Song *et al.*, 2016). Porém, a contribuição de cada fator para a composição e o rendimento das sementes não é bem elucidada (Assefa *et al.*, 2019).

A associação com algumas práticas, como rotação de culturas e impactos mistos de alguns fatores, como plantio direto, tratamento de sementes, aplicação de nutrientes foliares e fungicidas, também podem alterar a composição química das sementes (Assefa *et al.*, 2019). Assim, as mesmas variedades de soja podem variar na composição química do grão ao longo dos anos ou em diferentes ambientes no mesmo ano (Song *et al.*, 2016).

Um exemplo sobre a mudança na composição da semente em relação ao ambiente, foi relatada por Bellaloui *et al.* (2015), que realizaram um trabalho avaliando metabólitos neste tecido, incluindo teor de proteínas em época de semeadura. O plantio precoce resultou em menores concentrações de proteína enquanto o plantio tardio resultou em maiores, deixando claro que semeaduras em diferentes datas podem conferir diferenças na concentração de metabólitos nas plantas (Bellaloui *et al.* 2015).

Outro trabalho enfocando o efeito do ambiente na composição química do grão foi realizado por Mertz-Henning *et al.* (2017). Estes pesquisadores realizaram um estudo com indução de déficit hídrico em soja, durante os períodos vegetativos e reprodutivos, e analisaram seu efeito sobre o rendimento e no teor proteico do grão. Foi observado que, quando induzido em estágio reprodutivo, o teor de proteínas foi aumentado com o estresse hídrico na maioria dos genótipos utilizados, porém acarretando a diminuição significativa no óleo e na produtividade. Esse aumento possivelmente está correlacionado com a síntese de novas proteínas, além das de armazenamento, que estão relacionadas à proteção contra a seca, com a posterior redução no teor de proteínas devido à sua hidrólise e degradação (Mertz-Henning *et al.*, 2017). Estes trabalhos mostram que a composição dos grãos e sementes de soja podem ser fortemente afetados por diversas condições a qual a planta mãe é submetida, especialmente durante o período de enchimento das sementes,

quando ocorre o acúmulo de componentes químicos (Carrera *et al.*, 2009, Carrera *et al.*, 2011).

2.5.2 Fatores Nutricionais

O teor de nutrientes na planta materna também é importante para o teor proteico nos grãos, pois os nutrientes absorvidos podem interferir na assimilação de outros, além de ativar ou diminuir a expressão de genes que atuam na captação dos compostos assimilados pelos grãos (Gu *et al.*, 2010). Deficiência de macronutrientes como o fósforo e potássio, podem resultar em menor produção de flores e de vagens, além de diminuir o fluxo de metabólitos de formação de diversos tipos de proteínas. Micronutrientes, principalmente cobalto, boro e molibdênio, são importantes pela sua influência na fixação simbiótica de nitrogênio na soja (Ceretta *et al.*, 2005).

O macronutriente que mais influência no teor proteico em leguminosas é o nitrogênio (N), pois é constituinte de ácidos nucléicos, aminoácidos, nucleotídeos, coenzimas, vitaminas, alguns hormônios e clorofila (Epstein, 1999). Altas demandas de N são necessárias para obtenção de altos rendimentos de sementes e altos teores de proteínas nos grãos, uma lavoura com produtividade de 3.500 kg ha⁻¹ com teor de proteína no grão de 40% necessita para formação de grãos aproximadamente 224 kg ha⁻¹ de N (Hungria *et al.*, 2012).

Embora, o N seja encontrado abundantemente na atmosfera, este apresenta-se em sua forma estável (N₂), que é quimicamente inerte e inviável para absorção e utilização pela planta. Para se tornar assimilável é necessária a redução do N₂ atmosférico em N-amoniaco, processo viabilizado pela Fixação Biológica do Nitrogênio (FBN) (Alcantara; Reis, 2008). Devido a alta capacidade de associação simbiótica da soja com bactérias dos gêneros *Rhizobium*, *Bradyrhizobium*, *Azorhizobium*, *Photorhizobium*, *Sinorhizobium*, eleva a capacidade de suprir sua necessidade nutricional deste macronutriente (Taiz; Zieger, 2009).

Na soja, após a fixação de N₂, a amônia é finalmente reduzida em aminoácidos e em compostos nitrogenados denominados ureídeos (alantoína e

ácido alantóico) (Carter; Tegeder, 2016), estes representam as principais formas de transporte de N a longa distância (Carter; Tegeder, 2016). Deve-se considerar como parte do processo da formação de ureídeos, a primeira etapa do processo de FBN, momento em que a raiz excreta flavonóides para reconhecimento das bactérias endófitas, estas que irão colonizar as células vegetais formando os nódulos (Baral; Teixeira Da Silva; Izaguirre-Mayoral, 2016).

Os nódulos radiculares internos possuem morfologia caracterizada por possuir anatomia composta por um núcleo central de células rizobialmente infectadas e não infectadas (Downie, 2014). Após a infecção de algumas células vegetais pelo bacterióide, a planta hospedeira fornece carbono e nitrogênio, que são translocados pelo floema a partir da parte aérea, como fonte de energia para o crescimento do nódulo, e durante o crescimento e manutenção do tecido nodular, ocorre a redução de N_2 em NH_4^+ (Schubert, 1986), no interior dos endófitos que estão nas células nodulares infectadas (Ohyama *et al.*, 2013, **Figura 4**).

A maior parte do N fixado é imediatamente exportada do bacterióide para o citosol da planta e assimilada pela via glutamina sintetase (GS) / glutamato sintase (GOGAT) em vários aminoácidos por transaminação do glutamato. Este é transformado em glutamina e se move para a mitocôndria e para os plasmídeos, onde sintetiza a purina, passando pela via da biossíntese e degradação de purinas em xantina (Collier; Tegeder, 2012, **Figura 4**).

A xantina é liberada no citosol, e difunde-se de células de nódulos infectadas para não infectadas (Tegeder, 2014), então é oxidada em ácido úrico (Collier; Tegeder, 2012), que é convertido nos peroxissomos em alantoína. O ácido alantóico é produzido no retículo endoplasmático a partir da alantoína (Collier; Tegeder, 2012).

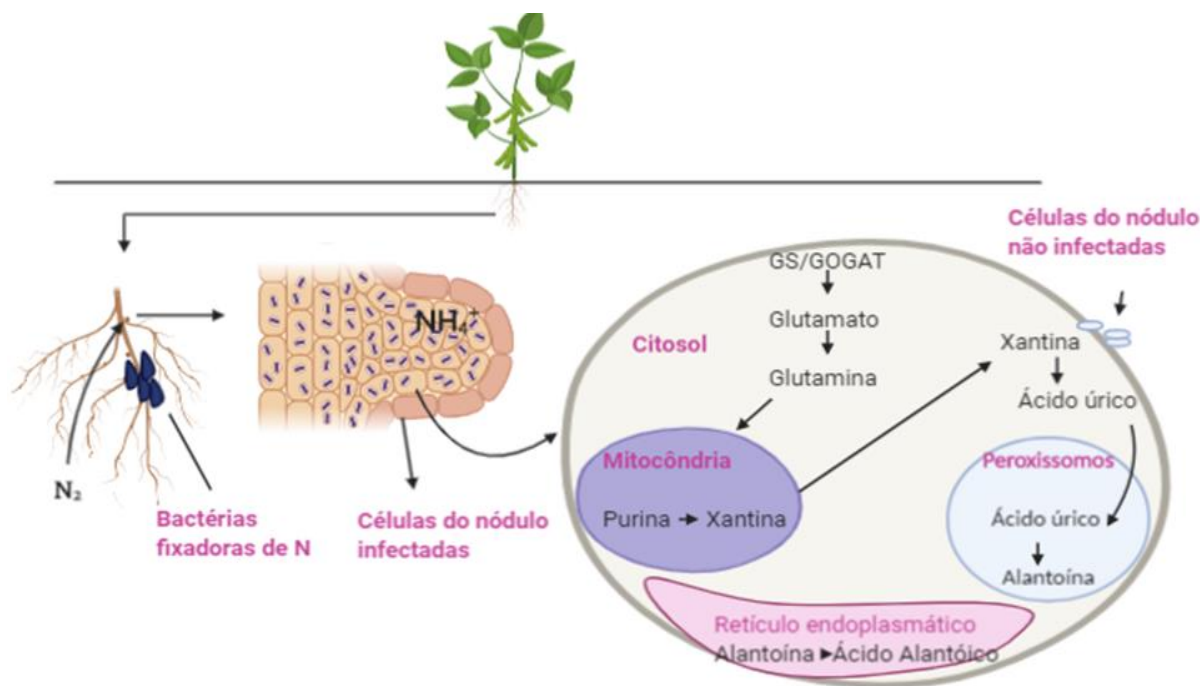


Figura 4 – Síntese de ureídeos a partir da fixação biológica do nitrogênio nas raízes de soja (Próprio autor).

Para alcançar o principal sistema vascular de uma leguminosa, os ureídeos devem se mover intercelularmente por todo o interior do periciclo do córtex, endoderme e parênquima das raízes noduladas (Tegeder, 2014). Quando os ureídeos começam a se acumular nas células não infectadas do nódulo, eles são carregados para o sistema vascular do xilema e são transportados aos caules e pecíolos até as folhas e posteriormente pelo floema para as vagens e sementes (Ohyama *et al.*, 2013).

Os ureídeos (alantoína e ácido alantóico) representam aproximadamente 90% do total de N transportados no xilema de leguminosas tropicais fixadoras de N_2 (Raso *et al.*, 2007). O ácido alantóico quando alcançam órgãos superiores como as folhas e sementes, sofre catálise e libera amônio, que entra na via de assimilação do nitrogênio, por meio do sistema GS/GOGAT, liberando aminoácidos e gerando um ciclo. Eles podem ser armazenados em grandes quantidades em diferentes órgãos vegetais (Figura 4) (Baral; Teixeira Da Silva; Izaguirre-Mayoral, 2016).

Durante o enchimento das sementes, Thomas e Schrades (1981) observaram que o ácido alantóico se acumula nas vagens durante o início do enchimento de sementes. As concentrações totais de ureia e ácido alantóico foi

maior em vagens do que em sementes em todas as fases de seu desenvolvimento, porém com o início da senescência houve diminuição nas concentrações de ácido alantóico em vagens. Em complemento, Ohyama (1984), observou que os ureídeos derivados do N₂ fixado pelos nódulos radiculares, são transportados diretamente para as vagens e utilizados no desenvolvimento de sementes, além de serem utilizados para a síntese de proteínas foliares.

Além dos ureídeos, cerca de 10% dos aminoácidos também são responsáveis pelo transporte de N em soja, e auxiliando no aumento do teor de proteínas em leguminosas. Quando o composto para formação de proteína entra em forma direta de aminoácido, a asparagina é dominante no cotilédone em desenvolvimento, representando até 50% do total de aminoácidos livres (Hernández-Sebastià *et al.*, 2005).

2.5.3 Fatores Genéticos

Mesmo sofrendo influência ambiental e nutricional, o teor de proteína da semente de soja é uma característica herdada quantitativamente e controlada por muitos genes e reguladores (Patil *et al.*, 2017). Os genes das proteínas das sementes são regulados temporalmente durante a embriogênese, podendo ser ativados ou reprimidos na planta madura. Além disso, são frequentemente expressos em diferentes níveis no cotilédone e eixo embrionário (Perez-Grau; Goldberg, 1989).

No trabalho de Jones, Gonzalez e Vodkin (2010) foi realizada a análise do transcriptoma de sementes de soja durante seu desenvolvimento, cobrindo os diferentes estágios de formação da semente. A partir destes dados, esses autores observaram que os genes envolvidos com as principais proteínas de armazenamento de sementes de soja, tem seus níveis de expressão aumentado significativamente entre a fase de 100-200mg e a fase de 400-500 mg em que a semente começa amarelar. A glicinina, teve o pico de expressão na fase de 100-200mg e a conglucina teve seu nível de expressão mais alto na fase em que a semente atinge seu peso fresco máximo (400-500 mg).

Jones e Vodkin (2013), realizaram uma análise transcricional via RNA-

Seq, de sementes de soja de dias após a fertilização até a maturidade sementes, observaram que os níveis de proteínas de armazenamento tiveram maiores expressões em cotilédones com 100–200mg e 400–500mg, comparados as outras fases de desenvolvimento da semente. Estas fases podem permitir um melhor entendimento da questão genética envolvida no enchimento de grãos, enfocando na compreensão de genes relacionados a transporte de aminoácidos e ureídeos (Schmidt *et al.*, 2011).

Na **Tabela 3**, estão descritos alguns trabalhos de prospecção de genes relacionados com o transporte de compostos nitrogenados, na espécie modelo *Arabidopsis thaliana* e em *Glycine max*. Em *arabidopsis*, Müller *et al.* (2015) estudaram os genes da família UmamiT (Usually Multiple Acids Move In and out Transporter). Ao verificar a localização subcelular de UmamiT 14, UmamiT 11, UmamiT 29 e UmamiT 28, descobriram que todas elas atuam principalmente na membrana plasmática. Essas famílias transportam aminoácidos em toda a planta. As famílias UmamiT 28 e UmamiT 29, foram sugeridas como de maior atuação em sementes. Mutantes com a perda da função desses genes resultaram em sementes menores e com maior acúmulo de aminoácidos nos frutos (Tabela 3).

Membros adicionais dessa mesma família foram estudados por Besnard *et al.* (2018). Esses autores mostraram que os membros UmamiT 24 e UmamiT 25 também atuam na transferência de aminoácidos para as sementes em desenvolvimento, sendo o UmamiT 24 expresso principalmente na calaza e localizado no tonoplasto, enquanto o UmamiT 25 é localizado na membrana plasmática e expresso nas células do endosperma. Esses autores sugerem que o UmamiT 24 promove um armazenamento temporário dos aminoácidos na calaza, já o UmamiT 25 mediará a exportação de aminoácidos das células do endosperma, disponibilizando-os para importação pelo embrião em desenvolvimento.

Em soja, Cheng *et al.* (2016), identificaram 189 genes da família AAT (Transportadores de aminoácidos) e os classificaram em 12 subfamílias distintas com base em sua composição de sequências e posições filogenéticas: **1.** transportadores de aminoácidos catiônicos (CAT), **2.** exportadores H + de poliamina (PHS), **3.** aminoácido /colina (ACT), **4.** transportadores de aminoácido

(TTP), **5.** transportadores de aminoácidos (ATLa), **6.** transportadores de aminoácidos (ATLb), **7.** transportadores de aminoácidos neutros e aromáticos (ANT), **8.** transportadores de auxina (AUX), **9.** aminoácido permease (AAP), **10.** transportadores de ácido c- aminobutírico (GAT), **11.** transportador de prolina (proT), **12.** transportadores de lisina/histidina (LHT). Estes autores também analisaram as distribuições cromossômicas, estruturas gênicas, padrões de duplicação, filogenia, padrões de expressão de tecidos desses 189 AAT na soja. Descobriram que um grande número de genes AAT na soja foi expandido via duplicação. Além disso, análises dos perfis de expressão desses genes em soja, em vários estágios de desenvolvimento vegetativo e reprodutivo, mostraram padrões de expressão distintos entre os diferentes tecidos.

Ainda em soja, Patil *et al.* (2017), realizaram um extenso trabalho de revisão, incluindo informações sobre diversos estudos de mapeamento de QTLs (locus de características quantitativas) e de estudos de mapeamento associativo, os quais identificaram regiões genômicas associadas ao teor de proteína em soja. Por meio da análise dessas regiões identificadas foram encontrados diversos genes associados ao transporte de aminoácidos e ureídeos (Tabela 3).

Tabela 3 - Transportadores de aminoácidos e ureídeos em Arabidopsis e soja.

Espécie de origem	Gene ID	Wm82.a2.v1	Anotação	Referência
Arabidopsis	At2G39510	Glyma.19G173800	EamA-like transporter	Müller, 2015
Soja	Glyma02g12980	Glyma.02G116400	Ureide permease	Carter; tegeder, 2012; 2016
Soja	Glyma02g19430	Glyma.02G160900	Amino acid transporter	Cheng <i>et al.</i> , 2016
Soja	Glyma05g02790	Glyma.05G043100	Amino acid transporter	Cheng <i>et al.</i> , 2016
Soja	Glyma05g11410	Glyma.05G089200	solute carrier family 7	Cheng <i>et al.</i> , 2016
Soja	Glyma08g11960	Glyma.08G113400	Amino acid transporter	Cheng <i>et al.</i> , 2016
Soja	Glyma08g44940	Glyma.08G336600	Amino acid transporter	Cheng <i>et al.</i> , 2016
Soja	Glyma09g01230	Glyma.09G009500	Cationic amino acid transporter 8	Cheng <i>et al.</i> , 2016
Soja	Glyma09g05580	Glyma.09G050600	Cationic amino acid transporter 5	Cheng <i>et al.</i> , 2016
Soja	Glyma09g34750	Glyma.09G214300	Polyamine transporter	Cheng <i>et al.</i> , 2016
Soja	Glyma09g37700	Glyma.09G242000	Cationic amino acid transporter 9	Cheng <i>et al.</i> , 2016
Soja	Glyma10g41560	Glyma.10G269700	Amini acid permeasse BAT	Cheng <i>et al.</i> , 2016
Soja	Glyma13g31880	Glyma.13G245000	Lysine Histidine transporter 8	Cheng <i>et al.</i> , 2016
Soja	Glyma14g22120	Glyma.14G144700	Amino acid transporter	Cheng <i>et al.</i> , 2016
Soja	Glyma14g24370	Glyma.14G153500	Amino acid permeasse 3	Cheng <i>et al.</i> , 2016
Soja	Glyma15g07440	Glyma.15G068700	Lysine Histidine transporter 8	Cheng <i>et al.</i> , 2016
Soja	Glyma16g04050	Glyma.16G036300	solute carrier family 7	Cheng <i>et al.</i> , 2016
Soja	Glyma17g26590	Glyma.17G192000	Amino acid permeasse 1	Cheng <i>et al.</i> , 2016
Soja	Glyma17g32240	Glyma.17G212800	Amino acid permeasse 7	Cheng <i>et al.</i> , 2016
Soja	Glyma18g06650	Glyma.18G059300	Transmembrane aa transporter	Cheng <i>et al.</i> , 2016
Soja	Glyma18g42640	Glyma.18G198400	Auxin transporter 1	Cheng <i>et al.</i> , 2016
Soja	Glyma19g29440	Glyma.19G116400	solute carrier family 7	Cheng <i>et al.</i> , 2016
Soja	Glyma19g39060	Glyma.19G203200	Aromatic and neutral transporter	Cheng <i>et al.</i> , 2016

Fonte: Próprio autor.

2.6. FERRAMENTAS DA BIOTECNOLOGIA PARA AUMENTO DO TEOR DE PROTEÍNA DA SOJA

O mercado de sementes de soja é estimado em R\$ 3,7 bilhões, com mais de 1.300 cultivares, das quais mais da metade são geneticamente modificadas. Estima-se que 95,7% da soja plantada na safra 2019/2020 no Brasil, sejam geneticamente modificadas (Céleres, 2019), o que demonstra a importância das ferramentas biotecnológicas no melhoramento genético da cultura e lançamento de cultivares elite no mercado, a maioria resistentes a herbicidas (Isaaa, 2018). Os avanços na biotecnologia viabilizaram o mapeamento do genoma de várias espécies de plantas, possibilitando a identificação de genes com características fundamentais ao seu desenvolvimento e produtividade, tornando possível manipulá-los afim de melhorar suas características genéticas via ferramentas biotecnológicas (Michalak *et al.*, 2016).

Uma das maiores dificuldades encontradas nos programas de melhoramento da soja é aumentar o conteúdo de proteínas sem diminuir o rendimento no campo (Torres, 2015). Por este motivo, pesquisadores em todo mundo vêm realizando pesquisas afim de gerar cultivares que apresentem ambas as características e por meio de melhoramento genético clássico, conseguiram gerar linhagens de soja (TN03-350 e TN04-5321) com teor elevado de proteínas (43,1% a 43,9%) sem sacrificar o rendimento das sementes (Panthee; Pantalone, 2006; Medic; Atkinson; Hurburgh, 2014). Isso sugere, que apesar de não estar ainda elucidada as causas da correlação negativa entre a produção de sementes e a conteúdo proteico, isso poderá ser superado no futuro, a partir de novos estudos sobre os mecanismos de síntese, transporte e armazenamento de proteínas na semente de soja (Medic; Atkinson; Hurburgh, 2014).

Para tal, estudos à nível genômico e transcricional vêm sendo cada vez mais utilizados para entender os mecanismos envolvidos nesses processos (Mardis, 2008). Entre as ferramentas biotecnológicas para análise transcricional mais utilizadas está o RNA-Seq (Mardis, 2008). Essa técnica se baseia no uso do sequenciamento de nova geração (NGS) para revelar a presença e a quantidade de genes diferencialmente expressos em um determinado momento

e/ou condição (Mardis, 2008; Michalak *et al.*, 2016).

Na **Tabela 4**, estão presentes genes com validação de função através de ferramentas da biotecnologia, que podem indicar famílias gênicas promissoras para identificação de novos genes candidatos a transportadores de aminoácidos e ureídeos. Em soja, Collier e Tegeder (2012), a partir da família UPS1 (ureídeo sintase) identificada anteriormente em feijão (*Phaseolus vulgaris*), encontraram o ortólogo dessa família em soja, para demonstrar as funções de duas proteínas de soja GmUPS1-1 (XP_003516366) e GmUPS1-2 (XP_003518768) como atuantes no transporte de alantoína e ácido alantóico para fora do nódulo. Ainda nesse trabalho os autores realizaram estudos de localização subcelular e silenciamento gênico através de RNA-interferente (RNAi). Os resultados mostraram que ambos os transportadores se localizaram na membrana plasmática, células do córtex dos nódulos e na endoderme vascular. A repressão dos genes, levou a um acúmulo de ureídeos e a uma menor exportação de nitrogênio nas raízes e na parte aérea, além de afetar negativamente o desenvolvimento de nódulos, a fixação de nitrogênio e o metabolismo dos nódulos. O estudo no entanto não avaliou a influência desse gene na concentração de proteína no grão.

Carter e Tegeder (2016), superexpressaram um transportador UPS1 de feijão comum, nas células do córtex e da endoderme dos nódulos de soja, resultando aumento no transporte de N dos nódulos para a parte aérea e aumento na produção de sementes. Além disso, houve aumento no número de nódulos e na fixação simbiótica de N₂ (Tabela 4).

Wang *et al.* (2019), obtiveram dois mutantes com transportadores de lisina/histidina em arroz (*Oryza sativa*) (OsLHT1), usando a tecnologia de edição do genoma CRISPR / Cas9. O silenciamento do OsLHT1 inibiu o crescimento da raiz e da parte aérea do arroz, reduzindo assim significativamente o rendimento de grãos. A análise RT-qPCR indicou que o OsLHT1 foi expresso em vários órgãos de arroz, incluindo raiz, caule, folha bandeira, bainha de folha bandeira e panícula jovem. A expressão transiente no protoplasto de arroz, sugeriu que o OsLHT1 foi localizado na membrana plasmática, o que é consistente com sua função como transportador de aminoácidos (Tabela 4).

Yang, Stierhof e Ludewig (2015), silenciaram através da inserção de T-DNA, o transportador de aminoácidos catiônicos 9 (CAT9), que realiza transporte de aminoácidos como glutamina e glutamato. Além disso, realizaram estudos de localização intracelular. O gene se encontra expresso principalmente em membranas vesiculares que estão envolvidas no tráfego vacuolar, e realiza o transporte dos aminoácidos principalmente para fora do vacúolo das células. Ao silenciar o CAT9, houve diminuição no teor total de aminoácidos solúveis presentes nas folhas, o que pode interferir no teor de aminoácidos nas sementes (Tabela 4).

A prospecção e validação de genes envolvidos no transporte e acúmulo de proteína no grão sem afetar a produtividade torna-se fundamental (Lu *et al.*, 2012), pois são poucas as variedades de soja que apresentam ambas as características sendo este um grande gargalo dos programas de melhoramento genético (Willis 2003; Patil *et al.*, 2017). Além disso, o farelo e o óleo de soja comercializados, tanto no mercado externo como no interno, devem atender às especificações contratuais de exportação (vide ANEC - Associação Nacional dos Exportadores de Cereais) e comercialização interna (vide MAPA - Ministério da Agricultura Pecuária e Abastecimento, Portaria 795). Se não foram geradas cultivares com maiores teores proteicos a única solução para atingir este percentual é pela retirada do tegumento dos grãos, aumentando custos de produção (Pípolo *et al.*, 2015).

Tabela 4- Genes candidatos a transportadores de aminoácidos e ureídeos com dados de validação funcional.

Gene Code	Gene ID	Anotação	Validação	Fenótipo	Referências
Glyma01g07120 Glyma02g12970	/ Glyma.01G058500/ Glyma.02G116300	Ureide permease (Ureide_permease)	Superexpressão; Silenciamento através de RNAi	Superexpressão aumentou transporte de ureídeos para fora dos nódulos, enquanto a repressão diminuiu exportação de N para parte aérea.	Collier; Tegeder, 2012
Glyma02g12980	Glyma.02G116400	Ureide permease (Ureide_permease)	Superexpressão	Aumento da entrega de N na parte aérea	Carter; Tegeder, 2016
Glyma16g06750	Glyma.16G062600	Amino acid transporter	Mutante Crispr/Cas9 para silenciamento do gene	Inibição do crescimento de raiz e parte aérea	Wang <i>et al.</i> , 2019
At1G25270	Glyma.08G182500	EamA-like transporter family (EamA)	Superexpressão;	Aumento da transferência de aminoácidos durante o desenvolvimento de sementes no revestimento e calaza das sementes; Nocaute de genes causou diminuição de genes durante Embriogênese	Besnard <i>et al.</i> , 2019
Glyma10g41560	Glyma.10G269700	AMINO-ACID PERMEASE BAT1	Silenciamento	Diminuição do teor de aminoácidos em folhas	Yang; Stierhof; Ludewig, 2015

Fonte: Próprio autor

3. OBJETIVO

3.1. OBJETIVO GERAL

Identificar por meio de ferramentas *in silico* e validação *in vivo* genes associados ao transporte de compostos nitrogenados em soja e que contribuem para o aumento no teor de proteína nos grãos.

3.2 OBJETIVOS ESPECÍFICOS

- Identificar genes candidatos a transportar aminoácidos e ureídeos em soja, a partir de informações da literatura e bancos de dados de espécies modelo;
- Avaliar a expressão desses genes em bibliotecas de transcriptoma (RNAseq) de grãos e legumes de soja;
- Validar os genes selecionados na fase de 25-100mg, 100-200mg e 400-500mg com expressões acima de 40 TPM obtidos no RNAseq, por meio de experimento *in vivo* através de PCR em tempo real, em genótipos de soja que apresentam alto (BRS 232) e baixo (BRS 284) teor de proteína no grão;
- Correlacionar os níveis de expressão desses transportadores ao fenótipo das cultivares com alto (BRS 232) e baixo (BRS 284) teor de proteína no grão;
- Indicar os genes mais promissores para serem utilizados em estratégias futuras por meio da engenharia genética e que visem o desenvolvimento de genótipos de soja com maior teor de proteína no grão.

REFERÊNCIAS BIBLIOGRÁFICAS

ABIOVE – ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE ÓLEOS VEGETAIS. Exportações do complexo soja – Março, 2020. Disponível em: <http://abiove.org.br/estatisticas/exportacoes-do-complexo-soja/>. Acesso em: 20 junho 2020.

ABRASEM – ASSOCIAÇÃO BRASILEIRA DE SEMENTES E MUDAS. **Anuário 2015**. Brasília, 2016. Disponível em: http://www.abrasem.com.br/wp-content/uploads/2013/09/Anuario_ABRASEM_2015_2.pdf. Acesso em: 14 nov. 2019.

AINSWORTH, E.A.; YENDREK, C.R.; SKONECZKA, J.A.; LONG, S.P. Accelerating yield potential in soybean: potential targets for biotechnological improvement. **Plant, cell & environment**, v. 35, n. 1, p. 38-52, 2012.

ALCANTARA, R. M. C. M., REIS, V. M. Metabolismo do Carbono nos Nódulos. 2008 Disponível em: <<http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/630376>>. Acesso em: 20 nov. 2019.

ANEC – ASSOCIAÇÃO NACIONAL DE EDUCAÇÃO CATÓLICA DO BRASIL. Portaria Nº 71, de 10 de junho de 2020. Disponível em: <<https://anec.org.br/wp-content/uploads/2020/07/PORTARIA-N%C2%BA-71-DE-10-DE-JUNHO-DE-2020.pdf>>. Acesso em: 14 Nov. 2019.

ASSEFA, Y., PURCELL, L. C., SALMERON, M., NAEVE, S., CASTEEL, S. N., KOVÁCS, P., LINDSEY, L. E. Assessing variation in US soybean seed composition (protein and oil). **Frontiers in plant science**, v.10, 298 p., 2019.

BARAL, B., TEIXEIRA, J. A., IZAGUIRRE-MAYORAL, M. L. Early signaling, synthesis, transport and metabolism of ureides. **Journal of Plant Physiology**, v. 193, n. 1, p. 97–109, 2016.

BARBIERI, A. P. P., MATTIONI, N. M., HAESBAERT, F. M., DE ANDRADE, F. F., CABRERA, I. C., MERTZ, L. M. Teste de condutividade elétrica individual em sementes de soja e a relação com emergência de plântulas a campo. **Interciência**, Caracas, v. 38, n. 4, p. 310-315, Abril 2013.

BELLALOU, N., SMITH, J. R., GILLEN, A. M., RAY, J. D. Effects of maturity, genotypic background, and temperature on seed mineral composition in near-isogenic soybean lines in the early soybean production system. **Crop Sci**. May, V. 51, p. 1161–1171, 2011.

BELLIENY-RABELO, D., DE OLIVEIRA, E. A. G., DA SILVA RIBEIRO, E., COSTA, E. P., OLIVEIRA, A. E. A., & VENANCIO, T. M. Transcriptome analysis uncovers key regulatory and metabolic aspects of soybean embryonic axes during germination. **Scientific reports**, v. 6: 36009, 2016.

BESNARD, J., ZHAO, C., AVICE, J. C., VITHA, S., HYODO, A., PILOT, G., OKUMOTO, S. Arabidopsis UMAMIT24 and 25 are amino acid exporters involved in seed loading. **Journal of Experimental Botany**, Oxford, v.69, n. 21, p. 5221-5232, Aug. 2018.

BORÉM, Aluízio; MIRANDA, Glauco Vieira; FRITSCHÉ-NETO, Roberto. Melhoria de plantas. [S.l.: s.n.], 2017.

BRUMM, Thomas J.; HURBURGH, Charles R. Changes in long-term soybean compositional patterns. **Journal of the American Oil Chemists' Society**, v. 83, n. 11, p. 981-983, 2006.

CARRERA, C.; MARTÍNEZ, M.J.; DARDANELLI, J.; BALZARINI, M. Water deficit effect on the relationship between temperature during the seed fill period and soybean seed oil and protein concentrations. **Crop Science**, Madson, v.49, n.6, p.990-998, 2009.

CARRERA, C.; MARTÍNEZ, M.J.; DARDANELLI, J.; BALZARINI, M. Environmental variation and correlation of seed components in non-transgenic soybeans: protein, oil, unsaturated fatty acids, tocopherols and isoflavones. **Crop Science**, Madson, v.51, n.2, p.800-809, 2011.

CARTER, Amanda M.; TEGEDER, Mechthild. Increasing nitrogen fixation and seed development in soybean requires complex adjustments of nodule nitrogen metabolism and partitioning processes. **Current Biology**, v. 26, n. 15, p. 2044-2051, 2016.

CÉLERES. IB 19.01 – Informativo de biotecnologia céleres – novembro 2019. Uberlândia: Céleres, 2019. Disponível em: < http://www.celeres.com.br/wp-content/uploads/2019/11/BoletimBiotecnologiaC%C3%A9leres_Novembro2019-2.pdf>. Acesso em: 11 abr. 2021.

CERETTA, C. A., PAVINATO, A., PAVINATO, P. S., MOREIRA, I. C. L., GIROTTO, E., & TRENTIN, É. E. Micronutrientes na soja: produtividade e análise econômica. **Ciência Rural**, Santa Maria, v.35, n.3, p.576-581, 2005.

CHEN, W., LI, X., RAHMAN, M. R. T., AL-HAJJ, N. Q. M., DEY, K. C., RAQIB, S. M. Emulsification properties of soy bean protein. **Nusantara Bioscience**, v. 6, n. 2, 2014.

CHENG, L., YUAN, H. Y., REN, R., ZHAO, S. Q., HAN, Y. P., ZHOU, Q. Y., WANG, L. Genome-wide identification, classification, and expression analysis of amino acid transporter gene family in Glycine Max. **Frontiers in plant science**, v. 7, 515 p., Abril. 2016.

COLLIER, R.; TEGEDER, M. Soybean ureide transporters play a critical role in nodule development, function and nitrogen export. *The Plant Journal*, v. 72, n. 3, p. 355-367, June 2012.

CONAB. COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento**

de safra brasileiro – grãos: Décimo levantamento, julho 2019 – safra 2018/2019. : Brasília: Companhia Nacional de Abastecimento. 2019. Disponível em: <<https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>>. Acesso em : 17 jul. 2019.

CONAB. COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento de safra brasileiro – grãos:** Segundo levantamento, novembro 2019 – safra 2019/2020. : Brasília: Companhia Nacional de Abastecimento. 2019. Disponível em: <<https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>>. Acesso em : 22 nov. 2020.

DAMODARAN, S.; PARKIN, K. L. **Química de alimentos de Fennema**. 5. Porto Alegre: Artmed Editora, 2018.

DE OLIVEIRA JUNIOR, E. F., CASAROTTO, E. L., MENDONÇA, J. C. A., BINOTTO, E., & DA SILVA, M. V. A. P. Exportação de Soja no Estado de Mato Grosso do Sul: Características da Comercialização. **Revista em Agronegócio e Meio Ambiente**, Maringá, v. 11, n. 1, p. 71-97, 2018.

DELARMELINO-FERRARESI, L. M.; VILLELA, F. A.; AUMONDE, T. Z. Desempenho fisiológico e composição química de sementes de soja. **Revista Brasileira de Ciências Agrárias**, Recife, v. 9, n. 1, p. 14-18, 2014.

DOWNIE, J. A. Legume nodulation. **Current Biology**, v. 24, n. 5, p. 184-190, 2014.

EPSTEIN, E. Plants and inorganic nutrients. **Introduction to plant physiology**. New York, v. 2, p. 61-67, 1999.

FABRE, F.; PLANCHON, C. Nitrogen nutrition, yield and protein content in soybean. **Plant Science**, v. 152, n. 1, p. 51-58, 2000.

FEHR, W. R.; CAVINESS; C. E.; BURMOOD D. T.; PENNINGTON, J. S. Stage of soybean development. **Special Report**, v. 80, p. 929-931, 1977.

FONTES, Amanda Vilaça. **Processo de industrialização da soja no Brasil**. 2019. 65 f. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) – Universidade Federal de Uberlândia, Uberlândia, 2020.

FUKUI, T. de J.; LOPES, R. L. A expansão do cultivo da soja na área de influência da BR 163 e os impactos nas regiões produtoras do estado do Paraná. Universidade Estadual de Maringá (UEM). p. 1-24 Disponível em: . Acesso em: 19 nov. 2019.

GAZZONI, D. L. A soja no Brasil é movida por inovações tecnológicas. **Cienc. Cult.**, São Paulo , v. 70, n. 3, p. 16-18, 2018.

GONÇALVES, L. L.; YADA, M. M. Desempenho brasileiro no mercado da soja e sua utilização na alimentação de bovinos. *In: SIMTEC - Simpósio de Tecnologia da Fatec Taquaritinga*, v. 5, n. 1, p. 189-201, 22 dez. 2019.

GU, C., PAN, H., SUN, Z., QIN, G. Effect of soybean variety on anti-nutritional factors content, and growth performance and nutrients metabolism in rat. **International Journal of Molecular Sciences**, v. 11, n. 3, p. 1048-1056, 2010.

GULERIA, H., KUMAR, P., JYOTI, B., KUMAR, A., PALIWAL, A., & PALIWAL, A. Genetic variability and correlation analysis in soybean (*Glycine max* (L.) Merrill) genotypes. **IJCS**, v.7, n. 1, p. 1928-1932, 2019.

HERNÁNDEZ-SEBASTIÀ, C., MARSOLAIS, F., SARAVITZ, C., ISRAEL, D., DEWEY, R. E., HUBER, S. C. Free amino acid profiles suggest a possible role for asparagine in the control of storage-product accumulation in developing seeds of low-and high-protein soybean lines. **Journal of Experimental Botany**, v. 56, n. 417, p. 1951-1963, 2005.

HIRAKURI, M. H.; LAZZAROTTO, J. J. O agronegócio da soja nos contextos mundial e brasileiro. **Embrapa Soja-Documents (INFOTECA-E)**, 2014. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/104753/1/O-agronegocio-da-soja-nos-contextos-mundial-e-brasileiro.pdf>. Acesso em: 11 nov. 2019.

HUNGRIA, M.; CAMPO, R. J.; NOGUEIRA, M. A. A pesquisa em Fixação Biológica do Nitrogênio na Embrapa Soja: passado, presente e perspectivas futuras. In: Anais da XVI Relare. Anais.. Londrina: 2012. Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/105092/1/A-pesquisa-em-fixacao-biologica-do-nitrogenio-na-Embrapa-Soja-passado-presente-e-perspectivas-futuras.pdf>>.

ISAAA. Global Status of Commercialized Biotech/GM Crops in 2018: Biotech Crops Continue to Help Meet the Challenges of Increased Population and Climate Change. ISAAA Brief No. 54, 2019.

JONES, S. I.; GONZALEZ, D. O.; VODKIN, L. O. Flux of transcript patterns during soybean seed development. **BMC genomics**, v. 11, n.1, p. 136, Feb. 2010.

JONES, S. I., VODKIN, L. O. Using RNA-Seq to profile soybean seed development from fertilization to maturity. **PLoS one**, v. 8, n.3, Mar. 2013.

LEMOS, M. L. F., GUIMARÃES, D. D., MAIA, G. B. D. S., & AMARAL, G. F. Agregação de valor na cadeia de soja, 2017. Disponível em: <<https://web.bndes.gov.br/bib/jspui/handle/1408/14138>>. Acesso em: 20 nov. 2019.

LI, L., HUR, M., LEE, J.-Y., ZHOU, W., SONG, Z., RANSOM, N. WURTELE, E. A systems biology approach toward understanding seed composition in soybean. **BMC Genomics**, v. 16, n. 3, 2015.

LU, X., XIONG, Q., CHENG, T., LI, Q. T., LIU, X. L., BI, Y. D., ZHANG, J. S. A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. **Molecular Plant**, v. 10, n. 5, p. 670-684, 2017.

MANDARINO, J. M. G., DE OLIVEIRA, M.A., BENASSI, V. T., LEITE R.S. **Qualidade de sementes e grãos comerciais de soja no Brasil-safra 2017/18**. 1.ed. Londrina: Embrapa Soja, 2019. p.163. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1057495>>. Acesso em: 19 nov. 2019.

MANDARINO, J. M. G.; DE OLIVEIRA, M. A.; BENASSI, V. T.; LORINI, I.; FRANÇA-NETO, J. B.; HENNING, A. A.; KRZYZANOWSKI, F.C.; HENNING, F. A.; HIRAKURI, M. H. Teor de Proteína em Grãos de Soja Coletados nas Principais Regiões Produtoras do Brasil nas Safras de 2014/15 A 2016/17. *In*: VII Conferência Brasileira de Pós-Colheita, Londrina, 2018. **Anais** [...]. Londrina: Associação brasileira de pós-colheita (ABRAPOS), p. 618-623.

MAPA – Ministério da agricultura, Pecuária e Abastecimento, Brasil. Portaria Nº 16, de 20 de julho de 2017. Aprova o ZARC para o Estado do Paraná. Brasília, Diário Oficial da União de 21 de julho de 2017. Disponível em: <http://www.in.gov.br/materia//asset_publisher/Kujrw0TZC2Mb/content/id/19188996/do1-2017-07-21-portaria-n-16-de-20-de-julho-de-2017-19188914>. Acesso em: 19 nov. 2019.

MARDIS, Elaine R. Next-generation DNA sequencing methods. **Annu. Rev. Genomics Hum. Genet.**, v. 9, p. 387-402, 2008.

MEDIC, JELENA; ATKINSON, CHRISTINE; HURBURGH, CHARLES R. Current knowledge in soybean composition. **Journal of the American oil chemists' society**, v. 91, n.3, p. 363-384, Jan. 2014.

MELLO FILHO, O. L. D., SEDIYAMA, C. S., MOREIRA, M. A., REIS, M. S., MASSONI, G. A., PIOVESAN, N. D. Grain yield and seed quality of soybean selected for high protein content. **Pesquisa Agropecuária Brasileira**, Brasília v.39, n. 5, p. 445-450, May 2004.

MERTZ-HENNING, L. M., FERREIRA, L. C., MANDARINO, J. M. G., SANTOS, E. D., DE OLIVEIRA, M. C. N., NEPOMUCENO, A., FARIAS, J. R. B, NEUMAIER, N. Effect of Water Deficit-Induced at Vegetative and Reproductive Stages on Protein and Oil Content in Soybean Grains. **Journal Agronomy**, v. 8, p. 3, dez. 2017.

MICHALAK, P., SOBRAL, B. W., ABEDI, V., KIM, Y. B., DENG, X., PHILIPSON, C., BASSAGANYA-RIERA, J. From Big Data Analytics and Network Inference to Systems Modeling. *In* **Computational Immunology**, p. 113-144. Academic Press. 2016.

MÜLLER, B., FASTNER, A., KARMANN, J., MANSCH, V., HOFFMANN, T., SCHWAB, W., BLECKMANN, A. Amino acid export in developing Arabidopsis

seeds depends on UmamiT facilitators. **Current Biology**, v. 25, n. 23, p.3126-3131, Dec. 2015.

MURPHY, Patricia A. Soybean proteins. In: **Soybeans**. AOCS Press, 2008. p. 229-267.

NYOKI, Daniel; BALTAZARI, Prosper I. A.; NDAKIDEMI, P. A. The Role of Agro-Inputs (Rhizobia, P and K) and Critical Stages of Application for Improved Soybean Yield and Biomass. 2019. Disponível em: <<http://www.avidscience.com/wp-content/uploads/2017/10/The-Role-of-Agro-Inputs-Rhizobia-P-and-K-and-Critical-Stages-of-Application-for-Improved-Soybean-Yield-and-Biomass.pdf>>. Acesso em: 22 Ago. 2020.

OHYAMA, Takuji. Comparative studies on the distribution of nitrogen in soybean plants supplied with N₂ and NO₃-at the pod filling stage: II. Assimilation and Transport of Nitrogenous Constituents. **Soil science and plant nutrition**, Niigata, v. 30, n. 2, p. 219-229, Mar.1984.

OHYAMA, T., MINAGAWA, R., ISHIKAWA, S., YAMAMOTO, M., HUNG, N. V. P., OHTAKE, N., TAKAHASHI, Y. Soybean seed production and nitrogen nutrition. In: A comprehensive survey of international soybean research-Genetics, physiology, agronomy and nitrogen relationships. **IntechOpen**, 2013.

OLDROYD, Giles ED; DOWNIE, J. Allan. Coordinating nodule morphogenesis with rhizobial infection in legumes. **Annu. Rev. Plant Biol.**, v. 59, p. 519-546, 2008.

PANDURANGAN, S., PAJAK, A., MOLNAR, S. J., COBER, E. R., DHAUBHADEL, S., HERNÁNDEZ-SEBASTIÀ, C., MARSOLAIS, F. Relationship between asparagine metabolism and protein concentration in soybean seed. **Journal of experimental botany**, v. 63, n. 8, p. 3173-3184, 2012.

PANTHEE, D. R., PANTALONE, V. R., SAXTON, A. M., WEST, D. R., & SAMS, C. E. Genomic regions associated with amino acid composition in soybean. **Molecular Breeding**, v. 17, n. 1, p. 79-89, 2006.

PATHAN, S. M., VUONG, T., CLARK, K., LEE, J. D., SHANNON, J. G., ROBERTS, C. A., NGUYEN, H. T. Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. **Crop Science**, v. 53, n.3, p. 765-774, 2013.

PATIL, G., MIAN, R., VUONG, T., PANTALONE, V., SONG, Q., CHEN, P., NGUYEN, H. T. Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. **Theoretical and Applied Genetics**, v. 130, n. 10, p. 1975-1991, 2017.

PERDIGÃO, D. T.; FERNANDES, Dos Santos C. H.; BURATTO, J. S. SOJA: FENOLOGIA, MORFOLOGIA E FATORES QUE INTERFEREM NA PRODUTIVIDADE. **Revista científica eletrônica da FAEF**. v. 14. n. 1, junho 2019.

PEREZ-GRAU, L.; GOLDBERG, R. B. Soybean Seed Protein Genes Are Regulated Spatially during Embryogenesis. **The Plant Cell**, v. 1, p. 1095-1109, 1989.

PÍPOLO, A E., MANDARINO, J. M. G. Os teores de proteína da soja e a qualidade para a indústria. **Embrapa Soja-Artigo em periódico indexado (ALICE)**, 2016.

PÍPOLO, A. E., HUNGRIA, M., FRANCHINI, J. C., BALBINOT JUNIOR, A. A., DEBIASI, H., & MANDARINO, J. M. G. (2015). Teores de óleo e proteína em soja: fatores envolvidos e qualidade para a indústria. **Embrapa Soja-Comunicado Técnico (INFOTECA-E)**, 2015. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1025298>>. Acesso em: 11 nov. 2019.

RASO, M. J., MUÑOZ, A., PINEDA, M., PIEDRAS, P. Biochemical characterisation of an allantoate-degrading enzyme from French bean (*Phaseolus vulgaris*): the requirement of phenylhydrazine. **Planta**, v. 226, n. 5, p. 1333-1342, 2007.

SALDIVAR, Xiaoyu et al. Changes in chemical composition during soybean seed development. **Food Chemistry**, v. 124, n. 4, p. 1369-1375, 2011.

SALES, V. H. G., PELUZIO, J. M., AFFÉRRI, F. S., JUNIOR, W. P. O., SALES, P. V. G. Teor de óleo e proteína em grãos de soja em diferentes posições da planta. **Revista Agro@mbienteOn-line**, v. 10, n. 1, p. 22-29, 2016.

SCHMIDT, M. A., BARBAZUK, W. B., SANDFORD, M., MAY, G., SONG, Z., ZHOU, W., HERMAN, E. M. Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. **Plant Physiology**, v. 156, n. 1, p. 330-345, 2011.

SCHUBERT, K. R. Products of Biological Nitrogen Fixation in Higher Plants: Synthesis, Transport, and Metabolism. **Annual Review of Plant Physiology**, v. 37, n. 1, p. 539-574, 1986.

SEDIYAMA, T. Tecnologias de Produção e Usos da Soja. Londrina, v. 1, p. 314, 2009. Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/95489/1/SP-16-online.pdf>>. Acesso em: 21 nov. 2019.

SONG, M., FUNG, T. T., HU, F. B., WILLETT, W. C., LONGO, V. D., CHAN, A. T., & GIOVANNUCCI, E. L. Association of animal and plant protein intake with all-cause and cause-specific mortality. **JAMA internal medicine**, v. 176, n. 10, p. 1453-1463, 2016.

TAKAHASHI, M., UEMATSU, Y., KASHIWABA, K., YAGASAKI, K., HAJIKA, M., MATSUNAGA, R., ISHIMOTO, M. Accumulation of high levels of free amino acids

in soybean seeds through integration of mutations conferring seed protein deficiency. **Plants**, v. 217, n. 4, p. 577-586, 2003.

TAIZ, L.; ZEIGER, E. **Fisiologia vegetal**. 4.ed. Porto Alegre: Artmed, 2009. p. 819.

TEGEDER, M. Transporters involved in source to sink partitioning of amino acids and ureides: opportunities for crop improvement. **Journal of Experimental Botany**, v. 65, n. 7, p. 1865-1878, 2014.

THOMAS, R. J., SCHRADER, L. E. The assimilation of ureides in shoot tissues of soybeans: 1. Changes in allantoinase activity and ureide contents of leaves and fruits. **Plant physiology**, v. 67, n. 5, p. 973-976, 1981.

WANG, X., YANG, G., SHI, M., HAO, D., WEI, Q., WANG, Z., XIA, J. Disruption of an amino acid transporter LHT1 leads to growth inhibition and low yields in rice. **BMC Plant Biology**, v. 19, n. 1, p. 268, June 2019.

WRIGHT, D.; LENSSEN, A. W. Staging soybean development. **Agriculture and Environment Extension Publications**. V. 191. 2013. Disponível em: <https://lib.dr.iastate.edu/extension_ag_pubs/191/?utm_source=lib.dr.iastate.edu%2Fextension_ag_pubs%2F191&utm_medium=PDF&utm_campaign=PDFCoverPages> Acesso em: 5 fev. 2020.

WILLIS, Sara. The use of soybean meal and full fat soybean meal by the animal feed industry. In: **12th Australian soybean conference. Soy Australia, Bundaberg**. 2003.

Xu, X. P., Liu, H., Tian, L., Dong, X. B., & Shen, S. H. Integrated and comparative proteomics of high-oil and high-protein soybean seeds. **Food Chemistry**, v. 172, p. 105-116, Apr. 2015.

YANG, H., STIERHOF, Y. D., & LUDEWIG, U. The putative Cationic Amino Acid Transporter 9 is targeted to vesicles and may be involved in plant amino acid homeostasis. **Frontiers in plant science**, v. 6, 212 p., Abril. 2015.

YU, X., YUAN, F., FU, X., & ZHU, D. Profiling and relationship of water-soluble sugar and protein compositions in soybean seeds. **Food Chemistry**, v. 196, p. 776-782, 2016.

ZHANG, Yaoyang et al. Protein analysis by shotgun/bottom-up proteomics. **Chemical Reviews**, v. 113, n. 4, p. 2343-2394, 2013.

ARTIGO 1: PROSPECÇÃO *IN SILICO* E AVALIAÇÃO DA EXPRESSÃO DE GENES ASSOCIADOS AO TEOR DE PROTEÍNA EM GRÃOS DE SOJA

Resumo

A soja é a principal fonte de proteína vegetal para produção de ração animal. Mas para que possa ser utilizada para tal finalidade, o teor de proteína mínimo do grão deve ser de 35%, viabilizando a produção de farelo com o mínimo de 46% de proteína. Em algumas situações, esse teor não é alcançado. Com o aumento da produtividade alcançado através do melhoramento genético, há redução no teor proteico. Os motivos dessa correlação negativa ainda permanecem desconhecidos. Para solucionar esse viés, torna-se fundamental entender o processo de produção, transporte e armazenamento de proteína na semente da soja. Para identificar genes associados a essas funções foi realizada a análise de 3 datasets de bibliotecas de RNA-Seq independentes, contendo informações transcricionais de tegumento, sementes em desenvolvimento, semente madura e legumes de soja. A partir destes dados, foi possível quantificar 55.589 genes, destes foram identificados 76 TCNs presentes no genoma de soja e *Arabidopsis*. Estes 76 genes foram divididos em dois grupos. O primeiro contendo 50 TCNs com altos níveis de expressão em transcritos por milhão-TPM (≥ 40) nas bibliotecas referentes aos estádios de maior síntese de proteínas reserva (100-200mg e/ou 400-500mg) e 26 TCNs com baixo nível de expressão (menor que 40 TPM) nestas fases. A partir dos 50 transportadores de alta contagem foram selecionados sete genes (UMAMIT25, CAT9, BAT1, AVT3, UPS2, LHT1, AAP7) para validação via RT-qPCR, em sementes de BRS 232 e BRS 284, com alto e baixo teor de proteína respectivamente. Em relação a cultivar de calibração BRS 284, é possível observar que apenas os genes AAP7, AVT3, CAT9 e UMAMIT25 apresentam-se *up*-regulados em todas fases de desenvolvimento da semente na cultivar de maior teor proteico (BRS 232). Em legume, os únicos genes que se apresentaram superexpresso em duas fases foram o gene CAT9 e o gene UMAMIT25. O gene CAT9 apresentou *up*-regulação nas fases de 25-100mg e 100-200mg. O gene UMAMIT25 apresentou *up*-regulação em legume de 100-200mg e 250-450mg. A superexpressão dos genes UMAMIT25 e CAT9 em futuros estudos, podem acarretar na solução da correlação negativa entre proteína e produtividade.

Palavras-chave: *Glycine max*, transportadores de compostos nitrogenados, RNA-Seq.

PAPER 1: *In silico* prospection and evaluation of the gene expression associated to protein content in soybeans.

ABSTRACT

Soybean grain is the main plant source of protein for animal food. But, to be used for that, soybeans need to have at least 35% of protein, making feasible a soybean meal with at least 46% of protein. However, in some situations, this content is not achieved. With the increase in productivity achieved through genetic improvement, there is a reduction in protein content. The reasons of this negative correlation remain unknown. To solve this bias, the understand of production process, transport and storage of protein in grain is essential. Based on this knowledge, it is possible to prospect and manipulate genes that may contribute to this negative correlation. In order to identify genes associated and their functions, 3 datasets of independent RNA-Seq libraries containing the transcription information about seed coat, seed development, mature seed and soybean pods were analysed. In these datasets were found 55,589 genes, of which 76 TCNs were identified in genomes of soybean and Arabidopsis. These 76 genes were divided in two groups. The first containing 50 TCNs with high levels of transcripts per million - TPM (≥ 40) in libraries from higher protein storage (100-200mg and/or 400-500mg) and 26 TCNs with low expression level (less than 40) in those phases. Considering the 50 transporters with higher counts, seven genes (UMAMIT25, CAT9, BAT1, AVT3, UPS2, LHT1, AAP7) were selected for validation via RT-qPCR, in seeds of BRS 232 and BRS 284, with high and low protein content respectively. Regarding the BRS 284 calibration cultivar, it is possible to observe that only the AAP7, AVT3, CAT9 and UMAMIT25 genes are up-regulated at all stages of seed development in the cultivar with the highest protein content (BRS 232). In pods, only the genes CAT9 and UMAMIT25 were overexpressed in two phases. The CAT9 gene was up-regulation in the 25-100mg and 100-200mg stages. While, the UMAMIT25 gene exhibited up-regulation in 100-200mg and 250-450mg pods. Overexpression of the UMAMIT25 and CAT9 genes, in further studies, may result in the solution of the negative correlation between protein and yield.

Introdução

A soja é uma cultura de fundamental importância sócio-econômica em nível global (Conab, 2019; Conab, 2020). No Brasil, é a principal commodity produzida e exportada, representando 48% da produção total de grãos (Conab, 2019). O País produz em torno de 124,8 milhões de toneladas em uma área aproximada de 37 milhões de hectares, ficando em primeiro lugar no ranking de produção mundial do grão (Conab, 2020), tendo como principais estados produtores, Mato Grosso, Paraná e Goiás (Conab, 2020).

Por apresentar proteína qualitativamente equivalente à proteína animal, é amplamente utilizada em forma de farelo para indústria de ração animal (Pípolo; Mandarino, 2016; Gazzoni, 2018). A produção de ração animal é estimada em pouco mais de um bilhão de toneladas por ano, movimentando ativamente a economia mundial (IFIF, 2019), porém para ser utilizada para este fim, o farelo deve conter um teor de proteína mínimo de 46%, segundo a ANEC-71 (Pípolo *et al.*, 2015).

O teor de proteína mínimo exigido pela indústria moageira de soja é de 35% nos grãos para atingir o mínimo de 46% exigido no farelo (PATIL *et al.*, 2017). O conteúdo de proteína nos grãos da soja geralmente atende aos padrões industriais, porém ao longo dos anos, em parte devido as alterações genéticas promovidas pelo melhoramento genético da cultura, a indústria moageira tem encontrado dificuldade em obter grãos que atendam ao padrão proteico estabelecido (Pípolo *et al.*, 2015). Na tentativa de alcançar o teor de proteína adequado, utiliza-se a estratégia de remover o tegumento da semente, entretanto, esta etapa a mais no processamento da soja aumenta consideravelmente os custos do processamento (Pathan, 2013; Pípolo *et al.*, 2015).

Linhas de soja com o mínimo de (Medic; Atkinson; Hurburgh, 2014), mas as linhas de alta proteína são caracterizados por um número menor de sementes por planta (Rotundo *et al.*, 2009; Medic; Atkinson; Hurburgh, 2014).

Por este motivo, um dos maiores problemas enfrentados neste nicho é a correlação negativa entre rendimento de grãos e teor de proteína (Pathan *et al.*, 2013). Isso resulta da ação do programa de melhoramento genético na cultura da soja que ao longo dos anos busca o desenvolvimento de cultivares mais produtivas, resultando em plantas com menores teores proteicos (Pathan *et al.*, 2013). Estudos realizados em um período de 10 anos, entre 1994 e 2004, mostraram que o teor proteico decresceu 0,1% neste período conforme o aumento de produtividade era alcançado (Brumm; Hurburgh 2006; Pathan, 2013). Segundo Pípolo *et al.* (2015), devido ao desempenho no aumento da produtividade da cultura a tendência é aumentar ainda mais a queda dos teores de proteína ao longo dos anos.

Na tentativa de evitar a perda de teor proteico, entender os mecanismos envolvidos na produção, transporte e armazenamento de proteínas no grão pode auxiliar como ferramenta para geração de cultivares elite que apresentem alto teor proteico no grão e que sejam ao mesmo tempo produtivas (Michalak *et al.*, 2016). Uma possibilidade é prospectar genes envolvidos na síntese de proteína no grão (Mardis, 2008). Para tal, ferramentas como RNA-Seq vem sendo amplamente utilizadas para estudar a nível transcricional, genes fundamentais a planta para melhorar tais fenótipos (Mardis, 2008; Michalak *et al.*, 2016). Entre os genes que podem estar envolvidos no aumento do teor de proteína no grão encontram-se os transportadores de compostos nitrogenados (TCNs) como aminoácidos e ureídeos (Ohyama, 1984), uma vez que em soja há uma alta demanda de nitrogênio para obtenção de produtividade e grãos com valor considerável de proteínas, por exemplo para uma produtividade de 3,500kg ha⁻¹ com teor de proteína no grão de 40%, necessita-se de aproximadamente 224 kg ha⁻¹ de N (Hungria *et al.*, 2012).

O aumento da disponibilidade de N para parte aérea, aumenta o teor de proteína das sementes (Fabre; Planchon, 2012). Trabalhos prévios em planta modelo (*Arabidopsis thaliana*) mostram que a superexpressão de alguns TCNs gera o aumento do transporte do N, aumentando também a produtividade (Carter e Tegeder, 2016; Besnard *et al.* 2018). No trabalho de Carter e Tegeder (2016), a superexpressão do gene UPS1 (Ureide permease 1) em *Arabidopsis thaliana* gerou um aumento em 36% no rendimento geral das sementes. Besnard *et al.*, 2018, também em *Arabidopsis*, perceberam aumento no número de sementes após superexpressarem o TCN UMAMIT25.

O avanço no conhecimento acerca dos mecanismos moleculares, através dos quais esses transportadores atuam, especialmente durante a formação e enchimento do grão, será útil no desenho de estratégias biotecnológicas que visem a obtenção de cultivares de soja com maior capacidade de síntese de proteínas de reserva resultando em grãos de alto teor de proteína e plantas de alta produtividade. Diante disso, esse trabalho foi desenvolvido com o objetivo de identificar por meio de estudos de transcriptoma, genes envolvidos no processo de transporte de aminoácidos e ureídeos nas sementes de soja, a fim de selecionar alvos promissores para o melhoramento genético da cultura através de ferramentas de edição.

Material e Métodos

Identificação de genes candidatos a transportadores de compostos nitrogenados expressos em sementes de soja

Os genes selecionados foram obtidos a partir de 3 datasets de bibliotecas de RNA-Seq independentes, cujas características quanto a plataforma de sequenciamento utilizada, tipo e tamanho dos reads, estão descritas na Tabela 1.

Tabela 1. Caracterização das Bibliotecas de RNA-Seq utilizadas para a obtenção de genes alvos, a partir de contagens em transcritos por milhão-TPM. **A.** Dataset 1. Bibliotecas de RNA-Seq de sementes em desenvolvimento e tegumento disponibilizadas no banco de dados SRA do NCBI através do bioprojeto PRJNA184365 (Jones e Vodkin, 2013). **B.** Dataset 2. Bibliotecas de RNA-Seq geradas pela Embrapa Soja a partir de 3 genótipos de soja (BRS MG715, DM 6563 e BRS 413RR) com teores variados de proteína em estágio de maturação plena. **C.** Dataset 3. Bibliotecas de RNA-Seq geradas pela Embrapa Soja em legumes de soja (BR16).

A. Dataset 1	Plataforma	Tipo	Tamanho do Read	Data	Tecido	Estágio
SRX212252	Illumina Genome Analyzer II	Single-end	40-75	2015	Cotilédone	100-200mg peso fresco
SRX212246	Illumina Genome Analyzer II	Paired-end	80-142	2015	Semente inteira	12-14 dias após o florescimento
SRX212244	Illumina Genome Analyzer II	Paired-end	80-142	2015	Semente inteira	4 após o florescimento
SRX212254	Illumina Genome Analyzer II	Paired-end	40-130	2015	Cotilédone	400-500mg peso fresco
SRX212250	Illumina Genome Analyzer II	Paired-end	80-142	2015	Semente inteira	5-6mg peso fresco
SRX212256	Illumina Genome Analyzer II	Paired-end	40-130	2015	Semente inteira	Semente madura
SRX212253	Illumina HiSeq 2000	Single-end	40-75	2013	Cotilédone	100-200mg peso fresco
SRX212247	Illumina HiSeq 2000	Single-end	40-75	2013	Semente inteira	12-14 dias após o florescimento
SRX212249	Illumina HiSeq 2000	Single-end	40-75	2013	Semente inteira	22-24 dias após o florescimento
SRX212248	Illumina HiSeq 2000	Single-end	40-75	2015	Semente inteira	22-24 dias após o florescimento
SRX212245	Illumina HiSeq 2000	Single-end	40-100	2013	Semente inteira	4 dias após o florescimento
SRX212255	Illumina HiSeq 2000	Single-end	40-75	2013	Cotilédone	400-500mg peso fresco
SRX212251	Illumina HiSeq 2000	Single-end	40-75	2013	Semente inteira	5-6mg peso fresco
SRX212258	Illumina HiSeq 2000	Single-end	40-75	2015	Cotilédone	5-6mg peso fresco
SRX212259	Illumina HiSeq 2000	Single-end	40-75	2015	Tegumento	5-6mg peso fresco
SRX212257	Illumina HiSeq 2000	Single-end	40-100	2013	Semente inteira	Semente madura
B. Dataset 2	Plataforma	Tipo	Tamanho	Data	Tecido	Estágio
51 - BRS 413 RR	Illumina Novaseq 6000	Paired-end	40-75	2019	Semente inteira	Semente madura
53 - BRS 413 RR	Illumina Novaseq 6000	Paired-end	40-75	2019	Semente inteira	Semente madura
62 - DM 6563	Illumina Novaseq 6000	Paired-end	40-75	2019	Semente inteira	Semente madura
63 - DM 6563	Illumina Novaseq 6000	Paired-end	40-75	2019	Semente inteira	Semente madura
81 - BRS MG75	Illumina Novaseq 6000	Paired-end	40-75	2019	Semente inteira	Semente madura
82 - BRS MG75	Illumina Novaseq 6000	Paired-end	40-75	2019	Semente inteira	Semente madura
C. Dataset 3	Plataforma	Tipo	Size	Data	Tecido	Estágio
Legumes	Illumina NextSeq 500	Paired-end	40-130	2017	Legume	R4
	Illumina NextSeq 500	Paired-end	40-130	2017	Legume	R4
	Illumina NextSeq 500	Paired-end	40-130	2017	Legume	R4
	Illumina NextSeq 500	Paired-end	40-130	2017	Legume	R4

Pipeline para análise das bibliotecas de RNA-Seq

A qualidade dos fragmentos antes e após a limpeza foi realizada pelo *software* FastQC versão 0.11.5 (Patel; Jain 2012; Andrews *et al.* 2017). Após a limpeza, foram resgatados apenas fragmentos com tamanho ≥ 40 pb contendo Phred Score ≥ 20 através do *software* Trimmomatic versão 0.36. Após o resgate de fragmentos com alta qualidade, os reads foram alinhados sobre o genoma de soja versão Wm82.a2.v1 através do *software* Hisat2 versão 2.1.0, com recuperação apenas de fragmentos contendo alinhamento único (KIM *et al.* 2015). O genoma de referência e sua anotação foram obtidos do banco de dados Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Gmax). Os artefatos de PCR foram removidos pelo *software* Samtools rmdup versão 1.5 (Li *et al.* 2009). O mapeamento e normalização entre as bibliotecas em TPM foi realizada pelo *software* Stringtie versão 1.3.3 (Pertea *et al.*, 2015).

Seleção de genes candidatos a transportadores de compostos nitrogenados (TCNs)

A identificação desses genes foi a partir de dados da literatura, em artigos que caracterizam famílias de transportadores em *Arabidopsis thaliana* (Müller *et al.*, 2015) e em soja (Cheng *et al.*, 2016; Patil *et al.*, 2017; Zhou *et al.*, 2020), buscando representar as principais famílias de TCNs (*Amino acid transporter* - CAT, *Lysine histidine transporter* - LHT, *Aminoacid permease* - AAP, UMAMIT – *Usually multiple acids move in and out transporters*, *Bidirectional amino acid transporter* - BAT, *Ureide permease* - UPS, *Aminoacid vacuolar transporter* - AVT) (Fujiki *et al.*, 2017; Zhou *et al.*, 2020), que passaram por validação prévia quanto a função, tendo como critério apresentar alta expressão (TPM ≥ 40) (Pomaznoy *et al.*, 2019), em pelo menos uma das bibliotecas de RNA-Seq referente ao estádio 100-200mg e 400-500mg, fases em que há maior acúmulo de proteína reserva no grão (Jones e Vodkin, 2013). A partir dos genes identificados nas bibliotecas de RNAseq, foram selecionados oito genes para serem validados por meio de PCR em tempo real, em sementes de soja de cultivares com variação no teor de proteína no grão.

Caracterização in silico dos genes alvos selecionados para validação via RT-qPCR

Foi realizada a caracterização dos genes quanto ao número de transcritos e cópias em soja e ortólogos em *Arabidopsis thaliana* (TAIR10) através do banco de dados phytozome (<https://phytozome.jgi.doe.gov/>) e do *software* Persephone (<https://persephonesoft.com/>). Para tal, os genomas TAIR10, W82.a2.v1 e W82.a4.v1 foram utilizados como referência. Apenas genes contidos nas duas versões de montagem e anotação do genoma de soja foram considerados, a fim de evitar selecionar genes provenientes de erros de sequenciamento. As análises filogenéticas foram realizadas a partir de sequências proteicas pelo *software* ItoI (<https://itol.embl.de/upload.cgi>), como alinhador foi utilizado o *software* Muscle (<https://www.ebi.ac.uk/Tools/msa/muscle/>). A análise de domínios proteicos foi realizada pelo *software* MotifScan (https://myhits.sib.swiss/cgi-bin/motif_scan) e o banco de dados Ensemble (<https://www.ensembl.org/index.html>). Os heatmaps foram gerados a partir do *software* ClustVis (Molinari *et al.*, 2021).

Desenho dos primers para RT-qPCR

O desenho dos primers foi realizado através do *software* Primer3plus (<https://primer3plus.com/cgi-bin/dev/primer3plus.cgi>). A checagem da formação de homodímeros e heterodímeros foi realizada pelo *software*

Multiple Primer Analyzer (<https://www.thermofisher.com/br/multiple-primer-analyzer.html>). Somente primers sem a formação de estruturas secundárias foram selecionados. Os iniciadores (*primers*) utilizados para validação estão descritos na Tabela 2.

Tabela 2. Identificação e sequência dos iniciadores F (Forward) e R (Reverse) dos genes codificadores de transportadores de compostos nitrogenados-TCN para RT-qPCR.

Gene ID	Anotação	Eficiência dos primers	Sequência dos Primers 5' - 3'	TM
Glyma.03G229700	EamA-like transporter family (EamA) – UMAMIT 25	94%	primer F 5'GCTAGTAGCTAAGGATTGACAAGT 3' primer R 3'ATATGTCTGCAATGCAACTG5'	60 60
Glyma.09G242000	Cationic Amino Acid Transporter 9 - CAT 9	95%	primer F 5'TTCAGTGAGTGCACCAAAA 3' primer R 3'AGAATTAGCGCGAAAATGAA5'	60 60
Glyma.10G269700	Amino-Acid Permease - BAT1	99%	primer F 5' GCCTGTTTCCTACCCAATAA 3' primer R 3'AATAAGCAAACATCCAACGG 5'	60 60
Glyma.16G062600	Amino Acid Transporter - LHT1	92%	primer F 5'CATCTGCAACTTGCTTGTTA3' primer R 3'TCTGATGTGAATTCAGTGG5'	60 60
Glyma.14G010300	Amino-Acid Permease 7 - AAP7	90%	primer F 5'GCTAAGCGAGCAACTAACTG3' primer R 3'CCTGTGCGTCTGTTGGAT5'	60 60
Glyma.11G086700	Amino Acid Transporter – AVT3A	95%	primer F 5' TCCTACTCACTCACAGCGTA3' primer R 3'GACGATGGAGATGAAGATGT5'	60 60
Glyma.02G116400	Ureide permeasse – UPS2	93%	primer F 5'GTTAGGGAAGAGCACTTTCA3' primer R 3'AACCCCTTTCTTCAAAGTGT5'	60 60

Experimento em casa de vegetação para coleta de material biológico para validação dos TCNs selecionados via RT-qPCR

Local do experimento e Material biológico

O experimento foi realizado em casa de vegetação com condições de cultivo controladas. Foram utilizadas sementes de duas cultivares convencionais de soja com alto potencial produtivo, sendo uma com teor mais alto de proteína no grão, a cultivar BRS 232 (Teor proteico: $38,24 \pm 0,56$, peso de 100 sementes:19.7g), e BRS 284 com menor teor proteico (Teor proteico: $34,27 \pm 0,49$; peso de 100 sementes:15.2g).

Delineamento experimental

O delineamento experimental foi realizado em blocos casualizados (DBC) com 2 cultivares convencionais de soja (BRS 232 e BRS 284) em 18 repetições biológicas, totalizando 36 vasos. Foram utilizados vasos de 10 litros com solo esterilizado, seco e peneirado, em uma composição de 1:1 solo/areia. Em cada vaso foi transplantado duas plântulas contendo radícula uniforme. Antes do plantio, as raízes das plântulas foram tratadas com inoculante *Bradyrhizobium japonicum* (Atmo®), $5,0 \times 10^9$ UFC/ mL. A irrigação foi realizada por sistema de gotejamento, para manter uma condição hídrica favorável.

As plântulas foram selecionadas a partir da germinação das sementes em rolos de papel Germitest®, umedecidos com teor de água 2,5 vezes o peso da massa seca do papel, de acordo com o manual de regras de análise de sementes (BRASIL, 2009). Foi utilizado um rolo de 50 sementes para cada cultivar. Os rolos foram mantidos em câmara de germinação em temperatura de $25 \pm 1^\circ\text{C}$ à 100% de umidade relativa durante quatro dias (BRASIL, 2009).

Coleta de amostras biológicas para análises moleculares

Para as análises moleculares, as 18 plantas de cada cultivar foram divididas em 3 bulks de 6 plantas (6 vagens cada planta), totalizando 36 vagens para cada bulk e em torno de 108 sementes para BRS 284 e 72 sementes para BRS 232, para cada um dos 3 estádios (25-100mg, 100-200mg, 350-450mg) separadamente. As vagens e suas respectivas sementes foram separadas e armazenadas em freezer a -80°C até o momento da extração do RNA.

Extração do RNA e validação via RT-qPCR

Para extração do RNA total, foram realizadas as macerações dos legumes e das sementes já separados em seus bulks. Após a maceração, o RNA total foi extraído usando o reagente Trizol®. Para cada 100 mg de amostra, foi utilizado 1 ml do reagente trizol, que é vortexado e incubado por 5 minutos em temperatura ambiente. Em seguida, foi adicionado 200 μL de clorofórmio, vortexado e colocado em centrifuga a 12000g/ 4°C / 15 minutos. Transfere-se a parte aquosa para um novo tubo e adiciona-se 500 μL de etanol, então a mistura foi agitada e centrifugada a 12000g/ 4°C / 10 minutos. O sobrenadante é descartado, e é adicionado etanol 75% centrifugado em 7500 g/5 minutos. O RNA é então ressuscitado em água livre de RNase.

O RNA foi tratado com o kit DNase I (Invitrogen, Carlsbad, CA) para remoção de DNA remanescente. Em seguida, foi realizada uma PCR convencional com o primer B-actina intron para confirmar a ausência de DNA genômico (Gdna). Depois disso, o cDNA foi sintetizado usando Super Script® III First-Strand Synthesis System

(Invitrogen, Carlsbad, CA) de acordo com as instruções do fabricante. Novamente foi realizada uma PCR para verificar a eficiência da síntese e ausência de gDNA. As reações de RT-qPCR foram compostas por cDNAs, primers F e R de 0,2 μ M e tampão de reação 1x Platinum® SYBR Green® qPCR SuperMix UDG (Invitrogen, Carlsbad, CA). A normalização de expressão gênica foi realizada utilizando os genes endógenos β -actina (iniciador F 5' TTCTGTCTTCTGCAAGTGGTG 3' / iniciador R 5' GATCCCTCATCCATACATTTCAG 3') e ELF1 β (primer F 5' GTTGAAAAGCCAGGGGACA 3' / primer R 5' TCTTACCCCTTGAGCGTGG 3') genes, selecionados a partir do trabalho de Wan *et al.* (2017). A quantificação relativa da expressão gênica foi realizada a partir de três repetições biológicas e três técnicas. As reações foram feitas em equipamento de Real Time PCR 7900HT (Applied Biosystems). As condições de ciclagem utilizadas foram 50°C por 2 min, desnaturação a 95°C por 10 min, seguidas de 40 ciclos de 95°C por 15 s, 60°C por 1 min, 95°C por 15 s, 60°C por 15s e 95°C por 15 s. As eficiências dos primers de PCR foram calculadas usando o *software* LinRegPCR v.2012.0 (Ruijter *et al.* 2009), considerando como primers eficientes aqueles apresentando valores \geq 85%. O nível de expressão foi determinado usando a fórmula $2^{-\Delta\Delta C_t}$ adaptada de acordo com as eficiências dos primers (Livak; Schmittgen, 2001).

Coleta de amostras biológicas para análises fisiológicas

A coleta das vagens e sementes de BRS 232 e BRS 284 foram coletadas nos estádios: R5.1/R5.2- referente a sementes com peso fresco de 25-100mg, R5.3- referente a sementes com peso fresco de 100-200mg e R6 – referente a sementes com peso fresco de 350-450mg. A seleção destes estádios se deu de acordo com o tipo de proteína sintetizadas durante o desenvolvimento da semente. Até 100 mg predomina-se a síntese de proteínas metabólicas, após este peso inicia-se a síntese de proteínas reserva.

Peso das vagens e sementes

Foi coletada 1 vagem de cada uma das 18 plantas para cada um dos 3 estádios separadamente. As vagens e suas respectivas sementes foram separadas e pesadas logo após a coleta, para obtenção do peso fresco. O peso seco foi obtido a partir das amostras secas durante 24 horas em estufa à 65°C.

O peso de sementes foi obtido a partir de uma amostra de 18 sementes para cada cultivar em cada um dos três estádios. O peso foi medido em miligramas (mg) em balança analítica.

Mensuração do teor de proteína

Para a quantificação dos teores de proteína, foi utilizado equipamento de espectroscopia de infravermelho próximo (FT-NIR) (Thermo Scientific, modelo Antaris II), dotado de esfera de integração e leitura na faixa entre 1100 e 2500 nanômetros. Para cada amostra de aproximadamente 20g de cada cultivar, foram coletados 32 scans com resolução de 4cm⁻¹, e background a cada leitura. Os espectros médios foram submetidos à predição em quatro modelos matemáticos distintos para cada constituinte. A modelagem foi baseada na regressão por mínimos quadrados parciais (PLS), com pré-tratamentos para normalização dos dados (SNV) e aplicação de derivada primeira de Savitsky-Golay.

Análises estatísticas

Os dados fisiológicos foram submetidos a teste de análise de variância (ANOVA) e comparação de médias pelo teste de Tukey ($p \leq 0.05$). Os dados moleculares foram submetidos a ANOVA e comparação de médias pelo teste T ($p \leq 0.05$).

Resultados

Identificação de genes transportadores de compostos nitrogenados-TCNs em transcriptoma de sementes, tegumento e legumes de soja

Foram mapeados nos 3 datasets um total de 55.589 genes em TPM. O direcionamento pela busca por anotações biológicas foi baseado nos artigos que caracterizaram as famílias de transportadores em soja e *Arabidopsis* (Müller *et al.*, 2015; Cheng *et al.*, 2016; Patil *et al.*, 2017; Zhou *et al.*, 2020).

Foram identificados 76 TCNs comuns em ambos os genomas (Arquivo adicional 1). Entre esses, 50 TCNs continham altos níveis de expressão (≥ 40 TPM) em sementes nos estádios em que ocorre a síntese de proteínas de reserva (C 100-200mg PF W82 e C 400-500mg PF W82) (Arquivo Adicional 1; Figura 1). A expressão em tegumento e legume em outras fases de desenvolvimento de sementes também foram avaliadas para auxiliar na compreensão dos mecanismos envolvidos no acúmulo de proteínas nos grãos. Os outros 26 TCNs apresentaram altas contagens em outras fases de desenvolvimento da semente e/ou outros tecidos (tegumento e legumes). As famílias observadas nos dois subgrupos (50 e 26) foram as mesmas (AAP, AAT, AVT, BAT, CAT, LAX, LHT, PROT, PUT, UMAMIT e UPS), porém os genes membros destas famílias foram diferentes. A família UMAMIT foi a única a conter genes presentes somente no subgrupo de 50 TCNs.

Na família AAP foram observados seis membros, sendo que AAP1, AAP2, AAP3 E AAP7 estavam presentes no subgrupo de 50 TCNs e AAP6 e AAP8 no subgrupo de 26 TCNs. Para a família AAT foi encontrado apenas um membro, com a presença de três genes codificantes para este. O Glyma.11G166900 e Glyma.18G59300 no subgrupo de 50 TCNs e o Glyma.06G019500 presente no subgrupo de 26 TCNs. Dentre a família AVT foram observados cinco membros (AVT1, AVT1C, AVT3, AVT6A e AVT6C). Os membros AVT1 e AVT1C encontram-se no subgrupo de 26 TCNs, enquanto os membros AVT3, AVT6A e AVT6C no subgrupo de 50 TCNs. Na família BAT foi observado um membro (BAT1), codificado por dois genes. O Glyma.10G269700 no subgrupo de 50 TCNs e o Glyma.20G12150026 no subgrupo de 26 TCNs. Dentre a família CAT, foram identificados seis membros, estando os membros CAT2, CAT4, CAT6 e CAT8 presentes no subgrupo de 50 TCNs, apenas o CAT5 encontra-se no grupo de 26 TCNs. A família LAX apresentou dois membros, o LAX2 no subgrupo de 50 TCNs e LAX1 presente no subgrupo de 26 TCNs (Arquivo adicional 1).

A família LHT, também apresentou dois membros, estando o membro LHT1 em ambos os subgrupos, pois seus codificantes, Glyma.16G062600 está presente no subgrupo de 50 TCNs e o Glyma.19G076800 presente no subgrupo de 26 TCNs. O segundo membro, referente a um LHT8 encontra-se no grupo de 50 TCNs. A família PROT, também apresentou dois membros, o PROT1 presente no subgrupo de 50 TCNs e o PROT3 no subgrupo de 26 TCNs. A família PUT está representada por três membros, dos quais o PUT2 e PUT5 encontram-se no subgrupo de 50 TCNs, enquanto o membro PUT4 apresenta genes em ambos os subgrupos, sendo Glyma.08G113400 e Glyma.14G177500 presentes no subgrupo de 50 TCNs e o Glyma.02G209400 presente no subgrupo de 26 TCNs. A família UPS apresentou apenas um membro (UPS2) que apresenta genes em ambos os

subgrupos, sendo o Glyma.02G116400 e Glyma.02G116300 presentes no subgrupo de 50 TCNs e o Glyma.01G058500 presente no subgrupo de 26 TCNs. A família UMAMIT apresenta cinco membros, todos estão presentes dentro do subgrupo de 50 TCNs. Os membros observados foram o UMAMIT2, UMAMIT3, UMAMIT12, UMAMIT40 e UMAMIT25 (Arquivo adicional 1).

Na figura 1, é possível observar a formação de cinco clusters obtidos a partir da correlação de Pearson entre as contagens em TPM dos 50 TCNs provenientes dos três datasets (Tabela 1). O cluster 1 foi formado pelo perfil de expressão observado em sementes maduras de quatro cultivares de soja (BRSMG 75; Williams 82; BRS 413 RR; DM 6563). O cluster 2 agrupou as sementes de soja com pesos frescos de 100-200 mg e 400-500 mg (C - cotilédone sem tegumento da cultivar W82). O Cluster 3 é composto por sementes inteiras – SI e cotilédone com peso fresco de 5-6 mg da cultivar W82. O Cluster 4 é composto por legumes de soja sem sementes - SS no estágio R4 da cultivar BR 16. Por fim, o Cluster 5 reuniu as sementes em desenvolvimento (quatro dias após o florescimento - DAF, 12-14 DAF e 22-24 DAF), agrupadas com tecido de tegumento com peso de 5-6 mg. Estes dados mostraram que o tecido de legume (cluster 4) apresenta perfil de expressão semelhante a sementes em estádios iniciais de desenvolvimento (cluster 5). Ainda nesta figura, é possível observar que as SI maduras das 4 cultivares de bibliotecas independentes de RNA-Seq se agruparam, apresentando perfil de contagem semelhantes para os 50 TCNs (Figura 1).

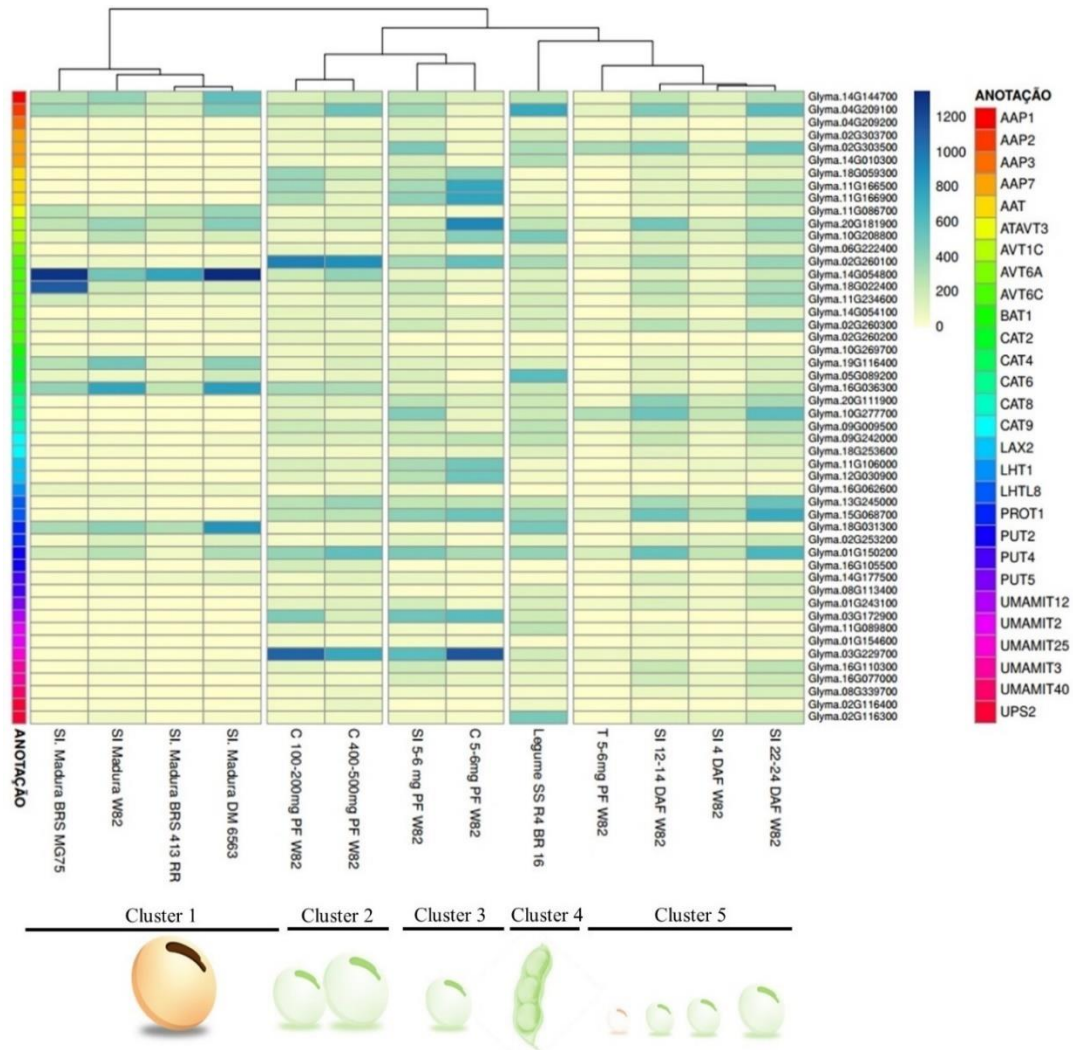


Figura 1. Heatmap ilustrando os 50 genes transportadores de compostos nitrogenados-TCNs com altos níveis de expressão (≥ 40 TPM) nas fases de acúmulo proteico de 100-200 mg e 400 mg. A clusterização das bibliotecas de RNA-Seq dos três datasets gerou cinco clusters. **Cluster 1:** Formado pelas SI (Sementes Inteiras) maduras das cultivares de soja BRS MG75, Willians 82-W82, BRS 413 RR e DM 6563. **Cluster 2:** Formado pelos estágios de desenvolvimento da semente sem o tegumento, contendo apenas o cotilédone-C, com peso entre 100-200 mg e 400-500 mg em W82. **Cluster 3:** Formado por Semente inteira - SI com peso fresco de 5-6 mg em W82. **Cluster 4:** Formado por legumes sem semente - SS no estágio R4 de BR16. **Cluster 5:** Formado por sementes inteiras-SI com 4, 12-14 e 22-24 dias após o florescimento-DAF e pelo tegumento de sementes com 5-6 mg de peso fresco em W82.

A partir do Cluster 2, momento de maior síntese de proteínas de reserva (Jones; Gonzalez; Vodkin, 2010; Jones; Vodkin, 2013), foram selecionados sete genes para validação via RT-qPCR (AAP7, AVT3, BAT1, CAT9, LHT1, UMAMIT25, UPS2). A seleção dos genes foi baseada em contagens ≥ 40 TPM (Pomaznoy *et al.*, 2019) nas fases de 100-200 mg e 400-500 mg. Os clusters (1, 3, 4 e 5) auxiliam na observação do comportamento destes genes em outros tecidos e estádios de desenvolvimento da semente (Figura 1, Figura 2).

Comportamento dos genes selecionados para validação via RT-qPCR (AAP7, AVT3, BAT1, CAT9, LHT1, UMAMIT25, UPS2) por cluster

O perfil de expressão obtido em cada cluster para os sete genes selecionados é apresentado na Figura 2, e pode ser observado quais os que apresentaram contagem acima e abaixo de 40 TPM. No cluster 1 (composto por sementes maduras inteiras de quatro cultivares de soja), os genes AVT3, BAT1 e LHT1 possuem contagem acima de 40 TPM, e os genes AAP7, CAT9, UMAMIT25 e UPS2, possuem contagem abaixo de 40 TPM. Já no cluster 2 (composto por cotilédones com pesos de 100-200 mg e 400-500 mg), todos os genes apresentam contagens acima de 40 TPM (Figura 2B). No cluster 3 (cotilédone e semente inteira com 5-6 mg), acima de 40 TPM para cotilédone estão os genes AAP7, CAT9 e UMAMIT25, e abaixo de 40 TPM permaneceram os genes AVT3, BAT1, LHT1 e UPS2, enquanto que para SI com 5-6 mg, os sete genes apresentaram contagens acima de 40 TPM (Figura 2C).

No cluster 4 (legume em R4) apenas o gene LHT1 encontra-se abaixo de 40 TPM (Figura 2D). E por fim, no cluster 5 (sementes inteiras em estágios iniciais de desenvolvimento e tegumento), acima de 40 TPMs estão os genes AAP7, AVT3, BAT1, CAT9 e LHT1. Abaixo de 40 TPM estão os genes UMAMIT25 e UPS2.

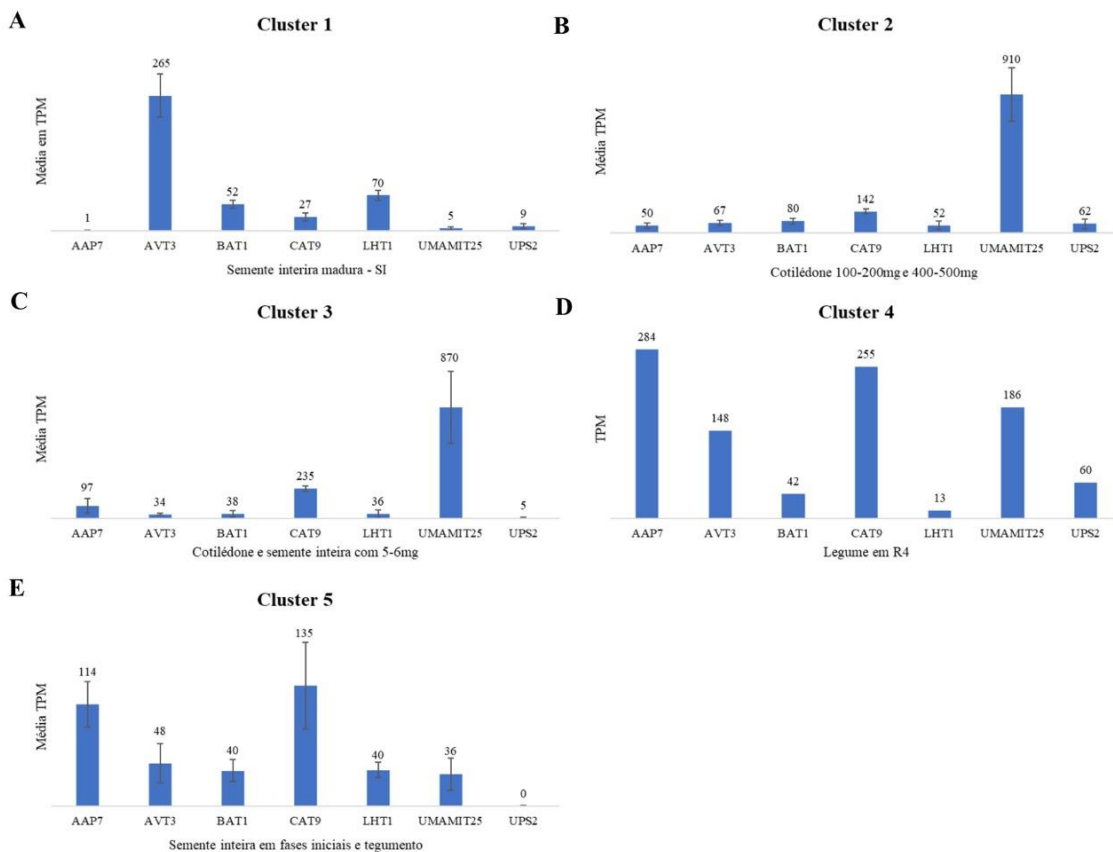


Figura 2. Gráficos ilustrando as médias das contagens em transcritos por milhão (TPM) dos sete genes selecionados para validação via RT-qPCR por cluster. Os valores representam a média de cada gene em relação a todas as bibliotecas clusterizadas juntas. A barra representa o desvio padrão amostral.

Os genes AAP7 e CAT9 apresentam alta contagem em cotilédone de 100-200 mg e 400-500 mg, cotilédone e semente inteira com 5-6 mg, legume e sementes em fases iniciais de desenvolvimento e tegumento.

Apenas em semente inteira madura esse gene apresentou contagem abaixo de 40 TPM. Esses dados mostram sua importância em quatro clusters avaliados (cluster 2, 3, 4 e 5). Os genes AVT3 e BAT1 são de alta contagem em semente inteira madura, cotilédone de 100-200 mg e 400-500 mg, em legume em R4, sementes inteiras em fases iniciais e tegumento. Apresentaram contagem abaixo de 40 TPM nas fases de cotilédone e semente inteira com 5-6 mg, esses dados mostram a importância desses genes em quatro clusters avaliados (cluster 1, 2, 4 e 5). O gene LHT1 é importante na semente inteira madura, cotilédone 100-200 mg e 400-500 mg, semente inteira em fases iniciais e tegumento, mostrando sua importância nos clusters 1, 2 e 5. O gene UMAMIT25 é expresso em cotilédone com 100-200 mg e 400-500 mg, cotilédones e semente inteira com 5-6 mg, legume R4, mostrando sua importância no cluster 2, 3 e 4. O gene UPS2 é expresso no cotilédone 100-200mg e 400-500mg e no legume em R4, ressaltando sua relevância nos clusters 2 e 4 (Figura 2).

Identificação de ortólogos em *Arabidopsis thaliana*, cópias e transcritos alternativos em soja dos 7 TCNs selecionados

Na figura 3 é possível observar que as seqüências proteicas dos genes UMAMIT25 e CAT9 se agruparam em clados próximos. Um clado maior foi formado a partir dos genes AVT3, AAP7 e LHT1 e dentre eles os genes AAP7 e LHT1 estão mais próximos. O gene BAT1 e UPS2 se agruparam em clados individuais.

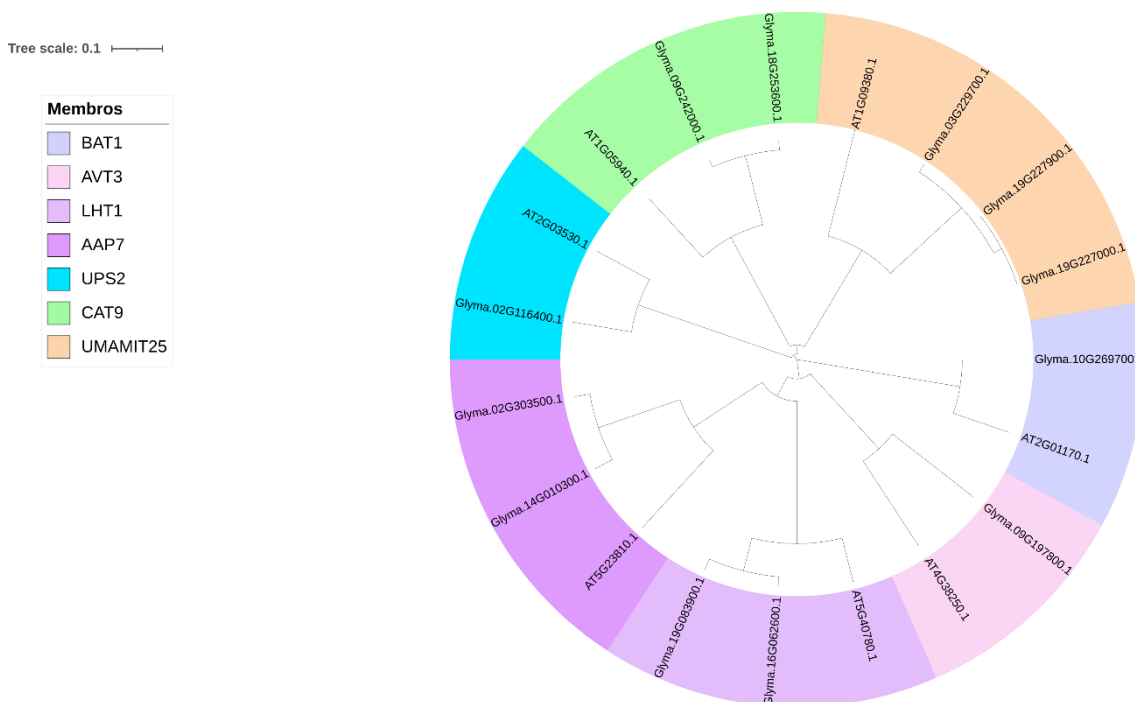


Figura 3. Dendrograma dos genes AAP7, AVT3, BAT1, CAT9, LHT1, UMAMIT25 e UPS2 a partir das seqüências proteicas. Na imagem está ilustrado a clusterização de cada gene com suas respectivas cópias em soja e ortólogos em *Arabidopsis thaliana*. Legenda: Roxo escuro: AAP7; Laranja: UMAMIT 25; Azul: BAT1; Verde claro: CAT9; Azul turquesa: UPS2; Rosa: AVT3A; Lilás: LHT1.

A fim de observar características dos sete genes selecionados, foi realizada uma busca pelo número de proteínas, cópias e transcritos alternativos. Na tabela 3A é possível observar que o gene AAP7 (Glyma.14G010300) apresenta 13 transcritos alternativos, que codificam para uma seqüência proteica de 378

aminoácidos (aa). Embora o tamanho da proteína seja igual para todos os transcritos, estes variam em tamanho (2361-3227 pares de bases - pb). O gene AAP7 apresenta como cópia o gene Glyma.02G303500 com 99% identidade (sequência proteica), que contém cinco transcritos alternativos. Entre eles, o transcrito 4 (360 aa) apresenta uma proteína menor que a dos outros transcritos (461 aa). Para este gene os transcritos variam de 1370 pb a 2495 pb. Este gene tem como ortólogo em *Arabidopsis* o gene AT5G23810 com 62% identidade. Em *Arabidopsis* esse gene é cópia única e apresenta dois transcritos alternativos com 1195 pb e 1573 pb. Estes codificam para duas proteínas de tamanhos diferentes, uma de 361 aa e a outra de 467 aa. Em soja esse gene sofreu uma duplicação e vários transcritos alternativos são codificados. O padrão de transcritos primários de ambas as espécies apresenta sete exons e todas as sequências proteicas apresentam o domínio conservado referente a família PF01490 -transportados de aminoácidos transmembrana.

Observa-se que o gene CAT9 (Glyma.09G242000) apresenta apenas um transcrito com 2173 pb, que codifica para uma proteína de 558 aa. Há uma segunda cópia no genoma de soja com 94% de identidade (Glyma.18G253600) que apresenta seis transcritos alternativos, que variam em tamanho (2202 pb a 2506 pb), assim como suas respectivas proteínas, que variam de 427 aa a 578 aa. Como ortólogo em *Arabidopsis thaliana* apresenta o gene AT1G05940 com 72% de identidade, que também apresenta apenas um transcrito (1955 pb) e uma proteína (569 aa). O padrão de exons gira em torno de oito entre as espécies. Além disso, este gene apresentou variação na presença de domínios proteicos. Todos os três genes apresentam o domínio referente as famílias de aminoácido permeases PF13520 e PF13906 em suas formas primarias (Tabela 3B).

Em relação ao gene AVT3 putativo em soja (Glyma.09G197800), observa-se que é um gene de cópia única, que apresenta apenas um transcrito de 1287 pb e uma proteína de 428 aa. Como ortólogo em *Arabidopsis thaliana* tem o gene AT4G38250 com 62% de identidade. Em *Arabidopsis* também é um gene de cópia única. Este, apresenta um transcrito com 1677pb e uma proteína de 436aa. Ambas as espécies apresentam o padrão de um exon e contém em suas proteínas o domínio conservado referente a família PF01490 de transportados de aminoácidos transmembrana (Tabela 3C).

O gene BAT1 (Glyma.10G269700) apresenta quatro transcritos alternativos, que variam entre 1601 pb a 1951 pb, com proteínas de 309 aa a 520 aa. Apresenta como ortólogo em *Arabidopsis* AT2G01170 com 79% de identidade e ambos apresentam cópia única. Em *Arabidopsis* apresenta dois transcritos alternativos com tamanho e proteínas de tamanhos diferentes. Ambas as espécies têm o padrão de oito exons para seus transcritos primários e apresentam o domínio conservado da família amino permeases PF13520 em suas proteínas (Tabela 3D).

Na tabela 3E, observamos que UPS2 também é cópia única nas duas espécies. Em soja, o gene UPS2 (Glyma.02G116400) apresenta três transcritos com tamanhos entre 1228 pb e 1639 pb e três proteínas com tamanho entre 226 aa a 395 aa. Como ortólogo em *Arabidopsis* está o gene AT2G03530 com 77% de identidade, que apresenta dois transcritos alternativos, porém que codificam para uma proteína de mesmo tamanho (398 aa). Em soja o transcrito primário apresenta sete exons e em *Arabidopsis* seis exons. Ambos apresentam o domínio proteico referente a família PF07168- Ureídeos permeases.

Tabela 3. Características dos sete genes selecionados. Legenda: T=Tamanho; N= Número.

A. AAP7-SOJA - ARABIDOPSIS (62% de identidade)					
GENE ID	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
Glyma.14G010300.1	3227	7	2	378aa	PF01490

Glyma.14G010300.2	2585	7	2	378aa	PF01490
Glyma.14G010300.3	2594	7	2	378aa	PF01490
Glyma.14G010300.4	2430	7	2	378aa	PF01490
Glyma.14G010300.5	2419	7	2	378aa	PF01490
Glyma.14G010300.6	2361	7	2	378aa	PF01490
Glyma.14G010300.7	2370	7	2	378aa	PF01490
Glyma.14G010300.8	2523	7	2	378aa	PF01490
Glyma.14G010300.9	2532	7	2	378aa	PF01490
Glyma.14G010300.10	2553	7	2	378aa	PF01490
Glyma.14G010300.11	2544	7	2	378aa	PF01490
Glyma.14G010300.12	3076	7	2	378aa	PF01490
Glyma.14G010300.13	2495	7	2	378aa	PF01490
COPIA (99%)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
Glyma.02G303500.1	2495	7	1	461aa	PF01490
Glyma.02G303500.2	1942	7	1	461aa	PF01490
Glyma.02G303500.3	1841	5	1	461aa	PF01490
Glyma.02G303500.4	1370	7	1	360aa	PF01490
Glyma.02G303500.5	1921	7	1	461aa	PF01490
ORTÓLOGO (Cópia Única)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
AT5G23810.1	1573	7	1	467aa	PF01490
AT5G23810.2	1194	6	1	361aa	PF01490

B. B. CAT9-SOJA - ARABIDOPSIS (72%)

GENE ID	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
Glyma.09G242000.1	2173	8	2	558aa	PF13520/PF13906
COPIA (94%)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
Glyma.18G253600.1	2379	9	1	578aa	PF13520/PF13906
Glyma.18G253600.2	2202	8	1	519aa	PF13520/PF13906
Glyma.18G253600.3	2325	8	1	560aa	PF13520
Glyma.18G253600.4	2506	7	1	533aa	PF13520
Glyma.18G253600.5	2329	7	1	427aa	PF13520/PF13906
Glyma.18G253600.6	2385	8	1	445aa	PF13520/PF13906
ORTÓLOGO (UNICO)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
AT1G05940.1	1955	8	2	569aa	PF13520/PF13906

C. AVT3-SOJA - ARABIDOPSIS (62%)

GENE ID (Cópia Única)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
Glyma.09G197800.1	1287	1	2	428aa	PF01490
ORTÓLOGO (Cópia Única)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
AT4G38250.1	1667	1	1	436aa	PF01490

D. BAT1-SOJA - ARABIDOPSIS (79%)

GENE ID (Cópia Única)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
------------------------------	-----------------	-----------------	-------------	----------------	-------------

Glyma.10G269700.1	1951	8	2	520aa	PF13520
Glyma.10G269700.2	1932	6	2	398aa	PF13520
Glyma.10G269700.3	1894	8	2	501aa	PF13520
Glyma.10G269700.4	1601	4	2	309aa	PF13520
ORTÓLOGO (Cópia Única)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
AT2G01170.1	1808	8	2	516aa	PF13520
AT2G01170.2	1825	7	2	437aa	PF13520

E. UPS2-SOJA - ARABIDOPSIS (77%)

GENE ID (Cópia Única)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
Glyma.02G116400.1	1639	7	1	395aa	PF07168
Glyma.02G116400.2	1425	7	1	321aa	PF07168
Glyma.02G116400.3	1228	5	1	226aa	PF07168
ORTÓLOGO (Cópia Única)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
AT2G03530.1	1368	6	1	398aa	PF07168
AT2G03530.2	1409	6	1	398aa	PF07168

F. LHT1-SOJA - ARABIDOPSIS (83%)

GENE ID	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
Glyma.16G062600.1	1999	8	2	450aa	PF01490
COPIA	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
Glyma.19G083900.1 (96%)	1865	8	1	450aa	PF01490
Glyma.19G083900.2	1620	8	1	436aa	PF01490
Glyma.19G083900.3	1703	7	1	398aa	PF01490
ORTÓLOGO (Cópia Única)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
AT5G40780.1	1712	8	1	446aa	PF01490
AT5G40780.2	1678	8	1	445aa	PF01490

G. UMMAMIT25-SOJA - ARABIDOPSIS (63%)

GENE ID	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
Glyma.03G229700.1	1518	7	1	366aa	PF00892/PF00892
Glyma.03G229700.2	1327	6	1	297aa	PF00892/PF00892
COPIA	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
Glyma.19G227900.1 (96%)	1563	7	2	367aa	PF00892/PF00892
Glyma.19G227900.2 (96%)	1396	6	2	328aa	PF00892/PF00892
Glyma.19G227000.1 (96%)	1532	7	1	367aa	PF00892/PF00892
ORTÓLOGO (Cópia Única)	T. TRANS	N. EXONS	FITA	T. PROT	PFAM
AT1G09380.1	1568	7	2	374aa	PF00892/PF00892

Na Tabela 3F, é possível observar que o gene LHT1 (Glyma.16G062600) apresenta duas cópias com 96% de identidade e não possui transcritos alternativos. Por outro lado, sua cópia (Glyma.19G083900) apresenta três transcritos alternativos que variam de 1620 pb a 1865 pb. Com proteínas que também variam em tamanho (398 aa a 450 aa). Tem como ortólogo em Arabidopsis o gene AT5G40780 com 83% de identidade, que apresenta cópia

única. Em *Arabidopsis* apresenta dois transcritos alternativos que codificam para uma proteína que se diferencia em apenas um aminoácido (446 aa e 445 aa). Em seus transcritos primários apresentam um padrão de oito exons em ambas as espécies. Ainda, ambos apresentam domínio proteico conservado referente a família PF01490-transportados de aminoácidos transmembrana.

Por fim, o gene *UMAMIT25* (Glyma.03G229700) apresenta dois transcritos alternativos de tamanhos diferentes que correspondem a duas proteínas também com tamanhos variados. Tem como cópia o gene Glyma.19G227900 e o gene Glyma.19G227000, ambos com 96% de identidade. Como ortólogo em *Arabidopsis* está o gene AT1G09380. Este não contém transcritos alternativos, ambas as espécies apresentam em seus transcritos primários um padrão de sete exons. Estes genes apresentam dois domínios conservados da mesma família (PF00892), que representa uma família de transportador de membrana.

Em síntese, foi possível observar a partir da Tabela 3 que os ortólogos dos sete genes avaliados não apresentam cópias em *Arabidopsis*. Em soja, apenas os genes *AVT3A*, *BAT1* e *UPS2* não apresentam cópias. Enquanto, os genes *AAP7*, *CAT9* e *LHT1* apresentam duas cópias e o gene *UMAMIT25* apresenta três cópias no genoma.

Ainda, esses dados mostram que os únicos genes que tiveram suas cópias também altamente expressas (≥ 40 TPM) nas fases de 100-200 mg e 400-500 mg, foram os genes *AAP7* (Glyma.02G303500) e *CAT9* (Glyma.18G253600). Em legume os genes *AAP7*, *AVT3*, *BAT1*, *CAT9*, *UMAMIT25* e *UPS2* apresentaram alta expressão, já o gene *LHT1* apresentou contagens abaixo de 40 TPM. As cópias de *AAP* e *CAT9* também apresentaram altas contagens em legume (Arquivo adicional 1, Tabela 3).

Na Figura 4, está representada a posição dos genes nos cromossomos.

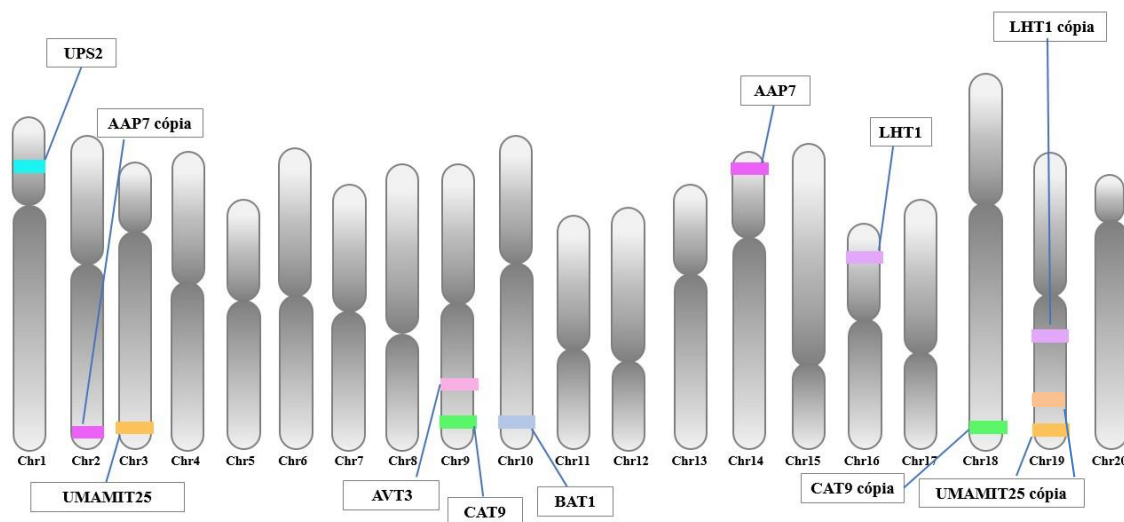


Figura 4 – Posicionamento de genes em seus respectivos cromossomos na soja. Legenda: Roxo escuro: *AAP7* (Glyma.14G010300) e cópia (Glyma.02G303500); Laranja: *UMAMIT 25* (Glyma.03G229700) e cópias (Glyma.19G229700 e Glyma19G229000); Azul: *BAT1* (Glyma.10G269700); Verde claro: *CAT9* (Glyma.09G242000) e cópia (Glyma.18G253600); Azul turquesa: *UPS2* (Glyma.02G116400); Rosa: *AVT3A* (Glyma.09G197800); Lilás: *LHT1* (Glyma.16G062600) e cópia (Glyma.19G083900).

Análises fisiológicas para Expressão diferencial via RT-qPCR

O peso seco e fresco das vagens e legumes de três estádios fenológicos (R5.1/2; R5.3; R6) foi avaliado em ambas as cultivares. Os pesos não apresentam variação significativa entre as cultivares, indicando que o experimento está padronizado e influências externas não irão interferir nos níveis de expressão gênica obtidos a partir dos tecidos avaliados, considerando que o experimento foi realizado sob condições controladas. O peso fresco das sementes para ambas as cultivares foi de 56.5 ± 3.22 mg no estádio R5.1/2, 180 ± 8.74 mg no estádio R5.3 e 385 ± 11.48 mg no estádio R6. O peso seco das sementes para ambas as cultivares foi de 8.5 ± 0.78 mg no estádio R5.1/2, 38.5 ± 2.19 mg no estádio R5.3 e 134 ± 3.40 mg no estádio R6.

Em legumes o peso fresco para ambas as cultivares foi de 556 ± 11.97 mg no estádio R5.1/2, 473 ± 11.64 mg no estádio R5.3 e 482 ± 16.53 mg no estádio R6. O peso seco dos legumes para ambas as cultivares foi de 101 ± 5.23 mg no estádio R5.1/2, 90 ± 3.14 mg no estádio R5.3 e 158 ± 7.38 mg no estádio R6.

Houve diferença estatística entre o teor de proteína dos grãos maduros entre as cultivares. O teor proteico foi de $38,98 \pm 0,46\%$ para a cultivar BRS 232 e $36,01 \pm 0,93$ mg para BRS 284. O peso de 100 sementes maduras foi de $20,21 \pm 0,73$ mg para BRS 232 e $15,15 \pm 0,71$ para BRS 284. Em conclusão os dados fisiológicos mostraram que o experimento em casa de vegetação estava padronizado. A partir dessas informações, foi realizado o RT-qPCR .

Na figura 5A, é possível observar que o gene AAP7 está *up*-regulado nas três fases de desenvolvimento da semente da cultivar de maior teor proteico (BRS 232). Por outro lado, em legume esses genes encontram-se *down*-regulado na cultivar BRS 232. O gene AVT3 apresentou *up*-regulação nas três fases de desenvolvimento em sementes e na fase de 25-100mg de legumes. Nas fases de 100-200 mg e nas fases de 350-450 mg, esse gene apresentou *down*-regulação em legume na cultivar BRS 232 (Figura 5B). O gene BAT1 apresentou *down*-regulação em ambos os tecidos da cultivar BRS 232 (Figura 5C). Na figura 5D, o gene CAT9 apresentou *up*-regulação em todos os níveis de desenvolvimento da semente, assim como nas fases de 25-100mg e 100-200mg de legumes. Somente na fase de 350-450mg em legume, esse gene foi *down*-regulado na cultivar BRS 232.

O gene LHT1 apresentou *up*-regulação apenas em legume na fase de 100-200mg. Em todas as outras fases no tecido de legumes e sementes encontram-se *down*-regulado na cultivar BRS 232 (Figura 5E). Na figura 5F, é possível observar que o gene UMAMIT25 encontra-se *up*-regulado em todas as fases de desenvolvimento de sementes da cultivar BRS 232, bem como em legume nas fases de 100-200mg e 350-450mg. Apresentou *down*-regulação apenas em legume na fase de 25-100mg. Na figura 5G, é possível observar que o gene UPS2 encontra-se *up*-regulado na cultivar BRS 232 apenas em semente na fase de 10-200mg. Em todas as outras fases de sementes e legumes apresentou-se *down*-regulado.

Em resumo, é possível observar que embora todos os genes avaliados estejam de fato presentes nesses tecidos como observado *in silico*, apenas os genes AAP7, AVT3, CAT9 e UMAMIT25 apresentam-se *up*-regulados em todas fases de desenvolvimento da cultivar de maior teor proteico. O gene UPS2, também apresentou uma correlação positiva com o teor de proteína na fase de 100-200mg. Embora os genes BAT1 e LHT1 estejam *down*-regulados na BRS 232, são genes importantes para cultivar BRS 284 nestes tecidos. Porém não apresentam correlação positiva com o teor proteico.

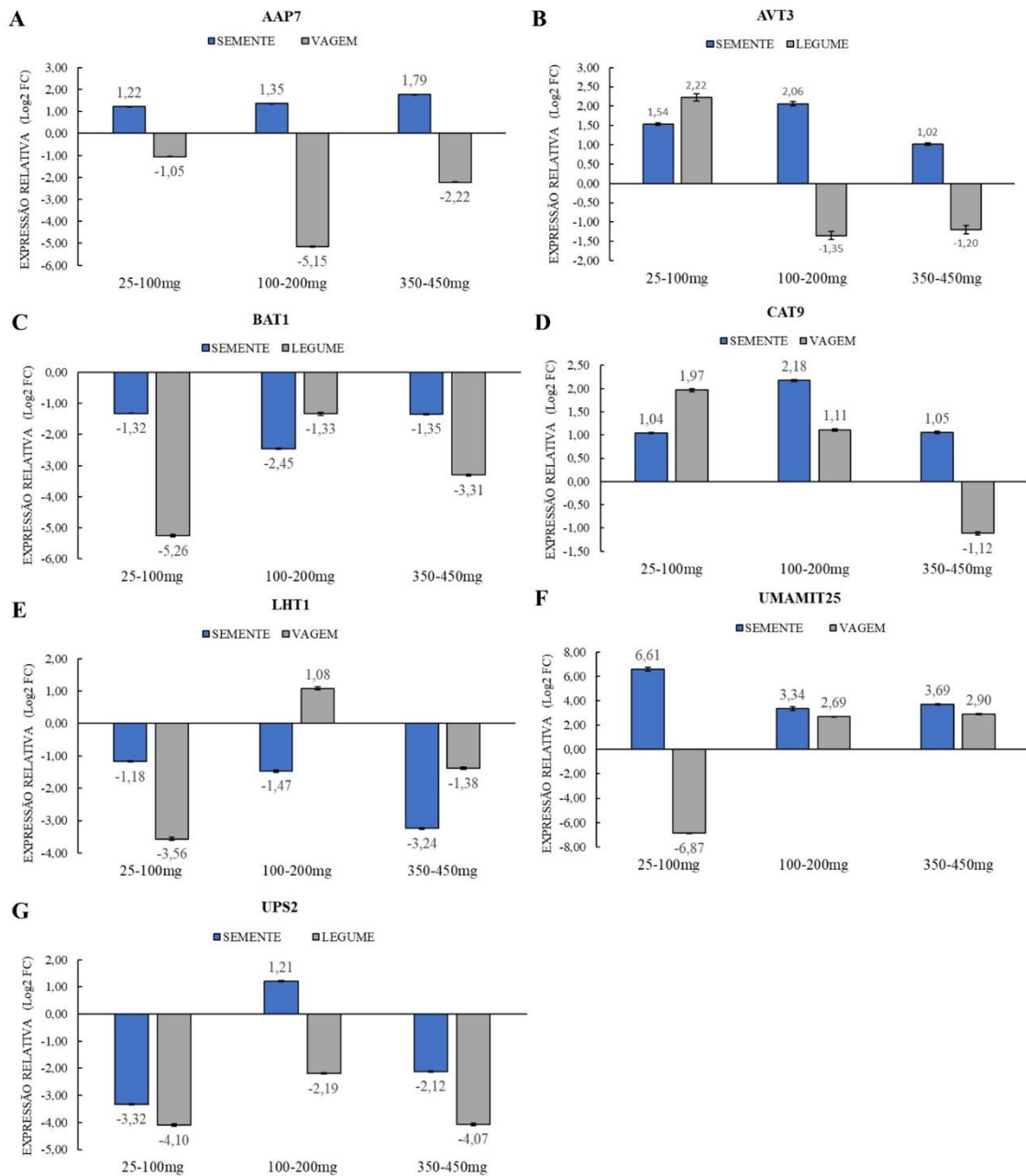


Figura 5 – Expressão relativa dos genes AAP7, AVT3, BAT1, CAT9, LHT1, UMAMIT25 e UPS2 em sementes e legumes nos estádios de desenvolvimento 25-100mg, 100-200mg e 400-500mg. Os níveis de expressão *up* e *down*-regulados representam os valores observados na cultivar BRS 232 de maior teor proteico em relação a cultivar de calibração BRS 284 de menor teor proteico. As barras representam o erro padrão de três repetições biológicas.

Discussão

Transportadores de compostos nitrogenados-TCNs presentes em sementes, tegumento e legume

A partir dos dados da figura 1 (Heatmap ilustrando os 50 genes transportadores de compostos nitrogenados-TCNs com altos níveis de expressão (≥ 40 TPM) nas fases de acúmulo proteico de 100-200 mg e 400 mg, foi possível observar que alguns dos tecidos da planta apresentam uma maior similaridade de expressão dos TCNs avaliados. Por este motivo, se agrupam dentro de um mesmo cluster ou de clusters muito próximos. No Cluster 1, independente da cultivar (BRS MG75, W82, BRS 413RR, DM 6563) é possível observar que a expressão dos TCNs é semelhante. As fases na qual a semente apresenta uma maior deposição de proteínas reserva (C 100-200mg, C 400-500mg) também se agrupou, indicando que as famílias de TCNs presentes nessas fases, apresentam expressão similar, algumas com correlação direta com o teor proteico e outras que podem estar atuando de maneira indireta nessa variável. A proximidade dos clusters 3, 4 e 5 indicam que sementes na fase de síntese de proteína metabólica, tegumento e legume apresentam expressão semelhante em relação aos TCNs avaliados. O que indica a participação desses outros tecidos, na síntese de proteína de desenvolvimento. Isso ocorre porque o desenvolvimento das sementes e armazenamento de proteínas de reserva dependem do tegumento e do legume (Herman 2014). Esses tecidos, tem papel crucial na captação de aminoácidos e ureídeos das folhas e raízes, antes da deposição nas sementes. A eficiência da atividade desses tecidos também é essencial para o desenvolvimento e qualidade da semente madura (Castro; Bradford; Hilhorst, 2004).

O tegumento além de proteção física é fundamental para a distribuição de nutrientes para o embrião (Abirached - Darmency *et al.*, 2005). Ele possui quatro camadas internas (cutícula encerada, epiderme, hipoderme e o parênquima interior) e quatro externas (hilo, micrópila, rafe e região calazal) (Herman 2014). Sendo a hipoderme composta de grandes células vacuoladas, onde podem ocorrer o metabolismo e a interconversão das amidas em outros aminoácidos, antes da liberação no apoplasto (Herman 2014). A absorção de assimilados do tegumento pelo embrião é também mediada por transportadores de membrana por um processo de transporte ativo, o que corrobora com a grande quantidade destes transportadores observados *in silico* no presente estudo (Tegeder e Rentsch, 2010).

Dentre os TCNs avaliados *in silico*, genes da família AAP (AAP1, AAP2, AAP6, AAP7, AAP8) demonstraram atividade nas sementes maduras (proteína de proteção) e em desenvolvimento (proteínas metabólicas e de reserva) (Jones, Gonzales e Vodkin, 2010; Jones e Vodkin, 2013). No geral, essa família não apresentou importância no tegumento, mas desempenha papel no legume. A família AAP transporta aminoácidos provenientes de outras partes da planta, via floema, para nutrição da semente, mesmo em condição de privação de N no solo, garantindo a produção de sementes através da remobilização do N (Müller; Karmann, Hammes, 2018; Zhou *et al.*, 2020; Liu *et al.*, 2020). Por outro lado, a superfamília de transportadores de aminoácidos (AAT) identificada, embora apresentou expressão significativa em sementes em desenvolvimento e legumes, apresenta características genéricas que não puderam ser identificadas mais a fundo, pois suas funções específicas ainda não foram caracterizadas na literatura (Cheng *et al.*, 2016) (Arquivo adicional 1).

A família AVT, tem papel importante na semente em desenvolvimento, madura e legume, porém em tegumentos não apresentou participação ativa, como observado *in silico*. Esse gene importa e exporta aminoácidos armazenados nos vacúolos, e essa atividade pode estar relacionada a liberação de aminoácidos essenciais e não essenciais, mantendo a homeostase desses tecidos, para que se desenvolvam e permitam um ambiente favorável

para acúmulo de proteínas, posteriormente contabilizadas no teor proteico final do grão (Fujiki *et al.*, 2017). A família CAT, desempenha um papel semelhante a família AVT, porém atua em todas as fases de desenvolvimento avaliadas (Arquivo adicional 1).

As famílias UMAMIT (UMAMIT2, UMAMIT3, UMAMIT12, UMAMIT25, UMAMIT40), LAX (LAX1, LAX2) e LHT (LHT1 e LHT8), apresentaram atividade nas sementes em desenvolvimento, em tegumento e legume. Todas essas famílias transportam aminoácidos para sementes e alguns deles (LHT e UMAMIT) aumentam significativamente o rendimento de sementes como observado em arroz e *Arabidopsis* (Besnard *et al.*, 2018; Wang *et al.*, 2019). A família LAX é uma possível transportadora de triptofano, um aminoácido essencial, que desempenha papel fundamental na qualidade do teor proteico final do grão (Swarup; Péret, 2012). A família UPS (UPS2), também atua nas sementes em desenvolvimento e legumes, porém, não está significativamente ativa em tegumento. Observar que estes transportadores estão presentes nos legumes confirma as teorias de que este tecido viabiliza a assimilação de aminoácidos e ureídeos para sementes (Benett; Roberts; Wagstaff, 2011; Carter e Tegeder, 2016). A família transportadora bidirecional BAT, tanto importa quanto exporta aminoácidos de diversos tecidos (Dundar e Bush, 2009). No presente estudo foi observado que embora a família BAT esteja presente nos tecidos de sementes em desenvolvimento, madura, tegumento e legumes, ele parece apresentar um papel indireto com o teor de proteína como observado nos dados de RT-qPCR (Arquivo adicional 1; figura 5C).

As famílias de transportador de aminoácido PUT (PUT2, PUT4 e PUT5) e ProT (ProT1), estão presentes nas sementes maduras e em desenvolvimento, tegumento e legume. A família ProT, é conhecida como transportadora de prolina e possível transportador de aminoácidos, envolvida no transporte a longas distâncias (Grallath *et al.*, 2005). A família de genes PUT, são transportadores de poliamina e possíveis transportadores de aminoácidos. As poliaminas, são pequenas amins alifáticas, sendo a mais conhecida putrescina, que são essenciais para o metabolismo de N e biossíntese de aminoácidos (Minguet *et al.* 2008, Silveira *et al.* 2013, Minocha *et al.* 2014, Majumdar *et al.* 2016; Page *et al.* 2016; Masson *et al.* 2017). Essas informações podem indicar seus papéis nestes tecidos como facilitadores das interconversões do N em outras formas ativas, o que auxilia o transporte de aminoácidos por outros transportadores (Arquivo adicional 1).

Comparação entre os TCNs em soja e Arabidopsis thaliana

Dentre os TCNs avaliados através de RT-qPCR o AAP7, CAT9, LHT1 apresentam duas cópias, o UMAMIT25 apresenta três cópias e AVT3, BAT1 e UPS2 apresentam cópia única. Na espécie modelo para dicotiledôneas (*Arabidopsis thaliana*) todos esses genes estão em cópias únicas. Durante a evolução essas espécies se diferenciaram, uma duplicando seu genoma (soja – 75% do genoma duplicado) e outra não. Ambas apresentam um clado comum denominado rosídea que após o processo evolutivo, se subdividiu em dois subclados, as fabídeas (soja) e as mavídeas (*Arabidopsis*) (Huang *et al.*, 2019; Huang *et al.*, 2020). Embora elas tenham sofrido essa subdivisão, muitos trabalhos na literatura mostram que ortólogos entre as espécies, desempenham em sua grande maioria a mesma função (Fuganti-Pagliarini *et al.* 2020). Como por exemplo, genes DREB1A, que em soja e em *Arabidopsis* desempenham o papel de proteção (Polizel *et al.*, 2011; Rolla *et al.*, 2013). Outros trabalhos com os genes DREB2A (Engels *et al.*, 2013), AREB1 (Barbosa *et al.*, 2012), GolS2 (Honna *et al.*, 2016) e NCED3 (Molinari *et al.*, 2020), também atuam da mesma maneira tanto em soja como em *Arabidopsis*.

Ainda que os ortólogos desempenham na maioria das vezes a mesma função, o processo de duplicação pode gerar a subfuncionalização desses genes (Schmutz *et al.*, 2010; Jung *et al.*, 2012). Segundo Takeshima *et al.*,

2019, ao estudar o gene de florescimento FT (flowering Locus T) em *Arabidopsis* e soja, observou que em *Arabidopsis* é um gene de cópia única e em soja apresenta duas cópias (FT2a e FT5a). Embora estes ortólogos desempenhassem a mesma função como indutor do florescimento, em soja foi observado que o FT5a era mais eficiente que o FT2a.

Estes trabalhos citados indicam que existe uma alta probabilidade dos TCNs avaliados no presente estudo (AAP7, AVT3, LHT1 e UMAMIT25), terem a mesma função observada em *Arabidopsis*. Principalmente pelo fato da porcentagem de identidade da sequência proteica das espécies, seja acima de 60% (Besnard *et al.*, 2018; Zhou *et al.*, 2020) (Arquivo adicional 1). Contudo no processo evolutivo o acúmulo de mutações nas cópias, podem também torná-las não funcionais (pseudogenização) e ainda assim, manter a atividade do gene, devido múltiplas cópias. Este fato corrobora com a grande variabilidade genética da soja e sua capacidade adaptativa (Schmuts *et al.*, 2010).

A duplicação genica pode acontecer em tandem ou de forma segmentada, ou seja, em locais diferentes do genoma (Ramsey; Schemske, 1998). Isso também pode ser um fator que influencia na atividade das cópias, como observamos nos genes de florescimento FT5a e FT2a (Schmuts *et al.*, 2010). Dependendo do local do cromossomo em que esse gene está localizado (eucromatina e heterocromatina), pode haver mais ou menos expressão, de uma ou outra cópia. No presente estudo foi observado que para o AAP7 apresentou cópia no cromossomo 2 e 14, CAT9 apresentou cópia nos cromossomos 9 e 15 e LHT1 nos cromossomos 16 e 19. Já para o gene UMAMIT, além de cópias segmentais (cromossomo 3 e 19), foi observado cópia em tandem no cromossomo 19 (Arquivo adicional 1).

TCNs envolvidos na síntese de proteínas metabólicas, de reserva e proteção da semente

Os genes AAP7 e UMAMIT25 são transportadores transmembrana, enquanto o AVT3 e CAT9, são transportadores vacuolares (Yang; Stierhof; Ludewig, 2015; Besnard *et al.* 2018; Sekito *et al.*, 2014; Zhou *et al.*, 2020). Todos eles transportam aminoácidos essenciais e não-essenciais, que possuem importância na síntese de todos os tipos de proteínas da soja (metabólica, reserva e proteção) (Jones e Vodkin, 2013; Patil *et al.*, 2017). Na tabela 4, é possível observar que diferentes genes podem transportar o mesmo tipo de aminoácido, o que pode inferir sua atividade em mecanismos semelhantes.

Tabela 4 - Aminoácidos transportado pelos genes AAP7, AVT3, CAT9, UMAMIT25 e sua classificação em essenciais e não essenciais.

Aminoácidos	Transportadores	Tipo
Serina	CAT9 ^a UMAMIT25 ^b AVT3 ^c	Não essencial
Fenilalanina	AAP7 ^d AVT3 ^c	Essencial
Fenilalanina	AAP7 ^d AVT3 ^c	Essencial
Glutamina	CAT9 ^a AVT3 ^c	Não essencial

Glutamina	CAT9 ^a	AVT3 ^c	Não essencial
Asparagina	UMAMIT25 ^b	AVT3 ^c	Não essencial
Asparagina	UMAMIT25 ^b	AVT3 ^c	Não essencial
Triptofano	UMAMIT25 ^b	AVT3 ^c	Essencial
Triptofano	UMAMIT25 ^b	AVT3 ^c	Essencial
Aspartato	AAP7 ^d	CAT9 ^a	Não essencial
Aspartato	AAP7 ^d	CAT9 ^a	Não essencial
Glutamato	CAT9 ^a	UMAMIT25 ^b	Não essencial
Glutamato	CAT9 ^a	UMAMIT25 ^b	Não essencial
Leucina	AAP7 ^d		Essencial
Lisina	AAP7 ^d		Essencial
Cistina	AVT3 ^c		Não essencial
Cistina	AVT3 ^c		Não essencial
Treonina	AVT3 ^c		Essencial
Treonina	AVT3 ^c		Essencial
Alanina	CAT9 ^a		Não essencial
Arginina	UMAMIT25 ^b		Não essencial
Arginina	UMAMIT25 ^b		Não essencial
Glicina	UMAMIT25 ^b		Não essencial
Valina	UMAMIT25 ^b		Não essencial

Legenda: ^a Yang; Stierhof; Ludewig, 2015; ^b Besnard *et al.* 2018; ^c Sekito *et al.*, 2014; ^d Zhou *et al.*, 2020; ^e Patil *et al.*, 2017.

Os quatro TCNs transportam pelo menos um dos aminoácidos essenciais durante a formação de proteínas de soja. Este fato, as torna de grande importância considerando que os aminoácidos essenciais são incapazes de ser sintetizados por humanos e animais, assim ao aumentar o transporte de aminoácidos essenciais nas plantas, podemos tornar o grão de soja mais adequado para indústria (Berry 1962; Wade 1985; Singer *et al.*, 2019). Os aminoácidos não essenciais, também são transportados por ambas as TCNs, além de grande importância para formação de proteínas e desenvolvimento da sementes (Herrmann *et al.*, 1983; Buttery *et al.*, 1994; D'Melo, 2012; Singer *et al.*, 2019).

Expressão diferencial via RT-qPCR

Os genes AAP7, AVT3, CAT9 e UMAMIT25 estão up-regulados na cultivar BRS 232 de maior teor de proteína no grão

Os quatro genes foram *up*-regulados em todas as fases de desenvolvimento das sementes avaliadas (25-100mg, 100-200mg e 400-500mg) na cultivar BRS 232. Essa informação indica que existe uma correlação positiva entre as expressões desses genes e o teor proteico. Visto que todos esses genes são transportadores de compostos nitrogenados, existe um indicativo de que o aumento dos níveis desses genes, podem influenciar diretamente na modulação do teor de proteína nos grãos de soja. O gene AAP7, está incluído em uma família multigênica envolvida em diversos processos biológicos como transporte a longas distâncias, modulação do transporte de aminoácidos em condições de estresse e remobilização de N em forma de amina. O desenvolvimento e o teor de proteína nas sementes, depende em grande parte do acúmulo de N, pois esse é um macro nutriente responsável tanto pelo crescimento da planta quanto pelo desenvolvimento de órgãos reprodutivos. O desenvolvimento das sementes requer um grande número de aminoácidos e a principal fonte é o N. A família AAP7, tem um papel fundamental nesse processo na fase de acúmulo de proteína, pois transporta aminoácidos para semente (Karmann; Müller; Hammes, 2018; Zhou *et al.*, 2020). Embora não tenha na literatura trabalhos de mutação ou

superexpressão deste gene, outro membro da família que apresenta mais de 60% de identidade de sequência proteica (AAP6) com o AAP7, mostrou-se fundamental na realocação de aminoácidos de outros tecidos para as sementes, quando a planta está em um ambiente de baixa concentração de N. A superexpressão desse gene, tanto em soja como em *Arabidopsis* torna a planta mais tolerante ao estresse na falta de N, melhorando a qualidade da semente de soja, atuando na reciclagem de N. Esses dados podem indicar o potencial do gene AAP7 no aumento do teor de proteína, mesmo em condições hostis de cultivo, principalmente pelo fato desse gene estar *up*-regulado na cultivar de maior teor proteico (BRS 232) (Liu *et al.*, 2020). Em legume, este gene apresentou-se *down*-regulado em todas as fases de desenvolvimento avaliadas. Isso indica seu papel em relação ao teor de proteína somente em sementes.

O gene UMAMIT25 está envolvido no transporte de aminoácido do floema para sementes (Besnard *et al.* 2018), o que corrobora com os dados obtidos no presente estudo, uma vez que esse gene se mantém *up*-regulado em todas as fases de desenvolvimento das sementes e também em legumes nas fases de 100-200mg e 400-500mg. Essa família tem papel importante na síntese de proteínas e metabólitos especializados. Em *Arabidopsis* foi observado uma alta expressão desses genes em sementes e legume em desenvolvimento (Schmid *et al.*, 2015), além disso, foi observado que ele medeia a exportação de aminoácidos do endosperma para nutrição do embrião. A superexpressão aumentou a produção de sementes. Esse fato aliado ao presente estudo, em que esse gene está altamente expresso em todos os níveis de desenvolvimento da semente de soja de maior teor proteico (Besnard *et al.*, 2018), é um indicativo de que este gene pode atuar como um agente importante para solucionar o viés da correlação negativa entre proteína e produtividade através da Engenharia Genética.

O CAT9 (transportador de aminoácido catiônico 9), apresentou-se *up*-regulado em todas as fases de desenvolvimento das sementes e em legume nas fases de 25-100mg e 100-200mg. Este gene também está envolvido na homeostase de aminoácidos em plantas. Em *Arabidopsis*, esse gene assim como AAP7, aumenta a reciclagem de aminoácidos entre tecidos da planta em meio de privação de N. A superexpressão desse gene aumenta a tolerância a falta de N no solo, justamente por remobilizar aminoácidos e assim afeta positivamente a biomassa e o desenvolvimento de sementes. A mutação desse gene atrasa o desenvolvimento e a capacidade da planta de sobreviver a este tipo de estresse (Yang; Stierhof; Ludwig, 2015).

O AVT3 está *up*-regulado em todas as fases de desenvolvimento das sementes e na fase de 25-100mg de legumes. Também é um transportador vacuolar que apresenta altos níveis de expressão em sementes de *Arabidopsis* (Fujiki *et al.*, 2017). Assim como CAT9 ele exporta aminoácidos para o meio intercelular. No presente estudo, foi observado a *up*-regulação deste gene em todas as fases de desenvolvimento avaliadas na cultivar BRS 232. Os vacúolos servem para armazenamento temporário de aminoácidos (Martinoia *et al.*, 2012; Fujiki *et al.*, 2017) e o fato deste gene estar altamente expresso em sementes, indica que este gene pode estar envolvido na remobilização de aminoácidos reservados no vacúolo para a síntese de proteínas metabólicas, de reserva e/ou proteção (Jones e Vodkin, 2013). E assim como os genes AAP7 e CAT9, sua superexpressão pode estar relacionada com a tolerância das plantas durante estresse por falta de N (Fujiki *et al.*, 2017).

Os genes BAT1 e LHT1 não apresentam correlação positiva com o teor proteico em sementes de soja

Os transportadores de aminoácidos transmembrana BAT1 e LHT1, embora estejam presentes em sementes e legumes *in silico* no RT-qPCR, esses genes não apresentam expressão em nenhuma das fases de desenvolvimento de sementes na cultivar de maior teor proteico BRS 232. O BAT1 também encontra-se *down*-

regulado, porém o gene LHT1 apresentou *up*-regulação em legumes na fase de 100-200mg. Assim como AAP7, CAT9 e AVT3, a superexpressão do LHT1 pode estar envolvida na remobilização de N da planta em condições de solo com baixa concentração de N (Yang; Stierhof; Ludewig, 2015; Fujiki *et al.*, 2017; Zhou *et al.*, 2020).

O gene LHT1, codifica um transportador de lisina e histidina do tipo 1. Este transportador está presente em todos os tecidos, com elevados níveis de expressão em folhas, flores e sílicas (Chen e Bush, 1997). Em arroz, o silenciamento desse gene diminui a taxa de formação das sementes e o peso do grão (Guo *et al.*, 2020). Outro trabalho também em arroz, observou que o silenciamento do gene LHT1 diminuiu o rendimento do grão em mais de 40% (Wang *et al.* 2019). Em soja o papel desse gene em relação ao teor proteico parece estar mais envolvido no tecido de legume. O legume é considerado um tecido de armazenamento ao longo do desenvolvimento (Rossato *et al.*, 2001), ele é capaz de armazenar nitrogênio e carbono remobilizados posteriormente para a semente para utilização na germinação. Os legumes também são capazes de fornecer aminoácidos as sementes (Bennett; Roberts; Wagstaff, 2011). Por este motivo, é um tecido essencial para produção de sementes que contenham altas concentrações de proteínas (Bennett; Roberts; Wagstaff, 2011). Ainda, na literatura já foi observado uma forte associação entre o desenvolvimento do legume e o tamanho da semente (Bennett; Roberts; Wagstaff, 2011).

Estudos usando legumes destacados mostraram que os tecidos fotossintéticos da parede do legume são capazes de gerar assimilados (Bennet; Roberts; Wagstaff, 2011), que são exportados do legume para as sementes durante o desenvolvimento, sendo as sementes bem adaptadas para a absorção de fotoassimilados translocados da parede do legume. Essas informações relacionadas a importância do legume no desenvolvimento e enchimento do grão, e a expressão de LHT1 em legume na fase de 100-200mg, pode indicar uma importância desse gene no aumento do teor de proteína dos grãos.

O gene UPS2 pode estar envolvido no aumento do teor de proteínas apenas na fase 100-200mg na cultivar BRS 232

O gene UPS2, é uma permease transmembrana carreadora de ureídeos (alantoína e ácido alantóico/alantoato) e pirimidina uracila (Schmidt *et al.*, 2004; Carter e Tegeder, 2016). Em *Arabidopsis*, foi observado que o silenciamento deste gene e do gene UPS1 (ambos apresentam 78% de identidade proteica) diminui os níveis de ácido alantóico (alantoato) afetando negativamente o crescimento reprodutivo da planta (Takagi *et al.*, 2018). Por outro lado, a superexpressão do UPS1, aumentou a fixação simbiótica de N₂, gerando um aumento de 36% no rendimento de sementes (Carter e Tegeder, 2016). Devido a identidade entre as sequências proteica desses genes, pode se aferir que o gene UPS2, pode ter papel importante tanto no aumento do teor proteico no grão, quanto no rendimento final da cultura. Diferente dos outros genes mencionados acima, seu papel no teor proteico está relacionado principalmente aos ureídeos na semente da BRS 232. Em legumes este gene encontra-se *down*-regulado em todas as fases de desenvolvimento das sementes.

Conclusões

Genes com anotação de transportadores de compostos nitrogenados que apresentam altas contagens *in silico*, podem não estar associados diretamente com alto teor proteico. Isso é observado nos genes BAT1 e LHT1, que apesar de altas contagens nas bibliotecas analisadas, quando validados através do RT-qPCR não apresentaram relação com o alto teor proteico de sementes. Os genes AAP7, AVT3, CAT9, e UMAMIT25 apresentam relação

com alto teor proteico de sementes, independentemente do tipo de proteína sintetizada. O gene UPS2 pode estar relacionado com maior teor proteico na fase de 100-200mg.

Em relação ao teor proteico de vagens, não apresentam correlação com teor proteico de sementes. Somente os genes AVT3, CAT9, LHT1 e UMAMIT estão relacionados com teor proteico em vagens.

Considerações finais

Os genes AAP7, AVT3, CAT9 e UMAMIT25 podem ser utilizados posteriormente para edição através de Engenharia Genética, a fim de aumentar o teor de proteína sem intervenção na produtividade. O gene UMAMIT25 é o melhor candidato a ser superexpressado em sementes para aumento de proteína.

Contribuição dos autores

PILJ: Concepção do artigo e condução de todas as etapas metodológicas e de escrita do arquivo. PILJ e MDCM: Condução das análises estatísticas, análises in silico, de expressão gênica, redação do manuscrito, edição e revisão linguística. SRRM: análises de expressão gênica. FAH: Obtenção das bibliotecas proveniente do dataset 2. LMMH e ALN: coordenadores do grupo de pesquisa e estiveram diretamente envolvidos em todas as etapas da pesquisa descrita neste manuscrito e na revisão do texto do artigo.

Agradecimentos

Agradecemos a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão de bolsa de Mestrado à PILJ e a Fundação Eliseu Alves pela concessão de bolsa de Pós-doutorado à MDCM.

Financiamento

Este trabalho foi financiado pela Empresa Brasileira de Pesquisa Agropecuária - Embrapa Soja.

Conflito de interesses

Os autores declaram não haver conflito de interesses.

Referências

- ABIRACHED-DARMENCY, M., ABDEL-GAWWAD, M. R., CONEJERO, G., VERDEIL, J. L., THOMPSON, R. In situ expression of two storage protein genes in relation to histo-differentiation at mid-embryogenesis in *Medicago truncatula* and *Pisum sativum* seeds. **Journal of Experimental Botany**, v. 56, n. 418, p. 2019-2028, 2005.
- BENNETT, E. J.; ROBERTS, J. A.; WAGSTAFF, C. The role of the pod in seed development: strategies for manipulating yield. **New Phytologist**, v. 190, n. 4, p. 838-853, 2011.
- BESNARD, J., ZHAO, C., AVICE, J. C., VITHA, S., HYODO, A., PILOT, G., OKUMOTO, S. Arabidopsis UMAMIT24 and 25 are amino acid exporters involved in seed loading. **Journal of Experimental Botany**, v. 69, n. 21, p. 5221-5232, 2018.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regras para Análise de Sementes. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Brasília, DF: Mapa/ACS, 2009. 398p.
- BRUMM, T. J.; HURBURGH, C. R. Changes in long-term soybean compositional patterns. **Journal of the American Oil Chemists' Society**, v. 83, n. 11, p. 981-983, 2006.
- CARTER, A. M.; TEGEDER, M.. Increasing nitrogen fixation and seed development in soybean requires complex adjustments of nodule nitrogen metabolism and partitioning processes. **Current Biology**, v. 26, n. 15, p. 2044-2051, 2016.
- CHEN, L.; BUSH, D. R. LHT1, a lysine-and histidine-specific amino acid transporter in Arabidopsis. **Plant Physiology**, v. 115, n. 3, p. 1127-1134, 1997.
- CHENG, L. et al. Genome-wide identification, classification, and expression analysis of amino acid transporter gene family in *Glycine Max*. **Frontiers in Plant Science**, v. 7, p. 515, 2016.
- COLLIER, R.; TEGEDER, M. Soybean ureide transporters play a critical role in nodule development, function and nitrogen export. **The Plant Journal**, v. 72, n. 3, p. 355-367, June 2012.
- CONAB. COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento de safra brasileiro – grãos: Segundo levantamento, novembro 2019 – safra 2019/2020**. : Brasília: Companhia Nacional de Abastecimento. 2019. Disponível em: <<https://www.conab.gov.br/info-agro/safra/safra-graos/boletim-da-safra-de-graos>>. Acesso em : 22 nov. 2019.
- CONAB. COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento de safra brasileiro – grãos: 12º Levantamento - Safra 2019/20** : Brasília: Companhia Nacional de Abastecimento. 2020. Disponível em: <<https://www.conab.gov.br/info-agro/safra/safra-graos/boletim-da-safra-de-graos>>. Acesso em : 12 dez. 2020.
- DÜNDAR, E.; BUSH, Daniel R. BAT1, a bidirectional amino acid transporter in Arabidopsis. **Planta**, v. 229, n. 5, p. 1047-1056, 2009.
- ENGELS, C., FUGANTI-PAGLIARINI, R., MARIN, S. R. R., MARCELINO-GUIMARÃES, F. C., OLIVEIRA, M. C. N., KANAMORI, N., NEPOMUCENO, A. L. Introduction of the rd29A: AtDREB2A CA gene into soybean (*Glycine max* L. Merrill) and its molecular characterization in leaves and roots during dehydration. **Genetics and Molecular Biology**, v. 36, n. 4, p. 556-565, 2013.
- FABRE, F.; PLANCHON, C. Nitrogen nutrition, yield and protein content in soybean. **Plant Science**, v. 152, n. 1, p. 51-58, 2000.
- FUJIKI, Y., TESHIMA, H., KASHIWAO, S., KAWANO-KAWADA, M., OHSUMI, Y., KAKINUMA, Y., SEKITO, T. Functional identification of AtAVT3, a family of vacuolar amino acid transporters, in Arabidopsis. **FEBS letters**, v. 591, n. 1, p. 5-15, 2017.
- FUGANTI-PAGLIARINI, R., MARIN, S. R. R., MOLINARI, M., BARBOSA, D. D. A., MOLINARI, H., MERTZ-HENNING, L. M., NEPOMUCENO, A. L. Drought-tolerant soybean development: evaluation of GM lines under greenhouse and field conditions. **Embrapa Soja-Capítulo em livro científico (ALICE)**, 2020.

- GAZZONI, D. L. A soja no Brasil é movida por inovações tecnológicas. **Cienc. Cult.**, v. 70, n. 3, p. 16-18, 2018.
- GRALLATH, S., WEIMAR, T., MEYER, A., GUMY, C., SUTER-GROTEMEYER, M., NEUHAUS, J. M., & Rentsch, D. The AtProT family. Compatible solute transporters with similar substrate specificity but differential expression patterns. **Plant Physiology**, v. 137, n. 1, p. 117-126, 2005.
- GUO, N., GU, M., HU, J., QU, H., XU, G. Rice OsLHT1 Functions in leaf-to-panicle nitrogen allocation for grain yield and quality. **Frontiers in Plant Science**, v. 11, p. 1150, 2020.
- HERMAN, E. M. Soybean seed proteome rebalancing. **Frontiers in Plant Science**, v. 5, p. 437, set. 2014.
- HUNGRIA, M.; CAMPO, R. J.; NOGUEIRA, M. A. A pesquisa em Fixação Biológica do Nitrogênio na Embrapa Soja: passado, presente e perspectivas futuras. In: Anais da XVI Relare. Anais.. Londrina: 2012. Disponível em: < <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/105092/1/A-pesquisa-em-fixacao-biologica-do-nitrogenio-na-Embrapa-Soja-passado-presente-e-perspectivas-futuras.pdf>>.
- HONNA, P. T., FUGANTI-PAGLIARINI, R., FERREIRA, L. C., MOLINARI, M. D., MARIN, S. R., DE OLIVEIRA, M. C., NEPOMUCENO, A. L. Molecular, physiological, and agronomical characterization, in greenhouse and in field conditions, of soybean plants genetically modified with AtGolS2 gene for drought tolerance. **Molecular Breeding**, v. 36, n. 11, p. 1-17, 2016.
- HUANG, J., WANG, C., LI, X., FANG, X., HUANG, N., WANG, Y., COPENHAVER, G. P. Conservation and divergence in the meicyte sRNAomes of Arabidopsis, soybean, and cucumber. **Plant physiology**, v. 182, n. 1, p. 301-317, 2020.
- IFIF. THE INTERNATIONAL FEED INDUSTRY FEDERATION. **World compound feed production is estimated at over one billion tonnes annually. Global commercial feed manufacturing generates an estimated annual turnover of over US \$400 billion.** 2019. Disponível em: <https://ifif.org/global-feed/statistics/#:~:text=World%20compound%20feed%20production%20is,over%20one%20billion%20tonnes%20annually..> Acesso em: 17 dez. 2020.
- JONES, S. I., VODKIN, L. O. Using RNA-Seq to profile soybean seed development from fertilization to maturity. **PLoS One**, v. 8, n.3, Mar. 2013.
- JONES, S. I.; GONZALEZ, D. O.; VODKIN, L. O. Flux of transcript patterns during soybean seed development. **BMC Genomics**, v. 11, n.1, p. 136, Feb. 2010.
- JUNG, C. H., WONG, C. E., SINGH, M. B., BHALLA, P. L. Comparative genomic analysis of soybean flowering genes. **PLoS One**, v. 7, n. 6, p. e38250, 2012.
- LIU, S., WANG, D., MEI, Y., XIA, T., XU, W., ZHANG, Y. WANG, N. N. LIU, Sheng et al. Overexpression of GmAAP6a enhances tolerance to low nitrogen and improves seed nitrogen status by optimizing amino acid partitioning in soybean. **Plant Biotechnology Journal**, v. 18, n. 8, p. 1749-1762, 2020.
- MARDIS, E R. Next-generation DNA sequencing methods. **Annu. Rev. Genomics Hum. Genet.**, v. 9, p. 387-402, 2008.
- MASSON, P. H.; TAKAHASHI, T.; ANGELINI, R. Molecular mechanisms underlying polyamine functions in plants. **Frontiers in Plant Science**, v. 8, p. 14, 2017.
- MEDIC, JELENA; ATKINSON, CHRISTINE; HURBURGH, CHARLES R. Current knowledge in soybean composition. **Journal of the American Oil Chemists' Society**, v. 91, n.3, p. 363-384, Jan. 2014.
- MICHALAK, P., SOBRAL, B. W., ABEDI, V., KIM, Y. B., DENG, X., PHILIPSON, C., BASSAGANYA-RIERA, J. From Big Data Analytics and Network Inference to Systems Modeling. In **Computational Immunology**. Academic Press, 2016. p. 113-144. 2016.
- MOLINARI, M. D. C., FUGANTI-PAGLIARINI, R., MARIN, S. R. R., FERREIRA, L. C., BARBOSA, D. D. A., MARCOLINO-GOMES, J., ... & NEPOMUCENO, A. L. Overexpression of AtNCED3 gene improved

drought tolerance in soybean in greenhouse and field conditions. **Genetics and Molecular Biology**, v. 43, n. 3, 2020.

MOLINARI, M.D.C; PAGLIARINI, R. F.; MENDONÇA, J.A.; BARBOSA, D.A.; MARIN, D.R.; MERTZ-HENNING, L.M.; NEPOMUCENO, A. L. Transcriptome analysis using RNA-Seq from experiments with and without biological replicates: a review. **Rev. Ciências Agrárias**, v. 64. 2021.

MÜLLER, B., FASTNER, A., KARMANN, J., MANSCH, V., HOFFMANN, T., SCHWAB, W., BLECKMANN, A. Amino acid export in developing Arabidopsis seeds depends on UmamiT facilitators. *Current Biology*, v. 25, n. 23, p.3126-3131, Dec. 2015.

OHYAMA, T. Comparative studies on the distribution of nitrogen in soybean plants supplied with N₂ and NO₃- at the pod filling stage: II. Assimilation and Transport of Nitrogenous Constituents. **Soil Science and Plant Nutrition**, Niigata, v. 30, n. 2, p. 219-229, Mar.1984.

PATHAN, S. M., VUONG, T., CLARK, K., LEE, J. D., SHANNON, J. G., ROBERTS, C. A., NGUYEN, H. T. Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. **Crop Science**, v. 53, n.3, p. 765-774, 2013.

PATIL, G., MIAN, R., VUONG, T., PANTALONE, V., SONG, Q., CHEN, P., & NGUYEN, H. T. Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. **Theoretical and Applied Genetics**, v. 130, n. 10, p. 1975-1991, 2017.

PÍPOLO, A E., MANDARINO, J. M. G. Os teores de proteína da soja e a qualidade para a indústria. **Embrapa Soja-Artigo em periódico indexado (ALICE)**, 2016.

PÍPOLO, A. E., HUNGRIA, M., FRANCHINI, J. C., BALBINOT JUNIOR, A. A., DEBIASI, H., & MANDARINO, J. M. G. (2015). Teores de óleo e proteína em soja: fatores envolvidos e qualidade para a indústria. **Embrapa Soja-Comunicado Técnico (INFOTECA-E)**, 2015. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1025298>>. Acesso em: 11 nov. 2019.

POMAZNOY, M., SETHI, A., GREENBAUM, J., & PETERS, B.. Identifying inaccuracies in gene expression estimates from unstranded RNA-seq data. **Scientific Reports**, v. 9, n. 1, p. 1-10, 2019.

POLIZEL, A. M., MEDRI, M. E., NAKASHIMA, K., YAMANAKA, N., FARIAS, J. R. B., DE OLIVEIRA, M. C. N., ... & NEPOMUCENO, A. L. Molecular, anatomical and physiological properties of a genetically modified soybean line transformed with rd29A: AtDREB1A for the improvement of drought tolerance. **Embrapa Soja-Artigo em periódico indexado (ALICE)**, 2011.

RAMSEY, J.; SCHEMSKE, D. W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. **Annual review of ecology and systematics**, v. 29, n. 1, p. 467-501, 1998.

ROSSATO, L., LAINÉ, P., OURRY, A. P. Nitrogen storage and remobilization in Brassica napus L. during the growth cycle: nitrogen fluxes within the plant and changes in soluble protein patterns. **Journal of Experimental Botany**, v. 52, n. 361, p. 1655-1663, 2001.

ROTUNDO, J. L., BORRÁS, L., WESTGATE, M. E., ORF, J. H. Relationship between assimilate supply per seed during seed filling and soybean seed composition. **Field Crops Research**, v. 112, n. 1, p. 90-96, 2009.

SCHMID, M., DAVISON, T. S., HENZ, S. R., PAPE, U. J., DEMAR, M., VINGRON, M., LOHMANN, J. U. A gene expression map of Arabidopsis thaliana development. **Nature Genetics**, v. 37, n. 5, p. 501-506, 2005.

SWARUP, R.; PÉRET, B. AUX/LAX family of auxin influx carriers—an overview. **Frontiers in Plant Science**, v. 3, p. 225, 2012.

TAKAGI, H., WATANABE, S., TANAKA, S., MATSUURA, T., MORI, I. C., HIRAYAMA, T., SAKAMOTO, A.. Disruption of ureide degradation affects plant growth and development during and after transition from vegetative to reproductive stages. **BMC Plant Biology**, v. 18, n. 1, p. 1-16, 2018.

TAKESHIMA, R., NAN, H., HARIGAI, K., DONG, L., ZHU, J., LU, S., ABE, J. Functional divergence between

soybean FLOWERING LOCUS T orthologues FT2a and FT5a in post-flowering stem growth. **Journal of Experimental Botany**, v. 70, n. 15, p. 3941-3953, 2019.

TEGEDER, M.; RENTSCH, D. Uptake and partitioning of amino acids and peptides. **Molecular Plant**, v. 3, n. 6, p. 997-1011, 2010.

WANG, X., YANG, G., SHI, M., HAO, D., WEI, Q., WANG, Z., XIA, J.. Disruption of an amino acid transporter LHT1 leads to growth inhibition and low yields in rice. **BMC Plant Biology**, v. 19, n. 1, p. 1-11, 2019.

YANG, H.; STIERHOF, Y.; LUDEWIG, U. The putative Cationic Amino Acid Transporter 9 is targeted to vesicles and may be involved in plant amino acid homeostasis. **Frontiers in Plant Science**, v. 6, p. 212, 2015.

ZHOU, T., YUE, C. P., HUANG, J. Y., CUI, J. Q., LIU, Y., WANG, W. M., HUA, Y. P. Genome-wide identification of the amino acid permease genes and molecular characterization of their transcriptional responses to various nutrient stresses in allotetraploid rapeseed. **BMC Plant Biology**, v. 20, p. 1-22, 2020.