JOURNAL OF PLANT INTERACTIONS
2022, VOL. 17, NO. 1, 309–318
https://doi.org/10.1080/17429145.2022.2039791

PLANT-ENVIRONMENT INTERACTIONS

OPEN ACCESS Check for updates

Nitrogen compounds transporters: candidates to increase the protein content in soybean seeds

Pamela I. L. Joaquim[a*], Mayla D. C. Molinari[b*], Silvana R. R. Marin[c], Daniel A. Barbosa[d], Américo J. C. Viana[b], Elibio L. Rech[e], Fernando A. Henning[c], Alexandre L. Nepomuceno[c] and Liliane M. Mertz-Henning[c]

[a]Biotechnology, Department, Londrina State University, Londrina, Brazil; [b]Arthur Bernardes Foundation, Embrapa Soja, Londrina, Brazil; [c]Empresa Brasileira de Pesquisa Agropecuária, Embrapa soja, Londrina, Brazil; [d]Department of General Biology, Londrina State University, Londrina, Brazil; [e]Embrapa Recursos Genéticos e Biotecnologia, National Institute of Science and Technology in Synthetic Biology, Brasília, Brazil

ABSTRACT

The yield and protein content of soybean seeds are challenging traits for breeding since they have a negative correlation. To solve this issue, it is essential to understand the protein transport in soybean seeds. Here we performed an analysis of 3 datasets containing transcriptional data from soybean seeds and pods, without any treatment, aiming to select genes related to the transport of organic nitrogen compounds. We identified 69 transporters, and among them, seven were chosen to be validated by qRT-PCR. Seeds and pods of soybean cultivars BRS 232 and BRS 284, respectively, presenting higher and lower protein content, were assayed under normal growth condition. Results showed in BRS 232 soybean cultivar, a positive correlation between seed protein content and gene expression for five out of the seven genes analyzed. These nitrogen compounds transporters can play an essential role in the storage of proteins, thus increases the protein content of soybean seeds and contributes to the decrease the negative correlation between yield and protein content of soybean seeds.

ARTICLE HISTORY
Received 6 October 2021
Accepted 4 February 2022

KEYWORDS
Glycine max; nitrogen compounds transporters; seed protein content; transcriptome; RNA-Seq; qRT-PCR

Introduction

Soybean is an important agricultural commodity globally since its production chain is 3% of the world's Gross Domestic Product, moving over 70 billion dollars (Franceschini 2017). The 2020/2021 crop season produced 362.95 million tons of soybean grain, and the estimate for the 2021/2022 harvest is an increase of approximately 5.86% in that production (USDA 2021). The leading producing countries are the USA, Brazil, and Argentina, which jointly account for over 80% of world production (http://www.soystats.com; Patil et al. 2017). The growing demand and supply of soybean are mainly due to its high protein content (∼38%) and quality (Pípolo and Mandarino 2016). The soybean seeds are rich in essential and sulfur amino acids, and for this reason, it is the dominant source of protein in animal feed (Michelfelder 2009; Pípolo and Mandarino 2016).

Over the years, breeding programs have focused on increasing soybean yield, which has led to a reduction in seed protein content due to the negative genetic correlation between these traits (Pathan et al. 2013). In this scenario, a study showed a decrease of 1.3 g Kg^{-1}, in protein content per Mg seed yield increase (Assefa et al. 2019). For mitigating that correlation, recent studies have employed molecular genetic approaches, such as identifying genomic regions, point mutations, and genes related to protein content in seeds. In this sense, 12 Quantitative Trait Loci (QTLs) (Seo et al. 2019) and 40 Single Nucleotide Polymorphisms (SNPs) localized in 17 genomic regions (Hwang et al. 2014), have been identified. From those, five SNPs, localized on chromosome 20, showed a strong positive correlation with protein content (Hwang et al. 2014). Despite the awareness of a set of molecular markers associated with soy protein content (Patil et al. 2017), their use in marker-assisted selection for genetic breeding is still not widespread. Among the difficulties for adopting such a strategy, is the complexity of that trait controlled by many genes, combined with the fact that many of those markers still need validation before they can get into the routine of breeding programs. Moreover, the negative genetic correlation between protein content and other important traits, such as yield and oil content, makes breeding for this trait even more challenging (Patil et al. 2017).

In such circumstances, genetic engineering approaches such as transgenics and gene editing can contribute to developing soybean genotypes with high seed protein content (Liu et al. 2020). For this, it is essential to expand the comprehension of the molecular mechanisms that control the synthesis, transport, and storage of proteins in the seeds (The et al. 2020; Yang et al. 2020). The nitrogen resources required for storage in seeds only partially derive from direct uptake but rather are derived from organic nitrogen, taken up before. Therefore, the seed protein content relies on the transport of organic nitrogen assigned as amino acids, ureides, and peptides from roots, nodules, and mature leaves (Tegeder and Masclaux-Daubresse 2018). Thus, among nitrogen transporters, those involved in the transport of organic nitrogen compounds are more important for protein accumulation in seeds than those involved in the transport of inorganic nitrogen compounds (Jones and Vodkin 2013; The et al. 2020; Thu et al. 2020).

CONTACT Liliane M. Mertz-Henning liliane.henning@embrapa.br Embrapa Soja, Londrina, PR, Brazil
*These authors contributed equally to this work
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Studies with soybean identified 189 genes coding for amino acid transporters. These transporters were classified into 12 subfamilies, among which Cationic Amino Acid Transporters (CATs), Amino Acid Permeases (AAPs), Lysine and Histidine Transporters (LHTs) are the best-studied (Cheng et al. 2016). Overexpression of genes related to organic nitrogen compounds transporters (NCTs) increased nitrogen transport and increase yield (Carter and Tegeder 2016; Besnard et al. 2018). In soybean plants overexpressing the common bean ureide transporter *PvUPS1* (Ureide Permease 1), a 36% increase in overall seed yield was observed (Carter and Tegeder 2016). In the model plant *Arabidopsis thaliana* (Arabidopsis), the overexpression of the amino acid transporter *UMAMIT25* (Usually Multiple Acids Move In And Out Transporters 25) induced an increase in the number of seeds (Besnard et al. 2018). In rice, the overexpression of NCT *OsAAP1* (Amino Acid Permease1) increased the uptake and reallocation of N, which positively affected the final grain yield (Ji et al. 2020).

Genes encoding NCTs identification helps to get a more accurate understanding of the processes involved in seed loading. For this reason, the objective of this work is to identify promising NCTs in 3 datasets of soybean seeds and pods transcriptome and validate their expressions by Quantitative Reverse Transcription PCR (qRT-PCR) in the same tissues of soybean cultivars with higher and lower seed protein content. Comparing the expression pattern of NCTs genes in productive soybean cultivars with contrasting seed protein content can assist in selecting target genes that are promising for strategies aimed at increasing the protein content of the seed without negatively affecting yield.

Material and methods

Identification of candidate genes for NCTs in RNA-Seq libraries

The selection of candidate genes for NCTs was carried out by analyzing three independent datasets of RNA-Seq libraries: Dataset 1, RNA-Seq of developing seeds and seed coats (Jones and Vodkin 2013); Dataset 2, RNA-Seq of mature seeds, generated by Embrapa Soybean from BRS MG715, DM 6563, and BRS 413RR genotypes; and Dataset 3, RNA-Seq of seedless pods, generated by Embrapa Soybean from the BR-16 genotype (Additional file 2). All the stages of biological material were described in the Additional file 4. In short, was evaluated mature seeds, seeds in different development stages, cotyledon, and pods.

For the *in silico* analysis, the quality of the fragments before and after cleaning by the FastQC software version 0.11.5 was evaluated (Andrews 2010; Patel and Jain 2012). After cleaning the reads, just fragments with ≥ 40 bp and Phred Quality Score ≥ 20 were retrieved using the Trimmomatic software version 0.36 (Bolger et al. 2014). Then, the reads were aligned on the soybean genome version Wm82.a2.v1 using the Hisat2 software version 2.1.0 (Kim et al. 2015; Kim et al. 2019). PCR artifacts using the Samtools v.1.5 software (Li et al. 2009) and mapping and normalization between libraries using Stringtie v.1.3.3 software in transcripts per million were carried out (TPM; Pertea et al. 2015). All genes annotated as amino acids transporters and with the TPM ≥ 10 (Machado et al. 2020) were selected for further evaluation. The bioinformatic analyses were performed according to the best practices of data analysis for RNA-Seq (Conesa et al. 2016; Molinari et al. 2021a).

In silico characterization of amino acid transporters

For the genes annotated as amino acid transporters, a characterization regarding the number of transcripts and copies in soybean and orthologs in Arabidopsis (TAIR10) by Phytozome database and Persephone software (https://persephonesoft.com/) was carried out. To perform this analyses, the annotations Araport11 (Cheng et al. 2017), Wm82.a2.v1 (Schmutz et al. 2010), and Wm82.a4.v1 (Lawrence-Dill 2019) served as references. Were validated by qRT-PCR only genes contained in the two versions of the soybean genome to avoid sequencing errors between both versions. Gene ontology using ShinyGO software version 0.66 (http://bioinformatics.sdstate.edu/go/) was carried out and analysis of protein domains using the MotifScan software (https://myhits.sib.swiss/cgi-bin/motif_scan) and the Ensemble database (https://www.ensembl.org/index.html) was performed. Heatmaps by the ClustVis software (Metsalu and Vilo 2015) and the Venn diagrams by the Jvenn software (Bardou et al. 2014) were generated.

Plant material

In the experiments carried out under controlled conditions in a greenhouse (Additional file 1A-D), two conventional soybean cultivars with high yield potential but having contrasting seed protein contents (BRS 232 with higher and BRS 284 with lower protein contents were used).

The experimental design used was randomized blocks, with 18 biological replications for each of the cultivars. After germinating the seeds on Germitest® paper rolls and according to the rules for seed analysis (Brasil 2009), we treated the seedling roots with 5.0×10^9 CFU/mL of *Bradyrhizobium japonicum* Atmo© inoculant and transplanted two of the seedlings to a 10 L pot containing sterilized soil and sand (1:1), in a total of 36 vessels, kept humid by a drip irrigation system. We collected one pod from each of the 18 plants in three stages of development (R5.1/R5.2 – seeds with a fresh weight of 25-100 mg; R5.3 – seeds with 100-200 mg and R6 – seeds with 350-450 mg; Additional file 1E) to determine the seed weight and protein content. The basis for the selection of these stages was on the type of protein synthesized during seed development, considering that, up to 100 mg, there is a predominance in the synthesis of metabolic proteins, and, beyond this weight, the synthesis of reserve proteins starts (Jones and Vodkin 2013).

In the gene expression experiments, we evaluated plants at the three stages mentioned above, and we divided the 18 plants of each cultivar into three bulks of six plants, with six pods collected from each plant, totaling 36 pods per bulk, around 108 seeds for BRS 284 and 72 for BRS 232. The pods and the seeds were separated and stored in a freezer at −80° C until the RNA extraction.

Rna isolation, cDNA synthesis, and qRT-PCR analysis

Total RNA extraction, conducted by using Trizol reagent, was done following the manufacturer’s recommendations

(Invitrogen, Carlsbad, CA, USA). After extraction, we treated the RNA was treated with DNAse I (Invitrogen, Carlsbad, CA, USA). Then, to confirm the absence of genomic DNA, a conventional PCR was performed using the β-actin intron primer. cDNA was synthesized from 1.5 µg total RNA using Super Script® III First-Strand Synthesis System (Invitrogen, Carlsbad, CA, USA) and oligo(dT)18 according to the manufacturer's instructions. Again, PCR to verify the efficiency of the synthesis and absence of genomic DNA was carried out. The qRT-PCR reactions were conducted on a Real-Time PCR 7900HT (Applied Biosystems, Foster City, CA, USA), using Platinum SYBR green (cat. no. 11733-038; Invitrogen), with *β-actina* and *ELF1β* (Wan et al. 2017) as reference genes (Czechowski et al. 2005). The relative quantification of genes was performed using three random biologicals and three technical replicates ($n = 9$). The reactions were made through 7900HT thermocycler equipment (Applied Biosystems). The cycling conditions used were denaturation at 95°C for 20 s (s) followed by 40 cycles of 95°C for 3 s, 60°C for 26 s, and 1 cycle for Melt curve at 95°C for 15 s, 60°C for 1 min and 95°C for 15 s. The expression level was determined using the formula $2^{-\Delta\Delta Ct}$ adapted according to the primer's efficiencies (Livak and Schmittgen 2001).

The qRT-PCR data was performed using the comparison between two groups by the two-tailed Student's t-test. Analyses were considered significant at $p < 0.05$ (*). These analyses as described previously (Czechowski et al. 2005; Molinari et al. 2021b).

Primer design

Primers for validation by qRT-PCR of genes from NCTs families best characterized, such as Cationic Amino Acid Transporter – CAT, Lysine Histidine Transporter – LHT, Amino Acid Permeasse – AAP, UMAMIT – Usually Multiple Acids Move In and Out Transporters, Bidirectional Amino Acid Transporter – BAT, Ureide Permease – UPS, Amino Acid Vacuolar Transporter – AVT (Cheng et al. 2016; Fujiki et al. 2017; Zhou et al. 2020), were designed using Primer3plus software (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi) and only primers without the formation of secondary structures were selected (Additional file 3).

Determination of weight and protein content in seeds and pods

Pods from both soybean cultivars, BRS 232 and BRS 284, separated from the seeds, were collected, and immediately assessed for fresh weight. The fresh weight of seeds was also assessed. The after drying these samples in an oven at 65° C, for 48 h, weighing was performed to obtain the dry weight. The fresh and dry weight of both tissues was performed separately in analytical balance.

Protein content determination was performed using a near-infrared spectroscopy equipment (FT-NIR) (Antaris II, Thermo Scientific, Waltham, MA, USA), equipped with an integrating sphere. Readings were assayed in the range between 1100 and 2500 nanometers. For every 30 g sample of each cultivar in biological triplicate, 32 scans were collected, with 4 cm^{-1} resolution and background at each reading. The average spectra went through prediction in four different mathematical models for each constituent. The basis for the modeling was on the partial least square's regression, with pre-treatments for data normalization and application of the first derivative of Savitzky–Golay.

Statistical analyses

Data on pod and seed weight and protein content were submitted to analysis of variance (ANOVA), with the means compared by the Tukey test. Analyses were considered significant at $P < 0.05$ (*).

Results and discussion

Expression of NCTs by tissue

We identified 69 genes for NCTs with high expression levels in soybean RNA-Seq libraries from seed coats, developing seeds, mature seeds, and pods tissues (Figure 1A; Additional file 4). Of those, 65 showed high expression in pods (Pod reproductive stage 4 – stage R4), 36 in seed coat with 5-6 mg, 58 in whole seed (WS) with 4–24 days after flowering (DAF, stage R5.1/5.2), 56 in WS with 5-6 mg, 50 in cotyledon with 100-200 mg (stage R5.3), 49 in cotyledon with 400-500 mg (stage R6) and 31 in mature WS (stage R8) (Additional file 4).

When evaluating the overall expression profile of NCTs by tissue through Pearson's correlation, a similar expression profile between NCTs of Pod R4, Seed Coat 5-6 mg, and WS 4–24 DAF we observed, so they were grouped in Cluster 1 while seeds with more than 5-6 mg until maturity were grouped in Cluster 2 (Figure 1A). We found that 36 (22.6%) among the 159 transcripts of Cluster 1 were common to all tissues (Figure 1B), while in Cluster 2, only 26 (13.9%) of the 186 transcripts were common to all tissues (Figure 1C). The clustering observed in Cluster 1 between the seed coat, pod, and seeds in the early development stages (WS 4–24 DAF), probably reflects the significant role of these tissues in controlling the embryo development (Jones and Vodkin 2013; Kim et al. 2017). Seeds also rely on the seed coat and pods to store proteins, and these tissues have the role of uptaking amino acids and ureides coming from the leaves and roots to allocate them to the seeds, and much of this transfer depends on membrane transporters (Castro et al. 2004; Abirached-Darmency et al. 2005; Herman 2014; Karmann et al. 2018). Pods, seed coats, and seeds shared 22.6% of NCTs in their early development stages. That observation, suggests that amino acids and ureides transported by pods and seed coats are focused on metabolic protein synthesis, whose function is to maintain the proper development of the embryo until the 24th4 DAF (Tegeder and Rentsch 2010; Jones and Vodkin 2013). Since, at this stage, the synthesized proteins are for metabolic use. After this stage, there is a change in the expression and types of NCTs, in which 75% of tissues initiate reserve-type protein synthesis (Tegeder and Rentsch 2010), which explains the grouping of these tissues in Cluster 2 (Figure 1A e 1C).

Among the 69 genes related to NCTs, 29% (20 genes) were identified as members of the AAP family (Amino Acid Permease), 18% (12 genes) from AVT family (Amino Acid Vacuolar Transporter, 13% (9 genes) from CAT family (Cationic Amino Acid Transporters), 10% (7 genes) from

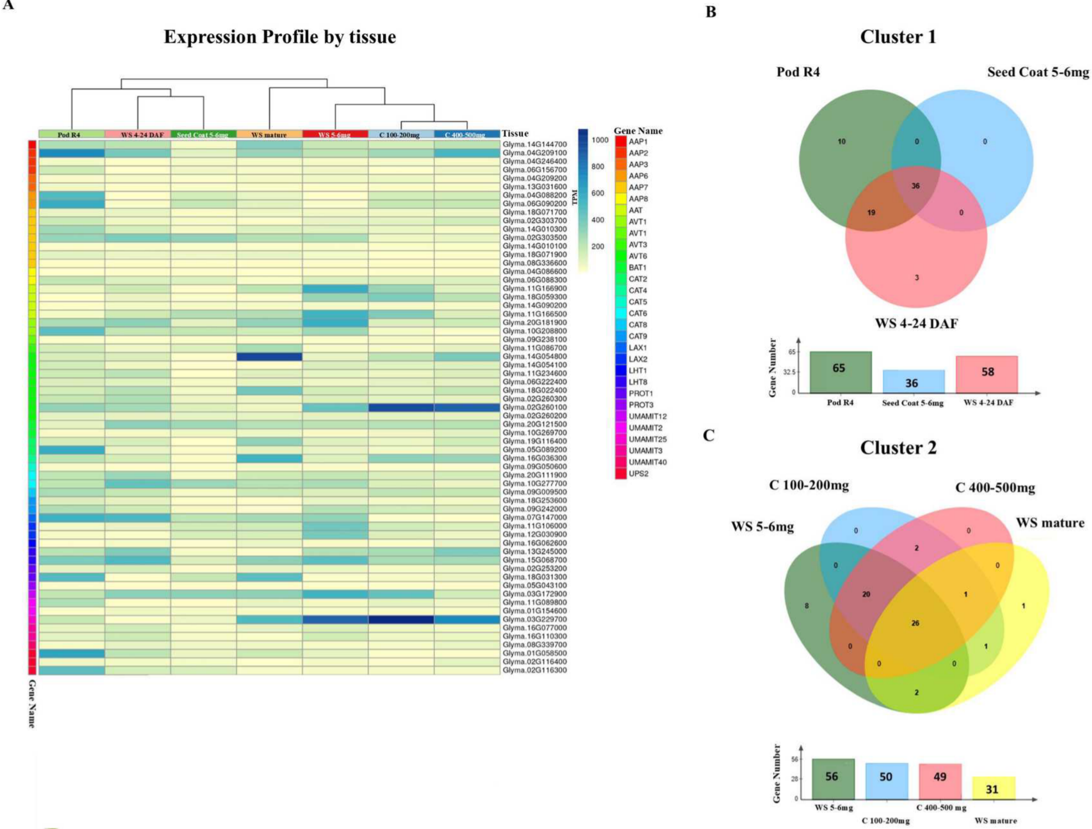

Figure 1. Expression profile of genes related to the nitrogen compounds transporter (NCTs) in soybean. **A**. Clustering of genes related to NCT. Cluster 1: Genes identified in pods (Pod R4), seeds in the early stages of development (WS 4–24 DAF), and seed coat (Seed Coat 5-6 mg). Cluster 2: Genes identified in seeds from 5-6 mg to maturity (WS 5-6 mg, WS 100-200 mg, WS 400-500 mg, and WS mature). Values represent expression levels in transcripts per million-TPM. **B.** Overlap between transcripts in Cluster 1. **C.** Overlap between in Cluster 2. WS: Whole seed; C: Cotyledon.

UMAMIT family (Usually Multiple Acids Move In and Out Transporters), 7% (5 genes) from AAT family (Amino Acid Transporters), 6% (4 genes) from LHT family (Lysine Histidine Transporter) and LAX family (Like Aux1), 4% (3 genes) from PROT family (Proline and Glycine Betaine Transporter) and UPS family (Ureide Permease), and 3% (2 genes) from BAT family (Bidirectional Amino Acid Transporter) Figure 2A, Additional file 4). Regarding the distribution of genes by tissue, we observed that all evaluated tissues shared 20 NCT-related genes (29%; Additional file 5).

Each type of transporter has an affinity for one or more substrates and can play different roles and functions in plants

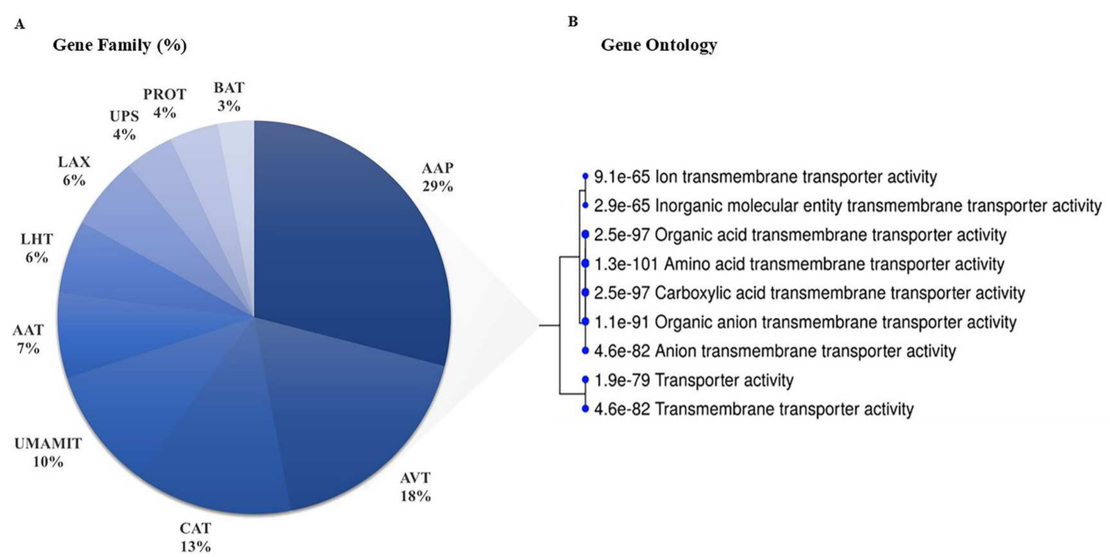

Figure 2. Characterization of genes related to the nitrogen compounds transport (NCTs) in soybean. **A**. Distribution by the family of the 69 NCT-related genes with TPM ≥ 10 identified in at least one of the seven RNA-seq libraries evaluated in this work. **B**. Molecular function of the 69 NCT-related genes. *AAP:* Amino Acid Permeasse*; AVT:* Amino Acid Vacuolar Transporter; *CAT:* Cationic Amino Acid Transporter*; UMAMIT:* Usually Multiple Acids Move In and Out Transporters; *AAT:* Amino Acid Transporters*; LHT*: Lysine Histidine Transporter; *LAX:* Like Aux1*; PROT*: Proline and Glycine Betaine Transporter*; UPS*: Ureide Permease; *BAT*: Bidirectional Amino Acid Transporter.

(Yang et al. 2020). In the early stages of seed development, the main requirements involve the development of the embryo and the accumulation of metabolic proteins for seed development (Jones and Vodkin 2013). As the seed grows, the type of protein synthesized changed, moving from metabolic to storage proteins (Jones and Vodkin 2013). Changes in gene expression and types of NCTs observed between Clusters 1 and 2 (Figure 1A) corroborate those observations. Previous results indicated that the activity of transporters in the pod and seed coat impact the quality of mature seed (Castro et al. 2004; Abirached-Darmency et al. 2005), which suggests the participation of these seed tissues in the storage of reserve-type proteins. The observation that all tissues evaluated in our work share 29% of NCTs support the findings available in the literature described above.

The gene ontology analysis indicated that the main molecular functions performed by NCT-related genes, in addition to amino acid transport, are the transport of organic acids, inorganic solutes, anionic ions, carboxylic acid, and organic anions (Figure 2B). Previous work had already suggested that NCTs can transport different substrates (Yang et al. 2020), and the gene ontology analysis of the present study confirmed this role (Figure 2B).

The NCTs also play a role in the transport of nitrogen compounds across cell membranes. The AAP family, for example, present in the plasma membrane, is characterized by transporting amino acids from other parts of the plant for seed nutrition, even in a condition of nitrogen deprivation in the soil, thus ensuring seed production through the remobilization of that element (Karmann et al. 2018; Liu et al. 2020; Zhou et al. 2020). The activity of the AVT family, also characterized by acting on the plasma membrane, is related to the importation and exportation of amino acids stored in vacuoles and may be related to the release of amino acids as required by the tissues. Therefore, contributing to the accumulation of proteins that will culminate later in the final seed protein content (Fujiki et al. 2017). The predominant export destination of amino acids from the vacuoles to the cytosol is protein synthesis, and members of the AVT family are among those primarily molecules responsible for that transport (Fujiki et al. 2017).

CAT family plays a similar role as AVT family, but besides acting in the vacuolar membrane, these transporters are also present in mitochondrial and chloroplast membranes (Su et al. 2004). According to Su et al. (2004), *CAT5* acts as a basic amino acid transporter, while *CAT3, CAT6,* and *CAT8* preferentially transport neutral and acidic amino acids. The UMAMIT and LHT families, in addition to transporting amino acids to seeds, significantly increase their yield (Besnard et al. 2018; Wang et al. 2019). The *UMAMIT* overexpression accelerated the pericarp maturation and resulted in an increase in total seed number and seed mass produced per plant in Arabidopsis (Besnard et al. 2018). While a loss-of-function of *OsLHT1* inhibited rice root and shoot growth, and markedly reduced grain yields suggesting that the overexpression may increase the yield (Wang et al. 2019). Finally, the ProT family, characterized as a carrier of proline, glycine, betaine, and GABA, is involved in long-distance transport, and it occurs too in the plasma membrane (Grallath et al. 2005; Lin et al. 2019; Yang et al. 2020)..

The activity of NCT-related genes has a positive relationship with seed yield and quality in several crops. In soybean, the overexpression of NCT *AAP6* increased tolerance to nitrogen starvation and the protein quality of soybean seeds, probably due to *AAP6* capacity to transport essential amino acids to seeds (Liu et al. 2020). Besides that, *APP6* overexpressed improves seed nitrogen status by optimizing amino acid partitioning in soybean resulting in seed yield increased (Liu et al. 2020). In rice, the overexpression of NCT *AAP1* positively increased the yield of this crop (Ji et al. 2020). *OsAAP1* may affect N transport and metabolism, and auxin, cytokinin, and strigolactone signaling in regulating rice tillering. These results support together with that the increase of neutral amino acid uptake and reallocation via *OsAAP1* could improve growth and grain yield in rice (Ji et al. 2020).

In Arabidopsis, the overexpression of NCTs *UMAMIT24* and *UMAMIT25* significantly increased the number of seeds and the essential amino acids in the seeds (Besnard et al. 2018).

Expression of NCTs in soybean cultivars with higher and lower protein content

We characterized the seeds and pods of the cultivars used for gene expression validation via qRT-PCR for weight and protein content in the development stages R5.1/2 (25-100 mg), R5.3 (100-200 mg), and R6 (350-450 mg). We observed a significant difference in seed protein content, 38.98 ± 0.46% for the cultivar BRS 232 and 36.01 ± 0.93% for BRS 284. These results follow previously observed data, where BRS 232 was classified as a higher protein content cultivar and BRS 284 as a lower one (EMBRAPA 2014). We highlighted that despite the variation in protein content, both cultivars are highly productive with weight of a thousand seeds of 185 grams to BRS 232 and 146 grams to BRS 284 (EMBRAPA 2014). As for the weight of seeds and pods, no significant variation between cultivars BRS 232 and BRS 284 was observed, both for fresh and dry weights. The means and standard deviations of fresh and dry weights of the seeds of both cultivars were, respectively, 56.5 ± 3.22 mg and 8.5 ± 0.78 mg for the stage R5.1/2, 180 ± 8, 74 mg and 38.5 ± 2.19 mg for stage R5.3 and 385 ± 11.48 mg and 134 ± 3.40 mg for stage R6. For pods, the fresh and dry weights of both cultivars were, respectively, 556 ± 11.97 mg and 101 ± 5.23 mg in R5.1/2, 473 ± 11.64 mg and 90 ± 3.14 mg in R5.3 and 482 ± 16.53 mg and 158 ± 7.38 mg in R6.

After characterization for weight and protein content, cultivars BRS 232 and BRS 284 were used to validate gene expression by qRT-PCR. To this end, we sought to represent the predominant families of NCTs described in the literature (Figure 2A). Thus, the following NCT-related genes were selected: *Glyma.14G010300, Glyma.09G242000, Glyma.09G197800, Glyma.10G269700, Glyma.02G116400, Glyma.16G062600*, and *Glyma.03G229700*, belonging to the AAP, CAT, AVT, BAT, UPS, LHT, and UMAMIT families, respectively. All NCT-related genes selected were expressed in seed and pods tissues at all three stages of development (R5.1/R5.2, R5.3, and R6), which confirmed data from transcriptome libraries. In seeds, for all development stages, genes *AAP7, AVT3, CAT9*, and *UMAMIT25* showed the highest expression levels in cultivar BRS 232 compared to BRS 284 (Figure 3). In this regard, the *UPS2* gene also showed higher expression in cultivar BRS 232 than in BRS

284, although the higher expression levels were only identified in R5.3 stage. On the other hand, at all three stages of development, the *LHT1* and *BAT1* genes showed lower expression in BRS 232 seeds were compared to BRS 284 (Figure 3).

Regarding gene expression in pods, an opposite pattern was observed, with genes *AAP7*, *BAT1* and *UPS2* being more expressed at all three stages in cultivar BRS 284, whose protein content is lower. Similarly, the *LHT1* gene presented a higher expression in BRS 284 at R5.1/R5.2 and R6 stages, as well as *AVT3* which, was more expressed in BRS 284 at R5.3 and R6 stages, and finally, *UMAMIT25* showed a markedly higher expression at R5.1/R5.2 stage in cultivar BRS 284 compared to BRS 232 (Figure 3).

Overall, these results suggest a positive correlation between higher expression for NCT-related genes *AAP7, AVT3, CAT9, UMAMIT25*, and *UPS2* with a higher seed protein content of cultivar BRS 232 (Figure 3). However, in pods, the expression of genes *AAP7, AVT3, BAT1, UPS2, LHT1*, and *UMAMIT25* did not correlate positively with the higher protein content of cultivar BRS 232, since these genes were more expressed in cultivar BRS 284, with lower protein content (Figure 3). The observation that five of the seven genes validated by qRT-PCR had higher expression in the seeds of the cultivar with the highest protein content reinforces the suggested role that NCTs have an essential role in increasing the protein content of soybean seeds (Liu et al. 2020), notably when expressed directly in seed tissues.

Copy number, transcripts, and orthologs in soybean and arabidopsis

In an *in silico* analysis carried out, we observed that five of the seven genes validated by qRT-PCR are present in multiple copies (*AAP7, CAT9, BAT1, LHT1*, and *UMAMIT25*) and two in single copies (*AVT3* and *UP2*) in the soybean genome. In Arabidopsis, all genes are present in single copies (Table 1). In soybean, in addition to higher copy numbers, higher numbers of alternative transcripts were observed if compared to orthologous in Arabidopsis (Table 1). The percentage of physicochemical similarity of protein sequences between soybean and Arabidopsis was higher than 62%, and all genes have conserved domains of transmembrane transporters (Table 1). The importance of assessing levels of gene duplication is that the duplication process can lead to pseudogenization (loss of function) during the evolutionary process (Schmutz et al. 2010; Jung et al. 2012). In the present study, although we observed that five of the seven genes in soybean are in multiple copies, all have conserved domains, even in alternative transcripts, indicating that the duplication of these genes did not negatively influence the performance of their biological expected roles. Instead, the increase in the number of copies of these genes can be advantageous for the overexpression process by genetic engineering.

In Arabidopsis, NCTs *AAP7, AVT3, LHT1*, and *UMAMIT25* are localized in the plasma membrane, while *CAT9, AVT3, BAT1, UPS2*, and *LHT1* occur in the membrane of

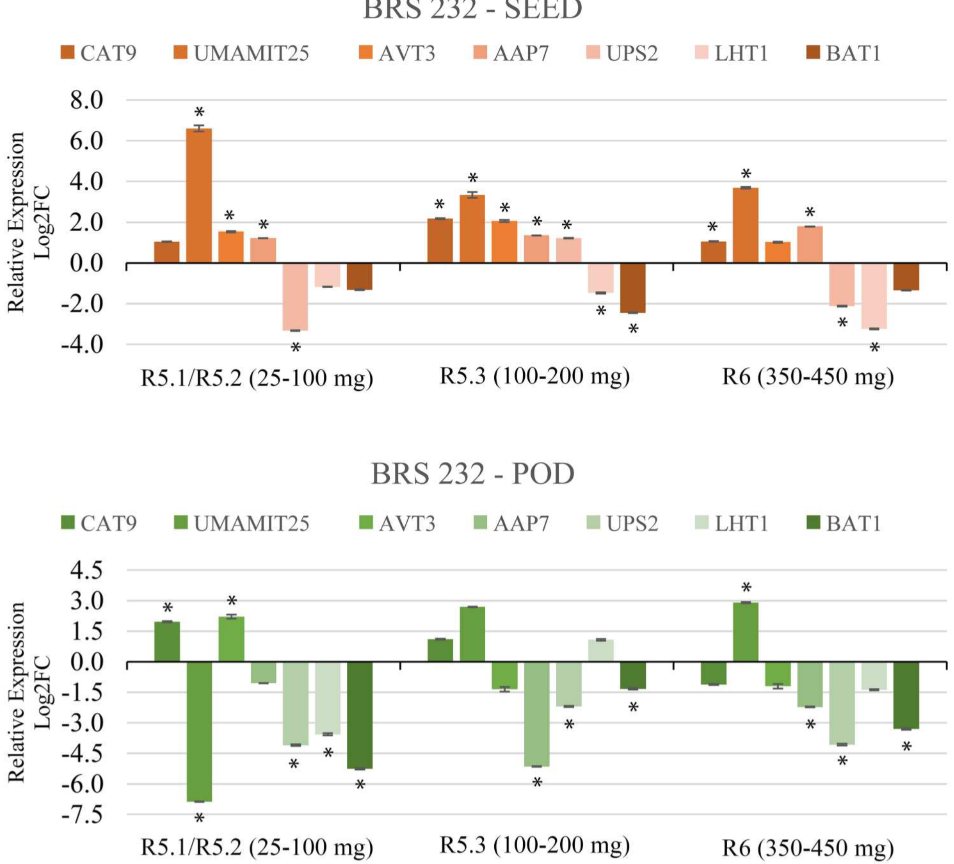

Figure 3. Relative gene expression of *AAP7, AVT3, BAT1, CAT9, LHT1, UMAMIT25*, and *UPS2* in soybean seeds and pods at development stages R5.1/R5.2, R5.3, and R6. The highest values (positive ones) and lowest expression (negative ones) represent the averages of three biological repetition observed in cultivar BRS 232 compared to the cultivar BRS 284. The bars represent the standard error of three biological repetitions. Two-tailed Student's *t*-test * $P < 0.05$. Orange bars represent the gene expression levels in seeds, and green bars represent the gene expression levels in pods. Upper bars represent up-regulated genes, and Lower bars represent down-regulated genes.

Table 1. Copy number and orthologue evaluation between Soybean and Arabidopsis.

Gene ID	SCN	Genetic copies	%S	STN	Annotation	Orthologue	ACN	ATN	%S	CL
Glyma.14G010300	2	Glyma.02G303500	99	13/5	AAP7	AT5G23810	1	4	62	Plasmatic
Glyma.09G242000	2	Glyma.18G253600	94	1/6	CAT9	AT1G05940	1	4	72	Chloroplast, Mitochondrion, Vacuole
Glyma.09G197800	1	NA	NA	NA	AVT3	AT4G38250	1	1	62	Plasmatic, Vacuole
Glyma.10G269700	2	Glyma.20G121500	91	4/1	BAT1	AT2G01170	1	2	79	Mitochondrion
Glyma.02G116400	1	NA	NA	NA	UPS2	AT2G03530	1	5	77	Mitochondrion
Glyma.16G062600	2	Glyma.19G083900	96	1/3	LHT1	AT5G40780	1	3	83	Plasmatic, Plastid
Glyma.03G229700	3	Glyma.19G227900/Glyma.19G227000	96/96	2/2/1	UMAMIT25	AT1G09380	1	1	63	Plasmatic

Legend: SCN: Soybean Copy Number. %S: Similarity Percentage. STN: Soybean Transcript Number. ACN: Arabidopsis Copy Number. ATN: Arabidopsis Transcript Number. CL: Cellular localization; NA: Not Available.

organelles such as mitochondria, chloroplast, vacuoles, and plastids (Figure 4). Considering the high (≥62%) physicochemical similarity of orthologs between soybean and Arabidopsis, it is possible to assume that, in soybean, these NCTs have identical subcellular localization (Table 1). Subcellular localization gives us clues about the function of a particular NCT since the synthesis of most amino acids occurs in organelles such as plastids, mitochondria, peroxisomes, and cytosol (Rentsch et al. 2007). Therefore, knowledge of the subcellular localization of these NCTs can tell us which type of nitrogenous compound is transported by a particular NCT. Whereas transmembrane transporters mediate the intra- and intercellular translocation of amino acids over short and long distances (Yang et al. 2020), the genetic manipulation of amino acid transporters has the potential to improve several aspects of the plant, such as biomass, yield, and seed quality (Yang et al. 2020). In other words, the genetic manipulation of these transporters constitutes a promising strategy for the betterment of several crops.

Conclusion

The NCT-related genes *AAP7, AVT3, CAT9, UMAMIT25* and *UPS2* are present higher expression in the seeds of the cultivar with higher protein content (BRS 232), having a positive correlation with this parameter. However, the *BAT1* and *LHT1* genes are more expressed in cultivar BRS 284, suggesting that their expression is not directly related to the high protein content of seeds. In pods, the expression of NCT-related genes *AAP7, BAT1, UPS2, LHT1*, and *UMAMIT25* did not correlate positively with the higher protein content of cultivar BRS 232, suggesting that, in pod tissues, the activity of these NCTs may has a secondary role in the increase of the protein content of soybean seeds.

Considering that, in the seed, the expression of the majority of NCTs evaluated was higher in the cultivar with higher protein content (BRS 232), this result suggests that the genetic manipulation of these transporters can increase both protein content and quality in soybean seeds.

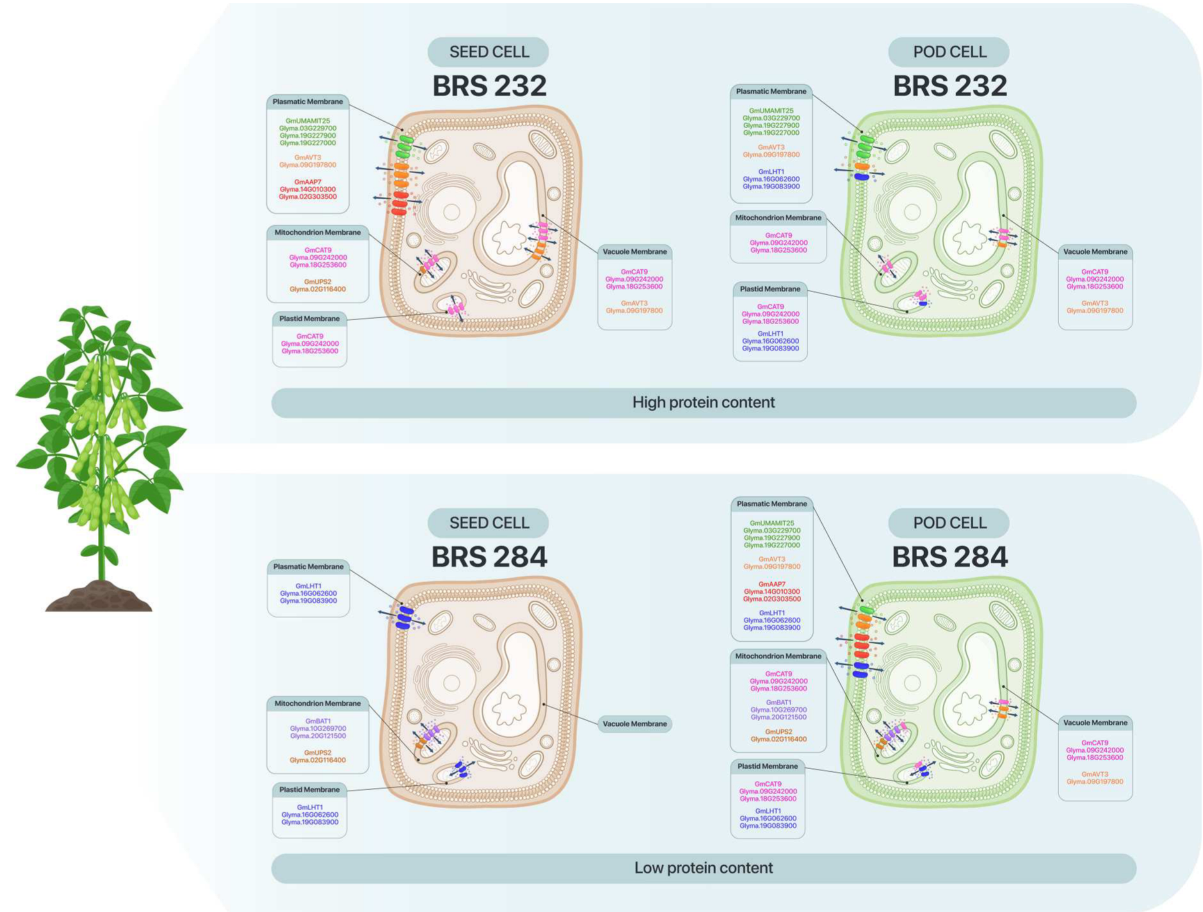

Figure 4. Subcellular localization of NCTs in seed and pod tissues of soybean cultivars with higher (BRS 232) and lower (BRS 284) seed protein content.

Author contributions

PILJ designed the experiments, collected, and analyzed all of the data and wrote the manuscript; MDCM performed *in silico* analysis, gene expression analysis, and helped writing the manuscript; SRRM e DAB performed gene expression analysis; FAH produced and made available to analysis the libraries from dataset 2; AJCV helped writing the manuscript and performed editing and linguistic revision; ELR performed editing and linguistic revision and obtained the funding; LMMH and ALN designed experiments, analyzed all of the data and obtained the funding.

Acknowledgments

The authors would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES) for granting a master's scholarship to PILJ and a Doctoral scholarship to DAB and Embrapa Soybean for granting a Postdoctoral Fellowship to MDCM.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by Empresa Brasileira de Pesquisa Agropecuária (Embrapa), INCT BioSyn (National Institute of Science and Technology in Synthetic Biology), CNPq (National Council for Scientific and Technological Development), CAPES (Coordination for the Improvement of Higher Education Personnel) and FAPDF (Research Support Foundation of the Federal District), Brazil.

Notes on contributors

Pamela I. L. Joaquim is graduated in Agronomy from the University of Western São Paulo (2015) and Master's in Biotechnology from State University of Londrina.

Mayla D. C. Molinari is graduated in Biomedicine from the State University of Londrina (2011) and Agronomy from the Philadelphia University Center (2015). Master's and Doctorate in Genetics and Molecular Biology with a focus on Biotechnology from the State University of Londrina. She was part of the doctorate at Plant Gene Expression Center in the United States. She is currently a postdoctoral researcher at Embrapa Soja at Plant Biotechnology Laboratory. She works in bioinformatics and genomic editing. Prospecting genes to the development of soybean cultivars with superior genetic characteristics.

Silvana R. R. Marin is graduated in Chemistry from the State University of Londrina (1994). Specialization in Applied Biochemistry, Master's and Doctorate in Genetics and Molecular Biology with a focus on Biotechnology from the State University of Londrina. She is currently an analyst senior at Embrapa Soja at Plant Biotechnology Laboratory. She has experience in plant molecular biology, genomics, transcriptomics, biochemistry, and plant physiology.

Daniel A. Barbosa is graduated in Biological Sciences from the State University of Northern Paraná (2016), Master's in Genetics and Molecular Biology from the State University of Londrina, and Ph.D. student in the Genetics and Molecular Biology at State University of Londrina. He has experience in plant molecular biology, entomology, plant physiology, experiments in greenhouse and field conditions, NBTs tools, also working on topics such as gene expression, abiotic stress (water deficit), and obtaining and characterization of genetically modified soybean plants.

Américo J. C. Viana is agronomist and Ph.D. in Genetics and Molecular Biology, have experience in research with genetics and plant molecular biology. Have solid knowledge in molecular biology approaches, such as plant genotyping, gene cloning, genetic transformation of plants (GMOs), Sanger and next-generation sequencing (NGS), determination of metabolites, etc. Have worked as a postdoctoral researcher and visiting professor at Unicamp and as a technician at the Mars Center for Cocoa Science/UESC. Currently participates in an Embrapa project in partnership with a private company to develop topical RNAi-based solutions for controlling corn and soybean pests.

Elibio L. Rech has pioneered studies to further understand the process of selecting germ-line-transformed leguminous plants by the use of recombinant DNA technology. He has elucidated how to select transgenic cells from the apical region by modulating the activity and derived the effect of specific molecules from developing a system for the production of high-frequency-transgenic leguminous plants. Over the last four decades, Rech is focused on basic and applied advances in science and technology. Rech's laboratory has engineered plants, opening new windows to the studies of basic cellular function and product development. Rech's current research activities include the development of complex genomes as a platform for editing and engineering, directed towards synthetic biology applications. Rech's current focus involves exploring the utilization of computational algorithms based on whole systems to simultaneously design, construct, and engineer functional synthetic genomes utilizing minimal cells of different pro/eukaryotic lineages. His techniques involve gene switches, genetic circuits, artificial chromosomes, and delivery of Mbp sequences to cells. Rech is applying his concept of the synthetic domestication of specific biodiversity traits as a foundation for conservation and a viable option to intensify and expand the utilization of recombinant DNA for food security and health care. In 2002, the President of the Federative Republic of Brazil awarded him the honorific title of 'Comendador' in the Order of Scientific Merit. In 2005, he was elected a Fellow of the Brazilian Academy of Sciences. In 2007, he was elected a Fellow of the World Academy of Sciences - for the Advancement of Science in Developing Countries (TWAS). In 2021, The President of the Superior Labor Court awarded him the honorific title of 'Comendador' in the Order of Scientific Merit. Elibio Rech is the Director of the National Institute of Science and Technology in Synthetic Biology.

Fernando A. Henning is graduated in Agronomy from the State University of Santa Catarina (2004), Master's and Doctorate in Seed Science and Technology from the Federal University of Pelotas. He is currently a researcher at Embrapa Soja. He works with Seed Science and Technology.

Alexandre L. Nepomuceno is graduated in Agronomy (1987) and Master's in Agronomy from the Federal University of Rio Grande do Sul, and Doctorate in Molecular Biology and Plant Physiology from University of Arkansas. Post-Doctorate at the Japan International Research Center for Agricultural Sciences (2000 and 2004). He is currently General *Head* and researcher at Embrapa Soja. Has experience in plant physiology and plant molecular biology, acting on the following subjects: gene expression, soybean physiology, Biosafety of GMO, and genetic engineering and gene editing (CRISPR technology).

Liliane M. Mertz-Henning is Graduated in Agronomy from the State University of Santa Catarina (2004), Master's and Doctorate in Seed Science and Technology with a focus on Biotechnology from the Federal University of Pelotas. She was Visiting Researcher at JIRCAS (Japan International Research Center for Agricultural Sciences) in 2013. She is currently a researcher at Embrapa Soja, a member of the Plant Biotechnology Laboratory. In 2019, she started to act as a Professor at the Postgraduate Program in Biotechnology at UEL (Universidade Estadual de Londrina). She is currently a member of the National Technical Commission on Biosafety (CTNBio). She works mainly in gene prospecting, functional validation, obtaining and characterization of genetically modified soybean plants, and gene editing via CRISPR for tolerance to abiotic stresses and the quality of seeds and grains.

References

Abirached-Darmency M, Abdel-gawwad M, Conejero G, Verdeil J, Thompson R. 2005. In situ expression of two storage protein genes in relation to histo-differentiation at mid-embryogenesis in medicago truncatula and pisum sativum seeds. J Exp Bot [Internet]. [accessed 2021 Sep 9];56(418):2019–2028. https://pubmed.ncbi.nlm.nih.gov/15983018/

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. FastQC a qual control tool high throughput seq

data [Internet]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Assefa Y, Purcell LC, Salmeron M, Naeve S, Casteel SN, Kovács P, Archontoulis S, Licht M, Below F, Kandel H, et al. 2019. Assessing variation in us soybean seed composition (protein and oil). Front Plant Sci. 10(March):1–13.

Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. 2014. Jvenn: an interactive venn diagram viewer. BMC Bioinforma 2014 151 [Internet]. [accessed 2021 Sep 6];15(1):1–7. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-293

Besnard J, Zhao C, Avice J-C, Vitha S, Hyodo A, Pilot G, Okumoto S. 2018. Arabidopsis umamit24 and 25 are amino acid exporters involved in seed loading. J Exp Bot [Internet]. [accessed 2021 Sep 6];69(21):5221–5232. https://academic.oup.com/jxb/article/69/21/5221/5078731

Bolger A, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics [Internet]. [accessed 2021 Sep 6];30(15):2114–2120. https://pubmed.ncbi.nlm.nih.gov/24695404/

Brasil. 2009. Regras para análise de sementes. Brasília, DF: Mapa/ACS, 399p. https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/arquivos-publicacoesinsumos/2946_regras_analise__sementes.pdf.

Carter A, Tegeder M. 2016. Increasing nitrogen fixation and seed development in soybean requires complex adjustments of nodule nitrogen metabolism and partitioning processes. Curr Biol [Internet]. [accessed 2021 Sep 6];26(15):2044–2051. https://pubmed.ncbi.nlm.nih.gov/27451897/

Castro RD, Bradford KJ, Hilhorst HWM. 2004. Embebição e reativação do metabolismo. In: Germinação do básico ao Apl. Porto Alegre: Artmed; p. 149–162.

Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. 2017. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. Plant J [Internet]. [accessed 2021 Sep 6];89(4):789–804. https://onlinelibrary.wiley.com/doi/full/10.1111/tpj.13415

Cheng L, Yuan H, Ren R, Zhao S, Han Y, Zhou Q, Ke D, Wang Y, Wang L. 2016. Genome-wide identification, classification, and expression analysis of amino acid transporter gene family in Glycine max. Front Plant Sci [Internet]. [accessed 2021 Sep 6];7(APR2016):1–14. https://pubmed.ncbi.nlm.nih.gov/27148336/.

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. 2016. A survey of best practices for RNA-seq data analysis. Genome Biol 2016 171 [Internet]. [accessed 2021 Sep 9];17(1):1–19. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0881-8

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R. 2005. Genome-wide identification and testing of superior reference genes for transcript normalization in arabidopsis. Plant Physiol. 139 (1):5–17.

EMBRAPA. 2014. Cultivares de soja. Londrina: Embrapa Soja. https://ainfo.cnptia.embrapa.br/digital/bitstream/item/114394/1/Cultivares-de-soja-macrorregioes-1-2-e-3-centro-sul-do-Brasil.pdf.

Franceschini A. 2017. Cadeia global de valor. Setor agronegócio: Soja [internet]. https://ri.espm.br/wp-content/uploads/2018/08/Setor-da-soja-1.pdf.

Fujiki Y, Teshima H, Kashiwao S, Kawano-Kawada M, Ohsumi Y, Kakinuma Y, Sekito T. 2017. Functional identification of AtAVT3, a family of vacuolar amino acid transporters, in arabidopsis. FEBS Lett [Internet]. [accessed 2021 Sep 6];591(1):5–15. https://pubmed.ncbi.nlm.nih.gov/27925655/

Grallath S, Weimar T, Meyer A, Gumy C, Suter-Grotemeyer M, Neuhaus J, Rentsch D. 2005. The AtProT family. compatible solute transporters with similar substrate specificity but differential expression patterns. Plant Physiol [Internet]. [accessed 2021 Sep 6];137(1):117–126. https://pubmed.ncbi.nlm.nih.gov/15618414/

Herman EM. 2014. Soybean seed proteome rebalancing. Front Plant Sci. 0(SEP):437.

Hwang E, Song Q, Jia G, Specht J, Hyten D, Costa J, Cregan P. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics [Internet]. [accessed 2021 Sep 6];15(1):1–12. https://pubmed.ncbi.nlm.nih.gov/24382143/.

Ji Y, Huang W, Wu B, Fang Z, Wang X. 2020. The amino acid transporter AAP1 mediates growth and grain yield by regulating neutral amino acid uptake and reallocation in oryza sativa. J Exp Bot [Internet]. [accessed 2021 Sep 6];71(16):4763–4777. https://academic.oup.com/jxb/article/71/16/4763/5850545

Jones SI, Vodkin LO. 2013. Using RNA-Seq to profile soybean seed development from fertilization to maturity. PLoS One [Internet]. [accessed 2021 Sep 6];8(3):e59270. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0059270

Jung C-H, Wong CE, Singh MB, Bhalla PL. 2012. Comparative genomic analysis of soybean flowering genes. PLoS One [Internet]. [accessed 2021 Sep 9];7(6):e38250. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0038250

Karmann J, Müller B, Hammes U. 2018. The long and winding road: transport pathways for amino acids in Arabidopsis seeds. Plant Reprod [Internet]. [accessed 2021 Sep 6];31(3):253–261. https://pubmed.ncbi.nlm.nih.gov/29549431/

Kim D, Langmead B, Salzberg S. 2015. Hisat: a fast spliced aligner with low memory requirements. Nat Methods [Internet]. [accessed 2021 Sep 6];12(4):357–360. https://pubmed.ncbi.nlm.nih.gov/25751142/

Kim D, Paggi J, Park C, Bennett C, Salzberg S. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol [Internet]. [accessed 2021 Sep 6];37(8):907–915. https://pubmed.ncbi.nlm.nih.gov/31375807/

Kim JA, Kim HS, Choi SH, Jang JY, Jeong MJ, Lee SI. 2017. The importance of the circadian clock in regulating plant metabolism. Int J Mol Sci. 18(12):1–11.

Lawrence-Dill C. 2019. GOMAP soybean JGI-Wm82.a4.v1. 1.0. CyVerse data commons. DOI:10.25739/59ec-1719 [Internet]. https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/Carolyn_Lawrence_Dill_GOMAP_Soybean_JGI-Wm82.a4.v1_April_2019.r1

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics [Internet]. [accessed 2021 Sep 6];25 (16):2078–2079. https://pubmed.ncbi.nlm.nih.gov/19505943/

Lin J-H, Xu Z-J, Peng J-S, Zhao J, Zhang G-B, Xie J, Yi Z-X, Zhang J-H, Gong J-M, Ye N-H, Meng S. 2019. Osprot1 and OsProT3 function to mediate proline- and γ-aminobutyric acid-specific transport in yeast and are differentially expressed in rice (oryza sativa L.). Rice 2019 121 [Internet]. [accessed 2021 Sep 6];12(1):1–10. https://link.springer.com/articles/10.1186/s12284-019-0341-7

Liu S, Wang D, Mei Y, Xia T, Xu W, Zhang Y, You X, Zhang X, Li L NNW. 2020. Overexpression of GmAAP6a enhances tolerance to low nitrogen and improves seed nitrogen status by optimizing amino acid partitioning in soybean. Plant Biotechnol J [Internet]. [accessed 2021 Sep 6];18(8):1749–1762. https://pubmed.ncbi.nlm.nih.gov/31945255/

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. 25:402–408. https://www.sciencedirect.com/science/article/abs/pii/S1046202301912629

Machado F, Moharana K, Almeida-Silva F, Gazara R, Pedrosa-Silva F, Coelho F, Grativol C, Venancio T. 2020. Systematic analysis of 1298 RNA-Seq samples and construction of a comprehensive soybean (Glycine max) expression atlas. Plant J [Internet]. [accessed 2021 Sep 6];103(5):1894–1909. https://pubmed.ncbi.nlm.nih.gov/32445587/

Metsalu T, Vilo J. 2015. Clustvis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. Nucleic Acids Res [Internet]. [accessed 2021 Sep 6];43(W1):W566–W570. https://pubmed.ncbi.nlm.nih.gov/25969447/

Michelfelder AJ. 2009. Soy: A complete source of protein. Am Fam Physician [Internet]. [accessed 2021 Sep 6];79(1):43–47. www.aafp.org/afp

Molinari MDC, Fuganti-Pagliarini R, Marcolino-Gomes J, de Amorim Barbosa D, Rockenbach Marin SR, Mertz-Henning LM, Rech Filho EL. 2021b. Flower and pod genes involved in soybean sensitivity to drought. Journal of Plant Interactions. 16(1):187–200. https://www.tandfonline.com/doi/full/10.1080/17429145.2021.1921293

Molinari MDC, Fuganti-Pagliarini R, Mendonça JA, Marin DR, Mertz-henning L, Nepomuceno AL. 2021a. Transcriptome Analysis Using RNA-Seq from Experiments with and Without Biological Replicates : A Review. 64:1–13.

Patel RK, Jain M. 2012. Ngs QC Toolkit: A Toolkit for quality control of next generation sequencing data. PLoS One [Internet]. [accessed 2021 Sep 6];7(2):e30619. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0030619

Pathan SM, Vuong T, Clark K, Lee JD, Grover Shannon J, Roberts CA, Ellersieck MR, Burton JW, Cregan PB, Hyten DL, et al. 2013. Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. Crop Sci. 53(3): 765–774.

Patil G, Mian R, Vuong T, Pantalone V, Song Q, Chen P, Shannon G, Carter T, Nguyen H. 2017. Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. Theor Appl Genet [Internet]. [accessed 2021 Sep 6];130(10):1975–1991. https://pubmed.ncbi.nlm.nih.gov/28801731/

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. Stringtie enables improved reconstruction of a transcriptome from RNA-Seq reads. Nat Biotechnol 2015 333 [Internet]. [accessed 2021 Sep 6];33(3):290–295. https://www.nature.com/articles/nbt.3122

Pípolo AE, Mandarino JMG. 2016. Os teores de proteina do soja e a qualidade para a indústria. Bol Inf da SBCS. 42(22):30–33.

Rentsch D, Schmidt S, Tegeder M. 2007. Transporters for uptake and allocation of organic nitrogen compounds in plants. FEBS Lett. 581(12):2281–2289.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, et al. 2010. Genome sequence of the palaeopolyploid soybean. Nat 2010 4637278 [Internet]. [accessed 2021 Sep 6];463(7278):178–183. https://www.nature.com/articles/nature08670

Seo JH, Kim KS, Ko JM, Choi MS, Kang BK, Kwon SW, Jun TH. 2019. Quantitative trait locus analysis for soybean (Glycine max) seed protein and oil concentrations using selected breeding populations. Plant Breed. 138(1):95–104.

Su Y-H, Frommer WB, Ludewig U. 2004. Molecular and Functional characterization of a family of amino acid transporters from arabidopsis. Plant Physiol [Internet]. [accessed 2021 Sep 6];136(2):3104. /pmc/articles/PMC523371/.

Tegeder M, Masclaux-Daubresse C. 2018. Source and sink mechanisms of nitrogen transport and use. New Phytol. 217(1):35–53.

Tegeder M, Rentsch D. 2010. Uptake and partitioning of amino acids and peptides. Mol Plant. 3(6):997–1011.

The SV, Snyder R, Tegeder M. 2020. Targeting nitrogen metabolism and transport processes to improve plant nitrogen Use efficiency. Front Plant Sci [Internet]. [accessed 2021 Sep 6];11:628366. /pmc/articles/PMC7957077/.

Thu SW, Lu M-Z, Carter AM, Collier R, Gandin A, Sitton CC, Tegeder M. 2020. Role of ureides in source-to-sink transport of photoassimilates in non-fixing soybean. J Exp Bot [Internet]. [accessed 2021 Sep 6];71(15):4495–4511. https://academic.oup.com/jxb/article/71/15/4495/5809926

USDA. 2021. World agricultural production [internet]. [Circular series: USDA]. https://apps.fas.usda.gov/psdonline/circulars/production.pdf

Wan Q, Chen S, Shan Z, Yang Z, Chen L, Zhang C, Yuan S, Hao Q, Zhang X, Qiu D, Chen H, Zhou X. 2017. Stability evaluation of reference genes for gene expression analysis by RT-qPCR in soybean under different conditions. PloS one [internet] [accessed 2021 Sep 6];12(12): e0189405. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0189405

Wang X, Yang G, Shi M, Hao D, Wei Q, Wang Z, Fu S, Su Y, Xia J. 2019. Disruption of an amino acid transporter LHT1 leads to growth inhibition and low yields in rice. BMC Plant Biol 2019 191 [Internet]. [accessed 2021 Sep 6];19(1):1–11. https://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-019-1885-9

Yang G, Wei Q, Huang H, Xia J. 2020. Amino acid transporters in plant cells: A brief review. PlantS 2020, Vol 9, Page 967 [Internet]. [accessed 2021 Sep 6];9(8):967. https://www.mdpi.com/2223-7747/9/8/967/htm

Zhou T, Yue C, Huang J, Cui J, Liu Y, Wang W, Tian C, Hua Y. 2020. 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 Biol 2020 201 [Internet]. [accessed 2021 Sep 6];20(1):1–22. https://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-020-02367-7