

# Genetic Architecture of Phenomic-Enabled Canopy Coverage in *Glycine max*

Alencar Xavier,\* Benjamin Hall,\* Anthony A. Hearst,<sup>†</sup> Keith A. Cherkauer,<sup>†</sup> and Katy M. Rainey\*<sup>1</sup>

\*Department of Agronomy and <sup>†</sup>Department of Agriculture and Biological Engineering, Purdue University, West Lafayette, Indiana 47907

**ABSTRACT** Digital imagery can help to quantify seasonal changes in desirable crop phenotypes that can be treated as quantitative traits. Because limitations in precise and functional phenotyping restrain genetic improvement in the postgenomic era, imagery-based phenomics could become the next breakthrough to accelerate genetic gains in field crops. Whereas many phenomic studies focus on exploratory analysis of spectral data without obvious interpretative value, we used field images to directly measure soybean canopy development from phenological stage V2 to R5. Over 3 years, we collected imagery using ground and aerial platforms of a large and diverse nested association panel comprising 5555 lines. Genome-wide association analysis of canopy coverage across sampling dates detected a large quantitative trait locus (QTL) on soybean (*Glycine max*, L. Merr.) chromosome 19. This QTL provided an increase in yield of 47.3 kg ha<sup>-1</sup>. Variance component analysis indicated that a parameter, described as average canopy coverage, is a highly heritable trait ( $h^2 = 0.77$ ) with a promising genetic correlation with grain yield (0.87), enabling indirect selection of yield via canopy development parameters. Our findings indicate that fast canopy coverage is an early season trait that is inexpensive to measure and has great potential for application in breeding programs focused on yield improvement. We recommend using the average canopy coverage in multiple trait schemes, especially for the early stages of the breeding pipeline (including progeny rows and preliminary yield trials), in which the large number of field plots makes collection of grain yield data challenging.

**KEYWORDS** SoyNAM; canopy; GWAS; phenomics

**I**NFORMATION on spectral reflectance obtained from digital imagery can indicate a plant's chemical composition and physical properties. Imagery also provides useful information about plant architecture (Yol *et al.* 2015). Image data can produce multiple time series and spectral indices (White *et al.* 2012), for which limiting factors include the frequency with which pictures are taken during the growing season, and the capabilities of the camera or sensor in a given wavelength spectrum (Andrade-Sanchez *et al.* 2014). Recent phenomic studies employing imaging have focused on multi- and hyper-spectral capabilities (Yang *et al.* 2013; Kumar *et al.* 2015), with minimal investigation of time series of images for capturing crop development and canopy architecture (Schunk *et al.* 2012). Yet plant architecture comprises a set of feature-

driven traits that can contribute to genetic improvement of field crops. For instance, dwarfing (an architectural trait) is well-known for contributing to the yield increases attained from the Green Revolution (Hammer *et al.* 2009; Swaminathan 2014).

Imagery-based field phenotyping with high-throughput platforms has the potential to drive the discovery of novel traits, and to facilitate routine quantification of architectural, developmental, physiological, and phenological crop characteristics (Araus and Cairns 2014). Because it allows inexpensive evaluation of large numbers of field plots with multiple measures in a short period of time (White *et al.* 2012), such information provides an efficient framework to tackle phenotyping bottlenecks in plant breeding (Furbank and Tester 2011; Araus and Cairns 2014; Kumar *et al.* 2015).

In contrast to point- or plant-based spectral reflectance, image analysis allows direct measurement of spatial or field-based traits that are known to be valuable, such as canopy coverage and vegetation indices, which are possible to collect from the field with high-throughput platforms, such as unmanned aircraft systems (UAS) (Cabrera-Bosquet *et al.* 2012;

Copyright © 2017 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.116.198713>

Manuscript received November 28, 2016; accepted for publication March 3, 2017; published Early Online March 30, 2017.

Available freely online through the author-supported open access option.

Supplemental material is available online at [www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.198713/-/DC1](http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.198713/-/DC1).

<sup>1</sup>Corresponding author: E-mail: [krainey@purdue.edu](mailto:krainey@purdue.edu)

Liebisch *et al.* 2015). Additionally, acquisition of imagery can be a robust approach, regardless of its collection platform, that is less sensitive to technical challenges and prevailing environmental conditions than other data acquisition methods. Thus, characterization and selection of phenotypes acquired from image analysis represents a low-hanging fruit for genetic improvement of yield potential in crops.

Increased crop yields often associate with greater biomass production (expressed as the product of total incident radiation throughout the duration of crop growth), the amount of light interception (LI) by crop canopies, and the conversion of chemical energy into plant dry matter (Richards 2000). Improved LI from rapid canopy development is a significant contributor to increased biomass, total photosynthesis, and yield, which makes it an important target when selecting for genetically superior crops. Richards (2000) proposed that genetic selection for improved LI has been occurring inadvertently since the beginning of crop domestication through the selection of young, competitive seedlings with fast crop growth rates, with this resulting in rapid canopy establishment, and, therefore, greater LI during early vegetative growth. Many canopy parameters play dominant roles in determining source-sink ratios (Board and Harville 1993) along with light-use efficiency of the plant's photosynthetic machinery (Tharakan *et al.* 2008).

From 1924 to 2012, on-farm soybean yields increased  $\sim 23$  kg ha<sup>-1</sup> annually, due to both improved genetics and improved agronomic practices (Rincker *et al.* 2014; Specht *et al.* 2014; Suhre *et al.* 2014). Since the introduction of genomic tools in the past two decades, the rate of yield increase has grown to 66 kg ha<sup>-1</sup> per year, where 50–66% of gain is associated with genetic improvement (Board and Kahlon 2011; Rowntree *et al.* 2013). Recently, evaluation of soybean cultivars released between 1923 and 2007 determined that the greater biomass production and yield potential of newer cultivars is driven by linear increases in LI, efficiency of radiation use, and partitioning efficiency (Koester *et al.* 2014). At present, much of the existing genetic variation in photosynthetic efficiency and related processes is confined to undomesticated germplasm (Sherman-Broyles *et al.* 2014).

Purcell (2000) developed a precise and rapid technique using ground-based digital imagery for evaluating seasonal variation in soybean canopy coverage. The methodology is based on the binary classification of individual pixels from plot images into canopy or background, thereby providing a direct quantification for the percent of canopy coverage on a continuous scale (Purcell 2000; Karcher and Richardson 2005). This research established a one-to-one relationship between fractional canopy coverage (determined through image analysis) and canopy light interception measurements acquired using a line quantum sensor, thereby facilitating the use of digital imagery to evaluate seasonal canopy interception of photosynthetically active radiation (PAR) (Purcell *et al.* 2002; Edwards *et al.* 2005). While these findings greatly

improved the efficiency with which we can assess LI, this method is impractical for use in large experiments that require phenotyping thousands of genetically distinct lines to determine gene-phenotype associations for complex traits (Myles *et al.* 2009).

Rapid canopy development in soybean optimizes growth dynamic parameters, providing a foundation for a greater biomass accumulation during the season, and ultimately greater grain yield potential (Hall 2015); it also favors early-season weed suppression (Jannink *et al.* 2010; Fickett *et al.* 2013). Canopy coverage can be observed for the entire season, and it is likely that observed variation has a genetic basis, and could be a target for genetic improvement. The objectives of this study were (1) to characterize imagery-based early-season canopy coverage in soybean as a trait with quantitative genetic control, (2) to describe the genetic architecture of canopy coverage via genome-wide association, and (3) to provide insight into the application of UAS quantification of canopy parameters for applications in soybean breeding aimed at improving yield.

The application of marker-assisted selection (MAS) in plant breeding represented a critical step in selection practices driven by realized genetic information (Muir 2007; Hayes *et al.* 2009). The applications of genomic tools evolved to whole-genome selection techniques, effective in selecting the favorable alleles in small-effect quantitative trait loci (QTL) that govern low-heritability traits (Heffner *et al.* 2009). However, poor phenotypic data quality is a major limitation for effective use of genomic prediction models (Cobb *et al.* 2013; de los Campos *et al.* 2013). More precise phenotypic information is necessary to accelerate the annual rate of genetic gain in field crops in the post-genomic era (Henryon *et al.* 2014). Nevertheless, phenomic platforms can be successful in crop improvement only if they satisfy basic requirements that include high-throughput at a low cost (White *et al.* 2012; Basu *et al.* 2015), and the ability to provide meaningful, heritable, and measurable trait information (Cobb *et al.* 2013; Liebisch *et al.* 2015).

Imagery acquired from ground and aerial platforms was used to characterize early-season canopy development for 3 years in a next-generation soybean genetic population, for which we also assessed agronomic performance. We developed methodology to automatically extract canopy data from the best quality aerial image of each plot, rather than using mosaics. Percentage canopy pixels was quantified from over 70,000 ground images and over 2000 aerial images. Methods necessary for the analyses include spatial statistics, generalized mixed linear models, and logistic growth function to model longitudinal data of canopy coverage for each genetic line. Genomic data for the population helped determine the genetic architecture of the newly-described soybean trait. Quantitative genetic properties were evaluated to assess the application of high-throughput field-based canopy phenotyping to increase the rate of genetic gain in soybean breeding pipelines.

## Materials and Methods

### Germplasm

The SoyNAM population ([www.soynam.org](http://www.soynam.org)) is a nested association mapping panel consisting of ~5600 F5-derived recombinant inbred lines (RILs) that originated as sets of 140 RILs descended from each of 40 biparental matings of 40 founder parents with one common, high yielding, hub parent (IA3023). The RILs exhibit maturities ranging from late MG II to early MG IV. The 40 founder parents include 17 cultivars or elite public breeding lines contributed by soybean breeders across the North Central Region, 15 unique breeding lines developed to have diverse ancestry, and eight diverse plant introductions selected for their performance in water-limited environments. Visual canopy field images of the SoyNAM parents are available at <http://www.soybase.org/SoyNAM/imagebrowser.php>. Genotyping of the lines employed an Illumina SoyNAM BeadChip SNP array designed for this population, using 5305 single nucleotide polymorphism (SNP) markers identified from the genomic sequences of all 41 parental lines. We imputed missing SNP locus calls using random forest, and removed SNPs with a minor allele frequency <0.15 (Xavier *et al.* 2016) using the R package NAM (Xavier *et al.* 2015). This left a final total of 4077 SNPs for the association analysis.

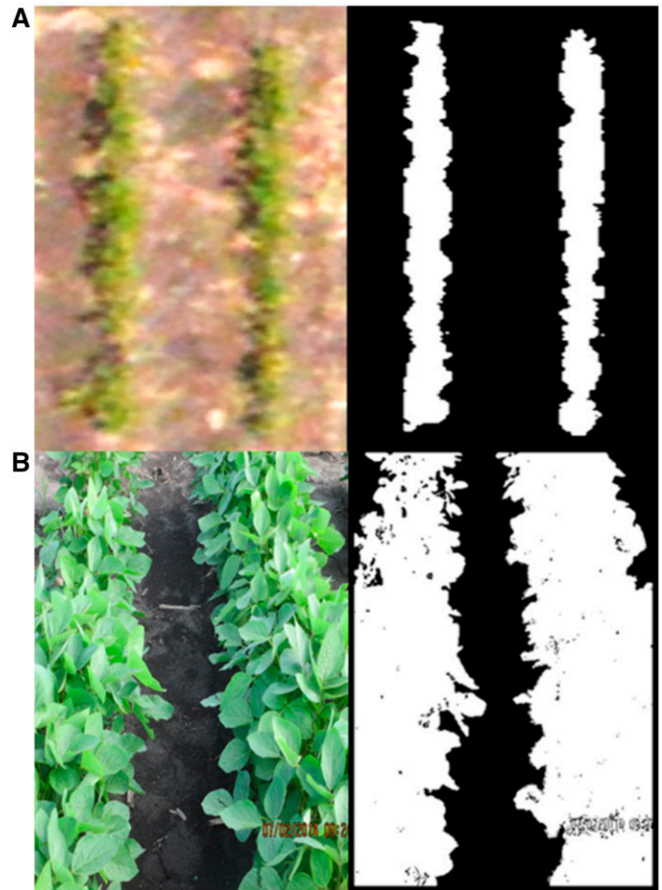
### Field design

In 2013, 2014, and 2015, we collected phenotypic data from the SoyNAM population in West Lafayette, Indiana. From 2012 to 2014, the experiment used a modified augmented design, and an augmented complete block design in 2015. Line planting took place on May 20, 2013; May 24, 2014; and May 23, 2015, at the Purdue University Agronomy Center for Research and Education (40°28'20.5''N 86°59'32.3''W). Experimental units were based on two-row plots (0.76 × 2.90 m) at a density of ~35 plants m<sup>-2</sup>. The 2012–2014 experiments grew all the SoyNAM entries, but the 2015 experiment grew only the six families (NAM5, NAM 9, NAM12, NAM15, NAM24, and NAM40) with the largest phenotypic variance for the traits of interest. The 2015 experiment experienced mild flood damage.

### Canopy coverage

The metric we used to measure canopy coverage was the percentage of image pixels that were classified as canopy pixels (Figure 1). We took images at regular intervals from 2 to 8 weeks after planting, which spanned soybean phenological stages V2–R5 (Fehr *et al.* 1971). The timeframe of data collection coincides with all but the final R5–R7 dry matter accumulation period (Egli and Leggett 1973). Collection of image data used ground-based (2013–2014) and aerial (2014–2015) platforms and images underwent a classification procedure to determine the number of pixels showing the soybean canopy.

Classification of ground-based images used SigmaScan Pro software according to the methods of Karcher and Richardson (2005). The Appendix (Supplemental Material, Figure S1) provides examples of classified ground imagery across the



**Figure 1** Example canopy imagery of a single plot, used to calculate a percentage canopy coverage on a given sampling date. (A, B) From aerial (above; A) or ground (below; B) platforms, with raw (left) and classified (right) imagery.

seasons. Classification of aerial images used a binomial model implemented in ENVI 5.0 and calibrated with manually labeled training data.

Using a logit link function for each field plot, we fit canopy coverage observations from multiple sampling dates into a logistic model, with the canopy coverage (cc) for any point in time (t) represented as

$$cc_t = \frac{e^{(\beta_0 + \beta_1 t)}}{1 + e^{(\beta_0 + \beta_1 t)}} \quad (1)$$

where  $\beta_0$  and  $\beta_1$  are the logistic regression coefficients for any given field plot. This function provided daily canopy coverage projections that were used to estimate canopy development parameters. The single-trait representation of canopy in this study is the average canopy coverage (ACC), defined as

$$ACC = \int_{t_0}^{t_n} f(cc_t) \partial(cc_t) = \frac{1}{N} \sum_{t_0}^{t_n} cc_t \quad (2)$$

where  $N$  is the number of observed days, and  $t_0$  and  $t_n$  represent the first and last day of canopy coverage data collection,

respectively. In essence, the ACC value is an arithmetical mean of multiple seasonally observed values of canopy coverage.

### Ground-based imagery

Collection of images used to infer canopy coverage employed the methodology described by Purcell (2000). We took Red-Green-Blue (RGB) pictures (680 × 480 pixels) at the center of each plot, ~1.5 m above the ground, at a 30° angle from the horizon, using a pocket camera (Canon PowerShot A400IS), then cropped these to 560 × 480 pixels to avoid capturing canopy from neighboring plots.

### Aerial imagery

In 2014, we collected aerial images using a Precision Hawk UAS equipped with a Nikon 1-J1 digital camera (3872 × 2592 pixels). It collected individual images at an altitude of 100 m, yielding a spatial resolution of 3.25 cm pixel<sup>-1</sup>. In 2015, it collected images at 50 m using a Nikon 1-J3 digital camera (4608 × 3072 pixels), yielding a spatial resolution of 1.50 cm pixel<sup>-1</sup>. These images had a 70% lateral and forward overlap, which allowed image stitching using Pix4Dmapper software. Image stitching outputs, including estimated camera positions and orientations, and plot map coordinates, helped to extract ortho-rectified image mosaics, digital surface models, camera positions, and orientations for each aerial photo, and camera model parameters. These outputs were registered to map coordinates (easting, northing, and altitude) using ground control points. We obtained the map coordinates of the individual plots by gridding the image mosaics. The map coordinates of the plots, positions, and orientations of the camera, camera model parameters, and a collinearity relationship, were used to automatically identify which plots appeared in which images and extract ortho-rectified images of the plots directly from the aerial images (Kasser 2002).

### Correspondence between platforms

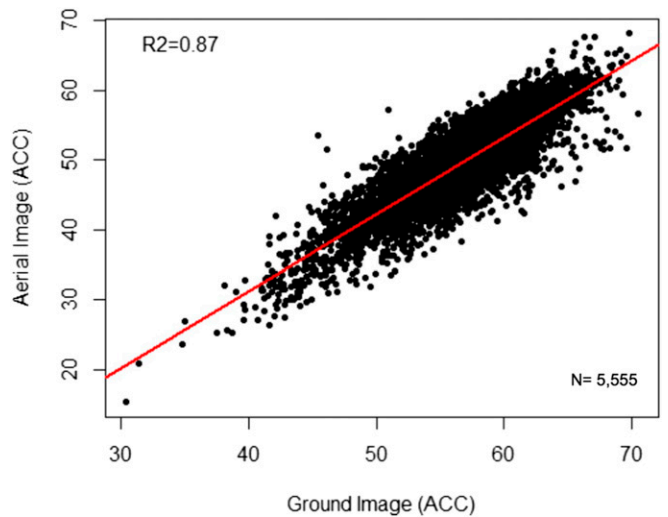
Ganey and Block (1994) reported that canopy coverage estimates from various methods are directly comparable. In 2014, we collected canopy data from both the ground and the UAS platforms, and calculated ACC values for both platforms. A paired data set comparison provided evidence of suitably sufficient correspondence between the two platforms (Figure 2), so we elected to collect data from the UAS platform only in 2015.

### Statistical model

Correlation and genome-wide association analysis used genetic values obtained from best linear unbiased prediction estimators, rather than observed phenotypes of canopy coverage from individual time points. The following mixed linear model was used to provide the genetic values

$$\mathbf{y} = 1\mu + f(x) + \mathbf{Zu} + \mathbf{Wg} + \mathbf{e} \quad (3)$$

where  $\mathbf{y}$  is the vector of observed phenotypes,  $\mu$  is the intercept, and  $f(x)$  is a nonlinear function that accounts for



**Figure 2** Correspondence between the average canopy coverage measured from ground or aerial imagery from 2014.

the spatial heterogeneity of field variation within the blocks imposed by the field design, where  $f(x)$  is computed as the average phenotypic value of neighbor plots (Lado *et al.* 2013).  $\mathbf{Z}$  is the incidence matrix of environment,  $\mathbf{u}$  is the vector of regression coefficients of environment effects,  $\mathbf{W}$  is the incidence matrix of genotypes,  $\mathbf{g}$  is the vector of genetic values, and  $\mathbf{e}$  is the vector of residuals. Environment and genotypes were treated as random effects. Random coefficients were assumed to be normally and independently distributed as  $\mathbf{u} \sim N(0, \mathbf{I}\sigma_u^2)$ ,  $\mathbf{g} \sim N(0, \mathbf{I}\sigma_g^2)$  and  $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$ .

### Association analysis

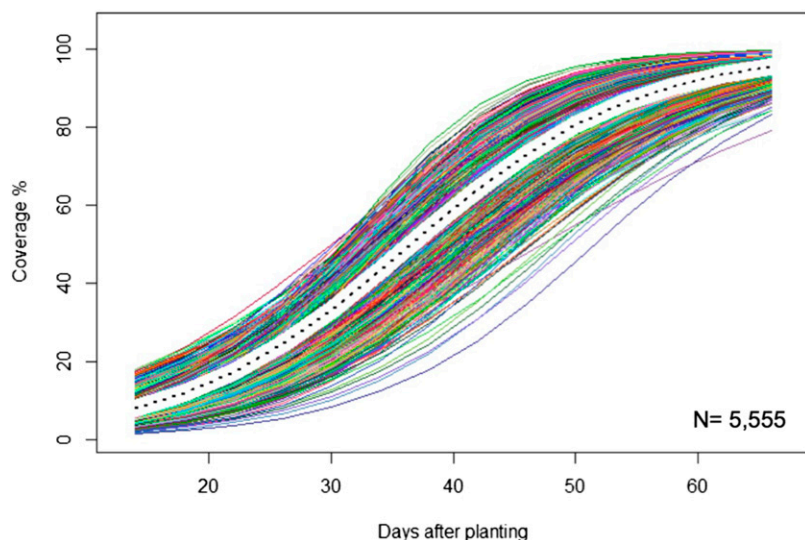
Genome-wide association analysis used the random effect model designed for multi-parental populations (Wei and Xu 2015), implemented in the function *gwas2* from the R package *NAM* (Xavier *et al.* 2015). Analysis were performed for individual canopy coverage measurement days spanning 14–56 days after planting, using a random linear effect model:

$$\mathbf{g} = 1\mu + \mathbf{X}\alpha + \boldsymbol{\psi} + \boldsymbol{\varepsilon} \quad (4)$$

where  $\mathbf{g}$  is the vector of genetic values of canopy coverage for a given point in time fitted in Equation 3,  $\mu$  is the intercept,  $\mathbf{X}$  is the incidence matrix of alleles generated from the marker data and family information,  $\alpha$  is the vector of regression coefficients corresponding to the allele effects,  $\boldsymbol{\psi}$  corresponds to the polygenic effect that accounts for population structure, and  $\boldsymbol{\varepsilon}$  is the vector of residuals. Random effect coefficients were assumed to be normal as  $\alpha \sim N(0, \mathbf{I}\sigma_\alpha^2)$ ,  $\boldsymbol{\psi} \sim N(0, \mathbf{K}\sigma_\psi^2)$  and  $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma_\varepsilon^2)$ , where  $\mathbf{K}$  represents the genomic relationship matrix.

We evaluated statistical significance using a likelihood ratio test by comparing the restricted log-likelihood of the model with the marker included ( $L_1$ ) to the reduced model ( $L_0$ ), not including the marker ( $\mathbf{X}\alpha$ ). Thus,

$$\text{LRT} = -2(L_1 - L_0). \quad (5)$$



**Figure 3** Development of canopy coverage in the Soy-NAM population described as logistic curves of canopy development for individual soybean lines over two or three seasons. Each curve describes an individual soybean line. The white dotted curve represents the mean curve.

In the random model, the LRT follows a mixture of chi-squared and binomial distributions (Xavier *et al.* 2015), with  $P$ -values computed using a chi-squared distribution with 0.5 degrees of freedom. The Bonferroni threshold that accounts for false positives under multiple hypothesis testing ( $\alpha = 0.05$ ) was used to define which markers were associated with the trait of interest. The expected threshold for 4077 SNPs markers was estimated  $4.91 -\log(P\text{-value})$ .

#### Phenotyping of agronomic traits

Grain yield was measured as grams of seed per plot, converted to  $\text{kg ha}^{-1}$  using harvest-timed seed moisture to adjust all plot values to 13% seed moisture. We collected maturity data twice a week as the number days after planting (DAP), back and forward scoring plots that flowered and matured between the intervals. The criterion for a plot to achieve maturity (R8) was 50% of the plants having 95% of their pods mature (Fehr *et al.* 1971).

#### Variance component analysis

We computed Pearson and Spearman correlations from the phenotypic data using built-in R functions. A multivariate mixed linear model estimated covariance components for each of the three traits: grain yield, days to maturity, and average canopy coverage. The linear model was presented on Equation 3, but environment was treated as fixed effect. Model fitting used the AI-REML algorithm (Gilmour *et al.* 1995) implemented in AIREMLF90 (Miszta *et al.* 2002) with the covariance structure set up as follows:

$$\text{Var}(\mathbf{Y}) = \left( \mathbf{K} \otimes \sum_g + \mathbf{I} \otimes \sum_e \right) \quad (6)$$

Here  $\mathbf{K}$  represents the realized genomic relationship matrix. Matrices  $\sum_g$  and  $\sum_e$  represent the genetic and residual covariance matrices among the three traits. These covariance components then helped to estimate the heritability of average canopy coverage, grain yield, and days to

maturity, as well as the genetic correlations among these traits.

#### Data availability

Phenotypes and genotypic data are available in the R package NAM (<https://CRAN.R-project.org/package=NAM>). Load the data using the following command: `data(met, package = "NAM")`.

## Results and Discussion

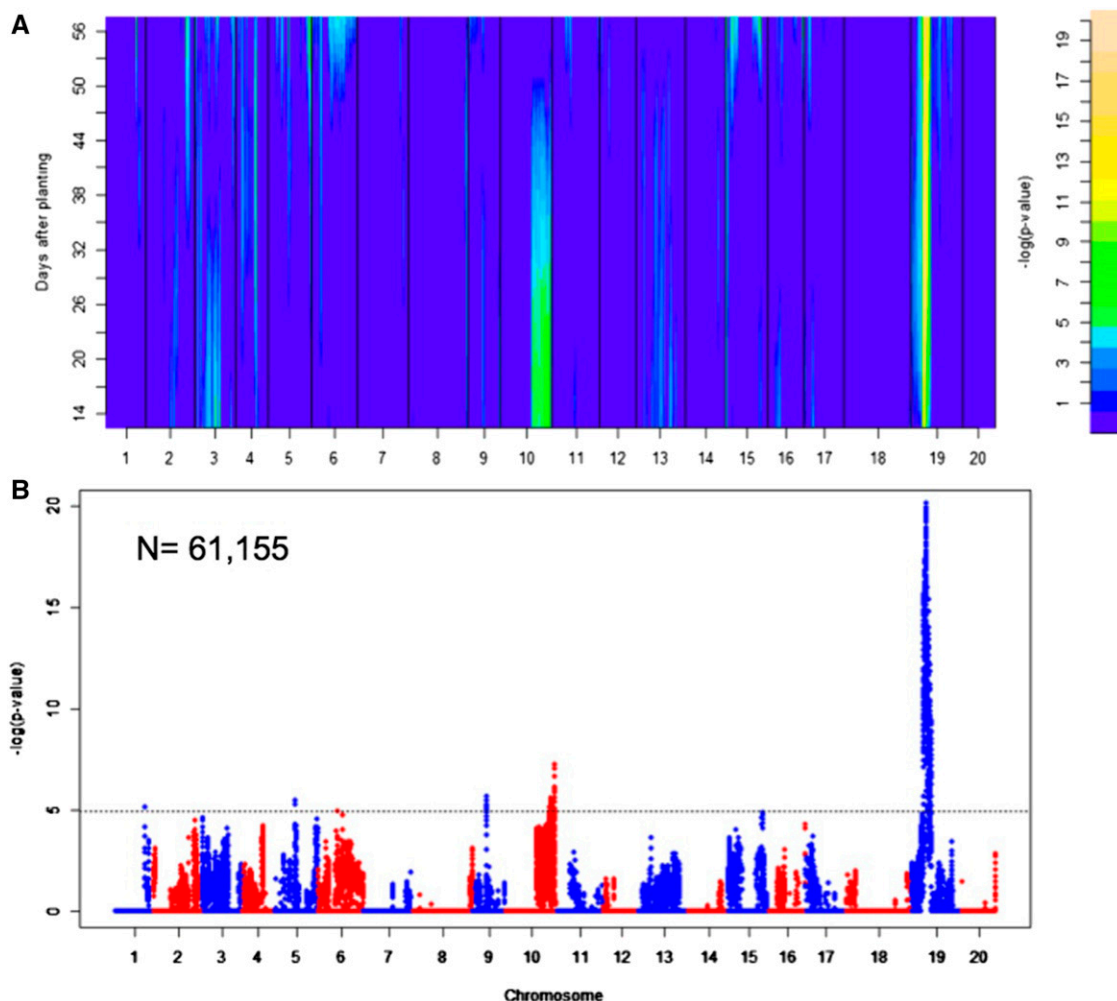
### Canopy behavior

Individual lines showed the potential to attain full canopy coverage (100%) as early as 50 DAP, while others did not reach full coverage within 64 DAP (Figure 3). In the latter case, lines seemed incapable of closing the canopy entirely in a production system employing a row spacing of 76 cm. Reports indicate that yield and canopy traits are under similar genetic control when grown using narrower row spacing, with slight changes in heritability (Weaver and Wilcox 1982; De Bruin and Pedersen 2008).

### Genome-wide association

Six genomic regions, on soybean chromosomes 1, 5, 6, 9, 10, and 19, associated with canopy coverage when measured or estimated across all days (Figure 4b). The largest effect QTL detected was on chromosome 19, and it associated with canopy coverage on all days within the survey period (Figure 4a). The second largest effect QTL was located on chromosome 10, and was significant for approximately two-thirds of the survey period, from 14 to 26 DAP. Table 1 summarizes the QTL detected, the period during which these QTL were significant, and their average allelic effects on ACC, grain yield, and maturity.

The existence of large-effect QTL along with several nearly significant associations indicate that ACC is controlled by major genes in the soybean genome. Another canopy trait,



**Figure 4** Results of the genome-wide association study for canopy coverage. Genomic regions significantly associated with early-season soybean canopy coverage for 2 or 3 years observed or estimated for (A) each day from 14 to 56 days after planting, and (B) across all days.

early-season canopy height, was reported to be heritable ( $h^2 = 0.64$ ) and genetically correlated to weed-suppression ability ( $\rho_{x,y} = 0.80$ ) (Jannink *et al.* 2000).

Genome-wide association presented by day (Figure 4a) shows that the chromosome 19 QTL was significant throughout the survey period. Other associations were significant either earlier or later in the season, or when the canopy

was nearly closed (Table 1). Early- and late-season canopy growth are controlled through independent genetic mechanisms and can be dissected into distinct traits (Jannink *et al.* 2001).

Assuming pleiotropy, the estimated increase in grain yield associated with the positive allele of the chromosome 19 QTL was 47.30 kg ha<sup>-1</sup>. Importantly, this increase did not also increase days to maturity (−0.24 days).

**Table 1** SNPs significantly associated with early-season soybean canopy coverage and the number of DAP during which they were significant

SNP	Period of Significant Association (DAP)	GY (kg ha <sup>-1</sup> )	R8 (DAP)	ACC (%)
Gm01_50911939_C_T	56	−26.04	−1.12	−0.36
Gm05_37467797_A_G	53–56	−44.71	0.08	−0.65
Gm06_14104090_T_C	56	99.58	0.50	0.55
Gm09_4034850_C_T	14–35, 59–64	12.24	−0.60	−0.44
Gm010_44120764_T_C	14–23	5.95	−0.59	0.04
Gm010_44630777_C_A	14–26	51.61	1.06	0.33
Gm019_1586092_T_C	14–64	47.30	−0.24	1.34

Mean change in phenotypic values associated with each QTL in terms of grain yield (GY) expressed as kilograms per hectare, number of DAP to maturity (R8), and the percentage of ACC, calculated as the allelic substitution of the homozygous founder parent genotype for the homozygous IA3023 genotype.

**Table 2** Variance components of the traits considered in this study

	GY	R8	ACC
Var(G)	78.80	208.36	25.54
Var(E)	56.12	9.39	8.10
$h^2$	0.58	0.96	0.76

Genetic (G) and environmental (E) variances and heritabilities for soybean grain yield (GY), and days to maturity (R8), and ACC.

### Quantitative genetic properties of average canopy closure

Variance components provide an insight about the level of genetic and environmental control of the traits of interest, necessary to estimate heritability (Table 2). At multivariate level, (co)variances indicate the nature of the relationship between multiple traits, uncovering whether traits are more associated at genetic or environmental level (Table 3).

For ACC to be an effective predictor of grain yield potential in the early stages of breeding pipelines, the phenotypic correlations, and, more importantly, the genetic correlations, must be high between the two traits. The genetic correlation between grain yield and ACC is quite high (Table 3), but so are the correlations between days to maturity and ACC, and between grain yield and days to maturity. However, as noted above, the main allelic effect of the chromosome 19 QTL affects ACC and yield, without negatively impacting days to maturity.

### Selection for canopy coverage

A faster rate of canopy coverage development (achieved through manipulation of crop management practices) increases light interception, suppresses weeds, and has other beneficial properties (Purcell 2000; Jannink *et al.* 2001; Martius *et al.* 2004; Campillo *et al.* 2008). As a target for artificial selection, the trait presents attractive genetic properties, such as high heritability and strong genetic correlation with grain yield that allows for indirect selection or prediction of grain yield potential.

For indirect selection to be effective, the response of grain yield potential when selecting for ACC must be greater than the response of selecting for yield directly. Very few cases of successful indirect selection have been reported (Bernardo 2010). The indirect selection equation indicates that efficient indirect selection for yield at the same selection intensity is possible with ACC because the correlated response (CR) to selection for grain yield is 14% greater than the direct response (R):

$$\rho_{cc,y} = \frac{CR}{R} = \frac{h_{cc}^2 \times r_{cc,y}}{h_y^2} = \frac{0.76 \times 0.88}{0.58} = 1.14. \quad (7)$$

In the equation above,  $h_{cc}^2$  represents the heritability of the secondary trait (*i.e.*, ACC),  $r_{cc,y}$  is the genetic correlation between primary and secondary traits, and  $h_y^2$  is the heritability of the primary trait (*i.e.*, grain yield potential). Thus, assuming identical selection intensities, indirect selection of grain

**Table 3** Phenotypic Pearson (P), phenotypic Spearman (S), genetic (G), and environmental (E) correlations among GY, ACC and days to maturity (R8)

	GY-ACC	R8-ACC	GY-R8
Cor(P)	0.63	0.31	0.42
Cor(S)	0.70	0.38	0.46
Cor(G)	0.88	0.77	0.72
Cor(E)	0.18	-0.06	0.23

yield using average canopy coverage would be expected to be better than direct selection for yield. Although the canopy traits reported by Jannink *et al.* (2001) did not show maturity-mediated tradeoffs, use of selection indices including both ACC and days to maturity may be necessary to mitigate indirect increases in maturity.

The use of average canopy coverage to predict lines with high grain-yield potential when collection of accurate yield data are not feasible or is challenging. This is often the case during early generations in a breeding pipeline when yield data are of low quality due to low seed number, or when limited resources are available for the evaluation of many lines, as in preliminary yield trials. ACC may also be useful in selection indices that include grain yield and other valuable traits.

### UAS phenotyping and phenomics

From the perspective of genetic improvement of crops, traits with known properties that can be collected *in situ* using UAS platforms are a low-hanging fruit, in particular canopy characteristics and vegetation indices (Cabrera-Bosquet *et al.* 2012; Liebisch *et al.* 2015). In addition to being robust, simple, and cost-effective, these traits are likely more valuable than measurements made in controlled environments for physiological characteristics that cannot be translated into field phenotypes, and which seldom provide economic advantages (White *et al.* 2012; Von Mogel 2013). Here, we have shown canopy coverage to be a highly heritable trait associated with grain yield potential, possibly the most promising phenomic trait reported so far.

### Accounting for environmental noise

This study used spatial statistics in its model, which provided a covariate term that accounted for micro-environmental field variation, or environmental noise, on a plot-by-plot basis (Lado *et al.* 2013). Without this adjustment, the heritability of ACC was estimated to be 0.46 (data not shown), though the magnitude of the genetic correlation of ACC with grain yield remained the same. Previous studies (Piepho *et al.* 2008; Oakey *et al.* 2016) described the importance of considering spatial structure among field observations to achieve proper estimation of statistically-estimated trait values, or breeding values. Many authors report the challenge of differentiating between signal and noise in phenomic data collected from the field (Cobb *et al.* 2013; Araus and Cairns 2014; Basu *et al.* 2015). We conclude that analysis of field phenomic traits should not ignore spatial information.

## Breeding values incorporating phenomic data

Research is beginning to explore the optimal application of phenomic traits in genetic improvement of crop plants, and how selection schemes and prediction models should accommodate phenomic information. Phenomic quantifications from field imagery are indicators of plant health and development (Liebisch *et al.* 2015), but inclusion of such measurements as covariates in genomic models can be controversial (Valente *et al.* 2015) because they share both a genetic and an environmental correlation with agronomic traits (Table 3). Instead, phenomic traits might be more suitably used in multi-trait models, to improve the accuracy of estimated breeding values by exploiting genetic correlations, and, therefore, preserve the genetic nature of the association between the target trait and phenomic trait. We showed that canopy coverage displays a high correlation with valuable soybean traits (Table 3), that favors multivariate models and enhances the accuracy of breeding values.

## Acknowledgments

We are grateful to Dr. Jim Specht for his critical insight and helpful review of the manuscript. We also thank the United Soybean Board for providing the funding to perform the experiment from 2012 to 2013, Dow AgroSciences for providing the funding to perform the experiment from 2014 to 2015, and the National Science Foundation for supporting the drone phenotyping. The authors declare no conflict of interest.

Author contributions: A.X. performed the statistical analysis and wrote the manuscript. B.H. collected and processed the ground imagery, and A.A.H. developed software for automatic extraction of quantitative information from aerial imagery. K.A.C. and K.M.R. contributed to the theoretical basis of the manuscript, and K.M.R. designed the experiment and edited the manuscript.

## Literature Cited

- Andrade-Sanchez, P., M. A. Gore, J. T. Heun, K. R. Thorp, A. E. Carmo-Silva *et al.*, 2014 Development and evaluation of a field-based high-throughput phenotyping platform. *Funct. Plant Biol.* 41: 68–79.
- Araus, J. L., and J. E. Cairns, 2014 Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19: 52–61.
- Basu, P. S., M. Srivastava, P. Singh, P. Porwal, R. Kant *et al.*, 2015 High-precision phenotyping under controlled vs. natural environments, pp. 27–40 in *The Phenomics in Crop Plants: Trends, Options and Limitations*, edited by J. Kumar, A. Pratap, and S. Kumar. Springer, India.
- Bernardo, R., 2010 *Breeding for Quantitative Traits in Plants*. Stemma Press, Woodbury, MN.
- Board, J. E., and B. G. Harville, 1993 Soybean yield component responses to a light interception during the reproductive period. *Crop Sci.* 33: 772–777.
- Board, J. E., and C. S. Kahlon, 2011 *Soybean Yield Formation: What Controls It and How It Can Be Improved*. INTECH Open Access Publisher, Rijeka, Croatia.
- Cabrera-Bosquet, L., J. Crossa, J. von Zitzewitz, M. D. Serret, and J. L. Araus, 2012 High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. *J. Integr. Plant Biol.* 54: 312–320.
- Campillo, C., M. H. Prieto, C. Daza, M. J. Moñino, and M. I. García, 2008 Using digital images to characterize canopy coverage and light interception in a processing tomato crop. *HortScience* 43: 1780–1786.
- Cobb, J. N., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch, 2013 Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* 126: 867–887.
- De Bruin, J. L., and P. Pedersen, 2008 Effect of row spacing and seeding rate on soybean yield. *Agron. J.* 100: 704–710.
- de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus, 2013 Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193: 327–345.
- Edwards, J. T., L. C. Purcell, and D. E. Karcher, 2005 Soybean yield and biomass responses to increasing plant population among diverse maturity groups. *Crop Sci.* 45: 1778.
- Egli, D. B., and J. E. Leggett, 1973 Dry matter accumulation patterns in determinate and indeterminate soybeans. *Crop Sci.* 13: 220–222.
- Fehr, W. R., C. E. Caviness, D. T. Burmood, and J. S. Pennington, 1971 Stage of development description for soybeans, *Glycine max* (L.) Merrill. *Crop Sci.* 11: 929–931.
- Fickett, N. D., C. M. Boerboom, and D. E. Stoltenberg, 2013 Soybean yield loss potential associated with early-season weed competition across 64 site-years. *Weed Sci.* 61: 500–507.
- Furbank, R. T., and M. Tester, 2011 Phenomics - technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16: 635–644.
- Ganey, J. L., and W. M. Block, 1994 A comparison of two techniques for measuring canopy closure a comparison of two techniques for measuring canopy closure. *West. J. Appl. For.* 9: 21–23.
- Gilmour, A. R., R. Thompson, and B. R. Cullis, 1995 Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51: 1440–1450.
- Hall, B., 2015 Quantitative characterization of canopy coverage in the genetically diverse SoyNAM population. Master's Dissertation, Purdue University, West Lafayette, IN.
- Hammer, G. L., Z. Dong, G. McLean, A. Doherty, C. Messina *et al.*, 2009 Can changes in canopy and/or root system architecture explain historical maize yield trends in the U.S. corn belt? *Crop Sci.* 49: 299–312.
- Hayes, B. J., P. M. Visscher, and M. E. Goddard, 2009 Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91: 47–60.
- Heffner, E. L., M. E. Sorrells, and J. L. Jannink, 2009 Genomic selection for crop improvement. *Crop Sci.* 49: 1.
- Henryon, M., P. Berg, and A. C. Sørensen, 2014 Animal-breeding schemes using genomic information need breeding plans designed to maximise long-term genetic gains. *Livest. Sci.* 166: 38–47.
- Jannink, J. L., J. H. Orf, N. R. Jordan, and R. G. Shaw, 2000 Index selection for weed suppressive ability in soybean. *Crop Sci.* 40: 1087–1094.
- Jannink, J. L., N. R. Jordan, and J. H. Orf, 2001 Feasibility of selection for high weed suppressive ability in soybean: absence of tradeoffs between rapid initial growth and sustained later growth. *Euphytica* 120: 291–300.
- Jannink, J. L., A. J. Lorenz, and H. Iwata, 2010 Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics* 9: 166–177.
- Karcher, D. E., and M. D. Richardson, 2005 Batch analysis of digital images to evaluate turfgrass characteristics. *Crop Sci.* 45: 1536–1539.



- Kasser, M., 2002 *Digital Photogrammetry*. Taylor & Francis, London, New York.
- Koester, R. P., J. A. Skoneczka, T. R. Cary, B. W. Diers, and E. A. Ainsworth, 2014 Historical gains in soybean (*Glycine max* Merr.) seed yield are driven by linear increases in light interception, energy conversion, and partitioning efficiencies. *J. Exp. Bot.* 65: 3311–3321.
- Kumar, J., A. Pratap, and S. Kumar, 2015 Plant phenomics: an overview, pp. 1–10 in *The Phenomics in Crop Plants: Trends, Options and Limitations*, edited by J. Kumar, A. Pratap, and S. Kumar. Springer, India.
- Lado, B., I. Matus, A. Rodriguez, L. Inostroza, J. Poland *et al.*, 2013 Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. *G3 (Bethesda)* 3: 2105–2114.
- Liebisch, F., N. Kirchgessner, D. Schneider, A. Walter, and A. Hund, 2015 Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. *Plant Methods* 11: 9.
- Martius, C., H. Höfer, M. V. B. Garcia, J. Römbke, B. Förster *et al.*, 2004 Microclimate in agroforestry systems in central Amazonia: does canopy closure matter to soil organisms? *Agrofor. Syst.* 60: 291–304.
- Misztal, L., S. Tsuruta, T. Strabel, B. Auvray, T. Druet *et al.*, 2002 BLUPF90 and related programs (BGF90). *Proc. 7th World Congr. Genet. Appl. to Livest. Prod.* 28: 21–22.
- Muir, W. M., 2007 Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J. Anim. Breed. Genet.* 124(6): 342–355.
- Myles, S., J. Peiffer, P. J. Brown, E. S. Ersoz, Z. Zhang *et al.*, 2009 Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 21(8): 2194–2202.
- Oakey, H., B. Cullis, R. Thompson, J. Comadran, C. Halpin *et al.*, 2016 Genomic selection in multi-environment crop trials. *G3 (Bethesda)* 6: 1313–1326.
- Piepho, H. P., J. Möhring, A. E. Melchinger, and A. Büchse, 2008 BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161(1–2): 209–228.
- Purcell, L. C., 2000 Soybean canopy coverage and light interception measurements using digital imagery. *Crop Sci.* 40: 834–837.
- Purcell, L. C., R. A. Ball, J. D. Reaper, and E. D. Vories, 2002 Radiation use efficiency and biomass production in soybean at different plant population densities. *Crop Sci.* 42: 172–177.
- Richards, R. A., 2000 Selectable traits to increase crop photosynthesis and yield of grain crops. *J. Exp. Bot.* 51: 447–458.
- Rincker, K., R. Nelson, J. Specht, D. Sleper, T. Cary *et al.*, 2014 Genetic improvement of U.S. soybean in maturity groups II, III, and IV. *Crop Sci.* 54: 1419–1432.
- Rowntree, S. C., J. J. Suhre, N. H. Weidenbenner, E. W. Wilson, V. M. Davis *et al.*, 2013 Genetic gain  $\times$  management interactions in soybean: I. Planting date. *Crop Sci.* 53(3): 1128–1138.
- Schunk, C. R., M. Eberius, Q. Y. Shu, B. P. Forster, and H. Nakagawa, 2012 Phenomics in plant biological research and mutation breeding, pp. 535–560 in *Plant Mutation Breeding and Biotechnology*. CABI, Wallingford, UK.
- Sherman-Broyles, S., A. Bombarely, A. F. Powell, J. L. Doyle, A. N. Egan *et al.*, 2014 The wild side of a major crop: soybean's perennial cousins from down under. *Am. J. Bot.* 101: 1651–1665.
- Specht, J. E., B. D. Diers, R. L. Nelson, J. F. F. de Toledo, J. A. Torrión *et al.*, 2014 Soybean (*glycine max* (L.) merr.), pp. 311–356 in *Yield Gains in Major US Field Crops CSSA Special Publication 33*, edited by S. Smith, B. W. Diers, J. Specht, and B. Carver. Crop Science Society of America, Madison, WI.
- Suhre, J. J., N. H. Weidenbenner, S. C. Rowntree, E. W. Wilson, S. L. Naeve *et al.*, 2014 Soybean yield partitioning changes revealed by genetic gain and seeding rate interactions. *Agron. J.* 106: 1631–1642.
- Swaminathan, M. S., 2014 Norman Borlaug and a hunger free world. *Resonance* 19: 109–115.
- Tharakan, P. J., T. A. Volk, C. A. Nowak, and G. J. Ofezu, 2008 Assessment of canopy structure, light interception, and light-use efficiency of first year regrowth of shrub willow (*Salix* sp.). *BioEnergy Res.* 1: 229–238.
- Valente, B. D., G. Morota, F. Peñagaricano, D. Gianola, K. Weigel *et al.*, 2015 The causal meaning of genomic predictors and how it affects construction and comparison of genome-enabled selection models. *Genetics* 200: 483–494.
- Von Mogel, K. H., 2013 Taking the phenomics revolution into the field. *CSA News* 58: 4.
- Weaver, D. B., and J. R. Wilcox, 1982 Heritabilities, gains from selection, and genetic correlations for characteristics of soybeans grown in two row spacings. *Crop Sci.* 22(3): 625–629.
- Wei, J., and S. Xu, 2015 A random-model approach to QTL mapping in multiparent advanced generation intercross (MAGIC) populations. *Genetics* 202: 471–486.
- White, J. W., P. Andrade-Sanchez, M. A. Gore, K. F. Bronson, T. A. Coffelt *et al.*, 2012 Field-based phenomics for plant genetics research. *F. Crop. Res.* 133: 101–112.
- Xavier, A., S. Xu, W. M. Muir, and K. M. Rainey, 2015 NAM: association studies in multiple populations. *Bioinformatics* 31: 3862–3864.
- Xavier, A., W. M. Muir, and K. M. Rainey, 2016 Assessing predictive properties of genome-wide selection in soybeans. *G3 (Bethesda)* 6: 2611–2616.
- Yang, W., L. Duan, G. Chen, L. Xiong, and Q. Liu, 2013 Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary technologies. *Curr. Opin. Plant Biol.* 16: 180–187.
- Yol, E., C. Toker, and B. Uzun, 2015 Traits for phenotyping, pp. 11–26 in *The Phenomics in Crop Plants: Trends, Options and Limitations*, edited by J. Kumar, A. Pratap, and S. Kumar. Springer, New Delhi, India.

Communicating editor: F. v. Eeuwijk