Aerial High-Throughput Phenotyping Enabling Indirect Selection for Grain Yield at the Early-generation Seed-limited Stages in Breeding Programs

Margaret R. Krause*, Suchismita Mondal†, José Crossa‡, Ravi P. Singh‡, Francisco Pinto‡, Atena Haghighattalab†§, Sandesh Shrestha**, Jessica Rutkoski††, Michael A. Gore*, Mark E. Sorrells*, Jesse Poland**

* Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, New York, 14853, USA
† Global Wheat Program, International Maize and Wheat Improvement Center (CIMMYT), Ciudad de México, 06600, México
‡ Stakman-Borlaug Center for Sustainable Plant Health, University of Minnesota Twin Cities, St. Paul, Minnesota, 55108, USA
§ Center for Applied Phenomics, University of Minnesota Twin Cities, St. Paul, Minnesota, 55108, USA
** Department of Plant Pathology, Kansas State University, Manhattan, Kansas, 66506, USA
†† Department of Crop Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, USA
ABSTRACT

Breeding programs for wheat and many other crops require one or more generations of seed increase before replicated yield trials can be sown. Extensive phenotyping at this stage of the breeding cycle is challenging due to the small plot size and large number of lines under evaluation. Therefore, breeders typically rely on visual selection of small, unreplicated seed increase plots for the promotion of breeding lines to replicated yield trials. With the development of aerial high-throughput phenotyping technologies, breeders now have the ability to rapidly phenotype thousands of breeding lines for traits that may be useful for indirect selection of grain yield. We evaluated early generation material in the irrigated bread wheat (*Triticum aestivum* L.) breeding program at the International Maize and Wheat Improvement Center to determine if aerial measurements of vegetation indices assessed on small, unreplicated plots were predictive of grain yield. To test this approach, two sets of 1,008 breeding lines were sown both as replicated yield trials and as small, unreplicated plots during two breeding cycles. Vegetation indices collected with an unmanned aerial vehicle in the small plots were observed to be heritable and moderately correlated with grain yield assessed in replicated yield trials. Furthermore, vegetation indices were more predictive of grain yield than univariate genomic selection, while multi-trait genomic selection approaches that combined genomic information with the aerial phenotypes were found to have the highest predictive abilities overall. A related experiment showed that selection approaches for grain yield based on vegetation indices could be more effective than visual selection; however, selection on the vegetation indices alone would have also driven a directional response in phenology due to confounding between those traits. A restricted selection
index was proposed for improving grain yield without affecting the distribution of
phenology in the breeding population. The results of these experiments provide a
promising outlook for the use of aerial high-throughput phenotyping traits to improve
selection at the early-generation seed-limited stage of wheat breeding programs.

INTRODUCTION

For small grains as well as many other crops such as soybean and rice, limited
seed yield per plant after deriving fixed lines can delay replicated yield testing until
sufficient quantities of seed are accumulated through one or more generations of seed
increase. As a result, breeding lines are sown in rows or small plots in the field for the
purpose of seed increase prior to yield testing. The generation at which this occurs
depends on the crop and breeding scheme followed, though in theory this may begin as
early as the F₃ with yield testing at the F₄. Breeding lines at this stage are also typically
subject to visual selection, largely for phenology and disease resistance, as more
extensive phenotyping is limited by time and costs. The ability to accurately cull breeding
lines with low potential in the early generation seed-limited stage is highly desirable
because it reduces subsequent expenditure on costly yield testing (Brennan, 1988).

Selection at the early generation seed-limited stage of a breeding program relies
on the assumption that the phenotype of a breeding line when sown in a small,
unreplicated plot is predictive of its eventual performance in a larger replicated yield trial.
Selection on small plots of unreplicated entries can be applied most effectively for
qualitative traits controlled by a limited number of genetic loci with major effects, high
heritability, and limited genotype-by-environment interaction (Snape & Simpson, 1984).
Selection on quantitative traits, however, becomes more challenging due to lower
heritability and difficulty in obtaining effective quantitative measurements (Bernardo, 2003).

Numerous studies have shown that plants exhibit plastic phenotypic responses to the competitive stress imposed by their neighbors (White & Harper, 1970; Edmeades & Daynard, 1979; Dornbusch et al., 2011). The level of competitiveness occurring in small plots does not completely mimic that which occurs in yield trials or farmers’ fields, though it will be more representative than single space-planted plants. In addition, the small plot size increases error in grain yield measurements. The yields of single wheat plants or small plots have been observed to exhibit low heritability (Fonseca & Patterson, 1968; Mitchell et al., 1982) and weak correlations with the yields of replicated yield trials (Syme, 1972; Knott & Kumar, 1975). Non-genetic factors can also drive variation in trait values when breeding material is evaluated in small plots (reviewed by Rebetzke et al., 2014). For example, variation in canopy height and architecture can lead to light competitiveness among different breeding lines sown in adjacent plots (Clarke et al., 1998). Finally, the large number of breeding lines at this stage of the program is effectively prohibitive for these desired measurements of grain yield. Therefore, it can be difficult to directly and accurately select for grain yield at the early generation seed-limited stages of breeding programs.

A potential strategy to improve selection accuracy at the early generation seed-limited stages of wheat breeding programs would be to perform indirect selection on a highly heritable secondary trait, namely traits that exhibit strong correlations with grain yield when measured in replicated yield trials (reviewed by Fischer & Rebetzke, 2018). Several measurable physiology traits in wheat have been shown to fit these criteria. For
example, Rebetzke et al. (2002) demonstrated that selection for reduced carbon isotope discrimination at the early generation stage increased grain yield and aboveground biomass in wheat when grown under drought stress. Quail et al. (1989) found significant correlations between wheat grain yield in the F7/F8 generation with fruiting efficiency at maturity when measured in the F3 generation. In Condon et al., (2008) stomatal aperture-related traits measured on 1.6 m² plots in the F4-F6 generations were as effective as visual selection in identifying wheat breeding lines with high yield potential.

While these and other studies demonstrate the potential for improving grain yield through indirect selection on secondary traits, the feasibility of applying these approaches within breeding programs remains limited due to the high cost and time required to evaluate these traits in the field on large numbers of breeding lines. For indirect selection at the early generation seed-limited stage to be feasible, the phenotyping approach must be amendable to high-throughput data collection to cover the large number of breeding lines sown, and the cost must not exceed the amount saved on subsequent yield trials by culling low potential lines based on secondary traits.

Recent applications of advances in the remote sensing industry to plant breeding have aimed to enable rapid, field-based phenotyping of physiological and other traits on large numbers of breeding lines while reducing labor, time, and cost with respect to traditional approaches (Cabrera-Bosquet et al., 2012; Araus & Cairns 2014; Pauli et al., 2016). In recent years, a range of ground- and aerial-based high-throughput phenotyping (HTP) systems utilizing various cameras and sensors have been developed and tested to expand the efficiency of phenotypic data collection (Andrade-Sanchez et al., 2014; Crain et al., 2016; Haghighattalab et al., 2016; Watanabe et al., 2017). These proof-of-concept
approaches have now reached a validation stage to be applied in breeding programs for improved selections.

Vegetation indices (VIs), which provide an integrated measurement of canopy structure and photosynthetic activity based on the amount of light reflected off of the crop canopy (Huete et al., 2000; Reynolds & Langridge, 2016), are highly amendable to HTP.

A number of studies have demonstrated strong correlations between VIs and grain yield in wheat (Babar et al., 2006; Prasad et al., 2007; Prasanna et al., 2013). In addition, spectral reflectance traits have been shown to improve the accuracy of grain yield prediction in wheat when integrated with genetic information in genomic selection (GS) (Rutkoski et al., 2016; Sun et al., 2017; Crain et al., 2018; Krause et al., 2019). These previous efforts have leveraged the replicated yield trial stage of breeding programs to evaluate the ability to predict grain yield using traits collected with aerial HTP systems.

While these approaches may serve to provide breeders with yield predictions earlier in the season, thereby facilitating greater resource-use efficiency at harvest, the potential cost savings may not justify the expense of collecting the HTP traits. Furthermore, once a field experiment with full-size, replicated plots has been established, it is in the best interest to complete harvesting for a direct measure of grain yield. In the pipeline of a breeding program, a greater benefit could come from applying HTP at the early stages of the breeding cycle when limited seed availability prevents the assessment of grain yield in replicated full-sized plots. If HTP traits can facilitate greater selection accuracy at early stages, programs can select at a higher intensity, thereby promoting fewer and better lines to more expensive replicated yield trials.
Plant breeding strategies in small grains – including pedigree selection, bulk
selection, single seed descent, and doubled haploids, among others – require one or more
generations of spaced plantings for seed increase prior to replicated yield testing, opening
up the possibility for indirect selection on secondary traits collected with HTP. The bread
wheat breeding program at the International Maize and Wheat Improvement Center
(CIMMYT) utilizes a “selected bulk” scheme from the F2 through F5 generations in
which selected spikes from a cross are harvested and bulked to form the next generation
(Singh et al., 1998). At the F5 or F6 generation, individual plants or spikes are selected
and promoted to the next stage as independent derived lines, which are sown in beds as
double-row 0.7m × 0.8m unreplicated small plots (SP) at the Campo Experimental
Norman E. Borlaug (CENEB) in Ciudad Obregón, Sonora, México. This method enables
tens-of-thousands of breeding lines to be planted within a small area for seed increase and
visual selection prior to the first year of replicated yield testing. In the field, lines are
selected and advanced based on uniformity, disease resistance, and other agronomic
characteristics. A second round of visual selection is then performed on the seed
following harvest to retain breeding lines with desirable grain characteristics for
promotion to the first year of replicated yield trials (YT) (van Ginkel et al., 2002).

Detailed phenotyping of the SP stage of CIMMYT’s irrigated bread wheat
breeding program is challenging due to the large number of lines sown each breeding
cycle. As an example, over 40,000 lines were sown as SP at CENEB during the 2016-17
cycle. Aerial-based HTP platforms, however, enable screening of thousands to tens-of-
thousands of plots within a short time span and with less labor, opening up the possibility
of selection on secondary HTP traits at the SP stage of the CIMMYT wheat breeding program.

A potential challenge that may limit the utility of aerial HTP for use in indirect selection for grain yield is the strong associations between VIs and phenology that are often observed (Magney et al., 2016; Duan et al., 2017). When this confounding of traits occurs, selection schemes based on the relationship between VIs and grain yield would indirectly affect the distribution of phenology, disproportionally favoring later maturing entries. Although late maturity often confers superior yields, it is not always favorable as it may expose the crop to terminal and continual high temperature stress and places constraints on cropping rotations within the following season (Joshi et al., 2007; Mondal et al., 2013). In addition, CIMMYT breeders aim to maintain a range of maturity within the germplasm to allow the development of elite varieties for different growing regions. Therefore, it is important to consider approaches that account for the influence of phenology on grain yield when selecting on VIs.

Selection indices represent a practical breeding strategy to account for these relationships between traits. Selection indices are linear combinations of traits that allow for selection on multiple traits simultaneously on the basis of a single value (reviewed by Céron-Rojas & Crossa, 2018). They are derived using the genetic and phenotypic correlations between traits in conjunction with weights assigned to each trait according to its importance. Restricted selection indices are a type of selection index that enables the improvement of one or more traits while holding other traits constant by setting their weights to zero (Kempthorn & Nordskog, 1959; Openshaw & Hadley, 1984). Restricted
selection indices could therefore be used with HTP and phenology records to select
indirectly for grain yield while avoiding driving changes in the distribution of phenology.

The primary objective of this study was to assess the potential for utilizing VIs
collected with aerial HTP systems at the early-generation seed-limited stage of the
breeding program to improve selection for grain yield in wheat.

MATERIALS AND METHODS

Field Trial Design

During the 2016-17 and 2017-18 breeding cycles, two sets of 1,008 entries from
CIMMYT's first-year yield trials (YT) were evaluated for agronomic and HTP traits at
ceneb in Ciudad Obregón, Mexico. A unique set of 1,008 entries was assessed in each
cycle with no overlap in breeding lines between cycles except for check varieties. Each
year, the 1,008 entries were arranged in 36 trials containing 28 entries each. Trials were
sown in an α-lattice design with 2 replicates and 6 incomplete blocks per replicate, with
each block containing 5 entries. Check varieties “Borlaug100 F2014” and “Kachu #1”
were included twice per replicate.

Each set of 1,008 entries contained full-sib families ranging in size from 1 to 34
full-sibs with an average of 4 full-sibs per family. In the first replicate of the YT, entries
were ordered by pedigree, such that full-sibs were sown adjacent to one another in
families in a serpentine order with the check varieties appearing as the first two plots of
the replicate. In the second replicate, entries and checks were completely randomized.
Plots in the YT consisted of two 80 cm beds containing three rows each for a total plot
size of 2.8m × 1.6m.
To empirically evaluate the correlation between small plot performance, including traits measured on these plots, and replicated yield trials, the same 1,008 entries in each breeding cycle were sown in parallel at CENEB as unreplicated 1m × 0.8m small plots (SP) containing three rows each and were evaluated for both agronomic and HTP traits. The 1,008 entries were sown in a serpentine pattern with the order reflecting that of the first replicate of each trial in the YT. During the 2016-17 breeding cycle, the two checks “Borlaug100 F2014” and “Kachu #1” appeared at the beginning of each set of 28 entries. In the 2017-18 breeding cycle, the checks were randomly dispersed within each set of 28 entries. Additional check varieties “MISR 1” and “Baj #1” were added during the 2017-18 breeding cycle and sown ten times each at random within the field.

The YT were sown in pre-irrigated fields on 29 November 2016 and 21 November 2017. The SP were sown in dry soil and received the first irrigation on 25 November 2016 and 23 November 2017. Plots were visually evaluated during the growing season for days to heading (DTHD) and days to maturity (DTMT). DTHD and DTMT were evaluated as the number of days to reach the heading or maturity stages starting from the date of sowing for the YT and from the date of first irrigation for the SP. In the YT, DTHD and DTMT were evaluated on the first replicate only, due to the moderate to high heritability of these traits. Although the breeding program does not typically assess grain yield at the SP stage, both the SP and YT were harvested using plot combine harvesters and then weighed manually for grain yield (GY).

In addition to this proof-of-concept experiment, the SP stage of the formal CIMMYT breeding program was assessed for HTP traits during the 2016-17 breeding cycle. These entries were sown as unreplicated 0.7m × 0.8m beds containing two rows
each. Plots were arranged in a serpentine pattern with full-sibs sown adjacent to one another in families. A check variety was sown after every 50 plots, alternating between varieties “Borlaug100 F2014” and “Kachu #1”. Those lines that were visually selected by breeders were promoted to first-year YT in 2017-18. Therefore, the breeding lines evaluated as YT and SP in 2017-18 as part of the proof-of-concept experiment had also been evaluated for HTP as part of the formal SP stage of the breeding program during the previous cycle. These plots are herein referred to as SPBP. The SPBP were visually assessed for DTHD and DTMT on a qualitative scale with respect to the nearest sown “Borlaug100 F2014” check. Plots were scored as “early”, “mid”, “late”, and “very late” for DTHD and DTMT. The check “Borlaug 100 F2014” was considered to have medium relative maturity with a score of “mid” for DTHD and DTMT. The first 300 plots of the SPBP were harvested and weighed manually for GY, enabling a direct comparison of the response to selection for visual versus HTP selection.

Aerial High-Throughput Phenotyping

Ground Control Points

The trials were prepared for aerial imaging by distributing ground control points (GCPs) within the field. The GCPs were 27.9 cm × 15.2 cm white polyvinyl sheets mounted on posts. Post heights were 1.2 m in the YT and 1.0 m in the SP. The GPS locations of each GCP were measured with a R4 RTK GPS (Trimble, Sunnyvale, CA) with a horizontal accuracy of 0.025m and a vertical accuracy of 0.035m.

Aerial HTP Equipment

A Matrice 100 quadcopter unmanned aerial vehicle (UAV) (DJI, Shenzhen, China) equipped with a MicaSense RedEdge camera (MicaSense, Seattle, WA) was
flown over both the YT and SP at multiple time-points during the grain filling developmental growth stage. The Matrice 100 has a maximum takeoff weight of 3600g and can hover for 22-28min with no payload. Following a modification to accommodate an additional battery pack, the maximum no payload hover time was increased to 33-40min. The MicaSense RedEdge weighs 150g and captures one image in 12-bit RAW format per second. Reflectance was recorded at five broadband wavelengths: blue (460-510nm), green (545-575nm), red (630-690nm), near infrared (820-860nm), and red edge (712-722nm). During the 2017-18 breeding cycle, a MicaSense Downwelling Light Sensor (MicaSense, Seattle, WA) integrated with the RedEdge was added on the top of the UAV to improve calibrations when ambient light conditions change throughout the flight.

**Calibration Panel**

Prior to each flight, images of a MicaSense Calibrated Reflectance Panel (MicaSense, Seattle, WA) were captured with the MicaSense RedEdge for use in post-processing of the imagery to account for the ambient light conditions at the time of flight. As provided by MicaSense, the absolute reflectance values of the panel were: 0.366 (blue), 0.366 (green), 0.391 (red), 0.411 (near infrared), and 0.399 (red edge).

** Acquisition of Aerial Imagery**

During the 2016-17 breeding cycle, the UAV was flown with a speed of 12 km/hr at an altitude of 30m, giving an image resolution of 2 cm/pixel. In the 2017-18 breeding cycle, the altitude and speed of the UAV were reduced to 16m and 7 km/hr, increasing the image resolution to 1 cm/pixel. Flights were conducted within 1h of solar noon to minimize variation due to the solar zenith angle (Gu et al., 1992).
Aerial Imagery Post-Processing and Data Extraction

A semi-automated data analysis pipeline presented in Haghighattalab et al., (2016) was utilized to analyze each image set and extract plot-level phenotypic values. In summary, the following steps were performed to process each HTP data collection time-point: 1) align the aerial images and building sparse point clouds, 2) import GCP GPS coordinates and geo-reference the images, 3) construct a dense point cloud, 4) create a digital elevation model (DEM), 5) generate an orthomosaic using the DEM, 6) calculate vegetation indices (VIs), and 6) extract the plot-level VI data. The VIs measured were: normalized difference vegetation index (NDVI), green NDVI (GNDVI), and red-edge NDVI (RENDVI). Each VI was calculated following:

$$\frac{\rho_{X} - \rho_{Red}}{\rho_{X} + \rho_{Red}}$$ (1)

where $\rho_{Red}$ represents the spectral reflectance recorded in the red region of the light spectrum and $\rho_{X}$ represents the spectral reflectance recorded in the near-infrared, green, and red-edge regions for NDVI, GNDVI, and RENDVI, respectively.

Genotyping and Relationship Matrix Development

The 2,016 entries were genotyped with genotyping-by-sequencing (GBS) using PstI-MspI (Poland et al., 2012). The Tassel 5 GBS v2 pipeline was used to process raw Illumina reads and call single nucleotide polymorphism (SNP) markers with minor allele frequency (MAF) of 0.01 (Glaubitz et al., 2014). The unique tags from the Tassel 5 GBS v2 pipeline were aligned to the International Wheat Genome Sequencing Consortium’s RefSeq v1.0 assembly of Chinese Spring (IWGSC, 2018) using Bowtie2 (Langmead & Salzberg 2012) to get a maximum number of tags with unique mapping. The tags with a
mapping quality of at least 20 were selected for SNP calling. The SNPs called from the production step were filtered with three criteria: inbreeding coefficient of at least 0.8, Fisher Exact Test ($p$-value<0.001) to determine biallelic single locus SNPs (Poland et al., 2012), and Chi-square test for biallelic segregation with 96 percent expected inbreeding. SNPs that passed at least one filtering criteria were retained, and the subsequent SNP set was filtered to remove those with greater than 50 percent missing data and less than 0.01 MAF. Entries with over 50 percent missing data were also removed from the analysis.

The final dataset consisted of 13,271 SNP markers for 1,817 entries. Missing markers were imputed with the marker mean. The genomic relationship matrix ($G$) between entries was calculated according to Endelman and Jannink (2012). The additive pedigree matrix ($A$) was calculated as twice the coefficient of parentage.

**Genetic Value Estimation**

Genetic values of each trait for each entry were estimated for the YT and SP experiments separately within each breeding cycle, as well as for the SPBP plots. For the YT, best linear unbiased predictors (BLUPs) were calculated for the agronomic traits and for each VI at each time-point by fitting the following mixed model:

$$y_{ijkl} = \mu + g_i + t_j + r_{k(j)} + b_{l(jk)} + \epsilon_{ijkl} \quad (2)$$

where $y_{ijkl}$ is the trait value; $\mu$ is the overall mean; $g_i$ is the random genetic effect for genotype $i$ (BLUP), which are assumed to be independently and identically distributed according to a normal distribution $g_i \sim iid \ N(0, \sigma_g^2)$; $t_j \sim iid \ N(0, \sigma_t^2)$ is the random effect for trial $j$; $r_{k(j)} \sim iid \ N(0, \sigma_r^2)$ is the random effect for replicate $k$ within trial $j$; $b_{l(jk)} \sim iid \ N(0, \sigma_b^2)$ is the random effect for block $l$ within replicate $k$ and trial $j$; and $\epsilon_{ijkl} \sim iid \ N(0, \sigma_\epsilon^2)$ is
the residual effect. For DTHD and DTMT, which were evaluated on the first replicate only, the random effect for replicate was removed from model (2).

For the SP and SPBP experiments, BLUPs were calculated for the agronomic traits and VIs as by fitting the model:

\[ y_i = \mu + g_i + \varepsilon_i \]  
(3)

where \( y_i \) is the trait value; \( \mu \) is the overall mean; \( g_i \sim iid \ N(0, \sigma_g^2) \) is the random genetic effect for genotype \( i \); and \( \varepsilon_i \sim iid \ N(0, \sigma_{\varepsilon_i}^2 R_i) \) is the residual variance for genotype \( i \), where \( R_i \) is the correlation matrix for residual effects parameterized as a one-dimensional autoregressive (AR1) process in the column direction to account for potential spatial correlation of observations. An AR1 process model was not applied in the direction of rows as full-sibs were sown adjacent to one another in rows, resulting in confounding between genetic and spatial variation in the row direction.

To account for the effect of phenology, BLUPs for GY and the VIs in the YT and SP were calculated a second time by adding a fixed effect covariate for DTHD to models (2) and (3). BLUPs for the VIs in the SPBP were corrected for DTHD including the qualitative DTHD scores as a fixed effect covariate in model (3).

Significant outliers (\( p \)-value < 0.05) were identified using Studentized Residuals and were removed from the analysis. In addition, the dataset was subset to remove two families that contained disproportionately high numbers of full-sibs. After further removing entries based on missing genotypic data, the final dataset included 839 and 920 entries in the 2016-17 and 2017-18 breeding cycles, respectively.

To avoid shrinkage on the same data twice (once during the calculation of iid BLUPs and again in the GS models), BLUPs were de-regressed by dividing by their
reliability \( 1 - \frac{\text{PEV}}{\sigma_g^2} \), where \( \text{PEV} \) is the prediction error variance of the BLUP and \( \sigma_g^2 \) is the genotypic variance (Garrick et al., 2009).

Weights for the error variances of the BLUPs were calculated as follows:

\[
\frac{1}{\sqrt{1 - H^2}} \cdot \frac{1 - r^2}{\sigma_g^2} \cdot \frac{1 - H^2}{\rho^2 + H^2} \]  

(4)

where \( H^2 \) is the proportion of the total variance explained by the genotypic variance component, \( \sigma_g^2 \), and \( r^2 \) are the reliabilities of the BLUPs (Garrick et al., 2009). These weights were used in downstream genomic prediction analyses to preserve information about heterogeneous variances driven by the differences in plot size and replication between the YT and SP. Without weights, this information would be otherwise ignored in a two-step genomic prediction procedure.

**Heritability, Trait Correlation, and Response to Selection**

To estimate the narrow-sense heritability, variance components for each of the agronomic and HTP traits were estimated according to the following model:

\[
y = Xb + Zg + e \]  

(5)

where \( y \) is a vector of de-regressed BLUPs calculated from models (2) or (3); \( X \) is the fixed effects design matrix; \( b \) is a vector of fixed effects; \( Z \) is the random effects design matrix; \( g \) is a vector of random effects for genotype estimated assuming \( g \sim N(0, G\sigma_g^2) \) where \( G \) is the genomic relationship matrix and \( \sigma_a^2 \) is the additive genetic variance; and \( e \) is a vector of residuals where \( e \sim \text{iid } N(0, I\sigma_e^2) \), \( \sigma_e^2 \) is the residual variance, and \( I \) is the identity matrix. Narrow-sense heritability was calculated as \( h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \).

Trait phenotypic correlations were calculated as the Pearson’s correlations between the iid BLUPs for each trait derived from models (2) and (3). Variance
components for deriving the genetic correlations between each pair of traits were
calculated according to the bivariate model:

\[
\begin{bmatrix}
  y_1 \\
  y_2
\end{bmatrix} =
\begin{bmatrix}
  X & 0 \\
  0 & X
\end{bmatrix}
\begin{bmatrix}
  u_1 \\
  u_2
\end{bmatrix} +
\begin{bmatrix}
  Z_1 & 0 \\
  0 & Z_2
\end{bmatrix}
\begin{bmatrix}
  a_1 \\
  a_2
\end{bmatrix} +
\begin{bmatrix}
  e_1 \\
  e_2
\end{bmatrix}
\]  

(6)

where \( y_1 \) and \( y_2 \) are vectors of de-regressed BLUPs for each trait calculated from models (1) or (2); \( X \) is the fixed effects design matrix, which is the same for both traits; \( u_1 \) and \( u_2 \) are vectors of fixed effects for each trait; \( Z_1 \) and \( Z_2 \) are the random effects design matrices for each trait; \( a_1 \) and \( a_2 \) are vectors of random effects for each trait; and \( e_1 \) and \( e_2 \) are vectors of residuals for each trait. The model was fit assuming \( \begin{bmatrix}
  a_1 \\
  a_2
\end{bmatrix} \sim N(0, G \otimes H) \), where \( G \) is the genomic relationship matrix and \( H \) is the variance-covariance matrix for the two traits. For the residual, \( \begin{bmatrix}
  e_1 \\
  e_2
\end{bmatrix} \sim N(0, I \otimes R) \), where \( I \) is the identity matrix and \( R \) is the residual variance-covariance matrix for the two traits. The resulting variance components were used to calculate the genetic correlation between each pair of traits according to:

\[
\rho_{g_{12}} = \frac{\text{cov}_{g_{12}}}{\sqrt{\text{var}_{g_{11}} \text{var}_{g_{22}}}}
\]  

(7)

where \( \rho_{g_{12}} \) and \( \text{cov}_{g_{12}} \) are the genetic correlation and covariance, respectively, between the traits, and \( \text{var}_{g_{11}} \) and \( \text{var}_{g_{22}} \) are the genetic variances for each trait.

Response to selection was assessed by identifying the top performing breeding lines according to traits measured in the SP and calculating the percent difference in YT trait values between the means of those superior breeding lines and the total population. Response to selection was estimated for selection intensities from 100 (no selection) to 20 percent of breeding lines. Response to visual and HTP selection in the SPBP was measured by calculating the percent difference in mean GY of the SPBP between the selected lines and the total population.
Selection Schemes

A train-test (TRN-TST) partitioning scheme was developed to estimate model accuracy within each breeding cycle. In each partition, 80 percent of the total number of lines within a breeding cycle was sampled to form the TRN set, while the remaining 20 percent of entries formed the TST set. Whole families were grouped during sampling to avoid full-sibs from the same family appearing in both the TRN and TST sets, which greatly inflates this prediction accuracy but does not represent a real-world implementation for forward selection into new breeding crosses.

High-Throughput Phenotyping Selection

To assess the ability to predict GY of the YT using solely HTP information from the SP stage, the following linear regression model was fit using GY and HTP information from the YT for the entries in the TRN set:

\[ y_i = \mu + z_i \alpha + \varepsilon_i \]  (8)

where \( y_i \) is the de-regressed BLUP of genotype \( i \) for GY in the YT calculated from model (2), \( \mu \) is the overall mean, \( z_i \) is the de-regressed BLUP of genotype \( i \) for an VI time-point observed in the YT calculated from model (2), \( \alpha \) is the regression coefficient for the HTP trait, and \( \varepsilon_i \) is the residual for genotype \( i \).

This model was then applied to predict GY for the breeding lines in the TST set. The de-regressed BLUPs for the same VI time-point observed in the SP calculated from model (3) were assigned to \( z_i \) while values for GY were “hidden.” Models were assessed for each VI on each HTP time-point.
Univariate Genomic Selection

As a basis for comparison, univariate GS models using genomic marker or pedigree information were developed to predict GY for each breeding cycle following the form:

\[ y_i = \mu + g_i + \varepsilon_i \] (9)

where \( y_i \) are the de-regressed BLUPs for GY in the YT calculated in model (2); \( \mu \) is the overall mean; \( g_i \) is the random effect for genotype \( i \) with \( g_i \sim N(0, G\sigma_g^2) \) or \( g_i \sim N(0, A\sigma_g^2) \), where \( G \) is the genomic relationship matrix and \( A \) is the pedigree matrix; and \( \varepsilon_i \sim iid N(0, R\sigma_e^2) \) is the residual error for genotype \( i \) where \( R \) is the residual variance matrix. Weights from model (4) were included in the diagonal of \( R \). These models contained no information on HTP traits.

Multi-Trait Genomic Selection

Finally, multi-trait GS approaches integrating HTP records with genomic/pedigree information to predict GY were developed following the bivariate model (6) where \( y_1 \) is a vector of de-regressed BLUPs for GY in the YT calculated from model (2); \( y_2 \) is a vector of de-regressed BLUPs for a VI time-point observed in the YT calculated from model (2); \( X \) is the fixed effects design matrix, which is the same for each trait; \( u_1 \) and \( u_2 \) are vectors of fixed effects for GY and the HTP time-point; \( Z_1 \) and \( Z_2 \) are the random effects design matrices for GY and the VI time-point; \( a_1 \) and \( a_2 \) are vectors of random effects for GY and the VI time-point; and \( \varepsilon_1 \) and \( \varepsilon_2 \) are vectors of residuals for GY and the VI trait. This model was fit assuming \( [a_1 | a_2] \sim N(0, G\otimes H) \) or

\( [a_1 | a_2] \sim N(0, A\otimes H) \), where \( G \) is the genomic relationship matrix, \( A \) is the pedigree matrix,
and \( H \) is the variance-covariance matrix for GY and the VI time-point. The weights from model (4) were included in the diagonal of \( R \).

This model was then applied to the breeding lines in the TST set, for which the de-regressed BLUPs for the same HTP time-point observed in the SP calculated from model (3) represented \( y_2 \) while the values for GY were masked. All possible combinations of HTP time-points, VIs, and relationship matrices were tested.

The same 20 TRN-TST partitions were evaluated for HTP selection, univariate GS, and multi-trait GS to ensure fair comparisons between methods. Predictive abilities were calculated by taking the Pearson’s correlation of the predicted values for the TST set lines with the iid BLUPs for GY in the YT from model (2). The reported predictive abilities are represented by the mean and standard deviation of the 20 TRN-TST partitions.

**Accounting for Phenology**

Due to the strong associations between GY, VIs, and phenology observed in this population, two approaches were evaluated to account for the influence of phenology on GY. In the first approach, iid BLUPs for DTHD observed in the YT calculated from model (2) were included in prediction models (8), (9), and (5) as a fixed effect covariate. For model validation, the predicted values were correlated to iid BLUPs for GY in the YT that had likewise been corrected for DTHD by including DTHD records observed in the YT as a fixed effect covariate in model (2). Prediction models (8), (9), and (5) were tested both with and without this correction for DTHD.

In the second approach, a desired gains selection index taking into account the genetic correlation between GY and DTHD was developed to assess the ability of
prediction models (8), (9), and (5) to identify breeding lines with high GY without selecting directionally on DTHD. The desired gains index takes the form:

\[ \mathbf{b} = \mathbf{G}^{-1}\mathbf{d} \] (10)

where \( \mathbf{b} \) is a vector of the selection index weights, \( \mathbf{G}^{-1} \) is the inverse of the genetic variance-covariance matrix among traits, and \( \mathbf{d} \) is a vector of the desired gains for the traits (Pešek & Baker, 1969).

To avoid directional selection on DTHD, the desired gain for DTHD, \( d_{DTHD} \), was assigned to zero. In each breeding cycle, the desired gain for GY, \( d_{GY} \), was set equal to one standard deviation of the iid BLUPs for GY in the YT. \( \mathbf{G}^{-1} \) was calculated by fitting the bivariate model (5) with the genetic relationship matrix \( \mathbf{G} \) and deregressed BLUPs for GY and DTHD in the YT. \( \mathbf{G}^{-1} \) was calculated separately for each breeding cycle, using records for all lines. The index weights, \( \mathbf{b} \), were then derived in model (10) using the resulting values for \( \mathbf{G}^{-1} \) and assigned values for \( \mathbf{d} \).

To validate the prediction models in the context of the restricted selection index, models (8), (9), and (5) were fit as described to predict GY for the TST set. The GY predictions were then combined with observed records for DTHD in the SP to obtain the “predicted” restricted gains index values:

\[ I_i = b_{GY}X_{GY_i} + b_{DTHD}X_{DTHD_i} \] (11)

where \( I_i \) is the index value for genotype \( i \); \( b_{GY} \) and \( b_{DTHD} \) are the index weights for GY and DTHD, respectively, derived from model (10); \( X_{GY_i} \) is the predicted value for GY of genotype \( i \) from prediction models (8), (9), and (5); and \( X_{DTHD_i} \) is the observed iid BLUP for DTHD in the SP for genotype \( i \) (Smith, 1936; Hazel, 1943).
For model validation, the observed index values for all breeding lines were calculated with model (11) using the observed iid BLUPs for GY and DTHD in the YT derived from model (2). Predictive abilities are represented as the Pearson’s correlations between the predicted and observed index values for the lines in the TST set.

**Software**

Image processing was conducted using Python, open source software QGIS (QGIS Development Team, 2019, QGIS Geographic Information System, Open Source Geospatial Foundation Project, http://qgis.osgeo.org), and Agisoft PhotoScan (Agisoft LLC, St. Petersburg, Russia). Data analyses were implemented in the R environment (R Core Team, 2016, Vienna, Austria) with the package “ASReml-R” (Gilmour et al., 2014) for R. The genomic relationship matrix \((G)\) was calculated using the A.mat() function within the R package “rrBLUP” (Endelman, 2011). The coefficients of parentage for calculating the pedigree relationship matrix \((A)\) were estimated with the “Browse” application of the International Crop Information System software (McLaren et al., 2000).

**Data Availability**

All phenotypic and genotypic data required to confirm the results presented in this study are available on CIMMYT Dataverse (http://hdl.handle.net/11529/10548379).

**RESULTS**

The distributions of BLUPs for GY in the YT were consistent across the two breeding cycles, while BLUPs for GY in the SP were more variable within and across cycles (Table 1). Both the YT and SP reached the heading and maturity stages around the same time during both breeding cycles (Fig. 1) and were therefore approximately at the
same stage of development when the VIs were recorded. The YT and SP were
phenotyped for VIs with the UAV at two time-points (TP-1: 13 Mar; TP-2: 20 Mar) in
2016-17 and at three time-points (TP-1: 07 Mar; TP-2: 12 Mar in the SP, 13 Mar in the
YT; TP-3: 19 Mar) in 2017-18 (Fig. 1). The SPBP was phenotyped with the UAV on the
same time-points as the YT and SP in 2016-17.

Trait Heritabilities

As expected, heritability estimates for GY were considerably greater for the YT
than for the SP in both the 2016-17 and 2017-18 cycles (Table 2). Accounting for
variation in DTHD resulted in a slight decrease in the heritability of GY in the YT and in
the SP from 2016-17 but had no effect in the SP from 2017-18. For a given VI time-
point, the heritability estimates for the three VIs differed from one another by an average
of 0.03. Due to the overall similarity between VIs for heritability, only the estimates for
NDVI are reported.

Heritability estimates for the NDVI time-points were similar across the YT and
SP and had an average value of 0.79, which was comparable to the heritabilities of
DTHD and DTMT and considerably exceeded heritability estimates for GY. As with GY,
accounting for DTHD in the NDVI time-points resulted in a slight reduction in
heritability. Despite the smaller plot size in the SPBP compared to the SP, heritability
estimates for the NDVI time-points in the SPBP experiment were similar to those for the
NDVI time-points recorded in the SP on the same time-points.

Trait Phenotypic and Genetic Correlations

Strong phenotypic and genetic correlations were observed among and between the
agronomic and VI traits within the YT, SP, and SPBP (Table 1). The phenotypic and
genetic correlations for each pair of traits largely followed the same trends, and therefore only the phenotypic correlations are discussed herein. The strongest correlations were observed between the three VIs recorded on the same time-point, which indicates that there was minimal difference in the information captured among the VIs. As a result, only correlations using NDVI, the most commonly used VI in the context of plant breeding among those tested, are reported and discussed. Very high correlations were also observed for NDVI recorded on different time-points, which suggests that changes in canopy reflectance during the period in which VIs were measured were marginal.

Despite the small plot size and lack of replication in the SP, the correlation for GY between the SP and YT was relatively moderate, though correlations between NDVI and GY were lower in the SP than in the YT. The correlations between NDVI in the SP and GY in the YT ranged from 0.37 to 0.44 and was only slightly reduced for SPBP, which had a smaller plot size than the SP and was measured in a different breeding cycle than GY in the YT.

Correcting for DTHD had a minimal impact on the correlation between NDVI and GY when measured both traits were measured in the SP. The effect was more pronounced for the YT as well as when NDVI and GY were evaluated in the SP and YT, respectively. Strong relationships were also observed between the NDVI time-points and DTHD/DTMT, averaging 0.76 for the YT and SP. Notably, DTHD and DTMT observed in both the YT and SP was moderately predictive of GY in the YT, with correlations ranging from 0.39 to 0.45.
Selection on High-Throughput Phenotyping Traits

By identifying the top breeding lines in terms of NDVI assessed in the SP and SPBP, results show that using NDVI as the selection criterion at the early generation seed-limited stage may result in considerable improvements in GY in the YT (Fig. 2), with projected increases of 2.17, 1.68, and 0.97 percent at selection intensities of 25, 50, and 75 percent, respectively. This approach was observed to more effective than selecting on GY measured in the SP during the 2016-17 cycle, though the two approaches showed similar results for the 2017-18 cycle. However, selecting on NDVI measured in the SP or SPBP would have resulted in significant directional response in DTHD in the YT, primarily favoring later-maturing breeding lines. Conversely, selecting on GY measured in the SP would have had no effect on DTHD in the YT.

When accounting for DTHD in the NDVI measurements in the SP, the response to selection for GY in the YT was reduced to below 1 percent at most levels of selection intensity. Accounting for DTHD in the GY records in the SP also reduced the projected response to selection for GY in the YT, however the decrease was less pronounced. As expected, selecting on DTHD-corrected NDVI or GY records from the SP would have had minimal impact on the distribution of DTHD in the YT.

Visual Versus High-Throughput Phenotyping Selection

Prior to harvest in 2016-17, the breeders made visual selections within the first 300 plots of the SPBP for promotion to the first-year YT in the 2017-18 breeding cycle. Of the 293 breeding lines (excluding checks), 36 were selected, representing a selection intensity of 12.3 percent. GY was measured on both the unselected and selected SPBP plots, and results showed that visual selection provided a 0.38 percent response to
selection. Among the total population of 293, over half were scored as “early” while the remaining were divided relatively evenly among “mid” and “late” scores for DTHD, with less than 2 percent scored as “very late”. The distribution for DTHD among the selected lines approximately reflected this, though all “very late” lines were eliminated.

In the SPBP, the NDVI time-points had correlations with GY of 0.13 for TP-1 and 0.05 for TP-2 (Fig. 3A). Despite these weak relationships, NDVI was effective in identifying SPBP breeding lines with low GY for culling. If the level of selection intensity that the breeders used for visual selection had been applied to the SPBP using the NDVI time-points as the selection criteria, the response to selection in GY of the SPBP would have been 1.54 and 1.40 percent for TP-1 and TP-2, respectively. Of the top 36 lines in terms of NDVI, 25 would have been selected by both TP-1 and TP-2. Only 4 and 5 of the lines visually selected by breeders were among the top 36 in terms of NDVI TP-1 and TP-2, respectively. However, over half of these lines were scored as “late” or “very late” for DTHD.

Since CIMMYT breeders attempt to avoid selecting directionally on DTHD when performing visual selection, a more appropriate comparison would be to identify the top breeding lines in terms of NDVI for each qualitative DTHD score. The breeders visually selected 22, 8, and 6 lines scored as “early,” “mid,” and “late” for DTHD, respectively. By taking subsets of the “early,” “mid,” and “late” lines and identifying the top 22, 8, and 6 lines, respectively, based on NDVI, the response to selection in GY for the SPBP would have been 2.91 percent for TP-1 and 4.23 percent for TP-2 (Fig. 3B). There was still minimal overlap between the lines selected by the breeders and the top lines for NDVI using this approach.
Predictive Abilities

The predictive abilities of the three VIs gave similar results with no individual VI providing a consistent, significant advantage. In addition, the predictive abilities of models using the pedigree relationship matrix were low, with an average of 0.16 for the univariate GS models, likely due to the large numbers of full-sibs and missing information in the pedigree records. Therefore, only the predictive abilities of models using NDVI and the genomic relationship matrix ($G$) are reported and discussed further. When using NDVI traits alone to predict yield in the absence of genomic marker or pedigree information, predictive abilities were moderate, averaging 0.43 and 0.38 in 2016-17 and 2017-18, respectively, with minimal variation observed between NDVI time-points (Fig. 4). This exceeded predictive abilities for univariate GS, though multi-trait GS models that incorporated both NDVI and genomic information provided the highest predictive abilities overall.

When estimating model predictive abilities, two approaches were taken to account for confounding of DTHD with both NDVI and GY. In the first approach, iid BLUPs for DTHD in the YT were included as a fixed effect in the prediction model. This correction had a negligible impact on univariate GS, but the predictive abilities of the HTP selection models were strongly reduced to the extent that they became less predictive than univariate GS. Following the correction, the predictive abilities of the multi-trait GS models were more similar to those of univariate GS.

The second approach was to develop a restricted selection index to identify superior breeding lines for GY without selecting directionally on DTHD while taking into account the genetic correlation between the traits (Fig. 5). The observed index values
showed phenotypic correlations of 0.89 and 0.85 with GY in 2016-17 and 2017-18, respectively, conferring a response to selection of 0.44 in both breeding cycles. As desired, the index had a negligible relationship with DTHD, with correlations of -0.05 and -0.10 in 2016-17 and 2017-18, respectively. The narrow-sense heritability of the index was 0.47 in 2016-17 and 0.64 in 2017-18. Genetic correlations between the index and the NDVI traits in the SP were moderate during the 2016-17 breeding cycle, with values 0.47 and 0.38 for TP-1 and TP-2, respectively. However, low correlations were observed in 2017-18, with values of 0.13, 0.14, and 0.05 for TP-1, TP-2, and TP-3, respectively.

GY predictions in the TST set for HTP selection, univariate GS, and multi-trait GS were used with observed iid BLUPs for DTHD in the SP and index weights to develop the index predictions. The predictive abilities for HTP selection and univariate GS were similar in 2016-17, while in 2017-18 the predictive ability of univariate GS exceeded that of HTP selection (Fig. 6). As with predictions of yield directly, multi-trait GS showed the highest predictive abilities for both breeding cycles.

DISCUSSION

The aerial VI traits were found to be more heritable than GY in both the YT and SP, and minimal differences were observed with respect to heritability among the multiple time-points on which the HTP traits were recorded. Several previous studies have found aerial VI traits to be heritable and repeatable for wheat (Rutkoski et al., 2014; Crain et al., 2017; Sun et al., 2017), though the traits measured in these works were evaluated on replicated trials with plot sizes sufficient for reliably assessing yield. To our knowledge, this is the first report aiming to assess the heritability of aerial VI traits
measured on small, unreplicated plots and estimate the extent to which they are predictive
of GY in replicated yield trials.

It is possible that the estimates of narrow-sense heritability for traits measured in
the SP are partially inflated. At the early stages of small grains breeding programs,
breeders often sow full-sibs derived from the same set of parents adjacent to one another
in the field for logistical ease and to enable visual comparisons among lines within
families. The disadvantage of this approach, however, is that the genetic and spatial
components of the phenotypic variance are confounded. When planting siblings together
in families, Magnussen (1993) showed that the estimates of additive genetic variance and
narrow-sense heritability were biased upward with positive spatial autocorrelation among
neighboring plots. In this study, despite the possible confounding of genetic and spatial
variance, the phenotypic and genetic correlations between the VI traits measured in the
SP and GY in the YT were moderate. The second replicate of the YT was randomized,
and the YT and SP also represented separate field experiments, each with independent
environmental effects. In addition, the HTP traits in the SPBP were also correlated with
GY in the YT, despite having been measured during the prior breeding cycle and on
smaller plot sizes than the SP. Together, these results suggest that the VI traits measured
in at the early-generation seed-limited stages may still be effective for improving GY at
the YT stage even if the genetic and spatial variance components are confounded as a
result of planting full-sibs adjacent to one another.

Harvesting 300 plots of the SPBP for GY enabled a direct comparison between
HTP and CIMMYT’s visual selection scheme because all breeding lines, both selected
and unselected by visual selection, were assessed for both GY and HTP. Although
correlations between NDVI and GY in the SPBP were weak, the NDVI traits were successful in identifying breeding lines with the lowest GY values for culling. The response to selection for GY was substantially greater using aerial NDVI as the selection criterion when compared to the current visual selection method. While these results are encouraging, the sample size of this experiment was relatively small. In addition, the GY values were based on the harvest of unreplicated 0.8m × 0.8m plots, and results from the YT and SP experiments showed that GY recorded in the SP is only moderately predictive of GY at the YT stage. Empirical validation is needed to confirm these results when GY is measured in replicated yield trials.

HTP traits have been shown to improve the ability to predict GY in wheat through multi-trait GS (Rutkoski et al., 2016; Sun et al., 2017; Crain et al., 2018); however, the ability to apply this approach to the seed-limited stage of breeding programs has not been previously evaluated. In this study, GY and HTP records assessed in the YT were integrated with genomic information to train multi-trait GS models. These were then applied to predict GY for the SP, where collecting HTP traits is possible but GY is not typically recorded. Overall, the multi-trait models showed the highest predictive abilities as compared to univariate GS and prediction with the HTP traits alone, though the increase over the latter was marginal. Therefore, it may be more cost-effective for breeding programs to invest in HTP alone as opposed to both HTP and genotyping for GS when selecting at the early-generation seed-limited stages. While the cost of genotyping has reduced greatly over the past decade, the large number of lines in breeding programs at the early stages can make applying GS expensive. However, if both
genomic markers and HTP traits are available, using both sources of information to build prediction models would be optimal.

Multiple findings from this study suggested that using NDVI as a selection criterion at early breeding stages would have driven marked changes in DTHD within the breeding population. Mason and Singh (2014) observed similar directional changes in phenology when selecting on canopy temperature for improving wheat GY under drought. The potential for undesired directional response to selection in confounded traits is not unique to HTP but is rather relevant to any selection scheme in which other traits are confounded with the trait of interest. To account for the confounding effects of phenology, GY and NDVI traits were corrected for phenology by including DTHD as a fixed effect covariate during prediction. Given that strong correlations were observed between GY, DTHD, and NDVI, the predictive abilities of HTP selection and the multi-trait GS models were greatly reduced after accounting for DTHD to the extent that the addition of NDVI traits in multi-trait GS provided no additional predictive power compared to the univariate GS. Estimates of response to selection for GY in the YT were low, less than 1 percent for most NDVI time-points, when using DTHD-corrected NDVI traits evaluated in the SP as the selection criteria.

A limitation of this approach of accounting for trait confounding is that the genetic correlations of DTHD with GY and NDVI are ignored. Any of the genetic variance that may have been contributing to both traits is partitioned into the fixed effect. Selection indices, which take into account genetic correlations between traits, may therefore be a more suitable approach when traits are highly related. In this study, visual selections made in the SPBP showed that the breeders are simultaneously accounting for
relative maturity and do not select directionally on DTHD. A restricted selection index can mimic this by enabling selection for increased GY while holding DTHD fixed. This approach was utilized to compare the predictive abilities of HTP selection, univariate GS, and multi-trait GS to identify breeding lines with high GY potential across the distribution of DTHD. The results from 2016-17 showed that HTP selection gave similar levels of predictive ability as univariate GS when predicting the index, and that the multi-trait GS models that combined genomic marker information with HTP traits further increased predictive abilities. However, in 2017-18, HTP selection performed more poorly than univariate GS, and multi-trait models were only marginal better than the univariate case. These differences in predictive abilities were likely due to the index showing stronger correlations with VI traits in the SP in 2016-17 than in 2017-18. Rutkoski et al. (2016) showed that the level of gain in GS predictive ability from the addition of secondary traits is significantly associated with the genetic correlation between the secondary and target traits.

While VIs alone may not show a sufficiently strong and consistent genetic correlation with a restricted selection index for GY and DTHD to provide significant increases in GS predictive ability, it is likely that breeders will have access to more extensive suites of phenotypic traits in addition to VIs as HTP systems become increasingly automated, customizable, and scalable. For example, hyperspectral reflectance phenotypes, which record spectral reflectance at a large range of wavelengths, were shown to increase prediction accuracy for GY when compared to VIs (Montesinos-López et al., 2017). Beyond spectral reflectance-related traits, deep learning algorithms have been developed to determine DTHD from proximal imagery of wheat canopies.
(Wang et al., 2019), UAV imagery has been used to estimate lodging in wheat as a function of changes in plant height throughout the growing season (Singh et al., 2019), and convolutional neural networks have been trained to identify foliar diseases in maize from aerial imagery (Wu et al, 2019). Looking ahead, these comprehensive suites of traits may allow confounding effects such as those observed in this study to be adequately addressed through integrative selection approaches that account for genetic correlations between traits.

Until that time, the most practical approach may be to leverage useful information from HTP traits without relying on them fully. For example, van Ginkel et al. (2008) compared elite wheat breeding lines for GY under drought and heat stress developed through visual selection and through an integrated approach combining visual selection with measurements of canopy temperature depressions. Although the lines with the highest GY for both methods were statistically, similar, the integrated approach was more effective in identifying and eliminating the lower yielding lines. In this study, we showed that the VI traits could be used in conjunction with qualitative DTHD scores to identify breeding lines with higher GY potential for the “early”, “mid”, and “late” heading subgroups.

Although recent advancements in remote sensing and robotic technologies have produced inexpensive cameras, sensors, UAV, smartphone applications, etc., the overall cost of deploying HTP platforms can be considerable (reviewed by Reynolds et al., 2019). Additional expenses may include the labor required to utilize HTP platforms in the field, training of personnel and dedicated bioinformaticians, the development and implementation of data extraction pipelines, propriety software, and the establishment
and maintenance of systems for storing and organizing large volumes of phenotypic data. Future improvements in the automation of HTP platforms and data analysis workflows as well as the development of open source software may help to alleviate some of the costs of implementing HTP, though they are not likely to decrease below the costs associated with visual selection. Therefore, the breeder must determine if the added investment in HTP can be offset by increases in the rate of genetic gain and/or the costs saved from sowing fewer replicated yield trials. Further studies are needed to compare empirical gains from selection per unit of cost using HTP versus current visual selection methods.

ACKNOWLEDGEMENTS

This work was supported through the US Agency for International Development (USAID) Feed the Future Innovation Lab for Applied Wheat Genomics (Cooperative Agreement No. AID-OAA-A-13-00051). It was also supported in part by the Delivering Genetic Gain in Wheat project, supported by aid from the U.K. Government’s Department of International Development (DFID) and the Bill & Melinda Gates Foundation (Grant No.: OPP1133199). We are thankful to the National Science Foundation Graduate Research Fellowship (Grant No. DGE-1650441) and the U.S. Borlaug Fellows in Global Food Security program for supporting the graduate studies of Margaret R. Krause.

REFERENCES

Andrade-Sanchez, P., Gore, M.A., Heun, J.T., Thorp, K.R., Carmo-Silva, A.E., French, A.N., Salvucci, M.E., & White, J.W. (2014). Development and evaluation of a field-based high-throughput phenotyping platform. *Func. Plant Biol.*, *41*, 68-79.
Araus, J.L. & Cairns, J.E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.*, 19, 52-61.

Babar, M.A., Reynolds, M.P., van Ginkel, M., Klatt, A.R., Raun, W.R., & Stone, M.L. (2006). Spectral reflectance indices as a potential indirect selection criteria for wheat yield under irrigation. *Crop Sci.*, 46, 578-588.

Bernardo, R. (2003). On the effectiveness of early generation selection in self-pollinated crops. *Crop Sci.*, 43, 1558-1560.

Brennan, J.P. (1988). *An economic investigation of wheat breeding programmes*. Agricultural Economics Bulletin 35, Department of Agricultural Economics and Business Management, University of New England, Armidale, NSW, Australia.

Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M.D., & Araus, J.L. (2012). High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. *J. Integr. Plant Biol.*, 54, 312-230.

Céron-Rojas, J.J., & Crossa, J. (2018). *Linear selection indices in modern plant breeding*. Springer, Cham, Netherlands.

Clarke, F.R., Baker, R.J., & DePauw, R.M. (1998). Interplot interference distorts yield estimates in spring wheat. *Crop Sci.*, 38, 62-66.

Condon, A.G., Reynolds, M.P., Brennan, J., van Ginkel, M., Trethowan, R., Rebetzke, G.J., Bonnett, D.G., Richards, R.A., & Farquhar, G.D. (2008). Stomatal aperture related traits and yield potential in bread wheat. In M.P. Reynolds, J. Pietragalla, & H.J. Braun (Eds.) *International Symposium on Wheat Yield Potential: Challenges to International Wheat Breeding* (pp. 126-133). CIMMYT: Mexico, DF.
Crain, J.L., Wei, Y., Barker, J., Thompson, S.M., Alderman, P.D., Reynolds, M., Zhang, N., & Poland, J. (2016). Development and deployment of a portable field phenotyping platform. *Crop Sci., 56*, 965-975.

Crain, J., Reynolds, M., & Poland, J. (2017). Utilizing high-throughput phenotypic data for improved phenotypic selection of stress-adaptive traits in wheat. *Crop Sci., 57*, 648-659.

Crain, J., Mondal, S., Rutkoski, J., Singh, R.P., & Poland, J. (2018). Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *Plant Genome, 11*.

Dornbusch, T., Baccar, R., Watt, J., Hillier, J., Bertheloot, J., Fournier, C., & Andrieu, B. (2011). Plasticity of winter wheat modulated by sowing date, plant population density and nitrogen fertilisation: dimensions and size of leaf blades, sheathes and internodes in relation to their position on a stem. *Field Crops Res., 121*, 116-124.

Duan, T., Chapman, S.C., Guao, Y., & Zheng, B. (2017). Dynamic monitoring of NDVI in wheat agronomy and breeding trials using an unmanned aerial vehicle. *Field Crops Res., 210*, 71-80.

Edmeades, G.O., & Daynard, T.B. (1979). The development of plant-to-plant variability in maize at different planting densities. *Can. J. Plant Sci., 59*, 561-576.

Endelman, J.B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome, 4*, 250-255.

Endelman, J.B., & Jannink, J.L. (2012). Shrinkage estimation of the realized relationship matrix. *G3 Genes Genom. Genet., 2*, 1405-1413.
Fischer, R.A., & Rebetzke, G.J. (2018). Indirect selection for potential yield in early-
generation, spaced plantings of wheat and other small-grain cereals: a review.
Crop Pasture Sci. 69, 439-459.

Fonseca, S., & Patterson, F.L. (1968). Yield component heritabilities and
interrelationships in winter wheat (Triticum aestivum L.). Crop Sci., 8, 614-617.

Garrick, D.J., Taylor, J.F., & Fernando, R.L. (2009). Deregressing estimated breeding
values and weighting information for genomic regression analyses. Genet. Sel.
Evol., 41, 55.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., & Thompson, R. (2014). ASReml
user guide release 4.1 structural specification. VSN International, Hemel
Hempstead, UK.

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., & Buckler,
E.S. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis
pipeline. PloS ONE, 9, e90346.

Gu, X.F., Guyot, G., & Verbrugghe, M. (1992). Evaluation of measurement errors in
ground surface reflectance for satellite calibration. Int. J. Remote Sense., 13,
2531-2546.

Haghighattalab, A., González-Pérez, L., Mondal, S., Singh, D., Schinstock, D., Rutkoski,
J., Ortiz-Monasterio, I., Singh, R.P., Goodin, D., & Poland, J. (2016). Application
of unmanned aerial systems for high throughput phenotyping of large wheat
breeding nurseries. Plant Methods, 12, 35.

Hazel, L.N. (1943). The genetic basis for constructing selection indexes. Genetics, 28,
476-490.
Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., & Ferreria, L.G. (2000). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ., 83, 195-213.

Joshi, A.K., Ferrara, O., Crossa, J., Singh, G., Sharma, R. Chand, R., Parsad, R. (2007). Combining superior agronomic performance and terminal heat tolerance with resistance to spot blotch (*Bipolaris sorokiniana*) in the warm humid Gangetic plains of south Asia. Field Crops Res., 103, 53-61

Kempthorne, O., & Nordskog, A.W. (1959). Restricted selection indices. Biometrics, 15, 10-19.

Knott, D.R., & Kumar, J. (1975). Comparison of early generation yield testing and a single seed descent procedure in wheat breeding. Crop Sci., 15, 295-299.

Krause, M.R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R.P., Dreisigacker, S., Poland, J., Rutkoski, J., Sorrells, M., Gore, M.A., & Mondal, S. (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. G3 Genes Genom. Genet., 9, 1231-1247.

Langmead, B., & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9, 357-359.

Magney, T.S., Eitel, J.U.H., Huggins, D.R. & Vierling, L.A. (2016). Proximal NDVI derived phenology improves in-season predictions of wheat quantity and quality. Agric. For. Meteorol, 217, 46-60.

Magnussen, S. (1993). Bias in genetic variance estimates due to spatial autocorrelation. Theor. Appl. Genet., 86, 349-355.
Mason, R.E., and R.P. Singh, 2014 Considerations when deploying canopy temperature
to select high yielding wheat breeding lines under drought and heat stress.
Agronomy 4: 191-201.

McLaren, C.G., Ramos, L., Lopez, C., & Eusebio, W. (2000). Application of the
genealogy management system. In C.G. McLaren, J.W. White, & P.N. Fox (Eds.)
*International Crop Information System. Technical Development Manual, version VI* (pp. 5.8-5.13). CIMMYT, México: CIMMYT and IRRI.

Mitchell, J.W., Baker, R.J., & Knott, D.R. (1982). Evaluation of honeycomb selection for
single plant yield in durum wheat. *Crop Sci.*, 22, 840-843.

Mondal, S., Singh, R.P., Crossa, J. Huerta-Espino, J., Sharma, I., Chatrath, R., Singh,
G.P., Sohu, V.S., Mavi, G.S., Sukuru, V.S.P., Kalappanavar, I.K., Mishra, V.K.,
Hussain, M., Gautam, N.R., Uddin, J., Barma, N.C.D., Hakim, A., & Joshi, A.K.
(2013). Earliness in wheat: a key adaptation under terminal and continual high
temperature stress in South Asia. *Field Crops Res.*, 151, 19-26.

Montesinos-López, O.A., Montesinos-López, A., Crossa, J., de los Campos, G. Alvarado,
G., Mondal, M., Rutkoski, J., González-Pérez, L., & Burgueño, J. (2017).
Predicting grain yield using canopy hyperspectral reflectance in wheat breeding
data. *Plant Methods*, 13, 4.

Openshaw, S.J., & Hadley, H.H. (1984). Selection indexes to modify protein
concentration of soybean seeds. *Crop Sci.*, 24, 1-4.

Pauli, D., Chapman, S.C., Bart, R., Topp, C.N., Lawrence-Dill, C.J., Poland, J., & Gore,
M.A. (2016). The quest for understanding phenotypic variation via integrated
approaches in the field environment. *Plant Physiol.*, 172, 622-634.
Pešek, J., & Baker, R.J. (1969). Desired improvement in relation to selection indices. *Can. J. Plant Sci.*, 49, 803-804.

Poland, J.A., Brown, P.J., Sorrells, M.E., & Jannink, J.L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS ONE*, 7, e32253.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., & Sorrells, M. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*, 5, 103-113.

Prasad, B., Carver, B.F., Stone, M.L., Babar, M.A., Raun, W.R., & Klatt, A.R. (2007). Potential use of spectral reflectance indices as a selection tool for grain yield in winter wheat under Great Plains conditions. *Crop Sci.*, 47, 1426-1440.

Prasanna, B.M., Araus, J.L., Crossa, J., Cairns, J.E., Palacios, N., Das, B., & Magorokosho, C. (2013). High-throughput and precision phenotyping for cereal breeding programs. In P.K. Gupta & R.K. Varshney (Eds.), *Cereal genomics II* (pp. 314-374). Springer: Dordrecht, Netherlands.

Quail, K.J., Fischer, R.A., & Woods, J.T. (1989). Early generation selection in wheat. I. Yield potential. *Aust. J. Agric. Res.*, 40, 1117-1133.

R Core Team. (2016). R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Rebetzke, G.J., Condon, A.G., Richards, R.A., & Farquhar, G.D. (2002). Selection for reduced carbon isotope discrimination increases aerial biomass and grain yield of rainfed bread wheat. *Crop Sci.*, 42, 739-745.
Rebetzke, G.J., Fischer, R.T.A., van Herwaarden, A.F., Bonnet, D.G., Chenu, K., Rattey, & Fettell, N.A. (2014). Plot size matters: interference from intergenotypic competition in plant phenotyping studies. *Func. Plant Bio.*, 41, 107-118.

Reynolds, D., Baret, F., Welcker, C., Bostrom, A., Ball, J., Cellini, F., Lorence, A., Chawade, A., Khafif, M., Noshita, K., Mueller-Linow, M., Zhou, J., & Tardieu, F. (2019). What is the cost of phenotyping? Optimizing costs for different scenarios. *Plant Sci.*, 282, 14-22.

Reynolds, M., & Langridge, P. (2016). Physiological breeding. *Curr. Opin. Plant Biol.*, 31, 162-171.

Rutkoski, J.E., Poland, J., Mondal, S., Autrique, E., González Pérez, L., Crossa, J., Reynolds, M., & Singh, R. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes Genom. Genet.*, 6, 2799-2808.

Singh, D., Wang, X., Kumar, U., Gao, L., Noor, M., Imtiaz, M., Singh, R.P., & Poland, J. (2019). High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. *Front. Plant Sci.*, 10, 394.

Singh, R.P., Rajaram, S., Miranda, A., Huerta-Espinoza, J., & Autrique, E. (1998). Comparison of two crossing and four selection schemes for yield, yield traits, and slow rusting resistance to leaf rust in wheat. In H.J. Braun, F. Altay, W.E. Kronstad, S.P.S. Beniwal, & A. McNab (Eds.) *Proceedings of the Fifth International Wheat Conference, 10-14 Jun, 1996, Ankara, Turkey* (pp. 93-101).

Smith, H.F. (1936). A discriminant function for plant selection. *Ann. Eugen.*, 7, 240-250.
Snape, J.W. & Simpson, E. (1984). Early generation selection and rapid generation advancement methods in autogamous crops. In W. Lange, A.C. Zeven, & N.G. Hogenboom (Eds.) Efficiency in plant breeding. Proc. 10th Cong. EUCARPIA, 19-24 June, 1983, Wageningen, Netherlands (pp. 82-86). Pudoc: Wageningen, Netherlands.

Sun, J., Rutkoski, J.E., Poland, J.A., Crossa, J., Jannink, J.L., & Sorrells, M.E. (2017). Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. Plant Genome, 10.

Syme, J.R. (1972). Single-plant characters as a measure of field plot performance of wheat cultivars. Aust. J. Agric. Res., 23, 753-760.

van Ginkel, M., Trethowan, R.M., Ammar, K., Wang, J., Lillemo, M. (2002). Guide to bread wheat breeding at CIMMYT. Wheat Program Special Report No. 5. CIMMYT: Mexico, D.F.

van Ginkel, M., Reynolds, M.P., Trethowan, R., & Hernandez, E. (2008). Complementing the breeder’s eye with canopy temperature measurements. In M.P. Reynolds, J. Pietragalla, & J.H. Braun (Eds.) International symposium on wheat yield potential: challenges to international wheat breeding (pp. 134-135). CIMMYT: Edo. de México, Mexico.

Wang, X., Xuan, H., Evers, B., Shrestha, S., Pless, R., & Poland, J. (2019). High-throughput phenotyping with deep learning gives insights into the genetic architecture of flowering time in wheat. GigaScience, 8, giz120.
Watanabe, K., Guo, W., Arai, K., Takanashi, H., Kajiya-Kanegae, H., Kobayashi, M., Yano, K., Tokunaga, T., Fujiwara, T., Tsutsumi, N., & Iwata, H. (2017). High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Front. Plant Sci.*, 8, 421.

White, J., & Harper, J.L. (1970). Correlated changes in plant size and number in plant populations. *J. Ecol.*, 58, 467-485.

Wu, H., Wisner-Hanks, T., Stewart, E.L., DeChant, C., Kaczmar, N., Gore, M.A., Nelson, R.J., & Lipson, H. (2019). Autonomous detection of plant disease symptoms directly from aerial imagery. *Plant Phenome J.*, 2.
|       | YT          | SP          |
|-------|-------------|-------------|
|       | 2016-17     | 2017-18     |
|       | GY          | DTHD        | DTMT        | NDVI TP-1 | NDVI TP-2 | GY          | DTHD        | DTMT        | NDVI TP-1 | NDVI TP-2 | GY          | DTHD        | DTMT        | NDVI TP-3 | NDVI TP-1 | NDVI TP-2 | NDVI TP-3 | NDVI TP-1 | NDVI TP-2 |
| GY    | 6.45 (0.32) | 0.41 (0.61) | 0.44 (0.62) | 0.50 (0.72) | 0.51 (0.67) | 0.32 (0.60) | 0.40 (0.63) | 0.42 (0.69) | 0.43 (0.72) | 0.44 (0.73) | 0.30 (0.78) | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           |
| DTHD  | -           | 74.5 (4.2)  | 0.78 (0.93) | 0.74 (0.82) | 0.80 (0.89) | 0.08 (0.12) | 0.82 (0.96) | 0.65 (0.85) | 0.60 (0.62) | 0.71 (0.81) | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           |
| DTMT  | -           | 122.5 (3.1) | 0.72 (0.80) | 0.80 (0.90) | 0.12 (0.17) | 0.73 (0.89) | 0.69 (0.91) | 0.59 (0.67) | 0.72 (0.86) | -           | 0.03 (0.19) | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           |
| NDVI TP-1 | 0.32 (0.48) | -           | -           | 0.82 (0.95) | 0.94 (0.79) | 0.17 (0.19) | 0.72 (0.75) | 0.69 (0.74) | 0.75 (0.78) | -           | -           | 0.03 (0.33) | -           | -           | -           | -           | -           | -           | -           | -           | -           |
| NDVI TP-2 | 0.34 (0.45) | -           | -           | 0.85 (0.79) | 0.66 (0.503) | 0.14 (0.05) | 0.78 (0.82) | 0.76 (0.86) | 0.76 (0.84) | -           | -           | 121.3 (3.5) | -           | -           | -           | -           | -           | -           | -           | -           | -           |
| GY    | 0.30 (0.78) | -           | -           | 0.14 (0.41) | 0.09 (0.21) | 5.20 (0.64) | 0.12 (0.05) | 0.14 (0.12) | 0.25 (0.27) | -           | -           | 78.2 (3.5)  | -           | -           | -           | -           | -           | -           | -           | -           | -           |
| DTHD  | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           |
| DTMT  | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           | -           |
| NDVI TP-1 | 0.22 (0.54) | -           | -           | 0.48 (0.89) | 0.44 (0.91) | 0.23 (0.33) | -           | -           | 0.774 (0.038) | -           | -           | 78.2 (3.5)  | -           | -           | -           | -           | -           | -           | -           | -           | -           |
| NDVI TP-2 | 0.17 (0.47) | -           | -           | 0.42 (0.74) | 0.49 (0.89) | 0.17 (0.21) | -           | -           | 0.82 (0.90)  | -           | -           | 0.89 (0.90) | -           | -           | -           | -           | -           | -           | -           | -           | -           |

The table shows trait correlations within and between lines in the YT, SP, and SP_{BP} variants for the years 2016-17 and 2017-18.
YT, yield trials; SP, small plots; SPBP, small plots from the 2016-17 breeding program; GY, grain yield; DTHD, days to heading; DTMT, days to maturity; NDVI, normalized difference vegetation index; TP, time-point. The diagonals contain the mean of the iid BLUPs for each trait derived from models (1) and (2). The values in parentheses along the diagonal are the standard deviations of the iid BLUPs for each trait. The upper triangles contain the Pearson’s correlation between each pair of traits. The lower triangles contain the Pearson’s correlations among and between GY and the NDVI time-points following correction for DTHD. The values in parentheses in the upper and lower triangles are the genetic correlations between each pair of traits.
Table 2. Narrow-sense heritability ($h^2$) estimates for the agronomic and NDVI traits

| Breeding Cycle | Trait     | YT $h^2$ | SP $h^2$ | SPBP $h^2$ |
|---------------|-----------|-----------|-----------|-------------|
| 2016-17       | GY        | 0.55 (0.46) | 0.39 (0.40) | -           |
|               | DTHD      | 0.79      | 0.82      | -           |
|               | DTMT      | 0.65      | 0.80      | -           |
|               | NDVI TP-1 | 0.77 (0.58) | 0.69 (0.75) | 0.79 (0.65) |
|               | NDVI TP-2 | 0.84 (0.57) | 0.79 (0.78) | 0.83 (0.65) |
| 2017-18       | GY        | 0.72 (0.65) | 0.37 (0.31) | -           |
|               | DTHD      | 0.82      | 0.85      | -           |
|               | DTMT      | 0.83      | 0.65      | -           |
|               | NDVI TP-1 | 0.79 (0.59) | 0.82 (0.74) | -           |
|               | NDVI TP-2 | 0.79 (0.58) | 0.78 (0.65) | -           |
|               | NDVI TP-3 | 0.83 (0.64) | 0.83 (0.59) | -           |

YT, yield trials; SP, small plots; SPBP, small plots from the 2016-17 breeding program; GY, grain yield; DTHD, days to heading; DTMT, days to maturity; NDVI, normalized difference vegetation index; TP, time-point. Values in parentheses are the narrow-sense heritability for the traits corrected for DTHD.
Figure 1. UAV phenotyping time-points in relationship to heading and maturity dates in each experiment.

YT, yield trials; SP, small plots; TP, time-point; UAV, unmanned aerial vehicle; VI, vegetation index. The histograms represent the heading and maturity dates of the SP and YT in 2016-17 and 2017-18. The dotted red lines indicate the time-points on which each experiment was phenotyped for VIs with the UAV.
Figure 2. Response to selection in the YT when selecting on traits measured in the SP and SP\textsubscript{BP}

SP\textsubscript{BP}, small plots from the 2016-17 breeding program; GY, grain yield; NDVI, normalized difference vegetation index; DTHD, days to heading. The x-axis represents the selection intensity expressed as the percentage of the population selected based on the SP or SP\textsubscript{BP} trait. The y-axis represents the corresponding percent change in the YT trait value.
Figure 3. Visual versus HTP selection in the SP_{BP}

NDVI

Grain Yield (t/ha)

NDVI TP-1

NDVI TP-2

DTHD Subgroup
- Early
- Mid
- Late
- Very Late

Selection Status
- Unselected
- Selected: Visual

NDVI

Grain Yield (t/ha)

NDVI TP-1

NDVI TP-2

Selection Status
- Unselected
- Selected: Visual
- Selected: NDVI
- Selected: Both
The x-axis shows NDVI values measured in the SP$_{BP}$. The y-axis represents GY values assessed in the SP$_{BP}$.

A. The color represents the qualitative DTHD subgroups the breeding lines were assigned to. The shape indicates which lines were visually selected by breeders. Points to the right of the dashed lines indicate lines that would have been selected using NDVI as the selection criteria, given the same level of selection intensity used by the breeders for visual selection.

B. Directional selection on DTHD is avoided by using NDVI to select the top 22, 8, and 6 lines in the “early”, “mid”, and “late” DTHD subgroups, respectively. The color and shape indicate if which lines were selected, and, if so, by which selection method(s).
Figure 4. Predictive abilities of univariate GS, HTP selection, and multi-trait GS uncorrected and corrected for phenology

GS, genomic selection; HTP, high-throughput phenotyping. Predictive ability is expressed as the Pearson’s correlations between predictions and observed iid BLUPs for GY in the YT across 20 random TRN-TST iterations. The univariate and multi-trait GS models were fit using the genomic relationship matrix (G). Corrected predictive abilities were estimated using genetic values that were calculated with DTHD as a fixed effect covariate.
Figure 5. Restricted selection index to increase GY while holding DTHD constant

Plot points are the iid BLUPs for DTHD and GY in the YT in 2016-17 and 2017-18. The color and shape of the points indicate which breeding lines were selected based on the restricted selection index at an intensity of 10 percent. The solid black line is the regression line. The dashed black lines indicate the trait means for GY and DTHD. The dashed red lines indicate the response to selection in both traits.
Figure 6. Predictive abilities of univariate GS, HTP selection, and multi-trait GS for a restricted selection index to increase GY while holding DTHD constant

Predicted index values for the TST set were calculated using the predicted values for GY and observed iid BLUPs for DTHD in the SP. The predictive ability is expressed as the Pearson’s correlations between predicted values and observed index values calculated using the observed iid BLUPs for GY and DTHD in the YT across 20 random TRN-TST iterations. The univariate and multi-trait GS models were fit using the genomic relationship matrix (G).