INTEGRATING EXPERIMENTS AND MODELS TO UNDERSTAND PLANT NUTRIENT FLUXES

By

Teresa Jane Clark

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Biology—Doctor of Philosophy
Quantitative Biology—Dual Major

2018

ABSTRACT

INTEGRATING EXPERIMENTS AND MODELS TO UNDERSTAND PLANT NUTRIENT FLUXES

By

Teresa Jane Clark

Improving the yield and efficiency of plants is an urgent goal of research, but it is difficult to understand the functioning and regulation of these complex traits by studying individual genes and phenotypes. Mathematical models are particularly useful in addressing multifactorial characteristics because they embody the relations of multiple system components and can predict the quantitative effect of changing these components. Here, I describe my collaborative work on integrating experimental data with models of nutrient fluxes in plant systems operating at three different biological scales: plant-microbe interactions, the central metabolism network, and within one biochemical pathway.

First, we analyzed carbon-for-nitrogen exchange in the legume-rhizobia mutualism using a trade-based model. We found that classical symmetric interactions could not explain the observed nutrient exchange rates and ratios. Rather, the plant had more influence on these nutrient exchange parameters than the rhizobia and the plant's influence rose as soil nitrogen became scarcer. This finding highlights the importance of environmental conditions for the functioning and evolution of mutualisms. To explore other ways in which experiments have been or could be integrated with mathematical models, we also wrote a synthesis review. In this review, we summarized mathematical approaches that could investigate the cellular, individual, population, and community scales of plant-microbe nutrient-exchange mutualisms. For each scale, we addressed the potentials of integration to investigate mutualism stability, and how resource availability and structured interactions influence these dynamics.

Second, we used isotopic labeling based metabolic flux analysis to determine the routes and rates of fluxes through central metabolism in developing embryos of *Camelina sativa*—a promising oil seed crop. By quantifying the major decarboxylation fluxes, we discovered that the oxidative phase of the

pentose phosphate pathway exhibited a very high flux. The OPP flux was also tightly correlated with the carbon use efficiencies of *C. sativa* embryos grown under different light levels. Together, these results indicate that this decarboxylation flux contributed substantially to the embryos having low carbon use efficiency. To test how bioengineering can affect central metabolism, we also performed metabolic flux analysis on transgenic *C. sativa* that had been genetically engineered to accumulate greater quantities of medium-chain fatty acids. We found that principal component analysis could use metabolite label content measurements to distinguish between embryos with different backgrounds, but there was much less distinction between the lines when the net fluxes were used. However, the lines could be moderately separated with hierarchal clustering, which could cleanly separate the transgenic lines from the wild-type grown under different light level. This shows that there are distinct metabolic differences between the transgenic lines that are not detected with principal component analysis, and that environmental variances can have a stronger effect on central metabolism than fatty acid bioengineering.

Lastly, I describe our recent progress towards elucidating the network topology of triacylglycerol synthesis in *C. sativa* embryos. We developed potential network topologies as systems of first-order rate equations. The models were fit to ^{14}C glyceryl time-course data of the label content in other lipid classes in the network. We compared how well the model explains the labeling data and found that the data was best explained by a network that included two active diacylglycerol pools that are successive precursor to triacylglycerol biosynthesis and the first diacylglyerol pool can exchange with the active phosphatidylcholine pool. In addition, we performed short-term ^{14}C acetate labeling experiments to track how newly synthesized or recently elongated fatty acids are incorporated into the different lipid classes. We found evidence for a large phosphatidylcholine pool that is used for acyl editing and is separate from the pool used for *de novo* triacylglycerol synthesis

Together, these endeavors demonstrate how models can be integrated with a broad range of data to address questions at ecological, organismal, and metabolic scales.

ACKNOWLEDGEMENTS

I would like to begin by acknowledging my mentor, Dr. Yair Shachar-Hill, who has been enormously supportive and inspiring. Under his guidance, I have learned to write more precisely and think more critically. I am especially grateful that he has encouraged me to integrate experimental and mathematical approaches in my research, a practice I intend to maintain in my upcoming postdoctoral position.

Secondly, I would like to thank my committee members for their instruction and support: Dr. Christoph Benning for reminding me to recognize the biological implications of my work, Dr. Maren Friesen for helping me learn how to be an effective collaborator, and Dr. Shinhan Shiu for introducing me to Python which has improved my research efficiency.

Next, I would like to acknowledge my collaborators Colleen Friel, Maren Friesen, and Emily Grman, as well as the labs who have provided guidance or have let us use their equipment, including the Friesen, Kramer, Ohlrogge, and Sharkey labs.

I would also like to acknowledge the past and present members of the Shachar-Hill lab for their contributions and camaraderie, including Lisa Carey, Karolina Czarnecki, Danielle Delamarter, Rahul Deshpande, Bradley Disbrow, Matt Juergens, Taghleb Al-Deeb, Michael Opperman, Mike Pollard, Shawna Rowe, Yuan Xu, and Danielle Young.

Lastly, I would like to thank my mother, who instilled in me high standards and has helped me financially; and my husband, who is always supportive, consistently shares his mechanical expertise, and keeps me from overstressing.

Teresa J. Clark

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

ACP	acyl carrier protein
acDAG	acetyl DAG
ATP	adenosine triphosphate
CPT	cholinephosphotransferase
DAF	days after flowering
DAG	diacylglycerol
DGAT	diacylglycerol acyltransferase
DPM	disintegrations per minute
FBA	flux balance analysis
G3P,	glyceraldehyde 3-phosphate
GC-FiD	gas chromatography with flame-ionization detection
GC-MS	gas chromatography–mass spectrometry
GPAT	glycerol-3-phosphate acyltransferase
HCA	hierarchical clustering analysis
LPAAT	lysophosphatidic acid acyltransferase
MFA	metabolic flux analysis
NMR	nuclear magnetic resonance
OPP	oxidative pentose phosphate
PA	phosphatidic acid
PAP	phosphatidic acid phosphatase
PC	phosphatidylcholine
PC	phosphatidylcholine

PCA	principal component analysis
PDAT	phospholipid:diacylglycerol acyltransferase
PDCT	phosphocholine transferase
PLA_2	phospholipase A_2
TAG	triacylglycerol
TCA	tricarboxylic acid cycle
TLC	thin-layer chromatography

INTRODUCTION

Plants are essential to human life. Not only do they provide food for people and their animals, fiber and other structural materials, fuels, and chemical feedstocks, but also their specialized metabolites can be used in medicines, dyes, fragrances, and other applications (Durrett *et al.*, 2008; Dyer *et al.*, 2008; Krishnaiah *et al.*, 2011; Mosiewicki & Aranguren, 2013). The work described in this dissertation is aimed at predicting how plants obtain and use nutrients to synthesize these or other biomass products, which could be expanded to rationally engineer plants.

It is obvious why there is great interest in designing plant metabolism to increase the production of compounds with significant economic value; however, plant nutrients can have other important, yet often overlooked, roles in their environment and ecosystems. For example, microbial communities are indirectly controlled by exudate and litter production because these compounds transfer carbon, nitrogen, and other nutrients from the plant to the rhizosphere (Brüggemann *et al.*, 2011). Plant nutrient levels can also internally influence the plant's physiology, such as potassium contributing to the effectiveness of plant defenses against fungal diseases, bacteria, insects, viruses, and other pests (Wang *et al.*, 2013), and nitrogen availability has been shown to affect how plants allocate their carbon to carbohydrates and organic acids (Nunes-Nesi *et al.*, 2010; Xu *et al.*, 2012).

Plant production of particular metabolites is often investigated with bioengineering. For example, "golden rice" is a transgenic rice that accumulates provitamin A (β-carotene) in its endosperm as a counter for vitamin A deficiency (Beyer *et al.*, 2002). This crop was developed from research on genes in the terpenoid pathway in daffodil and measuring precursor (e.g., phytoene) levels in experiments where daffodil genes were expressed in rice (Burkhardt *et al.*, 1997; Potrykus, 2001). Seed oils have also been engineered to improve human health (e.g., to decrease the risk of inflammatory diseases) by producing high levels of omega-3 and/or -6 long-chain polyunsaturated fatty acids (Haslam *et al.*, 2013). In one such endeavor, *Brassica napus* was transformed with fungal desaturases that led to increased accumulation of stearidonic acid (18:4 Δ 6,9,12,15) and γ-linolenic acid (18:3 Δ6,9,12; Liu *et al.*, 2001; Haslam *et al.*, 2013).

In addition, Siminszky (2005) identified a cytochrome P450 as being positively correlated with accumulation of a carcinogen precursor in tobacco plants by measuring transcript abundance in low and high carcinogen-producing plants. Knocking down this gene product was found to lower the accumulation in tobacco plants. A focus of much current work has been to develop non-leguminous crops to form mutualistic relationships with nitrogen-fixing bacteria in order to decrease the need for fertilizers (Saikia & Jain, 2007; Beatty & Good, 2011). This may require the crop to be engineered to produce specific exudates or other signals because, for example, rhizobia have been shown to form effective or ineffective nodules in response to exudates from different plants (Currier & Strobel, 1976).

Genetic engineering is a powerful tool, but it can also have unexpected metabolic effects. For example, *Arabidopsis thaliana* has been engineered to over-express a maize gene whose product down-regulates lignin biosynthesis (Fornalé *et al.*, 2010). The transgenic *A. thaliana* line was found to not only contain less lignin, but also accumulated anthocyanin. It was suggested that the accumulation was due to the potential lignin precursor being redirected toward anthocyanin biosynthesis. While this points to potential approaches for investigating the role of nutrient fluxes in plants, fluxes can be quantitatively studied in more depth with the use of isotopically labeled compounds (Brüggemann *et al.*, 2011; Christophe *et al.*, 2011; Shachar-Hill, 2013). For example, nitrogen labeling has been used to investigate nitrogen transfer in the plant-arbuscular mycorrhizal fungi symbiosis, which confirmed that the fungi are capable of taking up and transferring nitrogen to their hosts (Govindarajulu *et al.*, 2005). This study also used carbon labeling and revealed that the nitrogen is incorporated into amino acids and translocated to the plants without the transfer of carbon. Labeling experiments such as these are effective at analyzing nutrient fluxes, but these methods tend to be limited in how many branched fluxes they can simultaneously and quantitatively measure. Here, additional nutrition and metabolic insights that can be obtained with isotopic labeling include mapping the carbon fluxes necessary to provide the materials and energy required for fungal nitrogen uptake and transfer, or mapping how the plant obtains and uses the

traded nitrogen without requiring the simultaneous transfer of carbon. To investigate fluxes on a network scale, where there can be are numerous branched reactions involved, it can be efficient to integrate experimental data with a mathematical model.

Computational modeling approaches

Mathematical models can be designed to quantify nutrient flux networks and predict the effects of changing the abundance of nutrients, enzyme activities, or other factors. There is a wide variety of mathematical model types that can be applied to biological systems, each suited for different purposes, often operating at different scales, and with different capabilities for data integration. The two main types of computational models are statistical and mechanistic.

In general, statistical models allow biologists to identify significant trends in the data, and hence to generate hypotheses about causal relationships from the correlations. These models tend to start with experimental data and they try to construct a model to represent how biological processes and/or characteristics led to the generation of that data (Oberg & Mahoney, 2007). For example, multifactorial statistical models have been used to analyze published data on performance of plants inoculated with mycorrhizal fungi and those that were not inoculated (Hoeksema *et al.*, 2010). Among other trends, this meta-analysis found that the plant's response to inoculation was primarily explained by the plant's functional group (e.g., C_3 or C_4 grasses, a forb capable of mutualistic interactions with nitrogen-fixing bacteria) and the degree to which the soil had been fertilized with nitrogen. Although this model does not investigate the fluxes needed for mycorrhizal fungi to take up and translocate nitrogen, it points to variables to be tested with labeling-based methods.

In contrast, mechanistic modeling methods typically involve constructing a model to describe the system and gathering data to test the model's predictions and/or assumptions (Morgan & Rhodes, 2002). These models can be constructed to represent different biological scales (e.g., ecological, metabolic),

using different types of assumptions, such as 1) the processes that generate and maintain fitness advantages for two organisms interacting, 2) rules concerning how metabolites are moved and transformed in a particular network, and 3) principles of chemical reaction kinetics. This introduction gives an overview of how models with these assumptions have been combined with experimental measurements to study nutrient uptake, transformation, and exchange in plant systems. I then present some of the challenges and methods of statistically analyzing model fits to data.

Models representing processes that drive organisms to interact

Organisms interact in a myriad of ways, including being members of a population, competitors for the same resource, and part of a symbiosis where there are mutual benefits. These interactions can be studied with Lotka Volterra, consumer-resource, and other population dynamics models (Holland & DeAngelis, 2010; Elser *et al.*, 2012). Such studies have found, for example, that seasonal plant harvesting can prevent competitive exclusion between plant species and thus lead to coexistence of plants that are strongly competitive (Geijzendorffer *et al.*, 2011), as well as the potential for species densities alone inducing shifts between competitive, mutualistic, and parasitic interactions between populations (Holland & DeAngelis, 2010).

Game theory studies have also been fruitful in pointing to potential mechanisms influencing nutritional interactions, showing, for example, that in a mutualism, the partner that evolves slower is likely to benefit more from the interaction (Bergstrom & Lachmann, 2003), and that mutualisms could be maintained even if defectors ("cheaters") cannot be punished through a Public Gods game (Archetti & Scheuring, 2013). However, such interaction-based models can be difficult to parameterize and integrate or test with data because their parameters can be difficult to measure or even define, empirically (Clark *et al.*, 2017). Such cases include many game theory model components; for example, how long an individual will stay in the war-of-attrition contest (Akçay & Simms, 2011), and the steepness of the

sigmoidal function that relates the frequency of cooperators to the quantity of benefits received (Archetti & Scheuring, 2013). Due to this challenge, many of these models are considered theoretical rather than computational because their primary function is to understand how the interactions may in principle function rather than to quantify system traits. Game theory and other modeling approaches that have been applied to plant-microbe nutrient-exchange mutualisms are reviewed in Chapter 1.

Models representing metabolic fluxes in biochemical networks

At the level of the cell, tissue or organism, Metabolic Flux Analysis (MFA) and Flux Balance Analysis (FBA) are modeling methods to track how nutrients are transported and transformed at the level of biochemical reaction networks (Niedenführ *et al.*, 2015). MFA is heavily dependent on experimental data because the models track carbon isotopic labeling within metabolites and uses isotopomer measurements for model fitting (Wiechert *et al.*, 2001; Libourel & Shachar-Hill, 2008; Shachar-Hill, 2013; Antoniewicz, 2015). Due to the need for data for fitting, the modeled network is an abbreviated form of the biological network, and only includes reactions for which data has been measured and during which there was a measureable change in carbon positional labeling. In steady-state MFA models, the concentration of internal metabolites (pathway intermediates) are assumed to be constant during the time interval of interest, and the fluxes between these pools are also assumed to be constant. This type of model has been used, for example, to analyze the effects of oxygen availability on central metabolism in *A. thaliana* cell cultures (Williams *et al.*, 2008). It was found that, although there were significant difference in metabolic profiles, the proportions of carbon used for various reactions were unchanged between the oxygenation conditions, suggesting that there may not be a strong connection between metabolomics data and fluxes. ^{13}C MFA studies of developing oilseeds are the subject of Chapter 2.

FBA differs from MFA in several ways (García Sánchez & Torres Sáez, 2014; Junker, 2014). First, FBA networks aim to represent the entire (genome-scale) metabolic and transport network rather than a

condensed metabolic network or sub-network. This typically causes FBA networks to have hundreds more reactions than MFA. Second, because FBA does not track isotopic labeling, it uses much less data than MFA. It is common for plant FBA studies to only require steady-state substrate and light uptake rates, as well as products and biomass efflux rates. Third, due to the increased network size, FBA uses an "objective function" or operational target such as maximizing growth in order to transform the model solution task into a global optimization problem, which greatly reduces the very large number of potential routes and rates through which the metabolites can move (i.e., feasible space) to a subset of flux patterns that meet the objective. One crucial step in FBA modeling is therefore to identify the most appropriate objective function for the system being investigated. Potential objective functions include growth rate (Chapman *et al.*, 2015), carbon use efficiency (Rolletschek *et al.*, 2015), maximum production of a particular metabolite (Savakis & Hellingwerf, 2015), or minimum use of light (Boyle & Morgan, 2009). By integrating the network, data, and objective function, FBA can predict values for the network fluxes. FBA has been used to determine why mixotrophic *Chlamydomonas reinhardtii* cultures have lower photosynthesis yet higher growth rates than autotrophic cultures (Chapman *et al.*, 2015). To address this question, their objective function maximized growth rate. They found that the acetate provided to the mixotrophic cells was able to be used in a modified tricarboxylic acid cycle that bypassed the decarboxylation reactions (e.g., the glyoxylate cycle).

Models derived from chemical reaction kinetics

Chemical reaction kinetics have been well studied (Kramers, 1940; Hänggi *et al.*, 1990) and most biochemical network models that integrate labeling data are comprised of first-order rate equations. The rates of these reactions are proportional to the concentrations of a reactant, which allows the dynamics of metabolite pools and fluxes to be mathematically represented as ordinary differential equations. For example, in the pathway A → B → C, the rate of change of metabolite B can be written as

$\frac{dB}{dt} = k_1[A] - k_2[B]$, where k_1 and k_2 are rate constants for the first and second, respectively, reactions, and [A] and [B] are the metabolite concentrations. Such a network of first-order rate equations was constructed by (Hibino *et al.*, 2011) to model how a cytoplasmic serine/threonine kinase (RAF) is activated on the plasma membrane in response to a GTPase being activated. They integrated time-course single-molecule imaging data with the model to predict values of the pseudo-first order rate constants and to quantify the importance of the involvement of other enzymes in order to activate the GTPase. Values for pseudo-first order rate constants have also been estimated on a smaller scale to compare those for starch amylolysis for boiled potatoes, white bread, peas, and other starches to investigate if digestion rates change over time (Butterworth *et al.*, 2012). Chapter 3 describes how dynamic modeling constants can be quantified in order to analyze the network topology of oil biosynthesis.

Most biochemical network models of photosynthesis are constructed using enzyme kinetic equations. In these, all but the maximum rubisco and electron transport activity, daytime respiration and CO_2 diffusion rate parameters typically have *a priori* estimates (Von Caemmerer, 2013). One such model extended this approach to the leaf canopy scale to test how a leaf's canopy position and physiological traits affect its photosynthetic capacity (Prieto *et al.*, 2012). It was found that variations in the capacity due to canopy height were best predicted by leaf nitrogen content. (Nuccio *et al.*, 2000) used an enzyme kinetics model to test why glycine betaine synthesis (a metabolite which enhances plant stress resistance) is limited by the transport of its choline precursor into the plastid. By using ^{14}C choline to quantify kinetic parameters (K_m and V_{max}) for comparison between low and high glycine betaine-accumulating tobacco lines as well as between other plant species, they hypothesized that glycine betaine abundance may be correlated with the transporter's affinity for choline.

Challenges in statistically testing mechanistic models

In most biological studies, it is expected that the researchers perform statistical analyses on the results in

order to objectively assess if the finding is significant. These analyses often include statistic deviations, t-tests, ANOVA, regression analyses, chi-squared tests, Pearson or Spearman correlations, etc. (Saltelli & Marivoet, 1990) These analyses are challenging or impossible to perform on the results of many mechanistic models because the analyses assume a normal distribution of or independence in the results, but these characteristics are lacking in many mechanistic models:

Biological measurements are typically assumed to be normally distributed (Nakagawa & Schielzeth, 2010), including population growth rate, organismal biomass, and nutrient uptake rate per individual, which allows the calculation of averages and standard deviations for these traits. If a model requires these traits to be multiplied, such as calculating the population biomass growth rate from the population rate and organismal biomass measurements, then the results can still be expected to have normal distribution. However, if the model requires the division of normally distributed parameters, such as calculating the substrate uptake rate per individual biomass from the substrate uptake rate per individual and the organismal biomass, then the result can no longer be assumed to be normally distributed and thus is considered non-Gaussian (Nakagawa & Schielzeth, 2010).

MFA and FBA predict the value of all fluxes within the modeled network. If one reaction is changed, such as decreasing the flux from pyruvate to acetyl CoA in the tricarboxylic acid **cycle**, then other fluxes will be affected and changed in response, such as less pyruvate being produced from glycolytic hexose or more pyruvate serving as a substrate for alanine precursors. Models built from first-order rate equations will be similarly affected because the reactions are a network.

Statistically analyzing models with confidence intervals

An alternative statistical measure is confidence intervals. In some ways, they function similarly to standard deviations to identify a range, centered on the average, in which the prediction is statistically likely to be (Streiner, 1996; Nakagawa & Cuthill, 2007). For standard deviations of normally distributed systems, 68%

of the relevant values are expected to be within one standard deviation of the average and 95% are expected to be within two. Confidence intervals are derived from the observed standard deviation and samples size as well as the desired confidence interval. When a 90% confidence interval is obtained, then 90% of the relevant values are expected to be within one interval of the average (Streiner, 1996).

Confidence intervals are typically used to declare the statistical range of predictions, given the observed variation in model input data. This can be done by running the model for all input data (with deviations) and generating the set of reasonable output predictions, which would be used for determining the average and confidence interval of the predictions. A drawback to using confidence intervals as a statistical tool is that, unlike *p*-values, there is not yet an agreed-upon confidence level that is considered to be statistically significant (Nakagawa & Cuthill, 2007). The uncertain statistical power is a challenge for this statistical test, but I am not aware of other measures that could be used for this purpose in non-Gaussian mechanistic models.

Another challenge in statistically analyzing complex mechanistic models is identifying the input data to be used to generate confidence intervals. When all input parameters can be measured for each biological replicate, then the input data to be used is simply the measured datasets. However, when the model cannot be fully ran per individual replicate, then the input data must be used in a different way. An example of when this would be needed is if two destructive measurements are needed for one parameter, such as measuring a plant's growth rate from the total biomass at two time points (see Chapter 1). If measuring total biomass requires destruction of the plant, then at least two plants need to be used in order to measure the plant growth rate. Another example is when the biological system needs to be independently exposed to two or more labeled substrates at the same age (see Chapter 2). Because the organism (or tissues or cells) can only be independently exposed to one substrate at a given age, at least two biological samples would be needed to generate a complete dataset.

Consequently, in these situations, the input data for confidence interval determination needs to

be simulated in order to represent observed uncertainties, without implying that the uncertainties match up a certain way (e.g., in the above plant growth rate example, if one plant is measured to be small at the first time point, then should the growth rate be determined by comparing this small plant to the smallest plant at the second time point?). This can be done by generating pseudo datasets with Monte Carlo sensitivity analysis (Doubilet *et al.*, 1985; Saltelli & Marivoet, 1990; Alonso *et al.*, 2007). In this method, pseudo data is created for each of the measured values using the measured average and standard deviation. The measured values are usually assumed to have normal distribution, so a pseudo point for that measurement would be the measured average plus the standard deviation times a number randomly generated to have normal distribution, a standard deviation of 1, and be centered at 0. This would cause the pseudo data to mimic the measured because it would have the same average and standard deviation. Pseudo data would be generated for all measurements and then used to run the model. Therefore, each pseudo dataset would generate exactly 1 set of model predictions. By using a large number of pseudo datasets, only appropriate variations in model predictions should be observed. Using the growth rate example, if by chance, a small pseudo plant at the first time point is matched with a large pseudo plant at the second time point, the pseudo growth rate would appear very large. If enough pseudo datasets are used, then outliers like this would be outweighed by more reasonable pseudo predictions; thus the confidence range would properly represent reasonable variation.

In the following chapters, I describe my collaborative work to integrate mechanistic models with experimental data. First, I modeled the mutualistic relationship between legume *Medicago truncatula* and rhizobial bacteria to evaluate the effects of soil nitrogen content on nutrient allocation and trade. Second, I used MFA to determine why *Camelina sativa* developing embryos have low carbon use efficiency and how their metabolism was affected by light availability and genetic engineering. Lastly, I explored potential topologies of the triacylglycerol biosynthesis network by developing the network topologies as sets of first-order reactions and comparing model fits with ^{14}C glyceryl and ^{14}C acyl time-course data.

Together, these efforts demonstrate the potential of integrating experimental data with mathematical models to understand how plants use carbon, nitrogen, and other resources to produce oils, proteins, and other biomass products.

REFERENCES

REFERENCES

Akçay E, Simms EL. 2011. Negotiation, sanctions, and context dependency in the legume-rhizobium mutualism. *The American Naturalist* 178(1): 1-14.

Alonso AP, Goffman FD, Ohlrogge JB, Shachar-Hill Y. 2007. Carbon conversion efficiency and central metabolic fluxes in developing sunflower (Helianthus annuus L.) embryos. *The Plant Journal* 52(2): 296-308.

Antoniewicz MR. 2015. Methods and advances in metabolic flux analysis: a mini-review. *Journal of industrial microbiology & biotechnology* 42(3): 317-325.

Archetti M, Scheuring I. 2013. Trading public goods stabilizes interspecific mutualism. *Journal of theoretical biology* 318: 58-67.

Beatty PH, Good AG. 2011. Future prospects for cereals that fix nitrogen. *science* 333(6041): 416-417.

Bergstrom CT, Lachmann M. 2003. The Red King effect: when the slowest runner wins the coevolutionary race. *Proceedings of the National Academy of Sciences* 100(2): 593-598.

Beyer P, Al-Babili S, Ye X, Lucca P, Schaub P, Welsch R, Potrykus I. 2002. Golden rice: Introducing the β-carotene biosynthesis pathway into rice endosperm by genetic engineering to defeat vitamin A deficiency. *The Journal of nutrition* 132(3): 506S-510S.

Boyle NR, Morgan JA. 2009. Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii. *BMC systems biology* 3(1): 4.

Brüggemann N, Gessler A, Kayler Z, Keel S, Badeck F, Barthel M, Boeckx P, Buchmann N, Brugnoli E, Esperschütz J. 2011. Carbon allocation and carbon isotope fluxes in the plant-soil-atmosphere continuum: a review. *Biogeosciences* 8(11): 3457-3489.

Burkhardt PK, Beyer P, Wünn J, Klöti A, Armstrong GA, Schledz M, Von Lintig J, Potrykus I. 1997. Transgenic rice (Oryza sativa) endosperm expressing daffodil (Narcissus pseudonarcissus) phytoene synthase accumulates phytoene, a key intermediate of provitamin A biosynthesis. *The Plant Journal* 11(5): 1071-1078.

Butterworth PJ, Warren FJ, Grassby T, Patel H, Ellis PR. 2012. Analysis of starch amylolysis using plots for first-order kinetics. *Carbohydrate Polymers* 87(3): 2189-2197.

Chapman SP, Paget CM, Johnson GN, Schwartz J-M. 2015. Flux balance analysis reveals acetate metabolism modulates cyclic electron flow and alternative glycolytic pathways in Chlamydomonas reinhardtii. *Frontiers in plant science* 6: 474.

Christophe S, Jean-Christophe A, Annabelle L, Alain O, Marion P, Anne-Sophie V 2011. Plant N fluxes and modulation by nitrogen, heat and water stresses: A review based on comparison of legumes and non legume plants. *Abiotic Stress in Plants-Mechanisms and Adaptations*: InTech.

Clark TJ, Friel CA, Grman E, Shachar-Hill Y, Friesen ML. 2017. Modelling nutritional mutualisms: challenges and opportunities for data integration. *Ecology letters* 20(9): 1203-1215.

Currier WW, Strobel GA. 1976. Chemotaxis of Rhizobium spp. to plant root exudates. *Plant physiology* 57(5): 820-823.

Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. 1985. Probabilistic sensitivity analysis using Monte Carlo simulation: a practical approach. *Medical decision making* 5(2): 157-177.

Durrett TP, Benning C, Ohlrogge J. 2008. Plant triacylglycerols as feedstocks for the production of biofuels. *The Plant Journal* 54(4): 593-607.

Dyer JM, Stymne S, Green AG, Carlsson AS. 2008. High-value oils from plants. *The Plant Journal* 54(4): 640-655.

Elser JJ, Loladze I, Peace AL, Kuang Y. 2012. Lotka re-loaded: modeling trophic interactions under stoichiometric constraints. *Ecological Modelling* 245: 3-11.

Fornalé S, Shi X, Chai C, Encina A, Irar S, Capellades M, Fuguet E, Torres JL, Rovira P, Puigdomènech P. 2010. ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. *The Plant Journal* 64(4): 633-644.

García Sánchez CE, Torres Sáez RG. 2014. Comparison and analysis of objective functions in flux balance analysis. *Biotechnology progress* 30(5): 985-991.

Geijzendorffer I, Van der Werf W, Bianchi F, Schulte R. 2011. Sustained dynamic transience in a Lotka–Volterra competition model system for grassland species. *Ecological Modelling* 222(15): 2817-2824.

Govindarajulu M, Pfeffer PE, Jin H, Abubaker J, Douds DD, Allen JW, Bücking H, Lammers PJ, Shachar-Hill Y. 2005. Nitrogen transfer in the arbuscular mycorrhizal symbiosis. *Nature* 435(7043): 819.

Hänggi P, Talkner P, Borkovec M. 1990. Reaction-rate theory: fifty years after Kramers. *Reviews of modern physics* 62(2): 251.

Haslam RP, Ruiz-Lopez N, Eastmond P, Moloney M, Sayanova O, Napier JA. 2013. The modification of plant oil composition via metabolic engineering—better nutrition by design. *Plant Biotechnology Journal* 11(2): 157-168.

Hibino K, Shibata T, Yanagida T, Sako Y. 2011. Activation kinetics of RAF in the ternary complex of RAF, RASGTP, and kinase on the plasma membrane of living cells: single-molecule imaging analysis. *Journal of Biological Chemistry*: jbc. M111. 262675.

Hoeksema JD, Chaudhary VB, Gehring CA, Johnson NC, Karst J, Koide RT, Pringle A, Zabinski C, Bever JD, Moore JC. 2010. A meta-analysis of context-dependency in plant response to inoculation with mycorrhizal fungi. *Ecology letters* 13(3): 394-407.

Holland JN, DeAngelis DL. 2010. A consumer–resource approach to the density-dependent population

dynamics of mutualism. *Ecology* 91(5): 1286-1295.

Junker BH. 2014. Flux analysis in plant metabolic networks: increasing throughput and coverage. *Current opinion in biotechnology* 26: 183-188.

Kramers H. 1940. HA Kramers, Physica 7, 284 (1940). *Physica* 7: 284.

Krishnaiah D, Sarbatly R, Nithyanandam R. 2011. A review of the antioxidant potential of medicinal plant species. *Food and bioproducts processing* 89(3): 217-233.

Libourel IG, Shachar-Hill Y. 2008. Metabolic flux analysis in plants: from intelligent design to rational engineering. *Annu. Rev. Plant Biol.* 59: 625-650.

Liu J, Huang Y, DeMichele S, Bergana M, Bobik E, Hastilow C, Chuang L, Mukerji P, Knutzon D. 2001. Evaluation of the seed oils from a canola plant genetically transformed to produce high levels of c-linolenic acid. *Linolenic Acid: Recent Advances in Biotechnology and Clinical Applications. AOCS Press, Champaign, IL*: 61-71.

Morgan JA, Rhodes D. 2002. Mathematical modeling of plant metabolic pathways. *Metabolic engineering* 4(1): 80-89.

Mosiewicki MA, Aranguren MI. 2013. A short review on novel biocomposites based on plant oil precursors. *European Polymer Journal* 49(6): 1243-1256.

Nakagawa S, Cuthill IC. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews* 82(4): 591-605.

Nakagawa S, Schielzeth H. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews* 85(4): 935-956.

Niedenführ S, Wiechert W, Nöh K. 2015. How to measure metabolic fluxes: a taxonomic guide for 13C fluxomics. *Current opinion in biotechnology* 34: 82-90.

Nuccio ML, McNeil SD, Ziemak MJ, Hanson AD, Jain RK, Selvaraj G. 2000. Choline import into chloroplasts limits glycine betaine synthesis in tobacco: analysis of plants engineered with a chloroplastic or a cytosolic pathway. *Metabolic engineering* 2(4): 300-311.

Nunes-Nesi A, Fernie AR, Stitt M. 2010. Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Molecular plant* 3(6): 973-996.

Oberg AL, Mahoney DW 2007. Linear mixed effects models. *Topics in biostatistics*: Springer, 213-234.

Potrykus I. 2001. Golden rice and beyond. *Plant physiology* 125(3): 1157-1161.

Prieto JA, Louarn G, Perez Pena J, Ojeda H, Simonneau T, Lebon E. 2012. A leaf gas exchange model that accounts for intra-canopy variability by considering leaf nitrogen content and local acclimation to radiation in grapevine (Vitis vinifera L.). *Plant, Cell & Environment* 35(7): 1313-1328.

Rolletschek H, Grafahrend-Belau E, Munz E, Radchuk VV, Kartäusch R, Tschiersch H, Melkus G, Schreiber F, Jakob PM, Borisjuk L. 2015. Metabolic architecture of the cereal grain and its relevance to maximize carbon use efficiency. *Plant physiology*: pp. 00981.02015.

Saikia S, Jain V. 2007. Biological nitrogen fixation with non-legumes: An achievable target or a dogma? *Current science*: 317-322.

Saltelli A, Marivoet J. 1990. Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques. *Reliability Engineering & System Safety* 28(2): 229-253.

Savakis P, Hellingwerf KJ. 2015. Engineering cyanobacteria for direct biofuel production from CO2. *Current opinion in biotechnology* 33: 8-14.

Shachar-Hill Y. 2013. Metabolic network flux analysis for engineering plant systems. *Current opinion in biotechnology* 24(2): 247-255.

Siminszky B, Gavilano L, Bowen SW, Dewey RE. 2005. Conversion of nicotine to nornicotine in Nicotiana tabacum is mediated by CYP82E4, a cytochrome P450 monooxygenase. *Proceedings of the National Academy of Sciences* 102(41): 14919-14924.

Streiner DL. 1996. Maintaining standards: differences between the standard deviation and standard error, and when to use each. *The Canadian journal of psychiatry* 41(8): 498-502.

Von Caemmerer S. 2013. Steady-state models of photosynthesis. *Plant, Cell & Environment* 36(9): 1617-1630.

Wang M, Zheng Q, Shen Q, Guo S. 2013. The critical role of potassium in plant stress response. *International journal of molecular sciences* 14(4): 7370-7390.

Wiechert W, Möllney M, Petersen S, de Graaf AA. 2001. A universal framework for 13C metabolic flux analysis. *Metabolic engineering* 3(3): 265-283.

Williams TC, Miguet L, Masakapalli SK, Kruger NJ, Sweetlove LJ, Ratcliffe RG. 2008. Metabolic network fluxes in heterotrophic Arabidopsis cells: stability of the flux distribution under different oxygenation conditions. *Plant physiology* 148(2): 704-718.

Xu G, Fan X, Miller AJ. 2012. Plant nitrogen assimilation and use efficiency. *Annual review of plant biology* 63: 153-182.

CHAPTER 1

Plant-microbe nutrient exchange symbioses

PREFACE

This chapter presents the manuscripts resulting from a collaborative project with Colleen A. Friel (CF), Maren L. Friesen (MF), Emily Grman (EG), Yair Shachar-Hill (YSH, and myself (TJC). The project grew out of a proposal written in a class, *PLB 803: Integrative Topics in Plant Biology (Models in Plant Biology)*, by four graduate students (including CF and myself) that aimed to integrate experiments and mathematical modeling to answer a question that spanned molecular and ecological themes. Briefly, our proposal sought to understand how plant-microbe mutualisms affect plant community dynamics using a tripartite mutualism system (legume-rhizobia-mycorrhizal fungi), two types of models (one to predict the mutualistic nutrient trade rates, and one to predict the effects of a mutualistic relationship on a population's fitness and ability to compete), and a variety of experiments to measure the model parameters and test its predictions.

With the encouragement of the instructors (YSH and Chris Klausmeier, CK), I organized a collaboration of three groups—TJC. & YSH, CF & MF, and EG & CK—with the intent of pursuing these ideas. My roles in organizing the collaboration included coordinating our meetings the efforts of the collaborators, and keeping track of our progress.

In Section 1.1, I present the manuscript describing our experimental and modeling work, which is being revised for resubmission to *New Phytologist* where it received favorable reviews. In this study we made extensive measurements on the effects of soil nitrogen content on nutrient uptake and allocation, and partner growth and trade in the mutualistic relationship between legume *Medicago truncatula* and rhizobial bacteria and used these data to parameterize and test trade-based models of nutrient exchange. In Section 1.2, I present a published article we wrote to critically review how experiments have been and could further be integrated with mathematical models of plant-microbe nutritional symbioses at a range of biological scales.

Together, these endeavors demonstrate how mechanistic models can be used to better

understand the effects of environmental conditions and partner physiology on the functioning of plant-microbe nutrient-exchange mutualisms.

SECTION 1.1

Unfair trade underground revealed by integrating data with Nash bargaining models

---

This section includes a manuscript in revision after review at *New Phytologist*:
Clark TJ, CA Friel, E Grman, ML Friesen, and Y Shachar-Hill. Unfair trade underground revealed by integrating data with Nash bargaining models

Preface

As part of the collaboration of ecologists and molecular biologists, I adapted, tested, and refined a mathematical model to explore the trading principles underlying the effects of soil nitrogen content on nutrient allocation and exchange in the mutualistic relationship between legume *Medicago truncatula* and nitrogen fixing bacteria. The model was originally developed to make quantitative estimates of population growth and nutrient exchange rates in grass-arbuscular mycorrhizae mutualisms using ranges of literature values for model parameters. I refined and applied this model to the legume-rhizobia mutualism and our group identified experimental methods to measure the parameters.

Experiments were performed on *M. truncatula* grown in the presence or absence of rhizobia, and at low to high soil nitrogen availability to test model fit under high to low, respectively, plant investment in trade. CF and I contributed equally to the experimental work, including growing, maintaining, and harvesting the plants. In addition to shared work on growth experiments, I was responsible for measuring 1) plant carbon fixation and respiration measurements by designing (with YSH) and using custom-built whole-plant photosynthesis chambers and LI-COR systems, 2) sample preparation for plant nitrogen uptake measurements with ^{15}N labeling for analysis at the Stable Isotope Lab at Utah State University, and 3) sample preparation for carbon and nitrogen elemental composition analyses of roots, shoots, and nodules by the Robertson lab at the Kellogg Biological Station. Many of CF's experimental contributions were not included in this manuscript, such as leading efforts to identify four-to-six weeks as a suitable range for our measurements, and performing PhotosynQ and root imaging measurements.

I led and performed the modeling, with important input from the other collaborators. The model was written in Mathematica, so any tests I made to the model structure were performed in that program. These included testing 1) whether the parameters should be per biomass carbon (i.e., specific rates) or per plant system (i.e., organismal rates), 2) the effects of adding a nodule respiratory cost for nitrogen fixation, and 3) the effects of allowing partners to have unequal bargaining power. Some of these tests

included a scaling parameter, such as testing the plant's bargaining power as it varied from 1% to 99%. For these scaling tests, I used Python and a batch server to write and execute the Mathematica codes in parallel and at modest intervals (e.g., I tested bargaining power at 1% increments).

I integrated our measurements into the model using Excel. I first formatted the empirical data to be compatible with the model structure being tested. This formatting often required dividing one measurement by a biologically different one (e.g., biomass increase divided by biomass carbon content) and thus required analysis by Monte-Carlo type sampling to generate pseudo datasets for fitting. Some of my model tests included testing how the data should be formatted. For example, the plant carbon uptake parameter was primarily derived from whole-plant photosynthesis measurements. I tested how (if at all) I should reduce the measured photosynthesis rate by plant respiratory costs (either measured under dark conditions or extrapolated from literature findings). In addition, due to its original use, the model did not include environmental parameters. To test the model fit at the different soil nitrogen levels, I created separate pseudo datasets by Monte Carlo sampling for each nitrogen level.

The goodness of model fits were assessed using confidence intervals resulting from the pseudo datasets. As a group, our collaboration identified potential model refinements to be tested and evaluated whether they sufficiently improved the fit. The manuscript writing was a collaborative effort, with CF leading the empirical findings writing and formatting the figures, while I led writing about the computational findings and model context.

Our manuscript is currently being revised in response to reviewer feedback at *New Phytologist*. A recent copy of our manuscript is included in this section. As coordinator of the collaboration and (co)first author, I am handling the (re)submission process for this manuscript.

Summary

- Mutually beneficial resource exchange is foundational to global biogeochemical cycles and plant and animal nutrition. However, there is inherent potential conflict in mutualisms, as each organism benefits more when the exchange ratio ("price") minimizes its own costs and maximizes its benefits. Understanding the bargaining power that each partner has in these interactions is key to our ability to predict the exchange ratio and thus the functionality of the cell, organism, community, and ecosystem.
- We tested whether partners have symmetric ("fair") or asymmetric ("unfair") bargaining power in the legume-rhizobia nitrogen fixing symbiosis using measurements of carbon and nitrogen dynamics in a mathematical modeling framework derived from economic theory.
- A model of symmetric bargaining power was not consistent with our data. Instead, our data indicate that the growth benefit to the plant has greater weight in determining trade dynamics than the benefit to the bacteria. Quantitative estimates of the relative power of the plant reveal that the plant's influence rises as soil nitrogen availability decreases and trade benefits to both partners increase.
- Our finding that legumes have more bargaining power than rhizobia at lower nitrogen availabilities highlights the importance of context-dependence for the evolution of mutualism with increasing nutrient deposition.

Introduction

Mutualistic relationships abound in nature. They are rooted in the exchange of resources or services between different partners whose distinct capabilities allow them to perform better together than either could alone. Mutualisms involving the exchange of carbon for mineral nutrients between plants and microbes are ancient interactions that have shaped the evolution of land plants and play central roles in ecosystem functioning worldwide (Bronstein, 2015). Plants participating in these nutritional mutualisms with microbes such as mycorrhizal fungi or nitrogen-fixing rhizobium prokaryotes must optimize their

allocation of photosynthate between taking up nutrients directly and trading for them with mutualists (Bloom *et al.*, 1985). Indeed, plants exhibit considerable plasticity in partitioning carbon among shoots, roots, and mutualistic partners in response to environmental cues (Harris *et al.*, 1985; Wang *et al.*, 2011). This optimal allocation is determined by soil nutrient availability and the cost:benefit ratio of acquiring the nutrient through trade. However, it is in each partner's best interest to influence the carbon-for-nutrient exchange ratio ("price") to maximize the benefit to itself (Akçay & Roughgarden, 2007; Grman *et al.*, 2012), conditions that should lead to a power struggle over the price. Considerable effort has been devoted to applying economic principles to analyzing nutrient exchange, stability, and other aspects of mutualisms (Weyl *et al.*, 2010; Werner *et al.*, 2014; Clark *et al.*, 2017), but there is a major gap: we do not understand how the exchange ratio and the quantity traded between plant and microbe are determined.

This question has been explored using mathematical models in which partners have disparate abilities to acquire resources and divergent resource requirements (Akçay & Roughgarden, 2007; Grman *et al.*, 2012; Franklin *et al.*, 2014). In these models, mutualistic partners negotiate based on the principles of the Nash bargaining solution, an axiomatically derived result describing the expected distribution of benefits after bargaining between self-interested partners that are able to regulate their participation in trade in response to the benefits they receive from trade (Nash, 1950; Binmore *et al.*, 1986; Akçay & Roughgarden, 2007).

A central assumption underlying the Nash bargaining solution is symmetry in bargaining power between the partners, where the bargaining power of a partner is defined as the weight given to the benefit received by that partner in the determination of trade dynamics. The best indicator of symbiotic benefit is reproductive fitness in the field, but biomass is conventionally used as a proxy for fitness (Younginger *et al.*, 2017). Consequently, the symmetric Nash product is the product of the partners' growth gains from trade:

$$(gP_{trade} - gP_{notrade})\,(gR_{trade} - gR_{noTrade}) \qquad (1)$$

where gP_{trade} and gR_{trade} are the plant and rhizobial partner growth rates with trade, and $gP_{noTrade}$ and $gR_{noTrade}$ are the respective growth rates without trade. An extension of this framework allows for unequal power between partners through the asymmetric Nash product:

$$(gP_{trade} - gP_{notrade})^{\beta} (gR_{trade} - gR_{noTrade})^{1-\beta} \quad (2)$$

which arises when bargaining power differs between partners (Binmore *et al.*, 1986). β is a scaling exponent that assigns different weights to the gains from trade by the plant and microbe. Increases in β correlate with increases in plant bargaining power relative to the microbial symbiont, and bargaining is symmetrical when β = 0.5 (Binmore *et al.*, 1986).

Previous modeling analyses of nutrient exchange symbioses have been limited by the absence of quantitative experimental studies in which all the major relevant parameters were measured in a single study across a range of environmental conditions and used explicitly within a mathematical framework (Clark *et al.*, 2017). To address this knowledge gap, we measured biomass distributions, nitrogen uptake and exchange rates, photosynthetic carbon assimilation fluxes, and carbon and nitrogen compositions in the model *Medicago truncatula-Ensifer medicae* (legume-rhizobia) symbiosis under conditions ranging from low to high nitrogen availability. We used these measurements to quantitatively test whether trade in a mutualism follows the predictions of the Nash bargaining solution under different conditions of resource availability. To do this, we refined and parameterized a mechanistic model of resource trade between a plant and microbe (Fig 1.1.1; Grman *et al.*, 2012). This model assumes that the growth of each partner is limited by its ability to obtain carbon and/or a mineral nutrient, and determines the exchange ratio to be the one that maximizes the product of partner benefits from trade, consistent with the Nash bargaining solution (Nash, 1950; Akçay & Roughgarden, 2007; Grman *et al.*, 2012). The exchange ratio is then used to predict per capita partner growth rates and allocation to growth versus trade. We tested the assumption of symmetric bargaining by comparing the accuracy of predictions made by the model with experimental measurements. The experimental results were inconsistent with predictions based on the

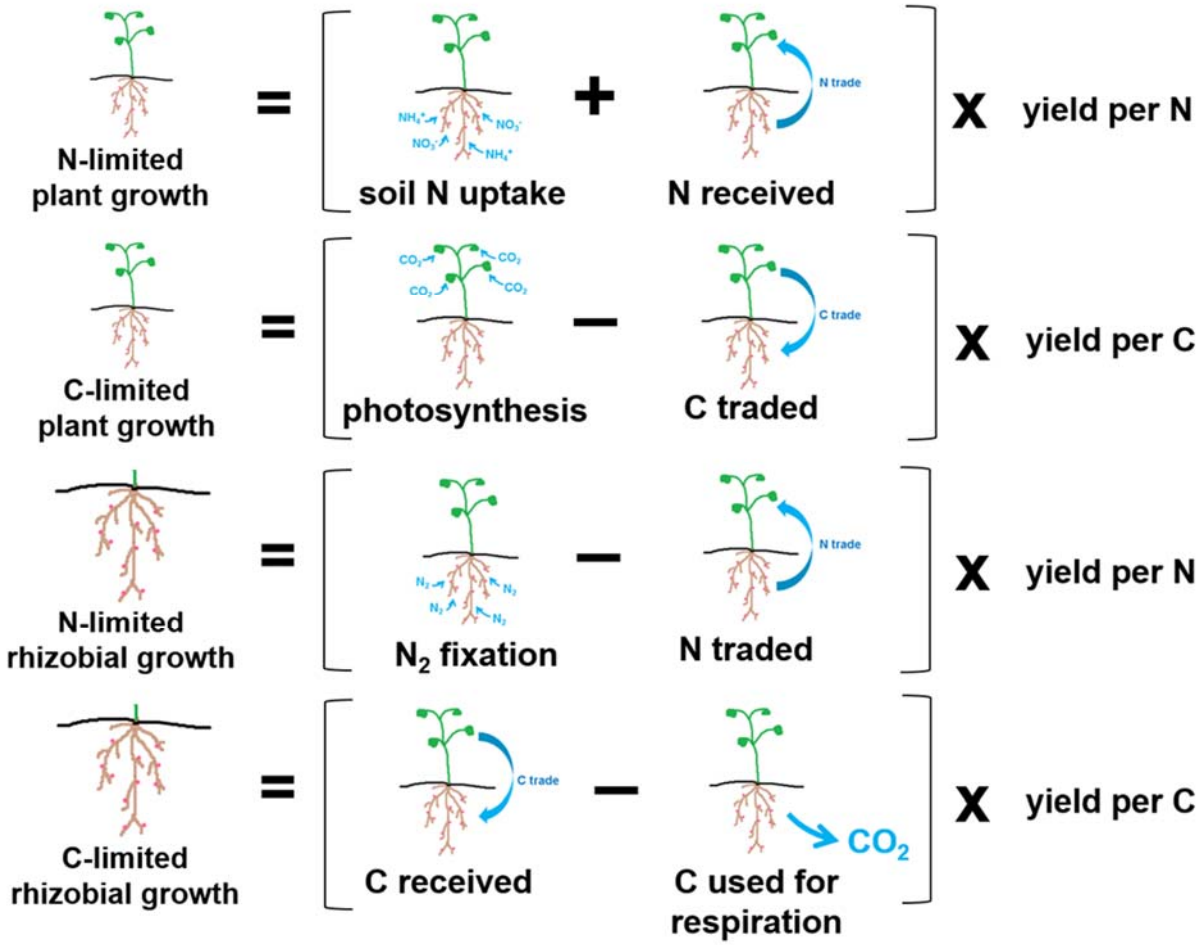

Figure 1.1.1. Pictorial representation of the legume-rhizobia nutrient exchange model. Plant and nodule growth are limited by the ability to obtain carbon (C) or nitrogen (N). In general, the growth rates are the rates of biomass carbon increase that complement via yield parameters the net amount of carbon or nitrogen obtained directly and/or from trade. Carbon and/or nitrogen can be lost during trade or respiration. More information about model construction is provided in the Methods.

symmetric Nash bargaining solution, so we explored how predictions based on asymmetries in relative bargaining power aligned with experimental measurements and determined the value of β for which model fit was greatest.

Materials and Methods

Experimental methods

SC10 Cone-Tainer pots (Steuwe and Sons Inc., Corvallis, OR, USA) were plugged with ⅜"diameter cotton wicks leading to opaque 50 mL reservoirs and filled with medium grain vermiculite. Pots were wetted with 25 mL deionized water, covered, and autoclaved for 45 minutes. After 24 hours, pots were wetted with a further 25 mL of deionized water and autoclaved again for 45 minutes. 24 hours later, the pots were wetted with 25 mL of Fahraeus nutrient solution (Fåhraeus, 1957) supplemented with 8, 24, 40, or 80 mg L^{-1} N in the form of NH_4NO_3.

Medicago truncatula A17 seeds (Young *et al.*, 2011) were scarified with 600 grit sandpaper, sterilized in commercial bleach (8.25% $NaHClPO_3$) for 3 minutes, and rinsed at least 6 times with sterile deionized water. Following 3 hours of incubation in sterile deionized water at room temperature, the seeds were re-sterilized in 0.825% $NaHClPO_3$ for 30 seconds and rinsed at least 6 times with sterile deionized water. Seeds were then incubated in sterile deionized water for 48 hours at 4°C. The water was replaced approximately every twelve hours during this incubation. The seedlings were then transferred to sterile petri dishes, sealed with Parafilm, and germinated at room temperature for 48 hours. Seedlings with 1 cm or longer radicles were aseptically transplanted into prepared pots. After planting, the plants were fully randomized and grown at 22°C with a 16 hour day/8 hour night cycle at approximately 250 µmol m^{-2} s^{-1}. Seedlings were misted daily with sterile deionized water for the first week. Throughout the growth period, plants were continuously supplied via their reservoirs with Fahraeus nutrient solution.

Ensifer medicae WSM419 (Reeve *et al.*, 2010) was grown for 48 hours in tryptone-yeast broth at 30°C with rotary shaking at 200 RPM. The OD600 of the culture was measured to estimate cell density. Half of the week old plants were inoculated with 1 mL of 10^6 CFU mL^{-1} inoculum in ½ x phosphate buffered saline (+ rhizobia treatment), while the other half were mock-inoculated with sterile buffer (- rhizobia treatment). At least 7 plants per nitrogen level and inoculation status were harvested after 4 or 6 weeks

of growth for biomass measurements (in total, 64 or 101 plants were harvested after 4 or 6 weeks, respectively, half of which were nodulated). Roots and shoots were separated, and roots were carefully washed in deionized water to remove vermiculite. Washed roots were checked for nodulation, and nodules were removed and counted. All tissue was dried at 60°C for at least one week. Of these plants, 3 plants per nitrogen level, inoculation status, and age were analyzed for carbon and nitrogen elemental compositions of roots, shoots, and nodules (48 plants in total). Dried plant tissues were ground using a NutriBullet, LLC household blender followed by a Retsch MM301 Mixer Mill. 2-5 mg dried tissue was weighed, packaged in tin capsules, and analyzed by the Robertson lab at Michigan State University's Kellogg Biological Station using a Costech ECS4010 elemental analyzer.

To measure direct nitrogen uptake with ^{15}N labeling, 24 plants (3 per nitrogen level and inoculation status) were grown as described above until 4.5 weeks, whereupon the pots and reservoirs were flushed with 500 mL N-free Fahraeus solution to remove soluble unlabeled nitrogen. The plants were then watered with 25 mL of Fahraeus solution containing the appropriate concentration of $^{15}NH_4{}^{15}NO_3$, supplied with this solution via their reservoirs for 1 week and harvested. Plants were harvested, weighed, dried, ground, and packaged into tin capsules as described above. The nitrogen content and ^{15}N abundance were then analyzed by the Stable Isotope Lab at Utah State University using a Europa Scientific SL-2020 system.

Photosynthesis rates were measured on 5-week-old nodulated and uninoculated plants (4-6 plants per nitrogen level and inoculation status; 41 plants total) using whole-plant (6"x6"x12") photosynthesis chambers connected to LI-COR LI-6400 apparatuses, illuminated with LED lights, and provided a constant airflow of 1000 µmol s^{-1} with 400 µmol CO_2 per mol of air (Fig 1.1.S1a,b). CO_2 from below ground was excluded using modeling clay. Assay conditions (e.g., light, humidity, temperature) matched the growth conditions. Steady-state photosynthetic rates were measured for at least 90 minutes after a pre-equilibration period in the chambers with the light on for at least 1 hour (Fig 1.1.S1c). The final

steady state CO_2 assimilation rate (µmol CO_2 sec^{-1}), was converted to a daily rate (mg C day^{-1}), assuming a constant rate throughout the 16h light period. After measuring photosynthesis rates, the plants were harvested, dried, and weighed as described above.

To test the effects of soil nitrogen and rhizobia on root and shoot biomass, we used a linear model ANOVA with Type II sum of squares (aov and car packages, R 3.3.1) with soil nitrogen, rhizobia, and the nitrogen by rhizobia interaction as fixed effects. Since we detected significant main effects of nitrogen and rhizobia, we conducted *post-hoc* testing with the Tukey test at a significance level of 0.05 (lsmeans package, R 3.3.1) to determine whether group means were significantly different.

Model construction

The legume-rhizobia model (Fig 1.1.1, Methods S1) was derived from the model of Grman *et al.* (2012) that assumes plants adjust carbon allocation to roots or shoots on a faster timescale than the carbon-for-nutrient exchange ratio is negotiated between the partners. Refinements to model structure equations were made using Wolfram Mathematica 11.3 (**Methods S1)**. Model predictions include plant and rhizobial specific growth rates (gP and gR, respectively) as functions of biomass carbon gain, the proportion of plant carbon allocated to roots (a_{NP}) or shoots, and the carbon-for-nitrogen exchange ratio (T). The model assumes organismal growth is limited by the ability to obtain nitrogen and/or carbon for biomass production. Consequently, for each partner, the predicted growth rate is the minimum growth predicted when nitrogen (gP_{Nlim}, gR_{Nlim}) or carbon (gP_{Clim}, gR_{Clim}) is limiting:

$$gP = \min(gP_{Clim}, gP_{Nlim})$$

$$gR = \min(gR_{Clim}, gR_{Nlim}).$$

When nitrogen is limiting, the organismal growth rates are represented as

$$gP_{Nlim} = (f_{np} + X/(p\ T))*y_{np}$$

$$gR_{Nlim} = (f_{nr} - X/(r\ T))*y_{nr},$$

where X is the rate of carbon traded from the plant to the rhizobia and, for the plant and rhizobia, respectively, p and r are the organismal carbon contents, f_{np} and f_{nr} are the rates of nitrogen uptake per organismal carbon content, and y_{np} and y_{nr} are the carbon biomass yields per unit nitrogen. Briefly in the vernacular, the growth rates are the rates of biomass carbon increase that complement (via the yield parameters) the amount of nitrogen obtained directly and/or from trade. Comparably, when carbon is limiting, the organismal growth rates are represented as

$$gP_{Clim} = f_{cp} - X/p$$

$$gR_{Clim} = X/r,$$

where f_{cp} is the rate of photosynthetic carbon uptake per organismal carbon content. Yield parameters are not necessary here because the growth rates are in units of biomass carbon gain.

Respiratory costs associated with nitrogen fixation were added to the model by reducing the rhizobial growth rate by the rate of nitrogen fixation (f_{nf}) multiplied by the biochemical stoichiometric trade constraints of 2.57 g C g^{-1} N (Phillips, 1980). Consequently, the rhizobial growth rate equations become

$$gR_{Nlim} = (f_{nr} - X/(r\,T))*y_{nr} - 2.57*f_{nr}$$

$$gR_{Clim} = X/r - 2.57*f_{nr}.$$

Plant respiratory costs are not as well-defined (Wardlaw, 1990) and thus were incorporated via the plant carbon uptake (f_{cp}) measurement as described below.

In addition to serving as model predictions, the growth rate equations provide the foundation for how the model predicts the rate of carbon trade, root-to-shoot allocation, and the carbon-for-nitrogen exchange ratio. Because rhizobia in nodules are unable to take up external carbon, we assumed the modeled trade is rhizobia-limited. Consequently, the rhizobial partner trades away all surplus nitrogen (i.e., nitrogen unnecessary for growth) in exchange for carbon from the plant. The traded nitrogen and that received by root uptake are used for plant growth and thus determine the plant surplus carbon (via

the yield parameter y_{np}) that is traded to the rhizobia (i.e., X is the plant surplus carbon). As described in Grman *et al.* (2012), the model predicts that the optimal root-to-shoot allocation is when, after trade, the total carbon uptake by the shoots complements (via the yield parameter) the total nitrogen uptake by the roots. Partner negotiations for the carbon-for-nitrogen exchange ratio are predicted to be consistent with economic modeling methods. When bargaining is symmetric, as in the Grman *et al.* (2012) model, the negotiated ratio is assumed to result in the Nash bargaining solution, in which the Nash product (Equation 1) is maximized. If bargaining is asymmetric, then negotiations should result in maximizing the asymmetric Nash product (Equation 2).

Computational methods

The model derivation (Grman et al. 2012) does not rely on the assumption of extended steady state because the solutions are instantaneous for any given set of input values for the plant (*Medicago truncatula*) and symbiont (*Ensifer medicae*). However, we first confirmed that over the period in which measurements were made (4-6 weeks), the plant per capita growth rate was near linear (Fig 1.1.1), thus its growth and nutrient uptake rates should change slowly compared to the negotiations. Measurements were either taken at 5 weeks of age or averaged between measurements at 4 and 6 weeks to represent 5-week-old systems. Goodness of fits for model predictions and parameters were assessed using 90% confidence intervals from the results of modeling 50 pseudo datasets per growth condition and age (Methods S2). The pseudo datasets were generated by Monte Carlo sampling of the experimental measurements; i.e., for each measurement, we generated random pseudo data points with normal distribution around the measured average with the measured standard deviation. To be consistent with biological reality, we assumed that biomasses and nutrient uptake rates could not be negative, so any randomly generated negative values were rounded up to zero. Each pseudo set contained one value for every measurement, and were used as model inputs to generate one set of model predictions. This

generated 50 sets of model predictions per nitrogen level, which were assessed using confidence intervals. This statistical method allowed us to evaluate how variations in measurements affected model predictions.

Because the model tracked how carbon was obtained and allocated, most model inputs (Table 1.1.S1) and predictions (Table 1.1.S2) were expressed per unit of biomass carbon content. Carbon and nitrogen contents were calculated using biomass and elemental composition measurements of carbon and nitrogen, respectively (Table 1.1.S3-S4). Both contents were used to calculate the carbon per nitrogen organismal yield parameters, and allocation to root biomass was calculated as the root carbon content divided by whole plant carbon content. Per capita growth rates were calculated as the carbon content gained between 4 and 6 weeks of age, divided by 5-week carbon content. For the plant, photosynthesis and soil nitrogen uptake rates were expressed per shoot carbon and root carbon, respectively, while rhizobial nitrogen fixation was per nodule carbon. Please note: when the abundance of nodule carbon is low (e.g., at 80 mg L^{-1} N), this requires dividing by a small number which amplifies the associated uncertainties.

Nitrogen elemental composition and ^{15}N enrichment measurements were used to differentiate between the rates of soil nitrogen uptake and nitrogen trade in nodulated plants. As described above, nodulated and uninoculated plants were labeled by flushing the soil with nitrogen-free media and then watering for 1 week with $^{15}NH_4{}^{15}NO_3$ (plants were labeled between 4.5 and 5.5 weeks of age). Labeling revealed that 96.7% of the variation in nitrogen uptake per total plant biomass was due to the abundance of nitrogen in the nutrient solution ($p < 0.001$; Fig 1.1.S2a). Consequently, we concluded that the average 5-week-old nodulated plant biomass can be used with the function shown in Fig 1.1.S2a to estimate the soil nitrogen uptake rate in nodulated plants and the remaining total nitrogen acquired can be attributed to trade. The total nitrogen acquisition rate was obtained by subtracting the average total nitrogen content of 4-week-old plants from that of 6-week-old plants and dividing by the time elapsed (i.e., 14

days). When analyzing the model with pseudo datasets, these rates were determined by generating pseudo data for uninoculated and nodulated plants using Monte Carlo sampling based on the experimental means and standard deviations. Because uninoculated plants can only obtain nitrogen from direct uptake and at 24, 40, and 80 mg L^{-1} N, there was no significant difference in ^{15}N uptake per total plant biomass between nodulated and uninoculated plants (Tukey *post-hoc* testing; $p = 0.940$, $p = 0.665$, and $p = 0.630$, respectively), pseudo 5-week-old uninoculated plant biomass data was used with the function shown in Fig 1.1.S2a to estimate the soil nitrogen uptake rate by nodulated plants at these nitrogen levels (each nodulated pseudo replicate had a corresponding uninoculated pseudo replicate, as shown in Methods S2). At 8 mg L^{-1} N, the average ^{15}N uptake rate was 2.4-fold greater in uninoculated plants than in nodulated plants. This average difference was incorporated into the model analysis by dividing the predicted nodulated soil nitrogen uptake rate (per pseudo replicate) by 2.4.

A minimum of 2.57 g C is biochemically required to produce the ATP and reductant necessary for rhizobia to fix 1 g N (Phillips, 1980). Using this biochemical minimum and the amount of nitrogen traded to the plant, we obtained a minimum for carbon to be traded for and consumed by rhizobial respiration. In addition to respiratory costs, the rhizobia used carbon for biomass production. We experimentally derived the quantity of biomass carbon using biomass and carbon composition measurements. Consequently, the predicted carbon-for-nitrogen exchange ratio was experimentally estimated as the amount of carbon traded to the rhizobia for biomass and respiration, divided by the amount of nitrogen traded to the plant. These estimates should be regarded as conservative (low) because they do not include any additional respiration costs for growth or maintenance; however, we found that moderate increases (e.g., 3.6 mg C mg^{-1} N, Ryle et al 1984) beyond the minimum had little effect on model predictions.

Measurements from uninoculated plants were used to test the effect of including plant respiration in the model because the presence of nodules prevented accurate measurement of below-ground (i.e., root) respiration in nodulated plants. It has been shown that when relative plant growth is

constant (as in the near linear legume-rhizobia system), the rate of plant respiration per unit of biomass is also constant (Lambers *et al.*, 1983). Furthermore, the rate of whole-plant respiration has been found to be a linear function of plant biomass and rate of gross photosynthesis. Therefore, we assumed that the rate of respiration per unit plant biomass was the same for plants with and without rhizobia. We found that in uninoculated plants, there was a significant correlation ($R^2 = 0.346$, $p < 0.001$) between biomass carbon content and the proportion of photosynthetic carbon used for biomass (Fig 1.1.S2b). We used this relationship to estimate the amount of carbon used for plant respiration in nodulated plants. Plant respiratory costs were included in the model by reducing the measured photosynthesis rate by the estimated respiration rate.

Overall model fits were quantified using the sum of squared differences between predicted and measured allocation to root biomass, plant growth, and nodule growth. The carbon-for-nitrogen exchange ratio predictions were not used in this assessment because the ratio was experimentally estimated (as described above) but not directly measured. The best fit value of β was identified as the one yielding the smallest normalized sum of squares. The sum of squares for each β in a pseudo dataset was normalized by dividing by the average sum of squares for that pseudo dataset. This allowed fit comparisons to be between the β values without bias from differences between pseudo datasets.

Results

Changes in soil nitrogen alter carbon and nitrogen uptake and the benefits of the legume-rhizobia mutualism

We determined the values of model input parameters (Table 1.1.S1) for the *Medicago truncatula-Ensifer medicae* (legume-rhizobia) symbiosis by measuring biomasses, nutrient uptake rates, and elemental compositions, and of the resulting model predictions, including growth rates, root-to-shoot allocation, and the carbon-for-nitrogen exchange ratio (Table 1.1.S2). The collective nodules of a single plant were

used as a proxy for the rhizobial partner because, although nodules contain plant cells as well as rhizobia, the plant's investment toward trade is the whole nodule. In addition, nodule biomass correlates well with rhizobial abundance (Ratcliff *et al.*, 2011), thus larger nodules indicate greater plant investment as well as greater rhizobial benefits from trade.

Plants were grown in the presence or absence of rhizobia and fertilized with 8, 24, 40, or 80 mg L^{-1} N (Fig 1.1.2a; Valladares *et al.*, 2002). Total plant biomass was significantly increased by soil nitrogen ($p < 0.001$), and inoculation with rhizobia significantly increased total plant biomass at 8, 24, and 40 mg L^{-1} N, but not at 80 mg L^{-1} N (Fig 1.1.2b), showing that rhizobia enhanced plant growth at low and intermediate nitrogen levels. Nodule biomass significantly decreased with increasing soil nitrogen ($p < 0.001$, Fig 1.1.2c). Together, these findings confirm that mutualistic relationships are more beneficial for both partners at lower nitrogen availabilities and that nitrogen is a key limiting resource in these experiments (Regus, J *et al.*, 2017).

Plants interacting with rhizobia have two options for acquiring nitrogen: taking it up directly from the soil or trading for it with rhizobia. These two nitrogen acquisition routes are rarely differentiated experimentally, but this differentiation is essential to quantifying symbiotic nutrient exchange. To do so, we measured direct nitrogen uptake using ^{15}N enriched nutrient solution and compared it to the total increase in plant nitrogen content. Our findings revealed that fixed nitrogen from trade adds to but does not replace nitrogen obtained by direct uptake. On a per-plant basis, plants with rhizobia obtained at least as much nitrogen by direct uptake as those without rhizobia at all nitrogen levels (Fig 1.1.3a). Furthermore, in comparison to uninoculated plants, the rate of direct nitrogen uptake per root carbon was slightly lower in nodulated plants at the lowest soil nitrogen, modestly greater at 24 and 40 mg L^{-1} N, and the same at 80 mg L^{-1} N (Fig 1.1.4b).

In addition to increasing plant growth rates and supplementing nitrogen acquisition, the presence of rhizobia significantly altered the pattern of carbon allocation within the plant (Voisin *et al.*, 2002; Goh

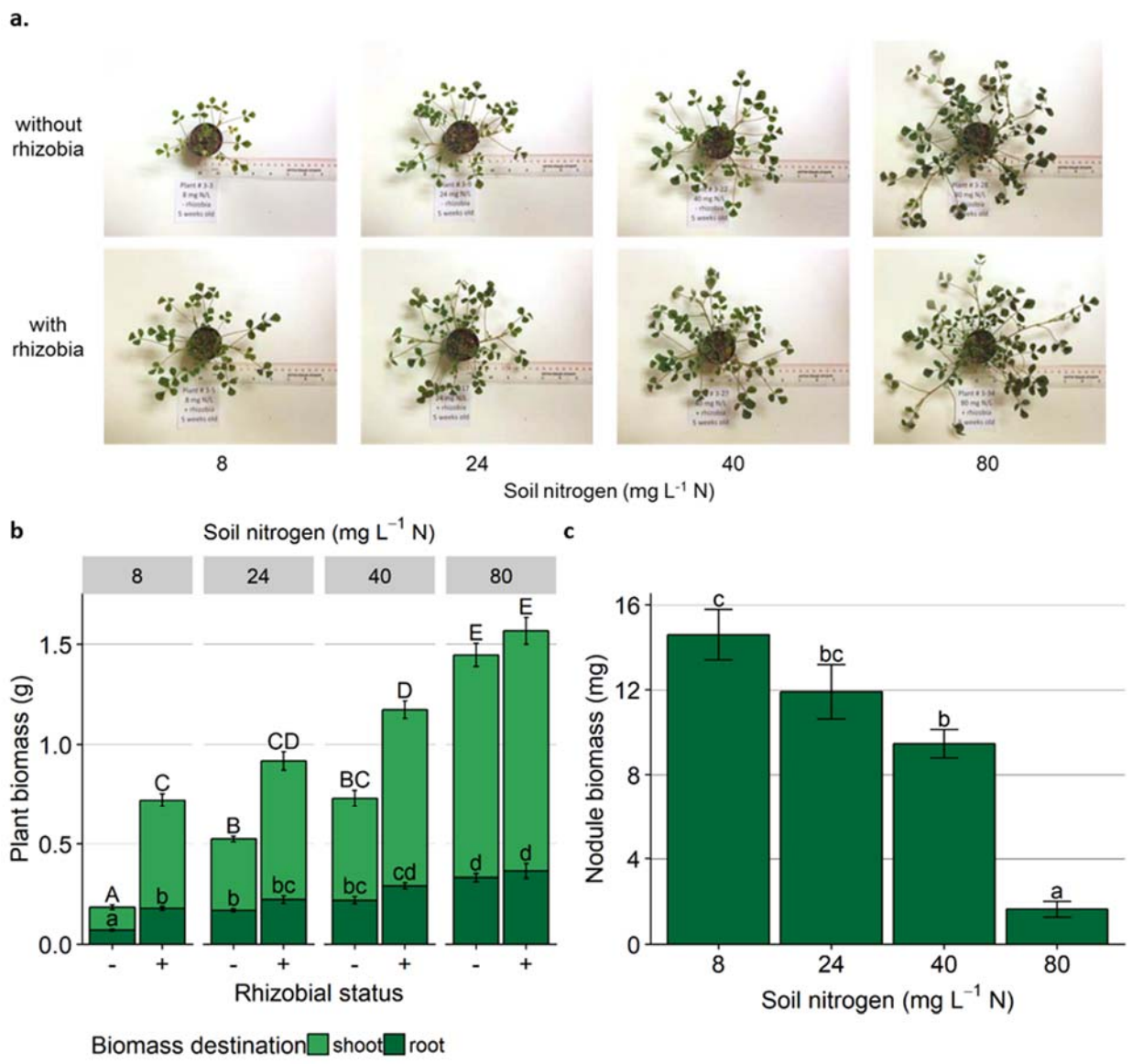

Figure 1.1.2. Both partners benefit more from trade at lower nitrogen availabilities. (a) Images of 5-week-old nodulated and uninoculated *M. truncatula* at 8, 24, 40, and 80 mg L^{-1} N. (b,c) Average 5-week root (b, dark green), shoot (b, light green), and nodule (c) biomasses. Error bars indicate standard error (n = 7-15), and bars with the same letter within the same panel do not differ significantly after *post hoc* testing with the Tukey test ($p < 0.05$). (b) Capital letters refer to shoot biomass and lowercase letters refer to root biomass.

et al., 2016). Increasing soil nitrogen significantly decreased root:shoot ratio in uncolonized plants (Fig 1.1.4a; Rufty Jr *et al.*, 1984; Paponov *et al.*, 2000), while the root:shoot ratio of nodulated plants did not vary with soil nitrogen (Fig 1.1.4a). This shows that plants adjust their relative allocation of biomass

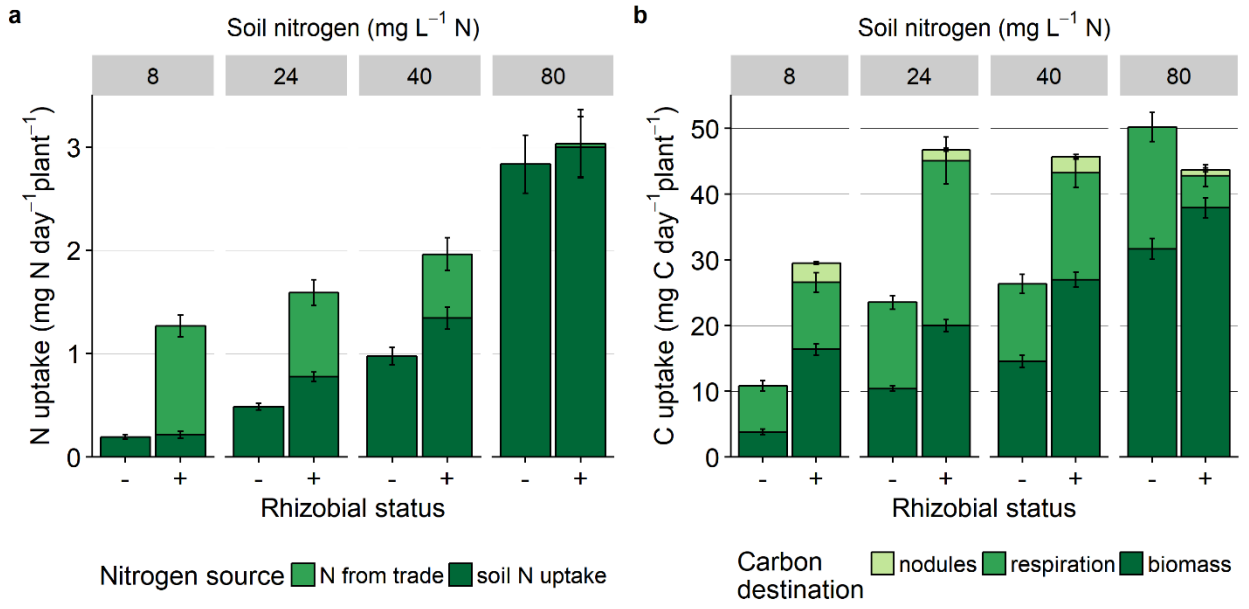

Figure 1.1.3. The effect of rhizobial inoculation on plant carbon and nitrogen budgets as functions of nitrogen availability. Whole-plant nitrogen (a) and carbon (b) uptake and allocation rates by nodulated (+) and uninoculated (-) *M. truncatula* at 8, 24, 40, and 80 mg L^{-1} N between 4 and 6 weeks of age (see Methods). (a) Direct nitrogen uptake by roots (dark green) and nitrogen received from trade with nodules (medium green). (b) Photosynthetic carbon allocated to plant biomass (dark green), plant respiration (medium green), and nodule growth or respiration (light green). Values and error bars represent the average and 90% confidence intervals, respectively, of 50 pseudo datasets generated by Monte Carlo sampling.

between roots and shoots depending on their ability to take up nitrogen directly from the soil or trade for it with rhizobia. However, root:shoot ratios are an incomplete measure of carbon allocation because they do not account for allocation to mutualists, respiration, exudates, and other carbon sinks (Wardlaw, 1990). To determine how organ allocation relates to the total carbon budget, we measured total plant carbon content and whole-plant photosynthetic carbon uptake. As expected from their larger shoot biomasses, we measured higher rates of carbon uptake per plant in nodulated plants grown at 8, 24, and 40 mg L^{-1} N than in uninoculated plants (Fig 1.1.3b, Fig 1.1.4c). Plant respiration ranged from 11-65% of total photosynthetic carbon, and carbon allocation to nodules ranged from 2-10% (Fig 1.1.3b). The magnitude of respiratory highlights the importance of accounting for respiration in models of plant carbon budgets (Table 1.1.S2); the strong dependence of respiration rates on nitrogen levels points to the value

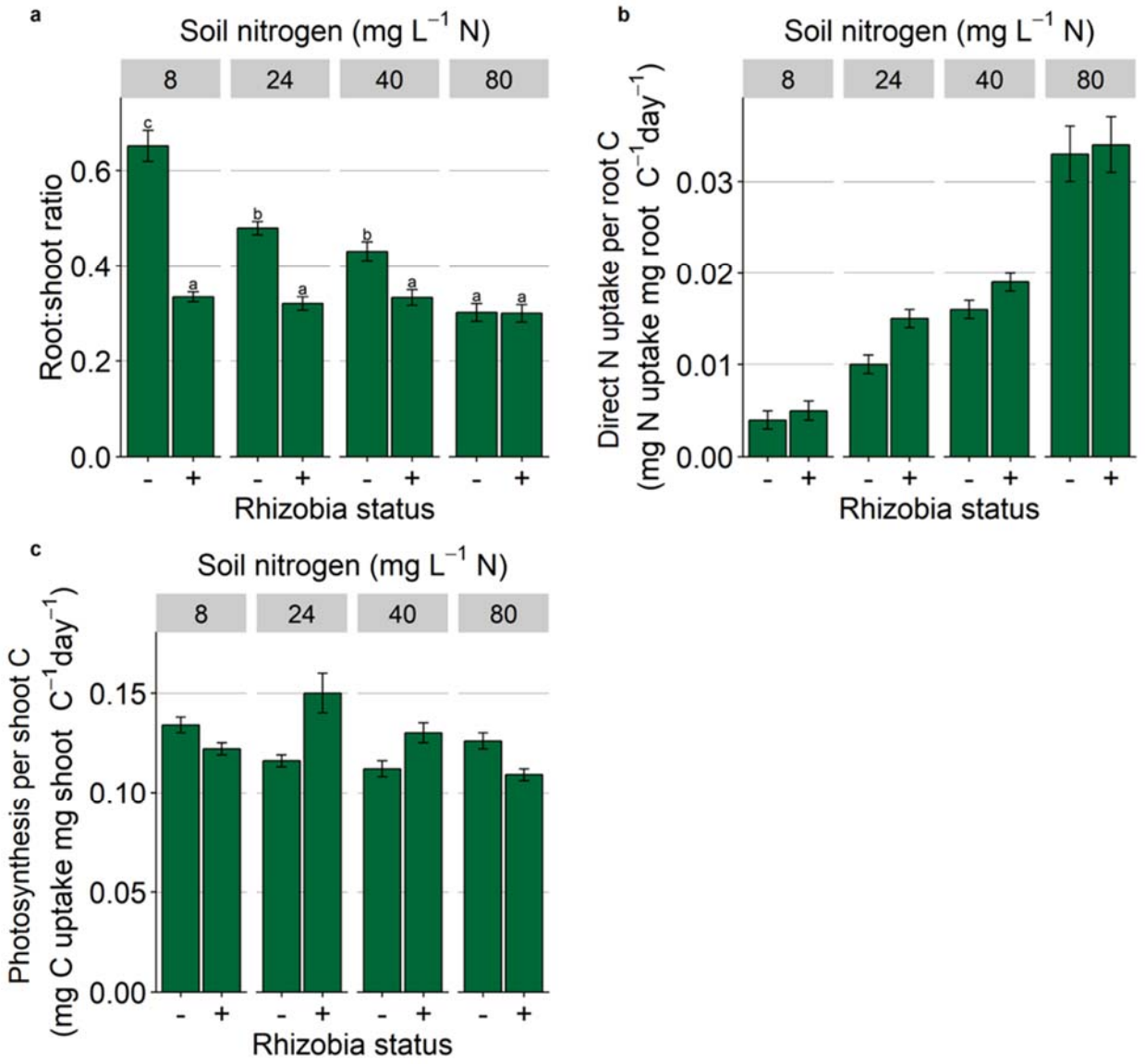

Figure 1.1.4. Plant carbon allocation to nutrient uptake. Root-to-shoot ratio (a), direct nitrogen uptake per root carbon (b), and photosynthesis per shoot carbon (c) of 5-week-old nodulated (+) and uninoculated (-) plants at 8, 24, 40, and 80 mg L^{-1} N. a, Values and error bars represent the average and standard error, respectively, of 7-15 biological replicates. Bars with the same letter do not differ significantly after *post hoc* testing with the Tukey test ($p > 0.05$). b, c, Values and error bars represent the average and 90% confidence intervals, respectively, of 50 pseudo datasets generated by Monte Carlo sampling.

of measuring them directly.

The benefits and exchange ratio of legume-rhizobia trade are better explained by asymmetric bargaining

We used the measurements described above to predict the carbon-for-nitrogen exchange ratio consistent with the Nash bargaining solution in the legume-rhizobium model. Using the biochemical stoichiometric trade constraints for nitrogen fixation of at least 2.57 mg C mg^{-1} N (Phillips, 1980) and measured volumes of nitrogen traded and nodule biomass carbon contents (Fig 1.1.2-3), we estimate that the nodules received 2.72-3.25 mg C mg^{-1} N for growth and respiration (Fig 1.1.5a). The model-predicted exchange ratios approached these estimates at the highest nitrogen level, but at the lowest, the predicted exchange ratio was 2.5-fold higher than experimentally estimated (Fig 1.1.5a). These estimates are consistent with measurements on the soybean-rhizobia mutualism, where nodules received 3.6 mg C mg^{-1} N (Ryle *et al.*, 1984). In addition to comparing experimentally estimated and model-predicted exchange ratios, we also examined how the exchange ratio influences partner benefits from trade, and thus how well the model predicts partner growth rates and the proportion of plant carbon allocated to roots or shoots. Using the predicted exchange ratio, the symmetric bargaining model predicted plant growth and carbon allocation to roots within 50% of measured (Fig 1.1.5b,c), but predicted nodule growth to be 3 to 7-fold higher than measured rates (Fig 1.1.5d). These discrepancies could not be resolved by modifications in respiratory cost interpretations, growth rate definitions (specific versus organismal), assuming that only 50% of the nodule biomass represents the rhizobial partner (Table 1.1.S2), or assumptions concerning the timescale of adjusting root-to-shoot allocation (Grman *et al.*, 2012), and thus suggest that this legume-rhizobia mutualism fails to meet fundamental conditions of the symmetric Nash bargaining solution (Nash, 1950).

Consequently, we investigated asymmetric bargaining using asymmetric Nash products in our model (Equation 2) and found that it could better explain the observed growth rates, nutrient allocations and trade across the range of growth conditions. We determined how values of β between 0 and 1 affected model fit for the carbon-for-nitrogen exchange ratio and other model predictions. We found that across the soil nitrogen levels, the best fit value of β was 0.70 (Fig 1.1.6). Using this estimate, model

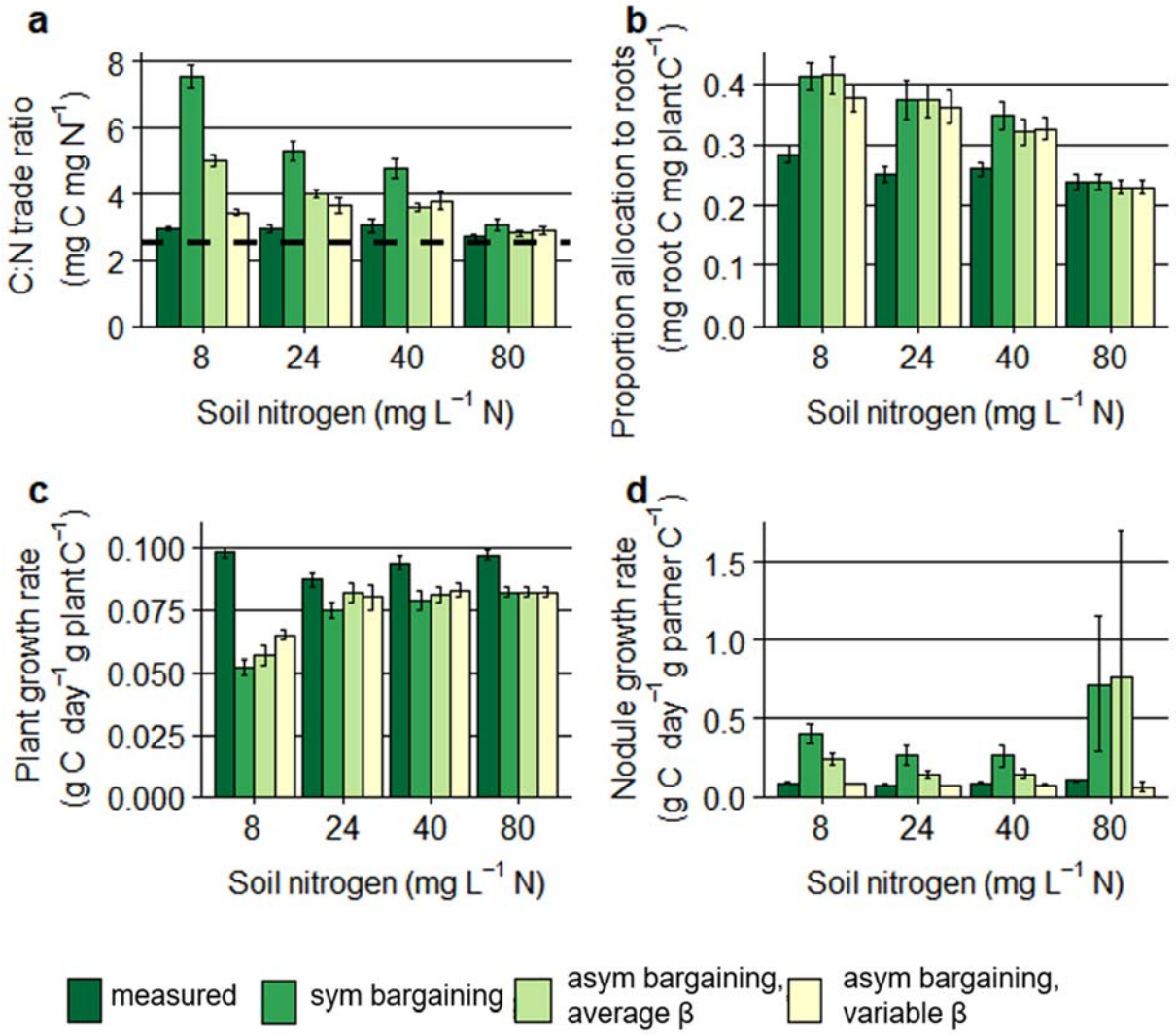

Figure 1.1.5. Model fit to experimentally estimated data is improved by allowing asymmetric bargaining power. Empirically estimated (measured) and model predicted carbon-for-nitrogen exchange ratios (a), percent allocations toward roots (b), plant growth rates (c), and nodule growth rates (d) of 5-week-old nodulated *M. truncatula* at 8, 24, 40, and 80 mg L^{-1} N. The symmetric (sym) bargaining predictions correspond to those predicted using symmetric Nash products, while the asymmetric (asym) bargaining predictions correspond to using asymmetric Nash products with either β fitted across the soil nitrogen levels (average β) or β fitted to individual soil nitrogen levels (variable β). The dotted line in (a) represents the biochemical minimum value of the exchange ratio. Values and error bars represent the average and 90% confidence intervals, respectively, of 50 pseudo datasets generated by Monte Carlo sampling.

predictions of nodule growth at 8, 24, and 40 mg L^{-1} N were improved up to 2-fold, but there was little improvement in plant growth, root:shoot allocation, or 80 mg L^{-1} N predictions (Fig 1.1.5). Next, we let β vary among the nitrogen levels and found that the best fit value of β increased from 0.57 to 0.86 as soil

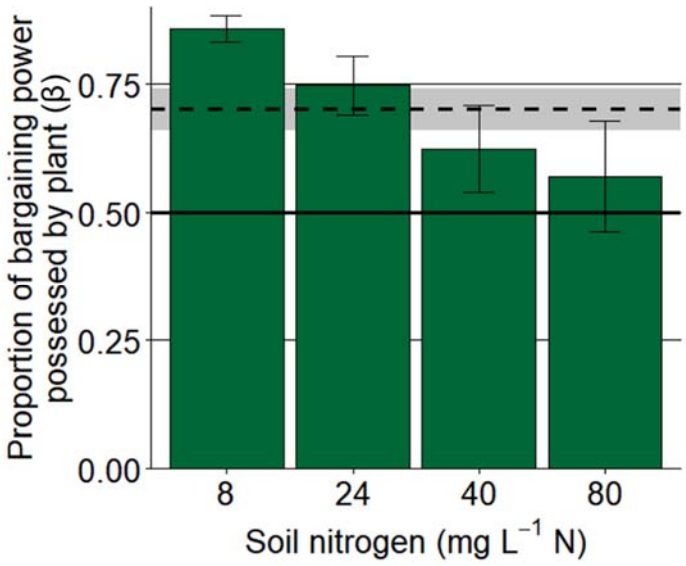

Figure 1.1.6. Plant bargaining power as a function of nitrogen availability. Best fit average β (dotted line; shaded area indicates the 90% confidence interval) and the best fit variable β at 8, 24, 40, and 80 mg L^{-1} N (bars; error bars indicate 90% confidence intervals of 50 pseudo datasets). Best fit β were those associated with the smallest sum of squared differences between predicted (using asymmetric Nash products, Equation 2) and measured allocation to root biomass, plant growth, and nodule growth in 5-week-old nodulated *M. truncatula*. The solid line represents symmetric bargaining.

nitrogen decreased (Fig 1.1.6, Table 1.1.S2). The 90% confidence intervals of β for 8 and 80 mg L^{-1} N were inconsistent with the value of 0.70 obtained assuming a constant value of β across nitrogen levels. Importantly, β for 8, 24, and 40 mg L^{-1} N indicate asymmetric trade (i.e., $\beta \neq 0.50$), and model fit was dramatically improved compared to symmetric predictions (Fig 1.1.5). For example, at 8 mg L^{-1} N, nodule growth was predicted within 3% of measured values (Fig 1.1.5d), and overall agreement between predicted and measured growth rate and root:shoot allocation predictions was increased 7-fold (Table 1.1.S2**)**. For all soil nitrogen levels, plant growth and root allocation predictions were improved, and the exchange ratio was predicted within 25% of experimental estimates (Fig 1.1.5b,c). Together, these findings indicate that plant bargaining power rises as its investment in trade and the benefits to both partners increase.

Discussion

Potential mechanisms underlying asymmetric bargaining

Our measurements of nutrient uptake, per capita growth, and partner composition in a legume-rhizobia system were more consistent with the plant having greater bargaining power than its symbiotic partner, than with a model of fair trade. Asymmetries in bargaining power can arise by three mechanisms that may be operating in the legume-rhizobia mutualism. First, one partner can have a lower effective "discount rate" (the rate at which benefits lose value to that partner if negotiations are prolonged), thus conferring greater bargaining power to the partner able to endure longer negotiations (Kawamori, 2014). In the legume-rhizobium mutualism, the plant's longer lifespan and potentially larger nutrient reserves could serve to provide a lower discount rate. Second, group bargaining dynamics can affect bargaining power. In "pure bargaining" situations, groups of individuals negotiating as a single partner have less apparent bargaining power than independent partners because group benefits would be divided among group members after negotiation (Chae & Heidhues, 2001). In this case, if the bacteroids that comprise a nodule (Udvardi & Poole, 2013) negotiate as a group (e.g., at the nodule level), then the rhizobial partner would have less apparent bargaining power because the carbon received by the nodule would be divided among the bacteroids. A third mechanism that can increase an individual's bargaining power is the ability to simultaneously negotiate with multiple trade partners (Chakraborty *et al.*, 2009; Chakraborty, 2011). In this case, the presence of multiple nodules on a plant and/or multiple bacteroids within a nodule could lead to bargaining power asymmetry. This mechanism could be particularly important if bacteroids within nodules bargain independently with the plant, which is consistent with the plant being able to interact differently with different bacteroids within a nodule (Daubech *et al.*, 2017; Regus, JU *et al.*, 2017). Each of these mechanisms can explain why the plant has more bargaining power than the rhizobia, but alone, they do not explain why plant bargaining power appears to be higher when nitrogen is scarce (e.g., 8 mg L^{-1} N). At lower soil nitrogen levels, the plant would be expected to have fewer nitrogen reserves and

nitrogen received from trade comprises a larger proportion of the plant's nitrogen budget (Fig 1.1.3a), which could increase its discount rate and lower its bargaining power. However, we found that as soil nitrogen decreases, total nodule biomass (Fig 1.1.2) increases. This apparent increase in the number of bacteroids (i.e., trade partners) could lead to increased plant bargaining power even as the plant's reliance on the rhizobia increases. Studies in other systems would be important for assessing the prevalence of asymmetric bargaining, and manipulating microbial numbers would allow the influence of partner number to be further investigated.

Relating trade conflict to the concept of cheating

Our finding that the plant host has a high degree of control over the carbon-for-nitrogen exchange ratio when soil nitrogen is limiting has important implications for conceptualizing conflict and cheating within mutualisms. Cheaters have been the focus of much empirical and theoretical work as they have the potential to lead to mutualism collapse, yet there has been debate in the literature regarding how to identify cheaters (Simms & Taylor, 2002; Frederickson, 2013; Ghoul *et al.*, 2014). A recent synthesis defines cheating as increasing one's own relative fitness while decreasing that of the partner (Jones *et al.*, 2015). However, cheating is not synonymous with conflict, and in fact Jones et al. (2015) show that under both fitness conflict and fitness alignment, cheating genotypes may be present within the population. Within the economic model that we use in our study, there is a fundamental conflict between plant and symbiont over the exchange ratio (Schwartz & Hoeksema, 1998; Grman *et al.*, 2012). Our finding of asymmetric bargaining can be interpreted as evidence of this conflict because it demonstrates that partners may not benefit equally from trade, even if trade is mutually beneficial. However, given that the exchange ratio seems to be determined almost entirely by the plant, this control could in effect force the fitness interests of the symbiont to align with the host. Consequently, we predict that with multiple genotypes of varying fixation abilities there would be a strong signal of fitness alignment—even in the

face of underlying conflict.

Limitations and future work

The nutrient-exchange model employed connects nutrient uptake rates to partner growth by predicting the allocation strategies that maximize growth, given the organismal stoichiometric compositions. More sophisticated models that include detailed chemical reactions can predict the metabolic processes involved (Resendis-Antonio *et al.*, 2011; Zhao *et al.*, 2012), but they also rely on observed stoichiometries and maximizing growth rates. Consequently, one limitation of our approach is that nutrient stoichiometries and rates are treated as constants, which is not necessarily consistent with biological systems (Näsholm *et al.*, 2009; Wolf *et al.*, 2017). We addressed this challenge by independently measuring model parameters and predictions at all four nitrogen levels. Another limitation is that stoichiometric models do not account for other benefits of mutualisms, such as rhizobia acting as biocontrol agents to protect the plant from fungal pathogens (Das *et al.*, 2017) or association with rhizobia leading to the induction of defense signaling pathways (Dean *et al.*, 2014). We sought to minimize the influence of these other benefits by protecting the plants from pathogens, and keeping them well-watered and under stable conditions.

To our knowledge, the symmetry of bargaining power has not yet been directly tested in other plant-microbe mutualisms, but researchers have proposed several mechanisms that influence the power dynamics, such as symbionts conferring greater competitive power to more cooperative plants (Bücking *et al.*, 2016), partner strategies for preventing imbalances in benefits received from trade (Kiers *et al.*, 2011), and outside negotiator options leading to joint control of the mutualism (Akçay & Simms, 2011). Further work in other plant-microbe symbioses is needed before it can be determined if asymmetric bargaining power is a broad feature of nutritional symbioses and how it relates to these mechanisms.

Conclusion

Our finding of experimental support for a model of asymmetric ("unfair") bargaining power between legumes and rhizobia versus a model of "fair" trade highlights the power of integrating quantitative models with data in the study of mutualisms (Clark *et al.*, 2017), which we believe will be broadly applicable to other systems. This work improves our understanding of the drivers of quantitative variation in symbiotic nitrogen fixation, a process that makes a major contribution to the global nitrogen cycle (Fowler *et al.*, 2013) and is critical for agricultural sustainability (Herridge *et al.*, 2008). Our finding of a reduction in plant bargaining power at higher nitrogen levels combined with the frequently high rates of fertilizer application to legume crops underscores the potential evolutionary danger of relaxed host control under anthropogenic inputs (Kiers *et al.*, 2007; Weese *et al.*, 2015). The ability to estimate bargaining power using this approach will allow bargaining strength to be compared with measurements of natural selection. We also believe these estimates contribute to investigating the genetic, metabolic, and physiological mechanisms underlying the regulation of resource exchange and could facilitate future efforts to breed crops that can maintain their own beneficial microbiomes (Busby *et al.*, 2017).

Acknowledgements

We wish to thank Sean Weise, Sarathi Wijetilleke, Tom Sharkey, Robert Zegarac, and Dave Kramer for assistance in developing the photosynthesis chambers and protocols. We would also like to thank Chris Klausmeier for contributing to the discussion that led to this work. We acknowledge funding support from NSF DEB 1354878 and NSF IOS 1342793 to MLF. This material was also supported by the National Science Foundation under Cooperative Agreement No. DBI-0939454 and partial support from a fellowship to CAF from Michigan State University under the National Institutes of Health Training Program in Plant Biotechnology for Health and Sustainability (T32- GM110523). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the

views of the National Science Foundation.

Author Contributions

All authors conceived the project, interpreted the results, and contributed to the writing. TJC and CAF carried out the experiments and, respectively, led the modeling and statistical analyses.

Supporting Information

Supplemental File 1 – Section 1.1 Methods S1

Supplemental File 2 – Section 1.1 Methods S2

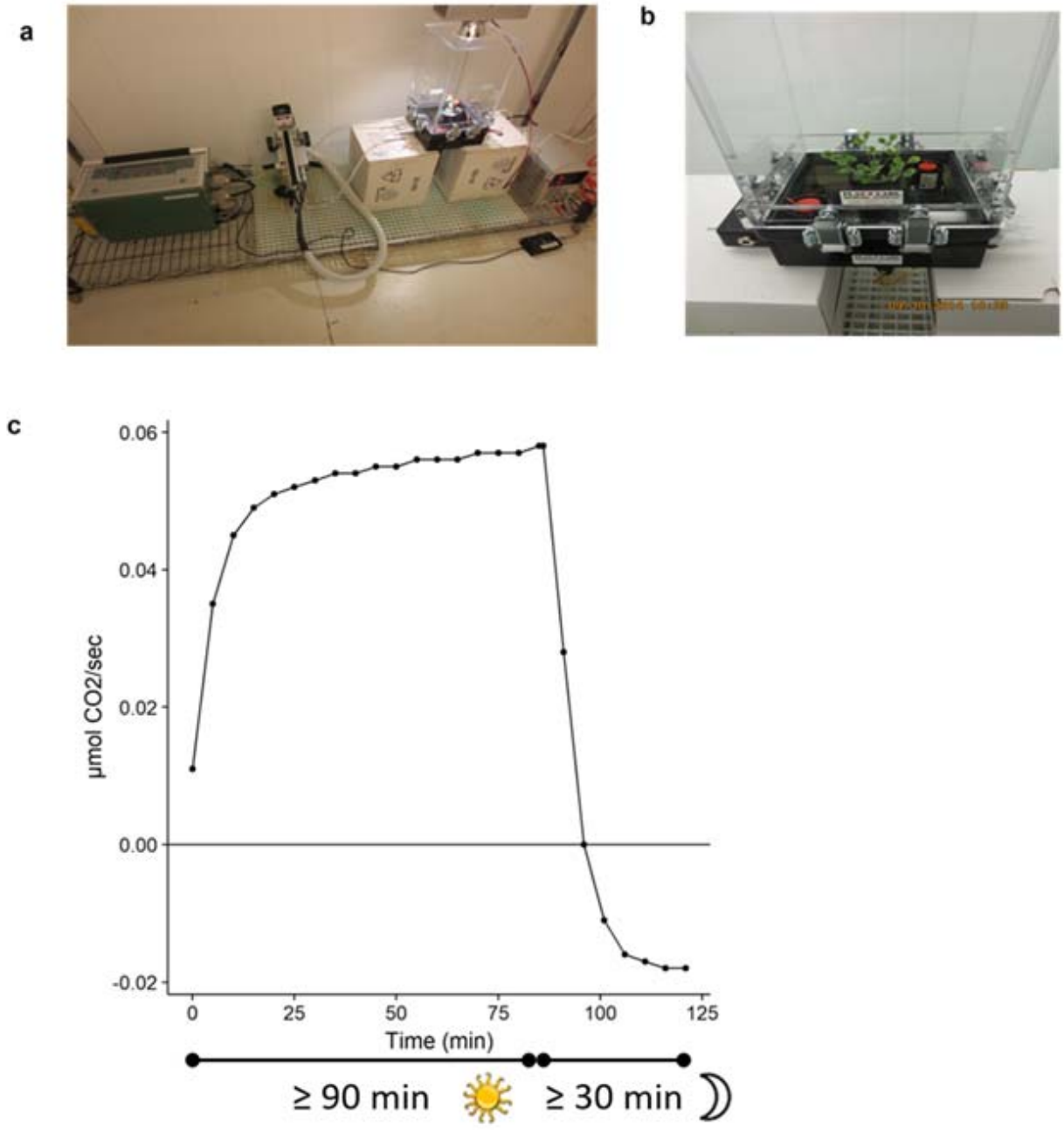

Figure 1.1.S1. Whole-plant photosynthesis chambers were used to measure carbon uptake rates. (a) Chambers were connected to LI-COR LI-6400 apparatuses, illuminated with LED lights, and used in a growth chamber to ensure assay conditions matched growth conditions. (b) Chambers included fans to promote air circulation. (c) Steady-state photosynthesis was measured by collecting readings every 5 min for at least 90 min in the light. Respiration was estimated with dark measurements. Data shown is from a nodulated plant grown at 40 mg L^{-1} N, but is representative of all measured plants.

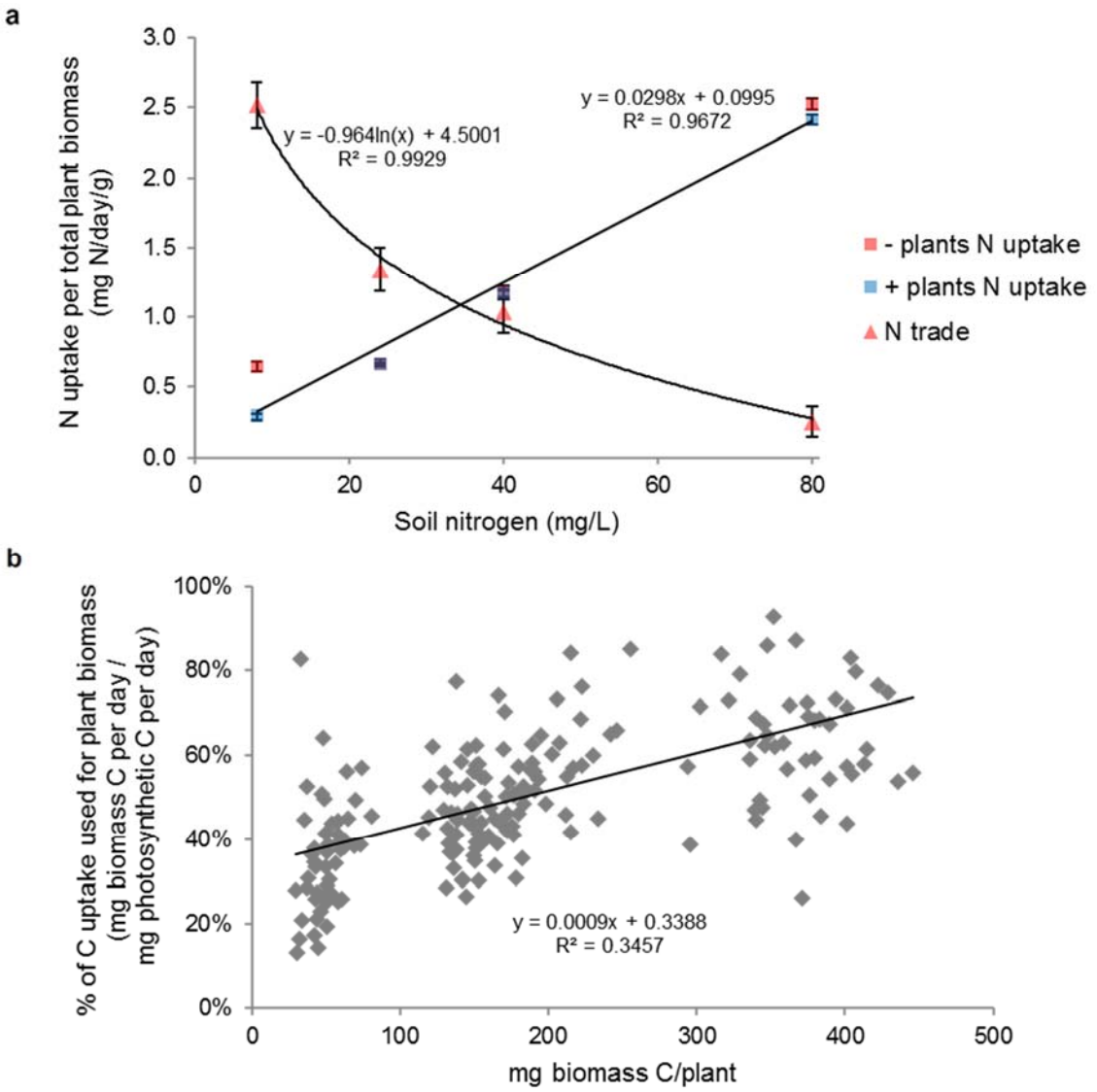

Figure 1.1.S2. Regression analyses used to calculate model parameters. (a) Direct nitrogen uptake (squares) and trade (triangles) rates per total plant biomass were experimentally estimated for uninoculated (blue) and/or nodulated (red) plants at 8, 24, 40, and 80 mg L^{-1} N. Linear or exponential regression was used to analyze the correlation between direct uptake or trade, respectively, and nitrogen availability. Values and error bars represent average and 90% confidence intervals, respectively, of 50 pseudo datasets generated by Monte Carlo sampling. (b) The percent of photosynthetic carbon uptake used for biomass is directly correlated with total carbon content in uninoculated plants. This correlation was analyzed with linear regression and used to estimate respiration rates in nodulated plants. Values represent 50 pseudo datasets per nitrogen treatment. (a,b) The linear equations were used in model analysis as described in the Methods

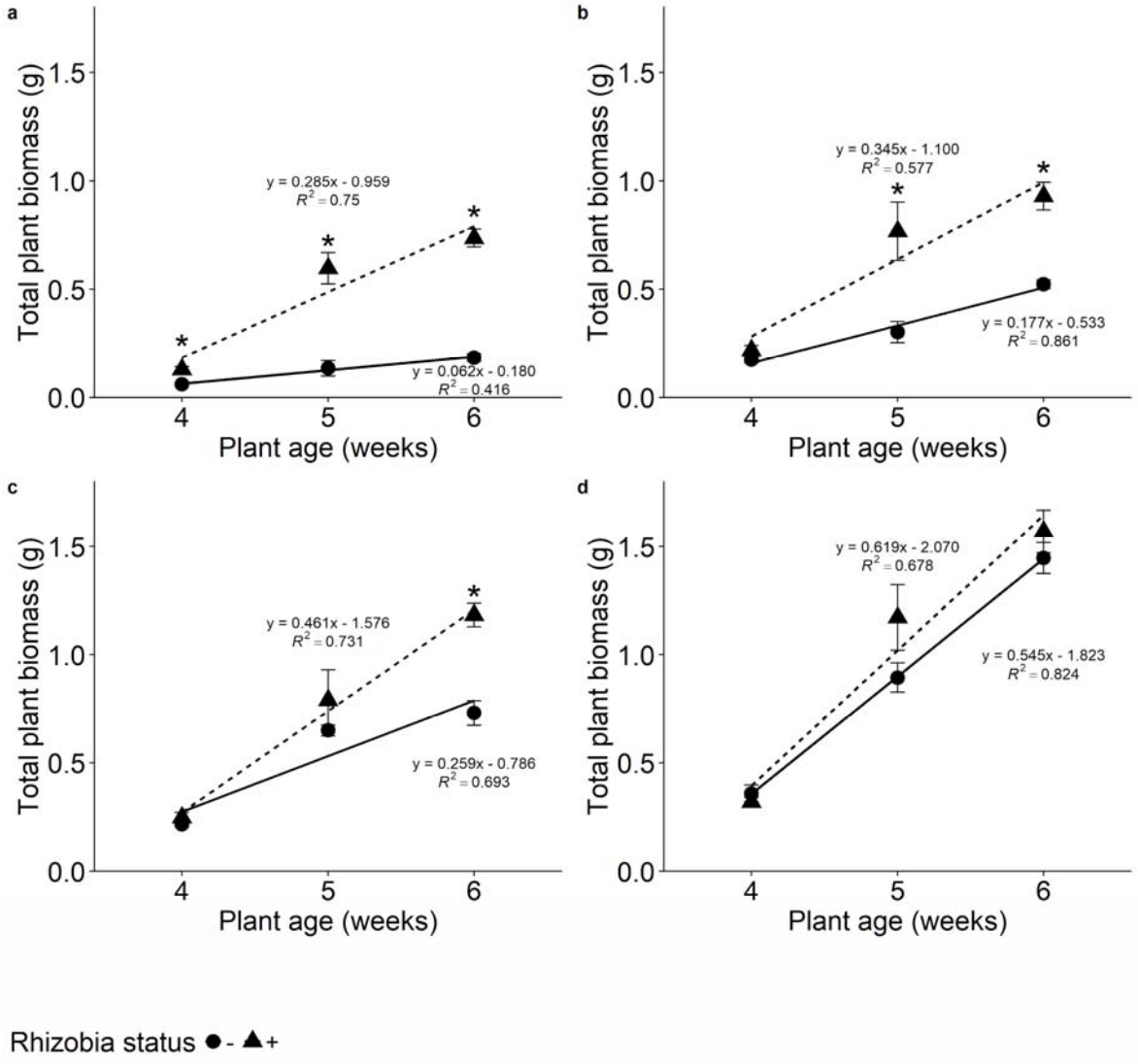

Figure 1.1.S3. Plant growth is near linear between 4 and 6 weeks. Nodulated (triangles, dashed lines) and uninoculated (circles, solid lines) plant dry weights were measured at 8 (a), 24 (b), 40 (c), and 80 (d) mg L^{-1} N. Growth rates were analyzed using linear regression. All rates were found to have $p < 0.001$, and equations and R^2 values as indicated. Error bars represent standard error (n = 7-15). Plant ages with asterisks indicate significant differences between nodulated and uninoculated plants at that age; plant ages without asterisks do not significantly differ after *post hoc* testing with the Tukey test ($p < 0.05$).

Table 1.1.S1. Model input parameters and measured values for nodulated plants

Parameter	Biological interpretation	Units	8 mg L^{-1} N	24 mg L^{-1} N	40 mg L^{-1} N	80 mg L^{-1} N
p	plant size	mg C	172.65 ± 1.94	229.82 ± 2.43	288.41 ± 2.39	392.34 ± 3.46
r	rhizobial partner size	mg C	3.92 ± 0.08	3.31 ± 0.07	2.64 ± 0.04	0.37 ± 0.01
ynp	plant carbon biomass yield per unit N	mg C/mg N	11.80 ± 0.07	11.47 ± 0.04	11.30 ± 0.06	10.35 ± 0.02
ynr	nodule carbon biomass yield per unit N	mg C/mg N	6.08 ± 0.01	6.21 ± 0.01	6.03 ± 0.01	6.18 ± 0.02
fcp'	photosynthetic carbon uptake rate	mg C/shoot C /day	0.13 ± 0.001	0.16 ± 0.003	0.14 ± 0.002	0.12 ± 0.001
fnp'	soil nitrogen uptake rate	mg N/root C /day	0.01 ± 0.000	0.02 ± 0.000	0.02 ± 0.000	0.04 ± 0.001
fnr'	nitrogen fixation rate	mg N/nod C/day	0.31 ± 0.01	0.26 ± 0.01	0.34 ± 0.02	2.27 ± 1.14

Parameters were measured and used in model construction as described in the Methods. The plant's collective nodules were used as a proxy for the rhizobial partner. Values correspond to the average ± 90% confidence interval by Monte Carlo sampling with 50 pseudo datasets.

Table 1.1.S2. Model predictions and corresponding values

8 mg/L N	T	aNP	gP	gR	β
experimentally estimated	2.96 ± 0.02	0.267 ± 0.004	0.098 ± 0.001	0.085 ± 0.002	-
symmetric, without respiration	8.86 ± 0.11	0.651 ± 0.007	0.066 ± 0.002	0.962 ± 0.044	50.00
symmetric, with respiration	7.97 ± 0.36	0.445 ± 0.011	0.054 ± 0.002	0.424 ± 0.022	50.00
average β, without respiration	5.60 ± 0.04	0.639 ± 0.007	0.073 ± 0.002	0.596 ± 0.027	69.00
average β, with respiration	5.13 ± 0.07	0.426 ± 0.009	0.060 ± 0.001	0.263 ± 0.013	69.00
variable β, without respiration	3.52 ± 0.06	0.626 ± 0.006	0.080 ± 0.002	0.164 ± 0.003	88.42 ± 0.58
variable β, with respiration	3.52 ± 0.06	0.405 ± 0.008	0.066 ± 0.001	0.083 ± 0.002	85.97 ± 0.75

24 mg/L N	T	aNP	gP	gR	β
experimentally estimated	3.48 ± 0.60	0.255 ± 0.004	0.086 ± 0.001	0.070 ± 0.002	-
symmetric, without respiration	6.69 ± 0.09	0.547 ± 0.009	0.106 ± 0.002	0.613 ± 0.029	50.00
symmetric, with respiration	5.42 ± 0.09	0.401 ± 0.009	0.083 ± 0.001	0.235 ± 0.012	50.00
average β, without respiration	4.71 ± 0.04	0.544 ± 0.009	0.108 ± 0.002	0.380 ± 0.018	69.00
average β, with respiration	4.12 ± 0.04	0.398 ± 0.009	0.085 ± 0.001	0.146 ± 0.008	69.00
variable β, without respiration	4.12 ± 0.21	0.541 ± 0.009	0.111 ± 0.002	0.136 ± 0.004	79.79 ± 1.63
variable β, with respiration	3.82 ± 0.13	0.395 ± 0.009	0.086 ± 0.001	0.067 ± 0.002	73.44 ± 1.94

40 mg/L N	T	aNP	gP	gR	β
experimentally estimated	3.51 ± 0.67	0.262 ± 0.003	0.093 ± 0.001	0.083 ± 0.002	-
symmetric, without respiration	5.78 ± 0.07	0.472 ± 0.006	0.106 ± 0.001	0.704 ± 0.043	50.00
symmetric, with respiration	4.78 ± 0.07	0.355 ± 0.006	0.084 ± 0.001	0.271 ± 0.018	50.00
average β, without respiration	4.29 ± 0.03	0.469 ± 0.006	0.108 ± 0.001	0.437 ± 0.026	69.00
average β, with respiration	3.80 ± 0.03	0.352 ± 0.006	0.086 ± 0.001	0.168 ± 0.011	69.00
variable β, without respiration	4.24 ± 0.19	0.466 ± 0.006	0.110 ± 0.001	0.153 ± 0.004	73.44 ± 2.19
variable β, with respiration	3.81 ± 0.11	0.350 ± 0.006	0.087 ± 0.001	0.074 ± 0.002	67.15 ± 2.35

80 mg/L N	T	aNP	gP	gR	β
experimentally estimated	2.89 ± 0.16	0.239 ± 0.004	0.095 ± 0.001	0.091 ± 0.005	-
symmetric, without respiration	3.54 ± 0.06	0.297 ± 0.005	0.108 ± 0.001	2.909 ± 1.561	50.00
symmetric, with respiration	3.09 ± 0.05	0.243 ± 0.004	0.089 ± 0.001	1.048 ± 0.559	50.00
average β, without respiration	3.13 ± 0.03	0.297 ± 0.005	0.108 ± 0.001	1.803 ± 0.968	69.00
average β, with respiration	2.88 ± 0.03	0.243 ± 0.004	0.089 ± 0.001	0.650 ± 0.347	69.00
variable β, without respiration	3.54 ± 0.08	0.296 ± 0.005	0.108 ± 0.001	0.120 ± 0.032	40.38 ± 3.35
variable β, with respiration	3.02 ± 0.04	0.243 ± 0.004	0.089 ± 0.001	0.050 ± 0.012	51.20 ± 3.42

Experimentally estimated or model predicted carbon-for-nitrogen exchange ratio (T; mg C mg^{-1} N), proportion of plant carbon allocated to roots (aNP; g root C g^{-1} plant C), plant per capita growth rate (gP; g C day^{-1} g^{-1} plant C), rhizobial partner growth rate (gR; g C day^{-1} g^{-1} rhizobial partner C), and proportion of bargaining power possessed by the plant (β). Model predictions include those predicted with or without plant respiration, and assume symmetric power (i.e., β = 0.50), average asymmetric power across the observed soil nitrogen range (i.e., β = 0.69), or asymmetric power which varies with soil nitrogen. Values correspond to the average ± 90% confidence interval by Monte Carlo sampling with 500 pseudo datasets (see Methods for further details).

Table 1.1.S3. Plant carbon and nitrogen elemental compositions

Age (wk)	[N] (mg/L)	Rhizobia	Root %C (%wt)	Shoot %C (%wt)	Root %N (%wt)	Shoot %N (%wt)	Plant Yield (mg C mg^{-1} N)
4	8	-	45.89 ± 1.90	39.18 ± 0.58	2.19 ± 0.14	3.46 ± 0.53	13.76 ± 1.69
4	24	-	42.89 ± 2.40	40.49 ± 0.76	2.31 ± 0.19	3.34 ± 0.13	13.74 ± 0.69
4	40	-	43.96 ± 0.84	39.80 ± 0.18	2.31 ± 0.26	3.45 ± 0.28	13.24 ± 1.20
4	80	-	44.19 ± 1.89	39.06 ± 1.06	2.71 ± 0.27	4.57 ± 0.55	10.00 ± 1.65
4	8	+	44.59 ± 1.85	41.06 ± 0.13	2.34 ± 0.23	4.41 ± 0.19	10.69 ± 0.46
4	24	+	44.13 ± 0.74	40.72 ± 1.42	2.40 ± 0.13	4.26 ± 0.08	10.86 ± 0.36
4	40	+	43.51 ± 0.49	39.81 ± 0.43	2.82 ± 0.08	4.38 ± 0.31	10.21 ± 0.59
4	80	+	45.63 ± 0.38	39.34 ± 1.21	3.23 ± 0.24	4.75 ± 0.35	9.23 ± 0.52
6	8	-	43.43 ± 1.41	38.40 ± 0.78	2.00 ± 0.36	2.60 ± 0.52	17.16 ± 3.12
6	24	-	44.32 ± 0.48	39.17 ± 0.43	1.99 ± 0.26	2.52 ± 0.09	17.48 ± 1.14
6	40	-	41.93 ± 4.68	39.45 ± 0.65	2.08 ± 0.21	2.89 ± 0.22	15.21 ± 1.47
6	80	-	44.43 ± 2.32	39.56 ± 1.02	2.99 ± 0.15	4.26 ± 0.83	10.25 ± 1.64
6	8	+	43.02 ± 3.15	39.30 ± 1.54	2.28 ± 0.09	3.47 ± 0.51	12.77 ± 1.90
6	24	+	42.10 ± 1.73	39.45 ± 0.56	2.16 ± 0.41	3.67 ± 0.31	12.06 ± 1.02
6	40	+	41.51 ± 2.64	39.98 ± 0.29	2.42 ± 0.21	3.65 ± 0.59	12.24 ± 1.68
6	80	+	42.86 ± 0.48	41.12 ± 0.10	2.98 ± 0.18	3.79 ± 0.06	11.49 ± 0.17

Carbon and nitrogen elemental compositions were measured in 4- and 6-week-old uninoculated (-) and nodulated (+) plants at 8, 24, 40, and 80 mg L^{-1} N. Values are average ± standard deviation.

Table 1.1.S4. Nodule carbon and nitrogen elemental compositions

Age (wk)	[N] (mg/L)	Nodule %C (%wt)	Nodule %N (%wt)	Nodule Yield (mg C mg^{-1} N)
4	8	44.30 ± 0.52	8.10 ± 0.11	5.47 ± 0.01
4	24	45.26 ± 0.06	7.85 ± 0.48	5.78 ± 0.36
4	40	43.02 ± 1.14	7.50 ± 0.42	5.74 ± 0.27
4	80	43.150	7.170	6.020
6	8	43.53 ± 1.14	6.52 ± 0.36	6.69 ± 0.36
6	24	44.56 ± 1.10	6.72 ± 0.24	6.64 ± 0.15
6	40	43.57 ± 1.07	6.89 ± 0.24	6.32 ± 0.09
6	80	40.47 ± 0.92	6.43 ± 0.68	6.32 ± 0.52

Carbon and nitrogen elemental compositions were measured in 4- and 6-week-old nodulated plant systems at 8, 24, 40, and 80 mg L^{-1} N. Values are average ± standard deviation.

SECTION 1.2

Modeling nutritional mutualisms: Challenges and opportunities for data integration

---

This section includes a manuscript that was published in *Ecology Letters*:
Clark, T. J., Friel, C. A., Grman, E., Shachar-Hill, Y., & Friesen, M. L. (2017). Modeling nutritional mutualisms: challenges and opportunities for data integration. *Ecology letters*, *20*(9), 1203-1215.

Preface

While performing experiments to test the effect of soil nitrogen availability on nutrient exchange in the legume-rhizobia mutualism (Section 1.1), our collaborative group read and discussed articles that used mathematical models to examine the stability or function of plant-microbe nutritional mutualisms. Due to the limited number of relevant review articles and our collaboration's expertise in different modeling approaches (MLF in game theory and adaptive dynamics; EG in population dynamics; YSH in metabolic network analyses), we decided to also write a review article on how these mutualisms have been addressed by mathematical models.

Our initial conception of the review was to introduce readers to five mathematical approaches commonly used to model nutritional mutualisms: metabolic networks, biological markets, game theory, population dynamics, and ecological networks. We summarized how these models are constructed and have addressed four foundational questions we identified concerning mutualism persistence, community stability, nutrient availability, and structured interactions. All collaborators contributed significantly to the writing. I took primary responsibility for the market approach section, and nutrient availability and structured interactions foundational questions, as well as co-authoring the sections on metabolic network and game theory approaches.

In addition to the above topics, we introduced how these modeling approaches could be integrated with data or combined with other approaches. However, due to space limitations, we were not able to thoroughly address these directions, and our review was initially rejected by *Ecology Letters* with encouragement to resubmit. In response to reviewer feedback, we decided to pursue writing two types of reviews: a conventional review to compile the findings of the different modeling approaches, and a synthesis review to elaborate on how these models could be integrated with experimental data to increase their utility. I led the revisions for, and am first author on, the synthesis review.

The synthesis review still included information about the five mathematical approaches, but was

restructured into how the different approaches function at four biological scales: cell, individual, population, and community. We also continued to address the foundational questions, but included more connections to experimental studies. For each of the biological scales and foundational questions, we included at least one reference to an experimental paper or review of experimental approaches that had been or could be connected to a specific model. I took responsibility for connecting empirical data to modeling approaches, selecting appropriate complementary studies from the literature, and exploring future possibilities. I coordinated these changes using drafts from our initial review and numerous discussions within our collaboration where we addressed how models and data could be integrated.

Our reconstructed review was published by *Ecology Letters* in 2017 and is included in this section. To our knowledge, this review was the first to explore the existing and potential integrations between experiments and theory-based computational methods across the full range of biological scales. The previous section is one of the very few studies to integrate between these approaches.

Abstract

Nutritional mutualisms are ancient, widespread, and profoundly influential in biological communities and ecosystems. Although much is known about these interactions, comprehensive answers to fundamental questions, such as how resource availability and structured interactions influence mutualism persistence, are still lacking. Mathematical modeling of nutritional mutualisms has great potential to facilitate the search for comprehensive answers to these and other fundamental questions by connecting the physiological and genomic underpinnings of mutualisms with ecological and evolutionary processes. In particular, when integrated with empirical data, models enable understanding of underlying mechanisms and generalization of principles beyond the particulars of a given system. Here, we demonstrate how mathematical models can be integrated with data to address questions of mutualism persistence at four biological scales: cell, individual, population, and community. We highlight select studies where data has been or could be integrated with models to either inform model structure or test model predictions. We also point out opportunities to increase model rigor through tighter integration with data, and describe areas in which data is urgently needed. We focus on plant-microbe systems, for which a wealth of empirical data is available, but the principles and approaches can be generally applied to any nutritional mutualism.

Introduction

Nutritional mutualisms are both ancient and widespread, shaping life on this planet as it exists today with innovations including the eukaryotic cell and the colonization of land by plants (Bronstein, 2015). Substantial attention has been focused on how these mutualisms have persisted over evolutionary time despite the potential benefits of defection (Sachs *et al.*, 2004; Ghoul *et al.*, 2014; Bronstein, 2015), yet this evolutionary threat stems largely from theoretical considerations rather than empirical data demonstrating that cheaters prosper within contemporary mutualisms (Jones *et al.*, 2015) and

phylogenetic analysis demonstrates that the evolution of parasites from within mutualistic clades occurs rarely (Sachs *et al.*, 2011). In wild populations, parasites of mutualisms often come from outside the focal interaction, as illustrated by non-fixing rhizobia that originated from non-symbiotic lineages (Sachs *et al.*, 2010), and mutualistic symbioses are highly stable over evolutionary time (Werner *et al.*, 2015). Together, these observations suggest the operation of mechanisms, potentially operating at multiple scales, that result in the evolutionary robustness of mutualisms as a whole. Further complicating our understanding of mutualisms is the observation that interactions may be mutualistic under some contexts but parasitic under others; this context dependence is often linked to the availability of external resources (Johnson *et al.*, 1997; Chamberlain *et al.*, 2014). Contemporary evolution of rhizobia under long-term nitrogen fertilization results in strains that fix less nitrogen, though whether these strains have higher fitness than their beneficial relatives and are thus cheaters is still unknown (Weese *et al.*, 2015). Critical empirical work addressing these issues is lacking (Friesen & Heath, 2013), but larger questions about mutualism remain.

Understanding the persistence and predicting the functioning of contemporary nutritional mutualisms cries out for mathematical modeling that is closely integrated with empirical data. Central to predicting mutualism persistence is determining the balance between fitness costs and benefits to each partner in natural ecosystems, which in turn depends upon the structure and operation of biochemical and physiological networks. While many open questions about mutualism remain, two that cut across scales of biological organization concern how mutualism is influenced by environmental nutrient ability and how mutualism is influenced by structured interactions. Varying nutrient availability in the environment can decisively alter the cost-benefit ratios of engaging in trade and thus lead to a mutualism-parasitism continuum correlated with low-to-high nutrient availability (Johnson *et al.*, 1997; Lau *et al.*, 2012). Therefore, nutrient availability alters the relative efficiency of direct nutrient uptake versus exchange between partners, but the molecular and physiological basis of this phenomenon are still being clarified. Translating the fitness consequences of varying environmental nutrients into consequences for

population stability and community diversity also remains an open challenge. In addition, partner identity can greatly impact an individual's fitness (Friesen, 2012), so structured interactions, where individuals associate with only a subset of potential partners, may influence mutualism persistence. These can arise through mechanisms operating at multiple scales, such as molecular compatibility, partner choice, or spatially structured populations (Noë & Hammerstein, 1994; Perret *et al.*, 2000; Paszkowski, 2006; Archetti *et al.*, 2011).

Mathematical models play a key role in addressing questions about mutualism because they allow researchers to qualitatively and/or quantitatively test the effects of hypothesized mechanisms, strategies, and interaction structures. Although incorporation of prior knowledge is inherent in the construction of all models, they vary widely. Models each fall somewhere in the space of generality, realism, and precision, as discussed by Levins (1966) for population biology models. Most of the progress to date in understanding mutualisms has leveraged general 'proof-of-concept' models. These models are extremely valuable in exploring the consequences of particular assumptions about reality (Servedio *et al.*, 2014) and enabling a general understanding of mutualisms. However, although they may use data to make appropriate assumptions, they are typically less realistic and therefore restricted in their parameterizability for a specific system. Consequently, they generate conceptual predictions that cannot be evaluated quantitatively. We argue that it is time for an increased focus on realistic, precise models to contribute to progress towards settling the larger questions in mutualisms. Modern advances in -omic and other analytical techniques enable close integration of models and empirical data. This tight integration of models and data can allow researchers to measure or infer parameter values and generate quantitative predictions which can be compared to measured values. Multiple models making contrasting predictions could be compared either qualitatively or quantitatively using information theoretic criteria through model selection (Burnham & Anderson, 2003) to evaluate the appropriateness of different assumptions, model structures, or modeled mechanisms. Consequently, we believe that greater efforts to obtain and

integrate high quality empirical data will accelerate progress in answering biological questions with mathematical models.

Plant-microbe nutritional mutualisms are amenable systems for addressing these issues because resource fluxes can be tracked in both directions and there are existing models at multiple biological scales in addition to relevant empirical datasets. These mutualisms are evolutionarily important and play major roles in global nutrient cycling. For example, mycorrhizal fungi colonize plant roots and use their hyphal networks to supply phosphorus and other minerals in exchange for roughly five billion tons of photosynthetically fixed carbon globally each year (Bago *et al.*, 2000). Similarly, rhizobia, soil bacteria that colonize plant roots, fix more than forty million tons of atmospheric nitrogen and exchange it for host carbon (Udvardi & Poole, 2013). Understanding these interactions thus has practical implications for the conservation and management of natural and agricultural ecosystems. We note, however, that the data integration approaches described here are broadly applicable beyond these systems.

In emphasizing the integration of empirical data in modeling mutualism at multiple biological scales, we present a synthetic review of how our knowledge of mutualism has expanded through mathematical modeling of three key questions and point out opportunities to go further. We explore (i) how mutualisms function, including molecular function, ecological dynamics and community structure, and evolutionary persistence; (ii) how the availability of nutrients outside of the interaction affects mutualism function; and (iii) how the existence of structured interactions influences mutualism function and dynamics across biological and temporal scales. While many other questions remain, these have received attention or are readily addressable at four biological scales. For each question we consider (1) cell scale models, which characterize the genes and enzymatic reactions in a single cell that underpin mutualisms; (2) individual scale models, which describe how individuals regulate trade and partner interactions; (3) population scale models, which focus on the persistence and coexistence of mutualistic host and symbiont guilds; and (4) community scale models, which assess how interactions with additional

guilds (e.g., a predator guild) alter mutualistic interactions. Finally, we discuss explicit recommendations to enhance integration of data and nutritional models and urge future collaborations with the aim of iterating between models and data collections. Given the extensive mutualism literature, models and empirical data presented are illustrative, not comprehensive, but we attempt to be as prescriptive as possible. The productive avenues for future research that we identify will facilitate data integration into mathematical models and further elucidate general principles governing mutualism function and stability.

How do mutualisms function?

The basic underpinning of nutrient-exchange mutualisms is the functionality whereby resources are exchanged between partners. There are multiple inter-related questions across scales concerning the factors underlying these fluxes and the consequences these exchanges have. We seek to understand mechanistically how enzymes and other cell components permit the physical exchange of nutrients, and how individuals decide when and how much to invest in trade. Nutrient exchanges between individuals can in turn influence both how populations interact and how these interactions impact multipartite communities.

Cell Scale

Nutritional mutualisms are founded on the operation of metabolic networks, which conduct biochemical fluxes of energy and matter within and between partners. Cell scale models analyze the functional capabilities of these networks, describing them at the level of enzymatic and transport reactions. Empirical data from annotated genomic (Resendis-Antonio *et al.*, 2007) and/or protein (Rodriguez-Llorente *et al.*, 2009) databases determines which biochemical and transport reactions to include in the network (Table 1.2.1). The completeness of these networks is tested by their ability to computationally use known substrates to produce known products (including cell components). Predictions about which genes are

Table 1.2.1. The relationship of mathematical modeling to empirical data at multiple scales.

Model scale	Model component	Data for building and/or parameterizing models	Data for testing model predictions
Cell	Reaction network	Annotated genome	Transcriptome, proteome, isotopic labeling, metabolic flux analysis maps
	Nutrient uptake/efflux	Gas exchange, chemical analysis, isotopic labeling	
Individual	Cost-benefit ratios	Fitness of each partner under different pairings	Dynamics of strategies across generations
	Nutrient uptake rate	Isotope labeling of external resource	Growth and nutrient content of each partner
	Allocation strategy	Tissue weights and elemental compositions; Isotope labeling	
Population	Interaction coefficient	Manipulation of partner densities and measurements of individual fitness or population growth rate	Long-term population dynamics
	Intrinsic growth rate	Individual fitness or population growth at low density	
	Dispersal	Typical distances partners disperse in the field; Population genetics	Spatial patterning of mutualism persistence through time
Community	Community network structure	Which species interact under field conditions	Long-term population dynamics, diversity, and coexistence
	Indirect effects	Abundances of partners in the presence or absence of additional community members	

The construction of models involves information about the system as well as the hypotheses to be evaluated, and the amount of data integrated in this process varies according to the modeling type. The use of quantitative empirical data from experimental or observational studies is also important in testing **Table 1.2.1 (cont'd).** model predictions, both qualitative outcomes such as persistence versus extinction

Table 1.2.1 (cont'd) but also quantitative outcomes such as individual metabolic fluxes or nutrient exchange ratios. Model components include the structure, constraints, predictions, and parameters.

essential for the mutualism can be tested by comparing model predictions to phenotypes of mutants that lack a particular reaction (Zhao *et al.*, 2012). Flux balance analysis (FBA; Box 1) of cell scale models can be used to test "objective functions" that embody hypotheses about metabolic strategies by comparing the model predictions to growth measurements or other empirical data. A series of studies on *Rhizobium etli* (Resendis-Antonio *et al.*, 2007; Resendis-Antonio *et al.*, 2011) found that an objective function of maximizing nitrogen fixation and production of known bacterial compounds (per input of carbon from the plant) yielded results consistent with observed fixation rates. However, Zhao et al. (2012) found that an objective function of strong coupling between nitrogen fixation and carbon-nitrogen exchange better agreed with published literature (e.g., proteomic and biochemical analyses) on which metabolic pathways have been reported as active or inactive in *Sinorhizobium meliloti*.

Although consistent with literature on the known biochemistry and gene involvement, the results of these modeling studies concerning internal metabolic fluxes should be regarded as hypotheses until they can be validated using experimental biochemical analyses of metabolic fluxes. Network-wide metabolic rates can be mapped using metabolic flux analysis, which models the results of isotopic labeling experiments to quantify metabolic and transport fluxes (Ratcliffe & Shachar-Hill, 2006). For example, Chen et al. (2011) applied FBA to aerobically and anaerobically growing bacteria, with an objective function of maximizing growth, and compared the predictions to findings from metabolic flux analysis. They found that although FBA could satisfactorily predict bacterial growth and metabolite secretion rates, it greatly under-estimated most oxidative pentose phosphate pathway fluxes. An analogous approach could be used to rigorously test the patterns of metabolic flux in plant-microbe mutualisms predicted by FBA modeling using different objective functions. Detailed quantitative model testing may reveal that one objective function serves to explain metabolic and transport fluxes across a wide range of nutritional

mutualisms, or that different partners or environmental conditions elicit different metabolic strategies.

Individual Scale

Mutualisms occur between individuals that exchange goods or services, and individual scale models have long been used to study these interactions over both physiological and evolutionary timescales. The general question of why mutualisms are not dominated by defectors has been a focus of the field for decades (reviewed in Jones et al. 2015) and many individual scale models have been aimed at answering this (Box 2). Numerous mechanisms have been proposed that result in net selection favoring mutualistic exchange, ranging from the cross-generation fitness coupling through partner fidelity feedback, preferential allocation/sanctions, and structured interactions mediated by spatial structure or partner choice (Sachs *et al.*, 2004). Individual scale modeling studies can differentiate between these mechanisms and/or determine conditions under which they can effectively promote mutualism (West *et al.*, 2002; Yamamura *et al.*, 2004; Doebeli & Hauert, 2005) For example, preferential allocation, where higher-quality partners are given more reward, promotes mutualism in two modeling studies of plants interacting with multiple symbionts, but only if adequate partner discrimination ensures the intended partner receives the benefits (Cowden & Peterson, 2009; Bever, 2015). Furthermore, Cowden and Peterson (2009) tested the effectiveness of different plant carbon regulation strategies. They found that being able to control whether or not to engage in trade with particular symbionts maximizes plant nutrient acquisition, but in order to maximize carbon use efficiency, this mechanism must be combined with the ability to control the total amount of carbon allocated to belowground organs.

These and other predictions concerning the allocation of nutrients can be empirically tested using labeling experiments (Table 1.2.1). For example, Zheng et al. (2015) investigated preferential allocation using carbon- and phosphorus-labeling in a plant-mycorrhizae split-root system. In abundant light where photosynthesis is maximized and carbon is not limiting, the plant allocates a larger fraction of its carbon

to high-quality symbionts, but if light is limiting, the plant does not preferentially allocate carbon. The authors hypothesize that this may provide opportunities for defectors to propagate, resulting in the coexistence of defectors and cooperators. A model of the system could make quantitative predictions regarding allocation strategies. Since exchanged nutrients can be tracked using isotopes, individual scale models are well-suited to examining regulatory mechanisms like preferential allocation, and how abiotic and biotic factors influence the ability to regulate trade.

Population Scale

Although cell and individual scale models can investigate mechanisms underlying mutualistic interactions, population scale models are needed to address the impact of mutualism on ecological processes. Early models of mutualism predicted unbounded population growth (Gause & Witt, 1935), which is clearly not observed in natural populations, so many models incorporate additional population regulatory mechanisms, such as density-dependence or the spread of defectors (Marco *et al.*, 2009; Pillai *et al.*, 2014). For example, symbionts compete for host resources and if cooperators are competitively inferior to defectors they can be excluded if their resource niches overlap too much. However, a model by Pillai et al. (2014) suggests that cooperators can persist if a subset of defectors' niches only intermediately overlap with the cooperators'. These defectors compete weakly with both cooperators and other defectors without excluding either. This has an indirect positive effect on cooperators by decreasing the abundance of defectors that compete strongly with cooperators. Consequently, overlap in resource niche may be able to serve as a proxy for the magnitudes of interaction coefficients (Box 3; Pillai et al., 2014).

The model of Pillai et al. (2014) can be tested qualitatively, but could also be extended to have higher realism in order to facilitate empirical parameterization and the testing of quantitative predictions (Table 1.2.1). To test the model, host fitness in association with symbiont monocultures or mixtures could be compared. An independent experimental study found that some rhizobial monocultures increased host

fitness, while others had no effect (Barrett *et al.*, 2015); these observations could be attributed to cooperator and defector monocultures, respectively. In mixed cooperator-defector cultures, host fitness was either slightly increased or not affected compared to fitness without symbionts (Barrett *et al.*, 2015); this relative lack of benefit from the mixture could be explained by intermediate or strong, respectively, overlap in resource niche between the cooperators and defectors. However, these observations could alternatively be explained by differences in rhizobial population sizes or the host benefits obtained from the cooperators being countered by investing in sanctions against defectors. Further empirical data to characterize species interactions would be required to parameterize and test a more realistic extension of the model of Pillai et al. (2014).

Community Scale

Considering mutualisms in the context of other guilds and interactions is key to understanding how they function in natural settings. Communities with multiple guilds (e.g., host, mutualist, and antagonist guilds) contain multiple direct and indirect interactions, including competition, predation, and commensalism (Mougi & Kondoh, 2012). Community scale models can assess how the sum of these interactions shapes community dynamics (Box 4). For example, mutualist species can have indirect negative effects on antagonists by enhancing host defenses (Bachelot *et al.*, 2015). Community scale models have found that the mixture of these interactions yields different community dynamics than the isolated interactions in host-mutualist subcommunities (Mougi & Kondoh, 2012; Bachelot *et al.*, 2015). Bachelot et al. (2015) investigated how communities are affected by the magnitudes of interaction effects in combination with species abundances. If mutualists only provide a modest benefit to the host, then intermediate mutualist abundance can stabilize coexistence of the three guilds by enhancing defense against the antagonist without leading to antagonist extinction (Bachelot *et al.*, 2015). This model could be empirically tested by measuring how host and antagonist fitnesses are affected by mutualist abundance (Table 1.2.1).

An independent empirical study on plant-mycorrhizal fungi-herbivore communities found that as fungal abundance increased, the production of plant defense compounds increased as expected, but the growth rate of the specialist herbivore unexpectedly increased as well (Vannette & Hunter, 2013). In order to ensure accurate measurements of all three variables (fungal abundance, defense compound production, herbivore growth rate), these experiments were performed under controlled conditions that limited the presence of other herbivores and predators (Vannette & Hunter, 2013). Although the mutualist did not have the expected indirect negative effect on the antagonist as predicted by the model (Bachelot *et al.*, 2015), the underlying mechanism in the model may still operate in other experimental settings because the plant defense traits may be more effective against generalist herbivores or may function to attract herbivore predators (Vannette & Hunter, 2013). Consequently, this illustrates the difficulty in obtaining empirical support for community scale models: controlling experimental variables while retaining the complexity of communities. Future studies would need to be conducted with natural herbivore communities and controlled manipulation of belowground symbionts, e.g., following methods developed by Simonsen and Stinchcombe (2014), who created a field system for inoculation with specific rhizobia.

How does nutrient availability affect mutualisms?

Although nutrient-exchange mutualisms are typically conceptualized as benefiting both partners, experiments have revealed that as nutrient availability increases, these interactions can become parasitic either due to plasticity (Johnson *et al.*, 1997; Lau *et al.*, 2012) or evolution (Weese *et al.*, 2015). Understanding the molecular and physiological bases of variation in resource exchange rates and linking them to individual fitness has the potential to inform both ecological and evolutionary dynamics of systems. Phylogenetic analysis demonstrates that even in high nutrient environments, the legume-rhizobia mutualism is evolutionarily stable (Werner *et al.*, 2015), suggesting the existence of strong

mechanisms that stabilize this interaction. Given the ongoing massive anthropogenic perturbations of nutrient cycling, it is crucial to able to predict population and community dynamics as a result of environmental nutrient levels; such predictions will require a mechanistic understanding of how partners interact.

Cell Scale

Because nutrient uptake rates are a key component of most cell scale models, it is important to understand how nutrient availability affects model predictions concerning the metabolic basis for trade. However, this has yet to be addressed in a plant-microbe mutualism model. Nutrient uptake rates usually correlate with nutrient availability in a manner similar to Michaelis-Menten kinetics (Rao *et al.*, 1993). Therefore, a feasible range of expected uptake rates can be generated and used to determine how nutrient availability affects internal metabolic fluxes. Alternatively, nutrient availability can be assessed by comparing the predictions of different uptake costs. For example, at low external nitrogen levels, ammonium can be taken up by active transport (Kraiser *et al.*, 2011), which has a higher carbon cost than the passive nitrogen transporters that can sustain uptake at higher nitrogen levels.

A widely used approach to assessing model predictions is to use changes in transcript or protein levels to explore molecular responses (Resendis-Antonio *et al.*, 2011) to environmental change. Because gene expression is markedly different at plant-microbe interfaces, a potentially productive future avenue would be to take advantage of localized transcriptomic or quantitative proteomic data from both host and microbial cells to test and improve cell level modeling. This raises the wider point that plants and some of their partners have differentiated cells and tissues with both short and long distance transport. Grafahrend-Belau et al. (2013) addressed this challenge by constructing separate FBA models for the leaves, stems, and seeds, and connecting these FBA models using phloem and root transport reactions. This multi-organ FBA model was used to investigate changing carbon source-sink interactions in barley

during senescence. This approach could be used to investigate the effect of nutrient availability on trade mechanisms by integrating a multi-organ plant model with a mutualistic microbe model.

Individual Scale

Individual scale models can investigate how nutrient availability affects investment in trade through nutrient uptake rates and other nutrient-dependent physiological parameters, in some cases with fluctuating resource environments (Moeller *et al.*, 2016). These physiological models typically assume some form of optimality, arguing that evolution will optimize the parameters to maximize performance. Some individual scale models (Grman *et al.*, 2012; Franklin *et al.*, 2014) used existing literature to obtain empirical values for these parameters and explore the sensitivity of model outcomes to environmental variation. For instance, the threshold of nutrient availability associated with switching between a strategy where one nutrient is acquired by both direct uptake and from trade, and a strategy of relying entirely on trade was determined for plant-mycorrhizae mutualisms (Grman *et al.*, 2012; Franklin *et al.*, 2014). Parameter values were taken from publications on different systems and/or experimental conditions, and varied drastically. For example, in the Grman et al. (2012) model, the fungal nutrient uptake rate varied by two orders of magnitude. Therefore, this method of data incorporation can allow model predictions of general patterns across diverse mutualisms to be quantitative. However, in order to make precise predictions for a specific mutualistic interaction that could enable the model to be refined or even falsified, and thus elucidate previously unknown features of the mutualism, data collected from a single study system is necessary.

To test the sufficiency of a model for a specific system, the majority of parameter values should be obtained from a single study, or at least on the same system, and care needs to be taken to use separate data for fitting and testing the model. Jamshidi et al. (2015) took this approach to validate a model of plant-fungus trade. Specifically, they compared the predicted and observed patterns of leaf

biomass and carbon trade in correlation with symbiont cooperativeness. After model validation, sensitivity analysis was performed to show that the thresholds of nutrient availability to elicit a change in plant allocation strategy from growth to storage have significant effects on plant biomass. The use of parameter value ranges from studies on a variety of systems has the advantage of generalizability but is less stringent for testing the falsifiability of a model. In particular, rigorous testing of parameterized models requires comprehensive physiological datasets that account for all model components, but there are few such sets currently available. Therefore, there is great need for generating these in additional plant-microbe systems and for testing additional models.

Population Scale

Population scale models can assess how resource availability affects the nutritive requirements of individuals, thus affecting the magnitudes of negative and positive interactions and hence population dynamics (De Mazancourt & Schwartz, 2010). The role of resource availability can also be tested through its influence on carrying capacity or intrinsic growth rate, both of which affect population dynamics and stability (Neuhauser & Fargione, 2004). Consequently, models at this scale can investigate the driving force for the mutualism-parasitism continuum correlated with low to high nutrient availability as well as the role of nutrient availability in species coexistence. These models have found that trade enables species to coexist at greater ranges of nutrient availability, but this cooperation may decrease the population density of one partner (De Mazancourt & Schwartz, 2010). At low host densities, the symbiont with the faster growth rate is favored, while at high host densities, the symbiont with the greater competitive advantage is favored (Neuhauser & Fargione, 2004).

These predictions could be empirically tested by determining host abundance, benefit from trade, and symbiont growth rates and competitive abilities, and assessing whether the shift from mutualism to parasitism correlates with the expected change in nutrient availability. Although host biomass is

sometimes used as a proxy for population density (Neuhauser & Fargione, 2004), this is not feasible for all systems (e.g., tree-ectomycorrhizal fungi systems). In systems where direct measurements of biomass or density are not good proxies for carrying capacity, it may be desirable to estimate carrying capacity using another model. Xia and Shao (2008) used a modeling approach to quantify carrying capacity for plants in arid and semi-arid environments by identifying the limiting resource (water) and quantifying processes that affect or are affected by the limiting resource (e.g., carbon allocation, photosynthetic rate) at different timescales (e.g., hourly, daily). If applied to plant-microbe nutritional mutualisms, the limiting resource could be the traded nutrients.

Community Scale

Similar to population scale models, community scale models have explored how nutrient availability influences community dynamics by altering the magnitudes of interspecific interactions (Gross, 2008; Georgelin & Loeuille, 2016). For example, because symbionts tend to be more mutualistic at lower nutrient availabilities, selection favors traits that attract symbionts when nutrients are scarce; however, these traits may also attract antagonists (López-Ráez *et al.*, 2011; Georgelin & Loeuille, 2016). If mutualists and antagonists are both strongly attracted to the host, then mutualists cannot invade a host-antagonist community because the antagonist prevents the host density from increasing, but antagonists can invade a host-mutualist community (Georgelin & Loeuille, 2016). As nutrient availability increases, it can be advantageous to lose the trait(s) that attract mutualists to lessen the negative effects of antagonists. This increases host abundance and enables mutualists to invade host-antagonist communities (Georgelin & Loeuille, 2016).

A system in which the predictions of this model could be empirically tested is plant production of strigolactones, a signaling compound that stimulates plant symbiosis with arbuscular mycorrhizal fungi (López-Ráez *et al.*, 2011). When nutrients are scarce, plants benefit from interactions with mycorrhizal

fungi (Johnson, 2010) and so they produce strigolactones to attract partners (Jamil *et al.*, 2011; López-Ráez *et al.*, 2011). However, strigolactones also attract parasitic plants, so as nutrients becomes more available and fungi become less beneficial, strigolactone production decreases (Jamil *et al.*, 2011; López-Ráez *et al.*, 2011). Together, these studies empirically validate the model assumption that the host has a trait (i.e., strigolactone production) that attracts both mutualists and antagonists, and correlates with nutrient availability. Consequently, this system could be used in long-term studies to evaluate model predictions such as conditions under which mutualists can invade host-antagonist communities.

How do structured interactions affect mutualisms?

Mutualisms occur across a heterogeneous landscape that encompasses both variation in externally-supplied nutrients, as explored above, but also variation in the identities and states of the partners that one associates with. The structuring of interactions can have immediate consequences for mutualism function, such as through variation in metabolic potential of individual cells, as well as longer-term ramifications for the evolutionary stability of mutualism, such as the effect of spatial structure in evolutionary game theory models.

Cell Scale

There is great opportunity for leveraging the mechanistic power of cell scale models to assess how spatially structured interactions influence mutualisms compared to the 'well-mixed' scenario. Each cell can be represented by a single FBA model, and cells can be linked by allowing the metabolic outputs of cells to diffuse into the environment and then serve as metabolic inputs for other cells. These networks have the potential to reveal trends in metabolism that could result from local competition for nutrients or metabolic effects from interacting in a structured manner with multiple partners. Although this method has yet to be applied to plant-microbe mutualisms, it has led to the illumination of a cross-feeding

relationship between bacteria growing in colonies (Cole *et al.*, 2015). Cole et al. (2015) simulated a bacterial colony using a three-dimensional lattice structure with literature values for parameters such as substrate uptake, efflux, and environmental diffusion rates. Simulations showed one group of bacteria growing in a hypoxic microenvironment using fermentation and releasing acetate, while a group not lacking oxygen predominantly consumed that acetate via the Krebs cycle. The existence of these groups was confirmed by imaging bacteria that express the green fluorescent protein when they consume acetate (Cole *et al.*, 2015).

A similar approach could be applied to rhizobial bacteria in a nodule or multiple nodules distributed in a heterogeneous soil. Spatially structured models of single nodules could address whether diffusion of gases through the nodule causes the formation of zones of metabolically distinct rhizobia, and whether a nutritional relationship exists between nitrogen-fixing differentiated rhizobial cells and dividing rhizobial cells in indeterminate nodules. The soil matrix varies in the levels of gases, water, and nutrients and we predict that nodules in distinct microenvironments may show contrasting metabolic functionality. FBA with multiple symbiont cells under varying conditions exchanging nutrients with a single host plant would shed light on the metabolic constraints and potential optimal distribution of host resources.

Individual Scale

Spatial structure in individual scale models can generate conflicting predictions regarding the evolutionary persistence of mutualistic strategies. Early individual-based simulations showed that spatial structure can promote the evolution of mutualism because limited dispersal structures interactions in space, so that offspring of one partner tend to occur near offspring of the other partner generating pseudo-vertical transmission (Yamamura *et al.*, 2004). If mutualists have increased fitness when interacting with other mutualists, limited dispersal allows mutualists to reproduce in clusters and maintain sufficient abundance to persist for a range of cost-benefit ratios (Yamamura *et al.*, 2004). However, spatial structure does not

always promote cooperative strategies: in models of intraspecific cooperation, the type of evolutionary game is critical and while in the Prisoner's Dilemma cooperative clusters can form, in the Snowdrift or Hawk-Dove game cooperators form dendritic patches and the cooperator strategy is often lower than under well-mixed conditions (Hauert & Doebeli, 2004).

Spatial structure may also negatively impact the evolutionary persistence of mutualism by restricting the range of partners that may be selected from in a biological market model or the types of partners that a given host has the option of preferentially allocating resources to. For example, Verbruggen et al. (2012) modeled the effectiveness of preferential allocation with high and low symbiont spatial structure and found that with high structure, less beneficial symbionts were favored by selection. The authors then qualitatively tested this prediction in a plant-mycorrhizae system and found that cooperation was more prevalent in well-mixed conditions. They attributed this outcome to the plants being able to enforce cooperation on a fine scale, i.e., through sanctions / preferential allocation. We suggest that spatial structuring could potentially be playing a role through the form of the payoff matrix in the interaction, similar to the Snowdrift game within species—unlike the Prisoner's Dilemma, in which spatial structure promotes cooperation, in the Snowdrift game, spatial structure leads to a lower equilibrium level of cooperation (Doebeli & Hauert, 2005). This alternative hypothesis could be tested by measuring the fitness of symbionts in a fine-scaled manner across generations to estimate the payoff structure of the interaction; a more detailed model could then be parameterized with this data and the relative importance of various factors assessed by comparing model outcomes to empirical data.

Population Scale

Mutualisms at a population scale can be considered across space using a meta-population approach, whereby populations interact locally but then undergo extinction-colonization dynamics across a landscape. A meta-population model grounded in the rhizobia-legume mutualism found that spatial

dynamics can lead to genetic mosaics for both partners, with specialists co-occurring in local patches (Parker, 1999). Stanton-Geddes and Anderson (2011) empirically investigated if legume dispersal distance is limited by the availability of rhizobia by transplanting legume seeds to areas within, at, or beyond the legume's natural range. They found that plants beyond the natural range were less likely to form nodules without manual inoculation than those within, and that nodulation aided plant growth. This suggests that host range is limited by rhizobia availability and that mutualism is promoted over ecological time-spans at shorter dispersal distances.

Population dynamics models can be used to make predictions about the role of dispersal rates, and subsequent spatial structure, on the prevalence of contrasting mutualistic strategies in wild populations. For example, Abbott et al. (2015) constructed a general model of invasive species dynamics in Californian grasslands wherein the native cooperator species is only competitively superior to the invading defector species when interacting with mutualistic soil microbes, and both plant species disperse between patches that either contain or lack such microbes. They found that higher dispersal rates can lead to exclusion by preventing source populations from maintaining sufficient density to replenish sink populations if the sink is outcompeted or otherwise disturbed. Consequently, mutualism persistence is promoted when cooperators disperse less than defectors. These qualitative predictions could be examined more quantitatively by altering the model to increase its realism and then incorporating data for parameter values and predictions.

Community Scale

Community scale models can utilize network theory to explicitly examine the consequences of interactions that occur between some pairs of populations but not others. These models can ask how the network structure of ecological interactions affects community stability (Thébault & Fontaine, 2010) and biodiversity (Bastolla *et al.*, 2009). Other models investigate the development of structured interactions

by evaluating how coevolution guides the formation of stable interaction network structure (Guimaraes Jr *et al.*, 2011). One key property of networks is their degree of nestedness, which occurs when specialists interact with a subset of those with which generalists interact. The role of nestedness in mutualistic communities is controversial, but community stability and biodiversity are often attributed to highly nested networks (Bastolla *et al.*, 2009; Thébault & Fontaine, 2010).

The relationship between network structure and community stability is supported by plant-pollinator mutualistic subcommunity networks, which tend to be highly nested (Vázquez *et al.*, 2009). However, Toju et al. (2015) recently analyzed interactions in three plant-fungal subcommunity network studies that utilized sequencing methods to detect interactions and found that plant-fungal subcommunity networks appear to be anti-nested; i.e., specialists interact with subsets of generalists less frequently than expected by random chance. The authors hypothesized that the degree of anti-nestedness is negatively correlated with the overlap in fungal host range, which may be due to interspecific competition or differences in habitat preference. Consequently, there is great need for community-scale modeling studies that focus on communities containing plant-microbe nutritional mutualisms in order to understand why these mutualistic subcommunity networks may be anti-nested and to explore the community-level consequences of anti-nestedness. Specific communities could be assembled under controlled conditions to parameterize and test these models. In addition, empirical data for other nutritional mutualisms is needed from diverse natural systems to understand the drivers of network properties.

Discussion

Mathematical models of mutualism vary along Levins' (1966) dimensions of realism, generality, and precision, which impact the extent to which data can be integrated. While we advocate for increased attention to realistic and precise—and hence parameterizable and empirically falsifiable—models, we

acknowledge the value of conceptual models in exploring the consequences of various simplifying assumptions. Servedio et al. (2014) describe the utility of such 'proof-of-concept' models in evolutionary biology and some of the ways in which data can be used in conjunction with models; these types of integration are mirrored in much of the mutualism modeling to date. Empirical data is commonly used in model construction and testing all biological levels, ranging from the data-rich genome-informed cell scale models to the natural history-based population and community scale models. Models that are high in the generality dimension are typically constructed to explain general features of mutualisms and primarily use empirical data to justify model assumptions (e.g., (Abbott *et al.*, 2015). These assumptions include both the exploratory and logistical assumptions that drive the construction of the model as well as the critical assumptions that the model is seeking to study (Servedio *et al.*, 2014). For example, if the host is an obligate mutualist, the host population should rapidly decline if the symbiont population drops below a threshold level (Marco *et al.*, 2009). This is a relatively simple principle, but the choice of threshold level and rate of host decline should be influenced by empirical studies for that system. If assumptions cannot be validated prior to model construction, modelers could declare these assumptions and suggest ways in which they can be empirically validated.

Empirical data can also be readily incorporated into realistic models that only contain parameters that are well-defined and can be directly measured (e.g., Franklin et al. 2014). These models can be quantitatively tested and thus in principle the models could be fully falsified. When models fail in some way, this indicates that one or more assumptions or deductive steps is unsound and prompts further model development or replacement. We acknowledge that this is an ambitious program, particularly at the community level where dozens of species are involved, but argue that lessons learned at lower biological scales can be translated to higher scales. Biological data can inform interconnections between models at different scales, and in some cases more abstract parameters, such as an individual's capacity to abstain from trade (Akçay & Simms, 2011), could potentially be translated into sub-models that only

contain biologically measurable parameters. We illustrate these principles and highlight avenues for the integration of data at each scale.

Integrating data into cell scale models

Cell scale models to date have been constructed for the rhizobial symbiosis, but the increasing number of well-annotated plant and microbial genomes (e.g., arbuscular mycorrhizal fungi, (Kuo *et al.*, 2013; Tisserant *et al.*, 2013) could be used to formulate cell scale models in other systems (Table 1.2.1). Rhizobial FBA models have thus far been validated by comparing the objective functions to empirical data on growth, e.g., maximizing nitrogen fixation versus maximizing biomass production (Resendis-Antonio *et al.*, 2011), but confirming specific predictions about patterns of metabolism and transport is most rigorously done by labeling studies (Chen *et al.*, 2011). Which specific hypothesis is being tested affects which biological scale and thus labeled substrate(s) is most appropriate to use. For example, to test the predicted carbon-phosphorus exchange rate in a plant-arbuscular mycorrhizal system, radioactive substrates could be used at an individual scale (e.g., (Zheng *et al.*, 2015). However, testing which metabolic pathways are active during trade should be performed at a cell scale with stable isotopes because together with nuclear magnetic resonance or mass spectrometry, they can yield sufficient information on the positional labeling of metabolites to quantify fluxes through different metabolic and transport routes. If labeling patterns at steady-state are not informative (e.g., carbon dioxide labeling in autotrophic systems yields uniform label distribution irrespective of the patterns of metabolic flux), then time-course labeling data can be used in isotopically dynamic flux analysis to obtain measurements of internal fluxes (Ma *et al.*, 2014). Dynamic flux analysis methods are an alternative modeling approach that could be used to explore non-steady-state conditions as well as potential regulatory mechanisms and physiological adaptation processes (Henson & Hanly, 2014). Finally, there is great potential for using cell scale models to predict metabolic changes in response to environmental or spatial conditions using individual scale

data. For example, nodule biomass has been measured in response to light and nutrient availability (Lau *et al.*, 2012). This data could serve as inputs for a validated cell scale FBA model to predict internal fluxes and the carbon-nitrogen exchange rate, thus linking the cellular and individual scales.

Integrating data into individual scale models

Many individual scale models are physiological models and contain biologically realistic parameters that have great potential to be integrated with detailed empirical data (Table 1.2.1), but doing so effectively requires appropriate quantification. For example, plant photosynthetic rates are a key parameter that determines the amount of carbon available for the plant to allocate to growth or trade (e.g., Grman et al. 2012). Photosynthetic rates are often quantified at a lower scale by measuring carbon uptake per leaf area and extrapolating to the individual scale, but this may grossly overestimate or underestimate carbon uptake depending on leaf age, angles, and self-shading. To avoid this scaling discrepancy, nutrient fluxes should be measured at an individual scale whenever possible, such as using photosynthetic chambers to measure whole-plant carbon uptake (Kölling *et al.*, 2015). Similarly, predictions concerning carbon-nutrient exchange rates are best tested at an individual scale with isotopic labeling, as described for cell scale models. However, measurements at the individual scale are not always appropriate for testing model predictions. For example, predictions of allocation strategies address metabolism within an individual and thus may require cell scale measurements (e.g., (Nord *et al.*, 2011). Conversely, predictions regarding mutualism stability are more suited to population scale testing because stability timescales exceed the lifespan of a single individual.

Integrating data into population scale models

Detailed empirical data is rarely explicitly incorporated into population scale models of mutualism, though this type of integration has a long history in predator-prey and competitive interactions (Laska & Wootton,

1998; Wootton & Emmerson, 2005). Interaction coefficients are key model components (Table 1.2.1), but are challenging to quantitatively test because they are not directly observable. The relative and qualitative strength of interaction coefficients could be assessed using the fairly simple method of growing different sets of interacting populations and estimating fitness. More quantitative estimations of interaction coefficients can be obtained by measuring intrinsic growth rate and population density in response to a range of partner densities (response-surface experimental design; (Inouye, 2001) followed by use of a general population dynamics model that mathematically relates these measurements to the interaction coefficient. Laska and Wootton (1998) evaluated the accuracy and empirical demand of four such models and found that the different models approach the idea of an interaction coefficient in different ways, including total population and per capita effects. Many population scale parameters (e.g., intrinsic growth rate) are best measured at a population scale, but some may require an individual scale approach. For example, population density can be measured at a population scale as the number of individuals in the population, but this approach may not be feasible for all systems (e.g., mycorrhizal fungi). Instead, modelers may recommend using the average individual biomass to estimate the population size (Neuhauser & Fargione, 2004) if the model applies to both interpretations (e.g., number of individuals and population biomass).

Integrating data into community scale models

To date, most community scale nutritional mutualism models were developed using assumptions derived from plant-animal mutualisms (e.g., plant-pollinators, (Georgelin & Loeuille, 2016). While some assumptions may hold true across mutualism types (e.g., only some hosts are obligate mutualists), others may differ in important ways. To have confidence in model predictions for a different system, it is important to empirically validate model components that may vary across systems, such as the cost-benefit ratios and the rates of dispersal. While independent validation of assumptions and predictions of

a model in many systems may be the ultimate goal, a more efficient approach might be to perform sensitivity analysis on the model component(s) in question. For example, this analysis may reveal that only the cost-benefit ratios in each system need to be measured in order to estimate differences in model predictions. Interaction coefficients are also key components in community scale models (Table 1.2.1) and can be quantitatively estimated in an approach similar to that proposed for population scale models (Laska & Wootton, 1998). At a community scale, the population of each species should be measured in response to independent fluctuations in the abundance of the other guilds. Alternatively, an appropriate allometric relationship, such as the correlation between interaction strength and body size, a metabolic rate, or another lower scale measurement, could be identified and then used as a proxy (Wootton & Emmerson, 2005). Although conducting such experiments in nature would integrate over all naturally occurring sources of variation, more rapid progress might be made by establishing carefully considered mesocosms that capture key aspects of the full diversity but allow for more precise control, measurement, and replication.

Conclusion

Parameterizable mathematical models can yield answers to fundamental questions about nutritional mutualism that are more comprehensive, specific, and mechanistic because they can examine questions from different biological perspectives. Together, data and modeling at different scales helps clarify how mutualisms function, both proximately through molecular and physiological mechanisms as well as ultimately through the selection pressures that shape evolutionary dynamics. The ecological implications of such nutrient exchanges can be dramatic—because individuals are more willing to trade when nutrients are scarce, mutualistic partners can persist in environments that are too harsh without trade (Neuhauser & Fargione, 2004; De Mazancourt & Schwartz, 2010). However, traits that attract potential partners can inadvertently attract antagonists, so community stability requires these positive and negative interactions

to be balanced (López-Ráez *et al.*, 2011; Georgelin & Loeuille, 2016). In addition, modeling at different scales reveals that limited dispersal can promote mutualism persistence because it enables cooperators to cluster, which increases both partners' fitness and maintains population densities that enable persistence in the face of competitively superior defectors (Yamamura *et al.*, 2004; Abbott *et al.*, 2015). Furthermore, in plant-microbe nutritional mutualisms characterized to date, spatially structured interactions lead to specialists interacting with subsets of generalists less frequently than expected by chance, which may be due to interspecific competition affecting the range of potential partners (Toju *et al.*, 2015).

Although 'proof-of-concept' models can be used to generate conceptual insights, data integration greatly strengthens models' predictive power by providing a means for rigorous, quantitative testing of the hypotheses they embody. We believe that the modeling and data integration approaches advocated here will deepen our understanding of a broad range of mutualistic systems. For example, models at different biological scales can give complementary insights into how nutrient availability, structured interactions, and other factors influence mutualism persistence and function. These scales range from the enzymatic reactions that underpin mutualism to the role of mutualism at community and ecosystem scales. We envision a future in which mathematical models are frequently developed with anticipation of empirical validation and are used iteratively to enable an integrative understanding of these crucial species interactions.

Acknowledgments

We wish to thank Erol Akçay, Gijsbert Werner, and one anonymous reviewer for thoughtful critiques and suggestions that improved the manuscript. In addition, Emily Jones and Jason Hoeksema gave helpful comments on an earlier version of the manuscript, and Chris Klausmeier provided helpful discussion that led to this review. We acknowledge funding support from NSF DEB 1354878 and NSF IOS 1342793 to MLF.

This material is based in part upon work supported by the National Science Foundation under Cooperative Agreement No. DBI-0939454 and partial support from a fellowship from Michigan State University under the National Institutes of Health Training Program in Plant Biotechnology for Health and Sustainability (T32-GM110523). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors do not have any conflicts of interest to declare.

Box 1 - Methods in Cell Scale Models

Cell scale models representing an organism's metabolism begin as systems of rate equations (i.e., coupled linear ordinary differential equations), with each equation representing the rate of change of the level of a metabolite in the system in terms of the rates of the biochemical and transport reactions that affect it (Palsson, 2015). The elemental composition of each metabolite is included, so that balances of carbon and mineral nutrients are quantified. Genome scale models covering the complete potential metabolic network of a cell are assembled using published literature along with annotated genomic and protein databases (Resendis-Antonio *et al.*, 2007; Rodriguez-Llorente *et al.*, 2009). In the steady-state, flux balance analysis (FBA) models can be used to investigate the potential of a metabolic network for growth, maintenance, or production of desired compounds when the system is constrained by substrate inputs and a hypothesized "objective" function, which is maximized or minimized. Because FBA models contain hundreds to thousands of reactions and comparatively few measured parameters (e.g., growth rate), if the system was not constrained by an objective function, there would be an almost unlimited range of possible metabolic patterns compatible with the measurements.

The most commonly assumed objective function in cell scale models of any system is maximal growth under carbon-limiting conditions, which corresponds to maximal substrate use efficiency for the system as a whole (Chen *et al.*, 2011). However, this objective function is unsuitable for organisms or

organs that cease growing at maturity. This is an important distinction in plant-microbe mutualisms because rhizobia occupy nodules that are either indeterminate or determinate, and thus either grow continuously or cease growing at maturity, respectively (Udvardi & Poole, 2013). Consequently, either the objective function needs to be specific for the type of nodule the rhizobia inhabits, or the objective function needs to be designed to not depend on the rhizobia growing through maturity. Cell scale models that use FBA commonly seek to identify an appropriate objective function to constrain the metabolic patterns for the system (Resendis-Antonio *et al.*, 2011; Zhao *et al.*, 2012). This can be used to focus the modeling results on processes of interest. For example, the objective function maximizing nitrogen fixation by the rhizobial symbiont and its transfer to the plant in return for a given amount of carbon-containing metabolites could be used in future modeling efforts to reveal the maximal carbon-nitrogen exchange rate, and the genes and pathway activities needed to realize it.

Box 2 - Methods in Individual Scale Models

Modeling mutualistic interactions on an individual scale is often conducted using game theory, which explores situations where the fitness costs and benefits of a particular strategy depend on the prevalence of other interacting strategies. Stemming from game theory is biological market theory, which emphasizes market mechanisms such as the ability to select partners (Noë & Hammerstein, 1994) and establish exchange rates based on bargaining (Akçay & Roughgarden, 2007; Akçay & Simms, 2011). Game theory models typically have at their core a payoff matrix describing the fitness outcome from interacting with a cooperator or defector that is derived from the costs and benefits of each type of interaction (Yamamura *et al.*, 2004); continuous strategies can also be modeled with a matrix of functions rather than fixed values (Doebeli & Knowlton, 1998; Doebeli & Hauert, 2005). Mutualisms can be modeled using a variety of games, including the Prisoner's Dilemma (Doebeli & Knowlton, 1998) and Public Goods Games (Archetti & Scheuring, 2013). Some individual scale models use optimality assumptions derived from evolutionary

reasoning to solve instantaneous decision problems with empirically defined parameters. These parameters are often physiological descriptions, such as the ability of each partner to obtain or use nutrients, which can be interpreted in terms of nutrient uptake and exchange rates (Cowden & Peterson, 2009; Franklin *et al.*, 2014). In some cases, the assumption that bargaining leads to a Nash equilibrium, where the product of partners' fitness is maximized, is used as a constraint to allow a unique solution for a set of equations (Grman *et al.*, 2012).

Box 3 - Methods in Population Scale Models

Most population scale mutualism models take a population dynamics mathematical approach by tracking species distribution and abundance through consideration of births, deaths, immigration, and emigration. Population dynamics can be captured using ordinary differential equations or discrete-time models, and adaptive dynamics extends these equations by enabling particular ecological parameters to evolve (Dercole & Rinaldi, 2008). Many population scale models modify Lotka-Volterra predator-prey models to include positive interaction coefficients due to resource exchange (Vandermeer & Boucher, 1978; Neuhauser & Fargione, 2004). Other population scale models build on resource ratio theory, i.e., R*-style models (Tilman, 1982), to explicitly track the depletion and exchange of resources. These models predict whether coexistence is feasible based on the availability of resources and nutritive demands of each species, thus providing an additional level of mechanistic detail that can be empirically measured (De Mazancourt & Schwartz, 2010). Spatially structured interactions can be modeled through the use of dual-lattice structure (Parker, 1999) or a patch environment (Abbott *et al.*, 2015). Under either mechanism, populations can interact based on dispersal distances and/or rates within a population dynamics model.

Box 4 - Methods in Community Scale Models

Population dynamics and/or network theory methods are typically used to answer questions related to

community stability and biodiversity (Golinski, 2006; Bascompte, 2009). Both methods rely on defining the types of interactions that connect pairs of species. Population dynamics methods often approach community scale questions by examining how the type and magnitude of interactions affect rates of change in population abundance in the presence of biotic and abiotic factors. At the community scale, this includes interactions between all species (Bachelot *et al.*, 2015). Furthermore, an adaptive dynamics approach, where population dynamics equations are incorporated into evolutionary models, can analyze the evolutionary basis for community assembly and diversity (Georgelin & Loeuille, 2016).

In an ecological network framework, the interactions connecting pairs of species can be visualized with each node representing a population and the links between these nodes representing interactions between species. Network theory methods analyze the interplay between interaction characteristics and varying network structure. This approach can use population dynamics models to represent links between two species, but in addition to including one equation per species (McQuaid & Britton, 2013), these models often include one equation per link in the community network to study network structure (Bastolla *et al.*, 2009). Consequently, by altering the network structure, the sum of interactions between species is changed. These models often examine network characteristics such as degree distribution (the number of links a given species is associated with) and nestedness, which occurs when specialists interact with a subset of those with which generalists interact (Fontaine *et al.*, 2011). Community network models can also be used in adaptive evolutionary frameworks with a relevant optimization criterion (e.g., maximize species abundance) to measure properties associated with optimized networks, such as network resilience or nestedness (Suweis *et al.*, 2013).

REFERENCES

REFERENCES

Abbott KC, Karst J, Biederman LA, Borrett SR, Hastings A, Walsh V, Bever JD. 2015. Spatial heterogeneity in soil microbes alters outcomes of plant competition. *PloS one* 10(5): e0125788.

Akçay E, Roughgarden J. 2007. Negotiation of mutualism: rhizobia and legumes. *Proceedings of the Royal Society of London B: Biological Sciences* 274(1606): 25-32.

Akçay E, Simms EL. 2011. Negotiation, sanctions, and context dependency in the legume-rhizobium mutualism. *The American Naturalist* 178(1): 1-14.

Archetti M, Scheuring I. 2013. Trading public goods stabilizes interspecific mutualism. *Journal of Theoretical Biology* 318: 58-67.

Archetti M, Úbeda F, Fudenberg D, Green J, Pierce NE, Douglas WY. 2011. Let the right one in: a microeconomic approach to partner choice in mutualisms. *The American Naturalist* 177(1): 75-85.

Bachelot B, Uriarte M, McGuire K. 2015. Interactions among mutualism, competition, and predation foster species coexistence in diverse communities. *Theoretical Ecology* 8(3): 297-312.

Bago B, Pfeffer PE, Shachar-Hill Y. 2000. Carbon metabolism and transport in arbuscular mycorrhizas. *Plant physiology* 124(3): 949-958.

Barrett LG, Bever JD, Bissett A, Thrall PH. 2015. Partner diversity and identity impacts on plant productivity in Acacia–rhizobial interactions. *Journal of Ecology* 103(1): 130-142.

Bascompte J. 2009. Mutualistic networks. *Frontiers in Ecology and the Environment* 7(8): 429-436.

Bastolla U, Fortuna MA, Pascual-García A, Ferrera A, Luque B, Bascompte J. 2009. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* 458(7241): 1018-1020.

Bever JD. 2015. Preferential allocation, physio-evolutionary feedbacks, and the stability and environmental patterns of mutualism between plants and their root symbionts. *New Phytologist* 205(4): 1503-1514.

Binmore K, Rubinstein A, Wolinsky A. 1986. The Nash bargaining solution in economic modelling. *The RAND Journal of Economics*: 176-188.

Bloom AJ, Chapin III FS, Mooney HA. 1985. Resource limitation in plants-an economic analogy. *Annual review of Ecology and Systematics* 16(1): 363-392.

Bronstein JL. 2015. *Mutualism*: Oxford University Press, USA.

Bücking H, Mensah JA, Fellbaum CR. 2016. Common mycorrhizal networks and their effect on the

bargaining power of the fungal partner in the arbuscular mycorrhizal symbiosis. *Communicative & integrative biology* 9(1): e1107684.

Burnham KP, Anderson DR. 2003. *Model selection and multimodel inference: a practical information-theoretic approach*: Springer Science & Business Media.

Busby PE, Soman C, Wagner MR, Friesen ML, Kremer J, Bennett A, Morsy M, Eisen JA, Leach JE, Dangl JL. 2017. Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS biology* 15(3): e2001793.

Chae S, Heidhues P. 2001. Nash bargaining solution with coalitions and the joint bargaining paradox: WZB Discussion Paper.

Chakraborty T. 2011. *Bargaining and pricing in networked economic systems.* University of Pennsylvania.

Chakraborty T, Kearns M, Khanna S 2009. Network bargaining: algorithms and structural results. *Proceedings of the 10th ACM conference on Electronic commerce*: ACM. 159-168.

Chamberlain SA, Bronstein JL, Rudgers JA. 2014. How context dependent are species interactions? *Ecology Letters* 17(7): 881-890.

Chen X, Alonso AP, Allen DK, Reed JL, Shachar-Hill Y. 2011. Synergy between 13 C-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in E. coli. *Metabolic engineering* 13(1): 38-48.

Clark TJ, Friel CA, Grman E, Shachar-Hill Y, Friesen ML. 2017. Modelling nutritional mutualisms: challenges and opportunities for data integration. *Ecology letters* 20(9): 1203-1215.

Cole JA, Kohler L, Hedhli J, Luthey-Schulten Z. 2015. Spatially-resolved metabolic cooperativity within dense bacterial colonies. *BMC systems biology* 9(1): 1.

Cowden CC, Peterson CJ. 2009. A multi-mutualist simulation: applying biological market models to diverse mycorrhizal communities. *Ecological Modelling* 220(12): 1522-1533.

Das K, Prasanna R, Saxena AK. 2017. Rhizobia: a potential biocontrol agent for soilborne fungal pathogens. *Folia microbiologica* 62(5): 425-435.

Daubech B, Remigi P, de Moura GD, Marchetti M, Pouzet C, Auriac M-C, Gokhale CS, Masson-Boivin C, Capela D. 2017. Spatio-temporal control of mutualism in legumes helps spread symbiotic nitrogen fixation. *Elife* 6.

Dean JM, Mescher MC, De Moraes CM. 2014. Plant dependence on rhizobia for nitrogen influences induced plant defenses and herbivore performance. *International journal of molecular sciences* 15(1): 1466-1480.

De Mazancourt C, Schwartz MW. 2010. A resource ratio theory of cooperation. *Ecology Letters* 13(3): 349-359.

Dercole F, Rinaldi S. 2008. *Analysis of evolutionary processes: the adaptive dynamics approach and its*

applications: the adaptive dynamics approach and its applications: Princeton University Press.

Doebeli M, Hauert C. 2005. Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. *Ecology Letters* 8(7): 748-766.

Doebeli M, Knowlton N. 1998. The evolution of interspecific mutualisms. *Proceedings of the National Academy of Sciences* 95(15): 8676-8680.

Fåhraeus G. 1957. The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique. *Microbiology* 16(2): 374-381.

Fontaine C, Guimarães PR, Kéfi S, Loeuille N, Memmott J, van Der Putten WH, van Veen FJ, Thebault E. 2011. The ecological and evolutionary implications of merging different types of networks. *Ecology Letters* 14(11): 1170-1181.

Fowler D, Coyle M, Skiba U, Sutton MA, Cape JN, Reis S, Sheppard LJ, Jenkins A, Grizzetti B, Galloway JN, et al. 2013. The global nitrogen cycle in the twenty-first century. *Phil. Trans. R. Soc. B* 368(1621): 20130164.

Franklin O, Näsholm T, Högberg P, Högberg MN. 2014. Forests trapped in nitrogen limitation–an ecological market perspective on ectomycorrhizal symbiosis. *New Phytologist* 203(2): 657-666.

Frederickson ME. 2013. Rethinking mutualism stability: cheaters and the evolution of sanctions. *The Quarterly review of biology* 88(4): 269-295.

Friesen ML. 2012. Widespread fitness alignment in the legume–rhizobium symbiosis. *New Phytologist* 194(4): 1096-1111.

Friesen ML, Heath KD. 2013. One hundred years of solitude: integrating single-strain inoculations with community perspectives in the legume–rhizobium symbiosis. *New Phytologist* 198(1): 7-9.

Gause G, Witt A. 1935. Behavior of mixed populations and the problem of natural selection. *The American Naturalist* 69(725): 596-609.

Georgelin E, Loeuille N. 2016. Evolutionary response of plant interaction traits to nutrient enrichment modifies the assembly and structure of antagonistic-mutualistic communities. *Journal of Ecology* 104(1): 193-205.

Ghoul M, Griffin AS, West SA. 2014. Toward an evolutionary definition of cheating. *Evolution* 68(2): 318-331.

Goh CH, Nicotra AB, Mathesius U. 2016. The presence of nodules on legume root systems can alter phenotypic plasticity in response to internal nitrogen independent of nitrogen fixation. *Plant, cell & environment* 39(4): 883-896.

Golinski MR. 2006. A review of theoretical approaches for studying the effects of interactions between mutualists and nonmutualists on community stability.

Grafahrend-Belau E, Junker A, Eschenröder A, Müller J, Schreiber F, Junker BH. 2013. Multiscale metabolic modeling: dynamic flux balance analysis on a whole-plant scale. *Plant physiology* 163(2): 637-647.

Grman E, Robinson TM, Klausmeier CA. 2012. Ecological specialization and trade affect the outcome of negotiations in mutualism. *The American Naturalist* 179(5): 567-581.

Gross K. 2008. Positive interactions among competitors can produce species-rich communities. *Ecology Letters* 11(9): 929-936.

Guimaraes Jr PR, Jordano P, Thompson JN. 2011. Evolution and coevolution in mutualistic networks. *Ecology Letters* 14(9): 877-885.

Harris D, Pacovsky R, Paul E. 1985. Carbon economy of soybean–Rhizobium–Glomus associations. *New phytologist* 101(3): 427-440.

Hauert C, Doebeli M. 2004. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* 428(6983): 643-646.

Henson MA, Hanly TJ. 2014. Dynamic flux balance analysis for synthetic microbial communities. *IET systems biology* 8(5): 214-229.

Herridge DF, Peoples MB, Boddey RM. 2008. Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil* 311(1-2): 1-18.

Inouye BD. 2001. Response surface experimental designs for investigating interspecific competition. *Ecology* 82(10): 2696-2706.

Jamil M, Charnikhova T, Cardoso C, Jamil T, Ueno K, Verstappen F, Asami T, Bouwmeester H. 2011. Quantification of the relationship between strigolactones and Striga hermonthica infection in rice under varying levels of nitrogen and phosphorus. *Weed research* 51(4): 373-385.

Johnson N, Graham JH, Smith F. 1997. Functioning of mycorrhizal associations along the mutualism–parasitism continuum. *New Phytologist* 135(4): 575-585.

Johnson NC. 2010. Resource stoichiometry elucidates the structure and function of arbuscular mycorrhizas across scales. *New Phytologist* 185(3): 631-647.

Jones EI, Afkhami ME, Akçay E, Bronstein JL, Bshary R, Frederickson ME, Heath KD, Hoeksema JD, Ness JH, Pankey MS, et al. 2015. Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualism. *Ecology Letters* 18(11): 1270-1284.

Kawamori T. 2014. A noncooperative foundation of the asymmetric Nash bargaining solution. *Journal of Mathematical Economics* 52: 12-15.

Kiers ET, Duhamel M, Beesetty Y, Mensah JA, Franken O, Verbruggen E, Fellbaum CR, Kowalchuk GA, Hart MM, Bago A, et al. 2011. Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *science* 333(6044): 880-882.

Kiers ET, Hutton MG, Denison RF. 2007. Human selection and the relaxation of legume defences against ineffective rhizobia. *Proceedings of the Royal Society of London B: Biological Sciences* 274(1629): 3119-3126.

Kölling K, George GM, Künzli R, Flütsch P, Zeeman SC. 2015. A whole-plant chamber system for parallel gas exchange measurements of Arabidopsis and other herbaceous species. *Plant methods* 11(1): 1.

Kraiser T, Gras DE, Gutiérrez AG, González B, Gutiérrez RA. 2011. A holistic view of nitrogen acquisition in plants. *Journal of Experimental Botany* 62(4): 1455-1466.

Kuo A, Kohler A, Martin FM, Grigoriev IV. 2013. Expanding genomics of mycorrhizal symbiosis. *Frontiers in microbiology* 5: 582-582.

Lambers H, Szaniawski RK, Visser R. 1983. Respiration for growth, maintenance and ion uptake. An evaluation of concepts, methods, values and their significance. *Physiologia plantarum* 58(4): 556-563.

Laska MS, Wootton JT. 1998. Theoretical concepts and empirical approaches to measuring interaction strength. *Ecology* 79(2): 461-476.

Lau JA, Bowling EJ, Gentry LE, Glasser PA, Monarch EA, Olesen WM, Waxmonsky J, Young RT. 2012. Direct and interactive effects of light and nutrients on the legume-rhizobia mutualism. *Acta oecologica* 39: 80-86.

López-Ráez JA, Pozo MJ, García-Garrido JM. 2011. Strigolactones: a cry for help in the rhizosphere. *Botany* 89(8): 513-522.

Ma F, Jazmin LJ, Young JD, Allen DK. 2014. Isotopically nonstationary 13C flux analysis of changes in Arabidopsis thaliana leaf metabolism due to high light acclimation. *Proceedings of the National Academy of Sciences* 111(47): 16967-16972.

Marco DE, Carbajal JP, Cannas S, Pérez-Arnedo R, Hidalgo-Perea Á, Olivares J, Ruiz-Sainz JE, Sanjuán J. 2009. An experimental and modelling exploration of the host-sanction hypothesis in legume–rhizobia mutualism. *Journal of Theoretical Biology* 259(3): 423-433.

McQuaid CF, Britton NF. 2013. Network dynamics contribute to structure: nestedness in mutualistic networks. *Bulletin of mathematical biology* 75(12): 2372-2388.

Moeller HV, Neubert MG, Klausmeier CA, Bronstein JL. 2016. Multiple Friends with Benefits: An Optimal Mutualist Management Strategy? *The American Naturalist* 187(1): E1-E12.

Mougi A, Kondoh M. 2012. Diversity of interaction types and ecological community stability. *Science* 337(6092): 349-351.

Nash Jr JF. 1950. The bargaining problem. *Econometrica: Journal of the Econometric Society*: 155-162.

Näsholm T, Kielland K, Ganeteg U. 2009. Uptake of organic nitrogen by plants. *New phytologist* 182(1):

31-48.

Neuhauser C, Fargione JE. 2004. A mutualism–parasitism continuum model and its application to plant–mycorrhizae interactions. *Ecological Modelling* 177(3): 337-352.

Noë R, Hammerstein P. 1994. Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral ecology and sociobiology* 35(1): 1-11.

Nord EA, Shea K, Lynch JP. 2011. Optimizing reproductive phenology in a two-resource world: a dynamic allocation model of plant growth predicts later reproduction in phosphorus-limited plants. *Annals of botany* 108(2): 391-404.

Palsson B. 2015. *Systems biology*: Cambridge university press.

Paponov I, Posepanov O, Lebedinskai S, Koshkin E. 2000. Growth and biomass allocation, with varying nitrogen availability, of near-isogenic pea lines with differing foliage structure. *Annals of botany* 85(4): 563-569.

Parker MA. 1999. Mutualism in metapopulations of legumes and rhizobia. *The American Naturalist* 153(S5): S48-S60.

Paszkowski U. 2006. Mutualism and parasitism: the yin and yang of plant symbioses. *Current opinion in plant biology* 9(4): 364-370.

Perret X, Staehelin C, Broughton WJ. 2000. Molecular basis of symbiotic promiscuity. *Microbiology and Molecular Biology Reviews* 64(1): 180-201.

Phillips DA. 1980. Efficiency of symbiotic nitrogen fixation in legumes. *Annual Review of Plant Physiology* 31(1): 29-49.

Pillai P, Gouhier TC, Vollmer SV. 2014. The cryptic role of biodiversity in the emergence of host–microbial mutualisms. *Ecology Letters* 17(11): 1437-1446.

Rao TP, Ito O, Matsunga R. 1993. Differences in uptake kinetics of ammonium and nitrate in legumes and cereals. *Plant and Soil* 154(1): 67-72.

Ratcliffe RG, Shachar-Hill Y. 2006. Measuring multiple fluxes through plant metabolic networks. *The Plant Journal* 45(4): 490-511.

Reeve W, Chain P, O'Hara G, Ardley J, Nandesena K, Bräu L, Tiwari R, Malfatti S, Kiss H, Lapidus A, et al. 2010. Complete genome sequence of the Medicago microsymbiont Ensifer (Sinorhizobium) medicae strain WSM419. *Standards in genomic sciences* 2(1): 77.

Regus J, Wendlandt C, Bantay R, Gano-Cohen K, Gleason N, Hollowell A, O'Neill M, Shahin K, Sachs J. 2017. Nitrogen deposition decreases the benefits of symbiosis in a native legume. *Plant and Soil* 414(1-2): 159-170.

Regus JU, Quides KW, O'Neill MR, Suzuki R, Savory EA, Chang JH, Sachs JL. 2017. Cell autonomous sanctions in legumes target ineffective rhizobia in nodules with mixed infections. *American*

Journal of Botany 104(9): 1299-1312.

Resendis-Antonio O, Hernández M, Salazar E, Contreras S, Batallar GM, Mora Y, Encarnación S. 2011. Systems biology of bacterial nitrogen fixation: high-throughput technology and its integrative description with constraint-based modeling. *BMC systems biology* 5(1): 120.

Resendis-Antonio O, Reed JL, Encarnación S, Collado-Vides J, Palsson BØ. 2007. Metabolic reconstruction and modeling of nitrogen fixation in Rhizobium etli. *PLoS Comput Biol* 3(10): e192.

Rodriguez-Llorente I, Caviedes MA, Dary M, Palomares AJ, Cánovas FM, Peregrín-Alvarez JM. 2009. The Symbiosis Interactome: a computational approach reveals novel components, functional interactions and modules in Sinorhizobium meliloti. *BMC systems biology* 3(1): 63.

Rufty Jr TW, Raper Jr CD, Huber SC. 1984. Alterations in internal partitioning of carbon in soybean plants in response to nitrogen stress. *Canadian Journal of Botany* 62(3): 501-508.

Ryle G, Arnott R, Powell C, Gordon A. 1984. N2 fixation and the respiratory costs of nodules, nitrogenase activity, and nodule growth and maintenance in Fiskeby soyabean. *Journal of Experimental Botany* 35(8): 1156-1165.

Sachs J, Ehinger M, Simms E. 2010. Origins of cheating and loss of symbiosis in wild Bradyrhizobium. *Journal of evolutionary biology* 23(5): 1075-1089.

Sachs JL, Mueller UG, Wilcox TP, Bull JJ. 2004. The evolution of cooperation. *The Quarterly Review of Biology* 79(2): 135-160.

Sachs JL, Skophammer RG, Regus JU. 2011. Evolutionary transitions in bacterial symbiosis. *Proceedings of the National Academy of Sciences* 108(Supplement 2): 10800-10807.

Schwartz MW, Hoeksema JD. 1998. Specialization and resource trade: biological markets as a model of mutualisms. *Ecology* 79(3): 1029-1038.

Servedio MR, Brandvain Y, Dhole S, Fitzpatrick CL, Goldberg EE, Stern CA, Van Cleve J, Yeh DJ. 2014. Not just a theory—the utility of mathematical models in evolutionary biology. *PLoS biology* 12(12): e1002017.

Simms EL, Taylor DL. 2002. Partner choice in nitrogen-fixation mutualisms of legumes and rhizobia. *Integrative and Comparative Biology* 42(2): 369-380.

Simonsen AK, Stinchcombe JR. 2014. Herbivory eliminates fitness costs of mutualism exploiters. *New Phytologist* 202(2): 651-661.

Stanton-Geddes J, Anderson CG. 2011. Does a facultative mutualism limit species range expansion? *Oecologia* 167(1): 149-155.

Suweis S, Simini F, Banavar JR, Maritan A. 2013. Emergence of structural and dynamical properties of ecological mutualistic networks. *Nature* 500(7463): 449-452.

Thébault E, Fontaine C. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* 329(5993): 853-856.

Tilman D. 1982. *Resource competition and community structure*: Princeton university press.

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing N, dit Frey NF, Gianinazzi-Pearson V. 2013. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proceedings of the National Academy of Sciences* 110(50): 20117-20122.

Toju H, Guimarães PR, Olesen JM, Thompson JN. 2015. Below-ground plant–fungus network topology is not congruent with above-ground plant–animal network topology. *Science advances* 1(9): e1500291.

Udvardi M, Poole PS. 2013. Transport and metabolism in legume-rhizobia symbioses. *Annual review of plant biology* 64: 781-805.

Valladares F, Villar-Salvador P, Domínguez S, Fernández-Pascual M, Peñuelas JL, Pugnaire FI. 2002. Enhancing the early performance of the leguminous shrub Retama sphaerocarpa (L.) Boiss.: fertilisation versus Rhizobium inoculation. *Plant and Soil* 240(2): 253-262.

Vandermeer JH, Boucher DH. 1978. Varieties of mutualistic interaction in population models. *Journal of Theoretical Biology* 74(4): 549-558.

Vannette RL, Hunter MD. 2013. Mycorrhizal abundance affects the expression of plant resistance traits and herbivore performance. *Journal of Ecology* 101(4): 1019-1029.

Vázquez DP, Blüthgen N, Cagnolo L, Chacoff NP. 2009. Uniting pattern and process in plant–animal mutualistic networks: a review. *Annals of botany* 103(9): 1445-1457.

Verbruggen E, El Mouden C, Jansa J, Akkermans G, Bücking H, West SA, Kiers ET. 2012. Spatial structure and interspecific cooperation: theory and an empirical test using the mycorrhizal mutualism. *The American Naturalist* 179(5): E133-E146.

Voisin A-S, Salon C, Munier-Jolain NG, Ney B. 2002. Effect of mineral nitrogen on nitrogen nutrition and biomass partitioning between the shoot and roots of pea (Pisum sativum L.). *Plant and Soil* 242(2): 251-262.

Wang X, Pan Q, Chen F, Yan X, Liao H. 2011. Effects of co-inoculation with arbuscular mycorrhizal fungi and rhizobia on soybean growth as related to root architecture and availability of N and P. *Mycorrhiza* 21(3): 173-181.

Wardlaw IF. 1990. Tansley Review No. 27 The control of carbon partitioning in plants. *New phytologist* 116(3): 341-381.

Weese DJ, Heath KD, Dentinger BT, Lau JA. 2015. Long-term nitrogen addition causes the evolution of less-cooperative mutualists. *Evolution* 69(3): 631-642.

Werner GD, Cornwell WK, Cornelissen JH, Kiers ET. 2015. Evolutionary signals of symbiotic persistence in

the legume–rhizobia mutualism. *Proceedings of the National Academy of Sciences*: 201424030.

Werner GD, Strassmann JE, Ivens AB, Engelmoer DJ, Verbruggen E, Queller DC, Noë R, Johnson NC, Hammerstein P, Kiers ET. 2014. Evolution of microbial markets. *Proceedings of the National Academy of Sciences* 111(4): 1237-1244.

West SA, Kiers ET, Simms EL, Denison RF. 2002. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proceedings of the Royal Society of London B: Biological Sciences* 269(1492): 685-694.

Weyl EG, Frederickson ME, Douglas WY, Pierce NE. 2010. Economic contract theory tests models of mutualism. *Proceedings of the National Academy of Sciences* 107(36): 15712-15716.

Wolf AA, Funk JL, Menge DN. 2017. The symbionts made me do it: legumes are not hardwired for high nitrogen concentrations but incorporate more nitrogen when inoculated. *New phytologist* 213(2): 690-699.

Wootton JT, Emmerson M. 2005. Measurement of interaction strength in nature. *Annual Review of Ecology, Evolution, and Systematics*: 419-444.

Yamamura N, Higashi M, Behera N, Wakano JY. 2004. Evolution of mutualism through spatial effects. *Journal of Theoretical Biology* 226(4): 421-428.

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H, et al. 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 480(7378): 520.

Younginger BS, Sirová D, Cruzan MB, Ballhorn DJ. 2017. Is biomass a reliable estimate of plant fitness? *Applications in plant sciences* 5(2): 1600094.

Zhao H, Li M, Fang K, Chen W, Wang J. 2012. In silico insights into the symbiotic nitrogen fixation in Sinorhizobium meliloti via metabolic reconstruction. *PloS one* 7(2): e31287.

Zheng C, Ji B, Zhang J, Zhang F, Bever JD. 2015. Shading decreases plant carbon preferential allocation towards the most beneficial mycorrhizal mutualist. *New Phytologist* 205(1): 361-368.

CHAPTER 2

Steady-state metabolic flux analysis

PREFACE

The Shachar-Hill lab is interested in rigorously testing the topology of networks in central metabolism and quantitatively analyzing how their fluxes are influenced by a variety of circumstances. Before I joined the lab, they had already started investigating why *Camelina sativa* embryos had poor carbon use efficiency and how transgenic *C. sativa* lines metabolically respond to the accumulation of medium-chain fatty acids. To answer these questions, my lab performed labeling experiments and other analyses on developing *C. sativa* embryos to obtain the data necessary for ^{13}C-derived metabolic flux analysis (MFA). MFA is largely centered on the use of an optimization program, where the researcher proposes a network topology and provides data that the program can use to calculate the most likely flux values for that network. I continued the computational MFA work that had been started by Rahul R. Deshpande (RD) and have incorporated other analyses to strengthen our conclusions.

In Section 2.1, I present the results of my contributions to using MFA on embryos grown at three light levels to understand why *C. sativa* embryos have low carbon use efficiency. In Section 2.2, I present the current draft of our experimental and modeling manuscript, where we Identified metabolic changes in *C. sativ*a embryos that had been engineered to accumulate medium-chain fatty acids.

These investigations highlight the potential of analytical models to reveal effects in complex metabolic networks due to environmental and genetic changes.

SECTION 2.1

Analysis of low biosynthetic efficiency in an oilseed using Metabolic Flux Analysis

---

This section includes text being drafted for the manuscript:
Carey LM, TJ Clark, RR Deshpande, JC Cocuron, EK Rustad, and Y Shachar-Hill. Analysis of low biosynthetic efficiency in an oilseed using Metabolic Flux Analysis.

Preface

When I joined the Shachar-Hill lab, my lab mates had already completed the experiments needed to use MFA to analyze carbon use efficiency in *Camelina sativa* developing embryos. They had also already found strong evidence that the inefficiency was due to a high oxidative pentose phosphate (OPP) decarboxylation flux.

My initial role was to perform MFA only on the four transgenic *C. sativa* lines that were engineered to accumulate medium-chain fatty acids. Due to some challenges in the transgenic MFA work (described in the Section 2.2 Preface), I added several constraints to our MFA model and re-analyzed the data for investigating the low carbon use efficiency. These constraints included 1) preventing the model from distinguishing between glucose substrates in the uptake reaction, 2) ensuring there is a mass balance of how much nitrogen the embryos are predicted to take up and use, and 3) loosely restricting the predicted stoichiometry of the glyceryl and acyl precursors synthesized to be consistent with that of the dominant lipid class in the embryo (i.e., triacylglycerol, which has a 1:3 ratio).

In addition to these constraints, I devised a method to determine flux confidence intervals for all embryo sets (wild-type under three light levels and four transgenics) using Monte Carlo sampling. Briefly, I modified an existing Perl script to execute the MFA program and, for each embryo set, I used MATLAB to generate potential starting points, Excel to generate pseudo datasets, a batch server to write and execute the MFA program scripts in parallel, and Python to collect the necessary output data. Even with the constraints, the model had a high chance of not being able to refine the stating point fluxes because the network was so complex. Therefore, I had to use 9,000 to 130,000 starting points per embryo set to generate at least 100 successful refinements to verify the global best fit was obtained. The best fit told us the flux values that best simultaneously explained the data and network. I then used the starting points for the top 100 fits per embryo set to model 50 pseudo datasets to obtain the range of flux predictions that could be derived from reasonable variations in our data.

RD had tested different network topologies before I joined the lab, but because I revised the model to have additional constraints, I re-tested the proposed topologies to verify our network was still supported by our method. The topologies I tested included if the hexose phosphate pool used in the OPP reaction was cytoplasmic or plastidic, and if there was an active rubisco bypass. I tested these topologies by comparing, respectively, the model fits with either sub-cellular pool location, and with and without the bypass.

In this section, I present one motivation for investigating carbon use efficiency in seeds, more details on how I contributed to strengthening our MFA model, and the resulting findings from my analyses on *C. sativa* embryos grown under three light levels.

Introduction

There is interest in engineering seeds to have greater biomass because larger seeds have been found to produce seedlings that grow faster or larger than those from smaller seeds (Howe & Richter, 1982; Stanton, 1984), and various seed compounds are of significant economic value to pharmaceutical and chemical companies (Howard, 2009; Howard, 2015). One factor that influences seed size is carbon use efficiency, i.e., the percentage of carbon received by the seed from the mother plant that is converted into biomass (Chen & Shachar-Hill, 2012). Oil seeds in particular are at risk for having low carbon use efficiency because their oil (e.g., triacylglycerol) requires substantial production of fatty acids from acetyl CoA, which is made from the decarboxylation of pyruvate (Bates *et al.*, 2013). Consequently, a third of the glycolytic carbon used for fatty acid synthesis is converted into carbon dioxide, and oilseed *Brassica napus* has been found to accumulate 2000-fold more carbon dioxide in its developing seeds in comparison to ambient air or concentrations found in leaves (Goffman *et al.,* 2004).

Internally produced carbon dioxide can be recovered by enzymes such as phosphoenolpyruvate carboxylase, which synthesizes oxaloacetate (Vennesland *et al.*, 1954). Green seeds have active photosystems (Eastmond *et al.*, 1996) and thus rubisco can operate without the Calvin cycle to assimilate carbon into triose precursors from ribulose-1,5-bisphosphate in a metabolic route known as the rubisco bypass (Schwender *et al.*, 2004a). The bypass likely contributes to green seeds generally having greater carbon use efficiency than other seeds (Chen & Shachar-Hill, 2012) because the bypass allows 40% less carbon dioxide to be lost during fatty acid synthesis (Schwender *et al.*, 2004a). However, *Camelina sativa*, a promising oilseed crop (Moser, 2010), has unusually low carbon use efficiency for a green seed (Chen & Shachar-Hill, 2012), which could be due to low rubisco activity, futile carbon cycling, or other deficiencies. The effects of different metabolic features on seed efficiencies have been assessed in maize (Alonso *et al.*, 2011) and brassica (Schwender *et al.*, 2006) by quantifying central metabolism fluxes with steady-state ^{13}C metabolic flux analysis (MFA).

The process by which MFA operates and its potential utility in rationally engineering plant systems have been described in recent reviews (Schwender *et al.*, 2004b; Ratcliffe & Shachar-Hill, 2006; Shachar-Hill, 2013). Briefly, MFA takes a mass balance approach to track how carbon moves through a network of nonlinear biochemical reactions, which is depicted as a stoichiometric matrix. In addition to this matrix, the model requires definitions of how the carbon moves within each reaction. For example, in the pyruvate decarboxylation reaction, the C1 carbon of pyruvate is removed to form carbon dioxide, while the C2-C3 carbons become part of acetyl CoA. A computer program (e.g., 13C-FLUX) uses the metabolic network and relevant experimental data to quantify the network fluxes that best fit the model. Experimental data inputs for MFA models include substrate uptake and metabolite production rates, as well as NMR and/or GC-MS measurements of these compounds. Various substrates, such as ^{13}C-glucose, ^{13}C-glutamine, and ^{13}C-alanine, can be used as the label source and give different information. For example, ^{13}C-glucose can be used by glycolytic enzymes in the cytosol to synthesize hexose- or triose-phosphates that are transported into the plastid and used in glycolysis or the pentose phosphate pathway (Plaxton, 1996; Kruger & von Schaewen, 2003). On the other hand, ^{13}C-glutamine enters the mitochondrial tricarboxylic acid cycle by being converted into glutamate and then alpha-ketoglutarate, while ^{13}C-alanine is converted into cytosolic pyruvate, which can serve as a precursor for plastidic or mitochondrial reactions (Wightman & Forest, 1978). Because of these different capabilities, label movement through parallel pathways can be more rigorously assessed by using multiple, distinct substrates to obtain the MFA data (Schwender *et al.*, 2004b).

In this work, we used MFA to investigate the low carbon conversion efficiency in developing *C sativa* seeds. We first identified metabolic sources of high CO_2 emission that could contribute to the low efficiency by performing MFA with four independent labeling treatments: U-^{13}C glutamine, U-^{13}C alanine, 1-^{13}C glucose, and 80% 1,2, 20% U-^{13}C glucose. We found that the oxidative pentose phosphate (OPP) decarboxylation reaction was highly active and was responsible for most of the CO_2 produced. Next, we

tested how strongly the OPP flux correlates with the carbon use efficiency by performing MFA on *C. sativa* embryos grown under dark and high light (50 µE) because embryos grown under these conditions are known to have lower and higher, respectively, efficiencies than embryos grown under physiological light (10 µE in the seed) (Chen & Shachar-Hill, 2012). This supported our conclusion that OPP decarboxylation is the major cause of *C. sativa* embryos having low carbon use efficiency.

Results & Discussion

Potential network topologies were quantitatively tested

Some metabolites (e.g., histidine, aromatic amino acids) are known to be synthesized in the plastid from pentose phosphate pathway intermediates (Umbarger, 1978; Maeda & Dudareva, 2012), but there has been uncertainty as to the subcellular location of the oxidative and nonoxidative reactions of this pathway. Previous studies have found evidence that at least some of the reactions are present in both the cytosol and plastid (Kruger & von Schaewen, 2003), and that the reversible nonoxidative reactions foster rapid exchange of pathway intermediates between the cytosol and plastid (Ratcliffe & Shachar-Hill, 2006). Furthermore, *Arabidopsis thaliana*, a close relative of *C sativa*, has been found to have phosphate-translocator proteins capable of transporting pentose phosphates into the plastid (Kruger & von Schaewen, 2003). Consequently, our model placed the nonoxidative reactions in the plastid to account for plastid-specific metabolite synthesis.

We tested whether OPP decarboxylation uses cytoplasmic or plastidic hexose phosphate as a substrate to synthesize plastidic ribulose 5-phosphate by comparing the model fits with either substrate. The 50 best model fits when the oxidative reactions were in the plastid were all at least 50% better than the best fit when the reactions were in the cytoplast (Fig 2.1.1a-b); therefore, we concluded that there was insufficient evidence for a significant cytosolic pentose phosphate pathway in these embryos.

Due to its observed importance in green seeds such as rapeseed (Schwender *et al.*, 2004a), we

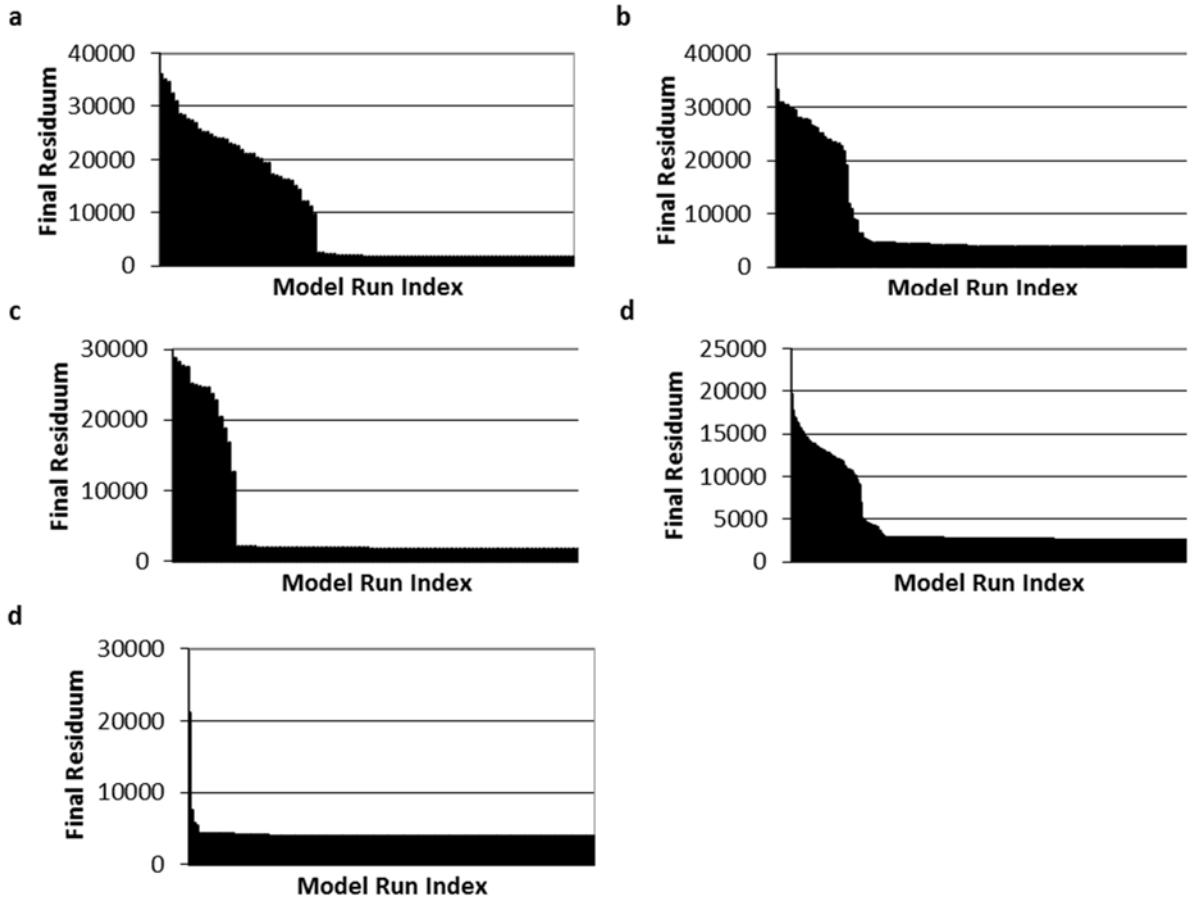

Figure 2.1.1. Testing alternative network topologies. (a- b) The sub-cellular location of the OPP decarboxylation hexose phosphate substrate was tested in *C. sativa* embryos grown under *in planta* light by determining the model fit (i.e., final residuum) when the substrate was in the plastid (a) or cytoplasm (b). We found that the final residuum of these tests was 1804 and 3790, respectively. (a, c) The presence of an active rubisco bypass was tested in embryos grown under *in planta* (a, c) and high light (d, e). We found that the final residuum without the bypass (a, d) was 1804 and 2601, and was 1794 and 3845 with the bypass (c, e) in the embryos grown under *in planta* and high light conditions, respectively.

tested if there was evidence for the rubisco bypass operating in developing *C sativa* embryos by comparing model fits with and without this reaction as an irreversible, free flux. In the rubisco bypass, carbon dioxide is added to the second carbon of ribulose 5-phosphate and results in two triose phosphates. At physiological and high light, permitting carbon flow through the rubisco bypass improved model fit by up to 5% (Fig 2.1.1), which could be explained by the model containing an extra free parameter. Together with the experimental results, we concluded that there was insufficient evidence for a significant rubisco

bypass flux in these embryos.

Most of the CO_2 in C. sativa embryos was produced via OPP decarboxylation

Most of the substrate carbon was taken up as glucose. MFA revealed that nearly all of this carbon was transported to the plastid as hexose phosphate and used for OPP decarboxylation (Fig 2.1.2a). The remaining pentose phosphate reactions led to a substantial production of hexoses, which were cycled back into the OPP decarboxylation reaction, and trioses, which were primarily converted into hexoses and then cycled into the OPP. The cycling of the carbon in this way led to an OPP decarboxylation flux of over 38,000 nmol C/embryo/day (Fig 2.1.2a). Because the decarboxylation reaction converts 1 of the 6 hexose phosphate carbons into CO_2, this high flux resulted in 86% of the total CO_2 produced (Fig 2.1.3a). In comparison, the TCA and fatty acid synthesis decarboxylations only accounted for 13% of the CO_2.

The OPP flux is strongly correlated with inefficiencies in carbon use

As expected from the differences in embryo growth rates and carbon conversion efficiencies, as light availability increased, the developing embryos had higher net fluxes toward synthesizing carbohydrates, amino acids, and fatty acids, as well as decreased carbon efflux (Fig 2.1.2). However, when these fluxes are normalized by total carbon uptake, we can see that the embryos use similar proportions of their carbon for synthesizing biomass products (Table 2.1.S1). In contrast, on average, the proportions of CO_2 efflux varied by 24% across the growth conditions, and the pentose phosphate reactions varied by at least 90%. The OPP decarboxylation flux increased from 415% of the total carbon in the inefficient dark embryos to 237% in the efficient high light embryos. This supports the hypothesis that OPP decarboxylation flux greatly influences the observed carbon use efficiency. We used linear regression to quantify how strongly the OPP flux and efficiency are correlated in the three light levels (Fig 2.1.3b). We found this correlation to explain 89% of the variation in carbon use efficiency. Consequently, we propose

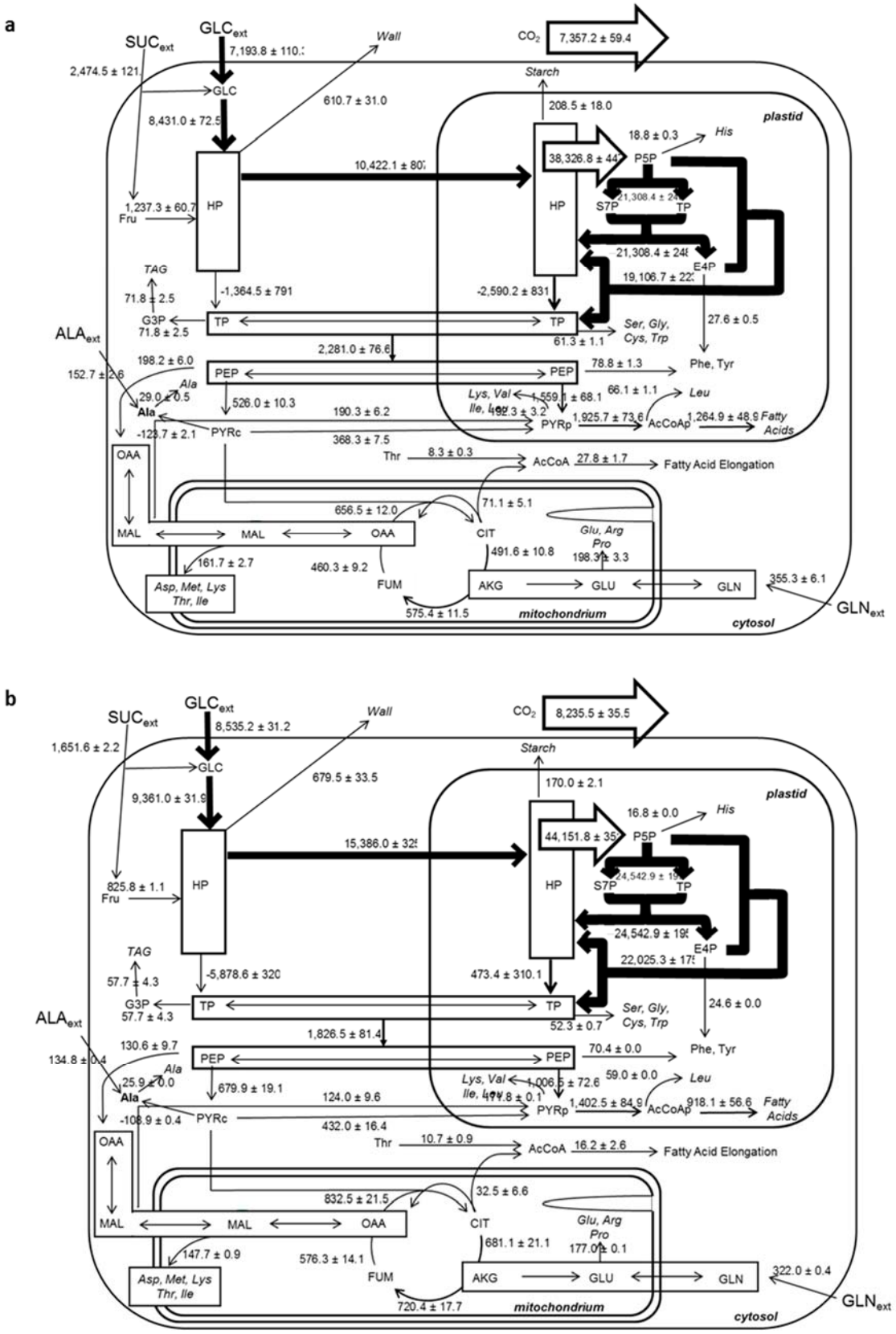

Figure 2.1.2. *C. sativa* flux maps showing total carbon flux.

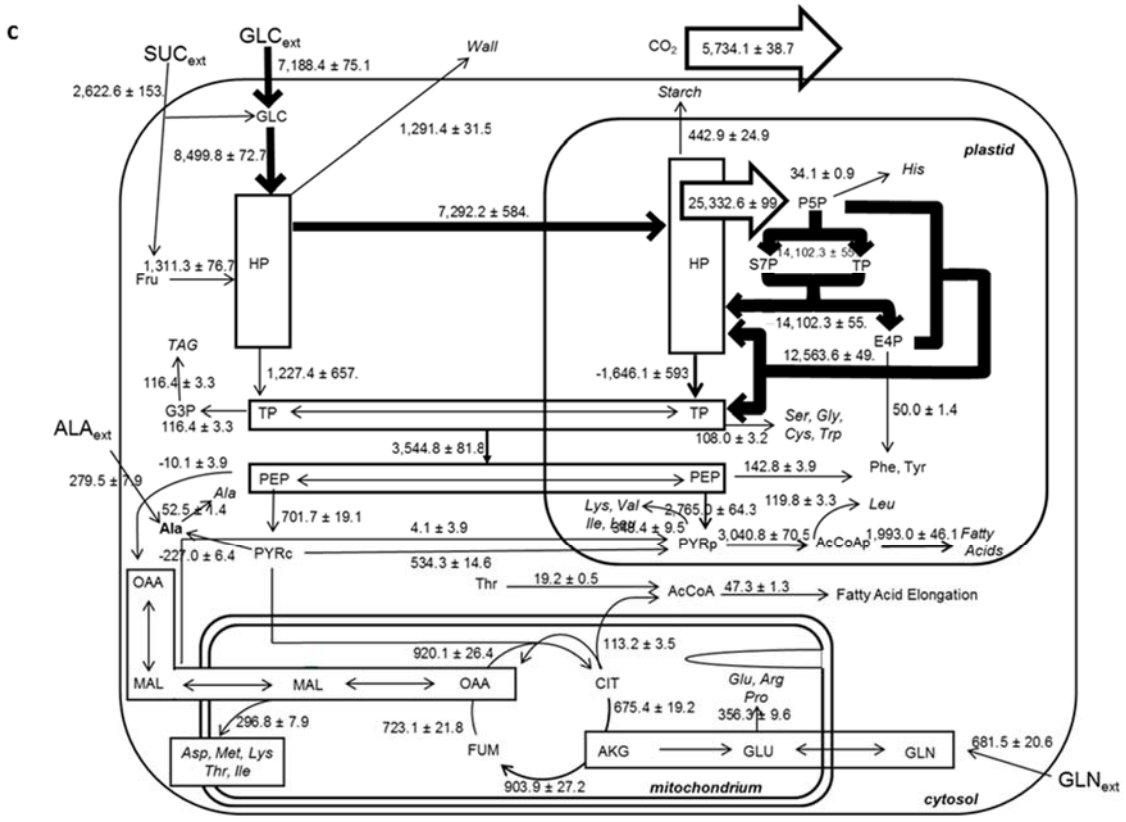

Figure 2.1.2 (cont'd). ***C. sativa*** **flux maps showing total carbon flux. E**mbryos were cultured for 6-10 days with U-^{13}C alanine, U-^{13}C glutamine, 1-^{13}C glucose , or 80% 1,2, 20% U-^{13}C glucose to obtain labeling data to quantify metabolic fluxes by ^{13}C-MFA for embryos grown under (a) *in planta*, (b) dark, and (c) high light conditions. Arrow sizes correlate with flux intensity and values are in units of nmol C/embryo/day. Values show the average net flux ± 90% confidence interval as determined from the 20 best fits derived from pseudo datasets.

that engineering *C. sativa* seeds to have decreased OPP decarboxylation activity will increase the seed size and oil content.

Methods

Model construction and constraints

The four labeling treatments were integrated into the MFA model by constructing four identical metabolic networks, each fitted to the data for one set of labeling measurements. By constraining the corresponding reactions for each network to be equal, the model could determine fluxes that best fit all four treatments

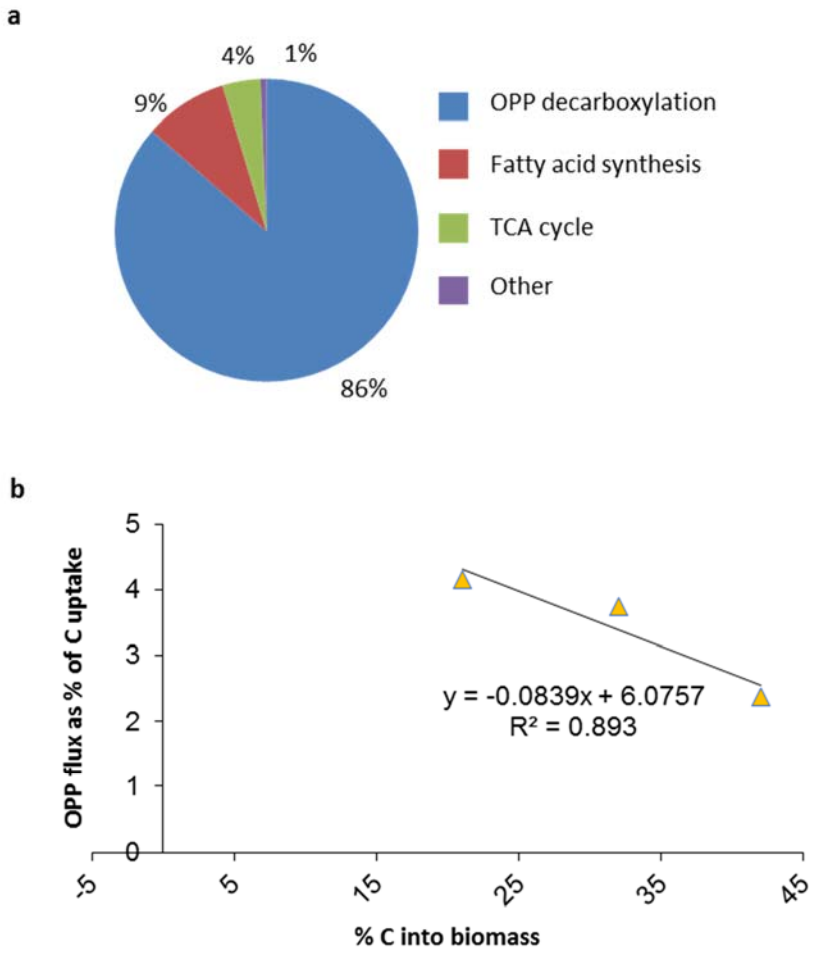

Figure 2.1.3. OPP decarboxylation produced most of the CO_2 and largely contributed to poor carbon use efficiency. (a) The proportion of CO_2, as determined by MFA, that is produced by OPP (blue), TCA cycle (green), fatty acid synthesis (red), and other (purple) decarboxylation reactions. (b) The proportion of embryo carbon uptake that was used for biomass production was determined experimentally and the OPP flux was determined by MFA and normalized by total carbon uptake. The correlation was analyzed with linear regression.

(Methods S1). The 80% 1,2, 20% U-^{13}C glucose treatment required two parallel glucose influx reactions. These parallel reactions were further constrained to be within 5% of the treatment ratio (i.e., 4:1) to disallow discrimination between the two glucose substrates. A network nitrogen balance was enforced by constraining the predicted total nitrogenous compound production (i.e., the sum of amino acid production fluxes, scaled by their nitrogen contents) to be within 15% of the nitrogen uptake (i.e., from

glutamine and alanine).

Select reactions were also constrained to better reflect biological restrictions or measurements. First, the biologically irreversible reactions, such as decarboxylase and kinase reactions, were constrained to have net fluxes greater than or equal to zero, and an exchange flux equal to zero. Second, to accommodate observed glycerolipid structures, the ratio of glyceraldehyde 3-phosphate to total (i.e., cytosolic and plastidic) acyl-CoA production fluxes was constrained to be within 30% of the measured glyceryl:acyl moieties ratio. Lastly, for fluxes that were experimentally derived (e.g., substrate influx, amino acid production), the predicted fluxes were constrained to be within 50% of the measured values.

Quantifying fluxes and their uncertainty

Metabolic fluxes were quantified using the 13C-FLUX software (Wiechert *et al.*, 2001; Wiechert *et al.*, 2015). Initial starting points were randomly generated with MATLAB R2016a to span a region with radius representing 10% of the total carbon influx surrounding the algebraic center of feasible space or previously analyzed optima. Starting points were optimized with the *Donlp2* program, where model fit was quantified by the residuum; i.e., the sum of squared differences between measured and predicted parameters, with deviations scaled by the measured standard deviation to ensure precise data weigh more heavily than scattered data. The residua of at least 100 optima per light level were analyzed to ensure a global minimum was reached (Fig 2.1.1), and the starting points corresponding to the 20 lowest residua were identified. 50 pseudo datasets were generated with Monte Carlo sampling using the average and standard deviation of all measurements. The 20 starting points were optimized with the 50 pseudo datasets to find the best fit flux values per pseudo dataset. These optima were used to calculate the averages and confidence intervals for the flux values (Fig 2.1.2).

Supporting Information

Supplemental File 3 – Chapter 2 Methods S1

Table 2.1.S1. Net fluxes normalized by total carbon uptake

flux	dark	physioloigcal	high
Vg1	0.14 ± 0.00	0.14 ± 0.00	0.12 ± 0.00
Vg2	0.66 ± 0.00	0.56 ± 0.02	0.53 ± 0.01
Vala1	0.01 ± 0.00	0.01 ± 0.00	0.03 ± 0.00
Va	0.03 ± 0.00	0.03 ± 0.00	0.06 ± 0.00
Vsuc	0.16 ± 0.00	0.26 ± 0.02	0.25 ± 0.01
Vsuc1	0.16 ± 0.00	0.26 ± 0.02	0.25 ± 0.01
Vhk1	0.88 ± 0.00	0.82 ± 0.01	0.78 ± 0.01
Vhk2	0.08 ± 0.00	0.13 ± 0.01	0.13 ± 0.01
Vald	-0.56 ± 0.03	-0.17 ± 0.09	0.16 ± 0.07
Vglyco	0.17 ± 0.01	0.23 ± 0.01	0.33 ± 0.00
Vfasa	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00
Vfasb	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00
Vpk	0.07 ± 0.00	0.05 ± 0.00	0.07 ± 0.00
Vg3p	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
Vgl	-0.03 ± 0.04	-0.02 ± 0.03	-0.02 ± 0.04
Vhpt	1.45 ± 0.04	1.06 ± 0.09	0.63 ± 0.07
Valdp	0.05 ± 0.03	-0.21 ± 0.10	-0.20 ± 0.06
Vpkp	0.10 ± 0.01	0.16 ± 0.01	0.26 ± 0.00
Vstsp	0.02 ± 0.00	0.02 ± 0.00	0.04 ± 0.00
Vpdhp	0.13 ± 0.01	0.19 ± 0.01	0.28 ± 0.00
Vfas1	0.09 ± 0.01	0.13 ± 0.00	0.18 ± 0.00
Vppp1	2.30 ± 0.03	2.08 ± 0.03	1.32 ± 0.03
Vppp2	2.07 ± 0.02	1.87 ± 0.02	1.17 ± 0.02
Vppp3	2.30 ± 0.03	2.08 ± 0.03	1.32 ± 0.03
Vopp	4.15 ± 0.05	3.74 ± 0.05	2.37 ± 0.05
Vakg	0.01 ± 0.00	0.02 ± 0.00	0.03 ± 0.00
Vcs	0.08 ± 0.00	0.06 ± 0.00	0.08 ± 0.00
Vca	0.07 ± 0.00	0.05 ± 0.00	0.06 ± 0.00
Vsfa1	0.03 ± 0.00	0.03 ± 0.00	0.04 ± 0.00
Vsfa2	0.03 ± 0.00	0.03 ± 0.00	0.04 ± 0.00
Vfum1	0.03 ± 0.00	0.02 ± 0.00	0.03 ± 0.00
Vfum2	0.03 ± 0.00	0.02 ± 0.00	0.03 ± 0.00
Vme	0.01 ± 0.00	0.02 ± 0.00	0.00 ± 0.00
Vpepc	0.01 ± 0.00	0.02 ± 0.00	0.00 ± 0.00
Vacl	0.00 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
Vpyrt	0.04 ± 0.00	0.04 ± 0.00	0.05 ± 0.00
Vco2	0.77 ± 0.01	0.72 ± 0.01	0.53 ± 0.01
Vwall	0.06 ± 0.00	0.06 ± 0.00	0.12 ± 0.00
Vsta	0.02 ± 0.00	0.02 ± 0.00	0.04 ± 0.00
Vglxn	0.06 ± 0.04	0.05 ± 0.03	0.08 ± 0.04
Vglxu	-0.05 ± 0.04	-0.04 ± 0.03	-0.07 ± 0.04
Vglueff	0.01 ± 0.00	0.01 ± 0.00	0.02 ± 0.00
Vasp	0.01 ± 0.00	0.01 ± 0.00	0.02 ± 0.00
Vaspeff	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00
Vser	0.00 ± 0.00	0.01 ± 0.00	0.01 ± 0.00

Table 2.1.S1 (cont'd)

flux	dark	physiological	high
Vsereff	0% ± 0%	0% ± 0%	0% ± 0%
Vgly	0% ± 0%	0% ± 0%	1% ± 0%
Vglyeff	0% ± 0%	0% ± 0%	0% ± 0%
Vcheff	0% ± 0%	0% ± 0%	0% ± 0%
Vala	-1% ± 0%	-1% ± 0%	-2% ± 0%
Valaeff	0% ± 0%	0% ± 0%	0% ± 0%
Varo1	0% ± 0%	0% ± 0%	0% ± 0%
Varo2	0% ± 0%	0% ± 0%	0% ± 0%
Vakiv	1% ± 0%	1% ± 0%	2% ± 0%
Vleu	1% ± 0%	1% ± 0%	1% ± 0%
Vleueff	0% ± 0%	1% ± 0%	1% ± 0%
Vthr	1% ± 0%	1% ± 0%	1% ± 0%
Vthrald	0% ± 0%	0% ± 0%	0% ± 0%
Vthreff	0% ± 0%	0% ± 0%	0% ± 0%
Vile	0% ± 0%	0% ± 0%	1% ± 0%
Vileeff	0% ± 0%	0% ± 0%	1% ± 0%
Vval	0% ± 0%	0% ± 0%	1% ± 0%
Vvaleff	0% ± 0%	0% ± 0%	1% ± 0%
Vphe	0% ± 0%	0% ± 0%	1% ± 0%
Vpheeff	0% ± 0%	0% ± 0%	1% ± 0%
Vtyr	0% ± 0%	0% ± 0%	1% ± 0%
Vtyreff	0% ± 0%	0% ± 0%	0% ± 0%
Vpro	0% ± 0%	0% ± 0%	1% ± 0%
Vproeff	0% ± 0%	0% ± 0%	1% ± 0%
Vlys1	0% ± 0%	0% ± 0%	1% ± 0%
Vlyseff	0% ± 0%	0% ± 0%	1% ± 0%
Vhis	0% ± 0%	0% ± 0%	0% ± 0%
Vhiseff	0% ± 0%	0% ± 0%	0% ± 0%
Vacoa	0% ± 0%	0% ± 0%	0% ± 0%
Vmet	0% ± 0%	0% ± 0%	0% ± 0%
Vmeteff	0% ± 0%	0% ± 0%	0% ± 0%
Varg	0% ± 0%	1% ± 0%	1% ± 0%
Vargeff	0% ± 0%	1% ± 0%	1% ± 0%

Net fluxes were determined with ^{13}C-MFA for *C. sativa* developing embryos grown under dark, physiological (i.e., *in planta*), and high light conditions. Values are the average proportions of total carbon uptake that is used in that flux ± 90% confidence intervals as determined from the 20 best fits derived from pseudo datasets.

Section 2.2

The effects of transgenic alteration of oil composition on central metabolic fluxes in developing oilseeds

---

This section includes a draft of the manuscript:
Clark TJ, LM Carey, RR Deshpande, JC Cocuron, EK Rustad, and Y Shachar-Hill. The effects of transgenic alteration of oil composition on central metabolic fluxes in developing oilseed.

Preface

While discovering that *Camelina sativa* embryos have poor carbon use efficiency due to a high oxidative pentose phosphate flux (Section 2.1), my lab sought to use the developed ^{13}C-metabolic flux analysis (MFA) method to investigate additional metabolic traits. A ^{14}C-acetate study on *Brassica napus* found that the label in transgenic embryos with increased medium-chain fatty acid content accumulated in water-soluble compounds rather than in the lipids as expected from the wildtype (Eccleston & Ohlrogge, 1998). Consequently, the authors hypothesized that this was due to the transgenic embryos having higher β-oxidation and glyoxylate cycle fluxes. My lab decided to test and quantify these flux increases using ^{13}C-MFA and transgenic *C. sativa* embryos with higher medium chain fatty acid contents.

Initially, my role in the MFA projects was to simply perform MFA on the transgenic lines using the existing model and experimental data that other lab members—led by Lisa M Carey (LC)—had generated. However, the model was not able to determine fluxes with data from the transgenic lines. YSH and I thought the model failures could be due to the model making infeasible predictions (e.g., the stoichiometry of the predicted products may have violated the network structure, thus the prediction must have been wrong). To keep the model predictions in feasible space, I added the constraints described in Section 2.1. Adding those constraints enabled the model to successfully determine fluxes for the transgenic lines, but did not significantly change its predictions for the wild-type embryos, so to be consistent, we chose to use the revised model for both MFA studies. As with the wild-type study (Section 2.1), I used the refined MFA model and the Monte Carlo method to determine flux maps with confidence intervals for each type of embryo.

I tested if there was an increase in the glyoxylate cycle flux in response to the transgenic medium-chain fatty acids by adding the glyoxylate cycle reactions to the model. I then compared the fits of the model with this cycle to the model without. Adding a free parameter inherently improves model fit, but only one of the tested transgenics showed a potentially significant improvement (8%). For that transgenic,

I further tested if there was an increase in the glyoxylate cycle flux by comparing the values of all fluxes determined by the models. I found no significant differences in the flux values, so we concluded that there was not sufficient evidence to support there being an increase in glyoxylate cycle flux in response to medium-chain fatty acids in these embryos.

In addition to the MFA, I also analyzed the transgenic findings with principal component analysis (PCA), hierarchical clustering analysis (HCA), and k-means clustering using the embryo biological replicates for most data-driven analyses, or pseudo replicates for flux or combinatorial data-driven analyses.

My PCA and HCA tests included testing how the embryos cluster in response to 1) the labeling data factorially (i.e., all substrates, all but one, all but two, etc.), 2) the net fluxes normalized by total substrate unit uptake or by total carbon uptake, and 3) the fatty acid species compositions. I also tested 4) different methods of dealing with missing data (e.g., exclude that data for all replicates, generate pseudo data using the MATLAB alternating least squares algorithm), 5) if the distance matrix could be manually created using confidence intervals for the distance between an object and itself (which is traditionally assumed to be zero), and 6) different distance algorithms and other MATLAB parameters (e.g., seuclidean linkage metric, Ward linkage method).

The manuscript on this project will include the experimental findings, MFA-determined fluxes, and PCA and HCA analyses. LC was responsible for the experimental work and I was responsible for the computational work. I am taking lead on the manuscript writing and will serve as first-author. The current draft of this manuscript is included in this section. LC drafted the experimental methods, and Figures 1 and S2, while I drafted the remaining text, figures, and tables.

Introduction

Either directly or indirectly as food for animals, plant seed crops provide the majority of human caloric intake; of these, rice, maize, wheat, and soybeans are by far the largest (Dyer *et al.*, 2008). During development, seeds receive carbon and nitrogen from the mother plant and convert these nutrients into biomass products, predominately carbohydrates, proteins, and triacylglycerols (TAG or oil; Borisjuk *et al.*, 2004; Weber *et al.*, 2005). There is great interest in designing these oils to be compatible with automotive applications because biodiesel use results in lower net emission of CO_2 and can replace petroleum, whose availability will become limited in the future (Durrett *et al.*, 2008; Gupta *et al.*, 2010; Kallio *et al.*, 2014). In particular, aircraft would benefit from a sustainable biofuel source because unlike the situation with terrestrial vehicles, it seems unlikely that electric power systems will be feasible for flight in the foreseeable future (Rosero *et al.*, 2007). Conventional jet fuels are typically comprised of C_9 to C_{15} hydrocarbons (Liu *et al.*, 2013), but TAG from most plants (including the world's major oil crop plants) predominately contain C_{16} and C_{18} fatty acids, with some longer ones (Durrett *et al.*, 2008). Consequently, significant research effort in recent decades has been directed at increasing the medium-chain fatty acid (C_8 to C_{14}) content of TAG in *Brassica napu*s (Eccleston & Ohlrogge, 1998; Stoll *et al.*, 2005), *Camelina sativa* (Kim *et al.*, 2015; Bansal *et al.*, 2018), A*rabidopsis thaliana* (Dörmann *et al.*, 2000) and other seeds (Thelen & Ohlrogge, 2002).

In plants, the first 16-18 fatty acid carbons are assembled in the plastid by fatty acid synthase (FAS), which adds successive two-carbon units to the growing chain that is attached to an acyl carrier protein (ACP; Li-Beisson *et al.*, 2013). In order to export the fatty acid out of the plastid to be used for TAG synthesis, a thioesterase hydrolyzes the acyl-ACP compound to release a free fatty acid (Durrett *et al.*, 2008; Li-Beisson *et al.*, 2013). The specificity of the thioesterase often determines the abundant fatty acid lengths. For example, the FatA thioesterases preferentially hydrolyze C_{18}-ACP compounds to release stearate (18:0) or oleate (18:1), while FatB thioesterase can hydrolyze 16:0-ACP and release palmitate

(Durrett *et al.*, 2008; Li-Beisson *et al.*, 2013; Kim *et al.*, 2015). Some plants, such as *Umbellularia californica* (California Bay), have alternative thioesterases that cause the plant to naturally accumulate medium-chain fatty acids (Pollard *et al.*, 1991; Voelker *et al.*, 1992). This provides an opportunity for researchers to engineer other plants to have these alternative thioesterases to increase their medium-chain levels (Voelker *et al.*, 1992; Eccleston & Ohlrogge, 1998; Kim *et al.*, 2015).

These four lines have increased levels of medium-chain fatty acids. While this has potential industrial benefits, *C sativa* does not naturally have these fatty acids, so they may induce other metabolic changes (Hooks *et al.*, 1995). Such effects have been observed by Eccleston and Ohlrogge (1998) in response to *B. napu*s being engineered to produce high levels of laureate (12:0). When wild-type developing embryos were incubated with ^{14}C-acetate, nearly all the label was incorporated into fatty acids. In contrast, in the transgenic high laureate embryos, 50% of the label was recovered in water-soluble compounds, including carbohydrates, without decreasing the total seed oil content. This suggests that much of the transgenically produced fatty acid into which the label was incorporated were broken down by β-oxidation and their carbon recycled through the glyoxylate cycle (Canvin & Beevers, 1961; Eccleston & Ohlrogge, 1998). In addition, isocitrate lyase and malate synthase, which are the enzymes unique to the glyoxylate cycle, and which typically have very low activity until the seeds mature and oil is broken down to facilitate germination (Beevers, 1980; Ettinger & Harada, 1990; Eastmond & Graham, 2001) showed activities that were elevated 6.6- and 30-fold, respectively, in the developing high laureate embryos (Eccleston & Ohlrogge, 1998). Consequently, it was concluded that the presence of laureate-synthesizing enzymes induced β-oxidation and the glyoxylate cycle, while also inducing fatty acid synthesis in order to maintain oil content. This would be expected to substantially lower carbon use efficiency in the developing seeds and limit the levels of desired fatty acids.

To address this question and to more broadly measure changes in central metabolism induced when fatty acid composition is altered, we quantitatively mapped metabolic fluxes in developing oilseed

embryos of *C. sativa* plants engineered to have increased medium-chain fatty acid content. *C. sativa* is a model oilseed crop that has gained attention for studying and implementing improvements in biofuel production because of its similarities to *A. thaliana* (e.g., short lifespan, ease of transformation, and the translatability of tools and knowledge developed in *A. thaliana*) and its ability to grow in conditions unsuitable for most oil-producing crop plants (Zubr, 1997; Lu & Kang, 2008; Iskandarov *et al.*, 2014). We analyzed four *C. sativa* lines that were genetically engineered in the Cahoon lab at the University of Nebraska–Lincoln to express genes specific to medium-chain fatty acid synthesis (Kim *et al.*, 2015). These lines are (A) a line designated as *ecthio14* that contains a *Cuphea hookeriana* FatB thioesterase (Thio14) and a *U. californica* FatB thioesterase (UcFatB1). These enzymes are specific for 12:0- and 14:0-ACP, and 12:0-ACP (Dehesh *et al.*, 1996; Tjellström *et al.*, 2013), respectively, which cause this line to accumulate high levels of 12:0 and 14:0. (B) a line designated as *3thio* that contains Thio14, UcFatB1, and a *Cuphea hookeriana* FatB thioesterase that is specific for 8:0- and 10:0-ACP (ChFatB2; Voelker *et al.*, 1992), with increased levels of 8:0, 10:0, 12:0, and 14:0. (C) a line designated as *echl* that contains UcFatB1 and a *Coco nucifera* lysophosphatidic acid acyltransferase (CnLPAT; (Knutzon *et al.*, 1999). The latter enzyme selectively attaches 12:0 to the *sn*-2 position of glyceraldehyde 3-phosphate (G3P), thus increasing the abundance of 12:0 fatty acids because the native LPAT is selective for C_{18} fatty acids (Kim *et al.*, 2015). (D) a line designated as *echm2* that contains Thio14 and CnLPAT, which results in high accumulation of 14:0.

We cultured and labeled developing embryos with ^{13}C-labeled glucose, -alanine, and -glutamine, and performed steady-state metabolic flux analysis (MFA) to quantify the central metabolic fluxes. We found significant changes in central metabolic fluxes in the transgenics compared to the wild type but no induction of significant glyoxylate cycle or β-oxidation fluxes. Modestly altered carbon conversion efficiency in the transgenic lines, as well as more substantial changes induced in the wild-type embryos by increasing or eliminating the exposure to light during culture, were strongly correlated with the size of fluxes through the oxidative pentose phosphate (OPP) pathway.

The range of genotypes and conditions studied as well as the number of measurements used in this study were significantly larger than in previous metabolic flux analyses in plant systems. This enabled the fluxes to be estimated within narrower ranges (90% confidence intervals within 10% of mean values). The large experimental dataset and well-defined flux maps provided the opportunity to test whether and how the measured data could be used to directly distinguish among genotypes and growth conditions without computational modeling of fluxes. Principal component analysis successfully distinguished between the transgenics using fatty acid or protein amino acid labeling data, but did not completely distinguish between them using the values of metabolic fluxes. These flux distributions were more effective in differentiating between the transgenics using hierarchical clustering analysis. We discuss how the combination of clustering analyses of labeling data and flux maps can explain metabolic differences between these transgenics.

Results

C. sativa developing embryos in culture imitated those in planta

C. sativa transgenic seeds were obtained from Edgar Cahoon (Kim *et al.*, 2015) and 10 day-after-flowering (DAF) embryos were cultured for 6 days under 10 uE in media containing 8 mM glutamine, 4 mM alanine, 130 mM glucose, and 12 mM sucrose, which are consistent with *in planta* light and carbon availabilities (Carey et al., in preparation). In addition, the media contained 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) as a buffering agent, 20% polyethylene glycol as an osmoticum (Schwender & Ohlrogge, 2002), and various vitamins (see Methods). After 6 days in culture, we found that the transgenic embryos presented similar fatty acid compositions as *in planta* (Fig 2.2.1a, S1) and as described by Kim *et al.* (2015). Furthermore, there was no significant difference in embryo dry weights between transgenics and wildtype (Fig 2.2.S2a). Consequently, we concluded that under these culture conditions, the growth and TAG production of embryos of the different lines were close.

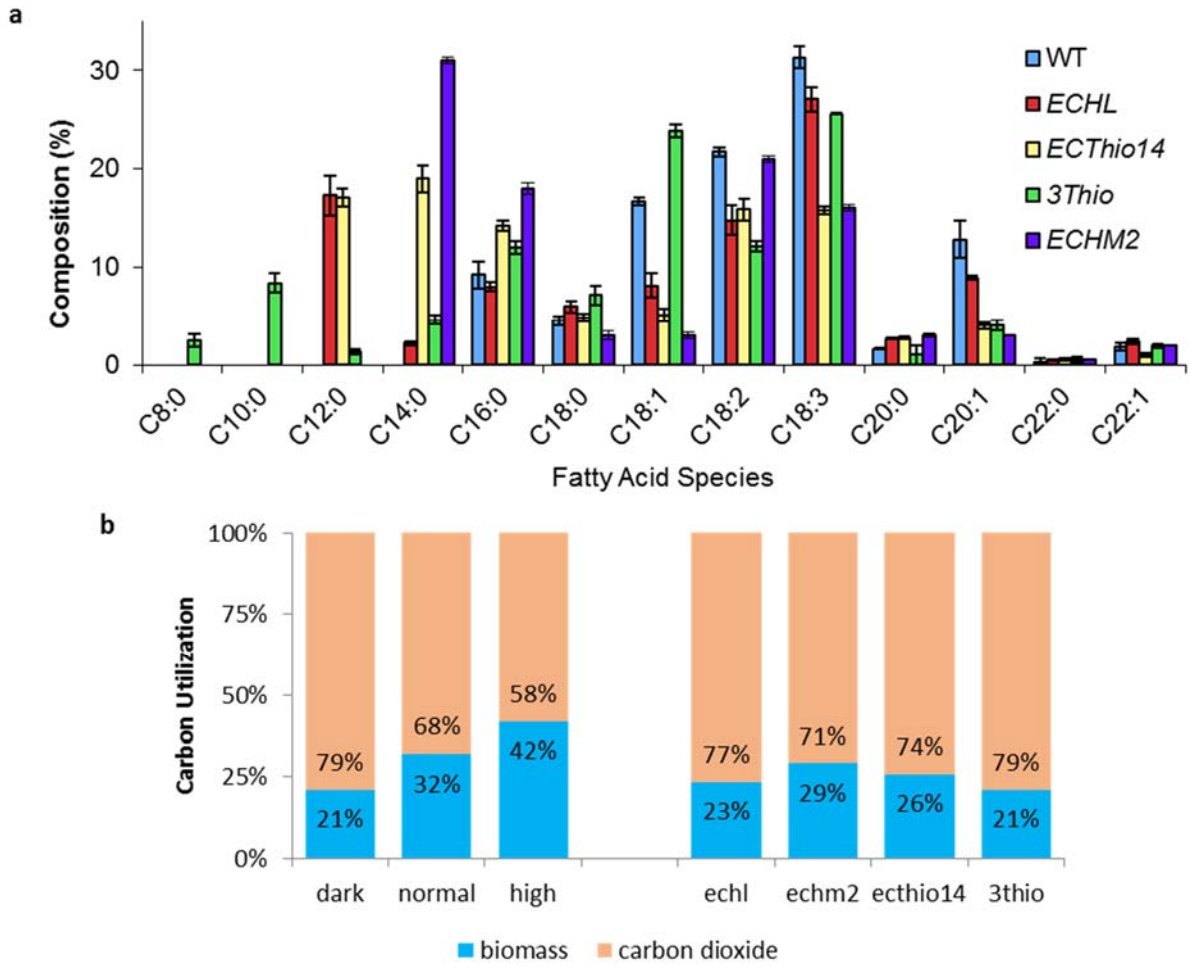

Figure 2.2.1. Physiological changes from wild-type to transgenic *C. sativa* embryos. (a) Fatty acid compositions were measured in wild-type or transgenic whole embryos after 6 days of culturing (10-16 DAF) with GC-FiD. Bars indicate average percent composition and standard deviation. (b) Proportion of carbon uptake used for biomass (blue) or emitted as CO_2 (red) in wild-type embryos grown under dark, physiological, or high light conditions, and in the four transgenics (*echl*, *echm2*, *ecthio14*, *3thio*). Carbon emission was determined by measuring total substrate (glucose, sucrose, alanine, glutamine) carbon uptake in comparison to the biomass carbon accumulated. The difference was assumed to have been emitted as to CO_2 based on ^{14}C labeling experiments on wild-type embryos where emitted $^{14}CO_2$ was trapped and quantified (Carey et al., in preparation)

Metabolite uptake rates were measured by performing ^{1}H NMR on the culture medium before and after culturing. We found that, in comparison to wild-type, transgenic embryos took up similar quantities of alanine (55-97 nmol/day/embryo), and at least as much glutamine and sucrose (97-197 and 147-215 nmol/embryo, respectively, Fig 2.2.S2c). In order to assess how the carbon was allocated to

biomass products in the different lines, we measured the total protein, fatty acid, and carbohydrate contents in cultured embryos. The transgenic lines contained 21-27% protein and 31-35% fatty acid per embryo biomass, and embryo starch contents ranged from 7% in *echm2* to 14% in *echl* (Fig 2.2.S2b).

Carbon uptake and allocation rates were used to calculate the transgenic embryo carbon conversion efficiencies, which were found to be 21-29% (Fig 2.2.1b). These values are between those of wild-type embryos grown under dark and *in planta* light levels (21 and 32%, respectively; Carey et al., in preparation). This indicates that the transgenic lines were less efficient at using carbon for biomass production than wild-type and could suggest that the transgenic oxidative pentose phosphate pathway fluxes are greater than the wild-type under *in planta* light.

Principal component analysis of labeling distinguished among transgenic lines

To assess metabolic differences between the transgenics and wildtype, we cultured the embryos under multiple labeling conditions and measured label contents in the fatty acids, amino acids, and carbohydrates. Embryos were labeled in four separate experiments with 100% U-^{13}C glutamine, 100% U-^{13}C alanine, 100% 1-^{13}C glucose, and 80% 1,2, 20% U-^{13}C glucose.

There was less fatty acid labeling when ^{13}C glutamine or ^{13}C alanine was provided to the embryos in comparison to 80% 1,2, 20% U-^{13}C glucose (Table 2.2.S1). However, both amino acid substrates were able to cleanly distinguish *echm2* from the other lines using principle component analysis (PCA, Fig 2.2.2). PCA is a statistical method of transforming the data using principal components to maximize differences between the data points are maximized. Different measurements contribute to different extents to each principal components. Measurements that are equal or vary very little across individual replicates contribute less to PCA components so that more variable measurements are emphasized. This allows PCA to resolve individuals based on the measures in which they have the mode biological differences; however, it also allows measurements that are less precise or otherwise less well determined to distort the components. The amino acid substrates were able to cleanly distinguish *echm2* largely due to the C_{20}

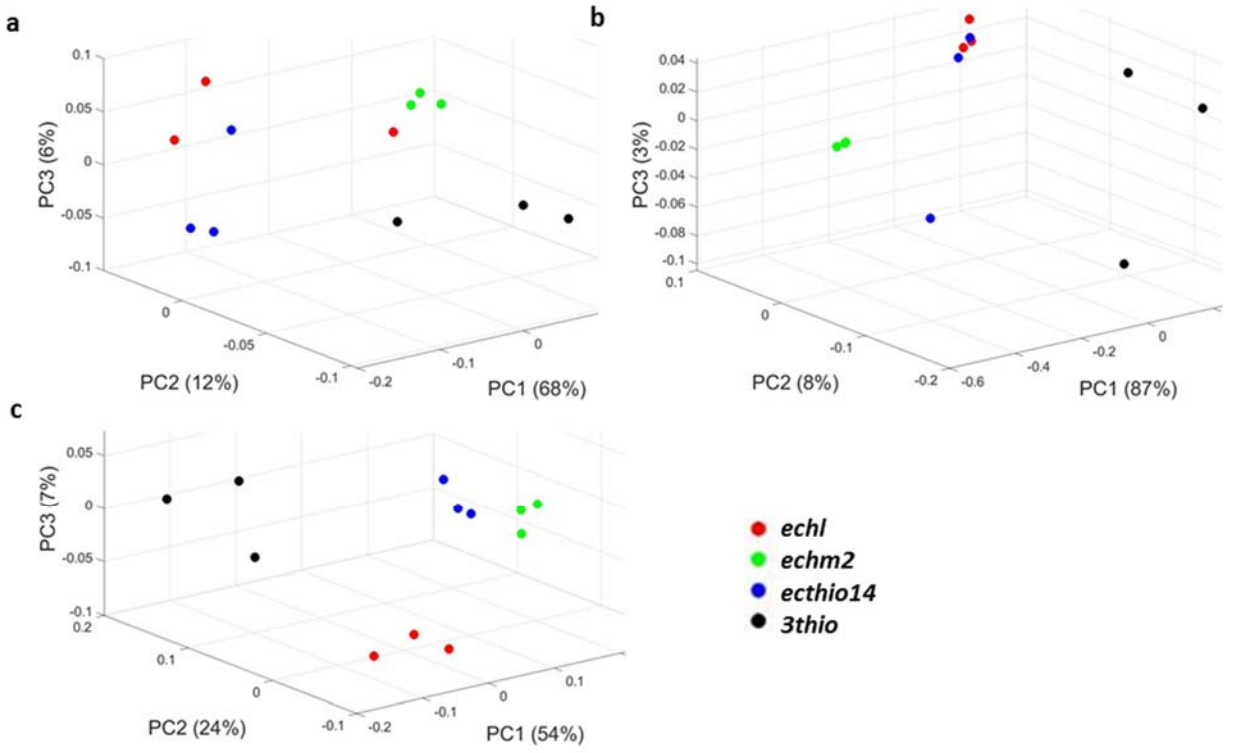

Figure 2.2.2. Clustering using PCA and fatty acid labeling data. Transgenic embryos were cultured for 6 days (10-16 DAF) with (a) U-^{13}C alanine, (b) U-^{13}C glutamine, or (c) 80% 1,2, 20% U-^{13}C glucose. PCA was performed on the fatty acid labeling profiles of *echl* (red), *echm2* (green), *ecthio14* (blue), and *3thio* (black). Fatty acid labeling was measured by transmethylating total lipid extracts of embryos after growth in the presence of labeled substrates and measuring label content in each fatty acid species with GC-MS. Circles represent biological replicates (n = 3).

to C_{22} fatty acids being more highly labeled in *echm2* (Table 2.2.S1-S2). Neither of the amino acid substrates could fully resolve *echl* and *ecthio14*, but all four transgenic lines could be separated when the embryos were provided 80% 1,2, 20% U-^{13}C glucose (Fig 2.2.2). The disparate abilities of the substrates to resolve the transgenic lines from their fatty acid labeling patterns point to there being significant differences in the fluxes under which the lines utilize hexoses and pyruvate.

PCA using the amino acid labeling data also distinguished between the lines, particularly when glucose was used as the substrate (Fig 2.2.S3). For example, when either 1-^{13}C glucose or 80% 1,2, 20% U-^{13}C glucose was used, proline labeling strongly influenced the first principle component, which accounted for over 40% of the PCA variation. In response to the 1-^{13}C glucose substrate, Proline 2345 fragment

labeling varied from 11% in *echm2* to 31% in *3thio* (Fig 2.2.3, Table 2.2.S3). This altered labeling pattern could be correlated with differences in proline synthesis fluxes and/or fluxes to synthesize its alpha-ketoglutarate precursor in the tricarboxylic acid cycle (TCA) cycle. Similarly, starch fragments in *echl* and *ecthio14* demonstrated higher percent labeling in comparison to cell wall fragments (Fig 2.2.S4), which could be explained by higher starch synthesis fluxes. Alternatively, this trend could be due to higher transport fluxes to move cytoplasmic hexose (which is a cell wall precursor) into the plastid (where hexose could be used for starch). Labeling alone could not discern why only some substrates permit PCA to distinguish between the transgenic lines or how genetically altering the fatty acid synthesis pathway could lead to differences in protein and carbohydrate labeling patterns. However, labeling data can be used with steady-state metabolic flux analysis (MFA) to quantify fluxes in central metabolism and answer these questions.

HCA distinguished between transgenic lines based on fluxes normalized by total carbon uptake

Following the method described in Carey et al. (in preparation) to analyze wild-type *C. sativa* embryos cultured across three light levels, ^{13}C MFA was performed on the four transgenic lines using metabolite labeling, uptake, and synthesis data from culturing with ^{13}C glutamine, ^{13}C alanine, 1-^{13}C glucose, and 80% 1,2, 20% U-^{13}C glucose. As indicated by the confidence intervals in the flux maps (Fig 2.2.4), the partitioning of hexose- and triose-phosphate metabolites between the cytosol and plastid was not well resolved. The triose-phosphate sub-cellular pools were treated as one pool, and the hexose transport flux was uncertain depending on what proportion of glycolytic flux was in each compartment.

Fatty acids are synthesized in the plastid from acetyl-CoA, which is made from the C2 and C3 of pyruvate. Labeling differences in fatty acids with 18 or fewer carbons therefore reflect differences in labeling of plastidic pyruvate, which can arise from differences in the origins of this pool and/or from differences in the labeling of the precursors of plastidic pyruvate. The contribution of malate

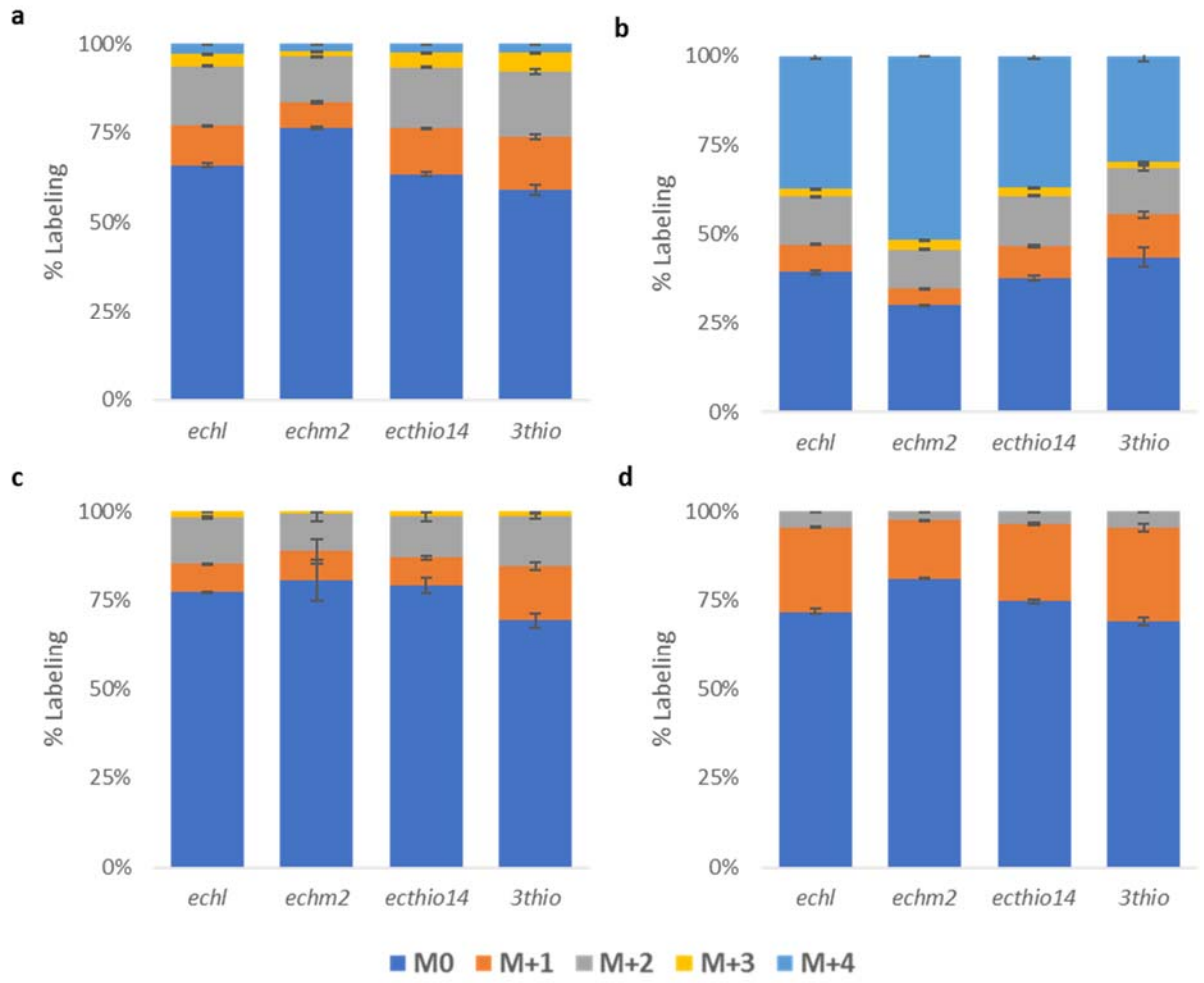

Figure 2.2.3. Variations in proline labeling across substrates and transgenics. Transgenic embryos were cultured for 6 days (10-16 DAF) with (a) 80% 1,2, 20% U-^{13}C glucose, (b) U-^{13}C glutamine, (c) U-^{13}C alanine, or (d) 1-^{13}C glucose. Labeling in proline and other amino acid was measured by hydrolyzing the total protein extract, cleanup of the resultant amino acids by ion exchange, and amino acid derivatization and GC-MS analysis. Bars represent average percent composition of the different mass isomers and the error bars are the standard deviation (n = 3-6 biological replicates).

decarboxylation to the production of plastidic pyruvate varies by ~30 fold among the lines. C_{20} and C_{22} fatty acids are made by elongation in the cytosol of C_{18} fatty acids from the plastid using cytosolic acetyl-CoA. Label measurements were made using the butyl amide derivatives of fatty acids and analysis of the resultant C2 and C3 fragments that are the last to be added to the fatty acid. Therefore, differences between the transgenic lines in the labeling in C_{20} and C_{22} were due to differences in labeling of cytosolic acetyl-CoA, which is made by ATP citrate lyase in oilseeds with high contents of long-chain fatty acids,

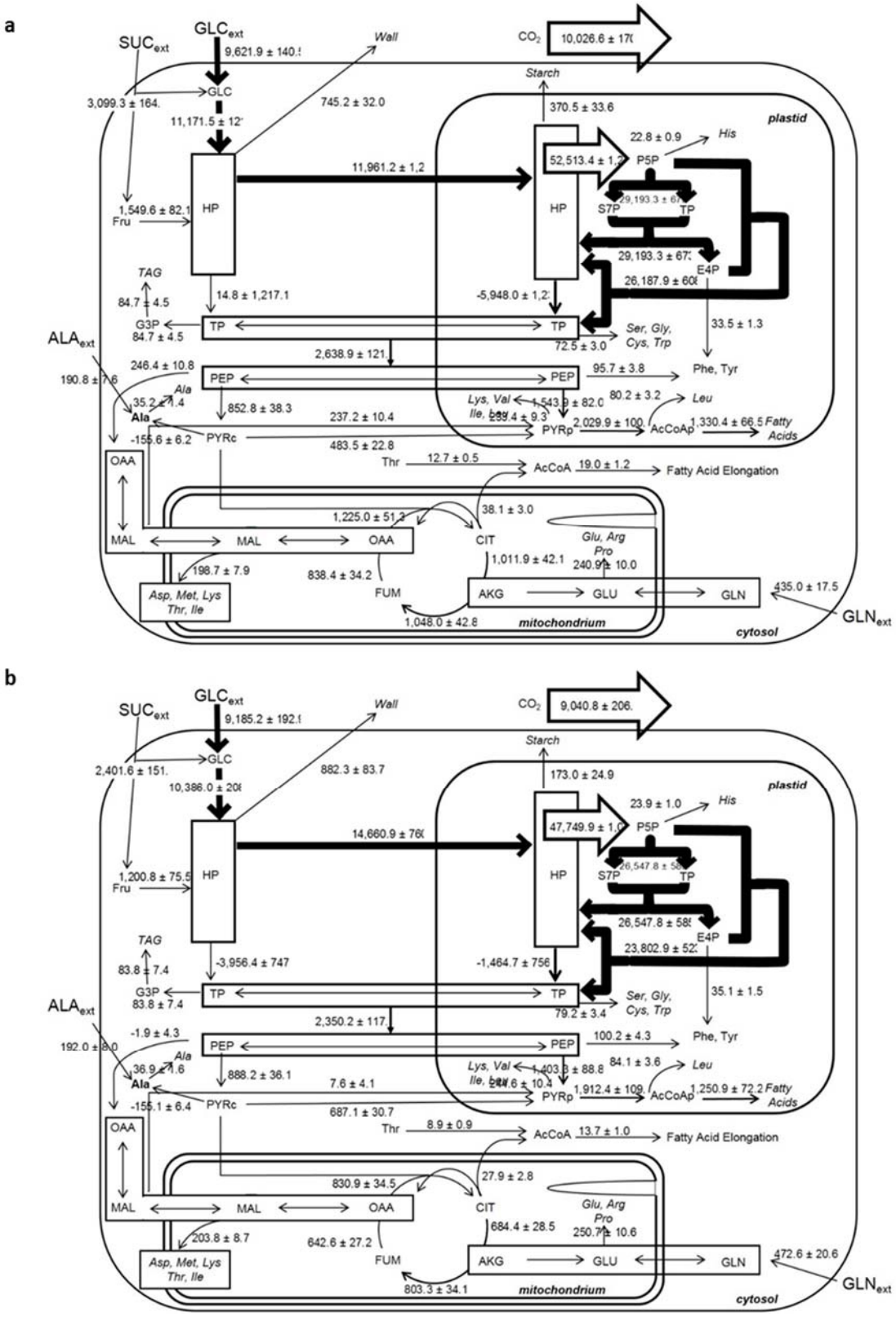

Figure 2.2.4. Transgenic flux maps showing total carbon flux.

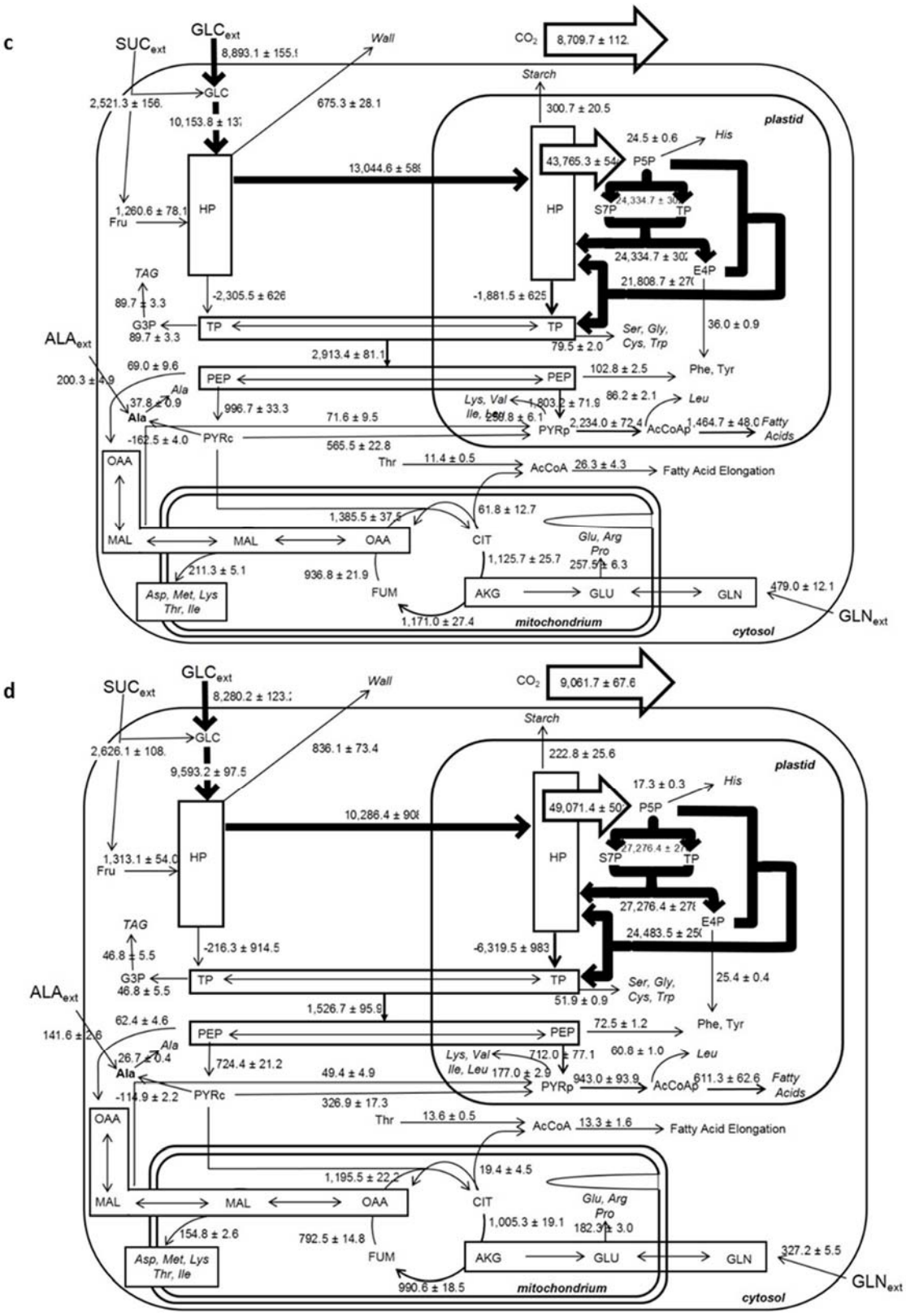

Figure 2.2.4 (cont'd). Transgenic flux maps showing total carbon flux. Embryos were cultured for 6-10 days with U-^{13}C alanine, U-^{13}C glutamine, 1-^{13}C glucose , or 80% 1,2, 20% U-^{13}C glucose to obtain labeling

Figure 2.2.4 (cont'd) data to quantify metabolic fluxes by ^{13}C-MFA for (a) *echl*, (b) *echm2*, (c) *ecthio14*, and (d) *3thio*. Arrow sizes correlate with flux intensity and values are in units of nmol C/embryo/day. Values show the average net flux ± 90% confidence interval as determined from the 20 best fits derived from pseudo datasets.

but which can also be the product of threonine aldolase activity in the cytosol. The proportion of cytosolic acetyl-CoA produced by these two routes (Fig. 4) does not vary as much as the origins of plastidic pyruvate and *echm2* does not appear to be different in this ratio from *echl*; both being intermediate between the other two transgenic lines. Consequently, it is likely that differences in labeling of citrate and/or threonine account for the observations.

As expected from the wild-type findings, only a moderate amount (4-7 % of the carbon taken up from the media) of carbon in the transgenic embryos was used in glycolysis or the TCA cycle (Fig 2.2.4, Table 2.2.S4). Most of the available carbon was taken up as glucose (Fig 2.2.S2c) and transported as hexose into the plastid, where it was used for OPP decarboxylation (Fig 2.2.4). *3thio* had the lowest CCE (21% to biomass) and highest OPP flux (91% of the carbon input) of the transgenic lines, which is consistent with the positive correlation between these traits predicted by wild-type *C. sativa* (Carey et al., in preparation, Fig 2.2.5). In contrast, *echm2* had a higher OPP flux (88%) yet higher CCE (29%) than either *ecthio14* (83%, 26%) or *echl* (87%, 23%; (Table 2.2.S4). This diversion from the expected trend suggests that other decarboxylation processes (e.g., pyruvate oxidation) are more active in *ecthio14* and *echl*. Indeed, the TCA cycle fluxes comprised a greater proportion of the available carbon in *ecthio14* and *echl* (6-7%) than *echm2* (4%) (Fig 2.2.4, Table 2.2.S4). Similarly, the flux toward cell wall synthesis represented 6% of the carbon input in both *ecthio14* and *echl*, but cell wall polysaccharides comprised a lower proportion of the embryo biomass in *ecthio14* (Fig 2.2.S2). This difference was correlated with the decarboxylation flux toward acetyl-CoA synthesis being greater in *ecthio14* (Fig 2.2.4, Table 2.2.S4), which would limit how much carbon would remain available to synthesize starch, protein, and lipids.

Because of previous reports (Eccleston & Ohlrogge, 1998), we tested whether the glyoxylate cycle

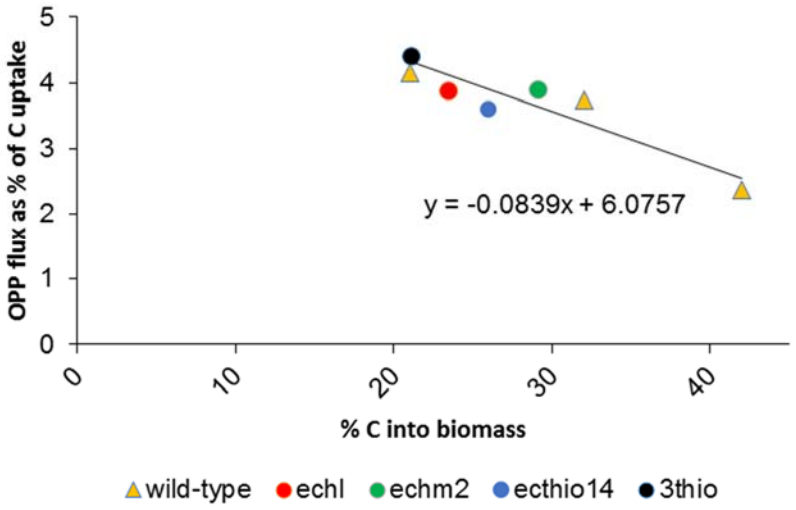

Figure 2.2.5. Correlation between carbon use efficiency and OPP decarboxylation. Carbon use efficiencies were determined by measuring total carbon uptake in comparison to total biomass carbon. ^{13}C-MFA was used quantify metabolic fluxes and normalized to represent the proportion of carbon uptake. The line of best fit was found by Carey *et al*., (in preparation) using linear regression and wild-type *C. sativa* embryos (triangles) grown under dark, *in planta*, or high-light conditions. For comparison to the transgenic lines, data points for *echl*, *echm2*, *ecthio14*, and *3thio* were added (indicated by circles), without modification to the linear curve equation. Data points represent the 10 best fits derived from pseudo datasets.

was activated in the transgenic lines due to their medium-chain fatty acids by comparing the model fit (i.e., total residuum) between the model with and without the glyoxylate cycle. Adding the glyoxylate cycle adds a free flux to the model and results in at least as good of a fit. Therefore, the modest increases in model fit that we observed for *echl*, *ecthio14*, and *3thio* (Fig 2.2.S5) were not sufficient evidence for the glyoxylate cycle to be prevalent in these lines. The model fit for *echm2* increased by 8%, which suggests a possible role for the glyoxylate cycle. However, the associated fluxes were not significantly changed (Table 2.2.S5) and conclude that any fluxes through the glyoxylate cycle in the transgenic lines are minor.

When PCA was performed using the net fluxes normalized by total carbon uptake (i.e., the substrate carbon input fluxes summed to 100%), there was a modest degree of clustering of the transgenic *C. sativa* lines (Fig 2.2.6a), but this was insufficient to separate them from one another. The first three principal components explained 98% of the variation in the normalized fluxes and were largely influenced

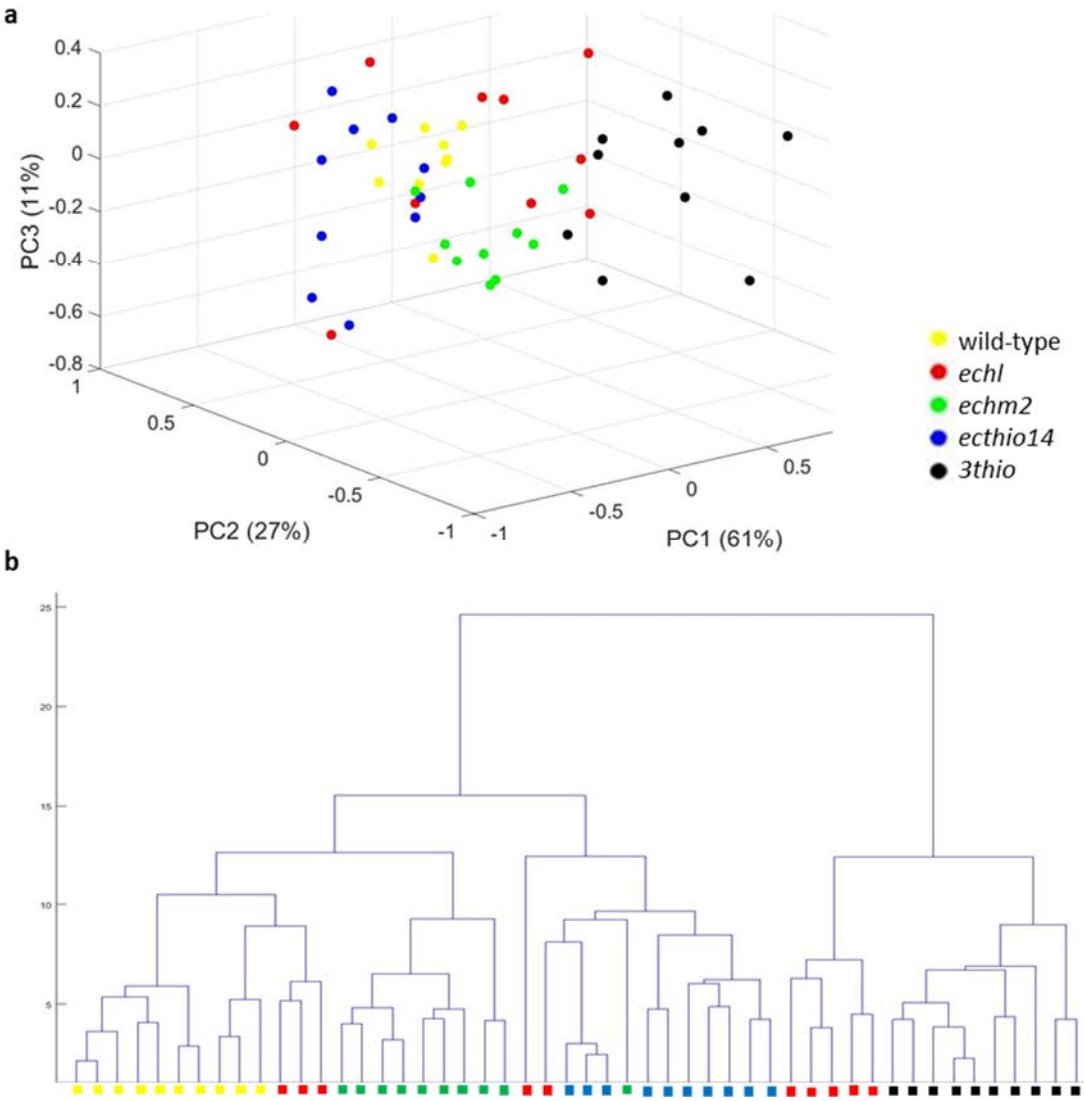

Figure 2.2.6. Clustering using PCA and HCA with normalize net fluxes. Embryos were cultured for 6-10 days with U-^{13}C alanine, U-^{13}C glutamine, 1-^{13}C glucose, or 80% 1,2, 20% U-^{13}C glucose to obtain labeling data to quantify metabolic fluxes by ^{13}C-MFA. Net fluxes were normalized to represent the proportion of total carbon input. PCA (a) and HCA (b) were performed on the normalized net fluxes of wild-type (yellow), *echl* (red), *echm2* (green), *ecthio14* (blue), and *3thio* (black). Colored shapes represent the 10 best fits derived from pseudo datasets.

by the glycolytic reactions interconverting hexose and triose, and the transport reaction to move hexose from the cytosol to the plastid (Table 2.2.S6). Because these fluxes could not be quantitatively well-defined and thus there was large variation within each line for these three fluxes (Vald, Valdp, Vhpt, Table 2.2.S4), we also performed PCA on the normalized fluxes with the hexose transport reaction excluded and

the conversion reactions summed (i.e., this analysis did not distinguish between cytosolic and plastidic glycolysis). With this revision, the PCA clustering was a little cleaner and the variation was largely influenced by the OPP and total glycolysis fluxes.

To further analyze the clustering relationships, in particular to explore which flux patterns are more closely related, we performed hierarchical clustering analysis (HCA) and found that the pseudo replicates of wild-type, *echm2*, *ecthio14*, and *3thio* lines all clustered quite cleanly within genotypes, regardless of whether the analysis distinguished between cytosolic and plastidic glycolysis (Fig 2.2.6b). In addition, when data from wild-type *C. sativa* grown under dark and high-light conditions (Carey et al., in preparation) were included, HCA cleanly distinguished them and the transgenics (Fig 2.2.S6). However, HCA was not able to cleanly distinguish *echl* from the other transgenic lines (Fig 2.2.6, S7), which could be due to higher variation between individual replicates in the *echl* fluxes (Table 2.2.S4).

Discussion

In this study, we supplied developing embryos of transgenic *C. sativa* developing embryos (i.e., *echl*, *echm2*, *ecthio14*, *3thio*) with U-^{13}C glutamine, U-^{13}C alanine, 1-^{13}C glucose, or a combination of 80% 1,2 and 20% U-^{13}C glucose and measured the resulting labeling in fatty acids, protein amino acids, and carbohydrates. In combination with steady-state MFA, this allowed us to better understand the metabolic effects of engineering *C. sativa* to have increased medium-chain fatty acid content (Fig 2.2.1).

As expected, we observed differences in metabolite labeling for the four substrates (Fig 2.2.3, S4, Table 2.2.S1). These differences led to PCA being more effective at cleanly separating the transgenics based on the combination of substrate and type of labeled metabolite. For example, PCA could not cleanly separate *echl* or *ecthio14* when either ^{13}C glutamine or ^{13}C alanine was used as the substrate, but the separation was more distinct when used with fatty acid labeling data than with amino acid data (Fig 2.2.2, S3). On the other hand, 80% 1,2, 20% U-^{13}C glucose enabled PCA to be quite effective at separating the

lines with either metabolite. This could have been simply due to the embryos taking up higher quantities of glucose than either alanine or glutamine (Fig 2.2.S2c), but the MFA flux maps also revealed that the amino acid substrates were largely restricted to the TCA cycle and fatty acid elongation (Fig 2.2.4). This limited how strongly plastidic metabolites could be labeled and thus how distinct the label measurements were from natural abundance levels.

In addition to testing how well PCA could cluster the transgenic lines based on their metabolite labeling, we also performed PCA and HCA on their net fluxes normalized by total carbon uptake (Table 2.2.S4). We found that PCA yielded moderate separation of the lines and HCA was able to extrapolate these patterns to cleanly cluster *echm2*, *ecthio14*, and *3thio* (Fig 2.2.6). Although HCA could not distinguish the *echl* line based on its normalized fluxes, it was able to separate wild-type grown under dark, *in planta*, and high-light conditions (Carey et al., in preparation, Fig 2.2.S6). This could suggest that we did not obtain sufficiently tight estimates for the normalized *echl* fluxes or that the genetic variation between the transgenics did no lead to significantly different fluxes. Regardless, this demonstrates that at times, an organism's environment (e.g., light availability) can lead to more significant changes in metabolic fluxes than genetic differences.

Eccleston and Ohlrogge (1998) previously found that β-oxidation and the glyoxylate cycle were induced by the presence of laurate (C12:0) in developing transgenic *B. napu*s embryos. Here, we measured the metabolic fluxes of *echl* and *ecthio14*, both of which accumulate high levels of laurate (Fig 2.2.1a), but we did not find sufficient support for an active glyoxylate cycle in either line (Fig 2.2.S5). It is possible that the glyoxylate cycle was not prevalent due to the TCA cycle representing small proportion of the total carbon flux (6-7%) and being overshadowed by OPP decarboxylation (over 300% of the total carbon uptake; Table 2.2.S4). Alternatively, the apparent lack of glyoxylate cycle activity could be due to the transgenic *B. napu*s being older than the *C. sativa* embryos (26-30 and 10 DAF, respectively). Poirier and Brumbley (2010) found that laureate does not induce increases in the expression of β-oxidation genes in

A. thaliana leaves, which suggests that the endogenous gene levels are adequate to break down the foreign fatty acid. Because 16-21 DAF *C. sativa* seeds have 1.6-fold ($p < 0.003$) more isocitrate lyase mRNA than 10-15 seeds (Abdullah *et al.*, 2016), it is possible that older *echl* and *ecthio14* embryos would have more active glyoxylate cycles than the early developmental staged embryos that we measured.

We found that differences in metabolite labeling patterns permitted PCA to cleanly cluster *C. sativa* developing embryos that have been engineered to have increased medium-chain fatty acid contents. PCA was able to cleanly separate the transgenic lines when used with fatty acid or amino acid labeling data obtained from labeling with 80% 1,2, 20% U-^{13}C glucose, but could only yield moderate separation when the data was obtained from 100% U-^{13}C glutamine, 100% U-^{13}C alanine (Fig 2.2.2, Fig 2.2.S3). We quantified metabolic fluxes using all labeling data with steady-state MFA (Fig 2.2.4). PCA could only weakly separate the lines using the normalized net fluxes, but HCA could moderately distinguish between the transgenics and cleanly separate the transgenics from wild-type embryos grown under different light availabilities (Fig 2.2.6). Together, these findings point to significant changes in fluxes through central metabolism resulting from genetic modifications of fatty acid composition and environmental variations

Methods

Plant growth

Camelina sativa var. Sunesson plants were grown in a growth chamber (BioChambers; Winnipeg, Manitoba, Canada) on a 16hr/8hr light/dark cycle with plants receiving 125 µmol m^{-2} s^{-1} of light. Seeds were placed in a 25% medium coarseness perlite (Sun Gro; Quincy, MI, USA) and 75% Sure-Mix Potting soil (Michigan grower's products Galesburg, MI, USA) that had been autoclaved when moist. Once seeds germinated, a 2:1 water : nutrient water (one-half strength Hoagland solution) mixture was added twice per week. Stems were tagged daily to track silique age and harvested into 10% chlorox solution to sterilize

siliques for culturing.

Light conditions for culture

In order to establish physiological light conditions for embryos in culture, silique walls at 10 DAF were split and arranged in a single layer within a spectrophotometer cuvette to measure the light transmission efficiency. The same was done with the coats of developing seeds. Percent transmittance was measured at wavelengths of 646nm and 663nm and light levels during culture were set to 10µE, which was estimated to be the level to which embryos are exposed *in planta* during daylight in the growing season.

Culturing conditions

Liquid endosperm composition was analyzed via ^{1}H NMR and GC Mass Spectrometry; media was made using the most abundant sugars and amino acids detected, and the proportions were varied so that the concentrations of carbon and nitrogen sources in prepared media were as follows: glutamine 8 mM, alanine 4 mM, glucose 130 mM, and sucrose 12 mM. 20 mM HEPES was added as a buffering agent, in addition to numerous stock solutions and vitamins with 20% polyethylene glycol (PEG 4000) provided as osmoticum (Schwender & Ohlrogge, 2002). 1 mL medium per well was added to a 6 well plate of 30 mm wells with two filter papers in each (EMD Millipore, USA). 5 embryos at 10 DAF were added to each well and the plates were incubated for 6 days at physiological (10 µE) or high (50 µE) of green light (white light from fluorescent bulbs was filtered through a layer of green celluloid filter). Other embryo cultures were incubated for 9 days in the dark. For labeling experiments, embryos were incubated in media with either U-^{13}C glutamine, U-^{13}C alanine, 1-^{13}C Glucose, or 80%1,2, 20%U-^{13}C glucose replacing the respective amino acid or glucose in unlabeled media at the same concentration.

Uptake rates

After culture period, embryos were removed from wells, rinsed with water and lyophilized. An internal standard of 10 mM methylphosphonic acid was added to the media in each well, the media was transferred to tubes, extracted twice with 2mL volumes of distilled water. Once dry, the PEG was extracted with chloroform. The samples were dissolved in D_2O for ^1H NMR analysis of substrate levels. Fresh samples were compared to "spent" media from culture to determine uptake rates. Carbon conversion efficiency (CCE) was calculated as previously described (Goffman *et al.*, 2005) as the amount of carbon assimilated into biomass (determined by elemental analysis of %C and the dry weight increase during culture) as a percentage of the carbon taken up from the media.

Determination of biomass components

Total lipids were extracted from lyophilized embryos using 2:1 hexane isopropanol kept at 4°C (Hara & Radin, 1978). Embryos were ground using 5 mm tungsten beads in a Retsch (Retsch GmbH, Germany) bead mill run at 30 Hz for three repetitions of 5 minutes. Resulting fractions from this extraction were analyzed for composition by gas chromatography with flame ionization detection (GC-FID).

Proteins were extracted by 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 1% SDS buffer warmed to 42°Cm as described by Schwender and Ohlrogge (2002). Total proteins were then quantified by a DC Protein (Bio-Rad Laboratories; Hercules, CA, USA) colorimetric assay.

Starch fractions were extracted using 0.1 M acetate buffer at 120°C for 1 hour. Starch was then hydrolyzed to glucose by alpha amylase and amyloglucosidase incubated at 50°C for 1 hour. D-glucose was quantified using a total starch assay kit (Megazyme International; Wicklow, Ireland).

Analysis of labeling

0.78 mg dry weight lipid was suspended in 700 µL of d-chloroform for analysis on a Varian Unity-Plus

500MHz spectrometer equipped with a 5 mm ^{13}C-^{1}H dual-purpose probe. ^{1}H spectra were obtained using a 90°C pulse angle and a relaxation delay sufficient for all detected carbon sources at 10 seconds to obtain labeling information. The main peaks of interest in this analysis were the glycerol and sucrose peaks. and labeling information was obtained by using gas chromatography mass spectrometry (GC-MS).

GC MS analysis

0.29 mg dry weight lipid was analyzed by butylamide derivatives for average labeling by GC-MS as described previously (Allen *et al.*, 2007). Butylamine derivatization was achieved by adding n-butylamine and hexane to each sample and incubating at 80°C for 48 hours. Samples were run on a Trace GC ultra/DSQII GC-MS (Thermo Fisher Scientific, USA) with an Agilent VF-23MS (30 m x 0.25 mm ID x 0.27 um film thickness) column and a ramp from 100°C to 260°C in 23 minutes. Resulting peaks were corrected for natural abundance and labeling was calculated for fragments m/z 115 and 128 in the most abundant fatty acids for *C. sativa*: C16:0, C18:0, C18:1, C18:2, C18:3, C20:0, C20:1, C22:0. Analysis of these fatty acids gives labeling coverage of the plastidic and cytosolic pools of acetyl-CoA for MFA compartmentation.

0.29 mg dry weight lipid was also analyzed for glycerol labeling by first transmethylating with 3 M HCl in methanol. Glycerol was separated in the aqueous phase then derivatized with TBDMS and run on a 6890N/5973 GC-MS (Agilent, USA) with a 30 m x 0.25 mm ID x 0.25 um film thickness Restek RTX-5MS column. Stationary phase: 5% diphenyl /95% dimethylpolsiloxane (Agilent GCMS – DB5) and ran using a temperature gradient of 135°C to 325°C at 5°C/minute.

Proteins were hydrolyzed to amino acids in 6 M HCl at 120°C for 24 hours then purified through a cation exchange column of dowex. Derivatization for GC MS analysis was accomplished using MTBSTFA + 1%TBDMCS (Neves & Vasconcelos, 1987; Dauner & Sauer, 2000) and samples were run on Agilent GCMS-DB5. Select ion monitoring (SIM) methods were customized each set of runs to detect only the amino acids and fragments of interest to increase sensitivity as previously described (Dauner & Sauer, 2000).

Peaks were integrated and quantified and natural abundance and labeling amounts were then calculated (Allen *et al.*, 2007).

Starch hydrolyzed as glucose was derivatized with trimethylsilyl (TMS; Sigma, USA) and analyzed by Agilent GCMS-DB5 for labeling in ion fragments m/z 160 (1-2C) and 364 (1-6C). After all extractions were complete, the remaining pellet contained predominantly cell wall material. This portion was hydrolyzed with 6 M HCl at 120°C for 24 hours to yield monosaccharides, and then TMS derivatives were analyzed via GC-MS as described for starch.

GC FID analysis

0.4 mg dry weight lipid was analyzed by GC-FID for quantitative compositional information. 3 M HCl in methanol was added to these fractions for transesterification and then incubated for 2 hours at 80°C. The samples were then neutralized and the methyl esters were extracted with hexane and run on a GCMS-Agilent 6890N/5973 GC-MS with a 30 m x 0.25 mm ID x 0.25 um film thickness Restek RTX-5MS column. Stationary phase: 5% diphenyl /95% dimethylpolsiloxane. Quantification was made possible by heptadecanoic acid (C17:0) internal standard at a known concentration.

Calculating network fluxes

Metabolic fluxes were quantified using the Metabolic Flux Analysis software 13C-FLUX (Wiechert *et al.*, 2001; Wiechert *et al.*, 2015) with the metabolic model developed by Carey *et al*., (in preparation; see Methods S1) for developing embryos of *C. sativa*. Initial starting points (sets of flux values) for determining best-fit fluxes were randomly generated with MATLAB R2016a to span a region with radius representing 10% of the total measured substrate influx surrounding the algebraic center of feasible space. Subsequent starting points were generated by random sampling of flux values within 10% distance of the best fit values determined in the first set of optimizations. This procedure was repeated in order to widely explore the feasible space for best-fit sets of flux values.

Starting flux values were optimized with the *Donlp2* program, where model fit was quantified by minimizing the total residuum, which is the sum of squared differences between measured and predicted parameters of labeling, substrate uptake rates, biomass composition, and growth rates. Each squared deviation was divided by the standard deviation of the respective measurements to ensure that more precisely determined data weighed more heavily. To reduce random effects in data weighting due to the effects of chance on the standard deviations of modest numbers of measurement replicates (typically n= 3-6, always >2) and also the potential for precise but inaccurate measurements, the measured standard deviations were increased before using as model input. The standard deviations of measured fluxes were increased by 10% of the mean, and the standard deviations of labeling data were increased by adding 1% to the measured deviation (mass isotopomers sum to 100% for each metabolite).

At least 100 optimizations per light level or genotype were obtained to maximize the likelihood that the best fit flux sets represented the global best fit. The 20 flux sets with the smallest residuum total were inspected to check that they all converged to similar residuum totals and similar flux values. The starting points corresponding to the 20 lowest residua were identified and for each one, the optimization was repeated using 50 pseudo datasets of measurements that were generated with Monte Carlo sampling using the average and standard deviation of each measurement. The resulting optima were used to obtain the mean and 90% confidence interval for each flux value.

The glyoxylate cycle includes two reactions in addition to reactions of the TCA cycle: gluconeogenesis:isocitrate lyase (ICL; generating glyoxylate and oxaloacetate from isocitrate) and malate synthase (MS; making malate from glyoxylate and acetyl-CoA). In non-photorespiratory tissues, glyoxylate is used to synthesize malate (e.g., in the C4 dicarboxylate pool). Therefore, for modeling purposes the ICL and MS reactions were represented by a single equation: citrate (ABCDEF) + AcCoA (ab) → fumarate (in rapid exchange with malate; CDEF) + oxaloacetate (abAB) (see Methods S1). In addition, we added transport reactions to allow either or both plastidic and cytosolic/mitochondrial AcCoA pools to be used

in the glyoxylate cycle reaction. The transport reactions were set to be irreversible free fluxes and the glyoxylate cycle reaction was set to be an irreversible dependent flux.

Clustering analyses

PCA was performed and plotted with MATLAB R2016a *pca* and *scatter3* function, and HCA was performed and plotted MATLAB R2016a *dendrogram*, and *linkage* functions. To make the HCA results be consistent with the PCA methods (i.e., maximize differences between samples), the linkage function was run with the 'complete' method, which maximizes the distance between objects in different clusters. In addition, to prevent individual data points from skewing the clusters, the linkage function was also run with the 'seuclidean' metric, which standardizes the Euclidean distance between objects by dividing by the standard deviations of their data set.

Supporting Information

Supplemental File 3 – Chapter 2 Methods S1

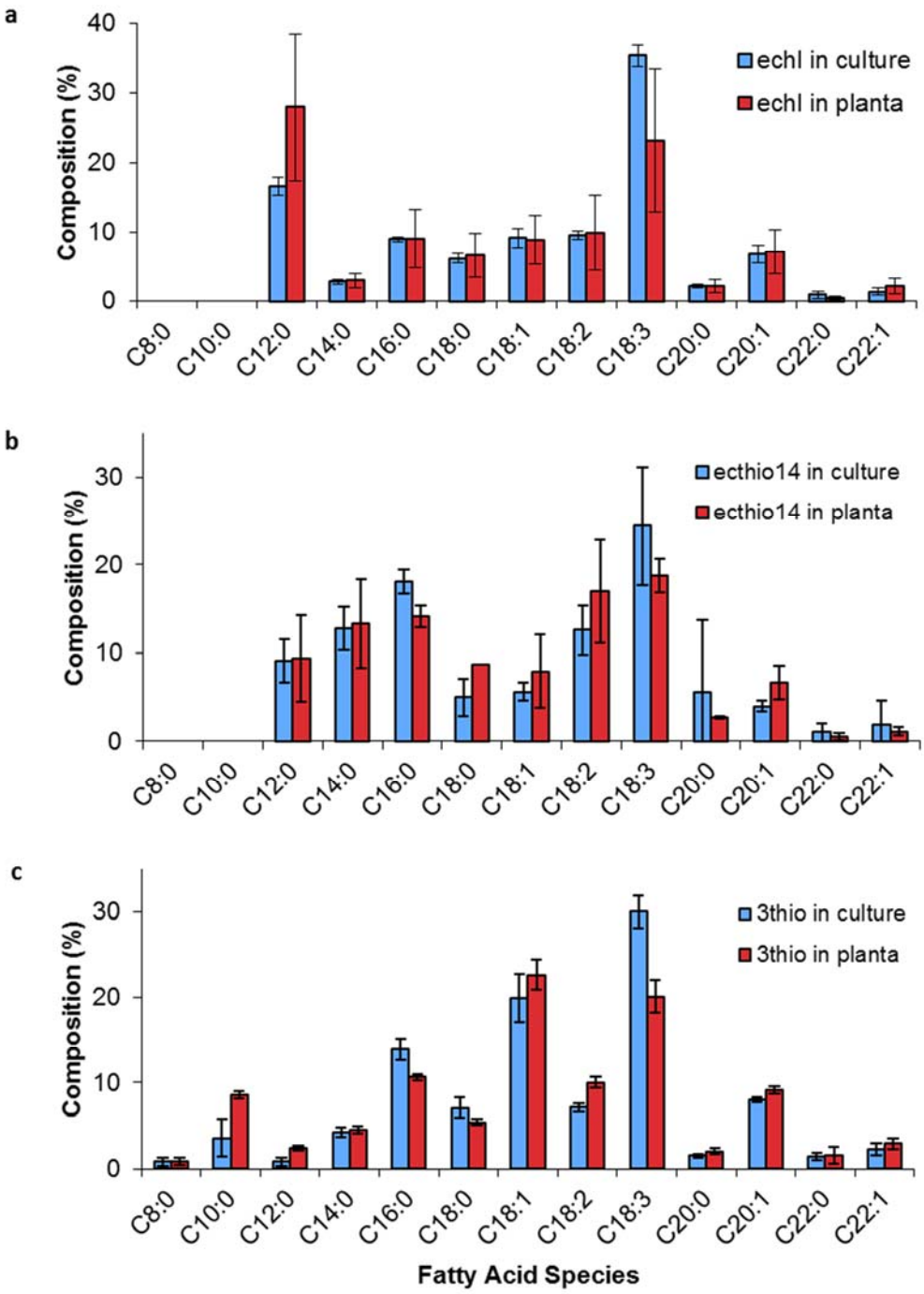

Figure 2.2.S1. Comparison of fatty acid compositions *in culture* vs. *in planta*. Fatty acid methyl esters were tranmethylated from whole *echl* (a), *ecthio14* (b), and *3thio* (c) embryos either after 6 days of culturing (10-16 DAF; *in culture*) or at 16 DAF (*in planta*). Fatty acid species were analyzed with GC-FiD. Bars represent average percent composition and standard deviation (n = 6-9).

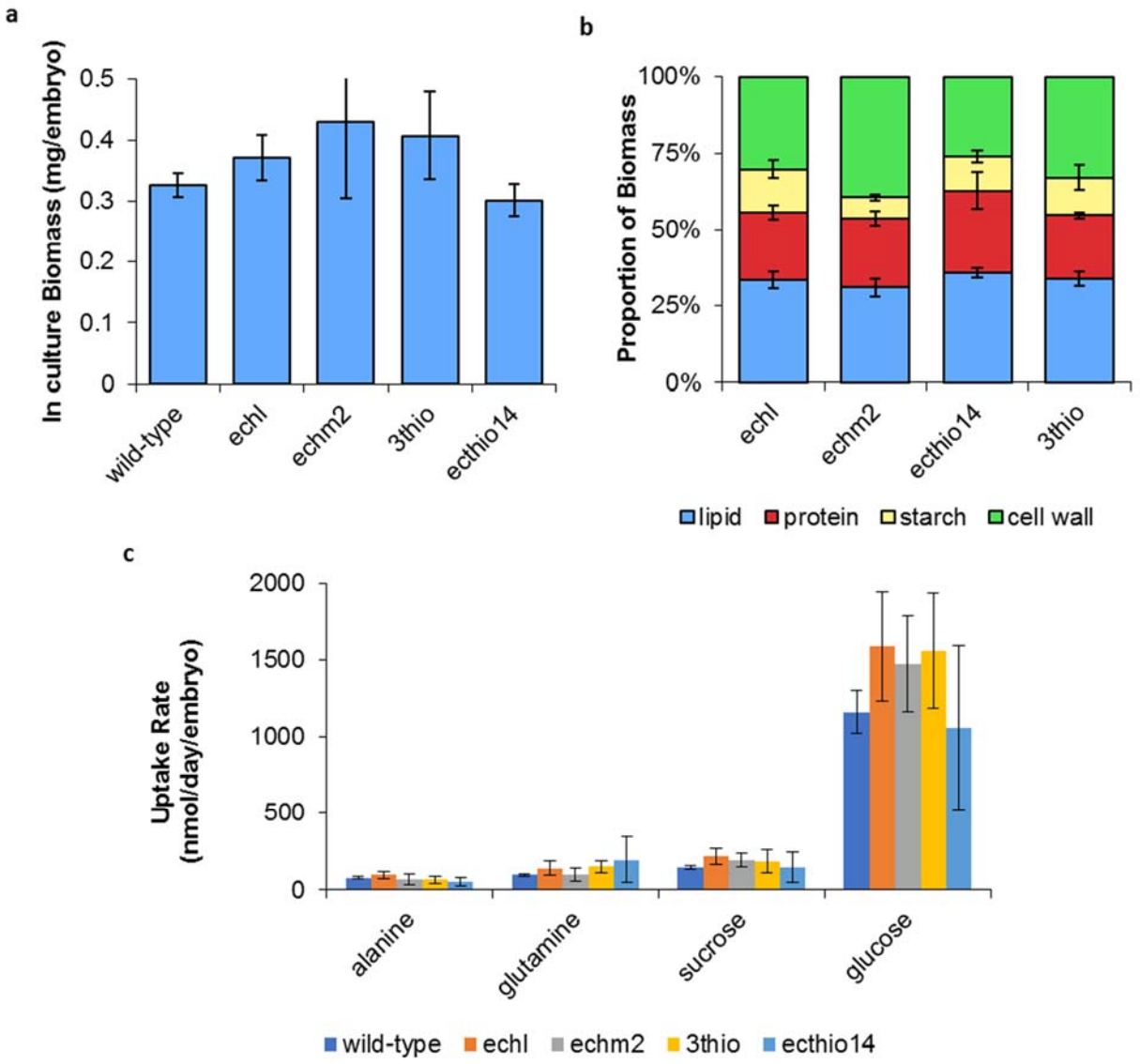

Figure 2.2.S2. Physiological characteristics of *in culture* embryos. (a) Embryos were grown in culture for 6 days (10-16 DPA) and then briefly rinsed in deionized water, frozen in liquid nitrogen, lyophilized, and weighed. (b) Biomass compositions were measured by extracting total lipids, transesterification, and GC-FID analysis (blue); protein was extracted and measured using a colorimeteric bradford assay (red); starch was measured as glucose released when biomass from which soluble metabolites had been extracted was autoclaved in buffer and treated with amylase and glycosylase (yellow). The biomass remaining after extraction of lipids, starch and protein was assumed to be cell wall components. (c) Substrate uptake rates were determined by measuring substrate contents with NMR spectroscopy of fresh and spent media.

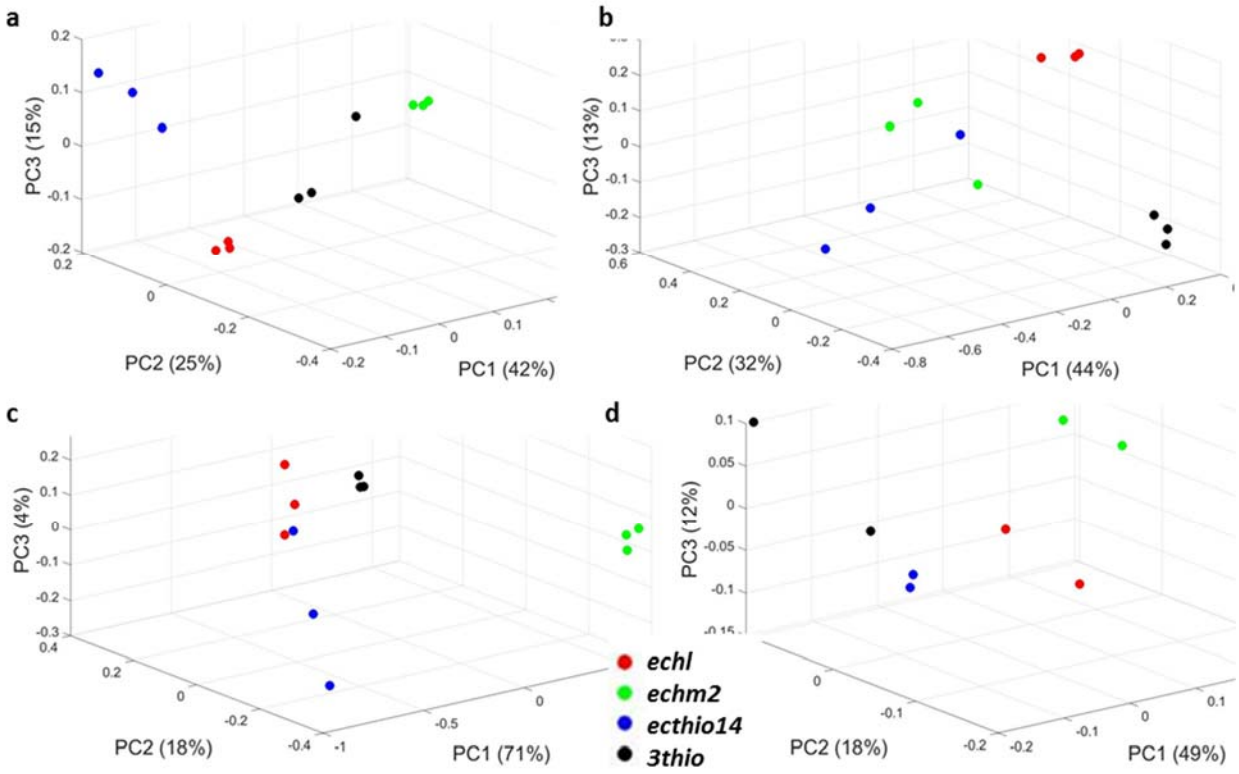

Figure 2.2.S3. Clustering using PCA and amino acid labeling data. Transgenic embryos were cultured for 6 days (10-16 DAF) with (a) 80% 1,2, 20% U-^{13}C glucose, (b) U-^{13}C glutamine, (c) U-^{13}C alanine, or (d) 1-^{13}C glucose. PCA was performed on the amino acid labeling profiles of *echl* (red), *echm2* (green), *ecthio14* (blue), and *3thio* (black). Amino acid labeling was measured by hydrolyzing proteins and measuring label content per amino acid fragment with GC-MS. Circles represent biological replicates (n = 2-3).

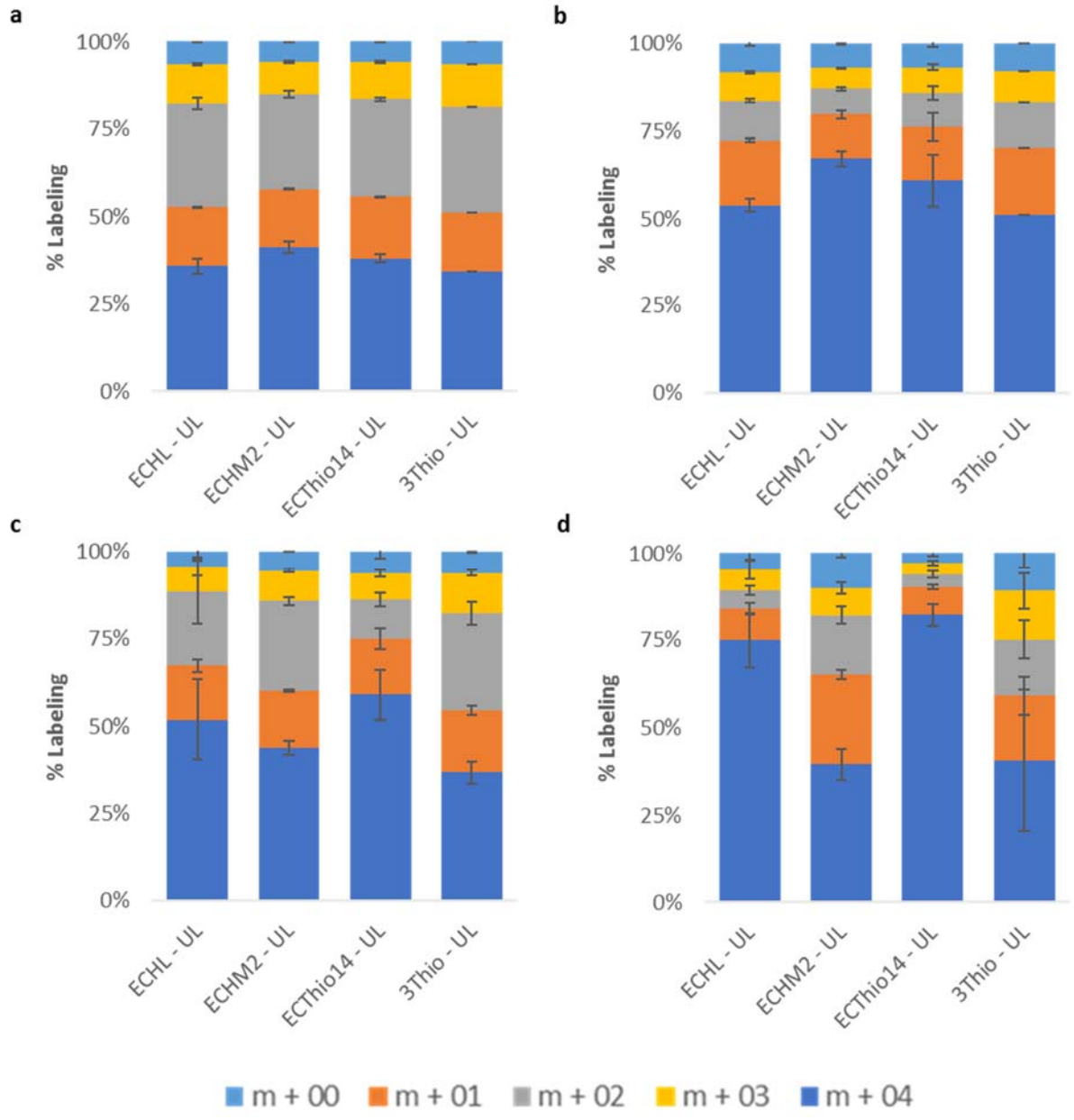

Figure 2.2.S4. Variations in carbohydrate labeling across substrates and transgenics. Transgenic embryos were cultured for 6 days (10-16 DAF) with 80% 1,2, 20% U-^{13}C glucose. Starch (a-b) glucose fragments were obtained by derivatizing the embryos with trimethylsilyl, while cell wall (c-d) glucose fragments were derived from the insoluble pellet remaining after all metabolites were extracted. Labeling in the 364 (a, c) and 319 (b, d) glucose fragments was assessed with GC-MS. Bars represent average percent composition and standard deviation (n = 3-6).

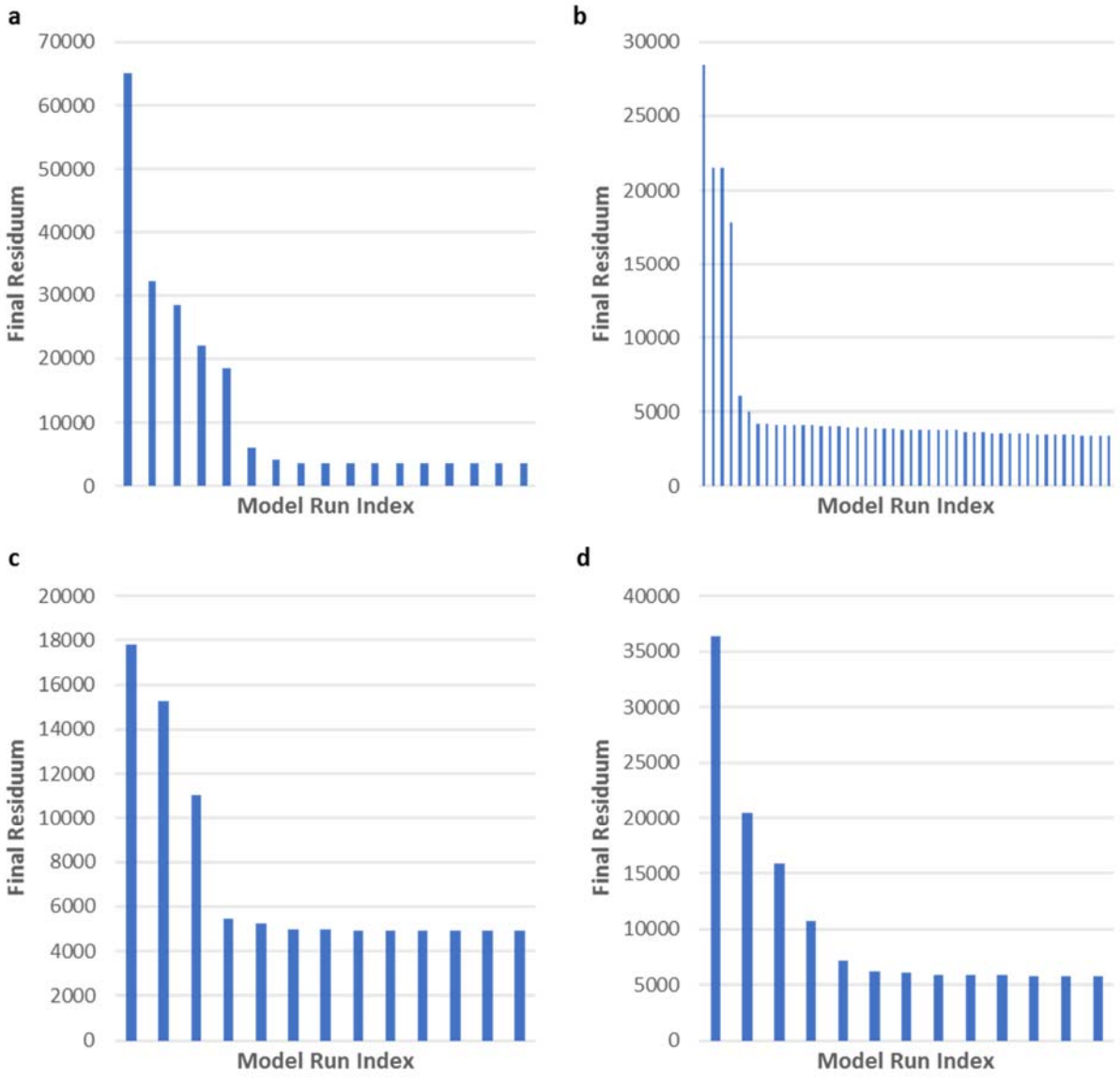

Figure 2.2.S5. Determining if there is an active glyoxylate cycle. Evidence of an active glyoxylate cycle would include its addition to the MFA model significantly improving model fit (i.e., beyond the modest fit inherent to adding a free parameter to the model). The bars represent the pseudo datasets that were able to remain in feasible space to refine the model at least once, starting from at least 10,000 starting points per transgenic. The final residuum in (a) echl decreased from 3676 to 3645, (b) echm2 decreased from 3711 to 3408, (c) ecthio14 decreased from 5014 to 4932, (d) 3thio decreased from 5780 to 5786.

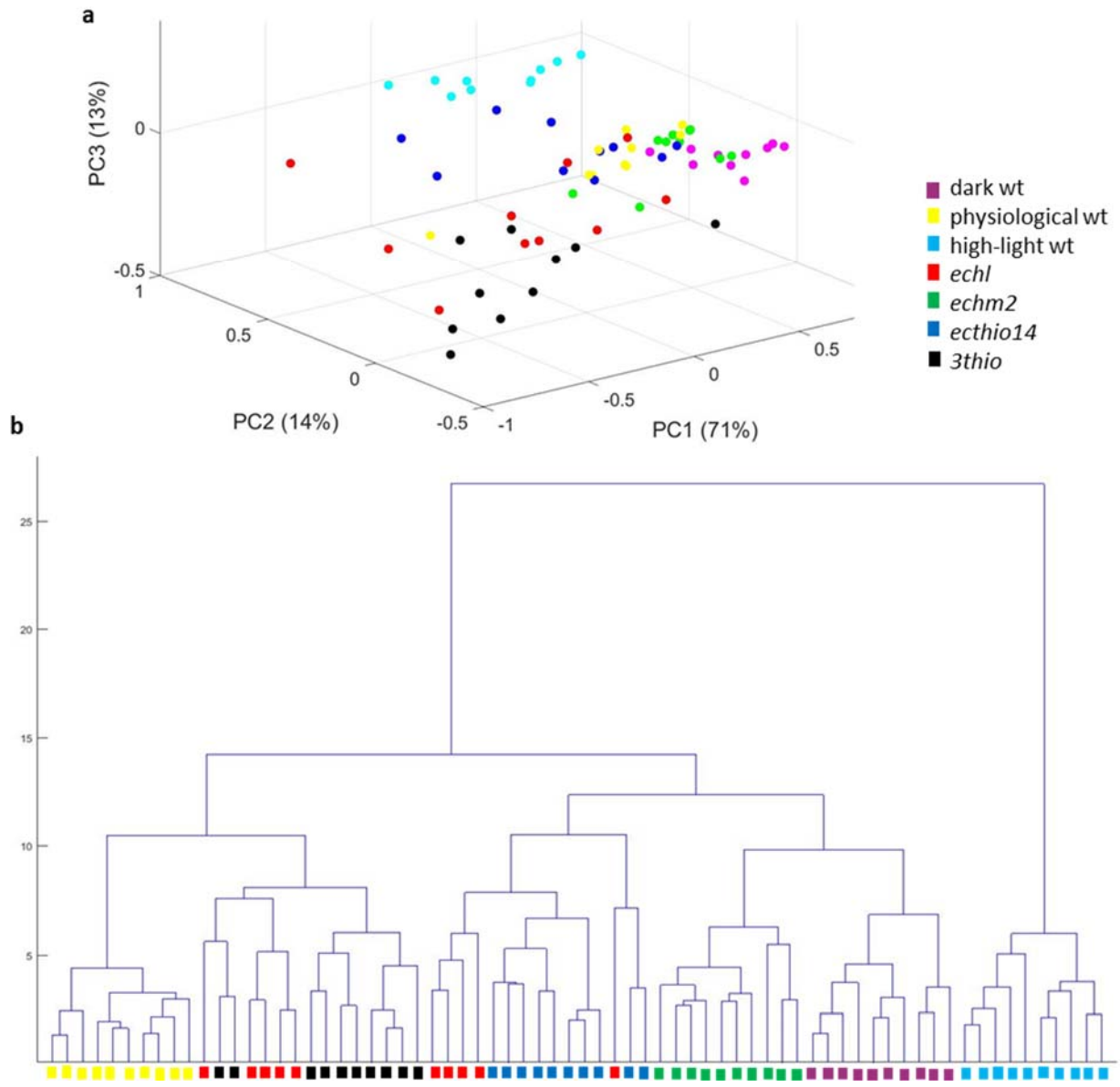

Figure 2.2.S6. Clustering transgenic and wild-type lines using PCA and HCA with normalize net fluxes. Embryos were cultured for 6-10 days with U-^{13}C alanine, U-^{13}C glutamine, 1-^{13}C glucose, or 80% 1,2, 20% U-^{13}C glucose to obtain labeling data to quantify metabolic fluxes by ^{13}C-MFA. Net fluxes were normalized to represent the proportion of total carbon input. PCA (a) and HCA (b) were performed on the normalized net fluxes of wild-type grown in the dark (magenta), *in planta* light (yellow), and high-light (cyan), as well as *echl* (red), *echm2* (green), *ecthio14* (blue), and *3thio* (black). Colored shapes represent the 10 best fits derived from pseudo datasets.

Table 2.2.S1. Isotopomer labeling of fatty acids from cultured embryos

		80% 1,2, 20% U-^{13}C Glucose				100% ^{13}C Glutamine				100% ^{13}C Alanine			
Fragment		*echl*	*echm2*	*cthio14*	*3thio*	*echl*	*echm2*	*cthio14*	*3thio*	*echl*	*echm2*	*cthio14*	*3thio*
						McLafferty fragments							
C8:0	M	-	-	-	-	-	-	-	-	-	-	-	0.94
	M+1	-	-	-	-	-	-	-	-	-	-	-	0.03
	M+2	-	-	-	-	-	-	-	-	-	-	-	0.03
C10:0	M	-	-	-	0.59	-	-	-	-	-	-	-	0.94
	M+1	-	-	-	0.10	-	-	-	-	-	-	-	0.03
	M+2	-	-	-	0.31	-	-	-	-	-	-	-	0.03
C12:0	M	0.60	-	0.58	0.59	0.96	-	0.95	-	0.93	-	0.93	0.91
	M+1	0.11	-	0.12	0.12	0.03	-	0.03	-	0.03	-	0.03	0.04
	M+2	0.30	-	0.30	0.28	0.01	-	0.01	-	0.04	-	0.03	0.05
C14:0	M	0.59	0.59	0.56	0.60	0.94	0.95	-	-	0.92	0.94	0.93	0.93
	M+1	0.11	0.10	0.11	0.11	0.04	0.04	-	-	0.03	0.03	0.03	0.04
	M+2	0.30	0.30	0.32	0.30	0.02	0.02	-	-	0.05	0.03	0.04	0.03
C16:0	M	0.61	0.62	0.60	0.62	0.93	0.94	-	0.94	0.90	0.92	0.91	0.92
	M+1	0.10	0.09	0.10	0.10	0.04	0.04	-	0.04	0.04	0.03	0.04	0.04
	M+2	0.29	0.29	0.30	0.28	0.02	0.02	-	0.02	0.06	0.04	0.05	0.05
C18:0	M	0.62	0.65	0.65	0.65	0.93	0.89	0.92	0.94	0.88	0.90	0.89	0.90
	M+1	0.09	0.08	0.09	0.09	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.04
	M+2	0.29	0.26	0.26	0.26	0.03	0.07	0.04	0.02	0.08	0.06	0.07	0.06
C18:1	M	0.58	0.59	0.58	0.58	0.92	0.92	0.92	0.92	0.88	0.90	0.88	0.89
	M+1	0.11	0.11	0.12	0.12	0.06	0.05	0.06	0.06	0.05	0.05	0.05	0.05
	M+2	0.30	0.30	0.30	0.30	0.02	0.03	0.02	0.02	0.07	0.05	0.06	0.06
C18:2	M	0.58	0.60	0.59	0.60	0.90	0.90	0.91	0.88	0.87	0.90	0.87	0.86
	M+1	0.12	0.12	0.13	0.13	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
	M+2	0.29	0.28	0.28	0.27	0.03	0.03	0.02	0.04	0.06	0.03	0.06	0.06
C18:3	M	0.58	0.76	0.61	0.66	0.88	0.86	0.88	0.88	0.86	0.88	0.85	0.84
	M+1	0.13	0.08	0.12	0.13	0.08	0.07	0.07	0.06	0.04	0.04	0.05	0.06
	M+2	0.29	0.17	0.27	0.22	0.04	0.07	0.04	0.06	0.09	0.08	0.10	0.10
C20:0	M	0.71	0.76	0.71	0.61	0.86	0.65	0.86	0.85	0.85	0.85	0.85	0.84
	M+1	0.09	0.09	0.09	0.12	0.05	0.06	0.05	0.08	0.08	0.05	0.08	0.08
	M+2	0.20	0.15	0.20	0.27	0.09	0.29	0.09	0.07	0.07	0.10	0.08	0.08
C20:1	M	0.67	0.62	0.70	0.63	0.85	0.71	0.84	0.86	0.85	0.85	0.84	0.84
	M+1	0.11	0.12	0.10	0.14	0.06	0.05	0.06	0.08	0.06	0.07	0.06	0.06
	M+2	0.22	0.27	0.20	0.23	0.09	0.25	0.11	0.06	0.09	0.08	0.10	0.10
C22:0	M	-	0.77	0.70	-	-	0.72	0.84	0.87	0.84	0.84	0.82	0.79
	M+1	-	0.08	0.10	-	-	0.05	0.06	0.07	0.06	0.08	0.07	0.08
	M+2	-	0.16	0.20	-	-	0.23	0.10	0.06	0.10	0.08	0.12	0.12
C22:1	M	0.65	0.73	0.67	0.65	0.84	0.71	0.86	0.84	0.83	-	0.82	0.86
	M+1	0.14	0.09	0.12	0.15	0.07	0.06	0.04	0.10	0.06	-	0.07	0.06
						Carboxymethoxy fragments							
C8:0	M	-	-	-	-	-	-	-	-	-	-	-	77%
	M+1	-	-	-	-	-	-	-	-	-	-	-	0.18
	M+2	-	-	-	-	-	-	-	-	-	-	-	0.03
	M+3	-	-	-	-	-	-	-	-	-	-	-	0.01

Table 2.2.S1 (cont'd)

		80% 1,2, 20% U-^{13}C Glucose				100% ^{13}C Glutamine				100% ^{13}C Alanine			
Fragment		*echl*	*echm2*	*cthio14*	*3thio*	*echl*	*echm2*	*cthio14*	*3thio*	*echl*	*echm2*	*cthio14*	*3thio*
C10:0	M	-	-	-	0.35	-	-	-	-	-	-	-	0.88
	M+1	-	-	-	0.28	-	-	-	-	-	-	-	0.08
	M+2	-	-	-	0.24	-	-	-	-	-	-	-	0.04
	M+3	-	-	-	0.13	-	-	-	-	-	-	-	0.01
C12:0	M	0.37	-	0.34	0.32	0.90	-	0.87	-	0.83	-	0.84	0.70
	M+1	0.27	-	0.28	0.37	0.08	-	0.09	-	0.11	-	0.12	0.06
	M+2	0.25	-	0.25	0.21	0.02	-	0.03	-	0.05	-	0.04	0.04
	M+3	0.11	-	0.13	0.10	0.00	-	0.00	-	0.01	-	0.01	0.20
C14:0	M	0.39	0.38	0.36	0.36	0.88	0.90	-	-	0.82	0.88	0.84	0.86
	M+1	0.27	0.27	0.28	0.26	0.09	0.08	-	-	0.11	0.09	0.10	0.09
	M+2	0.24	0.24	0.24	0.25	0.03	0.02	-	-	0.05	0.03	0.05	0.03
	M+3	0.10	0.11	0.12	0.12	0.00	0.00	-	-	0.01	0.00	0.00	0.01
C16:0	M	0.44	0.39	0.40	0.41	0.88	0.88	-	0.87	0.77	0.84	0.80	0.80
	M+1	0.25	0.26	0.27	0.26	0.09	0.09	-	0.09	0.14	0.11	0.13	0.13
	M+2	0.22	0.23	0.23	0.22	0.02	0.03	-	0.03	0.08	0.05	0.06	0.06
	M+3	0.10	0.11	0.10	0.11	0.00	0.00	-	0.01	0.01	0.01	0.01	0.01
C18:0	M	0.42	0.43	0.44	0.45	0.88	0.82	-	0.86	0.75	0.81	0.77	0.79
	M+1	0.26	0.26	0.26	0.24	0.09	0.10	-	0.10	0.14	0.12	0.13	0.13
	M+2	0.22	0.21	0.21	0.20	0.03	0.08	-	0.03	0.09	0.06	0.08	0.07
	M+3	0.11	0.10	0.09	0.11	0.00	0.01	-	0.01	0.02	0.01	0.02	0.01
C18:1	M	0.39	0.39	0.39	0.40	0.89	0.88	0.88	0.88	0.79	0.83	0.79	0.82
	M+1	0.27	0.27	0.27	0.26	0.08	0.09	0.09	0.09	0.13	0.11	0.13	0.12
	M+2	0.23	0.23	0.23	0.22	0.02	0.02	0.02	0.02	0.07	0.05	0.07	0.06
	M+3	0.11	0.11	0.11	0.12	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01
C18:2	M	0.41	0.41	0.41	0.44	0.89	0.88	0.89	0.86	0.80	0.83	0.80	0.81
	M+1	0.26	0.26	0.26	0.25	0.08	0.08	0.08	0.09	0.12	0.10	0.12	0.11
	M+2	0.23	0.22	0.22	0.21	0.02	0.03	0.02	0.04	0.06	0.05	0.07	0.06
	M+3	0.10	0.10	0.10	0.11	0.01	0.01	0.01	0.02	0.02	0.01	0.02	0.02
C18:3	M	0.41	0.53	0.45	0.44	0.89	0.86	0.88	0.75	0.72	0.83	0.73	0.75
	M+1	0.26	0.27	0.25	0.28	0.07	0.08	0.08	0.10	0.15	0.08	0.14	0.14
	M+2	0.22	0.15	0.21	0.19	0.02	0.04	0.02	0.10	0.10	0.06	0.11	0.10
	M+3	0.11	0.05	0.09	0.09	0.02	0.02	0.02	0.05	0.03	0.02	0.02	0.02
C20:0	M	0.46	0.50	0.49	0.44	0.80	0.61	0.79	0.83	0.78	0.79	0.78	0.79
	M+1	0.28	0.30	0.27	0.25	0.10	0.09	0.11	0.10	0.13	0.11	0.13	0.12
	M+2	0.18	0.14	0.18	0.20	0.09	0.28	0.09	0.03	0.06	0.09	0.07	0.06
	M+3	0.08	0.06	0.06	0.10	0.01	0.02	0.01	0.04	0.03	0.01	0.03	0.03
C20:1	M	0.44	0.43	0.48	0.44	0.82	0.62	0.80	0.84	0.78	0.81	0.76	0.78
	M+1	0.30	0.26	0.29	0.29	0.09	0.13	0.09	0.09	0.10	0.11	0.11	0.11
	M+2	0.18	0.21	0.16	0.19	0.09	0.22	0.10	0.06	0.11	0.05	0.12	0.09
	M+3	0.08	0.10	0.07	0.08	0.01	0.04	0.01	0.01	0.01	0.03	0.01	0.01
C22:0	M	-	0.60	0.54	-	-	0.53	0.71	0.65	0.68	0.81	0.64	0.76
	M+1	-	0.21	0.23	-	-	0.23	0.18	0.09	0.18	0.11	0.21	0.10
	M+2	-	0.15	0.17	-	-	0.16	0.09	0.14	0.12	0.05	0.13	0.11
	M+3	-	0.04	0.06	-	-	0.08	0.02	0.13	0.02	0.03	0.03	0.03
C22:1	M	0.49	0.59	0.53	0.44	0.74	0.51	0.71	0.72	0.68	-	0.69	0.77
	M+1	0.27	0.21	0.21	0.27	0.15	0.24	0.17	0.13	0.17	-	0.17	0.15
	M+2	0.19	0.16	0.19	0.19	0.09	0.17	0.09	0.11	0.11	-	0.12	0.07
	M+3	0.05	0.04	0.07	0.10	0.02	0.08	0.02	0.04	0.04	-	0.02	0.02

Table 2.2.S1 (cont'd). Fatty acid labeling was determined by GC-MS for echl, echm2, ecthio14, and 3thio transgenic *C. sativa* embryos labeled with 80% 1,2, 20% U-^{13}C glucose, ^{13}C glutamine, or ^{13}C alanine. Values indicate the average proportion of the fatty acid species with the associated molecular ion peak.

Table 2.2.S2. Contributions of fatty acid isotopomer labeling to PCA

		80/20 Glucose		^{13}C Glutamine		^{13}C Alanine	
Fragment		PC1	PC2	PC1	PC2	PC1	PC2
				McLafferty fragments			
C:14:0	M	0.004	0.099	-	-	0.043	-0.007
	M+1	-0.009	-0.024	-	-	-0.010	-0.022
	M+2	0.005	-0.075	-	-	-0.034	0.029
C:16:0	M	0.019	0.066	-	-	0.060	0.034
	M+1	-0.017	-0.017	-	-	-0.015	-0.012
	M+2	-0.002	-0.049	-	-	-0.045	-0.022
C:18:0	M	0.044	0.089	0.065	-0.025	0.068	0.007
	M+1	-0.020	-0.004	0.005	0.004	-0.020	-0.023
	M+2	-0.024	-0.086	-0.069	0.021	-0.048	0.016
C:18:1	M	0.038	0.009	-0.005	0.001	0.067	-0.010
	M+1	-0.015	0.000	0.007	0.000	-0.011	0.005
	M+2	-0.023	-0.010	-0.002	-0.002	-0.056	0.005
C:18:2	M	0.016	0.070	-0.011	0.108	0.082	0.157
	M+1	-0.022	0.027	0.007	-0.023	0.014	-0.006
	M+2	0.007	-0.097	0.004	-0.085	-0.096	-0.150
C:18:3	M	0.407	0.404	0.034	0.038	0.065	0.212
	M+1	-0.152	-0.052	-0.007	0.060	-0.028	-0.119
	M+2	-0.255	-0.352	-0.027	-0.097	-0.037	-0.093
C:20:0	M	0.318	-0.387	0.324	0.181	0.000	0.059
	M+1	-0.065	0.130	0.010	-0.175	-0.067	-0.106
	M+2	-0.253	0.257	-0.333	-0.006	0.067	0.047
C:20:1	M	-0.119	-0.268	0.230	0.023	0.013	0.062
	M+1	-0.011	0.149	0.034	-0.078	0.049	0.019
	M+2	0.129	0.118	-0.264	0.055	-0.062	-0.081
		-	-	-	-	0.041	0.275
		-	-	-	-	0.056	-0.017
		-	-	-	-	-0.098	-0.257
C:22:1	M	0.234	0.047	0.217	0.188	-	-
	M+1	-0.140	0.007	0.028	-0.276	-	-
	M+2	-0.094	-0.054	-0.245	0.088	-	-
C:14:0	M	0.032	-0.053	-	-	0.158	-0.032
	M+1	0.003	-0.052	-	-	-0.083	0.015
	M+2	-0.031	0.058	-	-	-0.059	0.057
	M+3	-0.003	0.047	-	-	-0.016	-0.040
C:16:0	M	-0.091	-0.037	-	-	0.167	0.164
	M+1	0.030	0.014	-	-	-0.071	-0.083
	M+2	0.037	-0.003	-	-	-0.075	-0.058
	M+3	0.024	0.026	-	-	-0.020	-0.023

Table 2.2.S2 (cont'd)

Fragment		80/20 Glucose PC1	80/20 Glucose PC2	^{13}C Glutamine PC1	^{13}C Glutamine PC2	^{13}C Alanine PC1	^{13}C Alanine PC2
Carboxymethoxy fragments							
C:18:0	M	-0.021	0.094	-	-	0.167	0.021
	M+1	0.031	-0.082	-	-	-0.060	-0.016
	M+2	0.010	-0.033	-	-	-0.082	-0.007
	M+3	-0.021	0.022	-	-	-0.024	0.001
C:18:1	M	-0.017	0.021	0.002	0.010	0.155	-0.031
	M+1	0.009	-0.022	-0.003	-0.009	-0.070	0.011
	M+2	0.012	-0.057	-0.003	0.014	-0.064	0.020
	M+3	-0.003	0.057	0.004	-0.016	-0.021	0.001
C:18:2	M	-0.037	0.103	-0.014	0.144	0.117	0.084
	M+1	0.016	-0.051	0.008	-0.011	-0.059	-0.045
	M+2	0.029	-0.080	0.002	-0.075	-0.052	-0.034
	M+3	-0.007	0.028	0.004	-0.058	-0.007	-0.005
C:18:3	M	0.292	0.120	-0.053	0.658	0.339	0.321
	M+1	0.004	0.096	0.014	-0.127	-0.178	-0.163
	M+2	-0.166	-0.126	0.027	-0.373	-0.136	-0.146
	M+3	-0.129	-0.090	0.013	-0.158	-0.025	-0.012
C:20:0	M	0.153	-0.112	0.317	-0.037	0.042	-0.020
	M+1	0.122	-0.087	0.013	0.035	-0.071	-0.034
	M+2	-0.166	0.083	-0.338	0.130	0.079	0.132
	M+3	-0.110	0.116	0.008	-0.129	-0.050	-0.078
C:20:1	M	-0.039	-0.095	0.331	-0.019	0.167	0.149
	M+1	-0.106	-0.057	-0.061	-0.056	0.003	-0.097
	M+2	0.098	0.095	-0.228	0.094	-0.210	-0.118
	M+3	0.046	0.058	-0.042	-0.019	0.040	0.067
C:22:0	M	-	-	-	-	0.570	-0.374
	M+1	-	-	-	-	-0.360	0.519
	M+2	-	-	-	-	-0.218	-0.141
	M+3	-	-	-	-	0.009	-0.004
C:22:1	M	0.389	-0.289	0.339	0.183	-	-
	M+1	-0.180	0.099	-0.151	0.092	-	-
	M+2	-0.076	-0.024	-0.104	-0.134	-	-
	M+3	-0.133	0.214	-0.083	-0.141	-	-

PCA was performed on fatty acid isotopomer labeling data from transgenic *C. sativa* lines that were cultured for 6 days (10-16 DAF) with U-^{13}C alanine, U-^{13}C glutamine, or 80% 1,2, 20% U-^{13}C glucose, as illustrated in Fig 2.2.2. Values are the coefficients for the first two principal components.

Table 2.2.S3. Isotopomer labeling of amino acids from cultured embryos

		80/20 Glc Labeling		Gln Labeling		Ala Labeling		C1-Glc Labeling	
Fragment		PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Ala123	M	0.09	0.07	0.03	0.01	0.05	0.06	-	-
	M+1	-0.05	-0.01	-0.01	-0.01	0.00	0.01	-	-
	M+2	-0.03	-0.05	-0.01	0.00	0.00	0.00	-	-
	M+3	-0.01	-0.01	-0.01	0.00	-0.04	-0.07	-	-
Ala23	M	0.07	0.05	0.02	0.02	0.04	0.06	-	-
	M+1	-0.05	0.02	-0.01	-0.02	0.00	0.01	-	-
	M+2	-0.02	-0.07	-0.01	0.00	-0.04	-0.07	-	-
Gly12	M	-0.01	-0.07	0.00	0.01	0.00	-0.01	-0.02	0.08
	M+1	0.04	0.08	0.00	0.00	0.00	0.00	-0.01	-0.05
	M+2	-0.03	-0.01	0.00	-0.01	0.00	0.01	0.03	-0.03
Gly2	M	-0.01	-0.05	0.01	0.00	0.00	-0.01	0.03	0.11
	M+1	0.01	0.05	-0.01	0.00	0.00	0.01	-0.03	-0.11
Val12345	M	-	-	0.01	0.00	0.04	0.00	-0.10	0.16
	M+1	-	-	0.00	-0.01	-0.01	0.00	0.11	-0.13
	M+2	-	-	0.00	0.01	-0.02	0.00	0.01	-0.05
	M+3	-	-	0.00	0.00	-0.01	0.00	0.01	-0.01
	M+4	-	-	0.00	0.00	0.00	0.00	-0.01	0.01
	M+5	-	-	0.00	0.00	0.00	0.00	-0.01	0.02
Val2345	M	0.00	0.01	0.02	0.00	0.04	0.00	-0.05	0.17
	M+1	-0.02	0.03	-0.01	-0.01	-0.01	0.00	0.00	-0.12
	M+2	0.03	-0.04	0.00	0.01	-0.03	0.00	0.03	-0.08
	M+3	-0.01	0.02	-0.01	0.00	0.00	0.00	0.00	0.01
	M+4	0.01	-0.02	0.00	0.00	0.00	0.00	0.02	0.02
Leu23456	M	-0.02	-0.02	-0.06	0.10	-0.01	-0.15	-0.03	0.29
	M+1	-0.04	0.02	0.04	-0.09	0.02	0.11	0.02	-0.13
	M+2	0.05	0.01	0.02	-0.01	0.00	0.04	0.01	-0.13
	M+3	-0.02	-0.01	0.00	0.00	0.00	0.00	0.00	-0.04
	M+4	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	M+5	0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00
Leu23456	M	-	-	-0.01	-0.01	0.03	0.01	0.00	0.27
	M+1	-	-	0.00	-0.01	-0.01	0.01	0.04	-0.06
	M+2	-	-	0.00	0.02	-0.01	-0.02	-0.03	-0.20
	M+3	-	-	0.00	0.00	-0.01	0.00	-0.01	0.00
	M+4	-	-	0.00	0.00	0.00	0.00	0.00	0.01
	M+5	-	-	0.00	0.00	0.00	0.00	0.00	0.00
Leu123456	M	-0.02	0.02	-0.04	0.04	0.02	-0.06	-	-
	M+1	-0.01	0.04	-0.01	0.00	-0.01	-0.01	-	-
	M+2	0.03	0.00	0.00	0.02	-0.03	-0.02	-	-
	M+3	-0.04	-0.06	0.04	-0.05	0.02	0.09	-	-
	M+4	0.03	0.00	0.01	-0.01	0.00	0.00	-	-
	M+5	0.00	0.00	0.00	0.00	0.00	0.00	-	-
	M+6	0.01	-0.01	0.00	0.00	0.00	0.00	-	-
Ile23456	M	0.00	0.02	0.03	0.01	0.02	-0.05	-0.01	0.10
	M+1	-0.04	-0.01	0.00	-0.06	0.00	0.03	0.04	-0.06
	M+2	0.03	0.01	0.00	0.02	-0.01	0.02	-0.03	-0.04

Table 2.2.S3 (cont'd)

Fragment		80/20 Glc Labeling		Gln Labeling		Ala Labeling		C1-Glc Labeling	
		PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
	M+3	-0.01	-0.01	-0.03	0.03	0.00	0.00	0.00	0.00
	M+4	0.02	-0.01	0.00	0.00	0.00	0.00	0.00	0.00
	M+5	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ile23456	M	-	-	0.03	0.01	0.03	-0.06	0.01	0.13
	M+1	-	-	0.00	-0.06	-0.01	0.03	0.02	-0.13
	M+2	-	-	0.00	0.02	-0.02	0.02	-0.02	0.01
	M+3	-	-	-0.02	0.03	0.00	0.00	0.00	0.01
	M+4	-	-	0.00	0.00	0.00	0.00	0.00	-0.01
	M+5	-	-	-0.01	0.00	0.00	0.00	0.00	0.00
Ile123456	M	-0.02	0.01	0.03	0.02	0.04	-0.06	-	-
	M+1	0.00	0.02	0.00	-0.04	-0.01	0.02	-	-
	M+2	-0.01	-0.01	-0.01	-0.04	-0.02	0.03	-	-
	M+3	0.03	0.00	0.00	0.06	0.00	0.01	-	-
	M+4	0.00	-0.01	-0.03	-0.01	0.00	0.00	-	-
	M+5	0.01	-0.01	0.00	0.00	0.00	0.00	-	-
	M+6	0.00	0.00	0.00	0.00	0.00	0.00	-	-
Pro12345	M	0.31	0.17	0.02	-0.16	0.04	-0.08	0.36	0.12
	M+1	-0.08	-0.04	0.00	-0.08	-0.01	0.03	-0.26	-0.03
	M+2	-0.13	-0.07	0.00	-0.02	-0.02	0.03	-0.12	-0.03
	M+3	-0.05	-0.04	-0.01	-0.03	0.00	0.01	0.01	0.01
	M+4	-0.03	-0.02	0.00	0.01	0.00	0.00	0.01	0.00
	M+5	-0.01	0.00	0.00	0.28	0.00	0.00	-0.01	-0.07
Pro2345	M	0.31	0.18	0.02	-0.17	-0.02	-0.15	0.32	0.11
	M+1	-0.13	-0.08	0.00	-0.09	0.03	0.09	-0.25	-0.07
	M+2	-0.11	-0.05	-0.01	-0.03	0.00	0.05	-0.05	-0.03
	M+3	-0.06	-0.05	0.00	0.01	0.00	0.01	-0.01	-0.01
	M+4	-0.01	0.00	0.00	0.28	0.00	0.00	0.00	-0.01
Pro2345	M	0.31	0.18	0.02	-0.18	0.03	-0.07	0.28	0.15
	M+1	-0.13	-0.08	0.00	-0.09	-0.01	0.04	-0.25	-0.06
	M+2	-0.11	-0.05	-0.02	-0.03	-0.02	0.02	-0.01	-0.11
	M+3	-0.06	-0.05	0.00	0.01	0.00	0.01	-0.01	0.02
	M+4	-0.01	0.00	0.00	0.28	0.00	0.00	-0.01	0.00
Met12345	M	0.00	-0.07	0.04	0.00	0.02	-0.05	0.09	0.04
	M+1	0.06	-0.01	0.01	-0.03	-0.01	0.02	-0.05	-0.06
	M+2	0.00	-0.04	0.00	-0.03	-0.01	0.03	-0.03	0.00
	M+3	0.02	-0.01	0.00	0.06	0.00	0.00	-0.01	0.04
	M+4	-0.04	0.10	-0.04	0.00	0.00	0.00	0.01	-0.02
	M+5	-0.03	0.04	-0.02	0.00	-0.01	0.00	0.00	0.00
Met2345	M	-0.03	0.02	0.00	0.03	0.01	-0.06	-0.01	0.08
	M+1	0.01	0.05	-0.01	-0.03	-0.01	0.03	0.07	-0.10
	M+2	0.02	-0.01	0.00	0.02	-0.01	0.03	-0.05	0.00
	M+3	-0.01	-0.07	0.00	-0.01	0.00	0.00	0.00	0.02
	M+4	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Met2345	M	0.06	-0.24	0.03	0.00	0.03	-0.05	-0.05	0.01
	M+1	-0.05	0.16	0.00	-0.04	-0.01	0.03	0.04	0.01
	M+2	-0.03	0.06	0.00	0.02	0.00	0.03	0.03	-0.03
	M+3	0.01	0.02	-0.03	0.03	0.00	0.00	0.00	0.00

Table 2.2.S3 (cont'd)

		80/20 Glc Labeling		Gln Labeling		Ala Labeling		C1-Glc Labeling	
Fragment		PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
	M+4	0.01	0.00	0.00	0.00	-0.01	-0.01	-0.01	0.01
Ser123	M	0.02	-0.06	0.01	-0.01	0.00	0.00	-0.06	-0.09
	M+1	-0.03	0.04	-0.01	0.00	0.00	0.00	0.07	0.10
	M+2	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	M+3	0.01	0.02	0.00	0.00	0.00	0.00	-0.01	-0.01
Ser23	M	0.01	-0.06	0.01	-0.01	0.00	-0.01	0.00	0.04
	M+1	-0.03	0.05	-0.01	0.00	0.00	0.01	0.01	-0.02
	M+2	0.02	0.01	0.00	0.00	0.00	0.00	-0.01	-0.03
Ser12	M	0.02	-0.01	0.00	0.01	0.00	0.00	-	-
	M+1	-0.01	0.01	0.00	-0.01	0.00	0.00	-	-
	M+2	-0.01	0.00	0.00	0.00	0.00	0.00	-	-
Thr1234	M	-0.04	0.05	0.03	0.01	0.03	-0.07	0.07	0.17
	M+1	0.01	-0.02	0.01	-0.05	-0.01	0.02	-0.04	-0.20
	M+2	-0.02	-0.04	-0.01	-0.04	-0.02	0.04	-0.02	0.01
	M+3	0.05	0.01	0.00	0.08	0.00	0.01	0.00	0.01
	M+4	0.00	0.00	-0.03	0.00	0.00	0.00	0.00	0.00
Thr234	M	-0.03	0.05	0.04	0.01	0.02	-0.07	0.10	0.08
	M+1	-0.07	-0.07	-0.01	-0.07	-0.01	0.04	-0.07	0.00
	M+2	0.06	0.01	-0.01	0.02	-0.01	0.03	-0.01	-0.06
	M+3	0.04	0.02	-0.02	0.04	0.00	0.00	-0.02	-0.03
Phe12	M	0.03	0.00	0.00	-0.01	0.00	0.00	-	-
	M+1	-0.02	-0.01	0.00	0.00	0.00	0.00	-	-
	M+2	-0.01	0.01	0.00	0.00	0.00	0.00	-	-
Phe123456789	M	0.04	-0.03	0.00	0.01	0.00	0.00	-0.01	0.05
	M+1	0.01	0.01	0.00	-0.01	0.00	0.00	0.05	-0.07
	M+2	0.03	0.01	0.00	0.00	0.00	0.00	0.01	-0.04
	M+3	-0.01	0.00	0.00	0.00	0.00	0.00	-0.03	0.00
	M+4	-0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.02
	M+5	-0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	M+6	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	M+7	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	M+8	-0.01	0.01	0.00	0.00	0.00	0.00	-0.01	0.02
	M+9	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.00
Phe23456789	M	0.02	-0.03	0.00	0.00	0.00	-0.01	0.01	0.05
	M+1	0.00	0.02	0.00	0.00	0.00	0.00	-0.05	-0.04
	M+2	0.03	0.00	0.00	0.00	0.00	0.00	0.03	-0.06
	M+3	-0.03	0.00	0.00	0.00	0.00	0.00	-0.03	0.01
	M+4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
	M+5	-0.02	0.00	0.00	0.00	0.00	0.00	0.00	-0.01
	M+6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	M+7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	M+8	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.03
Asp1234	M	-0.10	0.02	0.02	0.01	0.04	-0.07	-	-
	M+1	0.01	-0.02	0.01	-0.05	-0.01	0.03	-	-
	M+2	0.01	-0.02	-0.01	-0.04	-0.02	0.04	-	-
	M+3	0.07	0.01	0.01	0.09	0.00	0.01	-	-
	M+4	0.01	0.00	-0.02	-0.01	0.00	0.00	-	-

Table 2.2.S3 (cont'd)

		80/20 Glc Labeling		Gln Labeling		Ala Labeling		C1-Glc Labeling	
Fragment		PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Asp12	M	-0.04	-0.07	0.01	-0.04	0.00	-0.06	-	-
	M+1	0.37	-0.22	-0.11	-0.03	0.01	0.03	-	-
	M+2	-0.33	0.28	0.10	0.07	-0.01	0.03	-	-
Asp234	M	-0.13	0.02	-0.01	-0.02	0.05	-0.05	-	-
	M+1	-0.07	-0.06	-0.01	-0.06	-0.02	0.03	-	-
	M+2	0.14	0.03	0.05	0.02	-0.02	0.02	-	-
	M+3	0.06	0.02	-0.03	0.06	0.00	0.00	-	-
Glu12345	M	-	-	-	-	-0.01	-0.30	0.33	-0.15
	M+1	-	-	-	-	0.03	0.20	-0.21	0.14
	M+2	-	-	-	-	-0.01	0.09	-0.10	0.00
	M+3	-	-	-	-	-0.01	-0.01	-0.01	0.01
	M+4	-	-	-	-	0.00	0.00	0.00	0.01
	M+5	-	-	-	-	0.00	0.00	0.00	0.00
Glu2345	M	0.24	0.18	0.00	-0.29	0.04	-0.11	0.27	-0.09
	M+1	-0.14	-0.09	-0.02	-0.11	-0.03	0.07	-0.17	0.05
	M+2	-0.06	-0.04	-0.02	-0.03	-0.01	0.04	-0.10	0.05
	M+3	-0.06	-0.05	0.03	0.02	0.00	0.00	0.00	-0.01
	M+4	0.02	0.00	0.00	0.41	0.00	0.00	0.00	0.00
Lys123456	M	0.06	-0.10	0.10	0.04	-	-	0.04	-0.23
	M+1	-0.19	0.24	-0.10	-0.09	-	-	0.04	0.11
	M+2	0.06	-0.05	0.01	-0.02	-	-	-0.06	0.13
	M+3	0.04	-0.04	0.01	0.05	-	-	-0.02	-0.04
	M+4	0.01	-0.04	-0.01	0.01	-	-	-0.01	0.01
	M+5	0.02	-0.01	0.00	0.00	-	-	0.01	0.02
	M+6	0.00	0.00	0.00	0.00	-	-	0.00	0.00
Lys23456	M	0.01	0.00	0.04	0.01	-	-	0.00	0.23
	M+1	-0.05	0.00	0.00	-0.06	-	-	0.11	-0.23
	M+2	0.03	0.02	0.00	0.02	-	-	-0.09	0.04
	M+3	-0.01	-0.02	-0.02	0.03	-	-	0.03	0.00
	M+4	0.01	0.00	0.00	0.00	-	-	-0.02	-0.03
	M+5	0.01	0.00	0.00	0.00	-	-	-0.02	-0.02
His123456	M	0.00	-0.09	-	-	-0.17	0.46	-0.13	-0.01
	M+1	-0.04	0.08	-	-	0.10	-0.26	0.13	-0.11
	M+2	-0.11	0.00	-	-	0.07	-0.18	-0.01	0.05
	M+3	-0.05	-0.02	-	-	0.00	-0.01	0.00	0.07
	M+4	0.07	0.01	-	-	0.00	0.00	0.02	-0.02
	M+5	0.09	0.02	-	-	0.00	0.00	-0.01	0.00
	M+6	0.04	0.00	-	-	0.00	0.00	0.01	0.01
His23456	M	-	-	-0.08	-0.35	0.15	0.45	-0.09	-0.21
	M+1	-	-	0.04	0.24	0.08	-0.30	0.16	-0.06
	M+2	-	-	0.02	0.07	0.05	-0.07	0.02	0.15
	M+3	-	-	0.03	0.03	-0.11	-0.01	-0.03	0.08
	M+4	-	-	0.00	0.00	-0.02	0.00	-0.02	0.01
	M+5	-	-	-0.01	0.00	-0.15	-0.07	-0.03	0.04
His123456	M	-0.01	-0.06	0.02	-0.01	0.24	0.06	-	-
	M+1	0.02	0.03	-0.05	-0.02	0.00	0.02	-	-
	M+2	0.01	-0.01	0.00	0.00	-0.01	0.01	-	-
	M+3	0.00	0.01	0.01	0.02	-0.01	0.00	-	-
	M+4	0.03	0.00	0.00	0.00	-0.14	-0.06	-	-

Table 2.2.S3 (cont'd)

		80/20 Glc Labeling		Gln Labeling		Ala Labeling		C1-Glc Labeling	
Fragment		PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
	M+5	-0.06	0.01	0.01	0.00	-0.05	-0.02	-	-
	M+6	0.01	0.03	0.00	0.00	-0.02	-0.01	-	-
His23456	M	0.09	0.13	-0.26	0.13	0.55	0.04	-	-
	M+1	0.04	0.16	-0.01	0.03	0.01	0.04	-	-
	M+2	-0.02	0.14	0.01	0.00	-0.01	0.02	-	-
	M+3	-0.03	0.01	0.06	0.08	-0.07	0.05	-	-
	M+4	-0.02	-0.30	0.17	-0.23	-0.23	-0.09	-	-
	M+5	-0.05	-0.15	0.04	-0.02	-0.25	-0.07	-	-
Tyr23456789	M	-0.04	-0.19	0.75	0.01	0.53	-0.03	-	-
	M+1	0.00	-0.02	0.07	0.03	0.00	0.06	-	-
	M+2	0.09	-0.25	0.01	0.02	0.00	0.04	-	-
	M+3	0.03	-0.08	-0.02	0.00	-0.08	-0.03	-	-
	M+4	0.08	-0.05	-0.02	0.00	-0.04	0.00	-	-
	M+5	0.00	0.06	-0.03	-0.01	-0.14	-0.05	-	-
	M+6	0.03	0.04	-0.07	-0.01	-0.05	-0.03	-	-
	M+7	-0.05	0.16	-0.37	-0.07	-0.04	0.04	-	-
	M+8	-0.14	0.33	-0.32	0.02	-0.17	0.01	-	-
Tyr12	M	-	-	-	-	0.02	0.09	-	-
	M+1	-	-	-	-	-0.03	-0.11	-	-
	M+2	-	-	-	-	0.02	0.02	-	-
Tyr123456789	M	-	-	-	-	-	-	-0.02	0.05
	M+1	-	-	-	-	-	-	0.04	0.08
	M+2	-	-	-	-	-	-	-0.02	-0.10
	M+3	-	-	-	-	-	-	0.00	-0.04
	M+4	-	-	-	-	-	-	-0.01	0.00
	M+5	-	-	-	-	-	-	0.01	0.00
	M+6	-	-	-	-	-	-	-0.01	0.00
	M+7	-	-	-	-	-	-	0.00	0.01
	M+8	-	-	-	-	-	-	0.00	0.01
	M+9	-	-	-	-	-	-	0.01	0.00

Amino acid labeling was determined by GC-MS for echl, echm2, ecthio14, and 3thio transgenic *C. sativa* embryos labeled with U-^{13}C alanine, U-^{13}C glutamine, 1-^{13}C glucose, or 80% 1,2, 20% U-^{13}C glucose. Values indicate the average proportion of the amino acid with the associated molecular ion peak.

Table 2.2.S4. Net fluxes normalized by proportion of carbon uptake

flux	wild-type	echl	echm2	ecthio14	3thio
Vg1	0.14 ± 0.00	0.14 ± 0.00	0.13 ± 0.00	0.15 ± 0.00	0.14 ± 0.00
Vg2	0.57 ± 0.01	0.58 ± 0.01	0.62 ± 0.01	0.59 ± 0.01	0.58 ± 0.01
Vala1	0.02 ± 0.00	0.01 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.01 ± 0.00
Va	0.03 ± 0.00	0.03 ± 0.00	0.04 ± 0.00	0.04 ± 0.00	0.03 ± 0.00
Vsuc	0.24 ± 0.01	0.23 ± 0.01	0.20 ± 0.01	0.21 ± 0.01	0.23 ± 0.01
Vsuc1	0.24 ± 0.01	0.23 ± 0.01	0.20 ± 0.01	0.21 ± 0.01	0.23 ± 0.01
Vhk1	0.83 ± 0.01	0.84 ± 0.01	0.85 ± 0.01	0.84 ± 0.01	0.84 ± 0.00
Vhk2	0.12 ± 0.01	0.12 ± 0.01	0.10 ± 0.01	0.10 ± 0.01	0.12 ± 0.00
Vald	-0.13 ± 0.08	0.00 ± 0.09	-0.32 ± 0.06	-0.19 ± 0.05	-0.02 ± 0.08
Vglyco	0.22 ± 0.01	0.20 ± 0.01	0.19 ± 0.01	0.24 ± 0.00	0.13 ± 0.01
Vfasa	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vfasb	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vpk	0.05 ± 0.00	0.06 ± 0.00	0.07 ± 0.00	0.08 ± 0.00	0.06 ± 0.00
Vg3p	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Vgl	-0.04 ± 0.03	-0.03 ± 0.06	-0.04 ± 0.03	-0.12 ± 0.07	-0.03 ± 0.03
Vhpt	1.02 ± 0.08	0.90 ± 0.09	1.20 ± 0.06	1.08 ± 0.05	0.90 ± 0.08
Valdp	-0.26 ± 0.08	-0.44 ± 0.09	-0.12 ± 0.06	-0.15 ± 0.05	-0.56 ± 0.09
Vpkp	0.15 ± 0.01	0.12 ± 0.01	0.11 ± 0.01	0.15 ± 0.01	0.06 ± 0.01
Vstsp	0.02 ± 0.00	0.03 ± 0.00	0.01 ± 0.00	0.02 ± 0.00	0.02 ± 0.00
Vpdhp	0.19 ± 0.01	0.15 ± 0.01	0.16 ± 0.01	0.18 ± 0.00	0.08 ± 0.01
Vfas1	0.12 ± 0.00	0.10 ± 0.00	0.10 ± 0.00	0.12 ± 0.00	0.05 ± 0.01
Vppp1	2.09 ± 0.03	2.19 ± 0.05	2.17 ± 0.02	2.01 ± 0.02	2.40 ± 0.04
Vppp2	1.88 ± 0.03	1.96 ± 0.04	1.94 ± 0.02	1.80 ± 0.01	2.15 ± 0.03
Vppp3	2.09 ± 0.03	2.19 ± 0.05	2.17 ± 0.02	2.01 ± 0.02	2.40 ± 0.04
Vopp	3.77 ± 0.05	3.94 ± 0.08	3.90 ± 0.04	3.62 ± 0.03	4.32 ± 0.07
Vakg	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.01 ± 0.00
Vcs	0.06 ± 0.00	0.09 ± 0.00	0.07 ± 0.00	0.11 ± 0.00	0.11 ± 0.00
Vca	0.05 ± 0.00	0.08 ± 0.00	0.06 ± 0.00	0.09 ± 0.00	0.09 ± 0.00
Vsfa1	0.03 ± 0.00	0.04 ± 0.00	0.03 ± 0.00	0.05 ± 0.00	0.04 ± 0.00
Vsfa2	0.03 ± 0.00	0.04 ± 0.00	0.03 ± 0.00	0.05 ± 0.00	0.04 ± 0.00
Vfum1	0.02 ± 0.00	0.03 ± 0.00	0.03 ± 0.00	0.04 ± 0.00	0.03 ± 0.00
Vfum2	0.02 ± 0.00	0.03 ± 0.00	0.03 ± 0.00	0.04 ± 0.00	0.03 ± 0.00
Vme	0.02 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Vpepc	0.02 ± 0.00	0.02 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
Vacl	0.01 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Vpyrt	0.04 ± 0.00	0.04 ± 0.00	0.06 ± 0.00	0.05 ± 0.00	0.03 ± 0.00
Vco2	0.72 ± 0.01	0.75 ± 0.01	0.74 ± 0.01	0.72 ± 0.00	0.80 ± 0.01
Vwall	0.06 ± 0.00	0.06 ± 0.00	0.07 ± 0.01	0.06 ± 0.00	0.07 ± 0.01
Vsta	0.02 ± 0.00	0.03 ± 0.00	0.01 ± 0.00	0.02 ± 0.00	0.02 ± 0.00
Vglxn	0.07 ± 0.03	0.07 ± 0.06	0.08 ± 0.03	0.16 ± 0.07	0.06 ± 0.03
Vglxu	-0.06 ± 0.03	-0.06 ± 0.06	-0.07 ± 0.03	-0.15 ± 0.07	-0.05 ± 0.03
Vglueff	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
Vasp	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
Vaspeff	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Vser	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Vsereff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Table 2.2.S4 (cont'd)

flux	wild-type	echl	echm2	ecthio14	3thio
Vgly	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vglyeff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vcheff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vala	-0.01 ± 0.00	-0.01 ± 0.00	-0.01 ± 0.00	-0.01 ± 0.00	-0.01 ± 0.00
Valaeff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Varo1	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Varo2	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vakiv	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
Vleu	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
Vleueff	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Vthr	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
Vthrald	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vthreff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vile	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vileeff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vval	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vvaleff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vphe	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Vpheeff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vtyr	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vtyreff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vpro	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vproeff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vlys1	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vlyseff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vhis	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vhiseff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vacoa	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vmet	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Vmeteff	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Varg	0.01 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
Vargeff	0.01 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00

Net fluxes were determined with 13C-MFA for transgenic *C. sativa* developing embryos labeled with U-^{13}C alanine, U-^{13}C glutamine, 1-^{13}C glucose, or 80% 1,2, 20% U-^{13}C glucose. Values are the average net fluxes normalized by total carbon uptake ± 90% confidence intervals as determined from the 20 best fits derived from pseudo datasets.

Table 2.2.S5. Net fluxes in models with and without glyoxylate cycle reactions

flux	without	with	flux	without	with
Vg1	0.134	0.131	Vsta	0.016	0.015
Vg2	0.619	0.624	Vglxn	0.070	0.098
Vala1	0.015	0.015	Vglxu	-0.060	-0.087
Va	0.038	0.038	Vglueff	0.011	0.011
Vsuc	0.195	0.192	Vasp	0.011	0.011
Vsuc1	0.195	0.192	Vaspeff	0.005	0.005
Vhk1	0.849	0.851	Vser	0.006	0.006
Vhk2	0.097	0.096	Vsereff	0.003	0.003
Vald	-0.397	-0.119	Vgly	0.004	0.003
Vglyco	0.188	0.180	Vglyeff	0.003	0.003
Vfasa	0.003	0.003	Vcheff	0.001	0.001
Vfasb	0.003	0.003	Vala	-0.012	-0.012
Vpk	0.071	0.068	Valaeff	0.003	0.003
Vg3p	0.006	0.007	Varo1	0.003	0.003
Vgl	-0.032	-0.060	Varo2	0.003	0.003
Vhpt	1.268	0.986	Vakiv	0.011	0.011
Valdp	-0.050	-0.340	Vleu	0.007	0.007
Vpkp	0.112	0.110	Vleueff	0.006	0.006
Vstsp	0.016	0.015	Vthr	0.006	0.006
Vpdhp	0.153	0.152	Vthrald	0.001	0.001
Vfas1	0.100	0.101	Vthreff	0.003	0.003
Vppp1	2.166	2.183	Vile	0.004	0.004
Vppp2	1.942	1.957	Vileeff	0.003	0.003
Vppp3	2.166	2.183	Vval	0.004	0.004
Vopp	3.896	3.926	Vvaleff	0.004	0.004
Vakg	0.019	0.018	Vphe	0.005	0.005
Vcs	0.067	0.065	Vpheeff	0.004	0.004
Vca	0.055	0.026	Vtyr	0.003	0.003
Vsfa1	0.032	0.020	Vtyreff	0.003	0.003
Vsfa2	0.032	0.020	Vpro	0.004	0.004
Vfum1	0.026	0.020	Vproeff	0.004	0.004
Vfum2	0.026	0.020	Vlys1	0.005	0.004
Vme	0.001	0.005	Vlyseff	0.004	0.004
Vpepc	0.000	-0.003	Vhis	0.002	0.002
Vacl	0.002	0.018	Vhiseff	0.002	0.002
Vpyrt	0.055	0.052	Vacoa	0.001	0.001
Vbeta1	-	-0.001	Vmet	0.001	0.001
Vbeta2	-	0.006	Vmeteff	0.001	0.001
Vicl	-	0.016	Varg	0.005	0.005
Vco2	0.736	0.732	Vargeff	0.005	0.005
Vwall	0.076	0.080			

Net fluxes were determined with 13C-MFA for transgenic *C. sativa* developing embryos labeled with U-^{13}C alanine, U-^{13}C glutamine, 1-^{13}C glucose, or 80% 1,2, 20% U-^{13}C glucose. We tested for an active glyoxylate cycle by fitting the model with and without the cycle to the experimental data. In addition to comparing the resulting total model fits (Fig 2.2.S5), we compared the net fluxes normalized by total carbon intake to determine if there is a significant difference in how the model predicts a major flux to be used. This table lists the average normalized net fluxes for the 20 best fits derived from pseudo datasets.

Table 2.2.S6. Contributions of normalized net fluxes to PCA

flux	PC1	PC2	PC3	flux	PC1	PC2	PC3
Vg1	0.004	0.000	0.000	Vglxn	-0.023	0.158	-0.555
Vg2	-0.006	-0.034	-0.007	Vglxu	0.021	-0.158	0.555
Vala1	-0.003	0.002	0.000	Vglueff	-0.002	0.001	0.000
Va	-0.007	0.002	-0.001	Vasp	-0.002	0.001	0.000
Vsuc	0.012	0.029	0.007	Vaspeff	-0.001	0.000	0.000
Vsuc1	0.012	0.029	0.007	Vser	-0.001	0.000	0.000
Vhk1	0.004	-0.019	-0.003	Vsereff	0.000	0.000	0.000
Vhk2	0.006	0.015	0.004	Vgly	-0.001	0.000	0.000
Vald	0.320	0.475	0.127	Vglyeff	0.000	0.000	0.000
Vglyco	-0.073	0.066	0.025	Vcheff	0.000	0.000	0.000
Vfasa	-0.001	0.001	0.000	Vala	0.002	-0.001	0.000
Vfasb	-0.001	0.001	0.000	Valaeff	-0.001	0.000	0.000
Vpk	-0.009	0.006	-0.010	Varo1	0.000	0.000	0.000
Vg3p	-0.002	0.002	0.001	Varo2	0.000	0.000	0.000
Vgl	0.016	-0.156	0.554	Vakiv	-0.002	0.001	0.000
Vhpt	-0.310	-0.468	-0.119	Vleu	-0.001	0.000	0.000
Valdp	-0.487	-0.335	-0.078	Vleueff	-0.001	0.000	0.000
Vpkp	-0.062	0.052	0.029	Vthr	-0.001	0.000	0.000
Vstsp	-0.002	0.011	0.005	Vthrald	0.000	0.000	0.000
Vpdhp	-0.077	0.056	0.031	Vthreff	0.000	0.000	0.000
Vfas1	-0.051	0.037	0.021	Vile	-0.001	0.000	0.000
Vppp1	0.298	-0.241	-0.078	Vileeff	-0.001	0.000	0.000
Vppp2	0.270	-0.217	-0.070	Vval	-0.001	0.000	0.000
Vppp3	0.298	-0.241	-0.078	Vvaleff	-0.001	0.000	0.000
Vopp	0.537	-0.433	-0.140	Vphe	-0.001	0.000	0.000
Vakg	-0.004	0.001	-0.001	Vpheeff	-0.001	0.000	0.000
Vcs	0.010	0.022	-0.016	Vtyr	-0.001	0.000	0.000
Vca	0.011	0.017	-0.014	Vtyreff	0.000	0.000	0.000
Vsfa1	0.003	0.007	-0.006	Vpro	-0.001	0.000	0.000
Vsfa2	0.003	0.007	-0.006	Vproeff	-0.001	0.000	0.000
Vfum1	0.002	0.006	-0.005	Vlys1	-0.001	0.000	0.000
Vfum2	0.002	0.006	-0.005	Vlyseff	-0.001	0.000	0.000
Vme	-0.002	0.009	0.007	Vhis	0.000	0.000	0.000
Vpepc	-0.001	0.010	0.008	Vhiseff	0.000	0.000	0.000
Vacl	-0.002	0.002	0.001	Vacoa	-0.001	0.001	0.000
Vpyrt	-0.016	-0.002	-0.004	Vmet	0.000	0.000	0.000
Vco2	0.067	-0.044	-0.020	Vmeteff	0.000	0.000	0.000
Vwall	0.000	-0.011	-0.008	Varg	-0.001	0.000	0.000
Vsta	-0.002	0.011	0.005	Vargeff	-0.001	0.000	0.000

PCA was performed on net fluxes normailzed by total carbon uptake from transgenic *C. sativa* lines that were cultured for 6 days (10-16 DAF) with U-^{13}C alanine, U-^{13}C glutamine, 1-^{13}C glucose, or 80% 1,2, 20% U-^{13}C glucose, as illustrated in Fig 2.2.6. Values are the coefficients for the first three pricipal components.

REFERENCES

REFERENCES

Abdullah HM, Akbari P, Paulose B, Schnell D, Qi W, Park Y, Pareek A, Dhankher OP. 2016. Transcriptome profiling of Camelina sativa to identify genes involved in triacylglycerol biosynthesis and accumulation in the developing seeds. *Biotechnology for biofuels* 9(1): 136.

Allen DK, Shachar-Hill Y, Ohlrogge JB. 2007. Compartment-specific labeling information in 13C metabolic flux analysis of plants. *Phytochemistry* 68(16-18): 2197-2210.

Alonso AP, Val DL, Shachar-Hill Y. 2011. Central metabolic fluxes in the endosperm of developing maize seeds and their implications for metabolic engineering. *Metabolic engineering* 13(1): 96-107.

Bansal S, Kim HJ, Na G, Hamilton ME, Cahoon EB, Lu C, Durrett TP. 2018. Towards the synthetic design of camelina oil enriched in tailored acetyl-triacylglycerols with medium-chain fatty acids. *Journal of experimental botany* 69(18): 4395-4402.

Bates PD, Stymne S, Ohlrogge J. 2013. Biochemical pathways in seed oil synthesis. *Current opinion in plant biology* 16(3): 358-364.

Beevers H 1980. The role of the glyoxylate cycle. *Lipids: Structure and Function*: Elsevier, 117-130.

Borisjuk L, Rolletschek H, Radchuk R, Weschke W, Wobus U, Weber H. 2004. Seed development and differentiation: a role for metabolic regulation. *Plant Biology* 6(4): 375-386.

Canvin DT, Beevers H. 1961. Sucrose synthesis from acetate in the germinating castor bean: kinetics and pathway. *Journal of biological chemistry* 236(4): 988-995.

Chen X, Shachar-Hill Y. 2012. Insights into metabolic efficiency from flux analysis. *Journal of Experimental Botany* 63(6): 2343-2351.

Dauner M, Sauer U. 2000. GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing. *Biotechnology Progress* 16(4): 642-649.

Dehesh K, Jones A, Knutzon DS, Voelker TA. 1996. Production of high levels of 8: 0 and 10: 0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from Cuphea hookeriana. *The Plant Journal* 9(2): 167-172.

Dörmann P, Voelker TA, Ohlrogge JB. 2000. Accumulation of palmitate in Arabidopsis mediated by the acyl-acyl carrier protein thioesterase FATB1. *Plant Physiology* 123(2): 637-644.

Durrett TP, Benning C, Ohlrogge J. 2008. Plant triacylglycerols as feedstocks for the production of biofuels. *The Plant Journal* 54(4): 593-607.

Dyer JM, Stymne S, Green AG, Carlsson AS. 2008. High-value oils from plants. *The Plant Journal* 54(4): 640-655.

Eastmond PJ, Graham IA. 2001. Re-examining the role of the glyoxylate cycle in oilseeds. *Trends in plant science* 6(2): 72-78.

Eastmond P, Koláčá L, Rawsthorne S. 1996. Photosynthesis by developing embryos of oilseed rape (Brassica napus L.). *Journal of Experimental Botany* 47(11): 1763-1769.

Eccleston VS, Ohlrogge JB. 1998. Expression of lauroyl–acyl carrier protein thioesterase in Brassica napus seeds induces pathways for both fatty acid oxidation and biosynthesis and implies a set point for triacylglycerol accumulation. *The Plant Cell* 10(4): 613-621.

Ettinger WF, Harada JJ. 1990. Translational or post-translational processes affect differentially the accumulation of isocitrate lyase and malate synthase proteins and enzyme activities in embryos and seedlings of Brassica napus. *Archives of biochemistry and biophysics* 281(1): 139-143.

Goffman FD, Alonso AP, Schwender J, Shachar-Hill Y, Ohlrogge JB. 2005. Light enables a very high efficiency of carbon storage in developing embryos of rapeseed. *Plant Physiology* 138(4): 2269-2279.

Goffman, F. D., Ruckle, M., Ohlrogge, J., & Shachar-Hill, Y. (2004). Carbon dioxide concentrations are very high in developing oilseeds. Plant Physiology and Biochemistry, 42(9), 703-708.

Gupta K, Rehman A, Sarviya R. 2010. Bio-fuels for the gas turbine: A review. *Renewable and Sustainable Energy Reviews* 14(9): 2946-2955.

Hara A, Radin NS. 1978. Lipid extraction of tissues with a low-toxicity solvent. *Analytical biochemistry* 90(1): 420-426.

Hooks MA, Bode K, Couée I. 1995. Regulation of acyl-CoA oxidases in maize seedlings. *Phytochemistry* 40(3): 657-660.

Howard PH. 2009. Visualizing consolidation in the global seed industry: 1996–2008. *Sustainability* 1(4): 1266-1287.

Howard PH. 2015. Intellectual property and consolidation in the seed industry. *Crop Science* 55(6): 2489-2495.

Howe HF, Richter WM. 1982. Effects of seed size on seedling size in Virola surinamensis; a within and between tree analysis. *Oecologia* 53(3): 347-351.

Iskandarov U, Kim HJ, Cahoon EB 2014. Camelina: an emerging oilseed platform for advanced biofuels and bio-based materials. *Plants and bioenergy*: Springer, 131-140.

Kallio P, Pásztor A, Akhtar MK, Jones PR. 2014. Renewable jet fuel. *Current opinion in biotechnology* 26: 50-55.

Kim HJ, Silva JE, Vu HS, Mockaitis K, Nam J-W, 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(14): 4251-4265.

Knutzon DS, Hayes TR, Wyrick A, Xiong H, Davies HM, Voelker TA. 1999. Lysophosphatidic acid acyltransferase from coconut endosperm mediates the insertion of laurate at the sn-2 position of triacylglycerols in lauric rapeseed oil and can increase total laurate levels. *Plant Physiology* 120(3): 739-746.

Kruger NJ, von Schaewen A. 2003. The oxidative pentose phosphate pathway: structure and organisation. *Current opinion in plant biology* 6(3): 236-246.

Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, DeBono A, Durrett TP. 2013. Acyl-lipid metabolism. *The Arabidopsis book/American Society of Plant Biologists* 11.

Liu G, Yan B, Chen G. 2013. Technical review on jet fuel production. *Renewable and Sustainable Energy Reviews* 25: 59-70.

Lu C, Kang J. 2008. Generation of transgenic plants of a potential oilseed crop Camelina sativa by Agrobacterium-mediated transformation. *Plant cell reports* 27(2): 273-278.

Maeda H, Dudareva N. 2012. The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annual review of plant biology* 63: 73-105.

Moser BR. 2010. Camelina (Camelina sativa L.) oil as a biofuels feedstock: Golden opportunity or false hope? *Lipid Technology* 22(12): 270-273.

Neves HCD, Vasconcelos A. 1987. Capillary gas chromatography of amino acids, including asparagine and glutamine: sensitive gas chromatogaphic—mass spectrometric and selected ion monitoring gas chromatographic—mass spectrometric detection of the N, O (S)-tert.-butyldimethylsilyl derivatives. *Journal of Chromatography A* 392: 249-258.

Plaxton WC. 1996. The organization and regulation of plant glycolysis. *Annual review of plant biology* 47(1): 185-214.

Poirier Y, Brumbley SM 2010. Metabolic engineering of plants for the synthesis of polyhydroxyalkanaotes. *Plastics from bacteria*: Springer, 187-211.

Pollard MR, Anderson L, Fan C, Hawkins DJ, Davies HM. 1991. A specific acyl-ACP thioesterase implicated in medium-chain fatty acid production in immature cotyledons of Umbellularia californica. *Archives of biochemistry and biophysics* 284(2): 306-312.

Ratcliffe RG, Shachar-Hill Y. 2006. Measuring multiple fluxes through plant metabolic networks. *The Plant Journal* 45(4): 490-511.

Rosero J, Ortega J, Aldabas E, Romeral L. 2007. Moving towards a more electric aircraft. *IEEE Aerospace and Electronic Systems Magazine* 22(3): 3-9.

Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y. 2004a. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432(7018): 779.

Schwender J, Ohlrogge JB. 2002. Probing in vivo metabolism by stable isotope labeling of storage lipids

and proteins in developing Brassica napusembryos. *Plant Physiology* 130(1): 347-361.

Schwender J, Ohlrogge J, Shachar-Hill Y. 2004b. Understanding flux in plant metabolic networks. *Current opinion in plant biology* 7(3): 309-317.

Schwender J, Shachar-Hill Y, Ohlrogge JB. 2006. Mitochondrial metabolism in developing embryos of Brassica napus. *Journal of Biological Chemistry*.

Shachar-Hill Y. 2013. Metabolic network flux analysis for engineering plant systems. *Current opinion in biotechnology* 24(2): 247-255.

Stanton ML. 1984. Seed variation in wild radish: effect of seed size on components of seedling and adult fitness. *Ecology* 65(4): 1105-1112.

Stoll C, Lühs W, Zarhloul MK, Friedt W. 2005. Genetic modification of saturated fatty acids in oilseed rape (Brassica napus). *European journal of lipid science and technology* 107(4): 244-248.

Thelen JJ, Ohlrogge JB. 2002. Metabolic engineering of fatty acid biosynthesis in plants. *Metabolic engineering* 4(1): 12-21.

Tjellström H, Strawsine M, Silva J, Cahoon EB, Ohlrogge JB. 2013. Disruption of plastid acyl: acyl carrier protein synthetases increases medium chain fatty acid accumulation in seeds of transgenic Arabidopsis. *FEBS letters* 587(7): 936-942.

Umbarger HE. 1978. Amino acid biosynthesis and its regulation. *Annual review of biochemistry* 47(1): 533-606.

Vennesland B, Tchen T, Loewus FA. 1954. Mechanism of enzymatic carbon dioxide fixation into oxaloacetate. *Journal of the American Chemical Society* 76(12): 3358-3359.

Voelker TA, Worrell AC, Anderson L, Bleibaum J, Fan C, Hawkins DJ, Radke SE, Davies HM. 1992. Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. *Science* 257(5066): 72-74.

Weber H, Borisjuk L, Wobus U. 2005. Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* 56: 253-279.

Wiechert W, Möllney M, Petersen S, de Graaf AA. 2001. A universal framework for 13C metabolic flux analysis. *Metabolic engineering* 3(3): 265-283.

Wiechert W, Niedenführ S, Nöh K. 2015. A Primer to 13C Metabolic Flux Analysis. *Fundamental Bioengineering*: 97-142.

Wightman F, Forest JC. 1978. Properties of plant aminotransferases. *Phytochemistry* 17(9): 1455-1471.

Zubr J. 1997. Oil-seed crop: Camelina sativa. *Industrial crops and products* 6(2): 113-119.

CHAPTER 3

Dynamic modeling of seed oil biosynthesis

Preface

As part of an ongoing project, I collaborated with Dr. Mike Pollard (MP) to analyze triacylglycerol (TAG) biosynthesis in developing embryos of *Camelina sativa*. MP developed labeling and analytical methods to track carbon movement among lipid classes, sub-classes, and individual fatty acid species (Pollard et al 2015a,b). This chapter describes my work on this project, crediting MP as appropriate for the experimental results he provided for my modeling work.

Introduction

Seed oil is of great economic importance due to its use in food, lubrication, detergents, cosmetics, chemical feedstocks, and biofuels (Durrett *et al.*, 2008; Dyer *et al.*, 2008; Mosiewicki & Aranguren, 2013). The major seed storage lipid is triacylglycerol (TAG), which is composed of a glycerol backbone (i.e., glyceryl group) and three fatty acid hydrocarbon chains (i.e., acyl groups; Bates *et al.*, 2013; Li-Beisson *et al.*, 2013). The composition of these chains (*e.g.*, length and degree of saturation) influences the physical and chemical properties of the oil (Durrett *et al.*, 2008; Arab-Tehrany *et al.*, 2012). For example, saturated fatty acids have better stability to oxidation than unsaturated ones (Durrett *et al.*, 2008). This is important for oils used in frying and other high-temperature applications. Understanding lipid biosynthesis is an important step in engineering seeds to have increased oil content and/or altered fatty acid composition.

In the TAG *de novo* synthesis pathway (Chapman & Ohlrogge, 2012; Bates *et al.*, 2013; Li-Beisson *et al.*, 2013), two fatty acids are added in succession to glycerol 3-phosphate (G3P) by *sn*-1 glycerol-3-phosphate acyltransferase (GPAT) and lysophosphatidic acid acyltransferase (LPAAT)) using acyl CoA as the acyl donor to produce phosphatidic acid (PA). The phosphate group in the *sn*-3 position of PA is removed by phosphatidic acid phosphatase (PAP) to produce diacylglycerol (DAG), which can be acylated by diacylglycerol acyltransferase (DGAT) to produce TAG. In addition, other enzymes, such as phospholipid:diacylglycerol acyltransferase (PDAT), which acrylates DAG to make TAG using a

phospholipid as the acyl donor, and other intermediates, such as phosphatidylcholine (PC) and other phospholipids, are also involved in TAG biosynthesis in oilseeds (Chapman & Ohlrogge, 2012; Bates *et al.*, 2013; Li-Beisson *et al.*, 2013).

PC has an important role in TAG biosynthesis because the PC *sn*-2 position is the major site of fatty acid desaturation, hydroxylation, and other fatty acid modifications (Bates *et al.*, 2013). This site plays a critical role in acyl editing, where the PC *sn*-2 fatty acid is removed, yielding lyso-PC, and replaced by lyso-PC acyltransferase (LPCAT), which permits the initial PC fatty acid to be available for *de novo* TAG biosynthesis (e.g., G3P acylation; Bates *et al.*, 2012). Consequently, PC significantly contributes to the production of TAG with polyunsaturated or other modified fatty acids. Despite extensive research in recent decades on genes and enzymes involved in plant lipid metabolism (Mudd, 1980; Benning, 2009; Bates *et al.*, 2013; Li-Beisson *et al.*, 2013), it is still unclear how the PC and DAG pools are related in the biosynthetic network. PC is synthesized from DAG by the addition of the choline group by PC:DAG phosphocholine transferase (PDCT) or cholinephosphotransferase (CPT; Chapman & Ohlrogge, 2012). These reversible reactions have been well established, but there is not yet a clear consensus on whether there is rapid exchange between PC and DAG pools (Durrett *et al.*, 2008; Chapman & Ohlrogge, 2012), or if PC is synthesized from one DAG pool and used to produce a second, independent DAG pool (Bates *et al.*, 2013). Possible network topologies can be computationally tested with carbon labeling and mathematical models to determine which network topology best explains measurements.

Bates *et al.* (2009) constructed the first, and to date only, quantitative flux map of TAG biosynthesis in an oilseed (soybean). Using a simple kinetic model, and ^{14}C glycerol and ^{14}C acetate to track, respectively, the lipid backbone flux from G3P and the addition of newly synthesized or recently elongated fatty acids, they found that PC and DAG each had active and bulk pools. The active pools were predicted to be small and used to synthesize lipids, while the bulk pools were predicted to be large and used for storage or synthesis. Lastly, it was predicted that there were two active DAG pools, one before

and one after PC. These researchers also found that most newly formed fatty acids were used in acyl editing. The detached fatty acids were predicted to enter a recycled fatty acid (in the form of acyl-CoA) pool that also contained newly synthesized fatty acids and could be used for G3P acylation, while DAG acylation was predicted to selectively use newly synthesized fatty acids.

In this work, cultured *C. sativa* embryos were supplied with ^{14}C-glycerol to trace how glyceryl groups move through the biosynthesis network. Dynamic, first-order rate equation models were constructed to represent potential network topologies and tested using the labeling data to identify the likely network topology and to derive a flux map. Network topologies were compared using the sum of squared errors between predicted and measured ^{14}C contents in PC, DAG, and TAG over the time-course. The best fit was found when one DAG pool exchanges with the active PC pool and also feeds into a second DAG pool that serves as the immediate precursor pool for TAG synthesis (Fig 3.S2). In addition, we performed short-term ^{14}C-acetate labeling to assess how PC acyl editing is involved in this pathway and we explain how these findings can be integrated with the glyceryl flux map.

Results

Determining network topology with glyceryl labeling and mathematical models

Following the establishment of methods to culture *C. sativa* developing embryos (Chapter 2), and for isotopic labeling of lipid precursors and for analyzing the distribution of label in different lipids (Pollard et al 2015a,b), MP labeled cultured embryos with 1.5 mM U-^{14}C/$^{13}C_3$ glycerol for 0.5 to 24 hr, and measured ^{13}C and ^{14}C glyceryl and acyl label content in PC, DAG, and TAG. Exogenous glycerol is believed to serve as a substrate for a kinase that uses it to produce G3P and thus labels the glyceryl groups of lipids. In addition to labeling the glyceryl backbone of glycerolipids, glycerol also seems to serve as a substrate for G3P dehydrogenase, which produces a triose phosphate that can be used for acyl CoA synthesis. Consequently, ^{14}C-glycerol labels both glyceryl and acyl groups. MP isolated these two fractions and analyzed their

separate labeling patterns. The 0.5-24 hr time-course spanned the initial linear flux through steady-state asymptotic labeling (Fig 3.1) and could be used to estimate some relationships between active lipid pools and other network properties. Using first order kinetic modeling and MP's ^{14}C glyceryl labeling data (Table 3.3.S1), I extended these estimates into a flux map of glycerolipid metabolism during TAG biosynthesis by 1) finding a network topology that fits the glyceryl measurements, and 2) quantifying rate constants and pool sizes to determine reaction fluxes.

The model that best accounted for the glyceryl labeling data (Fig 3.1) contained two active and one bulk (e.g., storage) DAG pools (Fig 3.2). The first active DAG pool could rapidly exchange with an active PC pool, with a reversible flux of 0.074 μCi/hr/embryo, while the second active DAG pool could serve as a precursor for TAG synthesis (0.053 μCi/hr/embryo) or be slowly diverted into the bulk DAG pool. The network also contained two PC pools: one active and one bulk. The flux from active to bulk PC pools was much smaller (0.002 μCi/hr/embryo) than the *de novo* TAG synthesis fluxes.

Tracking fatty acids involved in de novo synthesis or acyl editing

In order to track newly synthesized fatty acids as well as those recently elongated, I performed short-term acyl labeling with ^{14}C-acetate. Because the labeling time-course was to be significantly shorter than the above experiment, more embryos per culture plate well were used than in the longer ^{14}C-glycerol experiment (where 5 embryos were cultured per well). To determine how many embryos could be simultaneously labeled in a single well without inducing changes in their lipid metabolism due to crowding stress, between 3 and 24 embryos per well were incubated for 6 hr with ^{14}C-acetate, and total lipid and fatty acid species labeling was measured. Between 3 and 18 embryos/well, there was no significant difference in the amount of label per embryo in the total lipid extract (Fig 3.3a) or within the more active fatty acid species (Fig 3.3b; Pollard *et al.*, 2015a). At 24 embryos/well, there was a marked decrease in total lipid content per embryo (Fig 3.3a), which indicated the embryos were too crowded. Consequently,

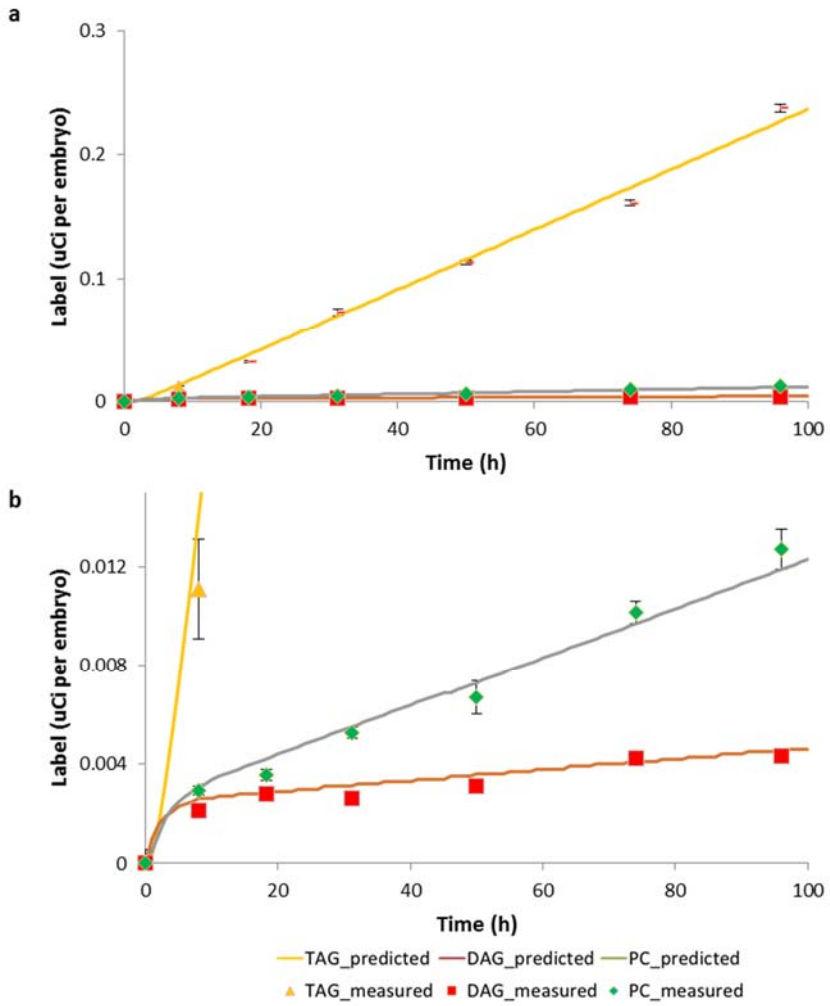

Figure 3.1. Comparison of model predictions to measurements. *C. sativa* embryos were cultured with ^{14}C glycerol to track the movement of the glyceryl backbone in the TAG biosynthesis pathway. Data points indicate the measured average ± standard deviation (n=3) glyceryl label content for TAG (yellow), DAG (red), and PC (blue). Data were obtained by MP. The curves represent how much glyceryl label was predicted for those lipid classes when the network topology in Fig 3.2 was modeled. (b) A zoomed-in view of Panel A highlighting the DAG and PC fits.

in the short-term ^{14}C-acetate labeling experiment, only 18 embryos were cultured per well.

Cultured *C. sativa* embryos were labeled with 1 mM ^{14}C-acetate for 0.5 to 16 min to track how newly synthesized or elongated fatty acids are incorporated into seed lipids. First, we measured the label content in total lipid extracts and found that ^{14}C-acetate is incorporated linearly during this timeframe (Fig 3.4). Furthermore, TAG accumulates label at a high linear rate (0.07 DPM/µL) that persists through 16

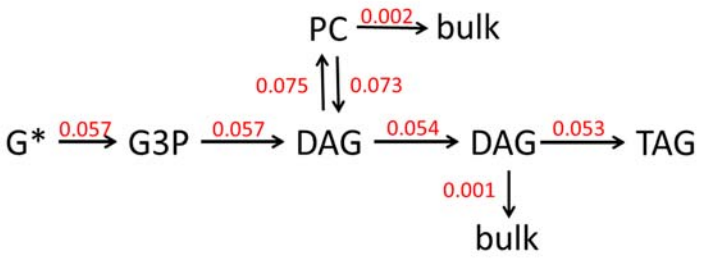

Figure 3.2. Current map of glyceryl backbone fluxes in the TAG biosynthesis network. Potential networks of the TAG biosynthesis pathway were tested with first-order reaction kinetic models. This is the topology and fluxes that had the best fit to the ^{14}C glyceryl time-course data; the model fit is shown in Fig 3.1. "Bulk" indicates end product PC and DAG pools that accumulate for storage and/or structural functions. Numbers indicate the fluxes in μCi/hr/embryo.

min, while PC and DAG have slower initial rates (0.03 and 0.02 DPM/μL, respectively). Neither DAG nor PC show significant saturation of their label content, as would be expected for precursor pools. The lack of saturation could be due to the time-course being too short, at least one of the pools being quite large, or the existence of bulk PC and DAG pools, as suggested by the ^{14}C glyceryl model (Fig 3.2). To further understand the route in which labeled fatty acids are being incorporated into the network, measured fatty acid positional labeling was also measured (Fig 3.S1).

The *de novo* synthesis of PC involves the addition of two fatty acids to G3P. These reactions are not known to be specific for labeled or unlabeled fatty acids, thus *de novo* PC should contain 1:1 positional fatty acid labeling. On the other hand, in PC acyl editing, LPCAT preferentially replaces the PC *sn*-2 fatty acid with a fatty acid that may have been recently elongated, which would result in higher labeling in the *sn*-2 position than in *sn*-1. Consequently, to differentiate between PC labeled in the *de novo* TAG pathway and PC labeled via acyl editing, fatty acid positional labeling in PC was measured using phospholipase A_2 (PLA_2). This enzyme hydrolyzes the *sn*-2 fatty of PC, resulting in a free fatty acid and lyso-PC, which can be separated with thin-layer chromatography (TLC, Fig 3.S1a). We found that labeling in *sn*-1 PC began to saturate by 4 min, while *sn*-2 labeling increased linearly through 16 min (Fig 3.5a). This suggested that the PC pool used for acyl editing was significantly larger than the active PC pool used for *de novo* TAG

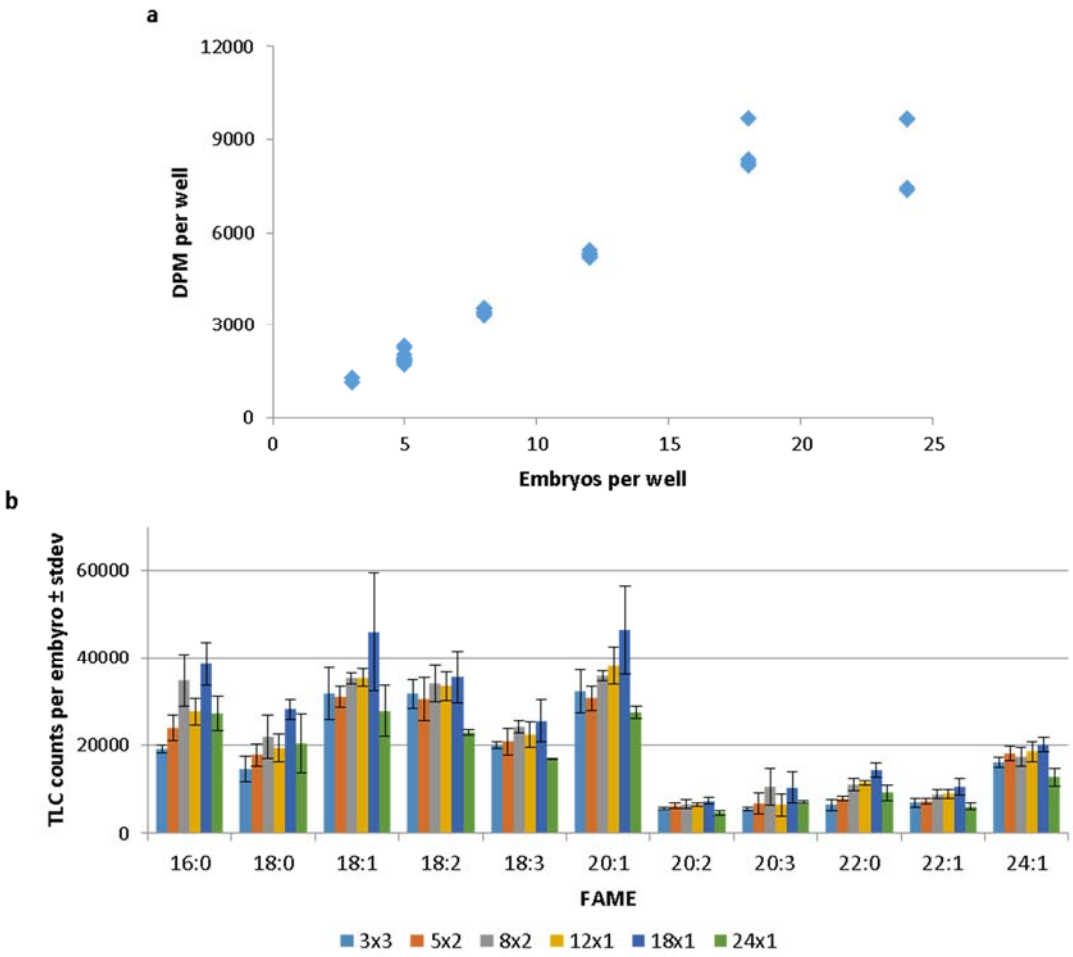

Figure 3.3. Effect of increasing embryo culturing density on lipid content. *C. sativa* embryos were labeled with ^{14}C acetate in culture wells containing 3 to 24 embryos. (a) The label content in total lipid extracts. Data points represent biological replicates (n = 2-4). (b) Average label content in fatty acid methyl ester species obtained from transmethylating the total lipid extract. Legend descriptions indicate the number of embryos per well x the number of wells per biological replicate; e.g., '8x2' indicates that 8 embryos were grown per well and two such well were combined before lipid extraction and thus represent one 16 embryo biological replicate.

synthesis.

Discerning how fatty acids are incorporated into TAG

To further understand how the PC acyl editing pool is related to *de novo* TAG synthesis, fatty acid positional labeling in DAG and TAG was measured. We found that labeling in both *sn*-1 and *sn*-2 DAG

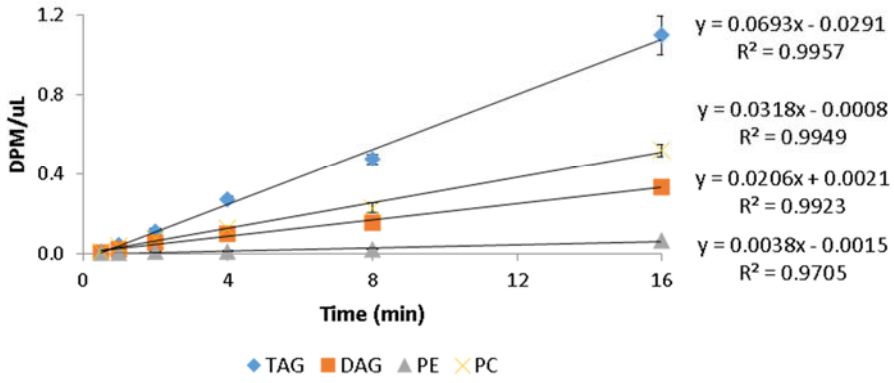

Figure 3.4. Short-term fatty acid labeling per lipid class. *C. sativa* embryos were labeled with ^{14}C acetate for 0.5 to 16 min. TAG (blue), DAG (orange), PC (yello), and phosphatidylethanolamine(PE, green) were separated with TLC and assayed for label content with scintillation counting. Data points indicate the measured average ± standard deviation (n=3). Lines and equations represent the results of linear regression analysis.

increased throughout the 16 min time-course (Fig 3.5b). DAG initially had more *sn*-2 label, but the ratio started approaching (though did not reach) a 1:1 ratio by 16 min (Fig 3.5b). The high initial *sn*-2 DAG labeling supports the acyl editing PC pool being a precursor to DAG. The trend toward 1:1 labeling was consistent with there also being a DAG pool that served as a precursor to PC and was significantly larger than the post-PC acyl editing DAG pool to mitigate the accumulation of *sn*-2 label. The precursor DAG pool may also be significantly larger than the active PC pool used for *de novo* TAG synthesis because *sn*-1 PC label leveled off at 3 DPM/µL, while *sn*-1 DAG label reached 15 DPM/µL at 16 min (Fig 3.5a-b). On average, only 4% of the TAG fatty acid labeling was in the *sn*-2 position (Fig 3.5c). Labeling in the *sn*-1 and *sn*-3 positions were indistinguishable due to the isometric nature of TAG, but it is reasonable to assume the *sn*-1 position contained at most 4% because both PC and DAG (the immediate precursor of TAG) had more *sn*-2 label than *sn*-1 and additions of fatty acids to G3P are not expected to be selective for labeled or unlabeled fatty acids (Fig 3.5a-b). Consequently, the remaining 92% of TAG label is expected to be in the *sn*-3 position, which suggests most of the TAG was derived from the addition of a labeled fatty acid to a

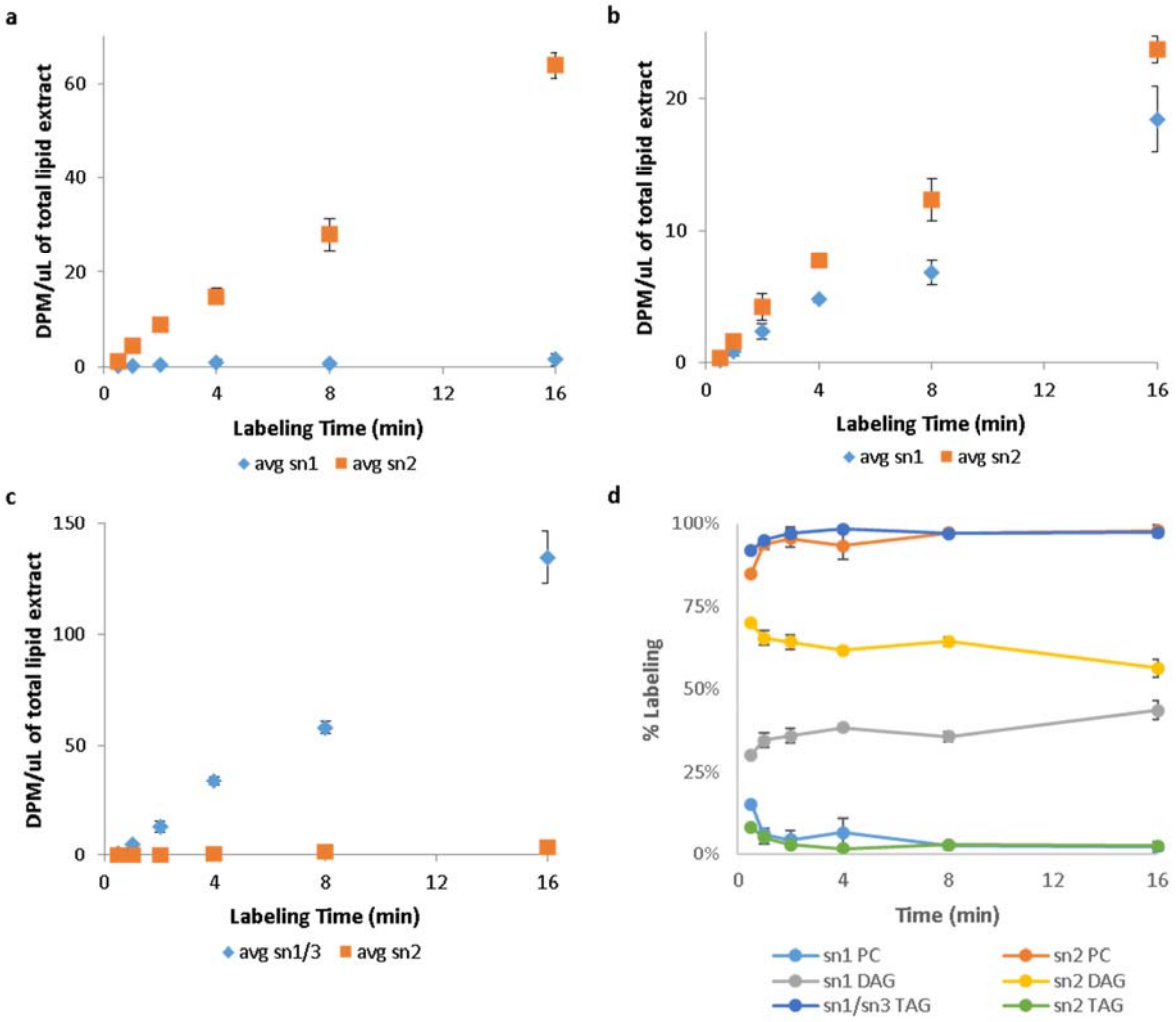

Figure 3.5. Fatty acid positional labeling in lipid classes. PC (a), DAG (b), and TAG (c) were isolated from the total lipid extract with TLC and then enzymatically treated (with PLA_2 or pancreatic lipase) to hydrolyze *sn*-2, or *sn*-1 and *sn*-3, respectively, fatty acids. Reaction products were separated by TLC and counted via scintillation counting. (c) Comparison of the proportions of fatty acid positional labeling in PC, DAG, and TAG, using the data shown in the other panels. The lines are provided for visualization and do not represent regression analysis. Data points represent the average ± standard deviation (n=3) for each measurement or calculation.

pre-existing, unlabeled DAG. Preliminary analysis of the labeled TAG fatty acid composition revealed that approximately 40% of the labeled fatty acids were newly synthesized (i.e., 16:0 or 18:1), and another 40% were newly elongated (Fig 3.6, Table 3.3.S2), both of which could be incorporated into the *sn*-3 position.

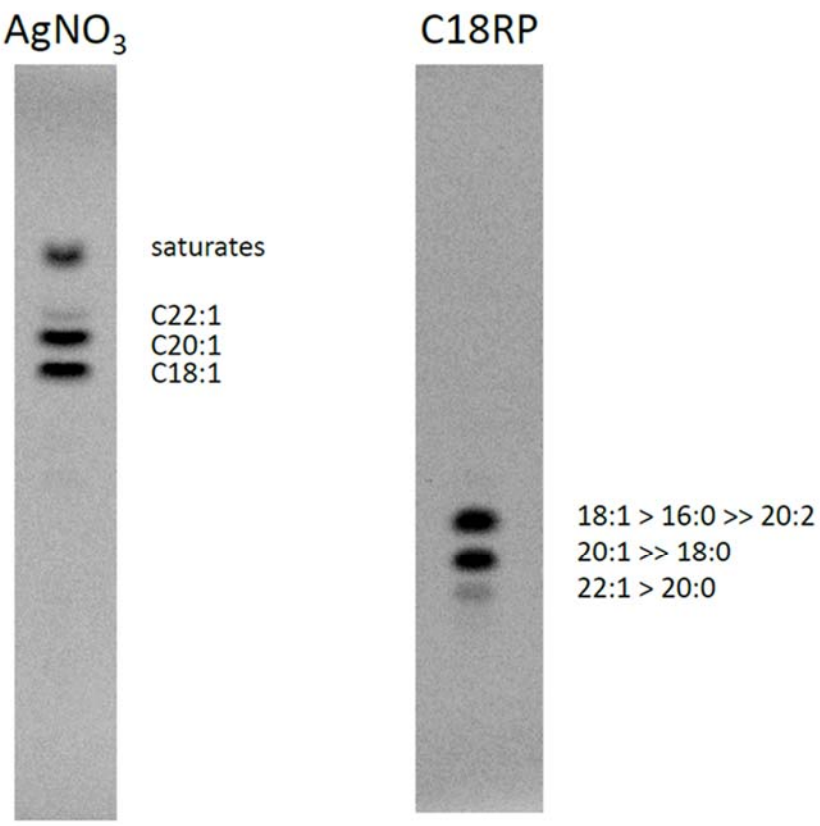

Figure 3.6. Preliminary test to differentiate between *de novo* and elongated fatty acid labeling. TAG was isolated from *C. sativa* embryos labeled with ^{14}C acetate for 10 min and transmethylated to separate the fatty acids. Fatty acids were separated by carbon number and/or saturation with TLC. Activities were measured with phosphorimaging. Percent labeling per fatty acid species is provided in Table 3.3.S2.

Discussion

In this work, ^{14}C glycerol and ^{14}C acetate time-course data were integrated with a model comprised of first-order rate equations in order to assay the topology of the TAG biosynthesis pathway in *C. sativa*. We found that a network containing two active DAG pools, with one being in exchange with the active PC pool, and bulk PC and DAG pools best fit our ^{14}C-glyceryl data (Fig 3.1-2). This topology was also consistent with our ^{14}C acetate data, in that the bulk PC pool could allow PC *sn*-2 positional labeling to significantly increase without an equal increase in DAG *sn*-2 labeling (Fig 3.5). We found that most of the TAG positional labeling was in the *sn*-1 or *sn*-3 position (Fig 3.5), 44% of which appeared to be in C_{20} or C_{22} fatty acids (Table 3.3.S2) and 41% in 18:1 or 16:0, which seems to likely be due to pre-existing, unlabeled DAG being

evenly acetylated with newly synthesized and recently elongated fatty acids.

A related study was performed by Yang *et al.* (2017) to test how the PC to DAG flux affects the accumulation of TAG. They expressed two *Arabidopsis thaliana* phospholipase Dζ (PLDζ) genes in *C. sativa*. As with the PLA_2 that we used to measure PC positional labeling, PLDζ hydrolyzes PC into PA and releases a fatty acid that could be used to make TAG. Consistent with our results, they found that both the transgenic and wild-type lines accumulated high levels of TAG *sn*-3 label. However, they also found that PC *sn*-2 labeling only accounted for approximately 60% of the ^{14}C acetate labeling in PC, whereas we estimated that *sn*-2 PC label comprises over 90% of the PC label. Our time-courses both represented short-term ^{14}C acetate labeling (0.5-16 min vs. 3-180 min) and it is possible that this difference is due to a difference in embryo age (15 vs. 20 DAF) because lipid content is beginning to saturate at the age of their embryos (Pollard *et al.*, 2015b), but we hypothesize that it is more likely due to differences in our culturing media (e.g., osmotic pressure) or pre-intubation times (overnight vs. 20 min). These would need to be tested by our lab imitating their methods and comparing the results.

We are currently working on methods to model acyl and glyceryl together. This could potentially be done by modeling both the acyl and glyceryl ^{14}C glycerol components, but this measurement requires the separation of the two compartments. Hence, while the components could both represent the same time-course; there would not be information on whether a molecule was simultaneously labeled in the glyceryl and acyl compartments. This could be observed with ^{13}C glycerol and mass spectrometry, which can quantify acyl (M+2), glyceryl (M+3), and simultaneous (M+3) labeling. However, there are limitations on how short ^{13}C time-courses can be to obtain enough label to clearly exceed natural abundance levels. Another direction that should be taken is to further elucidate how the fatty acid species are labeled within each lipid class. This would significantly increase the model complexity because there would need to be separate reactions for each class-species combination, but this is an important future direction because many of the lipid biosynthesis enzymes (e.g., LPCAT, PDAT, LPAAT) have some acyl substrate specificity

(Snyder *et al.*, 2009) and thus affect the network's function.,

This study represents a versatile method for analyzing the network topology and fluxes. Our long-term objective is to facilitate rational engineering of plant oils by developing a more complete and dynamic model of the TAG biosynthesis network that represents how the lipid classes and fatty acid species are related in oilseeds.

Methods

Growing and culturing C. sativa developing embryos

Camelina sativa var. Sunesson plants were grown in a growth chamber (BioChambers; Winnipeg, Manitoba, Canada) on a 16hr/8hr light/dark cycle with plants receiving 125 µmol m^{-2} s^{-1} of light. Seeds were placed in a 25% medium coarseness perlite (Sun Gro; Quincy, MI, USA) and 75% Sure-Mix Potting soil (Michigan grower's products Galesburg, MI, USA) that had been autoclaved when moist. Once seeds germinated, a 2:1 water:nutrient water (one-half strength Hoagland solution) mixture was added twice per week. Once flowering, stems were tagged to track silique age and harvested into 10% chlorox solution to sterilize siliques for culturing.

In order to mimic *in planta* endosperm composition as described in Carey *et al.* (in preparation, Chapter 2), media was made using the most abundant sugars and amino acids and consisted of 8 mM glutamine, 4 mM alanine, 130 mM glucose, 12 mM sucrose, 20 mM HEPES, various vitamins, and 20% polyethylene glycol (PEG 4000). Embryos were cultured in 30 mm wells of a 6 well plate (NEST Scientific, USA). 1.5 mL of media was added to the well before adding 15 DAF embryos. 2mL of water was added to the center of the plate in order to keep conditions humid. The plate was sealed with parafilm (Bemis NA, USA) and incubated with light shaking overnight at physiological (10 µE) green light (white light from fluorescent bulbs was filtered through a layer of green celluloid filter) to acclimate the embryos to the culturing conditions. After this initial incubation, media containing the appropriate amount of label (e.g.,

1.5 mM U-$^{14}C/^{13}C_3$ glycerol) was added to the embryos and they were inoculated as described above for the designated amount of time.

Lipid extraction, separation and label quantification

The labeling assays were terminated by collecting and rinsing the embryos with water to remove residual labeled media, adding hot isopropanol, and incubating the embryos at 85°C for 10-15 min. Once cooled, the embryos were ground in the isopropanol using glass homogenizer and transferred to a new tube using hexane. The lipids were extracted using a sodium sulfate phase a hexane/isopropanol phase separations.

Lipid classes were separated using thin-layer chromatography (TLC) on a silica plate (Analtech, Inc., USA). Polar lipids were separated with a full development with 85:15:5:2 (v/v/v/v) chloroform:methanol:glacial acetic acid:water, while neutral lipids were separated with a full development with 88:12 (v/v) toluene:ethyl acetate. Polar lipids were extracted from the silica with 5:5:1 (v/v/v) chloroform:methanol:water, followed by phase separation with 0.8% (w/v) potassium chloride. Neutral lipids were extracted from the silica with 2:1 (v/v) chloroform:methanol and phase separated with 0.8% (w/v) potassium chloride. Radioactivity were measured from extracted lipids in solution or separated lipids in silica with a liquid scintillation counter, or measured from silica with a phosphorimager.

Lipid hydrolysis

PC that had been isolated was treated with phospholipase A_2 (PLA_2) from porcine pancreas (P6534 Sigma) in conjunction with 50 µg cold carrier from soybean. The PC was dissolved in diethyl ether and 100mM sodium borate, pH 8.5, 5 mM calcium chloride, and incubated with 2 units PLA_2 for 30 min. The reaction was stopped with 2:1 (v/v) chloroform:methanol.

DAG that had been isolated was chemically acetylated by incubating with 3:1 (v/v) acetic anhydride:pyridine at 60°C for 2 hours, then at room temperature for 12 hours to form acetyl DAG

(acDAG). Both TAG and acDAG were treated with 1,3 lipase from porcine pancreas (L3126 Sigma) by dissolving in 1M Tris-HCl, pH 8, 2.2% (w/v) calcium chloride, and 0.5 mg/mL bile salts. The reaction was ran with 25 ug lipase for 60 min and stopped with the addition of cold 3 M hydrochloride acid and ethanol. Diethyl ether, 2:1 (v/v) chloroform:methanol, and water were then added to isolate the products via phase separation.

Enzyme products were separated with a full development with 88:12 (v/v) toluene:ethyl acetate and quantified as described above.

Testing potential network topologies

Different networks topologies were tested by writing the reactions as systems of first-order (i.e., ordinary differential) rate equations and using MATLAB to optimize model fit to the data in a method similar to that of (Hibino *et al.*, 2011). In this method, the rate equations, initial label content (i.e., 0 nmol label at 0 hr), and arbitrary estimates of the rate constants (e.g., $k_1 = 0.5$) were used as input for MATLAB *ode45* This script is a differential equations solver that numerically calculated the expected label content at the measured times (e.g., 24, 50 hr), given those rate constants. Initially, this resulted in a poor model fit because the rate constants were chosen arbitrarily. Next, the calculated label content and time-course data were used as input into MATLAB *lsqcurvefit* . This optimizer script is a nonlinear least-squares solver and was used to identify rate constants that permitted the model to better fit the data. The outputted rate constants were used as input into MATLAB *ode45* to obtain new calculated label content. This cycle was repeated until the optimizer could no longer identify a better fit for the rate constants, given the network equations. To identify a network topology to use for the preliminary flux map, we compared the sum of squared errors between predicted and measured ^{14}C contents in PC, DAG, and TAG over the time-course (Fig 3.S2).

Supporting Information

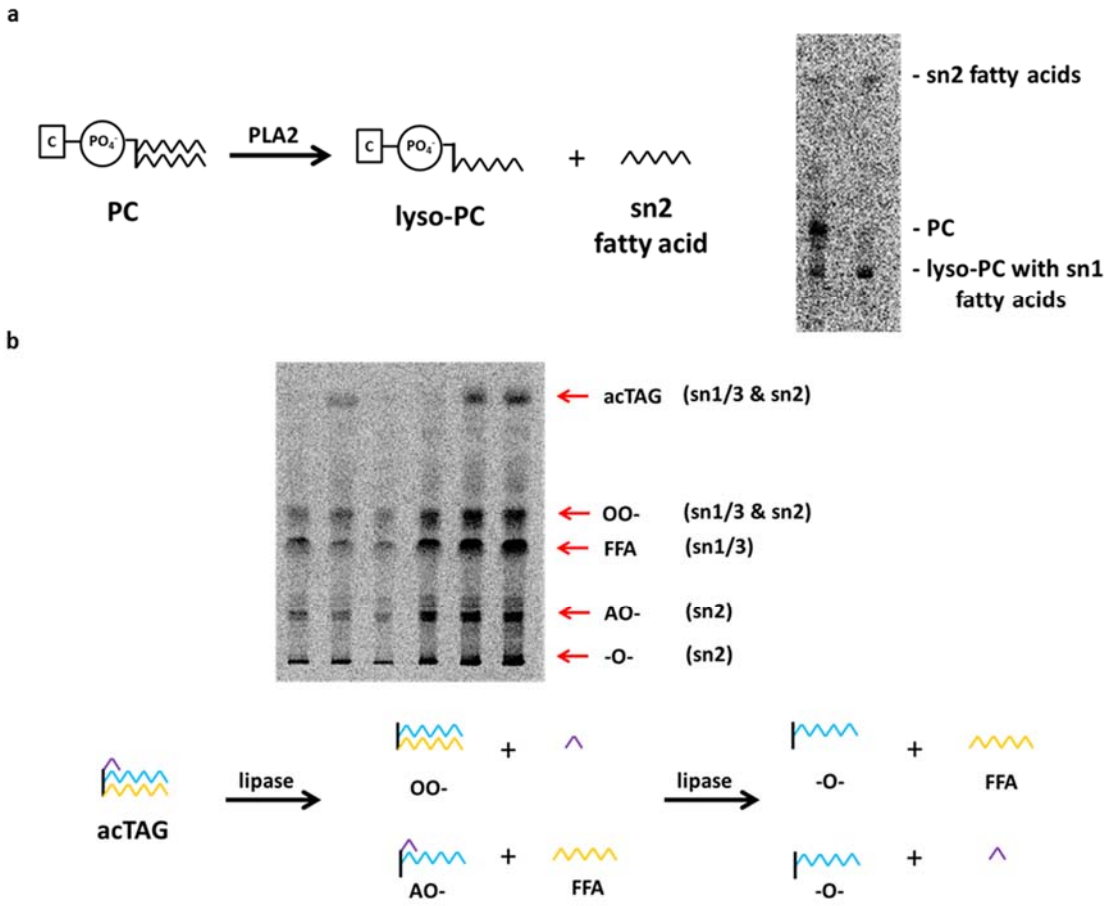

Figure 3.S1. Pictorial representation of the method used to measure fatty acid positional labeling. (a) PC was isolated from total lipid extracts and treated with PLA_2 to hydrolyze its *sn*-2 fatty acid, resulting in the free fatty acid and lyso-PC. (b) DAG and TAG were isolated from the neutral lipid component of the total fatty acid extracts and DAG was chemically acetylated (resulting in acDAG) to prevent acyl migration. TAG and acDAG were treated with pancreatic lipase in a two-step reaction to hydrolyze the *sn*-1 and *sn*-3 fatty acids. Pancreatic lipase does not distinguish between these positions so they were counted together (designated above as sn1/3). Enzymatic products were separated with TLC and counted with phosphorimaging, followed by scintillation counting.

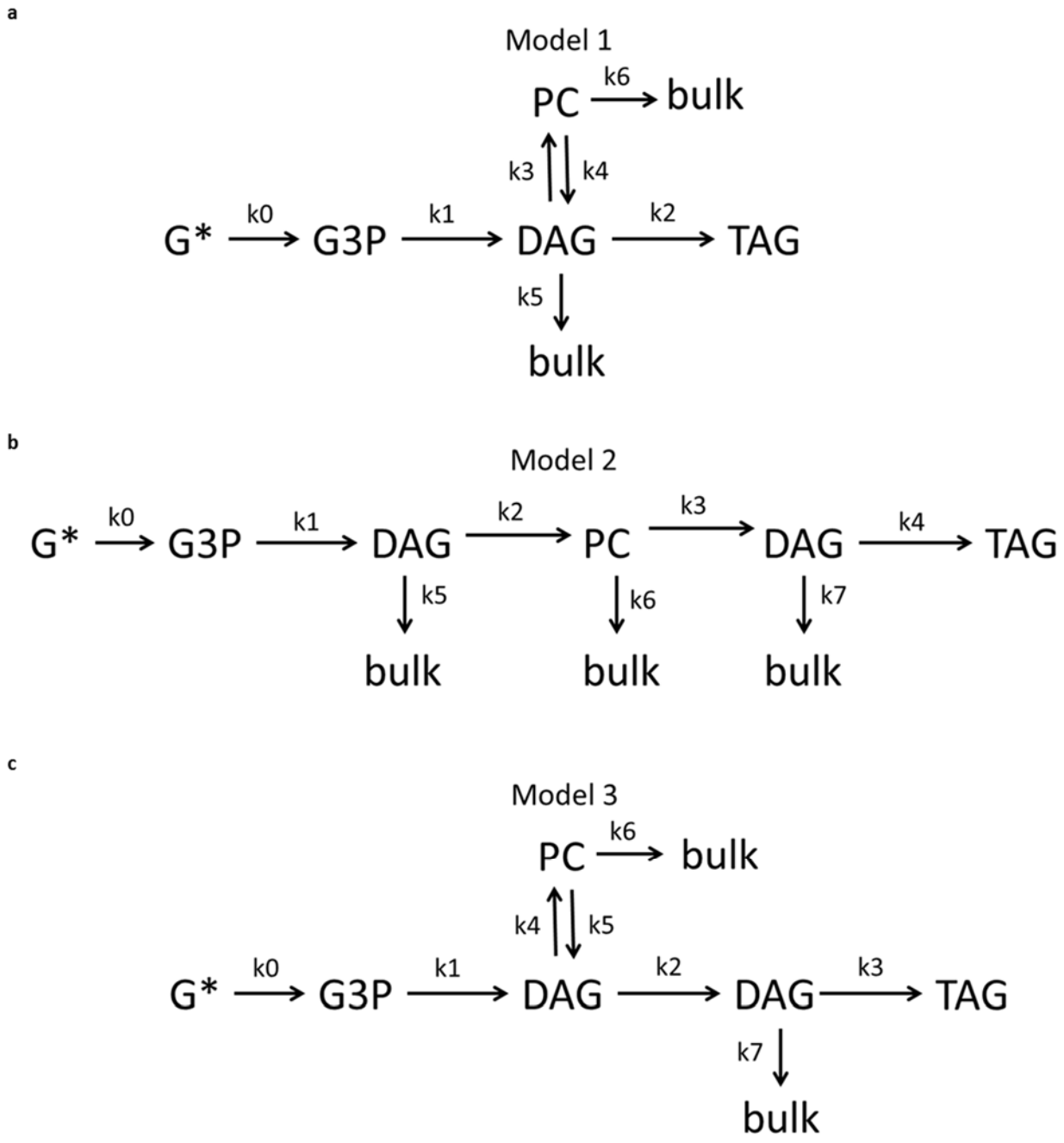

Figure 3.S2. Examples of models tested. Three of the networks tested for goodness of fit are illustrated, with G* indicating 14C glycerol content, "bulk" representing inactive lipid pools that may be alternatively used for storage, and the fitted rate constants denoted with k. Goodness of fit was determined using the sum of squared errors between model predicted and the average measured ^{14}C contents of PC, DAG, and TAG over the 24 hr time-course. The sum-of-squares for these models were 160 (a), 230 (b), and 67 (c).

Table 3.S1. Glyceryl label content per lipid class

Time (min)	TAG (nmol)	DAG (nmol)	PC (nmol)
0	0	0	0
24	0.89	2.19	1.46
50	2.02	3.62	2.7
95	5.68	5.35	4.95
183	14.24	6.62	5.96
365	37.62	8.96	6.44
669	81.23	10.98	8.93
1323	194.33	18.19	11.48

Data was measured and provided by Mike Pollard. *C. sativa* embryos were labeled with ^{14}C glycerol, transmethylated, and their glyceryl label content counted. Values indicate average ^{14}C glyceryl content per lipid class.

Table 3.S2. Preliminary distribution of label in fatty acid species

FAME	Proportion of Label
16:0	9%
18:0	8%
18:1	34%
20:0	5%
20:1	32%
20:2	3%
22:1	4%

TAG was isolated from *C. sativa* embryos labeled with ^{14}C acetate for 10 min, then transmethylated to separate the fatty acids in order to distinguish between label due to fatty acid synthesis and elongation. Fatty acids were separated by carbon number and/or carbon number with TLC. Activities were measured with phosphorimaging. Percent labeling per fatty acid species is provided in Table 3.3.S2.

REFERENCES

REFERENCES

Arab-Tehrany E, Jacquot M, Gaiani C, Imran M, Desobry S, Linder M. 2012. Beneficial effects and oxidative stability of omega-3 long-chain polyunsaturated fatty acids. *Trends in Food Science & Technology* 25(1): 24-33.

Bates PD, Durrett TP, Ohlrogge JB, Pollard M. 2009. Analysis of acyl fluxes through multiple pathways of triacylglycerol synthesis in developing soybean embryos. *Plant physiology* 150(1): 55-72.

Bates PD, Fatihi A, Snapp AR, Carlsson AS, Lu C. 2012. Acyl editing and headgroup exchange are the major mechanisms that direct polyunsaturated fatty acid flux into triacylglycerols. *Plant physiology*: pp. 112.204438.

Bates PD, Stymne S, Ohlrogge J. 2013. Biochemical pathways in seed oil synthesis. *Current opinion in plant biology* 16(3): 358-364.

Benning C. 2009. Mechanisms of lipid transport involved in organelle biogenesis in plant cells. *Annual Review of Cell and Developmental* 25: 71-91.

Chapman KD, Ohlrogge JB. 2012. Compartmentation of triacylglycerol accumulation in plants. *Journal of Biological Chemistry* 287(4): 2288-2294.

Durrett TP, Benning C, Ohlrogge J. 2008. Plant triacylglycerols as feedstocks for the production of biofuels. *The Plant Journal* 54(4): 593-607.

Dyer JM, Stymne S, Green AG, Carlsson AS. 2008. High-value oils from plants. *The Plant Journal* 54(4): 640-655.

Hibino K, Shibata T, Yanagida T, Sako Y. 2011. Activation kinetics of RAF in the ternary complex of RAF, RASGTP, and kinase on the plasma membrane of living cells: single-molecule imaging analysis. *Journal of Biological Chemistry*: jbc. M111. 262675.

Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, DeBono A, Durrett TP. 2013. Acyl-lipid metabolism. *The Arabidopsis book/American Society of Plant Biologists* 11.

Mosiewicki MA, Aranguren MI. 2013. A short review on novel biocomposites based on plant oil precursors. *European Polymer Journal* 49(6): 1243-1256.

Mudd J 1980. Phospholipid biosynthesis. *Lipids: Structure and Function*: Elsevier, 249-282.

Pollard M, Delamarter D, Martin TM, Shachar-Hill Y. 2015a. Lipid labeling from acetate or glycerol in cultured embryos of Camelina sativa seeds: a tale of two substrates. *Phytochemistry* 118: 192-203.

Pollard M, Martin TM, Shachar-Hill Y. 2015b. Lipid analysis of developing Camelina sativa seeds and cultured embryos. *Phytochemistry* 118: 23-32.

Snyder CL, Yurchenko OP, Siloto RM, Chen X, Liu Q, Mietkiewska E, Weselake RJ. 2009. Acyltransferase action in the modification of seed oil biosynthesis. *New biotechnology* 26(1-2): 11-16.

Yang W, Wang G, Li J, Bates PD, Wang X, Allen DK. 2017. Phospholipase Dζ enhances diacylglycerol flux into triacylglycerol. *Plant physiology*: pp. 00026.02017.

CHAPTER 4

Discussion

This dissertation depicts how experimental data can be integrated with mechanistic mathematical models to address questions about plant nutrient fluxes at three biological scales: 1) two mutualistic organisms exchanging nutrients, 2) one organism allocating carbon within its central metabolism, and 3) carbon flow within a biochemical network. These models provide a means of tracking how plants use their carbon in an ecological setting, for energy production, and to synthesize biomass products, but the different types of models employed are associated with their own capabilities and challenges.

Plant-microbe nutrient exchange symbioses

In Chapter 1, we tested how the model legume *Medicago truncatula* and its rhizobial partner benefit from exchanging photosynthetically fixed carbon for fixed nitrogen across a range of soil nitrogen availabilities. Using a biological market model that accounted for the movement and allocation of these nutrients, we found that when soil nitrogen was abundant and did not limit plant growth, the partners received equivalent—though small—benefits from this interaction. In contrast, when nitrogen was scarce and strongly limited plant growth, both partners significantly benefitted from the mutualism, yet the benefit to the plant had greater weight in determining the trade parameters than the rhizobial benefit. We hypothesize that this asymmetric bargaining power could be due to the plant having a lower discount rate due its longer lifespan and potentially larger nutrient reserves, or due to the rhizobia possessing group bargaining dynamics that limit their bargaining power.

The biological market model used in this study was adapted from a grass-mycorrhizal fungi model (Grman *et al.*, 2012). One of the first steps in the adaptation process was to design experiments to measure the input parameters and predictions. At times, this was straightforward, such as determining tissue nutrient yields in terms of carbon to nitrogen ratios. However, this process was more ambiguous for other parameters because the model was originally developed for ecological populations and we were applying it to a lower biological scale. For example, one parameter was the plant carbon uptake rate per

shoot carbon biomass. Photosynthesis rates are typically measured per leaf area, which do not consider the effects of leaf age or position, or any photosynthesis that may be performed by green stems. While such simplifications might not strongly affect total carbon uptake in a population, they can represent significant variations for individuals, especially of different sizes and allocation parameters. Consequently, since we intended for the entire plant partner to be represented in the model, we built chambers to measure whole-plant photosynthesis.

In addition to experimental challenges associated with applying models to different biological scales (e.g., ecological to organismal), this project also pointed to challenges in parameterizing theoretical models. Although the original model was parameterized in that it was fit to literature data (Grman *et al.*, 2012), it was derived from a theoretical framework of plant-microbe nutrient exchange mutualisms (Clark *et al.*, 2017), which affected how the mathematics operated. For example, a central assumption of the original model was that the mutualistic trade dynamics would be consistent with the Nash bargaining solution and that the partner benefits would be weighted equally. This is a common assumption in ecological models, but results from a game theory analysis that assumes that the partners are consistently self-interested, engage in bargaining, and are informed about the consequences of trade choices (Nash, 1950). Although the same bargaining solution can be obtained under other sets of assumptions, it is not clear how these traits can be measured, particularly for plant-microbe interactions, and thus how to test this assumption. The finding in our study that asymmetric bargaining theory better explains the observed trade parameters than Nash's original symmetric solution highlights the value of measuring as many physiological and nutritional parameters as possible in the same experimental system.

Steady-state metabolic flux analysis

In Chapter 2, metabolic flux analysis (MFA) was used to understand why *Camelina sativa* developing embryos have low carbon use efficiency and how genetic engineering these seeds to accumulate medium-

chain fatty acids affects their central metabolism. By comparing *C. sativa* decarboxylation fluxes, we identified the oxidative pentose phosphate (OPP) flux as being the major contributor to total carbon dioxide production. Furthermore, the OPP net fluxes were tightly correlated with embryo carbon use efficiencies across three light levels. Consequently, we concluded that the OPP flux was primarily responsible for low carbon use efficiency in *C. sativa* embryos. The transgenic lines were analyzed with MFA and clustering analyses. We found that the lines clustered so that the genotypes were cleanly resolved by principal component analysis (PCA) using metabolite labeling profiles resulting from labeling the embryos with ^{13}C glucose, while the clustering was less effective in distinguishing them when the embryos were labeled with ^{13}C alanine or ^{13}C glutamine. Regardless of the substrate, clustering analyses yielded cleaner separation between the transgenic lines and wild-type embryos grown under different light levels than clustering within the transgenic lines. This suggests that environmental variation had a greater effect on these embryos' central metabolism than the genetic modification of fatty acid composition.

Unlike the mutualism model described above, metabolic models are rooted in the chemical reactions that define the network; therefore, the primary assumption that requires testing (other than checking that the system is in metabolic and isotopic steady-state) is the proposed metabolic network, such as whether there is an active glyoxylate cycle in the transgenic *C. sativa* embryos. The model reactions specify how each of the substrate carbon atoms are transformed into specific carbons in the products. This allows the tracking of ^{13}C and ^{12}C to yield flux values, and the necessary data (isotopomer labeling, and metabolite uptake and efflux rates) can be measured directly with considerable precision and accuracy. However, if more than one labelled substrate is used in separate experiments (which is advantageous because different substrates are more informative about different processes), then the resulting labeling from each substrate needs to be separately measured and represented with a separate set of reactions in order to preserve the expected differences in isotopomer labeling. This greatly

increases the model complexity because increasing the number of network reactions constrains the feasible solution space, and leads to challenges in determining flux values that best fit the data. Briefly, we addressed this large network challenge by first randomly generating starting points within 10% of the total measured substrate influx from the center of feasible space. Next, we randomly selected starting points within 10% of the previously determined starting points that were able to be optimized. We repeated this process until at least 100 starting points yielded optimized (best fit) sets of flux values and we observed a clear asymptotic pattern in total model fit quality (sum of weighted squared deviations of model predictions from measured values). By exploring feasible space in this branching process, the starting points were more representative of feasible space rather than simply being a tight cluster. This is a crucial process in maximizing the likelihood that the flux maps obtained are true (global) best fits to the data rather than computational local optima, which is a significant risk in multi-parametric fitting of nonlinear systems to high dimensional datasets.

Optimized net fluxes that were normalized by total carbon uptake could not be cleanly clustered by PCA. This may have been due to a small number of fluxes comprising a significantly larger proportion of the carbon flux (e.g., OPP flux) and skewing the PCA results. We addressed this large network challenge using hierarchical clustering analysis (HCA) set to scale the differences between objects by dividing by the standard deviations. This would assign higher values to fluxes that varied more among the samples. We also set the HCA to maximize differences between samples in order to mimic PCA. With these settings, HCA was able to effectively, though not completely, separate the transgenic lines from each other, and to cleanly separate wild-type grown under different light levels from each other and from the transgenic lines.

Dynamic modeling of seed oil biosynthesis

In Chapter 3, we used a dynamic model comprised of first-order rate equations to examine the topology

of the triacylglycerol (TAG) biosynthesis pathway in *C. sativa* developing embryos. We fit the model to ^{14}C glyceryl time-course data to track how key lipid classes are made *de novo* and found that the data were best explained by a model in which there were two active diacylglycerol (DAG) pools, with one being in exchange with the active phosphatidylcholine (PC) pool, while the other serves as the immediate TAG precursor, and bulk PC and DAG pools. In addition to ^{14}C glyceryl modeling, we performed short-term ^{14}C acetate time-course experiments to track how newly synthesized or recently elongated fatty acids were added to lipid classes and moved through the network. These experimental results were consistent with the model's network topology, such as high PC *sn*-2 positional labeling in the absence of an equal increase in DAG *sn*-2 labeling, which is explained by the bulk PC pool being the substrate of acyl exchange at *sn*-2.

This type of modeling has the potential to overfit the data because it works by fitting rate constants and other parameters. Therefore, if the model is complex, the increased number of parameters could inherently enhance the model fit and/or cause the model to describe random errors in the data. This is a challenge for a variety of modeling types in situations where the complexity of the model is greater than the underlying information content of the experimental data. A common way to address this challenge is to broaden what is being measured in order to increase the information content and/or check that the model also fits to a second set of independent data (e.g., separate training and testing sets; Chang *et al.*, 2013). Computationally, these are relatively straightforward, but they may be experimentally challenging. In our study, we only obtained three biological replicates for the ^{14}C glyceryl time-course being modeled, so it would be beneficial if this experiment is repeated in the future in order to have testing and training sets available.

Additional experimental data can be utilized by modeling with ^{13}C glycerol time-course data instead of ^{14}C. Glycerol labels both the glyceryl and acyl groups, but these can only be distinguished for ^{14}C with transmethylation. ^{13}C, on the other hand, can be measured for M+2 (acyl), M+3 (glyceryl), and M+5 (both groups) labeling with mass spectrometry. We have ^{13}C glycerol time-course data and its

modeling analyses are in progress. An advantage to using ^{14}C over ^{13}C is that the label can be detected at smaller quantities and thus for shorter time-courses. A significant future direction for this project is to develop methods for integrating these types of labeling and different time-course.

Conclusion

Different biological scales have distinct challenges in how modeling analysis can be integrated with data. Ecological models tend to describe interacting populations and thus, due to system complexities and the longevity of populations, represent theoretical predictions. This makes these models difficult to parameterize and a method for extrapolating the model from the population scale to a scale that is easier to measure (e.g., individual) may not be obvious. Metabolic scale models are rooted in chemical principles, and modern analytical methods enable them to be highly parameterizable. However, these models can face computational challenges when applied to large networks, including optimization and potential biasing of results by a subset of the data.

Despite these challenges, the integration of models and experiments provides a valuable means for testing hypotheses about network structure and interactions, and for estimating the values of important biological parameters. In this dissertation, I have used different modeling approaches and presented evidence for *M. truncatula* having more bargaining power than its rhizobial partner, a high OPP flux being the cause of the low carbon use efficiency in *C. sativa*, environmental variation potentially having a greater impact on central metabolism than genetic engineering in oilseeds, and the potential existence of two active DAG pools in the oilseed TAG biosynthesis pathway. Together, modeling and experiments open approaches to track how plant nutrients are obtained and allocated.

REFERENCES

REFERENCES

Chang C-Y, Hsu M-T, Esposito EX, Tseng YJ. 2013. Oversampling to overcome overfitting: Exploring the relationship between data set composition, molecular descriptors, and predictive modeling methods. *Journal of chemical information and modeling* 53(4): 958-971.

Clark TJ, Friel CA, Grman E, Shachar-Hill Y, Friesen ML. 2017. Modelling nutritional mutualisms: challenges and opportunities for data integration. *Ecology letters* 20(9): 1203-1215.

Grman E, Robinson TM, Klausmeier CA. 2012. Ecological specialization and trade affect the outcome of negotiations in mutualism. *The American Naturalist* 179(5): 567-581.

Nash Jr JF. 1950. The bargaining problem. *Econometrica: Journal of the Econometric Society*: 155-162.