













High-throughput characterization, correlation, and mapping of leaf photosynthetic and functional traits in the soybean (*Glycine max*) nested association mapping population

Christopher M. Montes ¹, Carolyn Fox ², Álvaro Sanz-Sáez ³, Shawn P. Serbin ⁴, Etsushi Kumagai ⁵,
Matheus D. Krause ⁶, Alencar Xavier ^{7,8}, James E. Specht ⁹, William D. Beavis ⁶, Carl J. Bernacchi ^{1,10,11},
Brian W. Diers ^{2,*} and Elizabeth A. Ainsworth ^{1,10,11,*}

¹Global Change and Photosynthesis Research Unit, USDA ARS, Urbana, IL 61801, USA,

²Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA,

³Department of Crop, Soil, and Environmental Sciences, Auburn, AL 36849, USA,

⁴Environmental and Climate Sciences Department, Brookhaven National Laboratory, Upton, NY 11973, USA,

⁵Institute of Agro-environmental Sciences, National Agriculture and Food Research Organization, Tsukuba, Ibaraki 305-8604, Japan,

⁶Department of Agronomy, Iowa State University, Agronomy Hall, Ames, IA 50011, USA,

⁷Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA,

⁸Department of Biostatistics, Corteva Agrisciences, Johnston, IA 50131, USA,

⁹Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA,

¹⁰Carl R. Woese Institute for Genomic Biology, Urbana, IL 61801, USA,

¹¹Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

*Corresponding author: Department of Crop Sciences, University of Illinois at Urbana-Champaign, 1201 West Gregory Drive, 147 Edward R. Madigan Laboratory, Urbana, IL 61801, USA. Email: lisa.ainsworth@usda.gov; *Corresponding author: Global Change and Photosynthesis Research Unit, USDA ARS, 1101 West Peabody Drive, 268 National Soybean Research Center, Urbana, IL 61801, USA. Email: bdiers@illinois.edu

Abstract

Photosynthesis is a key target to improve crop production in many species including soybean [*Glycine max* (L.) Merr.]. A challenge is that phenotyping photosynthetic traits by traditional approaches is slow and destructive. There is proof-of-concept for leaf hyperspectral reflectance as a rapid method to model photosynthetic traits. However, the crucial step of demonstrating that hyperspectral approaches can be used to advance understanding of the genetic architecture of photosynthetic traits is untested. To address this challenge, we used full-range (500–2,400 nm) leaf reflectance spectroscopy to build partial least squares regression models to estimate leaf traits, including the rate-limiting processes of photosynthesis, maximum Rubisco carboxylation rate, and maximum electron transport. In total, 11 models were produced from a diverse population of soybean sampled over multiple field seasons to estimate photosynthetic parameters, chlorophyll content, leaf carbon and leaf nitrogen percentage, and specific leaf area (with R^2 from 0.56 to 0.96 and root mean square error approximately <10% of the range of calibration data). We explore the utility of these models by applying them to the soybean nested association mapping population, which showed variability in photosynthetic and leaf traits. Genetic mapping provided insights into the underlying genetic architecture of photosynthetic traits and potential improvement in soybean. Notably, the maximum Rubisco carboxylation rate mapped to a region of chromosome 19 containing genes encoding multiple small subunits of Rubisco. We also mapped the maximum electron transport rate to a region of chromosome 10 containing a fructose 1,6-bisphosphatase gene, encoding an important enzyme in the regeneration of ribulose 1,5-bisphosphate and the sucrose biosynthetic pathway. The estimated rate-limiting steps of photosynthesis were low or negatively correlated with yield suggesting that these traits are not influenced by the same genetic mechanisms and are not limiting yield in the soybean NAM population. Leaf carbon percentage, leaf nitrogen percentage, and specific leaf area showed strong correlations with yield and may be of interest in breeding programs as a proxy for yield. This work is among the first to use hyperspectral reflectance to model and map the genetic architecture of the rate-limiting steps of photosynthesis.

Keywords: photosynthesis; phenotyping; high-throughput; multiparental; SoyNAM; soybean; GWAS; heritability; PLSR

Introduction

The capacity to collect meaningful phenotypic data has not kept pace with increasingly extensive genomic data that are relatively easy and inexpensive to collect (Furbank and Tester 2011; Cabrera-Bosquet et al. 2012; Araus and Cairns 2014; Araus et al. 2018). Phenotyping functional traits of interest is often destructive, time consuming, and costly. Significant investment in developing high-throughput phenotyping technologies is seen as critical to screen the vast genetic diversity that is currently being underutilized in crop breeding (Mikel and Dudley 2006; Lam et al. 2010; Li et al. 2010, 2013; van Heerwaarden et al. 2012; Qiu et al. 2013). The coupling of phenotyping technologies with genetic improvement could help accelerate the development of plants adapted to a changing climate, thereby enhancing the annual rate of on-farm yield increases that will be necessary to feed a growing world population (Tester and Langridge 2010; Araus et al. 2018). Given that annual yield gains are projected to be insufficient to meet future demands (Ainsworth et al. 2008; Raines 2011; Ray et al. 2012, 2013; Hunter et al. 2017) this integration is especially important.

Improving photosynthesis has been suggested as a critical target for increasing crop yields (Long et al. 2006, 2015; von Caemmerer and Evans 2010; Raines 2011; Ort et al. 2015). One possibility for improving photosynthesis is to take advantage of natural variation in photosynthetic capacity and breed for genotypes with improved carbon assimilation (Parry et al. 2011; Raines 2011; Faralli and Lawson 2020). Photosynthesis is limited primarily by the maximum rate of carboxylation by Rubisco ($V_{c,max}$) or the rate of RuBP regeneration (J_{max}) (Farquhar et al. 1980; Farquhar and Sharkey 1982). In efforts to identify genetic variation in photosynthetic capacity, a means to rapidly screen $V_{c,max}$ and J_{max} in hundreds of different lines is essential. However, the current best methods for obtaining accurate estimates of both $V_{c,max}$ and J_{max} use infrared gas analysis, which is time-consuming. Measuring the response of photosynthetic carbon assimilation (A) to intercellular CO_2 concentration (c_i) to then estimate $V_{c,max}$ and J_{max} (Farquhar et al. 1980) takes 30–40 min for 1 sample. Thus, screening germplasm for natural variation of either limiting process of photosynthesis is not realistic using gas exchange (Lawson et al. 2012; Driever et al. 2014).

A promising alternative to gas exchange is high-throughput reflectance spectroscopy. Specific regions of the reflectance spectrum from 500 to 2,400 nm are known to correspond to different leaf components and have been used to study plants from the leaf to the ecosystem scale for decades (Curran 1989; Gamon et al. 1992; Penuelas et al. 1994; Asner 1998; Penuelas and Filella 1998; Carter and Knapp 2001; Slaton et al. 2001; Sims and Gamon 2002; Asner et al. 2003; Serbin and Townsend 2020). The signal captured by a spectroradiometer is largely dependent upon 3 components. The first is electromagnetic scatter, which includes surface reflection (attributable to leaf microstructures and roughness in our study). The next is the internal reflection of leaf structural components and diffraction arising from interference within the leaf. Water and intercellular airspace identify in the near and shortwave infrared regions of the spectrum (~700–2,400 nm). The last component of the reflectance signal is driven by chemical properties in the leaf. These properties include absorption of light in the visible spectrum (~500–750 nm) by chlorophyll and carotenoid pigments and absorption of light in the near and shortwave infrared regions resulting in molecular vibrations of functional groups (Walter-Shea and Norman 1991; Sandak et al. 2016).

Recent studies have demonstrated that leaf reflectance spectroscopy together with empirical statistical approaches, including partial least squares regression (PLSR) modeling, can be used to accurately predict photosynthetic parameters in C3 and C4 plants and other related leaf chemistry, morphology, water status, and isotopic composition traits as well as genetic diversity (Serbin et al. 2012, 2016; Ainsworth et al. 2014; Heckmann et al. 2017; Yendrek et al. 2017; Silva-Perez et al. 2018; Meacham-Hensold et al. 2019; Cotrozzi et al. 2020; Meireles et al. 2020). Once the predictive models are developed, estimates of these leaf phenotypes can be obtained in a rapid and nondestructive manner on large populations. Another major benefit is the derivation of multiple leaf traits from a single spectrum, further reducing the time, and costs associated with data collection. When applied to the correct populations, these predicted phenotypes can be used for linkage analysis and association mapping to elucidate the genetic architecture of the target traits.

Linkage analysis or quantitative trait locus (QTL) mapping traditionally relies on creating a mapping population of recombinant inbred lines (RILs) derived from the mating of 2 parental cultivars that have contrasting phenotypes for a quantitative trait of interest. QTL mapping identifies chromosomal segments that likely contain genes controlling the trait, but the degree of centimorgan (cM) resolution in the length of those segments is a function of the number of marker-detectable recombination events in the segregating population. That number can be limited by a lack of genetic diversity in parental choice and/or by an insufficient number of RILs (Beavis 1994). Aside from the time it takes to generate the RILs, a lack of sufficient resolution can make difficult a backcross-mediated introgression of the favorable QTL allele into elite cultivars. An alternative method of QTL identification, known as association mapping (Yu et al. 2006), involves screening a large population of accessions of a species for historical recombination events that thereby leads to the detection of linkage disequilibrium (i.e. nonrandom association of alleles between 2 or more marker loci). After taking into account population structure, this mapping strategy can provide a finer resolution of the genetic structure underlying a trait (Li et al. 2011). The correlation between genotype and phenotype in many unrelated individuals and dense genome-wide marker capabilities has allowed for finer genomic resolution in association mapping, but at a loss of power to detect QTL. Association mapping has also been shown to falsely identify the known location of causative loci or genes and an inability to detect QTL for which one of the alleles exists at a very low frequency in the chosen group of accessions (Korte and Farlow 2013; Vilhjalmsen and Nordborg 2013). Multiparental or next generation mapping populations have been proposed as a means of overcoming some of the shortcomings of linkage analysis and association mapping by pairing the 2 mapping strategies (Yu et al. 2008; Morrell et al. 2012).

Nested association mapping (NAM) is a mapping approach that attempts to fuse the best features of association- and segregation-based QTL mapping in a fashion that will more powerfully identify beneficial alleles possessed by a large number of elite and exotic germplasm parents (Yu et al. 2008; Stich 2009; Nice et al. 2016; Bouchet et al. 2017; Sharma et al. 2018; Brock et al. 2020). A NAM population was created in soybean by selecting a large and diverse number of parental lines that were all crossed to a common parent (Diers et al. 2018). The soybean NAM (SoyNAM) population has been a useful resource in the study of many agronomic traits and their response to the environment. These traits include grain yield, lodging, seed weight, length of

the reproductive period, node number, pods per node, internode length, plant height, canopy closure, disease resistance, steady state point estimates of A , and water use efficiency (Xavier et al. 2016, 2017, 2018; Diers et al. 2018; Lopez et al. 2019; Scott et al. 2019). The utilization of the SoyNAM population has provided a better understanding of these traits and how they affect yield, while also identifying heretofore unknown alleles that may provide new genetic sources for agronomic improvement. An inherent issue with NAM populations is the sheer number of RILs they contain. Traditional methods for scoring the leaf photosynthetic, biochemical, and structural traits are intractable at scale, which makes rapid leaf reflectance spectroscopy an attractive option to survey these large mapping populations.

This study is presented in 2 parts. In the first, we explored the capacity to build models relating leaf reflectance to gas exchange and other leaf traits using diverse soybean genotypes, including the NAM founders, with the aim of providing an accurate and high-throughput estimation of photosynthetic parameters and leaf biochemistry. In the second part of the study, we show the utility of these leaf reflectance derived models by applying them to the SoyNAM population of RILs to identify marker-trait associations (MTAs) for photosynthetic capacity and leaf-level biochemical traits. These leaf traits were also correlated with yield from the same plots to determine their relationship. This study shows the potential of spectroscopy as a high-throughput tool to not only provide new insights about difficult to measure trait variation in soybean leaves, but also provides novel understandings of how these traits could be useful in meeting the future demands on agriculture.

Materials and methods

Part 1: Reflectance spectroscopy and PLSR model development

Plant material and site description

Data used in the PLSR leaf reflectance model building and calibration component of this experiment were collected from 2011 to 2019 at the SoyFACE Research Facility (40°02'N, 88°14'W, <https://soyface.illinois.edu/>) and South Farms at University of Illinois, Urbana, IL (Supplementary Table 1). The facility is located in the middle of soybean maturity group (MG) zone III (Mourtzinis and Conley 2017). Best practices for building PLSRs models with the highest possible predictive power include sampling across the broadest possible trait space (Schweiger 2020; Burnett et al. 2021). To accomplish this goal, diverse soybean genotypes were planted and sampled to estimate photosynthetic and biochemical traits in 2015, and additional data collected from 2011 to 2019 were incorporated into the analyses (Supplementary Table 2). Ultimately, 114 lines including the 40 founder lines, the common parent of the SoyNAM population (Song et al. 2017), and several chlorophyll deficient soybean lines (Walker et al. 2018) were used to build PLSR models (Supplementary Table 2).

A/c_i curves and estimates of $V_{c,max}$ and J_{max}

The diverse soybean population was sampled to measure A/c_i curves and obtain estimates of $V_{c,max}$ and J_{max} using established protocols (Sanz-Sáez et al. 2017). Petioles of the youngest, most fully expanded leaf located at the top of the canopy were excised predawn, and quickly placed in water. Petioles were then cut again under water to maintain turgor. Sampled petioles were kept in low light conditions until ~20 min before gas exchange measurements, when they were placed in a diffuse light environment. Then, gas exchange was measured with a LI-COR

LI-6400XT Portable Photosynthesis System (LI-COR, Lincoln, NE) under high light ($1,750 \mu\text{mol m}^{-2} \text{s}^{-1}$), ambient leaf temperature, and relative humidity. Leaves were stabilized at a $[\text{CO}_2]$ of 400 ppm. The A/c_i curves were generated by exposures of a given leaf to these successive $[\text{CO}_2]$: 400, 300, 225, 150, 100, 50, 400, 400, 600, 800, and 1,000 ppm. Each stepwise measurement was limited to 60–180 s to minimize changes in the activation state of Rubisco. Measurements were logged once A and stomatal conductance stabilized. Estimates of $V_{c,max}$ and J_{max} were estimated as described in Ainsworth et al. (2014) and normalized to 25°C using the model of Farquhar et al. (1980) and temperature parameters from Bernacchi et al. (2001, 2003).

Leaf reflectance data were also collected with simultaneous estimates of $V_{c,max}$ obtained from gas exchange from the initial slope of the A/c_i curve from measurements made in the field at midday (~11–15 h). This was done to take advantage of the 2012 and 2013 plantings of the SoyNAM population at Illinois. Only the initial slope of the curve was used to quickly survey a subset of the founder lines for $V_{c,max}$. The gas exchange system cuvette was set to control the ambient temperature and relative humidity. Measurements were made under high light ($1,750\text{--}2,000 \mu\text{mol m}^{-2} \text{s}^{-1}$). Leaves were stabilized at a $[\text{CO}_2]$ of 400 ppm before initiating these stepwise gas exchange measurements at 400, 250, 150, 100, and 50 ppm to estimate $V_{c,max}$ and then normalized to 25°C as outlined above. Additional data of paired leaf reflectance and A/c_i curves from Ainsworth et al. (2014), Walker et al. (2018), and Kumagai et al. (2022) were also used in the PLSR model development portion of this study. This was done to include more trait variation captured by different growing seasons and experimental treatments including elevated and ambient ozone concentrations (Ainsworth et al. 2014) and elevated temperature treatments (Kumagai et al. 2022).

Leaf chemical and morphological sampling

Leaf tissues that were tested for a suite of phenotypes (leaf chemistry and morphology) were collected at midday from sunlit, fully developed leaves, and from the same leaves used for A/c_i curves. Small (~1.4 cm²) leaf punches were taken with a cork borer, placed in 2 ml screw-cap tubes, and flash frozen in liquid nitrogen. One punch was used to determine both chlorophyll concentration and starch content after ethanol extraction in 80% (v/v) buffered ethanol following the methods described in Koester et al. (2016). Glucose, fructose, sucrose, and protein were determined from separate leaf punches using established methods (Jones et al. 1977; Ainsworth et al. 2007). Three leaf punches were collected for specific leaf area (SLA; g m⁻²), leaf carbon (Cperc), and leaf nitrogen percentage (Nperc). These leaf punches were dried at 60°C for 5 days to determine SLA by dividing the dry mass weight of the leaf punches by the combined area. SLA (the inverse of leaf mass per area) is a proxy for leaf thickness and density. Dried leaf tissue was ground and combusted with oxygen to determine leaf nitrogen and carbon percentage using a Costech 4010 elemental analyzer (Costech Analytical Technologies, Inc., Valencia, CA). All leaf nitrogen and leaf carbon percentage were determined on a mass basis. Leaf temperatures were collected with an infrared thermometer (62 Max, Fluke, Everett, WA).

Leaf reflectance spectroscopy

Three relative reflectance (simply “reflectance” from herein) measurements of the soybean leaf were captured on the adaxial side of the same leaf sampled for gas exchange or leaf punches using an ASD FieldSpec 4 Standard-Res Spectroradiometer (Malvern Panalytical, Westborough, MA). The 3 leaf reflectance

measurements per leaf were processed and averaged using the FieldSpec R Package (<https://doi.org/10.5281/zenodo.6248237>) where built-in functions and default values were used for quality control (removal of spectra with likely artificial bias) and jump correction between sensors (maximum threshold for jump correction = 0.02). Reflectance was calculated by the spectroradiometer from leaf radiance divided by the radiance of a white reflectance standard (Spectralon, Labsphere Inc., North Dutton, NH). The spectroradiometer measures reflectance across the “full-range” (defined as 350–2,500 nm) using a 51-stranded fiber optic cable split evenly among 3 sensors with one sensor in the visible and near-infrared (350–1,000 nm) and 2 covering the shortwave infrared regions (SWIR1, 1,001–1,800 nm and SWIR2, 1,801–2,500 nm). The field of view for the fiber optic cable is 25°. We used the FieldSpec 4 Plant Probe (Malvern Panalytical) fitted with the ASD Leaf Clip (Malvern Panalytical) assembly. This system maintains a close and consistent distance, angle, and light intensity between the adaxial side of the leaf and the fiber optic cable. The plant probe has an offset and self-contained halogen light source (calibrated, 2900K) to limit light and temperature stress of the leaf during measurements. The leaf clip allows for precise reflectance measurements by limiting scattered or errant light affecting the spectral signal.

The manufacturer’s protocols for the spectroradiometer start-up and use were followed, including a ~30 min “warm-up” period prior to data collection as well as instrument optimization and use of the white reflectance standard. A new white reference measurement was taken every 10–20 min while samples were being collected to avoid instrument drift. This procedure was used throughout all reflectance collection in the model calibration stage of this manuscript and the reflectance measurements taken in the SoyNAM population.

Development of PLSR models for soybean leaf traits

PLSR modeling was used to estimate the rate limiting steps of photosynthesis (normalized to 25°C $V_{c,max25}$, J_{max25} and at the measured temperature $V_{c,max}$ and J_{max}), chlorophyll content (total chlorophyll, chlorophyll a, chlorophyll b, and the chlorophyll a to chlorophyll b ratio), carbon and nitrogen percentage, protein content, sucrose content, total nonstructural carbohydrate content, and leaf temperature. PLSR is a standard chemometric statistical approach often used to predict a continuous response variable from highly collinear predictor variables that outnumber the observations (Wold et al. 1984). This is accomplished by projecting the response and predictor variables into new orthogonal subspaces, while simultaneously reducing the dimensionality of the variables and maximizing the covariance between the subspaces, by latent variables. These new, independent latent variables are similar to a principal component, i.e. defined as linear combinations of the original variables (Li et al. 2002). The measured trait observations (response variable) were paired with leaf reflectance data (predictor variables) taken at the time of the trait measurement on fresh leaf tissue prior to processing. The reflectance data were subset to 500–2,400 nm to minimize instrument noise as was done in previous work (Yendrek et al. 2017).

The PLSR modeling followed the methods and code provided in Serbin et al. (2014) except for 2 additional preprocessing steps for potential outlier removal. Briefly, trait values greater than ± 2 SD from the mean were removed from the PLSR model development because they were assumed to be incorrect data (sampling error, etc.). The remaining measured trait observations and their paired leaf reflectance spectra for a given trait were randomly subset into calibration (80%) and validation (20%) datasets. The

validation data were not used in any model calibrations. An initial PLSR model run used 25 latent variables and a “leave-one-out” cross validation approach on the calibration dataset. Calibration data with predicted trait values from the initial PLSR model with residuals greater than ± 2.5 SD were removed from the calibration dataset. The number of removed observations was $\leq 5\%$ of the total observations used in model calibration.

To determine the number of latent variables to retain for a given PLSR model, an iterative approach with 50 iterations was used. The resulting plots of the predicted residual error sum of squares (PRESS) by PLSR latent variable were evaluated to identify the number of latent variables that produced the lowest PRESS value, while also weighting model parsimony (i.e. we selected the smallest number of latent variables that resulted in a low PRESS value). Plots of the root mean square of the cross-validation (RMSECV), root mean square error of the prediction (RMSE), and PRESS were used to determine the final number of latent variables to include in each PLSR model by minimizing error and maximizing the predictive ability of the validation data. The number of latent variables for each model varied by trait. A jack-knife resampling approach (1,000 resamples) was used to validate the model with the independent validation data, and to estimate confidence intervals for the modeled validation data and model bias. We determined a PLSR model successful by the model diagnostic outputs of the validation data with R^2 values > 0.50 and RMSE relative to the range of observed trait values ($\sim 10\%$) used in the model calibration (Table 1, model coefficients as Supplementary File 1, other trait diagnostic output as Supplementary File 2).

A second PLSR model for Chl was built without the chlorophyll deficient mutants to help determine if these observations were driving the fit or predictability of the PLSR models following the procedure outlined above. Chl was selected as a test case because the trait defined the mutants. Both PLSR models with and without chlorophyll deficient mutants and a commonly used reflectance derived Chl index (Richardson et al. 2002) were compared (Supplementary File 3). Bland-Altman plots showed little difference between the 2 PLSR models (data not shown); however, both PLSR models were more accurate than the Chl index, which overestimated Chl content. Because there was little observed difference between the models with and without the low Chl lines, we retained all lines in the PLSR models to maintain the broadest possible trait range (Burnett et al. 2021).

Part 2: Genetic mapping of reflectance derived leaf traits and yield in the SoyNAM population SoyNAM population and experimental design

Diers et al. (2018) provide a full description of the SoyNAM population and field design of these experiments. Briefly, 40 parental soybean lines representing contemporary soybean varieties, lines with diverse ancestry, and plant introductions (PI) that yielded well under drought conditions were crossed to a common parent, cultivar IA3023, to create 40 biparental families each consisting of approximately 140 F_5 -derived lines ($\sim 5,600$ total). The lines from each of the 40 NAM families were subdivided into 4 blocks of 35 lines. The parents of the family plus 3 checks were added to each block resulting in 40-entry incomplete blocks. Entries were randomized within each block and the blocks were randomized in the field. The entries were grown in 4 m long, 2-row wide plots with 0.76 m row spacing. Two site-year combinations, the University of Nebraska (Clay Center, NE, irrigated) in 2012 (NE12) and the University of Illinois at Urbana-Champaign (Urbana, rain-fed) in 2013 (IL13), were used in this experiment.

Table 1. Summary of reflectance-based PLSR model builds for leaf traits highlighted in this manuscript.

Trait	Factors	Cal N	Val N	Waveband (nm)	Train R ²	CV R ²	RMSECV	CV Bias	Val R ²	RMSE	Val bias
Chl	4	288	74	500–2,400	0.82	0.81	0.04	0	0.72	0.04	0.002
Cperc	11	487	126	500–2,400	0.9	0.89	0.54	−0.0019	0.9	0.58	−0.1119
J _{max25}	10	248	64	500–2,400	0.77	0.73	16.86	0.0383	0.56	25.32	−1.7283
Nperc	18	479	124	500–2,400	0.87	0.85	0.25	−5.00E−04	0.77	0.35	−0.0455
SLA	12	300	78	500–2,400	0.96	0.95	15.17	0.1546	0.96	15.48	−0.7754
V _{c,max25}	12	274	70	500–2,400	0.67	0.59	13.52	−0.0162	0.74	12.34	−2.3838

Presented data include the number of latent factors used in the final model build (Factors), the number of observations that were used to build the model (Cal N), the number of observations that were used to validate the model (Val N), the range of wavebands used for the final model build [Waveband (nm)], the coefficient of determination for the calibration model (Train R²), the coefficient of determination of the cross validation (CV R²), the RMSECV, the model bias from the cross validation (CV Bias), the coefficient of determination of the validation data (Val R²), the root mean square error of the validation data (RMSE), and the model bias of the validation data.

Chl, total chlorophyll content; Cperc, leaf carbon percentage; J_{max25}, the maximum rate of RuBP regeneration at 25°C; Nperc, leaf nitrogen percentage; SLA, specific leaf area; V_{c,max25}, the maximum rate of the carboxylation of Rubisco at 25°C.

Genomic information for the SoyNAM founders and RILs

Genetic marker information (Song et al. 2017) was accessed from the SoyNAM R package [data(soybase, package = "SoyNAM")], where missing and nonsegregating markers within a family were set to missing values (Diers et al. 2018; Xavier et al. 2018). The parental lines were deeply sequenced for SNP discovery and all lines were genotyped using the SoyNAM6K Beadchip (Anderson et al. 2014; Song et al. 2017). The genotypic data had previously been controlled for quality, which included the removal of lines that deviated from the expected marker segregation, ineffective crosses (likely self-pollinations), or segregation of alleles not associated with either parent allele. Additionally, all lines from NAM family N46 were removed following Diers et al. (2018) and Xavier et al. (2018) because of incorrect or failed crosses resulting in unexpected segregation ratios in the family.

In addition for this project, lines missing more than 50% of markers were removed from the dataset. Markers missing in >80% of the remaining lines were removed. Additional quality control and imputation of missing marker genotypes were performed using the snpQC function in the NAM R package (Xavier et al. 2015). Arguments in snpQC were set as follows: markers with minor allele frequencies <0.05 and monomorphic markers were removed. The default parameters for the other arguments were used. Missing data were imputed using the expected value. The final subset used in this analysis included 4,308 SNP markers, 5,171 F₅ derived lines, and 40 founder lines (Supplementary File 4).

Trait data collection: SoyNAM reflectance spectroscopy and yield

Reflectance spectra were collected from the SoyNAM lines planted at the NE12 and IL13 locations. The complete set of lines were grown at both locations. Leaf reflectance measurements were taken on three-fourths of the population and respective checks in 2012 (4,796 plots, 14,338 spectra), and the entire population in 2013 (6,400 plots, 19,200 spectra). Hyperspectral reflectance measurements were captured with a FieldSpec 4 Standard-Res Spectroradiometer as described above. Reflectance measurements were made during soybean reproductive stages R4 or R5 (seed fill) with a few exceptions due to the diversity of the SoyNAM population. Reflectance measurements were collected from an uppermost, fully expanded, sunlit leaf from 3 randomly sampled plants in each plot during midday (~11–15h) over ~2 weeks. Individual reflectance measurements per leaf were captured proximally from the adaxial side of the central leaflet using the ASD Plant Probe and Leaf Clip.

Reflectance measurements from the SoyNAM population (NE12 and IL13) were processed for quality, jump corrected, and averaged using the FieldSpec R Package (https://doi.org/10.5281/zenodo.6248237). The PLSR model coefficients (Supplementary File 1) for the PLSR models (Table 1; Supplementary File 2) were then applied to the averaged reflectance spectra (Supplementary File 5) to obtain the reflectance derived PLSR trait estimates for each plot from both environments separately (Supplementary File 6). Visual examination of histograms of the SoyNAM modeled trait estimates did not reveal significant departures from normality (Supplementary File 7). The yield data from the NE12 and IL13 SoyNAM lines were accessed from previously published work (Diers et al. 2018), using the SoyNAM R Package (Xavier et al. 2018), and subset to match the plots with available reflectance data.

Statistical analyses

Phenotypic models

A stage-wise linear mixed model similar to the method described in Diers et al. (2018) was used to incorporate the field design and environmental effects (location and year combination) into final estimates of the PLSR derived traits and yield. Statistical analyses used to estimate model parameters were conducted using restricted maximum likelihood (REML) (Patterson and Thompson 1971) implemented using the lme4 package (Bates et al. 2015) in R version 3.5.2 (R Core Team 2018). Specifically, incomplete blocks within the 2 environments (NE12 and IL13) were augmented with the common parent, IA3023, and 3 checks from the following list: IA2094, U06-100052, LD02-4485, LD04-11056, and LD04-13265, plus the founder parent of the lines in the block. The first-stage model used these checks to provide estimates of block effects to adjust genotypic values in the second-stage model for the leaf reflectance traits and yield. Empirical best linear unbiased predicted values (EBLUPs) of block effects were obtained using a linear mixed model (1):

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{C}\boldsymbol{\kappa} + \mathbf{B}\boldsymbol{\pi} + \boldsymbol{\varepsilon},$$

$$\boldsymbol{\pi} \sim \mathbf{N}(\mathbf{0}, \mathbf{I}\sigma_{\text{blk}}^2),$$

$$\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \mathbf{I}\sigma_{\varepsilon}^2),$$

$$\text{Cov}(\mathbf{B}\boldsymbol{\pi}, \boldsymbol{\varepsilon}) = \mathbf{0}.$$

For each trait, \mathbf{y} is a vector of phenotypes for the check varieties, $\boldsymbol{\mu}$ is the intercept, \mathbf{C} represents an incidence matrix for the check varieties, $\boldsymbol{\kappa}$ is a vector of fixed effects of the check varieties,

\mathbf{B} is an incidence matrix indicating block, $\boldsymbol{\pi}$ is the vector of random effects of incomplete blocks, \mathbf{I} is an identity matrix, σ_{blk}^2 is the variance among blocks, $\boldsymbol{\varepsilon}$ is a vector of residuals, and σ_{ε}^2 is the variance among the residual values.

The second-stage model (2) was applied to the PLSR derived trait values and yield for the lines and founders as follows:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\pi}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \boldsymbol{\varepsilon},$$

$$\mathbf{v} \sim N(\mathbf{0}, \mathbf{I}\sigma_g^2)$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_{\varepsilon}^2),$$

$$\text{Cov}(\mathbf{Z}\mathbf{v}, \boldsymbol{\varepsilon}) = \mathbf{0},$$

where \mathbf{y} is a vector of the phenotypes for the F_5 derived lines and founder lines for each trait (i.e. the PLSR derived trait values and yield). The length of \mathbf{y} , $\boldsymbol{\pi}$, and the dimension of the matrix \mathbf{Z} depend on the number of PLSR derived trait values from averaged reflectance spectra captured across the 2 years of the experiment. $\boldsymbol{\mu}$ is the intercept, $\boldsymbol{\pi}$ is a vector of shrunken EBLUP block effects from (1), considered a fixed effect covariate. $\boldsymbol{\beta}$ is the slope and intercept for the block effect. \mathbf{Z} is an incidence matrix for the lines and the founders indicating which block the measured trait value, \mathbf{y} , was obtained. \mathbf{v} is the vector of EBLUPs of genotypic values for the entries (lines and founders). \mathbf{I} is the identity matrix, σ_g^2 is the genotypic variance among the entries, and $\boldsymbol{\varepsilon}$ is the vector of residuals with variance, σ_{ε}^2 . EBLUPs, the estimated trait values across years and locations in this experiment, for the phenotypes are in [Supplementary Files 8 and 9](#). Pearson's correlations were performed on EBLUPs in a pairwise manner using the `cov.test` function in base R.

Variance components and heritability estimates

The estimated genetic variance (σ_g^2) and model error variance (σ_{ε}^2) were used to calculate the broad-sense heritability on an entry mean basis (H^2) over the 2 years: $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + (\sigma_{\varepsilon}^2/2)}$ (Table 2). Narrow-sense heritability based on marker data (h^2) for each trait was calculated by a multivariate mixed model approach using the MGREML function (default arguments used except for `-no-intercept` and `-ignore-collinearity`) from the MGREML Python package ([de Vlaming et al. 2021](#)) with a genomic relationship matrix (GRM) calculated via the `-make-grm` function in the GCTA software

Table 2. Estimates of genetic variance (σ_g^2) of the entry (RILs), the model error variance (σ_{ε}^2), and the broad-sense heritability (H^2)

with 95% confidence intervals (CI), where $H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{\varepsilon}^2/2}$ for the

EBLUPs of reflectance traits and yield in the SoyNAM population from data generated in Nebraska (2012) and Illinois (2013).

Trait	σ_g^2	σ_{ε}^2	H^2 (95% CI)
Chl	1.492E-04	4.347E-04	0.407 (0.386, 0.429)
Cperc	0.0214	0.0905	0.321 (0.304, 0.338)
$J_{\text{max}25}$	37.93	116.16	0.395 (0.375, 0.416)
Nperc	0.0133	0.057	0.320 (0.304, 0.337)
SLA	269.2	941	0.364 (0.345, 0.383)
$V_{\text{c,max}25}$	7.628	45.035	0.253 (0.240, 0.267)
Yield	95,044	365,325	0.342 (0.324, 0.361)

Chl, total chlorophyll content; Cperc, leaf carbon percentage; $J_{\text{max}25}$, the maximum rate of RuBP regeneration at 25°C; Nperc, leaf nitrogen percentage; SLA, specific leaf area; $V_{\text{c,max}25}$, the maximum rate of the carboxylation of Rubisco at 25°C.

P-value significance: ns, not significant; * < 0.05; ** < 0.01; *** < 0.001.

([Yang et al. 2011](#)). Collinearity between SLA and $J_{\text{max}25}$ resulted in a warning. Because of this collinearity and similar estimates of h^2 when random noise was added to the model allowing the MGREML function to run to completion (data not shown), the `-ignore-collinearity` argument was deemed appropriate and implemented to obtain the final variance estimates. Genetic correlations using marker data (GBLUPs) were estimated using the `-bivar` argument in GCTA for the traits shown in Table 3. A likelihood-ratio test (LRT) was used to determine significance (P value < 0.05) in GCTA by comparing the genetic correlation when set to 0 in the null hypothesis. An alternative REML algorithm (Fisher scoring approach) option was used when the null hypothesis would not converge. Additionally, the complete set of genetic correlation comparisons and h^2 were calculated using the MGREML approach outlined above and provided in [Supplementary File 10](#).

GWAS analyses of PLSR traits and yield in the SoyNAM population

GWAS detection followed the methods of [Diers et al. \(2018\)](#). EBLUPs of genotypic values from model 2 were used in genome wide associations to identify MTAs. The MTAs were identified as random effects that were dependent on family background. The parameter estimates were obtained via the NAM R package ([Xavier et al. 2015](#)) using an empirical Bayes algorithm ([Wei and Xu 2016](#)). The linear mixed model (3) is

$$\mathbf{v} = \boldsymbol{\mu} + \mathbf{W}\boldsymbol{\alpha} + \boldsymbol{\phi} + \boldsymbol{\varepsilon},$$

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{I}\sigma_{\alpha}^2),$$

$$\boldsymbol{\phi} \sim N(\mathbf{0}, \mathbf{K}\sigma_{\phi}^2),$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_{\varepsilon}^2),$$

$$\text{Cov}(\boldsymbol{\phi}, \boldsymbol{\varepsilon}) = \mathbf{0},$$

where \mathbf{v} are the EBLUPs of genotypic values from model (2), $\boldsymbol{\mu}$ is the intercept, \mathbf{W} is the incidence matrix containing the interaction between marker and family information—the haplotype representation of the common and founder parent for each marker. $\boldsymbol{\alpha}$ is a vector of regression coefficients associated with marker effects within family (i.e. the BLUP value for allelic substitution), $\boldsymbol{\phi}$ is the polygenic term accounting for the population structure among entries through the realized GRM \mathbf{K} , and $\boldsymbol{\varepsilon}$ is the vector of residuals. The regression fitted $\boldsymbol{\alpha}$ was interpreted as an estimate of marker substitution from each of the founder lines. There were potentially 40, 39 founders + 1 common parent, unique allelic substitution effects relative to the background ($\boldsymbol{\phi}$) for each marker loci. The polygene term was estimated with the genomic relationship \mathbf{K} that captures the additive relationship among individuals accounting for the population structure.

Specifically, the `gwas` function from the NAM R package ([Xavier et al. 2015](#)) was used for GWAS of all traits with the phenotypes coming from the PLSR derived traits and yield from model 2 (EBLUPs). The `gwas` function allows “linkage window” similar to what was outlined by [Xu and Atchley \(1995\)](#) to reduce the effect of tightly linked markers. A 5-cM window was used to match methods in [Diers et al. \(2018\)](#) as well as the significance threshold of $-\log_{10}$ (P value) ≥ 3 to determine significant MTAs. Two additional significance threshold estimates were provided for comparison, the

Table 3. Pairwise genetic correlations estimated from an REML approximated genetic relatedness matrix for a subset of modeled traits and the SE (in parenthesis) estimated by the bivar function in GCTA software (Yang et al. 2011).

	Chl	Cperc	J _{max25}	Nperc	SLA	V _{c,max25}	h ²
Chl							0.24 (0.028)
Cperc	0.45 (0.09)**						0.36 (0.028)
J _{max25}	0.34 (0.09)**	−0.80 (0.04)***					0.44 (0.028)
Nperc	0.45 (0.10)***	0.74 (0.05)***	−0.53 (0.08)***				0.28 (0.027)
SLA	−0.22 (0.10) ^{ns}	0.78 (0.05)***	−0.98 (0.004)***	0.56 (0.08)***			0.36 (0.028)
V _{c,max25}	0.11 (0.12) ^{ns}	−0.51 (0.09)***	0.36 (0.10)*	0.28 (0.12)*	−0.30 (0.09)*		0.19 (0.025)
Yield	0.45 (0.08)**	0.66 (0.06)***	−0.62 (0.06)***	0.83 (0.04)***	0.72 (0.05)***	−0.09 (0.10) ^{ns}	0.43 (0.030)

Significance testing used for log likelihood-ratio test (LRT) for each pairwise comparison was performed in GCTA with the genetic correlation of the null model fixed at 0. Multivariate genomic REML analysis (MGREML) (de Vlaming et al. 2021), was used for marker-based estimates and SEs of narrow-sense heritability (h²). Chl, total chlorophyll content; Cperc, leaf carbon percentage; J_{max25}, the maximum rate of RuBP regeneration at 25°C; Nperc, leaf nitrogen percentage; SLA, specific leaf area; V_{c,max25}, the maximum rate of the carboxylation of Rubisco at 25°C.

Bonferroni adjusted *P* value [$\alpha = 0.05$, 4,308 markers, $-\log_{10}(P \text{ value}) \geq 4.93$], and a trait specific *P* value outlined by Kaler and Purcell (2019) proposed as an alternative and less conservative approach based on marker heritability estimates (Supplementary File 11). The time required to perform a GWAS for a single trait was ~5 h (Intel Core i7-8650U CPU—1.90 GHz—16 GB of memory), so permutation was not considered because of the computational requirement. The GWAS was also performed without a linkage window. Similar MTAs were mapped using either method with the significance of the associations generally being greater without a linkage window (example in Supplementary File 12); but the 5-cM linkage window results are shown in alignment with Diers et al. (2018).

Multiple MTAs were assigned for a trait on a given chromosome (Chr) if the respective significant marker SNPs were at least 10 cM distant from each other, using map distances inferred by the physical position of the SNP on to the Williams 82 × *G. soja* PI 479752 (*W* × *P*) linkage map (Song et al. 2016). The most significant SNP in any given interval was assigned the MTA and its map position was determined by the *W* × *P* linkage map (Supplementary File 11). Summary information for significant SNP markers include the SoyNAM6K BeadChip SNP ID and the Wm82.a2.v1 position (Chr, base pair position), the reference and alternate allele for the SNP, the associated soybean gene models, and *Arabidopsis* annotation (Supplementary File 13). Gene models with “photosynthesis” GO Terms within ± 1.5 Mbp of the reported MTA interval for leaf traits were considered candidate genes and reported in Table 4 (Grant et al. 2010). SnpViz Version 2.0 (Langewisch et al. 2014) was used to view potential haplotypes across the Rubisco-small subunits (rbcS) identified by the GO Term on Chr 19 (Supplementary File 14). SnpViz settings were as follows (Data version: Wm82.a1.v1.1, Chromosome: Gm19, Range Type: Range Window, Starting point: 721600, Ending point: 731000, Clustering method: UPGMA, and Show Indel was marked). The mapping distance (cM) of the MTA intervals was estimated using the composite SoyNAM linkage map (Song et al. 2017) (Table 4).

For simplicity, tables and figures only present data for Chl, Cperc, J_{max25}, Nperc, SLA, and V_{c,max25}. Additional data of interest for these traits (i.e. confidence intervals for trait correlations from Fig. 4 are provided in Supplementary File 15), and related data for other traits (Chl a, Chl b, Chl a:b, J_{max}, and V_{c,max}) can be found in supplemental files.

Results

Part 1—Development of leaf reflectance models for soybean leaf traits

The soybean genotypes used for PLSR model-building showed variation in the measured leaf traits (Fig. 1; Supplementary Fig. 1).

Multiple traits were accurately predicted ($R^2 > 0.50$, RMSE ~10% of the range of the calibration data) from PLSR models (Fig. 1; Supplementary File 2) using the full-range (500–2,400 nm) of leaf reflectance data (Table 1; Supplementary Fig. 1). These traits included the rate-limiting steps of photosynthesis at the measured leaf temperature, V_{c,max} and J_{max}, and at a normalized leaf temperature of 25°C, V_{c,max25} and J_{max25}. The leaf reflectance data were also able to predict SLA, Chl, Chl a, Chl b, Chl a:b, Cperc, and Nperc. We determined that the leaf reflectance derived PLSR models did not accurately predict leaf protein content, leaf sucrose content, total leaf nonstructural carbohydrate content, or leaf temperature because of low R^2 values and high RMSE values relative to the range of the calibration data (Supplementary Fig. 1). Because of this limitation, these particular trait models were not used to estimate leaf reflectance derived traits in the SoyNAM population. The number of latent variables (or factors) used in the final PLSR models ranged from 4 for Chl to 18 for Nperc (Table 1; Supplementary File 2). The predictability of the different models varied when applied to the independent validation data. The best performing model was SLA ($R^2 = 0.96$, RMSE = 15.48), whereas the weakest model was J_{max25} ($R^2 = 0.56$, RMSE = 25.32), of all acceptable PLSR models (Fig. 1; Supplementary Fig. 1). VIP plots identified different spectral regions of importance for different physiological and biochemical traits. VIP plots showed that Chl was predominantly influenced by the visible spectrum (defined from 500 to ~750 nm for this experiment), while all the other models showed the importance of wavelengths in the near infrared and shortwave infrared regions (Supplementary Fig. 2). These leaf reflectance-based models were then used to predict leaf traits across the entire SoyNAM population.

Part 2—Application of soybean leaf reflectance models to the SoyNAM population

Variation in photosynthesis, leaf N, and SLA

Variation was observed in photosynthetic capacity within the SoyNAM population. The range of PLSR modeled trait values across the 2 years was 80 μmol CO₂ m^{−2} s^{−1} for V_{c,max25} and 134 μmol e m^{−2} s^{−1} for J_{max25} (Supplementary Files 6 and 7). Distributions of J_{max25} and V_{c,max25} EBLUPs by NAM family across both growing seasons also showed variation (Fig. 2). LD02-4485 was the founder parent with the greatest J_{max25} and V_{c,max25} values and NAM12, the family developed from a cross of common parent IA3023 with this line, had the highest median values for both traits across all NAM families (Fig. 2). Seven founder parents representing each background (elite, diverse ancestry, and PI) had greater J_{max25} values than IA3023 (Fig. 2). There were 12 founders that had a greater V_{c,max25} than IA3023, with 6 in the elite group, 5 in diverse ancestry, and 1 from the PI group (Fig. 2).

Table 4. List of gene models with GO term “photosynthesis” (GO:0015979) that have positions within ± 1.5 Mbp of SNP-based physical position of intervals containing PLSR modeled traits.

Interval	Chr	Left Mbp	Right Mbp	cM interval	Gene ID	Arabidopsis annotations	Traits in interval
2	10	43.2	46.7	3.6	Glyma.10g205500	Phosphoenolpyruvate carboxylase 4	Cperc, Nperc, $V_{c,max25}$ Chl, Chl a, Chl b J_{max} , J_{max25} , SLA, Yield
					Glyma.10g208900	Photosystem I subunit I	
					Glyma.10g225400	Inositol monophosphatase family protein	
3	10	47.1	50.1	10.6	Glyma.10g226400	Plastid-encoded CLP P	J_{max} , J_{max25} SLA Chl Nperc Chl a: b
					Glyma.10g243800	Chlorophyll a–b binding family protein	
					Glyma.10g249000	Photosystem I subunit D-2	
					Glyma.10g256000	Photosystem I reaction center subunit N	
4	12	34.2	37.2	6.7	Glyma.10g263000	Stress-enhanced protein 1	Nperc Chl a: b
					Glyma.12g193800	Aconitase 3	
					Glyma.12g200200	Photosystem II reaction center PSB29 protein	
5	13	29.5	32.5	3.1	Glyma.12g210600	Phosphoenolpyruvate carboxylase 2	Nperc
					Glyma.13g206700	Glucose-6-phosphate/phosphate translocator 2	
6	13	34.7	38.7	10.3	Glyma.13g270400	Phosphoenolpyruvate carboxylase 3	Chl, Chl a, Chl b, Chl a: b Cperc
					Glyma.13g282000	Light-harvesting chlorophyll b-binding protein 3	
					Glyma.13g286500	PsbQ-like 1	
7	13	37.4	40.4	8.7	Glyma.13g282000	Light-harvesting chlorophyll b-binding protein 3	Chl, Chl a, Chl b, Chl a: b Cperc $V_{c,max}$
					Glyma.13g286500	PsbQ-like 1	
					Glyma.13g290700	Phosphoenolpyruvate carboxylase 1	
					Glyma.13g302100	Photosystem II reaction center PSB29 protein	
					Glyma.13g308700	Aconitase 3	
8	13	39.9	42.9	11.9	Glyma.13g302100	Photosystem II reaction center PSB29 protein	Chl, Chl a, Chl b, Chl a: b Cperc, $V_{c,max}$
					Glyma.13g308700	Aconitase 3	
9	13	41.3	44.3	10.7	–	–	J_{max25}
10	15	3.0	6.0	6.2	Glyma.15g050100	Inositol monophosphatase family protein	J_{max}
					Glyma.15g052400	light harvesting complex photosystem II subunit 6	
13	18	0.2	3.6	4.1	Glyma.18g004600	Ferritin 4	Cperc Nperc Yield
					Glyma.18g011800	Photosystem II BY	
					Glyma.18g024600	Ferritin 4	
					Glyma.18g028400	Light harvesting complex photosystem II	
					Glyma.18g035300	Tetratricopeptide repeat (TPR)-like superfamily protein	
14	18	54.5	57.5	4.7	Glyma.18g262700	NAD(P)H dehydrogenase subunit H	Chl a J_{max} , J_{max25} SLA
					Glyma.18g286500	PsbP-like protein 1	
					Glyma.18g296900	Ribulose biphosphate carboxylase small chain 1A	
16	19	6.8	13.3	6.2	Glyma.19g046600	Ribulose biphosphate carboxylase (small chain) family protein	$V_{c,max25}$ Yield
					Glyma.19g046800	Ribulose biphosphate carboxylase (small chain) family protein	
					Glyma.19g046900	Ribulose biphosphate carboxylase (small chain) family protein	
					Glyma.19g047000	Ribulose biphosphate carboxylase (small chain) family protein	
					Glyma.19g051900	Ribosomal protein L2	
					Glyma.19g053400	Ribosomal protein L16	
					Glyma.19g054200	ATPase F0 complex subunit A protein	

Gene ID is given as the gene models described in Glyma.Wm82.a2.v1. Table shows the interval (provided in [Supplementary File 11](#)), chromosome, approximate range searched for gene models and the genetic distance (cM) of the interval, gene ID, the *Arabidopsis* annotation, and PLSR traits that mapped to the interval.

Ten founders, including LD02-4485 (NAM12), had greater leaf N content than the common parent with no founders from the PI drought tolerant group included ([Fig. 3](#), top). Only 6 founders, including the NAM12 parent, were predicted to have a lower SLA (i.e. a thicker leaf) than IA3023 ([Fig. 3](#), bottom).

Phenotypic and genetic correlations and trait heritability

Phenotypic correlations of EBLUPs between seed yield and all leaf traits were significant (P values <0.01), but with low correlation coefficients (r) ([Fig. 4](#); [Supplementary File 15](#)). The strongest

phenotypic correlation involving yield was with Chl ($r=0.255$, $CI_s=0.229$, 0.280), with $V_{c,max25}$, J_{max25} , and Chl all exhibiting moderate r values amongst themselves (0.492 – 0.517 ; see [Supplementary File 15](#) for CI_s). The strongest relationship between any 2 traits was J_{max25} and SLA ($r=-0.88$, $CI_s=-0.882$, -0.870). The broad-sense heritability estimates (H^2) on an entry mean basis were relatively low for most modeled leaf traits ranging from a low $H^2 = 0.253$ ($CI_s=0.240$, 0.267) for $V_{c,max25}$ to 0.407 ($CI_s=0.386$, 0.429) for Chl ([Table 2](#)). The marker-based genetic correlations among traits were generally higher than the

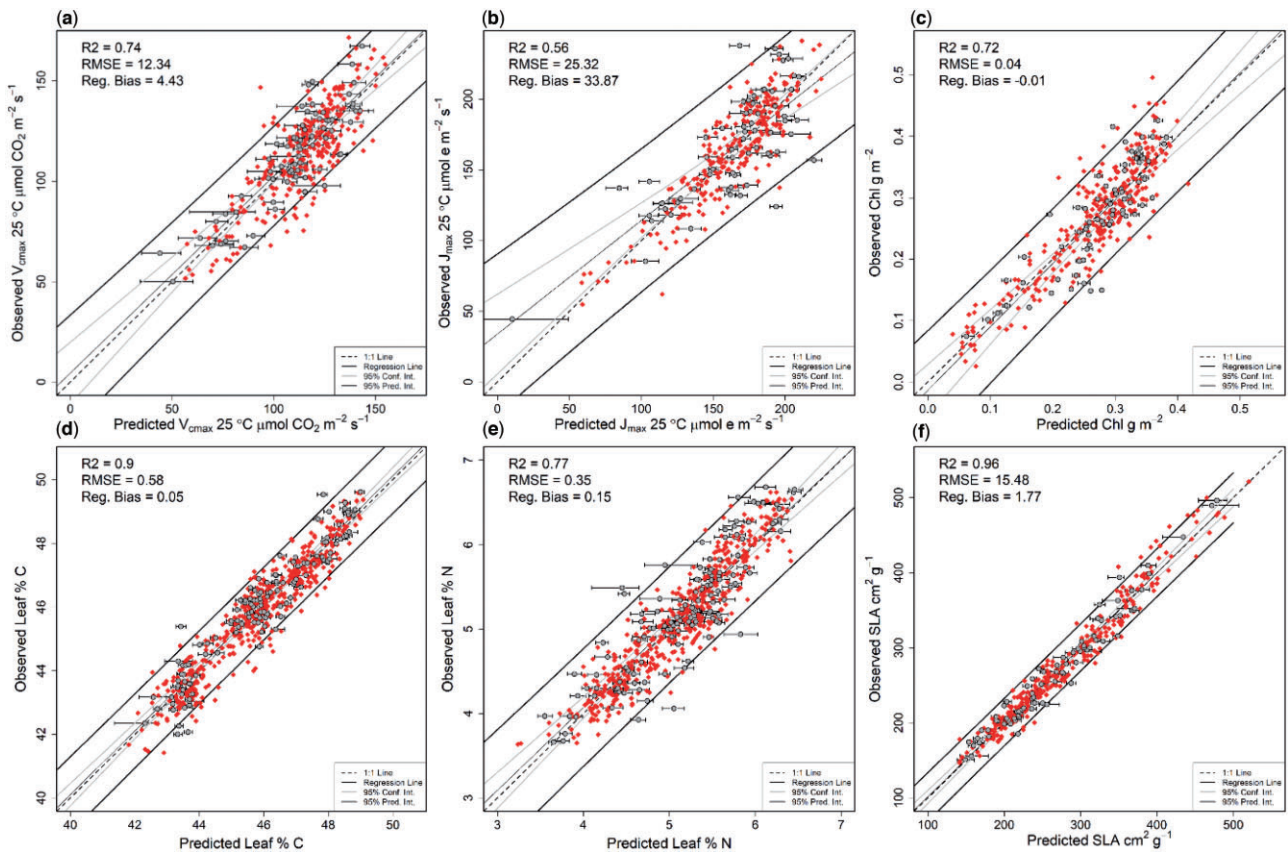


Fig. 1. PLSR model validations of (a) the maximum rate of the carboxylation of Rubisco at 25°C— $V_{c,max25}$; b) the maximum rate of RuBP regeneration at 25°C— J_{max25} ; c) total chlorophyll content—Chl; d) leaf carbon percentage (Cperc)—Leaf % C; e) leaf nitrogen percentage (Nperc)—Leaf % N; and f) specific leaf area—SLA. Red diamonds represent the data used in the calibration model. The y-axes represent measured values derived from A/c_i curves or standard laboratory procedures and the x-axes show the PLSR model predicted values. Light gray points represent validation data predicted by the PLSR model, with error bars representing the jack-knife confidence intervals of the predicted value, with both axes scaled in the specified PLSR unit of measurement. The 1:1 null hypothesis line is dashed, the linear regression line and its curved 95% confidence intervals for the observed values are solid gray, whereas the slightly curved 95% confidence intervals for the regression-predicted values are solid black curves.

phenotypic correlations (Table 3). Unlike the phenotypic correlations that were relatively weak, yield showed significant (P value < 0.001) positive genetic correlations with strong relationships with Cperc (0.66 ± 0.06), Nperc (0.83 ± 0.04), and SLA (0.72 ± 0.05). The genetic correlation of yield with $V_{c,max25}$ (-0.09 ± 0.10) was not significant, but was significantly negative with J_{max25} (-0.62 ± 0.06). Interestingly, the negative genetic correlation of SLA with J_{max25} was nearly unity (-0.98 ± 0.004), but was much less so with $V_{c,max25}$ (-0.30 ± 0.09). The narrow-sense heritability based on marker data (h^2) of these traits also increased relative to the H^2 estimates for most PLSR traits, with the lowest being $V_{c,max25}$ ($h^2 = 0.19 \pm 0.025$) and the highest being J_{max25} ($h^2 = 0.44 \pm 0.028$) (Table 3).

MTAs for photosynthetic traits and yield in soybean

GWAS identified a total of 49 significant MTAs ($-\log_{10} P > 3$), located on 8 of the 20 Chrs. On 3 of those 8 Chrs, a single MTA was detected (i.e. yield on Chr 3 and Chr 16, and Chl a:b ratio on Chr 12). Of the other 46 MTAs spread over 5 Chrs, 41 were specific for PLSR modeled traits, while 5 were for yield (Supplementary File 11). Multiple traits shared significant MTAs at a specific SNP marker or adjacent SNP markers, and most of these colocated MTAs were located on Chrs 10, 13, and 18. Notably, 10 traits were mapped to a small interval between 108.8 and 109.29 cM on Chr 10 (Table 4; Supplementary File 11). The MTAs for J_{max} , J_{max25} ,

and SLA were all highly significant ($-\log_{10} P > 25$) in this interval. Yield, Nperc, and Cperc MTAs were also colocated on Chr 18 (interval 13), and yield and $V_{c,max25}$ MTAs were colocated on Chr 19 (interval 16). The MTA on Chr 19 for yield occupied a position ca. 1.3 cM (ca. 3.5 Mbp) on the distal side of the $V_{c,max25}$ MTA. Positive effects for both traits (+/+ haplotype sign phase) were observed in the IA3023 parent. In the 12 founder families where effects were estimated for both traits (Supplementary File 11), the complementary phases of +/+ and -/- phase were evident in 6, whereas the cross-over phase types of +/- and -/+ were evident in the other 6. On no other Chr was there a close linkage of a MTA for yield and a MTA for $V_{c,max25}$. Both findings are consistent with a lack of correlation between the 2 traits (Tables 2 & 3). Four MTAs each were mapped for J_{max} and J_{max25} . Three of the 4 were colocated on Chr 10 and Chr 18. The common parent (IA3023) allelic effect was negative for these MTAs (Supplementary File 11). Most of the founder lines also had a negative allelic effect for the MTAs for J_{max} traits, but there were exceptions (Supplementary File 11). Only 2 MTAs were identified for $V_{c,max25}$, both of which were previously mentioned on Chr 10 and Chr 19 (Table 4). The common parent allelic effect was negative for the MTA on Chr 10 and positive for the MTA on Chr 19, with 16 and 8 founder parents with higher values for $V_{c,max25}$, respectively (Supplementary File 11). The non-normalized maximum rate of carboxylation, $V_{c,max}$, also had 2 MTAs, both

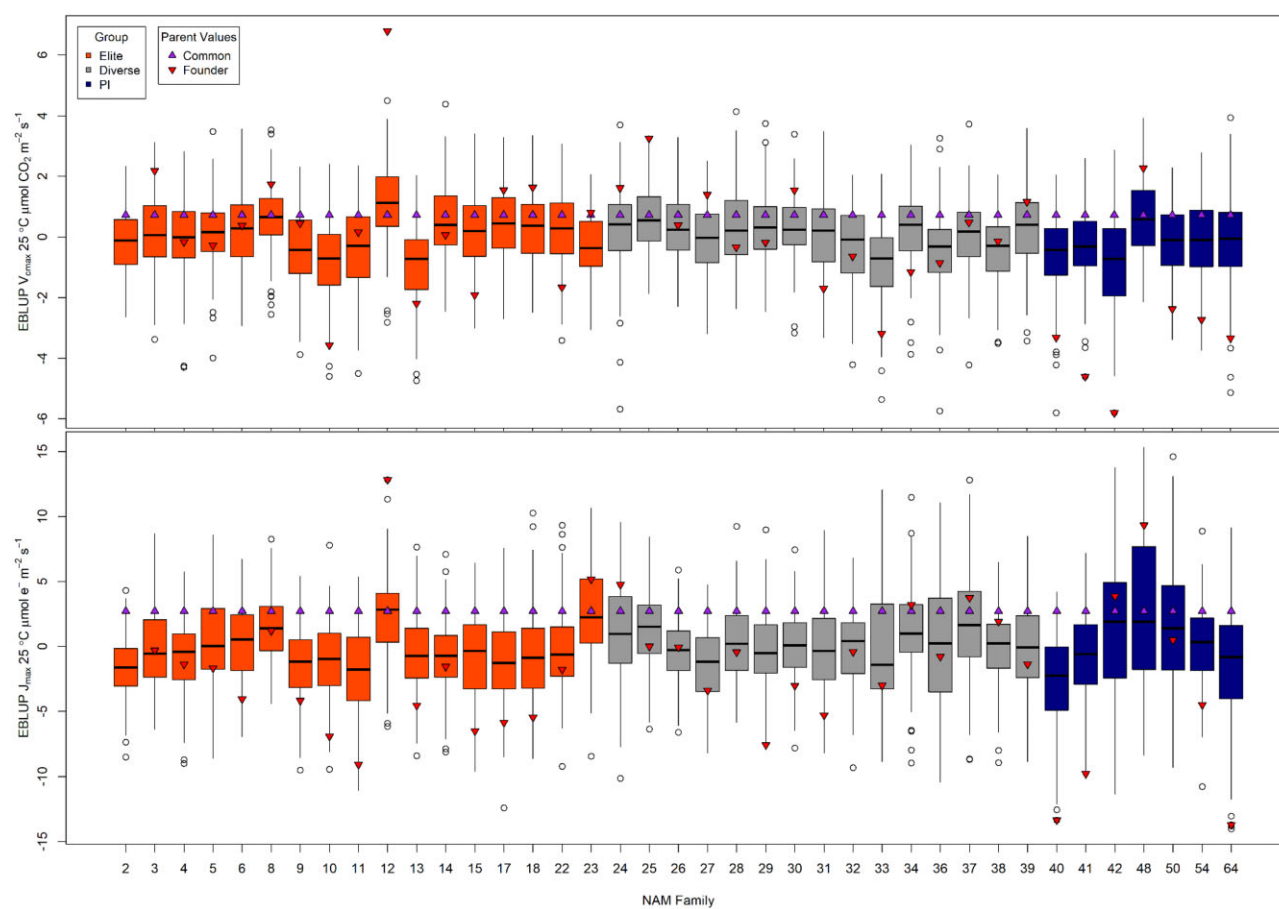


Fig. 2. Boxplots of EBLUPs for the maximum rate of carboxylation of Rubisco, $V_{c,max25}$ (top) and the maximum rate of RuBP regeneration, J_{max25} (bottom) for the SoyNAM families. Elite families (2-23), Diverse ancestry families (24-39), and PI families (40-64) are presented. The value for the common parent (IA3023) is shown in upward triangles. The founder parents for a given family are shown with downward triangles. Median values are the black horizontal bar, and outliers are indicated by open points.

located on Chr 13. The common allelic effect was positive for both with only a small number of founder lines providing a larger positive allelic effect (Supplementary File 11). A GO term analysis of all gene models within ± 1.5 Mbp of the MTAs associated with the rate-limiting steps of photosynthesis, $V_{c,max25}$ and J_{max25} identified at least 1 gene model associated with the GO term “photosynthesis” (Table 4).

Discussion

Improving photosynthesis is proposed as a next step toward increasing crop yield to meet the growing demands for feed, food, and fuel (Zhu et al. 2010; Ort et al. 2015). Developing high-throughput ways to effectively measure photosynthetic and other leaf traits in large populations would allow researchers to better understand not only the natural genetic variation and architecture of these traits, essential for improvement (van Bezouw et al. 2019), but also the relationships between leaf traits and yield. The goals of this work were (1) to develop a nondestructive, high-throughput method to quickly screen soybean germplasm to estimate the rate-limiting steps of photosynthesis and other leaf traits; (2) to determine if there was natural variation for these traits in soybean; and (3) to map these traits in the SoyNAM population thereby ultimately expanding the phenotypic toolbox available to researchers and soybean breeders.

High-throughput modeling of photosynthetic traits

To develop robust models of photosynthetic capacity and leaf traits, a deliberate effort was made to incorporate genotypic and phenotypic variation in the calibration dataset used to build the PLSR models (Fig. 1; Supplementary Table 2). This is recommended practice to capture as much of the expected trait space as possible (Burnett et al. 2021). We used diverse germplasm grown over several seasons to integrate differences in weather and genetics. Observations from different experiments at the SoyFACE Research Facility including variable ozone pollution (Ainsworth et al. 2014) and temperature treatments (Kumagai et al. 2022) were also used to expand the inference environment (Supplementary Table 1). In total, we developed 11 reflectance-based trait models (Table 1; Supplementary Files 1 and 2), which can be used to screen soybeans for the rate-limiting steps of photosynthesis, $V_{c,max}$ and J_{max} (including their 25°C normalized derivatives), chlorophyll traits (total chlorophyll content, chlorophyll a, chlorophyll b, and the chlorophyll a to chlorophyll b ratio), SLA, and leaf C and N percentages. Although the VIP scores indicate the predictive strength of different spectral regions for individual traits (Supplementary Fig. 2), limited mechanistic interpretation of the PLSR models is suggested (Tobias 1995).

The reflectance technique outlined in this manuscript takes ~30s to collect an average leaf reflectance spectrum from a soybean plot (3 separate leaves in this study). For comparison,

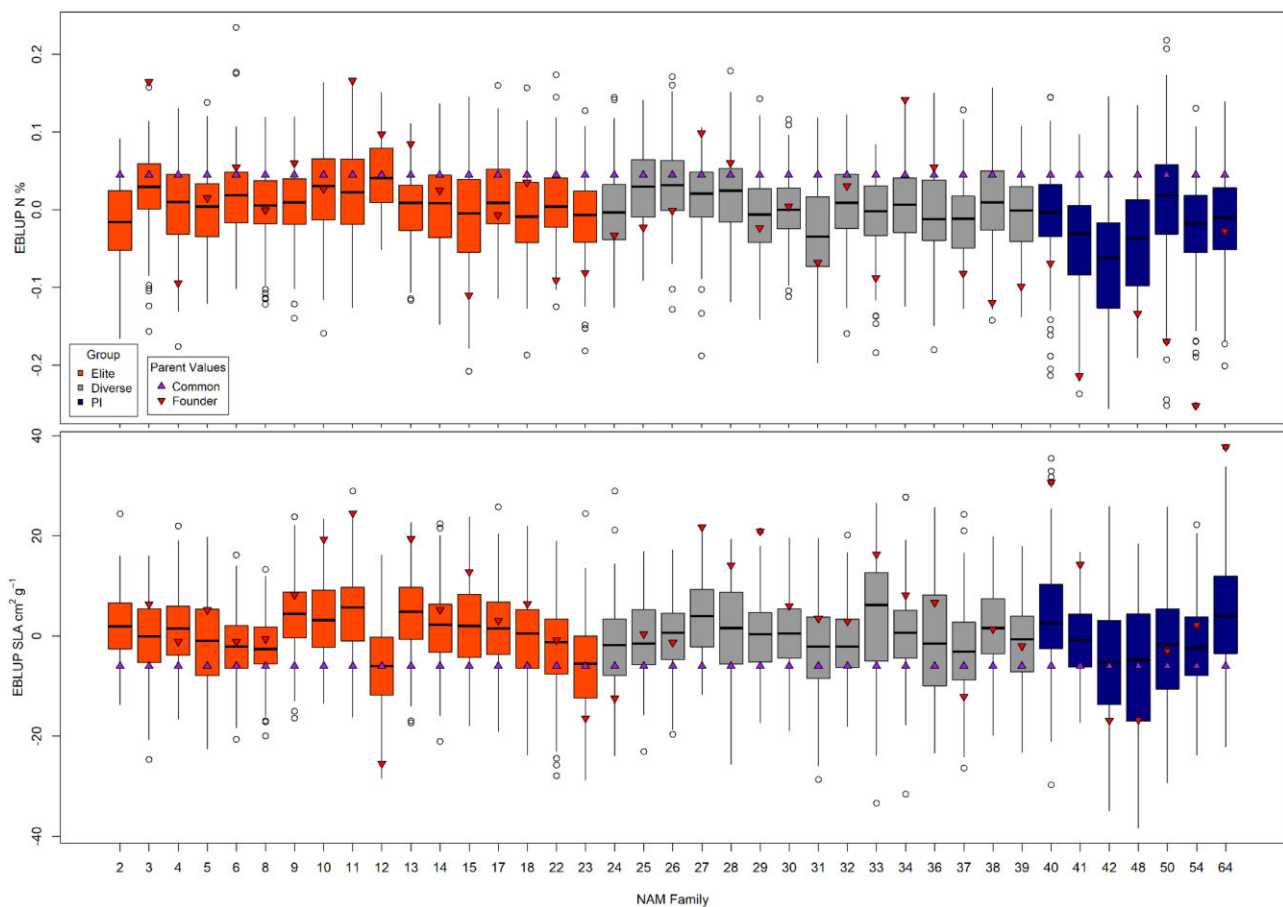


Fig. 3. Boxplots of EBLUPs for leaf nitrogen percentage (Nperc, N %) and specific leaf area (SLA) for the SoyNAM families. Elite families (2-23), Diverse ancestry families (24-39), and PI families (40-64) are presented. The value for the common parent (IA3023) is shown in upward triangles. The founder parents for a given family are shown with downward triangles. Median values are the black horizontal bar, and outliers are indicated by open points.

a standard A/Ci curve using a portable photosynthesis gas exchange system to estimate the rate limiting steps of photosynthesis takes ~30 min per leaf. The use of leaf reflectance to estimate these traits represents a 60× increase in the rate of data collection, enabling measurement of more than 5,000 plots at “midday” (11–15 h) over ~2 weeks with a single spectroradiometer. For comparison, it would take more than 100 continuous, 24-h days using a single gas exchange system to measure that number of plants. While this leaf-level reflectance approach may not be as high-throughput as aerial-based platforms (gantries, drones, satellites, etc.); there is much greater complexity and uncertainty in estimating photosynthetic traits remotely (Serbin et al. 2015; Meacham-Hensold et al. 2020). Leaf reflectance models provide an important benefit over traditional gas exchange approaches enabling investigation into the genetics of photosynthetic traits, albeit, with a tradeoff of accuracy as with virtually all high-throughput phenotyping (van Bezouw et al. 2019).

Natural variation of photosynthetic and leaf traits and their correlations in the SoyNAM population

Although the domestication bottleneck within US soybean germplasm has narrowed available genetic diversity for selection (Sedivy et al. 2017), significant variation in photosynthetic capacity was still evident within the NAM founder families (Fig. 2). This result was in agreement with previous research, which has also documented the existence of genetic variation in photosynthetic traits in soybean germplasm, but in a dramatically smaller

number of genotypes (Betzelberger et al. 2010; Gilbert et al. 2011; Koester et al. 2016; Sakoda et al. 2016; Soleh et al. 2017; Tomeo and Rosenthal 2017; Lopez et al. 2019; Wang et al. 2020b). Some recent studies have specifically concentrated on gas exchange traits within the SoyNAM population permitting a general comparison with our results. Soleh et al. (2017) found significant variation in photosynthetic induction, the gradual increase to maximum photosynthesis when transitioning from low to high light, in the founder lines. This variation was attributed to Rubisco activation. NAM founder line 12 (LD02-4485) had the slowest rate of photosynthetic induction, while NAM founder line 23 (U03-100612) had the fastest rate of induction. Notably rates of induction were not correlated with steady-state rates of CO₂ fixation (Soleh et al. 2017). Intrinsic water use efficiency (iWUE) and A were previously measured using gas exchange in a subset of RILs from the SoyNAM families, and the NAM12 family had a relatively high median rate of A in 2 of 3 environments (Lopez et al. 2019). In our study, the NAM12 family had higher $V_{c,max25}$ and J_{max25} , and the NAM23 family had higher J_{max25} compared with other families (Fig. 2). The agreement between our results and the previous study provides additional confidence in our high-throughput approach for estimating photosynthetic capacity, including the identification of families of interest for future study.

Phenotypic correlations between photosynthetic capacity ($V_{c,max25}$, J_{max25}) and yield were significant, but weak, across the NAM population (Fig. 4); however, the genetic correlations between these traits are likely to be of more interest to soybean

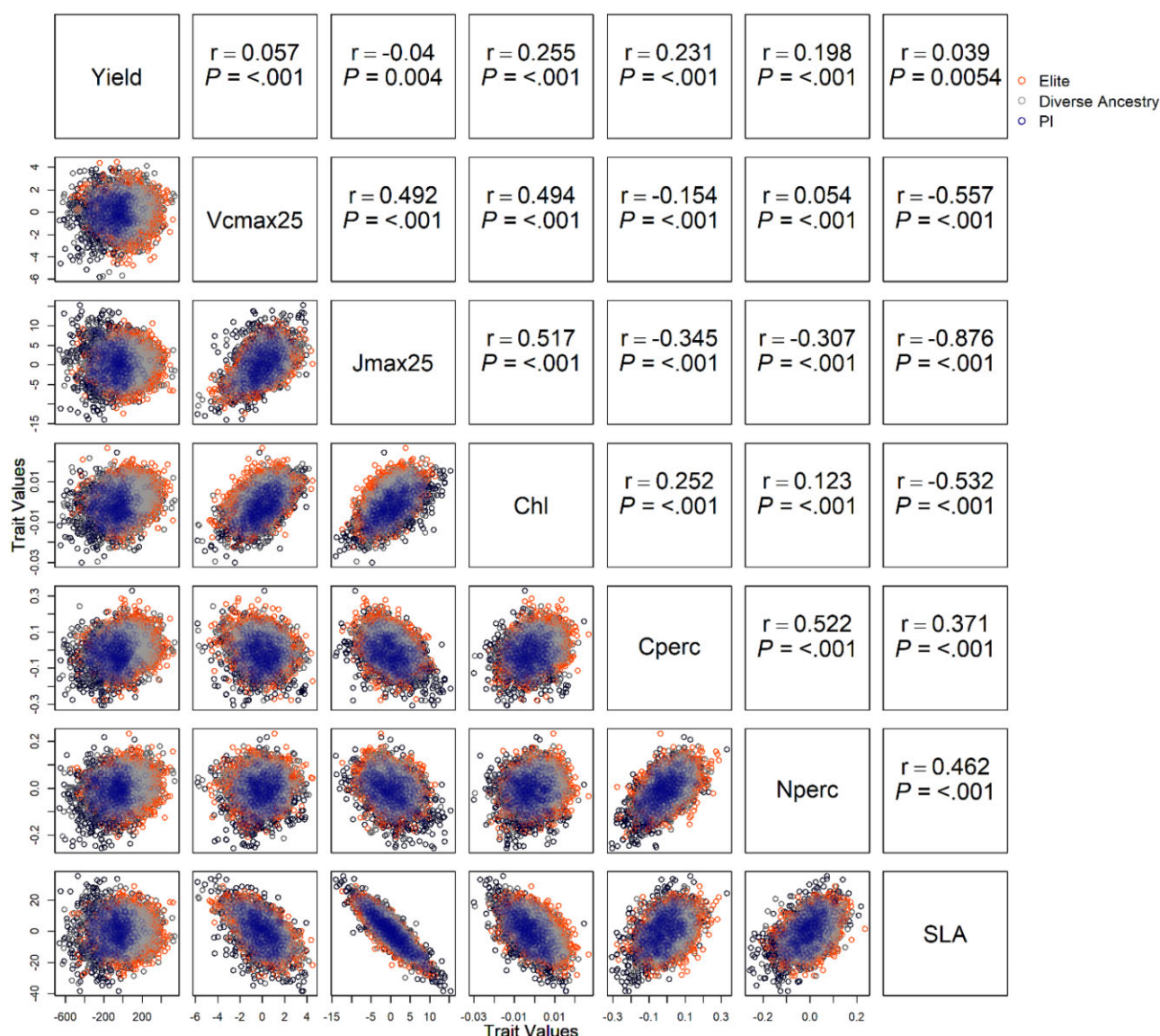


Fig. 4. Scatterplots of EBLUPs for yield and modeled PLSR traits are shown in the lower diagonal. The upper diagonal shows the Pearson correlation coefficient (r) and P values for phenotypic correlations. All pairwise trait comparisons were significant (P value < 0.01 , see [Supplementary File 15](#) for 95% confidence intervals and t -test statistics). Orange, gray, and blue dots represent RILs from Elite, Diverse Ancestry, and PI founders, respectively. Trait abbreviations: the maximum rate of the carboxylation of Rubisco at 25°C ($V_{c,max25}$), the maximum rate of RuBP regeneration at 25°C (J_{max25}), total chlorophyll content (Chl), leaf carbon percentage (Cperc), leaf nitrogen percentage (Nperc), and specific leaf area (SLA).

breeders (Falconer and Mackay 1996; Cheverud 2001) (Table 3). Notably, J_{max25} was moderately correlated with yield (-0.62 ± 0.06 , P value < 0.001), whereas $V_{c,max25}$ was not (-0.09 ± 0.10 , P value > 0.05). The latter result implies that the genetic basis of phenotypic variation in $V_{c,max25}$ and yield are not congruently concordant (i.e. direct selection for one is unlikely to result in any correlated change in the other). Direct selection targeting J_{max25} as a means of attaining an indirect correlated increase in yield is theoretically achievable; however, given their negative genetic correlation (Table 3), one must select for lower (not higher) J_{max25} values, and notably, selection for lower J_{max25} would also likely lead to lower $V_{c,max25}$ values due to their positive genetic correlation ($+0.36 \pm 0.10$, P value < 0.05). This finding is surprising and runs counter to the thesis that increased photosynthesis will result in increased yields (Long et al. 2006; Ort et al. 2015), as well as the results from a crop modeling simulation study performed by Wu et al. (2019). These authors showed that a

20% increase in J_{max25} would likely lead to respective 7% and 8% increases in the yields of the monocot C3 wheat and C4 sorghum crop species, whereas a 20% increase in $V_{c,max25}$ would likely lead to just a negligible 0.4% increase in sorghum yield, and zero change in wheat yield. The authors also showed that synchronous 20% increases in both $V_{c,max25}$ and J_{max25} would likely lead to somewhat greater increases of a respective 10.6% and 9.2% in wheat and sorghum yields, respectively. These model simulation results suggested that selection targeting J_{max25} alone was likely to be very effective in generating correlated yield improvement, but ultimately, targeted selection of both rate-limiting steps of photosynthesis was expected to be ultimately needed to enable the interactive synergism of both traits relative to monocot yield improvement. However, this thesis does not seem to be the case with respect to the dicot C3 soybean as shown in our results.

Other strong genetic and phenotypic correlations were identified (Table 3 and Fig. 4), and the -0.98 ± 0.004 (P value < 0.001)

genetic correlation between SLA and $J_{\max 25}$ was the strongest. This correlation suggests that if one is interested in increasing $J_{\max 25}$, then SLA (a trait accurately modeled by PLSR and easily measured with rudimentary equipment—a leaf punch) may serve as a proxy indicator to quickly select/identify breeding lines with contrasts in $J_{\max 25}$. Those lines could then be made available for more detailed physiological characterization. The traits Nperc and Cperc were also of interest, not because of their high genetic correlation ($+0.74 \pm 0.05$, P value <0.001), but because of their respective high correlations with yield ($+0.83 \pm 0.04$ and $+0.66 \pm 0.06$, P values <0.001) (Table 3). These high values may attract soybean breeders to use leaf reflectance derived estimates of Nperc and Cperc as potential proxy targets of direct selection that may bring about a correlated yield response. These findings correspond with the leaf economic spectrum, which describes the global relationship between leaf nitrogen, leaf thickness (SLA), and photosynthesis, and highlights the ability of the PLSR models to accurately estimate established trait relationships (Wright et al. 2004). However, the negative relationship between $J_{\max 25}$ and Nperc was not an expected result. This result may be driven by Nperc being estimated on a leaf mass basis instead of a leaf area basis, whereas $J_{\max 25}$ is measured on an area basis (Lloyd et al. 2013). Alternatively, the trait relationships outlined in the leaf economic spectrum are derived from very diverse species and may be less stable within a specific species.

Genetic mapping of modeled photosynthesis and leaf traits

Many of the MTAs for multiple traits were collocated to narrow intervals on Chrs 10, 13, and 18 (Table 4; Supplementary File 11). More specifically, all traits except for $V_{c,\max}$ and Chl a:b mapped to an interval on Chr 10 adjacent to the soybean maturity gene E2 (Glyma.10G221500) (Watanabe et al. 2011). The late maturity E2 allele was shown to be segregating in 7 of the NAM families; all other families were homozygous for the early maturity e2 allele (Langewisch et al. 2014; Diers et al. 2018). The E2/e2 locus influenced all of the measured yield and agronomic traits in Diers et al. (2018) as well as canopy coverage (Xavier et al. 2017) in the SoyNAM population. Variation in the traits modeled in this study was likely influenced by effects of this maturity gene that influences both plant development and leaf age (Kerstetter and Poethig 1998; Bielczynski et al. 2017). Throughout leaf development, the capacity of leaves to perform photosynthesis varies, as does the leaf size, thickness, pigment content from a host of different environmental cues (nutrient availability, water, light, and temperature) in response to feedback mechanisms within the plant. However, segregation at the E2/e2 locus certainly exerted pleiotropic control of at least some of the total variation measured for these traits (Falconer and Mackay 1996; Cheverud 2001). E2 is an ortholog of the *Arabidopsis* gene GIGANTEA, which is a known circadian clock gene associated with flowering and maturity (Watanabe et al. 2011). A growing body of research shows that GIGANTEA functions beyond flowering and maturity (Mishra and Panigrahi 2015) likely having roles in sink capacity via maintenance of the inflorescence architecture and limiting floral abortion (Brandoli et al. 2020), chlorophyll content, sucrose sensitivity (Dalchau et al. 2011; Mora-Garcia et al. 2017), stress tolerance (Riboni et al. 2013), carbon metabolism (Mugford et al. 2014; Krahmer et al. 2019), photosynthetic capacity, and plant growth (Dodd et al. 2005, 2015).

Because soybeans preferentially reproduce through self-pollination, the extended linkage disequilibrium found in both domesticated and wild soybean, and the relatively small number

of SNPs used in this analysis relative to other GWAS (Chung et al. 2014; Xavier et al. 2016), the specific marker SNPs and gene models directly associated with the MTAs identified for these PLSR traits should be viewed as putative targets for further study (i.e. this study lacks resolution) (Supplementary File 13). Given parental lines were deeply sequenced for SNP discovery, one strategy to overcome the lack of resolution in future work is parent-progeny genotype imputation (Technow and Gerke 2017; Gonen et al. 2021). Searching the adjacent regions of the genome identified some candidate genes that were more plausible controls of the PLSR modeled variation in photosynthetic traits (Table 4), but these should also be considered putative until additional research is performed.

A significant MTA for $V_{c,\max 25}$ was identified on Chr 19 (interval 16) within which there are 7 annotated gene models associated with “photosynthesis.” Most notably, 4 of these gene models (Glyma.19g046600 with Glyma.19g046800, Glyma.19g046900, and Glyma.19g04700 in tandem array within 50 kb, soybase.org) were identified as the *rbcS* (Table 4) and appear to have 2 haplotypes within the founder lines (Supplementary File 14). Rubisco is the initial site of carbon fixation for C3 plants. The rate at which Rubisco is able to effectively fix CO_2 ($V_{c,\max 25}$) is a major limitation of photosynthesis and has been proposed as a target for improvement (Spreitzer and Salvucci 2002; Cavanagh and Kubien 2014; Ort et al. 2015; Niinemets et al. 2017). The exact role of the *rbcS* is “enigmatic” with several proposed functions making the *rbcS* an active area of research (Cavanagh 2020). The *rbcS* genes in tandem array may result in greater transcript abundance or provide some novel mutation and function (Rosloski et al. 2010; Panchy et al. 2016; Das and Bansal 2019) ultimately leading to the modeled variation in $V_{c,\max 25}$.

Four significant MTAs were mapped for $J_{\max 25}$, the maximum rate of electron transport leading to the regeneration of RuBP, on Chrs 10, 13, and 18. Intervals for the MTAs on Chrs 10 and 18 contain gene models with GO terms for “photosynthesis” (Table 4). The most intriguing candidate gene is Glyma.10g225400, which encodes a fructose-1,6-bisphosphatase (FBPase). FBPase has been proposed as a target to improve photosynthesis (Tamoi et al. 2006; Zhu et al. 2008), and transgenic soybean overexpressing a cyanobacterial FBPase showed increased rates of J_{\max} (and $V_{c,\max}$) and improved yield at elevated temperature (Köhler et al. 2017). While this was the most significant MTA for $J_{\max 25}$, the proximity (~300 kilobase pair) of this MTA for $J_{\max 25}$ to the soybean maturity gene E2 may be problematic from a breeding perspective. The positive allelic effects of this MTA were predominately in the NAM families with the later maturity E2 allele (Supplementary File 11), in contrast to most adapted north central, USA germplasm which possesses the earlier maturity e2 allele. The other candidate genes for $J_{\max 25}$ were specific to chlorophyll a-b binding and the photosystems associated with the light reactions. These candidate genes likely have some role in light capture, linear electron flow, and ultimately the differences modeled for $J_{\max 25}$ in soybean (Yamori and Shikanai 2016; Nikkanen et al. 2018).

Most genetic mapping studies of photosynthetic traits are focused on pigments associated with photosynthetic efficiency (Dhanapal et al. 2016; Herritt et al. 2016). This is likely because of the difficulty in phenotyping photosynthetic efficiency and other gas exchange traits. With so few groups having mapped gas exchanged derived photosynthetic traits in soybean (Li et al. 2016; Lü et al. 2018; Lopez et al. 2019; Wang et al. 2020a), there are few, if any, direct comparisons that can be made between the MTAs for photosynthesis mapped in this study. Lopez et al. (2019) mapped

MTAs using a subset of 383 RILs and founder lines from the SoyNAM population for A and instantaneous water use efficiency (A divided by stomatal conductance). No significant MTAs identified in our research project collocated to intervals identified by Lopez *et al.* (2019) (Table 4; Supplementary File 11). The traits mapped between the 2 studies are related but different, so this is understandable. The MTAs identified in both studies provide a better understanding of the genetic controls of photosynthesis, multiple targets for improving photosynthetic traits, and opportunities for future analyses in soybean. Both studies also show low phenotypic correlations between the measured photosynthesis traits and final yield, and slightly stronger genetic correlations between these traits. This finding suggests that there may be a shared genetic component, albeit minor, and only for some photosynthetic traits and yield (Table 3). Because of the limited or negative genetic correlations between yield and photosynthetic traits identified in this study, multiple trait selection, and improvement is unlikely; however, other leaf traits may be better suited for concordant improvement (Neyhart *et al.* 2019). The heritability of photosynthetic traits in both studies are considered moderate to low (Holland *et al.* 2010) (Tables 2 and 3). This is likely because photosynthetic traits are complex, polygenic traits that are strongly influenced by changes in the phenology of the plant and the environment including air temperature, light, and water availability (Bernacchi *et al.* 2013; van Bezouw *et al.* 2019).

Conclusion

This work demonstrated the development and use of a high-throughput approach leveraging leaf reflectance data to determine photosynthetic and other leaf traits in soybean. We then used these high-throughput models to estimate leaf traits including the rate limiting steps of photosynthesis, a proposed means of improving crop yields, to the SoyNAM population gaining knowledge of the genetic architecture, natural variation, and potential targets for improvement. SoyNAM families with high photosynthetic capacity were identified. The respective low and negative correlations of $V_{c,max25}$ and J_{max25} with yield suggests that these rate-limiting steps of photosynthesis are not currently limiting soybean yield potential. We also showed that 3 modeled leaf traits: leaf carbon percentage, leaf nitrogen percentage, and SLA had positive genetic correlations with yield. While these relationships may not be entirely novel, the ability to estimate these traits in a high-throughput manner using leaf reflectance could be incorporated in a large breeding program as a proxy for yield. MTAs were mapped for all of the modeled leaf traits, and we identified some that co-located for photosynthetic capacity and yield. Additionally, some MTAs for the photosynthetic traits mapped to reasonable candidate genes, most notably $V_{c,max25}$ and the *rbcS*. This provided some additional confidence that the modeled traits are indeed predicting the measured traits. While we are quite enthusiastic about the approach outlined here, we were unable to address if we lose power or have differences in QTL effect size by mapping with the modeled traits over the measured traits. This will certainly need to be addressed in future research to establish confidence in the community that the mapped associations are in fact real and not simply error introduced by the high-throughput technique; however, the strides made in rapidly phenotyping these difficult to measure photosynthetic traits as well as multiple other leaf traits at once may outweigh, if even only temporarily, the tradeoff in loss of prediction accuracy.

Data availability

All of the genetic and yield data underlying this article are publicly available from the SoyNAM R package (Xavier *et al.* 2018; Diers *et al.* 2018) accessed from the Soybase.org website. Supplemental Files 1-18 are available at Figshare: <https://doi.org/10.25386/genetics.19394693>, including all other measured phenotype and leaf reflectance data (Supplementary Files 16 and 17) and a complete dataset to perform all of the genetic mapping and analyses (Supplementary File 18). The majority of lines and accessions used in this manuscript are available via GRIN (<https://www.ars-grin.gov/>) or by request from Soybase.org for the NAM lines (https://soybase.org/SoyNAM/SoyNAM_RIL_request.htm).

Supplemental material is available at GENETICS online.

Acknowledgments

We thank Troy Cary, Chris Moller, and Noah Mitchell for help in setting up and maintaining the experimental plots, collecting data, and processing samples.

EAA, CF, SPS, and BWD conceived the original project. CMM, EAA, and BWD wrote the manuscript. CMM, ASS, CF, SPS, and EK collected data and built PLSR models. CMM, AX, and MDK performed the statistical analysis for genomic mapping. EAA, CJB, WDB, JES, and BWD provided data, resources, and scientific guidance.

Funding

This work was supported by soybean checkoff funding from the United Soybean Board. ASS was supported by a postdoctoral fellowship granted by the Education, and Linguistic Policy Department of the Basque Country, Spain and is currently partially supported by the Alabama Agricultural Experiment Station and the Hatch Program of the National Institute of Food and Agriculture, US Department of Agriculture. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the US Department of Agriculture. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. USDA is an equal opportunity provider and employer.

Conflicts of interest

None declared.

Literature cited

- Ainsworth EA, Rogers A, Leakey ADB. Targets for crop biotechnology in a future high CO₂ and high O₃ world. *Plant Physiol.* 2008;147(1): 13–19.
- Ainsworth EA, Rogers A, Leakey ADB, Hady LE, Gibon Y, Stitt M, Schurr U. Does elevated atmospheric CO₂ alter diurnal C uptake and the balance of C and N metabolites in growing and fully expanded soybean leaves? *J Exp Bot.* 2007;58(3):579–591.
- Ainsworth EA, Serbin SP, Skoneczka JA, Townsend PA. Using leaf optical properties to detect ozone effects on foliar biochemistry. *Photosynth Res.* 2014;119(1–2):65–76.
- Anderson JE, Kantar MB, Kono TY, Fu F, Stec AO, Song Q, Cregan PB, Specht JE, Diers BW, Cannon SB, *et al.* A roadmap for functional

- structural variants in the soybean genome. *G3* (Bethesda). 2014; 4(7):1307–1318.
- Araus JL, Cairns JE. Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 2014;19(1):52–61.
- Araus JL, Kefauver SC, Zaman-Allah M, Olsen MS, Cairns JE. Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 2018;23(5):451–466.
- Asner GP. Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sens Environ.* 1998;64(3):234–253.
- Asner GP, Scurlock JMO, Hicke JA. Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Glob Ecol Biogeogr.* 2003;12(3):191–205.
- Bates D, Machler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 2015;67(1):48.
- Beavis W. The power and deceit of QTL experiments: lessons from comparative QTL studies. In: *Proceedings for the Forty-Ninth Annual Corn & Sorghum Industry Research Conference*. Washington (DC): American Seed Trade Association; 1994. p. 250–266.
- Bernacchi CJ, Bagley JE, Serbin SP, Ruiz-Vera UM, Rosenthal DM, Vanlooche A. Modelling C3 photosynthesis from the chloroplast to the ecosystem. *Plant Cell Environ.* 2013;36(9):1641–1657.
- Bernacchi CJ, Pimentel C, Long SP. In vivo temperature response functions of parameters required to model RuBP-limited photosynthesis. *Plant Cell Environ.* 2003;26(9):1419–1430.
- Bernacchi CJ, Singsaas EL, Pimentel C, Portis AR, Long SP. Improved temperature response functions for models of Rubisco-limited photosynthesis. *Plant Cell Environ.* 2001;24(2):253–259.
- Betzberger AM, Gillespie KM, McGrath JM, Koester RP, Nelson RL, Ainsworth EA. Effects of chronic elevated ozone concentration on antioxidant capacity, photosynthesis and seed yield of 10 soybean cultivars. *Plant Cell Environ.* 2010;33:1569–1581.
- Bielczynski LW, Łacki MK, Hoefnagels I, Gambin A, Croce R. Leaf and plant age affects photosynthetic performance and photoprotective capacity. *Plant Physiol.* 2017;175(4):1634–1648.
- Bouchet S, Olatoye MO, Marla SR, Perumal R, Tesso T, Yu J, Tuinstra M, Morris GP. Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. *Genetics.* 2017;206(2):573–585.
- Brandoli C, Petri C, Egea-Cortines M, Weiss J. The clock gene *Gigantea 1* from *Petunia hybrida* coordinates vegetative growth and inflorescence architecture. *Sci Rep.* 2020;10(1):275.
- Brock MT, Rubin MJ, DellaPenna D, Weinig C. A nested association mapping panel in *Arabidopsis thaliana* for mapping and characterizing genetic architecture. *G3* (Bethesda). 2020;10(10):3701–3708.
- Burnett AC, Anderson J, Davidson KJ, Ely KS, Lamour J, Li Q, Morrison BD, Yang D, Rogers A, Serbin SP. A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. *J Exp Bot.* 2021;72(18):6175–6189.
- Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Araus JL. High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. *J Integr Plant Biol.* 2012;54(5):312–320.
- Carter GA, Knapp AK. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *Am J Bot.* 2001;88(4):677–684.
- Cavanagh AP. Big progress for small subunits: new Rubisco mutants in *Arabidopsis*. *J Exp Bot.* 2020;71(19):5721–5724.
- Cavanagh AP, Kubien DS. Can phenotypic plasticity in Rubisco performance contribute to photosynthetic acclimation? *Photosynth Res.* 2014;119(1–2):203–214.
- Cheverud JM. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* (Edinb). 2001;87(Pt 1):52–58.
- Chung W-H, Jeong N, Kim J, Lee WK, Lee Y-G, Lee S-H, Yoon W, Kim J-H, Choi I-Y, Choi H-K, et al. Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. *DNA Res.* 2014;21(2):153–167.
- Cotrozzi L, Peron R, Tuinstra MR, Mickelbart MV, Couture JJ. Spectral phenotyping of physiological and anatomical leaf traits related with water status. *Plant Physiol.* 2020;184(3):1363–1377.
- Curran PJ. Remote-sensing of foliar chemistry. *Remote Sens Environ.* 1989;30(3):271–278.
- Dalchau N, Baek SJ, Briggs HM, Robertson FC, Dodd AN, Gardner MJ, Stancombe MA, Haydon MJ, Stan G-B, Gonçalves JM, et al. The circadian oscillator gene *GIGANTEA* mediates a long-term response of the *Arabidopsis thaliana* circadian clock to sucrose. *Proc Natl Acad Sci U S A.* 2011;108(12):5104–5109.
- Das S, Bansal M. Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. *PLoS One.* 2019;14(3):e0212678.
- de Vlaming R, Slob EAW, Jansen PR, Dagher A, Koellinger PD, Groenen PJ. Multivariate analysis reveals shared genetic architecture of brain morphology and human behavior. *Commun Biol.* 2021;4:1180.
- Dhanapal AP, Ray JD, Singh SK, Hoyos-Villegas V, Smith JR, Purcell LC, Fritschi FB. Genome-wide association mapping of soybean chlorophyll traits based on canopy spectral reflectance and leaf extracts. *BMC Plant Biol.* 2016;16(1).
- Diers BW, Specht J, Rainey KM, Cregan P, Song QJ, Ramasubramanian V, Graef G, Nelson R, Schapaugh W, Shannon G. Genetic architecture of soybean yield and agronomic traits. *G3* (Bethesda). 2018;8:3367–3375.
- Dodd AN, Belbin FE, Frank A, Webb AAR. Interactions between circadian clocks and photosynthesis for the temporal and spatial coordination of metabolism. *Front Plant Sci.* 2015;6.
- Dodd AN, Salathia N, Hall A, Kévei E, Tóth R, Nagy F, Hibberd JM, Millar AJ, Webb AAR. Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science.* 2005;309(5734):630–633.
- Driever SM, Lawson T, Andralojc PJ, Raines CA, Parry MAJ. Natural variation in photosynthetic capacity, growth, and yield in 64 field-grown wheat genotypes. *J Exp Bot.* 2014;65(17):4959–4973.
- Falconer DS, Mackay TFC. *Introduction to Quantitative Genetics*, 4th edn. Harlow, Essex (UK): Longmans Green; 1996.
- Faralli M, Lawson T. Natural genetic variation in photosynthesis: an untapped resource to increase crop yield potential? *Plant J.* 2020; 101(3):518–528.
- Farquhar GD, Caemmerer SV, Berry JA. A biochemical-model of photosynthetic CO₂ assimilation in leaves of C3 species. *Planta.* 1980; 149(1):78–90.
- Farquhar GD, Sharkey TD. Stomatal conductance and photosynthesis. *Annu Rev Plant Physiol.* 1982;33(1):317–345.
- Furbank RT, Tester M. Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 2011;16(12):635–644.
- Gamon JA, Penuelas J, Field CB. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens Environ.* 1992;41(1):35–44.
- Gilbert ME, Holbrook NM, Zwieniecki MA, Sadok W, Sinclair TR. Field confirmation of genetic variation in soybean transpiration response to vapor pressure deficit and photosynthetic compensation. *Field Crops Res.* 2011;124(1):85–92.
- Gonen S, Wimmer V, Gaynor RC, Byrne E, Gorjanc G, Hickey JM. Phasing and imputation of single nucleotide polymorphism data of missing parents of biparental plant populations. *Crop Sci.* 2021;61(4):2243–2253.

- Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* 2010;38(Database Issue):D843–D846.
- Heckmann D, Schluter U, Weber APM. Machine learning techniques for predicting crop photosynthetic capacity from leaf reflectance spectra. *Mol Plant.* 2017;10(6):878–890.
- Herritt M, Dhanapal AP, Fritsch FB. Identification of genomic loci associated with the photochemical reflectance index by genome-wide association study in soybean. *Plant Genome.* 2016;9.
- Holland JB, Nyquist WE, Cervantes-Martinez CT. Estimating and interpreting heritability for plant breeding: an update. *Plant Breed Rev.* 2010;22:9–122.
- Hunter MC, Smith RG, Schipanski ME, Atwood LW, Mortensen DA. Agriculture in 2050: recalibrating targets for sustainable intensification. *Bioscience.* 2017;67(4):386–390.
- Jones MGK, Outlaw WH, Lowry OH. Enzymatic assay of 10-7 to 10-14 moles of sucrose in plant-tissues. *Plant Physiol.* 1977;60(3):379–383.
- Kaler AS, Purcell LC. Estimation of a significance threshold for genome-wide association studies. *BMC Genomics.* 2019;20(1):618. doi:10.1186/s12864-019-5992-7.
- Kerstetter RA, Poethig RS. The specification of leaf identity during shoot development. *Annu Rev Cell Dev Biol.* 1998;14:373–398.
- Koester RP, Nohl BM, Diers BW, Ainsworth EA. Has photosynthetic capacity increased with 80 years of soybean breeding? An examination of historical soybean cultivars. *Plant Cell Environ.* 2016;39(5):1058–1067.
- Köhler IH, Ruiz-Vera UM, VanLoocke A, Thomey ML, Clemente T, Long SP, Ort DR, Bernacchi CJ. Expression of cyanobacterial FBP/SBPase in soybean prevents yield depression under future climate conditions. *J Exp Bot.* 2017;68(3):715–726.
- Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods.* 2013;9(1):29.
- Krahmer J, Goraloglia GS, Kubota A, Zardilis A, Johnson RS, Song YH, MacCoss MJ, Le Bihan T, Halliday KJ, Imaizumi T, et al. Time-resolved interaction proteomics of the GIGANTEA protein under diurnal cycles in *Arabidopsis*. *FEBS Lett.* 2019;593(3):319–338.
- Kumagai E, Burroughs CH, Pederson TL, Montes CM, Peng B, Kimm H, Guan K, Ainsworth EA, Bernacchi CJ. Predicting biochemical acclimation of leaf photosynthesis in soybean under in-field canopy warming using hyperspectral reflectance. *Plant Cell Environ.* 2022;45(1):80–94.
- Lam H-M, Xu X, Liu X, Chen W, Yang G, Wong F-L, Li M-W, He W, Qin N, Wang B, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet.* 2010;42(12):1053–1059.
- Langewisch T, Zhang H, Vincent R, Joshi T, Xu D, Bilyeu K. Major soybean maturity gene haplotypes revealed by SNPviz analysis of 72 sequenced soybean genomes. *PLoS One.* 2014;9(4):e94150.
- Lawson T, Kramer DM, Raines CA. Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. *Curr Opin Biotechnol.* 2012;23(2):215–220.
- Li BB, Morris J, Martin EB. Model selection for partial least squares regression. *Chemom Intell Lab Syst.* 2002;64(1):79–89.
- Li HH, Bradbury P, Ersoz E, Buckler ES, Wang JK. Joint QTL linkage mapping for multiple-cross mating design sharing one common parent. *PLoS One.* 2011;6.
- Li HY, Yang YM, Zhang HY, Chu SS, Zhang XG, Yin DM, Yu DY, Zhang D. A genetic relationship between phosphorus efficiency and photosynthetic traits in soybean as revealed by QTL analysis using a high-density genetic map. *Front Plant Sci.* 2016;7.
- Li Y-H, Li W, Zhang C, Yang L, Chang R-Z, Gaut BS, Qiu L-J. Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytologist.* 2010;188(1):242–253.
- Li Y-H, Zhao S-C, Ma J-X, Li D, Yan L, Li J, Qi X-T, Guo X-S, Zhang L, He W-M, et al. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics.* 2013;14(1).
- Lloyd J, Bloomfield K, Domingues TF, Farquhar GD. Photosynthetically relevant foliar traits correlating better on a mass vs an area basis: of ecophysiological relevance or just a case of mathematical imperatives and statistical quicksand? *New Phytol.* 2013;199(2):311–321.
- Long SP, Marshall-Colon A, Zhu XG. Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell.* 2015;161(1):56–66.
- Long SP, Zhu XG, Naidu SL, Ort DR. Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* 2006;29(3):315–330.
- Lopez MA, Xavier A, Rainey KM. Phenotypic variation and genetic architecture for photosynthesis and water use efficiency in soybean (*Glycine max* L. Merr). *Front Plant Sci.* 2019;10.
- Lü H, Yang Y, Li H, Liu Q, Zhang J, Yin J, Chu S, Zhang X, Yu K, Lv L, et al. Genome-wide association studies of photosynthetic traits related to phosphorus efficiency in soybean. *Front Plant Sci.* 2018;9.
- Meacham-Hensold K, Montes CM, Wu J, Guan KY, Fu P, Ainsworth EA, Pederson T, Moore CE, Brown KL, Raines C, et al. Plot-level rapid screening for photosynthetic parameters using proximal hyperspectral imaging. *J Exp Bot.* 2020;71(7):2312–2328.
- Meacham-Hensold K, Montes CM, Wu J, Guan KY, Fu P, et al. High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity. *Remote Sens Environ.* 2019;231.
- Meireles JE, Cavender-Bares J, Townsend PA, Ustin S, Gamon JA, Schweiger AK, Schaepman ME, Asner GP, Martin RE, Singh A, et al. Leaf reflectance spectra capture the evolutionary history of seed plants. *New Phytol.* 2020;228(2):485–493.
- Mikel MA, Dudley JW. Evolution of North American dent corn from public to proprietary germplasm. *Crop Sci.* 2006;46(3):1193–1205.
- Mishra P, Panigrahi KC. GIGANTEA—an emerging story. *Front Plant Sci.* 2015;6.
- Mora-Garcia S, de Leone MJ, Yanovsky M. Time to grow: circadian regulation of growth and metabolism in photosynthetic organisms. *Curr Opin Plant Biol.* 2017;35:84–90.
- Morrell PL, Buckler ES, Ross-Ibarra J. Crop genomics: advances and applications. *Nat Rev Genet.* 2012;13(2):85–96.
- Mourtzinis S, Conley SP. Delineating soybean maturity groups across the United States. *Agronomy J.* 2017;109(4):1397–1403.
- Mugford ST, Fernandez O, Brinton J, Flis A, Krohn N, Encke B, Feil R, Sulpice R, Lunn JE, Stitt M, et al. Regulatory properties of ADP glucose pyrophosphorylase are required for adjustment of leaf starch synthesis in different photoperiods. *Plant Physiol.* 2014;166(4):1733–1747.
- Neyhart JL, Lorenz AJ, Smith KP. Multi-trait improvement by predicting genetic correlations in breeding crosses. *G3 (Bethesda).* 2019;9:3153–3165.
- Nice LM, Steffenson BJ, Brown-Guedira GL, Akhunov ED, Liu C, Kono TJY, Morrell PL, Blake TK, Horsley RD, Smith KP, et al. Development and genetic characterization of an advanced backcross-nested association mapping (AB-NAM) population of wild x cultivated barley. *Genetics.* 2016;203(3):1453–1467.

- Niinimets Ü, Berry JA, Caemmerer S, Ort DR, Parry MAJ, Poorter H. Photosynthesis: ancient, essential, complex, diverse... and in need of improvement in a changing world. *New Phytol.* 2017; 213(1):43–47.
- Nikkanen L, Toivola J, Trotta A, Diaz MG, Tikkanen M, Aro Eva-Mari, Rintamäki E. Regulation of cyclic electron flow by chloroplast NADPH-dependent thioredoxin system. *Plant Direct.* 2018;2.
- Ort DR, Merchant SS, Alric J, Barkan A, Blankenship RE, Bock R, Croce R, Hanson MR, Hibberd JM, Long SP, et al. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc Natl Acad Sci U S A.* 2015;112(28):8529–8536.
- Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. *Plant Physiol.* 2016;171(4):2294–2316.
- Parry MAJ, Reynolds M, Salvucci ME, Raines C, Andralojc PJ, Zhu X-G, Price GD, Condon AG, Furbank RT. Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. *J Exp Bot.* 2011;62(2):453–467.
- Patterson HD, Thompson R. Recovery of inter-block information when block sizes are unequal. *Biometrika.* 1971;58(3):545–554.
- Penuelas J, Filella I. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends Plant Sci.* 1998; 3(4):151–156.
- Penuelas J, Gamon JA, Fredeen AL, Merino J, Field CB. Reflectance indexes associated with the physiological-changes in nitrogen-limited and water-limited sunflower leaves. *Remote Sens Environ.* 1994;48(2):135–146.
- Qiu L-J, Xing L-L, Guo Y, Wang J, Jackson SA, Chang R-Z. A platform for soybean molecular breeding: the utilization of core collections for food security. *Plant Mol Biol.* 2013;83(1–2):41–50.
- R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Raines CA. Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: current and future strategies. *Plant Physiol.* 2011;155(1):36–42.
- Ray DK, Mueller ND, West PC, Foley JA. Yield trends are insufficient to double global crop production by 2050. *PLoS One.* 2013;8.
- Ray DK, Ramankutty N, Mueller ND, West PC, Foley JA. Recent patterns of crop yield growth and stagnation. *Nat Commun.* 2012;3.
- Riboni M, Galbiati M, Tonelli C, Conti L. GIGANTEA enables drought escape response via abscisic acid-dependent activation of the florigens and SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1. *Plant Physiol.* 2013;162(3):1706–1719.
- Richardson AD, Duigan SP, Berlyn GP. An evaluation of noninvasive methods to estimate foliar chlorophyll content. *New Phytologist.* 2002;153(1):185–194.
- Rosloski SM, Jali SS, Balasubramanian S, Weigel D, Grbic V. Natural diversity in flowering responses of *Arabidopsis thaliana* caused by variation in a tandem gene array. *Genetics.* 2010;186(1): 263–U435.
- Sakoda K, Tanaka Y, Long SP, Shiraiwa T. Genetic and physiological diversity in the leaf photosynthetic capacity of soybean. *Crop Sci.* 2016;56(5):2731–2741.
- Sandak J, Sandak A, Meder R. Assessing trees, wood and derived products with near infrared spectroscopy: hints and tips. *J Near Infrared Spectrosc.* 2016;24(6):485–505.
- Sanz-Sáez Á, Koester RP, Rosenthal DM, Montes CM, Ort DR, Ainsworth EA. Leaf and canopy scale drivers of genotypic variation in soybean response to elevated carbon dioxide concentration. *Glob Chang Biol.* 2017;23(9):3908–3920.
- Schweiger AK. Spectral field campaigns: planning and data collection, pp. 385–423. In: JG Cavender-Bares, J Gamon, P Townsend, editors. *Remote Sensing of Plant Biodiversity.* Switzerland: Springer Nature; 2020.
- Scott K, Balk C, Veney D, McHale LK, Dorrance AE. Quantitative disease resistance loci towards *Phytophthora sojae* and three species of *Pythium* in six soybean nested association mapping populations. *Crop Sci.* 2019;59(2):605–623.
- Sedivy EJ, Wu FQ, Hanzawa Y. Soybean domestication: the origin, genetic architecture and molecular bases. *New Phytol.* 2017;214(2): 539–553.
- Serbin S, Townsend P. Scaling functional traits from leaves to canopies. In: JG Cavender-Bares, J Gamon, P Townsend, editors. *Remote Sensing of Plant Biodiversity.* Switzerland: Springer Nature; 2020. p. 43–82.
- Serbin SP, Dillaway DN, Kruger EL, Townsend PA. Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature. *J Exp Bot.* 2012;63(1):489–502.
- Serbin SP, Singh A, Desai AR, Dubois SG, Jablonski AD, Kingdon CC, Kruger EL, Townsend PA. Remotely estimating photosynthetic capacity, and its response to temperature, in vegetation canopies using imaging spectroscopy. *Remote Sens Environ.* 2015;167: 78–87.
- Serbin SP, Singh A, McNeil BE, Kingdon CC, Townsend PA. Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. *Ecol Appl.* 2014;24(7):1651–1669.
- Sharma R, Draicchio F, Bull H, Herzig P, Maurer A, Pillen K, Thomas WTB, Flavell AJ. Genome-wide association of yield traits in a nested association mapping population of barley reveals new gene diversity for future breeding. *J Exp Bot.* 2018;69(16): 3811–3822.
- Silva-Perez V, Molero G, Serbin SP, Condon AG, Reynolds MP, Furbank RT, Evans JR. Hyperspectral reflectance as a tool to measure biochemical and physiological traits in wheat. *J Exp Bot.* 2018;69(3):483–496.
- Sims DA, Gamon JA. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens Environ.* 2002; 81(2–3):337–354.
- Slaton MR, Hunt ER, Smith WK. Estimating near-infrared leaf reflectance from leaf structural characteristics. *Am J Bot.* 2001;88(2): 278–284.
- Soleh MA, Tanaka Y, Kim SY, Huber SC, Sakoda K, Shiraiwa T. Identification of large variation in the photosynthetic induction response among 37 soybean *Glycine max* (L.) Merr. genotypes that is not correlated with steady-state photosynthetic capacity. *Photosynth Res.* 2017;131(3):305–315.
- Song Q, Jenkins J, Jia G, Hyten DL, Pantalone V, Jackson SA, Schmutz J, Cregan PB. Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. *BMC Genomics.* 2016;17(1).
- Song Q, Yan L, Quigley C, Jordan BD, Fickus E, Schroeder S, Song B-H, Charles An Y-Q, Hyten D, Nelson R, et al. Genetic characterization of the soybean nested association mapping population. *Plant Genome.* 2017;10(2).
- Spreitzer RJ, Salvucci ME. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu Rev Plant Biol.* 2002; 53:449–475.
- Stich B. Comparison of mating designs for establishing nested association mapping populations in maize and *Arabidopsis thaliana*. *Genetics.* 2009;183(4):1525–1534.
- Tamoi M, Nagaoka M, Miyagawa Y, Shigeoka S. Contribution of fructose-1,6-bisphosphatase and sedoheptulose-1,7-bisphosphatase to the

- photosynthetic rate and carbon flow in the Calvin cycle in transgenic plants. *Plant Cell Physiol.* 2006;47(3):380–390.
- Technow F, Gerke J. Parent-progeny imputation from pooled samples for cost-efficient genotyping in plant breeding. *PLoS One.* 2017;12(12):e0190271.
- Tester M, Langridge P. Breeding technologies to increase crop production in a changing world. *Science.* 2010;327(5967):818–822.
- Tobias RD. An Introduction to Partial Least Squares Regression; 1995. <https://stats.oarc.ucla.edu/wp-content/uploads/2016/02/pls.pdf>.
- Tomeo NJ, Rosenthal DM. Variable mesophyll conductance among soybean cultivars sets a tradeoff between photosynthesis and water-use-efficiency. *Plant Physiol.* 2017;174(1):241–257.
- van Bezouw R, Keurentjes JJB, Harbinson J, Aarts MGM. Converging phenomics and genomics to study natural variation in plant photosynthetic efficiency. *Plant J.* 2019;97(1):112–133.
- van Heerwaarden J, Hufford MB, Ross-Ibarra J. Historical genomics of North American maize. *Proc Natl Acad Sci U S A.* 2012;109(31):12420–12425.
- Vilhjalmsson BJ, Nordborg M. The nature of confounding in genome-wide association studies. *Nat Rev Genet.* 2013;14(1):1.
- von Caemmerer S, Evans JR. Enhancing C3 photosynthesis. *Plant Physiol.* 2010;154(2):589–592.
- Walker BJ, Drewry DT, Slattery RA, VanLoocke A, Cho YB, Ort DR. Chlorophyll can be reduced in crop canopies with little penalty to photosynthesis. *Plant Physiol.* 2018;176(2):1215–1232.
- Walter-Shea EA, Norman JM. Leaf optical properties. In: RE Myneni, editor. *Photon-Vegetation Interaction*. Berlin: Springer-Verlag; 1991. p. 229–251.
- Wang L, Yang Y, Zhang S, Che Z, Yuan W, Yu D. GWAS reveals two novel loci for photosynthesis-related traits in soybean. *Mol Genet Genomics.* 2020a;295(3):705–716.
- Wang Y, Burgess SJ, Becker EM, Long SP. Photosynthesis in the fleeting shadows: an overlooked opportunity for increasing crop productivity? *Plant J.* 2020b;101(4):874–884.
- Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Takahashi R, Anai T, Tabata S, Kitamura K, et al. A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. *Genetics.* 2011;188(2):395–U260.
- Wei JL, Xu SZ. A random-model approach to QTL mapping in multiparent advanced generation intercross (MAGIC) populations. *Genetics.* 2016;202(2):471–486.
- Wold S, Ruhe A, Wold H, Dunn WJ. The collinearity problem in linear-regression – the partial least-squares (PLS) approach to generalized inverses. *Siam J Sci and Stat Comput.* 1984;5(3):735–743.
- Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, Cavender-Bares J, Chapin T, Cornelissen JHC, Diemer M, et al. The worldwide leaf economics spectrum. *Nature.* 2004;428(6985):821–827.
- Wu A, Hammer GL, Doherty A, von Caemmerer S, Farquhar GD. Quantifying impacts of enhancing photosynthesis on crop yield. *Nat Plants.* 2019;5(4):380–388.
- Xavier A, Hall B, Hearst AA, Cherkauer KA, Rainey KM. Genetic architecture of phenomic-enabled canopy coverage in *Glycine max*. *Genetics.* 2017;206(2):1081–1089.
- Xavier A, Jarquin D, Howard R, Ramasubramanian V, Specht JE, Graef GL, Beavis WD, Diers BW, Song Q, Cregan PB, et al. Genome-wide analysis of grain yield stability and environmental interactions in a multiparental soybean population. *G3 (Bethesda).* 2018;8(2):519–529.
- Xavier A, Muir WM, Rainey KM. Assessing predictive properties of genome-wide selection in soybeans. *G3 (Bethesda).* 2016;6(8):2611–2616.
- Xavier A, Xu SZ, Muir WM, Rainey KM. NAM: association studies in multiple populations. *Bioinformatics.* 2015;31(23):3862–3864.
- Xu SH, Atchley WR. A random model approach to interval mapping of quantitative trait loci. *Genetics.* 1995;141(3):1189–1197.
- Yamori W, Shikanai T. Physiological functions of cyclic electron transport around photosystem I in sustaining photosynthesis and plant growth. *Annu Rev Plant Biol.* 2016;67:81–106.
- Yang JA, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76–82.
- Yendrek CR, Tomaz T, Montes CM, Cao Y, Morse AM, Brown PJ, McIntyre LM, Leahey ADB, Ainsworth EA. High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance. *Plant Physiol.* 2017;173(1):614–626.
- Yu JM, Holland JB, McMullen MD, Buckler ES. Genetic design and statistical power of nested association mapping in maize. *Genetics.* 2008;178(1):539–551.
- Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet.* 2006;38(2):203–208.
- Zhu XG, Long SP, Ort DR. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr Opin Biotechnol.* 2008;19(2):153–159.
- Zhu XG, Long SP, Ort DR. Improving photosynthetic efficiency for greater yield. *Annu Rev Plant Biol.* 2010;61:235–261.

Communicating editor: T. Juenger