

**APPLICATIONS OF HIGH-THROUGHPUT PHENOTYPING IN
SOYBEAN (*Glycine max* L. Merr) BREEDING**

by

Fabiana Freitas Moreira

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Agronomy

West Lafayette, Indiana

May 2020

**THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL**

Dr. Katy Martin Rainey, Chair

Department of Agronomy

Dr. Luiz F. Brito

Department of Animal Sciences

Dr. Keith A. Cherkauer

Department of Agricultural and Biological Engineering

Dr. Mitchell R. Tuinstra

Department of Agronomy

Dr. Shaun N. Casteel

Department of Agronomy

Approved by:

Dr. Ronald F. Turco

To my family

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my advisor Dr. Katy Martin Rainey for the endless support, teachings, patience, and opportunities. Katy, thank you so much for all that you did for me, especially for believing me and giving me this life-changing opportunity to come to Purdue. I am very proud and lucky for being your student.

I thank Drs. Luiz F. Brito, Keith A. Cherkauer, Mitchell R. Tuinstra, Shaun N. Casteel, Karen Hudson, and Bill Muir for their guidance and intellectual contributions while serving on my committee. Your support and knowledge were fundamental to improve my research and career.

I am grateful to everybody that was part of the Soybean Breeding Lab at Purdue throughout my Ph.D.: Miguel, Diana, Lais, Thais, Rima, Alencar, Meng, Monica, Chris, Curtis, David, Vince, Ryan, and the numerous undergraduate assistants. Your support, hard work, and motivation made this research possible. I am grateful to have worked with such a great team.

I also thank Stuart, Bilal, Anthony, Ryan, and Aaron for your help during the UAS imagery collection and analyses.

I must also express my sincere gratitude to Dr. Luiz Brito and Dr. Hinayah Oliveira for generosity in assisting with statistical analyses and willingness to share your insightful ideas for the enhancement of this research. Thank you for doing your best to help me with everything that I needed.

I would like to thank the professors and staff of the Agronomy Department at Purdue for the teachings and support.

I also would like to say a big thank you to all the friends that I met during this journey: Marina, Luciana, Rodrigo, Debora, Guilherme, Isis, Chico, Miguel, Victor, Jorge, Sam, Valerie, Brad, Savannah, Seth, Blake, Rupesh, and many others who crossed my path at Purdue and through any word or action encouraged me to keep going.

And, of course, thank you to my family. I would never be able to accomplish anything without them.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	5
LIST OF TABLES.....	8
LIST OF FIGURES	9
LIST OF ABBREVIATIONS.....	11
ABSTRACT.....	14
CHAPTER 1. RESEARCH OVERVIEW	16
1.1 Introduction.....	16
1.2 Breeding Schemes and Genetic Gain.....	17
1.3 Field-based High-Throughput Phenotyping	19
1.4 Potential Targets for Soybean Improvement	22
1.5 Hypotheses and Objectives.....	23
CHAPTER 2. IMPROVING THE EFFICIENCY OF SOYBEAN BREEDING WITH HIGH-THROUGHPUT CANOPY PHENOTYPING.....	25
2.1 Abstract.....	25
2.2 Background.....	26
2.3 Methods.....	28
2.3.1 Description of Breeding Populations.....	28
2.3.2 Phenotypic Data.....	29
2.3.3 Statistical Data Analysis and Selection Methods of PR	30
2.3.4 Evaluation of PYT and AYT	32
2.4 Results.....	33
2.4.1 PR	33
2.4.2 PYT Selection Category Performance.....	34
2.4.3 AYT Yield Performance.....	35
2.5 Discussion.....	36
2.6 Conclusions.....	39
CHAPTER 3. INTEGRATING HIGH-THROUGHPUT PHENOTYPING AND STATISTICAL GENOMIC METHODS TO GENETICALLY IMPROVE LONGITUDINAL TRAITS IN CROPS.....	40

3.1	Abstract	40
3.2	Introduction.....	40
3.3	Phenotyping Longitudinal Traits	43
3.4	Modeling Longitudinal Traits	46
3.5	Statistical Genetic Models	49
3.5.1	Repeatability Model.....	50
3.5.2	Multiple-trait Model	51
3.5.3	Random Regression Model.....	53
3.6	Implementation of Genomic Selection for Longitudinal Traits.....	56
3.7	Detecting QTL and Causal Variants Associated with Longitudinal Traits	59
3.8	Challenges and Future Developments	61
3.8.1	Non-Additive Effects and GxE.....	62
3.8.2	Complementary “-omics” Technologies.....	63
3.8.3	Deep Learning	64
CHAPTER 4. HIGH-THROUGHPUT PHENOTYPING AND RANDOM REGRESSION		
MODELS REVEAL TEMPORAL GENETIC PATTERNS OF SOYBEAN BIOMASS.....		66
4.1	Abstract	66
4.2	Introduction.....	67
4.3	Material and Methods	70
4.3.1	Plant Materials, Field Experiments, and Genotypic Data.....	70
4.3.2	High-Throughput Phenotyping	71
4.3.3	Predicting Above-Ground Biomass.....	74
4.3.4	Random Regression Models	75
4.3.5	Genetic Parameters	77
4.3.6	Genomic Prediction of Breeding Values	77
4.3.7	Genome-Wide Association Study.....	78
4.4	Results.....	79
4.4.1	Predicting Above-Ground Biomass.....	79
4.4.2	Random Regression Models	81
4.4.3	Genetic Parameters	81
4.4.4	Genomic Prediction of Breeding Values	83

4.4.5	Genome-Wide Association Study.....	84
4.5	Discussion.....	86
4.5.1	High-Throughput Phenotyping of Soybean Above-Ground Biomass.....	86
4.5.2	Genetic Architecture of Soybean Temporal Above-Ground Biomass	88
4.5.3	Potential of Genomic Selection to Improve Soybean Temporal Above-Ground Biomass.....	91
4.6	Conclusions.....	93
CHAPTER 5. GENERAL CONCLUSIONS.....		95
APPENDIX A. CHAPTER 2.....		98
APPENDIX B. CHAPTER 4.....		100
REFERENCES		107

LIST OF TABLES

Table 2.1 Pearson’s correlations for PR 2015 (above diagonal) and 2016 (bellow diagonal) and narrow-sense heritability.....	33
Table 2.2 Progeny row selection categories choosing the ten top-ranked lines for advanced yield trials (AYT).....	36
Table 4.1 Descriptions and formulas of imagery features investigated in this study.	73
Table 4.2 Akaike’s information criterion (AIC) values calculated for the random regression models tested using homogenous and heterogeneous residual variances.	81
Table 4.3 Selected single nucleotide polymorphisms (SNPs) associated with above-ground biomass mapped inside potential candidate genes in the soybean genome.	86
Table A.1 Adjusted mean and standard deviation for yield (Kg/ha) and R8 (days to maturity) by selection criteria for preliminary yield trials (PYT) early and late in 2016 and 2017.	98
Table B.1 List of candidate genes for all selected SNP associated with above-ground biomass in soybean according to SoyBase (SoyBase.org).....	103

LIST OF FIGURES

Figure 2.1 Overview of data collection and processing to acquire average canopy coverage (ACC).	30
Figure 2.2 Number of lines selected from progeny rows (PR) 2015 and 2016 by each selection criteria.	32
Figure 2.3 A. Box plot of yield (Kg/ha) and B. Yield adjusted by R8 (Yield R8) distribution for all lines selected by each selection criteria (Yield, ACC and Yield ACC) for preliminary yield trials (PYT) early and late in 2016 and 2017. Diamond indicates mean for each selection criteria. The line crossing the box plots are representing the median for each group. No significative (ns): $p >$ 0.05 ; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$	35
Figure 3.1 Schematic workflow of longitudinal data analyses. Different remote-sensing tools most commonly used for high-throughput phenotyping monitoring crop growth and development. Comparative overview of the potential models for genomic analysis, together with examples of outputs and computational demand.	43
Figure 4.1 Data collection timeline by environment (2017_ACRE, 2018_ACRE, and 2018_Romney). Planting date in parentheses below environment. UAV: unmanned aerial vehicle. Phenological stages (Fehr and Caviness, 1977): R1: beginning bloom; R5: beginning seed; R7: beginning maturity; R8: full maturity.	71
Figure 4.2 Performance of above-ground biomass prediction for each environment. Predictions were performed using the least absolute shrinkage and selection operator (LASSO) regression, and the partial least squares regression (PLSR) methods. The performance of predictions was evaluated using the root mean squared error (RMSE) and coefficient of determination (R^2), using a 10-fold cross-validation set.....	79
Figure 4.3 Phenotypic distributions of predicted above-ground biomass (g/m^2) across environments by days after planting using the Least Absolute Shrinkage and Selection Operator (LASSO) Regression. Horizontal lines in the box indicate the median.	80
Figure 4.4 Narrow-sense heritability estimated for each day after planting	82
Figure 4.5 Estimated genetic correlations of above-ground biomass between different days after planting.	82
Figure 4.6 Genomic prediction accuracy based on Pearson's correlation coefficient (r) for each day after planting.	83
Figure 4.7 Regression coefficients' patterns of genomic estimated breeding values for each day after planting.	84
Figure 4.8 Effects for the selected single nucleotide polymorphisms (SNPs) across days after planting, in each duration category. Duration categories were defined as long-duration (SNPs present for more than 30 consecutive days), mid-duration (SNPs present for more than 10	

consecutive days but less than 30), short-duration (SNPs present for less than 10 consecutive days), and intermittent (SNPs at non-consecutive days). The grey color indicates a zero effect..... 85

Figure A.1 Box plot of adjusted R8 (days to maturity) distribution for lines selected by each selection categories (Yield, ACC and Yield|ACC) for preliminary yield trials (PYT) early and late in 2016 and 2017. Diamond indicates mean for each selection categories. The line crossing the box plots are representing the median for each class. No significant (ns): $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$ 99

Figure A.2 Distribution of average canopy coverage of the checks by days after planting for progeny rows 2015 and 2016..... 99

Figure B.1 Yield (Kg/ha) phenotypic distribution by environment..... 100

Figure B.2 Correlation (R) plots between observed and predicted above-ground biomass (g/m^2) for each environment for the least absolute shrinkage and selection operator (LASSO) regression and partial least squares regression (PLSR)..... 100

Figure B.3 Variable importance for each environment for the least absolute shrinkage and selection operator (LASSO) regression. 101

Figure B.4 Performance of above-ground biomass prediction leaving one environment out. Predictions were performed using the least absolute shrinkage and selection operator (LASSO) regression, and the partial least squares regression (PLSR) methods. The performance of predictions was evaluated using the root mean squared error (RMSE) and coefficient of determination (R^2)..... 101

Figure B.5 Phenotypic distribution of predicted above-ground biomass (g/m^2) using the Least Absolute Shrinkage and Selection Operator (LASSO) Regression by days after planting for each environment. 102

LIST OF ABBREVIATIONS

ΔG	Genetic gain
2017_ACRE	Location ACRE during year 2017
2018_ACRE	Location ACRE during year 2018
2018_Romney	Location Romney during year 2018
3D	Three-dimensional
3PL	Three-parameter logistic
4PL	Four-parameter logistic
5PL	Five-parameter logistic
ACC	Average canopy coverage
ACRE	Agronomy Center for Research and Education
AGB	Above-ground biomass
AIC	Akaike's information criterion
ANN	Artificial neural networks
ANOVA	Analysis of variance
AR-1	First-order autoregressive
AYT	Advanced yield trial
BIC	Bayesian's information criterion
BLUE	Best linear unbiased estimation
BLUP	Best linear unbiased prediction
CC	Canopy coverage
CIE 2.0	Crop image extraction version 2
CT	Canopy temperature
CT	Computed tomography
CV	Cross-validation
DAP	Days after planting
DL	Deep learning
FA	Factor analysis

GBLUP	Genomic BLUP
GS	Genomic selection
GWAS	Genome-wide association studies
GxE	Genotype by environment interactions
h^2	Narrow-sense heritability
HTP	High-throughput phenotyping
HTTP	High-throughput phenotyping platforms
i	Selection intensity
L	Breeding cycle
LAI	Leaf area index
LASSO	Least absolute shrinkage and selection operator
LD	Linkage disequilibrium
LI	Light interception
LiDAR	Laser-imaging detection and ranging
MANOVA	Multivariate analysis of variance
MAS	Marker assisted selection
MTM	Multiple-trait model
MTRRMS	Multiple-trait random regression models
NAM	Nested association mapping
PCA	Principal component analysis
PLSR	Partial least squares regression
PR	Progeny rows
PSA	Projected shoot area
PYT	Preliminary yield trial
QTL	Quantitative trait loci
R^2	Coefficient of determination
R8	Full maturity
REML	Restricted maximum likelihood
RGB	Red, green, blue

RMSE	Root mean squared error
RNN	Recurrent neural networks
RRM	Random regression model
S	Selection differential
SNP	Single nucleotide polymorphism
SoyNAM	Soybean nested association mapping
SR	Simple repeatability
ssGBLUP	Single-step GBLUP
TPAC	Throckmorton-Purdue Agricultural Center
UAS	Unmanned aerial systems
UAV	Unmanned aerial vehicle
VI	Vegetation indices
VID 1.0	Vegetation indices derivation version 1
YOR	Year of release

ABSTRACT

The rapid expansion of high-throughput phenotyping (HTP) platforms in agronomic research has led to a major shift in plant science towards time-series phenotyping that can track plant development through its life stages, providing an opportunity to dissect the genetic basis of longitudinal traits. Plant breeders can now phenotype large populations during the growing season and promote the desirable genetic gain for the traits of interest in specific time points within their breeding program. The biggest challenge is to use the various tools in a practical way to understand the many complexities of plant growth and development and breeding implications. This dissertation explores interdisciplinary frameworks to assess different applications of HTP for longitudinal traits in soybean breeding. We provide a review outlining the current analytical approaches in quantitative genetics and genomics to adequately use high-dimensional phenomic data. Examples, advantages, and pitfalls of each approach, and future research directions and opportunities are explored. Among longitudinal traits in soybean, average canopy coverage (ACC) and above-ground biomass (AGB) are promising traits to strategically improve yield gain. Soybean ACC is highly heritable, with a high genetic correlation to yield and can be effectively measured by unmanned aerial systems (UAS). This study reveals that progeny rows selection using yield given ACC (Yield|ACC) selected the most top-ranking lines in advanced yield trials, which emphasizes the value of HTP of ACC for selection in the early stages of soybean breeding. In addition, we developed a HTP methodology to predict soybean AGB over several days after planting (DAP) and assessed the quantitative genomic properties of temporal AGB using random regression models (RRM). Our results show that AGB narrow-sense heritability estimates fluctuated over time and the genetic correlation of AGB between DAP decreased as the days went further apart. Considering the trait heritability, high prediction accuracies suggest that AGB is a good indicator trait for genomic selection in soybean breeding. Different genomic regions were found to be associated with AGB over time with potential time-specific SNPs playing a role in the trait expression. Similarly, candidate genes were identified with potential different patterns of expression over time. This study presents novel genetic knowledge for longitudinal traits in soybean and may contribute to the development of new cultivars with high yield and optimized AGB. This is the first application of RRM for genomic evaluation of a longitudinal trait in soybean

and provides a foundation that RRM can be an effective approach to understand the temporal genetic architecture of a longitudinal trait in other crops.

CHAPTER 1. RESEARCH OVERVIEW

1.1 Introduction

Soybean [*Glycine max* (L.) Merrill] is one of the most important crops grown in the world, with 125.40 million hectares planted in the 2017/2018 cropping season, yielding 340.86 million tons worldwide (www.fas.usda.gov). Although soybean was domesticated in China, currently, the United States, Brazil, and Argentina account for almost 90% of the world production and exports (<http://www.fas.usda.gov>). In the past century, soybean yield has increased at an average rate of approximately 31 kg/ha annually (Ainsworth et al., 2012; Ray et al., 2013). The growing yield trend, along with the increase in cultivated area has more than doubled the global soybean production in the last decade (www.fas.usda.gov). These remarkable numbers indicate that soybean is a high value and profitable crop, since soybean's seeds are composed by approximately 40% protein and 20% oil. Consequently, soybean is the world's largest source of protein for animal feed and the second largest source of vegetable oil (USDA, 2018).

The success and expansion of soybean are due to a combination of factors, including improvement in farming technology and management, rising atmospheric CO₂ concentration and genetic improvement, with the development of cultivars more productive, adapted and resistant to various biotic and abiotic stress factors (Specht et al., 1999). However, the forecast for the coming years projects that soybean yield and production will have to grow in even larger rates. According to the United Nations, the world population will reach over nine billion people by 2050, which combined with increasing wealth, higher purchasing power and changes in diets will require between 70 to 100% increase in food production (Godfray et al., 2010; Alexandratos and Bruinsma, 2012). In this scenario, soybean average yield increases will have to shift from 1.3% per year to 2.4% per year (Ray et al., 2013).

There is no simple solution to this challenge, especially with limitation of agricultural land and environmental challenges. However, plant breeding will have a major role to sustain and enhance soybean yield for the future. It has been proposed that approximately half of the increase in soybean yield is due to genetic improvement (Specht et al., 1999; Board and Kahlon, 2011). Moving forward, plant breeders will have to strategically choose breeding targets and incorporate technological innovations to achieve maximum yield gain.

1.2 Breeding Schemes and Genetic Gain

Plant breeding for any crop can be summarized as the process of creating populations or assembling germplasm accessions with useful genetic variation and selecting individuals with superior phenotypes, with the objective of genetically improving the performance of cultivars in the most efficient way possible (Fehr, 1998). Increasing yield through breeding is a function of genetic gain (ΔG), which can be defined as the rate of increase in yield over a given period. The expression for genetic gain is $\Delta G = h^2 S/L$, where h^2 is narrow-sense heritability, S is the selection differential and L is the breeding cycle time. Considering that S is equivalent to the selection intensity (i) multiplied by the phenotypic standard deviation (σ_p), then $\Delta G = h^2 i \sigma_p / L$ (Falconer and Mackay, 1996). This equation comprises four key elements that influence breeding progress and that can be targets for improvement: (1) the amount of phenotypic variation in the population (σ_p); (2) the proportion of this phenotypic variation that can be attributed to genetic factors (h^2); (3) selection intensity (i); (4) and the breeding cycle (L) (Moose and Mumm, 2008; Xu et al., 2017). Thus, genetic gain can be improved by increasing σ_p , h^2 and i and decreasing L , providing the framework for breeding strategies.

Phenotypic variation depends on the population used and is defined as the total observed variance, represented by the sum of variance resulting from genotype (σ_g^2), variance due to environment (σ_e^2), as well the variance of their interaction. Narrow-sense heritability is the proportion genotypic variance due additive effects passed along generations and is represented as the ratio of additive variance (σ_a^2) by phenotypic variance (σ_p^2). Therefore, increasing h^2 and σ_p relies on two aspects: increasing genetic variation and decreasing environmental variation. In this context, a breeder carefully controls or manage environmental effects through experimental design, choosing uniform fields and cultural practices, *post hoc* statistical approaches, or both, resulting in a higher proportion of genetic variance represented by the phenotypic variance. In addition, in more advanced stages of a breeding program multi-environmental trials are added to the pipeline, in order to diminish environmental factors and detect genotype by environment interactions (GxE). In the other point of the equation, breeders aim to increase genetic variability in the population, which is determined at the time the population is formed and maintained in following cycles of selection (Hallauer et al., 2010). Genetic diversity relevant to the trait of interest is essentially obtained through crossing selected parental lines, either elite, exotic, interspecific material.

Although it seems simple, breeders use a range of methods to characterize and create favorable genetic variation in order to have potential genetic combinations and balanced elite performance. Molecular and genomic approaches including mutation, gene editing, transgenic, gene mapping and discovery, can help identify alleles or haplotypes or create new variations and add them to the existing genetic pool (Xu et al., 2017). In addition, the genetic variation on the target population must be adequately analyzed to obtain precise estimate of variance components and identify differences among genotypes.

Selection intensity can be defined as the percentage of genotypes selected from the original population. While increasing phenotypic variation and heritability is straightforward, increasing selection intensity comes with the burden of decreasing genetic variability in subsequent generations. Thereby, to maintain breeding populations for long-term gain, selection intensity must be reasonably chosen for maintaining high genetic variability (Hallauer et al., 2010). Usually, to increase selection intensity, breeders increase the population size. However, there is a balance between genetic variation and sample size, depending on how large populations can be managed and the cost involved (Cobb et al., 2019).

Breeding cycle is the last piece of the equation and there are several ways to accelerate the breeding process. First by increasing the efficiency and accuracy of selection, using approaches such as indirect selection, marker assisted selection (MAS) or genomic selection (GS) (Meuwissen et al., 2001; Mrode, 2014; Cobb et al., 2019). Yield is a very complex trait, being controlled by many genomic regions, and it can be measured only by the end of the growing season. Indirect selection by an early season correlated trait might select high yielding lines faster. MAS and GS enable the rapid selection of superior genotypes, skipping steps of phenotypic evaluation, which accelerates the breeding cycle (Crossa et al., 2017; Xu et al., 2017). Secondly, speeding up growth and development, using greenhouses, growth chambers or off-season nurseries are common practices that allow to manage more generations per year. New approaches that use extended photoperiods and controlled temperature regime or complete *in vitro* meiosis and mitosis system to achieve are overcoming the biological limits for a rapid generation cycling (De La Fuente et al., 2013; Yan et al., 2017; Li et al., 2018).

1.3 Field-based High-Throughput Phenotyping

Phenotyping refers to measure or observe any characteristic or trait of an organism, including morphological, developmental, biochemical and physiological properties. Phenotype is the result of combined effect of the genotype, the environment and their interaction, and can be envisioned as a dynamic and continuous stream of data changing with development and in response to different environment conditions (Cobb et al., 2013). Even before the advent of plant breeding as a science, farmers have been making selection based on phenotypes, choosing the better-performing plants or seeds. Phenotypic selection remains an important part of plant breeding (Cobb et al., 2019); however, plant breeders embrace new tools to understand heritability and genetic variation of valuable traits, improving power and efficiency of selection (Fehr, 1998). Genomic information have provided an economically feasible way to survey genetic variation, allowing breeders to: trace pedigree relationships, identify and select mutations, discover genes, gain insight into the genetic control of traits with genome wide association studies (GWAS), or QTL mapping, predict genomic values, perform genomic selection (GS), and perform indirect selection based on genetic correlations (Cobb et al., 2013). Most of these techniques relies on precision phenotyping data, having the potential to capitalize on advances in plant phenotyping. Therefore, continuing to develop new techniques available to breeders offer the potential to increase the rate of genetic improvement.

In recent years, plant phenotyping has been through a revolution, using remote sensing technologies and advanced software to achieve high-throughput, cost-effectiveness and reveal and dissect novel traits (Montes et al., 2007; Furbank and Tester, 2011; Fiorani and Schurr, 2013; Araus and Cairns, 2014; Coppens et al., 2017). Field-based high-throughput phenotyping mostly relies on imaging fields with different sensors, such as RGB (Xavier et al., 2017), multispectral (Zaman-Allah et al., 2015), hyperspectral (Gonzalez-Dugo et al., 2015) and light detection and ranging (LIDAR) (Sun et al., 2017b) cameras, mounted on unmanned aerial vehicles (UAV) (Yang et al., 2017) or ground-based vehicles (Salas Fernandez et al., 2017). In addition to the development and advance of sensors and platforms, data handling and processing became increasingly important for high-throughput phenotyping, in order to translate the imagery into useful information and knowledge in timely fashion (Tardieu et al., 2017).

The goal of imaging plants is to measure a phenotype quantitatively, which is possible due the interaction between radiant energy and plants (reflecting, absorbing or transmitting photons),

that is captured by a sensor as reflected radiation (Fiorani et al., 2012; Li et al., 2014). Thus, plants are optical objects and each component of a plant tissue has a characteristic spectral signature coming from wavelength specific properties of absorbance, reflectance and transmittance of the vegetation surface (Schowengerdt, 2012; Li et al., 2014). The spectral signature of plants changes during the life cycle, for instance, during senescence there is an increase of reflectance in the red region caused by a loss of chlorophyll (Schowengerdt, 2012). A broad range of imaging devices covering different regions of the electromagnetic spectrum is available to use in plant phenotyping (Tardieu et al., 2017).

RGB cameras are the most widely used sensor in plant science, providing information of aspects of plant architecture, such as canopy cover, biomass, height and canopy color (Baker et al., 1996; Casadesús et al., 2007; Lee and Lee, 2013; Bendig et al., 2014; Casadesús and Villegas, 2014; Holman et al., 2016). They are, basically, conventional digital cameras, that captures information in the red (~600 nm), green (~550 nm), and blue (~450 nm) spectral bands of visible light and transform them into a digital image. The use of an RGB camera in phenotyping is an attractive methodology for being robust, low cost, easy to operate, and simple to extract data (Zhao et al., 2019).

The number of detected spectral bands can be extended to complement the RGB bands. Depending on how many bands are added to system, the camera can be called multispectral (3 to 10 bands) or hyperspectral (more than 10 bands). In addition, spectral cameras differ from regular RGB cameras by being designed to give information of specific wavelengths and reflectance and typically are more sensitive for detection of differences in reflectance. Some bands have already been identified as being valuable for plant research, such as the transition from red to the infrared spectrum that are strongly reflected by plant tissue, but there is a lot of room for discovery (Li et al., 2014). Usually, the information from the spectral regions is extracted in the form of indices, which minimize disturbances caused by external factors, enhancing the variability due to vegetation characteristics (Peñuelas and Filella, 1998; Viña et al., 2011). These indices are created using composite arithmetic operations between spectral reflectance data of different wavelengths, such as the ratio of the difference and sum of reflected infrared and visible red measurements, known as the normalized difference vegetation index (NDVI) (Richards and Jia, 2006).

Plant spectral reflectance offers the same uses of a RGB camera, but brings a new range of applications, especially into physiological status of plants, such as green biomass, chlorophyll and

carotenoids content, senescence, disease detection. (Peñuelas and Filella, 1998; Pan et al., 2015; Mahlein, 2016; Potgieter et al., 2017). However, it also relies on the development of calibration models to transform the sensor data into spectral information. Despite the imaging method, it is necessary an understanding of the measurement principle in order to have reliable imagery to obtain meaningful biological information.

There are some key advances with high-throughput phenotyping platforms. Plant breeding is a number's game: the more genotypes you evaluate the higher the probability of identifying superior ones. With that, plant breeders want to phenotype a large number of lines, which usually is labor intensive, logistically unfeasible and subjective when humans assess the trait. High-throughput phenotyping enables screenings of thousands of genotypes in multi-locations field trials relatively effortless and with high precision. In addition, it is possible to capture several traits within a single image, including traits out of the range of human eyes. More interesting, it is a nondestructive and noninvasive procedure, making it possible to image the same experimental plot throughout the course of its lifecycle and reducing replication due sampling (Großkinsky et al., 2015; Zhao et al., 2019).

High-throughput phenotyping brings a new facet for plant genetic research, allowing time-series measurements that can track the progression of plant growth and its interaction with environment. Some traits, such as light interception, biomass accumulation, response to drought stress, were usually treated as static points (i.e. measured once in the growing season), ignoring their dynamic nature (Montes et al., 2007). In addition, phenotypes were mostly restricted to one trait at a time. As a consequence, information regarding plant functioning and activation of genes and interacting gene networks at different stages of plant development and in responses to environmental stimulus was lost (Wu and Lin, 2006; Montes et al., 2007). Now we are moving forward to high dimensional phenotype on an whole-plant scale, accounting for environmental and temporal variation, which might enhance our understanding and bridge the gap of genotype-phenotype relationships (Granier and Vile, 2014). To deal with this high dimensionally, advanced statistical approaches are needed and the set of traits might be treated as a function-value trait, since they are essentially a function, changing continuously in response to other variables (Stinchcombe and Kirkpatrick, 2012; Granier and Vile, 2014).

1.4 Potential Targets for Soybean Improvement

There are opportunities to increase yield in soybean by targeting each of the genetic gain equation factors, incorporating new biotechnology, statistical, phenotyping approaches. Here, we will focus on potential targets in improving soybean breeding using high-throughput phenotyping. The big question is what traits should we phenotype? Yield is a very complex trait, meaning that is controlled by different genomic regions and their interaction with environment, and it is the result of all the processes of growth and development. Historically, soybean genetic gain has been achieved by selecting desirable cultivars based on high and stable grain yield (Frederick et al., 1994). With the advance of molecular markers, several quantitative trait loci (QTL) associated with yield have been identified (Orf et al., 1999; Yuan et al., 2002; Chung et al., 2003; Kabelka et al., 2004; Kim et al., 2012). However, most of the QTLs identified in those studies haven't been introgressed into cultivars or being used in breeding programs, for reasons regarding their validity (Ainsworth et al., 2012). Genetic studies based on yield will continue to be implemented in breeding programs, but an alternative to increase yield is to identify strategic traits, understanding changes that have enable past improvements in soybean yield. With that, some studies focused on investigating the physiological basis of yield gain (Koester et al., 2014, 2016; Großkinsky et al., 2015, 2018; van Eeuwijk et al., 2018). However, most of these traits have not been used in breeding programs, due to the lack of techniques to rapidly and precisely assess them, and to do so nondestructively to preserve the breeding material.

Koester et al. (2014) showed that time to canopy closure did not change between older and newer cultivars, indicating that canopy needs to be optimized in a genotype to achieve its yield potential. Therefore, genotypes with lower canopy coverage may not have the potential to be a high yield line. This reflect the importance of canopy light interception (LI) for crop growth and yield. It has been demonstrated that soybean LI can be measured as a function of canopy coverage from images taken from above the plot (Purcell, 2000). Furthermore, soybean canopy can also be effectively derived from images acquired with UAS (Xavier et al., 2017; Hearst, 2019). Xavier et al. (2017) found that average canopy coverage (ACC) measured at early season is a highly heritable trait ($h^2 = 0.77$) with a promising genetic correlation with grain yield (0.87). In addition, early season canopy coverage improved the predictive ability for yield in genomic prediction models (Jarquin et al., 2018). Thus, early season canopy coverage has the ability to be used as a secondary trait for indirect selection for yield or as covariables to boost yield estimates.

When comparing soybean cultivars released between 1947-1971 with plant introductions, newer cultivars had enhanced biomass accumulation (Cregan and Yaklich, 1986). Another study with cultivars released 1911-1982, also demonstrated that yield improvement in the newer cultivars had increased biomass as a contributing factor (Frederick et al., 1991). Kumudini et al. (2001) showed that increased biomass accumulation and longer leaf area duration contributed most of the yield increase in the newer cultivars evaluated. Rate and total biomass accumulation were also found to have driven increased yield in another study and suggested to be used in future yield gains (De Bruin and Pedersen, 2009). More recently, Koester et al. (2014) investigated canopy light interception, light energy conversion into biomass, and partitioned biomass into seed in 24 soybean cultivars released 1923-2007 and found that energy conversion into biomass improved with year of release (YOR) and this gain is a result of increased biomass production for a given amount of light intercepted. In contrast, some studies showed inconsistent results with biomass from the ones just presented (Morrison et al., 2000; Jin et al., 2010). The contradicting results might be a result of the temporal relationship between biomass accumulation and yield (Kumudini et al., 2001), the differences of experiments' conditions and the cultivars chosen. Considering that harvest index in modern soybean cultivars is approaching theoretical maximum (Zhu et al., 2010; Koester et al., 2014), improving the temporal dynamics of biomass accumulation have the potential to contribute to future yield gains. Thus, canopy coverage and biomass are promising traits to strategic improve yield gain in soybean.

1.5 Hypotheses and Objectives

We hypothesize that high-throughput phenotyping can improve different aspects of soybean breeding. The specific hypothesis and respective objectives are:

1. Average canopy coverage acquired with UAS can aid selection of high yielding lines in early stage of soybean breeding. Our objective was to compare the yield performance of soybean lines selected from progeny rows based on a typical pipeline with yield vs those selected using additional average canopy coverage data.
2. Biomass accumulation can offer valuable insights into plant development and responses to the environment and can be a potential target to increase yield gain. As a dynamic trait changing over time, the phenotypic variation is associate with different regions on the soybean genome. In addition, longitudinal genomic prediction can be implemented to

accurately select biomass in soybean. Our objectives were to (a) to estimate soybean biomass throughout late vegetative and mid reproductive growth stages using HTPP; (b) reveal the genetic architecture of the dynamics of biomass accumulation in soybean and (c) assess the ability of genomic prediction of breeding values for this longitudinal trait.

CHAPTER 2. IMPROVING THE EFFICIENCY OF SOYBEAN BREEDING WITH HIGH-THROUGHPUT CANOPY PHENOTYPING

A version of this chapter was previously published by *Plant Methods*: Moreira, F.F., Hearst, A.A., Cherkauer, K.A. *et al.* Improving the efficiency of soybean breeding with high-throughput canopy phenotyping. *Plant Methods* 15, 139 (2019). <https://doi.org/10.1186/s13007-019-0519-4>

Fabiana Freitas Moreira¹, Anthony Ahau Hearst², Keith Aric Cherkauer², Katy Martin Rainey¹ *

¹Department of Agronomy, Purdue University, 915 West State Street, West Lafayette, Indiana, 47907.

²Department of Agricultural and Biological Engineering, Purdue University, 225 South University Street, West Lafayette, Indiana, 47907.

*Corresponding author: krainey@purdue.edu

Authors` contributions

Conceived and designed the study: FM, KR, KC; Developed the experiment, collected field data and conducted statistical analysis: FM; Collected the imagery and conducted the images analysis: AH, KC; Interpreted results: FM, KR. Authors read and approved the manuscript.

2.1 Abstract

Background: In the early stages of plant breeding programs high-quality phenotypes are still a constraint to improve genetic gain. New field-based high-throughput phenotyping (HTP) platforms have the capacity to rapidly assess thousands of plots in a field with high spatial and temporal resolution, with the potential to measure secondary traits correlated to yield throughout the growing season. These secondary traits may be key to select more time and most efficiently soybean lines with high yield potential. Soybean average canopy coverage (ACC), measured by unmanned aerial systems (UAS), is highly heritable, with a high genetic correlation with yield. The objective of this study was to compare the direct selection for yield with indirect selection using ACC and using ACC as a covariate in the yield prediction model (Yield|ACC) in early stages of soybean breeding. In 2015 and 2016 we grew progeny rows (PR) and collected yield and days to maturity (R8) in a typical way and canopy coverage using a UAS carrying an RGB camera. The best soybean lines were then selected with three parameters, Yield, ACC and Yield|ACC, and advanced to preliminary yield trials (PYT).

Results: We found that for the PYT in 2016, after adjusting yield for R8, there was no significant difference among the mean performances of the lines selected based on ACC and Yield. In the PYT in 2017 we found that the highest yield mean was from the lines directly selected for yield, but it may be due to environmental constraints in the canopy growth. Our results indicated that PR selection using Yield|ACC selected the most top-ranking lines in advanced yield trials.

Conclusions: Our findings emphasize the value of aerial HTP platforms for early stages of plant breeding. Though ACC selection did not result in the best performance lines in the second year of selections, our results indicate that ACC has a role in the effective selection of high-yielding soybean lines.

2.2 Background

Breeders are challenged to increase the rate of genetic gain. Genetic gain in a crop breeding program can be defined as $\Delta G = h^2 i \sigma_p / L$, where h^2 is the narrow-sense heritability, i is the selection intensity, σ_p is the phenotypic standard deviation and L is the breeding cycle time or generation (Falconer and Mackay, 1996). This equation translates theoretical quantitative genetics into parameters that breeders can manipulate in their breeding pipelines (Cobb et al., 2019). In this context genetic gain can be increased in a number of ways, including: increasing population size to increase selection intensity, shortening the breeding cycle, ensuring suitable genetic variation in the population, and obtaining accurate estimates of the genetic values (Moose and Mumm, 2008; Xu et al., 2017; Araus et al., 2018). Phenotyping directly or indirectly influences these parameters which emphasize the need for accurate, precise, relevant and cost-effective phenotypic data (Cobb et al., 2013).

Plant phenotyping has recently integrated new technology from the areas of computer science, robotics, and remote sensing, resulting in high-throughput phenotyping (HTP) (Cabrera-Bosquet et al., 2012; Cobb et al., 2013; Fiorani and Schurr, 2013; Li et al., 2014). Platforms have been developed based on high capacity for data recording and speed of data collection and processing in order to capture information on structure, physiology, development, and performance of large numbers of plants multiple times throughout the growing season (Cabrera-Bosquet et al., 2012; Tardieu et al., 2017). Compared with other platforms, imagery-based field HTP using unmanned aerial systems (UAS) has the advantage of high spatial and temporal resolution (Tattaris et al., 2016) and is non-destructive.

There are a number of applications of a trait that can be precisely phenotyped with an HTP platform in a breeding pipeline. Secondary traits may increase prediction accuracy in multivariate pedigree or genomic prediction models (Rutkoski et al., 2016; Crain et al., 2018; Jarquin et al., 2018). Alternately, traits measured with HTP can be used in selection indices or for indirect selection for yield (Prasad et al., 2007). Indirect selection may be preferable when the secondary trait is easier or less expensive to measure than yield and if it can be selected out-of-season or in earlier developmental stages or generations, accelerating decision-making steps, and consequently decreasing the breeding cycle (Richards, 2000; Bernardo, 2010).

In a typical soybean breeding program, after reaching desired homozygosity, a common procedure is to select individual plants and then grow the next generation in progeny rows (PR) trials (Orf et al., 2004). At this stage, there is usually a large number of entries but a small number of seeds, limiting the experiment to unreplicated one-row plots at one location (Sun et al., 2015). Due to these limitations, yield measurements in PR are inaccurate and may require a large investment of resources. In this scenario, HTP has the potential to remotely measure in a nondestructive manner traits correlated to yield in early stages of development, improving data quality and reducing time or cost, or, for selection (Montesinos-López et al., 2017; van Eeuwijk et al., 2018).

Several studies have demonstrated that attaining full canopy coverage, and thus maximum light interception (LI), during vegetative and early reproductive periods is responsible for yield increases in narrow-row culture due to enhanced early growth (Board et al., 1992; Shibles and Weber, 1996; Bullock et al., 1998). As management practices change over time, more recent studies using different plant populations found that rapid establishment of canopy coverage improves the interception of seasonal solar radiation, which is the foundation for crop growth and yield (Edwards and Purcell, 2005; Edwards et al., 2005). LI efficiency, measured as leaf area index (LAI), was significantly correlated to yield in a study comparing soybean cultivars released from 1923 to 2007 (Koester et al., 2014). In addition, the rapid development of canopy coverage can decrease soil evaporation (Boerma et al., 2004) and suppress weeds (Jannink et al., 2000, 2001; Fickett et al., 2013).

Purcell (2000) (Purcell, 2000) showed that soybean LI can be measured as a function of canopy coverage from images taken from above the plot using a digital camera. In addition, soybean canopy coverage can also be effectively extracted automatically from UAS-based digital

imagery (Xavier et al., 2017). Xavier *et al.*, (2017) (Xavier et al., 2017) observed that average canopy coverage (ACC) measured early season was highly heritable ($h^2 = 0.77$) and had a promising genetic correlation with yield (0.87), making it a valuable trait for indirect selection of yield. In the same study, they found a large effect quantitative trait locus (QTL) on soybean chromosome 19 that resulted in an estimated increase in grain yield of 47.30 kg ha^{-1} with no increase in days to maturity (-0.24 days). Candidate genes associated with growth, development, and light responses were found in genome-wide association analysis of imagery-based canopy coverage during vegetative development (Kaler et al., 2018). Jarquin *et al.* (2018) found that early season canopy coverage, used to calibrate genomic prediction models, improved the predictive ability for yield, suggesting that it is a valuable trait to assist selection of high yield potential lines. Thus, early season canopy coverage has the potential to be used as a secondary trait for indirect selection for yield or as covariables to improve yield estimations in quantitative genetic models (van Eeuwijk et al., 2018).

While several studies have shown the value of UAS to phenotype various traits for a number of crops (Pölönen et al., 2013; Bendig et al., 2014; Husson et al., 2014; Yang et al., 2014; De Souza et al., 2017; Madec et al., 2017), to our knowledge there is no study showing the use of UAS-derived phenotypes for applied breeding purposes. In addition, no empirical studies have reported on the efficacy of using canopy coverage phenotypes in a soybean breeding pipeline. Selection experiments are useful for comparing breeding methods by enabling the assessment of realized gains of different selection categories to identify the most effective method. Our aim was to perform a selection experiment to compare the yield performance of soybean lines selected from PR based on yield with those selected based on ACC from imagery acquired with UAS.

2.3 Methods

2.3.1 Description of Breeding Populations

This study used 2015 and 2016 $F_{4:5}$ progeny rows (PR) populations from the soybean breeding program at Purdue University. These trials were grown under a modified augmented design with replicated checks at the Purdue University Agronomy Center for Research and Education (ACRE) ($40^{\circ}28'20.5''\text{N } 86^{\circ}59'32.3''\text{W}$). Experimental units consisted of a one-row plot of size 1.83 m with 0.76 m row spacing and were planted on May 25, 2015, and May 24, 2016

(orientated South-North). In the 2015 PR experiment, we had 3311 plots with 2747 progenies and in 2016 PR we had 4220 plots with 4052 progenies. There was no overlap among the experimental lines in 2015 and 2016.

For both years, we advanced selected lines in early- and late- maturing preliminary yield trials (PYT early and PYT late) comprised of lines classified as earlier or later than the check IA3023. The lines selected from 2015 PR were advanced as 2016 PYT early and PYT late and the lines selected from 2016 PR were advanced as 2017 PYT early and PYT late. The PYTs were grown in two locations and with two replications using alpha-lattice designs. The experimental unit consisted of two rows plot of 2.9 m in length in 2016 and 3.4 m in length in 2017, with 0.76 m of row spacing. For both years, one of the locations was ACRE and the second location in 2016 was at the Throckmorton-Purdue Agricultural Center (TPAC) (40°17'49.1"N 86°54'12.8"W) and in 2017 was at Ag Alumni Seed (40°15'41.3"N 86°53'19.1"W), both in Romney, IN.

Lines selected from 2016 PYT and 2017 PYT were evaluated in an advanced yield trial (AYT) in 2017 and 2018, respectively. Both trials were grown in an alpha-lattice design in two locations with either three or four replications per location. The locations were the same as described for PYT 2017. AYT plots consisted of four rows of 3.4 m length and 0.76 m spacing among rows. AYT lines were classified as early and late in the same manner as PYT.

2.3.2 Phenotypic Data

For all trials, grain yield and days to maturity (R8) were collected for every plot. Grain yield (g/plot) was converted to $\text{kg}\cdot\text{ha}^{-1}$ using harvest-timed seed moisture to adjust all plot values to 13% seed moisture. R8 was expressed as days after planting when 50% of the plants in a plot had 95% of their pods mature (Fehr and Caviness, 1977).

For PR 2015 and 2016 we quantified canopy coverage from aerial images collected using a fixed-wing Precision Hawk Lancaster Mark-III UAS equipped with a 14-megapixel RGB Nikon 1-J3 digital camera. Flights were performed at an altitude of 50 m, which resulted in a spatial resolution of 1.5 cm per pixel. We used eight sampling dates of early-season canopy development, ranging from 15 to 54 DAP (15, 29, 34, 37, 44, 47, 51, 54 DAP) in 2015 PR, and seven sampling dates, ranging from 20 to 56 DAP (20, 27, 31, 37, 42, 52, 56 DAP) in 2016 PR. The trials were maintained free of weeds to ensure that the images captured only soybean canopy. Image analysis, plot extraction, and classification were performed using a multilayer mosaic methodology

described by Hearst (2019). This methodology allows for the extraction of the plots from ortho-rectified RGB images using map coordinates, resulting in several plot images of different perspectives from the same sampling date due to overlapping frame photos. The number of plot images from the same date varies from plot to plot. Image segmentation was done using Excess Green Index (ExG) and Otsu thresholding (Hearst, 2019) to separated canopy vegetation from the background. Canopy coverage was calculated as the percentage of image pixels classified as canopy pixels. Median of canopy coverage values from replicated plot images was calculated for each sampling date. For each plot, average canopy coverage (ACC) was obtained by averaging the median canopy coverage among sampling dates. Figure 2.1 summarizes the process from image acquisition to the calculation of ACC.

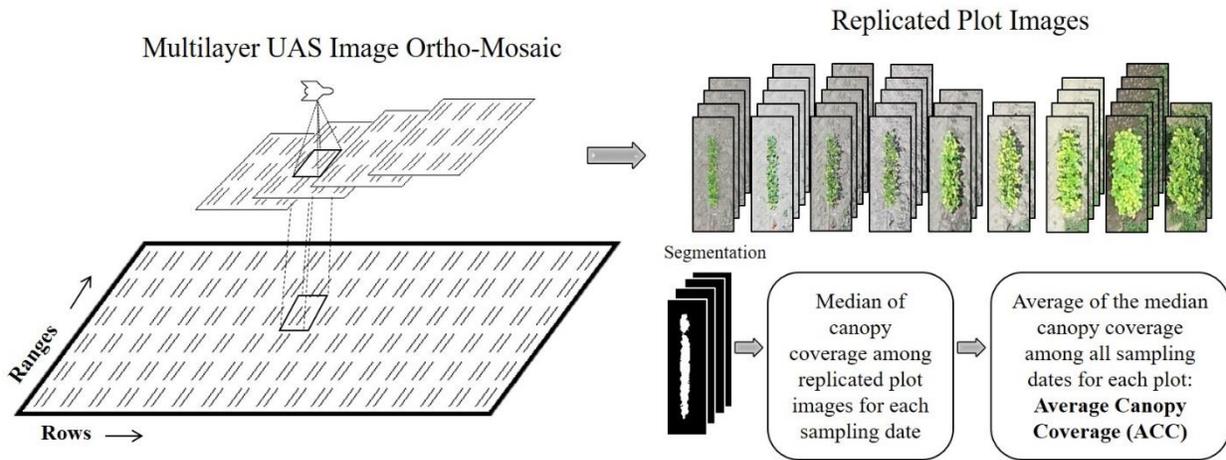


Figure 2.1 Overview of data collection and processing to acquire average canopy coverage (ACC).

2.3.3 Statistical Data Analysis and Selection Methods of PR

PR 2015 and 2016 yield, R8, and ACC phenotypes were fitted in a pedigree-based mixed model to estimate variance components and breeding values, using Gibbs sampling implemented in the R package NAM (Xavier et al., 2015), described as:

$$y_i = \mu + g_i + e_i \quad (\text{Equation 1})$$

where y_i is the phenotype, μ is the mean, g_i ($i = 1, \dots$, number of genotypes) is the random genotype effect with $g_i \sim N(0, A\sigma_a^2)$ where A is the relationship matrix calculated using pedigrees

that traced back to PR founders and σ_a^2 is the additive genetic variance, e_i is the residual term with $e_i \sim N(0, \mathbf{R}\sigma_e^2)$ where \mathbf{R} is a field correlation matrix considered to account for spatial variation in the field calculated as the average phenotypic value of neighbor plots (Lado et al., 2013) and σ_e^2 is the residual variance. For yield, an additional model was fitted in order to adjust for ACC (Yield|ACC), where the fixed ACC effect (aka covariate), β_i ($i = 1, \dots$, number of genotypes), was added to the previous model. Yield|ACC is considered a different trait than yield. The solutions for g_i for each trait here are defined as best linear unbiased predictors (BLUP). To estimate phenotypic correlations, we calculated Pearson's correlations among BLUPs for the different traits. Narrow-sense heritability (h^2) was calculated using the formula:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \quad (\text{Equation 2})$$

where σ_a^2 and σ_e^2 are described previously.

For the selection experiment, the selection categories or traits used in this study were yield BLUPs, as the traditional selection method, ACC BLUPs, and Yield|ACC BLUPs. Lines were selected based on BLUPs rankings within each selection category. For PR 2015 we selected approximately 9% of progenies for each selection category. Since some lines were selected by more than one selection category, the total lines selected was 523. In 2016, since we had more progeny lines, we decreased the selection to 7.5%. Due to the overlap of lines selected among the selection categories, we selected 705 lines. There was some deviation from the intended selection intensities due to seed limitations, field space, or logistics in the breeding pipeline. Figure 2.2 shows the summary of lines selected by each selection category for PR 2015 and 2016. As described above, selected lines were divided into early and late PYT.

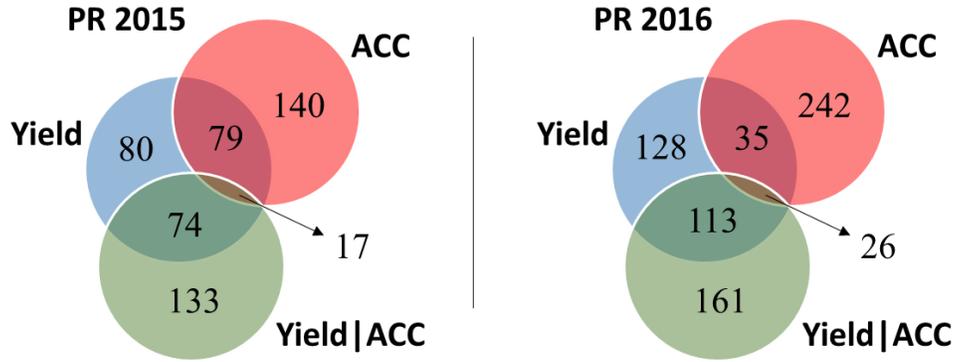


Figure 2.2 Number of lines selected from progeny rows (PR) 2015 and 2016 by each selection criteria.

2.3.4 Evaluation of PYT and AYT

To evaluate PYT line performance, yield and R8 phenotypes across locations were fitted using restricted maximum likelihood (REML) approach, implemented in the R package lme4 (Bates et al., 2015) :

$$y_{ijkl} = \mu + g_i + loc_j + r_{k(j)} + b_{l(k(j))} + (g * loc)_{ij} + e_{ijkl} \quad (\text{Equation 3})$$

where y_{ijkl} is the phenotype, μ is the mean, g_i ($i = 1, \dots$, number of genotypes) is the random genotype effect with $g_i \sim N(0, \sigma_g^2)$ where σ_g^2 is the genetic variance, loc_j ($j = 1, \dots$, number of environments) is the random location effect with $loc_j \sim N(0, \sigma_{loc}^2)$ where σ_{loc}^2 is the location variance, $r_{k(j)}$ is the random effect of k^{th} replication nested within j^{th} location with $r_{k(j)} \sim N(0, \sigma_r^2)$ where σ_r^2 is the replication within location variance, $b_{l(k(j))}$ is the random effect of the l^{th} incomplete block nested within the k^{th} replication and j^{th} location with $b_{l(k(j))} \sim N(0, \sigma_b^2)$ where σ_b^2 is the block variance, $(g * env)_{ij}$ is the random genotype by location interaction effect with $(g * loc)_{ij} \sim N(0, \sigma_{g*loc}^2)$ where σ_{g*loc}^2 is the genotype by location variance, and e_{ijkl} is the residual term with $e_{ijkl} \sim N(0, \sigma_e^2)$ where σ_e^2 is the residual variance. Adjusted values for yield and R8 were calculated as $\mu + g_i$, to express the phenotypes with units. Maturity is a confounding factor that influences yield, which may lead to misinterpretation of the yield potential of a line; therefore, we also calculated yield adjusted to R8 including R8 as a covariate in Equation 3.

In a breeding program, the method that increases the population mean the most from one generation to the next is the preferred method; therefore, to evaluate the performance of the lines

in the selected classes we performed two-sample *t*-tests to compare the adjusted yield means of lines in each selected class. The best selection category is the one producing the highest yield mean within an early or late trial, considering that all lines came from the same original populations. Although AYT was not part of the selection experiment, we wanted to evaluate how the top-ranked lines were selected. Lines were selected from PYT using rankings of yield BLUPs and advanced to AYT. For AYT data summary Equation 3 was used with the change of genotype to fixed effect. AYT lines were classified as early and late from R8 phenotypes.

2.4 Results

2.4.1 PR

Table 2.1 shows the estimated narrow-sense heritability and phenotypic Pearson's correlations for yield, ACC, Yield|ACC, and R8 for 2015 and 2016 PR. Positive correlations were observed among all traits with Yield, with the highest observed with Yield|ACC. ACC showed low (0.01) or negative (-0.1) correlation with R8 and negative correlation with Yield|ACC in both years. R8 and Yield|ACC were positively correlated. Narrow-sense heritability for Yield|ACC and R8 was higher than for Yield in both years. Narrow-sense heritabilities were low for ACC and Yield, but the heritability of ACC was higher than yield in 2017.

Table 2.1 Pearson's correlations for PR 2015 (above diagonal) and 2016 (below diagonal) and narrow-sense heritability

<i>r</i>	Yield	ACC	Yield ACC	R8
Yield	-	0.51	0.70	0.61
ACC	0.06	-	-0.14	0.01
Yield ACC	0.75	-0.20	-	0.69
R8	0.30	-0.10	0.20	-
<i>h</i> ²				
PR 2015	0.23	0.06	0.35	0.36
PR 2016	0.11	0.18	0.48	0.17

Yield (Kg/ha), average canopy coverage (ACC), yield given ACC (Yield|ACC) and R8 (days to maturity), progeny rows (PR). *r*: Person's correlation, *h*²: narrow-sense heritability.

2.4.2 PYT Selection Category Performance

The box plots presented in Figure 2.3 A show the distributions of adjusted yield values for lines in each selected class and adjusted R8 means are summarized in Appendix A (Table A.1). For PYT early 2016 the yield mean was not significantly different among the lines from different selected classes. For PYT late 2016 the lines selected by Yield had a statistically significantly higher mean yield, and there were no statistically significant differences in mean yield among the lines selected by ACC and Yield|ACC. The mean yield of the lines selected by ACC and Yield was not statistically significantly different in PYT late 2016 when considering yield adjusted by R8 (Fig. 2.3 B). For PYT early and late in 2017, the mean yield among lines from different selected classes was statistically significantly different, and the lines selected by Yield had a higher mean yield.

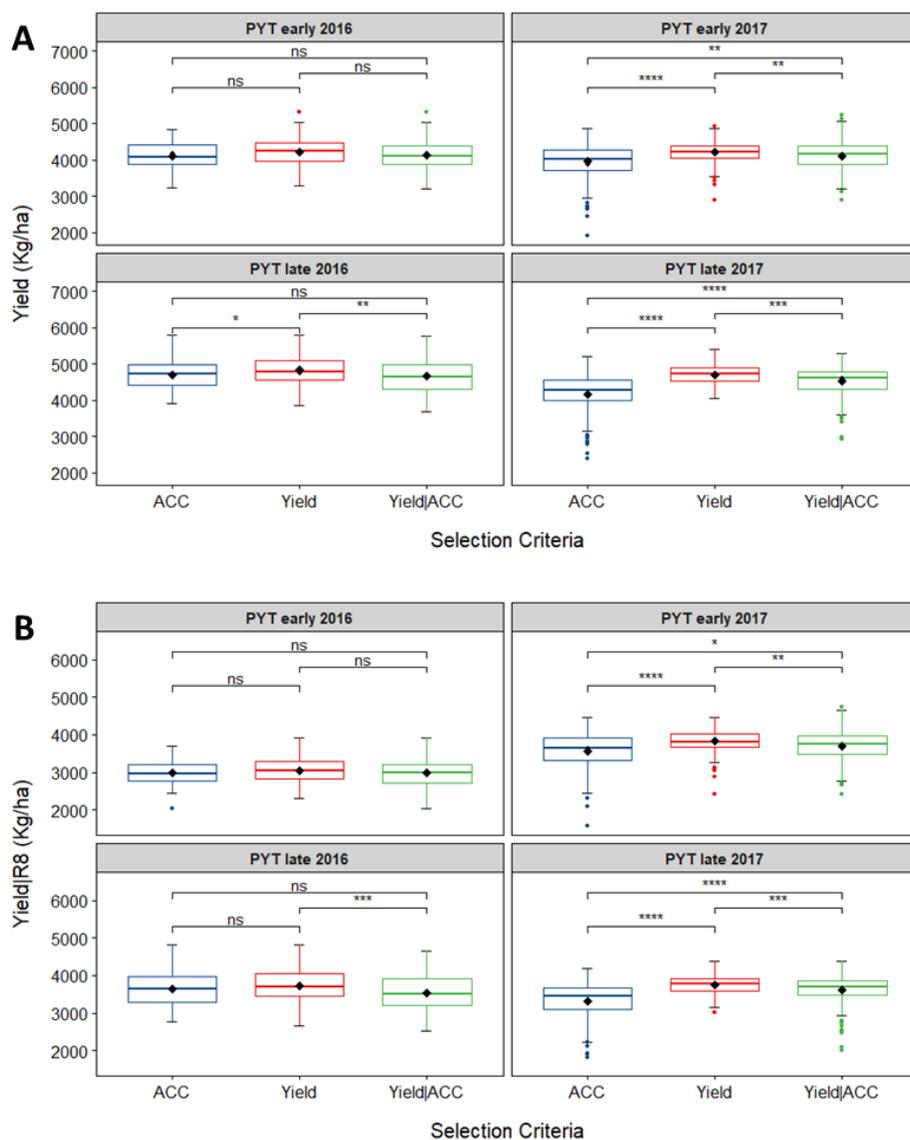


Figure 2.3 A. Box plot of yield (Kg/ha) and B. Yield adjusted by R8 (Yield|R8) distribution for all lines selected by each selection criteria (Yield, ACC and Yield|ACC) for preliminary yield trials (PYT) early and late in 2016 and 2017. Diamond indicates mean for each selection criteria. The line crossing the box plots are representing the median for each group. No significant (ns): $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$.

2.4.3 AYT Yield Performance

Table 2 summarizes the ten top-ranked lines in AYT 2017 and 2018. In both years, the lines were mostly selected by two selection categories. None of the ten top-ranked lines in the AYT early 2017 were selected by Yield alone in the PR stage. In the AYT late 2017 only one line

was selected by Yield alone in the PR stage, in rank position ten. In AYT 2018 early and late the Yield selection category alone selected just three and two of the ten top-ranked lines, respectively. Considering both years, the number of top-ranked lines selected using only ACC and/or Yield|ACC was greater (14 lines) than the lines selected by Yield alone (6 lines).

Table 2.2 Progeny row selection categories choosing the ten top-ranked lines for advanced yield trials (AYT).

AYT early 2017	AYT late 2017	Rank	AYT early 2018	AYT late 2018
Yield, Yield ACC	Yield ACC	1	Yield	Yield, Yield ACC
Yield, Yield ACC	Yield ACC	2	Yield, Yield ACC	Yield ACC
ACC, Yield	ACC, Yield ACC	3	Yield	Yield, Yield ACC
Yield, Yield ACC	ACC	4	Yield, Yield ACC	Yield ACC
ACC, Yield	Yield ACC	5	Yield	Yield
Yield ACC	ACC, Yield, Yield ACC	6	ACC, Yield	Yield, Yield ACC
Yield, Yield ACC	ACC, Yield, Yield ACC	7	ACC	Yield, Yield ACC
Yield, Yield ACC	Yield ACC	8	Yield, Yield ACC	Yield
ACC, Yield, Yield ACC	ACC, Yield ACC	9	Yield, Yield ACC	Yield ACC
Yield, Yield ACC	Yield	10	Yield ACC	Yield ACC

Average canopy coverage (ACC), yield given ACC (Yield|ACC).

2.5 Discussion

The positive phenotypic correlation found in this study among yield and ACC in PR 2015 (Table 1) is in agreement with other studies (Xavier et al., 2017; Jarquin et al., 2018; Kaler et al., 2018); however, this result was not repeated in PR 2016. Phenotypic correlation depends on genetic and environmental correlations, thus even when no phenotypic correlation can be estimated the traits may still be correlated genetically and environmentally (Falconer and Mackay, 1996). Considering that some studies showed a strong positive genetic correlation between ACC and yield, the lack of phenotypic correlation in PR 2016 may be the reflection of the genetic and environmental correlations acting in opposite directions between the two traits, as well as the interaction between genotype and environment (Falconer and Mackay, 1996; Hall, 2015; Xavier et al., 2017; Sodini et al., 2018).

We observed none to negative phenotypic correlations between ACC and R8 in PR 2015 and PR 2016, respectively, indicating that selection on ACC should not lead to indirect increases

in maturity. In both years, ACC and Yield|ACC were negatively correlated, which is expected since adjusting yield for ACC will correct the yield data to a baseline value of ACC, thus, simplistically, yield decreases for higher ACC and increases for lower ACC.

For PR 2015 and 2016 ACC heritabilities (Table 1) were lower when compared with other studies (Hall, 2015; Xavier et al., 2017), but these studies used multiple environments of replicated data, and we observed comparatively lower yield and R8 heritabilities as well. Generally, low heritabilities in PR trials are expected given unreplicated single row plot trials leading to challenges in the estimation of the genetic parameters of the tested lines.

It is generally accepted that maturity confounds yield estimates in soybeans and later maturing cultivars will generally out-yield earlier maturing cultivars. In soybean breeding, yield phenotypes are sometimes corrected for R8 to better estimate yield potential *per se* and avoid indirect selection for later maturity. In our study, PYT early 2016 was the best scenario to compare the selection categories due to the lack of statistically significant differences in R8 among the selected classes (Fig. A.1). For this trial, the mean yield among the selection categories was not significantly different (Fig. 2.3), indicating that indirect selection for yield based on ACC or using Yield|ACC would result in the same yield gain than direct selection on yield, considering that they derived from the same base population. Using ACC as a selection criterion in early stages of soybean breeding pipelines would provide advantages not only in the reduction of the time for selection but also in the cost associated with the trait measurement.

For the other three trials, PYT late 2016 and PYT 2017, there were differences in the mean R8 between at least among two of the selection categories (Fig. A.1). Therefore, differences in the mean yield among the selection categories may be associated with the differences in days to maturity. The yield correction for R8 changed the comparison among the selection categories Yield and ACC in PYT 2016 late, making them similarly efficient for selection (Fig. 2.3). Although ACC selection did not produce higher gains than Yield selection, both PYT in 2016 confirm findings from Xavier *et al.* (2017) (Xavier et al., 2017) that assuming identical selection intensities indirect selection for yield using ACC would have a relative efficiency for selection comparable to yield direct selection. In general, the findings from PYT 2016 did not hold in 2017 trials (Fig. 2.3). Even after adjusting for R8 the lines selected by Yield had a higher performance than the lines selected by the other selection categories; however, the differences among the yield mean from lines selected by Yield and Yield|ACC was small for both early (~120 kg/ha) and late

(~150 kg/ha) trials (Additional file 1, Supplementary Table 1), which may indicate that Yield|ACC is a valuable trait for selection.

This contrasting results in trait selection efficacy observed in 2016 and 2017 may be explained by differences in canopy coverage development in PR 2015 and PR 2016, as showed in the comparison of canopy coverage development over time of the common checks among years (Fig. A.2). In 2015 at around 53 days after planting (DAP) we observed an average of canopy coverage of 35% in the checks, while at the same DAP in 2016 the checks had an average of almost 80% canopy coverage. This abnormal growth in 2016 produced tall plants and increased lodging (data not shown), which has a great effect in unreplicated single row plot trials where every genotype is competing with both neighbor rows. Considering that taller and bigger plants do not result in higher yields when ranking the top BLUPs, several lines that were selected based on ACC may have had poor yield potential. In addition, the lack of correlation of yield and ACC in PR 2016 may have been a result of this unusual canopy growth. Therefore, despite the evidence that one trait can be used to indirect select for yield, the breeder needs to consider the environmental influence on the trait phenotypes at the time of selection. In our case, we could have used a threshold for ACC before doing the selections, avoiding the very high values of canopy coverage, or restricted selection dates to earlier points in development.

If we consider the top 40 lines from AYT in 2017 and 2018, direct selection for yield alone selected only 6 lines from the PR trials, compared to 14 lines selected using ACC and/or Yield|ACC. Thus, despite the difference in mean performance among the selection categories in the PYT stage, we have demonstrated that ACC alone or combined with yield (Yield|ACC) are valuable secondary traits for selection in the PR stage. Yield|ACC had the best selection result in the top 10 lines for the AYT. Poor yield measurements due to harvesting errors, weather, and plot damage, lead to inaccurate representations of yield potential. Adjusting yield for early season ACC compensates for these inadequacies and is a better predictor of the real yield potential. This is in agreement with Jarquin *et al.* (2018) (Jarquin et al., 2018) results showing that early season canopy coverage increased the predictive accuracy of yield in genomic predictions models. Additionally, digital canopy coverage has a one to one relationship to LI, which in turn is an important factor for yield potential equation (Monteith and Moss, 1977; Purcell, 2000; Xavier et al., 2017). Therefore, up to a certain point, increases in LI, through ACC, will result in increases in yield when the other parameters in the yield equation are kept the same.

In this study, we have shown that the efficiency of selecting high yielding soybean lines can be improved by taking advantage of an HTP trait. Field-based HTP using UAS is robust, simple, and cost-effective and can measure a wide range of phenotypes that can be converted into useful secondary traits (Cobb et al., 2019; Zhao et al., 2019). Breeding teams need to evaluate carefully the value of these secondary traits in increasing genetic gain either in a phenotypic selection or as part of pedigree or genomic prediction schemes (Rutkoski et al., 2016; Cobb et al., 2019). In addition, we recommend testing different scenarios to ensure if the greater response is using the secondary trait alone or in combination with yield. However, if not in the literature, an investigation of heritability and genetic correlation to yield should be carried out to evaluate the potential of the trait.

2.6 Conclusions

One of the most important tasks of a plant breeder is to find among the available selection criteria a combination that can promote the desirable genetic gain for the traits of interest within their breeding program. Field HTP must be integrated into a wider context in breeding programs than trait estimation, evaluation of platforms, and genetic association studies. We examined three different ways to select soybean lines from PR trials: Yield, ACC and Yield|ACC. We compared their performance in advancing selected lines in the following generations common in a soybean breeding program. We have demonstrated that the secondary trait ACC measured using an aerial HTP platform can be used for selection, alone or in combination with yield, in early stages of soybean breeding pipelines. This method may offer even more advantages when yield is low quality or can't be phenotyped due to the high cost or extreme weather events. Further studies are needed to assess environmental effects on canopy coverage phenotypic variation in order to have optimized recommendations on the use of ACC for selecting high yielding lines in different scenarios.

CHAPTER 3. INTEGRATING HIGH-THROUGHPUT PHENOTYPING AND STATISTICAL GENOMIC METHODS TO GENETICALLY IMPROVE LONGITUDINAL TRAITS IN CROPS

A version of this chapter has been accepted for publication by *Frontiers in Plant Science*.

Fabiana F. Moreira¹, Hinayah R. Oliveira², Jeffrey J. Volenec¹, Katy M. Rainey¹, Luiz F. Brito^{2*}

¹Department of Agronomy, Purdue University, West Lafayette, Indiana, 47907, USA

²Department of Animal Sciences, Purdue University, West Lafayette, Indiana, 47907, USA

* Corresponding author: britol@purdue.edu

Authors` Contributions

FM, KR, and LB conceptualized the manuscript. FM wrote the manuscript. HO, LB, and KR critically revised and improved the manuscript. JV edited section 3 “Modeling Longitudinal Traits”. All authors read and approved the manuscript.

3.1 Abstract

The rapid development of remote sensing in agronomic research allows the dynamic nature of longitudinal traits to be adequately described, which may enhance the genetic improvement of crop efficiency. For traits such as light interception, biomass accumulation, and responses to stressors, the data generated by the various high-throughput phenotyping methods requires adequate statistical techniques to evaluate phenotypic records throughout time. As a consequence, information about plant functioning and activation of genes, as well as the interaction of gene networks at different stages of plant development and in response to environmental stimulus can be exploited. In this review, we outline the current analytical approaches in quantitative genetics that are applied to longitudinal traits in crops throughout development, describe the advantages and pitfalls of each approach, and indicate future research directions and opportunities.

3.2 Introduction

Enhancing agricultural production efficiency by reducing yield gaps while also breeding more stress-resilient cultivars is the next challenge for plant breeders (Godfray et al., 2010; Foley et al., 2011; Ray et al., 2013; Challinor et al., 2014; Tai et al., 2014). The most feasible solutions are developing innovative approaches to speed up the genetic improvement of economically

important traits and characterizing novel traits and incorporating them into breeding programs (Duvick, 2005; Lange and Federizzi, 2009; Fischer and Edmeades, 2010; Nolan and Santos, 2012; Rogers et al., 2015).

Plant breeding was established as a science in the beginning of the 20th century, when new insights about the genetic basis of phenotypic variation for quantitative traits were integrated with the foundational theories of inheritance mechanisms and crop hybridization elucidated by Mendel and Darwin, respectively (Johannsen, 1909, 1911; East, 1911; Bradshaw, 2017). Since then, plant breeders have improved crop productivity by selecting for numerous traits. Breeding objectives are constantly refined to address new challenges, including adaptation to new production areas, addressing emerging pests and diseases, various end uses, advanced farming technologies, and climate change (Toenniessen, 2002; Baenziger et al., 2006; Tester and Langridge, 2010; Gilliam et al., 2017). For instance, in 1955 the focus of soybean breeding was to increase seed oil content, canopy ground cover, and ripening uniformity. Ten years later the focus shifted to reducing pod shattering and lodging, and then it changed again over the years to include quality and value-added traits (Baenziger et al., 2006). Similar shifts in breeding goals and phenotyping technologies have also occurred in animal breeding (Henryon et al., 2014; Miglior et al., 2017). However, throughout history, improving traits of interest depends on the ability to quantify phenotypes across genotypes replicated over multiple environments (Stoskopf et al., 1994). Therefore, potentially valuable traits may have been neglected due to costly phenotyping and technological limitations.

Plant phenotyping has always been paramount for genetic improvement. Recent advances in proximal remote sensing, paired with new sensors and computer science applications, has enabled cost-effective high-throughput phenotyping (**HTP**) and dissection of novel traits (Montes et al., 2007; Furbank and Tester, 2011; Fiorani and Schurr, 2013; Araus and Cairns, 2014; Coppens et al., 2017). High-throughput phenotyping provides time-series measurements that track the development of a crop through its life stages and as it responds to the environment. Information on gene function, the activation of genes, interaction of genes networks at different stages of plant development and in response to environmental stimulus can now be exploited (Wu and Lin, 2006; Montes et al., 2007). It is increasingly possible for plant breeders to consider light interception, biomass accumulation, and response to drought stress as dynamic traits, rather than static points in time (Montes et al., 2007). This analytical framework enhances our understanding of crop

development and bridges gaps in understanding the relationship between genotype and phenotype (Granier and Vile, 2014; Araus et al., 2018).

Traits that are expressed repeatedly or continuously over the lifetime of an individual can be defined as longitudinal traits (Yang et al., 2006; Oliveira et al., 2019a), infinite-dimensional traits (Kirkpatrick and Heckman, 1989), or function-valued traits (Promislow et al., 1996). The study of longitudinal traits can provide important insights into the genetic mechanisms underlying physiological responses to environmental stresses and developmental processes. This information can be used to improve predictive ability for complex polygenic traits under multivariate settings, and contribute to identifying overall (e.g., soybean yield) or time specific Quantitative Trait Loci (**QTL**) (Fahlgren et al., 2015; Campbell et al., 2017; Sun et al., 2017a). Such analysis enables assessment of the statistical association of genetic and environmental factors, such as the relationship between molecular markers and response to abiotic stress at different developmental stages (Langridge and Fleury, 2011). In this context, phenotypic data describes a function changing continuously in response to other variables (Stinchcombe and Kirkpatrick, 2012; Granier and Vile, 2014). These approaches generate a vast amount of data, which requires advanced statistical approaches to enable evaluation of the phenotypic data as a function of time. In this literature review, we will outline the current analytical approaches in quantitative genetics and genomics that can be applied to HTP quantified over time (Figure 3.1). In addition, we describe the advantages and pitfalls of each method and explore directions and opportunities for future research.

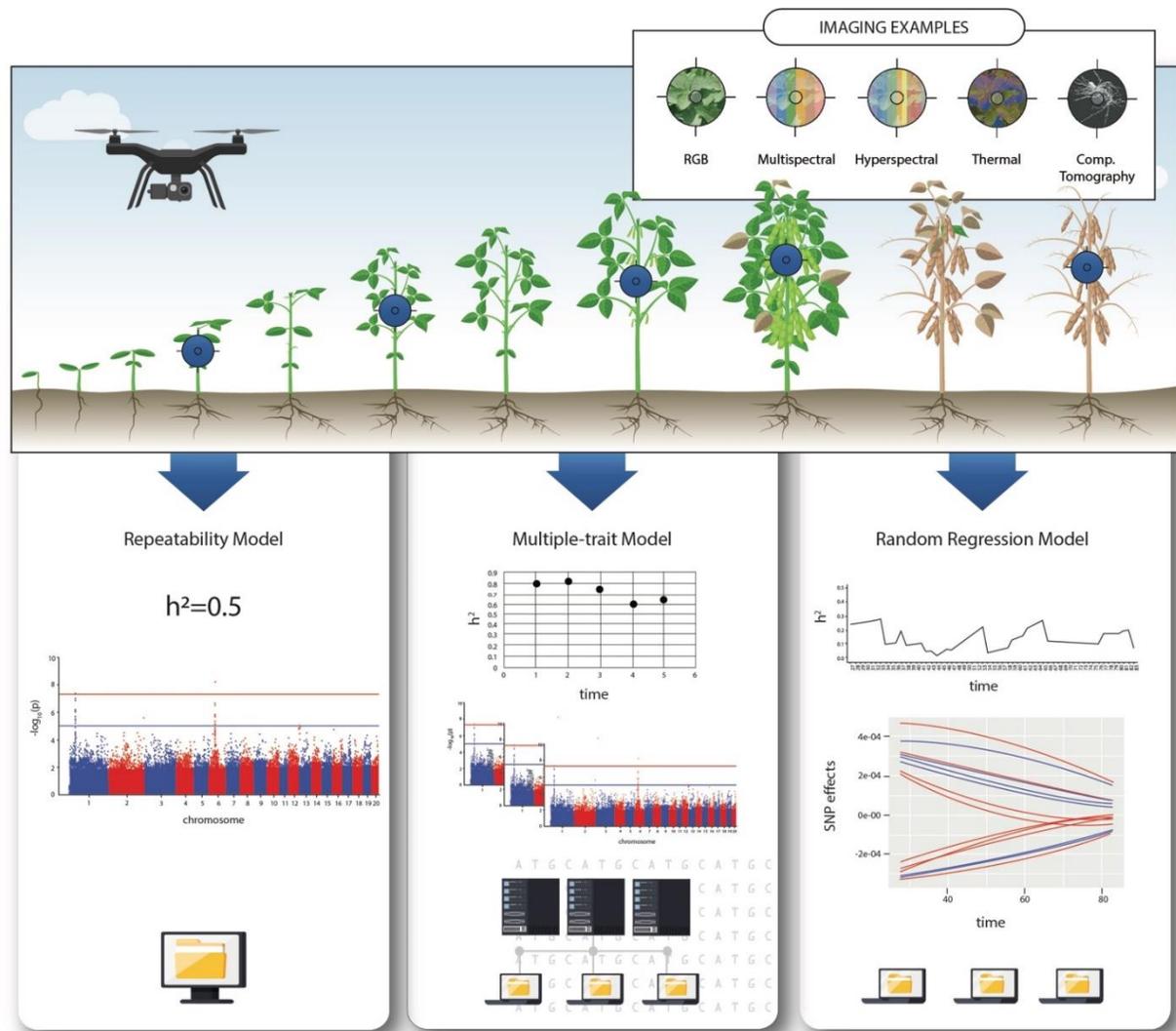


Figure 3.1 Schematic workflow of longitudinal data analyses. Different remote-sensing tools most commonly used for high-throughput phenotyping monitoring crop growth and development. Comparative overview of the potential models for genomic analysis, together with examples of outputs and computational demand.

3.3 Phenotyping Longitudinal Traits

Current HTP platforms, also referred to as “phenomic” tools, include a variety of methodologies that use remote sensing to obtain non-destructive phenotypic measurements, either in controlled environments or in the field (Pauli et al., 2016b). The most common types of sensors for crop phenotyping include red-green-blue (**RGB**; Xavier et al., 2017), multispectral (Xu et al., 2019), hyperspectral (Bodner et al., 2018), fluorescence (Pérez-Bueno et al., 2016), thermal (Sagan

et al., 2019), three-dimensional (**3D**; Topp et al., 2013), and laser-imaging detection and ranging (**LiDAR**) (Sun et al., 2018) devices. In general, these sensors rely on the interaction between electromagnetic radiation and plants (reflecting, absorbing, or transmitting photons), which is captured by the sensor as reflected radiation (Fiorani et al., 2012; Li et al., 2014). Thus, the sensors interpret the plants as optical objects, with each component of the plant displaying a characteristic spectral signature arising from wavelength-specific properties of absorbance, reflectance, and transmittance from the vegetation surface (Schowengerdt, 2012; Li et al., 2014).

The spectral signatures of plants change during the life cycle, giving rise to genotype-time-specific phenotypes. For instance, during senescence, there is an increase of reflectance in the red region caused by a loss of chlorophyll (Schowengerdt, 2012). For field-based phenotyping, these sensing tools are usually integrated into ground or aerial vehicles (Araus et al., 2018). Most HTP platforms have the spatial and temporal resolution needed to capture longitudinal traits. However, the needs and resources of the specific experiment should drive the choice of platform and sensor, as these choices directly impact the scale and type of research (Pauli et al., 2016b). Several reviews have focused on HTP and its nuances, such as data collection, data processing and types of sensors (Fahlgren et al., 2015; Rahaman et al., 2015; Singh et al., 2016; Tardieu et al., 2017; Yang et al., 2017; Zhao et al., 2019; Reynolds et al., 2020).

The genetic control of longitudinal traits captured with HTP was recently reported for different crops. In a greenhouse, Neilson et al. (2015) investigated the growth and dynamic phenotypic responses of sorghum to water-limited conditions and various levels of fertilizer over time. They defined and measured several traits, including leaf area, shoot biomass, height, tiller number, and leaf greenness using laser scanning, RGB, and near-infrared (**NIR**) cameras. In barley, multiple sensors captured daily images in a greenhouse over 58 days measuring the spectrum of visible light, fluorescence, and NIR in order to dissect phenotypic components of drought responses (Chen et al., 2014). Tiller growth in rice was examined using more than 700 traits extracted from an imaging system combining computed tomography (**CT**) and RGB imaging during the tillering process (Wu et al., 2019).

In the field, Sun et al. (2018) used LiDAR mounted on a tractor to quantify leaf area, stem height, and plant volume in cotton in the field from 43 to 109 days after planting. From these measurements, they generated genotype-specific growth curves and assessed variability in traits and their genetic correlation to yield over time. In soybean, an unmanned aerial vehicle (**UAV**, aka

drone) with an RGB camera measured canopy coverage in the field across several days after planting for genome-wide association analysis and genomic selection (Xavier et al., 2017; Hearst, 2019; Moreira et al., 2019). Blancon et al. (2019) characterized the dynamics of the green-leaf-area index in a diversity panel of maize under well-watered and water-deficient treatments using multispectral imagery acquired from a UAV throughout the growth cycle. Thermal and hyperspectral sensors mounted to a manned aircraft were used to extract canopy temperature and vegetation indexes of more than 500 lines of wheat in five field environments over a range of dates (Rutkoski et al., 2016).

Field-based HTP for roots has seen less progress than HTP of above-ground traits due to the difficulty of below-ground imaging (Atkinson et al., 2019). Shovelomics is a high-throughput method for root phenotyping and that has been used for crops, such as maize, common bean, cowpea, and wheat (Trachsel et al., 2011; BurrIDGE et al., 2016; York et al., 2018). It consists of extracting several traits from images. However, it is destructive and labor-intensive, requiring manual root excavation, which limits its ability to capture repeated records over time. Recently, geophysical techniques, including electrical resistance tomography, electromagnetic inductance, and ground-penetrating radar have contributed to identify and quantify roots in the field in a non-destructive manner (Shanahan et al., 2015; Whalley et al., 2017; Liu et al., 2018a; Atkinson et al., 2019). Nevertheless, HTP root phenotyping is more commonly performed in controlled environments that allow the use of alternative growth systems that enable root imaging, such as rhizotrons, growth pouches and transparent artificial growth media (Atkinson et al., 2019; Ma et al., 2019). To model the growth dynamics of maize, Hund et al. (2009) performed daily scans of root systems grown over blotting paper. Topp et al. (2013) used 3D imaging phenotyping of rice root architecture in a gellan gum medium on various days of growth to perform QTL detection analysis. Recent advances in the use of X-ray CT and magnetic resonance imaging in plant sciences have enabled monitoring of root system architecture and dynamic growth over time in soil (Metzner et al., 2015; Pfeifer et al., 2015; van Dusschoten et al., 2016; Pflugfelder et al., 2017; Gao et al., 2019). More details and additional techniques for root phenotyping are presented in Atkinson et al. (2019).

3.4 Modeling Longitudinal Traits

Plant growth and development are characterized by several phenotypic changes, which can only be studied by monitoring repeated phenotypes over time (Li and Sillanpää, 2015). High-throughput phenotyping platforms allow tracking of traits with a high temporal resolution, whether continuously over time or in discrete intervals (Furbank and Tester, 2011). Traditionally, mathematical functions are used to describe temporal trajectories of traits during the plant's life cycle (Paine et al., 2012). Analysis of longitudinal traits usually employs one of the following two techniques (Li and Sillanpää, 2013): 1) smooth functions (such as splines; e.g. Oliveira et al., 2019a, van Eeuwijk et al., 2018) or parametric functions (such as growth models; e.g. Paine et al., 2012) to fit the phenotypic records over time, providing interpolated values for all time points; or 2) the data is reparametrized by estimating the function's coefficients, which are then used in the analysis to represent the trait over time. Either way, it is necessary to select the function that best fits the shape of the trajectory of the trait to accurately estimate the curve parameters and results. Paine et al. (2012) provide a detailed review about growth models, highlighting basic functional forms, advantages and disadvantages. In this section, we will describe the main functions that have been successfully used to fit a variety of traits for the purposes of crop improvement.

Many of the complexities of plant growth are commonly represented using non-linear growth models that account for temporal variation in growth, capturing age and size-dependent growth (Paine et al., 2012). Usually, the growth pattern within a plant life cycle follows a sigmoid curve (S-shaped) characterized by an initial slow growth that then increases rapidly, approaching an exponential growth rate, and finally slows when it reaches a saturation phase (Yin et al., 2003). The S-shaped curve can be described by sigmoidal functions such as the logistic, Gompertz, Richards or β functions (Gompertz, 1815; Richards, 1959; Yin et al., 2003; Poorter et al., 2013). In this case, the Gompertz function is a special case of the Richard function; which is one of the oldest growth models frequently used to fit various biological processes across species (Tjørve and Tjørve, 2017). The Gompertz function has been used to describe biomass accumulation in maize kernels (Meade et al., 2013), barley biomass (Chen et al., 2014), and various longitudinal traits in wheat (Camargo et al., 2018) and sorghum (Neilson et al., 2015). The Logistic function is more commonly used in its asymptotic form to describe the time dependence of biological growth processes for traits such as biomass, canopy coverage, canopy size, volume, length, and area (Thornley et al., 2005; Paine et al., 2012). The Logistic function can have one, two, three, four, or

five-parameters (Tessmer et al., 2013). One- and two-parameter logistic models are simplistic and frequently do not fit the data well, but are still used in several studies (e.g. Paine et al., 2012; Tessmer et al., 2013). The three-parameter logistic function (**3PL**; also known as a Verhulst or autocatalytic growth function) is perhaps the most popular model for plant growth analysis. In a water-limitation experiment in sorghum, a 3PL model had the best performance of a variety of sigmoidal models to fit the projected leaf area (Neilson et al., 2015). Sun et al. (2018) used a 3PL model to fit growth curves for canopy height, projected canopy area, and plant volume obtained from HTP in cotton. In wheat, Baillot et al. (2018) estimated various grain-filling parameters by fitting a 3PL model. The four-parameter logistic (**4PL**) model is more flexible than the 3PL as it has fewer constraints (Pinheiro and Bates, 2000). Camargo et al. (2018) phenotyped the average of area, height, and senescence in wheat throughout its lifecycle and found that 4PL models best fit the longitudinal data. The five-parameter version (**5PL**) provides maximum flexibility and accommodates asymmetry (Gottschalk and Dunn, 2005), despite its higher complexity compared to the lower number of parameters.

Many biological curves cannot be described by sigmoidal functions. A Power Law (also known as allometry) function is a type of non-asymptotic, non-linear growth model that does not produce an S-shaped curve (Marquet et al., 2005). They are often used in ecology to predict relationships in plant communities (Chen and Shiyomi, 2019). It effectively captures temporal variation in growth as it allows the relative growth rate to slow down over time and with an increase in biomass (Paine et al., 2012). A Power Law function was used to fit projected leaf area data in sorghum receiving various levels of nitrogen (Neilson et al., 2015), as well as leaf length and rosette area in *Arabidopsis thaliana* (An et al., 2016).

Linear models, such as orthogonal polynomials and spline functions, are also used to fit longitudinal traits (Oliveira et al., 2019a). The use of polynomials in crop growth models started in the 1960s as a functional approach to fit growth data and provide a clear picture of ontogenetic drift (Vernon and Allison, 1963; Hughes and Freeman, 1967; Poorter, 1989). One of the advantages of these functions is that they do not require prior knowledge of the longitudinal shape of the phenotype. Therefore, they are useful for fitting biological data of any shape simply by choosing different orders of the polynomials. Although they are not linear in time, polynomial functions are linear in their parameters, and consequently, can take advantage of the inference methods available for linear models (Yang et al., 2006). For instance, cubic polynomial functions

have been used to describe grain growth in crops such as rice (Jones et al., 1979; Shi et al., 2015), wheat (Gebeyehou et al., 1982), barley (Leon and Geisler, 1994), and safflower (Koutroubas and Papakosta, 2010). One of the main difficulties with this approach is choosing the appropriate degree of polynomial to fit the data while avoiding spurious upward or downward trends or overfitting or underfitting the data (Paine et al., 2012).

Orthogonal polynomials are particularly popular for fitting biological curves because they have much lower correlations among their coefficients and provide estimates of the covariance matrices that tend to be more robust over a variety of data sets (Schaeffer, 2004). Legendre polynomials represent simple orthogonal polynomials and have been used successfully to fit longitudinal traits in livestock breeding programs (e.g. Albuquerque and Meyer, 2001; Oliveira et al., 2017; Brito et al., 2017; Oliveira et al., 2019b) and for plant research (e.g. Yang et al., 2006; Yang and Xu, 2007; Campbell et al., 2018; Momen et al., 2019).

Spline functions offer a more flexible alternative for modeling longitudinal traits compared to orthogonal polynomials (van Eeuwijk et al., 2018). Splines are piecewise polynomial functions, linked at specific points called knots (de Boor, 1980). For longitudinal data, these knots represent time points within the data collection interval (Li and Sillanpää, 2015). The greater flexibility of splines is due to the independence of each segment, which can have the same or different polynomial degrees, accommodating abrupt changes in the trajectory (Meyer, 2005b). A particular type of spline function is the basis spline, or B-spline (de Boor, 1980), extensively deployed in animal breeding (Meyer, 2005b; Oliveira et al., 2019a). Another version of spline is P-spline, which combines B-splines with different penalties on the coefficients of adjacent B-splines, resulting in smoother curves (Eilers and Marx, 1996; Meyer, 2005b).

Spline functions have recently been used to model longitudinal traits in crops. For instance, haulm senescence was assessed at several points during the growing season in a diploid potato mapping population and fitted using P-splines (Hurtado et al., 2012). Montesinos-López et al. (2017) used a B-spline function to fit wheat canopy hyperspectral bands in a yield prediction model. B-splines have also modeled rice temporal shoot biomass in a water-limited environment (Momen et al., 2019).

3.5 Statistical Genetic Models

Plant breeding is mostly based on the selection of new genetically superior cultivars from a large set of candidates. Simple arithmetic means of the phenotypic values, or Best Linear Unbiased Estimation (**BLUE**, treating genotypes as fixed effects), were used for selection prior to the development of Best Linear Unbiased Prediction (**BLUP**; Henderson, 1974). BLUPs are based on a mixed linear model and are now the most commonly used method for genetic evaluation of plant and livestock species (Piepho et al., 2008; Mrode, 2014). In the mixed-model framework, the genotypes are fitted as random and the genotypic effects are estimated by BLUP. The main advantage of BLUP over previous methods is its increased prediction accuracy for genetic effects. This is due to the shrinkage toward the mean that depends on the amount of information available (from the individual and/or its relatives), which will adjust extreme high and low performance toward the overall mean, and also to the incorporation of the genetic correlation between related genotypes from pedigree or genomic information (Piepho et al., 2008). The latter is not a requirement for the model, so the simplest case of BLUP uses no relationship matrix and the genotypes are considered to be independent random variables (Yan and Rajcan, 2003; Cullis et al., 2006). Piepho et al. (2008) present several examples of BLUP analyses in plant breeding.

Although rarely used, pedigree data is an easy and inexpensive source of information for plant breeders to leverage the relationship between individuals for a more accurate estimation of breeding values. Pedigree-based BLUPs have been successfully used in various crops (Bromley et al., 2000; Rutkoski et al., 2016; Basnet et al., 2018; Moreira et al., 2019), and have contributed to major advancements in the rates of genetic progress.

The inclusion of genomic information provides more accurate estimates of genetic relatedness among genotypes, especially with regards to the Mendelian sampling effects (Habier et al., 2007). Genomic information traces allele inheritance, capturing small segments of the genome shared among individuals, even when they are apparently unrelated through pedigree (Velazco et al., 2019). Plant breeders have widely adopted genomic-based BLUPs (**GBLUP**) for genomic selection (Aunger et al., 2016; Crossa et al., 2017; Schrag et al., 2019). Although genomic information is promising, in practice high-density genotyping is not always feasible for all genotypes within a breeding program due to genotyping costs, logistics, or both (Habier et al., 2009). An alternative is to construct a joint relationship matrix based on pedigree and genomic relationships to predict BLUPs for genotyped and non-genotyped material, which is called single-

step GBLUP (**ssGBLUP**; Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). This approach integrates both relationship matrices, connecting their different yet complementary information on genetic relatedness, and provides more reliable and accurate estimates of genetic similarities between genotypes. Genomic breeding values based on the ssGBLUP approach are commonly used in animal breeding (Aguilar et al., 2010; Legarra et al., 2014; Meuwissen et al., 2015; Guarini et al., 2019a, 2019b; Oliveira et al., 2019d), and their use has started to become popular in plant breeding as well (Ashraf et al., 2016; Cappa et al., 2019; Velazco et al., 2019). In sorghum, Velazco et al. (2019) demonstrated that this methodology improves the predictive ability for complex traits, especially for traits with low heritability estimates, measured late in the development stage, or those that are difficult or expensive to measure.

For longitudinal traits, one can calculate BLUPs for each time point separately as individual traits with unique phenotypes; however, these approaches do not directly investigate and compare trends over time (Littell et al., 1998). This makes it difficult to consider a large number of time points and inhibit data comparison when BLUPs shrink differently due to discrepancies in heritability estimates. The main goal of fitting curves and patterns for longitudinal traits is to consider variability in the developmental process across many points in time (e.g. growth). Analytical methods have been developed to better evaluate longitudinal traits using the BLUP context, a simple analysis of variance, or both (Littell, 1990; Meyer and Kirkpatrick, 2005; Mrode, 2014). We will discuss the main methods in this review.

3.5.1 Repeatability Model

Individual measurements recorded over time can be treated as repeated records of the same trait. This is known as the repeatability model. There are two critical assumptions implicit in this method: 1) the variances of different measurements within the same genotype (or individual) are always equal, regardless of the time interval between records; and 2) the genetic correlations between all measurements are equal to one, i.e. measurements at different time points are all influenced by the same genes (Falconer and Mackay, 1996; Meyer and Hill, 1997; Littell et al., 1998). In this scenario, simple repeatability models are the standard approach.

One of the simplest methods is the repeated-measurements analysis of variance (**ANOVA**) using a split-plot in time design, which treats the genotypes as a whole-plot unit and genotypes at particular times as a sub-plot unit (Rowell and Walters, 1976; Littell, 1990). It is important to

mention that as time is a factor in the experiment that cannot be randomized, this is not a true split-plot design. Also, this method assumes the data have equal variances (homoscedasticity) in all repeated measurements and that all pairs of measurements will have the same correlation (i.e. compound symmetry), which are unrealistic assumptions for most crop datasets. However, Huynh and Feldt (1970) showed that the equality of the variances of differences between any two treatment measurements assumed to be correlated was sufficient to perform a split-plot ANOVA. In this case, if the data violate the Huynh and Feldt condition, the F-statistics for the sub-plot unit and their interaction will be inflated. Thus, this method is prone to high Type I error rates, leading to conclusions that effects are statistically significant when they are not (Scheiner and Gurevitch, 2001; Fernandez, 2019).

In the context of mixed models, specifying the random and fixed effects in the model will depend on the study objectives, data structure, and the assumptions that can be made. Usually, time is considered as a fixed effect because it is not randomized in an experiment. Simple repeatability models have been used to calculate BLUPs and BLUEs of longitudinal traits derived from HTP for genomic prediction, such as in wheat (Rutkoski et al., 2016; Sun et al., 2017a).

3.5.2 Multiple-trait Model

Often, HTP platforms are used to generate phenotypes of plants in different “ages” or development stages, with the mean and variance of the phenotypes between measurements/assay dates changing over time. Thus, the assumption is that the genetic control of longitudinal traits will be different over time, characterizing the longitudinal records/phenotypes as different traits. A common approach to analyze longitudinal traits in this scenario is a multi-trait analysis that considers each time point as a different dependent variable (Sun et al., 2017a).

Multivariate analysis of variance (**MANOVA**) is an extension of ANOVA, mentioned earlier, that avoids the covariance structure problems raised in repeated-measures ANOVA. However, it still requires equality in covariance among the groups being compared and balanced data over time. In addition, MANOVA assumes a multivariate normal distribution. Alternative methods have been proposed to overcome these restrictions (Krishnamoorthy and Lu, 2010; Krishnamoorthy and Yu, 2012; Konietzschke et al., 2015), but MANOVA still has limited use in practice.

In the BLUP context, multiple-trait mixed models were first implemented by Henderson and Quaas (1976) to analyze two or more correlated traits making use of genetic and residual covariances among the traits (Speidel, 2011). Using this method, one can directly model the covariance structure of multiple dependent variables and efficiently handle missing data (Mrode, 2014). The main advantage of using a multiple-trait model (**MTM**) over a single-trait repeatability model is the improved evaluation accuracy for each trait arising from better connections in the data between the genetic and residual covariance (Colleau et al., 1999; Mrode, 2014). This data structure will benefit the prediction of traits with lower heritabilities when combined with highly heritable traits and genotypes with missing records for one or more traits (Mrode, 2014). In wheat, MTM was used to predict BLUPs for canopy temperature and normalized difference vegetation index (**NDVI**; Sun et al., 2017), and BLUEs for green NDVI (Juliana et al., 2018).

There are some disadvantages of multiple-trait mixed models. For instance, high-dimensional longitudinal data (e.g. traits recorded multiple times over a long period) can lead to over-parameterized models with high computational requirements (Speidel, 2011). There is also the potential for high correlations between consecutive measurements, which can reduce the power of the tests of significance (Foster et al., 2006). There are approaches to reduce the dimensionality of the MTM, which we are discussed below. It is worth noting that, when applying these approaches, the appropriate models should still be adequate to describe the data, accounting for the changes of mean and covariance over time, and estimate the necessary genetic parameters (Mrode, 2014).

Canonical transformation of phenotypic records is a common procedure to eliminate autocorrelation among traits through eigenvalue decomposition (Meyer and Hill, 1997). A set of highly correlated measures will provide eigenvalues close to zero. Under the framework of canonical transformed phenotypes, the original observations are transformed into a new set of response variables and the ones with the highest eigenvalues are selected to compose the new combination of traits. After fitting the MTM with the new values, the results are transformed back to the original scale (Mrode, 2014). Grosu et al. (2013) highlighted that canonical transformation can only be used if all individuals are recorded for all the traits and that the model needs to be the same for each trait, accommodating only two random effects: residual and genetic. Another strategy to fit MTM is referred to as ‘bending’ (Thompson and Meyer, 1986; Meyer, 2019). It does not require all traits to be measured in all individuals. This procedure forces a decreased

autocorrelation among traits by shrinking the covariance among traits by a bending factor, which creates a positive-definite covariance matrix.

The principal component analysis (**PCA**) and factor analysis (**FA**) methods are often more appropriate to reduce dimensionality for a large number of traits. FA identifies common factors, called latent variables, associated with the correlations between variables (Mrode, 2014). On the other hand, the PCA approach aims to create independent variables (principal components) that explain the maximum amount of variation in the dataset (Mrode, 2014). Thereafter, the principal components or latent variables become the new dependent variables in the MTM. Both methods have been used to reduce the dimensionality of longitudinal trait analysis in animals (Macciotta et al., 2017; Durón-Benítez et al., 2018; Vargas et al., 2018), and plants (Kwak et al., 2016; Yano et al., 2019).

As longitudinal traits are, by definition, taken along a time trajectory, the whole data set can be represented by parameters describing the shape of the trajectory curve (e.g. growth curve). These parameters can become the new dependent variables or integrated covariance structures in MTM (Speidel, 2011; Oliveira et al., 2019a); however, none of the approaches to analyze longitudinal data that we have discussed so far have considered that the genetic and environmental variances may change over time (Meyer, 1998, 2005a; Oliveira et al., 2019a). In addition, these approaches are limited to the time points at which traits were measured. Random regression models (**RRMs**) provide a way to overcome these limitations (Schaeffer, 2004).

3.5.3 Random Regression Model

A common property of longitudinal traits is that the covariance between repeated measures depends on the interval of time between them. In other words, measurements collected closer in time will be more correlated than measures collected farther apart. Kirkpatrick et al. (1990) presented the concept of analyzing longitudinal data using covariance functions by describing the covariance structure of the traits as a function of time. In essence, this approach fits a set of orthogonal functions to a given covariance matrix for the records taken over time (Meyer and Hill, 1997).

First-order autoregressive analysis (**AR-1**) is an appealing method for modeling covariance structure for phenotypes measured over time (Apiolaza and Garrick, 2001; Yang et al., 2006; Vanhatalo et al., 2019). It assumes homogenous variances and correlations that decline

exponentially as measurements are separated by greater time intervals. Thus, two measurements collected closer in time will be more correlated than those further apart (Wade et al., 1993; Littell et al., 2000; Piepho et al., 2004). The AR-1 structure is only applicable for measurements taken at equally spaced time points (Wang and Goonewardene, 2004). Though this is a difficult requirement to meet in agricultural research, especially in field trials, modeling the longitudinal trait as described in the previous section would make the data evenly spaced over time and validate the AR-1 method. An alternative is to use a spatial power covariance structure that allows for unequal intervals between time points (Wang and L. A. Goonewardene, 2004).

Note that so far, we are assuming homogenous variance over time. There are also covariance structures to handle heterogeneous variance, such as the first-order ante-dependence structure (Wolfinger, 1996). Thus, Legendre orthogonal polynomials and splines are more attractive covariance functions as they produce relatively small correlations among the regression parameters and adjust flexibly to the shape of the trajectory curve (Schaeffer, 2004; Meyer, 2005b, 2005a; Bohmanova et al., 2008; Pereira et al., 2013; Brito et al., 2018). In plants, different covariance structures have been assessed for a variety of traits (Apiolaza et al., 2011; Sun et al., 2017a; Campbell et al., 2019).

Meyer and Hill (1997) showed that covariance functions are equivalent to RRM. Schaeffer (2016) reported that covariance functions help to predict the change in variation over time, while RRM is a way to estimate covariance functions and determine individual differences in trajectories. RRM provide a robust framework for modeling trait trajectories using covariance at or between each time point with no assumptions of constant variances or correlations. RRM provide insights about the temporal genetic variation of developmental behavior underlying the studied traits (Oliveira et al., 2019a). Despite the increased computational cost, RRM result in more accurate breeding values compared to other methods (Sun et al., 2017a; Oliveira et al., 2019a).

The RRM were first introduced in animal breeding to overcome over-parameterized models in MTM and they have been used extensively since then (Jamrozik and Schaeffer, 1997; Schaeffer, 2004; van Pelt et al., 2015; Englishby et al., 2016; Oliveira et al., 2019a). In summary, RRM set the parameters of the function describing the trajectory of the trait as fixed and random effects in the model, resulting in fewer parameters than MTM (Schaeffer, 2016; Oliveira et al., 2019a). Consequently, in RRM the random parameters do not correspond directly to the individuals' genetic value for the longitudinal trait. Rather, they correspond to the genetic values

of sets of regression coefficients that represent the time trajectory of the longitudinal trait for each genotype (Turra et al., 2012). Estimates of genetic parameters and breeding values can be obtained for all time points within the interval analyzed from the genetic (co)variance matrices for the regression coefficients and the matrix of independent covariates for all time points associated with the function used (Oliveira et al., 2019a). When the same fixed effects are used in all models, it is appropriate to examine different covariance structures using real data and select the one that best fits the model based on a statistical methods, such as the Akaike Information Criterion (**AIC**, Wang and Goonewardene, 2004) or Bayesian Information Criterion (**BIC**, Neath and Cavanaugh, 2012). Finally, estimate the effects of interest using the selected covariance structure. In a general form, RRM can be described as follows (Oliveira et al., 2019a):

$$y_{gij} = \sum_{q=1}^Q b_{qg}z_{qg} + \sum_{r=1}^R a_{ri}z_{ri} + \sum_{s=1}^S p_{si}z_{si} + e_{gij}$$

where y_{gij} is the j^{th} repeated record of genotype i (e.g., canopy coverage at different days after planting); b_{qg} is the q^{th} fixed regression coefficient for the g^{th} group; a_{ri} is the r^{th} random regression coefficient for the additive genetic effect of the i^{th} genotype; p_{si} is the s^{th} random regression coefficient for permanent environmental effect of the i^{th} genotype; e_{gij} is the residual effect; and z_{qg} , z_{ri} and z_{si} are the covariates related to the function used to describe time (e.g., days after planting), assuming the same function (e.g., Legendre polynomial) with possible different orders Q , R , and S (e.g., linear, quadratic, cubic) (Oliveira et al., 2019a).

Random regression models have been shown to be the most effective choice to genetically evaluate longitudinal traits in numerous livestock breeding programs (as reviewed by Oliveira et al., 2019), but there are only a few examples of the applications of RRM in plant breeding, especially when incorporating genomic information. Sun et al. (2017) captured the change of HTP traits continually over wheat growth stages using RRM. Campbell et al. (2018) used RRM to predict shoot growth trajectories in a rice diversity panel and demonstrated an improvement in prediction accuracy compared to a single-time-point model. Based on the same rice dataset, Campbell et al. (2019) used RRM to identify QTL with time-specific effects. Multiple-trait RRM are also feasible and have been implemented in several settings in animal breeding programs (Nobre et al., 2003; Muir et al., 2007; Oliveira et al., 2019d, 2019b).

3.6 Implementation of Genomic Selection for Longitudinal Traits

Meuwissen et al. (2001) introduced the concept of genomic selection (GS) based on the idea that markers from dense genome-wide genotyping will be in linkage disequilibrium with QTLs that have an effect on the quantitative trait of interest. Thus, they can be used for selection without identifying the QTL or the functional polymorphism. This increased understanding of GS arose as it became known that markers would carry relationship information in addition to the signal captured by the linkage disequilibrium between markers and QTL (Habier et al., 2007; Meuwissen, 2009).

In GS, genomic and phenotypic data are combined in a training population to enable the development of prediction equations that can be used in a testing (or target) population of selection candidates consisting of individuals that were genotyped but not phenotyped (Cossa et al., 2017). Therefore, GS enables a more accurate selection of individuals at an early age (with no measurements). This increases the rate of genetic gain by reducing the time required for the variety development and the cost per cycle. High-throughput phenotyping is able to generate high-quality quantitative data and effectively characterizes large training populations during the growing season. The combination of GS and HTP has the potential to increase precision and efficiency while lowering costs and minimizing labor (Araus et al., 2018).

Under the longitudinal framework of GS, the prediction of temporal breeding values enables targeted selection on specific periods in the growing season or selection of individuals that exhibit desirable trait trajectories. In addition, the longitudinal trait can be used as secondary traits to improve the genomic selection of economic endpoint traits such as yields (Sun et al., 2017a). Campbell et al. (2018) used RRM with a second-order Legendre polynomial to perform pedigree and genomic predictions of shoot growth trajectories in a rice diversity panel. They demonstrated an improvement in prediction accuracy using the RRM compared to a single-time-point model. Furthermore, the authors reported that genomic RRMs were useful in predicting future phenotypes using a subset of early measurements. Another study in rice used RRMs to predict projected shoot area in controlled and water-limited conditions using Legendre polynomials and B-spline basis functions (Momen et al., 2019). Before fitting both functions, they adjusted raw phenotypic measurements to obtain BLUEs for downstream genetic analysis. Overall, RRMs produced higher prediction accuracy compared to the baseline multiple-trait model. In addition, B-splines performed slightly better than Legendre polynomials (Momen et al., 2019).

Currently, statistical models used in GS for plant breeding are most often single-trait (univariate) and do not take advantage of genetic covariance among traits or phenotypic records collected at different time points (Jia et al., 2012). However, MTM for GS was shown to outperform single-trait models by accounting for correlation among traits, thereby increasing prediction accuracy, statistical power, parameter estimation accuracy, and reducing trait selection bias (Jia et al., 2012; Guo et al., 2014a; Montesinos-López et al., 2016, 2019). These advantages are even more obvious for low-heritability traits, such as yield, that are genetically correlated with highly heritable traits (Guo et al., 2014a; Jiang et al., 2015). Recently, studies in the CIMMYT wheat breeding program (www.cimmyt.org/) have shown that the accuracy of GS is greatly improved by incorporating HTP longitudinal data from the so-called secondary traits measured with UAV (Rutkoski et al., 2016; Montesinos-López et al., 2017; Sun et al., 2017a, 2019), an approach that is relatively inexpensive to implement as HTP and genotyping have become more accessible (e.g. targeted genotyping-by-sequencing approach; Pembleton et al., 2016). In addition, secondary traits are also useful to predict the primary trait at early growth stages, since they can often be phenotyped ahead of a primary trait like grain yield (Sun et al., 2017a). Therefore, longitudinal traits can be used as secondary traits to improve the accuracy of GS and contribute to a better understanding of the biological mechanisms underlying stress responses and development. As described in the previous section, there are various ways to extract the genetic information from longitudinal traits and the methods employed will determine how they can be used in GS.

Rutkoski et al. (2016) used HTP canopy temperature (**CT**), green normalized difference vegetation index (**GNDVI**), and red normalized difference vegetation index (**RNDVI**) taken over time as secondary traits in GS for yield in wheat. First, they estimated BLUEs for the longitudinal traits using the repeatability model and used them in an MTM with yield, for pedigree and genomic predictions. They found that multiple-trait modeling with secondary traits increased accuracies for grain yield using both pedigree and genomic information, compared to the single-trait models. In another study, CT and NDVI also improved the ability to predict grain yield in wheat (Sun et al., 2017a). However, in addition to a repeatability model, the authors also used MTM and RRM to calculate BLUPs for the secondary traits in order to compare their efficiency. The predictive ability improved by 70%, on average, when including secondary traits, and the predictive ability of RRM and MTM were superior to the repeatability model. Also in wheat, Juliana et al. (2018) performed pedigree and genomic multi-trait prediction models using BLUEs of yield and GNDVI measured

at different dates. They found that including GNDVI increased prediction accuracies. Sun et al. (2019) used an RRM with a cubic smoothing spline to predict BLUPs for CT and GNDVI in wheat. In a second step, they used BLUPs for secondary traits and grain yield as the dependent variables in GS. The prediction accuracy using the secondary traits increased by an average of 146% for grain yield across cycles and the secondary traits measured in the early stages were optimal for enhancing the prediction accuracy. Montesinos-López et al. (2017) and Crain et al. (2018) obtained similar results in wheat. Howard and Jarquin (2019) modeled the genetic covariance between canopy coverage and yield using the SoyNAM dataset (Song et al., 2017; Diers et al., 2018) and demonstrated that, based on different cross-validation schemes, the predictive ability was the highest when both canopy and marker information were included in the model. Two other papers reported similar improvements with the same dataset (Xavier et al., 2017; Jarquin et al., 2018).

Given the capability of HTP to collect multiple temporal traits at the same time, multiple-trait RRM (MTRRM) can be powerful tools for joint genomic prediction of several longitudinal traits (Oliveira et al., 2016). In addition, MTRRM can incorporate different functions to describe different traits in the same model and estimate genetic correlations between different traits over time (Oliveira et al., 2016). In animals, MTRRM are a plausible alternative for joint genetic prediction of milk yield and milk constituents in goats (Oliveira et al., 2016), cattle (Oliveira et al., 2019d), and buffaloes (Borquis et al., 2013). Recently, MTRRM for projected shoot area and water-use recorded daily over a period of 20 days showed better predictive abilities compared to single-trait RRM (Baba et al., 2020).

In animal breeding and multiple-stage plant breeding analysis, it is common to use deregressed genetic values as the pseudo-phenotypes for genomic predictions. Oliveira et al. (2018) compared different deregression methods for longitudinal traits. However, this multiple-step approach may result in lower accuracies, bias, and loss of information (Legarra et al., 2009; Kang et al., 2017). Considering the advantages of ssGBLUP and RRM in genetic evaluation, integrating both approaches is an effective strategy to enhance the genomic prediction of longitudinal traits (Kang et al., 2017). Koivula et al. (2015) reported higher accuracy and less bias in the prediction of Nordic Red Dairy cows for milking performance using a ssGBLUP RRM compared to the traditional pedigree-based RRM. Kang et al. (2017) showed that ssGBLUP RRM achieved the highest accuracy and least bias under a variety of scenarios, including persistency of accuracy over generations, compared to other models. In summary, the use of ssGBLUP based on RRM can

increase the reliability of genomic predictions for test-day traits in dairy cattle (Koivula et al., 2015; Kang et al., 2018; Oliveira et al., 2019d), and possibly in crops.

3.7 Detecting QTL and Causal Variants Associated with Longitudinal Traits

One of the main goals in genomic research is to predict the phenotypic variation using genotypes, by identifying genetic variants. The development of an organism is the result of an interacting network of genes and environmental factors (Wu and Lin, 2006). Unlike single-time-point measurements, studying longitudinal traits as a function of time allows the comprehensive assessment of crop growth and development (e.g. age metabolic rate; Ma et al., 2002). However, in plants, the detection of QTL analysis or genome-wide association studies (**GWAS**) for longitudinal traits are still performed at each time point independently. For instance, Würschum et al. (2014) used linkage mapping at discrete time points separately to identify time-specific QTLs associated with plant height in triticale. In cotton, canopy-related traits were used separately for each of the several studied days to map additive QTL effects and their interaction with the environment (Pauli et al., 2016a). In soybeans, a GWAS was used to identify QTL for each individual canopy coverage measurement spanning 14–56 days after planting (Xavier et al., 2017). Zhang et al. (2017) performed QTL mapping for several growth-related traits at 16 time points separately in maize. Also in maize, analysis of individual time points found different and simultaneous QTLs controlling plant height at different growth stages (Wang et al., 2019). Using time point growth-related traits, Knoch et al. (2019) found evidence of temporal QTLs in canola. Although useful, these static examinations provide a simplified view of genetic control, neglecting temporal changes and developmental features of trait formation (Wu and Lin, 2006). In addition, in animals, it has been shown that neither the phenotypic nor additive polygenic effects of longitudinal traits are constant throughout the entire phenotypic expression (Szyda et al., 2014; Brito et al., 2018; Oliveira et al., 2019a).

As an alternative, Ma et al. (2002) proposed a dynamic model, called functional mapping, to map QTLs associated with the whole developmental process of longitudinal traits. As mentioned earlier, longitudinal traits can be represented as curves, described by a few parameters from a linear or non-linear function over a given time. The idea behind functional mapping is that the difference in curve parameters among genotypes may suggest temporal patterns of genetic control over the phenotypic trajectory (Ma et al., 2002). Thus, functional mapping allows testing of timing and the

duration of QTL expression (Wu et al., 2004). Several modeling strategies for functional mapping have been proposed and have been reviewed by Li and Sillanpää (2015). One of the approaches (the two-stage method) consists of modeling the whole phenotypic trajectory using linear and nonlinear models and using these parameters as latent-trait phenotypes for QTL detection (Li and Sillanpää, 2015). Often, researchers perform analyses for individual time points followed by this two-stage method to derive the curve parameters. Busemeyer et al. (2013) used a logistic function to fit high-throughput-derived biomass from different development stages of a large mapping population of 647 double-haploid triticale lines. In addition to GWAS for the individual days, they performed a multiple-trait functional GWAS using the parameters from the logistic curve to reveal temporal genetic patterns of biomass regulation. A similar approach was used to assess image-derived biovolume in maize lines (Muraya et al., 2017); digital biomass accumulation in spring barley (Neumann et al., 2017); and area, height, and senescence in wheat (Camargo et al., 2018). Campbell et al. (2017) calculated the projected shoot area in 360 rice accessions from 19 to 41 days after transplanting. They modeled the longitudinal phenotypes using a power function (Paine et al., 2012) and used the parameters as the pseudo-phenotypes in a multiple-trait GWAS. In order to reveal the temporal dynamics of senescence in potato, Hurtado et al. (2012) employed P-splines as a smoothing curve and used the curve parameters for identifying QTLs.

Kwak et al. (2014) proposed two simple regression-based methods to map QTL by analyzing each time point separately and then combining test statistics across time points to determine the overall significance. Later, Kwak et al. (2016) proposed an improved approach where the observed longitudinal traits were replaced by a smoothing approximation, followed by dimensional reduction via PCA. Multiple-trait QTL analysis was then performed on the reduced data (using principal components). Muraya et al. (2017) implemented the approach suggested by Kwak et al. (2016). They used B-splines to smooth the phenotype followed by PCA for variable reduction and performed a multiple-QTL analysis following the methods of Kwak et al. (2014) to reveal the underlying genetic variation of growth dynamics in maize. Temporal height QTLs in the model C4 grass *Setaria* were also revealed using this approach (Feldman et al., 2017). Animal breeders have been using PCA for longitudinal trait analysis for a long time to synthesize complex patterns and reduce computationally demanding multiple-trait QTL detection (Macciotta et al., 2006, 2015, 2017; Zhang et al., 2018b).

RRMs offer a better option to fit longitudinal traits and have been widely used in genetic evaluation of animals (Ning et al., 2017; Oliveira et al., 2019a). The random regression approach uncovers SNP effects over time because it is able to identify persistent and time-specific transient QTLs. Moreover, RRM have increased statistical power to detect QTLs over other approaches because they leverage the full set of raw longitudinal phenotypes (Ning et al., 2017) and can capture QTLs with significant effects in specific regions of the development curve, though the effects of these QTLs may be small overall. RRM are also useful for detecting QTL in gene-by-environment interactions (Lillehammer et al., 2007; Carvalheiro et al., 2019).

Das et al. (2011) proposed a method called functional GWAS (*f*GWAS), based on RRM, which integrates GWAS and mathematical models describing biological processes. In summary, *f*GWAS estimates the mean for different SNP effects for each genotype and time point and then performs hypothesis testing to determine whether the SNP has any additive or dominant effect during the time course. The main drawback of this method is that it only performs a single-locus analysis. Later, Ning et al. (2017) proposed a modification of *f*GWAS by estimating the time-dependent population mean and the SNP effects separately, instead of fitting them directly. They also extended the model to capture the time-varying polygenic effect of complex traits by treating SNPs as covariates (*f*GWAS-C) or factors (*f*GWAS-F). However, their method was shown to be computationally inefficient due to the high dimensionality of the mixed model equations compared to other models. Subsequently, Ning et al. (2018) proposed a rapid longitudinal GWAS method, transforming the covariance matrices to diagonal matrices using eigen-decomposition. This way the model can be solved by a weighted least squares model for each SNP test.

To the best of our knowledge, Campbell et al. (2019) were the first to use RRM GWAS in a major row crop. They took the genomic breeding values derived from RRM using Legendre orthogonal polynomials to assess the genetic architecture of rice shoot growth over a period of 20 days during early vegetative growth. They found both transient and persistent effects associated with shoot growth and more associations with the RRM when compared to single-time-point analysis.

3.8 Challenges and Future Developments

To capitalize on advances in phenotyping and molecular technologies, greater progress is needed in developing ways in which breeders can manipulate systems to understand the

relationships between genotype and phenotype. The underlying biological changes due to environmental systems and/or over time can be captured with longitudinal data. The major challenge is synthesizing the various layers of information together in a meaningful manner to understand the downstream effects of developmental stress and implications for breeding (Harfouche et al., 2019).

3.8.1 Non-Additive Effects and GxE

Non-additive genetic effects may significantly contribute to the total genetic variation of complex traits. Prediction models that include dominance effects represent an important component of breeding programs that focus on crossbred populations, hybrid production, and vegetatively-propagated species (Almeida Filho et al., 2019). There is also ample evidence of the importance of epistasis in the genetic architecture of complex traits in various crops (Guo et al., 2014b; Monir and Zhu, 2018). Integrating non-additive effects into statistical models may improve prediction accuracy and detect more QTLs than simple additive models, especially when the non-additive variance contributes to a large proportion of the genetic variance (Bouvet et al., 2016; Bonnafous et al., 2018; Liu et al., 2018b, 2019; Monir and Zhu, 2018; Varona et al., 2018). However, these studies are restricted to single-time-point traits. For longitudinal traits, it may be challenging to have a full genetic model (including both additive and non-additive effects), requiring dense marker panels to estimate the time-dependent (co)variances, as well as partitioning of the genetic variance components. Nevertheless, full genetic models of longitudinal traits may have the potential to impact future design and implementation of breeding strategies.

The temporal dynamics of longitudinal traits lead to interactions that change the phenotype over time. This may be because the gene-gene and gene-environment interactions (GxE) are time- or age-dependent and need to be properly modeled (Fan et al., 2012). In this case, environmental descriptors should be measured several times as the trait phenotypes. The resulting model is a multiple-trait, multiple-environment model with a variety of interactions, in which computational issues may arise due to the increase in the number of parameters being estimated. It has been shown that RRM can account simultaneously for the additive genetic effect and some degree of GxE in longitudinal traits in animal breeding by allowing for the estimation of genetic (co)variance components and breeding values over the whole trajectory of a time-dependent trait and environment-dependent covariate (Brügemann et al., 2011; Santana et al., 2016; Bohlouli et al.,

2019). In plant breeding, therefore, this model may provide considerable biological insights into the mechanisms determining performance in specific environments, making it a worthwhile method for study in future research.

3.8.2 Complementary “-omics” Technologies

The rapid advances in “-omics” technologies enable the generation of large-scale “-omics” datasets for many crop species, providing new opportunities to investigate and improve complex traits. The different approaches described in this review offer valuable tools to combine phenomics and genomics data to reveal the underlying genetic basis of longitudinal traits. However, one current challenge is integrating additional “-omics” technologies (e.g. transcriptomics, metagenomics, proteomics, metabolomics, epigenomics) to provide a holistic multi-omics approach to study biological mechanisms and their response to environmental stresses for important agronomic traits.

Recently, methods that combine “-omics” information have been used in some crops to study phenotypic networks for single-point traits, for example, to pinpoint candidate genes and/or loci and predict phenotypic variation (Acharjee et al., 2016; Li et al., 2016; Das et al., 2017; Sheng et al., 2017; Pandey et al., 2018; Jiang et al., 2019). A meta-analysis of the detailed “-omics” datasets regarding longitudinal traits in crops has been limited so far. Baker et al. (2019) characterized the mechanistic connections between the genomic architecture, transcriptomic expression networks, and phenotypic variation of growth curves that underlie the developmental dynamics of plant height in *Brassica rapa*. The combination of multi-omics approaches also seems promising to elucidate senescence processes in model and crop plants (Großkinsky et al., 2018). When joint modelling longitudinal “-omics” data (one or more type of “-omics” data measured over time) the statistical analysis becomes more challenging. Some key points can be found in Sperisen et al. (2015). In general, there is a need to adapt methodologies and experimental designs to explore processes related to the global evolution of biological processes such as growth and development. Despite all these challenges, integrative methods can increase analysis power to find true causal variants, regulatory networks, and pathways. These, in turn, could be incorporated into GS and breeding programs to speed up genetic gains (Suravajhala et al., 2016).

3.8.3 Deep Learning

Deep learning (DL) is a powerful and highly flexible class of machine learning algorithms based on representation-learning methods that incorporates multiple levels in a non-linear hierarchical learner (Lecun et al., 2015). Essentially, DL is an advanced version of artificial neural networks (ANN) with multiple hidden layers that aims to mimic the human brain functioning (Patterson and Gibson, 2017).

Deep learning has demonstrated its utility in different fields of biological sciences, such as disease diagnosis (Gao et al., 2018), multi-omics data integration (Chaudhary et al., 2018), predicting DNA- and RNA-binding specificity (Trabelsi et al., 2019), and, recently, in plant breeding genomic prediction (Ma et al., 2018; Montesinos-López et al., 2018a, 2018b, 2019). Zou et al. (2019) and Pérez-Enciso and Zingaretti (2019) provide a primer on DL in genomics. The growing interest in DL methods in plant breeding, especially for prediction, may arise from its powerful capability of learning complex non-linear relationships between predictors and responses hidden in big data, usually resulting in higher accuracy when compared with other methods (Montesinos-López et al., 2018a; Pérez-Enciso and Zingaretti, 2019; Zou et al., 2019). It is important to point out that even though DL can deal with complex scenarios and achieve state-of-the-art accuracy, it requires domain knowledge and large-scale datasets, while the interpretation of the underlying biology is more challenging than for standard statistical models (Zou et al., 2019).

Within the classes of DL, recurrent neural networks (RNN) are designed for sequential or time-series data (Lecun et al., 2015) and may be the most appropriate architecture to model longitudinal traits. An RNN can be thought of as a memory state that retains information on previous data the network has seen and updates its predictions in the light of new information. Thus, besides prediction, RNN has the ability to capture long-term temporal dependencies (Che et al., 2018). Recently, RNNs have achieved astonishing results in many applications with time series or sequential data, particularly in human sciences (Azizi et al., 2018; Che et al., 2018; Lee et al., 2019; Sung et al., 2019; Zhong et al., 2019). Despite its advantages, to our best knowledge RNNs have not been employed in genomic prediction or mapping QTL for longitudinal traits in plant breeding. In the context, versatile DL models for multiple-trait analysis (Montesinos-López et al., 2018b), multiple-environment analysis (Montesinos-López et al., 2018a), and performing simultaneous predictions of mixed phenotypes (binary, ordinal and continuous; Montesinos-López et al., 2019) have been successfully implemented. Such cases are not only encouraging but may

lead to future integration of DL and RNNs into the analysis of longitudinal traits in crops. DL is a powerful approach and is likely to transform many domains in plant breeding because it has the potential to handle all the complexities highlighted in this review. Needless to say, further innovation and technology assessment are required to fully enable DL to deal with the unique properties of planting breeding data.

CHAPTER 4. HIGH-THROUGHPUT PHENOTYPING AND RANDOM REGRESSION MODELS REVEAL TEMPORAL GENETIC PATTERNS OF SOYBEAN BIOMASS

A version of this chapter is pending publication.

Fabiana F. Moreira¹, Hinayah R. Oliveira², Miguel Lopez¹, Bilal J. Abughali³, Guilherme Gomes⁴, Keith Cherkauer³, Luiz F. Brito², Katy M. Rainey^{1*}

¹Department of Agronomy, Purdue University, 915 West State Street, West Lafayette, Indiana, 47907, United States of America.

²Department of Animal Sciences, Purdue University, 270 South Russell Street, West Lafayette, Indiana, 47907, United States of America.

³Department of Agricultural and Biological Engineering, Purdue University, 225 South University Street, West Lafayette, Indiana, 47907, United States of America.

⁴Department of Statistics, Purdue University, 250 North University Street, West Lafayette, Indiana, 47907, United States of America.

*Corresponding author: Katy M. Rainey (krainey@purdue.edu)

Authors` Contributions

FM developed the experiment, collected field data, conducted statistical analysis and wrote the manuscript. FM and KR conceived and designed the study. ML assisted with field data collection. KC and BA conducted multispectral images analysis. GG assisted with AGB phenotypic prediction. HO and LB assisted with the random regression model analyses. HO, LB, and KR critically revised and improved the manuscript. All authors read and approved the manuscript.

4.1 Abstract

Improving the temporal accumulation of above-ground biomass (AGB) has the potential to increase soybean yield. High-throughput phenotyping platforms allow plant breeders to measure phenotypic variability on the same genotype multiple times during a season, monitoring the development of a crop through its life stages, and how it responds to the environment. Random regression models (RRM) are routinely used in animal breeding for the genetic analysis of longitudinal traits and can provide insights about the temporal genetic variation of the studied traits. The main objectives of this study were: to develop a high-throughput phenotyping method to predict soybean AGB over time, and to reveal the quantitative genomic properties of temporal AGB, by using genomic prediction of breeding values and genome-wide association studies (GWAS) based on RRM. A random subset of 32 families from the SoyNAM population (n=383) was grown in field trials in three environments. Destructive AGB measurements were collected

along with multispectral and Red, Green and Blue (RGB) imaging with an unmanned aerial system from 27 to 83 days after planting (DAP). Two machine-learning methods were used to predict soybean AGB: Least Absolute Shrinkage and Selection Operator (LASSO) Regression and Partial Least Squares Regression (PLSR). Both methods performed similarly and resulted in AGB predictions with high R² values (0.92-0.94). When investigating different RRM, the most suitable to fit the data was the model using linear B-spline with 2 knots and heterogeneous residual variance. Narrow-sense heritabilities estimated over time using the mentioned RRM ranged from low to moderate (from 0.02 at 44 DAP to 0.28 at 33 DAP). In addition, adjacent DAP had the highest genetic correlations compared to those DAP further apart. The high accuracies and low biases of prediction suggest that genomic breeding values for AGB can be predicted over time using RRM. Different genomic regions associated with AGB were also found over time, and no genetic markers were significant in all time points. Thus, RRM seems a powerful tool for modeling the temporal genetic architecture of soybean AGB and can provide useful information for variety improvement. This study is the first application of RRM for genomic evaluation of a longitudinal trait in soybeans.

4.2 Introduction

Soybean [*Glycine max* (L.) Merr.] is one of the most economically important crops worldwide, being the primary source of plant-based protein, and the second largest source of vegetable oil available (USDA, 2018). Advances in plant breeding and enhanced agronomic methods have substantially improved soybean yield over time (Anderson et al., 2019). Yield potential in any environment or cropping system can be expressed as a function of biomass produced, and the partitioning of biomass to the seeds, or harvest index (Monteith, 1972, 1977). Assessments of historical soybean germplasm have shown that increases in soybean yield over the last several decades are associated with increases in biomass production (Cregan and Yaklich, 1986; Frederick et al., 1991; Kumudini et al., 2001; De Bruin and Pedersen, 2009; Koester et al., 2014; Balboa et al., 2018). For instance, using cultivars released between 1923 and 2007 and measuring above-ground biomass (AGB) every two weeks, Koester et al. (2014) found greater biomass production per unit of absorbed light with the release year. Additionally, information on temporal biomass production can give insights about crop development and response to multiple abiotic and biotic stressors (e.g., Bajgain et al., 2015; Jumrani and Bhatia, 2018). Increased temperatures and water stress imposed vegetative and reproductive stage reduced AGB

significantly and resulted in 28 and 74% reduction in soybean yield, respectively (Jumrani and Bhatia, 2018). Hence, understanding the genetic factors controlling the temporal dynamics of biomass accumulation may contribute to future soybean yield gains and the development of stress-resilient cultivars.

Measuring AGB at multiple developmental stages is laborious, involving cutting, drying, and weighing plants from a target area, and is subject to errors resulting from i) unrepresentative samples; ii) destructive sampling, which limits the number of samples that can be collected from a plot, and prevents longitudinal tracking of the same target area; and, iii) sampling and subsequent processing involves extensive manual handling, which may lead to sample loss, and can be restrictive in large experiments (Jimenez-Berni et al., 2018). High-throughput phenotyping platforms (HTPP) offer alternatives to ground-based AGB sampling, enabling collection of non-destructive data in large populations and experiments throughout the growing season under actual field conditions (van Eeuwijk et al., 2018; Zhao et al., 2019). In other crops, such as wheat, barley, rice, and dry beans, AGB accumulation has been recognized as a potential target to increase yield gain, and the success of image-based AGB phenotyping has been demonstrated (Serrano et al., 2000; Babar et al., 2006; Tilly et al., 2014; Cheng et al., 2017; Neumann et al., 2017; Yue et al., 2017; Sankaran et al., 2018). Using information from multiple sensors is a common practice to predict AGB because it improves trait estimation by combining the advantages of the spectral, spatial and structural metrics derived from different sensors (Bendig et al., 2015; Chen et al., 2016; Wang et al., 2017; Maimaitijiang et al., 2019; Li et al., 2020). For instance, spectral indices and plant height were used to predict barley, wheat, and potato AGB (Bendig et al., 2015; Yue et al., 2017; Li et al., 2020); and spectral and structural data fusion was applied for AGB estimation in maize (Wang et al., 2017). In soybean, Maimaitijiang et al. (2019) used Red, Green and Blue (RGB) imagery-derived metrics to predict AGB in agricultural fields; however, there are limited studies on the use of different HTPP to estimate soybean AGB in a plant breeding scenario.

High-throughput phenotyping (HTP) allows time-series measurements that monitor the development of a crop through its life stages, and how it responds to the environment. These measurements represent the crop in different “ages” or stages of development, with the mean and variance between measurements usually changing over time, characterizing the trait as longitudinal (Falconer and Mackay, 1996; Yang et al., 2006; Oliveira et al., 2019a). Different approaches can be utilized for genomic evaluation of longitudinal traits. A simple repeatability

(SR) model treats the individual measurements recorded over time as repeated records of the same trait (Meyer and Hill, 1997). This model assumes that the variances of different measurements are equal and the genetic correlations between all measurements are equal to one, which is unrealistic for most crop datasets (Falconer and Mackay, 1996; Meyer and Hill, 1997; Littell et al., 1998). An alternative method that overcomes these restrictions is a multiple-trait model (MTM), that considers each time point as a different trait. However, high-dimensional longitudinal data can lead to high correlations between consecutive measurements and over-parameterized models with high computational requirements, restricting the application of MTM (Foster et al., 2006; Speidel, 2011). Also, it has been shown that the phenotypic or additive polygenic effects of longitudinal traits are not constant during the entire trait expression in animals (Szyda et al., 2014; Brito et al., 2018; Oliveira et al., 2019a). Consequently, breeders need to adopt an amenable statistical framework for genetic and genomic analysis that accounts for the time-dependent genetic contribution to the phenotype.

Random regression models (RRM) provide a robust framework for estimating breeding values and identifying quantitative trait locus (QTL) with time-specific effects for longitudinal traits (Oliveira et al., 2019a). In summary, RRM use a given covariance function to describe the trajectory of the trait as a function of time (or environmental gradient), with no assumptions for constant variances and correlations (Kirkpatrick et al., 1990; Meyer and Hill, 1997; Schaeffer, 2016). When comparing with other models, RRM have some key advantages, such as 1) they are more computationally efficient, 2) enable prediction of breeding values for any time point within the range of data collection, and 3) provide more accurate breeding values (Oliveira et al., 2019a). The RRM were originally proposed for use in livestock breeding programs in order to overcome the problem of over-parameterization faced in MTM, and it has been successfully used for genetic evaluation of longitudinal traits since then (Jamrozik and Schaeffer, 1997; Schaeffer, 2004; van Pelt et al., 2015; Englishby et al., 2016; Oliveira et al., 2019a). However, just recently it started to be implemented in plant breeding (Sun et al., 2017; Campbell et al., 2018, 2019). Therefore, there is a great opportunity to extend the use of RRM to temporal measurements of complex polygenic phenotypes in major crops.

In this context, this study aimed to: (1) develop an HTTP methodology to estimate soybean AGB throughout the growing season; (2) reveal the genetic architecture and estimate time-dependent effects of single nucleotide polymorphisms (SNPs) associated with this longitudinal

trait using RRM, and; (3) investigate the feasibility of implementing genomic selection for longitudinal traits in soybean using RRM.

4.3 Material and Methods

4.3.1 Plant Materials, Field Experiments, and Genotypic Data

In this study, we used a set of 383 recombinant inbred lines (RILs) representing 32 families from the Nested Association Mapping (SoyNAM) population (~12 RILs per family) (Diers et al., 2018). The lines comprising the set were selected using breeding values for full maturity (R8) (Fehr and Caviness, 1977) and grain yield, calculated from experiments performed in Indiana and Illinois from 2011 to 2014, in order to have a maturity-controlled panel (Xavier et al., 2016; Lopez et al., 2019). More details about RIL panel selection and the full list of traits' collection and distribution are described in Lopez et al. (2019). Yield (kg/ha) phenotypic distribution by environment is presented in Appendix B Figure B.1.

The RILs were grown under a randomized complete block design with two replications at the Purdue University Agronomy Center for Research and Education (ACRE), West Lafayette, IN (40°28'20.5"N 86°59'32.3"W), and at Romney, IN (40°14'59.1"N 86°52'49.4"W). Planting occurred on May 31, 2017, and May 22, 2018 at ACRE, and May 17, 2018 at Romney. The combination of year and location where the experiment was grown was considered as an environment, resulting in three environments in this study (i.e., 2017_ACRE, 2018_ACRE, and 2018_Romney). Experimental units consisted of a six-row plot (3.35 m with 0.76 m) with a targeted seeding rate of 35 seeds m⁻². A total of 66 and 16 RILs were discarded in 2017 and 2018, respectively, as a consequence of poor emergence. In addition to the two full replications, we randomly selected 62 RILs in 2017 and 108 RILs in 2018 (the same 62 RILs in 2017 plus 46 others) to grow in a trail of eight-row plots (0.76m x 3.35m). This trail was defined as the biomass sampling panel and it was used as sampling plots for destructive AGB measurements throughout the growing season.

In the biomass sampling panel, AGB was collected approximately every 10 days during the growing season between 27 to 83 days after planting (DAP), from a linear section of 0.56 m in a row with borders. In 2017, we randomly picked plots to measure AGB in replication 1 for every sampling date, while in 2018 three full AGB sampling (~38, 58, and 84 DAP) were performed for

both locations in the two full replications. The fresh AGB was dried at 80°C using a dry-air system until achieving constant weight. Finally, we obtained the dry AGB weight and rescaled it to g/m². Figure 4.1 shows the data collection timeline for each environment, along with the planting date and the respective phenological stage periods.

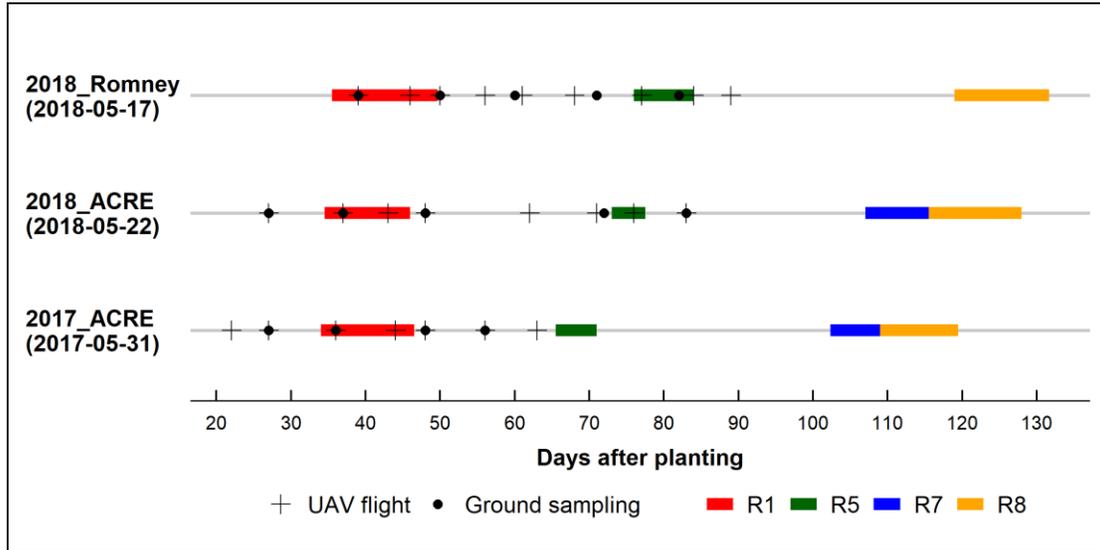


Figure 4.1 Data collection timeline by environment (2017_ACRE, 2018_ACRE, and 2018_Romney). Planting date in parentheses below environment. UAV: unmanned aerial vehicle. Phenological stages (Fehr and Caviness, 1977): R1: beginning bloom; R5: beginning seed; R7: beginning maturity; R8: full maturity.

The SoyNAM founder parents were genotyped by Song et al. (2013) using the SoySNP50K BeadChip resulting in 42,509 segregating SNP markers that were imputed into the SoyNAM RILS using the Williams 82 reference genome (Wm82.a2.v1) bp positions by Diers et al. (2018). For genotypic quality control, we excluded SNPs with minor allele frequency lower than 0.05 and call rate lower than 0.90, resulting in 40,110 SNPs for genome-wide analyses.

4.3.2 High-Throughput Phenotyping

Red, Green and Blue (RGB) and multispectral imagery were collected with fixed-wing SenseFly eBee unmanned aerial vehicle (UAV) equipped with two cameras. RGB imagery was collected by using a S.O.D.A. camera (senseFly Parrot Group, Switzerland). Multispectral imagery was collected with a 1.2 MP Parrot Sequoia camera (MicaSense Inc., Seattle, USA), which

captures four discrete spectral bands: green (wavelength = 550 nm, bandwidth = 40 nm), red (660 nm, 40 nm), red-edge (735 nm, 10 nm), and near-infrared (790 nm, 40 nm). Flights were performed close to solar noon at an altitude of approximately 120 m with both RGB and multispectral cameras. The forward and side overlap for flights were set to at least 85% and 70%, respectively. Ground control points (GCPs) were installed at the corners of the trials and their GPS coordinates were recorded using the TOPCON RTK (Topcon, Tokyo, Japan).

To process the multispectral imagery from this experiment, two pipelines were built in MATLAB: Crop Image Extraction version 2 (CIE 2.0) and Vegetation Indices Derivation version 1 (VID 1.0) (Lyu et al., 2019). The multispectral images were stitched using Pix4Dmapper (Pix4D SA, 2018) to produce a full ortho-mosaic of the experimental area. Individual plots were extracted from the ortho-mosaic using the CIE 2.0. Segmentation was performed to highlight the canopy of the vegetation using Otsu's method (Otsu, 1979). Radiometric calibration was done for every sampling date to remove atmospheric effects and potentially correct for any sensor sensitivity issues (Iqbal et al., 2018). During flight operations, we laid out four reflectance panels reflecting at a specific and consistent percentage of light (12%, 22%, 36%, and 48% reflectance) across the visible and near infra-red spectrum panels, spanning values expected for field crops. A handheld spectrometer ASD FieldSpec[®] 4 (ASD, Boulder, CO, USA) was used to measure the true reflectance of the panels while the multispectral images were collected. We used the reflectance values from the panels, along with radiance values of the panels, extracted from the generated ortho-mosaics, to generate an empirical line using the empirical line method (Smith and Milton, 1999). The empirical line fits the reflectance values measured using the spectrometer, over the reflectance panels, on the y-axis and the digital number values, for the reflectance panels, extracted from the ortho-mosaic on the x-axis for each color band separately. For dates where spectrometer data was unavailable, the expected reflectance values of the reflectance panels, mentioned earlier, were used instead. If saturation was captured in a specific band, the saturated points were removed and were not considered when calculating the slope and offset values. The offset and slope generated by the empirical line are the correction factors needed to correct the radiance values for the plots in the ortho-mosaics and are crucial in producing reflectance data over the plots. Generated values were used to calibrate the images for accurate reflectance and vegetation indices output for the designated rows in the plot experiment.

The reflectance from the calibrated images was used to calculate vegetation indices (VI) using the VID 1.0 pipeline. Several VI can be derived from the four bands of the multispectral images, such as the normalized difference vegetation index (NDVI). Vegetation indices are typically used to estimate crop biomass, and for this study, we selected fourteen VIs (Table 4.1) previously reported in the literature to correlate with crop biomass (Babar et al., 2006; Bendig et al., 2015; Wang et al., 2016; Yue et al., 2017; Sankaran et al., 2018).

From the RGB imagery, we calculated canopy coverage (CC) using the software Progeny® (Progeny Drone Inc., West Lafayette, IN) and the multilayer mosaic approach as described by Hearst (2019). As a result of image analysis, each plot was represented by a list of imagery features, resulting in a matrix $X_{n \times m}$, where n is the number of plants and m is the number of features. The list of the imagery features used in this study is in Table 4.1. All imagery features were calculated in intact and bordered plot rows not used for destructive biomass sampling.

Table 4.1 Descriptions and formulas of imagery features investigated in this study.

Acronym	Feature	Definition	Sensor	Citation
R_{red}	Reflectance of Red band	R_{660}	MSP	--
R_{green}	Reflectance of Green band	R_{550}	MSP	--
$R_{rededge}$	Reflectance of Red Edge band	R_{735}	MSP	--
R_{NIR}	Reflectance of Near Infra-Red band	R_{790}	MSP	--
NDVI	Normalized Difference Vegetation Index	$(R_{NIR} - R_{red}) / (R_{NIR} + R_{red})$	MSP	(Rouse Jr. et al., 1974)
SAVI	Soil-Adjusted Vegetation Index	$(R_{NIR} - R_{red}) / (R_{NIR} + R_{red} + 0.5) * 1.5$	MSP	(Huete, 1988)
MSAVI	Modified Soil-adjusted Vegetation Index	$(2 * R_{NIR} + 1 - (((2 * R_{NIR} + 1)^2 - 8 * (R_{NIR} - R_{red}))^{0.5}) / 2$	MSP	(Qi et al., 1994)

Table 4.1 continued.

GESAVI	Generalized Soil-adjusted Vegetation Index	$((R_{NIR}-1.18) * (R_{red}-0.012))/(R_{red}+0.35)$	MSP	(Gilbert et al., 1998)
GNDVI	Green-NDVI	$(R_{NIR} - R_{green}) / (R_{NIR} + R_{green})$	MSP	(Gitelson et al., 1996)
RVI	Ratio Vegetation Index	(R_{NIR} / R_{red})	MSP	(Jordan, 1969)
MSR	Modified Simple Ratio Index	$((R_{NIR} - R_{red}) - 1) / \sqrt{(R_{NIR} / R_{red} + 1)}$	MSP	(Chen, 1996)
RDVI	Re-normalized Difference Vegetation Index	$(R_{NIR} - R_{red}) / \sqrt{(R_{NIR} + R_{red})}$	MSP	(Roujean and Breon, 1995)
TVI	Transformational Vegetation Index	$\sqrt{(NDVI + 0.5)}$	MSP	(Deering et al., 1975)
GRVI	Green Ratio Vegetation Index	(R_{NIR} / R_{green})	MSP	(Inada, 1985)
EVI2	Enhanced Vegetation Index 2	$2.5 * ((R_{NIR} - R_{red}) / (R_{NIR} + 2.4 * R_{red} + 1))$	MSP	(Jiang et al., 2008)
VARI _{green}	Modified Visible Atmospherically Resistant Index- green	$(R_{green} - R_{red}) / (R_{green} + R_{red})$	MSP	(Gitelson et al., 2002)
VARI _{rededge}	Modified Visible Atmospherically Resistant Index- red edge	$(R_{rededge} - R_{red}) / (R_{rededge} + R_{red})$	MSP	(Gitelson et al., 2002)
REDEDGER	Red edge/green ratio	$(R_{rededge} / R_{green})$	MSP	--
CC	Canopy Coverage	Percentage of Green pixels/Total pixels	RGB	(Hearst, 2019)

RGB: Red, green, blue. MSP: multispectral. Ri: reflectance at band i (nanometer).

4.3.3 Predicting Above-Ground Biomass

In order to predict the AGB for all DAP, including days when ground truth data were not available, we considered a linear model using the imagery features as the predictor variables within each environment across all observed DAP. We observed that the distribution of the residuals was

highly asymmetric, suggesting that a linear model was not suitable to fit the data (Thoni et al., 1990). To correct the asymmetry we considered a Box-Cox transformation on the AGB, which led to the log-transformed values (data not shown, Box and Cox, 1964). The prediction of AGB was carried-out using two different machine-learning methods: Least Absolute Shrinkage and Selection Operator (LASSO) Regression (Tibshirani, 1996), and Partial Least Squares Regression (PLSR) (Wold et al., 2001). For the PLSR, 10 principal components were selected so that the root mean squared error (RMSE) was minimized.

The performance of the predictive models was evaluated using a 10-fold cross-validation strategy, in which the dataset was randomly divided into a training set (90% of the plots) and validation set (10% of the plots). The predictive accuracy of the model was measured by the coefficient of determination (R^2), which is equal to the fraction of AGB variance explained by the model, and by the RMSE, which measures the average error magnitude. Pearson's correlation coefficient (r) was also considered to quantify the linear correlation between the observations and their estimates, being an indication of model prediction ability. Both models were implemented in the R software (R Core Team, 2019), using the package *caret* (Kuhn, 2008).

4.3.4 Random Regression Models

Random regression models were used to model AGB across 27 to 83 DAPs. Seven different models were tested: third-, fourth- and fifth-order Legendre orthogonal polynomials (Kirkpatrick et al., 1990), and linear and quadratic B-splines (de Boor, 1980; Meyer, 2005b) with one (at 55 DAP) or two knots (at 44 and 66 DAPs). The general RRM can be described as:

$$y_{ijk} = Env_k + \sum_{m=1}^m b_m \phi_m(t_{ij}) + \sum_{m=1}^m a_{im} \phi_m(t_{ij}) + e_{ijk},$$

where y_{ijk} is the predicted AGB of the i^{th} RIL on DAP j within environment and replication combination k ; Env_k is the fixed effect of environment and replication combination; b_m is the m fixed regression coefficient for modeling the average curve of the population; a_{im} is the m random regression coefficient that describes the additive genetic effects for the i^{th} RIL; t_{ij} is the time of data collection (DAP j) for the i^{th} line; $\phi_m(t_{ij})$ is a regression function according to DAP j (using Legendre or B-spline polynomials); e_{ijk} is the random residual effect. The number of regression coefficients m varies according to the functions used for random regressions. For the Legendre

orthogonal polynomials, $\phi_m(t_{ij})$ is the m^{th} Legendre orthogonal polynomial coefficient for DAP j (standardized for the -1 to 1 interval) from RIL i . In the case of B-splines, $\phi_m(t_{ij})$ is the m^{th} interval given the previously mentioned knots associated with DAP from RIL i . According to Meyer (2005b), the basis function of degree $p=0$ have values of unity for all points in a given interval (t) and zero otherwise. For the m^{th} interval given by knots T_m and T_{m+1} , with $T_m \leq t < T_{m+1}$, $\phi_m(t_{ij}) = \begin{cases} 1, & \text{if } T_m \leq t < T_{m+1} \\ 0, & \text{otherwise} \end{cases}$. Basis function for $p > 0$, can be represented by $\phi_{m,p}(t_{ij}) = \left(\frac{t-T_m}{T_{m+p}-T_m} \right) \phi_{m,p-1}(t_{ij}) + \left(\frac{T_{m+p+1}-t}{T_{m+p+1}-T_{m+1}} \right) \phi_{m+1,p-1}(t_{ij})$. The individual segments were either linear or quadratic, with degree $p = 1$ or 2 , respectively. The joined knots allow the function to become continuous.

The models' assumptions are:

$$\text{var} \begin{bmatrix} \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} \otimes \mathbf{G}_0 & 0 \\ 0 & \mathbf{I} \otimes \mathbf{R} \end{bmatrix},$$

where \mathbf{G}_0 is the (co)variance matrix of the genomic random regression coefficients, \mathbf{G} is a genomic relationship matrix, \mathbf{I} is an identity matrix, \mathbf{R} represents a matrix containing residual variances, and \otimes is the Kronecker product between matrices. The \mathbf{G} matrix was calculated using the method presented by VanRaden (2008). The residual variances were allowed to be either homogeneous or heterogeneous. We defined a different residual variance for each of the eighteen DAP with AGB phenotypic data and grouped the remaining days based on their proximity to those DAP. The eighteen heterogeneous residual variances classes are as follow: 27-33, 34-36, 37, 38-41, 42-43, 44-45, 46, 47-49, 50-53, 54-58, 59-61, 62, 63-65, 66-71, 72-74, 75-76, 77-80, 81-82, and 83.

The AIREMLF90 and BLUPF90 software from the BLUPF90 family (Miszta et al., 2002) were used to estimate the variance components and the solutions of the mixed model equations, respectively. The BLUPF90 family programs perform by default the single-step GBLUP (Miszta et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010); however, as all RILs were genotyped, the program was adapted to perform the traditional GBLUP (VanRaden, 2008), by using a dummy pedigree file. Akaike's information criterion (AIC; Akaike, 1974) was used to compare the models' performance, in which models with lower AIC values were preferred.

4.3.5 Genetic Parameters

The genetic (co)variance matrix (Σ) for all DAP within the interval of AGB collection was obtained as (Oliveira et al., 2019a):

$$\Sigma = \mathbf{TGT}',$$

where \mathbf{T} is a matrix of covariates associated with the function assumed for RIL i , and \mathbf{G} is the genetic (co)variance matrix for the coefficients. The narrow-sense heritability for each DAP (h_j^2) was obtained as:

$$h_j^2 = \frac{\hat{\sigma}_{a_j}^2}{\hat{\sigma}_{a_j}^2 + \hat{\sigma}_e^2},$$

where $\hat{\sigma}_{a_j}^2$ is the additive genetic variance for DAP j , and $\hat{\sigma}_e^2$ is the residual variance, which depends on the residual variance classes previously mentioned (when using the heterogeneity of residual variance). The genetic correlation between different DAP ($r_{j,j'}$) was obtained as:

$$r_{j,j'} = \frac{\hat{\sigma}_{a_{j,j'}}}{\sqrt{(\hat{\sigma}_{a_j}^2 + \hat{\sigma}_{a_{j'}}^2)}},$$

where $\hat{\sigma}_{a_{j,j'}}$ is the genetic covariance between the DAP j and j' , and $\hat{\sigma}_{a_j}^2$ and $\hat{\sigma}_{a_{j'}}^2$ are the additive genetic variances for DAP j and j' , respectively. The vector of genomic estimated breeding values (\mathbf{GEBV}_i) for all DAP of RIL i was obtained as (Oliveira et al., 2019a):

$$\widehat{\mathbf{GEBV}}_i = \mathbf{T}\hat{\mathbf{g}}_i,$$

where $\hat{\mathbf{g}}_i$ is the vector of predicted genomic values for the coefficients, for each RIL i , and \mathbf{T} is a matrix of covariates associated with the assumed function.

4.3.6 Genomic Prediction of Breeding Values

The performance of the genomic prediction of breeding values for AGB was investigated using a five-fold cross-validation (CV) scheme. Briefly, all RILs were randomly separated into five equal-sized groups, where one group was retained as validation, and four groups were used as training. This procedure was repeated five times, with a unique group used exactly once as the validation set. Variance components and SNP marker effects were estimated based on the training set and used to predict GEBV in the validation set (reduced data). The prediction accuracy was measured using the Pearson's correlation coefficient (r) estimated between the GEBV predicted using the full data (i.e., data including all training and validation RIL) and the reduced data, only

for the validation RIL. To evaluate the genomic prediction bias, regression coefficients (b_l) were estimated using linear regression of the GEBV estimated based on the full dataset on the GEBV estimated based on the reduced dataset from each CV fold ($GEBV_{full} = b_0 + b_1 * GEBV_{reduced}$). Finally, prediction bias (b_l) was calculated as the average of CV folds for each DAP.

4.3.7 Genome-Wide Association Study

For the genome-wide association study (GWAS), SNP effects were derived from GEBVs for each additive random regression coefficient using the POSTGSF90 software (Aguilar et al., 2014). The prediction of SNP effects ($\hat{\mathbf{u}}_m$) for the m^{th} random regression coefficient was calculated as (Wang et al., 2012):

$$\hat{\mathbf{u}}_m = \mathbf{D}\mathbf{Z}'(\mathbf{Z}\mathbf{D}\mathbf{Z})^{-1}\widehat{\mathbf{GEBV}}_m$$

where \mathbf{D} is a diagonal matrix of weights accounting for variances of SNPs markers (assumed as an identity matrix in this study), \mathbf{Z} is a matrix relating genotypes of each locus, and $\widehat{\mathbf{GEBV}}_m$ is the vector of GEBV for the m^{th} random regression coefficient. Finally, the SNP effects for all DAP were obtained as (Oliveira et al., 2019c):

$$\widehat{\mathbf{SNP}}_s = \mathbf{T}\hat{\mathbf{u}}_s,$$

where $\widehat{\mathbf{SNP}}_s$ is the vector that contains the SNP effects estimated for each DAP of the s^{th} SNP, $\hat{\mathbf{u}}_s$ is the vector of SNP solutions for all random regression coefficients related to the s^{th} SNP, and \mathbf{T} is a matrix of covariates associated with the assumed function.

The SNPs were selected to be further investigated based on the magnitude of their effects, as suggested by Oliveira et al. (2019d). In this context, the top 10 SNPs that showed the highest magnitude of SNP effect in each DAP were selected as relevant SNPs. The exploration of candidate genes was carried out in the range of ± 25 kb from the location of the selected SNP. Potential candidate genes and their associated functional annotation were determined using the genomic position and gene models based on Glyma.Wm82.a2.v1 genome in the soybean database SoyBase (SoyBase.org).

4.4 Results

4.4.1 Predicting Above-Ground Biomass

We used two methods to quantify the ability of image-based features to statistically predict the AGB in soybean: LASSO regression and PLSR. Both methods were evaluated using a 10-fold CV method. Figure 4.2 shows the statistical distributions of R^2 and RMSE values for each CV fold, in each environment. In general, similar performance was observed for both methods, in all environments. It was found that LASSO and PLSR had the same R^2 averages for 2017_ACRE (0.94), 2018_ACRE (0.92), and 2018_Romney (0.94). However, the PLSR presented a smaller RMSE average for 2017_ACRE (0.23 vs 0.24 for PLSR and LASSO, respectively), and LASSO presented a smaller RMSE average for 2018_ACRE (0.28 and 0.29 for LASSO and PLSR, respectively). Both models presented the same RMSE average for 2018_ACRE (0.24).

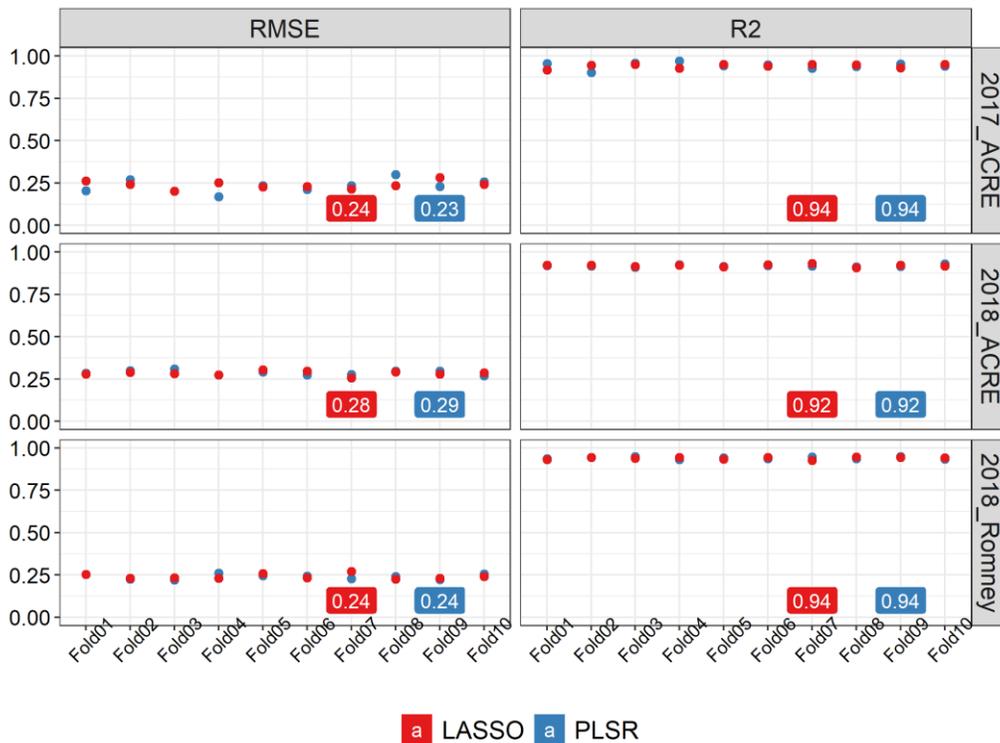


Figure 4.2 Performance of above-ground biomass prediction for each environment. Predictions were performed using the least absolute shrinkage and selection operator (LASSO) regression, and the partial least squares regression (PLSR) methods. The performance of predictions was evaluated using the root mean squared error (RMSE) and coefficient of determination (R^2), using a 10-fold cross-validation set.

The correlation between AGB predicted from UAV-based imagery and observed from ground samples was high ($r \geq 0.91$) in all environments for both methods, implying that the methods captured the relationship among image-based features and AGB (Appendix B Fig. B.2). Based on these findings, and because it makes a simpler and more direct connection between the response and predictor variables, the LASSO method was chosen to predict AGB for all plots of the two full replications on all flight dates in this study. Figure B.2 shows the relative importance of each predictor variable for the LASSO method, which indicates that the model utilized information from different predictor variables for each environment. In addition, we performed a CV leaving one environment out to assess the models' ability to predict AGB for a new environment. In this scenario, the performance of both methods declined greatly (Appendix B Fig. B.4). The phenotypic distribution of the predicted AGB across environments, by DAP, is presented in Figure 4.3. Figure B.5 (Appendix B) shows the distributions within each environment.

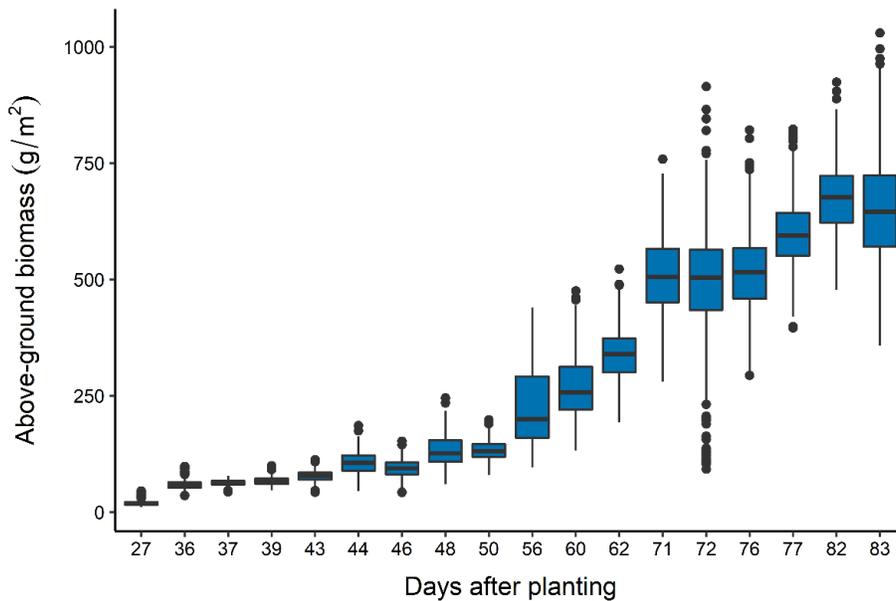


Figure 4.3 Phenotypic distributions of predicted above-ground biomass (g/m^2) across environments by days after planting using the Least Absolute Shrinkage and Selection Operator (LASSO) Regression. Horizontal lines in the box indicate the median.

4.4.2 Random Regression Models

Table 4.2 shows the AIC values calculated for all seven RRM using both homogeneous and heterogeneous residual variance. All models with homogeneity of residual variance showed worse fit compared to models with the heterogeneity of residual variance. The models using quadratic B-spline with 1 knot and homogeneous residual, fifth-order Legendre polynomial and quadratic B-spline with 2 knots and heterogeneous residual variance failed to converge. The third-order Legendre polynomial considering homogeneity of residual variance had the highest AIC, which suggests that this model is worse than the others to fit the analyzed data. The best model was using linear B-spline with 2 knots and heterogeneous residual variance and it was selected for subsequent genome-wide analyses.

Table 4.2 Akaike's information criterion (AIC) values calculated for the random regression models tested using homogenous and heterogeneous residual variances.

Model ^a	AIC	
	Homogeneous residual variance	Heterogeneous residual variance
Third-order Legendre Polynomial	165,795.94	149,108.17
Fourth-order Legendre Polynomial	164,767.63	148,437.23
Fifth-order Legendre Polynomial	163,865.17	-
Linear B-spline 1 knot	165,461.99	148,468.87
Linear B-spline 2 knots	164,922.42	147,386.23
Quadratic B-spline 1 knot	-	148,803.59
Quadratic B-spline 2 knots	163,851.66	-

^a Random regression models with respective Legendre Polynomial or B-spline for the fixed curve and for the additive genetic effect. - Model did not achieve the convergence criterium of 10^{-12} .

4.4.3 Genetic Parameters

The genetic architecture of AGB was assessed by estimating the narrow-sense heritabilities (h^2) across the 57 analyzed days (from 27 to 83 DAP; Fig. 4.4). Narrow-sense heritability estimates for AGB were low to moderate and varied over time (ranging from 0.02 at 44 DAP to 0.28 at 33 DAP). The genetic correlation between AGB on different DAP was also estimated, and it is

showed in Figure 4.5. Adjacent DAP showed the highest genetic correlations, while those further apart exhibited lower correlations. For instance, the lowest genetic correlation between 27 and 83 DAP was 0.16 and the highest genetic correlation between 48 to 50 DAP was 1.00.

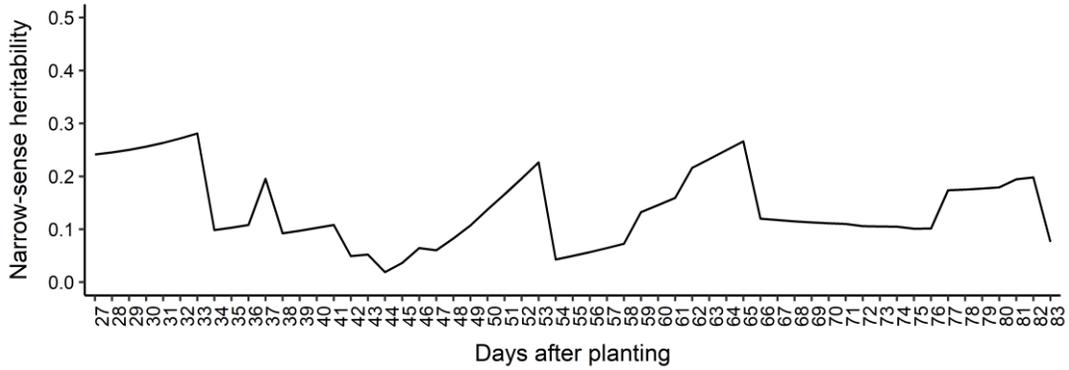


Figure 4.4 Narrow-sense heritability estimated for each day after planting.

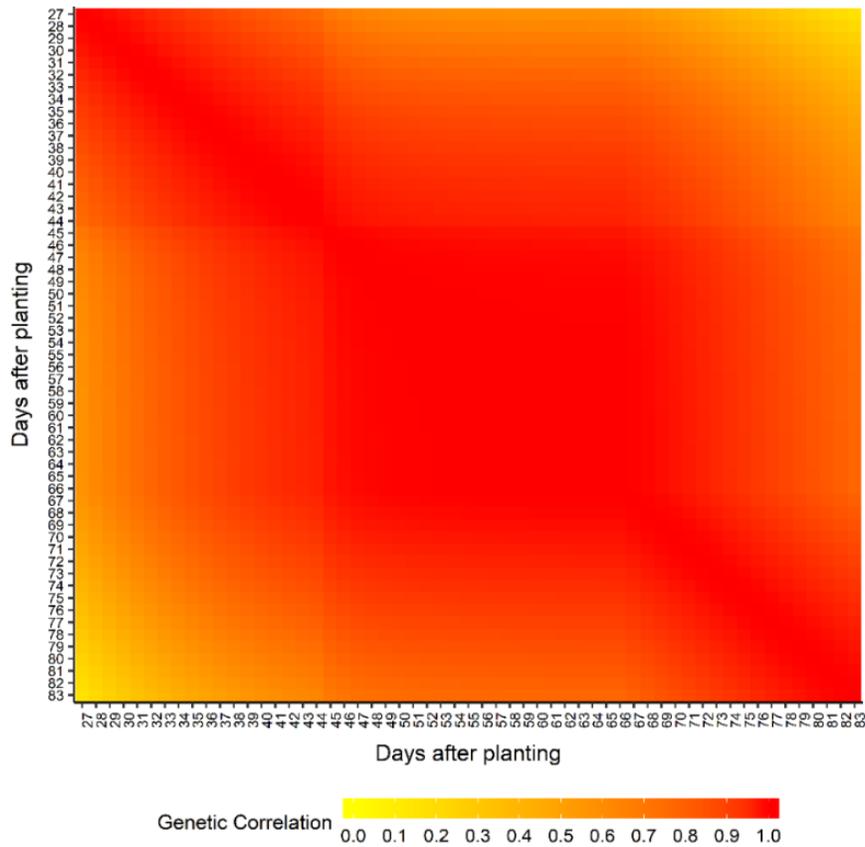


Figure 4.5 Estimated genetic correlations of above-ground biomass between different days after planting.

4.4.4 Genomic Prediction of Breeding Values

The genomic prediction accuracy for AGB over time is presented in Figure 4.6. Overall, the prediction accuracies were high considering the heritabilities estimated across all DAP, ranging from 0.21 at 83 DAP to 0.55 at 27 DAP. A decreasing trend was observed in prediction accuracy over time, indicating that it is more difficult to predict AGB for latter DAPs compared to early DAPs. From 27 DAP to 44 DAP the prediction accuracy steadily decreased, reaching a slight plateau between 44 to 66 DAP, and decreased again until the end of the surveyed time. These findings suggest that longitudinal phenotypes can be accurately predicted using RRM.

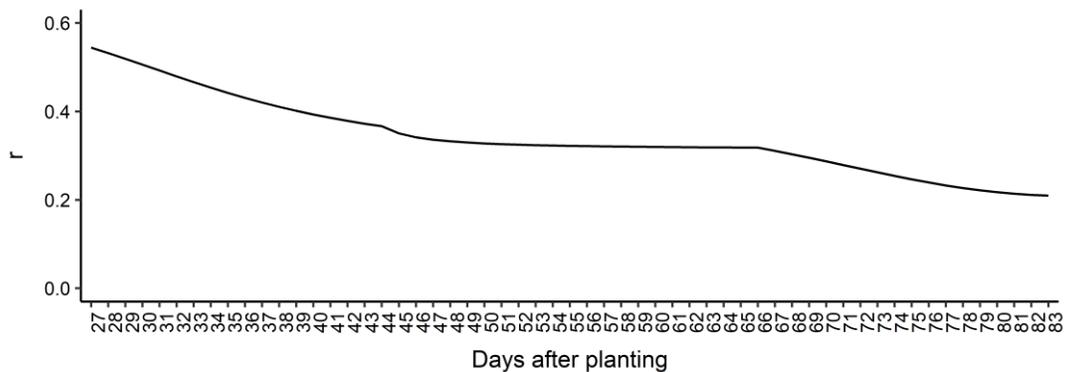


Figure 4.6 Genomic prediction accuracy based on Pearson's correlation coefficient (r) for each day after planting.

Regression coefficients' patterns were used to assess the bias of GEBV over DAP (Fig. 4.7). Overall, regression coefficients closer to 1.0 were found in earlier DAP. The most biased estimates with regression coefficients deviating from 1.0 were observed towards the end of the surveyed time.

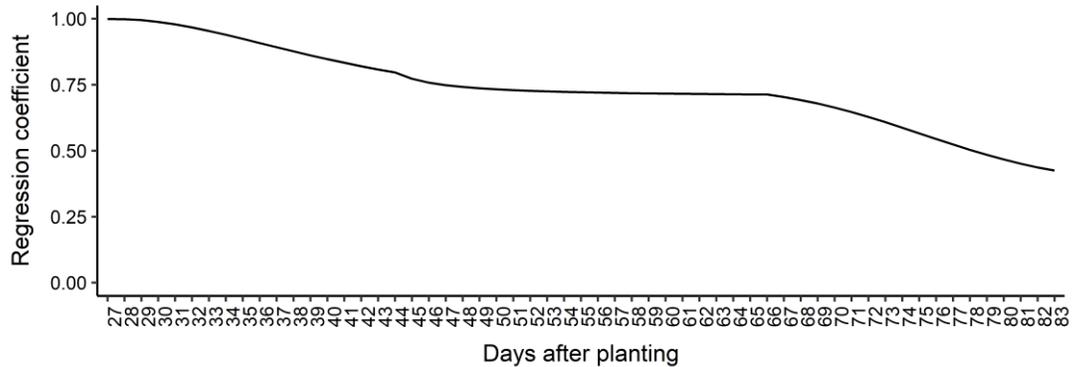


Figure 4.7 Regression coefficients' patterns of genomic estimated breeding values for each day after planting.

4.4.5 Genome-Wide Association Study

Thirty unique SNPs (i.e., 10 top SNPs for each DAP) were selected as the most relevant SNPs for AGB. Figure 4.8 shows the chromosome number, position, period of occurrence, and the SNP effects for selected SNPs. None of the SNPs selected were significant across all time points. In general, the magnitude of effects over time increased for most of the selected SNPs. According to the duration of the SNP effect across all 57 predicted days, the selected SNPs were classified as long-duration (they were considered as important SNPs for more than 30 consecutive days), mid-duration (they were considered as important SNPs for more than 10 consecutive days but less than 30), short-duration (they were considered as important SNPs for less than 10 consecutive days), and intermittent (they were considered as important SNPs on non-consecutive days) (Fig. 4.8). These SNP classes were nearly evenly distributed as long- (9 SNPs), mid- (8 SNPs) and short-duration (9 SNPs). The intermittent category had the lowest number of relevant SNPs (4 SNPs). The majority of mid-duration SNPs were detected towards the beginning of the DAP. Interestingly, the SNPs classified in the short- and mid-duration categories were found either towards the beginning or end of the studied time period.

Table 4.3 Selected single nucleotide polymorphisms (SNPs) associated with above-ground biomass mapped inside potential candidate genes in the soybean genome.

Duration Category	SNP	Chr.	Pos. (bp)	Selected candidate genes	Annotation Description
Long	2:5777782	2	5,777,782	<i>Glyma.02g064500</i>	rhombooid protein-related
				<i>Glyma.02g064600</i>	agenet domain-containing protein
Short	3:5150181	3	5,150,181	<i>Glyma.03g040800</i>	regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain
Long	7:6108702	7	6,108,702	<i>Glyma.07g067900</i>	disease resistance protein (TIR-NBS-LRR class), putative
Mid	7:6523718	7	6,523,718	<i>Glyma.07g071800</i>	cytidine/deoxycytidylate deaminase family protein
Short	7:15340513	7	15,340,513	<i>Glyma.07g128300</i>	-
Long	13:24980935	13	24,980,935	<i>Glyma.13g137200</i>	ROP interactive partner 3
Short	15:36306421	15	36,306,421	<i>Glyma.15g217500</i>	CTP synthase family protein
Long	16:4353954	16	4,353,954	<i>Glyma.16g046000</i>	DEAD/DEAH box helicase, putative

Chr: Chromosome, Pos (bp): position in base pair. - Annotation not available.

4.5 Discussion

4.5.1 High-Throughput Phenotyping of Soybean Above-Ground Biomass

Besides being an important yield component, plant biomass is a foundation for unraveling several complex processes of plant growth, development and environmental response (De Bruin and Pedersen, 2009; Koester et al., 2014; Balboa et al., 2018; Jumrani and Bhatia, 2018). The capacity to non-destructively estimate soybean AGB enables capturing these data in a temporal fashion leading to insights about AGB dynamics. Many different techniques and HTPP have been used to estimate AGB in different crops (Bendig et al., 2015; Wang et al., 2016; Zhang et al., 2017; Jimenez-Berni et al., 2018; Maimaitijiang et al., 2019). In this study, we compared two methods, LASSO regression and PLSR, combining nineteen imagery features (Table 4.1) captured with UAV to predict soybean AGB. Both approaches have been used extensively to generate predictive

models for HTP data (Montes et al., 2011; Bratsch et al., 2017; Wang et al., 2017; Vasseur et al., 2018; Fu et al., 2019). Our results showed that both methods presented similar performances in all environments (Fig. 4.2).

The LASSO regression is especially attractive because it shrinks regression coefficients, i.e., some coefficients can become zero and, therefore, some variables are eliminated from the model (Tibshirani, 1996). Such approach reduces overfitting and thus improves the prediction efficiency. The main reason to choose the LASSO regression over the PLSR for further analyses in this study is that it makes a simpler and more direct connection between the response and predictor variables (Cui and Wang, 2016). This advantage is important because it allows the direct inspection of the variables used to predict AGB, offering insights into the underlying processes that give rise to the observed AGB. When assessing the importance of the individual variables from the LASSO regression (Appendix B Fig. B3), we observed that this method used information from different predictor variables for each environment. For example, the relative importance of CC was higher for 2018_ACRE and 2018_Romney than 2017_ACRE. On the other hand, NDVI was only included in the model to predict AGB at 2017-ACRE. This is also supported by the results of the CV leaving one environment out which indicates that new environments could not be predicted accurately (Appendix B Fig. B4). These results provided a solid basis for constructing different models for each environment to enhance the strengths of each imagery feature by the environment.

To our best knowledge, this is the first study estimating soybean AGB in a plant breeding program, using experimental plots and diverse genotypes. Satellite-derived vegetation indices were used separately to predict soybean AGB reaching good predictive abilities (Kross et al., 2015; Richetti et al., 2019). However, both studies were performed in agricultural fields with no genetic variation. Recently, Maimaitijiang et al. (2019) used UAV-based RGB imagery-derived spectral, structural, volumetric information, and their combination in several methods to predict AGB in agriculture fields with three cultivars. Their best method yielded an R^2 of 0.91 and a relative RMSE of 14.1% using PLSR with several spectral indices and structural metrics combined. While they included different cultivars compared to the previous studies, it still might not represent the expected genetic diversity found in breeding programs. Also, the mentioned studies used extensive agronomic areas, which can lead to limitations of their results in small experimental plots equivalent to breeding programs.

In agreement with previous studies (Bendig et al., 2015; Chen et al., 2016; Wang et al., 2017; Maimaitijiang et al., 2019; Li et al., 2020), combining different image metrics resulted in models with high prediction performance for AGB estimation. The individual and yet complementary information between spectral and structural metrics can be key to enhance prediction methods and optimize the use of information from HTPP. We anticipate that our method could be further improved when new types of image metrics become available.

4.5.2 Genetic Architecture of Soybean Temporal Above-Ground Biomass

The identification of the genetic causes underlying phenotypic variation is a major step towards crop improvement. By implementing an HTPP that is capable of collecting non-destructive data in large populations throughout the season under actual field conditions, researchers and plant breeders are able to quantify and understand more thoroughly the dynamics of temporal variation of traits and thereby better optimize genotypes through selection in breeding programs (Pauli et al., 2016a). It is important to note that the effort and investment in HTPP demand equal effort to analyze the data properly. Nevertheless, the improvement of statistical methodologies to analyze image-based longitudinal phenotypes has not kept pace with the ability to generate high-throughput phenotypic data (Momen et al., 2019). Most of the studies using longitudinal traits mainly performed statistical genetic or genomic analysis for each time point independently (Würschum et al., 2014; Pauli et al., 2016a; Xavier et al., 2017; Zhang et al., 2017; Wang et al., 2019; Knoch et al., 2020), ignoring the existing temporal genetic correlation and dependency during trait development. RRM are deemed the most effective alternative to genetically evaluate longitudinal traits in numerous livestock breeding programs (Oliveira et al., 2019a). This approach uses the covariance between each time point with no assumptions of constant variances or correlations, resulting in more accurate breeding values compared to other methods (Sun et al., 2017a; Oliveira et al., 2019a). We combined HTP data, high-density genomic information and RRM to carry out longitudinal analysis and understand the genetics of the development of AGB in soybean. In this context, this study provides the first application of RRM for genomic analysis of longitudinal traits in soybean, as well as the first genetic study on soybean AGB.

Among the RRM tested here, the model using quadratic B-spline with 1 knot and homogeneous residual variance failed to converge, which indicates that this model did not fit the

data well. The models using fifth-order Legendre polynomial and quadratic B-spline with 2 knots also did not achieve the convergence when heterogeneous residual variance was used, probably because of the higher complexity of the models (i.e., they are more parameterized) and the data set size. Usually, more parametrized models require a higher number of observations to accurately estimate their parameters (Thoni et al., 1990). As the number of parameters increases, problems with convergence and estimation, as well as an increase in computational demand, can be expected. The model that seemed to be the most suitable to fit the data was the model using linear B-spline with 2 knots and heterogeneous residual variance. Hence, this model was selected to describe the genetic architecture of AGB over time in subsequent analyses.

As shown in Figure 4.4, heritability for AGB fluctuates over different DAP. Some of this variation is expected due to differential growth patterns across development, and it indicates that the proportion of genetic variation also changes across different DAP. Similar temporal variability has been reported for water usage heritabilities in rice using RRM (Baba et al., 2020). The residual variance structure applied in this study can also play a role in the heritability fluctuations. Days without phenotypic data were grouped with the closest DAP with phenotypic data, which may not reflect their true residual variance. However, all models with the heterogeneity of residual variance structure outperformed the models with homogeneous residual variance, agreeing with other studies (Brito et al., 2017a; Campbell et al., 2018). The residual variance is affected by many factors that can vary over DAP. For instance, as the plants grow, the scale of AGB phenotypes increases dramatically. Thus, it is crucial to assess the need of a heterogeneous residual variances structure over time points, since there can be improvements in the partition of the total variation, yielding better estimates of genetic parameters (Brito et al., 2017a). In this context, it is important to emphasize that this approach is often performed in studies using RRM (Brito et al., 2017a; Campbell et al., 2018).

Campbell et al. (2018) found heritabilities ranging from 0.60 to 0.77 for high-throughput phenotyped shoot biomass (projected shoot area, PSA) in rice using RRM. However, their study was performed in a greenhouse, with controlled environmental conditions across the entire phenotypic recording period. Additional studies in controlled-environments that carried out an independent analysis for each time point also found high broad-sense heritabilities for AGB in barley (Neumann et al., 2017), maize (Muraya et al., 2017), and canola (Knoch et al., 2020). This is probably a result of strict environmental control over the entire growing period. In the field,

environmental conditions are constantly fluctuating during the growing season, therefore, highly impacting the observed trait. Regarding the genetic correlation across days, Campbell et al. (2018) and Baba et al. (2020) found a similar trend, where the highest correlations were observed between adjacent time points.

For longitudinal traits, such as AGB, genetic effects are expected to vary over time. In fact, several studies have shown that the additive polygenic effects of longitudinal traits are not constant across the entire period (Brito et al., 2017b; Oliveira et al., 2019a). The analysis of longitudinal traits as a function of time enables the detection of time-specific significant SNPs. The RRM approach has improved statistical power to detect loci associated with longitudinal traits over other methods, as they exploit the entire collection of experimental raw phenotypes capturing the genetic differences that happened throughout the analyzed time (Ning et al., 2017; Oliveira et al., 2019a). In this study, time was introduced as an additional dimension to association studies enabling the observation of the effects of individual markers over 57 days of soybean AGB development from late vegetative up to mid reproductive stages between 27 and 83 DAP. Based on the top SNPs for each DAP, 30 unique SNPs were selected as most important for AGB across DAP and divided into four categories according to the length of the period they were detected (Fig. 4.8). The SNP effects were generally small and time-specific, and no SNPs were constant or continuous over time. These findings suggest a temporal pattern and that AGB is regulated mainly by small effect loci and their potential interactions. This highlights the importance of the temporal assessment of traits, as many associations could not have been discovered if AGB had been evaluated at the end of the experiment or at one time. Previous studies have explored the dynamic genetic architecture of AGB in other crops (Campbell et al., 2017, 2019; Muraya et al., 2017; Knoch et al., 2020), but none in field scale and with such high temporal resolution. Campbell et al. (2017) used power function parameters as the pseudo-phenotypes in a multiple-trait GWAS to study AGB in rice during early and active tillering stages. Using RRM, several loci with both transient and persistent effects were found controlling rice AGB during early vegetative development in a green-house (Campbell et al., 2019). Knoch et al. (2020) used time point data and relative growth rates for AGB GWAS of canola under controlled environmental conditions. The authors found that canola AGB is controlled by several medium and many small effect loci, most of which act only during short periods.

Among the selected SNPs positioned within candidate genes in soybean (Table 4.3), some may have a direct impact on AGB. The *Glyma.02g064600* candidate gene potentially codes a protein belonging to the Agenet domain family, which is known as chromatin remodeling proteins (Brasil et al., 2015). In *Arabidopsis thaliana*, Agenet/Tudor domain family proteins were associated with regulating gene expression by DNA methylation (Brasil et al., 2015; Zhang et al., 2018a). Interestingly, an Agenet domain-containing protein in *Arabidopsis* was highly expressed in reproductive tissues and its down-regulation delayed flower development timing (Brasil et al., 2015). In our study, the effect of the SNP associated with *Glyma.02g064600* started to be present at 43 DAP, which overlaps with the average beginning of the blooming (R1) period, and the magnitude of its effects increase with time. Also, in chromosome two, *Glyma.02g064500* possibly corresponds to rhomboid protein-related that in *Arabidopsis* is a putative cellular component in the Golgi apparatus with unknown function. Ban et al. (2019) reported that *Glyma.07g067900*, which codes a disease resistance protein, was upregulated when studying the regulation of genes in mutant dwarf soybeans related to plant growth. It is known that the over-expression of disease resistance and other immune-responsive genes tend to divert resources to generate protection metabolites, thus reducing overall growth (Ban et al., 2019). *Glyma.07g071800* is predicted to have biological functions involved in the riboflavin biosynthetic process. In plants, Riboflavin is known to be involved in disease defense (Nie and Xu, 2016), therefore *Glyma.07g071800* may be associated with the tradeoff between the defense response and plant growth as mentioned before. *Glyma.16g046000* is a putative DEAD/DEAH box helicase. Some proteins of this family are known to play a role in plant growth and development, and in response to stresses in plants (Wang et al., 2000; Zhu et al., 2015).

4.5.3 Potential of Genomic Selection to Improve Soybean Temporal Above-Ground Biomass

Genomic selection has been proved to be a powerful tool in plant and livestock breeding (Meuwissen et al., 2016; Crossa et al., 2017). It enables the rapid, cost-effective and more accurate selection of superior genotypes without phenotypes, accelerating breeding cycles and optimizing genetic gain over traditional approaches. Genomic selection requires accurate phenotyping to train prediction models that can predict the performance of new unphenotyped individuals based on their SNP markers (Meuwissen et al., 2001). HTPP allow crop scientists to generate high-quality

quantitative data and effectively characterize large training populations throughout the growing season. Thus, the combination of GS and HTPP have the potential to increase accuracy and throughput, while reducing costs and minimizing labor (Araus et al., 2018). Although several studies have already proven that RRM improved genomic prediction accuracy of longitudinal traits compared to single-time point and MTM in animals (Oliveira et al., 2019a), it was only demonstrated in plants more recently (Campbell et al., 2018; Momen et al., 2019). Here, the utility of RRM-based genomic selection for longitudinal soybean AGB was evaluated.

Using CV, we found that it was possible to model longitudinal AGB under a RRM approach (Fig. 4.6). As expected, prediction accuracy varied across DAP, with a decreasing trend over time. Accuracy of GS is dependent on many factors, such as the level of linkage disequilibrium (LD) in the population, effective population size, the number of markers, trait heritability, and the number of QTL influencing the trait (Lin et al., 2014; Wang et al., 2018). Since the LD, population size and number of markers were held constant in our study, the difference in prediction accuracy across DAP can be largely attributed to the differences in heritability. Considering the heritability values, in general, we obtained better prediction accuracy than predicted AGB in rice using RRM (Campbell et al., 2018). As well as the prediction accuracy, prediction bias for the GEBVs varied over DAP, suggesting that selection based on different days may have different results and usefulness (Fig. 4.7). This is in agreement with our GWAS results because it implies that different genes can be expressed by DAP and that selection based on different days can have distinct genetic implications on AGB (Oliveira et al., 2019b). One possible reason for the decreasing trend over time of prediction accuracy and bias can be the quality of the phenotypes as the season progress, because as the plot canopy closes it is harder to detect differences among plots. Enhancing imagery resolution, decreasing flight altitude, and adding volume and height metrics can potentially help eliminate this problem. Another reason may be our limited population size (383 genotypes). Increasing population and training set size can improve the accuracy of prediction, especially for traits with low heritability, because more samples with phenotypic and genotypic information will be used to estimate genetic effects (Goddard, 2009; Wang et al., 2018). Xavier et al. (2016) found that training population size was the most relevant factor in improving prediction accuracy in the SoyNAM population, with optimal populations size between 1000 and 2000 individuals. The rate of accuracy improvement decreased rapidly for

training populations above 2000 individuals. Thus, probably the accuracy and bias of our predictions would improve as we increase the training set size.

Even though we used a dummy pedigree matrix in the current study because all RILs were genotyped, the BLUPF90 family programs perform the ssGBLUP random regression. The ssGBLUP simultaneously combines phenotypic records, pedigree, and genomic information in the single-step genomic evaluation (Miszta et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). Although genomic information became more accessible over the years, in practice high density genotyping for all genotypes in a breeding program is not always feasible due to genotype costs, logistics or both. An alternative to improve genomic prediction performance is to construct a joint matrix based on pedigree and genomic relationships to predict BLUPs for genotyped and non-genotyped material (Legarra et al., 2009). Thus, the use of ssGBLUP based on RRM can lead to more accurate and less biased GEBV and increased the reliability of genomic predictions (Koivula et al., 2015; Kang et al., 2018; Oliveira et al., 2019d). Considering the benefits of ssGBLUP and RRM, incorporating both approaches is an effective strategy to enhance the genomic prediction of longitudinal traits in crops. Future research using genomic and pedigree information is necessary to assess the value of random regression ssGBLUP for predicting breeding values in soybean.

In summary, our results suggest that AGB is a potential candidate for genomic selection in soybeans. The ability to predict temporal-based GEBV allows targeting important periods in the growing season for selection. For example, using genomic selection to increase AGB during vegetative stages may lead to yield boosts or improved stress resilience. A second point of interest can be modifying the AGB trajectory along the growing season by selecting plants with a specific growth pattern. Moreover, even if the temporal characteristic itself may not be a target of selection, such data can be used to enhance the genomic selection of economical endpoint traits such as yield. Given HTPP's power to simultaneously collect multiple temporal traits, multiple-trait RRM may be powerful tools for joint genomic prediction of multiple longitudinal traits (Oliveira et al., 2016; Baba et al., 2020).

4.6 Conclusions

This study demonstrates the potential of HTPP to replace destructive ground measurements of soybean AGB over time. With these temporal phenotypes, RRM were used to leverage the

covariance across time points and to uncover the quantitative properties of AGB. Our results suggest evidence of differential sets of SNPs and respective candidate genes underlying the phenotypic variation of AGB over time. These findings highlight the importance of comprehensive time-resolved analyses to effectively unravel the dynamic contribution of genes influencing growth and developmental process. As in other association studies, additional experimental validation will be required to confirm if the identified loci are truly associated with AGB and to identify functional linked variants. Moreover, additional research on temporal growth and gene expression is needed to determine whether the candidate genes play a time-specific or a general role in soybean AGB regulation. Moreover, we demonstrated that AGB in soybeans is a good potential candidate for genomic selection. This research illustrated the promise of new emerging technologies and offer a basis for future studies to combine phenotyping and genomic analysis in order to understand the genetic architecture of complex longitudinal traits.

CHAPTER 5. GENERAL CONCLUSIONS

The main objective of plant breeding is to genetically improve the performance of cultivars in the most efficient way possible. One of the most important tasks of a plant breeder is to evaluate investment goals that can promote the desirable genetic gain for the traits of interest within their breeding program. The rapid development of remote sensing in agronomic research has contributed to a significant shift in plant sciences towards time-series phenotyping that can track the development of plants through their life stages, generating an opportunity to better understand the genetic basis underlying important traits over time. However, these new technologies also bring new challenges to adequately use high-dimensional phenomic data, through the integration of quantitative genetics, physiology, breeding, modeling, and engineering disciplines. The major challenge is synthesizing the various layers of information together in a meaningful manner to understand the many complexities of plant growth and developmental and implications for breeding. In this dissertation, we explored interdisciplinary frameworks to assess different applications of high-throughput phenotyping for longitudinal traits in plant breeding.

In chapter 2, we have shown that ACC measured using an aerial HTPP several times during the growing season can be used for selection, alone or in combination with yield, in the early stages of soybean breeding pipelines. Soybean progeny rows selection using yield given ACC (Yield|ACC) selected the most top-ranking lines in advanced yield trials. We observed considerable differences in canopy coverage development of the checks in 2015 and 2016 among which may have impacted some of our results. Thus, in order to provide tailored recommendations on the use of ACC for selecting high yielding lines in different scenarios, further studies are required to determine environmental effects on canopy coverage phenotypic variations. Furthermore, we suggest evaluating various scenarios to confirm that the greater response is either using the ACC alone or in combination with yield.

Chapter 3 is a review paper, where we outlined the current analytical approaches in phenotyping platforms, modeling and quantitative genetics that can be applied to high-throughput phenotyping quantified over time. We explore interdisciplinary frameworks, such as cutting-edge animal breeding models, that have important applications to crops longitudinal traits. Also, we described the advantages and pitfalls of each method and indicate future research directions and opportunities. We provided a comprehensive framework for researches to study how genetic

variability changes over time and impact the selection of crops with adaptation to future climate change environments as well as increased production efficiency.

Finally, we developed a whole framework to study soybean AGB in chapter 4. We used imagery features from RGB and multispectral cameras mounted on a UAS platform to predict AGB over time. Although this research obtained good prediction for AGB in soybean, additional research evaluating other sensors and metrics, including but not limited to volume, height, and reflectance of blue band, need to be considered to create a robust model that can be used in any environment. In addition, our research determined AGB during the vegetative and early reproductive stages, but phenotyping needs to be extended to phenological events occurring late in the season to better understand the contribution of AGB to the final yield.

With the remote estimate AGB we implemented RRM to study the genetic architecture of the trait. Genomic evaluations for various points in time can be simultaneously done using RRM, which enables the understanding of the complete pattern of AGB accumulation during the growing season. Here we tested several covariance structures to obtain adequate and parsimonious models for the estimation of genetic parameters for AGB. Although we tested the models with homogenous and heterogeneous residual variances, it would be of interest to further evaluated different numbers of classes of residual variances. Moreover, determining AGB on a daily bases under field conditions might help to estimate the ideal residual variance structure, improving the accuracy of genetic parameters.

Here we revealed the genetic architecture of AGB over time. Our results indicate that a set of different genomic regions is associate with AGB through the studied period. Evidence of a pattern of the SNP effects and different set of candidate genes were found across DAP. Further work is necessary to validate the associations found using independent populations and to explore differential gene expression over time. Potentially, some candidate genes can be found to be constant or transient expressed during AGB development in soybean. Our research also suggests that AGB is a good candidate for GS. Estimating GEBV using RRM has the potential to select lines for increased AGB in specific time points or to change the overall accumulation of the trait. However, this work needs to be complemented with validation to implement the GS model in new populations and to evaluate different factors influencing prediction accuracy over time, such as different training population size or the number of markers.

Although AGB has been associated with yield increases in soybean and other crops, the relationship between AGB and yield needs to be studied. We believe that AGB is genetically correlated to yield and can be used as a secondary trait to improve yield genomic prediction in a multivariate random regression model. Furthermore, soybean phenology plays a key role in determining the crop adaptation region and may contribute to the phenotypic variation of AGB development. Thus, future studies combining phenology models with AGB are necessary to understand their relationship over time.

APPENDIX A. CHAPTER 2

Table A.1 Adjusted mean and standard deviation for yield (Kg/ha) and R8 (days to maturity) by selection criteria for preliminary yield trials (PYT) early and late in 2016 and 2017.

Selection Criteria	Trial.	----- PYT 2016 -----		----- PYT 2017 -----	
		Yield	R8	Yield	R8
ACC	Early	4,121.1±361	117.5±3	3,954.9±486	115.5±2.2
Yield	Early	4,213.5±364	118.1±3	4,215.5±293	115.7±2
Yield ACC	Early	4,141.2±390	117.9±3.1	4,099.2±422	117±2.9
Check	Early	3,843.2±576	114.2±2.9	4,310.5±512	113.6±2.8
ACC	Late	4,704.4±397	127.3±4	4,170.7±542	120.9±3.6
Yield	Late	4,825.4±409	128.2±3.6	4,697±280	123.8±4.1
Yield ACC	Late	4,664.2±481	129.1±3.2	4,539.4±409	123.2±3.1
Check	Late	4,444.7±808	121.3±2.1	4,397.6±311	113.4±2.1

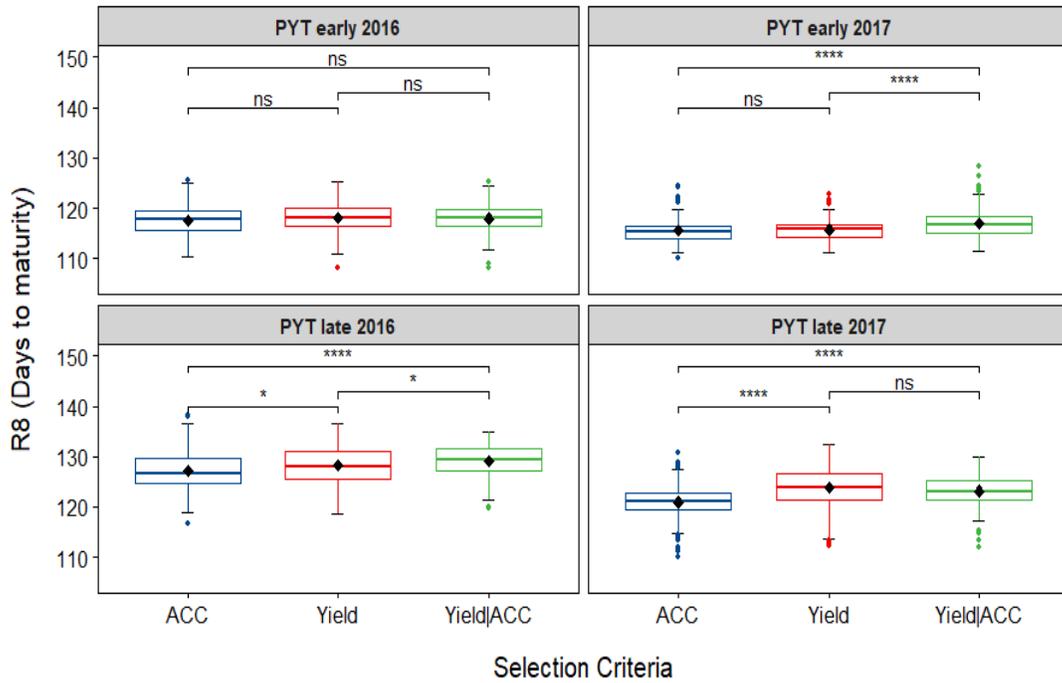


Figure A.1 Box plot of adjusted R8 (days to maturity) distribution for lines selected by each selection categories (Yield, ACC and Yield|ACC) for preliminary yield trials (PYT) early and late in 2016 and 2017. Diamond indicates mean for each selection categories. The line crossing the box plots are representing the median for each class. No significant (ns): $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$.

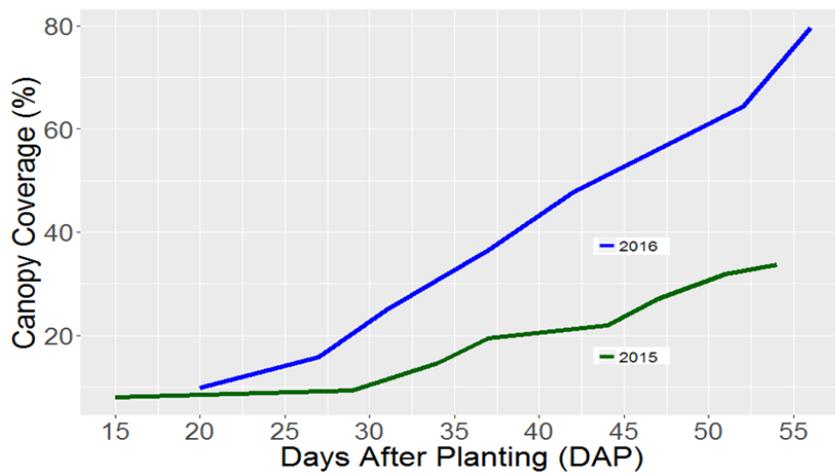


Figure A.2 Distribution of average canopy coverage of the checks by days after planting for progeny rows 2015 and 2016.

APPENDIX B. CHAPTER 4

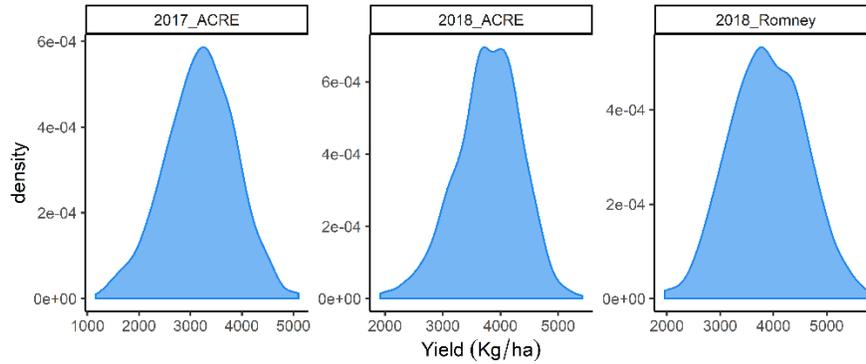


Figure B.1 Yield (Kg/ha) phenotypic distribution by environment.

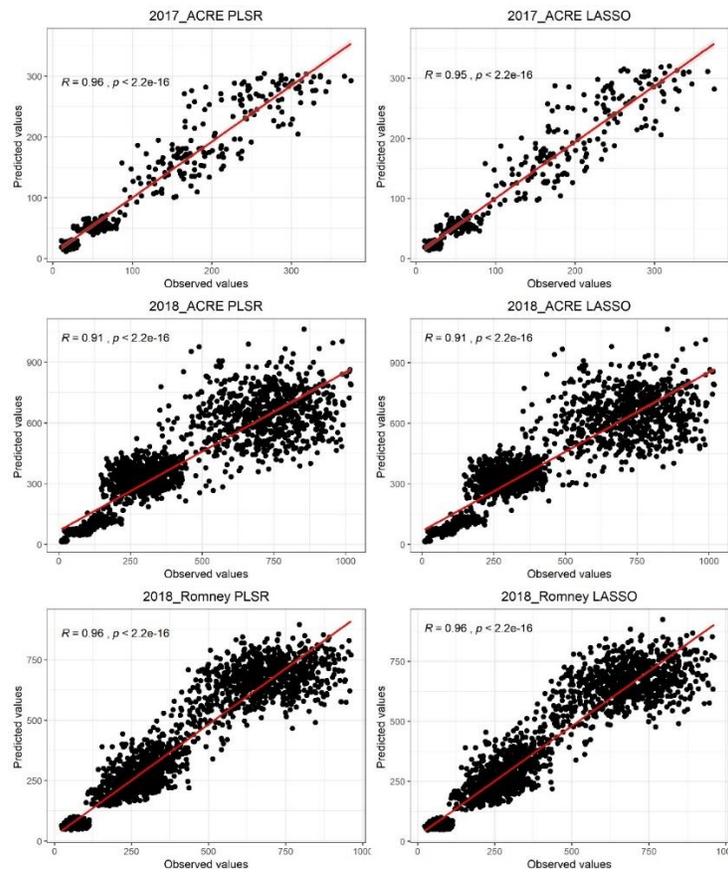


Figure B.2 Correlation (R) plots between observed and predicted above-ground biomass (g/m^2) for each environment for the least absolute shrinkage and selection operator (LASSO) regression and partial least squares regression (PLSR).

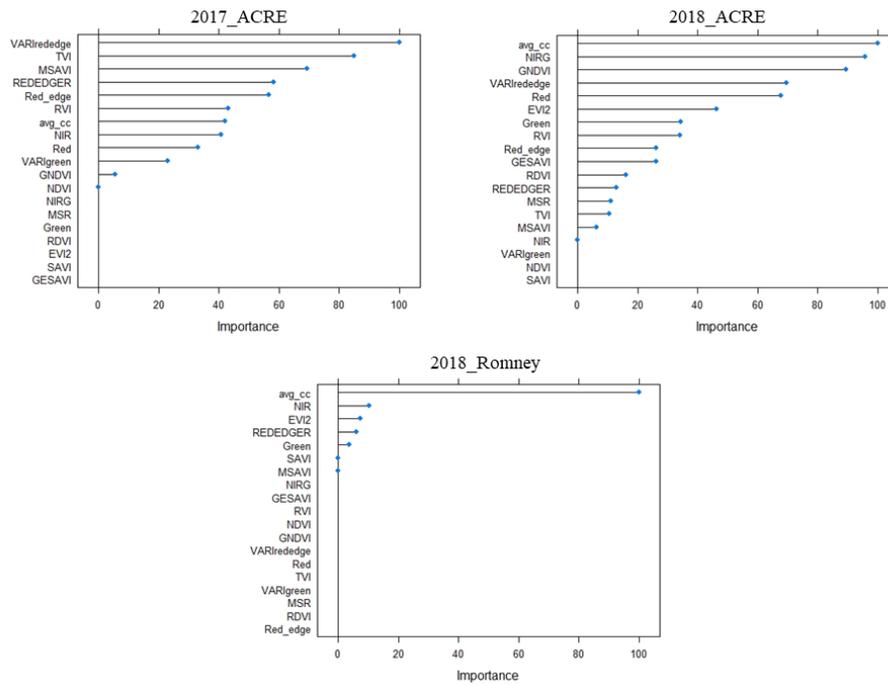


Figure B.3 Variable importance for each environment for the least absolute shrinkage and selection operator (LASSO) regression.

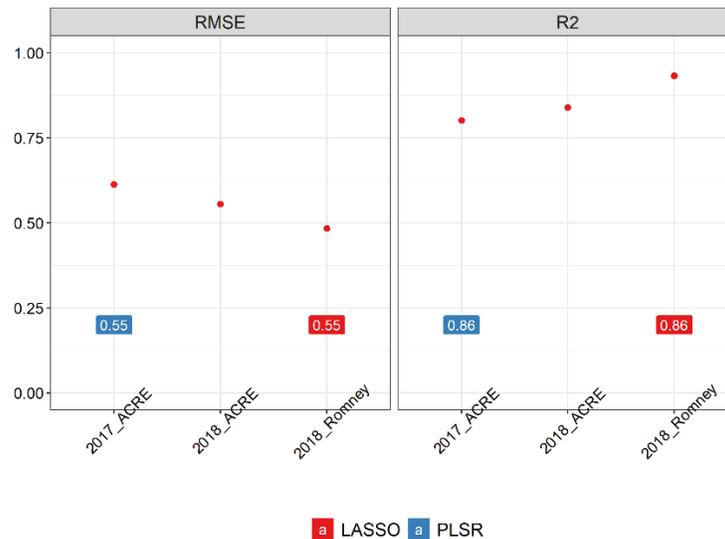


Figure B.4 Performance of above-ground biomass prediction leaving one environment out. Predictions were performed using the least absolute shrinkage and selection operator (LASSO) regression, and the partial least squares regression (PLSR) methods. The performance of predictions was evaluated using the root mean squared error (RMSE) and coefficient of determination (R²).

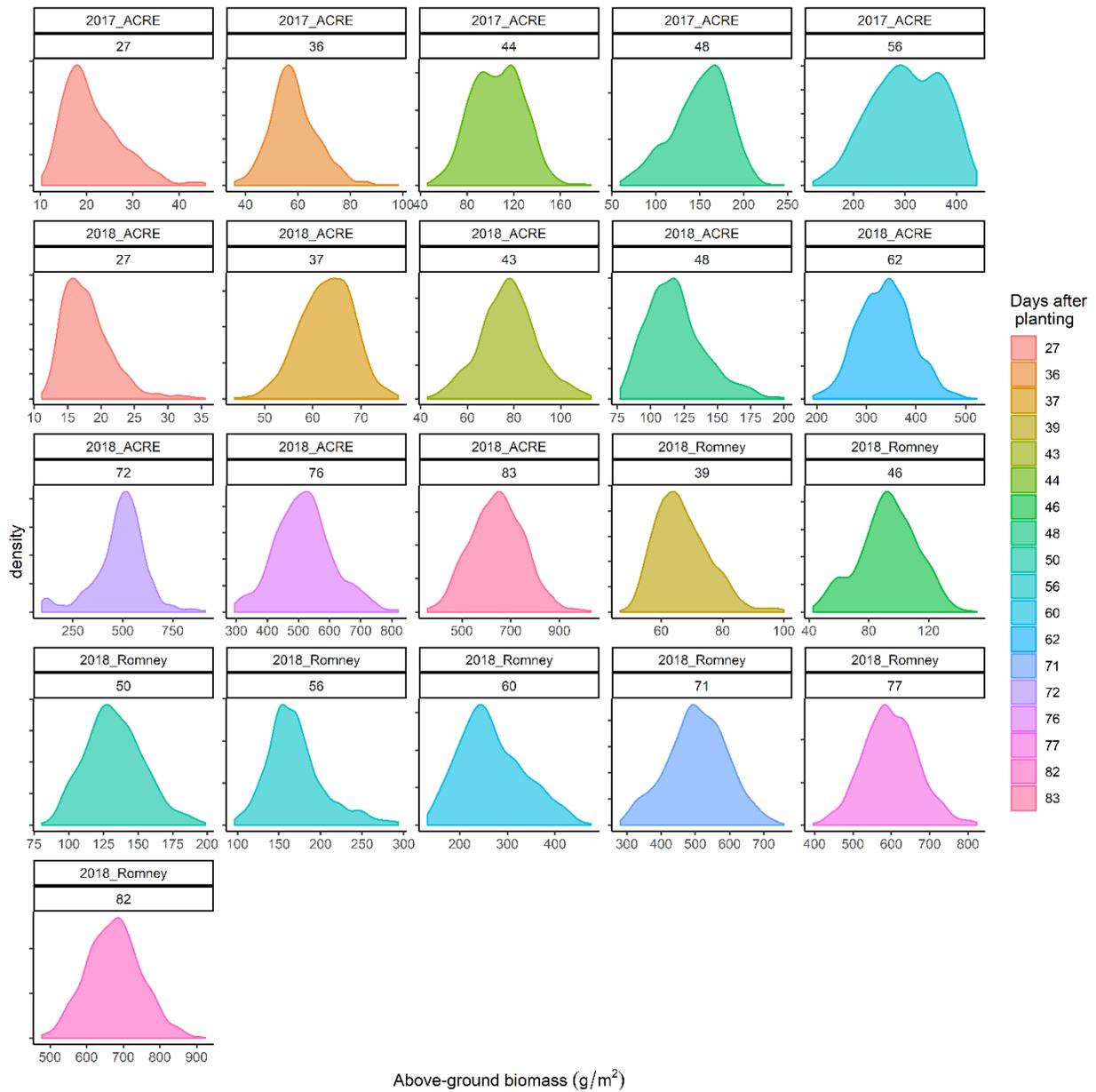


Figure B.5 Phenotypic distribution of predicted above-ground biomass (g/m²) using the Least Absolute Shrinkage and Selection Operator (LASSO) Regression by days after planting for each environment.

Table B.1 List of candidate genes for all selected SNP associated with above-ground biomass in soybean according to SoyBase (SoyBase.org).

Duration Category	SNP	Chr.	Pos. (bp)	Selected candidate genes	Database ID	Annotation Description
Long	2:5777782	2	5,777,782	<i>Glyma.02g064300</i>	AT2G02990.1	ribonuclease 1
				<i>Glyma.02g064400</i>	AT2G02990.1	ribonuclease 1
				<i>Glyma.02g064500</i>	AT3G07950.1	rhomboïd protein-related
				<i>Glyma.02g064600</i>	AT1G09320.1	agenet domain-containing protein
				<i>Glyma.02g064700</i>	AT3G54890.1	photosystem I light harvesting complex gene 1
				<i>Glyma.02g064800</i>	AT3G26935.1	DHHC-type zinc finger family protein
				<i>Glyma.02g064900</i>	AT1G48160.1	signal recognition particle 19 kDa protein, putative / SRP19, putative
Short	3:5150181	3	5,150,181	<i>Glyma.03g040800</i>	AT3G23270.1	regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain
				<i>Glyma.03g040900</i>	AT5G01750.2	protein of unknown function (DUF567)
Long	3:10014176	3	10,014,176	<i>Glyma.03g063400</i>	AT3G60470.1	plant protein of unknown function (DUF247)
Intermittent	3:14985662	3	14,985,662	NA	-	-
Short	4:10352467	4	10,352,467	NA	-	-
Mid	4:13022455	4	13,022,455	<i>Glyma.04g114400</i>	GO:0016021	integral component of membrane
				<i>Glyma.04g114500</i>	-	-
				<i>Glyma.04g114600</i>	-	-
				<i>Glyma.04g114700</i>	-	-
Short	4:14549891	4	14,549,891	NA	-	-
Short	5:3808314	5	3,808,314	<i>Glyma.05g042400</i>	AT5G50210.1	quinolinate synthase
				<i>Glyma.05g042500</i>	AT4G13310.1	cytochrome P450, family 71, subfamily A, polypeptide 20
				<i>Glyma.05g042600</i>	AT4G13310.1	cytochrome P450, family 71, subfamily A, polypeptide 20
				<i>Glyma.05g042700</i>	AT4G24740.1	FUS3-complementing gene 2
Long	7:6108702	7	6,108,702	<i>Glyma.07g067500</i>	AT5G41460.1	protein of unknown function (DUF604)
				<i>Glyma.07g067600</i>	AT3G30280.1	HXXXD-type acyl-transferase family protein
				<i>Glyma.07g067700</i>	AT1G01580.1	ferric reduction oxidase 2
				<i>Glyma.07g067800</i>	AT4G12010.1	disease resistance protein (TIR-NBS-LRR class) family
				<i>Glyma.07g067900</i>	AT5G17680.1	disease resistance protein (TIR-NBS-LRR class), putative
				<i>Glyma.07g068000</i>	AT3G61460.1	brassinosteroid-responsive RING-H2
Mid	7:6523718	7	6,523,718	<i>Glyma.07g071500</i>	AT3G61350.1	SKP1 interacting partner 4
				<i>Glyma.07g071600</i>	AT3G61350.1	SKP1 interacting partner 4

Table B.1 continued.

				<i>Glyma.07g071700</i>	-	-
				<i>Glyma.07g071800</i>	AT4G20960.1	cytidine/deoxycytidylate deaminase family protein
				<i>Glyma.07g071900</i>	AT3G61320.1	bestrophin-like protein
				<i>Glyma.07g072100</i>	AT4G00710.1	BR-signaling kinase 3
				<i>Glyma.07g119600</i>	PF01107	viral movement protein (MP)
Long	7:13739807	7	13,739,807	<i>Glyma.07g119700</i>	AT5G59540.1	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
				<i>Glyma.07g119800</i>	AT1G02180.1	ferredoxin-related
				<i>Glyma.07g119900</i>	AT1G22860.1	vacuolar sorting protein 39
				<i>Glyma.07g120000</i>	AT3G31430.1	
Short				<i>Glyma.07g128000</i>	AT4G26200.1	1-amino-cyclopropane-1-carboxylate synthase 7
				<i>Glyma.07g128100</i>	AT4G29510.1	arginine methyltransferase 11
				<i>Glyma.07g128200</i>	AT4G29520.1	-
				<i>Glyma.07g128300</i>	AT5G11780.1	-
				<i>Glyma.07g128400</i>	AT5G56750.1	N-MYC downregulated-like 1
				<i>Glyma.07g128500</i>	AT4G26180.1	mitochondrial substrate carrier family protein
				<i>Glyma.07g128600</i>	AT2G15690.1	tetratricopeptide repeat (TPR)-like superfamily protein
				<i>Glyma.07g128700</i>	AT5G56780.1	effector of transcription2
Long	7:27576963	7	27,576,963	NA	-	-
				<i>Glyma.08g247700</i>	AT3G29770.1	methyl esterase 11
Mid	8:21488582	8	21,488,582	<i>Glyma.08g247600</i>	AT4G26190.1	haloacid dehalogenase-like hydrolase (HAD) superfamily protein
				<i>Glyma.08g250700</i>	AT3G02310.1	K-box region and MADS-box transcription factor family protein
Mid	8:21970879	8	21,970,879	<i>Glyma.08g250800</i>	AT5G60910.1	AGAMOUS-like 8
				<i>Glyma.08g250900</i>	AT3G02300.1	regulator of chromosome condensation (RCC1) family protein
				<i>Glyma.11g067500</i>	AT5G66750.1	chromatin remodeling 1
Short	11:5076944	11	5,076,944	<i>Glyma.11g067700</i>	AT3G50660.1	cytochrome P450 superfamily protein
				<i>Glyma.11g067800</i>	AT5G66760.1	succinate dehydrogenase 1-1
				<i>Glyma.11g079500</i>	AT4G35920.1	PLAC8 family protein
Mid	11:6004142	11	6,004,142	<i>Glyma.11g079600</i>	AT1G75950.1	S phase kinase-associated protein 1
				<i>Glyma.11g079700</i>	AT3G51320.1	pentatricopeptide repeat (PPR) superfamily protein
				<i>Glyma.11g079800</i>	PF03966	Trm112p-like protein
				<i>Glyma.11g079900</i>	AT3G51325.1	RING/U-box superfamily protein

Table B.1 continued.

				<i>Glyma.11g080000</i>	AT4G35900.1	basic-leucine zipper (bZIP) transcription factor family protein
				<i>Glyma.11g080100</i>	AT4G35890.1	winged-helix DNA-binding transcription factor family protein
Mid	12:3072635	12	3,072,635	<i>Glyma.12g042300</i>	AT2G17080.1	<i>Arabidopsis</i> protein of unknown function (DUF241)
				<i>Glyma.12g042400</i>	AT4G39210.1	glucose-1-phosphate adenylyltransferase family protein
				<i>Glyma.12g042500</i>	AT2G16365.1	F-box family protein
				<i>Glyma.12g042600</i>	AT4G25730.1	FtsJ-like methyltransferase family protein
				<i>Glyma.12g042700</i>	AT4G34540.1	NmrA-like negative transcriptional regulator family protein
				<i>Glyma.12g042800</i>	AT2G18300.2	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
Intermittent	13:2112345	13	2,112,345	<i>Glyma.13g007700</i>	AT3G51690.1	PIF1 helicase
				<i>Glyma.13g007800</i>	AT1G02960.2	-
				<i>Glyma.13g007900</i>	AT4G12680.1	integral component of membrane
Long	13:24980935	13	24,980,935	<i>Glyma.13g137000</i>	AT3G09670.1	tudor/PWWP/MBT superfamily protein
				<i>Glyma.13g137100</i>	AT5G02960.1	ribosomal protein S12/S23 family protein
				<i>Glyma.13g137200</i>	AT2G37080.1	ROP interactive partner 3
Short	13:37339900	13	37,339,900	<i>Glyma.13g271100</i>	AT3G14980.1	acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein
				<i>Glyma.13g271200</i>	AT1G53280.1	class I glutamine amidotransferase-like superfamily protein
				<i>Glyma.13g271300</i>	AT1G53270.1	ABC-2 type transporter family protein
				<i>Glyma.13g271400</i>	AT3G15000.1	cobalt ion binding
				<i>Glyma.13g271500</i>	AT5G20040.1	isopenentenyltransferase 9
Long	14:47457673	14	47,457,673	<i>Glyma.14g208600</i>	AT2G40510.1	ribosomal protein S26e family protein
				<i>Glyma.14g208700</i>	AT5G37150.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein
				<i>Glyma.14g208800</i>	-	-
				<i>Glyma.14g208900</i>	AT1G17200.1	uncharacterised protein family (UPF0497)
				<i>Glyma.14g209000</i>	AT5G56550.1	oxidative stress 3
				<i>Glyma.14g209100</i>	GO:0015979	photosynthesis
				<i>Glyma.14g209200</i>	AT1G31240.1	bromodomain transcription factor
Intermittent	15:10511120	15	10,511,120	<i>Glyma.15g130600</i>	AT2G34730.1	myosin heavy chain-related
				<i>Glyma.15g130700</i>	GO:0016020	membrane
				<i>Glyma.15g130800</i>	AT1G30580.1	GTP binding
				<i>Glyma.15g130900</i>	AT4G20940.1	leucine-rich receptor-like protein kinase family protein
Long	15:13201754	15	13,201,754	<i>Glyma.15g157100</i>	AT5G44030.1	cellulose synthase A4

Table B.1 continued.

				<i>Glyma.15g157200</i>	AT3G27160.1	ribosomal protein S21 family protein
				<i>Glyma.15g157300</i>	AT4G20040.1	pectin lyase-like superfamily protein
				<i>Glyma.15g157400</i>	AT5G05280.1	RING/U-box superfamily protein
				<i>Glyma.15g157500</i>	AT2G32260.1	phosphorylcholine cytidyltransferase
				<i>Glyma.15g157700</i>	AT3G17365.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
Short	15:36306421	15	36,306,421	<i>Glyma.15g217500</i>	AT1G30820.1	CTP synthase family protein
Mid	15:36870472	15	36,870,472	NA	-	-
Long	16:4353954	16	4,353,954	<i>Glyma.16g045700</i>	AT3G26680.1	DNA repair metallo-beta-lactamase family protein
				<i>Glyma.16g045800</i>	AT5G13890.1	family of unknown function (DUF716)
				<i>Glyma.16g045900</i>	AT5G46860.1	syntaxin/t-SNARE family protein
				<i>Glyma.16g046000</i>	AT3G02060.2	DEAD/DEAH box helicase, putative
				<i>Glyma.16g046200</i>	AT3G02050.1	K ⁺ uptake transporter 3
Mid	16:29044334	16	29,044,334	<i>Glyma.16g133600</i>	AT3G50930.1	cytochrome BC1 synthesis
				<i>Glyma.16g133700</i>	ATMG01170.1	ATPase, F0 complex, subunit A protein
				<i>Glyma.16g133800</i>	AT1G19630.1	cytochrome P450, family 722, subfamily A, polypeptide 1
Short	18:6899324	18	6,899,324	<i>Glyma.18g072800</i>	AT1G55760.1	BTB/POZ domain-containing protein
				<i>Glyma.18g073100</i>	AT1G55790.1	domain of unknown function (DUF2431)
				<i>Glyma.18g073200</i>	AT4G26490.1	late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
				<i>Glyma.18g073300</i>	AT5G08520.1	duplicated homeodomain-like superfamily protein
				<i>Glyma.18g073400</i>	AT4G26510.1	uridine kinase-like 4
Intermittent	18:59148882	18	59,148,882	<i>Glyma.18g302000</i>	AT2G33510.1	-
				<i>Glyma.18g302100</i>	AT2G39290.1	phosphatidylglycerolphosphate synthase 1
				<i>Glyma.18g302200</i>	AT4G30860.1	SET domain group 4
				<i>Glyma.18g302300</i>	AT2G01910.2	microtubule associated protein (MAP65/ASE1) family protein

Chr: Chromosome, Pos (bp): position in base pair. NA: no positional candidate gene found within \pm 25kb. - Annotation not available.

REFERENCES

- Acharjee, A., Kloosterman, B., Visser, R. G. F., and Maliepaard, C. (2016). Integration of multi-omics data for prediction of phenotypic traits using random forest. *BMC Bioinformatics* 17, 180. doi:10.1186/s12859-016-1043-4.
- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93, 743–752. doi:10.3168/jds.2009-2730.
- Aguilar, I., Misztal, I., Tsuruta, S., Legarra, A. A., and Wang, H. (2014). PREGSF90 – POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. in *Proceedings of the World Congress on Genetics Applied to Livestock Production* (Vancouver, Canada: World Congress on Genetics Applied to Livestock Production), 680. doi:10.13140/2.1.4801.5045.
- Ainsworth, E. A., Yendrek, C. R., Skoneczka, J. A., and Long, S. P. (2012). Accelerating yield potential in soybean: Potential targets for biotechnological improvement. *Plant, Cell Environ.* 35, 38–52. doi:10.1111/j.1365-3040.2011.02378.x.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi:10.1109/TAC.1974.1100705.
- Albuquerque, L. G., and Meyer, K. (2001). Estimates of covariance functions for growth from birth to 630 days of age in Nelore cattle. *J. Anim. Sci.* 79, 2776–2789. doi:10.2527/2001.79112776x.
- Alexandratos, N., and Bruinsma, J. (2012). World agriculture towards 2030/2050. *Land use policy* 20, 375. doi:10.1016/S0264-8377(03)00047-4.
- An, N., Palmer, C. M., Baker, R. L., Markelz, R. J. C., Ta, J., Covington, M. F., 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.
- Anderson, E. J., Ali, M. L., Beavis, W. D., Chen, P., Clemente, T. E., Diers, B. W., et al. (2019). “Soybean [Glycine max (L.) Merr.] Breeding: History, Improvement, Production and Future Opportunities,” in *Advances in Plant Breeding Strategies: Legumes* (Cham: Springer International Publishing), 431–516. doi:10.1007/978-3-030-23400-3_12.
- Apiolaza, L. A., and Garrick, D. J. (2001). Analysis of Longitudinal Data from Progeny Tests: Some Multivariate Approaches. Available at: <https://academic.oup.com/forestscience/article-abstract/47/2/129/4617386> [Accessed June 4, 2019].

- Apiolaza, L. A., Gilmour, A. R., and Garrick, D. J. (2011). Variance modelling of longitudinal height data from a *Pinus radiata* progeny test. *Can. J. For. Res.* 30, 645–654. doi:10.1139/x99-246.
- Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi:10.1016/j.tplants.2013.09.008.
- Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating High-Throughput Phenotyping into Genetic Gain. Elsevier doi:10.1016/j.tplants.2018.02.001.
- Ashraf, B., Edriss, V., Akdemir, D., Autrique, E., Bonnett, D., Crossa, J., et al. (2016). Genomic prediction using phenotypes from pedigreed lines with no marker data. *Crop Sci.* 56, 957–964. doi:10.2135/cropsci2015.02.0111.
- Atkinson, J. A., Pound, M. P., Bennett, M. J., and Wells, D. M. (2019). Uncovering the hidden half of plants using new advances in root phenotyping. *Curr. Opin. Biotechnol.* 55, 1–8. doi:10.1016/j.copbio.2018.06.002.
- Auinger, H. J., Schönleben, M., Lehermeier, C., Schmidt, M., Korzun, V., Geiger, H. H., et al. (2016). Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (*Secale cereale* L.). *Theor. Appl. Genet.* 129, 2043–2053. doi:10.1007/s00122-016-2756-5.
- Azizi, S., Bayat, S., Yan, P., Tahmasebi, A., Kwak, J. T., Xu, S., et al. (2018). Deep recurrent neural networks for prostate cancer detection: Analysis of temporal enhanced ultrasound. *IEEE Trans. Med. Imaging* 37, 2695–2703. doi:10.1109/TMI.2018.2849959.
- Baba, T., Momen, M., Campbell, M. T., Walia, H., and Morota, G. (2020). Multi-trait random regression models increase genomic prediction accuracy for a temporal physiological trait derived from high-throughput phenotyping. *PLoS One* 15, e0228118. doi:10.1371/journal.pone.0228118.
- Babar, M. A., Reynolds, M. P., Van Ginkel, M., Klatt, A. R., Raun, W. R., and Stone, M. L. (2006). Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in wheat. *Crop Sci.* 46, 1046–1057. doi:10.2135/cropsci2005.0211.
- Baenziger, P. S., Russell, W. K., Graef, G. L., and Campbell, B. T. (2006). Improving lives: 50 Years of crop breeding, genetics, and cytology (C-1). in *Crop Science*, 2230–2244. doi:10.2135/cropsci2005.11.0404gas.
- Baillet, N., Girousse, C., Allard, V., Piquet-Pissaloux, A., and Le Gouis, J. (2018). Different grain-filling rates explain grain-weight differences along the wheat ear. *PLoS One* 13, e0209597. doi:10.1371/journal.pone.0209597.

- Bajgain, R., Kawasaki, Y., Akamatsu, Y., Tanaka, Y., Kawamura, H., Katsura, K., et al. (2015). Biomass production and yield of soybean grown under converted paddy fields with excess water during the early growth stage. *F. Crop. Res.* 180, 221–227. doi:10.1016/J.FCR.2015.06.010.
- Baker, B., Olszyk, D. M., and Tingey, D. (1996). Digital image analysis to estimate leaf area. *J. Plant Physiol.* 148, 530–535. doi:10.1016/S0176-1617(96)80072-1.
- Baker, R. L., Leong, W. F., Brock, M. T., Rubin, M. J., Markelz, R. J. C., Welch, S., et al. (2019). Integrating transcriptomic network reconstruction and eQTL analyses reveals mechanistic connections between genomic architecture and *Brassica rapa* development. *PLOS Genet.* 15, e1008367. doi:10.1371/journal.pgen.1008367.
- Balboa, G. R., Sadras, V. O., and Ciampitti, I. A. (2018). Shifts in Soybean Yield, Nutrient Uptake, and Nutrient Stoichiometry: A Historical Synthesis-Analysis. *Crop Sci.* 58, 43. doi:10.2135/cropsci2017.06.0349.
- Ban, Y. W., Roy, N. S., Yang, H., Choi, H. K., Kim, J. H., Babu, P., et al. (2019). Comparative transcriptome analysis reveals higher expression of stress and defense responsive genes in dwarf soybeans obtained from the crossing of *G. max* and *G. soja*. *Genes and Genomics* 41, 1315–1327. doi:10.1007/s13258-019-00846-2.
- Basnet, B. R., Crossa, J., Dreisigacker, S., Pérez-Rodríguez, P., Manes, Y., Singh, R. P., et al. (2018). Hybrid Wheat Prediction Using Genomic, Pedigree, and Environmental Covariables Interaction Models. *Plant Genome* 12, 0. doi:10.3835/plantgenome2018.07.0051.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *J. Stat. Softw.* 67, 1–48. doi:10.18637/jss.v067.i01.
- Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., and Bareth, G. (2014). Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging. *Remote Sens.* 6, 10395–10412. doi:10.3390/rs61110395.
- Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., et al. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* 39, 79–87. doi:10.1016/j.jag.2015.02.012.
- Bernardo, R. N. (2010). *Breeding for quantitative traits in plants*. Stemma Press.
- Blancon, J., Dutartre, D., Tixier, M.-H., Weiss, M., Comar, A., Praud, S., et al. (2019). A High-Throughput Model-Assisted Method for Phenotyping Maize Green Leaf Area Index Dynamics Using Unmanned Aerial Vehicle Imagery. *Front. Plant Sci.* 10, 685. doi:10.3389/fpls.2019.00685.

- Board, J. E., Kamal, M., and Harville, B. G. (1992). Temporal Importance of Greater Light Interception to Increased Yield in Narrow-Row Soybean. *Agron. J.* 84, 575. doi:10.2134/agronj1992.00021962008400040006x.
- Board, J., and Kahlon, C. (2011). Soybean yield formation: what controls it and how it can be improved. *Soybean Physiol. Biochem.*, 1–36. doi:10.5772/1006.
- Bodner, G., Nakhforoosh, A., Arnold, T., and Leitner, D. (2018). Hyperspectral imaging: A novel approach for plant root phenotyping. *Plant Methods* 14, 84. doi:10.1186/s13007-018-0352-1.
- Boerma, H. R., Specht, J. E., Purcell, L. C., Specht, J. E., Roger Boerma, H., Specht, J. E., et al. (2004). “Physiological Traits for Ameliorating Drought Stress,” in *Soybeans: improvement, production, and uses.*, eds. H. R. Boerma and J. E. Specht (Madison, WI: American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America), 569–620. doi:10.2134/agronmonogr16.3ed.c12.
- Bohlouli, M., Alijani, S., Naderi, S., Yin, T., and König, S. (2019). Prediction accuracies and genetic parameters for test-day traits from genomic and pedigree-based random regression models with or without heat stress interactions. *J. Dairy Sci.* 102, 488–502. doi:10.3168/jds.2018-15329.
- Bohmanova, J., Miglior, F., Jamrozik, J., Misztal, I., and Sullivan, P. G. (2008). Comparison of Random Regression Models with Legendre Polynomials and Linear Splines for Production Traits and Somatic Cell Score of Canadian Holstein Cows. *J. Dairy Sci.* 91, 3627–3638. doi:10.3168/jds.2007-0945.
- Bonafous, F., Fievet, G., Blanchet, N., Boniface, M.-C. C., Carrère, S., Gouzy, J., et al. (2018). Comparison of GWAS models to identify non-additive genetic control of flowering time in sunflower hybrids. *Theor. Appl. Genet.* 131, 319–332. doi:10.1007/s00122-017-3003-4.
- Borquis, R. R. A., Neto, F. R. de A., Baldi, F., Hurtado-Lugo, N., de Camargo, G. M. F., Muñoz-Berrocal, M., et al. (2013). Multiple-trait random regression models for the estimation of genetic parameters for milk, fat, and protein yield in buffaloes. *J. Dairy Sci.* 96, 5923–5932. doi:10.3168/JDS.2012-6023.
- Bouvet, J. M., Makouanzi, G., Cros, D., and Vigneron, P. (2016). Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping: Prediction accuracy implications. *Heredity (Edinb)*. 116, 146–157. doi:10.1038/hdy.2015.78.
- Box, G. E. P., and Cox, D. R. (1964). An Analysis of Transformations. *J. R. Stat. Soc. Ser. B* 26, 211–243. doi:10.1111/j.2517-6161.1964.tb00553.x.
- Bradshaw, J. E. (2017). *Plant breeding: Past, present and future*. Springer Netherlands doi:10.1007/978-3-319-23285-0.

- Brasil, J. N., Cabral, L. M., Eloy, N. B., Primo, L. M. F., Barroso-Neto, I. L., Grangeiro, L. P. P., et al. (2015). AIP1 is a novel Agenet/Tudor domain protein from Arabidopsis that interacts with regulators of DNA replication, transcription and chromatin remodeling. *BMC Plant Biol.* 15, 270. doi:10.1186/s12870-015-0641-z.
- Bratsch, S., Epstein, H., Buchhorn, M., Walker, D., and Landes, H. (2017). Relationships between hyperspectral data and components of vegetation biomass in Low Arctic tundra communities at Iivotuk, Alaska. *Environ. Res. Lett.* 12, 025003. doi:10.1088/1748-9326/aa572e.
- Brito, L. F., Gomes da Silva, F., Rojas de Oliveira, H., Souza, N., Caetano, G., Costa, E. V., et al. (2017a). Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. *Can. J. Anim. Sci.*, CJAS-2017-0019. doi:10.1139/CJAS-2017-0019.
- Brito, L. F., Silva, F. G., Oliveira, H. R., Souza, N. O., Caetano, G. C., Costa, E. V., et al. (2017b). Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. *Can. J. Anim. Sci.* 98, 73–83. doi:10.1139/cjas-2017-0019.
- Brito, L. F., Silva, F. G., Oliveira, H. R., Souza, N. O., Caetano, G. C., Costa, E. V., et al. (2018). Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. *Can. J. Anim. Sci.* 98, 73–83. doi:10.1139/cjas-2017-0019.
- Bromley, C. M., Van Vleck, L. D., Johnson, B. E., and Smith, O. S. (2000). Estimation of genetic variance in corn from F1 performance with and without pedigree relationships among inbred lines. *Crop Sci.* 40, 651–655. doi:10.2135/cropsci2000.403651x.
- Brügemann, K., Gernand, E., von Borstel, U. U., and König, S. (2011). Genetic analyses of protein yield in dairy cows applying random regression models with time-dependent and temperature x humidity-dependent covariates. *J. Dairy Sci.* 94, 4129–4139. doi:10.3168/jds.2010-4063.
- Bullock, D., Khan, S., and Rayburn, A. (1998). Soybean yield response to narrow rows is largely due to enhanced early growth. *Crop Sci.* 38, 1011–1016. doi:10.2135/cropsci1998.0011183X003800040021x.
- Burridge, J., Jochua, C. N., Bucksch, A., and Lynch, J. P. (2016). Legume shovelomics: High-Throughput phenotyping of common bean (*Phaseolus vulgaris* L.) and cowpea (*Vigna unguiculata* subsp. *unguiculata*) root architecture in the field. *F. Crop. Res.* 192, 21–32. doi:10.1016/j.fcr.2016.04.008.
- Busemeyer, L., Ruckelshausen, A., Möller, K., Melchinger, A. E., Alheit, K. V., Maurer, H. P., et al. (2013). Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. *Sci. Rep.* 3, 2442. doi:10.1038/srep02442.
- Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M. D., and Luis Araus, J. (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.

- Camargo, A. V., Mackay, I., Mott, R., Han, J., Doonan, J. H., Askew, K., et al. (2018). Functional Mapping of Quantitative Trait Loci (QTLs) Associated With Plant Performance in a Wheat MAGIC Mapping Population. *Front. Plant Sci.* 9, 887. doi:10.3389/fpls.2018.00887.
- Campbell, M., Momen, M., Walia, H., and Morota, G. (2019). Leveraging Breeding Values Obtained from Random Regression Models for Genetic Inference of Longitudinal Traits. *Plant Genome* 12, 435685. doi:10.3835/plantgenome2018.10.0075.
- Campbell, M. T., Du, Q., Liu, K., Brien, C. J., Berger, B., Zhang, C., et al. (2017). A Comprehensive Image-based Phenomic Analysis Reveals the Complex Genetic Architecture of Shoot Growth Dynamics in Rice (*Oryza sativa*). *Plant Genome* 10, 0. doi:10.3835/plantgenome2016.07.0064.
- Campbell, M., Walia, H., and Morota, G. (2018). Utilizing random regression models for genomic prediction of a longitudinal trait derived from high-throughput phenotyping. *Plant Direct* 2, e00080. doi:10.1002/pld3.80.
- Cappa, E. P., de Lima, B. M., da Silva-Junior, O. B., Garcia, C. C., Mansfield, S. D., and Grattapaglia, D. (2019). Improving genomic prediction of growth and wood traits in Eucalyptus using phenotypes from non-genotyped trees by single-step GBLUP. *Plant Sci.* 284, 9–15. doi:10.1016/j.plantsci.2019.03.017.
- Carvalho, R., Costilla, R., Neves, H. H. R., Albuquerque, L. G., Moore, S., and Hayes, B. J. (2019). Unraveling genetic sensitivity of beef cattle to environmental variation under tropical conditions. *Genet. Sel. Evol.* 51, 29. doi:10.1186/s12711-019-0470-x.
- Casadesús, J., Kaya, Y., Bort, J., Nachit, M. M., Araus, J. L., Amor, S., et al. (2007). Using vegetation indices derived from conventional digital cameras as selection criteria for wheat breeding in water-limited environments. *Ann. Appl. Biol.* 150, 227–236. doi:10.1111/j.1744-7348.2007.00116.x.
- Casadesús, J., and Villegas, D. (2014). Conventional digital cameras as a tool for assessing leaf area index and biomass for cereal breeding. *J. Integr. Plant Biol.* 56, 7–14. doi:10.1111/jipb.12117.
- Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., and Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nat. Clim. Chang.* 4, 287–291. doi:10.1038/nclimate2153.
- Chaudhary, K., Poirion, O. B., Lu, L., and Garmire, L. X. (2018). Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin. Cancer Res.* 24, 1248–1259. doi:10.1158/1078-0432.CCR-17-0853.
- Che, Z., Purushotham, S., Cho, K., Sontag, D., and Liu, Y. (2018). Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Sci. Rep.* 8, 6085. doi:10.1038/s41598-018-24271-9.

- Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., et al. (2014). Dissecting the Phenotypic Components of Crop Plant Growth and Drought Responses Based on High-Throughput Image Analysis. *Plant Cell Online* 26, 4636–4655. doi:10.1105/tpc.114.129601.
- Chen, D., Shi, R., Pape, J.-M., and Klukas, C. (2016). Predicting plant biomass accumulation from image-derived parameters. *bioRxiv* 7, 046656. doi:10.1101/046656.
- Chen, J. M. (1996). Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Can. J. Remote Sens.* 22, 229–242. doi:10.1080/07038992.1996.10855178.
- Chen, J., and Shiyomi, M. (2019). A power law model for analyzing spatial patterns of vegetation abundance in terms of cover, biomass, density, and occurrence: derivation of a common rule. *J. Plant Res.* 132, 481–497. doi:10.1007/s10265-019-01116-8.
- Cheng, T., Song, R., Li, D., Zhou, K., Zheng, H., Yao, X., et al. (2017). Spectroscopic estimation of biomass in canopy components of paddy rice using dry matter and chlorophyll indices. *Remote Sens.* 9, 319. doi:10.3390/rs9040319.
- Christensen, O. F., and Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42, 2. doi:10.1186/1297-9686-42-2.
- Chung, J., Babka, H. L., Graef, G. L., Staswick, P. E., Lee, D. J., Cregan, P. B., et al. (2003). The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci.* 43, 1053–1067. doi:10.2135/cropsci2003.1053.
- Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., and McCouch, S. (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.
- Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., et al. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theor. Appl. Genet.* 132, 627–645. doi:10.1007/s00122-019-03317-0.
- Colleau, J. J., Ducrocq, V., Boichard, D., and Larroque, H. (1999). Approximate multitrait BLUP evaluation to combine functional traits information. in, 151–160. Available at: file:///C:/Users/Fabiana/OneDrive - purdue.edu/PhD/Review paper/393-784-1-SM.pdf [Accessed June 3, 2019].
- Coppens, F., Wuyts, N., Inzé, D., and Dhondt, S. (2017). Unlocking the potential of plant phenotyping data through integration and data-driven approaches. *Curr. Opin. Syst. Biol.* 4, 58–63. doi:10.1016/j.coisb.2017.07.002.
- Crain, J., Mondal, S., Rutkoski, J., Singh, R. P., and Poland, J. (2018). Combining High-Throughput Phenotyping and Genomic Information to Increase Prediction and Selection Accuracy in Wheat Breeding. *Plant Genome* 11, 0. doi:10.3835/plantgenome2017.05.0043.

- Cregan, P. B., and Yaklich, R. W. (1986). Dry matter and nitrogen accumulation and partitioning in selected soybean genotypes of different derivation. *Theor. Appl. Genet.* 72, 782–786. doi:10.1007/BF00266545.
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de Los Campos, G., et al. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Elsevier Current Trends doi:10.1016/j.tplants.2017.08.011.
- Cui, C., and Wang, D. (2016). High dimensional data regression using Lasso model and neural networks with random weights. *Inf. Sci. (Ny)*. 372, 505–517. doi:10.1016/j.ins.2016.08.060.
- Cullis, B. R., Smith, A. B., and Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *J. Agric. Biol. Environ. Stat.* 11, 381–393. doi:10.1198/108571106X154443.
- Das, A., Rushton, P. J., and Rohila, J. S. (2017). Metabolomic profiling of soybeans (*Glycine max* L.) reveals the importance of sugar and nitrogen metabolism under drought and heat stress. *Plants* 6, 199–208. doi:10.3390/plants6020021.
- Das, K., Li, J., Wang, Z., Tong, C., Fu, G., Li, Y., et al. (2011). A dynamic model for genome-wide association studies. *Hum. Genet.* 129, 629–639. doi:10.1007/s00439-011-0960-6.
- de Almeida Filho, J. E., Guimarães, J. F. R., Fonsceca e Silva, F., Vilela de Resende, M. D., Muñoz, P., Kirst, M., et al. (2019). Genomic Prediction of Additive and Non-additive Effects Using Genetic Markers and Pedigrees. *G3 & Genes/Genomes/Genetics* 9, 2739–2748. doi:10.1534/g3.119.201004.
- de Boor, C. (1980). A Practical Guide to Splines. *Math. Comput.* 34, 325. doi:10.2307/2006241.
- De Bruin, J. L., and Pedersen, P. (2009). Growth, yield, and yield component changes among old and new soybean cultivars. *Agron. J.* 101, 124–130. doi:10.2134/agronj2008.0187.
- De La Fuente, G. N., Frei, U. K., and Lübberstedt, T. (2013). Accelerating plant breeding. *Trends Plant Sci.* 18, 667–672. doi:10.1016/j.tplants.2013.09.001.
- de Oliveira, H. R., e Silva, F. F., Barbosa da Silva, M. V. G., Barbosa Dias de Siqueira, O. H. G., Machado, M. A., Carmo Panetto, J. C. do, et al. (2017). Bayesian Models combining Legendre and B-spline polynomials for genetic analysis of multiple lactations in Gyr cattle. *Livest. Sci.* 201, 78–84. doi:10.1016/j.livsci.2017.05.007.
- de Oliveira, H. R., Silva, F. F., Brito, L. F., Guarini, A. R., Jamrozik, J., and Schenkel, F. S. (2018). Comparing deregression methods for genomic prediction of test-day traits in dairy cattle. *J. Anim. Breed. Genet.* 135, 97–106. doi:10.1111/jbg.12317.
- De Souza, C. H. W., Lamparelli, R. A. C., Rocha, J. V., and Magalhães, P. S. G. (2017). Height estimation of sugarcane using an unmanned aerial system (UAS) based on structure from motion (SfM) point clouds. *Int. J. Remote Sens.*, 1–13. doi:10.1080/01431161.2017.1285082.

- Deering, D. W., Rouse, J. W., Haas, R. H., and Schell, J. A. (1975). MEASURING "FORAGE PRODUCTION" OF GRAZING UNITS FROM LANDSAT MSS DATA. in *undefined*, 1169–1178. Available at: <https://www.semanticscholar.org/paper/Measuring-forage-production-of-grazing-units-from-Deering/bbdad3db3dc2c59ccf145407390a0771011c44b9> [Accessed October 3, 2019].
- Diers, B. W., Specht, J., Rainey, K. M., Cregan, P., Song, Q., Ramasubramanian, V., et al. (2018). Genetic Architecture of Soybean Yield and Agronomic Traits. *G3 (Bethesda)*. 8, 3367–3375. doi:10.1534/g3.118.200332.
- Durón-Benítez, A.-A., Weller, J. I., and Ezra, E. (2018). Using geometric morphometrics for the genetics analysis of shape and size of lactation curves in Israeli first-parity Holstein cattle. *J. Dairy Sci.* 101, 11132–11142. doi:10.3168/jds.2018-15209.
- Duvick, D. N. (2005). Genetic progress in yield of United States maize (*Zea mays* L.). *Maydica* 50, 193–202. Available at: <https://www.semanticscholar.org/paper/Genetic-progress-in-yield-of-United-States-maize-Duvick/218951bf5a95eb220fe06c28f335fa0a8d33331a> [Accessed September 30, 2019].
- East, E. M. (1911). The Genotype Hypothesis and Hybridization. *Am. Nat.* 45, 160–174. doi:10.1086/279203.
- Edwards, J. T., and Purcell, L. C. (2005). Soybean yield and biomass responses to increasing plant population among diverse maturity groups: I. Agronomic characteristics. *Crop Sci.* 45, 1770–1777. doi:10.2135/cropsci2004.0564.
- Edwards, J. T., Purcell, L. C., and Karcher, D. E. (2005). Soybean yield and biomass responses to increasing plant population among diverse maturity groups: II. Light interception and utilization. *Crop Sci.* 45, 1778–1785. doi:10.2135/cropsci2004.0570.
- Eilers, P. H. C., and Marx, B. D. (1996). Flexible Smoothing with B-splines and Penalties. Available at: https://projecteuclid.org/download/pdf_1/euclid.ss/1038425655 [Accessed May 24, 2019].
- Englishby, T. M., Banos, G., Moore, K. L., Coffey, M. P., Evans, R. D., and Berry, D. P. (2016). Genetic analysis of carcass traits in beef cattle using random regression models. *J. Anim. Sci.* 94, 1354–1364. doi:10.2527/jas.2015-0246.
- Fahlgren, N., Gehan, M. A., and Baxter, I. (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.
- Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. 4th Editio. Burnt Mill, England: Longman.

- Fan, R., Albert, P. S., and Schisterman, E. F. (2012). A discussion of gene-gene and gene-environment interactions and longitudinal genetic analysis of complex traits. *Stat. Med.* 31, 2565–8. doi:10.1002/sim.5495.
- Fehr, W. R. (1998). *Principles of cultivar development: Theory and technique, Voll.* Macmillan Publishing Company Available at: <https://www.cabdirect.org/cabdirect/abstract/19871663198> [Accessed March 15, 2018].
- Fehr, W. R., and Caviness, C. E. (1977). *Stages of Soybean Development.* Available at: <https://lib.dr.iastate.edu/specialreports/87> [Accessed February 21, 2018].
- Feldman, M. J., Paul, R. E., Banan, D., Barrett, J. F., Sebastian, J., Yee, M. C., et al. (2017). Time dependent genetic analysis links field and controlled environment phenotypes in the model C4grass *Setaria*. *PLoS Genet.* 13, e1006841. doi:10.1371/journal.pgen.1006841.
- Fernandez, G. C. J. (2019). Repeated Measure Analysis of Line-source Sprinkler Experiments. *HortScience* 26, 339–342. doi:10.21273/hortsci.26.4.339.
- Fickett, N. D., Boerboom, C. M., and Stoltenberg, D. E. (2013). Soybean Yield Loss Potential Associated with Early-Season Weed Competition across 64 Site-Years. *Weed Sci.* 61, 500–507. doi:10.1614/ws-d-12-00164.1.
- Fiorani, F., Rascher, U., Jahnke, S., and Schurr, U. (2012). Imaging plants dynamics in heterogenic environments. *Curr. Opin. Biotechnol.* 23, 227–235. doi:10.1016/j.copbio.2011.12.010.
- Fiorani, F., and Schurr, U. (2013). Future Scenarios for Plant Phenotyping. *Annu. Rev. Plant Biol.* 64, 267–291. doi:10.1146/annurev-arplant-050312-120137.
- Fischer, R. A., and Edmeades, G. O. (2010). Breeding and cereal yield progress. *Crop Sci.* 50, S-85-S-98. doi:10.2135/cropsci2009.10.0564.
- Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., et al. (2011). Solutions for a cultivated planet. *Nature* 478, 337–342. doi:10.1038/nature10452.
- Foster, J. J., Barkus, E., and Yavorsky, C. (2006). Understanding and using advanced statistics. *Choice Rev. Online* 43, 43-5938-43-5938. doi:10.5860/choice.43-5938.
- Frederick, J. R., Hesketh, J. D., and Slafer, G. A. (1994). Genetic improvement in soybean: physiological attributes. *Genet. Improv. F. Crop.* 237–286; B, lants, and. Available at: <http://agris.fao.org/agris-search/search.do?recordID=US201301488357> [Accessed March 22, 2018].
- Frederick, J. R., Woolley, J. T., Hesketh, J. D., and Peters, D. B. (1991). Seed yield and agronomic traits of old and modern soybean cultivars under irrigation and soil water-deficit. *F. Crop. Res.* 27, 71–82. doi:10.1016/0378-4290(91)90023-O.

- Fu, P., Meacham-Hensold, K., Guan, K., and Bernacchi, C. J. (2019). Hyperspectral leaf reflectance as proxy for photosynthetic capacities: An ensemble approach based on multiple machine learning algorithms. *Front. Plant Sci.* 10, 730. doi:10.3389/fpls.2019.00730.
- Furbank, R. T., and Tester, M. (2011). Phenomics - technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi:10.1016/j.tplants.2011.09.005.
- Gao, A. W., Schlüter, S., Blaser, S. R. G. A., Shen, J., and Vetterlein, D. (2019). A shape-based method for automatic and rapid segmentation of roots in soil from X-ray Computed Tomography images : Routine. *Plant Soil*, 1–13. doi:10.1007/s11104-019-04053-6.
- Gao, L., Pan, H., Liu, F., Xie, X., Zhang, Z., and Han, J. (2018). Brain disease diagnosis using deep learning features from longitudinal MR images. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Springer, Cham), 327–339. doi:10.1007/978-3-319-96890-2_27.
- Gebeyehou, G., Knott, D. R., and Baker, R. J. (1982). Rate and Duration of Grain Filling in Durum Wheat Cultivars1. *Crop Sci.* 22, 337. doi:10.2135/cropsci1982.0011183X002200020033x.
- Gilabert, M. A., Gonzalez Piqueras, J., Garcia-Haro, J., and Melia, J. (1998). Designing a generalized soil-adjusted vegetation index (GESAVI). in *Remote Sensing for Agriculture, Ecosystems, and Hydrology*, ed. E. T. Engman (International Society for Optics and Photonics), 396. doi:10.1117/12.332774.
- Gilliham, M., Able, J. A., and Roy, S. J. (2017). Translating knowledge about abiotic stress tolerance to breeding programmes. *Plant J.* 90, 898–917. doi:10.1111/tpj.13456.
- Gitelson, A. A., Kaufman, Y. J., and Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS- MODIS. *Remote Sens. Environ.* 58, 289–298. doi:10.1016/S0034-4257(96)00072-7.
- Gitelson, A. A., Kaufman, Y. J., Stark, R., and Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* 80, 76–87. doi:10.1016/S0034-4257(01)00289-9.
- Goddard, M. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi:10.1007/s10709-008-9308-0.
- Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: The challenge of feeding 9 billion people. *Science (80-.)*. 327, 812–818. doi:10.1126/science.1185383.
- Gompertz, B. (1815). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Proc. R. Soc. London* 2, 252–253. doi:10.1098/rspl.1815.0271.

- Gonzalez-Dugo, V., Hernandez, P., Solis, I., and Zarco-Tejada, P. J. (2015). Using high-resolution hyperspectral and thermal airborne imagery to assess physiological condition in the context of wheat phenotyping. *Remote Sens.* 7, 13586–13605. doi:10.3390/rs71013586.
- Gottschalk, P. G., and Dunn, J. R. (2005). The five-parameter logistic: A characterization and comparison with the four-parameter logistic. *Anal. Biochem.* 343, 54–65. doi:10.1016/j.ab.2005.04.035.
- Granier, C., and Vile, D. (2014). Phenotyping and beyond: Modelling the relationships between traits. *Curr. Opin. Plant Biol.* 18, 96–102. doi:10.1016/j.pbi.2014.02.009.
- Großkinsky, D. K., Jaya Syaifullah, S., Roitsch, T., Lim, P., Syaifullah, S. J., and Roitsch, T. (2018). Integration of multi-omics techniques and physiological phenotyping within a holistic phenomics approach to study senescence in model and crop plants. *J. Exp. Bot.* 69, 825–844. doi:10.1093/jxb/erx333.
- Großkinsky, D. K., Svensgaard, J., Christensen, S., and Roitsch, T. (2015). Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. *J. Exp. Bot.* 66, 5429–5440. doi:10.1093/jxb/erv345.
- Grosu, H., Schaeffer, L., Oltenacu, P. A., Norman, D., Powell, R. L., Kremer, V., et al. (2013). *History of genetic evaluation methods in dairy cattle*. Available at: https://books.google.com/books/about/History_of_Genetic_Evaluation_Methods_in.html?id=dNbsoQEACAAJ [Accessed June 6, 2019].
- Guarini, A. R., Lourenco, D. A. L., Brito, L. F., Sargolzaei, M., Baes, C. F., Miglior, F., et al. (2019a). Genetics and genomics of reproductive disorders in Canadian Holstein cattle. *J. Dairy Sci.* 102, 1341–1353. doi:10.3168/jds.2018-15038.
- Guarini, A. R., Lourenco, D. A. L., Brito, L. F., Sargolzaei, M., Baes, C. F., Miglior, F., et al. (2019b). Use of a single-step approach for integrating foreign information into national genomic evaluation in Holstein cattle. *J. Dairy Sci.* 102, 8175–8183. doi:10.3168/jds.2018-15819.
- Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L., and Su, G. (2014a). Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* 15, 30. doi:10.1186/1471-2156-15-30.
- Guo, T., Yang, N., Tong, H., Pan, Q., Yang, X., Tang, J., et al. (2014b). Genetic basis of grain yield heterosis in an “immortalized F2” maize population. *Theor. Appl. Genet.* 127, 2149–2158. doi:10.1007/s00122-014-2368-x.
- Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397. doi:10.1534/genetics.107.081190.

- Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2009). Genomic selection using low-density marker panels. *Genetics* 182, 343–353. doi:10.1534/genetics.108.100289.
- Hall, B. P. (2015). Quantitative characterization of canopy coverage in the genetically diverse SOYNAM population. *Theses Diss. Available from ProQuest*. Available at: <https://docs.lib.purdue.edu/dissertations/AAI10053967> [Accessed February 8, 2018].
- Hallauer, A. R., Carena, M., and Miranda-Fihlo, J. B. (2010). *Quantitative genetics in maize breeding*. Available at: <https://books.google.com/books?id=jJWw3C84e0YC&pg=PA273&lpg=PA273&dq=increasing+genetic+gain+by+decreasing+environmental+variance&source=bl&ots=OWwTHIBI6j&sig=-506gLviLjckTmOq77v9wyDZ5yc&hl=en&sa=X&ved=0ahUKEwi20aeAuf3ZAhVDR6wKHVi6AbwQ6AEIXjAH#v=onepage&q> [Accessed March 21, 2018].
- Harfouche, A. L., Jacobson, D. A., Kainer, D., Romero, J. C., Harfouche, A. H., Scarascia Mugnozza, G., et al. (2019). Accelerating Climate Resilient Plant Breeding by Applying Next-Generation Artificial Intelligence. *Trends Biotechnol.* 37, 1217–1235. doi:10.1016/j.tibtech.2019.05.007.
- Hearst, A. A. (2019). Remote Sensing of Soybean Canopy Cover, Color, and Visible Indicators of Moisture Stress Using Imagery from Unmanned Aircraft Systems. doi:<https://doi.org/10.25394/PGS.8023478.v1>.
- Henderson, C. R. (1974). General Flexibility of Linear Model Techniques for Sire Evaluation. *J. Dairy Sci.* 57, 963–972. doi:10.3168/jds.s0022-0302(74)84993-3.
- Henderson, C. R., and Quaas, R. L. (1976). Multiple Trait Evaluation Using Relatives' Records. *J. Anim. Sci.* 43, 1188–1197. doi:10.2527/jas1976.4361188x.
- Henryon, M., Berg, P., and Sørensen, A. C. (2014). Invited review: Animal-breeding schemes using genomic information need breeding plans designed to maximise long-term genetic gains. *Livest. Sci.* 166, 38–47. doi:10.1016/j.livsci.2014.06.016.
- Holman, F. H., Riche, A. B., Michalski, A., Castle, M., Wooster, M. J., and Hawkesford, M. J. (2016). High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sens.* 8, 1031. doi:10.3390/rs8121031.
- Howard, R., and Jarquin, D. (2019). Genomic Prediction Using Canopy Coverage Image and Genotypic Information in Soybean via a Hybrid Model. *Evol. Bioinforma.* 15, 117693431984002. doi:10.1177/1176934319840026.
- <http://www.fas.usda.gov> Home | USDA Foreign Agricultural Service. Available at: <https://www.fas.usda.gov/> [Accessed March 13, 2018].

<https://www.cimmyt.org/> CIMMYT: International Maize and Wheat Improvement Center. Available at: <https://www.cimmyt.org/> [Accessed October 1, 2019].

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25, 295–309. doi:10.1016/0034-4257(88)90106-X.

Hughes, A. P., and Freeman, P. R. (1967). Growth Analysis Using Frequent Small Harvests. *J. Appl. Ecol.* 4, 553. doi:10.2307/2401356.

Hund, A., Trachsel, S., and Stamp, P. (2009). Growth of axile and lateral roots of maize: I development of a phenotyping platform. *Plant Soil* 325, 335–349. doi:10.1007/s11104-009-9984-2.

Hurtado, P. X., Schnabel, S. K., Zaban, A., Veteläinen, M., Virtanen, E., Eilers, P. H. C., et al. (2012). Dynamics of senescence-related QTLs in potato. *Euphytica* 183, 289–302. doi:10.1007/s10681-011-0464-4.

Husson, E., Lindgren, F., and Ecke, F. (2014). Assessing Biomass and Metal Contents in Riparian Vegetation Along a Pollution Gradient Using an Unmanned Aircraft System. *Water, Air, Soil Pollut.* 225, 1957. doi:10.1007/s11270-014-1957-2.

Huynh, H., and Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *J. Am. Stat. Assoc.* 65, 1582–1589. doi:10.1080/01621459.1970.10481187.

Inada, K. (1985). Spectral Ratio of Reflectance for Estimating Chlorophyll Content of Leaf. *Japanese J. Crop Sci.* 54, 261–272.

Iqbal, F., Lucieer, A., and Barry, K. (2018). Simplified radiometric calibration for UAS-mounted multispectral sensor. *Eur. J. Remote Sens.* 51, 301–313. doi:10.1080/22797254.2018.1432293.

Jamrozik, J., and Schaeffer, L. R. (1997). Estimates of Genetic Parameters for a Test Day Model with Random Regressions for Yield Traits of First Lactation Holsteins. *J. Dairy Sci.* 80, 762–770. doi:10.3168/jds.S0022-0302(97)75996-4.

Jannink, J.-L., Jordan, N. R., and Orf, J. H. (2001). Feasibility of selection for high weed suppressive ability in soybean: Absence of tradeoffs between rapid initial growth and sustained later growth. Kluwer Academic Publishers doi:10.1023/A:1017540800854.

Jannink, J. L., Orf, J. H., Jordan, N. R., and Shaw, R. G. (2000). Index selection for weed suppressive ability in soybean. *Crop Sci.* 40, 1087–1094. doi:10.2135/cropsci2000.4041087x.

Jarquín, D., Howard, R., Xavier, A., and Das Choudhury, S. (2018). Increasing Predictive Ability by Modeling Interactions between Environments, Genotype and Canopy Coverage Image Data for Soybeans. *Agronomy* 8, 51. doi:10.3390/agronomy8040051.

- Jia, Y., Jannink, J.-L. L., and Goddard, M. E. (2012). Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. *Genetics* 192. doi:10.1534/genetics.112.144246.
- Jiang, J., Xing, F., Wang, C., Zeng, X., and Zou, Q. (2019). Investigation and development of maize fused network analysis with multi-omics. *Plant Physiol. Biochem.* 141, 380–387. doi:10.1016/j.plaphy.2019.06.016.
- Jiang, J., Zhang, Q., Ma, L., Li, J., Wang, Z., and Liu, J. F. (2015). Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. *Heredity (Edinb)*. 115, 29–36. doi:10.1038/hdy.2015.9.
- Jiang, Z., Huete, A. R., Didan, K., and Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* 112, 3833–3845. doi:10.1016/j.rse.2008.06.006.
- Jimenez-Berni, J. A., Deery, D. M., Rozas-Larraondo, P., Condon, A. (Tony) G., Rebetzke, G. J., James, R. A., et al. (2018). High throughput determination of plant height, ground cover, and above-ground biomass in wheat with LiDAR. *Front. Plant Sci.* 9, 237. doi:10.3389/fpls.2018.00237.
- Jin, J., Liu, X., Wang, G., Mi, L., Shen, Z., Chen, X., et al. (2010). Agronomic and physiological contributions to the yield improvement of soybean cultivars released from 1950 to 2006 in Northeast China. *F. Crop. Res.* 115, 116–123. doi:10.1016/j.fcr.2009.10.016.
- Johannsen, W. (1909). *Elemente der exakten erblichkeitslehre. [Elements of the Exact Theory of Inheritance]*. G. Fischer, Jena doi:10.5962/bhl.title.94247.
- Johannsen, W. (1911). The Genotype Conception of Heredity. *Am. Nat.* 45, 129–159. doi:10.1086/279202.
- Jones, D. B., Peterson, M. L., and Geng, S. (1979). Association Between Grain Filling Rate and Duration and Yield Components in Rice. *Crop Sci.* 19, 641. doi:10.2135/cropsci1979.0011183x001900050023x.
- Jordan, C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* 50, 663–666. doi:10.2307/1936256.
- Juliana, P., Montesinos-López, O. A., Crossa, J., Mondal, S., González Pérez, L., Poland, J., et al. (2018). Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. *Theor. Appl. Genet.* 132, 177–194. doi:10.1007/s00122-018-3206-3.
- Jumrani, K., and Bhatia, V. S. (2018). Impact of combined stress of high temperature and water deficit on growth and seed yield of soybean. *Physiol. Mol. Biol. Plants* 24, 37–50. doi:10.1007/s12298-017-0480-5.

- Kabelka, E. A., Diers, B. W., Fehr, W. R., LeRoy, A. R., Baianu, I. C., You, T., et al. (2004). Putative alleles for increased yield from soybean plant introductions. *Crop Sci.* 44, 784–791. doi:10.2135/cropsci2004.7840.
- Kaler, A. S., Ray, J. D., Schapaugh, W. T., Davies, M. K., King, C. A., and Purcell, L. C. (2018). Association mapping identifies loci for canopy coverage in diverse soybean genotypes. *Mol. Breed.* 38, 50. doi:10.1007/s11032-018-0810-5.
- Kang, H., Ning, C., Zhou, L., Zhang, S., Yan, Q., and Liu, J. F. (2018). Short communication: Single-step genomic evaluation of milk production traits using multiple-trait random regression model in Chinese Holsteins. *J. Dairy Sci.* 101, 11143–11149. doi:10.3168/jds.2018-15090.
- Kang, H., Zhou, L., Mrode, R., Zhang, Q., and Liu, J. F. (2017). Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. *Heredity (Edinb.)* 119, 459–467. doi:10.1038/hdy.2016.91.
- Kim, K. S., Diers, B. W., Hyten, D. L., Rouf Mian, M. A., Shannon, J. G., and Nelson, R. L. (2012). Identification of positive yield QTL alleles from exotic soybean germplasm in two backcross populations. *Theor. Appl. Genet.* 125, 1353–1369. doi:10.1007/s00122-012-1944-1.
- Kirkpatrick, M., and Heckman, N. (1989). A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *J. Math. Biol.* 27, 429–450. doi:10.1007/BF00290638.
- Kirkpatrick, M., Lofsvold, D., and Bulmer, M. (1990). Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124, 979–993. Available at: <https://www.genetics.org/content/genetics/124/4/979.full.pdf> [Accessed May 25, 2019].
- Knoch, D., Abadi, A., Grandke, F., Meyer, R. C., Samans, B., Werner, C. R., et al. (2020). Strong temporal dynamics of QTL action on plant growth progression revealed through high-throughput phenotyping in canola. *Plant Biotechnol. J.* 18, 68–82. doi:10.1111/pbi.13171.
- Koester, R. P., Nohl, B. M., Diers, B. W., and Ainsworth, E. A. (2016). Has photosynthetic capacity increased with 80 years of soybean breeding? An examination of historical soybean cultivars. *Plant Cell Environ.* 39, 1058–1067. doi:10.1111/pce.12675.
- Koester, R. P., Skoneczka, J. A., Cary, T. R., Diers, B. W., and Ainsworth, E. A. (2014). Historical gains in soybean (*Glycine max* Merr.) seed yield are driven by linear increases in light interception, energy conversion, and partitioning efficiencies. *J. Exp. Bot.* 65, 3311–3321. doi:10.1093/jxb/eru187.
- Koivula, M., Strandén, I., Pösö, J., Aamand, G. P., and Mäntysaari, E. A. (2015). Single-step genomic evaluation using multitrait random regression model and test-day data. *J. Dairy Sci.* 98, 2775–2784. doi:10.3168/jds.2014-8975.

- Konietschke, F., Bathke, A. C., Harrar, S. W., and Pauly, M. (2015). Parametric and nonparametric bootstrap methods for general MANOVA. *J. Multivar. Anal.* 140, 291–301. doi:10.1016/j.jmva.2015.05.001.
- Koutroubas, S. D., and Papakosta, D. K. (2010). Seed filling patterns of safflower: Genotypic and seasonal variations and association with other agronomic traits. *Ind. Crops Prod.* 31, 71–76. doi:10.1016/J.INDCROP.2009.09.014.
- Krishnamoorthy, K., and Lu, F. (2010). A parametric bootstrap solution to the MANOVA under heteroscedasticity. *J. Stat. Comput. Simul.* 80, 873–887. doi:10.1080/00949650902822564.
- Krishnamoorthy, K., and Yu, J. (2012). Multivariate Behrens-Fisher problem with missing data. *J. Multivar. Anal.* 105, 141–150. doi:10.1016/j.jmva.2011.08.019.
- Kross, A., McNairn, H., Lapen, D., Sunohara, M., and Champagne, C. (2015). Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *Int. J. Appl. Earth Obs. Geoinf.* 34, 235–248. doi:10.1016/j.jag.2014.08.002.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26. doi:10.18637/jss.v028.i05.
- Kumudini, S., Hume, D. J., and Chu, G. (2001). Genetic improvement in short season soybeans: I. Dry matter accumulation, partitioning, and leaf area duration. *Crop Sci.* 41, 391–398. doi:10.2135/cropsci2001.412391x.
- Kwak, I.-Y., Moore, C. R., Spalding, E. P., and Broman, K. W. (2016). Mapping Quantitative Trait Loci Underlying Function-Valued Traits Using Functional Principal Component Analysis and Multi-Trait Mapping. *G3 Genes, Genomes, Genet.* 6, 79–86. doi:10.1534/G3.115.024133.
- Kwak, I.-Y. Y., Moore, C. R., Spalding, E. P., and Broman, K. W. (2014). A simple regression-based method to map quantitative trait loci underlying function-valued phenotypes. *Genetics* 197. doi:10.1534/genetics.114.166306.
- Lado, B., Matus, I., Rodríguez, A., Inostroza, L., Poland, J., Belzile, F., et al. (2013). Increased Genomic Prediction Accuracy in Wheat Breeding Through Spatial Adjustment of Field Trial Data. *G3 & Genes/Genomes/Genetics* 3, 2105–2114. doi:10.1534/g3.113.007807.
- Lange, C. E., and Federizzi, L. C. (2009). Estimation of soybean genetic progress in the South of Brazil using multi-environmental yield trials. *Sci. Agric.* 66, 309–316. doi:10.1590/s0103-90162009000300005.
- Langridge, P., and Fleury, D. (2011). Making the most of “omics” for crop breeding. *Trends Biotechnol.* 29, 33–40. doi:10.1016/j.tibtech.2010.09.006.
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi:10.1038/nature14539.

- Lee, G., Nho, K., Kang, B., Sohn, K. A., Kim, D., Weiner, M. W., et al. (2019). Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Sci. Rep.* 9, 1952. doi:10.1038/s41598-018-37769-z.
- Lee, K. J., and Lee, B. W. (2013). Estimation of rice growth and nitrogen nutrition status using color digital camera image analysis. *Eur. J. Agron.* 48, 57–65. doi:10.1016/j.eja.2013.02.011.
- Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663. doi:10.3168/jds.2009-2061.
- Legarra, A., Christensen, O. F., Aguilar, I., and Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livest. Sci.* 166, 54–65. doi:10.1016/j.livsci.2014.04.029.
- Leon, J., and Geisler, G. (1994). Variation in Rate and Duration of Growth Among Spring Barley Cultivars1. *Plant Breed.* 112, 199–208. doi:10.1111/j.1439-0523.1994.tb00671.x.
- Li, B., Xu, X., Zhang, L., Han, J., Bian, C., Li, G., et al. (2020). Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS J. Photogramm. Remote Sens.* 162, 161–172. doi:10.1016/J.ISPRSJPRS.2020.02.013.
- Li, D., Huang, Z., Song, S., Xin, Y., Mao, D., Lv, Q., et al. (2016). Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. *Proc. Natl. Acad. Sci.* 113, E6026–E6035. doi:10.1073/pnas.1610115113.
- Li, H., Rasheed, A., Hickey, L. T., and He, Z. (2018). Fast-Forwarding Genetic Gain. *Trends Plant Sci.* 23, 184–186. doi:10.1016/j.tplants.2018.01.007.
- Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors (Switzerland)* 14, 20078–20111. doi:10.3390/s141120078.
- Li, Z., and Sillanpää, M. J. (2013). A Bayesian nonparametric approach for mapping dynamic quantitative traits. *Genetics* 194, 997–1016. doi:10.1534/genetics.113.152736.
- Li, Z., and Sillanpää, M. J. (2015). Dynamic Quantitative Trait Locus Analysis of Plant Phenomic Data. doi:10.1016/j.tplants.2015.08.012.
- Lillehammer, M., Ødegård, J., and Meuwissen, T. H. (2007). Random regression models for detection of gene by environment interaction. *Genet. Sel. Evol.* 39, 105. doi:10.1186/1297-9686-39-2-105.
- Lin, Z., Hayes, B. J., and Daetwyler, H. D. (2014). Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci.* 65, 1177. doi:10.1071/CP13363.
- Littell, R. C. . (1990). Analysis of repeated measures data. in *Conference on Applied Statistic in Agriculture* doi:10.1007/978-981-10-3794-8.

- Littell, R. C. ., Henry, P. R. ., and Ammerman, C. B. (1998). Statistical analysis of repeated measures data using SAS procedures. *J. Anim. Sci.* 76, 1216.
- Littell, R. C., Pendergast, J., and Natarajan, R. (2000). Tutorial in Biostatistics: Modelling covariance structure in the analysis of repeated measures data. *Stat. Med.* 19, 1793–1819. doi:10.1002/1097-0258(20000715)19:13<1793::AID-SIM482>3.3.CO;2-H.
- Liu, X., Dong, X., Xue, Q., Leskovar, D. I., Jifon, J., Butnor, J. R., et al. (2018a). Ground penetrating radar (GPR) detects fine roots of agricultural crops in the field. *Plant Soil* 423, 517–531. doi:10.1007/s11104-017-3531-3.
- Liu, X., Wang, H., Hu, X., Li, K., Liu, Z., Wu, Y., et al. (2019). Improving Genomic Selection With Quantitative Trait Loci and Nonadditive Effects Revealed by Empirical Evidence in Maize. *Front. Plant Sci.* 10, 1129. doi:10.3389/fpls.2019.01129.
- Liu, X., Wang, H., Wang, H., Guo, Z., Xu, X., Liu, J., et al. (2018b). Factors affecting genomic selection revealed by empirical evidence in maize. *Crop J.* 6, 341–352. doi:10.1016/j.cj.2018.03.005.
- Lopez, M. A., Xavier, A., and Rainey, K. M. (2019). Phenotypic Variation and Genetic Architecture for Photosynthesis and Water Use Efficiency in Soybean (*Glycine max* L. Merr). *Front. Plant Sci.* 10, 680. doi:10.3389/fpls.2019.00680.
- Ly, D., Huet, S., Gauffreteau, A., Rincant, R., Touzy, G., Mini, A., et al. (2018). Whole-genome prediction of reaction norms to environmental stress in bread wheat (*Triticum aestivum* L.) by genomic random regression. *F. Crop. Res.* 216, 32–41. doi:10.1016/J.FCR.2017.08.020.
- Lyu, B., Smith, S. D., Xue, Y., and Cherkauer, K. A. (2019). Deriving Vegetation Indices from High-throughput Images by Using Unmanned Aerial Systems in Soybean Breeding. in *2019 ASABE Annual International Meeting* doi:10.13031/aim.201900279.
- Ma, C. X., Casella, G., and Wu, R. (2002). Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 161, 1751–1762. Available at: <http://www.genetics.org/content/genetics/161/4/1751.full.pdf> [Accessed April 3, 2018].
- Ma, L., Shi, Y., Siemianowski, O., Yuan, B., Egner, T. K., Mirnezami, S. V., et al. (2019). Hydrogel-based transparent soils for root phenotyping in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 166, 11063–11068. doi:10.1073/pnas.1820334116.
- Ma, W., Qiu, Z., Song, J., Li, J., Cheng, Q., Zhai, J., et al. (2018). A deep convolutional neural network approach for predicting phenotypes from genotypes. *Planta* 248, 1307–1318. doi:10.1007/s00425-018-2976-9.

- Macciotta, N. P. P., Biffani, S., Bernabucci, U., Lacetera, N., Vitali, A., Ajmone-Marsan, P., et al. (2017). Derivation and genome-wide association study of a principal component-based measure of heat tolerance in dairy cattle. *J. Dairy Sci.* 100, 4683–4697. doi:10.3168/jds.2016-12249.
- Macciotta, N. P. P., Gaspa, G., Bomba, L., Vicario, D., Dimauro, C., Cellesi, M., et al. (2015). Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel. *J. Dairy Sci.* 98, 8175–8185. doi:10.3168/jds.2015-9500.
- Macciotta, N. P. P., Vicario, D., and Cappio-Borlino, A. (2006). Use of multivariate analysis to extract latent variables related to level of production and lactation persistency in dairy cattle. *J. Dairy Sci.* 89, 3188–3194. doi:10.3168/jds.S0022-0302(06)72593-0.
- Madec, S., Baret, F., de Solan, B., Thomas, S., Dutartre, D., Jezequel, S., et al. (2017). High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates. *Front. Plant Sci.* 8, 2002. doi:10.3389/fpls.2017.02002.
- Mahlein, A.-K. (2016). Present and Future Trends in Plant Disease Detection. *Plant Dis.* 100, 1–11. doi:10.1007/s13398-014-0173-7.2.
- Maimaitijiang, M., Sagan, V., Sidike, P., Maimaitiyiming, M., Hartling, S., Peterson, K. T., et al. (2019). Vegetation Index Weighted Canopy Volume Model (CVM VI) for soybean biomass estimation from Unmanned Aerial System-based RGB imagery. *ISPRS J. Photogramm. Remote Sens.* 151, 27–41. doi:10.1016/j.isprsjprs.2019.03.003.
- Marquet, P. A., Quiñones, R. A., Abades, S., Labra, F., Tognelli, M., Arim, M., et al. (2005). Scaling and power-laws in ecological systems. *J. Exp. Biol.* 208, 1749–1769. doi:10.1242/jeb.01588.
- Meade, K. A., Cooper, M., and Beavis, W. D. (2013). Modeling biomass accumulation in maize kernels. *F. Crop. Res.* 151, 92–100. doi:10.1016/j.fcr.2013.07.014.
- Metzner, R., Eggert, A., van Dusschoten, D., Pflugfelder, D., Gerth, S., Schurr, U., et al. (2015). Direct comparison of MRI and X-ray CT technologies for 3D imaging of root systems in soil: Potential and challenges for root trait quantification. *Plant Methods* 11, 17. doi:10.1186/s13007-015-0060-z.
- Meuwissen, T. H. (2009). Accuracy of breeding values of “unrelated” individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* 41, 35. doi:10.1186/1297-9686-41-35.
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi:11290733.
- Meuwissen, T. H. E., Svendsen, M., Solberg, T., and Ødegård, J. (2015). Genomic predictions based on animal models using genotype imputation on a national scale in Norwegian Red cattle. *Genet. Sel. Evol.* 47, 79. doi:10.1186/s12711-015-0159-8.

- Meuwissen, T., Hayes, B., and Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding. *Anim. Front.* 6, 6–14. doi:10.2527/af.2016-0002.
- Meyer, K. (1998). Estimating covariance functions for longitudinal data using a random regression model. *Genet. Sel. Evol.* 30, 221–240. doi:10.1186/1297-9686-30-3-221.
- Meyer, K. (2005a). Advances in methodology for random regression analyses. in *Australian Journal of Experimental Agriculture*, 847–858. doi:10.1071/EA05040.
- Meyer, K. (2005b). Random regression analyses using B-splines to model growth of Australian Angus cattle. *Genet. Sel. Evol.* 37, 473. doi:10.1186/1297-9686-37-6-473.
- Meyer, K. (2019). “Bending” and beyond: Better estimates of quantitative genetic parameters? *J. Anim. Breed. Genet.* 136, 243–251. doi:10.1111/jbg.12386.
- Meyer, K., and Hill, W. G. (1997). Estimation of genetic and phenotypic covariance functions for longitudinal or “repeated” records by restricted maximum likelihood. *Livest. Prod. Sci.* 47, 185–200. doi:10.1016/S0301-6226(96)01414-5.
- Meyer, K., and Kirkpatrick, M. (2005). Up hill, down dale: quantitative genetics of curvaceous traits. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 1443–1455. doi:10.1098/rstb.2005.1681.
- Miglior, F., Fleming, A., Malchiodi, F., Brito, L. F., Martin, P., and Baes, C. F. (2017). A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. *J. Dairy Sci.* 100, 10251–10271. doi:10.3168/jds.2017-12968.
- Misztal, I., Legarra, A., and Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92, 4648–4655. doi:10.3168/JDS.2009-2064.
- Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D. H. (2002). BLUPF90 and related programs (BGF90). in *Proc. 7th World Congress on Genetics Applied to Livestock Production* (Montpellier, France), 21–22. Available at: <https://orbi.uliege.be/handle/2268/84980> [Accessed February 12, 2020].
- Momen, M., Campbell, M. T., Walia, H., and Morota, G. (2019). Predicting Longitudinal Traits Derived from High-Throughput Phenomics in Contrasting Environments Using Genomic Legendre Polynomials and B-Splines. *G3 (Bethesda)*. 9, 3369–3380. doi:10.1534/g3.119.400346.
- Monir, M. M., and Zhu, J. (2018). Dominance and Epistasis Interactions Revealed as Important Variants for Leaf Traits of Maize NAM Population. *Front. Plant Sci.* 9, 627. doi:10.3389/fpls.2018.00627.
- Monteith, J. L. (1972). Solar radiation and productivity in tropical ecosystems. *J. Appl. Ecol.* 9, 747–766. doi:10.2307/2401901.

- Monteith, J. L. (1977). Climate and the efficiency of crop production in Britain. *Philos. Trans. R. Soc. London* 281, 277–294.
- Monteith, J. L., and Moss, C. J. (1977). Climate and the Efficiency of Crop Production in Britain [and Discussion]. *Philos. Trans. R. Soc. B Biol. Sci.* 281, 277–294. doi:10.1098/rstb.1977.0140.
- Montes, J. M., Melchinger, A. E., and Reif, J. C. (2007). Novel throughput phenotyping platforms in plant genetic studies. *Trends Plant Sci.* 12, 433–436. doi:10.1016/j.tplants.2007.08.006.
- Montes, J. M., Technow, F., Dhillon, B. S., Mauch, F., and Melchinger, A. E. (2011). High-throughput non-destructive biomass determination during early plant development in maize under field conditions. *F. Crop. Res.* 121, 268–273. doi:10.1016/J.FCR.2010.12.017.
- Montesinos-López, A., Montesinos-López, O. A., Gianola, D., Crossa, J., and Hernández-Suárez, C. M. (2018a). Multi-environment genomic prediction of plant traits using deep learners with dense architecture. *G3 Genes, Genomes, Genet.* 8, 3813–3828. doi:10.1534/g3.118.200740.
- Montesinos-López, O. A., Martín-Vallejo, J., Crossa, J., Gianola, D., Hernández-Suárez, C. M., Montesinos-López, A., et al. (2019). New Deep Learning Genomic-Based Prediction Model for Multiple Traits with Binary, Ordinal, and Continuous Phenotypes. *G3 (Bethesda)*. 9, 1545–1556. doi:10.1534/g3.119.300585.
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Gianola, D., Hernández-Suárez, C. M., and Martín-Vallejo, J. (2018b). Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction of plant traits. *G3 Genes, Genomes, Genet.* 8, 3829–3840. doi:10.1534/g3.118.200728.
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., los Campos, G., Alvarado, G., Suchismita, M., et al. (2017). Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods* 13, 4. doi:10.1186/s13007-016-0154-2.
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Toledo, F. H., Pérez-Hernández, O., Eskridge, K. M., et al. (2016). A Genomic Bayesian Multi-trait and Multi-environment Model. *G3 (Bethesda)*. 6, 2725–44. doi:10.1534/g3.116.032359.
- Moose, S. P., and Mumm, R. H. (2008). Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement. *PLANT Physiol.* 147, 969–977. doi:10.1104/pp.108.118232.
- Moreira, F. F., Hearst, A. A., Cherkauer, K. A., and Rainey, K. M. (2019). Improving the efficiency of soybean breeding with high-throughput canopy phenotyping. *Plant Methods* 15, 139. doi:10.1186/s13007-019-0519-4.
- Morrison, M. J., Voldeng, H. D., and Cober, E. R. (2000). Agronomic changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. *Agron. J.* 92, 780–784. doi:10.2134/agronj2000.924780x.

- Mrode, R. (2014). *Linear models for the prediction of animal breeding values.* , ed. R. Mrode Wallingford: CABI doi:10.1079/9781780643915.0000.
- Muir, B. L., Kistemaker, G., Jamrozik, J., and Canavesi, F. (2007). Genetic Parameters for a Multiple-Trait Multiple-Lactation Random Regression Test-Day Model in Italian Holsteins. *J. Dairy Sci.* 90, 1564–1574. doi:10.3168/jds.S0022-0302(07)71642-9.
- Muraya, M. M., Chu, J., Zhao, Y., Junker, A., Klukas, C., Reif, J. C., et al. (2017). Genetic variation of growth dynamics in maize (*Zea mays* L.) revealed through automated non-invasive phenotyping. *Plant J.* 89, 366–380. doi:10.1111/tpj.13390.
- Neath, A. A., and Cavanaugh, J. E. (2012). The Bayesian information criterion: Background, derivation, and applications. *Wiley Interdiscip. Rev. Comput. Stat.* 4, 199–203. doi:10.1002/wics.199.
- Neilson, E. H., Edwards, A. M., Blomstedt, C. K., Berger, B., Møller, B. L., and Gleadow, R. M. (2015). Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. *J. Exp. Bot.* 66, 1817–1832. doi:10.1093/jxb/eru526.
- Neumann, K., Zhao, Y., Chu, J., Keilwagen, J., Reif, J. C., Kilian, B., et al. (2017). Genetic architecture and temporal patterns of biomass accumulation in spring barley revealed by image analysis. *BMC Plant Biol.* 17, 137. doi:10.1186/s12870-017-1085-4.
- Nie, S., and Xu, H. (2016). Riboflavin-induced disease resistance requires the mitogen-activated protein kinases 3 and 6 in *Arabidopsis thaliana*. *PLoS One* 11, e0153175. doi:10.1371/journal.pone.0153175.
- Ning, C., Kang, H., Zhou, L., Wang, D., Wang, H., Wang, A., et al. (2017). Performance Gains in Genome-Wide Association Studies for Longitudinal Traits via Modeling Time-varied effects. *Sci. Rep.* 7, 590. doi:10.1038/s41598-017-00638-2.
- Ning, C., Wang, D., Zheng, X., Zhang, Q., Zhang, S., Mrode, R., et al. (2018). Eigen decomposition expedites longitudinal genome-wide association studies for milk production traits in Chinese Holstein. *Genet. Sel. Evol.* 50, 12. doi:10.1186/s12711-018-0383-0.
- Nobre, P. R. C., Misztal, I., Tsuruta, S., Bertrand, J. K., Silva, L. O. C., and Lopes, P. S. (2003). Analyses of growth curves of Nellore cattle by multiple-trait and random regression models. *J. Anim. Sci.* 81, 918–926. doi:10.2527/2003.814918x.
- Nolan, E., and Santos, P. (2012). The contribution of genetic modification to changes in corn yield in the United States. *Am. J. Agric. Econ.* 94, 1171–1188. doi:10.1093/ajae/aas069.
- Oliveira, H. R., Brito, L. F., Lourenco, D. A. L., Silva, F. F., Jamrozik, J., Schaeffer, L. R., et al. (2019a). Invited review: Advances and applications of random regression models: From quantitative genetics to genomics. *J. Dairy Sci.* doi:10.3168/jds.2019-16265.

- Oliveira, H. R., Brito, L. F., Silva, F. F., Lourenco, D. A. L., Jamrozik, J., and Schenkel, F. S. (2019b). Genomic prediction of lactation curves for milk, fat, protein, and somatic cell score in Holstein cattle. *J. Dairy Sci.* 102, 452–463. doi:10.3168/jds.2018-15159.
- Oliveira, H. R., Lourenco, D. A. L., Masuda, Y., Misztal, I., Tsuruta, S., Jamrozik, J., et al. (2019c). Single-step genome-wide association for longitudinal traits of Canadian Ayrshire, Holstein, and Jersey dairy cattle. *J. Dairy Sci.* 102, 9995–10011. doi:10.3168/jds.2019-16821.
- Oliveira, H. R. R., Lourenco, D. A. L. A. L., Masuda, Y., Misztal, I., Tsuruta, S., Jamrozik, J., et al. (2019d). Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. *J. Dairy Sci.* 102, 2365–2377. Available at: <https://www.sciencedirect.com/science/article/pii/S0022030219300451> [Accessed September 30, 2019].
- Oliveira, H. R., Silva, F. F., Siqueira, O. H. G. B. D. G. B. D., Souza, N. O., Junqueira, V. S., Resende, M. D. V. V., et al. (2016). Combining different functions to describe milk, fat, and protein yield in goats using Bayesian multiple-trait random regression models. *J. Anim. Sci.* 94, 1865–1874. doi:10.2527/jas.2015-0150.
- Orf, J. H., Chase, K., Adler, F. R., Mansur, L. M., and Lark, K. G. (1999). Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. *Crop Sci.* 39, 1652–1657. doi:10.2135/cropsci1999.3961652x.
- Orf, J. H., Diers, B. W., and Boerma, H. R. (2004). Genetic Improvement: Conventional and Molecular-Based Strategies. *Soybeans Improv. Prod. Uses* 8, 417–450. doi:10.2134/agronmonogr16.3ed.c9.
- Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man. Cybern.* 9, 62–66. doi:10.1109/TSMC.1979.4310076.
- Paine, C. E. T. T., Marthews, T. R., Vogt, D. R., Purves, D., Rees, M., Hector, A., et al. (2012). How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods Ecol. Evol.* 3, 245–256. doi:10.1111/j.2041-210X.2011.00155.x.
- Pan, W. J., Wang, X., Deng, Y. R., Li, J. H., Chen, W., Chiang, J. Y., et al. (2015). Nondestructive and intuitive determination of circadian chlorophyll rhythms in soybean leaves using multispectral imaging. *Sci. Rep.* 5, 11108. doi:10.1038/srep11108.
- Pandey, V., Singh, M., Pandey, D., and Kumar, A. (2018). Integrated proteomics, genomics, metabolomics approaches reveal oxalic acid as pathogenicity factor in *Tilletia indica* inciting Karnal bunt disease of wheat. *Sci. Rep.* 8, 7826. doi:10.1038/s41598-018-26257-z.
- Patterson, J., and Gibson, A. (2017). *Deep learning : a practitioner's approach*. O'Reilly Media.

- Pauli, D., Andrade-Sanchez, P., Carmo-Silva, A. E., Gazave, E., French, A. N., Heun, J., et al. (2016a). Field-Based High-Throughput Plant Phenotyping Reveals the Temporal Patterns of Quantitative Trait Loci Associated with Stress-Responsive Traits in Cotton. *G3 Genes/Genomes/Genetics* 6, 865. doi:10.1534/G3.115.023515.
- Pauli, D., Chapman, S. C., Bart, R., Topp, C. N., Lawrence-Dill, C. J., Poland, J., et al. (2016b). The quest for understanding phenotypic variation via integrated approaches in the field environment. *Plant Physiol.* 172, pp.00592.2016. doi:10.1104/pp.16.00592.
- Pembleton, L. W., Drayton, M. C., Bain, M., Baillie, R. C., Inch, C., Spangenberg, G. C., et al. (2016). Targeted genotyping-by-sequencing permits cost-effective identification and discrimination of pasture grass species and cultivars. *Theor. Appl. Genet.* 129, 991–1005. doi:10.1007/s00122-016-2678-2.
- Peñuelas, J., and Filella, L. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends Plant Sci.* 3, 151–156. doi:10.1016/S1360-1385(98)01213-8.
- Pereira, R. J., Bignardi, A. B., El Faro, L., Verneque, R. S., Vercesi Filho, A. E., and Albuquerque, L. G. (2013). Random regression models using Legendre polynomials or linear splines for test-day milk yield of dairy Gyr (*Bos indicus*) cattle. *J. Dairy Sci.* 96, 565–574. doi:10.3168/JDS.2011-5051.
- Pérez-Bueno, M. L., Pineda, M., Cabeza, F. M., and Barón, M. (2016). Multicolor Fluorescence Imaging as a Candidate for Disease Detection in Plant Phenotyping. *Front. Plant Sci.* 7, 1790. doi:10.3389/fpls.2016.01790.
- Pérez-Enciso, M., and Zingaretti, L. M. (2019). A Guide for Using Deep Learning for Complex Trait Genomic Prediction. *Genes (Basel)*. 10. doi:10.3390/genes10070553.
- Pfeifer, J., Kirchgessner, N., Colombi, T., and Walter, A. (2015). Rapid phenotyping of crop root systems in undisturbed field soils using X-ray computed tomography. *Plant Methods* 11, 41. doi:10.1186/s13007-015-0084-4.
- Pflugfelder, D., Metzner, R., Dusschoten, D., Reichel, R., Jahnke, S., and Koller, R. (2017). Non-invasive imaging of plant roots in different soils using magnetic resonance imaging (MRI). *Plant Methods* 13, 102. doi:10.1186/s13007-017-0252-9.
- Piepho, H. P., Bu'chse, A. B., Richter, C., Bu'chse, D. A. B., Büchse, A., and Richter, C. (2004). A mixed modelling approach for randomized experiments with repeated measures. *J. Agron. Crop Sci.* 190, 230–247. doi:10.1111/j.1439-037X.2004.00097.x.
- Piepho, H. P., Möhring, J., Melchinger, A. E., and Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161, 209–228. doi:10.1007/s10681-007-9449-8.

- Pinheiro, J., and Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag doi:10.1007/b98882.
- Pölönen, I., Saari, H., Kaivosoja, J., Honkavaara, E., and Pesonen, L. (2013). Hyperspectral imaging based biomass and nitrogen content estimations from light-weight UAV. in, eds. C. M. U. Neale and A. Maltese (International Society for Optics and Photonics), 88870J. doi:10.1117/12.2028624.
- Poorter, H. (1989). Plant growth analysis: towards a synthesis of the classical and the functional approach. *Physiol. Plant.* 75, 237–244. doi:10.1111/j.1399-3054.1989.tb06175.x.
- Poorter, H., Anten, N. P. R., and Marcelis, L. F. M. (2013). Physiological mechanisms in plant growth models: Do we need a supra-cellular systems biology approach? *Plant, Cell Environ.* 36, 1673–1690. doi:10.1111/pce.12123.
- Potgieter, A. B., George-Jaeggli, B., Chapman, S. C., Laws, K., Suárez Cadavid, L. A., Wixted, J., et al. (2017). Multi-Spectral Imaging from an Unmanned Aerial Vehicle Enables the Assessment of Seasonal Leaf Area Dynamics of Sorghum Breeding Lines. *Front. Plant Sci.* 8, 1532. doi:10.3389/fpls.2017.01532.
- Prasad, B., Carver, B. F., Stone, M. L., Babar, M. A., Raun, W. R., and Klatt, A. R. (2007). Genetic analysis of indirect selection for winter wheat grain yield using spectral reflectance indices. *Crop Sci.* 47, 1416–1425. doi:10.2135/cropsci2006.08.0546.
- Promislow, D. E. L., Tatar, M., Khazaeli, A. A., and Curtsinger, J. W. (1996). Age-specific patterns of genetic variance in *Drosophila melanogaster*. I. Mortality. *Genetics* 143, 839–848. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8725233> [Accessed May 22, 2019].
- Purcell, L. C. (2000). Soybean canopy coverage and light interception measurements using digital imagery. *Crop Sci.* 40, 834–837. doi:10.2135/cropsci2000.403834x.
- Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., and Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sens. Environ.* 48, 119–126. doi:10.1016/0034-4257(94)90134-1.
- R Core Team (2019). R: A language and environment for statistical computing. *Found. Stat. Comput. Vienna, Austria*. Available at: <https://www.r-project.org/> [Accessed October 3, 2019].
- Rahaman, M. M., Chen, D., Gillani, Z., Klukas, C., and Chen, M. (2015). Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front. Plant Sci.* 6, 619. doi:10.3389/fpls.2015.00619.
- Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS One* 8, e66428. doi:10.1371/journal.pone.0066428.

- Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., et al. (2020). Breeder friendly phenotyping. *Plant Sci.*, 110396. doi:10.1016/j.plantsci.2019.110396.
- Richards, F. J. (1959). A flexible growth function for empirical use. *J. Exp. Bot.* 10, 290–301. doi:10.1093/jxb/10.2.290.
- Richards, J. A., and Jia, X. (2006). *Remote Sensing Digital Image Analysis*. doi:10.1007/3-540-29711-1.
- Richards, R. A. (2000). Selectable traits to increase crop photosynthesis and yield of grain crops. *J. Exp. Bot.* 51, 447–458. doi:10.1093/jexbot/51.suppl_1.447.
- Richetti, J., Boote, K. J., Hoogenboom, G., Judge, J., Johann, J. A., and Uribe-Opazo, M. A. (2019). Remotely sensed vegetation index and LAI for parameter determination of the CSM-CROPGRO-Soybean model when in situ data are not available. *Int. J. Appl. Earth Obs. Geoinf.* 79, 110–115. doi:10.1016/j.jag.2019.03.007.
- Robert A. Schowengerdt (2012). *Remote Sensing: Models and Methods for Image Processing*. doi:10.1016/j.jenvman.2011.10.007.
- Rogers, J., Chen, P., Shi, A., Zhang, B., Scaboo, A., Smith, S. F., et al. (2015). Agronomic performance and genetic progress of selected historical soybean varieties in the southern USA. *Plant Breed.* 134, 85–93. doi:10.1111/pbr.12222.
- Roujean, J. L., and Breon, F. M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* 51, 375–384. doi:10.1016/0034-4257(94)00114-3.
- Rouse Jr., J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring vegetation systems in the great plains with erts. in *NASA SP-351, 3rd ERTS-1 Symposium*, 309–317. Available at: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740022614.pdf> [Accessed October 3, 2019].
- Rowell, J. G., and Walters, D. E. (1976). Analysing data with repeated observations on each experimental unit. *J. Agric. Sci.* 87, 423–432. doi:10.1017/S0021859600027763.
- Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., Crossa, J., et al. (2016). Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat. *G3 (Bethesda, Md.); Genes/Genomes/Genetics* 6, 2799–2808. doi:10.1534/g3.116.032888.
- Sagan, V., Maimaitijiang, M., Sidike, P., Eblimit, K., Peterson, K. T., Hartling, S., et al. (2019). UAV-based high resolution thermal imaging for vegetation monitoring, and plant phenotyping using ICI 8640 P, FLIR Vue Pro R 640, and thermomap cameras. *Remote Sens.* 11, 330. doi:10.3390/rs11030330.

- Salas Fernandez, M. G., Bao, Y., Tang, L., and Schnable, P. S. (2017). A High-Throughput, Field-Based Phenotyping Technology for Tall Biomass Crops. *Plant Physiol.* 174, 2008–2022. doi:10.1104/pp.17.00707.
- Sankaran, S., Zhou, J., Khot, L. R., Trapp, J. J., Mndolwa, E., and Miklas, P. N. (2018). High-throughput field phenotyping in dry bean using small unmanned aerial vehicle based multispectral imagery. *Comput. Electron. Agric.* 151, 84–92. doi:10.1016/j.compag.2018.05.034.
- Santana, M. L., Bignardi, A. B., Pereira, R. J., Menéndez-Buxadera, A., and El Faro, L. (2016). Random regression models to account for the effect of genotype by environment interaction due to heat stress on the milk yield of Holstein cows under tropical conditions. *J. Appl. Genet.* 57, 119–127. doi:10.1007/s13353-015-0301-x.
- Schaeffer, L. R. (2004). Application of random regression models in animal breeding. *Livest. Prod. Sci.* 86, 35–45. doi:10.1016/S0301-6226(03)00151-9.
- Schaeffer, L. R. (2016). Random Regression Models. 171.
- Scheiner, S. M., and Gurevitch, J. (2001). *Design and analysis of ecological experiments*. Oxford University Press.
- Schrag, T. A., Schipprack, W., and Melchinger, A. E. (2019). Across-years prediction of hybrid performance in maize using genomics. *Theor. Appl. Genet.* 132, 933–946. doi:10.1007/s00122-018-3249-5.
- Serrano, L., Filella, I., and Peñuelas, J. (2000). Remote sensing of biomass and yield of winter wheat under different nitrogen supplies. *Crop Sci.* 40, 723–731. doi:10.2135/cropsci2000.403723x.
- Shanahan, P. W., Binley, A., Whalley, W. R., and Watts, C. W. (2015). The Use of Electromagnetic Induction to Monitor Changes in Soil Moisture Profiles beneath Different Wheat Genotypes. *Soil Sci. Soc. Am. J.* 79, 459. doi:10.2136/sssaj2014.09.0360.
- Sheng, J., Zheng, X., Wang, J., Zeng, X., Zhou, F., Jin, S., et al. (2017). Transcriptomics and proteomics reveal genetic and biological basis of superior biomass crop *Miscanthus*. *Sci. Rep.* 7, 13777. doi:10.1038/s41598-017-14151-z.
- Shi, P., Tang, L., Lin, C., Liu, L., Wang, H., Cao, W., et al. (2015). Modeling the effects of post-anthesis heat stress on rice phenology. *F. Crop. Res.* 177, 26–36. doi:10.1016/j.fcr.2015.02.023.
- Shibles, R. M., and Weber, C. R. (1996). Interception of Solar Radiation and Dry Matter Production by Various Soybean Planting Patterns. *Crop Sci.* 6, 55. doi:10.2135/cropsci1966.0011183x000600010017x.

- Singh, A., Ganapathysubramanian, B., Singh, A. K. A., and Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends Plant Sci.* 21, 110–124. doi:10.1016/J.TPLANTS.2015.10.015.
- Smith, G. M., and Milton, E. J. (1999). The use of the empirical line method to calibrate remotely sensed data to reflectance. *Int. J. Remote Sens.* 20, 2653–2662. doi:10.1080/014311699211994.
- Sodini, S. M., Kemper, K. E., Wray, N. R., and Trzaskowski, M. (2018). Comparison of genotypic and phenotypic correlations: Cheverud's conjecture in humans. *Genetics* 209, 941–948. doi:10.1534/genetics.117.300630.
- Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2013). Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *PLoS One* 8, e54985. doi:10.1371/journal.pone.0054985.
- Song, Q., Yan, L., Quigley, C., Jordan, B. D., Fickus, E., Schroeder, S., et al. (2017). Genetic Characterization of the Soybean Nested Association Mapping Population. *Plant Genome* 10, 0. doi:10.3835/plantgenome2016.10.0109.
- SoyBase.org Available at: <https://soybase.org/> [Accessed March 16, 2020].
- Specht, J. E., Hume, D. J., and Kumudini, S. V. (1999). Soybean yield potential - A genetic and physiological perspective. in *Crop Science* (Crop Science Society of America), 1560–1570. doi:10.2135/cropsci1999.3961560x.
- Speidel, S. E. (2011). Random regression models for the prediction of days to finish in beef cattle. Available at: https://mountainscholar.org/bitstream/handle/10217/69294/Speidel_colostate_0053A_10777.pdf?sequence=1 [Accessed July 8, 2019].
- Sperisen, P., Cominetti, O., and Martin, F. P. J. (2015). Longitudinal omics modeling and integration in clinical metabonomics research: Challenges in childhood metabolic health research. *Front. Mol. Biosci.* 2, 44. doi:10.3389/fmolb.2015.00044.
- Stinchcombe, J. R., and Kirkpatrick, M. (2012). Genetics and evolution of function-valued traits: understanding environmentally responsive phenotypes. *Trends Ecol. Evol.* 27, 637–647. doi:10.1016/j.tree.2012.07.002.
- Stoskopf, N. C., Tomes, D. T. (Dwight T., and Christie, B. R. (Bertram R. . (1994). Plant breeding: theory and practice. *Choice Rev. Online* 32, 32-0301-32–0301. doi:10.5860/choice.32-0301.
- Sun, J., Poland, J. A., Mondal, S., Crossa, J., Juliana, P., Singh, R. P., et al. (2019). High-throughput phenotyping platforms enhance genomic selection for wheat grain yield across populations and cycles in early stage. *Theor. Appl. Genet.* 132, 1705–1720. doi:10.1007/s00122-019-03309-0.

- Sun, J., Rutkoski, J. E., Poland, J. A., Crossa, J., Jannink, J.-L., and Sorrells, M. E. (2017a). Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. *Plant Genome* 10, 0. doi:10.3835/plantgenome2016.11.0111.
- Sun, M., Goggi, S. A., Matson, K., Palmer, R. G., Moore, K., and Cianzio, S. R. (2015). Thin Plate Spline Regression Model Used at Early Stages of Soybean Breeding to Control Field Spatial Variation. *J. Crop Improv.* 29, 333–352. doi:10.1080/15427528.2015.1026623.
- Sun, S., Li, C., and Paterson, A. (2017b). In-Field High-Throughput Phenotyping of Cotton Plant Height Using LiDAR. *Remote Sens.* 9, 377. doi:10.3390/rs9040377.
- Sun, S., Li, C., Paterson, A. H., Jiang, Y., Xu, R., Robertson, J. S., et al. (2018). In-field High Throughput Phenotyping and Cotton Plant Growth Analysis Using LiDAR. *Front. Plant Sci.* 9, 16. doi:10.3389/fpls.2018.00016.
- Sung, J. M., Cho, I.-J., Sung, D., Kim, S., Kim, H. C., Chae, M.-H., et al. (2019). Development and verification of prediction models for preventing cardiovascular diseases. *PLoS One* 14, e0222809. doi:10.1371/journal.pone.0222809.
- Suravajhala, P., Kogelman, L. J. A., and Kadarmideen, H. N. (2016). Multi-omic data integration and analysis using systems genomics approaches: Methods and applications In animal production, health and welfare. *Genet. Sel. Evol.* 48, 38. doi:10.1186/s12711-016-0217-x.
- Szyda, J., Komisarek, J., and Antkowiak, I. (2014). Modelling effects of candidate genes on complex traits as variables over time. *Anim. Genet.* 45, 322–328. doi:10.1111/age.12144.
- Tai, A. P. K., Martin, M. V., and Heald, C. L. (2014). Threat to future global food security from climate change and ozone air pollution. *Nat. Clim. Chang.* 4, 817–821. doi:10.1038/nclimate2317.
- Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., and Bennett, M. (2017). Plant Phenomics, From Sensors to Knowledge. *Curr. Biol.* 27, R770–R783. doi:10.1016/j.cub.2017.05.055.
- Tattaris, M., Reynolds, M. P., and Chapman, S. C. (2016). A Direct Comparison of Remote Sensing Approaches for High-Throughput Phenotyping in Plant Breeding. *Front. Plant Sci.* 7, 1131. doi:10.3389/fpls.2016.01131.
- Tessmer, O. L., Jiao, Y., Cruz, J. A., Kramer, D. M., and Chen, J. (2013). Functional approach to high-throughput plant growth analysis. *BMC Syst. Biol.* 7, S17. doi:10.1186/1752-0509-7-S6-S17.
- Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science (80-.)*. 327, 818–822. doi:10.1126/science.1183700.

- Thompson, R., and Meyer, K. (1986). A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livest. Prod. Sci.* 15, 299–313. doi:10.1016/0301-6226(86)90071-0.
- Thoni, H., Neter, J., Wasserman, W., and Kutner, M. H. (1990). Applied Linear Regression Models. *Biometrics*. doi:10.2307/2531657.
- Thornley, J. H. M., France, J., and Cannell, M. G. R. (2005). An open-ended logistic-based growth function. *Ecol. Modell.* 184, 257–261. doi:10.1016/j.ecolmodel.2004.10.007.
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.
- Tilly, N., Hoffmeister, D., Cao, Q., Huang, S., Lenz-Wiedemann, V., Miao, Y., et al. (2014). Multitemporal crop surface models: accurate plant height measurement and biomass estimation with terrestrial laser scanning in paddy rice. *J. Appl. Remote Sens.* 8, 083671. doi:10.1117/1.JRS.8.083671.
- Tjørve, K. M. C., and Tjørve, E. (2017). The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. *PLoS One* 12, e0178691. doi:10.1371/journal.pone.0178691.
- Toenniessen, G. H. (2002). Crop Genetic Improvement for Enhanced Human Nutrition. *J. Nutr.* 132, 2943S–2946S. doi:10.1093/jn/132.9.2943s.
- Topcon, and Japan TOPCON. Available at: <https://www.topcon.co.jp/en/> [Accessed March 28, 2020].
- Topp, C. N., Iyer-Pascuzzi, A. S., Anderson, J. T., Lee, C.-R., Zurek, P. R., Symonova, O., et al. (2013). 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. *Proc. Natl. Acad. Sci.* 110, E1695–E1704. doi:10.1073/pnas.1304354110.
- Trabelsi, A., Chaabane, M., and Ben-Hur, A. (2019). Comprehensive evaluation of deep learning architectures for prediction of DNA/RNA sequence binding specificities. in *Bioinformatics* (Narnia), i269–i277. doi:10.1093/bioinformatics/btz339.
- Trachsel, S., Kaeppler, S. M., Brown, K. M., and Lynch, J. P. (2011). Shovelomics: High throughput phenotyping of maize (*Zea mays* L.) root architecture in the field. *Plant Soil* 341, 75–87. doi:10.1007/s11104-010-0623-8.
- Turra, E. M., de Oliveira, D. A. A., Valente, B. D., Teixeira, E. de A., Prado, S. de A., de Melo, D. C., et al. (2012). Estimation of genetic parameters for body weights of Nile tilapia *Oreochromis niloticus* using random regression models. *Aquaculture* 354–355, 31–37. doi:10.1016/j.aquaculture.2012.04.035.

- USDA (2018). USDA ERS - Soybeans & Oil Crops. *United States Dep. Agric. Econ. Res. Serv.* Available at: <https://www.ers.usda.gov/topics/crops/soybeans-oil-crops/> [Accessed March 13, 2018].
- van Dusschoten, D., Metzner, R., Kochs, J., Postma, J. A., Pflugfelder, D., Buehler, J., et al. (2016). Quantitative 3D Analysis of Plant Roots growing in Soil using Magnetic Resonance Imaging. *Plant Physiol.* 170, pp.01388.2015. doi:10.1104/pp.15.01388.
- van Eeuwijk, F. A., Bustos-Korts, D., Millet, E. J., Boer, M. P., Kruijer, W., Thompson, A., et al. (2018). Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Sci.* 282, 23–39. doi:10.1016/j.plantsci.2018.06.018.
- van Pelt, M. L., Meuwissen, T. H. E., de Jong, G., and Veerkamp, R. F. (2015). Genetic analysis of longevity in Dutch dairy cattle using random regression. *J. Dairy Sci.* 98, 4117–4130. doi:10.3168/jds.2014-9090.
- Vanhatalo, J., Li, Z., and Sillanpää, M. J. (2019). A Gaussian process model and Bayesian variable selection for mapping function-valued quantitative traits with incomplete phenotypic data. *Bioinformatics* 35, 3684–3692. doi:10.1093/bioinformatics/btz164.
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980.
- Vargas, G., Schenkel, F. S., Brito, L. F., Neves, H. H. de R., Munari, D. P., Boligon, A. A., et al. (2018). Unravelling biological biotypes for growth, visual score and reproductive traits in Nelore cattle via principal component analysis. *Livest. Sci.* 217, 37–43. doi:10.1016/j.livsci.2018.09.010.
- Varona, L., Legarra, A., Toro, M. A., and Vitezica, Z. G. (2018). Non-additive Effects in Genomic Selection. *Front. Genet.* 9, 78. doi:10.3389/fgene.2018.00078.
- Vasseur, F., Bresson, J., Wang, G., Schwab, R., and Weigel, D. (2018). Image-based methods for phenotyping growth dynamics and fitness components in *Arabidopsis thaliana*. *Plant Methods* 14, 63. doi:10.1186/s13007-018-0331-6.
- Velazco, J. G., Malosetti, M., Hunt, C. H., Mace, E. S., Jordan, D. R., and van Eeuwijk, F. A. (2019). Combining pedigree and genomic information to improve prediction quality: an example in sorghum. *Theor. Appl. Genet.*, 1–13. doi:10.1007/s00122-019-03337-w.
- Vernon, A. J., and Allison, J. C. S. (1963). A method of calculating net assimilation rate. *Nature* 200, 814. doi:10.1038/200814a0.
- Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., and Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sens. Environ.* 115, 3468–3478. doi:10.1016/j.rse.2011.08.010.

- Wade, K. M., Quaas, R. L., and Van Vleck, L. D. (1993). Estimation of the Parameters Involved in a First-Order Autoregressive Process for Contemporary Groups. *J. Dairy Sci.* 76, 3033–3040. doi:10.3168/JDS.S0022-0302(93)77643-2.
- Wang and L. A. Goonewardene, Z. (2004). The use of MIXED models in the analysis of animal experiments with repeated measures data. *Can. J. Anim. Sci.* 84, 1–11. doi:10.4141/a03-123.
- Wang, C., Nie, S., Xi, X., Luo, S., and Sun, X. (2017). Estimating the biomass of maize with hyperspectral and LiDAR data. *Remote Sens.* 9, 11. doi:10.3390/rs9010011.
- Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb)*. 94, 73–83. doi:10.1017/S0016672312000274.
- Wang, L., Zhou, X., Zhu, X., Dong, Z., and Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop J.* 4, 212–219. doi:10.1016/J.CJ.2016.01.008.
- Wang, X., Xu, Y., Hu, Z., and Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. *Crop J.* 6, 330–340. doi:10.1016/j.cj.2018.03.001.
- Wang, X., Zhang, R., Song, W., Han, L., Liu, X., Sun, X., et al. (2019). Dynamic plant height QTL revealed in maize through remote sensing phenotyping using a high-throughput unmanned aerial vehicle (UAV). *Sci. Rep.* 9, 3458. doi:10.1038/s41598-019-39448-z.
- Wang, Y., Duby, G., Purnelle, B., and Boutry, M. (2000). Tobacco VDL gene encodes a plastid DEAD box RNA helicase and is involved in chloroplast differentiation and plant morphogenesis. *Plant Cell* 12, 2129–2142. doi:10.1105/tpc.12.11.2129.
- Whalley, W. R., Binley, A., Watts, C. W., Shanahan, P., Dodd, I. C., Ober, E. S., et al. (2017). Methods to estimate changes in soil water for phenotyping root activity in the field. *Plant Soil* 415, 407–422. doi:10.1007/s11104-016-3161-1.
- Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58, 109–130. doi:10.1016/S0169-7439(01)00155-1.
- Wolfinger, R. D. (1996). Heterogeneous Variance-Covariance Structures for Repeated Measures. *J. Agric. Biol. Environ. Stat.* 1, 205–230.
- Wu, D., Guo, Z., Ye, J., Feng, H., Liu, J., Chen, G., et al. (2019). Combining high-throughput micro-CT-RGB phenotyping and genome-wide association study to dissect the genetic architecture of tiller growth in rice. *J. Exp. Bot.* 70, 545–561. doi:10.1093/jxb/ery373.
- Wu, R. L., and Lin, M. (2006). Functional mapping - how to map and study the genetic architecture of dynamic complex traits. *Nat. Rev. Genet.* 7, 229–237. doi:10.1038/nrg1804.

- Wu, R., Ma, C. X., Lin, M., Wang, Z., and Casella, G. (2004). Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model. *Biometrics* 60, 729–738. doi:10.1111/j.0006-341X.2004.00223.x.
- Würschum, T., Liu, W., Busemeyer, L., Tucker, M. R., Reif, J. C., Weissmann, E. A., et al. (2014). Mapping dynamic QTL for plant height in triticale. *BMC Genet.* 15, 59. doi:10.1186/1471-2156-15-59.
- Xavier, A., Hall, B., Hearst, A. A., Cherkauer, K. A., and Rainey, K. M. (2017). Genetic architecture of phenomic-enabled canopy coverage in Glycine max. *Genetics* 206, 1081–1089. doi:10.1534/genetics.116.198713.
- Xavier, A., Muir, W. M., and Rainey, K. M. (2016). Assessing Predictive Properties of Genome-Wide Selection in Soybeans. *G3 (Bethesda)*. 6, 2611–2616. doi:10.1534/g3.116.032268.
- Xavier, A., Xu, S., Muir, W. M., and Rainey, K. M. (2015). NAM: Association studies in multiple populations. *Bioinformatics* 31, 3862–3864. doi:10.1093/bioinformatics/btv448.
- Xu, R., Li, C., and Paterson, A. H. (2019). Multispectral imaging and unmanned aerial systems for cotton plant phenotyping. *PLoS One* 14, e0205083. doi:10.1371/journal.pone.0205083.
- Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., et al. (2017). Enhancing genetic gain in the era of molecular breeding. *Narnia* doi:10.1093/jxb/erx135.
- Yan, G., Liu, H., Wang, H., Lu, Z., Wang, Y., Mullan, D., et al. (2017). Accelerated Generation of Selfed Pure Line Plants for Gene Identification and Crop Breeding. *Front. Plant Sci.* 8, 1786. doi:10.3389/fpls.2017.01786.
- Yan, W., and Rajcan, I. (2003). Prediction of cultivar performance based on single- versus multiple-year tests in soybean. *Crop Sci.* 43, 549–555. doi:10.2135/CROPSCI2003.5490.
- Yang, G., Liu, J., Zhao, C., Li, Z. Z., Huang, Y., Yu, H., et al. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front. Plant Sci.* 8, 1111. doi:10.3389/fpls.2017.01111.
- Yang, R., Tian, Q., and Xu, S. (2006). Mapping quantitative trait loci for longitudinal traits in line crosses. *Genetics* 173, 2339–56. doi:10.1534/genetics.105.054775.
- Yang, R., and Xu, S. (2007). Bayesian shrinkage analysis of quantitative trait Loci for dynamic traits. *Genetics* 176, 1169–85. doi:10.1534/genetics.106.064279.
- Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5, 5087. doi:10.1038/ncomms6087.

- Yano, K., Morinaka, Y., Wang, F., Huang, P., Takehara, S., Hirai, T., et al. (2019). GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture. *Proc. Natl. Acad. Sci.* 116, 201904964. doi:10.1073/pnas.1904964116.
- Yin, X., Goudriaan, J., Lantinga, E. A., Vos, J., and Spiertz, H. J. (2003). A flexible sigmoid function of determinate growth. *Ann. Bot.* 91, 361–371. doi:10.1093/aob/mcg029.
- York, L. M., Slack, S., Bennett, M. J., and Foulkes, M. J. (2018). Wheat shovelomics I: A field phenotyping approach for characterising the structure and function of root systems in tillering species. *bioRxiv*, 280875. doi:10.1101/280875.
- Yuan, J., Njiti, V. N., Meksem, K., Iqbal, M. J., Triwitayakorn, K., Kassem, M. A., et al. (2002). Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. *Crop Sci.* 42, 271–277. doi:10.2135/cropsci2002.0271.
- Yue, J., Yang, G., Li, C., Li, Z., Wang, Y., Feng, H., et al. (2017). Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models. *Remote Sens.* 9, 708. doi:10.3390/rs9070708.
- Zaman-Allah, M., Vergara, O., Araus, J. L., Tarekegne, A., Magorokosho, C., Zarco-Tejada, P. J., et al. (2015). Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. *Plant Methods* 11, 35. doi:10.1186/s13007-015-0078-2.
- Zhang, C., Du, X., Tang, K., Yang, Z., Pan, L., Zhu, P., et al. (2018a). Arabidopsis AGDP1 links H3K9me2 to DNA methylation in heterochromatin. *Nat. Commun.* 9, 4547. doi:10.1038/s41467-018-06965-w.
- Zhang, W., Gao, X., Shi, X., Zhu, B., Wang, Z., Gao, H., et al. (2018b). PCA-based multiple-trait GWAS analysis: A powerful model for exploring pleiotropy. 8. doi:10.3390/ani8120239.
- Zhang, X., Huang, C., Wu, D., Qiao, F., Li, W., Duan, L., et al. (2017). High-Throughput Phenotyping and QTL Mapping Reveals the Genetic Architecture of Maize Plant Growth. *Plant Physiol.* 173, 1554–1564. doi:10.1104/pp.16.01516.
- Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., et al. (2019). Crop Phenomics: Current Status and Perspectives. *Front. Plant Sci.* 10, 714. doi:10.3389/fpls.2019.00714.
- Zhong, Z., Sun, A. Y., Yang, Q., and Ouyang, Q. (2019). A deep learning approach to anomaly detection in geological carbon sequestration sites using pressure measurements. *J. Hydrol.* 573, 885–894. doi:10.1016/j.jhydrol.2019.04.015.
- Zhu, M., Chen, G., Dong, T., Wang, L., Zhang, J., Zhao, Z., et al. (2015). SIDEAD31, a putative DEAD-Box RNA helicase gene, regulates salt and drought tolerance and stress-related genes in tomato. *PLoS One* 10. doi:10.1371/journal.pone.0133849.

Zhu, X.-G., Long, S. P., and Ort, D. R. (2010). Improving Photosynthetic Efficiency for Greater Yield. *Annu. Rev. Plant Biol.* 61, 235–261. doi:10.1146/annurev-arplant-042809-112206.

Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., and Telenti, A. (2019). A primer on deep learning in genomics. *Nat. Genet.* 51, 12–18. doi:10.1038/s41588-018-0295-5.