Soybean seed proteome rebalancing

Eliot M. Herman*

School of Plant Sciences, BIO5 Institute, University of Arizona, Tucson, AZ, USA

The soybean seed's protein content and composition are regulated by both genetics and physiology. Overt seed protein content is specified by the genotype's genetic framework and is selectable as a breeding trait. Within the genotype-specified protein content phenotype soybeans have the capacity to rebalance protein composition to create differing proteomes. Soybeans possess a relatively standardized proteome, but mutation or targeted engineering can induce large-scale proteome rebalancing. Proteome rebalancing shows that the output traits of seed content and composition result from two major types of regulation: genotype and post-transcriptional control of the proteome composition. Understanding the underlying mechanisms that specifies the seed proteome can enable engineering new phenotypes for the production of a high-quality plant protein source for food, feed, and industrial proteins.

Keywords: protein, proteome, seed, storage protein, soybean

SOYBEANS ARE A GLOBAL PROTEIN COMMODITY

Among the global commodity of crops, soybean has an almost unique role, being high enough in protein content to provide the nitrogen (N) needed for efficient large-scale animal feed production. Soybeans possess economically valued oil and protein and is an archetype seed used to dissect the processes that specify seed compositional output traits. Over the past decade, considerable public and industry funds have been invested to create soybean community resources, including genomic, transcript, SNP and SSR maps, proteomics, as well as supported a broad range of bioactivity, biochemical, nutritional, and agronomic projects.

PROTEIN CONTENT AS A GENOTYPE

The genome of soybean specifies the genetic framework for seed formation and maturation, and it controls the expression, mix, and timing for synthesis of the storage metabolite traits (Wilson, 2004 for general information; Hartwig and Kilen, 1991; Wilcox and Shibles, 2001; Chung et al., 2003; Nichols et al., 2006; Bolon et al., 2010). Plant breeding has shown that different soybean genotypes specify a standardized, often line/cultivar-specific protein content (Wilson, 2004). The genetic program that produces seeds is simultaneously manifested in the embryo, endosperm, and maternal plant. Genetic marker analysis, using SNPs, has identified QTLs that demonstrate the overt protein and oil content has a strong genetic determinant (see Diers et al., 1992; Cregan et al., 1999; Zhao-Ming et al., 2011 for examples); these traits have supported generations of breeders who have enhanced soybean as a crop (Brim and Burton, 1979; Carter et al., 1986; Cober and Voldeng, 2000; Wilson, 2004).

Of the three individual components that comprise seeds, two are reproductive progeny: the endosperm and the zygotic embryo, which result from the double fertilization and are enclosed in the maternal-origin seed coat that connects the maturing seed to the maternal plant. Each reproductive-phase soybean plant consists of a coordinated network of embryos and endosperms for the common goal of maximizing reproductive output. Historically, breeding programs have selected traits for enhancement of seed productivity, storage metabolite content, and important agronomic performance traits of the maternal plant. One way to view seed production by plants is as a population that produces (maternal) and distributes (maternal and endosperm) nutrient metabolites to the embryonic sink. During soybean seed development the endosperm undergoes progressive programmed cell death that is completed prior to the accumulation of stored metabolites. By the onset of protein and oil accumulation, only a single cell layer of aleurone remains from the endosperm that encapsulates the embryo, separating it from the inner surface of the maternal seed coat. But the physiological role, if any, of the aleurone in regulating nutrient flux to the developing embryo has not been investigated. Viewed in this way, the soybean plant's progeny are a population of aleurones and embryos interacting with the metabolite flux. The non-synchronized developing population of seeds must both compete and synchronize with the common maternal nutrient source. To assure the mature seeds are nearly all equivalent in composition, independent of their particular position on the maternal plant, their developmental program and physiological regulation must be coordinated along with the capacity of the maternal plant to nourish them.

The embryo's genotype specifies maturation-stages (Fehr et al., 1971) that are controlled by transcription factors that provide the developmental framework for storage substance accumulation (Hill and Breidenbach, 1974; Goldberg et al., 1981a,b; Mienke et al., 1981; Walling et al., 1986; Naito et al., 1988; Harada et al., 1989; Perez-Grau and Goldberg, 1989; Nielsen and Nam, 1999; To et al., 2006). A large number of soybean seed-specific DNA binding proteins have been identified (http://casp.rnet.missouri.edu/soydb/), and some of these have been shown to regulate specific seed maturation specific genes. (Chen et al., 1986; Jofuku et al., 1987; Allen et al., 1989; Lessard et al., 1991; Bäumlein et al., 1992; Kwong et al., 2003; Wang et al., 2007). A key role of some transcription factors is to regulate the metabolic and developmental processes that support

storage substance accumulation (Kroj et al., 2003; Gutierrez et al., 2007; Santos-Mendoza et al., 2008 for reviews). Understanding how cooperation between the embryo, endosperm, and maternal organs is integrated at the level of gene expression and cross-regulation of metabolism is important for creating models of the source-sink relationship of seed-fill.

NUTRIENT SOURCE DEFINES SEED PROTEIN ACCUMULATION

Whole plant physiological experiments demonstrate that nutrient distribution to seeds is highly regulated. The seed protein output trait is primarily regulated by controlling the composition of seeds, with the total number of seeds being a consequence of nutrient availability. The average size of seeds from small size plants compared with larger size plants, differs only a little, but the total number of seeds produced is directly related to the total biomass of the maternal plant and its mobilized metabolite source potential. In an agronomic context, this defines yield. From the perspective of the plant, the protein content genotype maximizes the potential of an individual seeds with the overall yield of seeds depending on the available biomass/growth conditions.

Although the maternal plant, endosperm, and embryo function in concert to form the seed, their metabolic interaction occurs without a direct, contiguous flow of nutrients, as each is apoplastically isolated from the other (Thorne, 1980; Thorne and Rainbird, 1983; Egli and Bruening, 2001; Patrick and Offler, 2001 for review). The metabolite flux from the maternal plant through the aleurone to the embryo results from coordinated secretion from the source and uptake by the sink, and this potentially determines the storage output trait (Borisjuk et al., 2003, 2004). In annual plants, such as soybean, the maternal plant must grow rapidly and produce nutrient-capture organs (i.e., roots and leaves).

The relationship between the maternal plant (source) and the seed (sink) has been investigated with increasingly more sophisticated tools and concepts for the past 40 years. For soybean, early studies focused on accumulation of vegetative proteins, primarily in foliage, as the nitrogen store that is mobilized to the seed and determines the accumulation of storage substances. This model of resource acquisition parallels that of most other seed plants, where metabolites in foliage are later mobilized to the seed. The amount of carbon fixed during photosynthesis is highly responsive to the environment, and the maternal plant manages carbon flux so it can distribute nutrients based on their availability and the demand of the seed (endosperm/embryo) sink (see Fellows et al., 1979; Borchers-Zampini et al., 1980 for early examples). A number of studies have shown that leaf proteins, predominantly Rubisco, accumulate over time (Schaefer et al., 1981a,b). In addition, soybean leaves accumulate a vegetative storage protein (VSP), a member of the vacuolar acid phosphatase family (DeWald et al., 1992; Staswick et al., 1994). VSP accumulation is highly responsive to nitrogen availability (Franceschi and Giaquinta, 1983; Franceschi et al., 1983; Staswick, 1989a,b), and it increases with depodding, i.e., removal of the seed sink. This observation led to proposals that VSP is a necessary adjunct that provides additional nitrogen resources for the seed. By silencing the VSP gene, it was later shown that VSP does not appear to make a difference in soybean seed protein content (Staswick et al., 2001).

Accumulated leaf proteins are mobilized by specific proteases (see Ragster and Chrispeels, 1979, 1981a,b) found in leaf cell vacuoles and plastids; the enzymes mediate the hydrolysis of Rubisco and VSP as well as other less abundant leaf proteins. Removal of all maturing seeds, except in one portion of the plant, leads to redistribution of source leaf nutrient flux (Carlson and Brun, 1984) indicating there must be (unknown) feedback-regulation between the seed and the nutrient source that is manifested through a long distance signal. Systems biology approaches could determine how the source size/composition is regulated, and how its mobilization is coordinated with the draw of the sink.

Leaf photosynthate and the products of protein hydrolysis produce a metabolite flux consisting predominantly of sucrose, glutamine and asparagine (Hsu et al., 1984; Rainbird et al., 1984; Krishnan et al., 2011); for nitrogen-fixing legumes in particular, there are also ureides from xylem fluid. With respect to the amino acid flux, the input from glutamine and asparagine has different characteristics. Skokut et al. (1982), using ^{15}N-NMR showed that there is no discrimination between the amino and amide N of glutamine, but for Asn the amino N is incorporated into protein twice as efficiently as the amide N, indicating a key role of asparagine in transamination. Asn may also be directly incorporated into proteins, with dual labeled ^{13}C and ^{15}N Asn being incorporated directly without scrambling the labels (Schaefer et al., 1981a,b). Within the seed, free Asn accounts for a larger fraction of amino acids (33–49%), with the fractional amount varying by genotype (Schaefer et al., 1981a,b). Asn dominates seed coat free amino acid efflux, assayed as apoplastic fluid, from the seed cup, i.e., the seed coat with embryo removed (Gifford and Thorne, 1985; Murray, 1987; De Jong et al., 1997; Hernández-Sebastià et al., 2005; Pandurangan et al., 2012). It is unclear whether the tissue source used for these experiments was derived from the inner surface of the aleurone or the maternal seed coat, or a combination of both, since the aleurone often adheres to the inner side of the seed coat. This shows that distinct from the embryo, the seed coat (perhaps comprising aleurone and seed coat) has an amino acid composition similar to the source (assayed as xylem sap), containing Asn as the dominant N-source, and is about 10-fold higher in abundance than Gln (Thorne and Rainbird, 1983; Lohaus et al., 1998; Zhang et al., 2010; Krishnan et al., 2011; Pandurangan et al., 2012).

PROTEIN COMPOSITION PLASTICITY AND THE SEED PROTEIN CONTENT GENOTYPE

The soybean seed protein output trait has two primary components: total protein content and the composition of individual proteins (the proteome). For soybeans, like many other seeds (Herman and Larkins, 1999 for review), the two major storage proteins, glycinin (11S legumin type) and conglycinin (7S vicilin type), dominate the proteome. The soybean seed proteome also includes many moderately abundant proteins that are bioactive and allergenic, such as the Kunitz and Bowman-Birk trypsin inhibitors, lectin, P34 allergen, sucrose binding protein, urease, oleosins (Herman and Burks, 2011) and several thousand low abundance proteins, including enzymes that mediate metabolism, synthesize storage substances, and create the structural framework of the cell. The specific mix of proteins and each protein's abundance

within the proteome determines the total amino acid composition trait.

Since the development of plant transformation techniques, there have been many attempts to express genes to induce accumulation of large quantities of foreign proteins in seeds. The goals of these projects were often to alter the nutritional quality of seeds, by increasing essential amino acids such as methionine in soybean, or to use the seeds as protein bioreactors. Even with strong storage protein promoters to regulate the transgene expression, the most frequent outcome of such experiments was to produce relatively small amounts of the heterologous protein (about 1% of total) and for the protein and amino acid composition to be little altered compared to control. These minor composition changes occurred whether the protein was targeted to the vacuole or accreted in ER-derived protein bodies, suggesting the limit of protein accumulation is independent of its deposition site. The observed heterologous protein production, compared to the expectations of the engineering design, indicated there must be seed regulatory mechanisms that limit foreign protein production so as not to significantly alter the seed protein content phenotype.

The converse experiment is to silence the intrinsic major storage protein genes and assess the impact on the seed's protein content. Kinney et al. (2001) showed total soybean protein content was conserved after silencing conglycinin, which constitutes about 20% of the total seed protein. The resulting seeds accumulated more glycinin, which apparently compensated for the missing conglycinin. Of the five glycinin-genes in the soybean genome, the protein encoded by the glycinin A4 gene tends to accrete in the ER, producing protein (ER)-bodies (Herman and Schmidt, 2004; Herman, 2008), which are not normally found in soybean (Kinney et al., 2001). Mori et al. (2004) showed similar observations in a conglycinin mutant obtained from screening a collection that exhibited the same phenotype of "glycinin rebalancing," and for some of this protein to remain as proglycinin and to accumulate ER-derived protein bodies. Schmidt and Herman (2008) showed that introducing a gene encoding a foreign protein into the conglycinin-glycinin rebalancing increased the accumulation of the heterologous protein. A GFP-HDEL gene, as a glycinin-gene mimic allele with a glycinin promoter and terminator, was constructed. The addition of the HDEL ER-retention sequence was intended to promote accretion of the protein in the ER to form ER-(protein)-bodies and mimic the accreting glycinin ORF by substituting the GFP-HDEL protein. The expression of GFP-HDEL in the standard cv *Jack* soybean background resulted in about 1% of the total seed protein accumulated as GFP in protein bodies, a level typical of experiments of heterologous protein production in seeds. However, introgressing the GFP-HDEL glycinin mimic allele into the conglycinin-silenced line resulted in about eightfold increased accumulation of GFP as the glycinin mimic was utilized to compensate for the conglycinin shortfall (**Figure 1**).

To further test regulation of the protein content genotype and its capacity to allow for proteome alterations, RNAi silencing of

FIGURE 1 | Variations of seed composition within the context of the protein content genotype are shown. The pie charts show the relative distribution of proteins comprising the seed proteome of the standard soybean line Jack, a line that silences β-conglycinin α,α′, another line that silences both β-conglycinin and glycinin, and a line where GFP-HDEL forming ER-derived protein bodies mimic of glycinin in the background of β-conglycinin silencing. Each of these seeds contains about the same total protein content while the proteome of each of these seeds varies. This demonstrates how the genomic framework controls the protein content genotype that can be rebalanced by protein composition plasticity.

both the glycinin and conglycinin storage proteins (*SP-* or storage protein minus) lines were created that eliminated over two thirds of the protein content of standard soybean seeds (Schmidt et al., 2011). *SP-* seeds exhibited a number of different phenotypes that included conserving the same seed protein content as the parental line, due to compensatory increases of other vacuolar proteins, including Kunitz Trypsin Inhibitor, Lectin, P34, and sucrose binding protein (**Figure 1**). These compensating proteins accumulated at levels up to 11X more than normal, and each protein's increase occurred without a parallel increase in its steady-state RNA transcript abundance. This suggests that the protein content trait is determined by genotype but the abundance of proteome members occurs at the translational level.

FREE ASN IS AN INDICATOR OF ALTERED PROTEIN CONTENT AND COMPOSITION

Free Asn is correlated with soybean seed protein content and composition. One line of experimental evidence has shown that in soybean seed cultivars with a range of protein contents there is a positive correlation between high free Asn and high protein content (Hernández-Sebastià et al., 2005; Pandurangan et al., 2012). For the *SP-* storage protein silenced soybean, free Asn increased 5.8X over the standard type (Schmidt et al., 2011). Perhaps in response to the elevated free Asn, the steady-state transcript (RNAseq) level for asparaginase increased 6.5X over the conventional controls. In standard lines the asparaginase level was previously correlated with protein content in standard lines (Wan et al., 2006). Together, these observations suggest there is a correlation of protein increase, whether by genotype selection for higher protein content or by increased abundance of individual proteins within the context of the protein genotype, with changes in free Asn and asparaginase. This suggests the free Asn level is a nitrogen status indicator (Miller et al., 2008), either as a regulator of, or as a component of, the processes that specify protein content and composition.

CULTURED SOMATIC AND ZYGOTIC EMBRYOS EXHIBIT AN EXCESSIVE GROWTH TRAIT

Ex vivo zygotic embryo and somatic embryo cultures are often used as proxies for seed maturation; however, there are significant differences in metabolic behavior of embryos that form *in vivo* and *in vitro* (see Thompson et al., 1977; Obendorf et al., 1983, 1984; Raper et al., 1984; Finer, 1988; Hayati et al., 1996; Santarem et al., 1997; Chanprame et al., 1998; Pipolo et al., 2004; Iyer et al., 2008; Nishizawa and Ishimoto, 2009; Allen and Young, 2013). RNA expression profiling showed that somatic embryos produce a relatively standard set of seed-specific transcripts (Thibaud-Nissen et al., 2003). *Ex vivo* cultures exhibited fidelity with *in planta* seeds, but exhibit differences in the content of accumulated reserve substances (Pipolo et al., 2004). Gln has been shown to be an effective N-input source for these cultures (Saravitz and Raper, 1995; Schmidt et al., 2005) and is often used as the experimental N-source in nutrition-flux studies (He et al., 2011; Allen and Young, 2013; Truong et al., 2013, for recent examples), even though it is Asn that accounts for the large majority of the actual maternal source N *in planta* (Lea and Miflin, 1980; Lohaus et al., 1998; Lima and Sodek, 2003). A recent paper by Allen and Young (2013)

showed in cultured zygotic soybean embryos that ^{14}C-Gln supplied 36–46% of the carbon of amino acids. In another study using somatic embryos, Truong et al. (2013) showed that increasing Gln in extrinsic culture media resulted in increased protein content, without greater oil content, showing that Gln is preferentially used to synthesize protein. This is consistent with older NMR observations on ^{13}C and ^{15}N that showed the amino and amido N for Gln as well as the carbon, is non-discriminatory when incorporated into the protein sink (Schaefer et al., 1981b; Skokut et al., 1982).

Taken together, these observations support a model where the maternal source supplies Asn (Pandurangan et al., 2012) as the N-source for zygotic embryos, but experimental *ex vivo* embryos can effectively use Gln. The difference between Asn and Gln may be important in the context of the morphological and compositional differences between *in planta* zygotic and cultured somatic and zygotic embryos. The media used for soybean culture varies, although Gln as the N-source dominates (Haga and Sodek, 1987), particularly in SHaM media (Schmidt et al., 2005), which was developed for transgenic embryo maturation (see Schmidt and Herman, 2008; Schmidt et al., 2011 for examples of its use). It is also used in some *ex vivo* nutritional studies (Truong et al., 2013). The media used for immature somatic embryo culture and transformation (Finer, 1988; Finer and McMullen, 1991; Walker and Parrott, 2001) has Asn as the N source. Tissue culture embryos used for transformation and regeneration, freed from the physical and metabolic constraints of the endosperm/aleurone and seed coat, exhibit aberrant growth (**Figures 2 and 3**), supporting

FIGURE 2 | The comparison of the size and morphology of a mature seed and a mature somatic embryo is shown. Note that the somatic embryo can exceed the seed's size, and the axis is enlarged compared to the cotyledon.
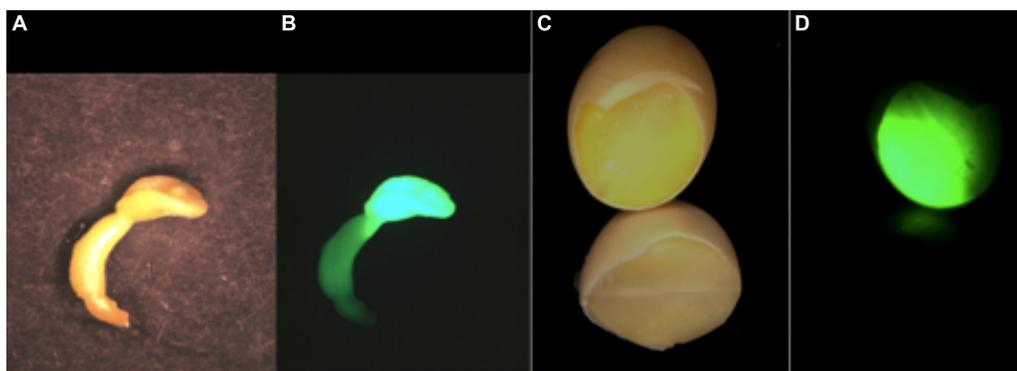
FIGURE 3 | A comparison of the expression of the storage protein trait in somatic embryos and seeds is shown as a GFP-HDEL glycinin mimic allele expressed as a storage protein proxy. The somatic embryo in white light **(A)** and UV fluorescence **(B)** shows that the primary GFP expression site is in the cotyledon that is reduced in size compared to cotyledons of seed. In seeds the cotyledons comprise the large majority of the seed's mass shown as a chipped seed illuminated with white light **(C)** and UV for GFP fluorescence **(D)**. The comparative expression the same glycinin promoter-regulated GFP in somatic embryos and seeds shows that while somatic embryos are a good proxy for seed expression there are challenges in interpreting somatic embryo results as an accurate proxy for a seed's in planta protein content and composition.

a regulatory role for the endosperm and/or seed coat in seed development (Garcia et al., 2003, 2005; Berger et al., 2006). In culture, somatic embryos form "monster" embryos with an enlarged embryonic axis and diminished, sometimes fused, cotyledons. Somatic embryos grown in the SHaM media are deemed "healthy," (i.e., large, green, well-formed), often exceeding in size a fully formed seed. Similar observations are obtained by culturing immature zygotic cotyledons that enlarge to a size that exceeds that of *in planta* seed cotyledons. This suggests the more an embryo is fed, the larger it grows, even beyond the size in a standard seed. Cultured embryos favor the accumulation of carbonaceous over nitrogenous metabolites, yielding less protein per mass than zygotic embryos. For *ex vivo* zygotic and somatic embryos, the genotype-specific protein content and its proteome phenotype appears to be less regulated, and instead the storage substance accumulation appears to have a direct relationship with nutrient input. The differences between *in planta* and *ex vivo* embryo development and storage substance accumulation indicates the significance of the *in planta* circumstance of each seed as an interactive member of a larger population.

SEED PROTEIN CONTENT AND ITS VARYING PROTEOME

For seed crops, of which soybean is a prominent example, historic and modern breeding has selected for enhanced storage substance accumulation. Generations of breeders have established that protein content is a genetically determined trait that can be selected. How protein content is regulated in relationship with protein composition appears to be a multilevel process, with the genotype establishing protein content. The rebalancing of proteome occurs in both dicots and monocots as shown by observations on soybean and maize (Wu and Messing, 2014). From the perspective of the seed that has to establish the next generation of an annual plant, its capacity to make compositional choices in response to altered metabolic circumstances has selective advantages. Understanding the processes that control proteome plasticity within the context of the protein content phenotype can enable biotechnologists to create enhanced soybeans optimized for specific end uses, such as species-specific feed or as protein bioreactors.

REFERENCES
Allen, D. K., and Young, J. D. (2013). Carbon and nitrogen provisions alter the metabolic flux in developing soybean embryos. *Plant Physiol.* 161, 1458–1475. doi: 10.1104/pp.112.203299

Allen, R. D., Bernier, F., Lessard, P. A., and Beachy, R. N. (1989). Nuclear factors interact with a soybean beta-conglycinin enhancer. *Plant Cell* 1, 623–631.

Bäumlein, H., Nagy, I., Villarroel, R., Inzé, D., and Wobus, U. (1992). Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *Plant J.* 2, 233–239.

Berger, F., Grini, P. E., and Schnittger, A. (2006). Endosperm: an integrator of seed growth and development. *Curr. Opin. Plant Biol.* 9, 664–670. doi: 10.1016/j.pbi.2006.09.015

Bolon, Y. T., Joseph, B., Cannon, S. B., Granham, M. A., Diers, B. W., Farmer, A. D., et al. (2010). Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. *BMC Plant Biol.* 10:41. doi: 10.1186/1471-2229-10-41

Borchers-Zampini, C., Glamm, A. B., Hoddinott, J., and Swanson, C. A. (1980). Alterations in source-sink patterns by modifications of source strength. *Plant Physiol.* 65, 1116–1120. doi: 10.1104/pp.65.6.1116

Borisjuk, L., Rolletschek, H., Radchuk, R., Weschke, W., Wobus, U., and Weber, H. (2004). Seed development and differentiation: a role for metabolic regulation. *Plant Biol. (Stuttg.)* 6, 375–386. doi: 10.1055/s-2004-817908

Borisjuk, L., Rolletschek, H., Wobus, U., and Weber, H. (2003). Differentiation of legume cotyledons as related to metabolic gradients and assimilate transport into seeds. *J. Exp. Bot.* 54, 503–512. doi: 10.1093/jxb/erg051

Brim, C. A., and Burton, J. W. (1979). Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. *Crop Sci.* 19, 494–498. doi: 10.2135/cropsci1979.0011183X001900040016x

Carlson, D. R., and Brun, W. A. (1984). Alteration of C-assimilate partitioning in leaves of soybeans having increased reproductive loads at one node. *Plant Physiol.* 75, 887–890. doi: 10.1104/pp.75.4.887

Carter, T. E., Burton, J. W., and Brim, C. A. (1986). Registration of NC101 to NC112 soybean germplasm lines contrasting in percent seed protein. *Crop Sci.* 26, 841–842. doi: 10.2135/cropsci1986.0011183X002600040057x

Chanprame, S., Kuo, T. M., and Widholm, J. M. (1998). Soluble carbohydrate content of soybean [*Glycine max* (L.) Merr.] somatic and zygotic embryos during development. *In Vitro Cell. Dev. Biol. Plant* 34, 64–68. doi: 10.1007/BF02823125

Chen, Z. L., Schuler, M. A., and Beachy, R. N. (1986). Functional analysis of regulatory elements in a plant embryo-specific gene. *Proc. Natl. Acad. Sci. U.S.A.* 83, 8560–8564. doi: 10.1073/pnas.83.22.8560

Chung, J., Babka, H. L., Graef, G. L., Staswick, P. E., Lee, D. J., Cregan, P. B., et al. (2003). The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci.* 43, 1053–1067. doi: 10.2135/cropsci2003.1053

Cober, E. R., and Voldeng, H. D. (2000). Developing high-protein, high-yield soybean populations and lines. *Crop Sci.* 40, 39–42. doi: 10.2135/cropsci2000.40139x

Cregan, P. B., Jarvik, T., Bush, A. L., Schoemaker, R. C., Lark, K. G., Kahler, A. L., et al. (1999). An integrated genetic linkage map of the soybean genome. *Crop Sci.* 39, 1464–1490. doi: 10.2135/cropsci1999.3951464x

De Jong, A., Koerselman-Kooij, J. W., Schuurmans, J. A. M. J., and Borstlap, A. C. (1997). The mechanism of amino acid efflux from seed coats of developing pea seeds as revealed by uptake experiments. *Plant Physiol.* 114, 731–736.

DeWald, D. B., Mason, H. S., and Mullet, J. E. (1992). The soybean vegetative storage proteins VSP alpha and VSP beta are acid phosphatases active on polyphosphates. *J. Biol. Chem.* 267, 15958–15964.

Diers, B. W., Keim, P., Fehr, W. R., and Shoemaker, R. C. (1992). RFLP analysis of soybean seed protein and oil content. *Theor. Appl. Genet.* 83, 608–612. doi: 10.1007/BF00226905

Egli, D. B., and Bruening, W. P. (2001). Source–sink relationship, seed sucrose levels and seed growth rates in soybean. *Ann. Bot.* 88, 235–242. doi: 10.1006/anbo.2001.1449

Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybean, *Glycine max* (L.) Merril. *Crop Sci.* 11, 929–931. doi: 10.2135/cropsci1971.0011183X001100060051x

Fellows, R. J., Egli, D. B., and Leggett, J. E. (1979). Rapid changes in translocation patterns in soybeans following source-sink alterations. *Plant Physiol.* 64, 652–655. doi: 10.1104/pp.64.4.652

Finer, J. J. (1988). Apical proliferation of embryonic tissue of soybean [*Glycine max* (L.) Merrill]. *Plant Cell Rep.* 7, 238–241. doi: 10.1007/BF00272532

Finer, J. J., and McMullen, M. D. (1991). Transformation of soybean via particle bombardment of embryogenic suspension culture tissue. *In Vitro Cell. Dev. Biol. Plant* 27, 175–182. doi: 10.1007/BF02632213

Franceschi, V. R., and Giaquinta, R. T. (1983). Specialized cellular arrangements in legume leaves in relation to assimilate transport and compartmentation: comparison of the paraveinal mesophyll. *Planta* 159, 415–422. doi: 10.1007/BF00392077

Franceschi, V. R., Wittenbach, V. A., and Giaquinta, R. T. (1983). A paraveinal mesophyll of soybean leaves in relation to assimilate transfer and compartmentation III. Immunohistochemical localization of specific glycopeptides in the vacuole after depodding. *Plant Physiol.* 72, 586–589. doi: 10.1104/pp.72.2.586

Garcia, D., Fitz Gerald, J. N., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis. Plant Cell* 17, 52–60. doi: 10.1105/tpc.104.027136

Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jurgens, G., and Berger, F. (2003). *Arabidopsis* haiku mutants reveal new controls of seed size by endosperm. *Plant Physiol.* 131, 1661–1670. doi: 10.1104/pp.102.018762

Gifford, R. M., and Thorne, J. H. (1985). Sucrose concentration at the apoplastic interface between seed coat and cotyledons of developing soybean seeds. *Plant Physiol.* 77, 863–868. doi: 10.1104/pp.77.4.863

Goldberg, R. B., Hoschek, G., Ditta, G. S., and Breidenbach, R. W. (1981a). Developmental regulation of cloned superabundant embryo mRNAs in soybean. *Dev. Biol.* 83, 218–231. doi: 10.1016/0012-1606(81)90468-1

Goldberg, R. B., Hoschek, G., Tam, S. H., Ditta, G. S., and Breidenbach, R. W. (1981b). Abundance, diversity, and regulation of mRNA sequence sets in soybean embryogenesis. *Dev. Biol.* 83, 201–217. doi: 10.1016/0012-1606(81)90467-X

Gutierrez, L., Van Wuytswinkel, O., Castelain, M., and Bellini, C. (2007). Combined networks regulating seed maturation. *Trends Plant Sci.* 12, 294–300. doi: 10.1016/j.tplants.2007.06.003

Haga, K. I., and Sodek, L. (1987). Utilization of nitrogen sources by immature soybean cotyledons in culture. *Ann. Bot.* 59, 597–601.

Harada, J. J., Barker, S. J., and Goldberg, R. B. (1989). Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell* 1, 415–425. doi: 10.1105/tpc.1.4.415

Hartwig, E. E., and Kilen, T. C. (1991). Yield and composition of soybean seed from parents with different protein, similar yield. *Crop Sci.* 31, 290–292. doi: 10.2135/cropsci1991.0011183X003100020011x

Hayati, R., Egli, D. B., and Crafts-Brandner, S. J. (1996). Independence of nitrogen supply and seed growth in soybean: studies using an in vitro culture system. *J. Exp. Bot.* 47, 33–40. doi: 10.1093/jxb/47.1.33

He, Y., Young, T., Clark, K., Kleppinger-Sparace, K., Bridges, W., and Sparace, S. (2011). Developmental profile of storage reserve accumulation in soybean somatic embryos. *In Vitro Cell. Dev. Biol. Plant* 47, 725–733. doi: 10.1007/s11627-011-9375-2

Herman, E. M. (2008). Endoplasmic reticulum bodies: Solving the insoluble. *Curr. Opin. Plant Biol.* 11, 672–679. doi: 10.1016/j.pbi.2008.08.004

Herman, E. M., and Burks, A. W. (2011). The impact of plant biotechnology on food allergy. *Curr. Opin. Biotechnol.* 22, 224–230. doi: 10.1016/j.copbio.2010.11.003

Herman, E. M., and Larkins, B. A. (1999). Protein storage bodies and vacuoles. *Plant Cell* 11, 601–614. doi: 10.1105/tpc.11.4.601

Herman, E. M., and Schmidt, M. A. (2004). ER to vacuole trafficking (ERvt) provides an alternate pathway for endomembrane progression for transfer to the vacuole. *Plant Physiol.* 136, 3440–3446. doi: 10.1104/pp.104.051722

Hernández-Sebastià, C., Marsolais, F., Saravitz, C., Israel, D., Dewey, R. E., and Huber, S. C. (2005). 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. *J. Exp. Bot.* 56, 1951–1963. doi: 10.1093/jxb/eri191

Hill, J. E., and Breidenbach, R. W. (1974). Proteins of soybean seeds. II. Accumulation of the major protein components during seed development and maturation. *Plant Physiol.* 53, 747–751. doi: 10.1104/pp.53.5.747

Hsu, F. C., Bennett, A. B., and Spanswick, R. M. (1984). Concentrations of sucrose and nitrogenous compounds in the apoplast of developing soybean seed coats and embryos. *Plant Physiol.* 75, 181–186. doi: 10.1104/pp.75.1.181

Iyer, V. V., Sriram, G., Fulton, D. B., Zhou, R., Westgate, M. E., and Shanks, J. V. (2008). Metabolic flux maps comparing the effect of temperature on protein and oil biosynthesis in developing soybean cotyledons. *Plant Cell Environ.* 31, 506–517. doi: 10.1111/j.1365-3040.2008.01781.x

Jofuku, K. D., Okamuro, J. K., and Goldberg, R. B. (1987). Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. *Nature* 328, 734–737. doi: 10.1038/328734a0

Kinney, A. J., Jung, R., and Herman, E. M. (2001). Cosuppression of the α-subunits of β-conglycinin in transgenic soybean seeds induces the formation of endoplasmic reticulum-derived protein bodies. *Plant Cell* 13, 1165–1178. doi: 10.1105/tpc.13.5.1165

Krishnan, H. B., Natarian, S. S., Bennett, J. O., and Sicher, R. C. (2011). Protein and metabolite composition of xylem sap from field-grown soybeans (*Glycine max*). *Planta* 233, 921–931. doi: 10.1007/s00425-011-1352-9

Kroj, T., Savino, G., Valon, C., Giraudat, J., and Parcy, F. (2003). Regulation of storage protein gene expression in *Arabidopsis. Development* 130, 6065–6073. doi: 10.1242/dev.00814

Kwong, R. W., Bui, A. Q., Lee, H., Kwong, L. W., Fischer, R. L., Goldberg, R. B., et al. (2003). LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* 15, 5–18. doi: 10.1105/tpc.006973

Lea, P. J., and Miflin, B. J. (1980). "Transport and metabolism of asparagine and other nitrogen compounds within the plant," in *The Biochemistry of Plant*, eds P. K. Stumpf and E. E. Conn (New York, NY: Academic Press), 569–607.

Lessard, P. A., Allen, R. D., Bernier, F., Crispino, J. D., Fujiwara, T., and Beachy, R. N. (1991). Multiple nuclear factors interact with upstream sequences of differentially regulated beta-conglycinin genes. *Plant Mol. Biol.* 16, 397–413. doi: 10.1007/BF00023991

Lima, J. D., and Sodek, L. (2003). N-stress alters aspartate and asparagine levels of xylem sap in soybean. *Plant Sci.* 165, 649–656. doi: 10.1016/S0168-9452(03)00251-6

Lohaus, G., Buker, M., Hussmann, M., Soave, C., and Heldt, H. W. (1998). Transport of amino acids with special emphasis on the synthesis and transport of asparagine in the Illinois Low Protein and Illinois High Protein strains of maize. *Planta* 205, 181–188. doi: 10.1007/s004250050310

Mienke, D. W., Chen, J., and Beachy, R. N. (1981). Expression of storage-protein genes during soybean seed development. *Planta* 153, 130–139. doi: 10.1007/BF00384094

Miller, A. J., Fan, X., Shen, Q., and Smith, S. J. (2008). Amino acids and nitrate as signals for the regulation of nitrogen acquisition. *J. Exp. Bot.* 59, 111–119. doi: 10.1093/jxb/erm208

Mori, T., Maruyama, N., Nishizawa, K., Higasa, T., Yagasaki, K., Ishimoto, M., et al. (2004). The composition of newly synthesized proteins in the endoplasmic reticulum determines the transport pathways of soybean seed storage proteins. *Plant J.* 40, 238–249. doi: 10.1111/j.1365-313X.2004.02204.x

Murray, D. R. (1987). Nutritive role of seed coats in developing legume seeds. *Am. J. Bot.* 74, 1122–1137. doi: 10.2307/2443954

Naito, S., Dube, P. H., and Beachy, R. N. (1988). Differential expression of con-glycinin alpha-subunit and beta-subunit genes in transgenic plants. *Plant Mol. Biol.* 11, 109–123. doi: 10.1007/BF00015664

Nichols, D. M., Glover, K. D., Carlson, S. R., Specht, J. E., and Diers, B. W. (2006). Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. *Crop Sci.* 46, 834–839. doi: 10.2135/cropsci2005.05-0168

Nielsen, N. C., and Nam, Y. W. (1999). "Soybean globulins," in *Seed Proteins,* eds P. R. Shewry and R. Casey (Dordrecht: Kluwer Academic Publishers), 285–313. doi: 10.1007/978-94-011-4431-5_13

Nishizawa, K., and Ishimoto, M. (2009). Maturation of somatic embryos as a model for soybean seed development. *Plant Biotechnol.* 26, 543–550. doi: 10.5511/plantbiotechnology.26.543

Obendorf, R. L., Rytko, G. T., and Byrne, M. C. (1983). Soya bean seed growth and maturation by in vitro pod culture. *Ann. Bot.* 51, 217–227.

Obendorf, R. L., Timpo, E. E., Byrne, M. C., Toai, T. V., Rytko, G. T., Hsu, F. C., et al. (1984). Soya bean seed growth and maturation in vitro without pods. *Ann. Bot.* 53, 853–863.

Pandurangan, S., Pajak, A., Molnar, S. J., Cober, E. R., Dhaubhadel, S., Hernandez-Sebastia, C., et al. (2012). Relationship between asparagine metabolism and protein concentration in soybean seed. *J. Exp. Bot.* 63, 3173–3184. doi: 10.1093/jxb/ers039

Patrick, J. W., and Offler, C. E. (2001). Compartmentation of transport and transfer events in developing seeds. *J. Exp. Bot.* 356, 551–564. doi: 10.1093/jexbot/52.356.551

Perez-Grau, L., and Goldberg, R. B. (1989). Soybean seed protein genes are regulated spatially during embryogenesis. *Plant Cell* 1, 1095–1109. doi: 10.1105/tpc.1.11.1095

Pipolo, A. E., Sinclair, T. R., and Camara, G. M. S. (2004). Protein and oil concentration of soybean seed cultured in vitro using nutrient solutions of differing glutamine concentration. *Ann. Appl. Biol.* 144, 223–227. doi: 10.1111/j.1744-7348.2004.tb00337.x

Ragster, L. E., and Chrispeels, M. J. (1981a). Autodigestion in crude extracts of soybean leaves and isolated chloroplasts as a measure of proteolytic activity. *Plant Physiol.* 67, 104–109. doi: 10.1104/pp.67.1.104

Ragster, L. V., and Chrispeels, M. J. (1981b). Hemoglobin-digesting Acid Proteinases in Soybean Leaves: characteristics and changes during leaf maturation and senescence. *Plant Physiol.* 67, 110–114. doi: 10.1104/pp.67.1.110

Ragster, L. V., and Chrispeels, M. J. (1979). Azocoll-digesting Proteinases in Soybean Leaves: characteristics and changes during Leaf Maturation and Senescence. *Plant Physiol.* 64, 857–862. doi: 10.1104/pp.64.5.857

Rainbird, R. M., Thorne, J. H., and Hardy, R. W. F. (1984). Role of amides, amino acids, and ureides in the nutrition of developing soybean seeds. *Plant Physiol.* 74, 329–334. doi: 10.1104/pp.74.2.329

Raper, C. D., Patterson, R. P., List, M. L., Obendorf, R. L., and Downs, R. J. (1984). Photoperiod effects on growth rate of in vitro cultured soybean embryos. *Bot. Gaz.* 145, 157–162. doi: 10.1086/337441

Santarem, E. R., Pelissier, B., and Finer, J. J. (1997). Effect of explant orientation, pH, solidifying agent and wounding on initiation of soybean somatic embryos. *In Vitro Cell. Dev. Biol. Plant* 33, 13–19. doi: 10.1007/s11627-997-0034-6

Santos-Mendoza, M., Dubreucq, B., Baud, S., Parcy, F., Caboche, M., and Lepiniec, L. (2008). Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *Plant J.* 54, 608–620. doi: 10.1111/j.1365-313X.2008.03461.x

Saravitz, C. H., and Raper, C. D. (1995). Responses to sucrose and glutamine by soybean embryos grown in vitro. *Physiol. Plant.* 93, 799–805. doi: 10.1111/j.1399-3054.1995.tb05134.x

Schaefer, J., Skokut, T. A., Stejskal, E. O., McKay, R. A., and Varner, J. E. (1981a). Estimation of protein turnover in soybean leaves using magic angle double cross-polarization nitrogen 15 nuclear magnetic resonance. *J. Biol. Chem.* 256, 11574–11579.

Schaefer, J., Skokut, T. A., Stejskal, E. O., McKay, R. A., and Varner, J. E. (1981b). Asparagine amide metabolism in developing cotyledons of soybean. *Proc. Natl. Acad. Sci. U.S.A.* 78, 5978–5982. doi: 10.1073/pnas.78.10.5978

Schmidt, M. A., Barbazuk, W. B., Stanford, M., May, G., Song, Z., Hong, W., et al. (2011). 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 Physiol.* 156, 330–345. doi: 10.1104/pp.111.173807

Schmidt, M. A., and Herman, E. M. (2008). The collateral protein compensation mechanism can be exploited to enhance foreign protein accumulation in soybean seeds. *Plant Biotechnol. J.* 6, 832–842. doi: 10.1111/j.1467-7652.2008.00364.x

Schmidt, M. A., Tucker, D. M., Cahoon, E. B., and Parrott, W. A. (2005). Towards normalization of soybean somatic embryo maturation. *Plant Cell Rep.* 24, 383–391. doi: 10.1007/s00299-005-0950-z

Skokut, T. A., Varner, J. E., Schaefer, J., Stejskal, E. O., and McKay, R. A. (1982). [N]NMR determination of asparagine and glutamine nitrogen utilization for synthesis of storage proteins I developing soybeans in culture. *Plant Physiol.* 60, 308–313. doi: 10.1104/pp.69.2.308

Staswick, P. E. (1989a). Developmental regulation and the influence of plant sinks on vegetative storage protein gene expression in soybean leaves. *Plant Physiol.* 89, 309–315. doi: 10.1104/pp.89.1.309

Staswick, P. E. (1989b). Preferential loss of an abundant storage protein from soybean pods during seed development. *Plant Physiol.* 90, 1252–1255. doi: 10.1104/pp.90.4.1252

Staswick, P. E., Papa, C., Huang, J.-F., and Rhee, Y. (1994). Purification of the major soybean leaf acid phosphatase that is increased by fruit removal. *Plant Physiol.* 104, 49–57.

Staswick, P. E., Zhang, Z., Clemente, T. E., and Specht, J. E. (2001). Efficient down-regulation of the major vegetative storage protein genes in transgenic soybean does not compromise plant productivity. *Plant Physiol.* 127, 1819–1826. doi: 10.1104/pp.010638

Thibaud-Nissen, F., Shealy, R. T., Kahanna, A., and Vodkin, L. O. (2003). Clustering of microarray data reveals transcript patterns associated with somatic embryogenesis in soybean. *Plant Physiol.* 132, 118–136. doi: 10.1104/pp.103.019968

Thompson, J. F., Madison, J. T., and Muenster, A. M. E. (1977). In vitro culture of immature cotyledons of soya bean (*Glycine max* L. Merr.). *Ann. Bot.* 41, 29–39.

Thorne, J. H. (1980). Kinetics of 14C-photosynthate uptake by developing soybean fruit. *Plant Physiol.* 65, 975–979. doi: 10.1104/pp.65.5.975

Thorne, J. H., and Rainbird, R. (1983). An in vivo technique for the study of phloem unloading in seed coats of developing soybean seeds. *Plant Physiol.* 72, 268–271. doi: 10.1104/pp.72.1.268

To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., Giraudat, J., et al. (2006). A network of local and redundant gene regulation governs *Arabidopsis* seed maturation. *Plant Cell* 18, 1642–1651. doi: 10.1105/tpc.105.039925

Truong, Q., Koch, K., Yoon, J. M., Everard, J. D., and Shanks, J. V. (2013). Influence of carbon to nitrogen ratios on soybean somatic embryo (cv. Jack) growth and composition. *J. Exp. Bot.* 64, 2985–2995. doi: 10.1093/jxb/ert138

Walker, D. R., and Parrott, W. A. (2001). Effect of polyethylene glycol and sugar alcohols on soybean somatic embryo germination and conversion. *Plant Cell Tissue Organ Cult.* 64, 55–62. doi: 10.1023/A:1010601402098

Walling, L., Drews, G. N., and Goldberg, R. B. (1986). Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. *Proc. Natl. Acad. Sci. U.S.A.* 83, 2123–2127. doi: 10.1073/pnas.83.7.2123

Wang, H. W., Zhang, B., Hao, Y. J., Huang, J., Tian, A. G., Liao, Y., et al. (2007). The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic *Arabidopsis* plants. *Plant J.* 52, 716–729. doi: 10.1111/j.1365-313X.2007.03268.x

Wan, T. F., Shao, G. H., Shan, X. C., Zeng, N. Y., and Lam, H. M. (2006). Correlation between AS1 gene expression and seed protein contents in different soybean (*Glycine max* [L.] Merr.) cultivars. *Plant Biol.* 8, 271–275. doi: 10.1055/s-2006-923876

Wilcox, J. R., and Shibles, R. M. (2001). Interrelationships among seed quality attributes in soybean. *Crop Sci.* 41, 11–14. doi: 10.2135/cropsci2001.41111x

Wilson, R. F. (2004). "Seed composition," in *Soybean: Improvement, Production and Uses*, 3rd Edn, eds H. R. Boerma and J. E. Specht (Madison, WI: American Society Of Agronomy-Crop Science Society), 621–677.

Wu, Y., and Messing, J. (2014). Proteome balancing of the maize seed for higher nutritional value. *Front. Plant Sci.* 5:240. doi: 10.3389/fpls.2014.00240

Zhang, L., Tan, Q., Lee, R., Trethewy, A., Lee, Y. H., and Tegeder, M. (2010). Altered xylem–phloem transfer of amino acids affects metabolism and leads to increased seed yield and oil content in *Arabidopsis*. *Plant Cell* 22, 3603–3620. doi: 10.1105/tpc.110.073833

Zhao-Ming, Q., Ya-nan, S., Qiong, W., Chun-yan, L., Guo-hua, H., and Qing-shan, C. (2011). A meta-analysis of seed protein concentration QTL in soybean. *Can. J. Plant Sci.* 91, 221–230. doi: 10.4141/cjps09193