Journal of Experimental Botany, Vol. 70, No. 6 pp. 1789–1799, 2019
doi:10.1093/jxb/erz061 Advance Access Publication 19 February 2019

RESEARCH PAPER

Spectroscopy can predict key leaf traits associated with source–sink balance and carbon–nitrogen status

Kim S. Ely, Angela C. Burnett, Wil Lieberman-Cribbin, Shawn P. Serbin and Alistair Rogers*

Environmental & Climate Sciences Department, Brookhaven National Laboratory, Upton, NY 11973-5000 USA

* Correspondence: arogers@bnl.gov

Received 10 October 2018; Editorial decision 5 February 2019; Accepted 5 February 2019

Editor: Susanne von Caemmerer, Australian National University, Australia

Abstract

Approaches that enable high-throughput, non-destructive measurement of plant traits are essential for programs seeking to improve crop yields through physiological breeding. However, many key traits still require measurement using slow, labor-intensive, and destructive approaches. We investigated the potential to retrieve key traits associated with leaf source–sink balance and carbon–nitrogen status from leaf optical properties. Structural and biochemical traits and leaf reflectance (500–2400 nm) of eight crop species were measured and used to develop predictive 'spectra–trait' models using partial least squares regression. Independent validation data demonstrated that the models achieved very high predictive power for C, N, C:N ratio, leaf mass per area, water content, and protein content (R^2>0.85), good predictive capability for starch, sucrose, glucose, and free amino acids (R^2=0.58–0.80), and some predictive capability for nitrate (R^2=0.51) and fructose (R^2=0.44). Our spectra–trait models were developed to cover the trait space associated with food or biofuel crop plants and can therefore be applied in a broad range of phenotyping studies.

Keywords: Amino acids, carbohydrates, carbon, leaf traits, metabolites, nitrogen, PLSR, remote sensing, source–sink, spectroscopy.

Introduction

To ensure future food security, a substantial increase in crop production is required (FAO *et al.*, 2017), and several steps are needed in order to achieve greater increases in yield (Reynolds *et al.*, 2012; Sreenivasulu and Schnurbusch, 2012; von Caemmerer *et al.*, 2012; Ort *et al.*, 2015). Gaining a better understanding of the factors limiting plant growth and yield is an essential milestone on the road to increased crop productivity (White *et al.*, 2016). Once plant traits are well understood in the context of yield, breeding programs can select for appropriate traits associated with greater yields, or those that confer optimal trade-offs between resilience to stress and yield. Physiological breeding employs this approach and relies on the identification of key physiological traits associated with increased yield to select higher yielding progeny, when parents with complementary traits are crossed (Reynolds and Langridge, 2016). In contrast to traditional breeding, which relies upon crossing elite crop varieties and sometimes introducing wild relatives or diverse landraces, physiological breeding incorporates phenomic and genomic screening of progeny resulting from well-characterized genetic resources, to develop physiologically characterized lines suited to the production environment. High-throughput trait measurement is necessary

Abbreviations: LMA, leaf mass per area (g leaf mass m^{-2} leaf area); NIR, near infrared; PLSR, partial least squares regression; RMSE, root mean square error; SWIR, shortwave infrared; TNCs, total non-structural carbohydrates (glucose+fructose+sucrose+starch); VIP, variable importance of projection.

to achieve rapid advances in yield via physiological breeding, since traditional methods of analyzing physiological traits are slow, labor-intensive, and often destructive. High-throughput approaches are therefore a critical element of ensuring the success and increasing the efficiency and effectiveness of physiological breeding programs so that breeders can achieve rapid gains in yield that will be required in the near future (FAO *et al.*, 2017).

High-throughput plant phenotyping takes many forms, and may be employed at a great range of scales from the molecular to the ecosystem level. At one end of the spectrum, high-throughput robotic assays may be used for fast estimates of the levels of key biochemical markers (Gibon *et al.*, 2004; Dai *et al.*, 2015; Burnett *et al.*, 2016; Emmett *et al.*, 2017). At a larger scale, whole plants may be measured on glasshouse conveyor belt systems (Al-Tamimi *et al.*, 2016; Ferdous *et al.*, 2017). At the canopy to landscape scale, passive monitoring, ground vehicle systems, unmanned aerial systems (UASs), or aircraft can be employed to retrieve spectral information, phenotypic plasticity, and plant traits (Madritch *et al.*, 2014; Singh *et al.*, 2015; Yang *et al.*, 2017), whilst at the largest scale, satellite imagery is a powerful tool for examining ecosystem-level phenotypes (Pettorelli *et al.*, 2018). High-throughput phenotyping using remote sensing is especially powerful since it is non-destructive, and enables measurements to be repeated on the same plants throughout a growing season, providing valuable insight into the response of key plant traits to environmental and developmental change.

Remote sensing has been used to predict a variety of plant traits, including the content of chlorophyll and other photosynthetic pigments (Asner and Martin, 2008; Yendrek *et al.*, 2017), nitrogen (Asner and Martin, 2008; Ainsworth *et al.*, 2014; Serbin et al., 2014; Dechant *et al.*, 2017; Yendrek *et al.*, 2017), phosphorus (Asner and Martin, 2008), sucrose (Yendrek *et al.*, 2017), fiber (Petisco *et al.*, 2006; Serbin et al., 2014), lignin (Petisco *et al.*, 2006; Serbin *et al.*, 2014), cellulose (Petisco *et al.*, 2006; Serbin *et al.*, 2014), water (Asner and Martin, 2008; Colombo *et al.*, 2008), and plant secondary metabolites such as phenolic compounds (Couture *et al.*, 2016). More recently, remote sensing has been used to measure physiological traits; for example, the maximum carboxylation rate in C_3 (Serbin *et al.*, 2012; Ainsworth *et al.*, 2014) and C_4 plants (Yendrek *et al.*, 2017), and the maximum rate of electron transport (Serbin *et al.*, 2012; Barnes *et al.*, 2017; Dechant *et al.*, 2017). Leaf structural traits such as leaf mass per unit area (LMA) (Asner *et al.*, 2011; Serbin *et al.*, 2014; Yendrek *et al.*, 2017) can also be readily obtained using remote sensing approaches.

It is particularly important to develop relationships between spectral data and traits associated with plant source–sink balance and carbon (C) and nitrogen (N) status. Source tissues have a net uptake of resources from the environment, providing the elements required for growth, while sink tissues have a net drawdown of these resources and are essential for plants to grow (Reekie *et al.*, 1998; Aranjuelo *et al.*, 2013; White *et al.*, 2016). For example, for C, a mature photosynthetically active leaf is a net source, whereas developing fruits or grains are a net sink. The balance between source and sink is a critical component of growth and therefore yield (Arp, 1991; Ainsworth *et al.*, 2003; Eyles *et al.*, 2013), while the balance between C and N is essential for tuning growth to environmental conditions and maximizing resource acquisition (Krapp *et al.*, 1991; Burnett *et al.*, 2016; White *et al.*, 2016). Key biochemical traits of interest for determining the C source–sink status include carbohydrate pools such as starch and sucrose, and the ratio of amino acids to sucrose which denotes the availability of excess N that in turn indicates C source limitation (Paul and Driscoll, 1997; Stitt and Krapp, 1999; Isopp *et al.*, 2000). For examining the N source–sink status, it is important to examine pools of nitrate, free amino acids, and protein.

This study aims to build relationships between leaf reflectance and key metabolites and traits related to plant source–sink balance and C–N status, by using reflectance spectroscopy to develop models validated by traditional, gold standard, laboratory measurements. The goals of this work are 2-fold: first, to define what is technically possible with respect to remote sensing of plant biochemistry associated with C and N metabolite pools; and, secondly, to develop relationships between leaf optical properties and biochemical traits that can enable and accelerate physiological breeding programs by facilitating rapid, inexpensive, non-destructive *in situ* plant measurements.

We sought to develop models that can be broadly applied across different vegetation and cover a large portion of the total trait space. In order to generate a suitable range of the traits of interest, we grew eight contrasting species in a range of conditions. In addition, we sampled at different times of the day to leverage expected diurnal cycles in key traits, and measured and sampled foliage at different stages of plant and leaf development to extend the range of measured traits further. We used partial least squares regression (PLSR) to build spectra–trait relationships; this is a well-established and robust technique for developing predictive models (Carrascal *et al.*, 2009; Asner and Martin, 2015; Silva-Perez *et al.*, 2018).

Materials and methods

Plant materials and growing conditions

Eight crop species were grown in a glasshouse at Brookhaven National Laboratory, with 16–21 individuals from each species available to sample. Species were selected to cover a broad range of leaf optical properties and an anticipated wide range in our traits of interest. Seven plant species, *Solanum lycopersicum* var. *lycopersicum*, *Cucumis sativus*, *Cucurbita pepo*, *Glycine max*, *Phaseolus vulgaris*, *Ocimum basilicum*, and *Helianthus annuus* (High Mowing Organic Seeds, VT, USA) were germinated from seed in trays and transplanted to pots 28 d after germination (see Supplementary Table S1 at *JXB* online). Poplar cuttings, *Populus deltoides* Bartr.×*Populus nigra* L. (Segal Ranch Hybrid Poplars, Grandview, WA, USA), were initially planted in 2 liter pots and transplanted to final pots after 40 d. Plants were grown in Pro-mix BX or Pro-mix BX mycorrhizae growing media (Premier Tech, Rivière-du-Loup, Québec, Canada) to provide a range of nutrient conditions to enhance the variation in N-containing metabolites. Greenhouse temperature and light conditions were controlled to a diurnal temperature range of 20–28 °C, and a 16 h day length. Plants were watered daily, or as required to prevent drying, and fertilized at 21 d intervals with Miracle-Gro soluble all-purpose plant food according to the manufacturer's directions (N 24:P 8:K 16, Scotts Miracle-Gro, Marysville, OH, USA).

Spectroscopy of fresh leaf material

Commencing 8 weeks after seed planting, plants were sampled during the hours 09.15–16.15 on 16 d over an 8 week period for a total of 187 individual leaves. Measurements were made on fully expanded leaves, and the number of leaves sampled from each species is recorded in Supplementary Table S1. Leaf spectra were measured using a Spectra Vista Corporation (SVC) HR1024i full range (350–2500 nm, spectral resolution of ≤3.5 nm, 700 nm; ≤9.5 nm, 1500 nm; ≤6.5 nm, 2100 nm; linearly interpolated to 1 nm) spectroradiometer, together with a leaf clip assembly with an internal, calibrated light source (SVC, Poughkeepsie, NY, USA). The spectroradiometer was calibrated using a LabSphere Spectralon® reflectance standard (LabSphere, Inc., North Sutton, NH, USA). We measured the spectra of each leaf five times on different areas of the adaxial surface of the leaf. Spectral discontinuities in the detector overlap areas were corrected using the SVC instrument software prior to sample averaging and other quality control steps, as described previously (Serbin *et al.*, 2014).

Determination of leaf traits

Following spectral measurements, each measured leaf was immediately removed from the plant and sampled using a circular punch of known area. A sample of leaf material was sealed in a foil packet and immediately flash-frozen in liquid nitrogen before transfer to a –80 °C freezer prior to biochemical analysis. The fresh mass of additional material from each leaf was determined using a precision balance. This additional material was dried at 70 °C for 3 d before weighing again to calculate LMA (g m^{-2}) and leaf water content (H_2O g m^{-2}).

Dried and frozen leaf samples were ground using a Spex CertiPrep 2000 Geno/Grinder (Spex SamplePrep, Metuchen, NJ, USA). The C and N contents of dried, ground samples were measured using a Perkin Elmer CHNO/S Series II 2400 elemental analyzer on two 1.50–2.50 mg replicates of each sample following the manufacturer's instructions. Ethanol-soluble metabolites were extracted from ~26 mg of frozen, ground leaf material. All subsequent biochemical analysis was conducted using 96-well microplates (Microtest Plate 96-well flat bottom, Sarstedt Inc., Newton, NC, USA), and a robotic liquid handling system (Evolution P3 Precision Pipetting Platform, Perkin Elmer, Waltham, MA, USA). A Biotek Synergy HT Multi-Detection Microplate Reader was used to determine amino acids, and a Biotek ELx808 Absorbance Microplate Reader (BioTek, Winooski, VT, USA) was used for all other determinations (Gibon *et al.*, 2004; Burnett *et al.*, 2016).

Simple carbohydrates (glucose, fructose, and sucrose) in the ethanol extract were measured using a continuous enzymatic substrate assay (Ainsworth *et al.*, 2007; Burnett *et al.*, 2016). Nitrate was quantified using the Griess reaction following the procedure described by Burnett *et al.* (2016). Total free amino acids were determined by the addition of fluorescamine and 0.1 M sodium borate buffer to the ethanol extract (Burnett *et al.*, 2016).

Proteins in the pellet resulting from the ethanol extraction were solubilized by adding 0.1 M NaOH and heating to 95 °C. The Pierce BCA protein assay kit (ThermoScientific, Rockford, IL, USA) was used to measure protein content against a BSA standard curve. Samples were then neutralized with HCl prior to starch determination. Starch was digested to glucose with an overnight incubation at 37 °C, and the resulting glucose was measured using a continuous enzymatic substrate assay that included a corn starch control to confirm digestion (Rogers *et al.*, 2004).

PLSR model fitting

Leaf traits were predicted on an area basis from spectra using a PLSR modeling approach with the 'pls' package (Mevik *et al.*, 2016) in the R open source software environment (R Core Team, 2018). We conducted this 'spectra–trait' modeling following the methods developed by Serbin *et al.* (2014). Spectral wavelengths of 500–2400 nm were used for all trait predictions. Analytical data for traits with highly skewed distributions (i.e. sugars and nitrate) were natural log transformed to achieve a more normal distribution prior to building the PLSR model. A small number of samples were rejected from the data set on the basis of analytical error or results with outlier residual errors; the final number of spectra–trait pairs used for each trait model was n=174–179. Observational data points were randomly assigned to a calibration data set comprising 80% of the data, or a validation data set (the remaining 20%), maintaining an even distribution from each species in each subset (see Table 1 for details of the sample number for each trait). The number of components used for each trait model was based upon the minimum required to minimize the PRESS statistic (Table 1). The model calibration performance was characterized using a 1000× permutation test on the calibration data set (refer to Serbin *et al.*, 2014 for a detailed explanation of this process). Models were created for each measured trait, and also the derived traits of C:N, total sugars, and total non-structural carbohydrates (TNCs, i.e. starch and

Table 1. *PLSR model input and results for each leaf trait*

Trait	Data treatment	Model comps	*n*		R^2		RMSE		%RMSE		Reg. bias	Res. bias
			Cal.	Val.	Cal.	Val.	Cal.	Val.	Cal.	Val.		
N (g m^{-2})	None	13	140	38	0.86	0.92	0.15	0.14	6.95	5.63	–0.07	0.00
C (g m^{-2})	None	12	140	38	0.89	0.95	1.18	1.13	5.66	5.94	0.65	0.00
C:N	None, 1	12	139	38	0.83	0.92	2.37	2.07	6.74	5.25	0.35	0.03
LMA (g m^{-2})	None	12	141	38	0.91	0.90	2.92	4.08	5.38	8.65	–3.84	0.00
H_2O (g m^{-2})	None	11	141	38	0.86	0.89	10.83	11.84	6.36	8.38	–12.23	–0.02
Protein (g m^{-2})	None, 2	12	140	37	0.82	0.85	0.76	0.76	7.36	8.98	0.19	0.00
Amino acids (μmol m^{-2})	None, 1	11	137	37	0.46	0.58	1.42	1.04	12.20	13.81	0.56	0.01
Nitrate (μmol m^{-2})	Log	10	141	38	0.32	0.51	1.09	1.20	16.12	17.68	0.08	–0.02
Starch (μmol C_6 m^{-2})	None, 3	12	136	38	0.75	0.80	5.25	5.70	9.81	10.59	2.78	0.11
TNC (μmol C_6 m^{-2})	None	12	139	38	0.67	0.70	6.29	7.53	8.73	14.21	2.5	0.05
Total sugars (μmol C_6 m^{-2})	Log	14	141	38	0.59	0.69	0.32	0.39	10.72	16.20	0.28	0.00
Sucrose (μmol C_6 m^{-2})	Log, 2	12	139	38	0.63	0.76	0.38	0.39	10.56	13.52	0.18	0.00
Glucose (μmol C_6 m^{-2})	Log, 2	8	139	38	0.56	0.59	0.45	0.43	12.96	19.06	0.02	0.01
Fructose (μmol C_6 m^{-2})	Log	10	141	38	0.29	0.44	0.80	0.87	11.81	14.78	0.12	0.00

Data treatment indicates if data were log transformed, and the number of residual outlier samples removed from the data set. Model comps is the number of components used in the PLSR model, chosen to minimize the PRESS statistic. *n* Cal. and *n* Val. indicate the number of samples included in the calibration and validation data subsets respectively, and the total of these values is the number of leaf spectra–trait pairings used for each trait. R^2, RMSE, and %RMSE are shown for the calibration and validation data sets. Regression bias (Reg. bias) is the regression intercept. The residual bias (Res. bias) is the mean of the difference between predicted and observed values in the calibration dataset.

sugars combined). The performance of each model was assessed using the R^2 and root mean square error (RMSE) of prediction of the validation data set. Qualitative evaluation of model predictor variables was assessed using the variable importance of projection (VIP) (Wold *et al.*, 2001).

Results

Leaf properties

Laboratory analysis revealed the anticipated large variation of leaf structural and biochemical properties across our sample set, with leaves containing 6.6–27.5 g m^{-2} C, 0.3–2.8 g m^{-2} N, 92–263 g m^{-2} H_2O, 1.3–12.1 g m^{-2} protein, 0.1–14.6 μmol m^{-2} amino acids, 0–8.9 μmol m^{-2} nitrate, 0.1–56.7 μmol C_6 m^{-2} starch, 0.5–17.5 μmol C_6 m^{-2} sucrose, 0.4–12.5 μmol C_6 m^{-2} glucose, 0–8.4 μmol C_6 m^{-2} fructose, and an LMA of 16–70 g m^{-2} (Fig. 1). A high degree of correlation exists between the trait pairs of LMA and C (r=0.98), N and protein (r=0.91), TNCs and starch (r=0.88), and, to a lesser extent, N and amino acids (r=0.71). Other trait pairs exhibited only weak (r<0.7) or no correlation across the whole data set.

Leaf spectral characteristics

The spectral reflectance across wavelength of 500–2400 nm exhibited a pattern typical of foliar reflectance, with a peak centering at ~550 nm in the visible region, high reflectivity throughout the near infrared (NIR, 800–1300 nm), and peaks in the shortwave infrared (SWIR) region centered at 1700, 1800, and 2200 nm (Fig. 2A). The largest ranges in reflectance generally coincide with these peaks (Fig. 2B). The greatest amount of variation across the spectral measurements, as shown by the coefficient of variation by wavelength (%CV), also occurred within the visible region, with up to 40%CV.

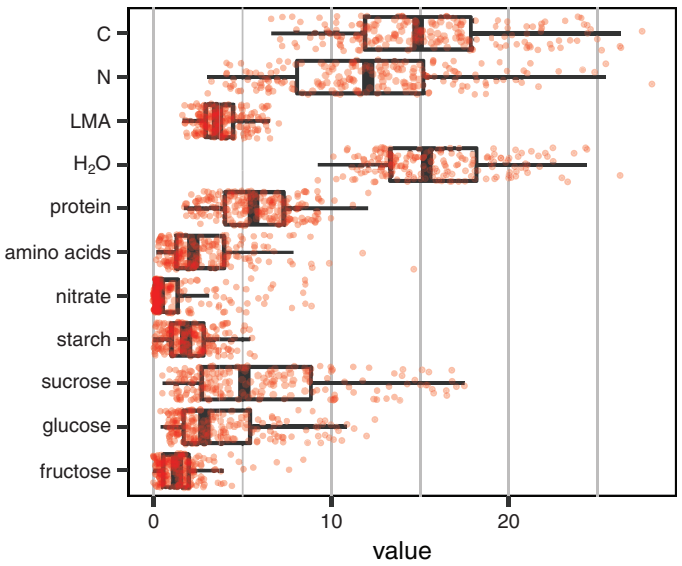

Fig. 1. Trait properties of sampled leaves from all species. Note units scaled as follows: C g m^{-2}, N g m^{-2}×10^{-1}, LMA g m^{-2}×10, H_2O g m^{-2}× 10, protein g m^{-2}, amino acids μmol m^{-2}, nitrate μmol m^{-2}, starch μmol C_6 m^{-2}×10, sucrose μmol C_6 m^{-2}, glucose μmol C_6 m^{-2}, and fructose μmol C_6 m^{-2}. Boxplots show the median, and the 25th and 75th percentiles. The maximum extent of boxplot whiskers is 1.5 times the interquartile range.

PLSR prediction model results

Regions of the reflectance spectrum that were most significant to each trait model were identified using the VIP metric. In the visible region, wavelengths centered around 550 nm and 700 nm showed high importance for all trait models with the exception of water (Fig. 3). Interestingly, while the 550 nm region is important for modeling individual sugars (Fig. 3L–N), it carried reduced weight for the combined sugars model (Fig. 3K). In the SWIR, a peak at 1400 nm is prominent for most traits. PLSR model loadings (Supplementary Fig. S1) and PLSR coefficient vectors (Supplementary Fig. S2) further illustrate the relative importance of different spectral regions to the prediction of each trait.

Utilization of a model based on mixed plant species demonstrated a very high level of predictability for the elemental and structural traits of C, N, C:N, LMA, and H_2O, all with a validation R^2 of 0.89–0.95 (Table 1; Fig. 4). Of the N-containing metabolites, the model performance was very high for predicting protein (R^2=0.85, RMSE=9%), and showed moderate predictive capability for amino acids (R^2=0.58, RMSE=14%) and nitrate (R^2=0.51, RMSE=18%; Fig. 5). For C-containing metabolites, the model performance was high for starch (R^2=0.80, RMSE=11%), sucrose (R^2=0.76, RMSE=14%), TNCs (R^2=0.70, RMSE=14%), and total sugars (R^2=0.69, RMSE=16%). The spectra–trait PLSR approach performed moderately well for glucose (R^2=0.59, RMSE=19%), and had more limited predictive capability for fructose (R^2=0.44,

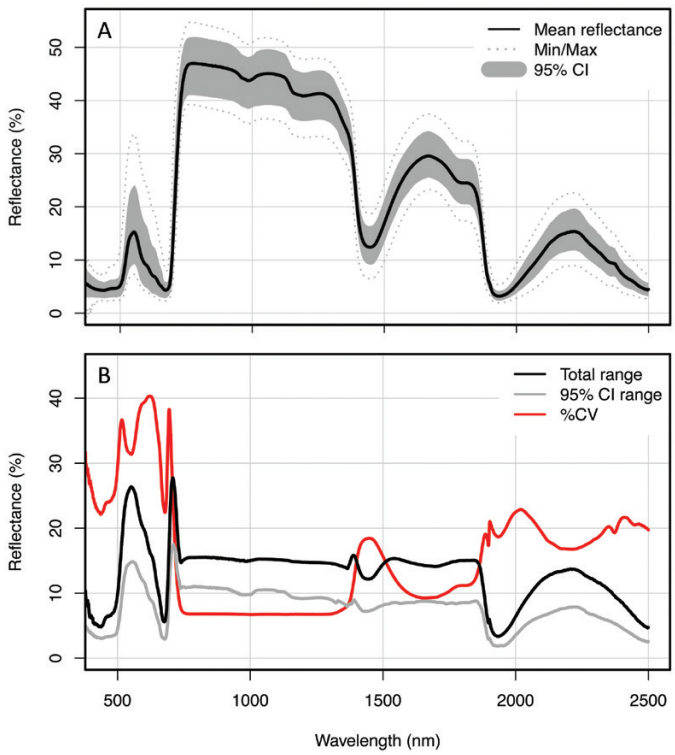

Fig. 2. Summary of leaf spectral reflectance measurements. (A) The mean, 95% confidence interval, minimum, and maximum spectral reflectance for all samples (n=179). (B) The range of reflectance measurements (black line), the range within the 95% confidence interval (gray line), and the percentage coefficient of variation (SD/mean) for each wavelength (red line).

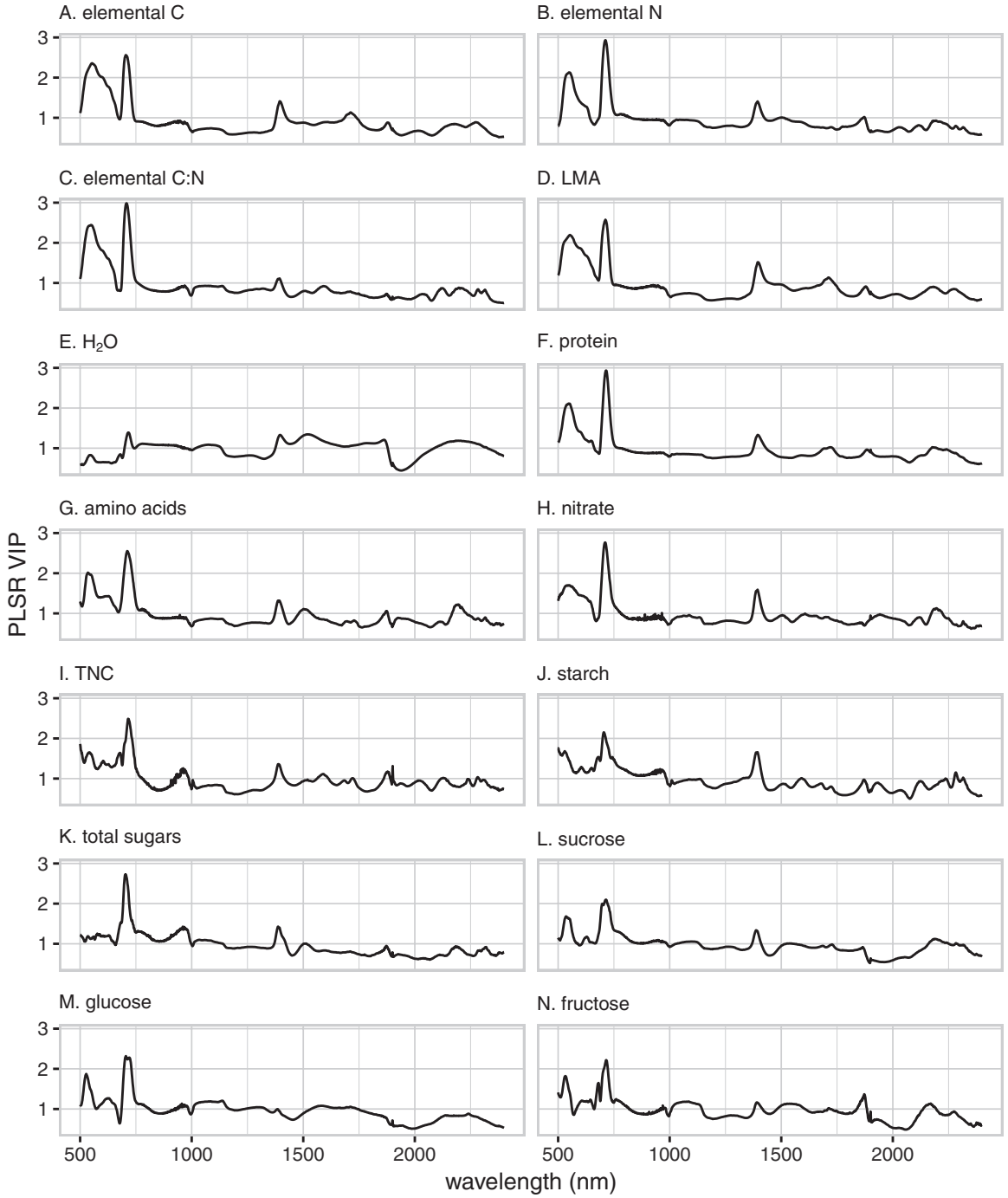

Fig. 3. PLSR model variable importance of projection (VIP) for all traits assessed in this study. (A) C g m^{-2}, (B) N g m^{-2}, (C) C:N, (D) LMA g m^{-2}, (E) H_2O g m^{-2}, (F) protein g m^{-2}, (G) amino acids μmol m^{-2}, (H) nitrate μmol m^{-2}, (I) total non-structural carbohydrate (TNC) μmol C_6 m^{-2}, (J) starch μmol C_6 m^{-2}, (K) log total sugars μmol C_6 m^{-2}, (L) log sucrose μmol C_6 m^{-2}, (M) log glucose μmol C_6 m^{-2}, and (N) log fructose μmol C_6 m^{-2}.

RMSE=15%; Fig. 6), though the prediction error (RMSE) was still within the range of some of the other traits.

Discussion

This study has demonstrated the technical capability to estimate key plant metabolite pools associated with source–sink balance and C–N status non-destructively using leaf spectroscopy, via the spectra–trait approach. Previous work has demonstrated the robustness of the spectra–trait approach, particularly when applied to elemental and structural traits (i.e. C_{area}, N_{area}, LMA, and H_2O%) (Singh *et al.*, 2015; Nunes *et al.*, 2017; Silva-Perez *et al.*, 2018). More recent studies have used this approach to predict some C-containing metabolites (Lohr *et al.*, 2017; Yendrek *et al.*, 2017; Das *et al.*, 2018), generally within a single species. Here the capacity to predict a wider range of plant traits accurately from fresh leaf spectra across a range of species is demonstrated for the first time, including the novel prediction of traits associated with N metabolism.

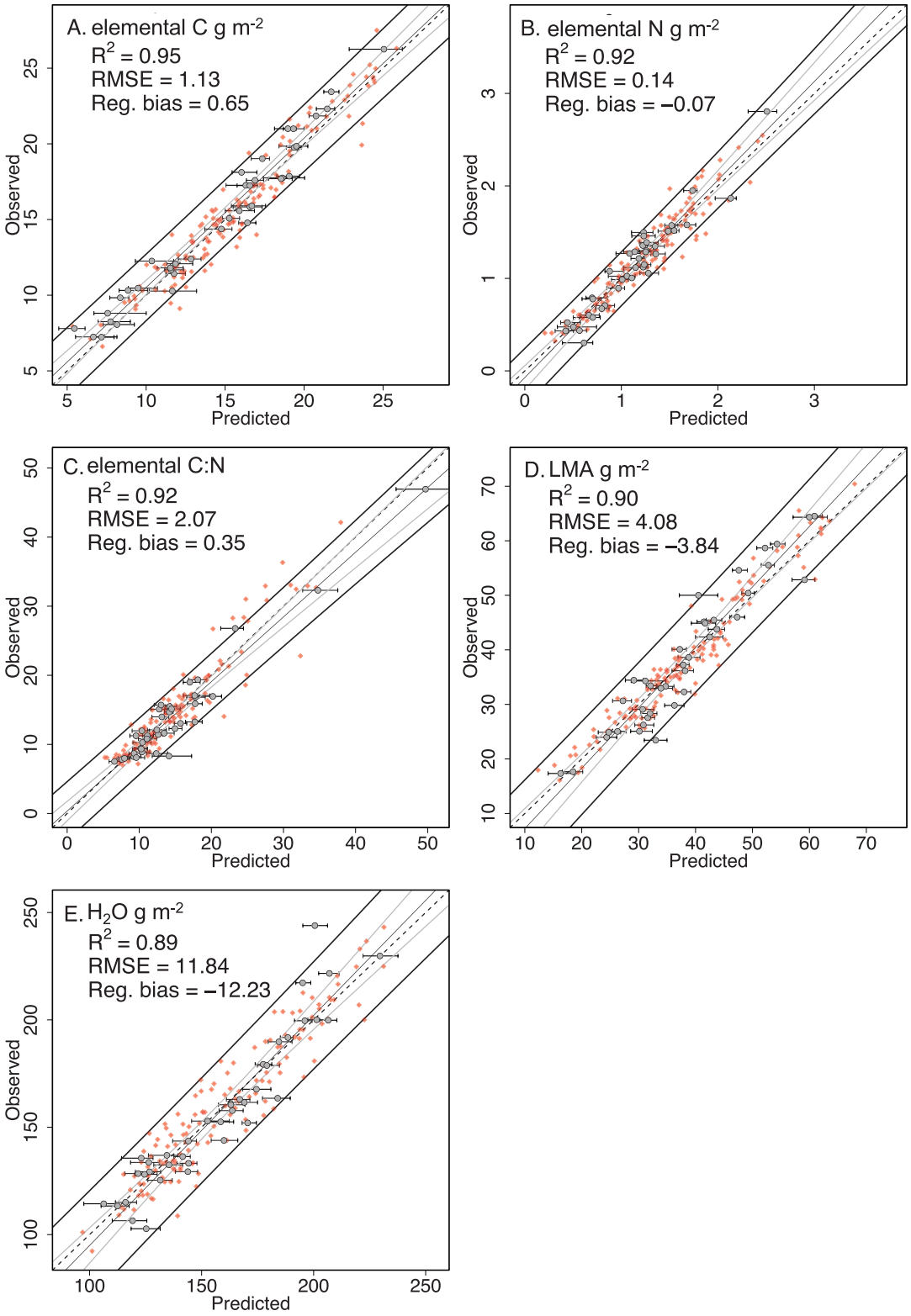

Fig. 4. Results for PLSR models of (A) carbon, (B) nitrogen, (C) C:N, (D) leaf mass per area (LMA), and (E) leaf water content. Calibration (red circles) and validation (gray circles; error bars show the 95% confidence interval) data points are shown. The 95% prediction interval (black lines), 95% confidence interval (gray lines), regression line (fine black line), and 1:1 line (dashed line) are shown. Statistics are for the validation results (refer to Table 1 for complete statistical reporting). All traits were modeled on a leaf area basis.

In order to develop a model with a broad potential application across many species and growth conditions, it is important that the model is developed from a data set that encompasses both a wide range of values for the given trait and a high diversity of leaf optical properties, so that the model covers as much of the potential trait and optical property space as possible. For elemental and structural traits in this study, this was assessed by comparison of our observed trait values with those

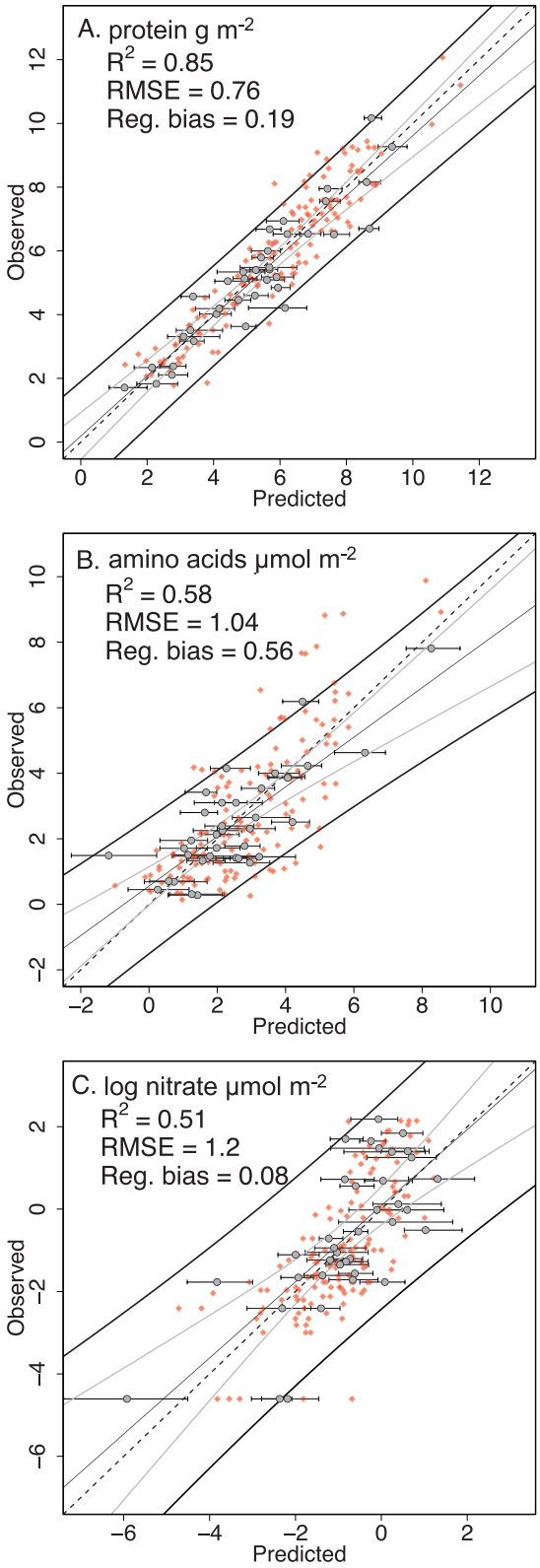

Fig. 5. Results for PLSR models of (A) protein g m^{-2}, (B) amino acids µmol m^{-2}, (C) log nitrate µmol m^{-2}. Calibration (red circles) and validation (gray circles; error bars show the 95% confidence interval) data points are shown. The 95% prediction interval (black lines), 95% confidence interval (gray lines), regression line (fine black line), and 1:1 line (dashed line) are shown. Statistics are for the validation results (refer to Table 1 for complete statistical reporting). All traits were modeled on a leaf area basis. Traits for which the raw data were highly skewed (i.e. nitrate) were natural log transformed prior to modeling.

available in the TRY plant trait database (Kattge *et al.*, 2011). The range of leaf N content in this study fully covers the range of N values reported in the TRY database, whereas leaf C content covers around half the range currently represented in the TRY database. The range of LMA in this study is representative of the lower 60% of the LMA range in TRY. This probably reflects the high species diversity in TRY including data from tropical forests where LMA is typically high. As recently highlighted (Martin and Isaac, 2015), there are no central databases for crop leaf traits, particularly for the physiological markers that we have investigated here, that would enable us to easily assess our success at covering the trait space for these C and N pools. However, databases such as TRY, the Biofuel Ecophysiological Traits and Yields (BETYdb) database (LeBauer *et al.*, 2018), or the Ecological Spectral Information System (EcoSIS) could be leveraged and expanded in the future to store this information.

For many metabolites, developing robust PLSR models is challenging because the *in vivo* pools of these metabolites are subject to regulation by central metabolism. For example, in the light, nitrate is rapidly metabolized and rarely accumulates to high levels (Heldt and Piechulla, 2010), although in low-light conditions a greater accumulation may be seen (EFSA, 2018), so sampling pre-dawn or under low light could provide high nitrate leaves to increase the range of nitrate available for PLSR model development. Storage of sugars is also controlled to avoid the build up of solutes since this can lead to osmotic stress, limiting the upper range for sugars. Excess sugars are typically stored as starch or fructan (Pollock and Cairns, 1991; Zeeman *et al.*, 2010), and therefore a wider range of starch pools was observed, resulting in a more robust PLSR model for starch than sugars. It may be possible to extend the natural range of these metabolites and improve the PLSR models through the use of extreme growth conditions, additional species that are known accumulators of metabolites of interest, or possible mutants with altered metabolism and storage (Müller-Röber *et al.*, 1992; Weichert *et al.*, 2010; Zuther *et al.*, 2011; EFSA, 2018).

Prediction of a wide range of plant traits from a single set of leaf spectra measurements provides insight into this modeling approach not apparent from less diverse studies. Comparison of traits with similar chemical properties suggests that higher predictive power using the PLSR approach may be achieved by building a model from a diverse data set. Of the simple sugars, glucose had both the widest compositional range, and the best predictive model, followed by sucrose; fructose had the narrowest range and least predictive power. Similarly, with the N-containing metabolites, amino acids had both a larger measured range and the resulting model was more accurate and has a lower error than the nitrate model, for which the measured data set included many values close to the lower detection limit of our assay. Improvement of the weaker models (i.e. nitrate and fructose) will require measurements from leaves with higher values of these metabolites.

The VIP results highlight the most influential spectral regions for predicting each trait, and enable the identification of commonalities between related traits. The most important spectral region for all traits occurs at ~702–715 nm, with C and C-containing metabolites peaking in the lower part of this

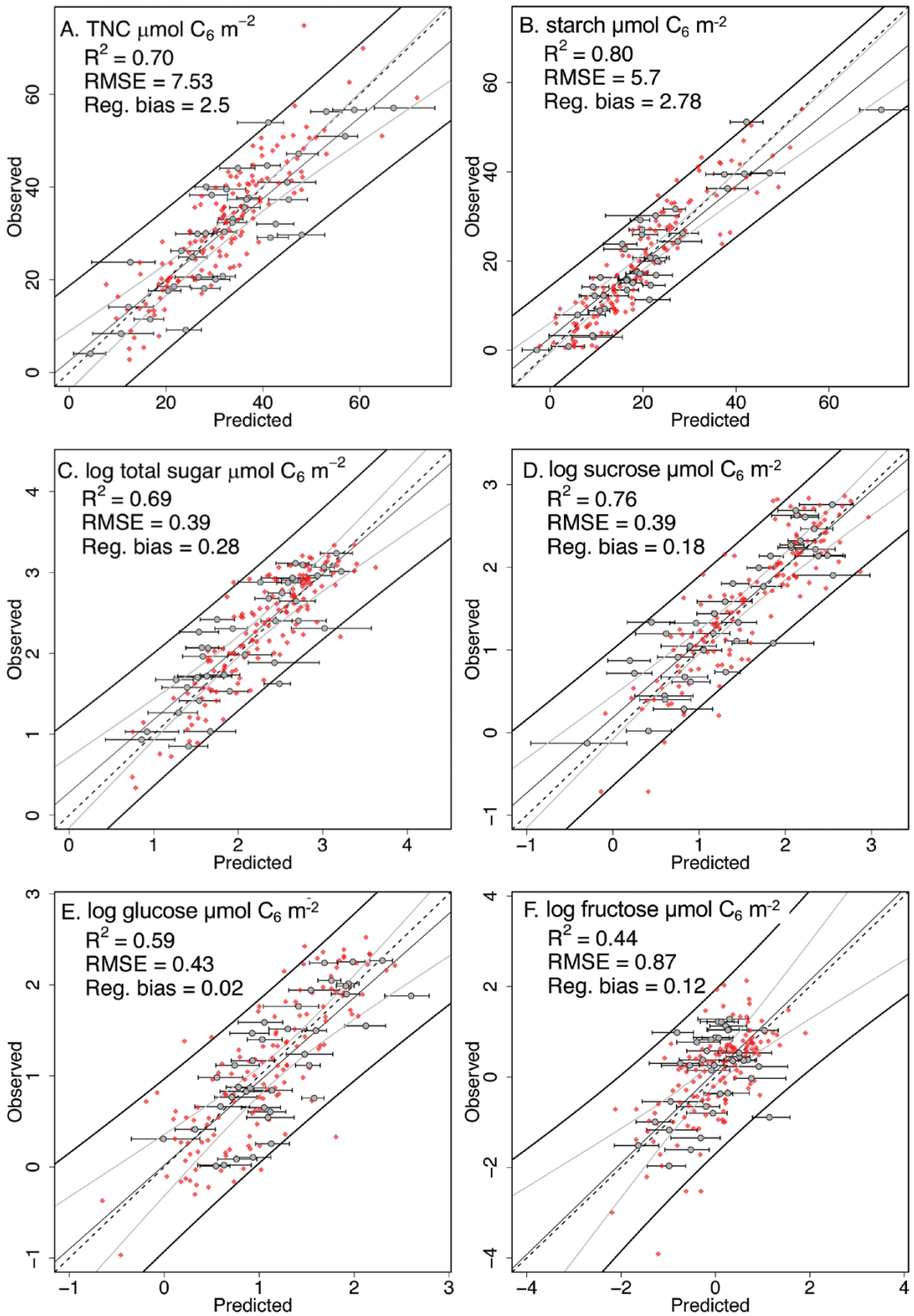

Fig. 6. Results for PLSR models of (A) total non-structural carbohydrates µmol C_6 m^{-2}, (B) starch µmol C_6 m^{-2}, (C) log total sugars µmol C_6 m^{-2}, (D) log sucrose µmol C_6 m^{-2}, (E) log glucose µmol C_6 m^{-2}, and (F) log fructose µmol C_6 m^{-2}. Calibration (red circles) and validation (gray circles; error bars show the 95% confidence interval) data points are shown. The 95% prediction interval (black lines), 95% confidence interval (gray lines), regression line (fine black line), and 1:1 line (dashed line) are shown. Statistics are for the validation results (refer to Table 1 for complete statistical reporting). All traits were modeled on a leaf area basis. Traits for which the raw data were highly skewed (i.e. sugars) were natural log transformed prior to modeling.

range and N and N-containing metabolites in the upper part, a common pattern observed in PLSR loadings for leaf physiological traits (e.g. Doughty *et al.*, 2011). This is on the 'red-edge' of the spectrum, at the transition between low reflectance of visible wavelengths and high reflectance in the NIR, which is associated with variation in leaf chlorophyll content and plant stress (Penuelas and Filella, 1998; Eitel *et al.*, 2011). Visible wavelengths centered on 550 nm form the second most important spectral region for C, N, LMA, protein, and nitrate, whereas for simple sugars and starch this peak centers around

530 nm. These wavelengths, which lie within the green part of the visible spectrum, have strong association with absorption features of the plant pigments anthocyanin and betacyanin (Peters and Noble, 2014). A third common region, with less influence, but a nonetheless well-defined peak, lies at 1394 nm for most traits, but slightly lower at 1387–1390 nm for sucrose, fructose, and the combined sugars and TNCs. This region lies at the transition from high to low reflectance; the water content of leaves is known to influence the slope of this transition (Kokaly *et al.*, 2009), and it is also a significant feature in the spectra of common leaf components including cellulose and polysaccharides (Elvidge, 1990).

The models presented here for C, N, LMA, and water content show equal or higher predictive power than previous studies on both single and multiple species (Asner *et al.*, 2017; Nunes *et al.*, 2017; Yendrek *et al.*, 2017; Silva-Perez *et al.*, 2018). Application of this technique for the prediction of N-containing metabolites measured in this study is novel. Only a few studies have used leaf spectra to predict simple sugars and starch. Das *et al.* (2018) and Yendrek *et al.* (2017) successfully predicted sucrose within a single species (R^2=0.78 and R^2=0.62, respectively). Das *et al.* (2018) also demonstrated good predictive capability for total sugars (R^2=0.72). Using a two-species model, Lohr *et al.* (2017) successfully predicted starch (R^2=0.84) and TNCs (R^2=0.83), but were not successful with glucose and fructose. Here we have demonstrated a moderate predictive ability for glucose and fructose.

Traditional techniques are often labor-intensive, time-consuming, and require destructive sampling, limiting the opportunity for frequent or repeated measurements; in addition, destructive sampling could confound the observed signals. The multispecies PLSR models described here can be used in future experiments and breeding programs that require high-throughput, inexpensive, and non-destructive measurements, and can enable the monitoring of key metabolic signals associated with source–sink balance and C–N status throughout a plant's life cycle without the need for repeated destructive harvesting. Efficient phenotyping platforms can accelerate breakthroughs in plant breeding (Furbank and Tester, 2011). The models presented here would enable advances in physiological breeding associated with source–sink balance and C–N status by facilitating the rapid plant phenotyping that is essential for this breeding strategy. By using spectral measurements rather than traditional techniques, data can be collected in a rapid and non-destructive manner enabling repeated analysis of plant metabolite status throughout the life cycle of a plant. Additionally, the rapid nature of this approach means that it could readily be used in combination with traditional breeding techniques. Further applications of this technique include uses in precision agriculture, and ecological monitoring such as understanding foliar carbohydrate content in the context of tree mortality (Dickman *et al.*, 2019).

In order to help advance further development of robust PLSR models that can be applied in many other species and environments, we have made all our biochemical trait data, leaf spectra, and code publicly available so that others may apply and evaluate our models and use our data in combination with their own to build new and improved models [leaf spectra, leaf trait data, and the R code for the PLSR model are available from the EcoSIS spectral library (ecosis.org), doi:10.21232/C2GM2Z].

Supplementary data

Supplementary data are available at *JXB* online.

Table S1. Detail of species, pot sizes, and growing media.

Fig. S1. PLSR model loadings for all traits assessed in this study.

Fig. S2. PLSR model coefficient vectors for all traits assessed in this study.

Acknowledgements

This work was supported by the United States Department of Energy contract no. DE-SC0012704 to Brookhaven National Laboratory.

References

Ainsworth EA, Davey PA, Hymus GJ, Osborne CP, Rogers A, Blum H, Nösberger J. 2003. Is stimulation of leaf photosynthesis by elevated carbon dioxide concentration maintained in the long term? A test with *Lolium perenne* grown for 10 years at two nitrogen fertilization levels under Free Air CO_2 Enrichment (FACE). Plant, Cell & Environment **26,** 705–714.

Ainsworth EA, Rogers A, Leakey AD, Heady LE, Gibon Y, Stitt M, Schurr U. 2007. Does elevated atmospheric [CO_2] alter diurnal C uptake and the balance of C and N metabolites in growing and fully expanded soybean leaves? Journal of Experimental Botany **58,** 579–591.

Ainsworth EA, Serbin SP, Skoneczka JA, Townsend PA. 2014. Using leaf optical properties to detect ozone effects on foliar biochemistry. Photosynthesis Research **119,** 65–76.

Al-Tamimi N, Brien C, Oakey H, Berger B, Saade S, Ho YS, Schmöckel SM, Tester M, Negrão S. 2016. Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping. Nature Communications **7,** 13342.

Aranjuelo I, Sanz-Sáez Á, Jauregui I, Irigoyen JJ, Araus JL, Sánchez-Díaz M, Erice G. 2013. Harvest index, a parameter conditioning responsiveness of wheat plants to elevated CO_2. Journal of Experimental Botany **64,** 1879–1892.

Arp WJ. 1991. Effects of source–sink relations on photosynthetic acclimation to elevated CO_2. Plant, Cell & Environment **14,** 869–875.

Asner GP, Martin RE. 2008. Spectral and chemical analysis of tropical forests: scaling from leaf to canopy levels. Remote Sensing of Environment **112,** 3958–3970.

Asner GP, Martin RE. 2015. Spectroscopic remote sensing of non-structural carbohydrates in forest canopies. Remote Sensing **7,** 3526–3547.

Asner GP, Martin RE, Anderson CB, *et al.* 2017. Scale dependence of canopy trait distributions along a tropical forest elevation gradient. New Phytologist **214,** 973–988.

Asner GP, Martin RE, Tupayachi R, Emerson R, Martinez P, Sinca F, Powell GV, Wright SJ, Lugo AE. 2011. Taxonomy and remote sensing of leaf mass per area (LMA) in humid tropical forests. Ecological Applications **21,** 85–98.

Barnes ML, Breshears DD, Law DJ, van Leeuwen WJD, Monson RK, Fojtik AC, Barron-Gafford GA, Moore DJP. 2017. Beyond greenness: detecting temporal changes in photosynthetic capacity with hyperspectral reflectance data. PLoS One **12,** e0189539.

Burnett AC, Rogers A, Rees M, Osborne CP. 2016. Carbon source–sink limitations differ between two species with contrasting growth strategies. Plant, Cell & Environment **39,** 2460–2472.

Carrascal LM, Galvan I, Gordo O. 2009. Partial least squares regression as an alternative to current regression methods used in ecology. Oikos **118,** 681–690.

Colombo R, Merom M, Marchesi A, Busetto L, Rossini M, Giardino C, Panigada C. 2008. Estimation of leaf and canopy water content in

poplar plantations by means of hyperspectral indices and inverse modeling. Remote Sensing of Environment **112,** 1820–1834.

Couture JJ, Singh A, Rubert-Nason KF, Serbin SP, Lindroth RL, Townsend PA. 2016. Spectroscopic determination of ecologically relevant plant secondary metabolites. Methods in Ecology and Evolution **7,** 1402–1412.

Dai Z, Plessis A, Vincent J, *et al*. 2015. Transcriptional and metabolic alternations rebalance wheat grain storage protein accumulation under variable nitrogen and sulfur supply. The Plant Journal **83,** 326–343.

Das B, Sahoo RN, Pargal S, Krishna G, Verma R, Chinnusamy V, Sehgal VK, Gupta VK, Dash SK, Swain P. 2018. Quantitative monitoring of sucrose, reducing sugar and total sugar dynamics for phenotyping of water-deficit stress tolerance in rice through spectroscopy and chemometrics. Spectrochimica Acta. Part A, Molecular and Biomolecular Spectroscopy **192,** 41–51.

Dechant B, Cuntz M, Vohland M, Schulz E, Doktor D. 2017. Estimation of photosynthesis traits from leaf reflectance spectra: correlation to nitrogen content as the dominant mechanism. Remote Sensing of Environment **196,** 279–292.

Dickman LT, McDowell NG, Grossiord C, *et al*. 2019. Homeostatic maintenance of non-structural carbohydrates during the 2015–2016 El Niño drought across a tropical forest precipitation gradient. Plant, Cell & Environment. doi:10.1111/pce.13501

Doughty CE, Asner GP, Martin RE. 2011. Predicting tropical plant physiology from leaf and canopy spectroscopy. Oecologia **165,** 289–299.

EFSA. 2018. Opinion of the Scientific Panel on Contaminants in the Food chain on a request from the European Commission to perform a scientific risk assessment on nitrate in vegetables. The EFSA Journal **269,** 1–79.

Eitel JUH, Vierling LA, Litvak ME, Long DS, Schulthess U, Ager AA, Krofcheck DJ, Stoscheck L. 2011. Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. Remote Sensing of Environment **115,** 3640–3646.

Elvidge CD. 1990. Visible and near infrared reflectance characteristics of dry plant materials. International Journal of Remote Sensing **11,** 1775–1795.

Emmett BD, Youngblut ND, Buckley DH, Drinkwater LE. 2017. Plant phylogeny and life history shape rhizosphere bacterial microbiome of summer annuals in an agricultural field. Frontiers in Microbiology **8,** 2414.

Eyles A, Pinkard EA, Davies NW, Corkrey R, Churchill K, O'Grady AP, Sands P, Mohammed C. 2013. Whole-plant versus leaf-level regulation of photosynthetic responses after partial defoliation in *Eucalyptus globulus* saplings. Journal of Experimental Botany **64,** 1625–1636.

FAO, IFAD, UNICEF, WFP, WHO. 2017. The State of Food Security and Nutrition in the World 2017. Building resilience for peace and food security. Rome: FAO.

Ferdous J, Whitford R, Nguyen M, Brien C, Langridge P, Tricker PJ. 2017. Drought-inducible expression of Hv-miR827 enhances drought tolerance in transgenic barley. Functional & Integrative Genomics **17,** 279–292.

Furbank RT, Tester M. 2011. Phenomics—technologies to relieve the phenotyping bottleneck. Trends in Plant Science **16,** 635–644.

Gibon Y, Blaesing OE, Hannemann J, Carillo P, Höhne M, Hendriks JH, Palacios N, Cross J, Selbig J, Stitt M. 2004. A robot-based platform to measure multiple enzyme activities in *Arabidopsis* using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. The Plant Cell **16,** 3304–3325.

Heldt H-W, Piechulla B. 2010. Plant biochemistry. New York: Academic Press.

Isopp H, Frehner M, Long SP, Nösberger J. 2000. Sucrose-phosphate synthase responds differently to source–sink relations and to photosynthetic rates: *Lolium perenne* L. growing at elevated p(CO_2) in the field. Plant, Cell & Environment **23,** 597–607.

Kattge J, Diaz S, Lavorel S, *et al*. 2011. TRY—a global database of plant traits. Global Change Biology **17,** 2905–2935.

Kokaly RF, Asner GP, Ollinger SV, Martin ME, Wessman CA. 2009. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. Remote Sensing of Environment **113,** S78–S91.

Krapp A, Quick WP, Stitt M. 1991. Ribulose-1,5-bisphosphate carboxylase-oxygenase, other Calvin-cycle enzymes, and chlorophyll decrease when glucose is supplied to mature spinach leaves via the transpiration stream. Planta **186,** 58–69.

LeBauer D, Kooper R, Mulrooney P, Rohde S, Wang D, Long SP, Dietze MC. 2018. BETYdb: a yield, trait, and ecosystem service database applied to second-generation bioenergy feedstock production. GCB Bioenergy **10,** 61–71.

Lohr D, Tillmann P, Druege U, Zerche S, Rath T, Meinken E. 2017. Non-destructive determination of carbohydrate reserves in leaves of ornamental cuttings by near-infrared spectroscopy (NIRS) as a key indicator for quality assessments. Biosystems Engineering **158,** 51–63.

Madritch MD, Kingdon CC, Singh A, Mock KE, Lindroth RL, Townsend PA. 2014. Imaging spectroscopy links aspen genotype with below-ground processes at landscape scales. Philosophical Transactions of the Royal Society B: Biological Sciences **369,** 20130194.

Martin AR, Isaac ME. 2015. Plant functional traits in agroecosystems: a blueprint for research. Journal of Applied Ecology **52,** 1425–1435.

Mevik B-H, Wehrens R, Hovde Liland K. 2016. Partial least squares and principal component regression. R package version 2.6-0.

Müller-Röber B, Sonnewald U, Willmitzer L. 1992. Inhibition of the ADP-glucose pyrophosphorylase in transgenic potatoes leads to sugar-storing tubers and influences tuber formation and expression of tuber storage protein genes. The EMBO Journal **11,** 1229–1238.

Nunes MH, Davey MP, Coomes DA. 2017. On the challenges of using field spectroscopy to measure the impact of soil type on leaf traits. Biogeosciences **14,** 3371–3385.

Ort DR, Merchant SS, Alric J, *et al*. 2015. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. Proceedings of the National Academy of Sciences, USA **112,** 8529–8536.

Paul MJ, Driscoll SP. 1997. Sugar repression of photosynthesis: the role of carbohydrates in signalling nitrogen deficiency through source:sink imbalance. Plant, Cell & Environment **20,** 110–116.

Penuelas J, Filella I. 1998. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. Trends in Plant Science **3,** 151–156.

Peters RD, Noble SD. 2014. Spectrographic measurement of plant pigments from 300 to 800 nm. Remote Sensing of Environment **148,** 119–123.

Petisco C, García-Criado B, Mediavilla S, Vázquez de Aldana BR, Zabalgogeazcoa I, García-Ciudad A. 2006. Near-infrared reflectance spectroscopy as a fast and non-destructive tool to predict foliar organic constituents of several woody species. Analytical and Bioanalytical Chemistry **386,** 1823–1833.

Pettorelli N, Schulte to Bühne H, Tulloch A, *et al*. 2018. Satellite remote sensing of ecosystem functions: opportunities, challenges and way forward. Remote Sensing in Ecology and Conservation **4,** 71–93.

Pollock CJ, Cairns AJ. 1991. Fructan metabolism in grasses and cereals. Annual Review of Plant Physiology and Plant Molecular Biology **42,** 77–101.

R Core Team. 2018. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Reekie EG, MacDougall G, Wong I, Hicklenton PR. 1998. Effect of sink size on growth response to elevated atmospheric CO_2 within the genus *Brassica*. Canadian Journal of Botany **76,** 829–835.

Reynolds M, Foulkes J, Furbank R, Griffiths S, King J, Murchie E, Parry M, Slafer G. 2012. Achieving yield gains in wheat. Plant, Cell & Environment **35,** 1799–1823.

Reynolds M, Langridge P. 2016. Physiological breeding. Current Opinion in Plant Biology **31,** 162–171.

Rogers A, Allen DJ, Davey PA, *et al*. 2004. Leaf photosynthesis and carbohydrate dynamics of soybeans grown throughout their life-cycle under Free-Air Carbon dioxide Enrichment. Plant, Cell & Environment **27,** 449–458.

Serbin SP, Dillaway DN, Kruger EL, Townsend PA. 2012. Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature. Journal of Experimental Botany **63,** 489–502.

Serbin SP, Singh A, McNeil BE, Kingdon CC, Townsend PA. 2014. Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. Ecological Applications **24,** 1651–1669.

Silva-Perez V, Molero G, Serbin SP, Condon AG, Reynolds MP, Furbank RT, Evans JR. 2018. Hyperspectral reflectance as a tool to measure biochemical and physiological traits in wheat. Journal of Experimental Botany **69,** 483–496.

Singh A, Serbin SP, McNeil BE, Kingdon CC, Townsend PA. 2015. Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. Ecological Applications **25,** 2180–2197.

Sreenivasulu N, Schnurbusch T. 2012. A genetic playground for enhancing grain number in cereals. Trends in Plant Science **17,** 91–101.

Stitt M, Krapp A. 1999. The interaction between elevated carbon dioxide and nitrogen nutrition: the physiological and molecular background. Plant, Cell & Environment **22,** 583–621.

von Caemmerer S, Quick WP, Furbank RT. 2012. The development of C_4 rice: current progress and future challenges. Science **336,** 1671–1672.

Weichert N, Saalbach I, Weichert H, *et al*. 2010. Increasing sucrose uptake capacity of wheat grains stimulates storage protein synthesis. Plant Physiology **152,** 698–710.

White AC, Rogers A, Rees M, Osborne CP. 2016. How can we make plants grow faster? A source–sink perspective on growth rate. Journal of Experimental Botany **67,** 31–45.

Wold S, Sjostrom M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems **58,** 109–130.

Yang GJ, Liu JG, Zhao CJ, *et al*. 2017. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. Frontiers in Plant Science **8,** 1111.

Yendrek CR, Tomaz T, Montes CM, Cao Y, Morse AM, Brown PJ, McIntyre LM, Leakey AD, Ainsworth EA. 2017. High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance. Plant Physiology **173,** 614–626.

Zeeman SC, Kossmann J, Smith AM. 2010. Starch: its metabolism, evolution, and biotechnological modification in plants. Annual Review of Plant Biology **61,** 209–234.

Zuther E, Hoermiller II, Heyer AG. 2011. Evidence against sink limitation by the sucrose-to-starch route in potato plants expressing fructosyltransferases. Physiologia Plantarum **143,** 115–125.