

# Developing Growth-Associated Molecular Markers Via High-Throughput Phenotyping in Spinach

Henry O. Awika, Renesh Bedre, Junho Yeom, Thiago G. Marconi, Juan Enciso, Kranthi K. Mandadi, Jinha Jung, and Carlos A. Avila\*

H.O. Awika, R. Bedre, T.G. Marconi, J. Enciso, K.K. Mandadi, C.A. Avila, Texas A&M Agrilife Research and Extension Center, Weslaco, TX 78596; J. Yeom, J. Jung, School of Engineering and Computing Sciences, Texas A&M-Corpus Christi, Corpus Christi, TX 78412; J. Yeom, (current address), Research Institute for Automotive Diagnosis Technology of Multi-scale Organic and Inorganic Structure, Kyungpook National Univ., Korea 37224; J. Enciso, Biological and Agricultural Engineering Dep., Texas A&M Univ., College Station, TX 77843; K.K. Mandadi, Dep. of Plant Pathology and Microbiology, Texas A&M Univ., College Station, TX 77843; C.A. Avila, Dep. of Horticultural Sciences, Texas A&M Univ., College Station, TX 77843.

**ABSTRACT** Despite advances in sequencing for genotyping, the lack of rapid, accurate, and reproducible phenotyping platforms has hampered efforts to use genetic analysis to predict traits of interest. Therefore, the use of high-throughput systems to phenotype traits related to crop growth, yield, quality, and resistance to biotic and abiotic stresses has become a major asset for breeding. Here, we assessed the efficacy of unmanned aircraft system (UAS)-based high-throughput phenotyping to obtain data for molecular marker development for spinach (*Spinacia oleracea* L.) improvement. We used a UAS equipped with a red–green–blue sensor to capture raw images of 284 spinach accessions throughout the crop cycle. Processed images generated orthomosaic and digital surface models for estimating canopy cover, canopy volume, and excess greenness index models. In addition, we manually recorded the number of days to bolting. Genome-wide association studies against a single-nucleotide polymorphism (SNP) panel obtained by ddRADseq identified 99 SNPs significantly associated with growth parameters. Some of these SNPs are in transcription factor and stress-response genes with possible roles in plant growth and development. The results underscore the utility of combining aerial imaging and genomic data analysis to optimize marker development. This study lays the foundation for the use of UAS-based high-throughput phenotyping for the molecular breeding of spinach.

**Abbreviations:** CC, canopy cover;  $CC_{max}$ , maximum canopy cover; chr, chromosome; CVol, canopy volume;  $CVol_{max}$ , maximum canopy volume; DAS, days after sowing;  $DAS_{bol}$ , days after sowing until early bolting;  $DAS_{CC}$ , days after sowing to maximum canopy cover;  $DAS_{CVol}$ , days after sowing to maximum canopy volume;  $DAS_{ExG}$ , days after sowing to maximum excess greenness index;  $DAS_{kf}$ , days after sowing until kernel filling;  $DAS_{poll}$ , days after sowing until pollination; DSM, digital surface model; ExG, excess greenness index;  $ExG_{max}$ , maximum excess greenness index; GVAS, genome-wide association studies;  $H^2$ , broad-sense heritability;  $h^2$ , narrow-sense heritability; K, the number of computation clusters used to determine the representative Q for the study population; k, kinship matrix; LD, linkage disequilibrium; MDS, multidimensional scaling; NPGS, National Plant Germplasm System; P3D, previously determined default population parameters; PCA, principal component analysis; PVE, phenotypic variance explained; Q, the number of computation clusters with a list of genetic distance probabilities for each accession in the population; RGB, red–green–blue; SNP, single nucleotide polymorphism; UAS, unmanned aircraft system; VIF, variance inflation factor.

## CORE IDEAS

- High-throughput imaging and genomic information can be combined to optimize marker development.
- Genome-wide association studies identified loci associated with plant growth traits.
- We identified candidate genes associated with plant growth and development.

**S**PINACH, a member of the Amaranthaceae family, is an economically important leafy green crop that is widely grown in the United States. Although the spinach production area has grown steadily during the past 7 yr ([https://www.nass.usda.gov/Quick\\_Stats/CDQT/chapter/1/table/36/state/US](https://www.nass.usda.gov/Quick_Stats/CDQT/chapter/1/table/36/state/US), accessed 28 June 2019), this crop must thrive in a dynamic environment that includes constant challenges by abiotic and biotic stresses, which can reduce yield and quality (Lyon et al., 2016; Agarwal et al., 2018; Feng et al., 2018; Min et al., 2018). Therefore, one current challenge in spinach production is to increase crop productivity by improving disease resistance and environmental stress tolerance.

Spinach is a highly diverse species (Christenhusz and Byng, 2016) and because of the dioecious nature of

Citation: Awika, H.O., R. Bedre, J. Yeom, T.G. Marconi, J. Enciso, K.K. Mandadi, J. Jung, and C.A. Avila\*. 2019. Developing growth-associated molecular markers via high-throughput phenotyping in spinach. *Plant Genome* 12:190027. doi: 10.3835/plantgenome2019.03.0027

Received 29 Jan. 2019. Accepted 22 Apr. 2019.

\*Corresponding author (Carlos.Avila@ag.tamu.edu).

© 2019 The Author(s). This is an open access article distributed under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

spinach (female and male organs on separate plants), breeding efficiency is very low with conventional methods. Moreover, the lack of good linkage maps and high-density molecular markers has hampered the application of modern molecular breeding methods in spinach. However, with advances in next-generation sequencing and the recent availability of a draft reference genome for spinach (Xu et al., 2017), as well as the identification of a large panel of SNPs by high-throughput genotyping (Shi et al., 2016a, 2016b, 2017; Qin et al., 2017), it is now possible to identify many markers.

To link these markers precisely to traits of interest, large, diverse populations for association studies and moderately sized biparental populations for targeted traits must be phenotyped. Phenotyping has traditionally been performed by taking manual measurements at one or a few time points, a laborious approach that is prone to high variability resulting from human error. Phenotypic information should reflect a continuous and complex set of responses to multi-dimensional endogenous and exogenous signals that change throughout the lifecycle of an individual plant and in response to environmental conditions (Tessmer et al., 2013; Moharana and Dutta, 2016). Therefore, complete phenomic characterization remains a long-term goal (Houle et al., 2010; Cobb et al., 2013).

In this area, horticultural crops such as spinach have lagged behind other field crops such as maize (*Zea mays* L.), rice (*Oryza sativa* L.), wheat (*Triticum aestivum* L.), and sorghum [*Sorghum bicolor* (L.) Moench.] (Tripodi et al., 2018). One challenge is to capture phenotypic information at a throughput and quality sufficient to mine usable information. Indeed, only a few reports describe the use of proximal instrumentation to phenotype plants (Tripodi et al., 2018), making it challenging to examine large numbers of plants under different environmental conditions, especially in a field environment. For instance, hyperspectral imaging has been limited to examining pigments and quality in plants under storage conditions (Diezma et al., 2013; Zhang et al., 2017), monitoring N content and biomass in a controlled environment (Corti et al., 2017), and using hyperspectral and ultraviolet fluorescence to detect contaminants on harvested leaves (Everard et al., 2014). Therefore, it is critical to develop high-throughput phenotyping methods that can reduce variability and ultimately improve breeding efficiency and cultivar development in spinach. Moreover, phenotyping of spinach must quickly evolve to be technology-driven and high-throughput (Berger et al., 2012), to be more accurate and less subjective, and to have improved phenotyping capabilities.

The UAS phenotyping platform, combined with digital image analysis, offers an emerging option for high-throughput phenotyping. Analysis of digital images taken via proximal or remote sensing phenotyping platforms allows us to examine plants in the ultraviolet, visible light, infrared, near infrared, and other thermal ranges or in the form of fluorescence. In this method, information about several traits can be extracted simultaneously

(Fahlgren et al., 2015; Tripodi et al., 2018). Digital aerial photography and multispectral and hyperspectral imaging from fixed-wing aircraft have been successfully used to discriminate and map long-range spatial features (Everitt et al., 2007; Sheard et al., 2010) but the details and resolution substantially vary and photographic cameras are limited to the visible light range. However, the use of image stacking (a top-view image and side views from multiple images taken with a side-mounted or angled photographic cameras) and algorithms to identify plant-derived pixels (Sheard et al., 2010; Chen et al., 2014; Neilson et al., 2015) have enabled the successful use of light imaging to measure the morphological (shape, structure), geometric (length, area), and color properties of individual plants or groups of plants. Such measurements in the visible light range have been used to estimate plant biomass, approximate plant volume, and measure total leaf area to model the fresh and dry weights of aboveground biomass in barley (*Hordeum vulgare* L.), sorghum, wheat, and rice (Golzarian et al., 2011; Hairmansis et al., 2014; Yang et al., 2014a; Fahlgren et al., 2015; Neilson et al., 2015). Similar methods have been used to model dynamic processes such as plant growth rates in sorghum, *Arabidopsis thaliana* (L.) Heynh., and barley (Tessmer et al., 2013; Yang et al., 2014b; Neilson et al., 2015). However, phenotypic imaging in the visible light range has not been widely used for vegetables, particularly spinach.

In *A. thaliana*, phenotypic traits such as leaf length, rosette area, and plant vegetation conditions have been successfully extracted from images obtained with low-cost, off-the-shelf digital visible light cameras. The phenotypic data have been successfully validated across two platforms: a stationary indoor platform and a movable platform in the field. However, many (108) cameras were used indoors to obtain a total field of view of less than 11 m<sup>2</sup>, whereas the two cameras used in the field had to be moved manually to seven field of view points along a metal frame to cover an area of less than 1 m<sup>2</sup> and with an 80% field of view overlap between cameras from a height of only 0.95 m (An et al., 2016). This method would be expensive and present logistical challenges when used in a larger experimental field because of the large number of cameras required, ground-level disturbance, shadowing from equipment, and the need for expensive facility fixtures. Furthermore, most such studies have mainly focused on obtaining phenotypic information coupled with physiological, chemical, and/or agronomic interpretations but have not been designed to model the underlying genetic architecture (Wang et al., 2018). The few studies that have attempted to model trait–genome associations have mainly used manually acquired, proximal, contact, or destructive phenotyping methods (Cabrera-Bosquet et al., 2012; Steidle Neto et al., 2017). Thus the potential advantages and power of remotely sensed data have not been adequately tested for field-grown vegetables.

In the present study, we used remote sensing technology to obtain agronomic growth metrics in field-grown spinach and used the resulting data to test for molecular

marker signals in a dense spinach genomic SNP map. The goal was to determine the utility of visible light imaging from a UAS platform for improving field phenotyping throughput for genomic studies in spinach and to motivate similar studies in other leafy greens. We used an autonomous aerial system with red–green–blue (RGB) sensors to capture images of 284 spinach accessions in the field over an 11-wk period. The digital signals were converted into numerical data and imputed into canopy cover (CC), canopy volume (CVol), and excess greenness index (ExG) values. We determined the maximum values of these phenotypes and the corresponding days after sowing at which the maximum observations were made. We also manually recorded the number of days to bolting. These observations were used as the traits that were subjected to association studies against SNP data obtained by high-throughput genotyping-by-sequencing. By using this approach, we identified 99 SNPs that were significantly associated with measured spinach growth parameters, thus demonstrating that it is possible to combine aerial imaging with information about genomic architecture to optimize marker development for spinach breeding.

## MATERIALS AND METHODS

### Plant Material

A total of 284 spinach accessions were used in this study, including 264 accessions provided by the USDA-National Plant Germplasm System (NPGS) and 20 lines from the Texas A&M AgriLife Research spinach vegetable breeding program. The countries of origin (USDA-NPGS records), regional groupings, and other categories used in this study are given in Supplemental Table S1. The origin of the USDA spinach accessions has also been described previously (Shi et al., 2016b, 2017). Detailed descriptions based on accession can be accessed at ([https://npgsweb.ars-grin.gov/gringlobal/view2.aspx?dv=web\\_site\\_taxon\\_accessionlist&params=:taxonomyid=35256::siteid=16](https://npgsweb.ars-grin.gov/gringlobal/view2.aspx?dv=web_site_taxon_accessionlist&params=:taxonomyid=35256::siteid=16), accessed 18 June 2019). For brevity, we grouped the accessions into continental subregions based on countries of origin including Eastern Africa (1), Eastern Europe (117), Far East (1), Middle East (44), South Asia (39), Texas A&M AgriLife Research breeding lines (20), US Central (2), US East Coast (19), US North (6), US Gulf Coast (12), Western Europe (22), and US East Coast (1). This diversity in germplasm origin was expected to provide sufficient genetic variation for the purposes of this study.

### Field Design

This study was conducted at the Texas A&M AgriLife Research and Extension Center located in Weslaco, TX at a latitude of 26° 9' 30" N and a longitude of 97° 57' 43". The 284 accessions were grown in a randomized complete block design with three replicates. Each plot was a double-row rectangular grid consisting of 14 plants spaced at ~10 cm between plants within a row of 10 plants and 15 cm between the two rows. The plots were arranged in field rows, each consisting of 40 plots with

1.22 m spacing between adjacent plots within a row and 1.22 m between rows. Conventional agronomic practices for spinach were performed from land preparation to the end date of data collection. The crop was fertilized with a generalized N-P<sub>2</sub>O<sub>5</sub>-K<sub>2</sub>O rate of 135–84–90 kg/ha.

### Unmanned Aircraft System Data Acquisition, Processing, and Extraction of Phenotypic Features

Unmanned aircraft system data acquisition missions were conducted, beginning at approximately 40 d after sowing (DAS) for 11 consecutive weeks until the end of the season. The UAS data were acquired with a DJI Phantom 4 Pro platform (DJI, Nanshan District, Shenzhen, China), which includes a RGB sensor with a 20-megapixel resolution. Images were captured at consecutive intervals of ~7 d (1 wk) for a total of ~12 wk. The UAS flight missions were designed with a Pix4D Capture application (Pix4D Inc., San Francisco, CA) with a 18-m flying altitude and 84% overlap. In addition to UAS flight, GPS surveying on six permanent and two portable ground control points was conducted for accurate georeferencing. Since the study area was approximately 4450 m<sup>2</sup> in size, the number and locations of ground control points were sufficient to remove geometrical errors from the UAS data (Mesas-Carrascosa et al., 2015). The coordinates of all ground control points were surveyed with a differential dual-frequency GPS manufactured by V-Map (<http://v-map.net/>, accessed 18 June 2019). The raw images were processed to generate orthomosaic images and a digital surface model (DSM) with Agisoft Photoscan Pro software (Agisoft LLC, St. Petersburg, Russia). The spatial resolutions of the orthomosaic image and DSM on each date were 0.4 and 0.8 cm, respectively.

All UAS measurements, including CC, canopy height, CVol, and ExG were derived for each plot. For the measurements, a geographical information system array consisting of small (0.9 by 0.9 m) grids was generated and the locations of each grid were manually fitted to each plant location. Canopy cover measurements indicated the percentage of green canopy per grid unit. The Canopeo algorithm (Patrignani and Ochsner, 2015) was used to differentiate canopy pixels from other background pixels, and CC was calculated as the ratio between the number of canopy pixels and the total number of pixels within the grid. Canopy height was calculated by subtracting the initial DSM before planting from the DSM obtained later. The maximum canopy height value of each grid was assigned as a representative value for each plant. Canopy volume was measured by multiplying each pixel's area by its DSM value. The accumulated volume of all pixels on each grid was used to obtain CVol measurements. Excess greenness is an RGB-based vegetation index known to be a good crop monitoring parameter (Woebbecke et al., 1995). The average value of ExG within each grid was used for further analysis.

### Plant Growth Models with UAS-Generated Data

Time series measurements of extracted UAS phenotypic features were used to model the growth patterns

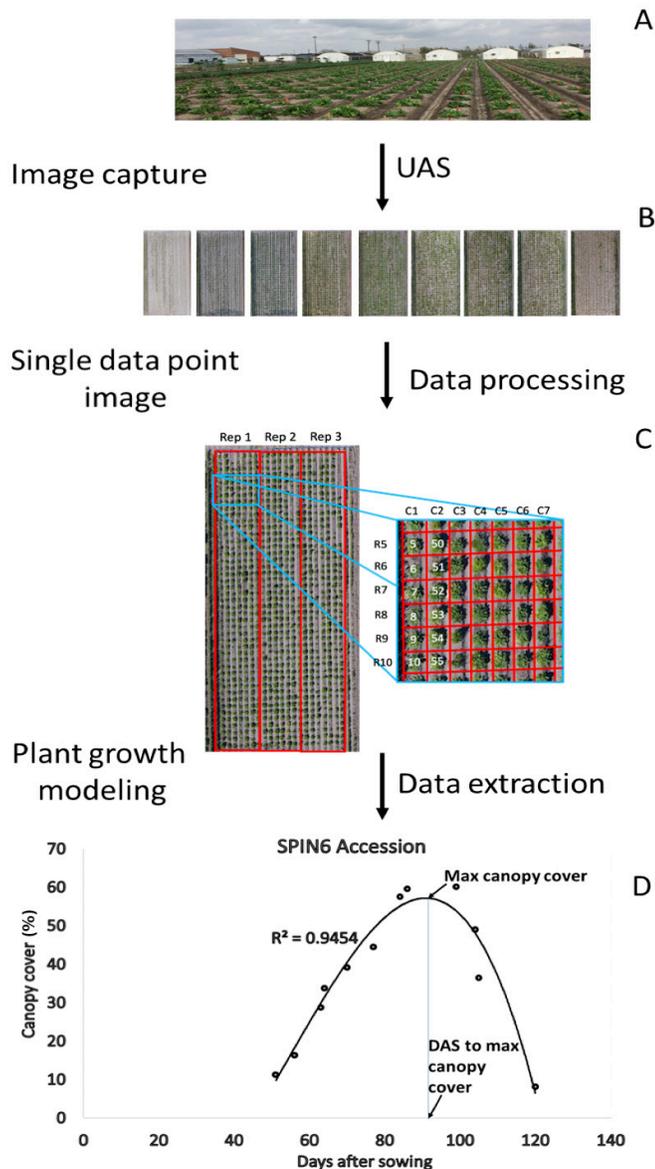


Fig. 1. Growth modeling of spinach using the unmanned aircraft system (UAS). (A) Ground image of field plots; (B) a series of orthomosaic aerial images from the UAS; (C) zoomed image showing replications and plot grids; (D) graph fitting canopy cover (one of six traits phenotyped by the UAS).

of individual spinach accessions by fitting them to the best polynomial curve function. The CVol, CC, and ExG curves were fitted to a univariate third-level polynomial function as shown in Fig. 1. Maximum growth values for each variable were calculated when the second derivative was equal to zero ( $F''(x) = 0$ , Fig. 1). The number of DAS to reach the phenotypic maximum values for CVol ( $CVol_{max}$ ), CC ( $CC_{max}$ ), and ExG ( $ExG_{max}$ ) were determined by estimating the corresponding DAS ( $x$ -axis) to  $CC_{max}$  ( $DAS_{CC}$ ),  $DAS$  to  $CVol_{max}$  ( $DAS_{CVol}$ ), and  $DAS$  to  $ExG_{max}$  ( $DAS_{ExG}$ ). The DAS-based traits were used as indicators of earliness, plant development, and timing for management practices aimed at exploiting maximum growth and photochemistry, such as harvest at peak

greenness and  $CVol_{max}$  before bolting and soil management based on CC. We tested for multicollinearity on correlation coefficients among the nine traits using the variance inflation factor (VIF) adjusted for population size (Allison, 1999; O'Brien, 2007).

## B Plant Bolting

Bolting and floral development data were collected by visual inspection as inferred from the classification of Eguchi and Ichikawa (1940) and Kim et al. (2000), beginning at flower cluster initiation. Additional observations were made to include the periods required to obtain visible, fully formed seed, for a total of seven stages. Briefly, the stages were Bolting 1 from inflorescence emergence up to 25 mm in length (equivalent to Eguchi and Ichikawa Stages 1 to 3, where 1 = flower cluster initiation stage, 2 = flower cluster differentiation stage, and 3 = flower cluster formation stage) (Eguchi and Ichikawa, 1940; Kim et al., 2000); Bolting 2, peduncle elongation to >25 mm; Bolting 3, formation of immature florets; Bolting 4, floret opening; Bolting 5, full flowering, Bolting 6, pollen (males) or mature stigmas (females) visible (Sherry et al., 1993); and Bolting 7, visible seed formation. The DAS to the observation date for each stage was recorded. Observations were recorded for each of the three replicates during two seasons when at least 50% of the male or female plants in each field plot reached each stage. For association studies, the stages were consolidated by averaging the DAS to Bolting 1 and Bolting 2 into early bolting ( $DAS_{bolt}$ ); Bolting 4, Bolting 5, and Bolting 6 into pollination (pollen maturation and pollen shedding,  $DAS_{pol}$ ); and Bolting 7 as the late bolting stage (seed kernel filling,  $DAS_{kf}$ ). All the UAS-imputed and in situ visualization data are included in Supplemental Table S2.

D

## Tissue Collection and DNA Isolation

Tissues from 10 plants per accession were pooled, lyophilized, ground into a fine powder, and stored at  $-80^{\circ}C$ . Genomic DNA was extracted from ~50 mg of lyophilized leaf tissue via the cetyl trimethylammonium bromide method described by Hoisington et al. (1994). DNA concentration and purity were estimated using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA). The DNA samples were subjected to electrophoresis in a 0.8% agarose gel to confirm their integrity prior to sequencing.

## Variant Calling and Genotyping

Genotyping was performed following the ddRADseq genotyping-by-sequencing protocol (Peterson et al., 2012). Illumina short-read sequencing (HiSeq 2500, Illumina Inc., San Diego, CA) and demultiplexing with individual indexes were performed by the Texas A&M AgriLife Genomics and Bioinformatics services. The paired-end raw sequencing reads (150 bp) were subjected to filtering to obtain high-quality reads for downstream analysis. Raw reads were filtered and trimmed for adaptor contamination and low-quality, ambiguous, and

uncalled nucleotide sequences. Sequence reads containing >5% uncalled bases and average quality scores of  $\leq 20$  were discarded. Filtering and trimming of the raw reads were performed with an in-house pipeline developed via Python programming (Python Language Reference version 2.7, Python Software Foundation; <http://www.python.org>, accessed 28 June 2019) (Bedre et al., 2015; Python Core Team, 2015). High-quality cleaned sequence data were aligned to the draft spinach reference genome (Version 1) (Xu et al., 2017) with the Bowtie2 alignment tool (Langmead and Salzberg, 2012).

The 284 spinach accessions were genotyped with Stacks (Version 1.48) (Catchen et al., 2013; Rochette and Catchen, 2017) on the basis of the available spinach reference genome (Version 1) (Xu et al., 2017). In brief, various modules of the Stack pipeline (pstacks, cstacks, sstacks, and rstacks) were used to identify and filter the genotypes (Catchen et al., 2013). The Ada cluster from the Texas A&M University High Performance Research Computing (<http://hprc.tamu.edu/>, accessed 18 June 2019) was used to perform the bioinformatics analysis. The SNP pipeline-end cleanup criteria also included removal of SNPs not anchoring onto the six published draft chromosomes (Xu et al., 2017). A minimum minor allele frequency of >0.05 was used in this study. The 6167 resulting biallelic SNPs in VCF file version 4.2 (Danecek et al., 2011) were used for downstream analyses. Genotypic data can be found in Supplementary Table S3.

### Population Stratification and Kinship Analysis

To account for as much population stratification bias as possible, three different approaches were used to control for population structure: principal component analysis (PCA) (Wold et al., 1987; Endelman and Jannink, 2012), multidimensional scaling (MDS) (Zhu and Yu, 2009), and the allelic ancestry-based admixture model (Falush et al., 2007; Alexander et al., 2009). Given the multiregional origin of the spinach accessions used in this study, PCA with five principal columns was used to account for any continuous spatial genetic variation that might show when the genotypes were separated by continent (Endelman and Jannink, 2012). This step aimed to removing patterns that did not reflect specific genetic drift events, thus helping correct for continuous population structure in this association study. Genotypes were filtered to eliminate monomorphic markers, and missing genotypes were imputed after numericalization as described in Endelman and Jannink (2012). The MDS was estimated with a genetic distance matrix based on the 6167 SNPs; missing genotypes were excluded without imputation.

We also used identity-by-state similarity, which assumes that two random alleles drawn from the same locus are the same. The distance of an individual from itself was set to 0. An admixture model was computed with STRUCTURE (Version 2.3.4) (Pritchard et al., 2000). STRUCTURE was run at 1000 burn-ins and 15 replications on all 6167 SNPs for each K (i.e., the number of the number of computation clusters) value from 1 to 12. The

1 to 12 K range was based on an assumed regional genetic pool from which the mapping collection was drawn (Supplemental Table S1). The optimal K of population structure (Q) was estimated using the Evanno method (Evanno et al., 2005). Briefly, to identify the most likely K value, we examined the rate of change ( $\Delta K$ ) in likelihoods between adjacent K values. To do this, we first determined the proportion of each subpopulation assigned to each cluster to identify the run with the highest log-likelihood among runs for the best K value. The optimal  $\Delta K$  was used to determine the subgroup membership of each accession by 10,000 iterations for each K from 2 to 12.

The graphics were visualized in STRUCTURE HARVESTER (Earl and vonHoldt, 2012), which also applies the CLUMPP algorithm (Jakobsson and Rosenberg, 2007) for label switching. A tree representing the genetic distances among K clusters was constructed with the NEIGHBOR program, which applies the neighbor-joining algorithm (Saitou and Nei, 1987) to the matrix of allele-frequency divergence among K clusters. The plot was produced by DRAWTREE in the PHYLIP package (Felsenstein, 1989). To account for possible hidden allele sharing (Blouin, 2003) that may bias associations within the study, a population cluster's kinship matrix ( $k$ ) was implemented under the nonshrunk (Bradbury et al., 2007) context of a realized relational matrix (Endelman and Jannink, 2012), as the number of markers was greater than the number of individuals genotyped. The genotypes were imputed prior to calculating kinship using the numerical genotype method (Endelman and Jannink, 2012).

### Genome-Wide Association Study

We used TASSEL Version 5.2.2 (Bradbury et al., 2007) (version released 1 July 2017) to implement four mixed linear model (Zhang, 2010) regression approaches for quantitative traits. The statistical hypothesis  $H_0$  (no association with the phenotype) was tested for each SNP and each trait-marker combination for all 6167 markers. We applied compression on the Q +  $k$ , PCA +  $k$ , and MDS +  $k$  models, and compression on Q +  $k$  on a per marker estimate (Q +  $k$  Per Marker) model. The Q, PCA, and MDS were treated as covariates and the  $k$  (kinship) matrix as the random coefficient. With compression and the previously determined default population parameters (P3D), the dimensionality of  $k$  and computational time are reduced and model fitting improves (Bradbury et al., 2007). We used MDS and PCA to test the markers from a purely population genetic distance model and from a purely statistical approach, respectively. The P3D were not implemented for the Q +  $k$  per marker mixed linear model, which allowed each taxon to belong to its own group and allowed us to test each marker independently with each of the nine traits. This also allowed for the estimation of genetic and residual variance for each marker.

For the models run on the P3D, the genetic and residual variances were estimated at the trait level. Effects were determined for each marker, independent of the compression model. For each compression level, likelihood, genetic

variance, and error variance were determined and the compression level with the lowest value of  $-2\text{Ln}$ -likelihood for each trait–compression combination (Supplemental Table S4) was used to test the markers.

Genome-wide significance thresholds were determined for each trait–marker combination via the Benjamini–Hochberg false discovery rate method (Benjamini and Hochberg, 1995) at  $\alpha = 0.05$  for critical  $P$ -values of  $1.60 \times 10^{-4}$  for  $\text{CC}_{\text{max}}$ ,  $4.60 \times 10^{-5}$  for  $\text{CVol}_{\text{max}}$ ,  $2.47 \times 10^{-5}$  for  $\text{EXG}_{\text{max}}$ ,  $2.6 \times 10^{-4}$  for  $\text{DAS}_{\text{CC}}$ ,  $1.05 \times 10^{-4}$  for  $\text{DAS}_{\text{CVol}}$ ,  $2.6 \times 10^{-4}$  for  $\text{DAS}_{\text{ExG}}$ ,  $1.69 \times 10^{-4}$  for  $\text{DAS}_{\text{bolt}}$ ,  $1.38 \times 10^{-4}$  for  $\text{DAS}_{\text{pol}}$ , and  $1.38 \times 10^{-4}$  for  $\text{DAS}_{\text{kf}}$ . Markers not meeting the cutoffs were not considered to be associated with the traits under study. Manhattan plots were drawn with the qqman R package (Turner, 2014).

### Linkage Disequilibrium and Linkage Disequilibrium Decay

To determine the extent of linkage disequilibrium (LD), LD decay, and LD blocks across the genome, we used TASSEL Version 5.2.2 to examine the 6167 markers used in this study (with a sliding window size of 500 and 29,582,250 comparisons), then a random 1000 markers (a sliding window size of 100 and 72,640 comparisons) and another random subset of 100 markers (a sliding window size of 40 and 3430 comparisons). Heterozygotes were set to missing, LD was estimated as squared allele frequency correlations ( $r^2$ ); LD decay was calculated on the basis of the correlation coefficient ( $r^2$ ) and displayed with  $P$ -values and positional information. We compared these results with the LD results that used the 99 markers with significant association signals (LD was determined with a sliding window size of 50 and 4525 comparisons). The extent of LD, LD decay, and LD blocks were strikingly similar in all four sets. The LD map plots were displayed with LDplot in TASSEL Version 5.2.2. Allele phasing was not performed for the markers in LD.

### Heritability Analysis and Variant (SNP) Effects

We used the restricted (residual) maximum likelihood to decompose the variance components on the phenotypic data with both the lines (accessions) and replicates (with blocks nested within replicates) treated as random effects in a full factorial regression to the mean. The best linear unbiased prediction method was used to obtain point estimates of these random effects in a mixed effect model in JMP Version 14.0 (SAS Institute, Cary, NC). Point prediction estimates were generated for each line. The best linear unbiased predictions have minimum mean squared error and thus provide the average value of the estimates, which is close to the mean linear function of the data (Robinson, 1991). We used correlation analysis to test the relationship between the predicted and the observed (or imputed) (Bernardo, 1996) values of the nine traits.

Broad-sense heritability ( $H^2$ ) was on a line-mean basis, with replicates nested in three blocks in a single location [Eq. 1]:

$$H^2 = \frac{V_l^2}{V_l^2 + \frac{Vl \times V_r^2 [\text{Block}]}{3} + \frac{V_e^2}{3}}, \quad [1]$$

where  $V_l$  is the variance within a line,  $r$  is the replicate, and  $e$  is the residual; these are divided by the replicates nested in three blocks.

Narrow-sense heritability ( $h^2$ ) was calculated for each marker from the outputs of the  $Q + k$  Per Marker model parameters. Per-trait  $h^2$  was obtained from the mean genetic variance and the mean residual variance from the compression models as follows:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \quad [2]$$

where  $\sigma_a^2$  is the additive variance of the marker allele and  $\sigma_e^2$  is the variance of the residual (error). For each compression level, the likelihood, genetic variance, and error variance were determined. Marker effects (dominance and additive for each marker) estimated by each of the four GWAS models were tested at  $\alpha = 0.05$ .

### Polymorphic Variants and Genomic Gene Context Assignment

In silico analysis of the tag SNPs was used to locate genomic features anchored in the draft spinach reference genome. Only markers showing insignificant LD decay and that met the significance thresholds in the GWAS were used. Putative functional annotations were determined through BLAST searches in SpinachBase ([www.spinachbase.org](http://www.spinachbase.org), accessed 18 June 2019) and Gramene BLAST (<http://www.gramene.org>, accessed 18 June 2019). For the purpose of this study, the gene at the association site itself was reported. However, when no anchor gene on the associated site was identified, the gene within 200 kb upstream or downstream of the site was reported, since it is possible that the SNP markers cosegregated with genes hundreds of kb apart (Meuwissen et al., 2014).

## RESULTS

### Phenotypic Distribution and Correlation among Traits

An autonomous aircraft vehicle equipped with an RGB sensor captured field images of 284 spinach accessions in three replications over a ~12-wk period (Fig. 1). Canopy volume, CC, and ExG values were estimated from orthomosaic images and fitted to univariate third-level polynomial functions, as shown in Fig. 1b. The coefficients of determination ( $r^2$ ) averaged  $0.87 \pm 0.09$  for CVol,  $0.88 \pm 0.11$  for CC, and  $0.94 \pm 0.17$  for ExG. The maximum crop growth values during the season ( $\text{CVol}_{\text{max}}$ ,  $\text{CC}_{\text{max}}$ , and  $\text{ExG}_{\text{max}}$ ) and DAS to reach the phenotypic maximum values ( $\text{DAS}_{\text{CVol}}$ ,  $\text{DAS}_{\text{CC}}$ , and  $\text{DAS}_{\text{ExG}}$ ) were determined for each spinach accession and replication. The  $\text{CVol}_{\text{max}}$ ,  $\text{ExG}_{\text{max}}$ , and  $\text{CC}_{\text{max}}$  values showed a normal distribution (Shapiro–Wilk test), whereas the associated DAS values did not (Fig. 2). The phenotypic variance for estimated maximum plant growth parameters ranged from 0.0004 for  $\text{ExG}_{\text{max}}$  (mean = 0.3) to 0.026 for  $\text{CC}_{\text{max}}$  (mean = 59.48%) and 0.026 for  $\text{CVol}_{\text{max}}$  (mean = 0.08  $\text{m}^3$ ). Ground-based data associated with flowering and

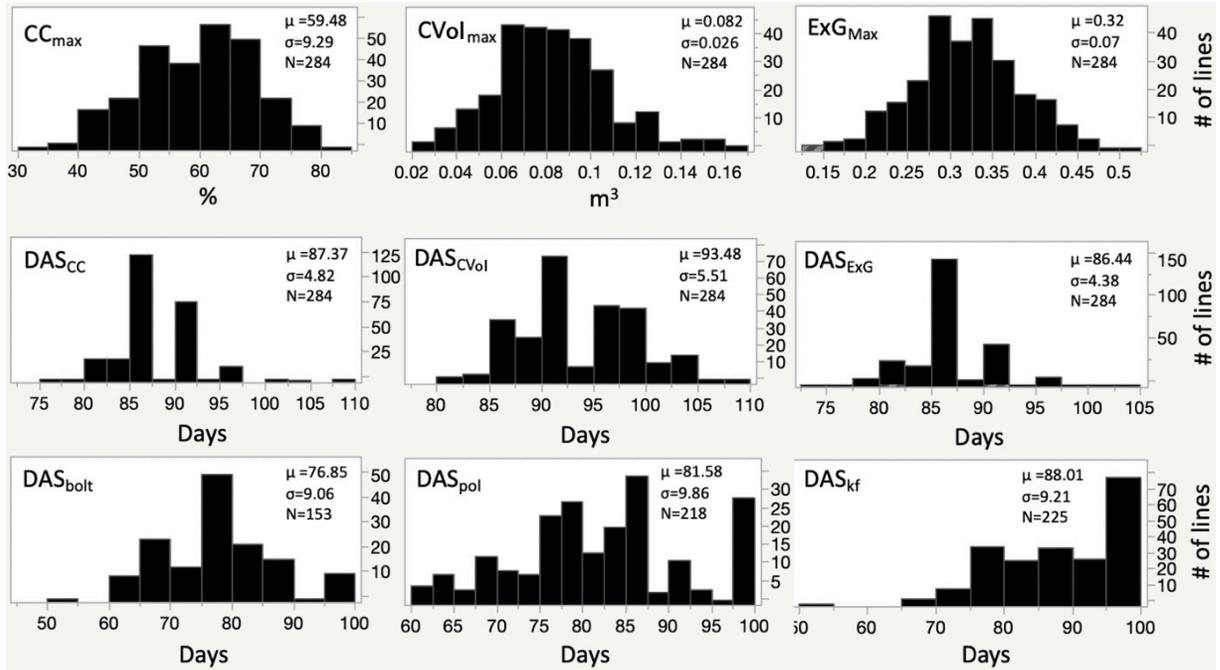


Fig. 2. Distribution of raw phenotypic means. Maximum seasonal values for spinach canopy cover ( $CC_{max}$ ), canopy volume ( $CVol_{max}$ ), and excess greenness index ( $ExG_{max}$ ); days after sowing to maximum seasonal values of canopy cover ( $DAS_{CC}$ ), canopy volume ( $DAS_{CVol}$ ), and excess greenness index ( $DAS_{ExG}$ ); and days after sowing for manually collected plant bolting stages: early bolting ( $DAS_{bolt}$ ), pollination ( $DAS_{pol}$ ), and kernel filling ( $DAS_{kf}$ ).

seed development such as  $DAS_{bolt}$ ,  $DAS_{pol}$ , and  $DAS_{kf}$  were visually collected and were skewed toward days of advanced plant development (Fig. 2). Their variance was the highest observed out of the nine traits measured.

Correlations between  $CC_{max}$  and  $CVol_{max}$  ( $r = 0.67$ ),  $CC_{max}$  and  $ExG_{max}$  ( $r = 0.95$ ),  $CVol_{max}$  and  $ExG_{max}$  ( $r = 0.68$ ), and  $DAS_{CVol}$  and  $DAS_{ExG}$  ( $r = 0.40$ ) were significant at  $\alpha = 0.05$  (Table 1). The rest of the UAS-collected variables did not show significant correlations. For the manually phenotyped flowering and seed development stages, there were strong correlations between  $DAS_{bolt}$  and  $DAS_{pol}$  ( $r = 0.85$ ),  $DAS_{bolt}$  and  $DAS_{kf}$  ( $r = 0.76$ ), and

$DAS_{pol}$  and  $DAS_{kf}$  ( $r = 0.88$ ). The correlations between the UAS-based data and the ground bolting data were tested but found to be nonsignificant. Because of the significant correlations between some traits, we tested collinearity by determining the VIF (Table 1) of their bivariate  $r^2$  (Lipovetsky and Conklin, 2001; Kock and Lynn, 2012). The VIF estimates how much the variance of a coefficient of correlation is inflated (Allison, 1999). As recommended by O'Brien (2007), we used multicollinearity testing to account for the effect of population size on variance inflation. No collinearity was found between variables, except for  $ExG_{max}$  and  $CC_{max}$  ( $r = 0.95$ ,  $VIF = 9.72776$ ).

Table 1. Correlation coefficients† and variance inflation factors (VIF)‡ among traits.

|              | $CC_{max}$ ¶ | $DAS_{CC}$ | $CVol_{max}$ | $DAS_{CVol}$ | $ExG_{max}$ | $DAS_{ExG}$ | $DAS_{bolt}$ | $DAS_{pol}$ | $DAS_{kf}$ |
|--------------|--------------|------------|--------------|--------------|-------------|-------------|--------------|-------------|------------|
| $CC_{max}$   | –            | 1.00107    | 1.79321      | 1.01177      | 9.72776§    | 1.00076     | 1.00031      | 1.00556     | 1.01348    |
| $DAS_{CC}$   | –0.03        | –          | 1.00005      | 1.00058      | 1.00042     | 1.00025     | 1.00238      | 1.00304     | 1.00019    |
| $CVol_{max}$ | 0.67*        | –0.01      | –            | 1.04629      | 1.87938     | 1.00000     | 1.00000      | 1.00148     | 1.00183    |
| $DAS_{CVol}$ | 0.11         | –0.02      | 0.21         | –            | 1.00399     | 1.18760     | 1.01000      | 1.00085     | 1.00030    |
| $ExG_{max}$  | 0.95*        | –0.02      | 0.68*        | 0.06         | –           | 1.00558     | 1.00040      | 1.00323     | 1.01016    |
| $DAS_{ExG}$  | –0.03        | –0.02      | 0.00         | 0.40*        | –0.07       | –           | 1.00017      | 1.00000     | 1.00010    |
| $DAS_{bolt}$ | 0.02         | 0.05       | 0.00         | 0.10         | 0.02        | 0.01        | –            | 3.70100     | 2.35319    |
| $DAS_{pol}$  | –0.07        | 0.06       | –0.04        | –0.03        | –0.06       | 0.00        | 0.85*        | –           | 4.50986    |
| $DAS_{kf}$   | –0.12        | –0.01      | –0.04        | –0.02        | –0.10       | –0.01       | 0.76*        | 0.88*       | –          |

\* Significant at the 0.05 probability level.

† Correlation coefficients are shown in the lower triangle.

‡ VIFs are shown in the upper triangle.

§ Significant collinearity at a tolerance level of 0.25.

¶  $CC_{max}$ , maximum canopy cover;  $CVol_{max}$ , maximum canopy volume;  $DAS_{bolt}$ , days after sowing until early bolting;  $DAS_{CC}$ , days after sowing until maximum canopy cover;  $DAS_{CVol}$ , days after sowing until maximum canopy volume;  $DAS_{ExG}$ , days after sowing until maximum excess greenness index;  $DAS_{pol}$ , days after sowing until pollination;  $DAS_{kf}$ , days after sowing until kernel filling;  $ExG_{max}$ , maximum excess greenness index.

Table 2. Single nucleotide polymorphism (SNP) discovery rate as a function of SNP density relative to the physical distance spanning the mapped region of each chromosome.

| Chromosome | SNPs     | Physical length spanned<br>bp | SNP† discovery rate per chromosome |
|------------|----------|-------------------------------|------------------------------------|
|            | <i>n</i> |                               | bp per SNP                         |
| 1          | 1027     | 50,544,189                    | 49,215.4                           |
| 2          | 1027     | 60,452,741                    | 58,863.4                           |
| 3          | 1060     | 112,131,948                   | 105,784.9                          |
| 4          | 1458     | 122,918,674                   | 84,306.4                           |
| 5          | 757      | 69,232,026                    | 91,455.9                           |
| 6          | 838      | 46,051,220                    | 54,953.7                           |

† Filtered SNPs.

### Single Nucleotide Polymorphism Density and Population Structure

At the genome level, we determined the total number of SNPs discovered per chromosome (chr) and their densities. Chromosome 4 had the highest number of SNPs discovered, whereas chr1, chr2, and chr3 had intermediate SNP counts and chr5 and chr6 had the lowest number of SNPs identified (Table 2). However, when we compared the corresponding physical genomic stretch spanned (bp for every SNP), chr1 had the greatest SNP density, followed by chr6 and chr2 (Table 2). We used the SNP allele distribution to determine the population structure of the spinach panel. The admixture model grouped the 284 accessions into five population clusters ( $k = 5$ ), based on the highest  $\Delta k$  (Fig. 3).

### Significant Markers, Marker Effects, and Heritability

In this study, associations based on the  $P$ -value of the  $F$ -test for the model were called when the Benjamini–Hochberg correction cutoff was met. We used  $P < 0.05$  for the  $F$ -test for the additive model and for the  $F$ -test for dominance after fitting the additive model to test the significance of the additive and dominance effects, respectively. We determined heritability ( $H^2$ ) on the basis of

the line mean and the marker effects, and per significant marker  $h^2$  (Table 3-7).

Ninety-nine SNPs were significantly associated with measured spinach growth parameters (Fig. 4, Table 4-7). Of these, 56% were associated with the UAS-generated traits; the rest (44%) were associated with the visually collected traits (Table 4-7). All identified UAS trait markers were associated with DAS variables except  $CC_{max}$ , for which the markers 42970\_33, 42970\_120, and 42970\_132 were identified (Table 4-7). These  $CC_{max}$ -associated makers are substitution polymorphisms <55 bp apart, consecutively on the same gene (i.e., *Spo18796*, with the locus segment type given in parentheses): A-G (exon), A-G (intron), and T-A (intron), respectively. Their effect alleles were G (additive), G (additive + dominance), and A (additive) (Table 4-7). These observations suggest that the three  $CC_{max}$ -associated SNP variants may be nonsynonymous. On the basis of a similar test criterion, other marker clusters in Table 4-7 may be classified. However, this conclusion may require follow-up studies to determine which amino acid residues are affected by the base changes, as well as the effect of those changes on the function of the translated protein.

Only a single locus colocalization was identified between markers associated with the manually collected traits during bolting and the markers associated with crop growth traits derived from UAS phenotyping. The marker 42970\_31 (Position 34,290,002) for  $DAS_{pol}$  and the three markers 42970\_33, 42970\_120, and 42970\_132 42970\_ (positioned between 34,290,043 and 34,290,103) for  $CC_{max}$  colocalized at the gene level on *Spo18796*, which is a transcription factor jumonji domain on chr6, suggesting that these markers may interact epistatically to affect the possible pleiotropic nature of *Spo18796* in the two traits.

In general, with a few exceptions,  $h^2$  on a trait-mean basis was higher for the visually (manually) collected traits than for the UAS-generated data in all the models.

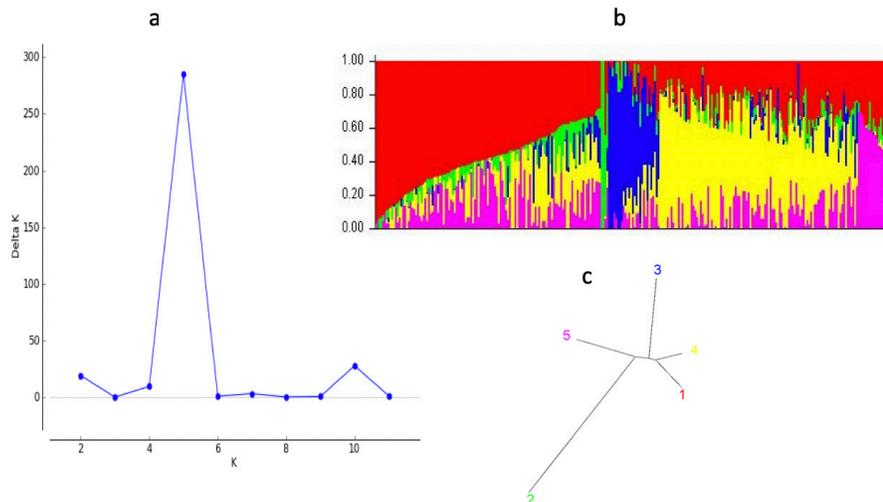


Fig. 3. Population structure estimates via the admixture model for the 284 spinach accessions. (a) Likelihood estimate plot showing the number of genetically distinct clusters ( $K$ ) for the 284 accessions. (b) Single-line estimates of genetically distinct clusters. Colors (red, yellow, pink, green, and blue) represent the five estimated population clusters used in this study. The horizontal axis shows the individuals genotyped and the vertical axis represents the allelic proportions defining the position in the cluster. (c) Representation of the average genetic distances among  $K$  clusters.

Table 3. Trait heritability calculated via the four association models.

| Trait               | Model statistics† |                |      | Line mean basis (REML) | Per marker basis (GWA)             | Per trait basis (GWA)   |                           |                           |
|---------------------|-------------------|----------------|------|------------------------|------------------------------------|-------------------------|---------------------------|---------------------------|
|                     | Least square mean | R <sup>2</sup> | CV   | H <sup>2</sup>         | h <sup>2</sup> (Q + k Per Marker)‡ | h <sup>2</sup> (Q + k)§ | h <sup>2</sup> (MDS + k)§ | h <sup>2</sup> (PCA + k)§ |
| CC <sub>max</sub> ¶ | 59.09             | 0.132          | 6.21 | 0.0173                 | 0.166                              | 0.163                   | 0.529                     | 0.168                     |
| CVol <sub>max</sub> | 0.08              | 0.069          | 1.44 | 0.0936                 | 0.513                              | 0.503                   | 0.641                     | 0.530                     |
| ExG <sub>max</sub>  | 0.32              | 0.098          | 8.04 | 0.0019                 | 0.211                              | 0.194                   | 0.580                     | 0.270                     |
| DAS <sub>CC</sub>   | 87.40             | 0.036          | 0.97 | 0.0000                 | 0.489                              | 0.459                   | 0.421                     | 0.533                     |
| DAS <sub>CVol</sub> | 93.42             | 0.153          | 2.53 | 0.0700                 | 0.013                              | 1.000                   | 0.471                     | 0.437                     |
| DAS <sub>ExG</sub>  | 86.35             | 0.018          | 0.34 | 0.0009                 | 0.277                              | 0.278                   | 0.565                     | 0.283                     |
| DAS <sub>bol</sub>  | 84.71             | 0.002          | 0.60 | 0.0704                 | 0.899                              | 0.898                   | 0.740                     | 0.905                     |
| DAS <sub>pol</sub>  | 86.86             | 0.134          | 2.87 | 0.0808                 | 0.481                              | 0.446                   | 0.641                     | 0.552                     |
| DAS <sub>kf</sub>   | 90.95             | 0.128          | 2.30 | 0.0034                 | 0.764                              | 0.772                   | 0.596                     | 0.776                     |

† Restricted (residual) maximum likelihood (REML) or its components.

‡ Compression model with variance components estimated for each marker separately.

§ Models used to calculate single compression mean genetic variance and single compression mean residual variance on a trait basis.

¶ CC<sub>max</sub>, maximum canopy cover; CVol<sub>max</sub>, maximum canopy volume; DAS<sub>bol</sub>, days after sowing until early bolting; DAS<sub>CC</sub>, days after sowing until maximum canopy cover; DAS<sub>CVol</sub>, days after sowing until maximum canopy volume; DAS<sub>ExG</sub>, days after sowing until maximum excess greenness index; DAS<sub>pol</sub>, days after sowing until pollination; DAS<sub>kf</sub>, days after sowing until kernel filling; ExG<sub>max</sub>, maximum excess greenness index; GWA, genome-wide association; H<sup>2</sup>, broad-sense heritability; h<sup>2</sup>, narrow-sense heritability; k, kinship matrix; MDS, multidimensional scaling; PCA, principal component analysis; Q, the number of computation clusters with a list of genetic distance probabilities for each accession in the population.

For example, the combined mean h<sup>2</sup> for the UAS-imputed traits were 27.8, 43.2, 53.4, and 37.0%, whereas those for the manually determined traits were 71.4, 70.5, 65.9, and 74.3% from the models Q + k Per Marker, Q + k, MDS + k, and PCA + k, respectively. The mean H<sup>2</sup> values on a line-mean basis were much lower for both the UAS-imputed and manually determined traits at 3.10% (mean

R<sup>2</sup>, 0.08; CV, 3.35%) and 5.15% (R<sup>2</sup>, 0.13; CV, 1.92%), respectively (Table 3). To achieve an experimental power of at least 0.8, assuming we had targeted H<sup>2</sup> of 0.75 for each of the traits at α = 0.05 (two-tailed) at the low R<sup>2</sup> of 0.08 (as in our results above) and a Type II error rate (β) of 0.2, the effective population size would be at least 194 accessions (plots). Since we tested 284 accessions, the

Table 4. Markers meeting the significance threshold in the Q + k model (per trait estimate).

| Trait              | Markers             | Representative marker | Chromosome | Position and/or range | Q + k                 |          |          |          |                |       |
|--------------------|---------------------|-----------------------|------------|-----------------------|-----------------------|----------|----------|----------|----------------|-------|
|                    |                     |                       |            |                       | Add.†                 | P-value  | Dom.     | P-value  | R <sup>2</sup> |       |
|                    | <i>n</i>            |                       |            |                       |                       |          |          |          |                |       |
| DAS <sub>CC</sub>  | 10                  | 38939_110             | 5          | 2,811,551–2,811,739   | 0.17851               | 1.48E-06 | −0.17839 | 4.47E-06 | 0.088          |       |
|                    | 1                   | 19651_36              | 1          | 3,846,589             | 0.13640               | 2.14E-05 | −0.11631 | 6.21E-04 | 0.081          |       |
|                    | 1                   | 17771_95              | 1          | 12,969,303            | −0.21635              | 1.67E-05 | −0.20997 | 1.04E-04 | 0.068          |       |
|                    | 2                   | 38455_121             | 5          | 21,285,477–21,285,566 | 0.17858               | 1.49E-06 | −0.16772 | 1.32E-05 | 0.090          |       |
|                    | 1                   | 44084_132             | 6          | 43,509,344            | 0.11502               | 9.24E-05 | −0.12028 | 7.07E-05 | 0.072          |       |
|                    | 1                   | 36324_127             | 4          | 80,322,196            | −0.22792              | 6.30E-06 | −0.18268 | 4.27E-04 | 0.102          |       |
|                    | DAS <sub>CVol</sub> | 2                     | 31710_51   | 4                     | 10,286,541–10,286,542 | 1.33003  | 1.79E-05 | −1.40687 | 9.78E-06       | 0.075 |
|                    |                     | 1                     | 26862_9    | 3                     | 15,991,151            | −0.77364 | 1.94E-05 | −0.81840 | 3.42E-05       | 0.070 |
|                    |                     | 3                     | 27988_137  | 3                     | 34,577,820–34,577,828 | −1.37576 | 1.01E-05 | −1.34997 | 1.83E-05       | 0.080 |
|                    |                     | 1                     | 22953_48   | 2                     | 36,249,915            | 1.33259  | 2.04E-05 | −1.39706 | 2.30E-05       | 0.075 |
| 1                  |                     | 44035_122             | 6          | 43,347,813            | 0.89791               | 4.17E-07 | −1.08089 | 4.47E-08 | 0.114          |       |
| 1                  |                     | 44180_7               | 6          | 44,500,215            | −1.30507              | 2.43E-05 | −1.46595 | 4.02E-06 | 0.080          |       |
| 1                  |                     | 34883_38              | 4          | 51,195,861            | 0.54333               | 2.43E-05 | −0.40724 | 6.86E-03 | 0.073          |       |
| 1                  |                     | 24363_107             | 2          | 51,681,433            | 0.64229               | 3.74E-05 | −0.57303 | 4.08E-03 | 0.071          |       |
| 1                  |                     | 29971_66              | 3          | 70,498,159            | −0.64269              | 4.38E-06 | −0.62923 | 8.85E-05 | 0.082          |       |
| DAS <sub>ExG</sub> |                     | 1                     | 40827_128  | 5                     | 7,581,055             | 0.02778  | 1.73E-05 | −0.03089 | 3.96E-06       | 0.080 |
|                    | 1                   | 34151_101             | 4          | 29,043,034            | 0.01501               | 5.56E-06 | −0.01512 | 2.38E-05 | 0.83           |       |
|                    | 1                   | 19887_120             | 1          | 40,388,453            | −0.03043              | 2.22E-06 | −0.02856 | 1.86E-05 | 0.085          |       |
|                    | 1                   | 20358_139             | 1          | 44,357,800            | 0.03056               | 2.64E-06 | −0.02740 | 3.32E-05 | 0.111          |       |
|                    | 1                   | 36231_148             | 4          | 78,521,968            | −0.02429              | 8.80E-08 | −0.02156 | 2.35E-05 | 0.114          |       |
|                    | 1                   | 26106_55              | 3          | 102,662,772           | 0.03053               | 2.35E-06 | −0.02952 | 5.83E-06 | 0.089          |       |
| DAS <sub>bol</sub> | 1                   | 36822_84              | 4          | 88,252                | 0.02191               | 2.24E-05 | −0.02012 | 5.57E-04 | 0.200          |       |
|                    | 1                   | 26785_87              | 3          | 1,532,247             | 0.02203               | 1.48E-05 | −0.01458 | 3.61E-03 | 0.214          |       |
|                    | 1                   | 30014_141             | 3          | 7,122,451             | −0.02087              | 1.41E-05 | −0.01949 | 9.06E-04 | 0.210          |       |

† Add., additive effect; DAS<sub>bol</sub>, days after sowing until early bolting; DAS<sub>CC</sub>, days after sowing until maximum canopy cover; DAS<sub>CVol</sub>, days after sowing until maximum canopy volume; DAS<sub>ExG</sub>, days after sowing until maximum excess greenness index; Dom., dominance effect; k, kinship matrix; Q, cluster of genetic distance probabilities for each accession in the population; R<sup>2</sup>, phenotypic variance explained by the effect model.

Table 5. Markers meeting the significance threshold in the MDS + k model (per trait estimate).

| Trait               | Markers            | Representative marker | Chromosome | Position and/or range | MDS + k             |          |          |          |                |       |
|---------------------|--------------------|-----------------------|------------|-----------------------|---------------------|----------|----------|----------|----------------|-------|
|                     |                    |                       |            |                       | Add.†               | P-value  | Dom.     | P-value  | R <sup>2</sup> |       |
|                     | <i>n</i>           |                       |            |                       |                     |          |          |          |                |       |
| DAS <sub>CC</sub>   | 10                 | 38939_110             | 5          | 2,811,551–2,811,739   | 0.18331             | 5.94E-06 | -0.17897 | 1.79E-05 | 0.078          |       |
|                     | 1                  | 19651_36              | 1          | 3,846,589             | 0.14825             | 1.75E-05 | -0.13275 | 1.92E-04 | 0.090          |       |
|                     | 2                  | 38455_121             | 5          | 21,285,477–21,285,566 | 0.18245             | 6.06E-06 | -0.17105 | 4.21E-05 | 0.081          |       |
|                     | 1                  | 22984_88              | 2          | 36,778,324            | -0.99955            | 2.72E-05 | -1.16185 | 1.07E-05 | 0.161          |       |
|                     | 1                  | 36324_127             | 4          | 80,322,196            | -0.22762            | 3.18E-05 | -0.18999 | 6.86E-04 | 0.082          |       |
|                     | DAS <sub>ExG</sub> | 6                     | 29372_33   | 3                     | 5,889,084–5,889,263 | -0.02973 | 4.58E-06 | -0.03112 | 7.09E-06       | 0.079 |
|                     |                    | 1                     | 40827_128  | 5                     | 7,581,055           | 0.02941  | 5.78E-06 | -0.03284 | 1.10E-06       | 0.088 |
|                     |                    | 1                     | 34151_101  | 4                     | 29,043,034          | 0.01551  | 4.13E-06 | -0.01598 | 1.23E-05       | 0.084 |
|                     |                    | 1                     | 19887_120  | 1                     | 40,388,453          | -0.03085 | 2.78E-06 | -0.03001 | 1.13E-05       | 0.081 |
|                     |                    | 1                     | 43709_72   | 6                     | 40,598,206          | -0.02969 | 4.71E-06 | -0.03105 | 5.41E-06       | 0.078 |
| 1                   |                    | 20358_139             | 1          | 44,357,800            | 0.03158             | 2.27E-06 | -0.02866 | 2.35E-05 | 0.109          |       |
| 1                   |                    | 44261_56              | 6          | 45,192,349            | 0.02986             | 4.06E-06 | -0.03088 | 7.13E-06 | 0.079          |       |
| 1                   |                    | 44291_137             | 6          | 45,352,029            | -0.03041            | 3.01E-06 | -0.02822 | 2.28E-05 | 0.092          |       |
| 1                   |                    | 28665_72              | 3          | 46,395,669            | -0.03009            | 3.47E-06 | -0.02688 | 7.51E-05 | 0.083          |       |
| 1                   |                    | 36231_148             | 4          | 78,521,968            | -0.02403            | 1.98E-07 | -0.02076 | 5.45E-05 | 0.109          |       |
| DAS <sub>bolr</sub> | 1                  | 36822_84              | 4          | 88,252                | 0.02100             | 4.84E-06 | -0.01954 | 1.80E-04 | -0.103         |       |
|                     | 1                  | 22357_36              | 2          | 27,309,513            | 0.00608             | 3.99E-02 | 0.00466  | 1.75E-01 | 0.090          |       |
| DAS <sub>Pol</sub>  | 2                  | 37716_63              | 5          | 11,878,605–11,878,606 | -1.02884            | 1.41E-05 | -0.93074 | 1.43E-04 | 0.142          |       |
|                     | 1                  | 39553_80              | 5          | 40,016,005            |                     |          |          |          |                |       |
| DAS <sub>kf</sub>   | 2                  | 43868_4               | 6          | 42,083,968–42,084,406 | -0.24693            | 1.30E-04 | -0.34147 | 8.62E-05 | 0.146          |       |
|                     | 1                  | 28312_106             | 3          | 4,026,915             | 0.27098             | 5.65E-05 | 0.20419  | 1.81E-02 | 0.068          |       |
|                     | 2                  | 35474_60              | 4          | 6,325,624–6,325,642   | -1.26120            | 6.40E-05 | -1.44458 | 9.38E-06 | 0.076          |       |
|                     | 1                  | 34672_102             | 4          | 46,706,902            | 0.79822             | 2.16E-06 | -1.00418 | 1.35E-07 | 0.104          |       |
|                     | 1                  | 30538_30              | 3          | 83,500,450            | 0.16537             | 1.52E-05 | -0.15352 | 3.36E-04 | 0.072          |       |
|                     | 1                  | 32020_29              | 4          | 107,271,059           | 0.05712             | 1.28E-04 | -0.01268 | 4.76E-01 | 0.080          |       |
|                     | 1                  | 32388_65              | 4          | 114,173,692           | 0.06281             | 8.37E-05 | -0.02055 | 2.66E-01 | 0.075          |       |

† Add., additive effect; DAS<sub>bolr</sub>, days after sowing until early bolting; DAS<sub>CC</sub>, days after sowing until maximum canopy cover; DAS<sub>ExG</sub>, days after sowing until maximum excess greenness index; DAS<sub>Pol</sub>, days after sowing until pollination; DAS<sub>kf</sub>, days after sowing until kernel filling; Dom., dominance effect; k, kinship matrix; MDS, multidimensional scaling; R<sup>2</sup>, phenotypic variance explained by the effect model.

Table 6. Markers meeting the significance threshold in the PCA + k model (per trait estimate).

| Trait               | Markers  | Representative marker | Chromosome | Position and/or range | PCA + k  |          |          |          |                |
|---------------------|----------|-----------------------|------------|-----------------------|----------|----------|----------|----------|----------------|
|                     |          |                       |            |                       | Add.†    | P-value  | Dom.     | P-value  | R <sup>2</sup> |
|                     | <i>n</i> |                       |            |                       |          |          |          |          |                |
| DAS <sub>CC</sub>   | 10       | 38939_110             | 5          | 2,811,551–2,811,739   | 0.17299  | 1.99E-06 | -0.17128 | 8.01E-06 | 0.084          |
|                     | 2        | 38455_121             | 5          | 21,285,477–21,285,566 | 0.17124  | 2.58E-06 | -0.15882 | 2.72E-05 | 0.085          |
|                     | 1        | 36324_127             | 4          | 80,322,196            | -0.22295 | 7.26E-06 | -0.18143 | 3.81E-04 | 0.095          |
| DAS <sub>CVol</sub> | 1        | 44035_122             | 6          | 43,347,813            | 0.91678  | 6.35E-06 | -1.09051 | 1.14E-06 | 0.116          |
| DAS <sub>ExG</sub>  | 1        | 40827_128             | 5          | 7,581,055             | -0.02451 | 8.58E-08 | -0.02132 | 2.73E-05 | 0.115          |
|                     | 1        | 40828_107             | 5          | 7,581,143             | -0.02796 | 2.40E-05 | -0.02894 | 1.70E-05 | 0.070          |
|                     | 1        | 34151_101             | 4          | 29,043,034            | 0.02779  | 2.60E-05 | -0.02916 | 1.92E-05 | 0.070          |
|                     | 1        | 19887_120             | 1          | 40,388,453            | -0.02960 | 3.49E-06 | -0.02717 | 4.16E-05 | 0.083          |
|                     | 1        | 43709_72              | 6          | 40,598,206            | -0.02829 | 1.61E-05 | -0.03038 | 1.19E-05 | 0.073          |
|                     | 1        | 20358_139             | 1          | 44,357,800            | 0.02962  | 4.23E-06 | -0.02634 | 5.39E-05 | 0.108          |
|                     | 1        | 44281_23              | 6          | 45,342,253            | 0.02867  | 1.42E-05 | -0.02748 | 4.63E-05 | 0.070          |
|                     | 1        | 24160_135             | 2          | 50,245,050            | -0.02897 | 1.05E-05 | -0.02566 | 1.67E-04 | 0.076          |
|                     | 1        | 36231_148             | 4          | 78,521,968            | 0.01485  | 4.83E-06 | -0.01480 | 3.00E-05 | 0.084          |
|                     | 1        | 26106_55              | 3          | 102,662,772           | 0.02984  | 3.27E-06 | -0.02860 | 9.45E-06 | 0.087          |
| DAS <sub>bolr</sub> | 1        | 26785_87              | 3          | 1,532,247             | 0.02779  | 2.30E-05 | -0.03096 | 4.95E-06 | 0.079          |
| DAS <sub>Pol</sub>  | 1        | 42970_31              | 6          | 34,290,002–34,290,103 | 0.52472  | 1.21E-04 | -0.33108 | 2.93E-02 | 0.062          |

† Add., additive effect; DAS<sub>Pol</sub>, days after sowing until pollination; DAS<sub>CC</sub>, days after sowing until maximum canopy cover; DAS<sub>CVol</sub>, days after sowing until maximum canopy volume; Dom., dominance effect; DAS<sub>ExG</sub>, days after sowing until maximum excess greenness index; DAS<sub>bolr</sub>, days after sowing until early bolting; k, kinship matrix; PCA, principal component analysis; R<sup>2</sup>, phenotypic variance explained by the effect model.

Table 7. Markers meeting the significance threshold in  $Q + k$  models (per marker estimate).

| Trait               | Markers | Representative marker | Chromosome | Position and/or range   | $Q + k$  |          |          |          |                |                |
|---------------------|---------|-----------------------|------------|-------------------------|----------|----------|----------|----------|----------------|----------------|
|                     |         |                       |            |                         | Add.†    | P-value  | Dom.     | P-value  | R <sup>2</sup> | h <sup>2</sup> |
|                     |         | <i>n</i>              |            |                         |          |          |          |          |                |                |
| CC <sub>max</sub>   | 3       | 42970_120             | 6          | 34,290,044–34,290,103   | 0.16891  | 7.23E-06 | -0.15721 | 1.79E-04 | 0.071          | 0.33           |
| EXG <sub>max</sub>  | 2       | 41834_137             | 6          | 21,986,964–21,990,581   | -0.03267 | 2.91E-04 | -0.02059 | 3.11E-02 | 0.188          | 1.00           |
| DAS <sub>CC</sub>   | 10      | 38939_110             | 5          | 2,811,551–2,811,739     | 0.17818  | 4.38E-07 | -0.17797 | 1.51E-06 | 0.089          | 0.52           |
|                     | 1       | 19651_36              | 1          | 3,846,589               | 0.13725  | 6.20E-06 | -0.11655 | 2.80E-04 | 0.080          | 0.37           |
|                     | 1       | 17771_95              | 1          | 12,969,303              | -0.21605 | 8.64E-06 | -0.20974 | 6.05E-05 | 0.069          | 0.51           |
|                     | 2       | 38455_121             | 5          | 21,285,477–21,285,566   | 0.17759  | 4.66E-07 | -0.16760 | 4.63E-06 | 0.093          | 0.57           |
|                     | 1       | 22984_88              | 2          | 36,778,324              | -1.00159 | 8.41E-06 | -1.14877 | 4.31E-06 | 0.147          | 0.26           |
|                     | 1       | 44084_132             | 6          | 43,509,344              | 0.11459  | 5.32E-05 | -0.12070 | 3.58E-05 | 0.072          | 0.50           |
|                     | 1       | 36324_127             | 4          | 80,322,196              | -0.22742 | 1.41E-06 | -0.18199 | 1.70E-04 | 0.104          | 0.54           |
| DAS <sub>CVol</sub> | 2       | 31710_51              | 4          | 10,286,541–10,286,542   | 1.33003  | 8.66E-06 | -1.40687 | 4.53E-06 | 0.075          | 0.00           |
|                     | 1       | 26862_9               | 3          | 15,991,151              | -0.77364 | 1.05E-05 | -0.81840 | 1.92E-05 | 0.070          | 0.00           |
|                     | 3       | 27988_137             | 3          | 34,577,820–34,577,828   | -1.37576 | 3.87E-06 | -1.34997 | 7.34E-06 | 0.080          | 0.00           |
|                     | 1       | 22953_48              | 2          | 36,249,915              | 1.33259  | 1.45E-05 | -1.39706 | 1.64E-05 | 0.075          | 0.00           |
|                     | 1       | 44035_122             | 6          | 43,347,813              | 0.89791  | 8.90E-08 | -1.08089 | 7.53E-09 | 0.114          | 0.00           |
|                     | 1       | 44180_7               | 6          | 4,450,0215              | -1.30507 | 1.17E-05 | -1.46595 | 1.70E-06 | 0.080          | 0.00           |
|                     | 1       | 34883_38              | 4          | 51,195,861              | 0.54333  | 1.54E-05 | -0.40724 | 5.61E-03 | 0.073          | 0.00           |
|                     | 1       | 24363_107             | 2          | 51,681,433              | 0.64229  | 1.42E-05 | -0.57303 | 2.47E-03 | 0.071          | 0.00           |
|                     | 1       | 29971_66              | 3          | 70,498,159              | -0.64269 | 1.69E-06 | -0.62923 | 4.34E-05 | 0.082          | 0.00           |
| DAS <sub>ExG</sub>  | 6       | 29372_33              | 3          | 5,889,084–5,889,263     | 0.02888  | 2.82E-06 | -0.03002 | 4.19E-06 | 0.079          | 0.13           |
|                     | 1       | 40827_128             | 5          | 7,581,055               | 0.02880  | 2.84E-06 | -0.03207 | 5.17E-07 | 0.090          | 0.11           |
|                     | 1       | 40828_107             | 5          | 7,581,143               | -0.02897 | 2.25E-06 | -0.03000 | 2.49E-06 | 0.079          | 0.13           |
|                     | 2       | 31191_149             | 3          | 9,539,283–9,539,285     | 0.02883  | 2.81E-06 | -0.02997 | 4.15E-06 | 0.079          | 0.13           |
|                     | 1       | 34151_101             | 4          | 29,043,034              | 0.01516  | 3.95E-06 | -0.01532 | 1.78E-05 | 0.086          | 0.18           |
|                     | 1       | 19887_120             | 1          | 40,388,453              | -0.03058 | 7.07E-07 | -0.02874 | 7.32E-06 | 0.087          | 0.19           |
|                     | 1       | 43709_72              | 6          | 40,598,206              | -0.02904 | 2.52E-06 | -0.03069 | 2.55E-06 | 0.081          | 0.13           |
|                     | 1       | 20358_139             | 1          | 44,357,800              | 0.03060  | 6.05E-07 | -0.02744 | 1.02E-05 | 0.112          | 0.27           |
|                     | 1       | 44261_56              | 6          | 45,192,349              | 0.02941  | 1.54E-06 | -0.02926 | 5.23E-06 | 0.081          | 0.13           |
|                     | 1       | 44281_23              | 6          | 45,342,253              | 0.02938  | 1.61E-06 | -0.02840 | 9.51E-06 | 0.080          | 0.13           |
|                     | 1       | 44291_137             | 6          | 45,352,029              | -0.02900 | 2.34E-06 | -0.02803 | 9.33E-06 | 0.086          | 0.17           |
|                     | 1       | 28665_72              | 3          | 46,395,669              | -0.00606 | 3.06E-04 | -0.00709 | 1.94E-04 | 0.063          | 0.32           |
|                     | 1       | 23822_144             | 2          | 47,797,116              | -0.01898 | 1.80E-05 | -0.01879 | 4.05E-05 | 0.066          | 0.24           |
|                     | 1       | 24160_135             | 2          | 50,245,050              | -0.02966 | 1.17E-06 | -0.02643 | 3.47E-05 | 0.086          | 0.12           |
|                     | 1       | 40520_4               | 5          | 63,027,301              | 0.01043  | 2.96E-05 | -0.00990 | 6.73E-04 | 0.064          | 0.10           |
|                     | 1       | 36231_148             | 4          | 78,521,968              | -0.02439 | 9.70E-09 | -0.02171 | 5.96E-06 | 0.118          | 0.17           |
|                     | 1       | 26106_55              | 3          | 102,662,772             | 0.00911  | 1.45E-05 | -0.00871 | 2.53E-04 | 0.072          | 0.17           |
| DAS <sub>bolr</sub> | 1       | 36822_84              | 4          | 88,252                  | 0.02189  | 3.02E-06 | -0.02015 | 1.41E-04 | 0.200          | 0.84           |
|                     | 1       | 26785_87              | 3          | 1,532,247               | 0.02231  | 9.44E-07 | -0.01444 | 1.22E-03 | 0.222          | 0.64           |
|                     | 1       | 30014_141             | 3          | 7,122,451               | -0.02098 | 1.12E-06 | -0.01985 | 1.50E-04 | 0.215          | 1.00           |
|                     | 1       | 36995_21              | 4          | 9,056,952               | -0.01453 | 1.75E-05 | -0.01323 | 3.01E-04 | 0.193          | 0.75           |
|                     | 1       | 18134_18              | 1          | 21,587,981              | 0.00658  | 2.04E-02 | 0.00271  | 4.18E-01 | 0.077          | 0.14           |
|                     | 1       | 22357_36              | 2          | 27,309,513              | 0.00565  | 4.58E-02 | 0.00506  | 1.31E-01 | 0.086          | 0.24           |
|                     | 1       | 22477_19              | 2          | 28,651,041              | -0.02488 | 2.41E-05 | -0.02386 | 1.54E-04 | 0.169          | 0.94           |
|                     | 2       | 39130_111             | 5          | 30,116,760–30,216,896   | 0.01591  | 1.21E-06 | -0.01294 | 1.00E-03 | 0.222          | 0.50           |
|                     | 1       | 39815_64              | 5          | 43,754,662              | -0.02237 | 1.08E-05 | -0.02521 | 1.89E-05 | 0.202          | 0.58           |
|                     | 1       | 39876_92              | 5          | 44,808,159              | -0.01694 | 3.79E-03 | -0.02851 | 4.92E-06 | 0.199          | 0.67           |
|                     | 1       | 40099_56              | 5          | 50,109,435              | -0.01866 | 7.86E-06 | -0.01712 | 8.64E-03 | 0.186          | 0.28           |
|                     | 2       | 32988_7               | 4          | 119,357,657–119,357,896 | -0.00626 | 9.14E-02 | -0.01871 | 2.78E-05 | 0.184          | 1.00           |

(cont'd)

most likely cause of the lost heritability is the large confounding error variance (Zou and Zuo, 2006), as shown by the large coefficients of variation observed as being associated with the environmental effect.

### Linkage Disequilibrium Decay Estimates

Of the 99 associated SNPs, 82 loci were in LD decay, including bolting and seed development-associated markers 36231\_148 (chr4; DAS<sub>kf</sub>) and 37180\_66 (chr4; DAS<sub>pol</sub>)

Table 7. Continued.

| Trait              | Markers           | Representative marker | Chromosome | Position and/or range | Q + k               |          |          |          |                |                |      |
|--------------------|-------------------|-----------------------|------------|-----------------------|---------------------|----------|----------|----------|----------------|----------------|------|
|                    |                   |                       |            |                       | Add.†               | P-value  | Dom.     | P-value  | R <sup>2</sup> | h <sup>2</sup> |      |
| DAS <sub>pol</sub> | 2                 | 37716_63              | 5          | 11,878,605–1,187,606  | -1.00261            | 2.44E-06 | -0.91319 | 3.71E-05 | 0.147          | 0.00           |      |
|                    | 1                 | 39553_80              | 5          | 40,016,005            | -1.00527            | 4.88E-06 | -0.92718 | 8.70E-05 | 0.130          | 0.13           |      |
|                    | 2                 | 43868_4               | 6          | 42,083,968–42,084,406 | -0.31990            | 3.51E-06 | -0.35847 | 1.05E-05 | 0.172          | 0.60           |      |
|                    | 1                 | 42970_31              | 6          | 34,290,002–34,290,103 | -1.00683            | 8.93E-07 | -0.86663 | 9.29E-05 | 0.155          | 0.00           |      |
|                    | 1                 | 20296_75              | 1          | 43,713,731            | -0.11412            | 3.89E-05 | 0.03664  | 3.47E-01 | 0.084          | 0.97           |      |
|                    | 1                 | 23894_148             | 2          | 48,350,859            | -0.05019            | 1.18E-01 | -0.04262 | 1.84E-01 | 0.099          | 1.00           |      |
|                    | 1                 | 40125_127             | 5          | 50,582,753            | -1.00450            | 3.80E-06 | -0.92234 | 7.89E-05 | 0.133          | 0.08           |      |
|                    | 1                 | 40446_103             | 5          | 59,288,499            | -1.00064            | 2.74E-06 | -0.96605 | 1.31E-05 | 0.132          | 0.00           |      |
|                    | 1                 | 29529_92              | 3          | 62,540,090            | -1.00398            | 2.34E-06 | -1.12204 | 3.77E-06 | 0.137          | 0.00           |      |
|                    | 1                 | 37179_94              | 4          | 93,704,388            | 0.54122             | 1.00E-05 | -0.54581 | 3.03E-05 | 0.137          | 0.26           |      |
|                    | DAS <sub>kt</sub> | 1                     | 28312_106  | 3                     | 4,026,915           | -0.10500 | 1.13E-02 | 0.19638  | 2.86E-05       | 0.086          | 0.92 |
|                    |                   | 2                     | 35474_60   | 4                     | 6,325,624–6,325,642 | 0.14813  | 2.97E-22 | 0.16997  | 4.00E-21       | 0.317          | 1.00 |
|                    |                   | 2                     | 37037_28   | 4                     | 9,101,643           | -0.14895 | 1.69E-12 | 0.17268  | 1.11E-15       | 0.307          | 0.99 |
|                    |                   | 1                     | 41288_57   | 6                     | 13,952,350          | 0.14635  | 2.85E-11 | -0.15646 | 1.26E-14       | 0.305          | 0.99 |
| 1                  |                   | 42302_61              | 6          | 27,292,261            | -0.02385            | 5.89E-02 | 0.05703  | 7.49E-26 | 0.125          | 1.00           |      |
| 1                  |                   | 28330_57              | 3          | 40,573,808            | -0.10384            | 1.25E-02 | 0.19991  | 2.10E-05 | 0.087          | 0.92           |      |
| 1                  |                   | 39773_89              | 5          | 42,984,604            | -0.09142            | 7.84E-03 | 0.16994  | 2.72E-05 | 0.092          | 0.94           |      |
| 1                  |                   | 39858_80              | 5          | 4,459,8648            | -0.09142            | 7.84E-03 | 0.16994  | 2.72E-05 | 0.092          | 0.94           |      |
| 1                  |                   | 34672_102             | 4          | 46,706,902            | 0.10765             | 2.41E-04 | -0.12735 | 4.19E-03 | 0.119          | 0.93           |      |
| 1                  |                   | 30538_30              | 3          | 83,500,450            | 0.06862             | 3.41E-02 | 0.17537  | 1.22E-04 | 0.095          | 0.93           |      |
| 2                  |                   | 25958_90              | 3          | 101,055,049           | 0.00291             | 1.11E-03 | -0.00460 | 1.20E-05 | 0.066          | 0.23           |      |
| 1                  |                   | 32020_29              | 4          | 107,271,059           | 0.06685             | 9.16E-02 | 0.21706  | 2.27E-05 | 0.086          | 0.90           |      |
| 1                  |                   | 32388_65              | 4          | 114,173,692           | 0.15934             | 2.31E-13 | 0.10482  | 1.11E-03 | 0.229          | 0.99           |      |

† Add., additive effect; CC<sub>max</sub>, maximum canopy cover; Dom, dominance effect; DAS<sub>bol</sub>, days after sowing until early bolting; DAS<sub>cc</sub>, days after sowing until maximum canopy cover; DAS<sub>cvol</sub>, days after sowing until maximum canopy volume; DAS<sub>exg</sub>, days after sowing until maximum excess greenness index; DAS<sub>pol</sub>, days after sowing until pollination; DAS<sub>kt</sub>, days after sowing until kernel filling; ExG<sub>max</sub>, maximum excess greenness index; h<sup>2</sup>, trait mean narrow-sense heritability; k, kinship matrix; Q, cluster of genetic distance probabilities for each accession in the population; R<sup>2</sup>, phenotypic variance explained by the effect model.

and the CC<sub>max</sub> growth-associated marker 42970\_132. The rest of the markers, including markers 38940\_6 (chr6; DAS<sub>cc</sub>) and 30538\_30 (chr3; DAS<sub>kt</sub>) were above the LD decay threshold, representing LD block sizes of 27 and ~132 kb, respectively, indicating that under our experimental conditions, the UAS was able to detect phenotypic signals associated with traits across various LD blocks and in LD decay (Fig. 5). The LD block sizes ranged from 20 bp to as much as 120 Mb, suggesting that the spinach accessions used in this study have undergone varying degrees of selection pressure over time, thus providing the possibility of large selective sweeps being available to move traits of interest into breeding materials.

### Causative SNP-Associated Putative Genes and Genomic Features

We used the 99 associated SNP marker positions to query the genomic contexts of genes. The 99 SNPs are spread across 69 genes. Of these markers, 53 are on the leading strand, though the majority (51%) are harbored in the non-coding intronic regions and 27% in the exons. The rest are located in the intergenic regions, 19% within 100-kb gene proximity and the other 5% outside at least 200 kb but within 850 kb (Table 8). Six of these SNPs are located in segments that are important for the transcriptional regulation of multiple genes implicated in the regulation and signaling of important growth parameters. For example, the DAS<sub>cc</sub>-associated SNP 36324\_127 is located in *Spo15232*,

which belongs to the basic leucine zipper transcription factor family proteins, members of which are involved in light and stress signaling influencing plant growth and development (Xiang et al., 2008). The marker 44291\_137 (chr6, Position 45,352,029) resides in an intronic segment of the gene *Spo11400*, which codes for cytochrome P450 family protein, whose orthologs have been implicated in stay-green, photosynthesis, defense, and plant development (Cooper, 2000; Xu et al., 2015; Awika et al., 2017).

## DISCUSSION

### Unmanned Aircraft System-Based Phenotyping can be Used to Develop Marker–Trait Associations across Different Modeling Methods

We tested the robustness of four compression models, with three models based on the P3D algorithm, which estimated the variance components on a trait basis (Q + k, PCA + k, and MDS + k), and one model (not implementing P3D) that estimated the variance components for each marker effect model (Q + k Per Marker). These models were used to evaluate the association between remotely sensed phenotypic signals and filtered SNP markers. Genome-wide association study models have been extensively evaluated in other studies (Korte et al., 2012; Yang et al., 2014a); the details are beyond the scope of the current study. Our preliminary observation is that in general, the trait-mean variance-based models Q + k, PCA + k, and MDS + k were

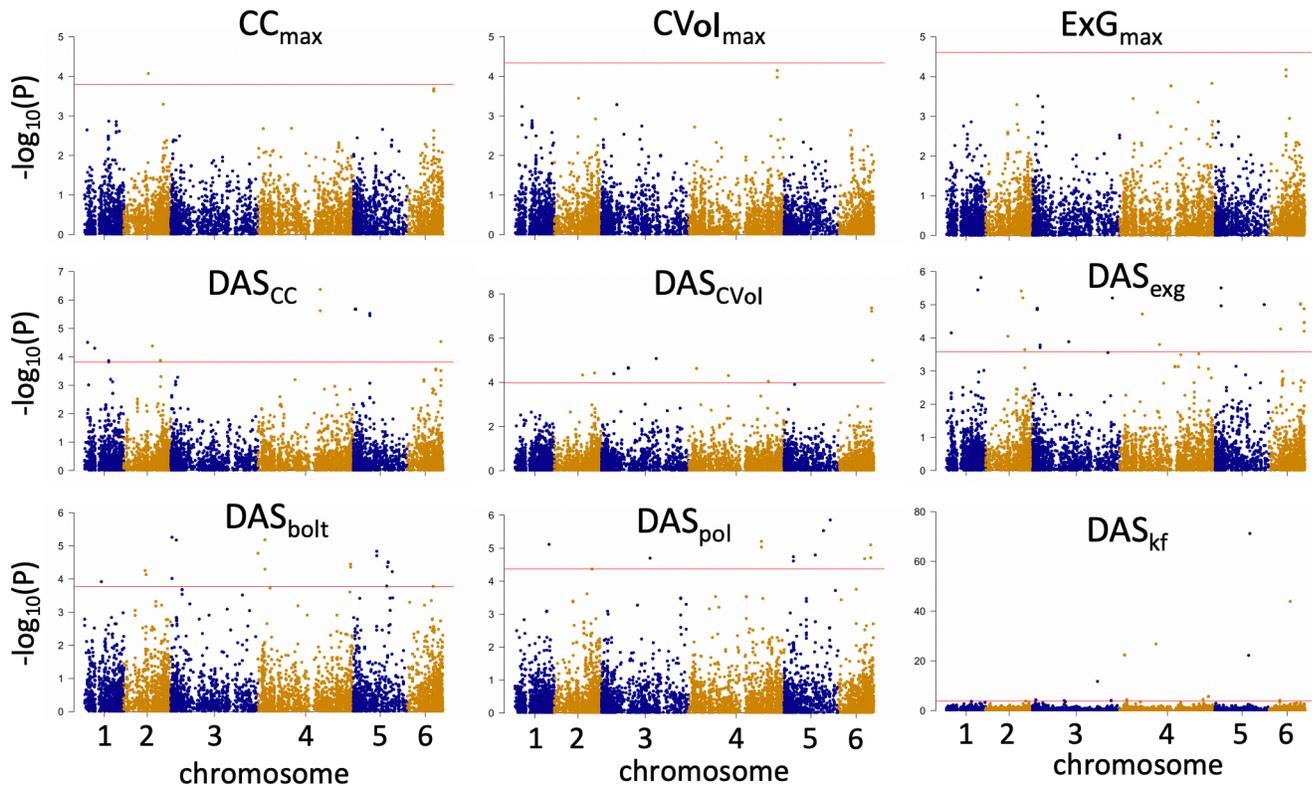


Fig. 4. Manhattan plots of association tests for 6167 markers and nine growth parameter traits. The vertical axes show negative  $\log_{10}P$  transformed values of significance tests for each marker (one single nucleotide polymorphism = one dot). Chromosome designations are on the horizontal axis (six chromosomes, 1–6, harbored the SNP markers tested with each trait). The marker–trait association significance thresholds (based on Benjamini and Hochberg’s false discovery rate) are shown by the solid horizontal red lines.

less conservative and returned higher mean  $h^2$  (43.2, 53.4, and 37.2%, respectively) across the UAS-derived traits than the  $Q + k$  Per Marker model ( $h^2 \sim 27.8\%$ , Table 3). There was no significant difference in the  $h^2$  of the manually quantitated traits. By contrast, at the marker level (Table 4-7), the reverse was true for per-marker phenotypic variance explained (PVE), with the  $Q + k$ , PCA +  $k$ , and MDS +  $k$  models being generally more conservative (lower but not significantly different mean  $R^2$  values of  $\sim 9.5$ ,  $8.3$ , and  $8.4$  respectively) than the corresponding  $Q + k$  Per Marker model (mean  $R^2 = \sim 11.7$ ). We calculated per-marker  $h^2$  via the  $Q + k$  Per Marker model only. No specific pattern was detected for the effect sizes of additive and dominance effects across all models (Table 4-7). Although any of the methods would appear to suffice, the compression  $Q + k$  Per Marker model detected more significant association signals than the other models in the current study.

### Unmanned Aircraft System and Visual Data GWAS Signals

We detected associated markers for four of the six UAS-measured traits. Most of these associated markers were related to the DAS measurements ( $DAS_{CVol}$ ,  $DAS_{CC}$ , and  $DAS_{ExG}$ ), whereas we only uncovered significant markers for one maximum growth variable ( $CC_{max}$ ). Under the assumption that that SNP marker saturation on chromosomes was not a limiting factor, these results suggest that the DAS variables are more robust than the maximum

value ones for detecting differences among accessions, perhaps because they are less affected by small changes in plant density (e.g., missing single plants within a plot). By contrast, the effect of missing plants may have had stronger effects on the maximum CC, CVol, and ExG measurements because of the small plot size used in this study. The use of larger plots with uniform planting densities could potentially improve the chances of detecting additional marker–trait associations that we failed to detect in the present study.

On the other hand, analysis of the associated signals for the manually collected bolting-related measurements resulted in the identification of several significant markers at 13 loci for  $DAS_{kf}$ , 10 for  $DAS_{pol}$ , and 12 for  $DAS_{bolt}$ . Chitwood et al. (2016) identified three SNP markers related to bolting by using a similar population derived from the USDA-NPGS. However, those markers were not localized at the same loci as the markers identified in the current study, perhaps because environmental conditions such as photoperiod might have influenced the bolting rates at different locations (Chun et al., 2000). Interestingly, no correlation was observed between time to bolting stages and maximum plant growth (Table 1). Therefore, alternative UAS measurements need to be evaluated to screen accessions for bolting resistance.

Associated individual SNPs explained low levels of PVE (6.18–22.9%) across all mixed linear models (Table 4-7). Low PVE has also been reported for GWAS of

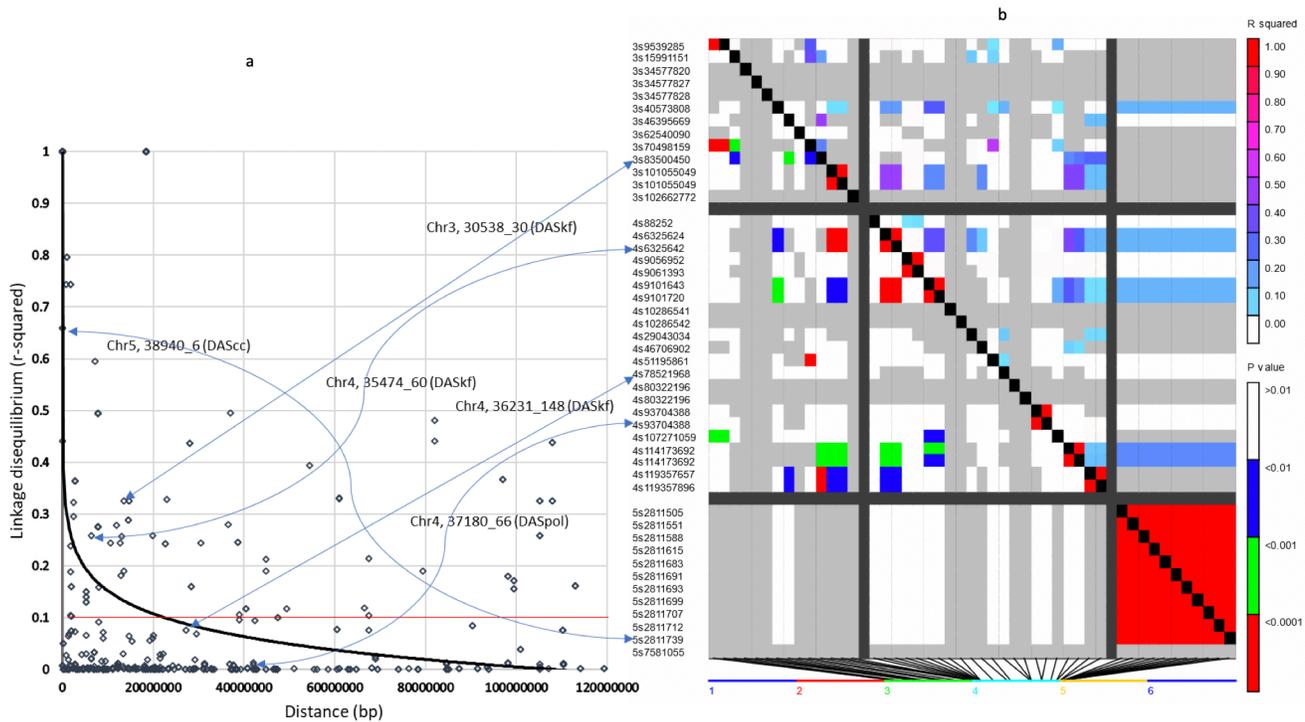


Fig. 5. Linkage disequilibrium (LD) and LD decay analysis of the 99 associated polymorphic variants. (a) The LD decay obtained from LD values of trait-associated sites, providing a measure of the average LD block sizes. The inner fitted trend line (thick black curve) is a nonlinear logarithmic regression curve of  $r^2$  on genetic distance. Linkage disequilibrium decay is considered to lie below the  $r^2 = 0.1$  threshold (cutoff, horizontal red line). Marker positions (diamond shapes) above the cutoff line are in LD. The blue arrows connect the marker positions on the LD decay plot (only a few marker identities are shown) to their corresponding chromosomal positions shown in the pairwise grid. For some of the markers, the traits with which they are associated (in parenthesis) are shown above the LD decay threshold. (b) Triangle plot for pairwise values ( $r^2$ ) plotted against physical genomic distance between marker sites in a genome fragment harboring 57 of the 117 associated sites. The pairwise LD values of polymorphic sites are plotted on both the  $x$ - and  $y$ -axes. The  $r^2$  values are above the diagonal; the corresponding  $P$ -values are below the diagonal. Each cell represents a comparison of two pairs of marker sites and the cells are color-coded with respect to the presence of significant LD. A colored barcode for the significance threshold levels in both diagonals is shown. Sites were selected on the basis of the strength of their association with the traits according to Benjamini and Hochberg's false discovery rate (see Methods).

spinach as well as other plant and mammalian systems (Slatkin, 2008; Xu et al., 2017), which may be related to the high within-accession heterogeneity found in the USDA-NPGS collection (e.g., the open pollinated nature of spinach). However, the low threshold of associated tag SNPs might be validated at much higher significance levels in follow-up studies. For example, Yeager et al. (2007) reported associated SNPs from initial GWAS at a  $P$ -value of only 0.042, although the significance dramatically increased in a follow-up study, with a  $P$ -value (above the Bonferroni threshold) of  $7.31 \times 10^{-13}$  (Thomas et al., 2008). Therefore, marker validation must be performed before using these associated SNPs in marker-assisted selection programs.

Finally, on the basis of the marker information, we calculated the additive and dominance effects and  $h^2$  of each trait to determine the potential breeding value of selection for the plant growth parameters based on RGB sensor sensitivity. The line-mean heritability across the different models for UAS and visually collected traits was low ( $H^2 \leq 9.46\%$ ), indicating missing heritability, which is typical of many quantitative traits phenotyped in the field (Brachi et al., 2011). However, when we considered

heritability on marker mean basis,  $h^2$  ranged from 16.30% ( $CC_{\max}$ ,  $Q + k$  model) to as high as ~100% ( $CC_{\max}$ ,  $Q + k$  model).

### The Detection of Large LD Blocks and Low Heritability Offers Insights into the Functionality of High-Throughput UAS Phenotyping in Spinach

Among the significantly associated markers, the conserved LD blocks consisted of small physical distances of ~20 bp to large sweeps of ~112 Mb (Fig. 5a). Some of these markers, their home chromosomes, and the trait with which they are associated (in parenthesis) are shown above the LD decay threshold in Fig. 5a, along with their relative physical positions in Fig. 5b. Some of the markers are associated with UAS-imputed phenotypes (only one is shown in Fig. 5a because of the genomic segment selected in Fig. 5b). This finding suggests that the few associated markers detected in the current study provide a good representation of LD blocks harboring possible causative (including small effect) variants, despite the low per-marker heritability and PVE on most of these markers.

Table 8. Genomic context and resident genes harboring candidate single nucleotide polymorphisms (SNPs).

| Trait               | Marker    | SNP alleles | Chr | Position    | Locus ID          | Segment type† | Strand | Annotation  |
|---------------------|-----------|-------------|-----|-------------|-------------------|---------------|--------|---|
| CC <sub>max</sub> ‡ | 42970_120 | A G         | 6   | 34,290,091  | <i>Spo18796.1</i> | Intron        | +      | Transcription factor jumonji (JmjC) domain protein  |
| DAS <sub>cc</sub>   | 38939_110 | T C         | 5   | 2,811,505   | <i>Spo02370.1</i> | Intron        | +      | NADH dehydrogenase (ubiquinone) flavoprotein 2  |
|                     | 19651_36  | G T         | 1   | 2,811,739   | <i>Spo03627.1</i> | Intron        | +      | DNA repair protein RAD50, putative  |
|                     | 17771_95  | A G         | 1   | 12,969,303  | <i>Spo03824.1</i> | Intron        | -      | IKI3 family elongator complex protein   |
|                     | 38455_121 | C A         | 5   | 21,285,477  | <i>Spo15202.1</i> | Exon          | +      | UPF0503 protein Ar3g09070, chloroplastic (precursor)  |
|                     | 22984_88  | T C         | 2   | 36,778,324  | <i>Spo12315.1</i> | Intron        | -      | Kinesin-like protein  |
|                     | 44084_132 | C A         | 6   | 43,509,344  | <i>Spo23411.1</i> | Exon          | +      | Disease resistance protein  |
|                     | 36324_127 | A G         | 4   | 80,322,196  | <i>Spo15232.1</i> | Intron        | +      | Basic leucine zipper transcription factor family protein  |
|                     | 31710_51  | G A         | 4   | 10,286,541  | <i>Spo14212.1</i> | Intron        | -      | Transmembrane protein, putative   |
|                     | 26862_9   | C T         | 3   | 15,991,151  | <i>Spo10226.1</i> | Exon          | -      | Flavin-containing monooxygenase (1...)  |
|                     | 27988_137 | C G         | 3   | 34,577,827  | <i>Spo13590.1</i> | -20 kb down   | +      | Calcium-transporting ATPase (3.6.3.8)   |
| DAS <sub>cvl</sub>  | 22953_48  | T C         | 2   | 36,249,915  | <i>Spo00454.1</i> | Intron        | -      | Peptidyl-prolyl <i>cis-trans</i> isomerase (5.2.1.8)  |
|                     | 44035_122 | G A         | 6   | 43,347,813  | <i>Spo23418.1</i> | Intron        | +      | Peroxidase (1.11.1.7)   |
|                     | 44180_7   | A C         | 6   | 44,500,215  | <i>Spo23246.1</i> | ~3.4 kb up    | -      | Small ubiquitin-related modifier  |
|                     | 34883_38  | T C         | 4   | 51,195,861  | <i>Spo11659.1</i> | Intron        | -      | tRNA (guanine(37)-M1)-methyltransferase (2.1.1.228)<br>(M1G-methyltransferase) (tRNA [GM37] methyltransferase) (tRNA methyltransferase 5 homolog) |
|                     | 24363_107 | G A         | 2   | 51,681,433  | <i>Spo23576.1</i> | Exon          | +      | Gibberellic Acid Insensitive (GAI), Repressor of GAI, and Scarecrow family transcription factor   |
|                     | 29971_66  | A G         | 3   | 70,498,159  | <i>Spo07147.1</i> | Exon          | -      | Zinc finger family protein  |
|                     | 29372_33  | G A         | 3   | 5,889,084   | <i>Spo02574.1</i> | Intron        | -      | PP2A regulatory subunit TAP46-like protein  |
|                     | 40827_128 | G C         | 5   | 7,581,055   | <i>Spo25275.1</i> | Exon          | +      | F-box protein interaction domain protein  |
|                     | 31191_149 | C A         | 3   | 9,539,283   | <i>Spo26660.1</i> | Intron        | +      | -tocopherol methyltransferase   |
|                     | 34151_101 | T A         | 4   | 29,043,034  | <i>Spo00815.1</i> | ~5 kb down    | +      | Chromatin assembly factor 1 subunit FAS1, putative  |
| DAS <sub>exg</sub>  | 19887_120 | A G         | 1   | 40,388,453  | <i>Spo10629.1</i> | Exon          | +      | Tetraspanin family protein  |
|                     | 43709_72  | T C         | 6   | 40,598,206  | <i>Spo26018.1</i> | Exon          | +      | Basic helix loop helix family transcription factor  |
|                     | 20358_139 | G T         | 1   | 44,357,800  | <i>Spo09940.1</i> | Intron        | +      | Nuclear transport factor 2 family protein with RNA binding domain, putative isoform 1   |
|                     | 44261_56  | A G         | 6   | 45,192,349  | <i>Spo11356.1</i> | ~400 bp up    | +      | Serine/threonine-protein kinase STE20, putative   |
|                     | 44281_23  | G T         | 6   | 45,342,253  | <i>Spo11398.1</i> | Intron        | -      | Cytochrome P450, putative (1.14.13.88)  |
|                     | 44291_137 | T A         | 6   | 45,352,029  | <i>Spo11400.1</i> | Intron        | -      | Cytochrome P450 family protein  |
|                     | 28665_72  | G A         | 3   | 46,395,669  | <i>Spo03641.1</i> | Intron        | +      | DNA repair and recombination protein  |
|                     | 23822_144 | G T         | 2   | 47,797,116  | <i>Spo18839.1</i> | Intron        | +      | Transducin/WD40 repeat-like superfamily protein   |
|                     | 24160_135 | A T         | 2   | 50,245,050  | <i>Spo01675.1</i> | Intron        | -      | Arid/Bright DNA-binding domain, ELM2 domain protein, putative   |
|                     | 40520_4   | G T         | 5   | 63,027,301  | <i>Spo26541.1</i> | ~810 kb down  | -      | Transmembrane protein, putative   |
| DAS <sub>bol</sub>  | 36231_148 | C T         | 4   | 78,521,968  | <i>Spo22543.1</i> | ~16 kb up     | -      | MATE efflux family protein  |
|                     | 26106_55  | G C         | 3   | 102,662,772 | <i>Spo15533.1</i> | ~80 kb up     | +      | Jasmonate-induced protein homolog   |
|                     | 36822_84  | G A         | 4   | 88,252      | <i>Spo16530.1</i> | Exon          | +      | Zinc-binding protein  |
|                     | 26785_87  | T G         | 3   | 1,532,247   | <i>Spo12930.1</i> | Exon          | -      | Protein Longifolia 2 (Protein TONT Recruiting Motif 1)  |
|                     | 30014_141 | A T         | 3   | 7,122,451   | <i>Spo22366.1</i> | Intron        | -      | DNA binding/transcription factor  |
|                     | 36995_21  | T A         | 4   | 9,056,952   | <i>Spo11986.1</i> | 201 bp down   | +      | Probable E3 ubiquitin-protein ligase AR17 (6.3.2.-) (Ariadne-like protein AR17) (Protein Ariadne homolog 7)                                       |
|                     | 37001_48  | T C         | 4   | 9,061,393   | <i>Spo12066.1</i> | Intron        | -      | γ-tubulin complex component, putative   |
|                     | 18134_18  | C T         | 1   | 21,587,981  | <i>Spo02028.1</i> | Exon          | +      | EEIG1/EHBP1 protein amino-terminal domain protein   |
|                     | 22357_36  | T A         | 2   | 27,309,513  | <i>Spo02637.1</i> | Intron        | -      | Ankyrin repeat family protein   |
|                     | 22477_19  | A G         | 2   | 28,651,041  | <i>Spo15819.1</i> | 353 bp down   | -      | RNA-binding protein-related   |
| DAS <sub>bol</sub>  | 39130_111 | G T         | 5   | 30,216,896  | <i>Spo14477.1</i> | Intron        | -      | Auxin-responsive family protein; Expressed protein; SAUR-like auxin-responsive protein  |
|                     | 39815_64  | G A         | 5   | 43,754,662  | <i>Spo06032.1</i> | Intron        | -      | B3 domain-containing protein  |
|                     | 39876_92  | A T         | 5   | 44,808,159  | <i>Spo26692.1</i> | Intron        | -      | BTB/POZ domain protein  |
|                     | 40099_56  | C T         | 5   | 50,109,435  | <i>Spo16236.1</i> | ~6.2 kb up    | -      | Plant protein of unknown function (DUF641)  |
|                     | 32988_7   | C G         | 4   | 119,357,657 | <i>Spo19883.1</i> | Intron        | +      | Glycosyl hydrolase family 10 family protein   |

(cont'd)

### The UAS-Mounted Sensor Captures Growth Phenotypes Associated with Candidate Tag SNPs

Several traits examined by our aerial phenotyping platform were associated with tag SNPs in genes known to participate in important plant growth functions. Table 8 provides a rich list of putatively involved genes that could

be targeted in future functional analysis studies. For example, DAS<sub>pol</sub> and CC<sub>max</sub> were associated with the SNP loci 42970\_31 (C-T, exon) and 42970\_120 (A-G, intron). The markers are harbored only 89 bp apart on chr6. The makers 42970\_73 (A/G, exon) and 42970\_132 (T/A, intron) (Supplemental Table 3) are also significantly associated

Table 8. Continued.

| Trait              | Marker    | SNP alleles | Chr | Position    | Locus ID          | Segment type† | Strand | Annotation  |
|--------------------|-----------|-------------|-----|-------------|-------------------|---------------|--------|---|
| DAS <sub>Pol</sub> | 37716_63  | A T         | 5   | 11,878,605  | <i>Spo08215.1</i> | ~6 kb up      | –      | Unknown protein   |
|                    | 39553_80  | C T         | 5   | 40,016,005  | <i>Spo15859.1</i> | ~32 kb up     | –      | Gamete expressed 2  |
|                    | 43868_4   | C G         | 6   | 42,083,968  | <i>Spo09739.1</i> | Intron        | –      | Bax inhibitor   |
|                    | 42970_31  | C T         | 6   | 34,290,002  | <i>Spo18796.1</i> | Exon          | +      | Transcription factor JmjC domain protein  |
|                    | 20296_75  | G T         | 1   | 43,713,731  | <i>Spo07638.1</i> | ~3.3 kb up    | –      | UDP glycosyltransferase   |
|                    | 23894_148 | C A         | 2   | 48,350,859  | <i>Spo00444.1</i> | Exon          | +      | Histone h1/h5, putative   |
|                    | 40125_127 | C T         | 5   | 50,582,753  | <i>Spo04571.1</i> | Intron        | +      | DNA polymerase III subunit /  |
|                    | 40446_103 | A G         | 5   | 59,288,499  | <i>Spo04785.1</i> | ~532 kb up    | +      | Unknown protein   |
|                    | 29529_92  | A G         | 3   | 62,540,090  | <i>Spo05146.1</i> | Intron        | +      | Hydroxyproline-rich glycoprotein family protein                                   |
|                    | 37179_94  | A G         | 4   | 93,704,388  | <i>Spo14170.1</i> | ~14 kb down   | –      | Ethylene-responsive transcription factor, putative                                |
| DAS <sub>Kt</sub>  | 28312_106 | A G         | 3   | 4,026,915   | <i>Spo16027.1</i> | Exon          | –      | Multidrug resistance protein ABC transporter family                               |
|                    | 35474_60  | T C         | 4   | 6,325,624   | <i>Spo15727.1</i> | Exon          | +      | Oxygen-evolving enhancer protein; Photosystem II oxygen-evolving enhancer protein |
|                    | 37037_28  | G A         | 4   | 9,101,643   | <i>Spo12068.1</i> | ~423 bp up    | –      | Receptor-like protein kinase  |
|                    | 41288_57  | A C         | 6   | 13,952,350  | <i>Spo08623.1</i> | Intron        | –      | Rubisco methyltransferase family protein  |
|                    | 42302_61  | G C         | 6   | 27,292,261  | <i>Spo16801.1</i> | ~52 kb down   | +      | Nijmegen breakage syndrome 1  |
|                    | 28330_57  | A G         | 3   | 40,573,808  | <i>Spo11338.1</i> | ~6 kb down    | –      | BnaC07g26200D protein   |
|                    | 39773_89  | A G         | 5   | 42,984,604  | <i>Spo22580.1</i> | Exon          | –      | C2H2 zinc finger protein  |
|                    | 39858_80  | T A         | 5   | 44,598,648  | <i>Spo26678.1</i> | ~23 kb up     | +      | Xyloglucan xylosyltransferase   |
|                    | 34672_102 | C T         | 4   | 46,706,902  | <i>Spo00605.1</i> | Intron        | –      | Uroporphyrinogen-III C-methyltransferase  |
|                    | 30538_30  | C T         | 3   | 83,500,450  | <i>Spo08712.1</i> | ~63 kb down   | +      | AT5g16110/T21H19_30   |
|                    | 25958_90  | G C         | 3   | 101,055,049 | <i>Spo11151.1</i> | Exon          | +      | UDP-glycosyltransferase   |
|                    | 32020_29  | A G         | 4   | 107,271,059 | <i>Spo15437.1</i> | Exon          | +      | Nuclear pore complex protein Nup188b  |
|                    | 32388_65  | G A         | 4   | 114,173,692 | <i>Spo03506.1</i> | Exon          | –      | Hexokinase  |

† The part of the gene or the physical proximity of the gene in at which the candidate SNP was detected.

‡ CC<sub>max</sub>, maximum canopy cover; chr, chromosome; CVol<sub>max</sub>, maximum canopy volume; DAS<sub>bol</sub>, days after sowing until early bolting; DAS<sub>cc</sub>, days after sowing until maximum canopy cover; DAS<sub>cvol</sub>, days after sowing until maximum canopy volume; DAS<sub>exg</sub>, days after sowing until maximum excess greenness index; DAS<sub>pol</sub>, days after sowing until pollination; DAS<sub>kt</sub>, days after sowing until kernel filling; ExG<sub>max</sub>, maximum excess greenness index.

with CC<sub>max</sub> and are also a few bases from each of the other two previously mentioned, suggesting that this quantitative trait locus plays a role in growth and reproduction (Fig. 5, Table 8). Indeed, these tag SNP are located within the gene *Spo18796*, encoding a transcription factor jumonji domain protein whose family members, such as Early Flowering 6, Relative Of Early Flowering 6, and Jumonji Domain Containing 5 play divergent roles in the regulation of flowering and plant growth rhythm and tissue expansion (Noh et al., 2004; Jones et al., 2010).

The gene *Spo11400*, encoding a cytochrome P450 family protein (whose orthologs have been implicated in stay-green, photosynthesis, defense, and plant development) (Xu et al., 2015; Awika et al., 2017; Cooper, 2000) harbors the SNP marker 44291\_137 (T-A; chr6, Position 45,352,029). This SNP may have a role in determining the timing of maximum greenness, as observed in the present study, indicating the ability to phenotype accurately for greenness level in spinach with a UAS-mounted visible light sensor.

Another example is DAS<sub>cvol</sub> was moderately associated with SNP 24363\_107 (G-A; chr2), which is located in an exon of *Spo23576*, a Gibberellic Acid Insensitive, Repressor of Gibberellic Acid Insensitive, and Scarecrow transcription factor; members of this family play important roles in gibberellin, jasmonate, and light signaling and influence plant growth and development (Zhang et al., 2012). Further functional genomic analysis will be needed to validate

the roles of these genes in plant growth and development. Functional analysis will need to be performed to determine the possible effects of these SNPs after transcription in the host amino acid residues' polypeptides.

### Aerial Imaging and Genomic Architecture Information can be Combined to Optimize Marker Development for Spinach Breeding

Determining whether the UAS captures useful phenotypic data for marker development to enhance spinach breeding is essential for exploring future applications and improvements to UAS-based phenotyping. Here, we used 6167 markers on chr1 to chr6 of the draft spinach reference genome. We then used the best linear unbiased predictions of plant growth spectral signals obtained with the UAS autonomous platform in conjunction with these SNPs to identify polymorphic variants with association signals that might warrant further investigation. We also mapped these markers to various genomic features, such as proteins that affect leaf growth, chlorophyll content, and functionality and genes that affect plant height through cell elongation. This study is among the first to use high-throughput phenotyping via a UAS and high-throughput genotyping-by-sequencing of spinach. This study demonstrates the potential of using remote sensing platforms, image-analysis, and genotyping pipelines to screen leafy green vegetables for traits of interest,

such as growth rate, soil nutrient use efficiency, disease and pest resistance, and abiotic stress tolerance.

### Supplemental Information

Supplementary Table S1. Origin and grouping of spinach accessions used in this study.

Supplementary Table S2. Raw UAS and manually collected phenotypic data.

Supplementary Table S3. Genotypic data by spinach accession.

Supplementary Table S4. Compression levels used to optimize and test markers for the GWA models.

### Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

### Author Contributions

HOA performed population and genomic analyses; TGM conducted field experiments and nucleic acid isolation; RB and KKM performed SNP variant call analysis; JE, JY, and JJ performed UAS data acquisition, extraction, and analysis; and CAA designed and supervised the experiments. All the authors contributed to writing and reviewing the manuscript.

### ACKNOWLEDGMENTS

The authors thank Samantha Serna and Julianna Kurpis (Texas A&M AgriLife Research, Weslaco, TX) for technical support with plot field maintenance. This study was supported in part by funds from USDA–National Institute of Food and Agriculture Specialty Crops Research Initiative 2017-51181-26830 and Texas A&M AgriLife Research Vegetable Seed Grant FY17 and FY18 124353-96181 to CAA, and Texas A&M AgriLife Research Insect-vectored Disease Seed Grants (114190-96210) to KM.

### REFERENCES

- Agarwal, A., S. Dutta Gupta, M. Barman, and A. Mitra. 2018. Photosynthetic apparatus plays a central role in photosensitive physiological acclimations affecting spinach (*Spinacia oleracea* L.) growth in response to blue and red photon flux ratios. *Environ. Exp. Bot.* 156:170–182. doi:10.1016/j.envexpbot.2018.09.009
- Alexander, D.H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664. doi:10.1101/gr.094052.109
- Allison, P.D. 1999. *Multiple regression: A primer*. Pine Forge Press, Thousand Oaks, CA.
- An, N., C.M. Palmer, R.L. Baker, R.J.C. Markelz, J. Ta, M.F. Covington, et al. 2016. Plant high-throughput phenotyping using photogrammetry and imaging techniques to measure leaf length and rosette area. *Comput. Electron. Agric.* 127:376–394. doi:10.1016/j.compag.2016.04.002
- Awika, H.O., D.B. Hays, J.E. Mullet, W.L. Rooney, and B.D. Weers. 2017. QTL mapping and loci dissection for leaf epicuticular wax load and canopy temperature depression and their association with QTL for staygreen in *Sorghum bicolor* under stress. *Euphytica* 213:207. doi:10.1007/s10681-017-1990-5
- Bedre, R., K. Rajasekaran, V. Mangu, L. Sanchez Timm, D. Bhatnagar, and N. Baisakh. 2015. Genome-wide transcriptome analysis of cotton (*Gossypium hirsutum* L.) identifies candidate gene signatures in response to aflatoxin producing fungus *Aspergillus flavus*. *PLoS One* 10:E0138025. doi:10.1371/journal.pone.0138025
- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57:289–300.
- Berger, B., B. de Regt, and M. Tester. 2012. High-throughput phenotyping of plant shoots. In: J. Normanly, editor, *High-throughput phenotyping in plants: Methods and protocols*. Humana Press, Totowa, NJ. p. 9–20.
- Bernardo, R. 1996. Best linear unbiased prediction of maize single-cross performance. *Crop Sci.* 36:50–56. doi:10.2135/cropsci1996.0011183X003600010009x
- Blouin, M.S. 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends Ecol. Evol.* 18:503–511. doi:10.1016/S0169-5347(03)00225-8
- Brachi, B., G.P. Morris, and J.O. Borevitz. 2011. Genome-wide association studies in plants: The missing heritability is in the field. *Genome Biol.* 12:232. doi:10.1186/gb-2011-12-10-232
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. doi:10.1093/bioinformatics/btm308
- 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. doi:10.1111/j.1744-7909.2012.01116.x
- Catchen, J., P.A. Hohenlohe, S. Bassham, A. Amores, and W.A. Cresko. 2013. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* 22:3124–3140. doi:10.1111/mec.12354
- Chen, D., K. Neumann, S. Friedel, B. Kilian, M. Chen, T. Altmann, et al. 2014. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *Plant Cell* 26:4636–4655. doi:10.1105/tpc.114.129601
- Chitwood, J., A. Shi, B. Mou, M. Evans, J. Clark, D. Motes, et al. 2016. Population structure and association analysis of bolting, plant height, and leaf erectness in spinach. *HortScience* 51:481–486. doi:10.21273/HORTSCI.51.5.481
- Christenhusz, M.J.M., and J.W. Byng. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa* 261(3):201–217. doi:10.11646/phytotaxa.261.3.1
- Chun, C., T. Kozai, C. Kubota, and K. Okabe. 2000. Manipulation of bolting and flowering in spinach (*Spinacia oleracea* L.) transplant production system using artificial light. *Acta Hort.* 515:201–206. doi:10.17660/ActaHortic.2000.515.25
- 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. doi:10.1007/s00122-013-2066-0
- Cooper, G.M. 2000. *The cell: A molecular approach*. 2nd ed. Sinauer Associates, Sunderland, MA.
- Corti, M., P. Marino Gallina, D. Cavalli, and G. Cabassi. 2017. Hyperspectral imaging of spinach canopy under combined water and nitrogen stress to estimate biomass, water, and nitrogen content. *Biosyst. Eng.* 158:38–50. doi:10.1016/j.biosystemseng.2017.03.006
- Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, et al. 2011. The Variant Call Format and VCFtools. *Bioinformatics* 27:2156–2158. doi:10.1093/bioinformatics/btr330
- Diezma, B., L. Lleó, J.M. Roger, A. Herrero-Langreo, L. Lunadei, and M. Ruiz-Altisent. 2013. Examination of the quality of spinach leaves using hyperspectral imaging. *Postharvest Biol. Technol.* 85:8–17. doi:10.1016/j.postharvbio.2013.04.017
- Earl, D.A., and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4:359–361. doi:10.1007/s12686-011-9548-7
- Eguchi, T., and H. Ichikawa. 1940. Studies on flower differentiation and bolting of spinach. *J. Jpn. Soc. Hortic. Sci.* 11:13–56. doi:10.2503/jjshs.11.13
- Endelman, J.B., and J.-L. Jannink. 2012. Shrinkage estimation of the realized relationship matrix. *G3 (Bethesda)* 2:1405–1413.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x
- Everard, C.D., M.S. Kim, and H. Lee. 2014. A comparison of hyperspectral reflectance and fluorescence imaging techniques for detection

- of contaminants on spinach leaves. *J. Food Eng.* 143:139–145. doi:10.1016/j.jfoodeng.2014.06.042
- Everitt, J.H., C. Yang, and M.R. Davis. 2007. Mapping wild taro with color-infrared aerial photography and image processing. *J. Aquat. Plant Manage.* 45:106–110.
- Fahlgren, N., M.A. Gehan, and I. Baxter. 2015. Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. *Curr. Opin. Plant Biol.* 24:93–99. doi:10.1016/j.pbi.2015.02.006
- Falush, D., M. Stephens, and J.K. Pritchard. 2007. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Mol. Ecol. Notes* 7:574–578. doi:10.1111/j.1471-8286.2007.01758.x
- Felsenstein, J. 1989. PHYLIP—Phylogeny inference package (version 3.2). *Cladistics* 5: 164–166.
- Feng, C., B. Bluhm, A. Shi, and J.C. Correll. 2018. Development of molecular markers linked to three spinach downy mildew resistance loci. *Euphytica* 214:174. doi:10.1007/s10681-018-2258-4
- Golzarian, M.R., R.A. Frick, K. Rajendran, B. Berger, S. Roy, M. Tester, et al. 2011. Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods* 7:2. doi:10.1186/1746-4811-7-2
- Hairmansis, A., B. Berger, M. Tester, and S.J. Roy. 2014. Image-based phenotyping for non-destructive screening of different salinity tolerance traits in rice. *Rice (N. Y.)* 7:16. doi:10.1186/s12284-014-0016-3
- Hoisington, D., M. Khairallah, and D. González-de-León. 1994. Laboratory protocols. CIMMYT Applied Molecular Genetics Laboratory, CIMMYT, Texcoco, México.
- Houle, D., D.R. Govindaraju, and S. Omholt. 2010. Phenomics: The next challenge. *Nat. Rev. Genet.* 11:855–866. doi:10.1038/nrg2897
- Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806. doi:10.1093/bioinformatics/btm233
- Jones, M.A., M.F. Covington, L. DiTacchio, C. Vollmers, S. Panda, and S.L. Harmer. 2010. Jumonji domain protein JMJD5 functions in both the plant and human circadian systems. *Proc. Natl. Acad. Sci. USA* 107:21623–21628. doi:10.1073/pnas.1014204108
- Kim, H.-H., C. Chun, T. Kozai, and J. Fuse. 2000. The potential use of photoperiod during transplant production under artificial lighting conditions on floral development and bolting, using spinach as a model. *HortScience* 35:43–45. doi:10.21273/HORTSCI.35.1.43
- Kock, N., and G.S. Lynn. 2012. Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *J. Assoc. Inf. Syst.* 13:546–580.
- Korte, A., B.J. Vilhjálmsson, V. Segura, A. Platt, Q. Long, and M. Nordborg. 2012. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* 44:1066–1071. doi:10.1038/ng.2376
- Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–359. doi:10.1038/nmeth.1923
- Lipovetsky, S., and M. Conklin. 2001. Analysis of regression in game theory approach. *Appl. Stochastic Models Data Anal.* 17:319–330. doi:10.1002/asmb.446
- Lyon, R., J. Correll, C. Feng, B. Bluhm, S. Shrestha, A. Shi, et al. 2016. Population structure of *Peronospora effusa* in the Southwestern United States. *PLoS One* 11:E0148385. doi:10.1371/journal.pone.0148385
- Mesas-Carrascosa, F.-J., J. Torres-Sánchez, I. Clavero-Rumbao, A. García-Ferrer, J.-M. Peña, I. Borra-Serrano, et al. 2015. Assessing optimal flight parameters for generating accurate multispectral orthomosaics by UAV to support site-specific crop management. *Remote Sens.* 7:12793–12814. doi:10.3390/rs71012793
- Meuwissen, T.H.E., J. Odegard, I. Andersen-Ranberg, and E. Grindflek. 2014. On the distance of genetic relationships and the accuracy of genomic prediction in pig breeding. *Genet Sel Evol.* 46:49. doi:10.1186/1297-9686-46-49
- Min, K., L. Showman, A. Perera, and R. Arora. 2018. Salicylic acid-induced freezing tolerance in spinach (*Spinacia oleracea* L.) leaves explored through metabolite profiling. *Environ. Exp. Bot.* 156:214–227. doi:10.1016/j.envexpbot.2018.09.011
- Moharana, S., and S. Dutta. 2016. Spatial variability of chlorophyll and nitrogen content of rice from hyperspectral imagery. *ISPRS J. Photogramm. Remote Sens.* 122:17–29. doi:10.1016/j.isprsjprs.2016.09.002
- Neilson, E.H., A.M. Edwards, C.K. Blomstedt, B. Berger, B.L. Møller, and R.M. Gleadow. 2015. Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C<sub>4</sub> cereal crop plant to nitrogen and water deficiency over time. *J. Exp. Bot.* 66:1817–1832. doi:10.1093/jxb/eru526
- Noh, B., S.-H. Lee, H.-J. Kim, G. Yi, E.-A. Shin, M. Lee, et al. 2004. Divergent roles of a pair of homologous jumonji/zinc-finger-class transcription factor proteins in the regulation of *Arabidopsis* flowering time. *Plant Cell* 16:2601–2613. doi:10.1105/tpc.104.025353
- O’Brien, R.M. 2007. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* 41:673–690. doi:10.1007/s11135-006-9018-6
- Patrignani, A., and T.E. Ochsner. 2015. Canopeo: A powerful new tool for measuring fractional green canopy cover. *Agron. J.* 107:2312–2320. doi:10.2134/agronj15.0150
- Peterson, B.K., J.N. Weber, E.H. Kay, H.S. Fisher, and H.E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7:E37135. doi:10.1371/journal.pone.0037135
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Qin, J., A. Shi, B. Mou, M.A. Grusak, Y. Weng, W. Ravelombola, et al. 2017. Genetic diversity and association mapping of mineral element concentrations in spinach leaves. *BMC Genomics* 18:941. doi:10.1186/s12864-017-4297-y
- Robinson, G.K. 1991. That BLUP is a good thing: The estimation of random effects. *Stat. Sci.* 6:15–32. doi:10.1214/ss/1177011926
- Rochette, N.C., and J.M. Catchen. 2017. Deriving genotypes from RAD-Seq short-read data using Stacks. *Nat. Protoc.* 12:2640–2659. doi:10.1038/nprot.2017.123
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425. doi:10.1093/oxfordjournals.molbev.a040454
- Sheard, L.B., X. Tan, H. Mao, J. Withers, G. Ben-Nissan, T.R. Hinds, et al. 2010. Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. *Nature* 468:400–405. doi:10.1038/nature09430
- Sherry, R.A., K.J. Eckard, and E.M. Lord. 1993. Flower development in dioecious *Spinacia oleracea* (Chenopodiaceae). *Am. J. Bot.* 80:283–291. doi:10.1002/j.1537-2197.1993.tb13801.x
- Shi, A., B. Mou, J. Correll, S.T. Koike, D. Motes, J. Qin, et al. 2016a. Association analysis and identification of SNP markers for *Stemphylium* leaf spot (*Stemphylium botryosum* f. sp. *spinacia*) resistance in spinach (*Spinacia oleracea*). *Am. J. Plant Sci.* 7(12):1600–1611. doi:10.4236/ajps.2016.712151.
- Shi, A., B. Mou, and J.C. Correll. 2016b. Association analysis for oxalate concentration in spinach. *Euphytica* 212:17–28. doi:10.1007/s10681-016-1740-0
- Shi, A., J. Qin, B. Mou, J. Correll, Y. Weng, D. Brenner, et al. 2017. Genetic diversity and population structure analysis of spinach by single-nucleotide polymorphisms identified through genotyping-by-sequencing. *PLoS One* 12:E0188745. doi:10.1371/journal.pone.0188745
- Slatkin, M. 2008. Linkage disequilibrium—Understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* 9:477–485. doi:10.1038/nrg2361
- Steidle Neto, A.J., D.C. Lopes, F.A.C. Pinto, and S. Zolnier. 2017. Vis/NIR spectroscopy and chemometrics for non-destructive estimation of water and chlorophyll status in sunflower leaves. *Biosyst. Eng.* 155:124–133. doi:10.1016/j.biosystemseng.2016.12.008
- Tessmer, O.L., Y. Jiao, J.A. Cruz, D.M. Kramer, and J. Chen. 2013. Functional approach to high-throughput plant growth analysis. *BMC Syst. Biol.* 7(Suppl 6):S17. doi:10.1186/1752-0509-7-S6-S17
- Thomas, G., K.B. Jacobs, M. Yeager, P. Kraft, S. Wacholder, N. Orr, et al. 2008. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* 40:310–315. doi:10.1038/ng.91

- Tripodi, P., D. Massa, A. Venezia, and T. Cardi. 2018. Sensing technologies for precision phenotyping in vegetable crops: Current status and future challenges. *Agronomy (Basel)* 8(4):57. doi:10.3390/agronomy8040057
- Turner, S.D. 2014. qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. bioRxiv 005165. doi:10.1101/005165
- Wang, X., D. Singh, S. Marla, G. Morris, and J. Poland. 2018. Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. *Plant Methods* 14:53. doi:10.1186/s13007-018-0324-5
- Woebecke, D.M., G.E. Meyer, K. Von Bargen, and D.A. Mortensen. 1995. Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. ASAE* 38:259–269. doi:10.13031/2013.27838
- Wold, S., K. Esbensen, and P. Geladi. 1987. Principal component analysis. *Chemom. Intell. Lab. Syst.* 2:37–52. doi:10.1016/0169-7439(87)80084-9
- Xiang, Y., N. Tang, H. Du, H. Ye, and L. Xiong. 2008. Characterization of *OsZIP23* as a key player of the basic leucine zipper transcription factor family for conferring abscisic acid sensitivity and salinity and drought tolerance in rice. *Plant Physiol.* 148:1938–1952. doi:10.1104/pp.108.128199
- Xu, C., C. Jiao, H. Sun, X. Cai, X. Wang, C. Ge, et al. 2017. Draft genome of spinach and transcriptome diversity of 120 *Spinacia* accessions. *Nat. Commun.* 8:15275. doi:10.1038/ncomms15275
- Xu, J., X.-Y. Wang, and W.-Z. Guo. 2015. The cytochrome *P450* superfamily: Key players in plant development and defense. *J. Integr. Agric.* 14:1673–1686. doi:10.1016/S2095-3119(14)60980-1
- Yang, J., N.A. Zaitlen, M.E. Goddard, P.M. Visscher, and A.L. Price. 2014a. Advantages and pitfalls in the application of mixed model association methods. *Nat. Genet.* 46:100–106. doi:10.1038/ng.2876
- Yang, W., Z. Guo, C. Huang, L. Duan, G. Chen, N. Jiang, et al. 2014b. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi:10.1038/ncomms6087
- Yeager, M., N. Orr, R.B. Hayes, K.B. Jacobs, P. Kraft, S. Wacholder, et al. 2007. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* 39:645–649. doi:10.1038/ng2022
- Zhang, C., Q. Wang, F. Liu, Y. He, and Y. Xiao. 2017. Rapid and non-destructive measurement of spinach pigments content during storage using hyperspectral imaging with chemometrics. *Measurement* 97:149–155. doi:10.1016/j.measurement.2016.10.058
- Zhang, D., L.M. Iyer, and L. Aravind. 2012. Bacterial GRAS domain proteins throw new light on gibberellic acid response mechanisms. *Bioinformatics* 28:2407–2411. doi:10.1093/bioinformatics/bts464
- Zhang Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42:355–360. doi:10.1038/ng.546
- Zhu, C., and J. Yu. 2009. Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. *Genetics* 182:875–888. doi:10.1534/genetics.108.098863
- Zou, G., and Y. Zuo. 2006. On the sample size requirement in genetic association tests when the proportion of false positives is controlled. *Genetics* 172:687–691. doi:10.1534/genetics.105.049536