AAAS
Plant Phenomics
Volume 2022, Article ID 9768502, 18 pages
https://doi.org/10.34133/2022/9768502

Research Article

Estimating Photosynthetic Attributes from High-Throughput Canopy Hyperspectral Sensing in Sorghum

Xiaoyu Zhi,[1] **Sean Reynolds Massey-Reed**,[1] **Alex Wu**,[2] **Andries Potgieter**,[3] **Andrew Borrell**,[1] **Colleen Hunt**,[1,4] **David Jordan**,[1] **Yan Zhao**,[2,3] **Scott Chapman**,[2,5] **Graeme Hammer**,[2] **and Barbara George-Jaeggli**[1,4]

[1]*The University of Queensland, Queensland Alliance for Agriculture and Food Innovation (QAAFI), Hermitage Research Facility, Warwick, QLD, Australia*
[2]*The University of Queensland, Queensland Alliance for Agriculture and Food Innovation (QAAFI), St Lucia, QLD, Australia*
[3]*The University of Queensland, Queensland Alliance for Agriculture and Food Innovation (QAAFI), Gatton, QLD, Australia*
[4]*Agri-Science Queensland, Department of Agriculture and Fisheries (DAF), Hermitage Research Facility, Warwick, QLD, Australia*
[5]*School of Agriculture and Food Sciences, The University of Queensland, Gatton, QLD, Australia*

Correspondence should be addressed to Xiaoyu Zhi; x.zhi@uq.net.au and Barbara George-Jaeggli; b.georgejaeggli@uq.edu.au

Received 26 October 2021; Accepted 25 February 2022; Published 8 April 2022

Sorghum, a genetically diverse C_4 cereal, is an ideal model to study natural variation in photosynthetic capacity. Specific leaf nitrogen (SLN) and leaf mass per leaf area (LMA), as well as, maximal rates of Rubisco carboxylation (V_{cmax}), phosphoenolpyruvate (PEP) carboxylation (V_{pmax}), and electron transport (J_{max}), quantified using a C_4 photosynthesis model, were evaluated in two field-grown training sets ($n = 169$ plots including 124 genotypes) in 2019 and 2020. Partial least square regression (PLSR) was used to predict V_{cmax} ($R^2 = 0.83$), V_{pmax} ($R^2 = 0.93$), J_{max} ($R^2 = 0.76$), SLN ($R^2 = 0.82$), and LMA ($R^2 = 0.68$) from tractor-based hyperspectral sensing. Further assessments of the capability of the PLSR models for V_{cmax}, V_{pmax}, J_{max}, SLN, and LMA were conducted by extrapolating these models to two trials of genome-wide association studies adjacent to the training sets in 2019 ($n = 875$ plots including 650 genotypes) and 2020 ($n = 912$ plots with 634 genotypes). The predicted traits showed medium to high heritability and genome-wide association studies using the predicted values identified four QTL for V_{cmax} and two QTL for J_{max}. Candidate genes within 200 kb of the V_{cmax} QTL were involved in nitrogen storage, which is closely associated with Rubisco, while not directly associated with Rubisco activity *per se*. J_{max} QTL was enriched for candidate genes involved in electron transport. These outcomes suggest the methods here are of great promise to effectively screen large germplasm collections for enhanced photosynthetic capacity.

1. Introduction

Sorghum (*Sorghum bicolor* L. Moench), a C_4 pathway species and the world's fifth most produced cereal [1], is adapted to a range of environments and retains high photosynthetic efficiency in diverse conditions [2–4]. These characteristics make it a crop of interest for the dual challenge of meeting increasing demands for food and adapting to the effects of climate change [5, 6]. In addition to the C_4 pathway, which confers adaptation to hot and dry environments, the natural genetic diversity of sorghum provides potential to identify genotypes or genetic loci associated with greater photosynthetic capacity [7]. However, in order to select the photosynthetically favourable genotypes adapted to contrasting environments, tools are required to quantify the biochemical parameters underpinning photosynthetic capacity in a high-throughput manner, removing the phenotyping bottleneck with the traditional gas exchange approach.

Photosynthesis is the process of converting captured solar radiation into chemical energy by fixing carbon dioxide (CO_2) to form carbohydrates and biomass. Improving photosynthetic capacity is seen as a major target to further improve crop yields [2, 3, 8]. Screening germplasm to directly breed for improved photosynthetic responses to environment conditions is constrained by the complexity

of measuring such responses and requires development of higher-throughput indirect phenotyping techniques.

In the C_4 photosynthetic pathway, the biochemical processes in the mesophyll cells are coordinated with a CO_2 concentrating mechanism in the bundle-sheath cells [9, 10]. In the mesophyll, CO_2 is initially fixed by phosphoenolpyruvate (PEP) carboxylase into C_4 acids, which are then decarboxylated in the bundle sheath cells leading to high CO_2 levels and hence more efficient carboxylation of Ribulose-1,5-bisphosphate (RuBP) by Ribulose 1,5-bisphosphate carboxylase-oxygenase (Rubisco) [11, 12]. The energy for the regeneration of RuBP in the bundle sheath and PEP in the mesophyll comes from chloroplast electron transport [11]. Due to their key roles in the photosynthetic pathway, the maximal rates of Rubisco carboxylation (V_{cmax}, μmol $m^{-2}s^{-1}$), PEP carboxylation (V_{pmax}, μmol $m^{-2}s^{-1}$), and maximal electron transport rate (J_{max}, μmol $m^{-2}s^{-1}$) largely determine photosynthetic capacity of C_4 plants and therefore underpin crop productivity. Simulations using a diurnal canopy photosynthesis model predict that canopy growth rate of C_4 cereals responds largely to changes in J_{max} [13]. Quantification of these biochemical parameters is hence of value for selecting enhanced photosynthesis and growth. This is traditionally achieved by conducting gas exchange measurements and fitting observed photosynthetic responses to CO_2 or light with the Rubisco-activity or electron-transport limited equations in the C_4 photosynthesis model [11, 14]. However, this method is very time-consuming and not suitable for high-throughput screening of large germplasm collections.

The capacity of leaves to convert absorbed CO_2 and radiation into biomass also depends on key leaf physiological and structural properties [15]. Two such properties are specific leaf nitrogen (SLN, g m^{-2}) and leaf mass per leaf area (LMA, g m^{-2}), and both of these are known to be closely associated with photosynthetic capacity [16, 17]. Because nitrogen is a key element in photosynthetic machinery, such as chloroplasts, plant nitrogen status closely links with leaf photosynthetic rates and canopy radiation use efficiency [18–20] and is hence an important parameter in canopy performance modelling [13, 21]. The relationship between leaf nitrogen content and maximal net photosynthesis rate is influenced by LMA which is strongly associated with leaf lifespan and thus affecting the rates of the photosynthetic parameters [15, 16, 22]. However, conventional measurements of SLN and LMA are destructive and slow, limiting their potential to identify germplasm with higher photosynthetic capacity in large breeding programs.

High-throughput plant phenotyping technologies enable the collection of plant biochemical and physiological traits rapidly and nondestructively at large scale [23–26]. Various vegetation indices, which are usually calculated using a few selected wavelengths, have been correlated with plant structural traits (e.g., leaf area index and biomass) or leaf pigment concentration (e.g., chlorophyll). Typical canopy size indicators include normalized difference vegetation index (NDVI) [27, 28] and optimized soil adjusted vegetation index (OSAVI) [29]. Chlorophyll content, on the other hand, has been indicated by indices, such as normalized difference red edge (NDRE) [30] and chlorophyll vegetation index (CVI), which is an indirect measure of nitrogen content [31]. Adjustments to these vegetation indices have also been reported. For example, replacing red bands with red edge when calculating some indices exhibited better performance in estimating chlorophyll content [32].

More recently, hyperspectral imaging sensors with wavelengths in the visible (400-700 nm), near infrared (700-1000 nm), and shortwave infrared (1000-2500 nm) domain have advanced the development of high-resolution spectroscopy techniques. This has led to significant increases in the accuracy and the types of physiological properties that can be retrieved [26, 33]. The linkage between photosynthetic capacity and hyperspectral features therefore constitutes a promising avenue to predict photosynthetic performance of plants across broad scales [20, 34–36]. Various studies have exploited the plethora of bands (>270) and the much narrower band width (<6 nm) available from current hyperspectral sensors to better quantify biochemical and physiological properties in crops [35, 37]. However, most of the studies so far use hyperspectral reflectance to estimate leaf photosynthetic capacity in C_3 crops [34, 35, 37–41], and similar studies are much rarer for C_4 crops. At least one study focused on V_{cmax}, V_{pmax}, leaf nitrogen content, and specific leaf area from whole spectra reflectance (500-2400 nm) using partial least square regression (PLSR) in C_4 crop maize [42]. However, J_{max} that quantifies the rate of electron-transport limited photosynthetic rate [11] is also important in determining daily biomass growth [13], but has not previously been targeted.

A more comprehensive study on quantifying the key parameters of photosynthesis V_{cmax}, V_{pmax}, and J_{max} in a C_4 crop species is proposed. In addition, a high-throughput method to predict key parameters linked to photosynthetic capacity from canopy-level hyperspectral measurements will aid in the selection of genetic material with improved photosynthetic capacity at a large scale. To our knowledge, there are no published previous attempts to estimate the full set of key parameters known to limit C_4 photosynthesis, at canopy level, using hyperspectral reflectance. Additionally, next generation sequencing techniques have provided a high-throughput and cost-efficient tool for detecting genomic regions associated with crop traits of interest via genome-wide association studies (GWAS) [43–45]. Combining the techniques of hyperspectral sensing and GWAS would greatly facilitate the improvement of photosynthetic capacity and ultimate crop performance, which to date has rarely been explored.

The main objective of this study was to estimate traits associated with photosynthetic capacity from proximal hyperspectral sensing of sorghum canopies. Specifically, we aimed to (i) develop algorithms to predict photosynthetic parameters (V_{cmax}, V_{pmax}, and J_{max}), SLN, and LMA from proximal hyperspectral canopy reflectance captured with a spectrometer attached to a mobile phenotyping platform in two field-grown training sets; (ii) extrapolate the algorithms to GWAS trials grown adjacent to the training sets using a fully genotyped sorghum diversity panel; (iii) evaluate the heritability of the predicted traits; and (iv) undertake GWAS to detect genomic loci associated with the key

photosynthetic parameters and identify potential candidate genes to assess the usefulness and robustness of the approaches used in this study.

2. Materials and Methods

2.1. GWAS Trials. Two field experiments were conducted during two consecutive summer seasons (2019 and 2020) at Gatton Research Station (GAT), Gatton, Queensland, Australia (27°33′ S, 152°20′E, 94 m above sea level). GAT1 and GAT2 were sown on 14 January 2019 and 12 November 2019, respectively. Both trials were designed using partial replication with spatially randomised genotypes arranged in rows and columns. There were 875 plots, including 650 genotypes in GAT1, and 912 plots, including 634 genotypes in GAT2, with 70 genotypes in common between trials (Table 1). The genotypes in GAT1 were all inbred lines ($n = 649$) from a sorghum diversity panel comprising world-wide collections [43], and one hybrid was also included. In GAT2, 89% genotypes were hybrids from the Queensland breeding program, and the rest were inbred lines from the sorghum diversity panel. Each plot (4.5 m length and 3 m width) sown to a genotype consisted of four rows. Both trials were planted with a GPS precision planter at a population density of 108,000 plants ha^{-1}. For both trials, 150 kg of nitrogen per hectare was applied preplanting, and plots were irrigated regularly to provide nutrient and water nonlimiting conditions. The temperature, photosynthetic photon flux (PPF), and relative humidity (RH) from 6 am to 6 pm for the duration of each trial are shown in Table 1.

2.2. Training Sets. Adjacent to each of the GWAS trials, a training set comprising a representative sample of the lines in the GWAS trials was used to collect ground truth data for association with hyperspectral measurements. Completely randomised block designs (row-column) were also used in the training sets. The middle two rows (0.63 m row spacing) of each four-row plot were used for the ground truth data collection while the outside two rows (0.75 m row spacing) were guard rows. The training set in 2019 (TS1) consisted of 80 plots comprising 60 genotypes which were all inbred lines and also included in GAT1. In the training set of 2020 (TS2), there were 108 plots with 93 genotypes of which 63 (68%) were hybrids. There were 19 genotypes in common between TS1 and TS2. Due to differences in germination and vigour of the diverse germplasm used, there was substantial variability in final plant establishment in both trials. The ground truth measurements were only taken from plots which had good establishment, which reduced the number of possible observations that could be used to develop the models. To maximise the number and the range of observations, the ground truth data from TS1 and TS2 were pooled.

2.3. Ground Truth Measurements in the Training Sets. In both trials, gas exchange measurements were taken under mostly cloudless conditions (between 9 am and 12 pm) between 35 and 50 days after sowing (DAS)), which was during the active vegetative growth period for all genotypes and hence before the switch to reproductive growth which may introduce physiological and metabolic changes, but after full canopy closure. This period is known to be the most critical period for grain production in sorghum [46]. In total, 75 CO_2 (ACi) and 75 light (Ai) response curves were collected across TS1 ($n = 31$ plots comprising 29 inbred lines) and TS2 ($n = 44$ plots comprising 30 hybrid and 10 inbred lines) with six inbred lines in common between TS1 and TS2. One plant per plot was randomly selected for gas exchange measurements. The ACi curves were performed on the last or second last fully expanded leaf using a LI-6400 (LI-COR, Inc., Lincoln, Nebraska USA) with a 6400-02B Red/Blue LED light source illuminating a leaf chamber of 6 cm^2. To measure ACi curves, photosynthetically active radiation (PAR) was set at 1800 μmol photons $m^{-2}s^{-1}$, flow rate through the chamber at 500 μmol mol^{-1}, and temperature was set to leaf temperature measured at the commencement of each curve. Vapour-pressure deficit (VPD) was generally held at around 3.0 kPa, by adjusting the scrubbing of the incoming air via the desiccant. For each ACi curve, the reference CO_2 levels were set to the sequences of 200, 100, 50, 250, 400, 650, 800, 1000, 1200, and 1400 ppm, with a duration of 1-5 min for each step. Measurements were made at each CO_2 supply point when gas exchange had equilibrated, at which point, the coefficient of variation for the CO_2 concentration differential between the sample and reference analysers was below 1%. The light levels for the Ai curves were set at 2000, 1500, 1000, 500, 250, 120, 60, 30, 15, and 0 μmol $m^{-2}s^{-1}$. The other controls were set as follows: reference CO_2 (constant at 400 μmol mol^{-1}), flow (500 μmol mol^{-1}), temperature was set to leaf temperatures, and humidity was controlled by scrubbing incoming air to maintain a VPD around 3.0 kPa. The duration for every light level was 1-3 min. Sample and reference analysers were matched before each data point was logged.

A small square section of the leaf (1.6 cm^2) was collected with a leaf punch from the same leaf section as was used for gas exchange measurements. The leaf sections were dried at 80°C and weighed to calculate LMA (g m^{-2}). Percent nitrogen of each sample was determined with a continuous flow isotope ratio mass spectrometer (CF-IRMS), and SLN (g m^{-2}) was calculated by multiplying percent nitrogen with LMA. Across the two training sets, 129 SLN and 169 LMA observations (plots) were obtained, involving 124 unique genotypes.

To generate a maximised dataset and enhance robustness of associating the ground truth data taken in a plot with hyperspectral measurements obtained from the same plot, individual plots, rather than genotypes, were considered as an observational unit.

2.4. Canopy Hyperspectral Measurements. Hyperspectral data captured before anthesis and around the same time as the ground-truthing data (at 58 and 52 DAS in 2019 and 2020, respectively) was used to associate with the ground truth data. At this stage of sorghum crop growth, canopies are fully closed and nitrogen content of individual leaves is expected to be at a maximum as all mainstem leaves are fully expanded, but, prior to any translocation of nitrogen during senescence [47]. A tractor-based field phenotyping platform

Table 1: Top: mean and maximum daily temperatures, mean daily photosynthetic photon flux, and relative humidity during the two GWAS trials and two training sets in 2019 and 2020; bottom: number of plots and genotypes used in each experiment; and the genotypes in common between trials are in italic.

Year	Temperature (°C)		PPF (μmol $s^{-1}m^{-2}$)	RH (%)
	Mean	Maximum	Mean	
2019	26.84	38.98	743.11	62.86
2020	29.22	38.52	1000.95	56.1
Trials	TS1	TS2	GAT1	GAT2
TS1	80 plots (60 genotypes)	19 *genotypes*	60 *genotypes*	36 *genotypes*
TS2		108 plots (93 genotypes)	30 *genotypes*	92 *genotypes*
GAT1			875 plots (650 genotypes)	70 *genotypes*
GAT2				912 plots (634 genotypes)

Note: photosynthetic photon flux (PPF) and relative humidity (RH); the trials in 2019 including the training set TS1 and the GWAS trial GAT1; the trials in 2020 including the training set TS2 and the GWAS trial GAT2.

(GECKO; developed at The University of Queensland) which enables simultaneous crop canopy proximal sensing was used [48]. The tractor moves at a constant 1.1 metres per second and is integrated with a GPS real-time kinematic system with 2 cm accuracy to locate sampling plots (individual size of 4.5 × 3 m). A microhyperspectral imager (Micro-Hyperspec VNIR model, Headwall Photonics, Fitchburg, MA, USA) mounted on this phenotyping platform (3 m above ground and~1.7 m above the canopy) was used to obtain the spectral response of each pixel (5 × 5 mm) at 272 spectral wavelengths between 395 and 997 nm (visible and near infrared). The resolution was approximately 2.2 nm with 6.0 nm Full Width Half Maxima. A radiometric calibration (dark signal calibration) of the hyperspectral camera was performed weekly. A spectral calibration using the nominal white and spectral diffusers with specific band sets focused on the highest possible spectral resolution was conducted every three months by comparing their respective responses in almost identical illumination conditions. An automated software data calibration pipeline was used to convert raw digital numbers to reflectance values at each pixel. Pixel reflectance was calculated by the ratio between pixel radiance from the microhyperspectral imager and the reference pixel radiance from an upward sensor measuring incoming radiance. To segment plants from soil and remove background noise from lower canopy levels, a threshold of NDVI > 0.5 was applied for each pixel based on the fractional vegetation cover [27, 36, 49], which could ensures only spectral information from green leaves is retained for the reflectance calculations and shadows and other background noise are excluded from the hyperspectral images. After masking by NDVI > 0.5, plant pixels within a plot were averaged to calculate reflectance of each plot. All hyperspectral data was collected from 9 am to 12 pm to minimise the effects of relative orientation of the sun, and no adjustments were made for the sensor or the distribution of leaf angles in the masking. As an example, images, radiance, and reflectance pre- and postmasking by NDVI > 0.5 for plot 361 in 2020 are shown in Figure 1.

A set of hyperspectral vegetation indices known to be associated with photosynthesis was computed from the plot reflectance involving 16 wavelengths as shown in Figure 1. The equations used to calculate the indices in this study were summarised in Table 2.

Note: Wavelengths with black bars show the wavelengths used for calculating the set of vegetation indices known to be associated with photosynthesis; wavelengths with red bars indicate the wavelengths involved in the stepwise linear regression (referring to 2.2).

2.5. Determining V_{cmax}, V_{pmax}, and $J_{\max}$ from ACi and Ai Curves. For quantifying the actual photosynthetic parameters, we applied the C_4 photosynthesis model to the measured ACi and Ai response curves [11, 14]. The CO_2 assimilation rate (A) in the bundle sheath is given by the minimum of either Rubisco carboxylation limited (A_c) or electron transport limited (A_j) rates:

$$A = \min\left(A_c, A_j\right), \tag{1}$$

where,

$$A_c = \frac{(C_s - \gamma * O_s)V_{\text{cmax}}}{(C_s + K_c(1 + (O_s/K_o)))} - R_d, \tag{2}$$

$$A_j = \frac{(1 - \gamma * O_s/C_s)(1 - x)J_t}{3(1 + 7\gamma * O_s/(3C_s))} - R_d, \tag{3}$$

where O_s is the O_2 partial pressure in the bundle sheath, $\gamma *$ is the half of the reciprocal of Rubisco specificity, K_c and K_o are the Michaelis-Menten constant of Rubisco for CO_2 and O_2, respectively, and R_d is the mitochondrial respiration rate in the light. All enzymatic constants and variables in the equations above were detailed in a previous study [8].

The C_s (CO_2 partial pressure in the bundle sheath) is modelled by ambient CO_2 (C_a) entering the leaf via stomata and being diffused into the mesophyll, converted into C_4 acids then decarboxylated, and released as CO_2 in the bundle sheath. The supply of CO_2 to the mesophyll (C_m) depends

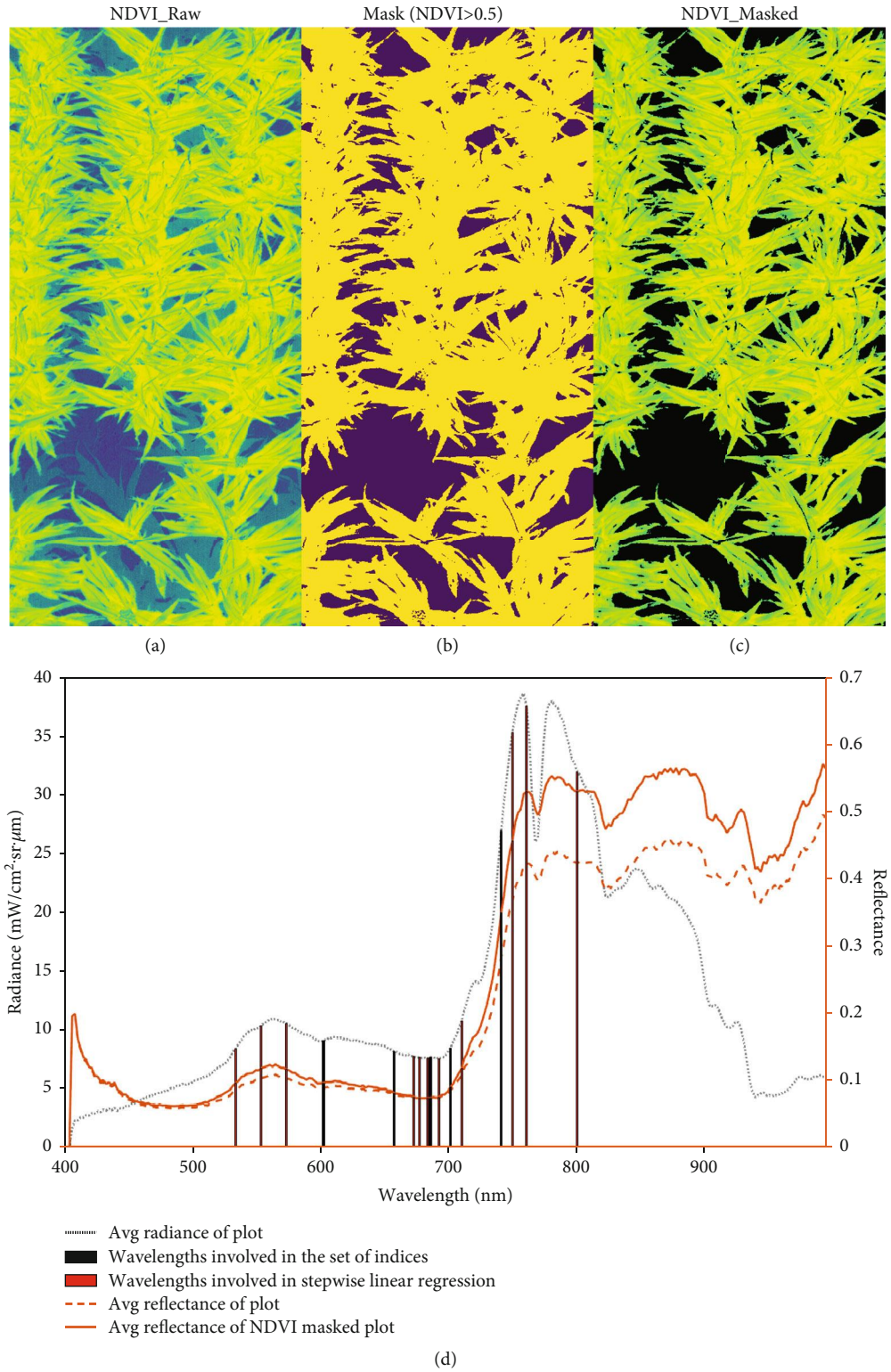

Figure 1: An example (plot 361 in training set 2) of plant canopy area (a) before and (c) after masking by (b) NDVI > 0.5; averaged plot radiance and reflectance before and after masking by NDVI > 0.5 (d).

Table 2: Summary of the equations for the set of vegetation indices associated with photosynthesis.

Acronym	Indices	Traits associated	Equations	References
Curvature	Curvature between red and NIR	Chlorophyll content	p683^2/(p675 × p690)	[50]
CVI	Chlorophyll vegetation index	Chlorophyll content	(p750/p550) × (p670/p550)	[31]
NDRE	Normalized difference red edge	Chlorophyll content	(p750 − p710)/(p750 + p710)	[30]
NDVI	Normalized difference vegetation index	Leaf area index	(p800 − p670)/(p800 + p670)	[28]
PRI	Photochemical reflectance index	Photosynthetic efficiency	(p531 − p570)/(p531 + p570)	[51]
r685_r655		Chlorophyll fluorescence	p685/p655	[52]
r690_r600			p690/p600	[52]
r740_r700			p740/p700	[52]
r760_r750			p760/p750	[52]
r760_r750index			$(p760 - p750)/(p760 + p750)$	[53]
Red_edge		Chlorophyll content/leaf area index	$p750/p710$	[54]
OSAVI	Optimized soil adjusted vegetation index	Leaf area index	(1 + 0.16) × (p800 − p670)/(p800 + p670 + 0.16)	[29]
r750		Vcmax	p750	[34]
r760		Chlorophyll fluorescence	$p760$	[55]
TVI	Transformed vegetation index	Leaf area index	$0.5 \times (120 \times (p750 - p550)) - 200 \times (p670 - p550))$	[56]

on the intercellular CO_2 partial pressure (C_i), the mesophyll conductance (g_m), and the demand term, which is the CO_2 assimilation rate A:

$$C_m = C_a \times \frac{C_i}{C_a} - \frac{A}{g_m}. \quad (4)$$

Here, the effects of the leaf boundary layer and stomatal conductance are incorporated into the C_i/C_a term.

The supply of CO_2 to the bundle sheath (C_s) can be limited by enzymatic capacity or chemical energy from the photosynthetic electron transport chain. For the enzyme-limited case, C_s is given by

$$C_s = C_m + \frac{V_p - A - R_m}{g_{bs}}, \quad (5)$$

where,

$$V_p = C_m V_{\text{pmax}} / \left(C_m + K_{\text{p}}\right), \quad (6)$$

where g_{bs} is the bundle sheath conductance to CO_2, R_m is the mitochondrial respiration in the mesophyll, and K_p is the Michaelis-Menten constant for CO_2 associated with PEP carboxylation. Equations (5) and (6) assume carboxylation of CO_2 by PEP is rate limiting.

The electron transport rate limited CO_2 supply is given by the same equation structure as in (5), but with the "V_p" term replaced:

$$C_s = C_m + \frac{xJ_t/\varphi - A - R_m}{g_{bs}}, \quad (7)$$

where,

$$J_t = \frac{I_2 + J_{\max} - \sqrt{(I_2 + J_{\max})^2 - 4\theta I_2 J_{\max}}}{2\theta}, \quad (8)$$

where x is a partitioning factor of electron transport rate between the C_4 and C_3 cycles (~0.4) and φ is the ATP requirement of the C_4 cycle (~2 ATP). I_2 is the photosynthetically useful light absorbed by PSII ($PS_{\text{abs}} \times$ incident light) and θ is an empirical curvature factor assumed as 0.3 [11].

Equations (3), (4), (7), and (8) were rearranged and fitted to measured Ai curve to infer $J_{\max}$, θ, and PS_{abs}, which were fed into ACi curve fitting using Equations (2), (4), (5), and (6). Overall, this allows prediction of the Rubisco (V_{cmax}), PEP (V_{pmax}), and electron transport ($J_{\max}$) limited CO_2 assimilation. The fitting was performed using the numerical solver option in Excel which minimises the sum of square errors of A between observed and predicted. The Excel spreadsheet for calculation is shown in Table S1, which

shows ACi and Ai fitting with predicted V_{cmax}, V_{pmax}, and J_{max} for plot 272 in TS2.

2.6. Association of Ground Truth Data with Hyperspectral Measurements

2.6.1. Approach 1: Stepwise Multilinear Regression Using the Vegetation Indices. Stepwise regression consists of iteratively adding and removing predictors used in the predictive model, in order to find the subset of variables in the dataset resulting in the best performing model that lowers prediction error. It has been used to select spectral wavelengths highly related to leaf nitrogen, lignin, and cellulose concentrations in diverse species [57, 58]. Stepwise multilinear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data [59]. Input variables (vegetation indices) are eliminated according to the Pearson correlation coefficient with dependent variables (leaf properties and photosynthetic parameters), which should indicate the most relevant indices to photosynthesis. However, stepwise multilinear regression often suffers from multicollinearity existing in the predictors [58, 60]. In this study, before undertaking stepwise multilinear regression, principal component analysis (PCA) was conducted for the set of hyperspectral vegetation indices in Table 2 to reduce collinearities among them. This resulted in a subset of vegetation indices with reduced correlation between each other which were used in stepwise multilinear regression. The wavelengths used to calculate all the vegetation indices in Table 2 and involved in the subset of vegetation indices are indicated in Figure 1(d). Stepwise multilinear regression using the "MASS" package in R (v 4.0.3) [61] was then conducted to detect the best models for photosynthetic parameters (V_{cmax}, V_{pmax}, and J_{max}) and key leaf properties (SLN and LMA). The best models for each trait were selected, based on Akaike's Information Criteria (AIC) which is commonly used in model selection with lower values indicating a more parsimonious model than a model with a higher AIC [62]. Coefficient of determination (R^2) and root mean squared error (RMSE) were used for model assessment.

2.6.2. Approach 2: Partial Least Square Regression (PLSR) Derived from Spectral Reflectance. In this approach, PLSR was used to correlate the spectra reflectance of all available wavelengths with the photosynthetic parameters (V_{cmax}, V_{pmax}, and J_{max}) and key leaf properties (SLN and LMA) across TS1 and TS2. PLSR has been commonly used in remote sensing spectroscopy to predict plant biochemical and physiological parameters, being able to handle highly correlated predictors and the case of more predictors than observations [60, 63, 64]. The "pls" package in R (v 4.0.3) predicted the traits of interest from reflectance of all the 272 wavelengths, via decomposing the predictor matrix into a set of loadings and scores with the objective of maximising covariance between the scores and response [65, 66]. This process is repeated for a given number of latent variables as the number of loadings and scores necessary to explain sufficient variance in response. The optimal number of latent variables was taken as the minimum number required to minimise the root mean squared error of prediction while not significantly decreasing the cross-validation error, with a maximum of 25 latent variables being considered.

The evaluation of the PLSR models was performed by a leave-one-out cross-validation approach, by training the model on all but one observation and then predicting for the remaining observations [67]. The benefit of many iterations of fitting and evaluating during this cross-validation is that it results in a more robust estimate of model performance as each row of data is given an opportunity to represent the entirety of the test dataset, which is appropriate for a small dataset given the computational cost [68, 69]. This cross-validation approach has been applied in remote sensing of wheat leaf area index, maize and tobacco biochemical traits, crop yield forecasting, and poplar tree photosynthetic capacity predicting from spectral measurements [40, 42, 70–73]. The performances of these regression models were assessed using R^2 and RMSE.

2.7. Extrapolating the PLSR Models Built across the Training Sets to the GWAS Trials. To further test the accuracy of the PLSR models built across the training sets, the PLSR models for V_{cmax}, V_{pmax}, J_{max}, SLN, and LMA were used to estimate these traits for each line in the GWAS trials GAT1 and GAT2. Subsequently, GWAS analyses for the two most important photosynthetic parameters (V_{cmax} and J_{max}) in GAT1 were conducted to identify the underlying genetic loci.

2.7.1. BLUPs for the Traits of Interest in the GWAS Trials. To minimise environmental and special effects within trials and perform GWAS, the best linear unbiased predictors (BLUPs) of the predicted traits in the GWAS trials were calculated using a restricted maximum likelihood (REML) by fitting a linear mixed model using the ASReml-R package (Equation (9)) [74, 75].

$$y = X\beta + Zu + \varepsilon, \tag{9}$$

where the response vector y is modelled by all the fixed effects β, random effects u, and all the residual effects ε. The matrix X represents the design matrix for the fixed effects, and the matrix Z is the design matrix for the random effects. The fixed effects were composed of main effects for each trial plus any effects associated with linear changes along the rows and columns. The random effects contained sources of error within each trial including replication and any trial specific random row and column effects. The residual effects included trial specific residual effects and first order autoregressive (AR1) effects in both the row and column directions for each trial. The model included genotype as a random effect to predict genotype BLUPs within trials. All possible sources of variation in the BLUPs were allowed for in the linear mixed model [75]. A generalised measure of heritability was calculated due to the complex variance structure, of which the equation is given by (Equation (10)).

$$H^2 = 1 - \overline{SED}^2 / \left(2\sigma_g^2\right), \quad (10)$$

where H^2 is the generalised heritability, σ_g^2 represents the genetic variance, and $\overline{SED}$ is the average standard error of difference [76].

2.7.2. GWAS for V_{cmax} and J_{max} in the GWAS Trial GAT1. All genotypes from the diversity panel used in the GWAS trial GAT1 were resequenced by Diversity Arrays Technology Pty Ltd (http://www.diversityarrays.com). The sequence data was aligned to version v3.1 of the sorghum reference genome sequence [77] to identify SNPs (Single Nucleotide Polymorphisms), resulting in 414,899 SNPs. GWAS analyses were conducted using BLUPs of V_{cmax} and J_{max} predicted by extrapolating the PLSR models from the training sets to the GWAS trial GAT1. Software FarmCPU [78] was used to conduct GWAS, using 302,631 filtered SNPs (minor allele frequency (MAF) > 0.01). A significant threshold was set as Bonferroni-corrected 0.05/number of effective SNPs [79, 80], resulting in a threshold of p value < 1.6e-7.

2.7.3. Pathway Enrichment Analyses Based on Genes within 200 kb from the QTL of V_{cmax} and J_{max}. To further evaluate the reliability of extrapolating the PLSR models for V_{cmax} and J_{max} from the training sets to the GWAS trials, pathways enriched for genes around the QTL of V_{cmax} and J_{max} were analysed using *PhytoMine* of *Phytozome* v13 (https://phytozome-next.jgi.doe.gov/phytomine/begin.do), by inputting genes within 200 kb of each QTL detected from the Sorghum_bicolor.Sorghum_bicolor_NCBIv3.47.chr.gff3. Genes identified as enriched in the pathways via *PhytoMine* were defined as candidate genes.

3. Results

3.1. Variation in Ground Truth V_{cmax}, V_{pmax}, J_{max}, SLN, and LMA across the Two Training Sets. Substantial variation for all traits measured by ground truthing was observed in the two training sets (Figure 2). In the training set in 2019 (TS1), plot values of V_{cmax} had an average of 51.1 μmol $m^{-2}s^{-1}$ and ranged from 40.3 to 65.5 μmol $m^{-2}s^{-1}$, V_{pmax} varied between 123 and 922 μmol $m^{-2}s^{-1}$ with a mean of 408 μmol $m^{-2}s^{-1}$, and J_{max} had an average of 409 with a range of 280 to 773 μmol $m^{-2}s^{-1}$. In the training set in 2020 (TS2), V_{cmax} varied from 36.8 to 85.6 μmol $m^{-2}s^{-1}$ with a mean of 50.9 μmol $m^{-2}s^{-1}$, V_{pmax} had an average of 410 μmol $m^{-2}s^{-1}$ and ranged from 105 to 952 μmol $m^{-2}s^{-1}$, and J_{max} ranged from 227 to 673 μmol $m^{-2}s^{-1}$ with a mean of 383. No significant differences were observed in the photosynthetic parameters between the training sets in two years (ANOVA, $p > 0.05$), and pooled data of observations from individual plots across TS1 and TS2 were used to enrich the results. With the pooled data, a total of 75 ACi and 75 Ai curves were used for fitting V_{cmax},V_{pmax}, and J_{max}. However, eight ACi curves could not be fitted sensibly with the C_4 photosynthesis model, possibly due to low data quality caused by high air temperature (> 38°C, Table 1). Given the possible errors from confounding environmental factors in the fittings of V_{cmax}, V_{pmax}, and J_{max}, extreme values ($V_{cmax} > 65$, $V_{pmax} > 750$, and $J_{max} > 700$ μmol $m^{-2}s^{-1}$) were treated as outliers and excluded from further analyses as shown in Figures 2(a)–2(c), based on their average values. In total, 67 V_{cmax}, 60 V_{pmax}, and 74 J_{max} plot observations were effective for further analyses.

SLN varied from 1.6 to 2.4 g m^{-2} with a mean of 2.0 g m^{-2} in TS1 and ranged from 1.3 to 2.5 g m^{-2} with a mean of 1.9 g m^{-2} in TS2 (Figure 2(d)). Pooled data across the two training sets was used for the estimation of SLN ($n = 129$ plots) (Table 1). LMA ranged from 36.0 to 63.5 g m^{-2} ($n = 169$ plots) and did not significantly differ between TS1 and TS2 (Figure 2(e)), and data from the two trials were pooled together. No outliers of SLN or LMA were removed from the following analyses, given no extreme values were observed (Figures 2(d) and 2(e)). Thus, in total, 129 SLN and 169 LMA observations were used for association with hyperspectral data.

3.2. Approach 1: Stepwise Multilinear Regression Using the Vegetation Indices. The first two components of the PCA captured about 80% of the variation in the set of indices, showing strong collinearities among them (Figure 3). For example, strong correlations were observed among NDRE, Red_edge, and r740_r700, as indicated by large positive loadings on component 1. Similarly, NDVI highly correlated with several indices, such as r760_r750, r760_r750index, and CVI, indicated by large negative loadings on component 1. To reduce the collinearities, a subset of vegetation indices (Red_edge, CVI, OSAVI, r760, curvature, and PRI) was selected as predictors for the traits of interest in the stepwise multilinear regression models, based on the correlations among the indices and their loadings on the first two principal components (Figure 3).

The best models based on the AIC criteria are given in Table 3. All models were significant ($p < 0.05$) for estimating the photosynthetic parameters, despite the low R^2 of around 0.20 (Table 3). The RMSEs for predicting V_{cmax}, V_{pmax}, and J_{max} were 9%, 35%, and 18% of the mean, respectively, suggesting a modest accuracy in estimations of the photosynthetic parameters from the proximal hyperspectral vegetation indices. Moreover, the vegetation indices detected in the best models for V_{cmax}, V_{pmax}, and J_{max} were mostly based on near infrared (~800 nm), red edge (~710-750 nm), and green (~550 nm) portions of the spectrum (Figure 1(d)), such as CVI, curvature, and OSAVI, which have previously mostly been used as indicators for variation in nitrogen status and canopy size [28–31, 52]. Interestingly, significant association of V_{pmax} and J_{max} with an oxygen-A band based index (r760) was observed, which has been used to predict chlorophyll fluorescence [81], suggesting sensitivity of this region to photosynthesis. An indicator of light use efficiency, PRI (based on 531 and 570 nm), showed a high coefficient in the estimators of V_{cmax}, consistent with the physiological linkages between maximum Rubisco activity and electron transport processes. Red_edge and curvature, known to be sensitive to chlorophyll content [52], were commonly detected in the best stepwise multilinear regression models for SLN and LMA.

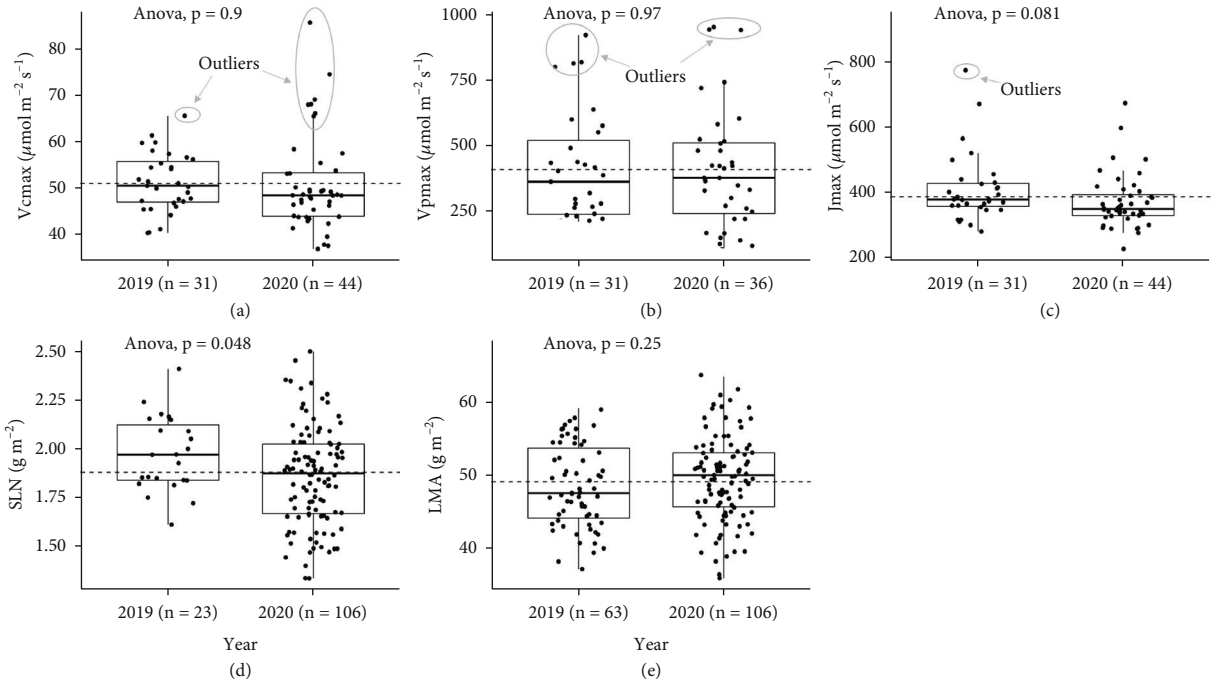

Figure 2: Boxplots showing range of maximal Rubisco carboxylation ((a) V_{cmax}), maximal PEP carboxylation ((b) V_{pmax}), maximal electron transport ((c) J_{max}), specific leaf nitrogen ((d) SLN), and leaf mass per area ((e) LMA) in training set 1 (2019) and training set 2 (2020).

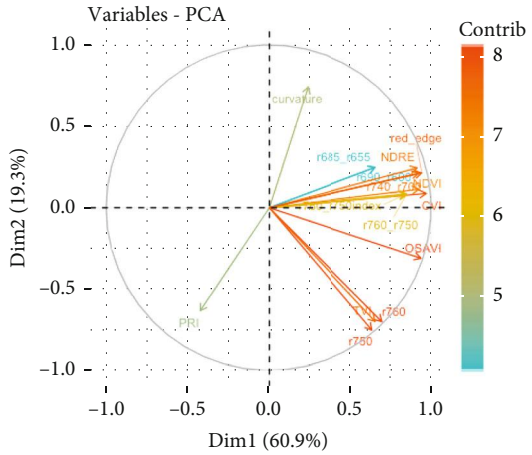

Figure 3: PCA for hyperspectral vegetation indices in Table 1 across two training sets. Note: Length of each arrow represents loading of each variable on dimension 1 and 2; contrib: contribution of each variable to dimension 1 and 2.

3.3. Approach 2: PLSR Derived from Reflectance at All Available Wavelengths. Compared with the stepwise multilinear regression models derived from the set of indices, PLSR using reflectance across all the available wavelengths was much more robust for the estimations of V_{cmax}, V_{pmax}, and J_{max}, with R^2 of 0.83, 0.93, and 0.76, respectively (Figures 4(a)–4(c)). The RMSEs for estimating V_{cmax}, V_{pmax}, and J_{max} were reduced to 4%, 12%, and 10% of the mean, respectively (Figures 4(a)–4(c)). Model loadings, (Figures 4(d)–4(f)) which indicate the contribution of the wavelengths in a specific PLSR model, highlighted the red edge (685-750 nm) and near infrared (a major peak around 950-960 nm) region as important regions for predicting photosynthetic capacity.

Using PLSR derived from reflectance of all wavelengths, the predictions of SLN and LMA improved in both R^2 and RMSE compared with the models developed by stepwise multilinear regression using vegetation indices (Figures 5(a) and 5(b)). For SLN and LMA, the RMSE was reduced to 5% and 6% of the mean, respectively. The R^2 reached 0.82 for SLN and 0.68 for LMA. In the models for SLN and LMA, the wavelengths with high loadings largely fell in the near infrared regions with peaks around 722-769 nm and 922-956 nm (Figures 5(c) and 5(d)).

3.4. Extrapolating the PLSR Models Built Using the Training Sets to the GWAS Trials

3.4.1. Variation and Heritability of V_{cmax}, V_{pmax}, and J_{max}, and SLN and LMA in GAT1 and GAT2. When using the PLSR models built across the two training sets to estimate the traits in the GWAS trials, reasonable ranges and heritability were observed for all the traits, especially for the two key photosynthetic parameters V_{cmax} and J_{max} (Table 4). The ranges of the predicted V_{cmax} (46-65 μmol $m^{-2}s^{-1}$) and J_{max} (317-595 μmol $m^{-2}s^{-1}$) in GAT1 were particularly comparable with the ground truth measurements in the training sets (Figure 2), suggesting a reasonable accuracy of the extrapolations. This was also supported by the high heritability (around 0.90) of V_{cmax} and J_{max} in GAT1 (Table 4). The heritabilities of the predictions in GAT2 were lower than in GAT1, because most of the genotypes in GAT2

Table 3: The best models chosen by AIC in stepwise multilinear regression for traits of interest.

Traits of interest	No. observations	Vegetation indices	Coefficients	R^2	p value	RMSE
		Red_edge	-34.3			
V_{cmax}	67	CVI	2.3	0.3	<0.01	4.7
		PRI	-521.0			
		OSAVI	-5077.0			
V_{pmax}	60	Curvature	6682.0	0.2	<0.05	143.9
		r760	3899.0			
J_{max}	74	r760	-1766.0	0.2	<0.05	75.5
		Red_edge	0.6			
SLN	129	OSAVI	-6.1	0.2	<0.01	0.2
		Curvature	-15.8			
		Red_edge	30.5			
LMA	169	CVI	-2.4	0.2	<0.01	5.0
		Curvature	-258.1			

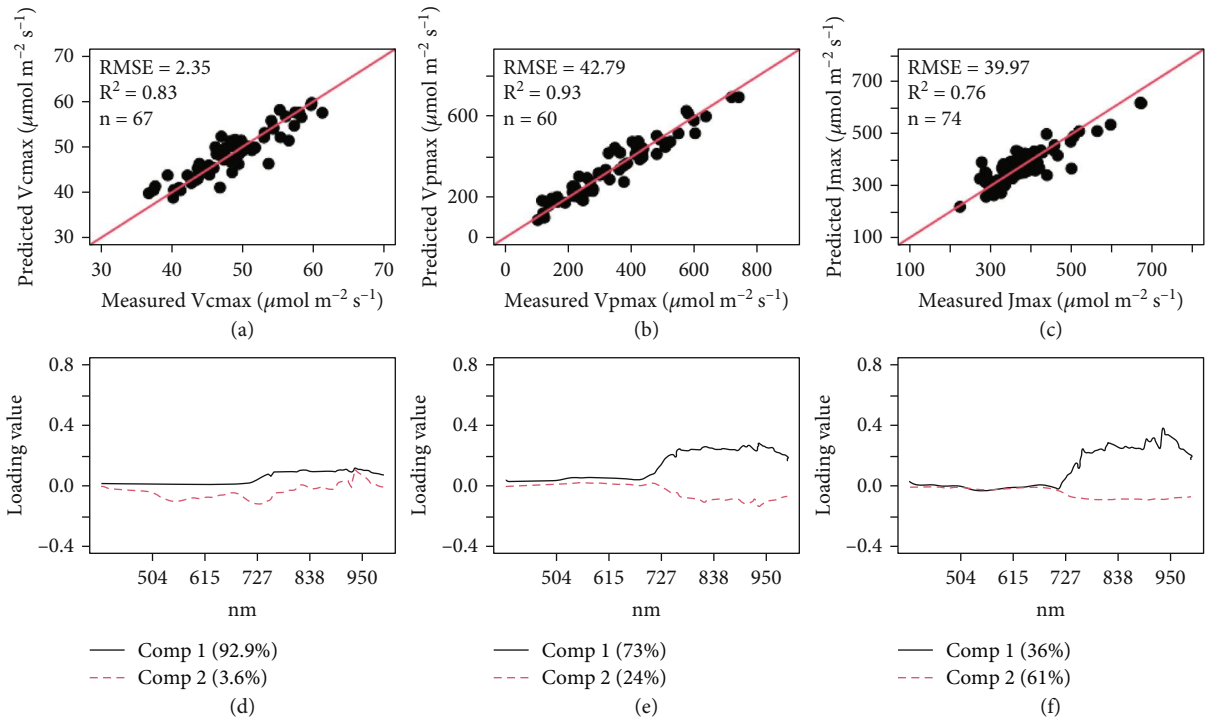

Figure 4: Cross-validated predictions of V_{cmax} (a), V_{pmax} (b), and J_{max} (c) and corresponding loadings with principal components 1 and 2 for V_{cmax} (d), V_{pmax} (e), J_{max}(f) using partial least square regression (PLSR) and reflectance values at various wavelengths between 395 and 997 nm. Note: Bottom panels (d, e, and f) show model loadings which represent the relative importance of a given spectral wavelength in each model (a, b, and c, respectively); values in brackets indicate the percentage of variance explained by the first two principal components.

were hybrids which have less genetic diversity (Tables 1 and 4).

3.4.2. GWAS Based on the Predictions of V_{cmax} and J_{max} in GAT1. To further evaluate the predictivity of the PLSR models, GWAS analyses were performed on BLUPs of V_{cmax} and J_{max} predictions in GAT1 ($n = 649$ inbred lines), and given V_{cmax} and J_{max} have been identified to be the two key photosynthetic parameters for determining net rate of canopy photosynthesis [13]. Four QTL were detected to be associated with the variation in V_{cmax} (Figure 6 and Table 5), were located on chromosome 6, 9, and 10, suggesting likely genomic regions associated with the processes of CO_2 assimilation. In terms of J_{max}, two QTL located on chromosomes 4 and 5 were identified, providing likely chromosomal regions relevant to the processes of electron transport.

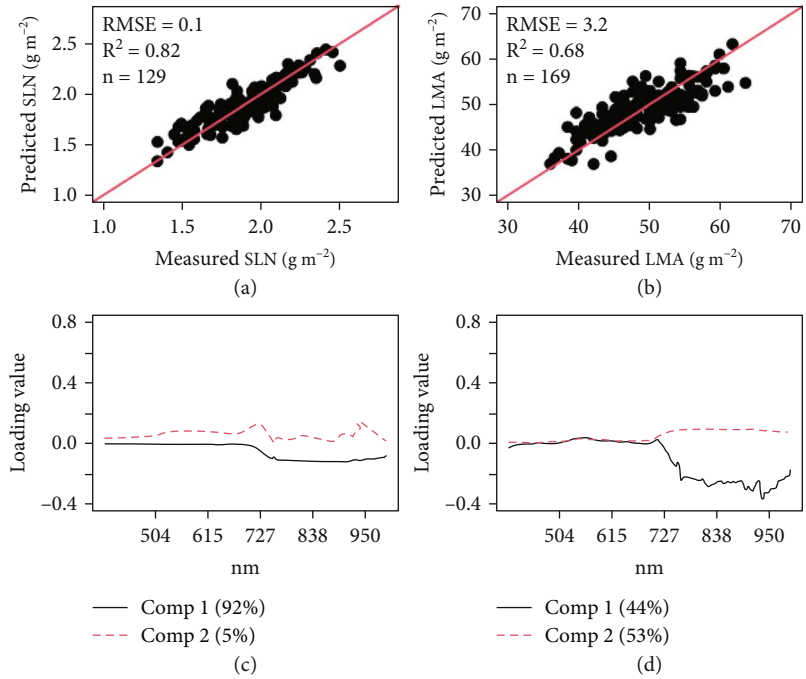

Figure 5: Cross-validated predictions of SLN (a) and LMA (b) and corresponding loadings with principal components 1 and 2 for SLN (c) and LMA (d) using partial least square regression (PLSR) with reflectance values at different wavelengths between 395 and 997 nm. Note: Bottom panels (c and d) show model loadings which represent the relative importance of a given spectral wavelength in each model (a and b, respectively); values in brackets indicate the percentage of variance explained by the first two principal components.

Table 4: Range and heritability of predicted SLN, LMA, V_{cmax}, V_{pmax}, and J_{max} in the GWAS trials.

Site	Trait	Max	Min	Mean	Std.error	H^2
	Pred.SLN	2.6	1.4	2.0	0.1	0.85
	Pred.LMA	73.1	50.3	58.8	2.0	0.69
GAT1	Pred.Vcmax	64.6	45.7	53.8	1.1	0.87
	Pred.Vpmax	811.4	97.7	399.8	35.3	0.89
	Pred.Jmax	595.2	317.2	457.4	17.4	0.90
	Pred.SLN	2.1	1.8	1.9	0.1	0.53
	Pred.LMA	65.7	58.8	61.8	1.8	0.52
GAT2	Pred.Vcmax	48.6	40.2	43.2	0.9	0.73
	Pred.Vpmax	579.5	258.4	406.8	37.7	0.59
	Pred.Jmax	514.6	443.1	472.9	14.3	0.56

Note: Pred.: predictions for traits in the GWAS trials from the PLSR models built using the pooled training sets; V_{cmax} (μmol $m^{-2}s^{-1}$): maximal Rubisco carboxylation; V_{pmax} (μmol $m^{-2}s^{-1}$): maximal PEP carboxylation; J_{max} (μmol $m^{-2}s^{-1}$): maximal electron transport rate; SLN (g m^{-2}): specific leaf nitrogen content: LMA (g m^{-2}): leaf mass per area; H^2: generalised heritability.

3.4.3. Pathways Enriched for Genes within 200 kb from the QTL of V_{cmax} and J_{max}. To further assess the accuracy of the PLSR models from the training sets, the genes within 200 kb [43] from the QTL detected for V_{cmax} and J_{max} were analysed by *PhytoMine* (https://phytozome-next.jgi.doe.gov/). One pathway was enriched for five candidate genes of V_{cmax}, which has been annotated to be associated with UDPG-glucosyl transferase (Table 6). Another pathway, enriched for four candidate genes of J_{max}, was found to be involved in metabolic processes resulting in the removal or addition of electrons (iron ion binding).

4. Discussion

In this study, five key photosynthesis related variables were investigated and predicted from canopy hyperspectral reflectance data, providing an efficient and nondestructive tool to screen genotypes for improved photosynthetic capacity at large scale. Maximal Rubisco carboxylation rate (V_{cmax}), PEP carboxylation rate (V_{pmax}), and electron transport rate (J_{max}), which are the main rate-limiting processes in C_4-carbon assimilation, were quantified in a diverse set of sorghum genotypes across the two training sets ($n = 75$ plots including 63 genotypes). To date, this is the first attempt to correlate hyperspectral reflectance to detailed fittings of these three parameters from both ACi and Ai curves in C_4 pathway photosynthesis. The obtained V_{cmax} and V_{pmax} values were comparable with those reported previously in sorghum [82]. Compared with stepwise multilinear regression, PLSR models improved the prediction accuracy for the three photosynthetic parameters and the other two key leaf properties (SLN and LMA, $n = 169$ plots including 124 genotypes), based on R^2 (~0.80) and RMSE (less than 12% of mean). Subsequently, these PLSR models were extrapolated to two GWAS trials (875 plots with 650 genotypes in GAT1; 912 plots with 634 genotypes in GAT2), with the resulting predictions for both photosynthetic parameters and key leaf properties (SLN and LMA) showing medium to high

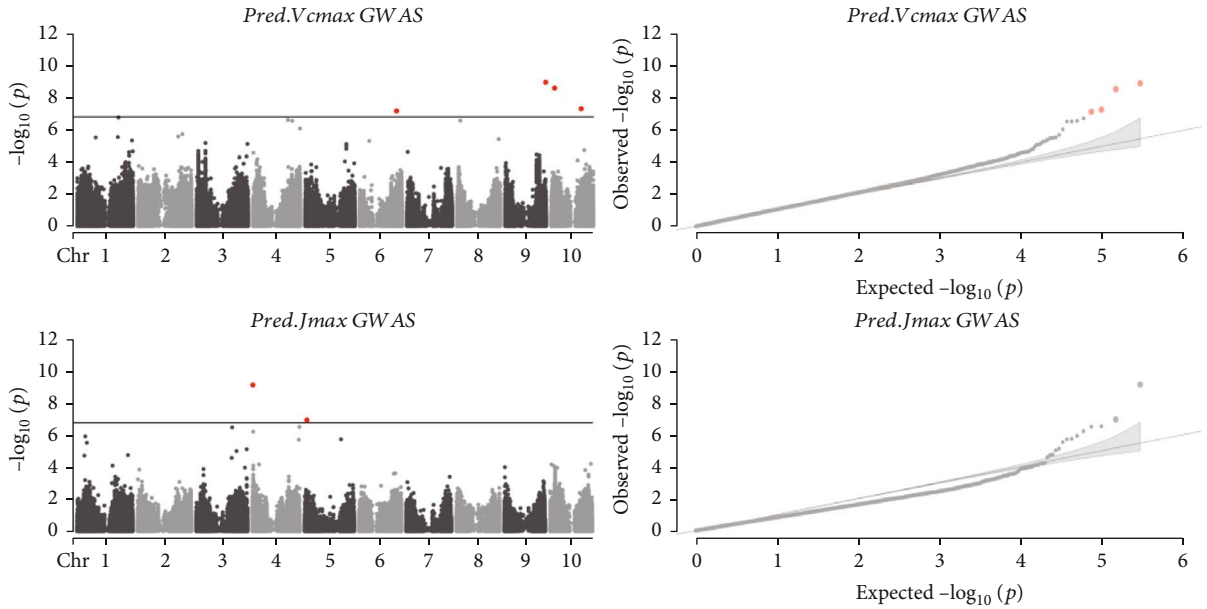

Figure 6: Manhattan and Q-Q plots of GWAS for V_{cmax} and J_{max} in GAT1. Note: Pred.Vcmax: maximal Rubisco carboxylation rate predicted by the PLSR model for V_{cmax} using the pooled training sets; Pred.Jmax: maximal electron transport rate predicted by the PLSR model for J_{max} from the pooled training sets; in the Manhattan plots: solid black line showing Bonferroni corrected significant threshold (p value <1.6e-7); in both Manhattan and Q-Q plots: SNPs in red passed the significant threshold.

Table 5: QTL identified for V_{cmax} and J_{max} in GAT1.

Trait	QTL	Chromosome	Position (bp)	p value	MAF
Pred.Jmax	qJmax4.1	4	747956	6.86E-10	0.3
Pred.Jmax	qJmax5.1	5	3363160	1.08E-07	0.1
Pred.Vcmax	qVcmax6.1	6	53165713	6.76E-08	0.1
Pred.Vcmax	qVcmax9.1	9	58600798	1.12E-09	0.2
Pred.Vcmax	qVcmax10.1	10	5271782	2.56E-09	0.2
Pred.Vcmax	qVcmax10.2	10	43584867	4.95E-08	0.1

Note: Pred.Vcmax: maximal Rubisco carboxylation rate predicted by the PLSR model for Vcmax using the pooled training sets; Pred.Jmax: maximal electron transport rate predicted by the PLSR model for Jmax from the pooled training sets; Position (bp): the physical positions of QTL identified on the sorghum reference genome v3.1; MAF: minor allele frequency.

Table 6: Pathway enrichment analyses for candidate genes within 200 kb from the QTL of V_{cmax} and J_{max}.

Candidate genes	Chr	bp_start	bp_end	Distance to QTL	Closest QTL	Pathway	GO annotation
Sobic.006G174000	6	52,994,972	52,996,405	169,308	qVcmax6.1	PWY-2902	UDPG-glucosyl transferase
Sobic.006G174300	6	53,001,457	53,004,943	160,770			
Sobic.006G174400	6	53,005,433	53,007,221	158,492			
Sobic.006G174500	6	53,012,401	53,014,032	151,681			
Sobic.006G174600	6	53,021,800	53,023,200	142,513			
Sobic.004G008300	4	733,525	734,761	13,195	qJmax4.1	PWY-5129	Iron ion binding, SUR2 and oxidation-reduction process
Sobic.004G008800	4	763,419	764,817	15,463			
Sobic.004G008900	4	769,869	771,113	21,913			
Sobic.004G008200	4	724,142	725,780	22,176			

Note: Chr: chromosome; bp_start: the start point of the gene in the reference genome; bp_end: the end point of the gene in the reference genome; distance to QTL: distance of the gene to the closest QTL in bp; closest QTL: the closest QTL to the candidate gene.

heritability. Furthermore, the genomic regions associated with V_{cmax} and J_{max} that were detected by GWAS in GAT1 ($n = 649$ inbred lines) revealed candidate genes involved in the pathways of UDPG-glucosyl transferase and removal or addition of electrons, respectively.

4.1. Plot-Based Hyperspectral Reflectance Can Be Used to Predict Leaf Photosynthetic Capacity

4.1.1. Models for V_{cmax}, V_{pmax}, and J_{max}. Hyperspectral reflectance using leaf clips has shown promise for predicting photosynthetic capacity in a variety of plant species [42, 69, 70, 83–85]. However, measurements requiring leaf clips are not practical for screening thousands of breeding lines. To fully achieve high-throughput phenotyping, rather than using handheld spectroradiometers on a leaf-by-leaf basis, estimations of photosynthetic capacity from automated proximal or remote sensing at the canopy level are needed. Apart from greater throughput, canopy measurements also better reflect the whole-plant, which integrates photosynthetic activities measured at the leaf level.

Canopy hyperspectral reflectance has shown promise for estimating V_{cmax} and net canopy photosynthetic rate through different approaches, such as airborne-based model inversion in wheat [34]. Another study using canopy hyperspectral reflectance also successfully predicted V_{cmax} and J_{max} with a ground-based phenotyping platform in tobacco [86]. Moreover, these authors compared three different PLSR approaches including reflectance-based, index-based, and model inversion-based methods, indicating better performance in models based on reflectance and indices than model inversion [86]. A comparison based on leaf- and plot-level PLSR models confirmed the capability of plot-level hyperspectral imaging to predict photosynthetic parameters in transgenic tobacco plants expressing C_4 photosynthesis pathway genes [40]. In the present study, across 63 sorghum varieties in the training sets, V_{cmax}, V_{pmax}, and J_{max} were predicted with reasonably high accuracy (R^2 around 0.80 and RMSE within 12% of mean) using PLSR models built from canopy hyperspectral data collected via a proximal phenotyping platform (~1.7 m from canopy). The index-based stepwise multilinear regression models for V_{cmax} and J_{max} could also estimate the photosynthetic parameters with a reasonably small RMSE around 13% of mean, although with much less percentage of variance explained (R^2 around 0.20). The results from the present study demonstrate the promise of utilising hyperspectral sensing at a canopy level in selective breeding for photosynthetic capacity at large scale and put forward a high-throughput tool to explore genotype by environment interactions of photosynthetic capacity related traits.

4.1.2. Models of SLN and LMA. Nitrogen content has been one of the most successfully predicted traits in crops from both leaf and canopy spectral measurements [20, 87, 88]. In addition to the biochemical parameters, and given the strong associations of nitrogen and LMA with photosynthesis, remote sensing of the key leaf properties has also previously been explored, [16, 89, 90]. Among the estimations from PLSR models in this study, a high coefficient of determination was consistently observed in SLN predictions ($R^2 = 0.82$), which also had a low RMSE in stepwise multilinear models (10% of mean SLN), demonstrating the effectiveness and suitability of approaches applied in this study.

Another key leaf property, LMA, has been identified as a proxy of photosynthetic capacity in maize [89]. Robust models for predicting LMA from leaf-level hyperspectral reflectance have been reported for wheat and soybean [18, 35, 91]. Additionally, lower RMSE at canopy level than leaf level has been reported for LMA estimations, as multiple scattering in the upper canopy leaf layers could strengthen the expression of key leaf properties in a closed canopy compared with leaf-level measurements [83]. A more recent study in the C_3 crop zucchini using both leaf- and canopy-level hyperspectral reflectance and PLSR has successfully predicted LMA with R^2 of 0.91 and 0.60, respectively [92]. In the present study, low RMSE (6% of mean LMA) and medium to high R^2 of 0.68 were found in the LMA estimations from canopy hyperspectral reflectance using PLSR. This was also supported by LMA predictions from the stepwise multilinear regression with an acceptable RMSE, 10% of mean. These results indicate that proximally sensed and canopy-based hyperspectral reflectance measurements provide a rapid and robust measure of key leaf properties related to photosynthetic efficiency.

4.1.3. Potential Strategies to Train Robust Models for Predicting Leaf Traits from Canopy-Based Sensing. When using canopy-level hyperspectral data to train leaf-level measurements, shadows, soil background, and canopy structure could be complicating factors that affect the robustness of the model. To address some of the issues with using canopy reflectance, a NDVI > 0.5 mask was applied to each pixel used in the reflectance calculation. This masked out the soil background reflectance and thus minimising the variation in spectral responses from effects associated with canopy heterogeneity (e.g., light or temperature) at the plot level. In addition, some of the noise from canopy structural factors was also minimised in this study by operating within one critical growth stage. However, for future application, developing models suitable for different stages or less sensitive to the variation of canopy structure within a time window would improve utility of the method developed here. Additionally, an automatic thresholding technique (e.g., Otsu) fused with canopy height from LiDAR could be applied in canopy delineation which should be more accurate [93] in delineating the exact canopy areas within a plot. This could reduce spurious reflectance values and thus increase the signal measured from proximal sensing at the canopy level, depending on agricultural contexts (e.g., species or canopy size). Alternatively, combining relevant models that improve the relationships between canopy hyperspectral reflectance and leaf photosynthetic parameters could be useful [34]. Increasing the number of ground truth samples can also improve model performance; however, simply increasing the size of the dataset not only leads to highly complex models but is also affected by the high costs associated with additional measurements, especially in the case of gas exchange

measurements which are notoriously slow to obtain [25, 94]. To date, gas exchange measurements are the only realistic measurement of photosynthesis; however, given the confounding factor of variation in photosynthetic capacity within crop canopies of the same genotype [40], this is not ideal.

Reducing the confounding environmental factors (e.g., light or temperature) will also improve model strength when using canopy-based hyperspectral sensing methods to estimate key leaf traits. In this study, all ground truth and sensing data was collected between 9 am and 12 pm, which minimised the effects of sun angle, temperature, and light on canopy reflectance and on photosynthetic rates. Further improvement could be made by incorporating temperature at the time of image capture and tentatively correcting photosynthetic parameters to a standard temperature, as it is one of the most important environmental factors influencing both hyperspectral reflectance and photosynthesis. This was not considered here due to scarce documentation of temperature responses of Vcmax, Vpmax and Jmax in C_4 crops [39].

4.2. PLSR Derived from Entire Wavelength Spectrum Strengthens Model Performance. Compared with the models developed using stepwise multilinear regression, PLSR models were more robust and demonstrated a higher cross validated R^2 and lower RMSE. This is attributed to the fact that additional spectral information was incorporated in the PLSR models using the complete wavelength range compared with the stepwise multilinear regression models [36, 63, 83, 86]. Based on peak loadings (red edge and near infrared), the wavelengths that explained most of the variance in the PLSR models aligned closely with the locations of the wavelength bands selected to develop the best-performing multilinear vegetation index approach. Compared with the published indices that correlate with nitrogen content, a strong overlap was found around the red edge (~710-750 nm) in the present study, consistent with the finding that leaf nitrogen content is linearly correlated with the first derivatives of reflectance at the red edge region around 730 nm [20]. The most important parts of the spectrum for predicting photosynthesis have been shown to be in the visible (400-700 nm) and red edge (710-750 nm) range [83]. In this study, the spectral loadings used to predict photosynthetic parameters had similar peaks to the spectral loadings of SLN and LMA, likely attributed to these features being interdependent [89]. These results provide useful information for selecting relevant wavelengths to predict the traits of interest in further studies.

4.3. PLSR Models Built across the Training Sets Can Be Extrapolated to the GWAS Trials. In this study, the PLSR models were extrapolated to the GWAS trials, demonstrating comparable variation and high heritability (~0.80) for the predicted biochemical (V_{cmax}, V_{pmax}, and J_{max}) and key leaf properties (SLN and LMA) in the GWAS trial (GAT1), including predominantly inbred lines. Based on the predictions for these traits in the GWAS trial (GAT2), comprising mostly hybrid lines, relatively lower heritability (~0.5) was observed, as expected, due to similarity among the hybrids both at the molecular and phenotypic level. This suggests hyperspectral sensing is a promising avenue to screen large populations for such traits that have previously been out of reach of crop breeding programs [34, 42, 95]. However, the capacity of green leaves to convert CO_2 into biomass varies throughout the season mainly due to interactions among genotypes, plant phenological stage and environment [13, 96]. This is likely to further influence predictive skill especially in cases where there is a high within-population variation as a result of the genotype by environment interactions.

The models built across the training sets show sufficient skill to estimate key determinants of photosynthesis in large sorghum mapping populations, grown adjacent to these ground-truth trials, despite potential challenges of predicting leaf photosynthetic capacity from canopy-based hyperspectral sensing. This would not only enable the screening for materials with improved photosynthetic capacity, following identification of genetic loci and potential candidate genes for photosynthetic capacity in the C_4 crop sorghum but also benefit the quantification of the association between photosynthetic capacity and ultimate biomass improvement in crops. In further applications, it is important to select the best phenology stage for data collection, when the degree of canopy development expressed by leaf area index has more consistent levels of pigment concentration per unit area and more similar spectral response for reducing the impact of such confounding effects associated with plant growth processes (e.g., canopy structure and nitrogen status), [31]. Additionally, further studies to test temporal stability of relationships between canopy reflectance spectra and leaf photosynthetic capacity are needed before extrapolated associations from a specific hyperspectral measurement through the growing season can be made in other crops or agricultural contexts.

Here, GWAS analyses for the two photosynthetic parameters, V_{cmax} and J_{max}, provided useful information for further fine mapping to identify potential candidate genes controlling CO_2 assimilation and electron transport in sorghum. This is one of the significant and novel outcomes from this study, as this is the first attempt to quantify the genetic basis of the key photosynthetic parameters using hyperspectral sensing in hundreds of lines. Additionally, pathway enrichment analysis for genes within 200 kb from J_{max} QTL detected four candidate genes involved in the process of electron transport and light signalling [97]. This means the PLSR model for J_{max} built across the training sets was able to capture the genomic loci associated with its phenotypic variation in the sorghum diversity panel. The pathway enriched for genes within 200 kb from the V_{cmax} QTL is known to catalyse the transfer of a hexosyl group from one compound to another, as well as function in nitrogen storage [98]. While this is not directly associated with Rubisco activity *per se*, plant nitrogen status is closely associated with Rubisco and leaf photosynthetic rates [18–20]. Additionally, the photosynthetic capacity is colimited by Rubisco activity (V_{cmax}) and RuBP regeneration, which depends on electron transport (J_{max}) and the coordination of Calvin cycle enzymes [11, 93]. Enzyme interactions in the Calvin cycle are highly complex [92], and further studies

are needed to explore the relevance of the V_{cmax} QTL detected here.

5. Conclusions

Being able to map crop traits associated with improved resource use efficiency (e.g., nitrogen, light, and water) will contribute to further understanding of the natural variation in photosynthetic processes and enable the exploration of opportunities to modify photosynthesis. This study developed a model using PLSR to estimate maximal Rubisco activities (V_{cmax}, $R^2 = 0.83$), maximal PEP activities (V_{pmax}, $R^2 = 0.93$), maximal electron transport activities (J_{max}, $R^2 = 0.76$), specific leaf nitrogen (SLN, $R^2 = 0.82$), and leaf mass per leaf area (LMA, $R^2 = 0.68$) from proximal hyperspectral sensing using two combined training sets ($n = 169$ plots). Further, extrapolating the PLSR models built across the training sets to the GWAS trials including hundreds of lines demonstrates that the predictions of the traits of interest are heritable. GWAS analyses for J_{max} in the inbred lines detected genomic regions comprising candidate genes controlling the process of electron transport. While the V_{cmax} candidate genes identified here are not associated directly with Rubisco activity *per se*, they are involved in nitrogen storage which is closely associated with Rubisco. These results suggest that the PLSR models from the training sets were able to capture the phenotypic variation in the photosynthetic parameters allowing the discovery of the underlying genetic basis of these important traits.

Data Availability

All phenotypic data used to develop the models presented in this manuscript is available here: https://doi.org/10.48610/acbe0df. Genotypic marker data used for GWAS is available upon request to the corresponding author.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this article.

Authors' Contributions

B.G.J., X.Z., A.P., and G.H. conceived the research plans; X.Z., S.M.R., and B.G.J. performed the experiments; A.W., S.M.R., S.C., G.H., C.H., and A.P. provided technical assistance to X.Z.; X.Z. analyzed the data and wrote the article; and all authors provided input into the interpretation of results and the final version of the article.

Acknowledgments

We would like to thank Glen Roulston, Kate Jordan, Janet Roberts, Jane Heron, and James Heron for assistance with data collection and the farm staff at Gatton Research Facilities and staff from the Queensland prebreeding for experiment management and Prof Susanne von Caemmerer for advice with collecting and interpreting gas exchange data. X. Z. was financially supported through a University of Queensland Research Training Scholarship. This study was partially funded by the Centre of Excellence for Translational Photosynthesis, Australian Research Council (grant CE140100015) and the Bill & Melinda Gates Foundation (grant OPPGD1197 iMashilla "A targeted approach to sorghum improvement in moisture stress areas of Ethiopia").

Supplementary Materials

Table S1: to show the Excel spreadsheet for ACi and Ai fitting with predicted V_{cmax}, V_{pmax}, and J_{max} for plot 272 in TS2. *(Supplementary Materials)*

References

[1] P. S. Belton and J. R. N. Taylor, "Sorghum and millets: protein sources for Africa," *Trends in Food Science & Technology*, vol. 15, no. 2, pp. 94–98, 2004.

[2] J. R. Evans, "Improving photosynthesis," *Plant Physiology*, vol. 162, no. 4, pp. 1780–1793, 2013.

[3] R. T. Furbank, R. Sharwood, G. M. Estavillo, V. Silva-Perez, and A. G. Condon, "Photons to food: genetic improvement of cereal crop photosynthesis," *Journal of Experimental Botany*, vol. 71, no. 7, pp. 2226–2238, 2020.

[4] C. Ishikawa, T. Hatanaka, S. Misoo, C. Miyake, and H. Fukayama, "Functional incorporation of sorghum small subunit increases the catalytic turnover rate of rubisco in transgenic Rice," *Plant Physiology*, vol. 156, no. 3, pp. 1603–1611, 2011.

[5] FAO, *Global agriculture towards 2050*, High Level Expert Forum-How Feed World, 2009.

[6] B. I. Haussmann, H. Fred Rattunde, E. Weltzien-Rattunde, P. S. Traoré, K. Vom Brocke, and H. K. Parzies, "Breeding strategies for adaptation of pearl millet and sorghum to climate variability and change in West Africa," *Journal of Agronomy and Crop Science*, vol. 198, no. 5, pp. 327–339, 2012.

[7] R. E. Boyles, Z. W. Brenton, and S. Kresovich, "Genetic and genomic resources of sorghum to connect genotype with phenotype in contrasting environments," *The Plant Journal*, vol. 97, no. 1, pp. 19–39, 2019.

[8] A. Wu, G. L. Hammer, A. Doherty, S. Caemmerervon, and G. D. Farquhar, "Quantifying impacts of enhancing photosynthesis on crop yield," *Nature Plants*, vol. 5, no. 4, pp. 380–388, 2019.

[9] J. Ehleringer and R. W. Pearcy, "Variation in quantum yield for CO_2Uptake among C3and C4Plants," *Plant Physiology*, vol. 73, no. 3, pp. 555–559, 1983.

[10] M. D. Hatch, "C4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure," *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics*, vol. 895, no. 2, pp. 81–106, 1987.

[11] S. von Caemmerer, *Biochemical Models of Leaf Photosynthesis*, Csiro Publishing, 2000.

[12] R. F. Sage, "The evolution of C4photosynthesis," *New Phytologist*, vol. 161, no. 2, pp. 341–370, 2004.

[13] A. Wu, A. Doherty, G. D. Farquhar, and G. L. Hammer, "Simulating daily field crop canopy photosynthesis: an integrated software package," *Functional Plant Biology*, vol. 45, no. 3, pp. 362–377, 2017.

[14] S. von Caemmerer and R. T. Furbank, *Modeling C4 photosynthesis*, C4 plant biology, 1999.

[15] I. J. Wright, P. B. Reich, M. Westoby et al., "The worldwide leaf economics spectrum," *Nature*, vol. 428, no. 6985, pp. 821–827, 2004.

[16] P. B. Reich, D. S. Ellsworth, and M. B. Walters, "Leaf structure (specific leaf area) modulates photosynthesis–nitrogen relations: evidence from within and across species and functional groups," *Functional Ecology*, vol. 12, no. 6, pp. 948–958, 1998.

[17] T. R. Sinclair and T. Horie, "Leaf nitrogen, photosynthesis, and crop radiation use efficiency: a review," *Crop science*, vol. 29, no. 1, pp. 90–98, 1989.

[18] M. Ecarnot, F. Compan, and P. Roumet, "Assessing leaf nitrogen content and leaf mass per unit area of wheat in the field throughout plant cycle with a portable spectrometer," *Field Crops Research*, vol. 140, pp. 44–50, 2013.

[19] A. L. Fletcher, P. R. Johnstone, E. Chakwizira, and H. E. Brown, "Radiation capture and radiation use efficiency in response to N supply for crop species with contrasting canopies," *Field Crops Research*, vol. 150, pp. 126–134, 2013.

[20] D. Zhao, K. R. Reddy, V. G. Kakani, and V. R. Reddy, "Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of sorghum," *European Journal of Agronomy*, vol. 22, no. 4, pp. 391–403, 2005.

[21] G. L. Hammer and G. C. Wright, "A theoretical analysis of nitrogen and radiation effects on radiation use efficiency in peanut," *Australian Journal of Agricultural Research*, vol. 45, no. 3, pp. 575–589, 1994.

[22] A. K. Borrell and G. L. Hammer, "Nitrogen dynamics and the physiological basis of stay-green in sorghum," *Crop Science*, vol. 40, no. 5, pp. 1295–1307, 2000.

[23] M. Kitao, Y. Yasuda, E. Kodani et al., "Integration of electron flow partitioning improves estimation of photosynthetic rate under various environmental conditions based on chlorophyll fluorescence," *Remote Sensing of Environment*, vol. 254, article 112273, 2021.

[24] E. Piegari, J. Gossn, F. Grings et al., "Estimation of leaf area index and leaf chlorophyll content in *Sporobolus densiflorus* using hyperspectral measurements and PROSAIL model simulations," *International Journal of Remote Sensing*, vol. 42, no. 4, pp. 1181–1200, 2021.

[25] L. M. York, "Functional phenomics: an emerging field integrating high-throughput phenotyping, physiology, and bioinformatics," *Journal of Experimental Botany*, vol. 70, no. 2, pp. 379–386, 2019.

[26] Y. Zhang, M. Migliavacca, J. Penuelas, and W. Ju, "Advances in hyperspectral remote sensing of vegetation traits and functions," *Remote Sensing of Environment*, vol. 252, article 112121, 2021.

[27] A. B. Potgieter, B. George-Jaeggli, S. C. Chapman et al., "Multi-Spectral Imaging from an Unmanned Aerial Vehicle Enables the Assessment of Seasonal Leaf Area Dynamics of Sorghum Breeding Lines," *Frontiers in Plant Science*, vol. 8, p. 8, 2017.

[28] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sensing of Environment*, vol. 8, no. 2, pp. 127–150, 1979.

[29] G. Rondeaux, M. Steven, and F. Baret, "Optimization of soil-adjusted vegetation indices," *Remote Sensing of Environment*, vol. 55, no. 2, pp. 95–107, 1996.

[30] G. J. Fitzgerald, D. Rodriguez, L. K. Christensen, R. Belford, V. O. Sadras, and T. R. Clarke, "Spectral and thermal sensing for nitrogen and water status in rainfed and irrigated wheat environments," *Precision Agriculture*, vol. 7, no. 4, pp. 233–248, 2006.

[31] M. Vincini, E. Frazzi, and P. D'Alessio, "A broad-band leaf chlorophyll vegetation index at the canopy scale," *Precision Agriculture*, vol. 9, no. 5, pp. 303–319, 2008.

[32] Y. Miao, F. Yuan, S. Yue et al., "Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices," *Field Crops Research*, vol. 157, pp. 111–123, 2014.

[33] P. S. Thenkabail, R. B. Smith, and E. De Pauw, "Hyperspectral vegetation indices and their relationships with agricultural crop characteristics," *Remote Sensing of Environment*, vol. 71, no. 2, pp. 158–182, 2000.

[34] C. Camino, V. Gonzalez-Dugo, P. Hernandez, and P. J. Zarco-Tejada, "Radiative transfer Vcmax estimation from hyperspectral imagery and SIF retrievals to assess photosynthetic performance in rainfed and irrigated plant phenotyping trials," *Remote Sensing of Environment*, vol. 231, article 111186, 2019.

[35] V. Silva-Perez, G. Molero, S. P. Serbin et al., "Hyperspectral reflectance as a tool to measure biochemical and physiological traits in wheat," *Journal of Experimental Botany*, vol. 69, no. 3, pp. 483–496, 2018.

[36] V. Sobejano-Paz, T. N. Mikkelsen, A. Baum et al., "Hyperspectral and thermal sensing of stomatal conductance, transpiration, and photosynthesis for soybean and maize under drought," *Remote Sensing*, vol. 12, no. 19, p. 3182, 2020.

[37] K. Meacham-Hensold, C. M. Montes, J. Wu et al., "High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity," *Remote Sensing of Environment*, vol. 231, article 111176, 2019.

[38] C. Camino, V. Gonzalez-Dugo, P. Hernandez, J. C. Sillero, and P. J. Zarco-Tejada, "Improved nitrogen retrievals with airborne-derived fluorescence and plant traits quantified from VNIR-SWIR hyperspectral imagery in the context of precision agriculture," *International Journal of Applied Earth Observation and Geoinformation*, vol. 70, pp. 105–117, 2018.

[39] H. A. Khan, Y. Nakamura, R. T. Furbank, and J. R. Evans, "Effect of leaf temperature on the estimation of photosynthetic and other traits of wheat leaves from hyperspectral reflectance," *Journal of Experimental Botany*, vol. 72, no. 4, pp. 1271–1281, 2021.

[40] K. Meacham-Hensold, P. Fu, J. Wu et al., "Plot-level rapid screening for photosynthetic parameters using proximal hyperspectral imaging," *Journal of Experimental Botany*, vol. 71, no. 7, pp. 2312–2328, 2020.

[41] N. Vilfan, C. Tolvan der, and W. Verhoef, "Estimating photosynthetic capacity from leaf reflectance and Chl fluorescence by coupling radiative transfer to a model for photosynthesis," *New Phytologist*, vol. 223, no. 1, pp. 487–500, 2019.

[42] C. R. Yendrek, T. Tomaz, C. M. Montes et al., "High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance," *Plant Physiology*, vol. 173, no. 1, pp. 614–626, 2017.

[43] Y. Tao, X. Zhao, X. Wang et al., "Large-scale GWAS in sorghum reveals common genetic control of grain size among cereals," *Plant Biotechnology Journal*, vol. 18, no. 4, pp. 1093–1105, 2020.

[44] F. Tian, P. J. Bradbury, P. J. Brown et al., "Genome-wide association study of leaf architecture in the maize nested association mapping population," *Nature Genetics*, vol. 43, no. 2, pp. 159–162, 2011.

[45] Y. Xiao, H. Liu, L. Wu, M. Warburton, and J. Yan, "Genome-wide association studies in maize: praise and stargaze," *Molecular Plant*, vol. 10, no. 3, pp. 359–374, 2017.

[46] E. J. van Oosterom and G. L. Hammer, "Determination of grain number in sorghum," *Field Crops Research*, vol. 108, no. 3, pp. 259–268, 2008.

[47] E. J. van Oosterom, S. C. Chapman, A. K. Borrell, I. J. Broad, and G. L. Hammer, "Functional dynamics of the nitrogen balance of sorghum. II. Grain filling period," *Field Crops Research*, vol. 115, no. 1, pp. 29–38, 2010.

[48] A. B. Potgieter, J. Watson, M. Eldridge et al., "Determining crop growth dynamics in sorghum breeding trials through remote and proximal sensing technologies," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 8244–8247, Valencia, Spain, 2018.

[49] S. P. Serbin, A. Singh, A. R. Desai et al., "Remotely estimating photosynthetic capacity, and its response to temperature, in vegetation canopies using imaging spectroscopy," *Remote Sensing of Environment*, vol. 167, pp. 78–87, 2015.

[50] A. Tillack, A. Clasen, B. Kleinschmit, and M. Förster, "Estimation of the seasonal leaf area index in an alluvial forest using high-resolution satellite-based vegetation indices," *Remote Sensing of Environment*, vol. 141, pp. 52–63, 2014.

[51] J. A. Gamon, J. Penuelas, and C. B. Field, "A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency," *Remote Sensing of environment*, vol. 41, no. 1, pp. 35–44, 1992.

[52] P. J. Zarco-Tejada, J. A. J. Berni, L. Suárez, G. Sepulcre-Cantó, F. Morales, and J. R. Miller, "Imaging chlorophyll fluorescence with an airborne narrow-band multispectral camera for vegetation stress detection," *Remote Sensing of Environment*, vol. 113, no. 6, pp. 1262–1275, 2009.

[53] M. Meroni, M. Rossini, L. Luis Guanter, U. Alonso, R. C. Rascher, and J. Moreno, "Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications," *Remote Sensing of Environment*, vol. 113, no. 10, pp. 2037–2051, 2009.

[54] Q. Xie, J. Dash, W. Huang et al., "Vegetation Indices Combining the Red and Red-Edge Spectral Information for Leaf Area Index Retrieval," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 5, pp. 1482–1493, 2018.

[55] O. Perez-Priego, P. J. Zarco-Tejada, J. R. Miller, G. Sepulcre-Canto, and E. Fereres, "Detection of water stress in orchard trees with a high-resolution spectrometer through chlorophyll fluorescence in-filling of the O/sub 2/-A band," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 12, pp. 2860–2869, 2005.

[56] A. J. Richardson and C. L. Wiegand, "Distinguishing vegetation from soil background information," *Photogrammetric Engineering and Remote Sensing*, vol. 43, no. 12, pp. 1541–1552, 1977.

[57] R. F. Kokaly and R. N. Clark, "Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression," *Remote Sensing of Environment*, vol. 67, no. 3, pp. 267–287, 1999.

[58] O. Mutanga, A. K. Skidmore, and H. H. T. Prins, "Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features," *Remote Sensing of Environment*, vol. 89, no. 3, pp. 393–408, 2004.

[59] O. Satir and S. Berberoglu, "Crop yield prediction under soil salinity using satellite derived vegetation indices," *Field Crops Research*, vol. 192, pp. 134–143, 2016.

[60] R. Darvishzadeh, A. Skidmore, M. Schlerf, C. Atzberger, F. Corsi, and M. Cho, "LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 63, no. 4, pp. 409–426, 2008.

[61] W. N. Venables and B. D. Ripley, *Modern applied statistics*, S. Fourth, Ed., Springer, New York, 2002.

[62] T. Yamashita, K. Yamashita, and R. Kamimura, "A stepwise AIC method for variable selection in linear regression," *Communications in Statistics-Theory and Methods*, vol. 36, no. 13, pp. 2395–2403, 2007.

[63] M. A. Cho, A. Skidmore, F. Corsi, S. E. van Wieren, and I. Sobhan, "Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression," *International Journal of Applied Earth Observation and Geoinformation*, vol. 9, no. 4, pp. 414–424, 2007.

[64] X. Li, Y. Zhang, Y. Bao et al., "Exploring the best hyperspectral features for LAI estimation using partial least squares regression," *Remote Sensing*, vol. 6, no. 7, pp. 6221–6241, 2014.

[65] A. Singh, S. P. Serbin, B. E. McNeil, C. C. Kingdon, and P. A. Townsend, "Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties," *Ecological Applications*, vol. 25, no. 8, pp. 2180–2197, 2015.

[66] S. Wold, A. Ruhe, H. Wold, and I. W. J. Dunn, "The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses," *SIAM Journal on Scientific and Statistical Computing*, vol. 5, no. 3, pp. 735–743, 1984.

[67] B. Efron and G. Gong, "A leisurely look at the bootstrap, the jackknife, and cross-validation," *The American Statistician*, vol. 37, pp. 36–48, 1983.

[68] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer New York, New York, NY, 2013.

[69] M. Shu, M. Shen, J. Zuo et al., "The Application of UAV-Based Hyperspectral Imaging to Estimate Crop Traits in Maize Inbred Lines," *Plant Phenomics*, vol. 2021, article 9890745, pp. 1–14, 2021.

[70] M. L. Barnes, D. D. Breshears, D. J. Law et al., "Beyond greenness: detecting temporal changes in photosynthetic capacity with hyperspectral reflectance data," *PLoS One*, vol. 12, no. 12, article e0189539, 2017.

[71] A. B. Potgieter, Y. L. Everingham, and G. L. Hammer, "On measuring quality of a probabilistic commodity forecast for a system that incorporates seasonal climate forecasts," *International Journal of Climatology*, vol. 23, no. 10, pp. 1195–1210, 2003.

[72] A. B. Potgieter, G. L. Hammer, A. Doherty, and P. de Voil, "A simple regional-scale model for forecasting sorghum yield across North-Eastern Australia," *Agricultural and Forest Meteorology*, vol. 132, pp. 143–153, 2005.

[73] B. Siegmann and T. Jarmer, "Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data," *International*

Journal of Remote Sensing, vol. 36, no. 18, pp. 4519–4534, 2015.

[74] D. G. Butler, B. R. Cullis, A. R. Gilmour, and B. J. Gogel, *ASReml-R 4 Reference Manual: Mixed Models for S Language Environments*, Queensland Department of Primary Industries and Fisheries, 2018.

[75] A. R. Gilmour, B. R. Cullis, A. P. Verbyla, and A. P. Verbyla, "Accounting for natural and extraneous variation in the analysis of field experiments," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 2, no. 3, pp. 269–293, 1997.

[76] B. R. Cullis, A. B. Smith, and N. E. Coombes, "On the design of early generation variety trials with correlated data," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 11, no. 4, pp. 381–393, 2006.

[77] R. F. McCormick, S. K. Truong, A. Sreedasyam et al., "The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization," *The Plant Journal*, vol. 93, no. 2, pp. 338–354, 2018.

[78] X. Liu, M. Huang, B. Fan, E. S. Buckler, and Z. Zhang, "Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies," *PLoS Genetics*, vol. 12, no. 2, article e1005767, 2016.

[79] P. Duggal, E. M. Gillanders, T. N. Holmes, and J. E. Bailey-Wilson, "Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies," *BMC Genomics*, vol. 9, no. 1, p. 516, 2008.

[80] M.-X. Li, J. M. Y. Yeung, S. S. Cherny, and P. C. Sham, "Evaluating the effective numbers of independent tests and significant p -value thresholds in commercial genotyping arrays and public imputation reference datasets," *Human Genetics*, vol. 131, no. 5, pp. 747–756, 2012.

[81] I. Moya, L. Camenen, S. Evain et al., "A new instrument for passive remote sensing: 1. Measurements of sunlight-induced chlorophyll fluorescence," *Remote Sensing of Environment*, vol. 91, pp. 186–197, 2004.

[82] B. V. Sonawane, R. E. Sharwood, S. von Caemmerer, S. M. Whitney, and O. Ghannoum, "Short-term thermal photosynthetic responses of C4 grasses are independent of the biochemical subtype," *Journal of Experimental Botany*, vol. 68, no. 20, pp. 5583–5597, 2017.

[83] C. E. Doughty, G. P. Asner, and R. E. Martin, "Predicting tropical plant physiology from leaf and canopy spectroscopy," *Oecologia*, vol. 165, no. 2, pp. 289–299, 2011.

[84] D. Heckmann, U. Schlüter, and A. P. M. Weber, "Machine learning techniques for predicting crop photosynthetic capacity from leaf reflectance spectra," *Molecular Plant*, vol. 10, no. 6, pp. 878–890, 2017.

[85] S. P. Serbin, D. N. Dillaway, E. L. Kruger, and P. A. Townsend, "Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature," *Journal of Experimental Botany*, vol. 63, no. 1, pp. 489–502, 2012.

[86] P. Fu, K. Meacham-Hensold, K. Guan, J. Wu, and C. Bernacchi, "Estimating photosynthetic traits from reflectance spectra: a synthesis of spectral indices, numerical inversion, and partial least square regression," *Plant, Cell & Environment*, vol. 43, no. 5, pp. 1241–1258, 2020.

[87] T. M. Blackmer, J. S. Schepers, G. E. Varvel, and E. A. Walter-Shea, "Nitrogen deficiency detection using reflected shortwave radiation from irrigated corn canopies," *Agronomy Journal*, vol. 88, no. 1, pp. 1–5, 1996.

[88] B. L. Ma, M. J. Morrison, and L. M. Dwyer, "Canopy light reflectance and field greenness to assess nitrogen fertilization and yield of maize," *Agronomy Journal*, vol. 88, no. 6, pp. 915–920, 1996.

[89] G. L. Miner and W. L. Bauerle, "Seasonal responses of photosynthetic parameters in maize and sunflower and their relationship with leaf functional traits," *Plant, Cell & Environment*, vol. 42, no. 5, pp. 1561–1574, 2019.

[90] P. B. Reich and M. B. Walters, "Photosynthesis-nitrogen relations in Amazonian tree species. II. Variation in nitrogen Vis-a-Vis Specific leaf area influences Mass- and Area-Based expressions," *Oecologia*, vol. 97, pp. 73–81, 1994.

[91] E. A. Ainsworth, S. P. Serbin, J. A. Skoneczka, and P. A. Townsend, "Using leaf optical properties to detect ozone effects on foliar biochemistry," *Photosynthesis Research*, vol. 119, no. 1-2, pp. 65–76, 2014.

[92] A. C. Burnett, S. P. Serbin, and A. Rogers, "Source:sink imbalance detected with leaf- and canopy-level spectroscopy in a field-grown crop," *Plant, Cell & Environment*, vol. 44, no. 8, pp. 2466–2479, 2021.

[93] J. Torres-Sánchez, F. López-Granados, and J. M. Peña, "An automatic object-based method for optimal thresholding in UAV images: application for vegetation detection in herbaceous crops," *Computers and Electronics in Agriculture*, vol. 114, pp. 43–52, 2015.

[94] R. T. Furbank and M. Tester, "Phenomics – technologies to relieve the phenotyping bottleneck," *Trends in Plant Science*, vol. 16, no. 12, pp. 635–644, 2011.

[95] T. Zheng, J. Chen, L. He et al., "Inverting the maximum carboxylation rate (Vcmax) from the sunlit leaf photosynthesis rate derived from measured light response curves at tower flux sites," *Agricultural and Forest Meteorology*, vol. 236, pp. 48–66, 2017.

[96] R. T. Furbank, J. A. Jimenez-Berni, B. George-Jaeggli, A. B. Potgieter, and D. M. Deery, "Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops," *New Phytologist*, vol. 223, no. 4, pp. 1714–1727, 2019.

[97] K. J. Halliday, J. F. Martínez-García, and E.-M. Josse, "Integration of light and auxin signaling," *Cold Spring Harbor Perspectives in Biology*, vol. 1, p. a001586, 2009.

[98] B. S. J. Winkel, "Metabolic channeling in plants," *Annual Review of Plant Biology*, vol. 55, no. 1, pp. 85–107, 2004.

[99] J. L. Araus, R. Sanchez-Bragado, and R. Vicente, "Improving crop yield and resilience through optimization of photosynthesis: panacea or pipe dream?," *Journal of Experimental Botany*, vol. 72, no. 11, pp. 3936–3955, 2021.