



Temporal changes in metabolism late in seed development affect biomass composition

Shrikaar Kambhampati ¹, Jose A. Aznar-Moreno,^{2,†} Sally R. Bailey,³ Jennifer J. Arp ¹, Kevin L. Chu ^{1,3,4}, Kristin D. Bilyeu ⁵, Timothy P. Durrett ² and Doug K. Allen ^{1,3,*,‡}

¹ Donald Danforth Plant Science Center, St. Louis, Missouri 63132, USA

² Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, Kansas 66506, USA

³ United States Department of Agriculture, Agricultural Research Service, St. Louis, Missouri 63132, USA

⁴ Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164, USA

⁵ United States Department of Agriculture, Agricultural Research Service, Columbia, Missouri 65211, USA

*Author for communication: Doug.Allen@ars.usda.gov

†Present address: Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid, Pozuelo de Alarcón, Madrid 28223, Spain.

‡Senior author.

D.K.A., T.P.D., and S.K. designed the research; J.A.A.-M., performed RNA extraction and droplet digital PCR; J.J.A. performed PEPCK assay; K.L.C. quantified starch; S.R.B. performed CO₂ experiments and provided technical assistance to S.K.; K.D.B. developed the Jack-derived soybean germplasm. S.K. performed and analyzed data for all other experiments, and wrote the article with contributions from all authors; D.K.A. and T.P.D. supervised the research.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (<https://academic.oup.com/plphys/pages/general-instructions>) is: Doug K. Allen (Doug.Allen@ars.usda.gov).

Abstract

The negative association between protein and oil production in soybean (*Glycine max*) seed is well-documented. However, this inverse relationship is based primarily on the composition of mature seed, which reflects the cumulative result of events over the course of soybean seed development and therefore does not convey information specific to metabolic fluctuations during developmental growth regimes. In this study, we assessed maternal nutrient supply via measurement of seed coat exudates and metabolite levels within the cotyledon throughout development to identify trends in the accumulation of central carbon and nitrogen metabolic intermediates. Active metabolic activity during late seed development was probed through transient labeling with ¹³C substrates. The results indicated: (1) a drop in lipid contents during seed maturation with a concomitant increase in carbohydrates, (2) a transition from seed filling to maturation phases characterized by quantitatively balanced changes in carbon use and CO₂ release, (3) changes in measured carbon and nitrogen resources supplied maternally throughout development, (4) ¹³C metabolite production through gluconeogenic steps for sustained carbohydrate accumulation as the maternal nutrient supply diminishes, and (5) oligosaccharide biosynthesis within the seed coat during the maturation phase. These results highlight temporal engineering targets for altering final biomass composition to increase the value of soybeans and a path to breaking the inverse correlation between seed protein and oil content.

Introduction

The commercial value of a seed is established by its composition including protein, oil, and carbohydrate levels. At a

value of ~\$40 billion per year, soybean [*Glycine max* (L.) Merr.] production is second only to that of corn (*Zea mays*) in contribution of a crop to the US economy, as reported

by United States Department of Agriculture (USDA) and National Agriculture Statistical Services. In soybean, storage protein accounts for 35%–40% of seed dry weight (DW), with lipids (i.e. oil) accounting for 18%–20%, predominantly as triacylglycerol (TAG; Adams et al., 1983; Collakova et al., 2013; Li et al., 2015). Other biomass components, including carbohydrates, possess less market value, and a subset, such as the raffinose family oligosaccharides (RFOs) produced late in development, cannot be metabolized for energy by monogastric animals. The RFOs include raffinose and stachyose and are considered anti-nutritional components of livestock feed; therefore, detracting from seed value. In soybean, breeding efforts that increased protein content have resulted in lower yields (Mello Filho et al., 2004; Singh et al., 2016; Assefa et al., 2018) and the production of less TAG, indicating a tradeoff between protein and both yield and oil in mature seeds. Breaking the inverse correlations to improve the total seed value without compromising yields are unrealized goals of most breeding and biotechnological efforts.

Central carbon metabolic pathways are responsible for the production of storage reserves including lipids, proteins, and carbohydrates in plants. Though the network of primary metabolism is highly conserved across species, there is great diversity in biomass compositions within plant tissues, indicating that flux through the metabolic pathways can vary extensively. For example, the level of lipids in reproductive organs can range from <1% in peas (*Pisum sativum*) and lentils (*Lens culinaris*) to >70% in pecans (*Carya illinoensis*) and walnuts (*Juglans* sp.) and up to 88% in mesocarp tissues, such as palm (Dyer et al., 2008; Bates and Browse, 2012; Allen et al., 2015). Other organs such as leaves have low levels of lipids (<5%) in the forms of phospho- and galactolipids for membranes and very little storage lipid in the form of TAG (Lin and Oliver, 2008; Chapman et al., 2012). This variation indicates that steps in the metabolic network are pliable with throughput being context-specific across organs, species, and environments (Allen et al., 2015; Allen, 2016). However, resources available to a developing organ, such as a seed, are finite, being constrained by the supply and form of exudates from the seed coat of the maternal plant, usually comprised of sugars (i.e. sucrose, glucose, fructose) and amino acids (glutamine, asparagine, alanine; Rainbird et al., 1984; Fabre and Planchon, 2000; Schwender and Ohlrogge, 2002; Pipolo et al., 2004; Hernández-Sebastià et al., 2005; Allen et al., 2009). Thus, final seed composition, including oil and protein quantities, is a consequence of the availability of received assimilates and flux through enzymatic steps in metabolic pathways (Allen and Young, 2013; Truong et al., 2013). Understanding the differences in metabolic network flux and operation in tissues and species provides a template to engineer seeds or other organs with value-added compositions.

A quantitative description of temporal changes in metabolic network operation requires experimental methods that can probe stages of seed metabolism precisely and

dynamically. Metabolite levels of primary intermediates, such as amino acids, sugars, and organic acids, decline throughout development, while the storage components that include RFOs, lipids, and proteins increase (Fait et al., 2006; Collakova et al., 2013; Li et al., 2015). These levels, however, are routinely reported on a “per gram” basis. As the content of storage components that are considered “inactive/inert pools” in developing seeds increases, the primary metabolite pools are diluted as indicated by the hypothetical description in Figure 1A. Hence, reports of metabolite levels must properly account for dilution due to reserve accumulation when comparing trends over development. Metabolite amounts when compared on a “per seed” basis take into account the increase in storage reserves and may portray more accurately the transient changes in accumulation (Figure 1B).

One understudied developmental phase of metabolism is seed maturation. The process of desiccation involves more than drying, as indicated by enhanced enzyme activities and gene expression levels (Angelovici et al., 2010), and has important consequences on final reserve composition. However, hypotheses suggested by gene expression and final compositions require validation. Seed maturation represents ~40% of the entire seed developmental progression (Leprince et al., 2016) during which 10%–15% of TAGs are turned over (Chia et al., 2005; Baud and Lepiniec, 2009; Baud et al., 2009) and concomitantly, carbohydrates such as RFOs and cell wall polysaccharides (CWPs) continue to accumulate. The metabolic fate of turned over lipid carbon is not clear, but as the supply of exogenous substrates from the maternal plant ceases, sources of carbon are needed to support the biosynthetic demands of the seed (Baud et al., 2002; Baud and Graham, 2006; Angelovici et al., 2010). Genes involved in fatty acid oxidation and the glyoxylate cycle are expressed at higher levels late in seed development (Chia et al., 2005; Fait et al., 2006), suggesting that altered tricarboxylic acid (TCA) cycle metabolism, which can vary extensively in seeds (Schwender et al., 2006; Alonso et al., 2007; Allen et al., 2009), might be necessary to meet differing demands (Rolletschek et al., 2003; Rolletschek et al., 2005; Tschiersch et al., 2011) when seed-based photosynthetic contributions decline (Borisjuk et al., 2005; Fait et al., 2006; Angelovici et al., 2010). Whether changes in mitochondrial respiration (Chia et al., 2005) and peroxisomal metabolism (Salon et al., 1988; Raymond et al., 1992; Eastmond et al., 2000; Eastmond and Graham, 2001) could explain repartitioning of carbon for demand late in seed development and support RFO and CWP production (Kuo et al., 1988; Sánchez-Mata et al., 1998; Fait et al., 2006; Collakova et al., 2013; Gawłowska et al., 2017) is unknown.

The result of lipid decreases and production of RFOs during maturation metabolism is a less valuable seed. Experimental results presented here suggest that turned over reserves, including lipids, provide the carbon necessary to sustain the production of RFOs and CWPs late in seed development. In this study, we monitored temporal changes in seed biomass components, the maternal nutrient

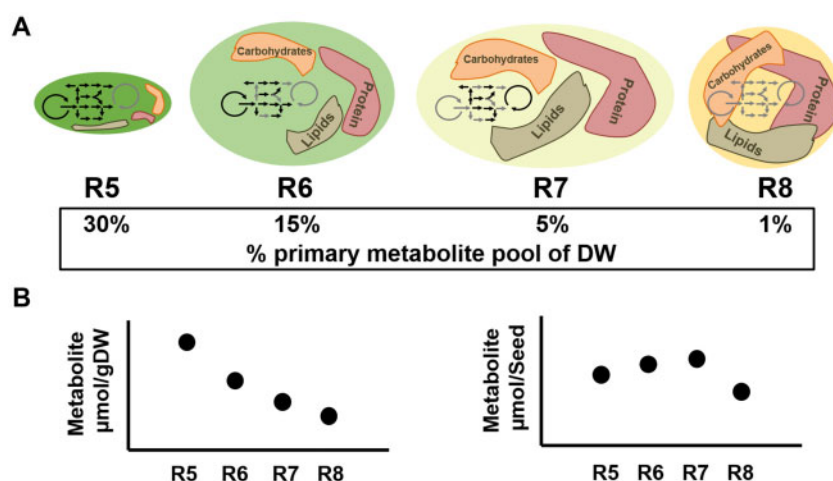


Figure 1 Description of temporal changes in metabolite content throughout seed development. A, Representation of the increasing accumulation of inactive/inert pools that constitute key storage reserves throughout the course of development (R5–R8) diluting the primary metabolite pool. In the box, the decrease in primary metabolite content as a percent of biomass (DW basis) by developmental stage is shown in B. Comparison of the trend of the primary metabolite content as evaluated on a “per gram DW” basis or as evaluated on a “per seed DW” basis.

contribution, and concentrations of central carbon and nitrogen metabolic intermediates to study the changing operation of the metabolic network during seed development. Stable isotopes were used to probe cotyledon and seed coat metabolism specific to the maturation phase to describe changes in carbon partitioning late in development. In addition, we compared biomass component accumulation and degradation between an ultra-low RFO line “Jack *rs2 rs3*” and wild-type “Jack” and assessed transcript patterns over development to probe the mechanisms of carbon turnover during maturation. The results are discussed in the context of improving seed composition.

Results

Changes in soybean seed biomass composition during maturation decrease seed value

We harvested seeds of the soybean cultivar “Williams 82” (Wm 82) according to size based on contemporary developmental stage descriptions (Naeve, 2005; Licht, 2014; Figure 2A), removed the seed coats, and quantified the fresh weight (FW). Cotyledons were subsequently dried to determine DW and moisture content (Figure 2B). R5 seeds were comprised of $81.5 \pm 2.5\%$ moisture, with seed desiccation events reducing this amount to $53.4 \pm 0.5\%$ in R7, $45.9 \pm 1.6\%$ in R7.5, and $12.9 \pm 0.8\%$ in R8 (maturity). Further loss of moisture continued in mature seed over time to $<9\%$ DW categorized as R8b and R8c. The net CO_2 release from cotyledons was quantified and indicated a peak in CO_2 release at R6 followed by a rapid decline (Figure 2C). The measured CO_2 spike was consistent with differences in storage reserve production, including substantial CO_2 generation when flux from pyruvate to acetyl-CoA enables fatty acid biosynthesis. The lipid production between R6 and R7 over a 2-week period accounted for 64% of the CO_2 generated during this time interval based on the accumulation of lipid in seeds (Figure 3).

During maturation of cotyledons there was a small but insignificant drop in total protein accumulation from $67.6 \pm 3.6 \text{ mg seed}^{-1}$ at R7 to $59.7 \pm 4 \text{ mg seed}^{-1}$ at maturity ($P = 0.11$) and a more dramatic significant change in lipid content from $40 \pm 1.1 \text{ mg seed}^{-1}$ at R7 down to $30.4 \pm 0.2 \text{ mg seed}^{-1}$ at maturity ($P = 0.004$; Figure 3). Starch production peaked at R6 ($4.3 \pm 0.2 \text{ mg seed}^{-1}$) and declined sharply before leveling off at $0.55 \pm 0.1 \text{ mg seed}^{-1}$. Sucrose accumulation peaked at $2.2 \pm 0.2 \text{ mg seed}^{-1}$ during R7 and remained at that level until maturity, while the RFOs raffinose and stachyose accumulated between R6 and R7 to a maximum of 1.6 and 5.9 mg seed^{-1} , respectively. Total free amino acids increased modestly throughout development (to 0.6 mg seed^{-1} at R6), and the residual biomass largely attributed to CWP reached $62.3 \pm 7 \text{ mg seed}^{-1}$ at R8. All biomass components increased between R5 and R6, indicating that inverse correlations between individual components are not obligatory when sufficient resources were present. Starch turned over between R6 and R7, at the same time that lipids (67.5%), protein (53.6%), and RFOs (84.1%) were being synthesized. The changes suggested that the turned-over carbon from starch was likely incorporated into lipids, protein, and/or RFOs.

Vegetative carbon and nitrogen sources diminish prior to seed maturation in soybeans

Vegetative parts of the plant are the source of sugars and amino acids for seeds during much of development (Hsu et al., 1984; Rainbird et al., 1984; Gifford and John, 1985; Egli and Bruening, 2001; Hernández-Sebastià et al., 2005) and impact final soybean composition (Allen and Young, 2013); however, the provisions change as seeds mature. Unlike the significant liquid endosperm present in *Brassicaceae*, soybean seed coat exudate is barely detectable at any given stage in development and is present as a shiny wet surface on cotyledons that amounts to a few microliters and diminishes with development. To recover the maximum amount of

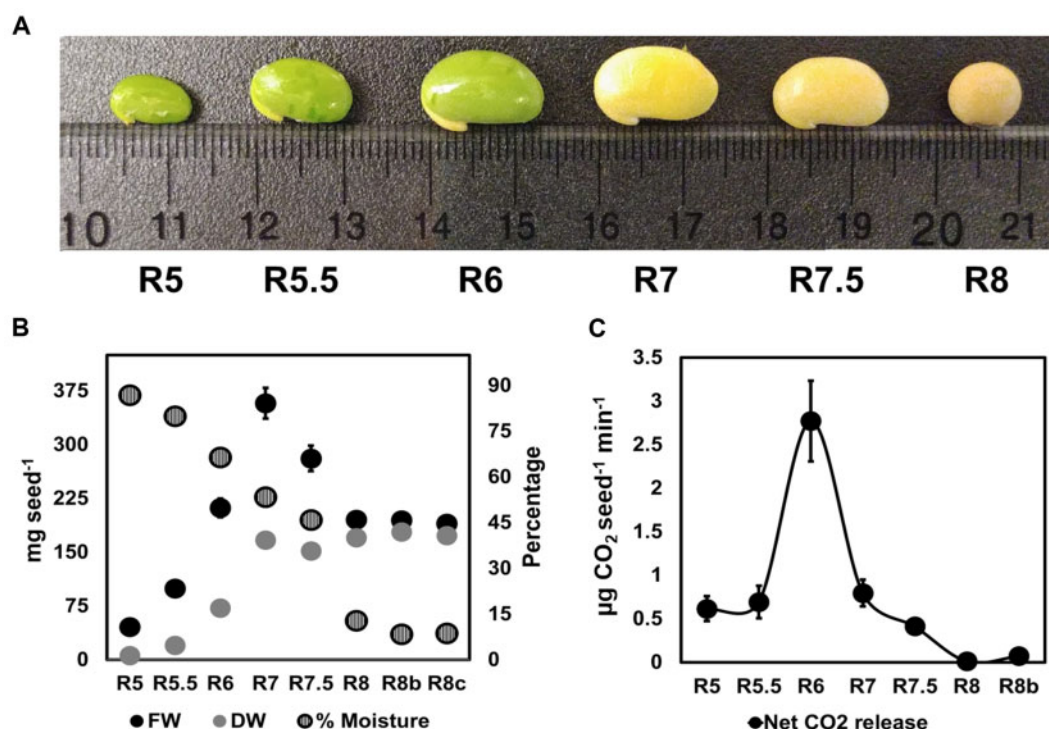


Figure 2 Soybean seed developmental stage descriptions. A, Image of representative cotyledons excised from the seed coat used for analyses from R5 (seed filling stage) to R8 (maturity). B, FW and DW measurements of cotyledon pairs represented as mg seed⁻¹, with moisture content calculated as loss of water from cotyledons upon drying, represented in percentages. Error bars represent standard errors of the mean (SEM), $n = 6$. C, Net CO₂ released presented in μg seed⁻¹ min⁻¹. Error bars represent SEM, $n = 6$ where each of the replicates represents an average of 10 measurements for a single cotyledon to overcome instrument drift.

seed coat exudate for measurements and minimize extraction from within the testa, the surface contents of the interior of the seed coat were briefly extracted with an isotonic solution of 20 mM ammonium acetate, pH 6.5. The exterior surface of the developing cotyledon was also briefly immersed in the same buffer to capture and quantify the major contents of the exudate. An overall decreasing trend of total exudate metabolites (Figure 4) was observed during seed development. The total metabolite levels in the exudate decreased significantly as seeds approached the maturation phase (R7 and R7.5), consistent with the trends in reserve accumulation (Figure 3) where no new storage proteins or lipids are made past R7, thus indicating a metabolic transition prior to maturation.

Nine amino acids were measured at detectable levels in the exudate of all developmental stages. Alanine and lysine were detected in R5, R5.5, and R6. Methionine, threonine, tryptophan, serine, glycine, cysteine, tyrosine, and valine were not detected and must be generated from seed-based metabolism during development (Figure 4; Supplemental Table S1). The nitrogen-rich amino acids asparagine, glutamine, arginine, and histidine were among the highest in content during peak protein accumulation stages (R5–R6), consistent with prior studies that indicated glutamine and asparagine are important sources of nitrogen for filling soybeans at least during early stages of development (Rainbird et al., 1984; Hernández-Sebastià et al., 2005). The

accumulation of nitrogen-rich amino acids including aspartate, glutamate, and arginine during the last stage of development (from R7.5 to R8) could hint at the importance of nitrogen provision for amino acid biosynthesis or other nitrogen-rich processes during germination.

Sucrose was also a significant carbon source through R6 before decreasing (Figure 4; Supplemental Table S1), possibly due to raffinose and stachyose production from the seed coat at stages R7 and R7.5. Prior reports that investigated RFOs in young developing seeds (Gomes et al., 2005; Kosina et al., 2009) suggested that the exudate contains precursors to RFO biosynthesis, that is, sucrose, myo-inositol, chiro-inositol, and pinitol, but did not consider stages of development beyond R6. Data in the current study indicate that RFOs are produced and exuded from the seed coat during the maturation phase, that is, R7 and R7.5 (Figure 4; Supplemental Table S1).

In planta levels of metabolites in cotyledons change throughout seed development

The pathway intermediates that are characteristic of stage-specific metabolism were investigated through pool size quantification (metabolite quantities) in cotyledons. Measurements were first obtained on a “per mg DW” basis and converted to amounts per dry seed to account for the increasing inert pools (lipids, protein, and carbohydrates) over the course of development and enable comparison

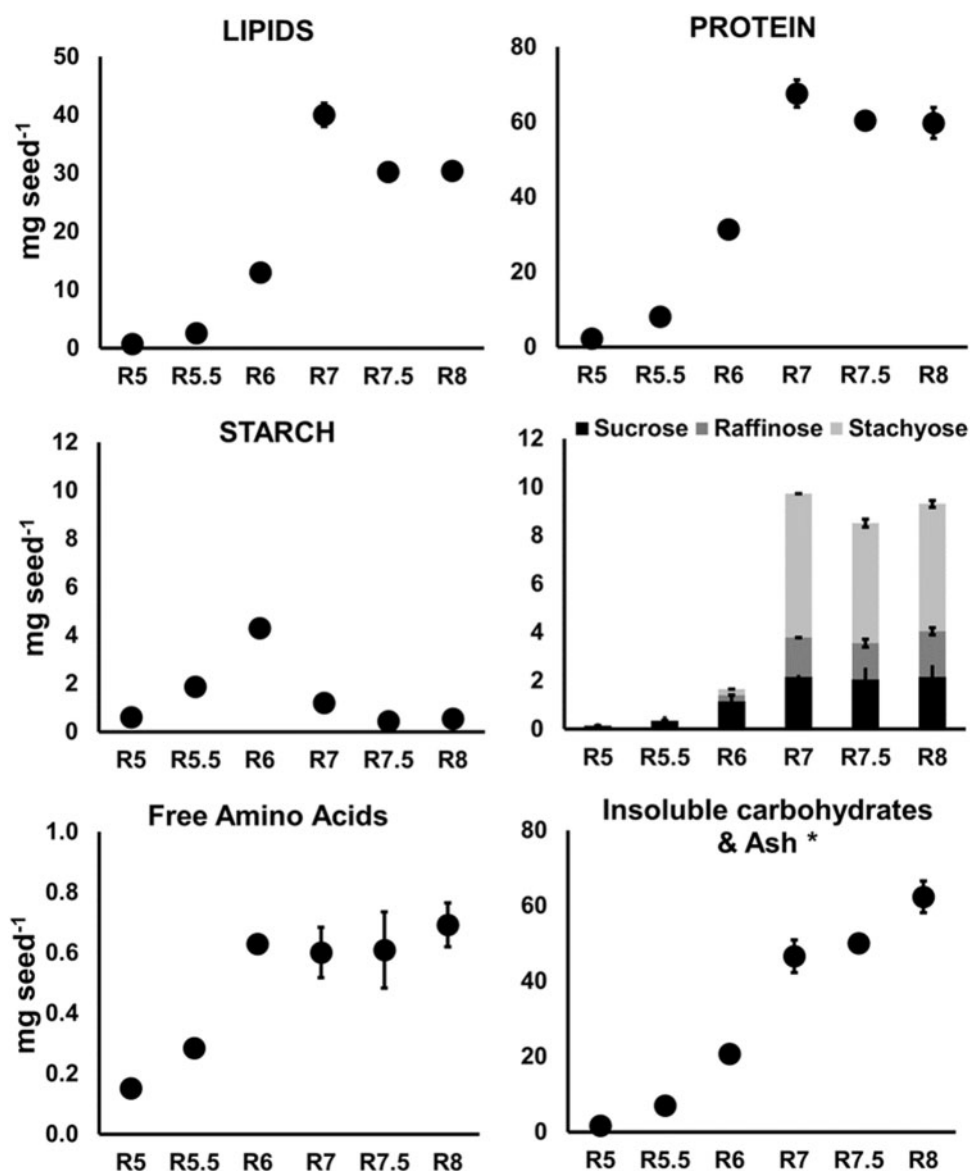


Figure 3 Trends in biomass component accumulation during seed development. Levels of individual biomass components were quantified at different stages of seed development, on a mg per seed basis. Values are based on cotyledons and do not include the seed coat except for R8. *Insoluble carbohydrates and ash that comprise much of the residual fraction were calculated by subtracting all other components from total seed biomass. Error bars represent SEM ($n = 3$).

between different stages (Supplemental Table S2). A *k*-means clustering approach was used to compare changes in trends of metabolite pools throughout development (Figure 5). All measured amino acids except glutamine clustered into groups 1 and 5, consistent with protein accumulation and a demand for storage protein synthesis plateauing between R7 and R7.5 when storage protein accumulation stopped (Figure 3). The steep increase in amino acid content between R7.5 and R8 (more pronounced for cluster 5) along with the concomitant decrease in storage protein (Figure 3) suggested proteolytic activity during maturation. Cluster 2 consisted of only two metabolites, glucose and fructose, which were elevated at R5 and dropped by R5.5. The variation within this cluster past R5.5 was high,

likely due to an increase in glucose content past R6 that was not observed for fructose (Supplemental Table S2). The levels of glucose and fructose in R5 may be a consequence of sucrose breakdown, whereas the increase in glucose at R6 and R7 occurred when starch was turning over (Figure 3), which was supported by the presence of the starch degradation product maltose detected in R6 and R7 (Supplemental Table S2). Maltose, in cluster 3, was similar to the organic acids 2-oxoglutaric acid (2OG), malate (Mal), succinate (Suc), and fumarate involved in the TCA cycle and the sugar phosphates 6-phosphogluconate (6PG), ribose 5-phosphate (R5P), and sedoheptulose 7-phosphate (S7P) involved in the oxidative and reductive steps of pentose phosphate pathway (PPP) metabolism. Cluster 3 increased to R6 then declined,

similar to measured net CO₂ release (Figure 2C), suggesting that CO₂ released in R6 resulted from the TCA cycle and oxidative PPP (OPPP) activity along with fatty acid biosynthesis, which is necessary to produce lipids during the same time frame.

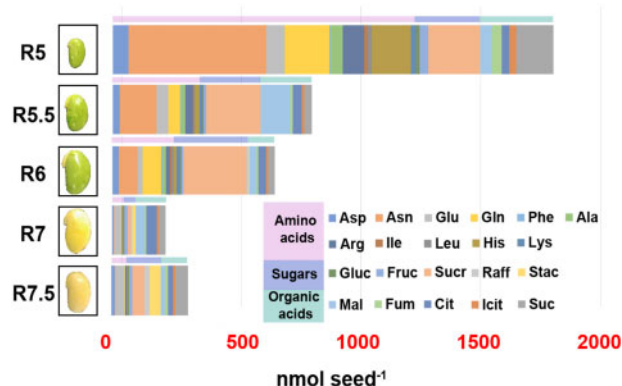


Figure 4 Levels of quantifiable metabolites in the exudate of developing seeds. Three metabolite classes: amino acids (purple), sugars (blue), and organic acids (green) were quantified and presented as nmol per seed amounts. Amino acids represented a major supply in R5 before decreasing significantly (by ~60%) at R5.5. The supply of sugars remained relatively consistent between R5 and R6 and decreased considerably by the time seeds reached the maturation phase (R7). Of the measured metabolites, asparagine (Asn), glutamine (Gln), histidine (His), sucrose (Sucr), malate (Mal), and succinate (Suc) were present in abundance at all stages.

Metabolites in cluster 3 and cluster 4 (Figure 5) overlapped significantly based on the 2D representation of clusters with principal component analysis (PCA). The common metabolites included those from glycolytic/gluconeogenic pathways and the Calvin–Benson cycle. Nonoverlapping metabolites (sucrose, galactinol, stachyose) were associated with RFO accumulation. The decrease in cluster 4 late in development was consistent with RFO production giving way to CWP biosynthesis (Supplemental Table S2).

Lipid turnover supports biosynthesis of insoluble carbohydrates late in development

Both the supply of resources from the exudate and the metabolic events as indicated from the pool size comparisons and altered storage reserve profiles changed as seeds developed. Pool sizes are suggestive of changes in metabolism but the differences in pool sizes cannot be strictly attributed to altered biosynthetic or turnover rates. Given that protein and lipid levels decrease nearing maturity while CWPs and undesirable RFOs accumulate (Figure 3), we used isotope tracers to probe the movement of carbon into and out of metabolic pools. After validation of the culturing approach (Supplementary Data S1), labeled ¹³C₃ glycerol was provided to cotyledons for up to 30 min to examine metabolism specific to lipid degradation at the beginning of the maturation phase (R7).

R7 seeds incorporated glycerol to produce triose and hexose phosphates over the course of 30 min (Figure 6A), consistent with the capacity of operating enzymes gluconeogenically to convert trioses into carbohydrates.

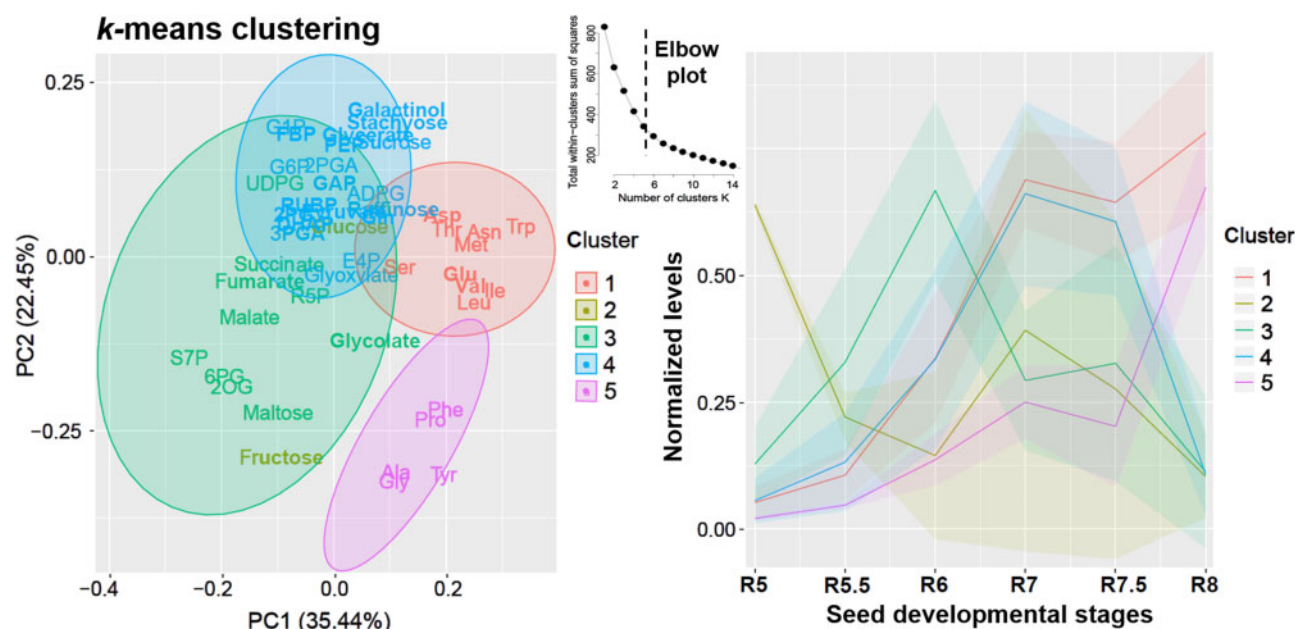


Figure 5 *k*-means clustering using central carbon and nitrogen metabolism intermediates representing trends over seed development stages. A total of 47 metabolites that include central carbon intermediates, organic, and amino acids as well as sugars were used for clustering. Metabolite levels were first calculated as nmol seed⁻¹ (Supplemental Table S2). For the clustering analysis, each metabolite was normalized using its maximum value at any stage (which was given a value of 1) in order to enable comparison of trends over the course of seed development. The optimal number of clusters was determined using the elbow method and was set at *k* = 5, as the within-cluster sum of squared distances reduced past five clusters. Metabolites that clustered together were represented on a 2D space using PCA (left panel) and the trends over development for each cluster were presented on the right panel. Abbreviations are defined in Supplemental Table S2.

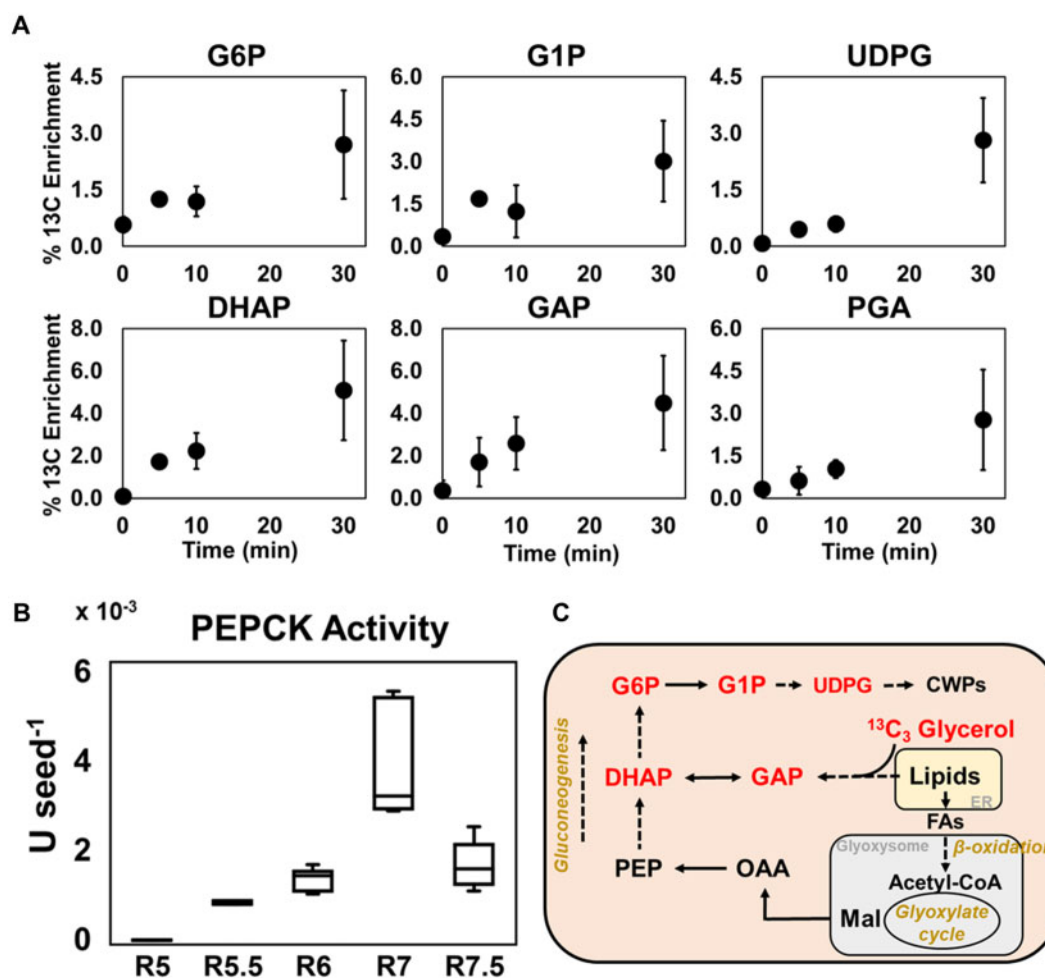


Figure 6 Carbon from turned over lipids is used to make hexose phosphates during R7. A, ^{13}C enrichment in intermediates of gluconeogenesis within a 30-min time-course pulse-labeling experiment using $^{13}\text{C}_3$ glycerol. Error bars represent SD ($n = 3$). B, PEPCK activity, as a signature of gluconeogenesis, over the course of seed development. Box-plot center lines indicate the medians; box limits, 25th and 75th percentiles; whiskers, 1.5 times the interquartile range from the 25th and 75th percentiles ($n = 6$). C, Schematic representation of carbon movement at R7 through central carbon metabolism involved in shuttling carbon from degrading lipids toward carbohydrate metabolism. Intermediates of gluconeogenic and carbohydrate metabolism that were labeled by $^{13}\text{C}_3$ glycerol are highlighted in red. ER, endoplasmic reticulum; FA, fatty acids; G1P, glucose 1-phosphate; G6P, glucose 6-phosphate; GAP, glyceraldehyde 3-phosphate; Mal, malic acid; OAA, oxaloacetic acid; PGA, 3-phospho glyceric acid; UDPG: uridine diphosphate glucose.

Glycerol was chosen in part because entry into metabolism as dihydroxyacetone phosphate (DHAP) would mimic the source of DHAP from the glycerol backbone that remains after lipolysis late in development. By 30 min, labeled carbon originating from $^{13}\text{C}_3$ glycerol was present at measurable levels in DHAP, GAP, PGA, G6P, G1P, and UDPG (Figure 6A), suggesting that gluconeogenesis may occur late in seed development. Labeling results were further confirmed considering the activity of phosphoenolpyruvate carboxykinase (PEPCK), a key enzyme in gluconeogenic metabolism. PEPCK activity was highest at R7 (Figure 6B), consistent with gluconeogenic activity to supply carbon for carbohydrate metabolism at this stage (Figure 6C).

The soybean cultivar “Jack” and a near-isogenic ultra-low RFO line (Jack *rs2 rs3*) containing natural variations that affect the function of the two *raffinose synthase* (RS) genes

RS2 and RS3 (Hagely et al., 2020) were used to compare carbon allocation and turnover between biomass components. Both Wm 82 and Jack cultivars accumulated RFOs predominantly between R6 and R7, while the Jack *rs2 rs3* mutant line had $<0.7\%$ total RFOs ($0.6 \pm 0.1 \text{ mg seed}^{-1}$; Figure 7A). Temporal trends for protein and lipids were similar in Wm 82 (Figure 3), Jack and Jack *rs2 rs3* lines (Figure 7A). Protein and lipids accumulated up to R7 and then declined ($P < 0.05$, $n = 4$ for lipid and $P < 0.001$, $n = 6$ for protein) in both genotypes during the maturation phase (R7–R8). No difference was observed in the amount of lipid accumulated between the two genotypes; however, a significantly higher amount of protein was accumulated between R6 and R7 in the ultra-low RFO line compared to the wild-type ($P = 0.0002$, $n = 6$; Figure 7A). This suggests that, though part of the excess carbon unused for RFO synthesis was

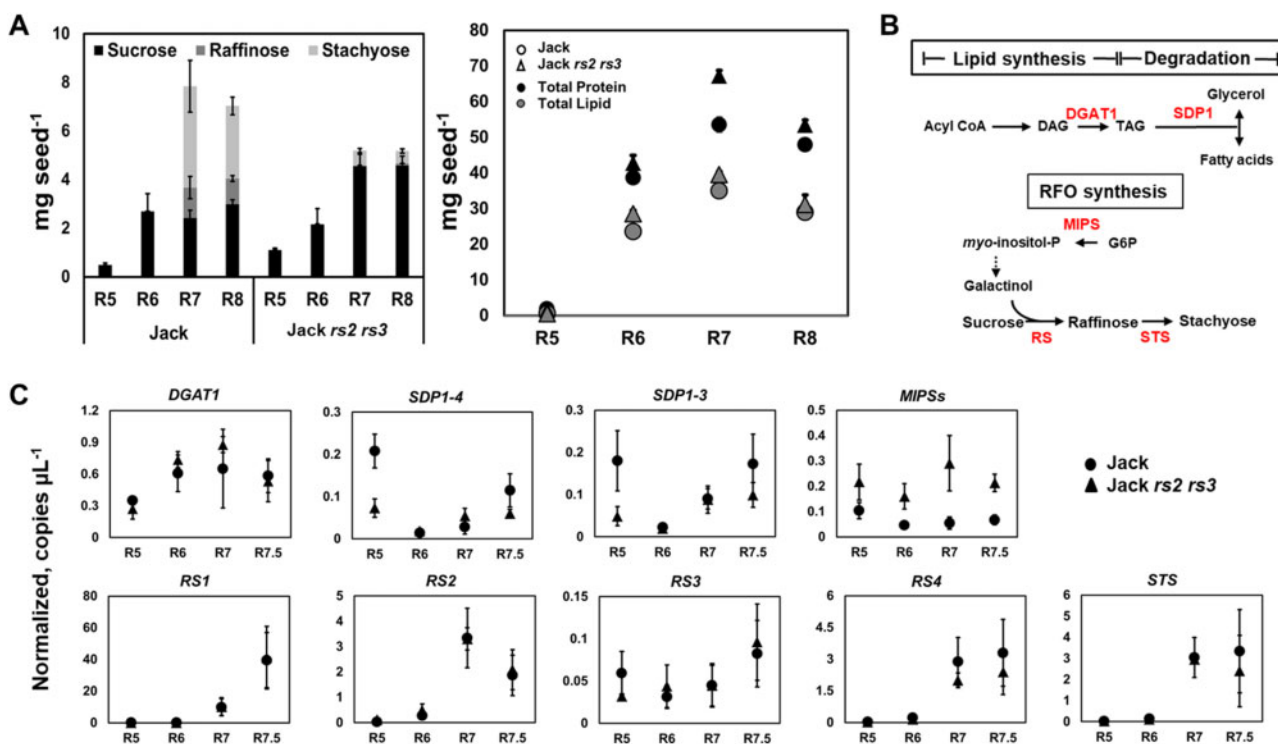


Figure 7 Comparison of biomass component and transcript accumulation between Jack and the ultra-low RFO line, Jack *rs2 rs3*. A, Accumulation of sucrose, raffinose, stachyose, lipids, and protein in the two genotypes throughout development. Error bars represent SEM ($n = 6$). B, Pathway description of critical enzymes involved in lipid synthesis (diacylglycerol acyltransferase; DGAT1), degradation (Sugar Dependant 1; SDP1), and RFO metabolism. C, Transcript abundance of genes highlighted in B quantified with droplet digital PCR using RNA extracted from cotyledons of developing soybean seeds from the cultivars Jack (wild-type) and Jack *rs2 rs3*. The Y-axis represents normalized (to reference genes, *ATP synthase subunit 1* and *SKIP16*) copies per microliter. For DGAT1 and MIPS, the expression levels of multiple paralogs is summed. Error bars represent SEM ($n = 3$). MIPS, myo-inositol phosphate synthetase.

accumulated as sucrose (1.5-fold higher in Jack *rs2 rs3*, $P = 0.0034$, $n = 6$), a portion of carbon was likely allocated to make protein.

Temporal changes in the expression patterns of genes involved in lipid biosynthesis, degradation, and RFO accumulation (Figure 7B) were investigated in the Jack wild-type cultivar and the ultra-low RFO Jack *rs2 rs3* mutant. Of the four paralogs of RS genes previously reported for soybean (Dierking and Bilyeu, 2008), RS1, RS2, and RS4 had similar patterns of expression with a major increase between R6 and R7 (Figure 7C), in concordance with raffinose metabolite accumulation (Figure 7A). No significant difference was observed between R7 and R7.5 in these three genes. The gene RS3 had the lowest expression and stayed at similar levels throughout development. Stachyose synthase (STS) also displayed a similar temporal trend as RS with a spike in transcript abundance from R6 to R7 when the majority of stachyose is accumulated, followed by no significant difference between R7 and R7.5 (Figure 7C). The transcript abundance and temporal expression pattern in the ultra-low RFO line, Jack *rs2 rs3*, did not show a difference compared to Jack. The mutation in *rs2* is a single codon omission within the gene due to a 3-bp deletion and *rs3* is a characterized variant allele previously reported to have no transcriptional changes (Hagely et al., 2020). This suggests that the

phenotypic difference (ultra-low RFO content) was likely a consequence of alteration at the protein level. In both Jack and Jack *rs2 rs3*, paralogs of *diacylglycerol acyltransferase1* (DGAT1) involved in TAG biosynthesis showed an increasing transcript abundance from R5 through R7 (Figure 7C) coinciding with lipid accumulation (Figure 7A). Similarly, the *sugar depending protein1* (SDP1) genes involved in lipid degradation increased from R7 to R7.5 (Figure 7C), coinciding with the lipid degradation period. Possibly SDP1 could be targeted to enhance seed oil content.

Isotopic labeling of seed coats indicates production of oligosaccharides that are partitioned to the surface of maturing cotyledons

The presence of RFOs in the exudate was unanticipated and may reflect biosynthetic activities in the seed coat. RFOs in seeds are not required for desiccation tolerance or germination (Dierking and Bilyeu, 2009; Valentine et al., 2017), and metabolism of the seed coat during development has not been previously described; thus, the role of these oligosaccharides remains obscure. During development, the seed coat DW is reduced with age, indicating that it may be partially remobilized as the last filial tissue that provides reserves to the seed or could be helpful as an osmotic regulator during germination and the imbibition process. To test

the contribution of the seed coat, detached seed coats from the R7 stage were labeled with ^{13}C -Sucr, resulting in the production of ^{13}C -raffinose (see methods for culture system set up; Figure 8, A and B) using the soybean line Jack and the near-isogenic ultra-low RFO line Jack *rs2 rs3*. As shown in Figure 8C, a significant amount of ^{13}C incorporation into raffinose within the seed coat was observed in the wild-type line Jack relative to the ultra-low RFO line (Jack *rs2 rs3*). Similar to the expression patterns observed in cotyledons, transcripts of all RS genes increased in abundance in the seed coats as the seed entered later stages of seed development, with RS1 being the most active gene. The two genotypes, however, did not show a significant difference in their expression patterns similar to cotyledons except for RS3, which had a higher expression in Jack compared with Jack *rs2 rs3* at R7 ($n = 2$, $P = 0.03$). These results indicated unequivocally that RFOs result in part from seed coat

metabolism and may be an important engineering target to favorably alter soybean seed composition.

Discussion

The composition of mature seed is the cumulative effect of events throughout seed development and is a consequence of: (1) the supply of carbon, nitrogen, and other resources from the maternal plant and (2) seed-based metabolism. In this study, we probed the changes in seed composition throughout seed development to explain the reduction in oil and protein levels and the accumulation of oligosaccharides late in development. Though inverse protein-to-oil relationships have been reported (reviewed in Clemente and Cahoon, 2009; Patil et al., 2017), the levels of these two reserves are not at odds during development (Kambhampati et al., 2019). Levels of both lipid and protein in the seed are highest during the initiation of the maturation phase and

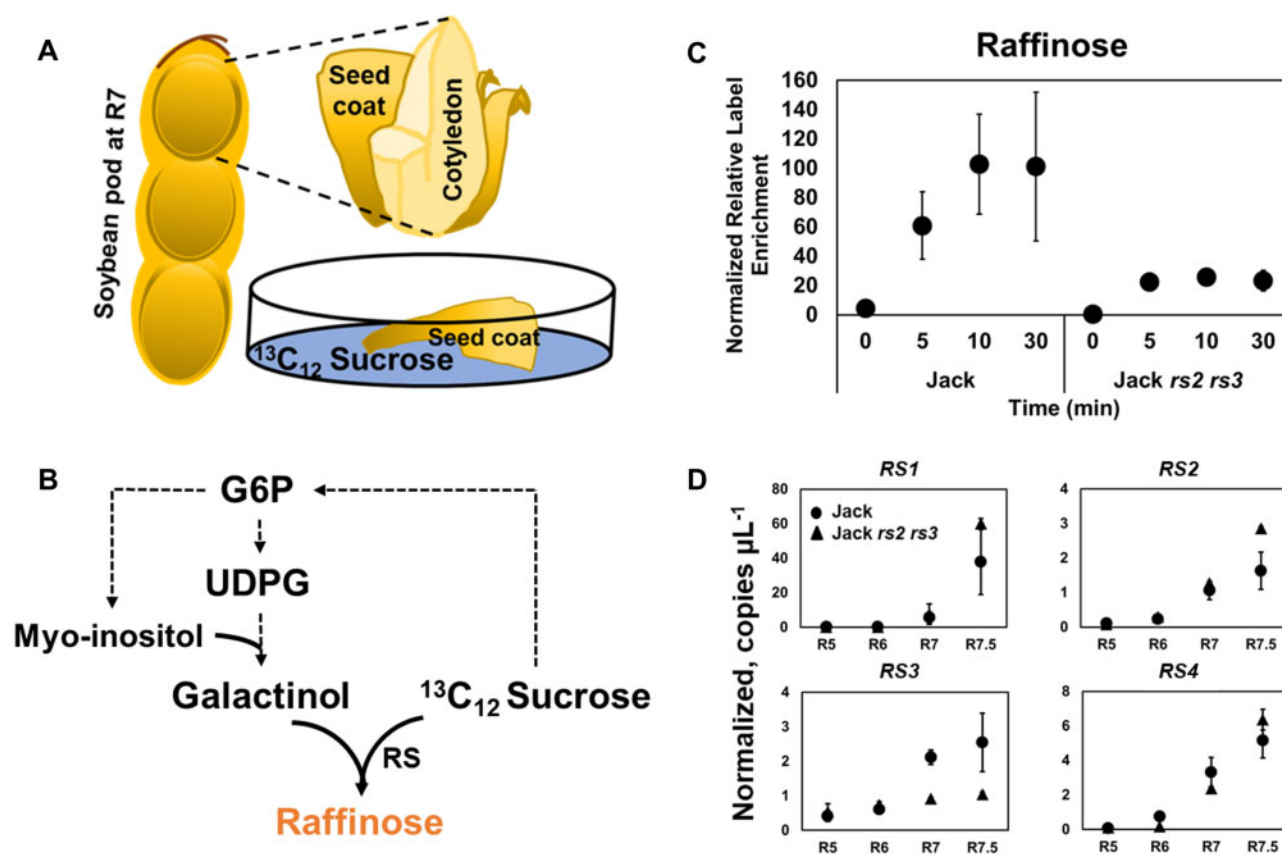


Figure 8 $^{13}\text{C}_{12}$ sucrose labeling in seed coats of R7 seeds. A, Depiction of an R7 pod with an expanded view of the seed coat and cotyledons. Seed coats were excised and cultured with $^{13}\text{C}_{12}$ sucrose as a sole source of carbon over 30 min. B, Biochemical route for $^{13}\text{C}_{12}$ sucrose incorporation for raffinose biosynthesis. Sucrose is used for the production of G6P followed by myo-inositol. G6P enters carbohydrate metabolism to produce UDPG. UDPG and myo-inositol together produce galactinol, which is combined with sucrose to produce raffinose via RS. C, A 30 min pulse labeling experiment using seed coats of the soybean line “Jack” and a near-isogenic ultra-low RFO line, Jack *rs2 rs3* (Hagely et al., 2020) at the initiation of maturation stage (R7) incubated with $^{13}\text{C}_{12}$ sucrose indicated significant label enrichment in raffinose. The y-axis represents arbitrary values normalized for pool size comparisons (see Supplemental Table S9) due to significantly different pool sizes of raffinose allowing for direct comparison of label (^{13}C) enrichment between the two genotypes. Error bars represent SE of the mean ($n = 3$). D, Transcript abundance of the four RS genes from seed coat tissue of developing soybean seeds of the cultivars Jack (wild-type) and Jack *rs2 rs3* were quantified using droplet digital PCR. The y-axis represents normalized (to reference genes, *ATP synthase subunit 1*, and *SKIP16*) copies per microliter. Gene abbreviations and primers used are described in Supplemental Table S4. Error bars represent SEM ($n = 2$).

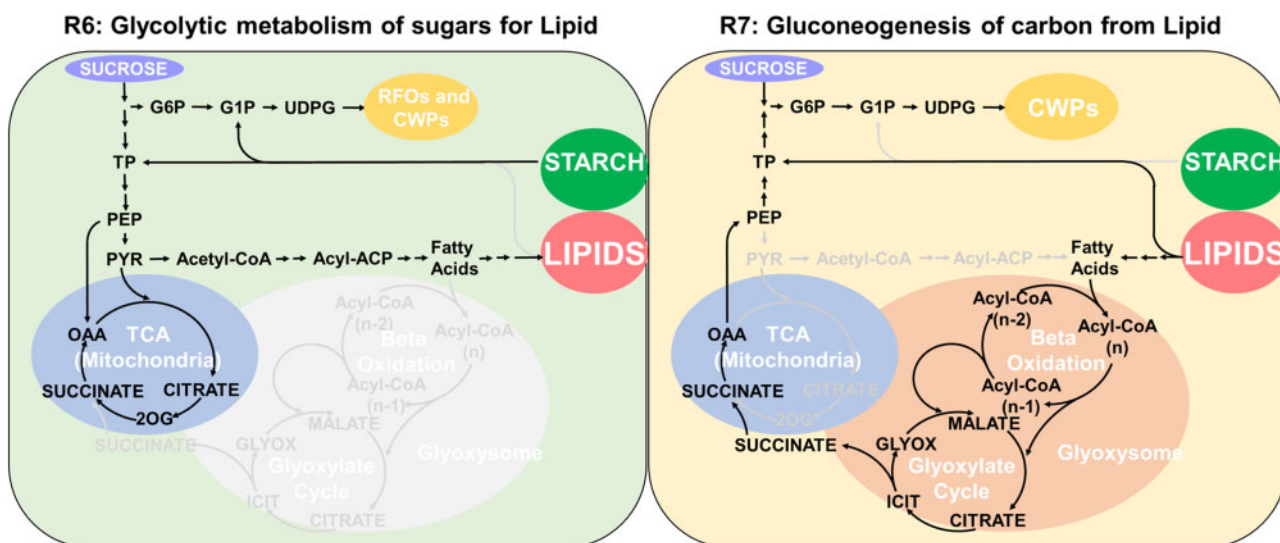


Figure 9 Proposed model for the metabolic switch between R6 and R7 to shuttle carbon from starch and lipid breakdown to oligosaccharide and cell wall polysaccharide biosynthesis. As described in the text, sources of carbon in R6 used for reserve production and energy metabolism decline late in development and result in some storage reserves being turned over to support biosynthesis of others. 2OG, 2-oxoglutarate; ACP, acyl carrier protein; GLYOX, glyoxylate; ICIT, isocitrate; PEP, phosphoenol pyruvate; PYR, pyruvate; TP, triose phosphate.

coincidentally decline at the time when residuals, mostly in the form of CWPs, are the only biomass component being accumulated (Figure 3) and when little exogenous carbon is available (Figure 4) to support biosynthetic demands. Thus, the pull for carbon is not exclusive to oil and protein and their turnover is likely a direct or indirect source for production of other reserves.

Seed-based carbon use efficiency indicates the redistribution of reserves to support metabolism late in development

Carbon conversion efficiency or carbon use efficiency (CUE) has been described in seeds with flux analyses to account for the production of CO_2 relative to substrates taken up (Schwender et al., 2004; Alonso et al., 2007; Allen et al., 2009; O'Grady et al., 2012). Though the description provides an indication of carbon lost relative to that converted to biomass, the calculation also reflects the composition of the biomass. Seeds that make large amounts of lipid produce more CO_2 as a part of fatty acid biosynthesis than seeds that predominantly make starch. Green seeds capitalize on photosynthesis to improve carbon efficiency (Schwender et al., 2004; Allen et al., 2009). Thus, the CUE calculation must take into consideration the metabolic context that differs amongst seeds and tissues.

Analogously, the temporal dynamics of seed metabolism are self-evident from the dramatic change in seed appearance with development. Soybean seeds are green at R5 (Figure 2A) and are capable of productively using available sunlight (Ruuska et al., 2004; Borisjuk et al., 2005; Rolletschek et al., 2005; reviewed in Angelovici et al., 2010) due in part to the contribution of Rubisco-based CO_2 fixation (Schwender et al., 2004; Allen et al., 2009). CUEs for stages in metabolism were calculated based on differences in

composition and CO_2 production over developmental stages, resulting in values of 90% and 76% between R5–R6 and R6–R7, respectively (Supplemental Table S3). Prior reports (Allen et al., 2009) that focused exclusively on seed filling indicated reasonable agreement with these early stages. During this time, the spike in CO_2 production occurs when lipid production is high and seeds are starting to transition from green to yellow in color (Figure 2C). TCA cycle metabolism and elevated oxidative PPP were also supported during this interval based on related metabolite clustering analysis (Figure 5). Later in development, differences in CO_2 production correlated with parallel drops in TCA cycle and OPPP metabolite levels (cluster 3 of Figure 4).

As seeds continue to develop, the CUE calculation is no longer applicable because the supply of exudate from the maternal plant is exhausted. Instead, the turnover of one reserve should equate with production of other reserves and CO_2 , to avoid violation of mass conservation. Developmental staging showed that starch levels drop (Figure 2) to supply other needs including RFO, lipid, and/or protein biosynthesis. From R7 to maturation, the balance of carbon turned over as lipid and protein must account for new production of carbohydrates and CO_2 generated. The reported changes in storage reserves (Supplemental Table S3) indicated a balance that was 94% closed. The production of some CO_2 late in development may suggest a slight decrease in final seed biomass with desiccation; however, the change in seed weight was not statistically significant.

Interestingly, the results indicated biosynthesis of RFOs at multiple locations, with carbon supplied from turned over storage reserves in the cotyledon and also as a result of the withering of the seed coat during desiccation. $^{13}\text{C}_{12}$ sucrose incubation (i.e. a precursor to raffinose) with seed coats indicated production of RFOs (Figure 8) and suggests that the

reduction in seed coat biomass during maturation may be analogous to a senescence process where the carbohydrate in the seed coat is converted to RFOs at the surface of the cotyledon.

Gluconeogenic activity is temporally synchronized with lipid degradation to supply turned over carbon for carbohydrate biosynthesis

Steady-state metabolic flux analysis using developing soybean seeds previously indicated that gluconeogenic metabolism does not occur when seeds receive adequate supplies of sugars (Allen et al., 2009). However, the composition of seed exudate late in development indicates that seeds do not receive an extensive supply of sugars from the maternal plant at these stages (Figure 4). As shown with $^{13}\text{C}_3$ glycerol labeling experiments (Figure 5A), enzymes involved in shuttling carbon to hexose phosphates can operate effectively late in development so that carbon can be used to produce insoluble carbohydrates (CWP). ^{13}C enrichment occurred in intermediates of glycolysis (DHAP, PEP), hexose phosphates (G6P, G1P), and the nucleotide sugar UDPG, which are precursors to carbohydrate production (Figure 6). From the balance of biomass components (Supplemental Table S3), the CO_2 loss of 3.3 mg (calculated) could come from PEPCK activity (61%; see Supplemental Table S3 for details) which was highest at R7.

The carbon required for PEPCK activity is likely obtained via the glyoxylate cycle, which utilizes acetyl-CoA derived from repeated deacylation of lipids beginning at R7 and continuing throughout the maturation phase by β -oxidation. The key enzymes required for glyoxylate cycle and β -oxidation, isocitrate lyase, malate synthase, 3-ketoacyl-CoA thiolase, and the multifunctional enzyme of β -oxidation, were previously shown to increase in activity during the maturation phase of embryo development in rapeseed (*Brassica napus*), characterized by lipid degradation (Chia et al., 2005). We observed that the glyoxylate levels increased over the course of development, peaking at R7 and R7.5 (cluster 4 of Figure 5, Supplemental Table S2). Further, activity of the glyoxylate cycle and gluconeogenesis are supported by prior measurements of transcript abundance throughout soybean development (Collakova et al., 2013; Li et al., 2015). The differences in carbon movement between the R6 stage of seed filling and initiation of maturation at R7 are stark and are summarized in Figure 9. Carbon received as sucrose via maternal contribution in R6 is channeled into lipid biosynthesis at R6. As the maternal contribution decreases between R6 and R7, starch turnover may contribute carbon to protein, lipid, and oligosaccharide biosynthesis while TCA cycle operation can meet the energy requirement for biosynthetic processes. The contribution of carbon from starch to lipids and oligosaccharides has been previously suggested based on transcript expression patterns during seed development in oilseeds (reviewed in Baud et al., 2008) and in tobacco (*Nicotiana tabacum*) leaves that over accumulate lipids (Grimberg et al., 2015). As the seeds reach the R7

maturation phase, turnover of lipids is initiated and the carbon from the glycerol backbones of lipids as well as degraded acyl chains is channeled into CWPs via gluconeogenic and β -oxidative pathways, respectively, or consumed to make energy within the TCA cycle. Although the focus of the current work is centered around the genetic and biochemical targets involved in primary metabolism, it is important to note that several transcriptional regulators and phytohormones play critical roles in determining temporal changes in metabolic processes and have been discussed elsewhere (Collakova et al., 2013; Jones and Vodkin, 2013; Li et al., 2015; Pereira Lima et al., 2017).

Recent efforts to improve seed quality have targeted lipases to reduce lipid breakdown late in development (Kanai et al., 2019) and RFO biosynthesis steps (Valentine et al., 2017; Hagely et al., 2020) to improve seed compositional traits without significant phenotypic consequences to maturation or germination. Carbohydrates as a whole (RFOs and CWPs) constitute $\sim 35\%$ or more of final seed composition and are an important sink that partitions carbon away from the production of oil and protein. Analysis of biomass composition in the ultra-low RFO line, Jack *rs2 rs3*, indicated a significant increase in protein between R6 and R7 instead of to RFOs as in the wild-type cv Jack. Thus, future studies to further elucidate the carbohydrate and protein tradeoff and focus on manipulating key carbon partitioning pathway nodes amongst all three biomass components, may indeed break the perceived inverse correlation between oil and protein. Temporally improving channeling of carbon from malate toward lipids using malic enzyme (Allen and Young, 2013) and increasing the sink strength of developing seeds (Rolletschek et al., 2020) by manipulating the hormone status (Quoc Thien et al., 2016; Kambhampati et al., 2017) represent unrealized potential targets for future soybean improvement.

Materials and methods

Plant growth conditions and tissue collection for in planta and seed coat exudate measurements

Soybean (*Glycine max*) cultivar Williams 82 was grown in a greenhouse as previously described (Kambhampati et al., 2019). Germinating seeds were transferred to one-gallon pots containing Fafard 4M and grown at 25°C – 27°C / 21°C – 23°C day/night temperatures with $> 35\%$ humidity and sunlight supplemented by ~ 400 – $1,000 \text{ W m}^{-2}$ to establish a 14 h d/10 h night photoperiod. Plants were watered daily and received Jack's 15-16-17 (JR Peters) fertilizer three times a week. Developing seeds were grouped based on FW and visual appearance to determine the developmental stage (Figure 1). At the time of harvest, seeds that were used as controls representing in planta conditions were dissected from pods, the seed coat was removed, and cotyledons were flash frozen with liquid nitrogen and stored at -80°C until further use. Cotyledons collected from a single plant were treated as a single biological replicate for all stages of development. Tissue that belonged to each replicate was ground

individually using a machined home-made stainless-steel hammer-crushing pestle and mortar design. Ground and frozen tissue was then lyophilized and aliquoted for individual biomass components and metabolite measurements.

For exudate experiments, an isotonic solution of 20 mM ammonium acetate pH 6.8 was placed on the interior side of excised seed coats and pipetted up and down repeatedly for 10 s. In addition, the surface of the corresponding cotyledon was rinsed with the same solution to collect any surface contents. The extracts from each stage were dried using a speed vacuum centrifuge and resuspended in water and filtered using 0.45 μ m cellulose acetate centrifuge filters (costar, Corning Inc.) prior to metabolite measurements using liquid chromatography–tandem mass spectrometry (LC–MS/MS).

Moisture content, FW, DW, and net CO₂ measurements

Moisture content was determined using FW and DW measurements. Seeds were weighed immediately after harvesting to obtain FWs, then sliced and dried in an oven. Dried seeds were measured at least three times over the course of several weeks to ensure no moisture remained and the weights were stable. CO₂ measurements were taken from whole soybean cotyledons, after excising the pod walls and seed coats, using a LI-COR 6400 with an attached insect respiration chamber (#6400-89) following the manufacturers protocol. Five replicates with three cotyledons each and ten measurements were used with readings taken every 20–30 s; 30 μ E of light was maintained throughout the measurement period to simulate light received by the cotyledons within the pods (Allen et al., 2009).

Normal seed RFO soybean cultivar “Jack” and the ultra-low RFO derivative “Jack rs2 rs3”

An ultra-low RFO version of soybean cultivar “Jack” (Jack rs2 rs3) was developed by backcrossing variant alleles of the rs2 and rs3 genes into Jack (Hagely et al., 2020). The ultra-low RFO phenotype has been defined as having a raffinose and stachyose content <0.70% of seed DW (Hagely et al., 2013; Schillinger et al., 2013, 2018).

¹³C labeled culture system set up and conditions

For culturing system development, cotyledons from specific stages were excised from seed coats under sterile conditions and immediately placed flat face down for each cotyledonary half into 300 μ L of sterile culture medium in a 24-well plate. A modified Linsmaier and Skoog medium (Thompson et al., 1977; Hsu and Obendorf, 1982) with Gamborg’s vitamins (Sigma) and 5 mM MES buffer adjusted to pH 5.8 was used as the culture medium and contained 200 mM U-¹³C₆ glucose as the exclusive carbon source. Culturing was performed under 30–35 μ E continuous light at 26°C, consistent with Allen et al. (2009) and tissue was collected at 5, 10, and 30 min, in triplicates, for all time course studies described. Untreated samples were also taken in triplicates for 0 timepoint (in planta) measurements. At the conclusion of

labeling experiments, the metabolism was quenched by a very brief rinse of the cotyledon surface with water prior to slicing off layers of the cotyledon to assess label uptake, metabolism, and heterogeneity (see [Supplementary Data S1](#) for details). Slices were rapidly frozen in liquid nitrogen and stored at –80°C until extraction.

To investigate carbon turnover from lipids to carbohydrates, we substituted U-¹³C₆ glucose with ¹³C₃ glycerol (15 mM) as the sole source of carbon. ¹³C₁₂ sucrose (100 mM) was used as the carbon source for data presented in [Figure 8](#). The salts and vitamins in all individual labeling experiments remained the same as described above. The bottom slice of labeled seeds was used for metabolite extraction and measurements of isotopologue distribution (see [Supplementary Data S1](#) for rationale).

RNA extraction and transcript analysis for verification of the culturing system

Total RNA was extracted from soybean cotyledons, seed slices, or seed coats depending on the experiment, using the RNeasy Plant Mini Kit (Qiagen) according to the supplier’s instructions. cDNA was synthesized using the qScript cDNA SuperMix (Quantabio) from 1 μ g total RNA previously treated with DNase I (Merck). For droplet generation, 20 μ L of the PCR reaction (cDNA, primers and the Bio-Rad ddPCR supermix) and 70 μ L of droplet generation oil were transferred to the middle and to the bottom rows, respectively, of a DG8 Cartridge before insertion into a QX200 Droplet Generator. The genes used and the primer sequences are included in [Supplemental Table S4](#). Droplets were transferred to a 96-well plate for PCR amplification in a C1000 Touch thermal cycler. The cycling protocol was 95°C enzyme activation for 5 min followed by 40 cycles of a two-step cycling protocol of 95°C for 30 s and T_m ([Supplemental Table S4](#)) for 1 min, then 4°C for 5 min and 90°C for 5 min. Following PCR amplification, the plate containing the droplets was placed in a QX200 droplet reader. Droplet digital PCR (ddPCR) data was analyzed with Bio-Rad QuantaSoft Analysis Pro Software. The *Glycine max* ATP synthase subunit S1 (ATP; Glyma12g020500) and SKP1/ASK-interacting protein 16 (SKIP16; Glyma12g051100) were used as internal references (Hu et al., 2009).

Extraction of polar and nonpolar metabolites

Metabolite extraction was carried out following the protocol described in Czajka et al. (2020) and Kambhampati et al. (2019) with a few modifications. Briefly, the stored samples were removed from –80°C and two metal beads were added to each tube along with 1 mL 7:3 methanol/chloroform (–20°C) and a piperazine-N,N’-bis(2-ethanesulfonic acid; PIPES), norvaline, and ribitol mixed standard (66, 17, and 130 pmol μ L^{–1} final concentration, respectively). Samples were kept on ice throughout extraction unless otherwise noted. Samples were pulverized using a ball mill at 28 Hz for 5 min or until fully ground. The mixtures were then incubated at –20°C for 3 h, with intermittent vortexing to ensure complete extraction. A 500 μ L of ddH₂O (4°C) was

added to each sample and vortexed vigorously before being centrifuged at 14,000 rpm at 4°C for 10 min, during which the samples phase-separated. The upper aqueous phase containing water-soluble metabolites was transferred to a 1.5 mL eppendorf tube with a 0.45 µm centrifugal filter (Costar, Corning Inc.) and spun at 14,000 rpm at 4°C for 2 min. This solution was then transferred to glass vials (Agilent, Xpertek) for LC–MS/MS analysis to detect soluble sugars, free amino acids, and sugar phosphates.

Quantification of proteins, lipids, and starch

A total of 3–5 mg of ground lyophilized tissue was subjected to liquid hydrolysis and protein was measured using amino acid compositional analysis as described in [Kambhampati et al. \(2019\)](#). In brief, 20 µL of 1 mM cell-free ¹³C-labeled amino acid standard mix (Sigma) was added to the biomass pellet and dried using a speed vacuum centrifuge. Then, 50 µL of 4M methanesulfonic acid containing 0.2% (w/v) tryptamine was added to the dried product and incubated at 110°C for 22 h. Upon completion of hydrolysis, the samples were neutralized using 50 µL of 4 M sodium hydroxide, briefly vortexed and dried. Upon drying, the samples were resuspended in 1 mL ultra-pure water and vortexed to recover the hydrolyzed amino acids and then filtered using 0.45 µm centrifugal filters. Amino acids were detected using LC–MS/MS (described below) and quantified using isotopic dilution based on peak areas obtained from known concentrations of internal standards. The sum of the concentrations of all 20 amino acids, in milligrams, was used to establish the concentration of protein ([Supplemental Table S5](#)).

Lipid content was analyzed according to an adapted version of the method described in [Allen and Young \(2013\)](#) by converting total lipids into fatty acid methyl esters (FAMES). In brief, freshly prepared 5% (v/v) sulfuric acid:methanol was added to ~20 mg of ground lyophilized tissue along with 25 µL 0.2% (w/v) butylated hydroxytoluene in methanol to prevent oxidation and two internal standards, triheptadecanoin and tripentadecanoin, before heating at 110°C for 3 h, vortexing hourly. After cooling to room temperature, 0.9% NaCl (w/v) was added to each sample to quench the reaction. The FAMES were then extracted using hexane and quantified by gas chromatography–flame ionization detection (GC–FID) using a DB23 column (30 m, 0.25-mm i.d., 0.25-µm film; J&W Scientific). The GC was operated in a split mode (30:1). The flame ionization detector was operated with a temperature of 250°C with an oven temperature ramp profile from 180°C to 260°C at a rate of 20°C min⁻¹ followed by a hold time of 7 min. Comparisons of peak areas to the two internal standards were used for quantification.

Starch measurements were performed on ~20 mg of ground lyophilized tissue, directly without prior extraction. Total starch content in cotyledons throughout reproductive development was determined, in triplicates, using the Megazyme starch assay kit (Megazyme International Ireland), using the AOAC Official Method 996.11 (Approved Methods of the AACC; [McCleary et al., 1997, 2019](#)),

modified to adjust the final assay volume for 96-well plate reader compatibility. Briefly, the ground lyophilized tissue was washed twice with 80% (v/v) ethanol at 85°C prior to heating at 110°C for 10 min with DMSO. The samples were then treated with α-amylase at 110°C for 12 min (vortexing every 4 min) followed by amyloglucosidase at 50°C for 1 h. The samples were then centrifuged; the supernatant was collected, and incubated with the GOPOD reagent at 50°C for 20 min. The absorbance at 510 nm was measured using a spectrophotometer, and the starch content was determined by comparison with a standard curve generated using a serial dilution of starch standards treated the same way as biological samples. Quantities of all biomass component measurements presented in [Figure 2](#) are provided as [Supplemental Table S6](#).

Quantification of soluble sugars, amino acids, and sugar phosphates using high-performance liquid chromatography–MS/MS

Sugars and sugar phosphates were analyzed from the water-soluble fraction using a Shimadzu (UFLCXR) high-performance liquid chromatography (HPLC) system connected to an AB Sciex triple quadrupole MS equipped with a Turbo V electrospray ionization (ESI) source using the method described in [Czajka et al. \(2020\)](#). Negative ion mode was used to monitor sugar and sugar phosphate fragments. A 5 µL sample was injected on the Infinity Lab poroshell 120 Z-HILIC column (2.7 µm, 100 × 2.1 mm²; Agilent technologies, Santa Clara, CA, USA) and the metabolites were eluted with an increasing gradient of acetonitrile: 10 mM ammonium acetate (90:10, v/v) and 5 µM medronic acid, pH 9.0 (A) and 10 mM ammonium acetate in water, pH 9.0 (B). The flow rate was 0.25 mL/min. Sugars and sugar phosphates were separated using a binary gradient of 95%–70% B over 8 min then to 50% B over the next 4 min followed by a hold at 25% B for 1.5 min. The gradient was then decreased to 30% B over 0.5 min followed by a hold for 1 min before returning to 95% B to re-equilibrating the column for 6 min. The HPLC eluent was introduced into an ESI source with the following conditions: ion spray voltage, 4.5 kV (ESI-); ion source temperature, 550°C; source gas 1, 45 psi; source gas 2, 40 psi; curtain gas, 35 psi; and entrance potential, 10. Ions were detected and monitored using a targeted Multiple Reaction Monitoring (MRM) approach with the parameters optimized by direct infusions, provided in [Supplemental Table S7](#), for accurate quantification. The value for entrance potential was the default (–10) for all analytes. For absolute quantification, data were analyzed using the quantitation wizard available in Analyst (v. 1.6.2) software (AB SCIEX, Concord, Canada). Metabolite concentrations were calculated based on calibration curves. Recoveries were assessed using ribitol and PIPES as internal standards for sugars and sugar phosphates, respectively.

Amino acids were measured using the same instrumentation and column as described above, with the following changes in mobile phases, gradient and ionization conditions. Mobile phase A consisted of 20 mM ammonium formate in water, pH 3.0, and B was composed of 90% acetonitrile and 10% water with a final concentration of 20 mM ammonium formate, pH 3. Three μL from each sample were injected and a flow rate of 0.25 μL was used for separation of amino acids on the HPLC column. A binary gradient composed of 100%–90% B over 2 min, 90%–50% B over the next 6 min followed by returning to 100% B over 30 s and re-equilibration of 5.5 min was used to separate the analytes. The HPLC eluent was introduced into an ESI source with the following conditions: ion spray voltage, 4.5 kV (ESI+ and ESI–); ion source temperature, 400°C; source gas 1, 45; source gas 2, 40; curtain gas, 35; and entrance potential, 10. Ions were detected and monitored using a targeted MRM approach using parameters included in [Supplemental Table S7](#). Data were analyzed similar to the sugars and sugar phosphates described above, except novaline was used as an internal standard to assess recoveries. All statistical analysis and data visualization were performed using Microsoft Excel (2013) or R programming language ([R CoreTeam, 2013](#)) using base functions and the package ggplot2 ([Wickham, 2016](#)).

Quantification of isotopologue abundance and average labeling

The LC–MS/MS conditions used for label detection are identical to the ones described above with the exception of MRM transitions used ([Supplemental Table S8](#)) that were selected based on [Kappelmann et al. \(2017\)](#). Peaks were manually integrated and the natural abundance was corrected using the R package IsoCorrector ([Heinrich et al., 2018](#)). Fractional enrichment of the corrected isotopologues (M0–Mn) obtained from IsoCorrector was used to calculate average labeling. Average labeling was calculated as described in [Buescher et al. \(2015\)](#) using the following equation;

$$\% \text{ }^{13}\text{C enrichment} = \frac{\sum_{i=0}^n i \cdot S_i}{n} \cdot 100$$

where i denotes the isotopologue, n is the number of possible ^{13}C carbons, and S is the fraction of the labeled isotopologue. For detailed calculations, see [Supplemental Table S9](#).

PEPCK enzyme activity assay

PEPCK enzyme activity was measured using the method of [Walker et al. \(1999\)](#) detailed on protocols.org ([Osorio et al., 2014](#)). Briefly, crude protein was extracted from 100 to 250 mg FW of seed tissue in a buffer containing 0.5 M bicine-KOH (pH 9.0), 0.2 M KCl, 3 mM EDTA, 5% (w/v) PEG-4000, 25 mM DTT, and 0.4% (w/v) bovine serum albumin. The extract was centrifuged for 20 min at 13,000g at 4°C, and the supernatant was added to a buffer containing 0.5 M bicine-KOH (pH 9.0), 3 mM EDTA, 55% (w/v) PEG-4000, and 25 mM DTT, then incubated for 10 min on ice and centrifuged at 13,000g at 4°C for 20 min. The supernatant was

discarded, and the pellet was resuspended in 10 mM bicine-KOH (pH 9.0) containing 25 mM DTT. PEPCK activity was measured in the direction of the carboxylation reaction by coupling the reaction with malate dehydrogenase (EC 1.1.1.37; Sigma Aldrich 10127914001) and following the oxidation of NADH at 340 nm at room temperature using a spectrophotometer (SpectraMax M2^e, Molecular Devices). Total protein was measured using the protein extract for PEPCK activity using Bradford reagent (Millipore Sigma; Cat: B6916) and a standard curve using commercial bovine serum albumin standards (Thermo Fisher, cat: 23208).

Accession numbers

Accession numbers for all genes used in RT-qPCR experiments are listed within the legend of [Supplemental Table S4](#).

Supplemental data

Supplemental Data S1. Establishing and assessing short-time pulse-labeling conditions for nonperturbed in planta temporal assessment of seed development

Supplemental Figure S1. $^{13}\text{C}_6$ glucose labeling using cotyledons of R7 seeds for establishing the culturing system.

Supplemental Table S1. Composition of seed coat exudate, quantities represented in nmol per seed.

Supplemental Table S2. Total pool size (accurate quantities) of the metabolites detected via LC-MS/MS.

Supplemental Table S3. Temporal changes in carbon conversion efficiency and loss of CO_2 as PEPCK.

Supplemental Table S4. Primers used for droplet digital PCR and their annealing temperatures.

Supplemental Table S5. Amino acid concentration (mg seed^{-1}) from hydrolyzed protein at different developmental stages.

Supplemental Table S6. Biomass component measurements for soybean seed developmental stages.

Supplemental Table S7. MRM parameters used for absolute quantification.

Supplemental Table S8. MRM parameters used for isotope labeling experiments.

Supplemental Table S9. Calculations for isotopologue distribution and average labeling.

Acknowledgments

The authors thank the director of the Proteomics and Mass Spectrometry (PMSF) facility at the Donald Danforth Plant Science Center (DDPSC), Dr Bradley S. Evans for his support and advice throughout this work. In addition, the authors acknowledge the Plant Growth Facility (PGF) at DDPSC for their help with plant growth and maintenance.

Funding

This work was supported by the United States Department of Agriculture, Agricultural Research Service, the United States Department of Agriculture, National Institute of Food

and Agriculture (award number USDA-NIFA #2017-67013-26156, #2016-67013-24585), the Multistate Research Project (NC1203), the United Soybean Board (#1820-152-0134), National Institutes of Health (U01 CA235508), and the National Science Foundation (NSF-IOS-1812235). Support for the acquisition of the 6500 QTRAP LC–MS/MS was also provided by the National Science Foundation (NSF-DBI #1427621).

Conflict of interest statement. None declared.

References

- Adams CA, Fjerstad MC, Rinne RW (1983) Characteristics of soybean seed maturation: necessity for slow dehydration. *Crop Sci* **23**: 265–267
- Allen DK (2016) Quantifying plant phenotypes with isotopic labeling & metabolic flux analysis. *Curr Opin Biotechnol* **37**: 45–52
- Allen DK, Bates PD, Tjellström H (2015) Tracking the metabolic pulse of plant lipid production with isotopic labeling and flux analyses: past, present and future. *Prog Lipid Res* **58**: 97–120
- Allen DK, Ohlrogge JB, Shachar-Hill Y (2009) The role of light in soybean seed filling metabolism. *Plant J* **58**: 220–234
- Allen DK, Young JD (2013) Carbon and nitrogen provisions alter the metabolic flux in developing soybean embryos. *Plant Physiol* **161**: 1458–1475
- Alonso AP, Goffman FD, Ohlrogge JB, Shachar-Hill Y (2007) Carbon conversion efficiency and central metabolic fluxes in developing sunflower (*Helianthus annuus* L.) embryos. *Plant J* **52**: 296–308
- Angelovici R, Galili G, Fernie AR, Fait A (2010) Seed desiccation: a bridge between maturation and germination. *Trends Plant Sci* **15**: 211–218
- Assefa Y, Bajjalieh N, Archontoulis S, Casteel S, Davidson D, Kovács P, Naeve S, Ciampitti IA (2018) Spatial characterization of soybean yield and quality (amino acids, oil, and protein) for United States. *Sci Rep* **8**: 14653
- Bates P, Browse J (2012) The significance of different diacylglycerol synthesis pathways on plant oil composition and bioengineering. *Front Plant Sci* **3**: 147
- Baud S, Boutin J-P, Miquel M, Lepiniec L, Rochat C (2002) An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. *Plant Physiol Biochem* **40**: 151–160
- Baud S, Dubreucq B, Miquel M, Rochat C, Lepiniec L (2008) Storage reserve accumulation in *Arabidopsis*: metabolic and developmental control of seed filling. *Arabidopsis Book* **6**: e0113
- Baud S, Graham IA (2006) A spatiotemporal analysis of enzymatic activities associated with carbon metabolism in wild-type and mutant embryos of *Arabidopsis* using *in situ* histochemistry. *Plant J* **46**: 155–169
- Baud S, Lepiniec L (2009) Regulation of *de novo* fatty acid synthesis in maturing oilseeds of *Arabidopsis*. *Plant Physiol Biochem* **47**: 448–455
- Baud S, Wuillème S, To A, Rochat C, Lepiniec L (2009) Role of WRINKLED1 in the transcriptional regulation of glycolytic and fatty acid biosynthetic genes in *Arabidopsis*. *Plant J* **60**: 933–947
- Borisjuk L, Nguyen TH, Neuberger T, Rutten T, Tschiersch H, Claus B, Feussner I, Webb AG, Jakob P, Weber H, et al. (2005) Gradients of lipid storage, photosynthesis and plastid differentiation in developing soybean seeds. *New Phytol* **167**: 761–776
- Buescher JM, Antoniewicz MR, Boros LG, Burgess SC, Brunengraber H, Clish CB, DeBerardinis RJ, Feron O, Frezza C, Ghesquiere B, et al. (2015) A roadmap for interpreting ¹³C metabolite labeling patterns from cells. *Curr Opin Biotechnol* **34**: 189–201
- Chapman KD, Dyer JM, Mullen RT (2012) Biogenesis and functions of lipid droplets in plants: thematic review series: lipid droplet synthesis and metabolism: from yeast to man. *J Lipid Res* **53**: 215–226
- Chia TYP, Pike MJ, Rawsthorne S (2005) Storage oil breakdown during embryo development of *Brassica napus* (L.). *J Exp Bot* **56**: 1285–1296
- Clemente TE, Cahoon EB (2009) Soybean oil: genetic approaches for modification of functionality and total content. *Plant Physiol* **151**: 1030–1040
- Collakova E, Aghamirzaie D, Fang Y, Klumas C, Tabataba F, Kakumanu A, Myers E, Heath LS, Grene R (2013) Metabolic and transcriptional reprogramming in developing soybean (*Glycine max*) Embryos. *Metabolites* **3**: 347–372
- Czajka JJ, Kambhampati S, Tang YJ, Wang Y, Allen DK (2020) Application of stable isotope tracing to elucidate metabolic dynamics during *Yarrowia lipolytica* α -ionone fermentation. *iScience* **23**: 100854
- Dierking EC, Bilyeu KD (2008) Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *Plant Genome* **1**: 135–145
- Dierking EC, Bilyeu KD (2009) Raffinose and stachyose metabolism are not required for efficient soybean seed germination. *J Plant Physiol* **166**: 1329–1335
- Dyer JM, Stymne S, Green AG, Carlsson AS (2008) High-value oils from plants. *Plant J* **54**: 640–655
- Eastmond PJ, Germain V, Lange PR, Bryce JH, Smith SM, Graham IA (2000) Postgerminative growth and lipid catabolism in oilseeds lacking the glyoxylate cycle. *Proc Natl Acad Sci USA* **97**: 5669–5674
- Eastmond PJ, Graham IA (2001) Re-examining the role of the glyoxylate cycle in oilseeds. *Trends Plant Sci* **6**: 72–78
- Egli DB, Bruening WP (2001) Source-sink relationships, seed sucrose levels and seed growth rates in soybean. *Ann Bot* **88**: 235–242
- Fabre F, Planchon C (2000) Nitrogen nutrition, yield and protein content in soybean. *Plant Sci* **152**: 51–58
- Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G (2006) *Arabidopsis* seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiol* **142**: 839–854
- Gawłowska M, Świącicki W, Lahuta L, Kaczmarek Z (2017) Raffinose family oligosaccharides in seeds of *Pisum* wild taxa, type lines for seed genes, domesticated and advanced breeding materials. *Genet Resour Crop Evol* **64**: 569–578
- Gifford RM, John HT (1985) Sucrose concentration at the apoplastic interface between seed coat and cotyledons of developing soybean seeds. *Plant Physiol* **77**: 863–868
- Gomes CI, Obendorf RL, Horbowicz M (2005) myo-Inositol, D-chiro-Inositol, and D-Pinitol synthesis, transport, and galactoside formation in soybean explants. *Crop Sci* **45**: 1312–1319
- Grimberg Á, Carlsson AS, Marttila S, Bhalerao R, Hofvander P (2015) Transcriptional transitions in *Nicotiana benthamiana* leaves upon induction of oil synthesis by WRINKLED1 homologs from diverse species and tissues. *BMC Plant Biol* **15**: 192
- Hagely KB, Jo H, Kim J-H, Hudson KA, Bilyeu K (2020) Molecular-assisted breeding for improved carbohydrate profiles in soybean seed. *Theor Appl Genet* **133**: 1189–1200
- Hagely KB, Palmquist D, Bilyeu KD (2013) Classification of distinct seed carbohydrate profiles in soybean. *J Agric Food Chem* **61**: 1105–1111
- Heinrich P, Kohler C, Ellmann L, Kuerner P, Spang R, Oefner PJ, Dettmer K (2018) Correcting for natural isotope abundance and tracer impurity in MS-, MS/MS- and high-resolution-multiple-tracer-data from stable isotope labeling experiments with IsoCorrector. *Sci Rep* **8**: 17910
- Hernández-Sebastià C, Marsolais F, Saravitz C, Israel D, Dewey RE, Huber SC (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
- Hsu FC, Bennett AB, Spanswick RM** (1984) Concentrations of sucrose and nitrogenous compounds in the apoplast of developing soybean seed coats and embryos. *Plant Physiol* **75**: 181–186
- Hsu FC, Obendorf RL** (1982) Compositional analysis of in vitro matured soybean seeds. *Plant Sci Lett* **27**: 129–135
- Hu R, Fan C, Li H, Zhang Q, Fu Y-F** (2009) Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Mol Biol* **10**: 93
- Jones SI, Vodkin LO** (2013) Using RNA-seq to profile soybean seed development from fertilization to maturity. *PLoS One* **8**: e59270
- Kambhampati S, Aznar-Moreno JA, Hostetler C, Caso T, Bailey SR, Hubbard AH, Durrett TP, Allen DK** (2019) On the inverse correlation of protein and oil: examining the effects of altered central carbon metabolism on seed composition using soybean fast neutron mutants. *Metabolites* **10**: 18
- Kambhampati S, Kurepin LV, Kisiala AB, Bruce KE, Cober ER, Morrison MJ, Emery RJN** (2017) Yield associated traits correlate with cytokinin profiles in developing pods and seeds of field-grown soybean cultivars. *Field Crops Res* **214**: 175–184
- Kambhampati S, Li J, Evans BS, Allen DK** (2019) Accurate and efficient amino acid analysis for protein quantification using hydrophilic interaction chromatography coupled tandem mass spectrometry. *Plant Methods* **15**: 46
- Kanai M, Yamada T, Hayashi M, Mano S, Nishimura M** (2019) Soybean (*Glycine max* L.) triacylglycerol lipase GmSDP1 regulates the quality and quantity of seed oil. *Sci Rep* **9**: 8924
- Kappelmann J, Klein B, Geilenkirchen P, Noack S** (2017) Comprehensive and accurate tracking of carbon origin of LC-tandem mass spectrometry collisional fragments for ¹³C-MFA. *Anal Bioanal Chem* **409**: 2309–2326
- Kosina SM, Castillo A, Schnebly SR, Obendorf RL** (2009) Soybean seed coat cup unloading on plants with low-raffinose, low-stachyose seeds. *Seed Sci Res* **19**: 145–153
- Kuo TM, VanMiddlesworth JF, Wolf WJ** (1988) Content of raffinose oligosaccharides and sucrose in various plant seeds. *J Agric Food Chem* **36**: 32–36
- Leprince O, Pellizzaro A, Berriri S, Buitink J** (2016) Late seed maturation: drying without dying. *J Exp Bot* **68**: 827–841
- Li L, Hur M, Lee J-Y, Zhou W, Song Z, Ransom N, Demirkale CY, Nettleton D, Westgate M, Arendsee Z, et al.** (2015) A systems biology approach toward understanding seed composition in soybean. *BMC Genomics* **16**: S9
- Licht M** (2014) Soybean Growth and Development. Iowa State University Extension and Outreach, Iowa, USA
- Lin W, Oliver DJ** (2008) Role of triacylglycerols in leaves. *Plant Sci* **175**: 233–237
- McCleary BV, Charmier LMJ, McKie VA** (2019) Measurement of starch: critical evaluation of current methodology. *Starch-Stärke* **71**: 1800146
- McCleary BV, Gibson TS, Mugford DC** (1997) Measurement of total starch in cereal products by amyloglucosidase- α -amylase method: collaborative study. *J AOAC Int* **80**: 571–579
- de Mello Filho OL, Sediyaama CS, Moreira MA, Reis MS, Massoni GA, Piovesan ND** (2004) Grain yield and seed quality of soybean selected for high protein content. *Pesq Agropec Bras* **39**: 445–450
- Naeve SL** (2005) Soybean Growth Stages. University of Minnesota Extension, Minnesota
- O'Grady J, Schwender J, Shachar-Hill Y, Morgan JA** (2012) Metabolic cartography: experimental quantification of metabolic fluxes from isotopic labelling studies. *J Exp Bot* **63**: 2293–2308
- Osorio S, Vallarino JG, Szczowka M, Ufaz S, Tzin V, Angelovici R, Galili G, Aarabi F** (2014) Extraction and measurement of the activities of cytosolic phosphoenolpyruvate carboxykinase (PEPCK) and plastidic NADP-dependent malic enzyme (ME) on tomato (*Solanum lycopersicum*). *Bio-protocol* **4**: e1122
- Patil G, Mian R, Vuong T, Pantalone V, Song Q, Chen P, Shannon GJ, Carter TC, Nguyen HT** (2017) Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theor Appl Genet* **130**: 1975–1991
- Pereira Lima JJ, Buitink J, Lalanne D, Rossi RF, Pelletier S, da Silva EAA, Leprince O** (2017) Molecular characterization of the acquisition of longevity during seed maturation in soybean. *PLoS One* **12**: e0180282
- Pipolo AE, Sinclair TR, Camara GMS** (2004) Protein and oil concentration of soybean seed cultured in vitro using nutrient solutions of differing glutamine concentration. *Ann Appl Biol* **144**: 223–227
- Quoc TN, Anna K, Peter A, Emery RJN, Suresh N** (2016) Soybean seed development: fatty acid and phytohormone metabolism and their interactions. *Curr Genomics* **17**: 241–260
- Rainbird RM, Thorne JH, Hardy RWF** (1984) Role of amides, amino acids, and ureides in the nutrition of developing soybean seeds. *Plant Physiol* **74**: 329–334
- Raymond R, Spiteri A, Dieuaide M, Gerhardt B, Pradet A** (1992) Peroxisomal β -oxidation of fatty acids and citrate formation by a particulate fraction from early germinating sunflower seeds. *Plant Physiol Biochem* **30**: 153–161
- Rolletschek H, Radchuk R, Klukas C, Schreiber F, Wobus U, Borisjuk L** (2005) Evidence of a key role for photosynthetic oxygen release in oil storage in developing soybean seeds. *New Phytologist* **167**: 777–786
- Rolletschek H, Schwender J, König C, Chapman KD, Romsdahl T, Lorenz C, Braun HP, Denolf P, Van Audenhove K, Munz E, et al.** (2020) Cellular plasticity in response to suppression of storage proteins in the *Brassica napus* embryo. *Plant Cell* **32**: 2383–2401
- Rolletschek H, Weber H, Borisjuk L** (2003) Energy status and its control on embryogenesis of legumes. Embryo photosynthesis contributes to oxygen supply and is coupled to biosynthetic fluxes. *Plant Physiol* **132**: 1196–1206
- Ruuska SA, Schwender J, Ohlrogge JB** (2004) The capacity of green oilseeds to utilize photosynthesis to drive biosynthetic processes. *Plant Physiol* **136**: 2700–2709
- Salon C, Raymond P, Pradet A** (1988) Quantification of carbon fluxes through the tricarboxylic acid cycle in early germinating lettuce embryos. *J Biol Chem* **263**: 12278–12287
- Sánchez-Mata MC, Peñuela-Teruel MJ, Cámara-Hurtado M, Díez-Marqués C, Torija-Isasa ME** (1998) Determination of mono-, di-, and oligosaccharides in legumes by high-performance liquid chromatography using an amino-bonded silica column. *J Agric Food Chem* **46**: 3648–3652
- SchillingerJA, Dierking EC, Bilyeu KD**, Schillinger Genetics, Inc., The United States of America as represented by the secretary of agriculture (2013) Soybeans having high germination rates and ultra-low raffinose and stachyose content. US Patent Application No. 8,471,107
- SchillingerJA, Dierking EC, Bilyeu KD** (2018) Soybeans having high germination rates and ultra-low raffinose and stachyose content. US Patent Application No. 10081814
- Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y** (2004) Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* **432**: 779–782
- Schwender J, Ohlrogge JB** (2002) Probing in vivo metabolism by stable isotope labeling of storage lipids and proteins in developing (*Brassica napus*) embryos. *Plant Physiol* **130**: 347–361
- Schwender J, Shachar-Hill Y, Ohlrogge JB** (2006) Mitochondrial metabolism in developing embryos of *Brassica napus*. *J Biol Chem* **281**: 34040–34047
- Singh SK, Barnaby JY, Reddy VR, Sicher RC** (2016) Varying response of the concentration and yield of soybean seed mineral elements, carbohydrates, organic acids, amino acids, protein, and oil to phosphorus starvation and CO₂ enrichment. *Front Plant Sci* **7**: 1967
- Team RC (2013) A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

- Thompson JF, Madison JT, Muenster A-ME** (1977) In vitro culture of immature cotyledons of soya bean (*Glycine max* L. Merr.). *Ann Bot* **41**: 29–39
- Truong Q, Koch K, Yoon JM, Everard JD, Shanks JV** (2013) Influence of carbon to nitrogen ratios on soybean somatic embryo (cv Jack) growth and composition. *J Exp Bot* **64**: 2985–2995
- Tschiersch H, Borisjuk L, Rutten T, Rolletschek H** (2011) Gradients of seed photosynthesis and its role for oxygen balancing. *Biosystems* **103**: 302–308
- Valentine MF, De Tar JR, Mookkan M, Firman JD, Zhang ZJ** (2017) Silencing of soybean raffinose synthase gene reduced raffinose family oligosaccharides and increased true metabolizable energy of poultry feed. *Front Plant Sci* **8**: 692
- Walker RP, Chen Z-H, Técsi LI, Famiani F, Lea PJ, Leegood RC** (1999) Phosphoenolpyruvate carboxykinase plays a role in interactions of carbon and nitrogen metabolism during grape seed development. *Planta* **210**: 9–18
- Wickham H** (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer Publishing Company, New York