Research

New Phytologist

Expression of malic enzyme reveals subcellular carbon partitioning for storage reserve production in soybeans

Stewart A. Morley[1,2], Fangfang Ma[2], Mazen Alazem[2], Cheryl Frankfater[1,2], Hochul Yi[2], Tessa Burch-Smith[2], Tom Elmo Clemente[3], Veena Veena[2], Hanh Nguyen[4] and Doug K. Allen[1,2]

[1]United States Department of Agriculture, Agricultural Research Service, 975 N Warson Rd, St Louis, MO 63132, USA; [2]Donald Danforth Plant Science Center, 975 N Warson Rd, St Louis, MO 63132, USA; [3]Department of Agronomy & Horticulture, University of Nebraska-Lincoln, 202 Keim Hall, Lincoln, NE 68583, USA; [4]Center for Plant Science Innovation, University of Nebraska, N300 Beadle Center, 1901 Vine St., Lincoln, NE 68588, USA

Author for correspondence:
Doug K. Allen
Email: doug.allen@usda.gov

Received: *13 July 2022*
Accepted: *9 February 2023*

New Phytologist (2023) **239:** 1834–1851
doi: 10.1111/nph.18835

Key words: carbon partitioning, central carbon metabolism, lipid production, malic enzyme, metabolic flux, soybean seed composition, subcellular compartmentation.

Summary

- Central metabolism produces amino and fatty acids for protein and lipids that establish seed value. Biosynthesis of storage reserves occurs in multiple organelles that exchange central intermediates including two essential metabolites, malate, and pyruvate that are linked by malic enzyme. Malic enzyme can be active in multiple subcellular compartments, partitioning carbon and reducing equivalents for anabolic and catabolic requirements. Prior studies based on isotopic labeling and steady-state metabolic flux analyses indicated malic enzyme provides carbon for fatty acid biosynthesis in plants, though genetic evidence confirming this role is lacking. We hypothesized that increasing malic enzyme flux would alter carbon partitioning and result in increased lipid levels in soybeans.
- Homozygous transgenic soybean plants expressing Arabidopsis malic enzyme alleles, targeting the translational products to plastid or outside the plastid during seed development, were verified by transcript and enzyme activity analyses, organelle proteomics, and transient expression assays. Protein, oil, central metabolites, cofactors, and acyl-acyl carrier protein (ACPs) levels were quantified overdevelopment.
- Amino and fatty acid levels were altered resulting in an increase in lipids by 0.5–2% of seed biomass (i.e. 2–9% change in oil).
- Subcellular targeting of a single gene product in central metabolism impacts carbon and reducing equivalent partitioning for seed storage reserves in soybeans.

Introduction

The market value for soybeans (*Glycine max*) is derived from high-quality protein and oil seed reserves. Soybeans contain *c.* 40% protein by weight, the highest of any major crop (Egli, 1998), and the amino acid profile of soybean meal is the compositional standard for animal feed (Messina, 1999; Friedman & Brandon, 2001). Soybeans also contain *c.* 20% oil by weight, whereas other legumes such as peas and dry beans produce a greater percentage of starch. Though protein-enriched meal is the primary market driver for soybeans, oil has greater value on a per pound basis because of its utility in cooking applications and as a raw material for polymers, plastics, surfactants, and biofuel. In combination, soybean protein and oil are used to meet many food, feed, and industrial applications, supporting an annual production in the United States that approaches 100 million metric tons (USDA reports; usda.gov/media/agency-reports). The US soybean exports exceeded $27B in 2021 (https://www.fas.usda.gov/commodities/soybeans), and price per pound for soybean oil has more than doubled since 2019 (Ates & Bukowski, 2022). With such large demand, small improvements in seed protein and oil content have significant impacts on agriculture, export markets, and the US and other global economies (Clemente & Cahoon, 2009).

Improving soybean composition has proved challenging because oil and protein levels are negatively correlated across genotypes (Hymowitz *et al.*, 1972; Marega *et al.*, 2001; Truong *et al.*, 2013; Patil *et al.*, 2017; Kambhampati *et al.*, 2020), and protein is negatively associated with yield (Assefa *et al.*, 2018, 2019; de Borja Reis *et al.*, 2020). Lipid production in seeds and leaves has been augmented, for example, altering carbon provision, fatty acid synthesis, or triacylglycerol assembly by employing a 'push' or 'pull' approach (Vanhercke *et al.*, 2019); modulating the WRINKLED1 (WRI1) transcription factor, or acetyl-CoA carboxylase (Ye *et al.*, 2020; Wang *et al.*, 2022), or diacylglycerol acetyltransferase (DGAT) genes (Cernac & Benning, 2004; Shen *et al.*, 2010; Vanhercke *et al.*, 2013; Hofvander *et al.*, 2016; Roesler *et al.*, 2016; Chen *et al.*, 2020; Torabi *et al.*, 2021; Arias *et al.*, 2022). Strategies to increase oil have also included 'protecting' accumulated lipids by inhibiting lipases during seed maturation (Kanai *et al.*, 2019; Aznar-Moreno *et al.*, 2022) or strengthening the oil body packaging structure and formation through oleosins, seipins, and other lipid droplet associated proteins in seeds and leaves (Winichayakul *et al.*,

2013; Pyc *et al.*, 2017; Cai *et al.*, 2019; Beechey-Gradwell *et al.*, 2020; Greer *et al.*, 2020).

Soybean seed storage reserve biosynthesis requires a significant source of carbon and nitrogen that are supplied as sucrose, glucose, glutamine, and asparagine from vegetative source tissues (Hsu *et al.*, 1984; Rainbird *et al.*, 1984; Pipolo *et al.*, 2004; Hernández-Sebastià *et al.*, 2005; Allen & Young, 2013; Truong *et al.*, 2013; Kambhampati *et al.*, 2020, 2021). Imported sugars and amino acids are converted to other amino and fatty acids for protein and lipid production. Through isotopic labelling and metabolic flux analysis in soybeans (Sriram *et al.*, 2004; Iyer *et al.*, 2008; Allen *et al.*, 2009a; Allen & Young, 2013), differences in resource partitioning in central metabolism can quantitatively explain the final storage reserve composition. The high protein levels involve the transfer of glutamine and glutamate nitrogen to other carbon skeletons for amino acid biosynthesis coproducing 2-oxoglutarate, that is partially used for fatty acid biosynthesis (Allen, 2016a). The pool of 2-oxoglutarate is converted through multiple steps to acetyl-CoA and results in production of reducing equivalents including NADH, NADPH, and $FADH_2$, some of which are used to make ATP in the mitochondria. Two key enzymes, malic enzyme, and pyruvate dehydrogenase have been linked specifically to fatty acyl chain biosynthesis. Evidence from isotopic labeling-based flux maps of soybean has suggested malic enzyme is the source of a significant amount of pyruvate that is the precursor of acetyl-CoA used in lipid biosynthesis. The involvement of this step produces 10–20% of lipid carbon that originates from glutamine (Fig. 1; Allen *et al.*, 2009b; Allen & Young, 2013). Hence, increasing the activity of malic enzyme may be an effective strategy to 'push' carbon into fatty acid synthesis for oil and could additionally impact the 'pull' of carbon and nitrogen supply from vegetatively supplied glutamine in soybeans impacting protein content.

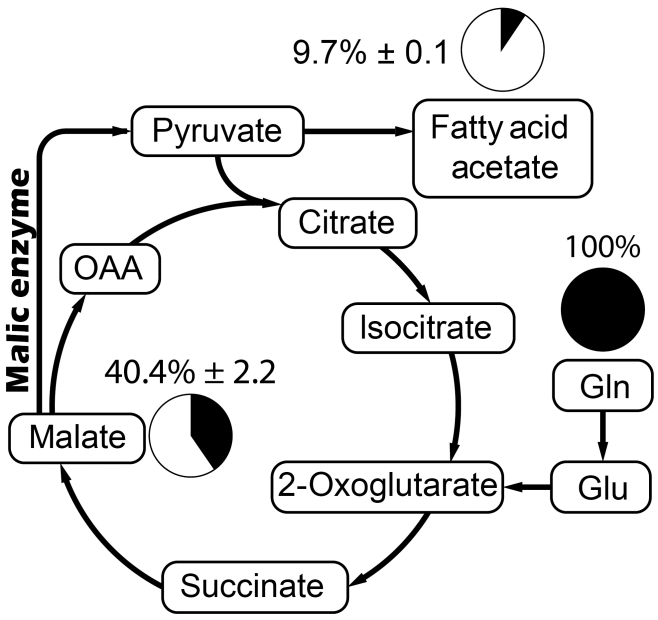


Fig. 1 Incorporation of U-^{13}C glutamine into malate and fatty acid acetate. When *Glycine max* seeds were cultured with U-^{13}C glutamine nearly 40% of malate and 10% of carbon in fatty acids became labeled (inspired by Allen *et al.*, 2009b). ± represents 1 SD of measured replicates. OAA, oxaloacetate.

In fungi and algae, research has demonstrated that malic enzyme contributes significantly to lipid accumulation. Mutants of the fungus *Aspergillus nidulans* lacking malic enzyme activity accumulate approximately one-half the lipid observed in wild-type (Wynn & Ratledge, 1997). Lipid levels in filamentous fungi have been directly correlated with malic enzyme activity (Wynn *et al.*, 1999; Zhang *et al.*, 2007) though not other lipogenic enzymes (Wynn *et al.*, 1999) suggesting that malic enzyme might be a bottleneck or control point for metabolism. In algae, $NADP^+$-dependent malic enzyme overexpression resulted in *c.* 1.5-fold increase in lipid content (Zhu *et al.*, 2018); however, subcellular evaluation of the activity was not established and it remains less clear which isoforms and subcellular location in plants are crucial to oilseed metabolism (Allen *et al.*, 2009a; Allen, 2016a).

In the current study, malic enzyme was enhanced within or outside of the chloroplast in soybean. The seed-specific promoter from the beta-conglycinin storage protein was used to regulate expression of transgenic alleles, targeting the translational products encoding the enzymatic activity to the respective organelles. Soybean events carrying the transgenic alleles were examined over the course of seed development for effects on metabolism including the production of oil and protein. The results indicated changes in lipid levels, and fatty and amino acid content, dependent on the subcellular location of expressed proteins. Carbon partitioning was substantially altered indicating a single pathway step in central metabolism can significantly affect plant storage reserve metabolism in oilseeds with important consequences for metabolic engineering efforts.

Materials and Methods

Materials

Unless referenced otherwise, solvents, buffers, and chemicals were obtained from MilliporeSigma, Burlington, MA, USA.

Plant growth and glasshouse conditions

Glycine max (cv Thorne (McBlain *et al.*, 1993)) was grown in glasshouses with 14-h photoperiods in one-gallon pots with Berger BM7 bark soil (10121500; Hummert, Earth City, MO, USA), watered twice daily, and fertilized three times weekly with 15–16–17 nitrogen–phosphate–potash. Supplemental light was provided to ensure a 14 h : 10 h, day : night period with corresponding temperatures of 22–25°C : 20–23°C and a relative humidity of > 40%.

In vivo culturing of *G. max* seeds with U-^{13}C Glutamine

Metabolites from previously cultured embryos, described in Allen & Young (2013), were extracted and analyzed with an LC–MS/MS ion pairing approach using an AB Sciex QTRAP 4000 connected to an LC-20AD XR chromatogram from Shimadzu. Ion pairing-based metabolite analysis was identical to that of Ma *et al.* (2014, 2017).

Ortholog gene alignment, phylogeny, and structure comparison

Arabidopsis ME2 and *ME4* gene orthologs (referred to as *AtME2* and *AtME4*) were identified in *G. max* using NCBI BLAST and SOYBASE database resources (blast.ncbi.nlm.nih.gov and soybase.org respectively). *Arabidopsis* and orthologous *G. max* genes were aligned using the MAFFT alignment program (Katoh *et al.*, 2018; Supporting Information Fig. S1a). The sequence alignment was used to create a phylogenetic tree with RAxML (Stamatakis, 2006) resulting in an evolutionary model that was unpartitioned, and based on a LG substitution matrix, stationary base frequencies, a single gamma mean category rate for the proportion of invariant sites among-site rate heterogeneity, and no ascertainment bias correction as described in the software documentation (Fig. S1b). Analysis parameters including topology, branch lengths, and model were optimized with 10 trees for parsimony and randomness. Bootstrapping was performed by bootstopping with a cutoff of 0.03. AlphaFold protein structures (Jumper *et al.*, 2021) from Uniprot of *Arabidopsis* and *G. max* malic enzyme were used for structural and active site comparison against the crystal structures of *H. sapiens* NAD^+-dependent and *Z. mays* $NADP^+$-dependent orthologs (Tao *et al.*, 2003; Alvarez *et al.*, 2019) using the Matchmaker function included in the protein visualization software CHIMERA (Pettersen *et al.*, 2004) (Fig. S2).

Vectors construction and plant transformation

AtME2 (At4g00570, GenBank accession no. NP_191966.2) and *AtME4* (At1g79750, GenBank accession no. NP_178093.1) sequences were codon-optimized for *G. max* expression and synthesized by GenScript (Piscataway, NJ USA). The *AtME2* and *AtME4* expression cassettes were under the control of the *G. max* seed-specific beta-conglycinin promoter (Allen *et al.*, 1989), fused to the tobacco etch virus translational enhancer element (Carrington & Freed, 1990), and terminated with the 35S cauliflower mosaic virus polyadenylation signal. Two gene cassettes were assembled into separate vectors, referred to as pPTN1062 and pPTN1061 respectively, harboring the *bar* gene (Thompson *et al.*, 1987) under control of the Agrobacterium Pnos promoter for herbicide selection (Fig. S3). The vectors were mobilized into *Agrobacterium tumefaciens* strain EHA101 (Hood *et al.*, 1986) via triparental mating and transformed into cv Thorne (Zhang *et al.*, 1999; Xing *et al.*, 2000). Transformants were identified through monitoring the expression of the *bar* gene via leaf painting (Zhang *et al.*, 1999) and genomic DNA isolated following a modified CTAB method (Springer, 2010), which was used for PCR genotyping transgenic plants (primers used, Table S1). Transgenic allelic insertion numbers were verified by qPCR Taqman assays with probes designed to anneal to the *bar* gene (Grohmann *et al.*, 2009). Combined with previous screens, the observation of two allelic insertions confirmed homozygosity in *AtME2* and *AtME4* plants.

Digital droplet PCR (ddPCR)

Primers were generated to target a housekeeping gene, *Elf1b* (Glyma02g276600) (Jian *et al.*, 2008; Yim *et al.*, 2015) for normalization, endogenous $NAD(P)^+$-dependent *G. max* malic enzyme genes, and transgenic *Arabidopsis* malic enzyme genes (Table S2). RNA was harvested from plants at stages R5-R8 of development, extracted using TRIzol reagent (15596026; Thermo Fisher, Waltham, MA, USA), and quantified by Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA). Biological replicates of RNA (1 μg) were converted to cDNA using the SuperScript IV Vilo kit from Thermo Fisher Scientific (11756050) and qPCR performed to obtain C_t values and melt curves using a Bio-Rad CFX384 (Bio-Rad) to optimize cDNA concentration for ddPCR and verify generation of a single PCR product. ddPCR was performed on a Bio-Rad QX200 and analyzed with QUANTASOFT software (v.1.7.4; Bio-Rad). Copy numbers were averaged across biological replicates and then normalized against expression of *Elf1b*. Final expression values represent each gene compared with the expression of *Elf1b*.

Malate and pyruvate pool size quantitation

Lyophilized tissue obtained over development was spiked with isotopic standards of malate (641049; MilliporeSigma) and pyruvate (490717; MilliporeSigma) before extraction. Metabolites were extracted with 7 : 3 methanol : chloroform and phase-separated by addition of 30% water. Phenylhydrazine derivatization allowed for more sensitive detection of some metabolites by mass spectrometry (Zimmermann *et al.*, 2014). Separations were performed with an Imtakt Intrada Amino Acid column (Imtakt, Portland, OR, USA) using buffers and gradient conditions as described in Koley *et al.* (2022) and Methods S1.

Quantification of oxidized and reduced cofactors

Levels of NAD^+, $NADP^+$, NADH, and NADPH were measured by mass spectrometry, adapting the method of Lu *et al.* (2018) for plant tissue. In brief, seed tissue was harvested and weighed before being homogenized in a solution of 40 : 40 : 20 acetonitrile : methanol : water with 0.1 M formic acid. Samples were neutralized with 15% NH_4HCO_3, incubated on ice, and then centrifuged to remove solid debris. Samples were then analyzed with mass spectrometry to obtain peak areas for each metabolite. LC–MS/MS details are provided in Methods S2.

Transient expression of AtME2 and AtME4 in *Nicotiana benthamiana*

Subcellular predictions of *AtME2* and *AtME4* were made by uploading amino acid sequences to the TARGETP-2.0 server (https://services.healthtech.dtu.dk/service.php?TargetP-2.0). *AtME2* and *AtME4* were cloned by restriction digest/ligation into a binary vector carrying GFP for C-terminal fusions (pXM82). Leaves from 4-wk-old *N. benthamiana* plants were infiltrated with *A. tumefaciens* GV3101 strains carrying the fusion constructs at OD_{600} = 0.5. Samples were imaged with a Leica SP8 confocal microscope (Leica Microsystems, Deerfield, IL, USA) 2 d after infiltration. Mitotracker Red CMXRos (Cell Signaling Technology, Danvers, MA, USA) and chlorophyll autofluorescence were

used to image mitochondria and chloroplasts, respectively. Leaf sections were infiltrated with 1 : 1000 dilution of Mitotracker Red stock solution made with DMSO for 10 min before imaging. Images were captured using A ×40 lens (HC PL APO CS2 ×40/1.10 water), with bidirectional scanning at a speed of 400 Hz and a resolution of 2048 × 2048. Excitation wavelengths for GFP and Mitotracker Red were 488 and 580 nm, respectively. Emission wavelengths of 505–515 nm for GFP and 590–630 nm for Mitotracker Red were collected. Chlorophyll autofluorescence was collected after excitation at 580 nm. The Leica application suite X (LASX) package was used to analyze images.

Organelle isolation and proteomic analysis

Mitochondria were purified from *AtME2-1*, *AtME2-2*, and *AtME2-3* seeds using a density gradient centrifugation strategy (Huang *et al.*, 2014). Briefly, *c.* 50 g of R6 and R7 seeds were harvested and blended in homogenization buffer (0.4 M sucrose, 5 mM EGTA, 50 mM sodium pyrophosphate–KOH, pH 7.5, 0.5% (w/v) BSA, 10 mM ascorbate, and 1% (w/v) PVP-40). Following homogenization, crude lysate was strained through Miracloth and centrifuged to remove cellular debris and enriched for mitochondria by differential centrifugation. This enriched pellet was resuspended and layered on top of a Percoll step gradient that after centrifugation resulted in an isolated mitochondrial band. Fractions of crude lysate, enriched mitochondria, and purified mitochondria were collected along the purification process and proteins were evaluated by proteomics analysis as described in Methods S3.

Enzyme activity assay

Flash-frozen seeds from R5-R8 stages of development were homogenized in extraction buffer (100 mM Tris–HCl pH 8, 5 mM $MgCl_2$, 2 mM EDTA, 10% v/v glycerol, 10 mM β-mercaptoethanol) by bead mill (MM 400 Mixer Mill; Restch, Haan, Germany). The combination of *c.* 50 mg of fresh tissue and 200 µl of extraction buffer produced consistent homogenate that was clarified by centrifugation at 16 000 ***g*** for 15 min, and supernatant was removed and centrifuged a second time. Enzymatic activity was measured immediately, or samples were flash frozen in liquid nitrogen and stored at −80°C. Malic enzyme activity was measured spectrophotometrically using a 96-well microplate reader by recording the change in absorbance at 340 nm over time corresponding to NAD(P)H production as detailed in Methods S4.

Fatty acid methyl ester analysis for oil content

Four to six seeds from each plant across development were flash frozen, crushed to a fine powder, and lyophilized. 15–30 mg of tissue was weighed into 8 ml glass vials with 0.2 and 1 mg of C15:0 and C17:0 TAG standards (T4257 and T2151; MilliporeSigma), respectively. Two milliliters of a mixture of methanol, concentrated sulfuric acid, and 0.2% w/v butylated hydroxytoluene (101162; MP Biomedicals, Irvine, CA, USA) were combined by volume (19 : 1 : 0.250, v/v/v) and aliquoted per tube. Fatty acid methyl esters (FAMEs) were generated by incubating at 110°C for 3 h with hourly vortexing, then cooled, % 0.9 w/v NaCl (3 ml) was added to quench the reaction, and extracted with hexane. The FAMEs in the hexane organic layer were quantified by GC-FID (Thermo Fisher Focus GC) with Thermo Fisher XCALIBUR software (v.4.0; Thermo Fisher) relative to C15 and C17 standards.

Acyl carrier protein analysis

Acyl carrier proteins (ACPs) were measured by following the method described by Nam *et al.* (2020). Briefly, fresh *G. max* seed tissue was homogenized in a solution of 5% trichloroacetic acid (TCA). Precipitated proteins were pelleted by centrifugation, washed with 1% TCA, and resuspended in MOPS buffer (50 mM, pH 7.6). Proteins were precipitated again by adding TCA to a concentration of 10%, followed by centrifugation and washing of the pellet with 1% TCA. Precipitated proteins were then resuspended in 50 µl of MOPS buffer and digested with Endoproteinase AspN (P3303; MilliporeSigma). Digestions were quenched with the addition of 50 µl methanol, centrifuged at 15 000 ***g***, and acyl-ACPs in the supernatant analyzed using a QTRAP 6500 mass spectrometer from Sciex (Framingham, MA, USA) connected to an LC-20AD XR Shimadzu liquid chromatography pump with settings as described in Jenkins *et al.* (2021).

Total protein and free amino acid analysis

Three methods were employed to measure total protein. NIR was performed on mature seeds using a Perten DA 7250 (PerkinElmer, Waltham, MA, USA). Total protein was estimated using instrument default settings calibrated for *G. max* (PerkinElmer). As a second approach, C : N analysis was performed by the Center for Applied Isotope Studies at the University of Georgia, using the Dumas combustion analysis method (Kirsten, 1983). Total protein was calculated assuming a Jones factor of 6.25. A third strategy for protein quantification included acid hydrolysis that was performed by reacting lyophilized tissue with 4 M methanesulfonic acid at 110°C for 22 h. Samples contained an isotope internal standard mixture of amino acids (MSK-A2-1.2; Cambridge Isotope Laboratories, Andover, MA, USA). Following digestion, the acid was neutralized, and the sample concentrated to a final concentration of 10 µM internal standard amino acid mixture. Quantification was performed on a Sciex QTRAP 6500 mass spectrometer connected to an LC-20AD XR Shimadzu liquid chromatography pump. Mass spectrometry conditions were the same as described previously (Kambhampati *et al.*, 2019).

Free amino acids (*c.* 20 mg of lyophilized tissue) were quantified in the presence of an equimolar U-^{13}C internal isotope standard amino acid mixture (MSK-CAA-1; Cambridge Isotope Laboratories). Samples were extracted with 7 : 3 (v/v) chloroform : methanol using a bead mill (MM 400 Mixer Mill; Restch) then inverted for 2 h at 4°C on a rotating platform.

Water (400 µl) was added, and the aqueous phase concentrated to a final concentration of 10 µM internal standard amino acid mixture before analysis using the previously described mass spectrometry approach. Quantitation of free amino acids was performed identical to amino acids resulting from protein hydrolysis with additional *m/z* transitions for asparagine and glutamine, using a standard curve isotope dilution strategy (Kambhampati *et al.*, 2019).

Results

Culturing of soybean seeds with U-$^{13}C_5$ Glutamine indicates a role for malic enzyme in seeds

Prior flux studies (Allen *et al.*, 2009b; Allen & Young, 2013) with U-$^{13}C_5$-glutamine resulted in labeling in fatty acids and branched-chain amino acids, but not aromatic amino acids, all of which are synthesized in the plastid. The investigations inferred that pyruvate but not phosphoenolpyruvate (PEP) was isotopically enriched; however, more complicated descriptions including label disequilibrium across subcellular pool locations could not be exhaustively ruled out. To more conclusively evaluate the pathway of labeling that led to labeled lipids, an LC–MS/MS approach (Ma *et al.*, 2014, 2017) was used to quantify central intermediate isotopologs from U-$^{13}C_5$-glutamine cultured soybeans with varied carbon-to-nitrogen ratios of 13, 21, and 37 (Allen & Young, 2013). ^{13}C was incorporated into pyruvate at significant levels (15–20% when the carbon-to-nitrogen ratio was 21) indicating that glutamine carbon supplied up to 1/5th of the total carbon for this pool in developing soybeans (Fig. 2; Table S3). PEP and other glycolytic intermediates including fructose-6-phosphate and dihydroxyacetone phosphate were unlabeled, indicating that phosphoenolypyruvate carboxykinase (PEPCK) is not actively shuttling carbon during seed filling stages of metabolism.

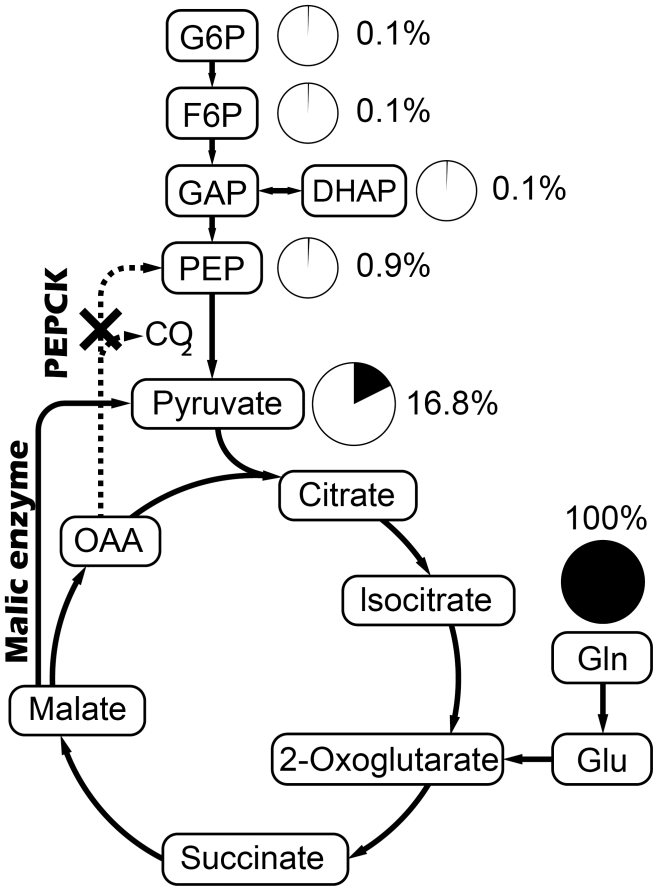


Fig. 2 Incorporation of ^{13}C-glutamine into pyruvate but not PEP indicated PEPCK was not active in *Glycine max* seeds. Nine samples, three from each of C : N 13, C : N 21, and C : N 37 were used to inspect labeling of phosphorylated sugars from the supply of ^{13}C-glutamine by LC–MS/MS. C : N 21 presented, C : N 13, and 37 in Supporting Information Table S3 with statistics, labeling corrected for natural abundance. DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde-3-phosphate; Gln, glutamine; Glu, glutamate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PEPCK, phosphoenolypyruvate carboxykinase.

Generation of homozygous transgenic soybeans and expression of transgenes

Transgenic soybean alleles were generated carrying the well-annotated *Arabidopsis* At4g00570 (*AtME2*, NAD^+ dependent or NAD-ME) and At1g79750 (*AtME4*, $NADP^+$ dependent or NADP-ME) seed expression cassettes, noted as mitochondrial and plastidic, respectively. Arabidopsis genes were chosen based on available annotations at the time of cloning and subsequently compared with sequenced soybean genome orthologs. Though it is possible other nonplant orthologs could have further enhanced activities, the potential incompatibilities and low throughput of transgenic studies in soybean precluded consideration of options with more risk, in this first study in plants. For simplicity, the beta subunit of NAD-ME that is known to be active was expressed (Tronconi *et al.*, 2008).

Multiple sequence alignment of *Arabidopsis AtME2* and *AtME4* with *Glycine* orthologs indicated 62–81% homology for NAD type ME and 51–79% for NADP type ME (Fig. S1a). Phylogenetic analysis revealed distinct clades for each cofactor type with high similarity between soybean and Arabidopsis orthologs (Fig. S1b). Additionally, multiple sequence alignment of *Arabidopsis* and *G. max* malic enzymes with solved structures of human NAD^+-dependent (Tao *et al.*, 2003) and maize $NADP^+$-dependent (Alvarez *et al.*, 2019) malic enzyme showed high conservation of residues involved in the active site binding of malate or pyruvate (Fig. S2a). Structural comparison of solved crystal structures with predicted structures of *Arabidopsis* and *G. max* orthologs also showed high similarity (e.g. in Fig. S2b,c). Open reading frames were codon-optimized for soybean. For each transformation event, at least 18 plants were screened for homozygous patterns of herbicide resistance. After herbicide screening, events were selected for in-depth analysis. Homozygosity in five events was verified based on Taqman assays designed to confirm the presence of two allelic insertions of bar gene (i.e. one per chromosome) per transgenic plant. The combined assays provided a conclusive screen that ensured single transgenic insertions in three *AtME2* and two *AtME4* homozygous events (Tables 1, S4; Fig. S4).

Gene expression of transgenic malic enzyme was assessed over seed development with digital droplet PCR (ddPCR). Endogenous

Table 1 Homozygous transgenic *Glycine max* selected for study.

Event	Gene	Predicted localization	NAD(P)$^+$ dependence
AtME2-1	At4g00570	Mitochondria	NAD^+
AtME2-2	At4g00570	Mitochondria	NAD^+
AtME2-3	At4g00570	Mitochondria	NAD^+
AtME4-1	At1g79750	Plastid	$NADP^+$
AtME4-2	At1g79750	Plastid	$NADP^+$

NAD-ME genes (*Glyma03g24630* and *Glyma07g08110*) and endogenous NADP-ME genes (*Glyma13g43130* and *Glyma15g02230*) were compared with the expression of the transgenic alleles (Table S5). Elongation factor 1-beta (*Elf1b*) served as a reference gene to normalize expression values (Jian *et al.*, 2008; Reddy *et al.*, 2013; Yim *et al.*, 2015). At the R5 stage of development, neither *AtME2* nor *AtME4* events displayed differences in expression relative to endogenous genes; however, peak expression at R6 or R7 (Fig. 3b) was 4–40 times more than native levels (Table S5). Transgene expression dropped to wild-type levels near maturity (i.e. R8 stage).

Transient expression assays indicate chloroplast and extraplastidic locations

Though genes were chosen based on evidence for subcellular targeting to chloroplast and mitochondria from literature (Macrae, 1971; Day & Wiskich, 1974; Hatch *et al.*, 1974), partial cytosolic localization for the *AtME2* gene product has been suggested by a more recent study (Ito *et al.*, 2011). TargetP v.2.0 software was used to predict subcellular localization of each gene product (Armenteros *et al.*, 2019). Results supported mitochondrial and chloroplast localizations for *AtME2* and *AtME4* respectively, with much stronger mitochondrial signals for *AtME2* relative to chloroplast signals for *AtME4* (Fig. S5).

Plastidic and extraplastidic subcellular locations for *AtME4* and *ATME2* were further explored through transient expression in *N. benthamiana* leaves. Subcellular localization was determined through confocal fluorescence microscopy by infiltrating *N. benthamiana* leaves with *Agrobacterium*-carrying constructs expressing GFP fused to the C-terminus of *AtME2* or *AtME4* under the control of the 35S *CaMV* promoter. Mitotracker Red CMXRos and chlorophyll autofluorescence were used to visualize mitochondria and chloroplasts, respectively. Free GFP was also expressed in *N. benthamiana* leaves as a positive control and displayed expected nuclear and cytosolic localization patterns. *AtME2-GFP* was detected in the cytoplasm and mitochondria but not the chloroplast (Fig. 3c). *AtME4-GFP* was localized to the chloroplast stroma with a minor fraction of signal detected in the cytoplasm.

Proteomic evaluation of organelles indicates enriched expression in subcellular locations

As transient expression with a constitutive promoter in *N. benthamiana* leaves may not completely reflect the cells of filling soybean seeds, organelle proteomic analysis of *AtME2's* subcellular localization was performed with enriched mitochondria from *AtME2-1*, *AtME2-2*, and *AtME2-3* plants. Fractions collected at three different points along the purification process were analyzed by proteomics resulting in the identification of 2909 unique proteins. Proteins that were not identified in all three fractions or that had low false discovery rate confidence were excluded from analysis. Remaining proteins were classified as cytosolic or mitochondrial using DeepLoc2.0 prediction algorithms. Within this group of proteins, known mitochondrial and cytosolic proteins were specifically identified from established literature (Millar *et al.*, 2001; Oehrle *et al.*, 2008; Ito *et al.*, 2011) and Uniprot annotations. A heatmap showing normalized abundances of mitochondrial and cytosolic proteins from all three fractions indicated an enrichment of mitochondrial proteins and a dilution of cytosolic proteins (Fig. 4) through the mitochondrial isolation process. *AtME2* abundance increased with isolation of mitochondria. The sequence differences between the native soybean malic enzyme and the *AtME2* proteomic products avoided potential ambiguity between identified proteins. Apart from the alpha and beta subunits of ATP synthase that function as part of the electron transport chain, *AtME2* was the third most abundant protein and greater than common mitochondrial marker proteins including but not limited to aconitate hydratase (2.5-fold), NADH ubiquinone reductase (45-fold), and formate dehydrogenase (400-fold) (Tables S6, S7). A number of plastid proteins were enriched during mitochondrial isolation, possibly from thylakoid membranes of broken chloroplasts that can co-enrich with the mitochondria. Soybean seed chloroplasts have storage plastids (Borisjuk *et al.*, 2005) that can rupture chloroplasts during purification and result in contaminated organelle fractions (Steinmüller & Tevini, 1985). Nonetheless, the combined observations from transient expression and proteomics confirmed the localization of *AtME4* to chloroplasts and indicated that *AtME2* protein was outside the chloroplast and some protein was localized to the mitochondria.

Enzyme assays confirm activity of transgenic malic enzyme in the plastid

To confirm functionality of the malic enzyme transgenic alleles, NADP-ME, NAD-ME, and NAD-malate dehydrogenase enzyme activities were measured in wild-type and transgenic seeds across development. Seeds were collected from multiple plants for each event of each genotype per developmental stage. Wild-type levels of NAD-ME, NADP-ME, and NAD-MDH activity were highest at the R5 stage for all enzymes and decreased as the seed matured (Fig. 5a; Table S8). Compared with wild-type, *AtME4* events displayed roughly 5–10 times the activity of NADP-ME beginning at the R6 stage in development and lasting until maturity. NAD-ME activity could not be reliably determined due to interference from MDH that favors the generation of NAD (Guynn *et al.*, 1973; Bowman & Ikuma, 1976) and can be in excess of 100-fold that of NADP-ME (Gerrard Wheeler *et al.*, 2016). Attempts to inhibit the malate dehydrogenase activity using historical inhibitors phenol, albendazole, and mebendazole (Henneke & Wedding, 1975; Tejada *et al.*, 1987) were unsuccessful and not pursued further as they can result in artifacts.

New Phytologist

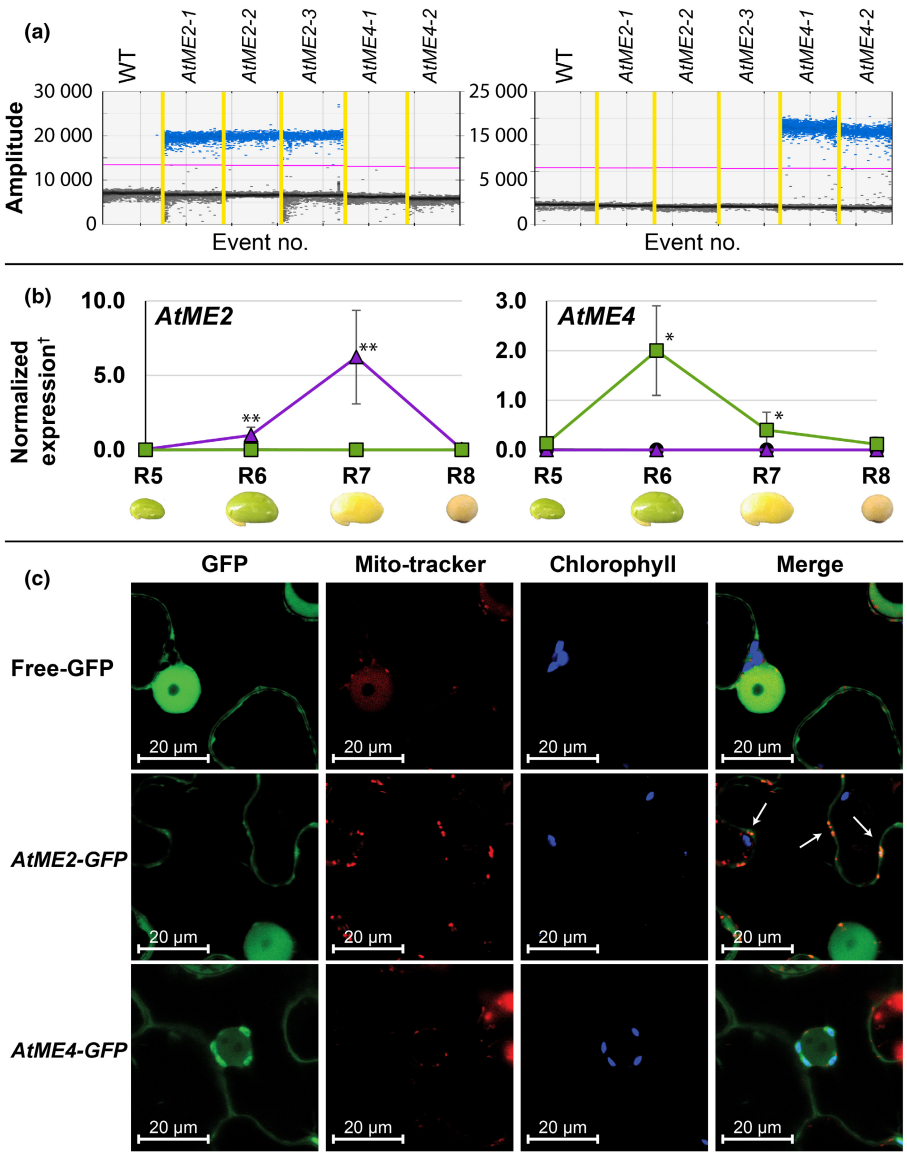


Fig. 3 Gene and transient expression of *AtME2* and *AtME4* transgenes. (a) Representative ddPCR plot of positive droplets for multiple *Glycine max* plants from each independent event (*AtME2*, $n = 6$; *AtME4*, $n = 4$) at R6 stage (other developmental stages confirm expression differences in Supporting Information Table S5). (b) Normalized expression of combined *AtME2* (▲, $n = 6$, ±1 SD) and *AtME4* events (■, $n = 4$, ±1 SD) in *AtME2* or *AtME4* transgenic plants overdevelopment. †Normalized expression relative to *Elf1b*. α: **, < 0.01; *, < 0.05 based on two-tailed *t*-test assuming unequal variance with $n \geq 4$ replicates. (c) Confocal imaging of transient expression of *AtME2* and *AtME4* in *Nicotiana benthamiana*. GFP linked to the C-terminal end of *AtME2* and *AtME4* under the control of the constitutive 35S *CaMV* promoter was infiltrated into *N. benthamiana* using *Agrobacterium tumefaciens*. Mitotracker Red CMXRos and chlorophyll autofluorescence were used for mitochondria and chloroplast visualization. Unlinked GFP was infiltrated as a positive control of GFP. White arrows in the *AtME2-GFP* merge panel indicate instances of colocalization of GFP and mitotracker signals.

Malate and pyruvate pools changed substantially in transgenic soybean alleles

The direct substrate and product of malic enzyme, that is, malate and pyruvate, were quantified during reproductive development stages in transgenic soybeans. Consistent with other measurements, no changes in malate or pyruvate level were observed at the R5 stage of development (Fig. 5b); however, beginning at R6, malate levels were reduced in all transgenic events at all stages of development. Absolute changes in pyruvate were less pronounced due to the smaller pool size though increased pyruvate levels were easily visualized at R6 for *AtME4* and R6 and R7 for *AtME2*.

Seed sink strength was unchanged

Seed mass is a strong indicator of overall sink strength (Ho, 1988). To determine whether sink strength in the transgenic malic enzyme plants was altered, seeds from multiple plants, generations, and times of year were weighed. Transgenic seed weight did not increase or decrease relative to wild-type (Fig. 6); though variation of up to 50 mg was observed due to environmental conditions that vary with seasons.

Enhanced malic enzyme activity in the plastid resulted in increased seed oil levels and altered fatty acid profiles

Seed oil levels were quantified overdevelopment by transmethylation that is commonly used for seed oil analysis (Li *et al.*, 2006). Changes in *AtME2* and *AtME4* events trended similarly; though *AtME2* events contained less oil at mid-development (Fig. 7a; Table S9) and were not significantly different from wild-type at maturity. *AtME4* events had increased oil levels that were measured over additional generations (0.5–2% increase as a fraction of seed biomass, that is 2–9% change as a fraction of total oil; Fig. 7b; Table S9). FAME analysis indicated *AtME4* events

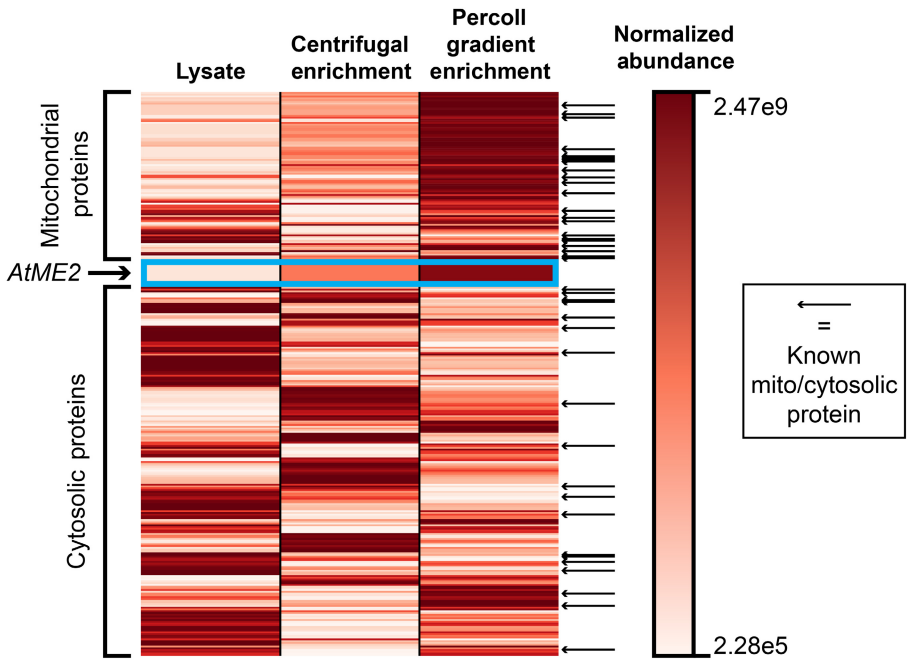


Fig. 4 Heatmap of normalized protein abundances from purified *Glycine max* seed mitochondria isolated through Percoll density gradient centrifugation show *AtME2* enrichment. Successive protein fractions were obtained from total lysate, centrifugal enrichment of mitochondria, and mitochondrial purification via Percoll density gradient centrifugation. Each fraction was analyzed by LC–MS proteomic workflows to determine protein abundances in each successive fraction of mitochondrial purification. The results show an enrichment of mitochondrial proteins (top section) and a dilution of cytosolic proteins (bottom section). *AtME2* enrichment correlates with overall enrichment of mitochondrial proteins (middle section). Mitochondrial and cytosolic proteins were identified using DeepLoc2.0 and known mitochondrial and cytosolic proteins from published literature (indicated by black arrows on the right-hand side of the heat map). Identified mitochondrial and cytosolic proteins are provided in Supporting Information Table S6 with corresponding references from Millar *et al*. (2001), Oehrle *et al*. (2008), and Ito *et al*. (2011).

contained 10% more oleic acid and 10% less linoleic acid as a percentage of total lipid content starting at R6 and persisting until maturity (Fig. 7c). Enhanced malic enzyme activity in *AtME2* events over seed development resulted in slightly depressed levels of oleic acid and higher levels of linoleic acid. The greatest changes in fatty acid profiles for *AtME2* and *AtME4* events coincided with peak gene expression at R7 and R6 stages respectively (Fig. 3b; Table S5).

Ratios of reduced to oxidized cofactors indicate changes in seed metabolism

Given the coproduction of NAD(P)H cofactors with malic enzyme activity, the levels of reduced and oxidized cofactors were measured by LC–MS/MS. Unlike malate and pyruvate that have a select number of alternative reactions through which they are produced or consumed, the production and turnover of reduced and oxidized forms of cofactors involves many reactions that could reduce the measurable impact of the genetic changes. Nonetheless, the levels of reduced cofactors were increased consistent with the function of malic enzyme (Fig. 8a); though some differences were not statistically significant and may reflect metabolic homeostasis, malate shuttles, or the impact of other cofactor-requiring reactions. The pairwise ratios of reduced to oxidized cofactors enable anabolic processes (i.e. NADPH : $NADP^+$), and a currency supply for electron transport (i.e. NADH : NAD^+) (Heldt, 2005) and were calculated directly from measured values. The ratio of NADPH : $NADP^+$ was increased *c.* twofold at R6 and R7 in both *AtME4* and *AtME2* lines relative to WT and the ratio of NADH : NAD^+ was increased *c.* threefold at R6 and twofold at R7 in *AtME2* lines relative to WT (Fig. 8b) comparable to ranges reported in the literature for other systems (Heineke *et al.*, 1991; Henry *et al.*, 2007).

Relative comparisons of acyl-ACP intermediates in transgenic tissue

Acyl-acyl carrier proteins (ACPs) provide the scaffold for fatty acid biosynthesis and are an essential component of lipid metabolism. Following methods that were established to quantify absolute levels of acyl-ACPs (Nam *et al.*, 2020; Jenkins *et al.*, 2021), we assessed whether increases in malic enzyme activity resulted in changes in ACP levels. Relative ratios of each acyl-ACP were determined by comparing peak areas from mass spectrometry. Individual acyl-ACPs have not been previously reported for soybean temporally and could only be measured at R5 and R6, due to reduced levels with developmental progression. No difference in acyl-ACP levels was observed between transgenic and WT plants at R5 or R6; however, longer chain acyl-ACPs, specifically C14 chain length, increased at the R6 stage relative to R5 (Fig. 8c) for all plants. This increase in C14 acyl-ACP may be unique to soybean and appears to differ from recently quantified levels in other species (Nam *et al.*, 2020; Chu *et al.*, 2022).

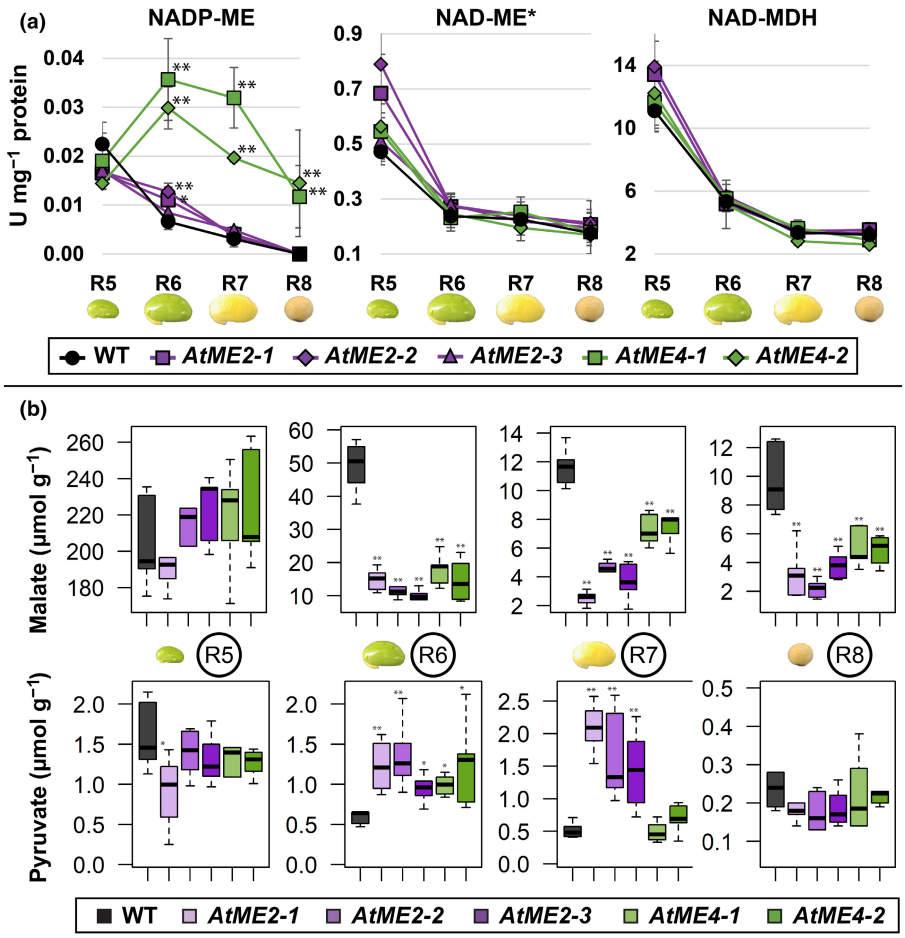


Fig. 5 Transgenic malic enzyme activity and altered metabolite pools in *Glycine max* seeds. (a) Activity of $NADP^+$ or NAD^+-dependent malic enzyme was quantified by measurement of NADPH or NADH generation from crude tissue extract. *NAD-ME activity was measured but could not be accurately represented due to contaminating activity from malate dehydrogenase (NAD-MDH). Error bars represent sample SD with $n \geq 5$ degrees of freedom. Two-tailed *t*-test: **, < 0.01, assuming unequal variance with $n \geq 6$ replicates. (b) Malate and pyruvate levels were measured by LC–MS/MS overdevelopment. The resulting pools are displayed as μmol g^{-1} of fresh tissue. Results significant at α: **, < 0.01; *, < 0.05 based on two-tailed *t*-test assuming unequal variance with $n \geq 5$ degrees of freedom from six replicates. The box vertical dimensions represent the interquartile range (IQR) between upper and lower quartiles with the median designated by the thick line within the box. Boxplot whiskers represent ±1.5 IQR.

Changes in free amino acid indicated rebalancing of carbon partitioning in seeds

We hypothesized that malic enzyme might impact amino acid profile by affecting levels of oxaloacetate and pyruvate that are precursors for aspartate and pyruvate-derived amino acids. Free amino acid levels were measured across development using LC–MS/MS with inclusion of ^{13}C-labeled internal standards (Tables S10–S12) and grouped into one of four families based on the origin of metabolic precursors to better assess changes in metabolism. The four families included amino acids derived from: pentose, triose phosphates, and aromatic amino acid precursors, henceforth referred to as the phosphorylated intermediate family; aspartate; glutamate; and pyruvate (Fig. 9a). In WT at developmental stage R5, the aspartate family of free amino acids accounted for over 58% of total, whereas the phosphorylated intermediate family comprised 17.5%, the glutamate family 21%, and the remaining 3.5% was associated with pyruvate-derived amino acids.

By R6, the proportion of the pyruvate-derived and phosphorylated intermediate-derived amino acids had increased at the expense of the aspartate family in both *AtME2* and *AtME4* but was more pronounced in *AtME2* events. The trends continued to R7 with further increases in the pyruvate and phosphorylated intermediate families of amino acids and further reductions in the aspartate family. The increases in the pyruvate-derived and decreases in aspartate-derived amino acids were maintained to maturity with modest rebalancing of glutamate and phosphorylated intermediate families in later stages of development (Fig. 9b; Table S11). Inspection of the pyruvate family amino acids indicated that increases in alanine accounted for the largest proportional change in *AtME2* and are consistent with evidence for alanine production outside of the chloroplast (Fig. 9c) (Biekmann & Feierabend, 1982; Miyashita *et al.*, 2007; Allen *et al.*, 2012).

In *AtME4* events, the differences were more subtle; however, the branched-chain amino acid leucine was elevated significantly by R7 (Table S12), consistent with enhanced enzymatic activity in the plastid where branched-chain amino acids are produced (Hagelstein *et al.*, 1997). There was also a large increase in alanine at R6 for all events including WT, however beginning at R7, the WT levels of alanine dropped. *AtME2* events maintained an elevated level of alanine and *AtME4* shifted to more significant leucine production. Within the aspartate family of amino acids, aspartate levels were significantly depressed in all transgenic alleles leading to overall depression of the entire family.

Protein content was unchanged in malic overexpression events

Given the significant changes in profile of free amino acids, we questioned whether this could impact final proteinogenic amino

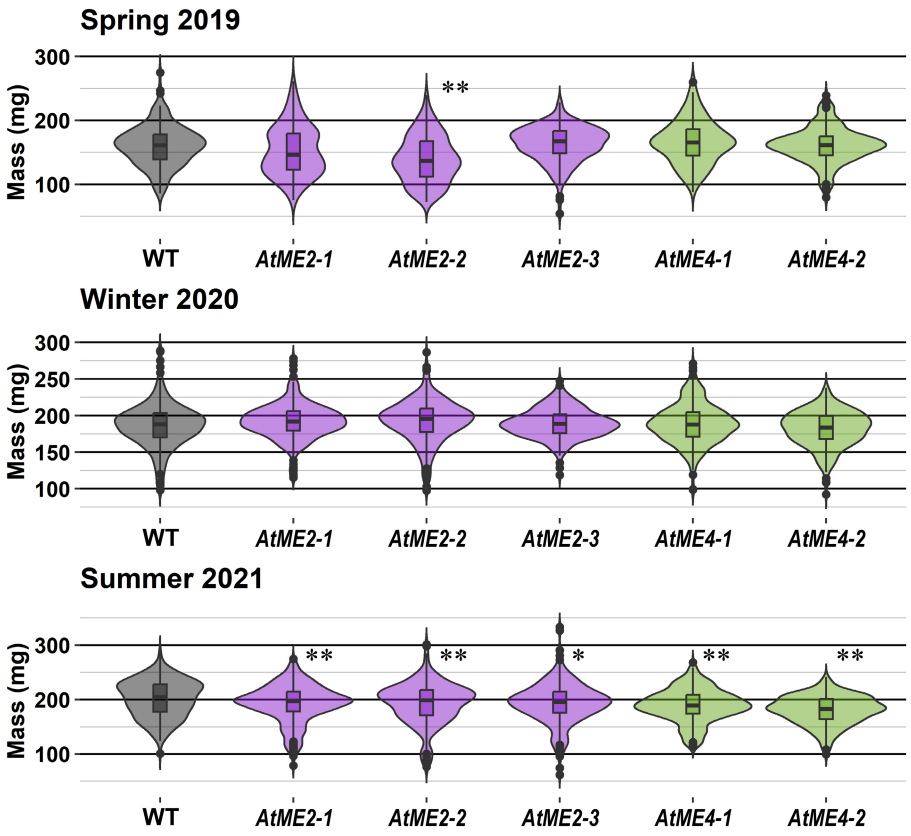


Fig. 6 Violin plots of *Glycine max* seed biomass. Mature seed weight of seeds did not consistently vary per generation (148–317 measurements combined from independent biological replicates) but was distinct in all lines as a result of seasonal growing conditions. Significance at α: **, < 0.01; *, < 0.05 based on a two-sample *z*-test compared with WT population mean. The violin plot area represents kernel density estimation of measured seed weights. Boxes within the violin plots define the interquartile range (IQR) with upper and lower quartiles separated by the median. Lines extending from the inner box plot define ±1.5 IQR.

acid composition. Total protein was measured through a combination of complementary approaches. Near-infrared spectroscopy (NIR) indicated a modest, but statistical decrease in total protein in *AtME2-1* and *AtME4-2* events though increases or decreases were not observed in the other transgenic alleles (Fig. S6). Crude protein assessed overdevelopment with the Dumas method (Mariotti *et al.*, 2008; Allen & Young, 2013) did not indicate differences in total protein between events. Lyophilized tissue was also subjected to acid hydrolysis using an isotope dilution-based mass spectrometry approach (Kambhampati *et al.*, 2019) to quantify total protein and proteinogenic amino acid composition and similarly indicated no significant differences (Fig. 10; Table S13). The combination of methods indicated levels of total protein and amino acids was generally unaltered.

Discussion

Targeted expression of malic enzyme impacts composition in developing soybean seeds by altering subcellular carbon, nitrogen, and reducing equivalent provision

Organelles are a hallmark of eukaryotic metabolism (Allen *et al.*, 2009a, 2012). The current study established the significant consequences of modulating a single enzymatic step in central metabolism principally in two subcellular locations (Fig. 11). The outcomes reflect the impact of spatial considerations on metabolic engineering efforts (Allen *et al.*, 2009a; Allen, 2016b). When malic enzyme was localized outside the chloroplast, that is *AtME2* events, pyruvate-derived amino acid levels increased, especially alanine, but did not result in more protein. Disproportionate production of one amino acid would not be expected to support an increase in total protein content unless that amino acid was limiting. To increase protein requires stoichiometric increases in amino acids corresponding to mRNA templates; 'pulling' on central metabolism from many precursors. This is in contrast to lipids or carbohydrates that are principally derived from one or two precursors (Allen & Young, 2013; Allen, 2016b). Fatty acid biosynthesis in plants occurs in the chloroplast (Ohlrogge *et al.*, 1979; Rawsthorne, 2002) and requires a balanced supply of carbon and reducing equivalents that are provided by the transgenic allele in *AtME4* events resulting in enhanced levels of oleic acid and an overall increase of 0.5–2% oil in mature seeds (i.e. 2–9% change as a fraction of total oil).

Malate is intimately tied to the balance of carbon and reducing equivalents, due to roles in oxidation–reduction reactions and the shuttling of reducing equivalents across organelles through well-described malate valves (Selinski, 2019; Dao *et al.*, 2021). Thus, enhanced malic enzyme activity resulted in more pyruvate (Wedding, 1989) and influenced the cyclic shuttle of malate and oxaloacetate and altered the subcellular balance of NAD(P)H available for biosynthetic processes. Malic enzyme and pyruvate dehydrogenase activities in plastids produce one NADPH, one NADH, and one acetyl-CoA molecule for fatty acid biosynthesis. NADPH can further be used to replenish ferredoxin-based reducing equivalents used by the stearoyl-ACP desaturase in the same location (Nagai & Bloch, 1968; Jacobson *et al.*, 1974; Ohlrogge *et al.*, 1979; Schultz *et al.*, 2000) and thus increased

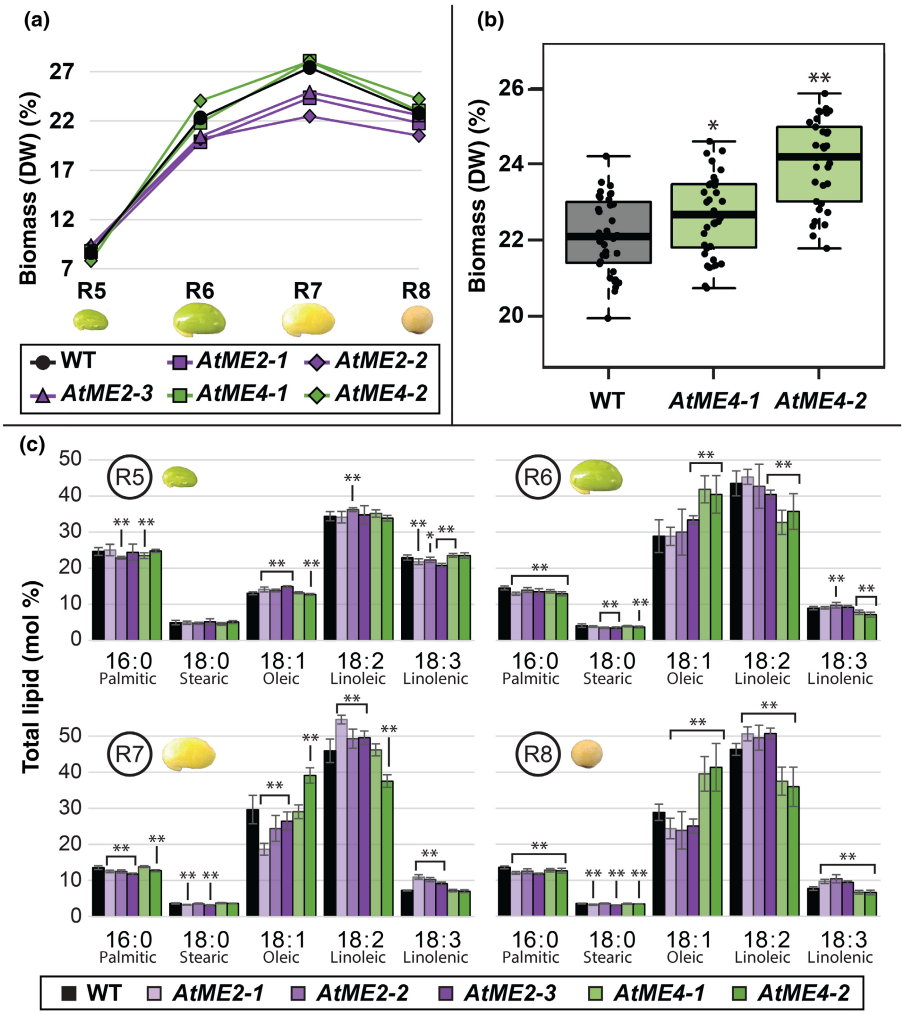


Fig. 7 Lipid profile of transgenic *Glycine max* seeds. (a) Lipid production overdevelopment as a percent of total biomass dry weight (DW) in transgenic malic enzyme seeds (full statistics Supporting Information Table S9) (b) *AtME4-1* and *AtME4-2* display 0.5% and 2% increases in lipid respectively as a percent of biomass (i.e. 2–9% change as a fraction of total oil). Results significant at α: **, < 0.01; *, < 0.05 based on two-tailed *t*-test assuming unequal variance with $n \geq 64$ degrees of freedom. The box vertical dimensions represent the interquartile range (IQR) between upper and lower quartiles with the median designated by the thick line within the box. Boxplot whiskers represent ±1.5 IQR. Black circles represent individual observations. (c) Fatty acid profiles overdevelopment. Error bars represent sample SD with $n = 11$ degrees of freedom. α: **, < 0.01; *, < 0.05 based on two-tailed *t*-test assuming unequal variance with $n = 12$ replicates. Brackets over adjacent columns denote the same level of statistical significance.

NADPH : $NADP^+$ ratios maintain enhanced desaturation rates to oleoyl chains that are produced within the chloroplast for the measured increase in lipid levels. By contrast, the *AtME2* events result in pyruvate and NADH in the cytosol and mitochondria that can be respired through the electron transport chain for increased ATP synthesis or used to sustain polyunsaturation of fatty acids by replenishing cytochrome b5 (Bonnerot *et al.*, 1985; Kumar *et al.*, 2012).

Amino acid distributions were also impacted by enhancing organellar malic enzyme (Fig. 9b,c). We hypothesized that aspartate family amino acids would be reduced, and pyruvate-based amino acids increased, due to the repartitioning of malate carbon from oxaloacetate to pyruvate. The effect was more pronounced in the *AtME2* events suggesting that aspartate biosynthesis is more dependent on the mitochondrial or cytosolic supply of oxaloacetate and malate. Given that aspartate biosynthesis has been hypothesized to occur in multiple subcellular locations (Schultz & Coruzzi, 1995; Miesak & Coruzzi, 2002), efforts to modify the production of aspartate-derived amino acids, including some that limit soybean meal quality for animal feed (Fernandez *et al.*, 1994; Abe *et al.*, 1998), should carefully consider the subcellular source of carbon. Consistent with this thought, increased alanine in the *AtME2* events, and leucine in the *AtME4* events, indicated available pyruvate supplies outside and inside the plastid, respectively. The supply of pyruvate generated from subcellular locations of malic enzyme supported proposed organellar biosynthetic origins of amino acids (Hagelstein *et al.*, 1997; Binder *et al.*, 2007).

Acyl-ACP profiles indicate that fatty acid biosynthesis can accommodate an increased supply of carbon resulting in gains in lipid content

Whether the additional supply of carbon from malate could trigger a new bottleneck at the point of fatty acid synthesis was assessed by examining the acyl-ACP pools. The structures, compositions, and regulation of acyl-ACPs that are located in the chloroplast (Ohlrogge *et al.*, 1979) and to a lesser extent in the mitochondria of plants (Chuman & Brody, 1989) have been studied in oilseeds to better comprehend lipid metabolism (Majerus & Vagelos, 1967; Post-Beittenmiller *et al.*, 1991, 1992; Jaworski *et al.*, 1993; Ohlrogge *et al.*, 1995; Andre *et al.*, 2012; Bates *et al.*, 2014; Kim *et al.*, 2015; Nam *et al.*, 2020; Msanne *et al.*, 2021; Chu *et al.*, 2022). Pooled acyl-ACP levels in soybeans from

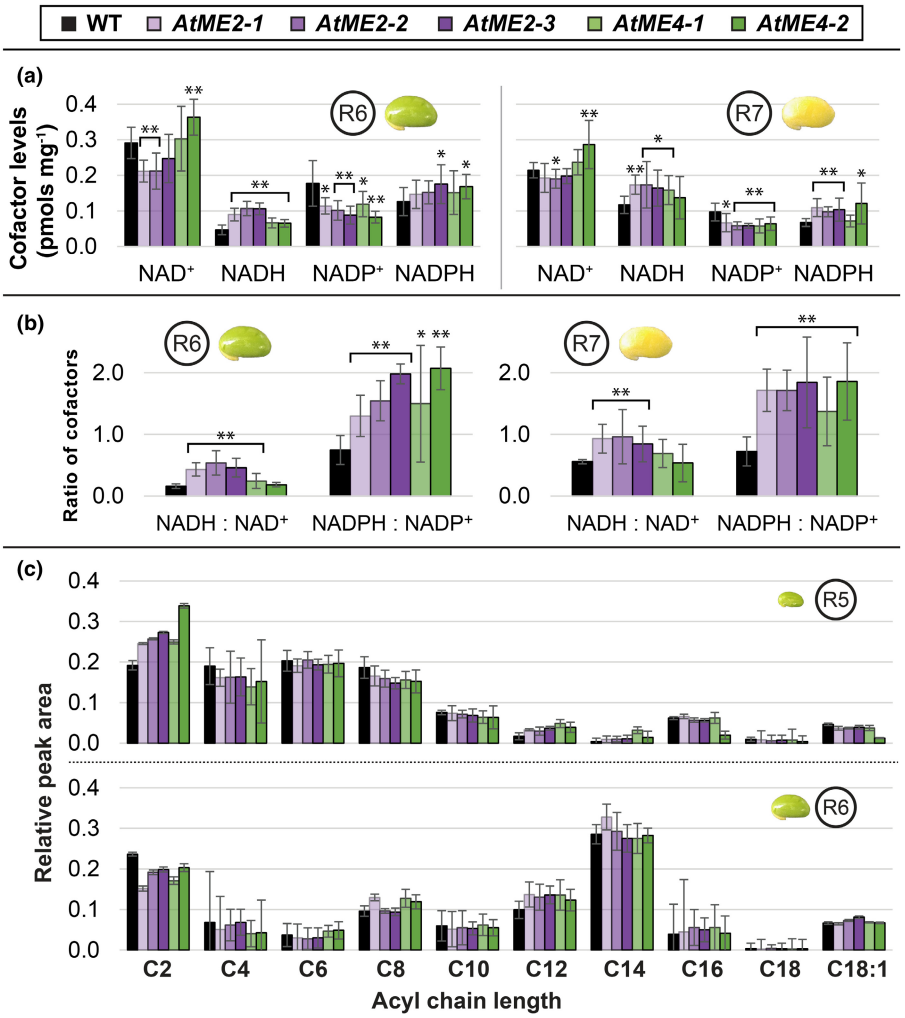


Fig. 8 Levels of NAD(P)H and acyl carrier proteins (ACPs) in transgenic *Glycine max* seeds. (a) Absolute quantity of NAD(P)H from R6 and R7 seeds. Significant differences in the levels of NAD^+, $NADP^+$, NADH, and NADPH can be detected in all transgenic lines. Results significant at α: **, < 0.01; *, < 0.05 based on two-tailed *t*-test assuming unequal variance with $n \geq 10$ degrees of freedom. Error bars represent sample SD with $n = 9$ degrees of freedom. (b) Ratios of NADH : NAD^+ and NADPH : $NADP^+$ in R6 and R7 transgenic seeds were calculated from pairwise measured levels in (a) for enhanced comparison. *AtME2* lines display a much greater increase in NADH : NAD^+ ratio compared with *AtME4* lines. Results significant at α: **, < 0.01; *, < 0.05 based on two-tailed *t*-test assuming unequal variance with $n \geq 10$ degrees of freedom. Error bars represent sample SD with $n = 9$ degrees of freedom. (c) Acyl-ACP levels were measured in R5 and R6 developmental stages of seeds. Relative levels of each acyl-ACP were made using the ratio of peak area to initial tissue mass ($n = 6$, ±1 SD). Later stages of development did not have quantifiable levels.

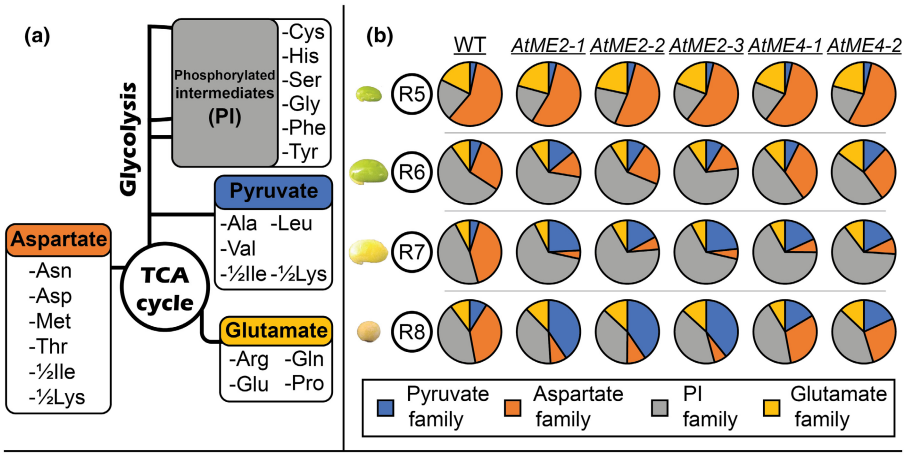
enzymatic and immunochemical assays have indicated that ACPs as a class increase over early development, during fatty acid biosynthesis (Ohlrogge & Kuo, 1984). The analyses of individual acyl-ACPs (Nam *et al.*, 2020) performed here provided a snapshot of the specific ACP levels that correlate with the accumulation and leveling off of lipid production. The results indicated that the ACP levels did not differ significantly between events. Furthermore, the increased levels of pyruvate and reduction in malate, driven by increased malic enzyme activity, suggested that enhancing supplies of central intermediates can lead to a measured increase in lipid in soybeans (i.e. 0.5–2% of total biomass, i.e. 2–9% change in total oil) without requiring changes to fatty acid biosynthetic machinery. Whether further increases in oil content could be obtained by pairing a malic enzyme line with, for example, enhanced acetyl-CoA carboxylase activity (Wilson & Thelen, 2018) or other proteins that increase oil production in soybean would be intriguing to consider.

Though fatty acid synthesis may provide some flexibility for increased oil content in soybean, the ACP profile observed in soybeans differs from other reports and varied considerably between R5 and R6 stages of development. R5 seeds showed higher levels of short-chain ACPs but upon reaching the R6 stage of development there was a larger accumulation of C14 ACP. This is relevant when considering C14 ACP is the longest acyl-ACP without a chain-specific thioesterase in soybean (Jones *et al.*, 1995; Tran *et al.*, 2019; Zhou *et al.*, 2021) and possibly results from increased nonspecific FATA/B thioesterase activity when lipid production is starting to decline as seed development progresses. Total acyl-ACPs and relative level of oil production dropped concomitantly indicating that the quantities of these fatty acid intermediates may be a marker for active lipid production.

Changes in malate partitioning through a single gene effect impacts subcellular resource availability

Reports of significant changes to seed composition from a single gene effect in central metabolism are limited because central metabolism operates in a pliable, context-specific way and provides multiple routes to get from a substrate to a product (Allen *et al.*, 2015; Allen & Young, 2020). In addition, metabolic control analyses suggest that altering a single flux limiting step may only shift the bottleneck to another enzyme in the same pathway (Heinrich & Rapoport, 1974; Kacser *et al.*, 1995; de Castro,

New Phytologist



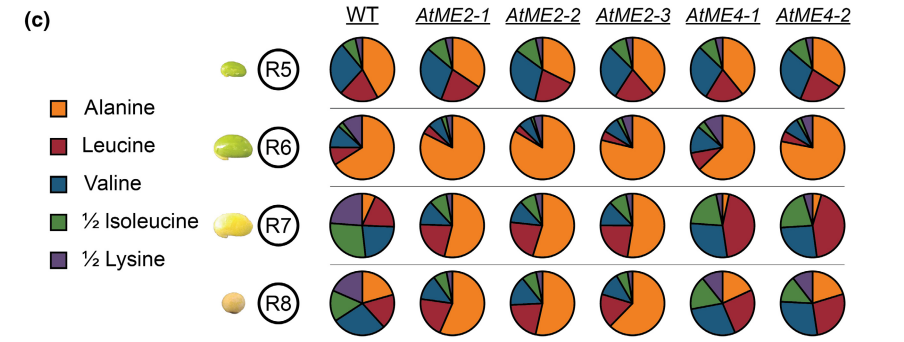
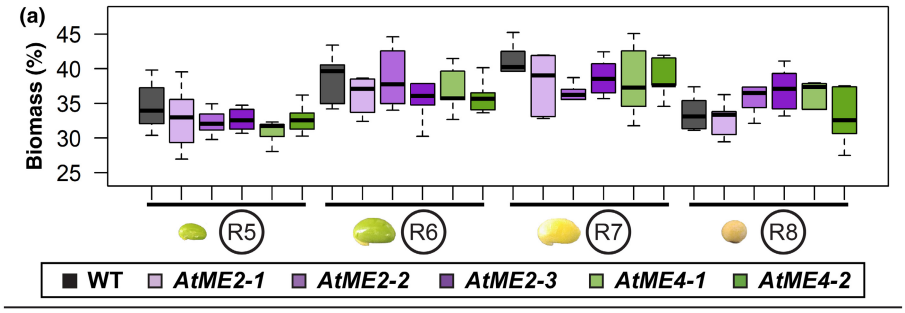


Fig. 9 Free amino acid level comparison in malic enzyme lines. Free amino acids were quantified in transgenic *Glycine max* seeds overdevelopment. (a) Simplified diagram of central carbon metabolism illustrating amino acid families. (b) Comparison of free amino acid families indicating repartitioning of carbon from aspartate to pyruvate family. Statistical information supplied in Supporting Information Table S10. (c) Ratios of individual amino acids within the pyruvate family overdevelopment. Statistical information supplied in Table S11.



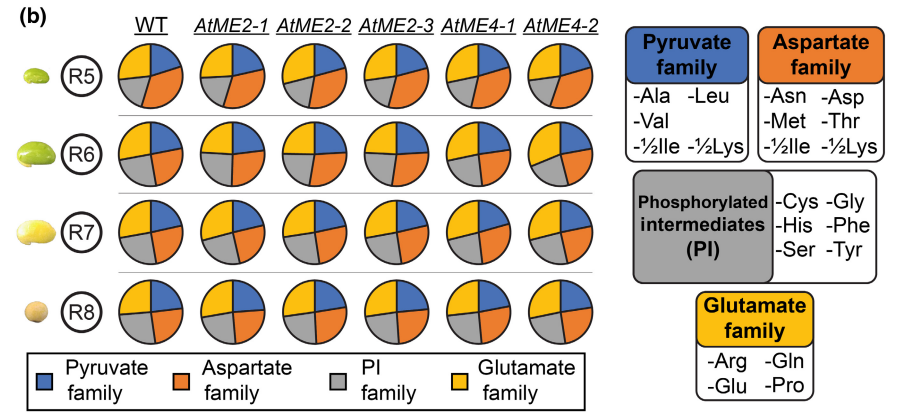


Fig. 10 Total protein in malic enzyme lines. (a) Acid hydrolysis of *Glycine max* seed biomass overdevelopment was quantified and did not show significant changes in total protein levels as a percent of total biomass. The box vertical dimensions represent the interquartile range (IQR) between upper and lower quartiles with the median designated by the thick line within the box. Boxplot whiskers represent ±1.5 IQR. Statistical information provided in Supporting Information Table S13. (b) Ratios of amino acids in total protein were similar between WT and transgenic tissue.

1997). While this latter point likely limits the extent of changes to storage reserve composition observed here, malic enzyme is uniquely positioned in the network as a conduit to shuttle carbon, originating from a vegetative source of glutamine, to seed oil biosynthesis. This concomitantly affects cofactor distribution within organelles in the developing seeds. Prior flux studies of

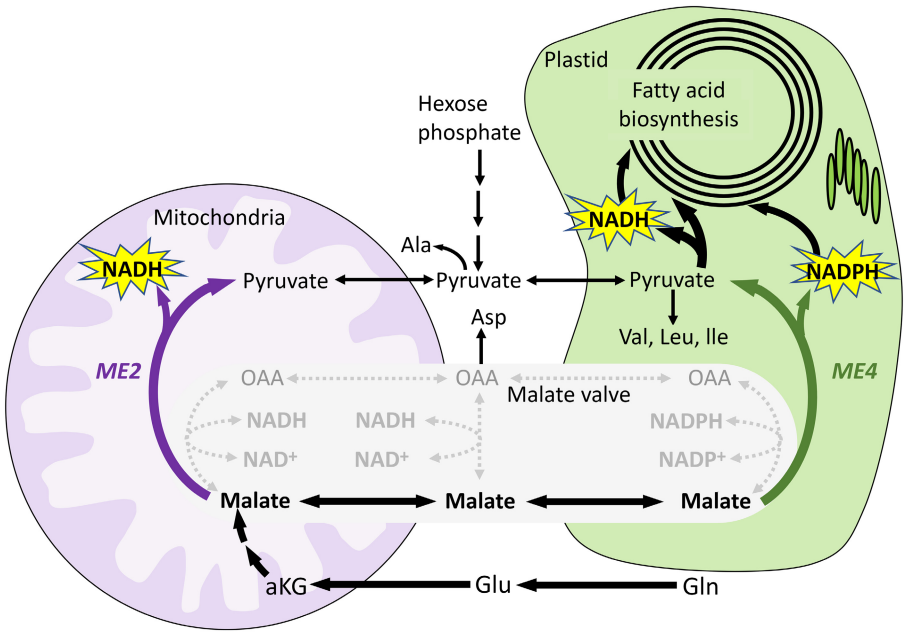


Fig. 11 Summary of malic enzyme activity in overexpressing transgenic *Glycine max* lines. Changes in the *AtME2* and *AtME4* transgenics are displayed by purple■ and green■ arrows and describe the network relationship to malic enzyme. Thick solid lines emphsize major metabolic pathways and substrates directly affected by the action of malic enzyme. Thin solid lines show alternate routes of pyruvate, alanine, and aspartate production. Dashed lines show the interplay of oxaloacetate (OAA), malate, and reducing equivalents due to malate valve activity.

wild-type soybean seed metabolism indicated that malic enzyme, not PEPCK, supported fatty acid biosynthesis (Allen *et al.*, 2009b; Allen & Young, 2013) which was further supported here by labeling in pyruvate but not in PEP from LC–MS/MS analysis of U-^{13}C glutamine cultured soybeans. The results are consistent with recent studies that indicate gluconeogenic pathways including PEPCK-based PEP production are only active late in development when sugar reserves are largely depleted (Kambhampati *et al.*, 2021). The current effort provides genetic confirmation of the impact of subcellular and temporal changes in expression of malic enzyme, can significantly alter central metabolism. Malic enzyme alters the partitioning of carbon and reducing equivalents, that results in enhanced total seed lipid levels, affecting free amino acids but without negatively impacting total protein in the crop.

Acknowledgements

The authors acknowledge support from the United States Department of Agriculture-Agricultural Research Service and the Nebraska Soybean Board. We are grateful to Elizabeth Boedecker (Saint Louis Community College) digital droplet PCR analysis; Brad Evans, Russell Williams, and Shin-Cheng Tzeng (Proteomics and Mass Spectrometry Facility/Donald Danforth Plant Science Center) for LC–MS/MS instrumentation access and assistance; Thomas Maddox (Center for Applied Isotope Studies/University of Georgia) for C : N elemental analysis; Saba Gill for assistance with metabolite derivatization for mass spectrometry; and Kevin Reilly, Kris Haines, Stephen Kraeuter, and Aileen Wok (Donald Danforth Plant Science Center) for assistance with plant growth. The 6500 QTRAP LC–MS/MS was acquired through the National Science Foundation Major Research Instrumentation grant (DBI-1427621) and the Thermo Lumos Fusion Orbital trap was similarly acquired through (DBI-1827534).

Competing interests

None declared.

Author contributions

DKA conceived the project. DKA and SAM designed the experiments which were principally performed and analyzed by SAM unless indicated elsewhere. FM performed initial labeling analysis. TEC and HN assembled constructs and generated soybean transformants that were grown and initially screened and evaluated by CF. HY and VV developed the Taqman assay to assess copy number and confirm single gene insertion and homozygosity of the transgenic plants. MA, SAM, and TB-S developed and performed the transient expression assays, and organelle proteomics was completed by SAM with the help of the PMSF Core Facility. SAM and DKA wrote the manuscript which was edited and approved by all authors.

ORCID

Mazen Alazem https://orcid.org/0000-0003-3690-7459
Doug K. Allen https://orcid.org/0000-0001-8599-8946
Tessa Burch-Smith https://orcid.org/0000-0002-5994-4343
Fangfang Ma https://orcid.org/0000-0001-7220-0675
Stewart A. Morley https://orcid.org/0000-0001-5340-7945

Data availability

Additional data that support the findings of this study are available within the Supporting Information and raw files deposited in Zenodo public repository doi: 10.5281/zenodo.7434125. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol

et al., 2022) partner repository with the dataset identifier PXD039994 and doi: 10.6019/PXD039994.

References

Abe M, Iriki T, Funaba M, Onda S. 1998. Limiting amino acids for a corn and soybean meal diet in weaned calves less than three months of age. *Journal of Animal Science* **76**: 628–636.

Allen DK. 2016a. Assessing compartmentalized flux in lipid metabolism with isotopes. *Biochimica et Biophysica Acta – Molecular and Cell Biology of Lipids* **1861**: 1226–1242.

Allen DK. 2016b. Quantifying plant phenotypes with isotopic labeling & metabolic flux analysis. *Current Opinion in Biotechnology* **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. *Progress in Lipid Research* **58**: 97–120.

Allen DK, Laclair RW, Ohlrogge JB, Shachar-Hill Y. 2012. Isotope labelling of Rubisco subunits provides *in vivo* information on subcellular biosynthesis and exchange of amino acids between compartments. *Plant, Cell & Environment* **35**: 1232–1244.

Allen DK, Libourel IGL, Shachar-Hill Y. 2009a. Metabolic flux analysis in plants: coping with complexity. *Plant, Cell & Environment* **32**: 1241–1257.

Allen DK, Ohlrogge JB, Shachar-Hill Y. 2009b. The role of light in soybean seed filling metabolism. *The Plant Journal* **58**: 220–234.

Allen DK, Young JD. 2013. Carbon and nitrogen provisions alter the metabolic flux in developing soybean embryos. *Plant Physiology* **161**: 1458–1475.

Allen DK, Young JD. 2020. Tracing metabolic flux through time and space with isotope labeling experiments. *Current Opinion in Biotechnology* **64**: 92–100.

Allen RD, Bernier F, Lessard PA, Beachy RN. 1989. Nuclear factors interact with a soybean beta-conglycinin enhancer. *Plant Cell* **1**: 623–631.

Alvarez CE, Bovdilova A, Höppner A, Wolff CC, Saigo M, Trajtenberg F, Zhang T, Buschiazzo A, Nagel-Steger L, Drincovich MF *et al.* 2019. Molecular adaptations of NADP-malic enzyme for its function in C_4 photosynthesis in grasses. *Nature Plants* **5**: 755–765.

Andre C, Haslam RP, Shanklin J. 2012. Feedback regulation of plastidic acetyl-CoA carboxylase by 18:1-acyl carrier protein in *Brassica napus*. *Proceedings of the National Academy of Sciences, USA* **109**: 10107–10112.

Arias CL, Quach T, Huynh T, Nguyen H, Moretti A, Shi Y, Guo M, Rasoul A, Van K, McHale L *et al.* 2022. Expression of AtWRI1 and AtDGAT1 during soybean embryo development influences oil and carbohydrate metabolism. *Plant Biotechnology Journal* **20**: 1327–1345.

Armenteros JJA, Salvatore M, Emanuelsson O, Winther O, Von Heijne G, Elofsson A, Nielsen H. 2019. Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance* **2**: e201900429.

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. *Scientific Reports* **8**: 14653.

Assefa Y, Purcell LC, Salmeron M, Naeve S, Casteel SN, Kovács P, Archontoulis S, Licht M, Below F, Kandel H *et al.* 2019. Assessing variation in us soybean seed composition (protein and oil). *Frontiers in Plant Science* **10**: 00298.

Ates A, Bukowski M. 2022. *Oil crops outlook: March 2022 global oilseed prices soar as South American soybean supply decreases.* U.S. Department of Agriculture, Economic Research Service. [WWW document] URL https://www.ers.usda.gov/publications/pub-details/?pubid=103475, 1–11.

Aznar-Moreno JA, Mukherjee T, Morley SA, Duressa D, Kambhampati S, Chu KL, Koley S, Allen DK, Durrett TP. 2022. Suppression of SDP1 improves soybean seed composition by increasing oil and reducing undigestible oligosaccharides. *Frontiers in Plant Science* **13**: 863254.

Bates PD, Johnson SR, Cao X, Li J, Nam JW, Jaworski JG, Ohlrogge JB, Browse J. 2014. Fatty acid synthesis is inhibited by inefficient utilization of unusual fatty acids for glycerolipid assembly. *Proceedings of the National Academy of Sciences, USA* **111**: 1204–1209.

Beechey-Gradwell Z, Cooney L, Winichayakul S, Andrews M, Hea SY, Crowther T, Roberts N. 2020. Storing carbon in leaf lipid sinks enhances perennial ryegrass carbon capture especially under high N and elevated CO_2. *Journal of Experimental Botany* **71**: 2351–2361.

Biekmann S, Feierabend J. 1982. Subcellular distribution, multiple forms and development of glutamate-pyruvate (glyoxylate) aminotransferase in plant tissues. *Biochimica et Biophysica Acta – Molecular Cell Research* **721**: 268–279.

Binder S, Knill T, Schuster J. 2007. Branched-chain amino acid metabolism in higher plants. *Physiologia Plantarum* **129**: 68–78.

Bonnerot C, Galle AM, Jolliot A, Kader JC. 1985. Purification and properties of plant cytochrome b5. *The Biochemical Journal* **226**: 331–334.

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 Phytologist* **167**: 761–776.

de Borja Reis AF, Tamagno S, Moro Rosso LH, Ortez OA, Naeve S, Ciampitti IA. 2020. Historical trend on seed amino acid concentration does not follow protein changes in soybeans. *Scientific Reports* **10**: 17707.

Bowman EJ, Ikuma H. 1976. Regulation of malate oxidation in isolated mung bean mitochondria. *Plant Physiology* **58**: 438–446.

Cai Y, Whitehead P, Chappell J, Chapman KD. 2019. Mouse lipogenic proteins promote the co-accumulation of triacylglycerols and sesquiterpenes in plant cells. *Planta* **250**: 79–94.

Carrington JC, Freed DD. 1990. Cap-independent enhancement of translation by a plant potyvirus 5′ nontranslated region. *Journal of Virology* **64**: 1590–1597.

de Castro IN. 1997. Understanding the control of metabolism. *Biochemical Education* **25**: 250–251.

Cernac A, Benning C. 2004. WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *The Plant Journal* **40**: 575–585.

Chen B, Zhang G, Li P, Yang J, Guo L, Benning C, Wang X, Zhao J. 2020. Multiple GmWRI1s are redundantly involved in seed filling and nodulation by regulating plastidic glycolysis, lipid biosynthesis and hormone signaling in soybean (*Glycine max*). *Plant Biotechnology Journal* **18**: 155–171.

Chu KL, Koley S, Jenkins LM, Bailey SR, Kambhampati S, Foley K, Arp JJ, Morley SA, Czymmek KJ, Bates PD *et al.* 2022. Metabolic flux analysis of the non-transitory starch tradeoff for lipid production in mature tobacco leaves. *Metabolic Engineering* **69**: 231–248.

Chuman L, Brody S. 1989. Acyl carrier protein is present in the mitochondria of plants and eucaryotic micro-organisms. *European Journal of Biochemistry* **184**: 643–649.

Clemente TE, Cahoon EB. 2009. Soybean oil: genetic approaches for modification of functionality and total content. *Plant Physiology* **151**: 1030–1040.

Dao O, Kuhnert F, Weber APM, Peltier G, Li-Beisson Y. 2021. Physiological functions of malate shuttles in plants and algae. *Trends in Plant Science* **27**: 488–501.

Day DA, Wiskich JT. 1974. The oxidation of malate and exogenous reduced nicotinamide adenine dinucleotide by isolated plant mitochondria. *Plant Physiology* **53**: 104–109.

Egli DB. 1998. *Seed biology and yield of grain crops.* New York, NY, USA: CAB International.

Fernandez SR, Aoyagi S, Han Y, Parsons CM, Baker DH. 1994. Limiting order of amino acids in corn and soybean meal for growth of the chick. *Poultry Science* **73**: 1887–1896.

Friedman M, Brandon DL. 2001. Nutritional and health benefits of soy proteins. *Journal of Agricultural and Food Chemistry* **49**: 1069–1086.

Gerrard Wheeler MC, Arias CL, Righini S, Badia MB, Andreo CS, Drincovich MF, Saigo M. 2016. Differential contribution of malic enzymes during soybean and castor seeds maturation. *PLoS ONE* **11**: 0158040.

Greer MS, Cai Y, Gidda SK, Esnay N, Kretzschmar FK, Seay D, McClinchie E, Ischebeck T, Mullen RT, Dyer JM *et al.* 2020. SEIPIN isoforms interact with the membrane-tethering protein VAP27-1 for lipid droplet formation. *Plant Cell* **32**: 2932–2950.

Grohmann L, Nieweler CB, Nemeth A, Waiblinger HU. 2009. Collaborative trial validation studies of real-time PCR-based GMO Screening methods for detection of the bar gene and the ctp2-cp4epsps construct. *Journal of Agricultural and Food Chemistry* **57**: 8913–8920.

Guynn RW, Gelberg HJ, Veech RL. 1973. Equilibrium constants of the malate dehydrogenase, citrate synthase, citrate lyase, and acetyl coenzyme A hydrolysis

reactions under physiological conditions. *Journal of Biological Chemistry* **248**: 6957–6965.

Hagelstein P, Sieve B, Klein M, Jans H, Schultz G. 1997. Leucine synthesis in chloroplasts: leucine/isoleucine aminotransferase and valine aminotransferase are different enzymes in spinach chloroplasts. *Journal of Plant Physiology* **150**: 23–30.

Hatch MD, Mau SL, Kagawa T. 1974. Properties of leaf NAD malic enzyme from plants with C_4 pathway photosynthesis. *Archives of Biochemistry and Biophysics* **165**: 188–200.

Heineke D, Riens B, Grosse H, Hoferichter P, Peter U, Flügge UI, Heldt HW. 1991. Redox transfer across the inner chloroplast envelope membrane. *Plant Physiology* **95**: 1131–1137.

Heinrich R, Rapoport TA. 1974. A linear steady-state treatment of enzymatic chains: general properties, control and effector strength. *European Journal of Biochemistry* **42**: 89–95.

Heldt H. 2005. *Biochemistry*. Amsterdam, the Netherlands; Boston, MA, USA: Elsevier Academic Press.

Henneke CM, Wedding RT. 1975. NAD-phenol complex formation, the inhibition of malate dehydrogenase by phenols, and the influence of phenol substituents on inhibitory effectiveness. *Archives of Biochemistry and Biophysics* **168**: 443–449.

Henry CS, Broadbelt LJ, Hatzimanikatis V. 2007. Thermodynamics-based metabolic flux analysis. *Biophysical Journal* **92**: 1792–1805.

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. *Journal of Experimental Botany* **56**: 1951–1963.

Ho LC. 1988. Metabolism and compartmentation of imported sugars in sink organs in relation to sink strength. *Annual Review of Plant Physiology and Plant Molecular Biology* **39**: 355–378.

Hofvander P, Ischebeck T, Turesson H, Kushwaha SK, Feussner I, Carlsson AS, Andersson M. 2016. Potato tuber expression of Arabidopsis WRINKLED1 increase triacylglycerol and membrane lipids while affecting central carbohydrate metabolism. *Plant Biotechnology Journal* **14**: 1883–1898.

Hood EE, Helmer GL, Fraley RT, Chilton MD. 1986. The hypervirulence of *Agrobacterium tumefaciens* A281 is encoded in a region of pTiBo542 outside of T-DNA. *Journal of Bacteriology* **168**: 1291–1301.

Hsu FC, Bennett AB, Spanswick RM. 1984. Concentrations of sucrose and nitrogenous compounds in the apoplast of developing soybean seed coats and embryos. *Plant Physiology* **75**: 181–186.

Huang S, Jacoby RP, Millar AH, Taylor NL. 2014. Plant mitochondrial proteomics. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S, eds. *Plant proteomics: methods and protocols*. Totowa, NJ, USA: Humana Press, 499–525.

Hymowitz T, Collins FI, Panczner J, Walker WM. 1972. Relationship between the content of oil, protein, and sugar in soybean seed 1. *Agronomy Journal* **64**: 613–616.

Ito J, Batth TS, Petzold CJ, Redding-Johanson AM, Mukhopadhyay A, Verboom R, Meyer EH, Millar AH, Heazlewood JL. 2011. Analysis of the Arabidopsis cytosolic proteome highlights subcellular partitioning of central plant metabolism. *Journal of Proteome Research* **10**: 3309.

Iyer VV, Sriram G, Fulton DB, Zhou R, Westgate ME, Shanks JV. 2008. Metabolic flux maps comparing the effect of temperature on protein and oil biosynthesis in developing soybean cotyledons. *Plant, Cell, & Environment* **31**: 506–517.

Jacobson BS, Jaworski JG, Stumpf PK. 1974. Fat metabolism in higher plants: LXII. Stearl-acyl carrier protein desaturase from spinach chloroplasts. *Plant Physiology* **54**: 484–486.

Jaworski JG, Post-Beittenmiller D, Ohlrogge JB. 1993. Acetyl-acyl carrier protein is not a major intermediate in fatty acid biosynthesis in spinach. *European Journal of Biochemistry* **987**: 981–987.

Jenkins LM, Nam J-W, Evans BS, Allen DK. 2021. Quantification of acyl-acyl carrier proteins for fatty acid synthesis fatty acid synthesis using LC-MS/MS. In: Bartels D, Dörmann P, eds. *Plant lipids: methods and protocols*. New York, NY, USA: Springer US, 219–247.

Jian B, Liu B, Bi Y, Hou W, Wu C, Han T. 2008. Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Molecular Biology* **9**: 59.

Jones A, Davies HM, Voelker TA. 1995. Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. *Plant Cell* **7**: 359–371.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A *et al.* 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583–589.

Kacser H, Burns JA, Fell DA. 1995. The control of flux. *Biochemical Society Transactions* **23**: 341–366.

Kambhampati S, Aznar-Moreno JA, Bailey SR, Arp JJ, Chu KL, Bilyeu KD, Durrett TP, Allen DK. 2021. Temporal changes in metabolism late in seed development affect biomass composition. *Plant Physiology* **186**: 874–890.

Kambhampati S, Aznar-Moreno JA, Hostetler C, Caso T, Bailey SR, Hubbard AH, Durrett TP, Allen DK. 2020. 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**: 1–15.

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**: 1–12.

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. *Scientific Reports* **9**: 1–10.

Katoh K, Rozewicki J, Yamada KD. 2018. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20**: 1160–1166.

Kim HJ, Silva JE, Vu HS, Mockaitis K, Nam JW, Cahoon EB. 2015. Toward production of jet fuel functionality in oilseeds: identification of FatB acyl-acyl carrier protein thioesterases and evaluation of combinatorial expression strategies in *Camelina* seeds. *Journal of Experimental Botany* **66**: 4251–4265.

Kirsten WJ. 1983. *Organic elemental analysis*. Amsterdam, the Netherlands; Boston, MA, USA: Elsevier Academic Press.

Koley S, Chu KL, Gill SS, Allen DK. 2022. An efficient LC-MS method for isomer separation and detection of sugars, phosphorylated sugars, and organic acids. *Journal of Experimental Botany* **73**: 2938–2952.

Kumar R, Tran LSP, Neelakandan AK, Nguyen HT. 2012. Higher plant cytochrome b5 polypeptides modulate fatty acid desaturation. *PLoS ONE* **7**: e44743.

Li Y, Beisson F, Pollard M, Ohlrogge J. 2006. Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. *Phytochemistry* **67**: 904–915.

Lu W, Wang L, Chen L, Hui S, Rabinowitz JD. 2018. Extraction and quantitation of nicotinamide adenine dinucleotide redox cofactors. *Antioxidants and Redox Signaling* **28**: 167–179.

Ma F, Jazmin LJ, Young JD, Allen DK. 2014. Isotopically nonstationary ^{13}C flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. *Proceedings of the National Academy of Sciences, USA* **111**: 16967–16972.

Ma F, Jazmin LJ, Young JD, Allen DK. 2017. Isotopically nonstationary metabolic flux analysis (INST-MFA) of photosynthesis and photorespiration in plants. In: Fernie AR, Bauwe H, Weber APM, eds. *Photorespiration: methods and protocols*. New York, NY, USA: Springer, 167–194.

Macrae AR. 1971. Malic enzyme activity of plant mitochondria. *Phytochemistry* **10**: 2343–2347.

Majerus PW, Vagelos PR. 1967. Fatty acid biosynthesis and the role of the acyl carrier protein. *Advances in Lipid Research* **5**: 1–33.

Marega M, Destro D, Miranda LA, Spinosa WA, Carräo-Panizzi MC, Montalván R. 2001. Relationships among oil content, protein content and seed size in soybeans. *Brazilian Archives of Biology and Technology* **44**: 23–32.

Mariotti F, Tomé D, Mirand PP. 2008. Converting nitrogen into protein – beyond 6.25 and Jones' factors. *Critical Reviews in Food Science and Nutrition* **48**: 177–184.

McBlain BA, Fioritto RJ, St. Martin SK, Calip-Dubois AJ, Schmitthenner AF, Cooper RL, Martin RJ. 1993. Registration of 'Thorne' soybean. *Crop Science* **33**: 1406.

Messina MJ. 1999. Legumes and soybeans: overview of their nutritional profiles and health effects. *American Journal of Clinical Nutrition* **70**: S439–S450.

Miesak BH, Coruzzi GM. 2002. Molecular and physiological analysis of Arabidopsis mutants defective in cytosolic or chloroplastic aspartate aminotransferase. *Plant Physiology* **129**: 650–660.

Millar AH, Sweetlove LJ, Giegé P, Leaver CJ. 2001. Analysis of the Arabidopsis mitochondrial proteome. *Plant Physiology* **127**: 1711–1727.

Miyashita Y, Dolferus R, Ismond KP, Good AG. 2007. Alanine aminotransferase catalyses the breakdown of alanine after hypoxia in *Arabidopsis thaliana*. *The Plant Journal* **49**: 1108–1121.

Msanne J, Vu HS, Cahoon EB. 2021. Acyl-acyl carrier protein pool dynamics with oil accumulation in nitrogen-deprived *Chlamydomonas reinhardtii* microalgal cells. *Journal of the American Oil Chemists' Society* **98**: 1107–1112.

Nagai J, Bloch K. 1968. Enzymatic desaturation of stearyl acyl carrier protein. *Journal of Biological Chemistry* **243**: 4626–4633.

Nam JW, Jenkins LM, Li J, Evans BS, Jaworski JG, Allen DK. 2020. A general method for quantification and discovery of acyl groups attached to acyl carrier proteins in fatty acid metabolism using LC-MS/MS. *Plant Cell* **32**: 820–832.

Oehrle NW, Sarma AD, Waters JK, Emerich DW. 2008. Proteomic analysis of soybean nodule cytosol. *Phytochemistry* **69**: 2426–2438.

Ohlrogge J, Savage L, Jaworski J, Voelker T, Post-Beittenmiller D. 1995. Alteration of acyl-acyl carrier protein pools and acetyl-coa carboxylase expression in *Escherichia coli* by a plant medium-chain acyl-acyl carrier protein thioesterase. *Archives of Biochemistry and Biophysics* **317**: 185–190.

Ohlrogge JB, Kuhn DN, Stumpf PK. 1979. Subcellular localization of acyl carrier protein in leaf protoplasts of *Spinacia oleracea*. *Proceedings of the National Academy of Sciences, USA* **76**: 1194–1198.

Ohlrogge JB, Kuo T-M. 1984. Control of lipid synthesis during soybean seed development: enzymic and immunochemical assay of acyl carrier protein. *Plant Physiology* **74**: 622–625.

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. *Theoretical and Applied Genetics* **130**: 1975–1991.

Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M *et al.* 2022. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research* **50**: D543–D552.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera – a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**: 1605–1612.

Pipolo AE, Sinclair TR, Camara GMS. 2004. Protein and oil concentration of soybean seed cultured *in vitro* using nutrient solutions of differing glutamine concentration. *Annals of Applied Biology* **144**: 223–227.

Post-Beittenmiller D, Jaworski JG, Ohlrogge JB. 1991. *In vivo* pools of free and acylated acyl carrier proteins in spinach. Evidence for sites of regulation of fatty acid biosynthesis. *Journal of Biological Chemistry* **266**: 1858–1865.

Post-Beittenmiller D, Roughan G, Ohlrogge JB. 1992. Regulation of plant fatty acid biosynthesis: analysis of acyl-coenzyme and acyl-acyl carrier protein substrate pools in spinach and pea chloroplasts. *Plant Physiology* **100**: 923–930.

Pyc M, Cai Y, Gidda SK, Yurchenko O, Park S, Kretzschmar FK, Ischebeck T, Valerius O, Braus GH, Chapman KD *et al.* 2017. Arabidopsis lipid droplet-associated protein (LDAP) – interacting protein (LDIP) influences lipid droplet size and neutral lipid homeostasis in both leaves and seeds. *The Plant Journal* **92**: 1182–1201.

Rainbird RM, Thorne JH, Hardy RWF. 1984. Role of amides, amino acids, and ureides in the nutrition of developing soybean seeds. *Plant Physiology* **74**: 329–334.

Rawsthorne S. 2002. Carbon flux and fatty acid synthesis in plants. *Progress in Lipid Research* **41**: 182–196.

Reddy DS, Bhatnagar-mathur P, Cindhuri KS, Sharma KK. 2013. Evaluation and validation of reference genes for normalization of quantitative real-time PCR based gene expression studies in peanut. *PLoS ONE* **8**: e78555.

Roesler K, Shen B, Bermudez E, Li C, Hunt J, Damude HG, Ripp KG, Everard JD, Booth JR, Castaneda L *et al.* 2016. An improved variant of soybean type 1 diacylglycerol acyltransferase increases the oil content and decreases the soluble carbohydrate content of soybeans. *Plant Physiology* **171**: 878–893.

Schultz CJ, Coruzzi GM. 1995. The aspartate aminotransferase gene family of Arabidopsis encodes isoenzymes localized to three distinct subcellular compartments. *The Plant Journal* **7**: 61–75.

Schultz DJ, Suh MC, Ohlrogge JB. 2000. Stearoyl-acyl carrier protein and unusual acyl-acyl carrier protein desaturase activities are differentially influenced by ferredoxin. *Plant Physiology* **124**: 681–692.

Selinski J. 2019. Malate valves: old shuttles with new perspectives. *Plant Biology* **21**: 21–30.

Shen B, Allen WB, Zheng P, Li C, Glassman K, Ranch J, Nubel D, Tarczynski MC. 2010. Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. *Plant Physiology* **153**: 980–987.

Springer NM. 2010. Isolation of plant DNA for PCR and genotyping using organic extraction and CTAB. *Cold Spring Harbor Protocols* **2010**: pdb.prot5515.

Sriram G, Fulton DB, Iyer VV, Peterson JM, Zhou R, Westgate ME, Spalding MH, Shanks JV. 2004. Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional ^{13}C labeling, two-dimensional [^{13}C, ^{1}H] nuclear magnetic resonance, and comprehensive isotopomer balancing. *Plant Physiology* **136**: 3043–3057.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Steinmüller D, Tevini M. 1985. Composition and function of plastoglobuli. *Planta* **163**: 201–207.

Tao X, Yang Z, Tong L. 2003. Crystal structures of substrate complexes of malic enzyme and insights into the catalytic mechanism. *Structure* **11**: 1141–1150.

Tejada P, Sanchez-Moreno M, Monteoliva M, Gomez-Banqueri H. 1987. Inhibition of malate dehydrogenase enzymes by benzimidazole anthelmintics. *Veterinary Parasitology* **24**: 269–274.

Thompson CJ, Movva NR, Tizard R, Crameri R, Davies JE, Lauwereys M, Botterman J. 1987. Characterization of the herbicide-resistance gene bar from *Streptomyces hygroscopicus*. *EMBO Journal* **6**: 2519–2523.

Torabi S, Sukumaran A, Dhaubhadel S, Johnson SE, LaFayette P, Parrott WA, Rajcan I, Eskandari M. 2021. Effects of type I *diacylglycerol O-acyltransferase (DGAT1)* genes on soybean (*Glycine max* L.) seed composition. *Scientific Reports* **11**: 2556.

Tran HTD, Le NT, Khuat VLU, Nguyen TTH. 2019. Identification and functional characterization of a soybean (*Glycine max*) thioesterase that acts on intermediates of fatty acid biosynthesis. *Plants* **8**: 1–21.

Tronconi MA, Fahnenstich H, Gerrard Weehler MC, Andreo CS, Flügge UI, Drincovich MF, Maurino VG. 2008. Arabidopsis NAD-malic enzyme functions as a homodimer and heterodimer and has a major impact on nocturnal metabolism. *Plant Physiology* **146**: 1540–1552.

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. *Journal of Experimental Botany* **64**: 2985–2995.

Vanhercke T, Dyer JM, Mullen RT, Kilaru A, Rahman MM, Petrie JR, Green AG, Yurchenko O, Singh SP. 2019. Metabolic engineering for enhanced oil in biomass. *Progress in Lipid Research* **74**: 103–129.

Vanhercke T, El Tahchy A, Shrestha P, Zhou XR, Singh SP, Petrie JR. 2013. Synergistic effect of WRI1 and DGAT1 coexpression on triacylglycerol biosynthesis in plants. *FEBS Letters* **587**: 364–369.

Wang M, Garneau MG, Poudel AN, Lamm D, Koo AJ, Bates PD, Thelen JJ. 2022. Overexpression of pea α-carboxyltransferase in Arabidopsis and camelina increases fatty acid synthesis leading to improved seed oil content. *The Plant Journal* **4**: 1–12.

Wedding RT. 1989. Malic enzymes of higher plants: characteristics, regulation, and physiological function. *Plant Physiology* **90**: 367–371.

Wilson RS, Thelen JJ. 2018. *In vivo* quantitative monitoring of subunit stoichiometry for metabolic complexes. *Journal of Proteome Research* **17**: 1773–1783.

Winichayakul S, William Scott R, Roldan M, Bertrand Hatier JH, Livingston S, Cookson R, Curran AC, Roberts NJ. 2013. *In vivo* packaging of triacylglycerols enhances Arabidopsis leaf biomass and energy density. *Plant Physiology* **162**: 626–639.

Wynn JP, Hamidt A, Ratledge C. 1999. Lipid accumulation in filamentous fungi. *Microbiology* **145**: 1911–1917.

Wynn JP, Ratledge C. 1997. Malic enzyme is a major source of NADPH for lipid accumulation by *Aspergillus nidulans*. *Microbiology* **143**: 253–257.

Xing A, Zhang Z, Sato S, Staswick P, Clemente TE. 2000. The use of the two T-DNA binary system to derive marker-free transgenic soybeans on JSTOR. *In Vitro Cellular & Developmental Biology* **36**: 456–463.

Ye Y, Nikovics K, To A, Lepiniec L, Fedosejevs ET, Van Doren SR, Baud S, Thelen JJ. 2020. Docking of acetyl-CoA carboxylase to the plastid envelope membrane attenuates fatty acid production in plants. *Nature Communications* **11**: 1–14.

Yim AK, Wong JW, Ku Y, Qin H. 2015. Using RNA-Seq data to evaluate reference genes suitable for gene expression studies in soybean. *PLoS ONE* **10**: e0136343.

Zhang Y, Adams IP, Ratledge C. 2007. Malic enzyme: the controlling activity for lipid production? Overexpression of malic enzyme in *Mucor circinelloides* leads to a 2.5-fold increase in lipid accumulation. *Microbiology* **153**: 2013–2025.

Zhang Z, Xing A, Staswick P, Clemente TE. 1999. The use of glufosinate as a selective agent in Agrobacterium-mediated transformation of soybean. *Plant Cell, Tissue, and Organ Culture* **56**: 37–46.

Zhou Z, Lakhssassi N, Knizia D, Cullen MA, El Baz A, Embaby MG, Liu S, Badad O, Vuong TD, AbuGhazaleh A *et al.* 2021. Genome-wide identification and analysis of soybean acyl-ACP thioesterase gene family reveals the role of GmFAT to improve fatty acid composition in soybean seed. *Theoretical and Applied Genetics* **134**: 3611–3623.

Zhu BH, Zhang RH, Lv NN, Yang GP, Wang YS, Pan KH. 2018. The role of malic enzyme on promoting total lipid and fatty acid production in *Phaeodactylum tricornutum*. *Frontiers in Plant Science* **9**: 1–8.

Zimmermann M, Sauer U, Zamboni N. 2014. Quantification and mass isotopomer profiling of α-keto acids in central carbon metabolism. *Analytical Chemistry* **86**: 3232–3238.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Phylogenetic and structural analysis of Arabidopsis and *Glycine max* malic enzyme proteins.

Fig. S2 Active site and structural comparison of ME2 and ME4 orthologs.

Fig. S3 Vector maps used in *Agrobacterium tumefaciens* to create transgenic *Glycine max*.

Fig. S4 PCR of transgenic *Glycine max* confirmed presence of Arabidopsis malic enzyme genes.

Fig. S5 *In silico* prediction of AtME2 and AtME4 subcellular localization.

Fig. S6 Near-infrared measurement of total protein in mature *Glycine max* seeds.

Methods S1 Chromatography conditions and multiple reaction monitoring parameters for detection of metabolites by mass spectrometry.

Methods S2 Chromatography conditions and multiple reaction monitoring parameters for detection of NAD^+, $NADP^+$, NADH, and NADPH by mass spectrometry.

Methods S3 Organelle proteomic sample preparation and LC–MS/MS analysis.

Methods S4 Enzyme activity assay description and calculations.

Table S1 Primers used to confirm presence of transgenic Arabidopsis genes.

Table S2 Primers used for digital droplet PCR.

Table S3 Percent incorporation of U-^{13}C glutamine into central carbon metabolites using media with varying carbon : nitrogen (C : N) ratios.

Table S4 Allelic copies of the bar gene as determined by Taqman assay in transgenic *Glycine max* lines.

Table S5 ddPCR expression of endogenous and transgenic malic enzyme genes in *Glycine max*.

Table S6 Identified mitochondrial and cytosolic proteins from *Glycine max* seed proteomics.

Table S7 All protein identifications from *Glycine max* seed proteomics.

Table S8 Measured enzyme activity of transgenic *Glycine max*.

Table S9 Total oil as a % biomass in transgenic *Glycine max* seeds.

Table S10 Free amino acid levels in transgenic *Glycine max*.

Table S11 Ratio of total free amino acids grouped by precursor families in transgenic *Glycine max*.

Table S12 Ratio of amino acids within the pyruvate family.

Table S13 Proteinogenic amino acid levels.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

See also the Commentary on this article by Schwender, **239**: 1539–1541.