Assessment and implementation of new breeding methods in the Kansas State winter wheat breeding program

by

Megan Calvert

B.Sc. Genetics, University of Pretoria, 2011
B.Sc. (Hons) Genetics, University of Pretoria, 2012
M.Sc. Genetics, University of Pretoria, 2017

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Abstract

Growing populations and shifting climatic conditions are placing constraints on global food security that have not previously been experienced. Novel technologies are being developed to combat this challenge at many levels including crop germplasm improvement. It is unlikely that all of these technologies will provide a significant benefit to crop breeding programs whilst still being economically viable. The use of high-throughput phenotyping (HTP) in crop breeding programs is becoming more commonplace due in part to the relatively low establishment costs for several of the technologies. Genomic prediction models are also becoming more common in crop breeding programs due to their success in animal breeding schemes and decreases in sequencing costs for genotyping. Here we examine the utility of several HTP technologies and genomic prediction in wheat breeding in Kansas. An uncrewed aerial system (UAS) measuring reflectance values at different bandwidths was used to formulate vegetation indices (VIs) which are known to correlate with economically valuable phenotypes. The genetic architecture of these VIs was examined in an association mapping population of winter wheat (*Triticum aestivum*) and their use for genomic prediction was examined in the Kansas State University (KSU) winter wheat breeding program. Several economic and population parameters were determined under which genomic prediction would be favored in the breeding program. Based on simulated and empirical observations of model accuracy the KSU breeding program could potentially make larger genetic gains using genomic prediction than traditional phenotypic selection methods.

Assessment and implementation of new breeding methods in the Kansas State winter wheat breeding program

by

Megan Calvert

B.Sc. Genetics, University of Pretoria, 2011
B.Sc. (Hons) Genetics, University of Pretoria, 2012
M.Sc. Genetics, University of Pretoria, 2017

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Approved by:

Major Professor
Dr. Jesse A. Poland

Copyright

MEGAN CALVERT

2020

Abstract

Growing populations and shifting climatic conditions are placing constraints on global food security that have not previously been experienced. Novel technologies are being developed to combat this challenge at many levels including crop germplasm improvement. It is unlikely that all of these technologies will provide a significant benefit to crop breeding programs whilst still being economically viable. The use of high-throughput phenotyping (HTP) in crop breeding programs is becoming more commonplace due in part to the relatively low establishment costs for several of the technologies. Genomic prediction models are also becoming more common in crop breeding programs due to their success in animal breeding schemes and decreases in sequencing costs for genotyping. Here we examine the utility of several HTP technologies and genomic prediction in wheat breeding in Kansas. An uncrewed aerial system (UAS) measuring reflectance values at different bandwidths was used to formulate vegetation indices (VIs) which are known to correlate with economically valuable phenotypes. The genetic architecture of these VIs was examined in an association mapping population of winter wheat (*Triticum aestivum*) and their use for genomic prediction was examined in the Kansas State University (KSU) winter wheat breeding program. Several economic and population parameters were determined under which genomic prediction would be favored in the breeding program. Based on simulated and empirical observations of model accuracy the KSU breeding program could potentially make larger genetic gains using genomic prediction than traditional phenotypic selection methods.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This degree would not have been completed without a great deal of support and encouragement. Even living on the other side of the world my family provided invaluable support and encouragement through some of the hardest moments. They are truly incredible, and I can never thank them enough.

No scientific endeavor is ever completed alone. I would like to thank all of the Undergraduate student research assistants, Quentin Ediger, Kyle Laessig, Jacob Ramsey, Jamie Clark, Molly Smith, Erika Kringen, Nicholas VanPelt, Bryce Teaford, and other lab members for the hours spent in fields and labs performing the same measurement hundreds of times. This work would have taken significantly longer if not for their help. A special thank you to Byron Evers, for coordinating the research tasks and resources, and to Xu Wang for showing endless patience, kindness, and massive amounts of computational analysis.

I would not be here if it were not for Kansas State University Counselling Services. The service that they provide allowed me to become the person that I am today. I cannot explain the impact that they had on my life.

Chapter 1 - Adapting New Technologies to Crop Development

Abbreviations

GEBV, genomic-estimated breeding value; HTP, High-throughput phenotyping; MAS, Marker-assisted selection; QTL, quantitative trait loci; UAS, uncrewed aerial systems; VIs, vegetation indices

Introduction

Changing climates and growing populations are increasing the need for stable and sustainable food production in challenging environmental conditions. This improvement also needs to occur at faster rates than are currently being achieved (Tester and Langridge 2010; Lenaerts, Collard, and Demont 2019). It is estimated that the current yield gains will need to approximately double to account for the increase in global population alone (Ray et al. 2013; Lenaerts, Collard, and Demont 2019). Changing and advancing technologies have the potential to close this short fall in yield gain if implemented efficiently and effectively. Yet the validation, implementation, and dissemination of these technologies to regions of the world where they may have the most impact is still proving to be a challenge (Lenaerts, Collard, and Demont 2019).

While these are noble goals that are for the greater global good, it is also important to remember that the main purpose of breeding programs is to be financially successful. The commercial value of the trait and delivery of improved varieties gives the determination of the breeding investment and priority for selection (Moose and

Mumm 2008). The new technologies may be available, but if they are too expensive to implement or don't provide the return on investment that is achieved through other methods, they will not be used regardless of the potential yield gain. Technological developments that promise revolutionary solutions will always be present, yet these technologies must be effectively integrated into the production systems to advance breeding programs. Therefore, critical assessment of the implementation of new technologies and selection methodologies into the breeding program are needed.

Genetic Gain per Cost

One of the ways that the success of a breeding program is measured is through the determination of the genetic gain of the program. The genetic gain is the change in the mean of a population that occurs through artificial selection (Falconer and Mackay 2009; Moose and Mumm 2008). Genetic gain is calculated by four main components that influence the progress of a breeding program: the phenotypic variation present in the population for a trait (σ_p), the heritability of the trait (h^2), the selection intensity(i), and the length of the breeding cycle (L , Equation 1, Falconer and Mackay 2009).

$$Genetic\ Gain = \Delta G = \frac{h^2 \sigma_p i}{L} \qquad \text{Equation 1}$$

These components all impact genetic gain and can be optimized with good breeding practices, improved selection methodologies and new technologies. A decrease in the length of the cycle causes a proportional increase in the genetic gain. The heritability of the trait can often be improved with good experimental design and a sufficient number of replicates. Whilst it may seem easy to increase genetic gain by

increasing the phenotypic variation, not all variation is beneficial, and a large increase in genetic variation due to the introduction of new germplasm often comes at the cost of an overall decrease in the population mean. The intensity of selection is typically a balance of achieving the desired genetic gain with a higher selection intensity, while not eliminating needed genetic variation through too high a selection intensity. All of these measures to improve the genetic gain then need to be optimized for the budgetary requirements of the specific breeding program. Determining the optimal number of replicates, trials, locations, technologies, phenotypes, and genotypes requires a fine balance of budgetary and breeding requirements.

Molecular Breeding Approaches

One of the new methodologies that can have an influence are those that use molecular breeding approaches. Molecular breeding approaches are those that make use of the genotypic or genetic marker information. Molecular breeding approaches seek to exploit the relationship between the genetic markers and a phenotype of economic interest (Lenaerts, Collard, and Demont 2019; Moose and Mumm 2008). Genotyping technologies have become more economically viable and widespread globally, making these approaches more available to plant breeders. This has resulted in the development of several techniques to optimally use the genetic marker/sequence information.

Marker-Assisted Selection

Marker-assisted selection (MAS) is a molecular breeding strategy that uses a single, or few, genetic markers which can be used to select for a target phenotype due to their linkage with genes of large effect that are underlying that phenotype (Moose and Mumm 2008). The justification for MAS has been reviewed multiple times and falls into four broad categories: 1) the ability to select for traits that are difficult to phenotype, 2) traits whose expression depends on environmental pressures that may not always be present, 3) selecting the desired alleles during backcrossing, and 4) the pyramiding of several genes for a specific phenotype (Tester and Langridge 2010; Xu and Crouch 2008; Álvarez, Mosquera, and Blair 2014). MAS and its associated techniques have been adapted for certain phenotypes in both the public and private breeding sector (Eathington et al. 2007).

One of the major drawbacks of using traditional MAS is the need for validated genes of a quantitative trait loci (QTL) or a qualitative phenotype (Lenaerts, Collard, and Demont 2019). This requires investment in research that may not directly reflect germplasm in the breeding program and a redistribution of resources that may not be feasible (Dekkers and Hospital 2002). This additional work can occasionally be mitigated through the use of association mapping to find QTL in breeding program data that is already available, or the integration of outside information from molecular laboratory studies. Unfortunately this often results in the detection of few QTL with inflated effects due to low-heritability, confounding population structure, small population size, and statistical error thresholds (Beavis 1998).

Genomic Prediction

Genomic Prediction is, at least conceptually, a form of MAS that estimates the effects of all of the markers at the same time which are summed to calculate a genomic-estimated breeding value (GEBV), (Meuwissen, Hayes, and Goddard 2001). The most explicit difference from MAS is that all of the genome-wide markers are used, as opposed to the identification and selection of only significant markers. This allows for the incorporation of many small-effect loci to be captured in the estimation of GEBVs, leading to a more accurate estimate of the genetic merit of each individual.

The estimation of GEBVs for breeding lines requires a training population that is phenotyped and genotyped. This training population is used to build a model which estimates the effect for all markers in the population and subsequent calculation of GEBVs (Meuwissen, Hayes, and Goddard 2001). From this point the breeding lines for selection do not need to be phenotyped and can be predicted directly from the genomic information. This provides the benefit of being able to predict phenotypes that are only observable under specific environmental conditions, e.g. disease pressure or abiotic stress, and has the potential to shorten the amount of time taken for each breeding cycle as favorable lines can be prioritized.

Non-Molecular Breeding Approaches

Traditional or non-molecular breeding approaches rely on phenotypic observation of the lines under evaluation in the environments that they will be grown in. When breeding for a large geographical area that may comprise several mega-

environments this can require an immense amount of resources. The correct investment of those resources can be crucial for the success of a breeding program. Accurate estimates of the phenotype of interest, or component phenotypes, is essential to traditional breeding approaches. Technologies that can assist with more rapid and accurate phenotyping are becoming available and more common place.

Phenomics

Phenomics involves the collection of high-dimensional phenotype data on an organism wide scale or over the life time of the organism (Houle, Govindaraju, and Omholt 2010). As the economic value of a crop is based on a phenotype or index of phenotypes, accurate and high dimensional data can be of great value to most breeding programs to aid in selection. There have been several areas in the field of phenomics that have seen recent advances which may be of use for crop improvement.

High-throughput phenotyping (HTP) covers a variety of technologies and software that work to speed up or ease the task of taking a large number of phenotypic measurements. HTP technologies have branched into two major systems, those systems that are automated and greenhouse based, and those systems that are used in field conditions (Dhondt, Wuyts, and Inzé 2013).

The use of automated systems in greenhouse experiments allows for precise phenotypes to be taken throughout the growth cycle. These systems are capable of measuring the phenotype at a finer resolution than can normally be achieved without the investment of significant personnel hours (Furbank and Tester 2011; Dhondt, Wuyts, and Inzé 2013). These systems are often not high throughput enough to use directly in

breeding program selections being limited in the number of plants that can be evaluated and hence limiting the population size. However, they contribute significantly to the understanding of underlying plant biology.

HTP systems that are to be used under field conditions fall into a variety of categories. These include systems that use ground vehicles, uncrewed aerial systems (UAS), and those that are human powered (Crain et al. 2018; Andrade-Sanchez et al. 2014; Araus and Cairns 2014). These systems are also used to take phenotypes throughout the course of the growth cycle but they are not of as fine a resolution as those produced by the greenhouse systems (Dhondt, Wuyts, and Inzé 2013).

A common set of phenotypes that are taken by the field based HTP systems are reflectance values per plant or plot. The reflectance values that are gathered cover the visible and invisible light spectrum range. Either these reflectance values themselves are correlated with a phenotype of economic interest, or they can be used to generate vegetation indices (VIs) which have been shown to be correlated with the phenotype (Araus and Cairns 2014; White et al. 2012).

Shuttle and Speed Breeding

Shuttle and speed breeding are both techniques that increase the number of generations that can be produced each year. As repeated phenotypic testing in different environments requires an inbred line, 5 or 6 generations are required to produce a fixed line. Shortening this interval in terms of years can allow for rapid advancement of lines.

Shuttle breeding is used to produce up to 2 or more generations a year depending on the crop species by using two separate sites with contrasting growing

seasons (Ortiz et al. 2007). This method may produce other benefits due to the exposure of the test varieties to different soil conditions and disease pressures. It can however be difficult to implement logistically if international borders need to be crossed and the various countries' differing regulations have to be accounted for. The expense of moving large amounts of plant material and the potential for loss during transport also make shuttle breeding a daunting investment for all but the largest breeding programs or most valuable crop species (Ortiz et al. 2007; Lenaerts, Collard, and Demont 2019).

Speed breeding is a recently developed technique that exploits an extended photoperiod and supra optimal temperatures to accelerate the development of a plant (Watson et al. 2018). This technique has produced remarkable results in several crop species already. For four of the major crop species, wheat, barley, chickpea, and canola, the time to anthesis was approximately half for the crops grown under speed breeding protocols when compared to those under glasshouse conditions (Watson et al. 2018). This has the potential to allow for up to 6 generations to be produced in one year for some crop species. Speed breeding does require an investment of resources for the initial set up of the system, but it has been shown that comparable results can be achieved with a small home-made system if required.

A Balanced Approach for the Future

Advances are being made to assist plant breeders in the production of varieties that can accommodate future global needs. The applicability of these advances to a specific breeding program depends on multiple factors and considerations. As the cost

of phenotyping a given trait increases, it is likely that the economics and return of investing in genomics and predicting larger populations than can be phenotyped is favorable. If the phenotype that is of most economic value is cheap and easy to obtain, the most economic value may come from investing in phenotypic selection and high-throughput phenotyping. There are other considerations such as the gain that the program is currently achieving, the established workflow, or the changing consumer desires that will also impact the decision of which techniques are best suited to the individual program.

References

Álvarez, María F., Teresa Mosquera, and Matthew W. Blair. 2014. "The Use of Association Genetics Approaches in Plant Breeding." In *Plant Breeding Reviews: Volume 38*, 17–68. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118916865.ch02.

Andrade-Sanchez, Pedro, Michael A. Gore, John T. Heun, Kelly R. Thorp, A. Elizabete Carmo-Silva, Andrew N. French, Michael E. Salvucci, and Jeffrey W. White. 2014. "Development and Evaluation of a Field-Based High-Throughput Phenotyping Platform." *Functional Plant Biology* 41 (1): 68. https://doi.org/10.1071/FP13126.

Araus, José Luis, and Jill E. Cairns. 2014. "Field High-Throughput Phenotyping: The New Crop Breeding Frontier." *Trends in Plant Science* 19 (1): 52–61. https://doi.org/10.1016/j.tplants.2013.09.008.

Beavis, William D. 1998. "QTL Analyses: Power, Precision, and Accuracy." *Molecular Dissection of Complex Traits*, 145–62.

Crain, Jared L., Yong Wei, Jared Barker, Sean M. Thompson, Phillip D. Alderman, Matthew Reynolds, Naiqian Zhang, and Jesse Poland. 2016. "Development and Deployment of a Portable Field Phenotyping Platform." *Crop Science* 56 (3): 965–75. https://doi.org/10.2135/cropsci2015.05.0290.

Dekkers, Jack C. M., and Frédéric Hospital. 2002. "The Use of Molecular Genetics in the Improvement of Agricultural Populations." *Nature Reviews Genetics* 3 (1): 22–32. https://doi.org/10.1038/nrg701.

Dhondt, Stijn, Nathalie Wuyts, and Dirk Inzé. 2013. "Cell to Whole-Plant Phenotyping: The Best Is yet to Come." *Trends in Plant Science* 18 (8): 428–39. https://doi.org/10.1016/j.tplants.2013.04.008.

Eathington, Sam R., Theodore M. Crosbie, Marlin D. Edwards, Robert S. Reiter, and Jason K. Bull. 2007. "Molecular Markers in a Commercial Breeding Program." *Crop Science* 47 (S3): S-154-S-163. https://doi.org/10.2135/cropsci2007.04.0015IPBS.

Falconer, Douglas S., and Trudy F. C. Mackay. 2009. *Introduction to Quantitative Genetics*. 4. ed., [16. print.]. Harlow: Pearson, Prentice Hall.

Furbank, Robert T., and Mark Tester. 2011. "Phenomics – Technologies to Relieve the Phenotyping Bottleneck." *Trends in Plant Science* 16 (12): 635–44. https://doi.org/10.1016/j.tplants.2011.09.005.

Haghighattalab, Atena, Lorena González Pérez, Suchismita Mondal, Daljit Singh, Dale Schinstock, Jessica Rutkoski, Ivan Ortiz-Monasterio, Ravi Prakash Singh, Douglas Goodin, and Jesse Poland. 2016. "Application of Unmanned Aerial Systems for High Throughput Phenotyping of Large Wheat Breeding Nurseries." *Plant Methods* 12 (1): 35. https://doi.org/10.1186/s13007-016-0134-6.

Houle, David, Diddahally R. Govindaraju, and Stig Omholt. 2010. “Phenomics: The next Challenge.” *Nature Reviews Genetics* 11 (12): 855–66. https://doi.org/10.1038/nrg2897.

Lenaerts, Bert, Bertrand C.Y. Collard, and Matty Demont. 2019. “Review: Improving Global Food Security through Accelerated Plant Breeding.” *Plant Science* 287 (October): 110207. https://doi.org/10.1016/j.plantsci.2019.110207.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. “Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps.” *Genetics* 157 (4): 1819–29.

Moose, Stephen P., and Rita H. Mumm. 2008. “Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement.” *Plant Physiology* 147 (3): 969–77. https://doi.org/10.1104/pp.108.118232.

Ortiz, Rodomiro, Richard Trethowan, Guillermo Ortiz Ferrara, Masa Iwanaga, John H. Dodds, Jonathan H. Crouch, Jose Crossa, and Hans-Joachim Braun. 2007. “High Yield Potential, Shuttle Breeding, Genetic Diversity, and a New International Wheat Improvement Strategy.” *Euphytica* 157 (3): 365–84. https://doi.org/10.1007/s10681-007-9375-9.

Ray, Deepak K., Nathaniel D. Mueller, Paul C. West, and Jonathan A. Foley. 2013. “Yield Trends Are Insufficient to Double Global Crop Production by 2050.” *PLoS ONE* 8 (6): e66428. https://doi.org/10.1371/journal.pone.0066428.

Sankaran, Sindhuja, Lav R. Khot, and Arron H. Carter. 2015. “Field-Based Crop Phenotyping: Multispectral Aerial Imaging for Evaluation of Winter Wheat Emergence and Spring Stand.” *Computers and Electronics in Agriculture* 118 (October): 372–79. https://doi.org/10.1016/j.compag.2015.09.001.

Sankaran, Sindhuja, Lav R. Khot, Carlos Zúñiga Espinoza, Sanaz Jarolmasjed, Vidyasagar R. Sathuvalli, George J. Vandemark, Phillip N. Miklas, et al. 2015. “Low-Altitude, High-Resolution Aerial Imaging Systems for Row and Field Crop Phenotyping: A Review.” *European Journal of Agronomy* 70 (October): 112–23. https://doi.org/10.1016/j.eja.2015.07.004.

Tester, M., and P. Langridge. 2010. “Breeding Technologies to Increase Crop Production in a Changing World.” *Science* 327 (5967): 818–22. https://doi.org/10.1126/science.1183700.

United Nations, Department of Economic and Social Affairs, and Population Division. 2019. *World Population Prospects 2019: Highlights*. Vol. (ST/ESA/SER.A/423).

Watson, Amy, Sreya Ghosh, Matthew J. Williams, William S. Cuddy, James Simmonds, María-Dolores Rey, M. Asyraf Md Hatta, et al. 2018. “Speed Breeding Is a Powerful Tool to Accelerate Crop Research and Breeding.” *Nature Plants* 4 (1): 23–29. https://doi.org/10.1038/s41477-017-0083-8.

White, Jeffrey W., Pedro Andrade-Sanchez, Michael A. Gore, Kevin F. Bronson, Terry A. Coffelt, Matthew M. Conley, Kenneth A. Feldmann, et al. 2012. “Field-Based

Phenomics for Plant Genetics Research." *Field Crops Research* 133 (July): 101–12. https://doi.org/10.1016/j.fcr.2012.04.003.

Xu, Yunbi, and Jonathan H. Crouch. 2008. "Marker-Assisted Selection in Plant Breeding: From Publications to Practice." *Crop Science* 48 (2): 391–407. https://doi.org/10.2135/cropsci2007.04.0191.

Chapter 2 - OneKK: An Android application for rapid assessment of seed size, shape, and thousand kernel weight

Megan Calvert[a,b] , Trevor W. Rife[a], Siddharth Amaravadi[d], Cameron Amos[a], Chaney Courtney[d], Shravan D. Gangadhara[d], Sandra Dunckel[e], Mitchell Neilsen[d], and Jesse A. Poland[a,c]

a Wheat Genetics Resource Center, Department of Plant Pathology, Kansas State University, Manhattan, KS, 66506

b Interdepartmental Genetics, Kansas State University, Manhattan, KS, 66506

c Department of Agronomy, Kansas State University, Manhattan, KS, 66506

d Department of Computer Science, Kansas State University, Manhattan, KS 66506

e Longreach Plant Breeding, Australia

Abbreviations

DH double haploid; HTP, high-throughput phenotyping; QTL, quantitative trait loci; RIL, recombinant inbred line; SynOpDH, Synthetic W7984 X Opata M85 doubled haploid reference population; UAV, uncrewed aerial vehicle

Abstract

Background

A major limitation to the study of plants and cultivar development in breeding is the challenge of quickly obtaining large amounts of accurate phenotypic measurements across large populations. The development of new high-throughput phenotyping tools is enabling studies to be conducted faster and with larger populations than was previously possible. An important target for HTP is seed size and morphology which is a critical yield component and a strong determinant of end-use quality and value.

Results

We have developed a mobile Android application, OneKK, that simultaneously captures a photo of seeds and performs the analysis to extract morphological characteristics. OneKK was validated by comparing digital measurements to manual measurements and SmartGrain, and showed high concordance with both approaches. We then used the app to characterize the Synthetic/Opata DH wheat reference mapping population and identify a QTL for grain length on chromosome 2D at 69.2 cM that explained approximately 17.48% of the variance.

Conclusions

OneKK is a fast and effective means of obtaining information about seed size and shape in different plant species. Implementing OneKK into breeding programs will make

it possible to collect data about seed size and use those newly collected phenotypes to make informed and rapid selection decisions.

Keywords

High-throughput phenotyping, mobile application, seed size

Background

A major limitation to the study of plants is the challenge of quickly obtaining large amounts of accurate phenotypic measurements. The development of new high-throughput phenotyping (HTP) tools and systems is enabling studies to be conducted faster and with larger populations than was previously possible (Crain et al. 2016). These new tools allow for more phenotypic characteristics to be examined and a more 'complete' picture of the organism to be obtained while simultaneously reducing many of the inaccuracies associated with manual measurements (Dhondt, Wuyts, and Inzé 2013). Combining HTP tools with new genotyping and sequencing methods can greatly improve our understanding of the mechanisms underlying traits of interest in plants. Utilizing these new approaches for plant breeding will shorten breeding cycles by allowing for faster, earlier, and more accurate phenotyping which can then be utilized for direct selection, indirect phenotypic selection, and training improved genomic prediction models.

Recent developments in HTP have led to the creation of both field-based systems and controlled environment systems (White et al. 2012). The field-based HTP developments have focused on 1) arrays of sensors and cameras with high-precision

GPS to position the measurements at a plot level (Busemeyer et al. 2013; Andrade-Sanchez et al. 2014; Barker et al. 2016), 2) gantry type systems carrying arrays of sensors (Subramanian, Spalding, and Ferrier 2013), and 3) uncrewed aerial vehicle (UAV) systems with thermal and spectral cameras (Haghighattalab et al. 2016; Sankaran et al. 2015; Sankaran, Khot, and Carter 2015). Controlled environment systems involve growing the plants in greenhouses and growth chambers. This allows all variables to be controlled and are monitored via a range of phenotyping tools such as imaging or thermography during growth (Chaerle and Van Der Straeten 2001; Yang et al. 2013).

While high-throughput and powerful, the challenge with many of the field and controlled environment systems is the complex engineering and high equipment costs. An opportunity to create a highly scalable and simple phenotyping platform exists using mobile apps that utilize consumer grade smartphones and tablets. This approach is highly effective for developing countries due to the availability of inexpensive hardware and the minimal technical expertise required. Recently developed apps, such as Field Book are rapidly being implemented across the world (Rife and Poland 2014, >10K downloads, Rife, personal correspondence) helping find utility in breeding programs and genetic research. In particular, adoption in developing countries parallels the rapid increase in adoption of mobile phones throughout Africa and South Asia (Bajarin 2014; Anand 2019; Schwieters and Saleem 2013).

Seed size and morphology is a critical yield component and strong determinant of end-use quality and value, and thus an important target for HTP (Moles 2005). Grain size was one of the characteristics originally used for selection when domesticating

plant species for food sources and is still a primary target of selection in modern breeding programs. The ability to understand the genetic characteristics underlying these traits, and the fast inclusion of those characteristics into breeding programs, is an important objective that can be assisted by HTP methods.

There are many studies examining seed characteristics in crop species such as rice, wheat and maize as they influence potential end use products (Liu et al. 2014; Breseghello and Sorrells 2007; Zhang et al. 2014; Li et al. 2004; Huang et al. 1997). For wheat, in particular, seed size and shape have a strong influence on overall yield and milling (Pask et al. 2012; Marshall et al. 1986). Reflecting this, these traits have been extensively studied with a focus on identifying the genomic regions in wheat that control seed quality (Dholakia et al. 2003; Ammiraju et al. 2001; Breseghello and Sorrells 2006; Giura and Saulescu 1996; Ramya et al. 2010; Rasheed et al. 2014; Williams and Sorrells 2014; Campbell et al. 1999). The majority of past studies relied on manual or destructive measurements to determine wheat kernel morphological characteristics. Manual measurements are time consuming and prone to personal bias and inaccuracies, resulting in fewer lines or kernels from each line being measured (Ramya et al. 2010). Destructive measurements, such as those provided by the single kernel characterization system (SKCS, Bean et al. 2006), are often more comprehensive than the manual measurements as more information is provided about hardness yet they are still low throughput. However, destructive phenotyping approaches are limited in that they cannot be used for early generation material where seed from each line is used for advancement.

Newer studies have utilized digital imaging software to extract seed characteristics, partially removing bias and allowing for larger sample sizes to be analyzed (Campbell et al. 1999; Breseghello and Sorrells 2006; 2007; Williams and Sorrells 2014; Rasheed et al. 2014). This approach is still limited in that it is a two-stage process of acquiring images with either a scanner or camera and then utilizing software to analyze the images, as is done in the desktop programs SmartGrain and GrainScan (Tanabata et al. 2012; Whan et al. 2014). While an improvement over manually measuring seeds, this approach is limited based on access to effective hardware, time available to collect data, and ease of use for the user.

To improve the process of rapid seed phenotyping, we have developed a mobile application, OneKK, which simultaneously captures a photo of seeds and performs the analysis to extract morphological characteristics. Derived from its namesake of one thousand (1K) kernels, OneKK was validated with concurrent manual measurements in contrasting samples from a heat stress experiment in wheat (*Triticum aestivum*) and was then used in a quantitative trait loci (QTL) analysis of a synthetic wheat double haploid (DH) population.

Methods and Materials

OneKK Android Application

To enable high-throughput phenotyping of seed size and shape, we developed the 'OneKK' app. OneKK was created for the Android mobile operating system and utilizes the Open Computer Vision Library (OpenCV, Bradski 2000) for image processing.

For this analysis, OneKK was run on a Nexus 7 tablet using the built-in camera for image capture. Source code for OneKK is readily available for further development on GitHub (https://github.com/trife/OneKK).

To simplify image processing and extract absolute measurements, we created a 'green-screen' background with contrasting blue reference circles of known size. For image collection, seeds are spread on the background and imaged using the built-in camera on the Android device (Figure 1). To associate an image with a biological sample (e.g. seed packet), the app supports external text input from barcode scanners and triggers the camera upon sample entry (e.g. upon barcode scan) (https://www.youtube.com/watch?v=i3614N3bMzI). The collected image is subsequently associated with the input text / scanned barcode and stored on the device for later access. In the absence of barcoding, the sample name can be manually input with the camera triggered upon entry of the text. Once an image is taken, the raw image is saved onto the device in the ~/OneKK/Photos directory. The raw image is then converted to HSB color format for all subsequent processing. Next, a color threshold (H>120) is applied to segment the blue reference circles from the background and the image is converted to binary. The 'findContours' function in OpenCV was used with a defined threshold of circularity > 0.90 and height to width ratio < 1.1 to identify the round reference circles and discard partially covered circles. The actual size of the reference circles is set by the user and stored internally for the duration of the phenotyping session. The average size of measured reference circles in pixels and the

input size of the reference panel circles (e.g. in millimeters) is used to calculate the pixel to empirical conversion value for subsequent analysis of seeds.

To measure the seeds, a threshold (H<60) is used to segment all objects (seeds) from the green/blue background and the image is then converted to a binary matrix. For size and shape measurements, only single, individual, isolated seeds are measured. To identify seeds that are not touching other seeds, a non-parametric algorithm identifies individual seeds in the image. Briefly, this method estimates the number of objects by calculating ratios between object clusters for a given criterion and then using the calculated ratio to determine the number of objects in each cluster. In OneKK, we classify an object as being an individual seed if it is found to be an individual object when using width, length, perimeter, and area as the specified criteria. Morphology characteristics are then computed from each individual seed. Native OpenCV functions contourArea and arcLength were used to measure area and perimeter (in pixels) of the individual seeds, respectively. For length and width measurements, we implemented the algorithms described by Tanabata *et al*. (2012) as implemented in SmartGrain. To reduce computational time of the perimeter search, a convex hull is applied to each contour (seed) to reduce the search space and time needed to compute length and width (in pixels). All measurements taken on individual seeds are then converted to millimeters using the previously measured reference circles and pixel-to-empirical conversion factor (Figure 2).

To estimate the number of seeds in each image, an algorithm that extends the traditional watershed segmentation was used (Neilsen, Gangadhara, and Rife 2016).

Briefly, this algorithm calculates a Euclidean Distance Map, finds the ultimate eroded points (UEPs), dilates each UEP until it reaches the edge of the seed or the edge of another growing UEP, and then recursively merges and splits the seeds that are identified. Groups of seeds are split by identifying points in the outline that are not convex. Smaller seeds adjacent to one another with edges artificially added by the threshold process are also identifiable and easily merged.

OneKK is freely available for download and can be found at https://github.com/PhenoApps/OneKK\.

Plant Material Used for Validation and Comparison

The winter wheat lines used for the accuracy and validation of the app were obtained from a previous heat trial conducted by Dunckel et al. (2015). Two $F_{5:6}$ recombinant inbred line (RIL) mapping populations, U6019 and U6020, obtained from "Overley" by "Jefimija" crosses and advanced through single seed decent. The populations consisted of 103 and 100 lines in the U6019 and U6020 populations respectively from which 100 lines were selected at random and used for manual validation. The seeds were grown under optimal conditions in the greenhouse until 14 days after heading when they were randomly assigned to one of four growth chambers where one of two temperature treatments was applied: an optimal control temperature, 21/17oC +/- 1.5oC day/night temperatures, or high temperature to simulate heat stress. The plants under heat stress were exposed to day/night temperatures of 36/30oC +/- 2.0oC, with sufficient water to avoid dehydration and possible drought stress until the plants were ready for harvest (Dunckel 2015).

Manual Validation

Length and width for each wheat kernel in a line, ranging between 2 and 80 kernels per line, were measured across the longest and widest sections of the kernel by hand using a digital caliper (OriginCal IP54 Digital Calipers, iGaging, California USA) for 80 lines. The average kernel length and width for lines was then calculated in the R software environment V3.2 (Core Team 2020).

Comparison with Available Software

Seed length and width were also measured using SmartGrain (Tanabata et al. 2012). The SmartGrain analysis was performed from the original images captured by the Nexus 7 tablet's built-in camera. The averages for length and width for each line were determined using SmartGrain's color segmentation method, after which results were checked and parameters were optimized before performing the analysis on all images. These results were then compared to the measurements obtained from OneKK and those obtained through manual hand measurement. The coefficient of determination (r^2) was determined for length and width using the statistical software R.

Plant Material for QTL Mapping

The plant material used for the QTL mapping analysis is from the Synthetic W7984 X Opata M85 doubled haploid reference population (SynOpDH) recently reconstructed by Sorrells et al. (2011) with the pedigree Synthetic W7984(Altar 84/Ae. tauschii (219) CIGM86.940)/Opata M85. These lines were grown out under optimal conditions in the greenhouse for seed increase.

Genotyping

Two-enzyme genotyping-by-sequencing was used to generate the genome-wide marker data for the SynOpDH population (Elshire et al. 2011). The generation of this data is described in Dunckel (2015). Briefly, a map with 1485 single nucleotide polymorphisms (SNPs) was constructed based on protocols laid out in (Poland et al. 2012). The alleles were coded as "A" if they were from the female Synthetic W7984 parent or "B" if they were from the male parent Opata M85.

Phenotyping for QTL Analysis

From the SynOpDH population, 167 lines were phenotyped using OneKK. The average length, width, and area of kernels in each line was obtained from OneKK, tested for univariate normality, and used for QTL analysis. The median kernel length, width and area was then calculated from the individual seed measurements OneKK data in the R software environment.

QTL Analysis

QTL analysis for each trait was performed using the R software package R/qtl in the R software environment (Broman et al. 2003). A Composite Interval Mapping approach utilizing Haley-Knott regression with forward selection of 3 marker covariates and a window size of 10 cM was used for the sample average length, average width, and average area. 1000 permutations were used to determine the genome-wide logarithm of the odds value (LOD) for declaring a significant QTL.

Results

To enable high-throughput phenotyping of seed morphology, we developed OneKK, an android app that leverages the robust OpenCV image processing features. The development of this app creates a low-cost and scalable platform for breeding and research programs to rapidly obtain detailed measurements of individual seeds and descriptive statistics (e.g. average).

Manual Validation and Comparison with Available Software

For implementation and deployment of OneKK with measurements for length, width, and number of seeds, validation to hand and current image analysis software is needed. In this study, we validated the accuracy of OneKK on wheat RILs by comparing manual measurements and those obtained from SmartGrain, a PC based software program for similar image analysis. For comparison, 80 RILs that were grown under heat stress conditions were measured using OneKK and SmartGrain, and measured for length and width using a digital caliper (Supplementary Table 1). The coefficient of determination for the manual measurements and the OneKK measurements were relatively high (length $r^2 = 0.598$, width $r^2 = 0.845$, Figure 3A and 3I) and comparable to those obtained between the manual measurements and SmartGrain (length $r^2 = 0.599$, width $r^2 = 0.892$, Figure 3D and 3H). Likewise, we tested the association between two image analysis programs, OneKK and SmartGrain, and found measurements were consistent with the overall same level of correlation for both length ($r^2 = 0.617$) and width ($r^2 = 0.892$) (Figure 3B and 3F).

QTL Mapping

The SynOpDH lines used for QTL mapping showed a normal distribution and univariate normality for seed length, width and area as determined in by a Shapiro-Wilk's Normality test (Length p-value = 0.69, width p-value = 0.62, area p-value = 0.063 (Korkmaz, Goksuluk, and Zararsiz 2014). We implemented CIM in r/QTL and found one QTL for kernel length on chromosome 2D at 69.2 cM (GBS marker synopGBS745 at 69.52 cM, Figure 4). This QTL on chromosome 2D explained approximately 17.5% of the variance seen in this trait with an effect size of -0.246 mm (SE +/- 0.044) for the allele contributed by the Opata M85 parent (Figure 5).

Discussion

The use of HTP technology in plant breeding and genetic studies is becoming more common. This is concurrent with the increasing push to implement HTP approaches to alleviate the critical phenotyping bottleneck facing most programs. The advancement of this field is required to fully take advantage of the developments that have been made in genome sequencing and marker development while bringing phenotypic data to the level of available genomic data. At the same time, HTP platforms that are easily implemented with a low cost and technology barrier are needed to scale these approaches in breeding and genetics programs around the world. With this goal in mind, the newly developed OneKK app is an accurate, easy to use, inexpensive, and high throughput tool for measuring seed size, shape, and number. It is comparable to already-available HTP techniques for seed characteristics (Figure 3B and 3F) yet is more

accessible and user-friendly by leveraging advances made in cellphone and tablet camera technology.

Cellphone and tablet camera technology is constantly advancing and becoming more sensitive to light and movement. When using these tools for image analysis there is a need to ensure consistent lighting of the area, as shadows or reflectance can affect the ability of the application to run optimally. This can be obtained by using an even light source or non-glossy paper for the background. Another area of optimization for the utilization of OneKK can come in the form of a designated stand on which all the analysis is performed (Figure 1). This ensures that multiple samples are subjected to the same lighting conditions and that the tablet or cellphone is at a constant height from the background. This is not essential to the use of OneKK, but it does remove some potential causes of variability.

The width measurements for OneKK are more accurate than the length measurements yet there is a slight bias towards underestimation of the measurements. This can also be seen in the results obtained for SmartGrain. The lower accuracy for the HTP length measurements in comparison to the hand measurements may be caused by the compression of the fragile seeds during hand measurements or by the over estimation and inclusion of awns by the HTP systems. Overall OneKK appears to show a relatively high accuracy with some bias towards an under estimation of the measurements.

Utilizing OneKK we identified a QTL for length on chromosome 2D at 69.2 cM in the SynOpDH wheat reference mapping population. The estimated allele effect of this

QTL indicates that the allele from the synthetic parent gives an increase in the length of the wheat kernels. The SynOpDH population has a sister population of recombinant inbred lines (SynOpRIL) derived from the same parents on which a similar study was performed (Breseghello and Sorrells 2007; Sorrells et al. 2011). In this study they also determined that chromosome 2D was influential on kernel characteristics and the synthetic allele had a similar effect on the lateral dimensions of the kernel. This is in line with what we have observed in our study, as a decrease in length of the kernel will result in a shorter and wider kernel structure. This result validates OneKK as a tool for genomic studies as well as for plant breeding selections.

OneKK has a versatility that other similar tools for HTP measurement of seeds do not, as it is capable of measuring various crops including cassava and potato. As such, OneKK will likely find utility in breeding and genetics programs for diverse crops, enabling the rapid measurements of these target traits.

Conclusion

OneKK is a fast and effective means of obtaining information about seed characteristics of size and shape in different plant species. Implementing OneKK into breeding programs will make it possible to collect data about seed size and use those newly collected phenotypes to make informed and rapid selection decisions. OneKK is easy to use and can be incorporated into established breeding programs with minimal training, as well as in plant genetics research for use in genetic mapping and building genomic prediction models for different seed characteristics.

Competing interests

The authors declare that they have no competing interests.

Funding

Support for T. Rife was through the USDA-NIFA funded Triticeae Coordinated Agriculture Project (T-CAP) (2011-68002-30029). Support for M. Calvert was provided through the Kansas State University Interdepartmental Genetics Fellowship. This work was supported by the U.S. Agency for International Development Feed the Future Innovation Lab for Applied Wheat Genomics (Cooperative Agreement No. AID-OAA-A-13-00051), the U.S. National Science Foundation Plant Genome Research Program (IOS-1238187), the U.S. National Science Foundation Basic Research to Enable Agriculture Development (BREAD) (FAIN-1543958), the Kansas Wheat Alliance, and the Kansas Wheat Commission. This work was completed under the auspices of WGRC I/UCRC partially funded by NSF grant contract (IIP-1338897).

Authors' contributions

MC collected and analyzed all data. TR designed and programmed the OneKK. SD provided individual seed stocks and analyzed data. JP conceived and planned the app and experiments. All authors read and approved the final manuscript.

References

Ammiraju, J. S. S., B. B. Dholakia, D. K. Santra, H. Singh, M. D. Lagu, S. A. Tamhankar, H. S. Dhaliwal, V. S. Rao, V. S. Gupta, and P. K. Ranjekar. 2001. “Identification of Inter Simple Sequence Repeat (ISSR) Markers Associated with Seed Size in

Wheat:" Theoretical and Applied Genetics 102 (5): 726–32. https://doi.org/10.1007/s001220051703.

Anand, Shivani. 2019. "India Smartphone Market Registers Highest Second-Quarter Shipments in 2Q19, IDC India Reports." IDC: The Premier Global Market Intelligence Company. August 13, 2019. https://www.idc.com/getdoc.jsp?containerId=prAP45442419.

Andrade-Sanchez, Pedro, Michael A. Gore, John T. Heun, Kelly R. Thorp, A. Elizabete Carmo-Silva, Andrew N. French, Michael E. Salvucci, and Jeffrey W. White. 2014. "Development and Evaluation of a Field-Based High-Throughput Phenotyping Platform." Functional Plant Biology 41 (1): 68. https://doi.org/10.1071/FP13126.

Bajarin, Ben. 2014. "Why India Will Be the World's Second Biggest Smartphone Market." Fortune. December 2, 2014. https://fortune.com/2014/12/02/why-india-will-be-the-worlds-second-biggest-smartphone-market/.

Barker, Jared, Naiqian Zhang, Joshua Sharon, Ryan Steeves, Xu Wang, Yong Wei, and Jesse Poland. 2016. "Development of a Field-Based High-Throughput Mobile Phenotyping Platform." Computers and Electronics in Agriculture 122 (March): 74–85. https://doi.org/10.1016/j.compag.2016.01.017.

Bean, S. R., O. K. Chung, M. R. Tuinstra, J. F. Pedersen, and J. Erpelding. 2006. "Evaluation of the Single Kernel Characterization System (SKCS) for Measurement of Sorghum Grain Attributes." Cereal Chemistry Journal 83 (1): 108–13. https://doi.org/10.1094/CC-83-0108.

Bradski, Gary. 2000. The OpenCV Library. https://www.drdobbs.com/open-source/the-opencv-library/184404319.

Breseghello, Flavio, and Mark E. Sorrells. 2006. "Association Mapping of Kernel Size and Milling Quality in Wheat (Triticum Aestivum L.) Cultivars." Genetics 172 (2): 1165–77. https://doi.org/10.1534/genetics.105.044586.

Breseghello, Flavio, and Mark E. Sorrells. 2007. "QTL Analysis of Kernel Size and Shape in Two Hexaploid Wheat Mapping Populations." Field Crops Research 101 (2): 172–79. https://doi.org/10.1016/j.fcr.2006.11.008.

Broman, Karl W., Hao Wu, Saunak Sen, and Gary A. Churchill. 2003. "R/Qtl: QTL Mapping in Experimental Crosses." Bioinformatics (Oxford, England) 19 (7): 889–90. https://doi.org/10.1093/bioinformatics/btg112.

Busemeyer, Lucas, Daniel Mentrup, Kim Möller, Erik Wunder, Katharina Alheit, Volker Hahn, Hans Peter Maurer, et al. 2013. "BreedVision — A Multi-Sensor Platform for Non-Destructive Field-Based Phenotyping in Plant Breeding." Sensors 13 (3): 2830–47. https://doi.org/10.3390/s130302830.

Campbell, Kim Garland, Christine J. Bergman, Daisy G. Gualberto, James A. Anderson, Micheal J. Giroux, Gary Hareland, R. Gary Fulcher, Mark E. Sorrells, and Patrick L. Finney. 1999. "Quantitative Trait Loci Associated with Kernel Traits in a Soft ×

Hard Wheat Cross." Crop Science 39 (4): 1184. https://doi.org/10.2135/cropsci1999.0011183X003900040039x.

Chaerle, L., and D. Van Der Straeten. 2001. "Seeing Is Believing: Imaging Techniques to Monitor Plant Health." Biochimica Et Biophysica Acta 1519 (3): 153–66. https://doi.org/10.1016/s0167-4781(01)00238-x.

Core Team, R. 2020. R: A Language and Environment for Statistical Computing (version 4.0.0). Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Crain, Jared L., Yong Wei, Jared Barker, Sean M. Thompson, Phillip D. Alderman, Matthew Reynolds, Naiqian Zhang, and Jesse Poland. 2016. "Development and Deployment of a Portable Field Phenotyping Platform." Crop Science 56 (3): 965–75. https://doi.org/10.2135/cropsci2015.05.0290.

Dholakia, B. B., J. S. S. Ammiraju, H. Singh, M. D. Lagu, M. S. Roder, V. S. Rao, H. S. Dhaliwal, P. K. Ranjekar, and V. S. Gupta. 2003. "Molecular Marker Analysis of Kernel Size and Shape in Bread Wheat." Plant Breeding 122 (5): 392–95. https://doi.org/10.1046/j.1439-0523.2003.00896.x.

Dhondt, Stijn, Nathalie Wuyts, and Dirk Inzé. 2013. "Cell to Whole-Plant Phenotyping: The Best Is yet to Come." Trends in Plant Science 18 (8): 428–39. https://doi.org/10.1016/j.tplants.2013.04.008.

Dunckel, Sandra M., Eric L. Olson, Matthew N. Rouse, Robert L. Bowden, and Jesse A. Poland. 2015. "Genetic Mapping of Race-Specific Stem Rust Resistance in the Synthetic Hexaploid W7984 × Opata M85 Mapping Population." Crop Science 55 (6): 2580. https://doi.org/10.2135/cropsci2014.11.0755.

Dunckel, Sandra Margarita. 2015. "Whole Genome Approaches for Characterizing and Utilizing Synthetic Wheat." https://krex.k-state.edu/dspace/handle/2097/20518.

Giura, A., and N. N. Saulescu. 1996. "Chromosomal Location of Genes Controlling Grain Size in a Large Grained Selection of Wheat (Triticum Aestivum L.)." Euphytica 89 (1): 77–80. https://doi.org/10.1007/BF00015722.

Haghighattalab, Atena, Lorena González Pérez, Suchismita Mondal, Daljit Singh, Dale Schinstock, Jessica Rutkoski, Ivan Ortiz-Monasterio, Ravi Prakash Singh, Douglas Goodin, and Jesse Poland. 2016. "Application of Unmanned Aerial Systems for High Throughput Phenotyping of Large Wheat Breeding Nurseries." Plant Methods 12 (1): 35. https://doi.org/10.1186/s13007-016-0134-6.

Huang, Ning, Arnold Parco, Teresita Mew, Gerard Magpantay, Susan McCouch, Emmanuel Guiderdoni, Jichen Xu, Prasanta Subudhi, Enrique R. Angeles, and Gurdev S. Khush. 1997. "RFLP Mapping of Isozymes, RAPD and QTLs for Grain Shape, Brown Planthopper Resistance in a Doubled Haploid Rice Population." Molecular Breeding 3 (2): 105–13. https://doi.org/10.1023/A:1009683603862.

Korkmaz, Selçuk, Dincer Goksuluk, and Gokmen Zararsiz. 2014. “MVN: An R Package for Assessing Multivariate Normality.” https://doi.org/10.32614/RJ-2014-031.

Li, Jiming, Jinhua Xiao, Silvana Grandillo, Longying Jiang, Yizhen Wan, Qiyun Deng, Longping Yuan, and Susan R. McCouch. 2004. “QTL Detection for Rice Grain Quality Traits Using an Interspecific Backcross Population Derived from Cultivated Asian (O. Sativa L.) and African (O. Glaberrima S.) Rice.” Genome 47 (4): 697–704. https://doi.org/10.1139/g04-029.

Liu, Ying, Liwei Wang, Chuanlong Sun, Zuxin Zhang, Yonglian Zheng, and Fazhan Qiu. 2014. “Genetic Analysis and Major QTL Detection for Maize Kernel Size and Weight in Multi-Environments.” Theoretical and Applied Genetics 127 (5): 1019–37. https://doi.org/10.1007/s00122-014-2276-0.

Marshall, D. R., D. J. Mares, H. J. Moss, and F. W. Ellison. 1986. “Effects of Grain Shape and Size on Milling Yields in Wheat. II. Experimental Studies.” Australian Journal of Agricultural Research 37 (4): 331–42. https://doi.org/10.1071/ar9860331.

Moles, A. T. 2005. “A Brief History of Seed Size.” Science 307 (5709): 576–80. https://doi.org/10.1126/science.1104863.

Neilsen, Mitchell, Shravan D Gangadhara, and Trevor W Rife. 2016. “Extending Watershed Segmentation Algorithms for High-Throughput Phenotyping.” In. Denver USA. https://www.researchgate.net/publication/322369315_Extending_Watershed_Segmentation_Algorithms_for_High-Throughput_Phenotyping.

Pask, A., J. Pietragalla, D. Mullan, and M. P. Reynolds. 2012. “Physiological Breeding II: A Field Guide to Wheat Phenotyping.” https://repository.cimmyt.org/handle/10883/1288.

Poland, Jesse A., Patrick J. Brown, Mark E. Sorrells, and Jean-Luc Jannink. 2012. “Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach.” Edited by Tongming Yin. PLoS ONE 7 (2): e32253. https://doi.org/10.1371/journal.pone.0032253.

Ramya, P., A. Chaubal, K. Kulkarni, L. Gupta, N. Kadoo, H. S. Dhaliwal, P. Chhuneja, M. Lagu, and V. Gupt. 2010. “QTL Mapping of 1000-Kernel Weight, Kernel Length, and Kernel Width in Bread Wheat (Triticum Aestivum L.).” Journal of Applied Genetics 51 (4): 421–29. https://doi.org/10.1007/BF03208872.

Rasheed, Awais, Xianchun Xia, Francis Ogbonnaya, Tariq Mahmood, Zongwen Zhang, Abdul Mujeeb-Kazi, and Zhonghu He. 2014. “Genome-Wide Association for Grain Morphology in Synthetic Hexaploid Wheats Using Digital Imaging Analysis.” BMC Plant Biology 14 (1): 128. https://doi.org/10.1186/1471-2229-14-128.

Rife, Trevor W., and Jesse A. Poland. 2014. “Field Book: An Open-Source Application for Field Data Collection on Android.” Crop Science 54 (4): 1624–27. https://doi.org/10.2135/cropsci2013.08.0579.

Sankaran, Sindhuja, Lav R. Khot, and Arron H. Carter. 2015. “Field-Based Crop Phenotyping: Multispectral Aerial Imaging for Evaluation of Winter Wheat Emergence and Spring Stand.” Computers and Electronics in Agriculture 118 (October): 372–79. https://doi.org/10.1016/j.compag.2015.09.001.

Sankaran, Sindhuja, Lav R. Khot, Carlos Zúñiga Espinoza, Sanaz Jarolmasjed, Vidyasagar R. Sathuvalli, George J. Vandemark, Phillip N. Miklas, et al. 2015. “Low-Altitude, High-Resolution Aerial Imaging Systems for Row and Field Crop Phenotyping: A Review.” European Journal of Agronomy 70 (October): 112–23. https://doi.org/10.1016/j.eja.2015.07.004.

Schwieters, T, and H Saleem. 2013. “Smartphone Uptake Gaining Pace in Africa as IDC Tips Shipments to Double over next Four Years.” October 25, 2013. http://itersnews.com/?p=55430.

Sorrells, Mark E., J. Perry Gustafson, Daryl Somers, Shiaoman Chao, David Benscher, Gina Guedira-Brown, Eric Huttner, et al. 2011. “Reconstruction of the Synthetic W7984 × Opata M85 Wheat Reference Population.” Edited by A. Van Deynze. Genome 54 (11): 875–82. https://doi.org/10.1139/g11-054.

Subramanian, Ram, Edgar P. Spalding, and Nicola J. Ferrier. 2013. “A High Throughput Robot System for Machine Vision Based Plant Phenotype Studies.” Machine Vision and Applications 24 (3): 619–36. https://doi.org/10.1007/s00138-012-0434-4.

Tanabata, Takanari, Taeko Shibaya, Kiyosumi Hori, Kaworu Ebana, and Masahiro Yano. 2012. “SmartGrain: High-Throughput Phenotyping Software for Measuring Seed Shape through Image Analysis.” Plant Physiology 160 (4): 1871–80. https://doi.org/10.1104/pp.112.205120.

Whan, Alex P, Alison B Smith, Colin R Cavanagh, Jean-Philippe F Ral, Lindsay M Shaw, Crispin A Howitt, and Leanne Bischof. 2014. “GrainScan: A Low Cost, Fast Method for Grain Size and Colour Measurements.” Plant Methods 10 (1): 23. https://doi.org/10.1186/1746-4811-10-23.

White, Jeffrey W., Pedro Andrade-Sanchez, Michael A. Gore, Kevin F. Bronson, Terry A. Coffelt, Matthew M. Conley, Kenneth A. Feldmann, et al. 2012. “Field-Based Phenomics for Plant Genetics Research.” Field Crops Research 133 (July): 101–12. https://doi.org/10.1016/j.fcr.2012.04.003.

Williams, Keith, and Mark E. Sorrells. 2014. “Three-Dimensional Seed Size and Shape QTL in Hexaploid Wheat (L.) Populations.” Crop Science 54 (1): 98. https://doi.org/10.2135/cropsci2012.10.0609.

Yang, Wanneng, Lingfeng Duan, Guoxing Chen, Lizhong Xiong, and Qian Liu. 2013. “Plant Phenomics and High-Throughput Phenotyping: Accelerating Rice Functional Genomics Using Multidisciplinary Technologies.” Current Opinion in Plant Biology 16 (2): 180–87. https://doi.org/10.1016/j.pbi.2013.03.005.

Zhang, Zhanhui, Zonghua Liu, Yanmin Hu, Weihua Li, Zhiyuan Fu, Dong Ding, Haochuan Li, Mengmeng Qiao, and Jihua Tang. 2014. “QTL Analysis of Kernel-Related Traits in Maize Using an Immortalized F2 Population.” Edited by James C. Nelson. PLoS ONE 9 (2): e89645. https://doi.org/10.1371/journal.pone.0089645.

Figures

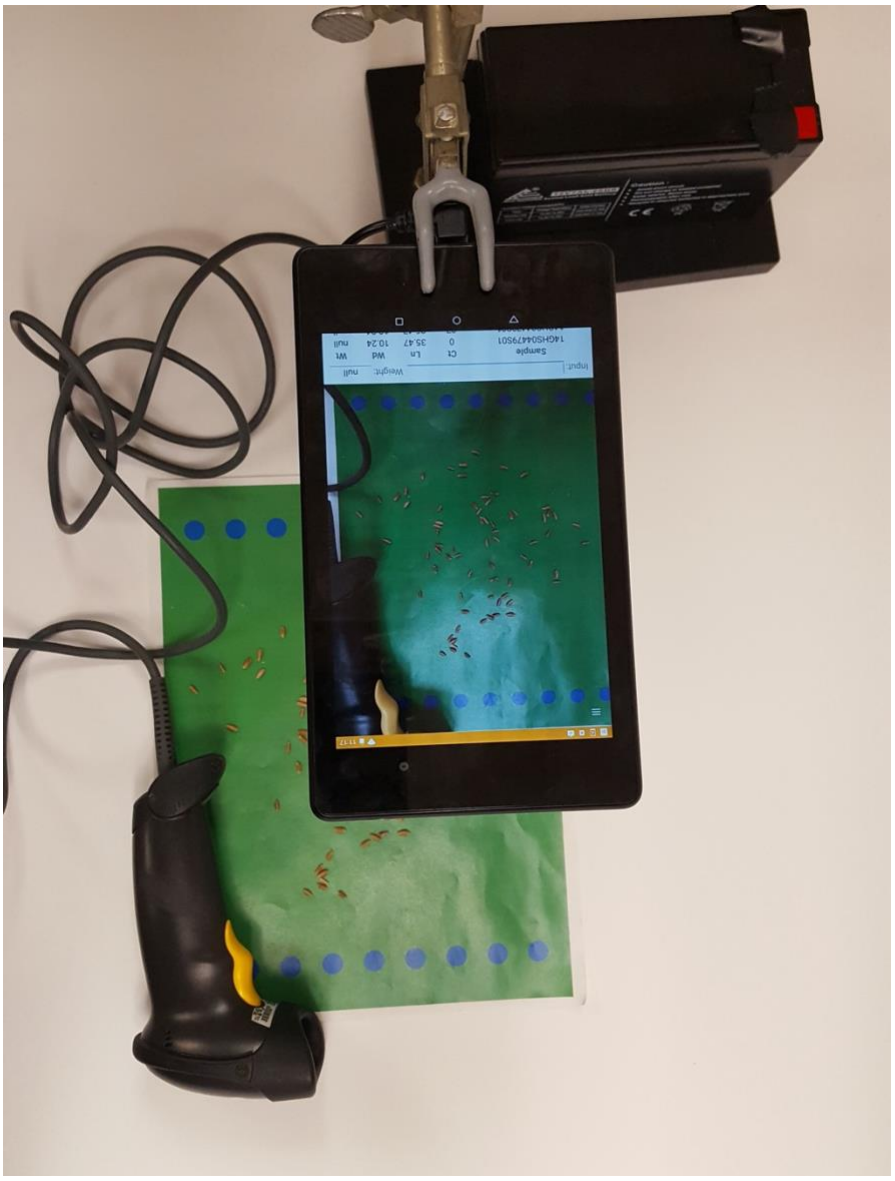


Figure 2.1 Set-up of OneKK used in this study.
A flask stand was used to hold the tablet above the printed background.

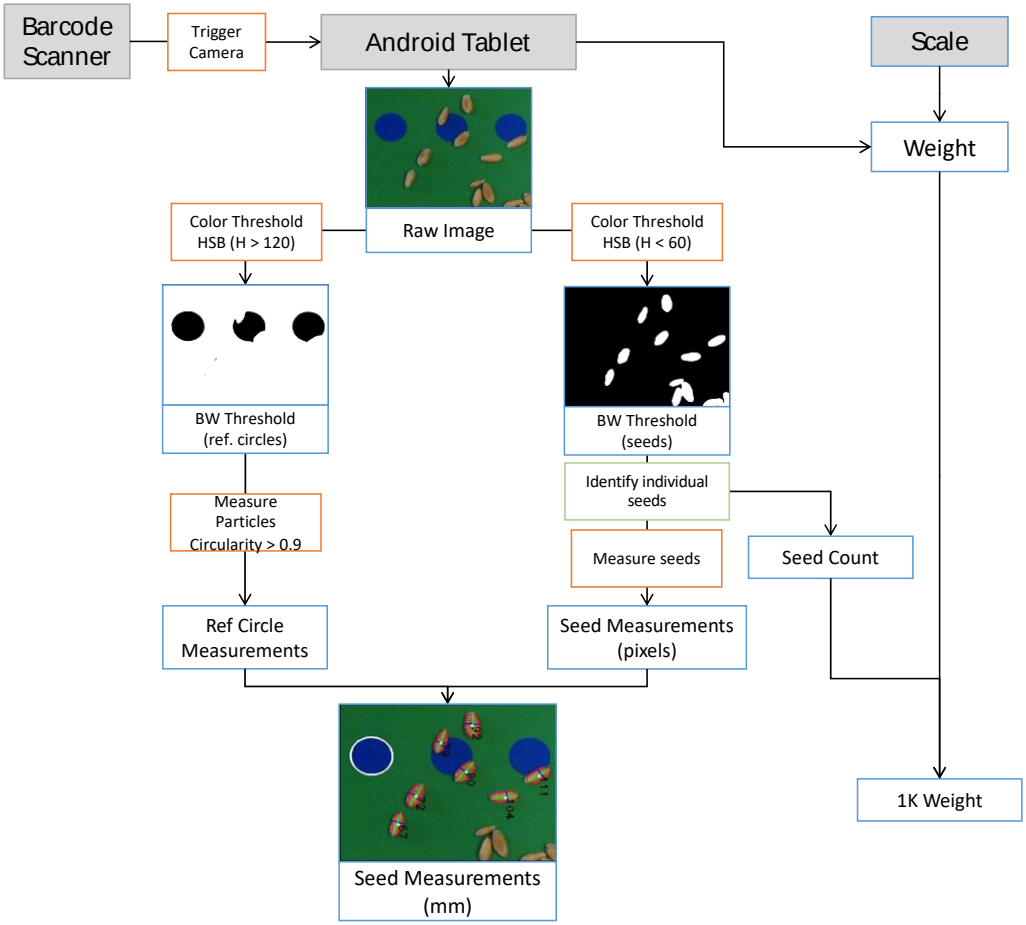


Figure 2.2. Schematic representation of the OneKK workflow.

Blue is data entity, red is process/action. Hardware is in bold with a grey filled box. The barcode scanner is used to trigger the device camera on the Android tablet and the weight measurement given by the connected scale. The raw image is then taken, and two separate color thresholds are applied. The first color threshold (H > 120) is used to identify the reference circles, which are then checked for circularity so that only complete circles are used, and then allow for the reference circle measurements to be related to pixel size. The second color threshold (H < 60) is then used to identify the seeds. Individual seeds are then identified and measured in pixels, which is then related back to the reference circle pixel size so that actual measurements are obtained. Once individual seeds are identified, they are counted and then the thousand kernel weight is determined from the weight measured by the scale and the number of seeds.

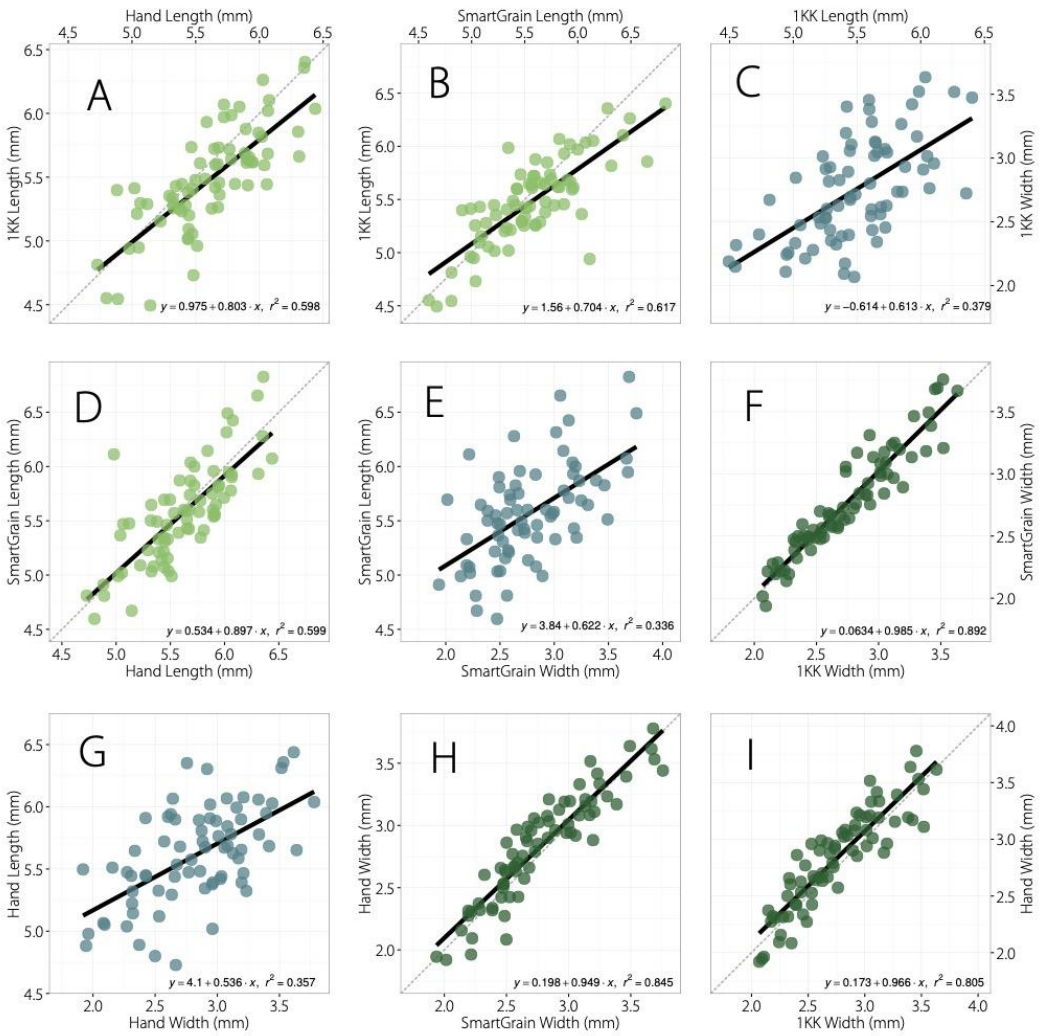


Figure 2.3. Correlation between the manual (hand) and HTP measurement procedures on 70 lines from a heat trial wheat population (Overley/Jefimija). All measurements are in millimeters (mm).

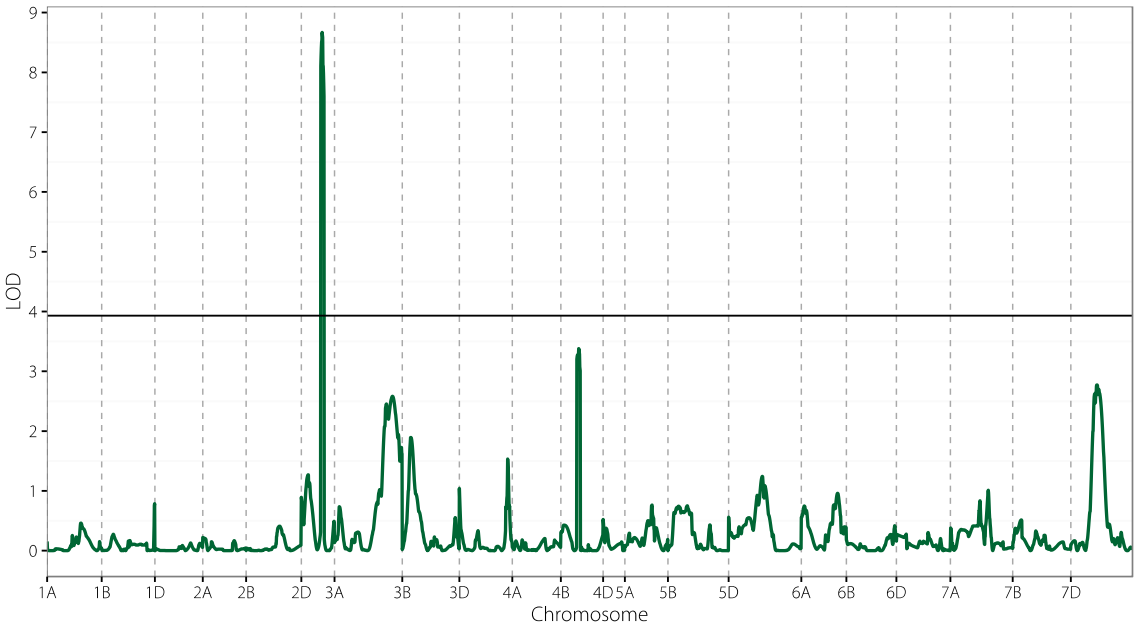


Figure 2.4. The logarithm of the odd (LOD) profile testing for marker-trait association for seed length in the wheat SynOpDH population. Horizontal lines show experiment-wise significance threshold determined by 1000 permutations at LOD = 3.93 for overall p-value < 0.05.

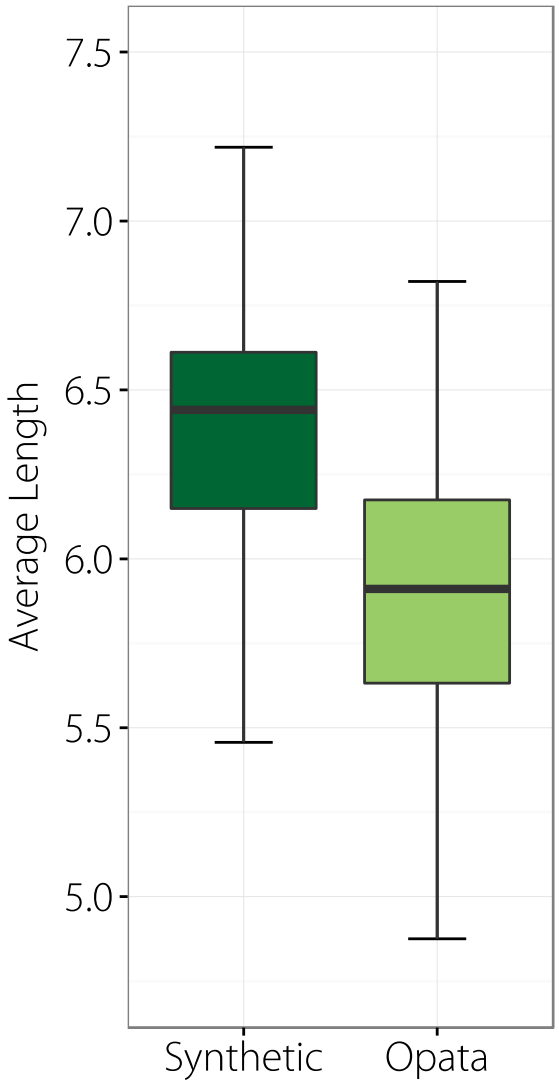


Figure 2.5. Estimated allele effect at identified length quantitative trait loci (QTL) identified on chromosome 2D at 69.2 cM.

Chapter 3 - Validation of vegetation indices for use in winter wheat (*Triticum aestivum*) genomic selection.

Megan Calvert[a,b], Byron Evers[b], Xu Wang[b], and Jesse Poland[b]

AFFILIATION

a Interdepartmental Genetics, Kansas State University, Manhattan Kansas 66502

b Department of Plant Pathology, Kansas State University, Manhattan, KS, 66506

KEYWORDS

Triticum aestivum, Genetic architecture, High-throughput phenotyping, Vegetation Indices

Abbreviations

BLUEs, Best Linear Unbiased Estimates; GEBVs, genomic estimated breeding values; GNDVI, Green Normalized Difference Vegetation Index; GRWT, Grain weight; GRVI, green-red vegetation indices; HDDT, heading date; HTP, high-throughput phenotyping; IWGSC, International Wheat Genome Sequencing Consortium; KS, Kansas; LD, linkage disequilibrium; MAF, minor allele frequency; MOIST, moisture; NDRE, normalized difference red edge; NDVI, normalized difference vegetation indices; NIR, near-infrared; PCTHEAD, percentage headed; PTHT, plant height; SNP, single nucleotide polymorphism; TESTWT test weight; VIs, Vegetation Indices

Abstract

The challenge of feeding growing populations has resulted in the generation of new technologies and methodologies that have potential in crop germplasm improvement. Precise, high-throughput phenotyping is one of the fields which has seen many developments that can influence germplasm improvement. Before these technologies are deployed on a large-scale it needs to be determined if the benefits of the technology out weight the costs of changing the current improvement system. In this study we examined the utility of vegetation indices, derived from spectral reflectance values, in winter wheat germplasm improvement. It is shown that the vegetation indices are correlated with grain yield and have potential as a covariate in genomic selection for yield.

Introduction

The global populations is expected to reach 9.7 billion in 2050 (United Nations, Department of Economic and Social Affairs, and Population Division 2019). This population growth, together with changing global climates and finite arable land resources have placed an increased pressure on crop production systems to do more with less (Tester and Langridge 2010; Ray et al. 2013; Lenaerts, Collard, and Demont 2019). While this challenge is going to require new practices in multiple areas, an improvement in crop germplasm is going to be a necessity of any solution moving forward.

One of the advances that is being investigated for its use in crop germplasm improvement is the use of high-throughput phenotyping (HTP) technologies, which are capable of phenotyping tens-of-thousands of plots in a day (White et al. 2012; Araus and Cairns 2014). These technologies have been developed to rapidly collect specific phenotypic measurements that are related to the trait of economic value. This makes them suitable for use in plant breeding programs due to the need for accurate phenotypic measurements on a large number of individual lines or plots in a small amount of time (Sankaran, Khot, and Carter 2015; Sankaran et al. 2015; Haghighattalab et al. 2016).

The HTP technologies that are currently being deployed in the field record several phenotypic measurements. One of the most common being some form of multispectral reflectance data such as red, green, blue, red-edge, and near-infrared (NIR) reflectance values (Reynolds et al. 2020). These reflectance bands are regularly formulated into Vegetation Indices (VIs), such as the Green Normalized Difference Vegetation Index – GNDVI (Gitelson, Kaufman, and Merzlyak 1996). The VIs have been used to assess line performance due to their correlation with desired agricultural traits such as grain yield and green biomass. Recently these VIs are also being incorporated in genomic prediction methods (Cabrera-Bosquet et al. 2012; Rutkoski et al. 2016; Crain et al. 2018). These studies have shown the potential of HTP in increasing the accuracy of prediction models for grain yield, both without and with additional genotyping.

The underlying genetic architecture of the VIs is not yet known, as less work has been done on using HTP technologies to perform association mapping studies. Breeding

programs are not optimally designed for these types of analyses but have been the priority when deploying and studying HTP technologies due to the larger potential impact. It is also unlikely that the VIs are influenced by a few large effect loci that are easily detected in breeding populations, as these kinds of significant differences in phenotypes are not common in breeding populations. It is more likely that a large number of loci of small effects are influencing the phenotypic variation. This sort of genetic architecture is more suited to genomic prediction, especially in wheat where large linkage blocks allow for the majority of loci to be in linkage with at least one genetic marker. An association mapping panel may allow for an examination of the genetic architecture of the VIs in wheat.

To further wheat breeding, substantial investment has led to multiple public sources of genotypic and phenotypic information, such as the release of an annotated wheat genome and The Triticeae Toolbox (Blake et al. 2016; International Wheat Genome Sequencing Consortium et al. 2018). The Triticeae Toolbox contains genotypic information from hundreds of released and elite lines. These lines are found in many of the pedigrees of the elite lines currently planted across the winter wheat growing region. By establishing a population of these lines, regions of the genome that are associated with the HTP phenotypes already in wheat breeding programs may be identified.

This study looks at the genetic architecture of some of the VIs that are used in breeding programs, in elite lines of winter wheat. This was done by analyzing the estimated effects of the genetic markers that potentially influence grain yield and the

VIs, the presence of any large effect loci, and the accuracy of whole-genome prediction models.

Materials and Methods

Plant Material

A diverse panel of 317 elite varieties and breeding lines of winter wheat were grown in locations around Manhattan KS for 3 seasons (Supplementary Table 1). In each year plots were planted on 2016/10/19, 2017/10/26, 2018/10/23, in 6-row plots (1.8m by 0.8m) in a randomized complete block design and harvested on the 2017/06/23, 2018/06/27, and 2019/07/15 respectively. For the 2017 and 2018 seasons the panel was grown at Ashland Bottoms Research Farm, KS and at Rock Ford Research Center, KS for the 2019 season.

Marker information

The genetic marker information was downloaded from the T3 Wheat database (Blake et al. 2016) filtering out markers that have a minor allele frequency (MAF) less than 0.05 and more than 20% missing data, anchored to the physical map of Chinese Spring RefSeq v1.0 from the IWGSC (International Wheat Genome Sequencing Consortium et al. 2018). The markers were generated using the Infinium 90K SNP chip (S. Wang et al. 2014). This resulted in 14523 markers being used in the final analysis. Missing marker data was imputed with Beagle V2.3 (Browning and Browning 2016). Marker information is available for 299 lines of which 252 were also planted in all years. For consistency only the 252 lines planted in all years were used in further analysis.

Phenotyping

Manual measurements of plant height (PTHT), heading date (HDDT), percentage headed (PCTHEAD) were taken using the FieldBook Android Application (Rife and Poland 2014). Grain weight (GRWT), test weight (TESTWT), and moisture (MOIST) measurements were obtained at harvest for each plot. Grain yield was calculated as tons per hectare standardized to a moisture content of 12%.

The HTP phenotypes were taken utilizing a quadcopter DJI Matrice100 (DJI, USA) equipped with a 5-channel multispectral RedEdge camera (MicaSense Inc., USA) following the protocols established by Wang et al. (2018). The multispectral camera captures 5 spectral bands from which several VIs can be calculated. These indices include green normalized difference vegetation indices (GNDVI), normalized difference red edge (NDRE), and normalized difference vegetation index (NDVI) (Supplementary Table 2). High-throughput UAS measurements were taken throughout the course of the growing season, with at least one flight in the Fall prior to winter season and vernalization. In the Spring, flights were conducted once every 7-10 days between 10am and 2pm after green up until harvest. Plot height was calculated from the orthographic images generated from the UAS system following the approach of Wang et al. (2018).

Data Analysis

Correlations between phenotypes were calculated with the Psych package (Revelle 2019), with a Holm's adjustment for multiple tests (alpha = 0.05), while linear regressions and Welch two-sample t-tests were performed with the base R functionality.

Best Linear Unbiased Estimates (BLUEs) were calculated within year using ASReml-R (Butler 2020), as follows:

$$y_{irl} = \mu + G_i + M_r + B_{r(l)} + \varepsilon_{irl} \quad \text{Equation 1}$$

where y_{irl} is the phenotypic response variable, μ is the fixed overall mean, G_i is the fixed effect of genotype i, M_r is the random effect of the replicate r distributed as iid $M_r \sim N(0, \sigma_r^2)$, $B_{r(l)}$ is the random effect of the experimental block l nested within replicate r distributed as iid $B_{r(l)} \sim N(0, \sigma_l^2)$, and ε_{irl} is the residual effect distributed as iid $\varepsilon_{irl} \sim N(0, \sigma_\delta^2)$.

Broad-sense heritability, H2, was calculated using Equation 1, with the genotypes as a random effect in the model. The estimated variance components for each random effect were used to calculate broad-sense heritability within years as (equation 2):

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{error}^2}{r}} \quad \text{Equation 2}$$

where σ_g^2 is the genetic variance, σ_{error}^2 is the residual variance and r is the number of replicates.

Principal component analysis was performed with the pcaMethods package in R on the genetic markers to identify underlying population structure (Stacklies et al. 2007).

Association Mapping

The BLUEs from Equation 1 for each genotype were estimated for each phenotype and used in association mapping. Initial genome-wide association mapping was performed with rrBLUP by implementation of a mixed model (Yu et al. 2006). The

realized relationship matrix was calculated within rrBLUP using the shrinkage method and included in the model to account for kinship (Endelman and Jannink 2012). An additional four principal components were included as fixed effects to account for population structure in the mixed model, and the EMMA with REML option was implemented (Kang et al. 2008).

Genomic Prediction

Genomic prediction for grain yield was performed by ridge regression in rrBLUP as follows:

$$y = \mu + Zu + \varepsilon \qquad \text{Equation 3}$$

where y is the BLUE for grain yield or the VI from equation 1, μ is the overall mean, Z is an $(n\ x\ m)$ matrix assigning markers to genotypes, u is a $(1\ x\ n)$ array of the random effects of the markers and ε is the residual error.

No covariates were included initially. Prediction accuracy was determined by cross-validation. This was done by 100 replications of randomly dividing the population into an 80% training and an 20% validation set, generating the model with the training data, predicting the validation set, and finally determining the accuracy of the prediction by correlation between the predicted and the measured phenotype. The same procedure for genomic prediction was also followed for each of the VIs. Each individual year was used to determine genomic estimated breeding values (GEBVs) for grain yield, which were then compared to the BLUEs determined for grain yield for the same individual year.

The VIs were then used as covariates in the genomic prediction model to predict grain yield as follows:

$$y = \mu + X\beta + Zu + \varepsilon \quad \text{Equation 4}$$

where y is the BLUE for grain yield from equation 1, μ is the overall mean, X is a $(n\ x\ 1)$ matrix of the individual observations of the VI, β is the fixed effects of the VI measurements, Z is an $(n\ x\ m)$ matrix assigning markers to genotypes, u is a $(1\ x\ n)$ array of the random effects of the markers and ε is the residual error.

The same cross-validation procedure as before was followed with 100 replicates. As covariates, the VIs were not masked in the validation set when predicting grain yield.

All data analysis was done within the R software environment unless otherwise stated (Core Team 2020), visualized with ggplot2 (Wickham 2016) and can be found at: https://github.com/megzcalvert/AMtoo.

Results

Heading Period

The AM panel has a broad range of relative maturities which may need to be accounted for when assessing grain yield. In the experimental location, Manhattan, KS, lines that reach the heading stage earlier are expected to have a longer grain fill period to take advantage of before a possible summer heat event causes senescence of the plant. This may unfairly bias the results of the analysis and, as such, was tested by examining the correlation between grain yield and heading date. It was determined that the heading date did have a significantly negative correlation with grain yield and

explained a significant proportion of the variance in the 2017 and 2019 season, but not in the 2018 season (Figure 1).

VIs and Grain Yield

The VIs are expected to change over the growth cycle of the plant, initially increasing until a peak is reached at approximately the flowering/heading date, before then decreasing due to senescence of the plant, capturing the overall development of green biomass cycle of the plants throughout the growing season. As plants that have more green biomass are expected to have higher grain yield, plants with higher VIs should likewise have higher grain yield.

The different VIs that were measured show a very similar pattern of correlations with grain yield over the season. Each of the Vis had changing correlation between the same VI and grain yield, across different dates, rising over the course of the growing season, peaking at approximately heading, before decreasing again There is not one VI that was observed as being more significantly correlated or as explaining a larger proportion of the variance. The window of dates in which heading/flowering occurs appears to have the highest correlation with grain yield (Figure 2).

Heritability

For the VIs to be useful for selection in the breeding program, they need to measure a genetic component that is repeatable and hence heritable in future generations. This can be evaluated with association mapping or genomic selection to determine the genetic components and underlying genetic architecture of the indexes.

We evaluated broad-sense heritability over the season and found that the Vis generally had a moderate to strong heritability that changes over the growing season (Figure 3).

Population Structure

The lines that make up the association mapping panel are derived from a variety of different breeding programs. Under these conditions some form of population structure can be expected due to different sets of germplasm and the different selection criteria in various regions. However, as these programs do share material among themselves, there is possibility that there will be limited population structure. Reflecting this, we did not observe strong differentiation or clusters of different breeding lines based on program of origin. With the strong pedigree structure of breeding programs, we did observe several pinnacle points on the PCA which reflect several “founder” lines and important parents that were identified and form the extremes of the population structure (Figure 4).

Linkage Disequilibrium

The linkage disequilibrium (LD) of the panel was examined to determine the presence of any large linkage blocks, and the expected power and resolution of association mapping. As is expected in wheat, and other self-pollinating crops, and in breeding programs with small effective population sizes, there were large blocks of LD. The majority of these were found in the expected low-recombination region around the centromere. There were several other large linkage blocks that represented fixed haplotypes in the population that were outside of the low-recombination regions

expected around the centromeric regions. These likely have been formed by diversity reduction and haplotype fixation through selection (Supplementary Figure 1).

Association Mapping

To assess the hypothesis that the genetic architecture of the VIs and grain yield are highly polygenic, an association mapping analysis was performed. We predict that a GWAS for highly heritable but highly polygenic traits will find few if any associations with large effect alleles. However there were several significant associations for the VIs in all years and one for grain yield in the 2018 season (Supplementary Figure 2). The association for grain yield corresponds to the position of Rht-B1 and Rht-D1 which is known to be segregating in the population (Peng et al. 1999; Wilhelm et al. 2013). None of the other associations are in regions of known genes nor do they have a significant effect on grain yield. We therefore observe that the only major genetic determinants of yield found in the population are the large effect dwarfing loci, and that within a single elite germplasm only semi-dwarf lines would be selected. Overall the limited marker-trait associations found with GWAS support the hypothesis that grain yield and VIs in this panel are highly polygenic and not conditioned by any large effect loci other than the Rht loci.

Genomic Prediction

Genomic prediction was used to determine how much of the additive genetic effect could be captured by all loci modeling small genetic effects. When using 100 replications for cross-validation and an 80:20 split between training and test

populations, the average prediction accuracy for grain yield ranges between -0.05 in 2018 and 0.73 in 2017 (Figure 5). The VIs were able to be predicted using genomic prediction with a reasonable accuracy (Figure 5). The observed moderate to high prediction accuracy for the VIs supports the hypothesis that there is strong genetic control underlying these indexes, but that it is highly polygenic with many loci of small effect spread across the genome. There are also large environmental differences that contribute to changes over the season and different years.

As the VIs are correlated with grain yield, they have the potential to increase the correlation between the GEBVs and the grain yield BLUPs when used as a covariate in genomic prediction. When using the same cross-validation procedure as previously described there is no benefit or penalty to using the VIs as a covariate as the correlation between the GEBVs and the grain yield BLUPs do not change (Figure 6). There are small but not significant changes when different VIs are used as the covariate. The growth period during which the VI was taken does not appear to influence the accuracy of predictions when the VI is used as a covariate.

Discussion

VIs are being investigated for use in genomic prediction models due to the ability to rapidly phenotype thousands of plants. This would be highly advantageous in breeding programs where phenotyping is often the most labor-intensive and rate limiting step.

There are no major differences between the various VIs over the course of the season. They show very similar trends over the season and environmental effects have a

larger influence than the different VIs. As such it may not be necessary to calculate multiple VIs and a single one could be chosen for future analysis to optimize future workflows.

The correlation between VIs and grain yield changes across the course of the season, with the highest correlations observed during or just after the heading period. This period of time would be the most advantageous to obtain the VIs for breeding programs as it is likely to have the largest influence on grain yield. This is also the period of time during which grain yield and the VIs are likely to share the most underlying genetic components which can be influenced by selection.

VIs have a heritable genetic component which can be influenced by selection, as can be seen by the medium to high broad-sense heritability. The VIs in the AM panel lines show a highly polygenic genetic architecture with few loci that have a statistically detectable effect. This sort of genetic architecture is most suited to genomic prediction, as it does not require the identification and validation of large effect loci. The accuracy of the genomic prediction for the VIs followed similar changes over the course of the season as seen in the VIs phenotypic measurements and correlations with grain yield. This again emphasizes the importance of the period in which the VIs are obtained.

There are no additional benefits or penalty's when including VIs as a covariate in genomic prediction for grain yield. The fluctuations over the season that are seen in the genomic prediction for the VIs are not observed when using them as a covariate for grain yield prediction. Before the VIs are deployed in breeding programs for use in making selections, optimization of the methods and analysis used would be required. As

the VIs do not penalize the genomic prediction of grain yield when used as a covariate, there is still potential for their use in breeding programs once further development of the procedures has occurred.

Authors' contributions

MC and BE collected data. MC and XW analyzed data. JP conceived and planned the experiments. All authors read and approved the final manuscript.

Acknowledgements

We would like to acknowledge all of the undergraduate assistants who assisted with data collection. Daljit Singh, Richard Brown and Grant Williams who assisted with UAS flights and data collection.

MC was supported in part by fellowship funds from the Interdepartmental Genetics program at Kansas State University. This material is based upon work supported by the National Science Foundation under Grant No. 1238187 project “A Field-Based High-Throughput Phenotyping Platform for Plant Genetics” and the USDA NIFA International Wheat Yield Partnership grant no. 2017-67007-25933/project accession no. 1011391 project “Wheat Yield Prediction and Advanced Selection Methodologies through Field-Based High-Throughput Phenotyping with UAVs”. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the U.S. Dept. of Agriculture.

References

Araus, José Luis, and Jill E. Cairns. 2014. “Field High-Throughput Phenotyping: The New Crop Breeding Frontier.” Trends in Plant Science 19 (1): 52–61. https://doi.org/10.1016/j.tplants.2013.09.008.

Blake, Victoria C., Clay Birkett, David E. Matthews, David L. Hane, Peter Bradbury, and Jean-Luc Jannink. 2016. “The Triticeae Toolbox: Combining Phenotype and Genotype Data to Advance Small-Grains Breeding.” The Plant Genome 9 (2): 0. https://doi.org/10.3835/plantgenome2014.12.0099.

Browning, Brian L., and Sharon R. Browning. 2016. “Genotype Imputation with Millions of Reference Samples.” The American Journal of Human Genetics 98 (1): 116–26. https://doi.org/10.1016/j.ajhg.2015.11.020.

Butler, David. 2020. Asreml: Fits the Linear Mixed Model. www.vsni.co.uk.

Cabrera-Bosquet, Llorenç, José Crossa, Jarislav von Zitzewitz, María Dolors Serret, and José Luis Araus. 2012. “High-Throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding Converge: High-Throughput Phenotyping and Genomic Selection.” Journal of Integrative Plant Biology 54 (5): 312–20. https://doi.org/10.1111/j.1744-7909.2012.01116.x.

Core Team, R. 2020. R: A Language and Environment for Statistical Computing (version 4.0.0). Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Crain, Jared, Suchismita Mondal, Jessica Rutkoski, Ravi P. Singh, and Jesse Poland. 2018. “Combining High-Throughput Phenotyping and Genomic Information to Increase Prediction and Selection Accuracy in Wheat Breeding.” The Plant Genome 11 (1): 0. https://doi.org/10.3835/plantgenome2017.05.0043.

Endelman, Jeffrey B., and Jean-Luc Jannink. 2012. “Shrinkage Estimation of the Realized Relationship Matrix.” G3: Genes|Genomes|Genetics 2 (11): 1405–13. https://doi.org/10.1534/g3.112.004259.

Gitelson, Anatoly A., Yoram J. Kaufman, and Mark N. Merzlyak. 1996. “Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS.” Remote Sensing of Environment 58 (3): 289–98. https://doi.org/10.1016/S0034-4257(96)00072-7.

Haghighattalab, Atena, Lorena González Pérez, Suchismita Mondal, Daljit Singh, Dale Schinstock, Jessica Rutkoski, Ivan Ortiz-Monasterio, Ravi Prakash Singh, Douglas Goodin, and Jesse Poland. 2016. “Application of Unmanned Aerial Systems for High Throughput Phenotyping of Large Wheat Breeding Nurseries.” Plant Methods 12 (1): 35. https://doi.org/10.1186/s13007-016-0134-6.

International Wheat Genome Sequencing Consortium, Michael Alaux, Jane Rogers, Thomas Letellier, Raphaël Flores, Françoise Alfama, Cyril Pommier, et al. 2018.

"Linking the International Wheat Genome Sequencing Consortium Bread Wheat Reference Genome Sequence to Wheat Genetic and Phenomic Data." Genome Biology 19 (1): 111. https://doi.org/10.1186/s13059-018-1491-4.

Kang, Hyun Min, Noah A. Zaitlen, Claire M. Wade, Andrew Kirby, David Heckerman, Mark J. Daly, and Eleazar Eskin. 2008. "Efficient Control of Population Structure in Model Organism Association Mapping." Genetics 178 (3): 1709–23. https://doi.org/10.1534/genetics.107.080101.

Lenaerts, Bert, Bertrand C.Y. Collard, and Matty Demont. 2019. "Review: Improving Global Food Security through Accelerated Plant Breeding." Plant Science 287 (October): 110207. https://doi.org/10.1016/j.plantsci.2019.110207.

Peng, Jinrong, Donald E. Richards, Nigel M. Hartley, George P. Murphy, Katrien M. Devos, John E. Flintham, James Beales, et al. 1999. "'Green Revolution' Genes Encode Mutant Gibberellin Response Modulators." Nature 400 (6741): 256–61. https://doi.org/10.1038/22307.

Ray, Deepak K., Nathaniel D. Mueller, Paul C. West, and Jonathan A. Foley. 2013. "Yield Trends Are Insufficient to Double Global Crop Production by 2050." PLoS ONE 8 (6): e66428. https://doi.org/10.1371/journal.pone.0066428.

Revelle, William. 2019. Psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois: Northwestern University. https://CRAN.R-project.org/package=psych.

Reynolds, Matthew, Scott Chapman, Leonardo Crespo-Herrera, Gemma Molero, Suchismita Mondal, Diego N. L. Pequeno, Francisco Pinto, et al. 2020. "Breeder Friendly Phenotyping." Plant Science, Food Security under Climate Change, 295 (June): 110396. https://doi.org/10.1016/j.plantsci.2019.110396.

Rife, Trevor W., and Jesse A. Poland. 2014. "Field Book: An Open-Source Application for Field Data Collection on Android." Crop Science 54 (4): 1624–27. https://doi.org/10.2135/cropsci2013.08.0579.

Rutkoski, Jessica, Jesse Poland, Suchismita Mondal, Enrique Autrique, Lorena González Pérez, José Crossa, Matthew Reynolds, and Ravi Singh. 2016. "Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat." G3: Genes|Genomes|Genetics 6 (9): 2799–2808. https://doi.org/10.1534/g3.116.032888.

Sankaran, Sindhuja, Lav R. Khot, and Arron H. Carter. 2015. "Field-Based Crop Phenotyping: Multispectral Aerial Imaging for Evaluation of Winter Wheat Emergence and Spring Stand." Computers and Electronics in Agriculture 118 (October): 372–79. https://doi.org/10.1016/j.compag.2015.09.001.

Sankaran, Sindhuja, Lav R. Khot, Carlos Zúñiga Espinoza, Sanaz Jarolmasjed, Vidyasagar R. Sathuvalli, George J. Vandemark, Phillip N. Miklas, et al. 2015. "Low-Altitude, High-Resolution Aerial Imaging Systems for Row and Field Crop Phenotyping: A

Review." European Journal of Agronomy 70 (October): 112–23. https://doi.org/10.1016/j.eja.2015.07.004.

Stacklies, W., H. Redestig, M. Scholz, D. Walther, and J. Selbig. 2007. "PcaMethods a Bioconductor Package Providing PCA Methods for Incomplete Data." Bioinformatics 23 (9): 1164–67. https://doi.org/10.1093/bioinformatics/btm069.

Tester, M., and P. Langridge. 2010. "Breeding Technologies to Increase Crop Production in a Changing World." Science 327 (5967): 818–22. https://doi.org/10.1126/science.1183700.

United Nations, Department of Economic and Social Affairs, and Population Division. 2019. World Population Prospects 2019: Highlights. Vol. (ST/ESA/SER.A/423).

Wang, Shichen, Debbie Wong, Kerrie Forrest, Alexandra Allen, Shiaoman Chao, Bevan E. Huang, Marco Maccaferri, et al. 2014. "Characterization of Polyploid Wheat Genomic Diversity Using a High-Density 90 000 Single Nucleotide Polymorphism Array." Plant Biotechnology Journal 12 (6): 787–96. https://doi.org/10.1111/pbi.12183.

Wang, Xu, Daljit Singh, Sandeep Marla, Geoffrey Morris, and Jesse Poland. 2018. "Field-Based High-Throughput Phenotyping of Plant Height in Sorghum Using Different Sensing Technologies." Plant Methods 14 (1): 53. https://doi.org/10.1186/s13007-018-0324-5.

White, Jeffrey W., Pedro Andrade-Sanchez, Michael A. Gore, Kevin F. Bronson, Terry A. Coffelt, Matthew M. Conley, Kenneth A. Feldmann, et al. 2012. "Field-Based Phenomics for Plant Genetics Research." Field Crops Research 133 (July): 101–12. https://doi.org/10.1016/j.fcr.2012.04.003.

Wickham, Hadley. 2016. Ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag. https://ggplot2.tidyverse.org.

Wilhelm, Edward P., Ian J. Mackay, Robert J. Saville, Andrey V. Korolev, Francois Balfourier, Andy J. Greenland, Margaret I. Boulton, and Wayne Powell. 2013. "Haplotype Dictionary for the Rht-1 Loci in Wheat." Theoretical and Applied Genetics 126 (7): 1733–47. https://doi.org/10.1007/s00122-013-2088-7.

Yu, Jianming, Gael Pressoir, William H Briggs, Irie Vroh Bi, Masanori Yamasaki, John F Doebley, Michael D McMullen, et al. 2006. "A Unified Mixed-Model Method for Association Mapping That Accounts for Multiple Levels of Relatedness." Nature Genetics 38 (2): 203–8. https://doi.org/10.1038/ng1702.

Tables

Table 3.1. Relationship between grain yield and heading date. The Pearson's correlation coefficient is given as well as the coefficient of determination adjusted for multiple tests.

Year	Pearson Correlation		Linear Regression	
	r	p-value	Adjusted R^2	p-value
2016/2017	-0.399	< 0.0001	0.158	< 0.0001
2017/2018	-0.053	0.239	0.001	0.239
2018/2019	-0.491	< 0.0001	0.240	< 0.0001

Figures

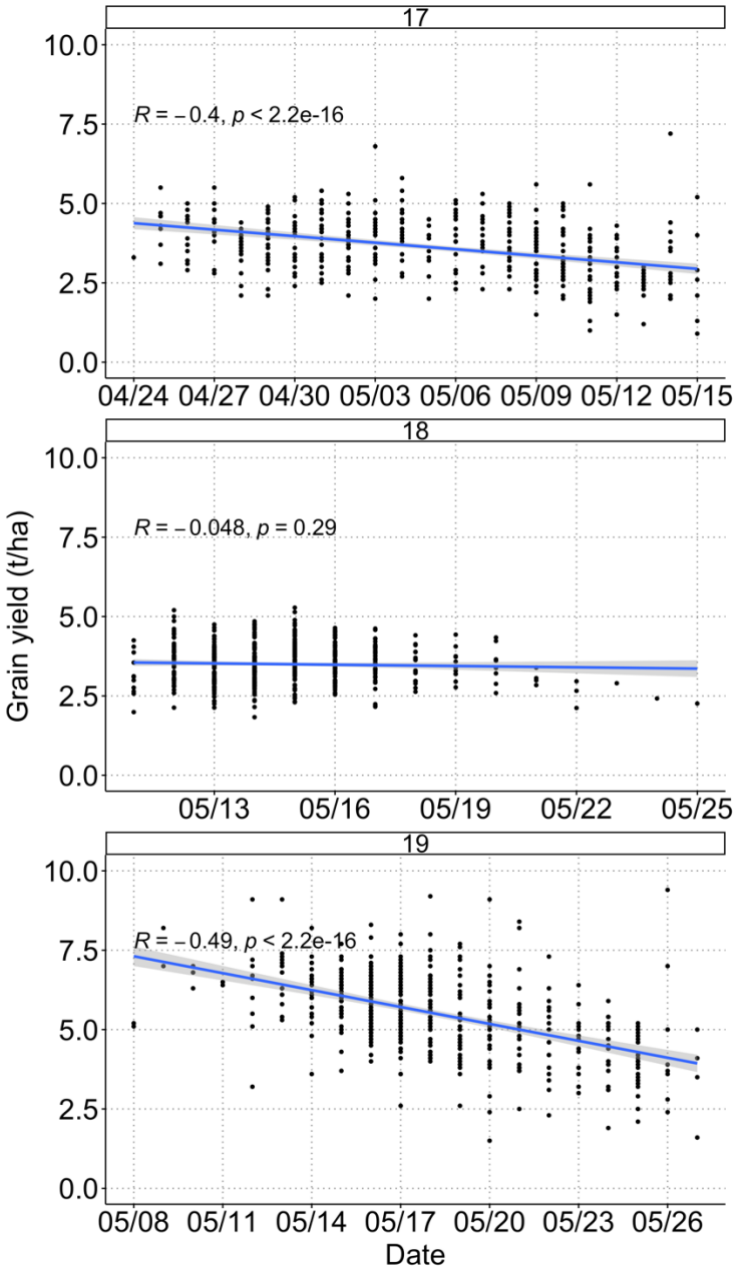


Figure 3.1. Relationship of grain yield and heading date on a per plot basis.

The plot gives the grain yield for each plot on the date at which the plot was determined to have headed. The year is given in the plot title with the Pearson correlation coefficient as well as the line of best fit.

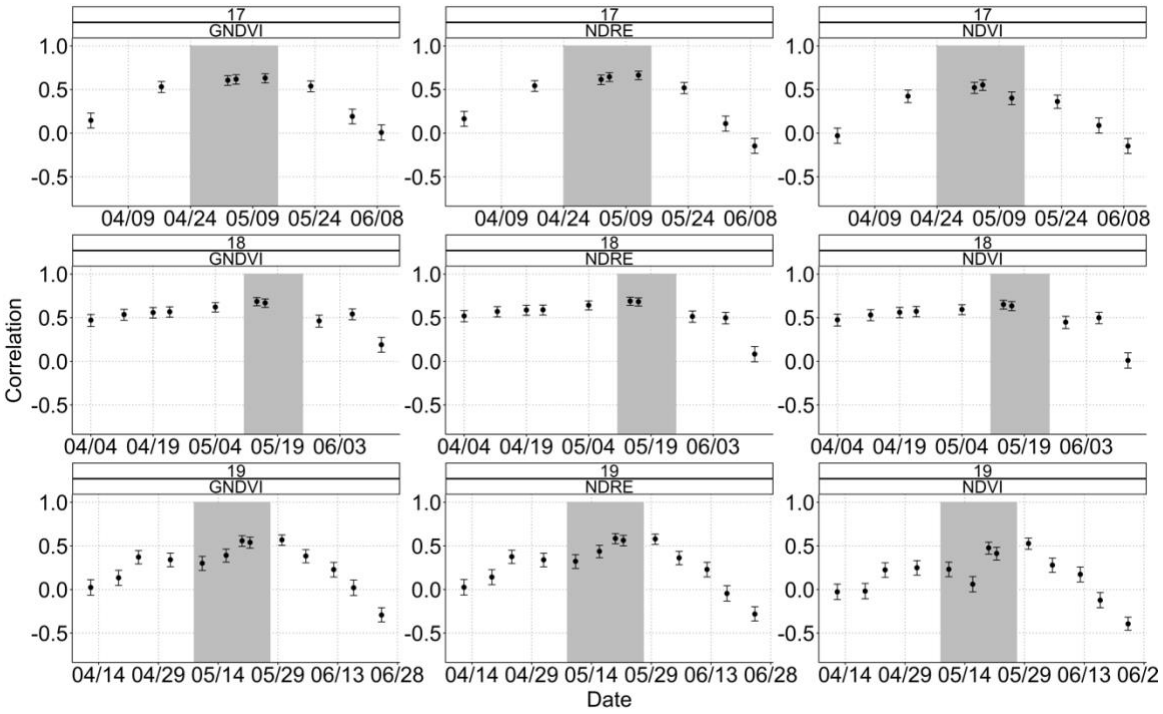


Figure 3.2. Time series of correlation between grain yield and vegetation indices. The correlation of the vegetation indices to grain yield are given over the dates measured with the season-VI combination given in the strip title. 95% confidence interval of each correlation shown with bars. Grey-shaded region show the period during which heading occurred for the population. NDVI = Normalized Difference Vegetation Index, GNDVI = Green Normalized Difference Vegetation Index, and NDRE = Red Edge Normalized Difference Vegetation Index.

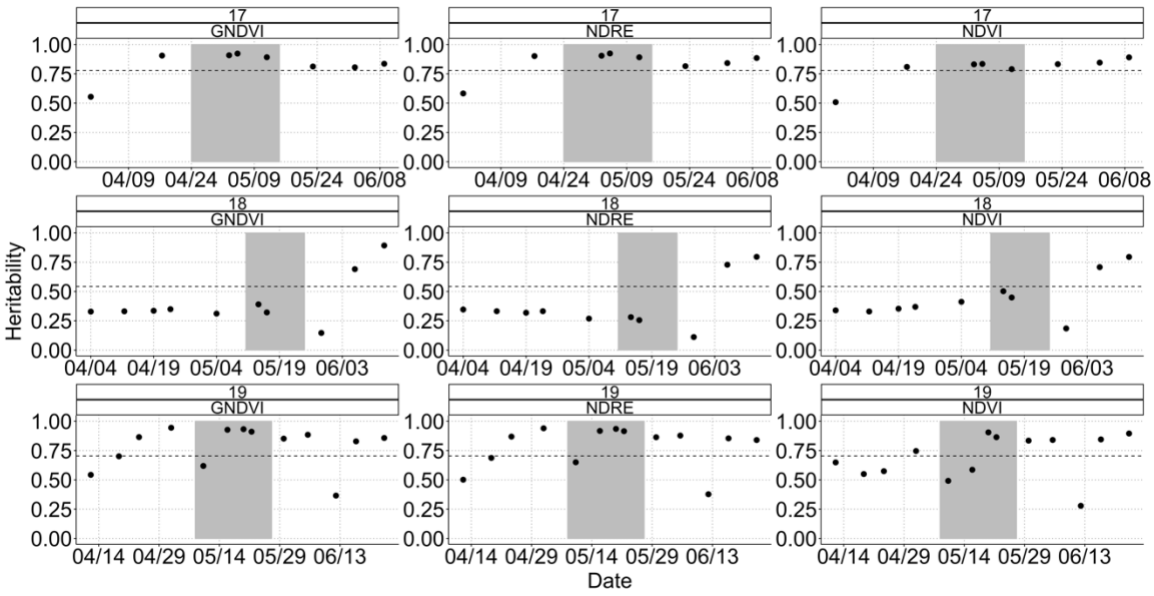


Figure 3.3. Broad-sense heritability over dates.

The broad-sense heritability of each season-VI shown corresponding to the observation date. The dotted line shows the heritability of grain yield for that year. Grey-shaded region show the period during which heading occurred for the population. NDVI = Normalized Difference Vegetation Index, GNDVI = Green Normalized Difference Vegetation Index, and NDRE = Red Edge Normalized Difference Vegetation Index.

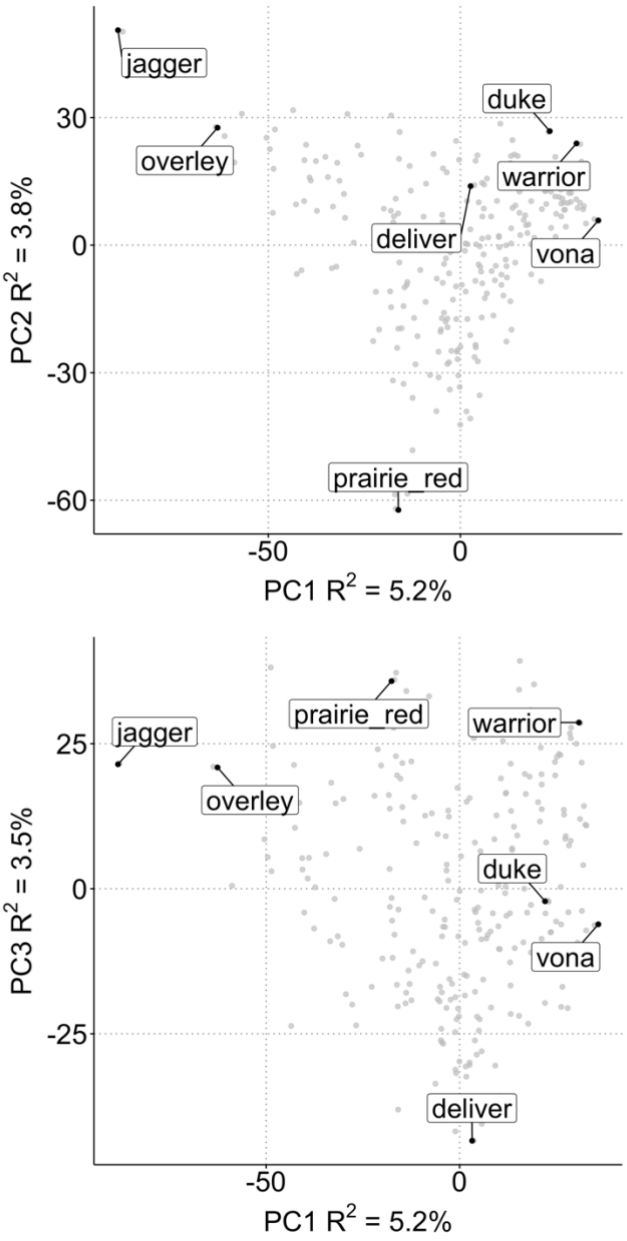


Figure 3.4. Principal component analysis from molecular markers in the association mapping panel.

A principal component analysis conducted on the genetic markers of the individuals in the population showing the first, second and third principle components. The percentage of the variation explained is given in the axis title. Identified founder lines are highlighted by name.

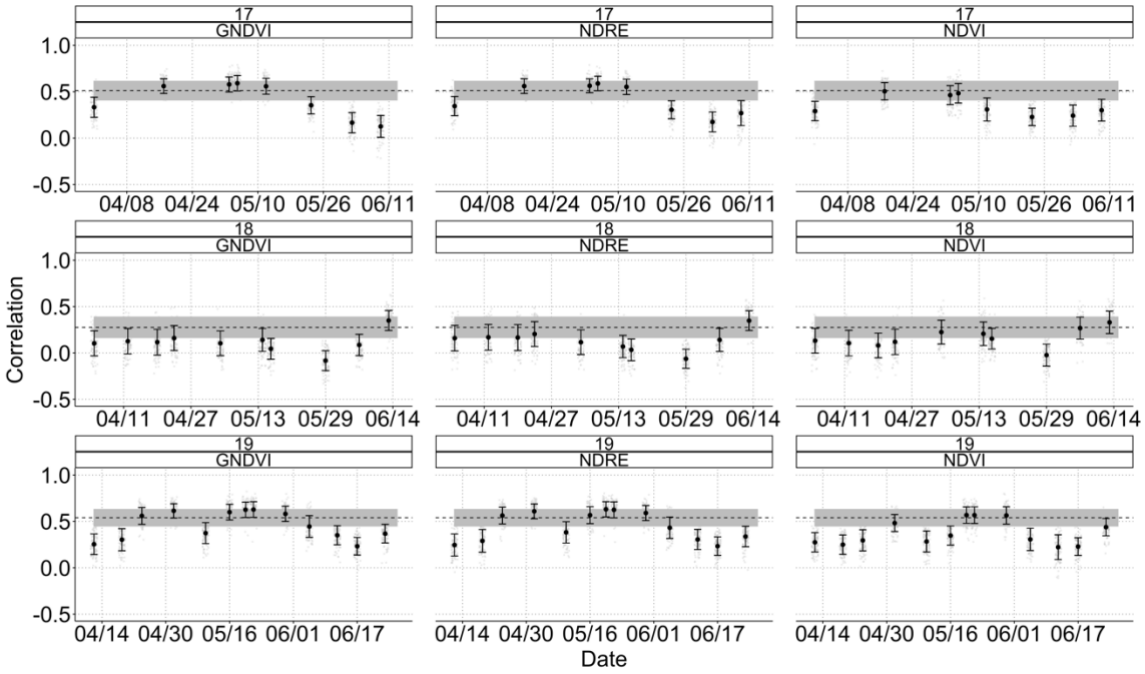


Figure 3.5. Time series of correlation between GEBVs and observed phenotypes for grain yield and VI.

The title of the plot gives the year-VI combination, with the date on which the VI was taken given on the x-axis. The dotted line is the mean correlation for the GEBVs for grain yield and the observed grain yield values, with the grey region being one standard deviation. The large point is the mean correlation for the GEBVs for the VI with the error bars showing one standard deviation. NDVI = Normalized Difference Vegetation Index, GNDVI = Green Normalized Difference Vegetation Index, and NDRE = Red Edge Normalized Difference Vegetation Index.

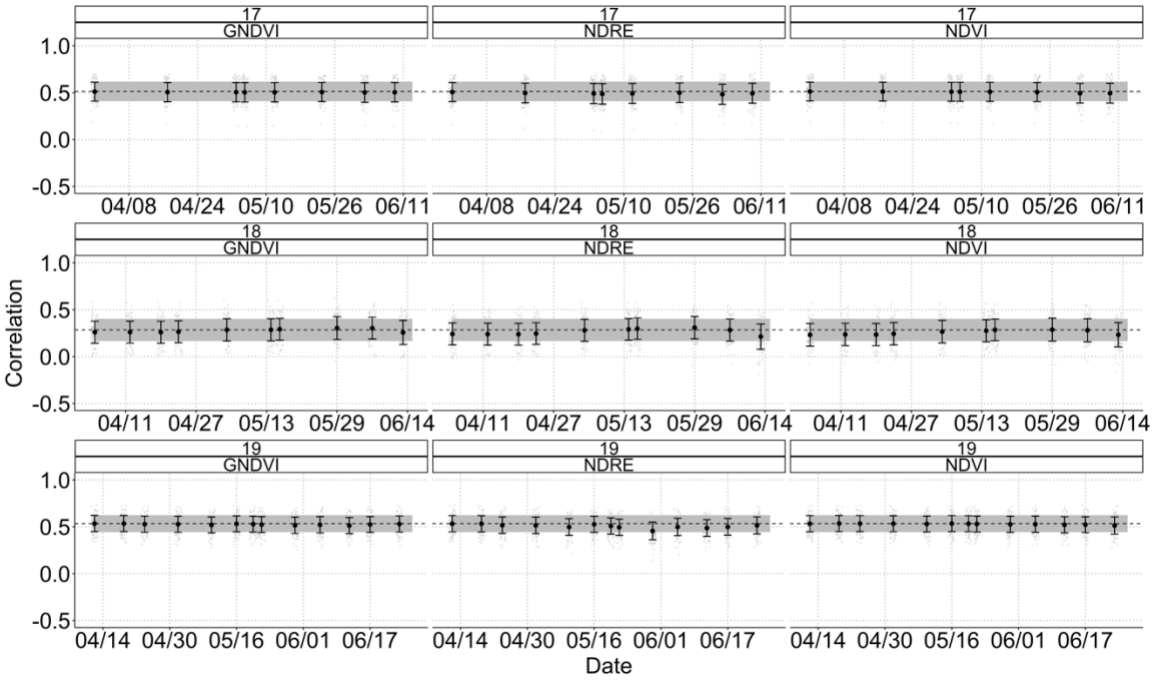


Figure 3.6. Time series of correlations between GEBVs and observed phenotypes for grain yield when using the VI as covariates.

The title of the plot gives the year-VI combination, with the date on which the VI was taken given on the x-axis. The dotted line is the mean correlation for the GEBVs for grain yield with no covariate and the observed grain yield values, with the grey region being one standard deviation. The large point is the mean correlation for the GEBVs with the VI as a covariate with the error bars showing one standard deviation. NDVI = Normalized Difference Vegetation Index, GNDVI = Green Normalized Difference Vegetation Index, and NDRE = Red Edge Normalized Difference Vegetation Index.

Chapter 4 - Breeding Program Optimization for Genomic Selection in Winter Wheat

Megan Calvert[a], Byron Evers[b], Xu Wang[b], Allan Fritz[c], Jesse Poland[b]

AFFILIATION

a Interdepartmental Genetics, Kansas State University, Manhattan Kansas 66506

b Department of Plant Pathology, Kansas State University, Manhattan, KS, 66506

c Department of Agronomy, Kansas State University, Manhattan, KS, 66506

KEYWORDS

Triticum aestivum, Plant Breeding, Genomic Prediction, Optimization

Abbreviations

AYN, advanced yield nurseries; BEL, Belleville; BLUPs, Best Linear Unbiased Predictions; GEBVs, genomic estimated breeding values; GNDVI, Green Normalized Difference Vegetation Index; GP, genomic prediction; GRVI, green-red vegetation indices; GRYLD, grain yield; GYP, Gypsum; GxE, genotype x environment; HTP, high-throughput phenotyping; HUTCH, Hutchinson; KSU, Kansas State University; MANH, Manhattan; MP, McPherson; NDRE, normalized difference red edge; NDVI, normalized difference vegetation indices; PTHT, plant height; PC, principal component; PYN, preliminary yield nurseries; RKHS, reproducing kernel Hilbert spaces regression; RR,

ridge regression; SNPs, Single nucleotide polymorphisms; UAS, uncrewed aerial system; VI, vegetation indices

Abstract

Developing methodologies in the fields of phenomics and genomic prediction have the potential to increase the production of crop species by accelerating germplasm improvement. The integration of these technologies into germplasm improvement and breeding programs requires evidence that there will be a direct economic benefit to the program. We determined a basic set of parameters, such as prediction accuracy greater than 0.3, the ability to genotype over 7 lines for the cost of one phenotypic evaluation, and heritability levels below 0.4, at which the use of genomic selection would be of economic benefit in terms of genetic gain and operational costs to the Kansas State University (KSU) winter wheat breeding program. The breeding program was then examined to determine whether the parameters benefitting genomic selection were observed or achievable in a practical sense. Our results show that the KSU winter wheat breeding program is at a decision point with regards to their primary means of selection. A few operational changes to increase prediction accuracy would place the program in the parameter space where genomic selection is expected to outpace the current phenotypic selection methodology at a parity of the operation cost and would be of greatest benefit to the program.

Introduction

Technologies are constantly evolving and growing in today's ever-changing world. The introduction of new technologies into an essential service and the development of new crop varieties can be challenging and costly (Moose and Mumm 2008). New technologies must be carefully evaluated to determine if they provide a significant enough advantage to adjust proven current practice. There are many technologies that appeared to have great promise, such as quantitative trait loci (QTL), and marker-assisted selection (MAS), and yet were not beneficial in a practical sense to breeding programs at the time (Bernardo 2008). Yet with a growing global population, erratic environmental conditions, and a finite-amount of arable land, the introduction of new technologies into our crop development systems is essential. The determination of which technologies to implement and the most efficient way to do so is a constant challenge facing plant breeders.

Wheat (*Triticum aestivum*) is one of the top three field crops planted in the USA, behind soybeans and corn, with 1.9 billion bushels produced in 2019 (Bond 2020). Wheat acres and production in the USA have been in a decline since the 1980s as a result of international competition, changing economic conditions and production practices. This is despite the increasing demand for wheat due to growing global populations, which is only expected to increase over the coming century. It is estimated that a 2% increase in yearly grain yields are required to meet these demands (Bassi et al. 2016). Currently there are no reports of wheat breeding programs achieving this level of

gain, making wheat an important candidate for breeding technologies designed to accelerate variety development.

Genomic prediction (GP) is one of the new technologies that is showing a great deal of promise to assist in crop-variety production. GP involves the use of genome-wide markers to predict the breeding value of individuals in a population (Meuwissen, Hayes, and Goddard 2001). This is done by genotyping and phenotyping a training population which is used to establish a model for the trait of interest. This model is then used to calculate genomic estimated breeding values (GEBVs) of a population for which only genotype information is available. GP is already a common technique used in animal breeding due to the benefit of being able to predict a phenotype without having to observe the phenotype, for example the milk yield of a bull's offspring (Hayes et al. 2009; Georges, Charlier, and Hayes 2019).

Plant breeding programs have yet to fully utilize GP for a number of reasons. Breeding programs have to test the same experimental line in multiple locations due to the need to select varieties that are stable across environments. This negates some of the benefit that GP supplies to animal breeding, in which the same genotype cannot be tested under multiple conditions. The genotype x environment (GxE) variation is often high in plant breeding populations due in part to the large weather differences experienced between locations. This significantly decreases the accuracies of most GP models (Dawson et al. 2013; Heslot et al. 2012). Previous GP models did not account for GxE interactions which limited inference and selection decisions in plant breeding. Newer genomic prediction models can now take into account these GxE interactions

and are showing greater accuracy in plant breeding situations (Burgueño et al. 2012; Pérez-Rodríguez et al. 2017; Montesinos-López et al. 2016).

The accuracy of GP models has shown some improvement when a covariate is included (Crain et al. 2018; Rutkoski et al. 2016). This has had some success when including high-throughput phenotyping (HTP) data that is taken throughout the course of the growing season. HTP techniques have the ability to measure thousands of phenotypes accurately in a short period of time. Many of the HTP techniques used in breeding programs take advantage of new developments in remote sensing, uncrewed aerial system (UAS), and sensor technology to measure reflectance and temperature phenotypes. These are all phenotypes that have been shown to be correlated with yield and as such could be effective secondary targets for high-throughput phenotyping and indirect selection (Elliott and Regan 1993; Blackmer et al. 1996; Curran et al. 1983).

In addition to the use of correlated traits, the efficiency of GP is also affected by the stage of the breeding program when it is implemented. Primarily, GP is used to predict the breeding value which is comprised of the additive genetic variation. The greatest advantages of GS will be found when the selection candidates still encompass the most additive genetic variation for the phenotype of interest (Bassi et al. 2016). This is more likely to occur in earlier stages of a breeding program as it is easily selected for, while later stages of a breeding have progressed through strong selection and are more likely to select for other epistatic genetic variation on the performance of the line *per se*.

Another factor in addition to stage of implementation that has hindered the adoption of GP in plant breeding programs are the costs and complex logistics

associated with it. The development and establishment of genotyping practices in a plant breeding program requires an investment in infrastructure and training (Moose and Mumm 2008). Along with genetic marker costs, the establishment and maintenance of an adequate training population adds additional costs to the GP protocols. These costs may not be offset by the gains that the program could potentially achieve using GP (Bassi et al. 2016; Jarquín et al. 2017).

In this study we examined the implementation of GP in a wheat breeding program to specifically examine 1) what parameters/infrastructure is required to implement GP as a primary selection strategy, and 2) are those parameters being met in the KSU public breeding program?

Method and Materials

Cost-Benefit Simulation

A simulated breeding program was used to determine the expected genetic gain when basing selection of material exclusively on observed line performance *per se*, or exclusively on prediction of breeding values. The simulation assumes that a breeding program does not use a combination of phenotypic selection and genomic selection at the same time. In consultation with the Kansas State Hard Winter Wheat Breeding Program, referred to as the KSU breeding program from here on, an appropriate range of cost estimates were determined, as well as sizes of the program as determined by the number of lines developed and evaluated (Supplementary Table 1). For simplicity looking at a single stage of selection in the program, the simulated values assume that

the cycle length for each selection scheme is the same and covers the period of the program before the Advanced Yield Nursery (AYN) stage. It also assumes that the number of lines that will be selected for the AYN stage are the same regardless of which selection method is used.

The cost of line developments was estimated to be between $4-$30 per line which covers the initial cross and formation of an inbred line. This can either be done by several years of inbreeding or by the formation of double haploid lines. This was applied as a fixed cost needed to develop the initial population for selection regardless of which selection method was used.

Phenotypic Selection

The costs of phenotypic observation plots are estimated to be between $12-$40 per plot and evaluated within this range at $5 increments. To obtain an accurate phenotypic measurement, replications of the experimental line need to be planted and phenotyped. Depending on the structure of the program this may mean several replications at a single site, or fewer replications at several sites. The maximum number of experimental breeding lines was calculated as:

$$ExperimentalLines = {Budget}\Big/{(LineDevelopment + (PhenoCost * rep))} \quad (1)$$

where $ExperimentalLines$ is the number of lines that would be advanced to the phenotyping stage of selection, the $Budget$ is the total monetary budget for that stage of the breeding program, $LineDevelopment$ is the cost of advancing a single cross to an inbred experimental line for evaluation, $PhenoCost$ is the total cost to

obtain the phenotype of interest including labor and other miscellaneous operation costs, and rep is the number of replications of each experimental line that are planted in field and will require phenotyping.

The number of experimental breeding lines was used as the population size when estimating other population parameters.

Genomic Selection

The cost of genotyping a single line for prediction was estimated to be between $1-$10 and evaluated within this range at $1 increments. The maximum number of experimental lines that can be genotyped for prediction was calculated as:

$$GenotypedLines = {}^{Budget}/_{LineDevelopment + GenoCost} \quad (2)$$

Where $GenotypedLines$ is the number of lines that would be evaluated by genotyping, the $Budget$ is the total monetary budget for that stage of the breeding program, $LineDevelopment$ is the cost of advancing a single cross to an inbred experimental line for evaluation, and $GenoCost$ which is the cost of genotyping a single line including labor and other operational costs. It is assumed that the genotyping is only performed once, and that replication is not required.

The number of possible genotyped lines was then used as the population size when estimating other population parameters.

Simulation Details

The selection methods were compared for every possible combination of number of experimental lines phenotyped and number of experimental lines genotyped,

based on the ratio between the correlated response to selection, and the response to selection (Falconer and Mackay 2009).

The expected response to selection for the primary trait was calculated by:

$$R = ih\sigma_A \tag{3}$$

Where i is the intensity of selection, h is the square-root if the narrow-sense heritability of the primary trait, and σ_A is the standard deviation of the additive genetic variance (Falconer and Mackay 2009).

For this study, the GEBVs are assumed to be the secondary trait that is correlated to the primary trait, which would be through experimental observation plots. The correlated response to selection is calculated by:

$$CR = i * \sqrt{h_x^2} * \sqrt{h_y^2} * r_g * \sigma_{py} \tag{4}$$

Where i is the intensity of selection, $\sqrt{h_x^2}$ is the square-root of the narrow-sense heritability of the response trait, $\sqrt{h_y^2}$ is the square-root of the narrow-sense heritability of the secondary trait, r_g is the additive genetic correlation between the traits, and σ_{py} is the standard deviation of the phenotypic variation for the secondary trait (Falconer and Mackay 2009).

A comparison between the indirect response to selection and the expected response to selection is best demonstrated as (Falconer and Mackay 2009):

$$\frac{CR_x}{R_x} = \frac{i_Y r_g \sqrt{h_y^2}}{i_x \sqrt{h_x^2}} \tag{5}$$

Where i_Y is the selection intensity on the secondary trait, r_g is the additive genetic correlation between the primary and secondary trait, $\sqrt{h_y^2}$ is the square-root of

the narrow-sense heritability of the secondary trait, i_x is the selection intensity on the direct trait, and $\sqrt{h_x^2}$ is the square-root of the narrow-sense heritability of the response trait.

It was assumed that the individual phenotypes were made up of a genetic portion and an environmental portion. The distribution of each of these portions was assumed to be a random normal with a mean of 0, and a standard deviation of $\sqrt{h^2}$ or $\sqrt{(1 - h^2}$ for the genotypic and environmental proportions, respectively. A sample the size of the number of lines that were possibly genotyped was taken from this distribution for each individual to give the overall phenotypic distribution. This overall phenotypic distribution was used to determine the intensity of selection (i) in terms of the standard deviation and the selection differential with the msm package (Jackson 2011).

The narrow sense heritabilities for the response trait and the additive genetic correlation were set, with testing the ranges of 0 and 1 at increments of 0.1 each. The narrow-sense heritability of the secondary trait, the genotyping, was assumed to be 0.95. This is under the assumption that the genotypes are inherited almost exactly as they are sequenced and that there are only a few genotyping errors. The ratio between the correlated response and the expected response to selection was plotted against the ratio between number of experimental lines phenotyped and number of lines genotyped.

Plant Material

The KSU Breeding Program breeds hard red winter wheat for a large area which contains different mega-environmental conditions. A subset of 5 locations within the

same Kansas mega-environment based on breeder knowledge, Belleville (BEL), Gypsum (GYP), McPherson (MP), Hutchinson (HUTCH) and Manhattan (MANH), were selected for analysis in Kansas between 2016 and 2019. This resulted in 1989 experimental lines being examined over the course of 4 years. These sites contain trials from the Preliminary Yield Trials (PYN, primary $F_{5:7}$) and the advance yield nursery (AYN, primarily $F_{5:8}$). The planting, harvest dates and trial size are provided in Table 1. These locations, excluding Manhattan and Hutchinson, are located in farmers' fields under typical grower management practices. The PYN and AYN trials were all planted in six-row plots of 1.5m by 4.5m. The PYN are planted in a modified augmented design with one replicate of the experimental line per location (Federer and Raghavarao 1975). Plant checks are planted across whole rows and columns in the trial, and sub-block checks are assigned randomly within each block. The AYN is made up of lines selected from the PYN trials. The lines are planted using two replicated α-lattice designs (Patterson and Williams 1976). All 5 locations were planted each year but if a site experienced extreme environmental variation from the normal climate it was not harvested, providing an unbalanced set of data.

Phenotyping

Phenotypic information was collected either by combine for grain yield (GRYLD), by UAS for vegetation indices (VIs), or by hand using the Field Book application (Rife and Poland 2014) for plant height (PTHT). A DJI Matrice100 (DJI, USA) quadcopter UAS was equipped with a 5-band multi-spectral RedEdge camera (MicaSense Inc. USA) to collect plot-level reflectance values for each year, based on the standard protocols

developed by The Wheat Genetics Lab at Kansas State University as laid out in Wang et al. (2018). The UAS data was collected throughout the course of the growing season, once every 7-10 days depending on weather conditions. Plant height was collected manually during the grain ripening stage before harvest. VIs were calculated from the reflectance values based on the protocols laid out in Wang et al. (2018) and given in Table 2.

Genotyping

The 1989 lines in the study years were sequenced using genotype-by-sequencing on an Illumina Hi Seq2000 or Hi Seq2500 (Elshire et al. 2011). Single nucleotide polymorphisms (SNPs) were called with the Tassel software with the Chinese Spring wheat assembly v1.0 as a reference (Bradbury et al. 2007; Glaubitz et al. 2014; International Wheat Genome Sequencing Consortium et al. 2018). The final data set included 8182 SNPs that were selected for use passed one of three filtering criteria optimized for the wheat genome by Shrestha et al. (2020) that include Chi-square, Fisher's test for independence, and the inbreeding coefficient, as well as having a minor allele frequency greater than 0.05 and missing less than 20% of the data. Missing SNPs were imputed with Beagle 5.1 (Browning, Zhou, and Browning 2018).

Data Analysis

The Best Linear Unbiased Predictions (BLUPs) for each line were calculated for each individual year and for multiple years, including or excluding locations as needed.

Where multiple years and locations are included, such as for GRYLD, the BLUPs were calculated by:

$$y_{ijklt} = \mu + G_i + S_j + M_{l(j)} + GM_{il(j)} + R_{k(lj)} + P_{t(lj)} + \varepsilon_{ijklt} \quad (6)$$

where y_{ijklt} is the phenotypic response variable, μ is the fixed overall mean, G_i is the random genotype effect for line i distributed as iid $G_i \sim N(0, \sigma_i^2)$, S_j is the random effect for year *j* distributed as iid $S_j \sim N(0, \sigma_j^2)$, $M_{l(j)}$ is the random effect for location l within year *j* distributed as iid $M_{l(j)} \sim N(0, \sigma_l^2)$, $GM_{il(j)}$ is the random genotype by location effect nested within year *j* distributed as iid $GM_{il(j)} \sim N(0, \sigma_{il}^2)$, $R_{k(lj)}$ is the random effect of replication *k* within location and year distributed as iid $R_{k(lj)} \sim N(0, \sigma_k^2)$, $P_{t(lj)}$ is the random fungal treatment effect *t* nested within year-location distributed as iid $P_{t(lj)} \sim N(0, \sigma_t^2)$, and ε_{ijklt} as the residual effect distributed as iid $\varepsilon_{ijklt} \sim N(0, \sigma_\delta^2)$.

When only a single year with multiple locations is included the BLUPs are calculated by:

$$y_{ijkl} = \mu + G_i + L_j + GL_{ij} + R_{k(j)} + T_{l(j)} + \varepsilon_{ijkl} \quad (7)$$

where y_{ijkl} is the phenotypic response variable, μ is the fixed overall mean, G_i is the random genotype effect for line i distributed as iid $G_i \sim N(0, \sigma_i^2)$, L_j is the random effect of location *j* distributed as iid $L_j \sim N(0, \sigma_j^2)$, GL_{ij} is the random effect of genotype by location distributed as iid $GL_{ij} \sim N(0, \sigma_{ij}^2)$, $R_{k(j)}$ is the random effect of replication *k* nested in location *j* distributed as iid $R_{k(j)} \sim N(0, \sigma_k^2)$, $T_{l(j)}$ is the random

effect of fungal treatment l nested within location *j* distributed as iid $T_{l(j)} \sim N(0, \sigma_T^2)$, , and ε_{ijkl} is the residual effect distributed as iid $\varepsilon_{ijkl} \sim N(0, \sigma_\delta^2)$.

The broad-sense heritability for all years and multiple year-locations was calculated as:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gxe}^2}{y*l} + \frac{\sigma_e^2}{y*l*r}} \quad (8)$$

Where σ_g is the total genetic variance, σ_{gxe} is the variance contributed by the location and variety nested within year, σ_e is the residual environmental variance. As the data is unbalanced, y is the harmonic mean of the number of years planted, l is the harmonic mean of the number of locations planted within year, and r is the harmonic mean of the number of replications per location per year (Holland, Nyquist, and Cervantes-Martínez 2010).

The broad-sense heritability for individual year AYN and individual year-multiple location PYN was calculated by:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gxe}^2}{l} + \frac{\sigma_e^2}{l*r}} \quad (9)$$

where σ_g^2 is the total genetic variance, σ_{gxe}^2 is the variance contributed by the genotype-location combination, and σ_e^2 is the residual environmental variance. Similar to equation 8, l is the harmonic mean of the number of locations planted and r is the harmonic mean of the number of replications per location (Holland, Nyquist, and Cervantes-Martínez 2010).

The VI phenotypes were mainly considered on a year-location-trial basis. As such the BLUPs for the VI phenotypes for the AYN trials are calculated by:

$$y_{irb} = \mu + G_i + M_r + B_{r(b)} + \varepsilon_{irb} \quad (10)$$

where y_{irb} is the phenotypic response variable, μ is the fixed overall mean, G_i is the random genotypic effect for line i distributed as iid $G_i \sim N(0, \sigma_i^2)$, M_r is the random effect of the replicate *r* distributed as iid $M_r \sim N(0, \sigma_r^2)$, $B_{r(b)}$ is the random effect of the experimental block nested within replicate distributed as iid $B_{r(b)} \sim N(0, \sigma_b^2)$, and ε_{irb} is the residual effect distributed as iid $\varepsilon_{irb} \sim N(0, \sigma_\delta^2)$. The broad-sense heritability was calculated for each year-location's AYN that had VI phenotypes by:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{error}^2}{r}} \quad (11)$$

where σ_g^2 is the genetic variance, σ_{error}^2 is the residual environmental variance and r is the number of replicates.

The BLUPs for the VI phenotypes for the PYN trials are calculated by:

$$y_{ib} = \mu + G_i + B_b + \varepsilon_{ib} \quad (12)$$

where y_{ib} is the phenotypic response variable, μ is the fixed overall mean, G_i is the random genotypic effect N(0,σ^2), $B_b$is the random effect of the experimental block N(0,σ^2), and $\varepsilon_{ib}$is the residual effect.

Genome-wide Association Analysis

A principal component (PC) analysis of the genotypic information was conducted with the pcaMethods package (Stacklies et al. 2007). A genome-wide association analysis was performed for the VI BLUPs for each year-location-trial using the rrBLUP package (Endelman 2011). The kinship matrix was calculated using rrBLUP was included with 4 PCs to account for kinship and population structure respectively (Endelman and

Jannink 2012). A Bonferroni correction was applied with α = 0.05 to determine significance.

Genomic Prediction

Genomic prediction was performed with the rrBLUP package and the BGLR statistical package (Pérez and de los Campos 2014) in the R software environment. The rrBLUP package performs ridge regression (RR), and the BGLR package was used to perform a BayesC based prediction and a reproducing kernel Hilbert spaces regression (RKHS). A cross-validation strategy of 100 replications with 80% of the population in the training data set and 20% of the population in the predicted data set was used. The accuracy of the prediction was determined by the correlation between the predicted and observed values. Across year and location predictions were also performed where years, year-trial, and year-location combinations that were not included in the training population are predicted from the remaining data. An example of this would be to use all years excluding the 2016 season as the training data set and the 2016 season as the prediction data set, or all years excluding 2016 season except for the 2016-HUTCH data as the training data set and the remaining three 2016 locations as the prediction data set.

For the VI data only rrBLUP was used for genomic prediction. The VI were used as a cofactor when predicting grain yield for that specific year-location-trial combination as follows:

$$y = \mu + X\beta + Zu + \varepsilon \quad (13)$$

where y is the BLUP for grain yield, μ is the overall mean, X is a $(n\ x\ 1)$ matrix of the individual observations of the VI, β is the fixed effects of the VI measurements, Z is an $(n\ x\ m)$ matrix assigning markers to genotypes, u is a $(1\ x\ n)$ array of the random effects of the markers and ε is the residual error. The same cross-validation procedure as before was used.

Unless otherwise stated, all analysis took place in the R software environment using the Tidyverse suite of packages (Core Team 2020; Wickham et al. 2019) and visualised with ggplot2 (Wickham 2016). The required code can be found at: https://github.com/megzcalvert/ProgramBreeding

Results

Simulation

To determine the optimal parameters for the operation of the KSU winter wheat breeding program comparing current phenotypic selection methodology verse genomic prediction a simulation was created based on economic decisions. We observed that at low heritabilities of the primary trait, prediction is favored even at low prediction accuracies (Figure 1). Once the ratio between the correlated response and the expected response is greater than 1, then selection on the correlated trait is expected to give greater genetic gain than selection on the primary trait. For the KSU wheat breeding program a primary focus on genotyping is beneficial in terms of genetic gain when the narrow-sense heritability of grain yield is below 0.4, approximately 7.5 lines can be

genotyped for every one line that can be phenotyped, and the prediction accuracy of the genomic prediction models is greater than 0.3.

Heritability of traits

Heritability was calculated to identify the phenotypic variance that could be partitioned to the genetic component across the experimental trials. The broad sense heritabilities for each of the seasons ranged from 0.022 for the AYN to 0.441 for the PYN in the 2016 season (Figure 2). The broad-sense heritability for the 2019 PYN cannot be calculated as only one rep was planted at one location (HUTCH).

The VIs show a moderate heritability across the seasons, with the majority being above 0.5 (Figure 3) The heritabilities are variable across the course of the season and locations but the same date-location combination shows similarity across the various VIs, indicating a large influence of ambient conditions of various days within the growing season on the accuracy of VI measurements.

The correlations between grain yield for a location-season-trial combination and the VI's for that location-season-trial combination over the course of the season are given in Figure 4. The correlations ranged between -0.6 in the 2018 PYN at HUTCH for NDRE, to 0.69 in the 2018 AYN at MP for NDRE.

GWAS

To try and determine if any of the VI had genomic regions associated with a large effect size, a GWAS was conducted. As population structure is known to have an effect on GWAS results a PCA of the 1989 experimental lines genotypic data was performed.

The PCA shows no distinct population structure and the plot is bounded by the expected founder genotypes, Figure 5. There were no major QTLs for grain yield in any year, yet several of the VIs showed significant associations with regions of the genome. When these associations were further examined it was shown that they had an influence on the VI value but not GRYLD (Supplementary Figure 1). The regions that showed an association with the VI were often shared between different VI (3/4 in the 2018 season and 3/3 in the 2018 season, Supplementary Figure 1).

Genomic Prediction

Based on the GWAS, GP was evaluated to predict the highly quantitative trait of grain yield. All of the tested genomic prediction models produced similar accuracies (Supplementary Table 2), we therefore focused on rrBLUP models for computational efficiently for further analysis which will be reported. The genomic prediction accuracies based on cross-validation range between 0.311 (SD = 0.079) for the 2018 season and 0.469 (SD = 0.105) for the 2017 season (Figure 6, Supplementary Table 2).

When making forward predictions the strongest correlation, -0.164, was achieved using all seasons excluding the 2017 season as the training population, and the 2017 season as the prediction population (Table 4). When predicting all locations in a single season except for HUTCH, using the data from other seasons and the data from HUTCH as the training population, the greatest accuracy achieved was 0.572 (95% CI [0.503, 0.634]) for 2019 PYN in MANH and BEL. HUTCH was chosen as the location to include in the forward predictions as it is the first location that is normally harvested in

Kansas. This would be similar to harvesting one location and then using the predictions to make selections.

To evaluate a prediction approach using secondary traits, the VI were used as covariates in the prediction of grain yield. The prediction of grain yield was not improved by the addition of a covariate (Figure 7). The prediction of grain yield without a covariate was performed at the same time for comparison and was comparable to the best estimate using VI as covariate.

Discussion

Changing global conditions require current food production systems to be resilient against unexpected events in the face of global population growth. This will require the development of crop varieties that are adapted to hotter and dryer climates. The speed at which these varieties are developed will require the adoption of new technologies that have been proven to show a positive economic investment.

This study examined the validity of using GP and other HTP techniques in the Kansas State University winter wheat breeding program based on several population and model parameters, such as heritability of the primary trait, the prediction accuracy and the selection intensity. These parameters give an indication of whether a new GP method would provide more genetic gain than traditional phenotypic selection equal resource expenditures.

Heritability

The heritability of a trait has a significant impact on whether the trait can be selected for and the possibility for genetic gain of that trait. Traits with a lower heritability are difficult to select for regardless of which selection method is used. However, our simulation showed that traits with a low to moderate narrow-sense heritability (0-0.4) favor the use of GP. For the Kansas State University winter wheat breeding program, the overall broad-sense heritability of grain yield is below 0.3. As the narrow-sense heritability is the always less than the broad-sense heritability, the heritability favors GP under the operation cost parameters for the program.

Prediction Accuracy

The prediction accuracy for GP when making forward predictions in the breeding program do not meet the criteria to favor GP unless very large populations are used. This is a possibility in a breeding program setting as historical grain yield data and genotypes are available for previous seasons. The experimental lines in these seasons are likely to be less related to the current lines as the parental lines in the crossing block are updated which may lower prediction accuracy, however, this is the critical assessment of prediction accuracy that is needed for implementation in the breeding program as all selections will be focused on new breeding lines into new year's.

HTP Prediction Accuracy

Utilizing the VI to increase the prediction accuracy of the GP models requires more optimization before it is commonly adopted. The VI need to be able to be used

across locations and years for them to really have an impact on the prediction accuracy of the GP models. This will require either the manual measurement of growth stages to standardize measurements to growth stage across locations and years, or the use of another method such as thermal units to account for differences in growth stage in the training populations and prediction populations. This additional labor makes the use of VIs in GP models more costly than utilizing just the GP model. It is an additional factor that needs to be taken into account before the decision to transition to GP is made.

Conclusion

When all parameters are considered it appears that the Kansas State University winter wheat breeding program is on the edge of a large decision. The heritability of grain yield in the breeding program as well as the cost of phenotyping compared to genotyping favor genomic prediction as the way forward for the breeding program. Yet the low accuracy of forward predictions favors the use of phenotypic selection. The forward prediction accuracy can be increased as seen in the 2019 season (Table 4), but this requires a much larger training population. With a few adjustments to the experimental design such as allowing for more replicates of the training population, and changes to the program workflow that allows for the loss of the PYN, genomic prediction could allow the KSU winter wheat breeding program to make larger genetic gains.

Authors' contributions

MC and BE collected data. MC and XW analyzed data. JP conceived and planned the experiments. All authors read and approved the final manuscript.

Acknowledgements

We would like to acknowledge the undergraduate assistants, Quentin Ediger, Kyle Laessig, Jacob Ramsey, Jamie Clark, Molly Smith, Erika Kringen, Nicholas VanPelt, Bryce Teaford, who assisted with data collection. Daljit Singh, Richard Brown and Grant Williams who assisted with UAS flights and data collection.

MC was supported in part by fellowship funds from the Interdepartmental Genetics program at Kansas State University. This material is based upon work supported by the National Science Foundation under Grant No. 1238187 project "A Field-Based High-Throughput Phenotyping Platform for Plant Genetics" and the USDA NIFA International Wheat Yield Partnership grant no. 2017-67007-25933/project accession no. 1011391 project "Wheat Yield Prediction and Advanced Selection Methodologies through Field-Based High-Throughput Phenotyping with UAVs". Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the U.S. Dept. of Agriculture.

References

Bassi, Filippo M., Alison R. Bentley, Gilles Charmet, Rodomiro Ortiz, and Jose Crossa. 2016. "Breeding Schemes for the Implementation of Genomic Selection in Wheat (*Triticum* Spp .)." Plant Science 242 (January): 23–36. https://doi.org/10.1016/j.plantsci.2015.08.021.

Bernardo, Rex. 2008. “Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years.” Crop Science 48 (5): 1649. https://doi.org/10.2135/cropsci2008.03.0131.

Blackmer, Tracy M., James S. Schepers, Gary E. Varvel, and Elizabeth A. Walter-Shea. 1996. “Nitrogen Deficiency Detection Using Reflected Shortwave Radiation from Irrigated Corn Canopies.” Agronomy Journal 88 (1): 1–5. https://doi.org/10.2134/agronj1996.00021962008800010001x.

Bond, Jennifer K. 2020. “USDA ERS - Wheat Sector at a Glance.” June 26, 2020. https://www.ers.usda.gov/topics/crops/wheat/wheat-sector-at-a-glance/.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. “TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples.” Bioinformatics 23 (19): 2633–35. https://doi.org/10.1093/bioinformatics/btm308.

Browning, Brian L., Ying Zhou, and Sharon R. Browning. 2018. “A One-Penny Imputed Genome from Next-Generation Reference Panels.” The American Journal of Human Genetics 103 (3): 338–48. https://doi.org/10.1016/j.ajhg.2018.07.015.

Burgueño, Juan, Gustavo de los Campos, Kent Weigel, and José Crossa. 2012. “Genomic Prediction of Breeding Values When Modeling Genotype × Environment Interaction Using Pedigree and Dense Molecular Markers.” Crop Science 52 (2): 707–19. https://doi.org/10.2135/cropsci2011.06.0299.

Core Team, R. 2020. R: A Language and Environment for Statistical Computing (version 4.0.0). Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Crain, Jared, Suchismita Mondal, Jessica Rutkoski, Ravi P. Singh, and Jesse Poland. 2018. “Combining High-Throughput Phenotyping and Genomic Information to Increase Prediction and Selection Accuracy in Wheat Breeding.” The Plant Genome 11 (1): 0. https://doi.org/10.3835/plantgenome2017.05.0043.

Curran, P. J., M. D. Steven, John Theodore Houghton, Arthur Herbert Cook, and Henry Charnock. 1983. “Multispectral Remote Sensing for the Estimation of Green Leaf Area Index.” Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences 309 (1508): 257–70. https://doi.org/10.1098/rsta.1983.0039.

Dawson, Julie C., Jeffrey B. Endelman, Nicolas Heslot, Jose Crossa, Jesse Poland, Susanne Dreisigacker, Yann Manès, Mark E. Sorrells, and Jean-Luc Jannink. 2013. “The Use of Unbalanced Historical Data for Genomic Selection in an International Wheat Breeding Program.” Field Crops Research 154 (December): 12–22. https://doi.org/10.1016/j.fcr.2013.07.020.

Elliott, G. A., and K. L. Regan. 1993. “Use of Reflectance Measurements to Estimate Early Cereal Biomass Production on Sandplain Soils.” Australian Journal of Experimental Agriculture 33 (2): 179–83. https://doi.org/10.1071/ea9930179.

Endelman, Jeffrey B. 2011. “Ridge Regression and Other Kernels for Genomic Selection with R Package RrBLUP.” The Plant Genome Journal 4 (3): 250. https://doi.org/10.3835/plantgenome2011.08.0024.

Falconer, Douglas S., and Trudy F. C. Mackay. 2009. Introduction to Quantitative Genetics. 4. ed., [16. print.]. Harlow: Pearson, Prentice Hall.

Federer, W. T., and D. Raghavarao. 1975. “On Augmented Designs.” Biometrics 31 (1): 29–35. https://doi.org/10.2307/2529707.

Georges, Michel, Carole Charlier, and Ben Hayes. 2019. “Harnessing Genomic Information for Livestock Improvement.” Nature Reviews Genetics 20 (3): 135–56. https://doi.org/10.1038/s41576-018-0082-2.

Glaubitz, Jeffrey C., Terry M. Casstevens, Fei Lu, James Harriman, Robert J. Elshire, Qi Sun, and Edward S. Buckler. 2014. “TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline.” Edited by Nicholas A. Tinker. PLoS ONE 9 (2): e90346. https://doi.org/10.1371/journal.pone.0090346.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. “Invited Review: Genomic Selection in Dairy Cattle: Progress and Challenges.” Journal of Dairy Science 92 (2): 433–43. https://doi.org/10.3168/jds.2008-1646.

Heslot, Nicolas, Hsiao-Pei Yang, Mark E. Sorrells, and Jean-Luc Jannink. 2012. “Genomic Selection in Plant Breeding: A Comparison of Models.” Crop Science 52 (1): 146–60. https://doi.org/10.2135/cropsci2011.06.0297.

Holland, James B., Wyman E. Nyquist, and Cuauhtemoc T. Cervantes-Martínez. 2010. “Estimating and Interpreting Heritability for Plant Breeding: An Update.” In Plant Breeding Reviews, 9–112. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470650202.ch2.

International Wheat Genome Sequencing Consortium, Michael Alaux, Jane Rogers, Thomas Letellier, Raphaël Flores, Françoise Alfama, Cyril Pommier, et al. 2018. “Linking the International Wheat Genome Sequencing Consortium Bread Wheat Reference Genome Sequence to Wheat Genetic and Phenomic Data.” Genome Biology 19 (1): 111. https://doi.org/10.1186/s13059-018-1491-4.

Jackson, Christopher. 2011. “Multi-State Models for Panel Data: The Msm Package for R.” Journal of Statistical Software 38 (1): 1–28. https://doi.org/10.18637/jss.v038.i08.

Jarquín, Diego, Cristiano Lemes da Silva, R. Chris Gaynor, Jesse Poland, Allan Fritz, Reka Howard, Sarah Battenfield, and Jose Crossa. 2017. “Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype × Environment Interactions in Kansas Wheat.” The Plant Genome 10 (2): plantgenome2016.12.0130. https://doi.org/10.3835/plantgenome2016.12.0130.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. “Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps.” Genetics 157 (4): 1819–29.

Montesinos-López, Osval A., Abelardo Montesinos-López, José Crossa, Fernando H. Toledo, Oscar Pérez-Hernández, Kent M. Eskridge, and Jessica Rutkoski. 2016. “A Genomic Bayesian Multi-Trait and Multi-Environment Model.” G3: Genes|Genomes|Genetics 6 (9): 2725–44. https://doi.org/10.1534/g3.116.032359.

Moose, Stephen P., and Rita H. Mumm. 2008. “Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement.” Plant Physiology 147 (3): 969–77. https://doi.org/10.1104/pp.108.118232.

Patterson, H. D., and E. R. Williams. 1976. “A New Class of Resolvable Incomplete Block Designs.” Biometrika 63 (1): 83–92. https://doi.org/10.2307/2335087.

Pérez, Paulino, and Gustavo de los Campos. 2014. “Genome-Wide Regression and Prediction with the BGLR Statistical Package.” Genetics 198 (2): 483–95. https://doi.org/10.1534/genetics.114.164442.

Pérez-Rodríguez, Paulino, José Crossa, Jessica Rutkoski, Jesse Poland, Ravi Singh, Andrés Legarra, Enrique Autrique, Gustavo de los Campos, Juan Burgueño, and Susanne Dreisigacker. 2017. “Single-Step Genomic and Pedigree Genotype × Environment Interaction Models for Predicting Wheat Lines in International Environments.” The Plant Genome 10 (2): 0. https://doi.org/10.3835/plantgenome2016.09.0089.

Poland, Jesse A., Patrick J. Brown, Mark E. Sorrells, and Jean-Luc Jannink. 2012. “Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach.” Edited by Tongming Yin. PLoS ONE 7 (2): e32253. https://doi.org/10.1371/journal.pone.0032253.

Rife, Trevor W., and Jesse A. Poland. 2014. “Field Book: An Open-Source Application for Field Data Collection on Android.” Crop Science 54 (4): 1624–27. https://doi.org/10.2135/cropsci2013.08.0579.

Rutkoski, Jessica, Jesse Poland, Suchismita Mondal, Enrique Autrique, Lorena González Pérez, José Crossa, Matthew Reynolds, and Ravi Singh. 2016. “Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat.” G3: Genes|Genomes|Genetics 6 (9): 2799–2808. https://doi.org/10.1534/g3.116.032888.

Stacklies, W., H. Redestig, M. Scholz, D. Walther, and J. Selbig. 2007. “PcaMethods a Bioconductor Package Providing PCA Methods for Incomplete Data.” Bioinformatics 23 (9): 1164–67. https://doi.org/10.1093/bioinformatics/btm069.

Wang, Xu, Daljit Singh, Sandeep Marla, Geoffrey Morris, and Jesse Poland. 2018. “Field-Based High-Throughput Phenotyping of Plant Height in Sorghum Using Different Sensing Technologies.” Plant Methods 14 (1): 53. https://doi.org/10.1186/s13007-018-0324-5.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the

Tidyverse.” Journal of Open Source Software 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Figures

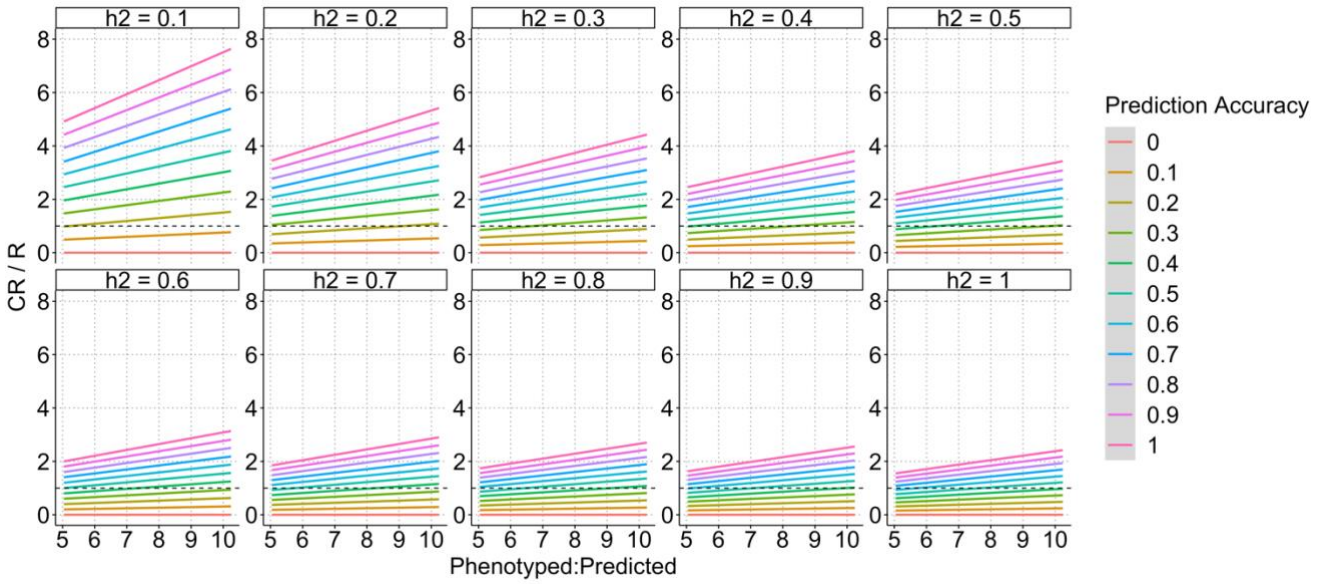


Figure 4.1. Simulation results comparing phenotypic selection as a direct trait to selection on the genotype as a secondary trait.

The narrow-sense heritability of the direct trait is in the panel title and the y-axis is the Correlated Response / Response of selection for the direct trait. When this ratio is above 1, represented by the dotted line, selection on the secondary trait is favored. The x-axis is the number of experimental lines that can be genotyped and predicted for every line that is phenotyped based on estimated costs. The trend of each prediction accuracy is given by the color of the line.

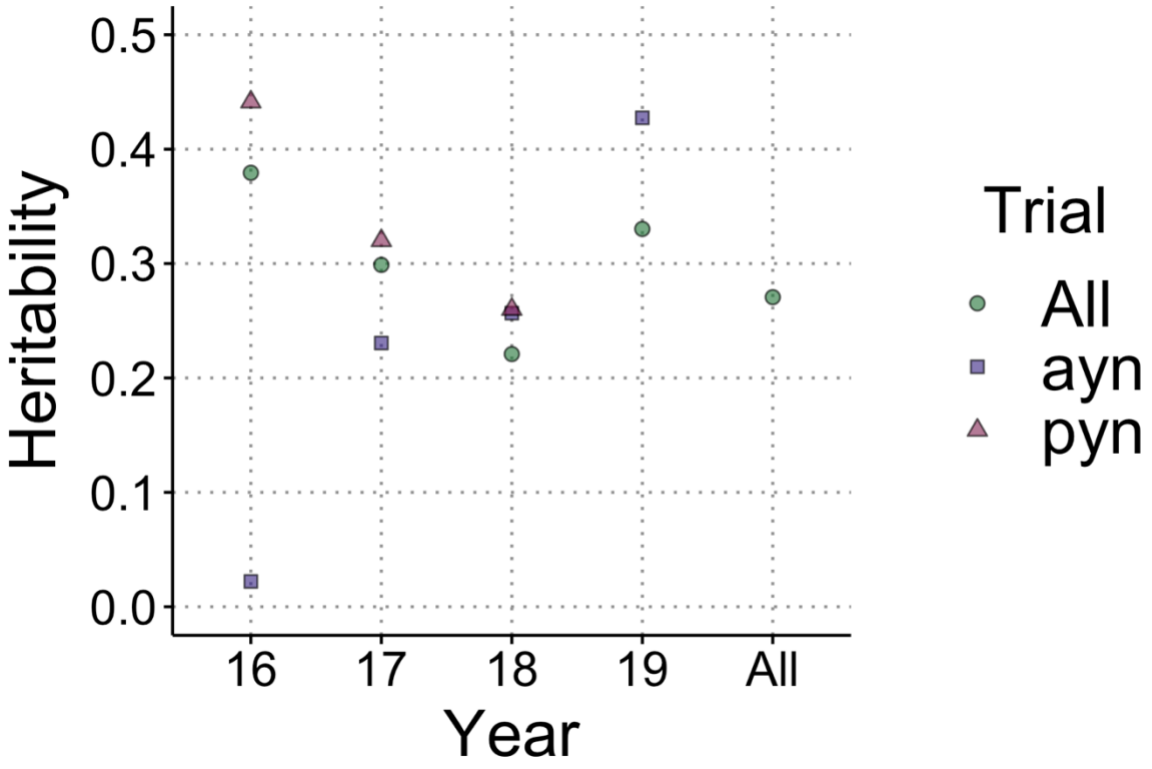


Figure 4.2. Broad sense heritabilities for grain yield across 4 years of phenotypic data. The broad sense heritabilities are presented by season on the x-axis with each trial type denoted by the character shape. There is no PYN for the 2019 season as it was only planted in one location with one replication.

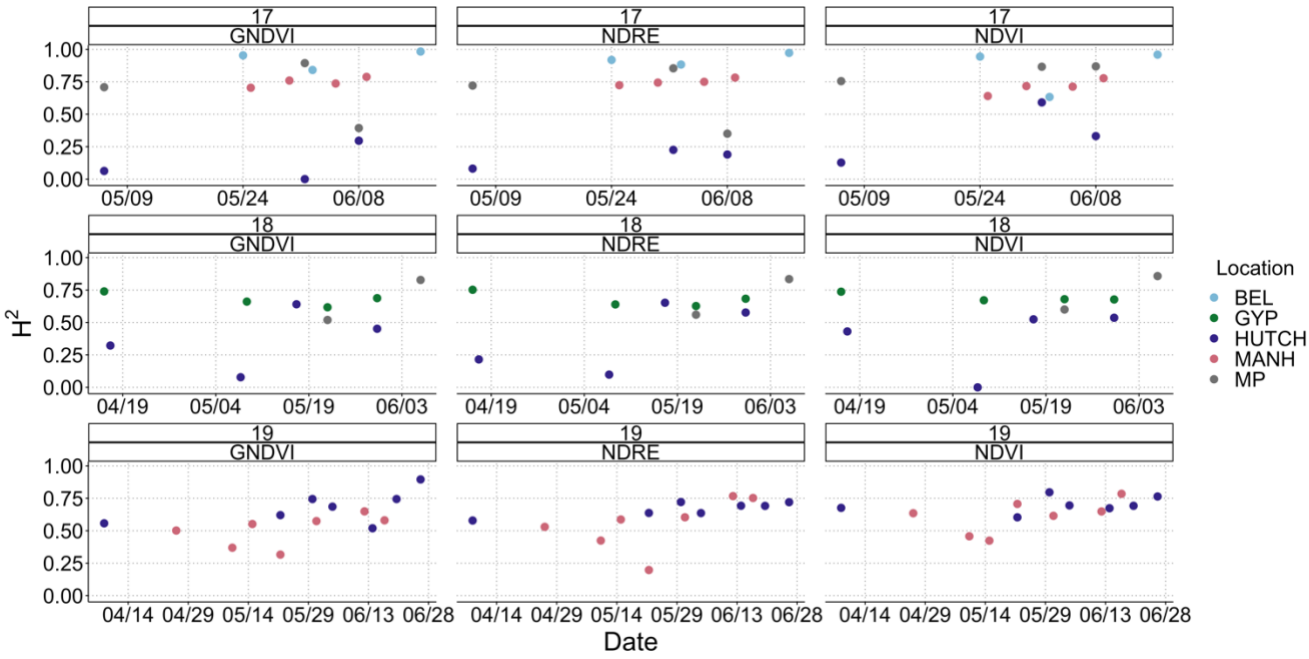


Figure 4.3. The broad sense heritabilities for the VIs for the AYN trials planted at each location.

The title of each plot gives the season-VI combination, with the date on which the VI was taken given by the x-axis, the y-axis is the broad sense heritability, while the color of the point determines the location. NDVI = Normalized Difference Vegetation Index, GNDVI = Green Normalized Difference Vegetation Index, and NDRE = Red Edge Normalized Difference Vegetation Index.

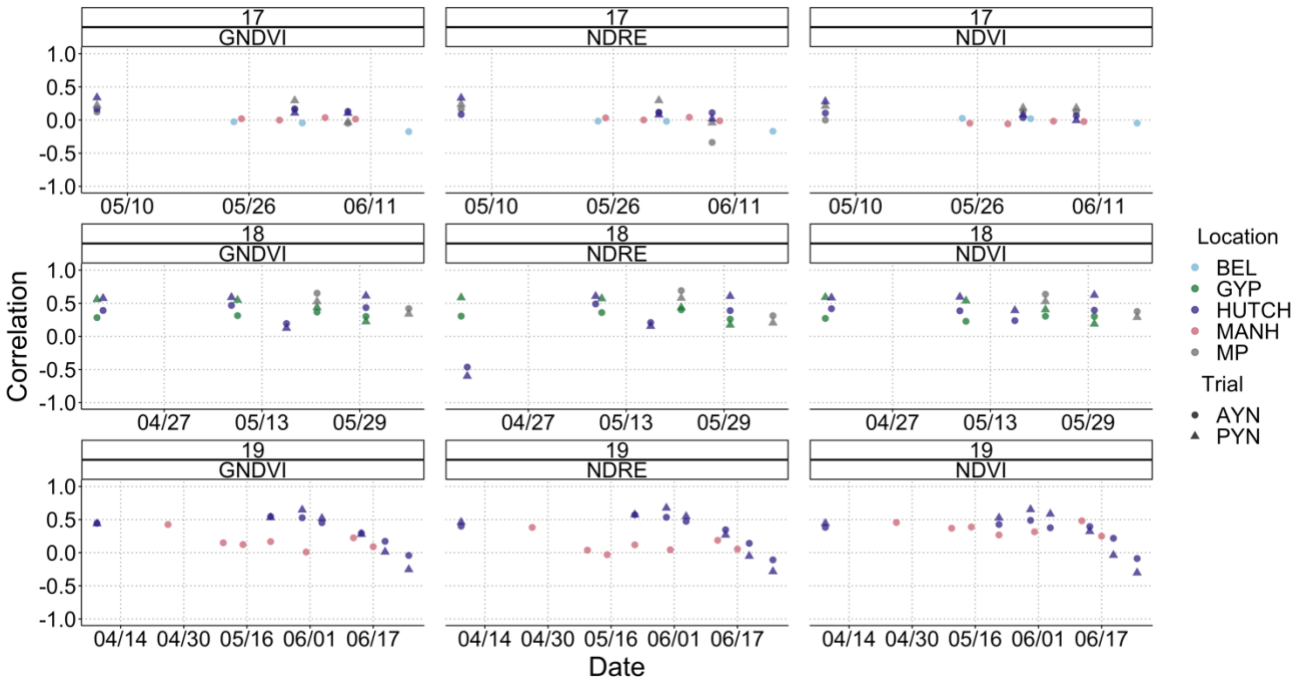


Figure 4.4. The correlations between grain yield and VI by year-trial combinations. The title of each plot gives the season-VI combination, with the date on which the VI was taken given by the x-axis, the y-axis is the Pearson correlation coefficient, while the color of the point determines the location. NDVI = Normalized Difference Vegetation Index, GNDVI = Green Normalized Difference Vegetation Index, and NDRE = Red Edge Normalized Difference Vegetation Index.

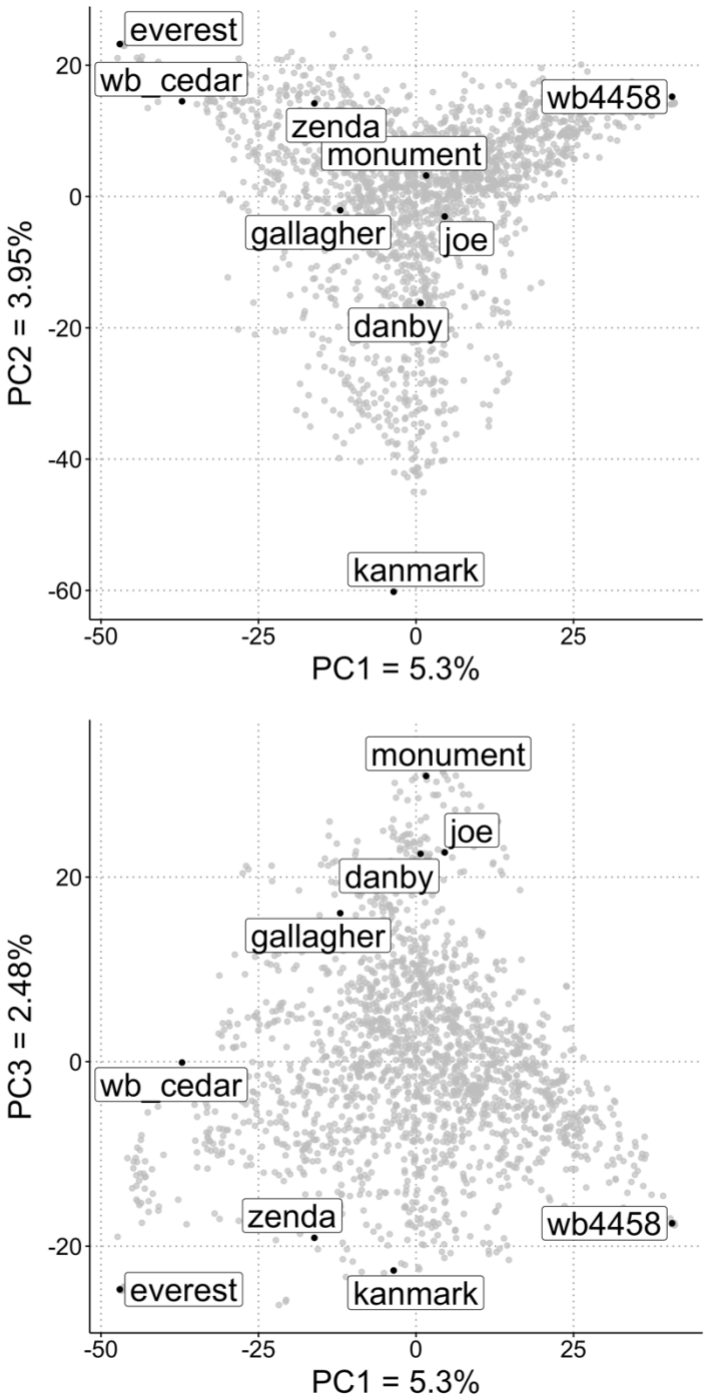


Figure 4.5. Population structure based on genotypic information in the Kansas State winter wheat breeding program nurseries between the 2016-2019 seasons.

The top panel shows the first two principal components (PC), while the bottom panel displays the 2nd and 3rd PC. The variance contributed by each PC is given next to the PC name. Several "founder" lines are highlighted.

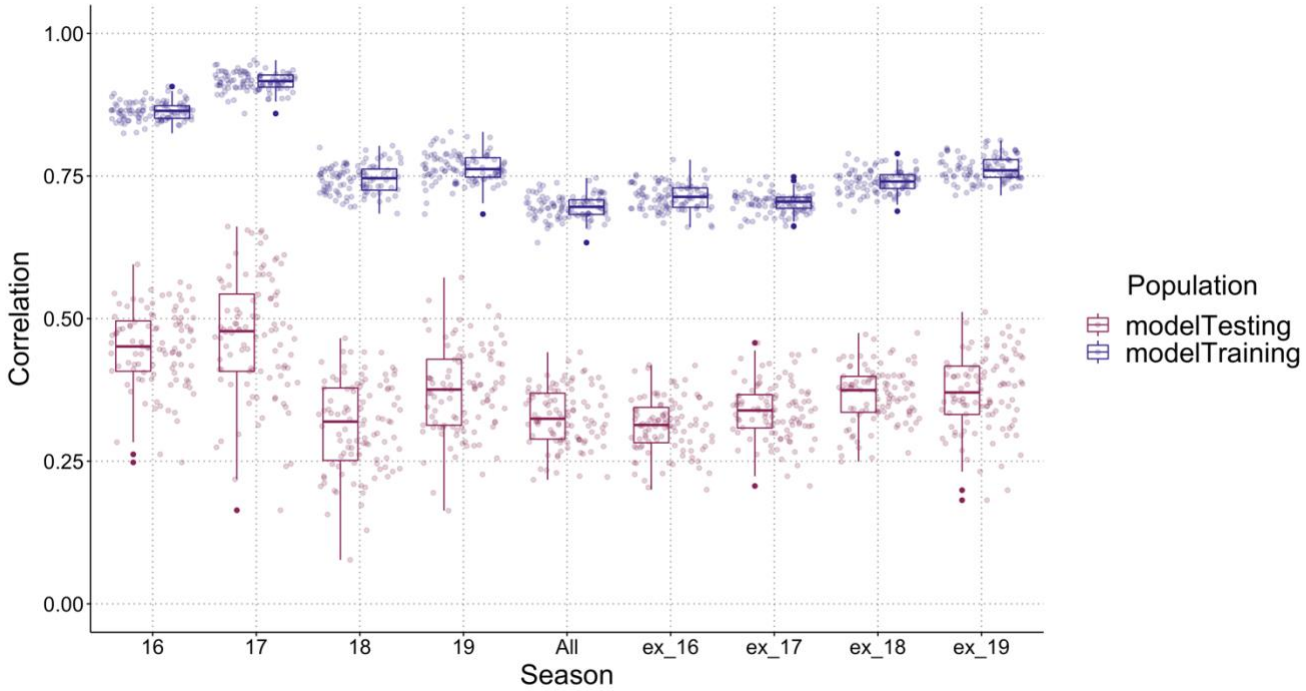


Figure 4.6. Genomic prediction accuracies for grain yield determined by cross-validation.

The season or seasons used in each analysis are given on the x-axis. The y-axis shows the range of possible correlations between the predicted phenotype and the observed phenotype. The color of the point determines if the individual line had been in the training or testing population of the analysis. 100 cross-validations were performed in each analysis. The modelTraining results give an indication of the model-fit whereas the modelTesting results give an indication of the predictive ability of the model. The AYN and PYN trials are included in the analysis.

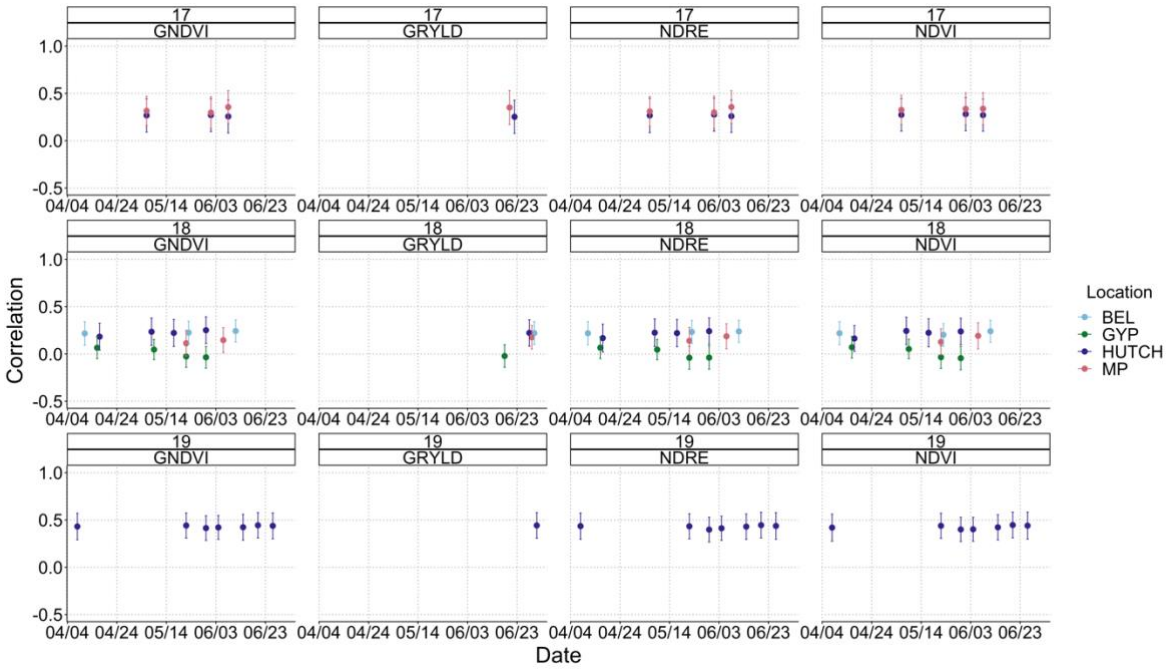


Figure 4.7. Genomic prediction accuracies for PYN grain yield when a VI is used as a covariate determined by cross-validation.

The season and VI are given in the strip title and the points are colored by the location. The GRYLD measurements are those for the prediction of grain yield without a covariate. The large point is the mean and the error bars give the standard deviation. NDVI = Normalized Difference Vegetation Index, GNDVI = Green Normalized Difference Vegetation Index, and NDRE = Red Edge Normalized Difference Vegetation Index.

Tables

Table 4.1. Summary of trials planted between the 2016 and 2019 seasons across locations. A summary is given by season, trial and location with the number of individual plots, experimental lines, date of planting, date of harvest, the mean grain yield in t/ha and the standard deviation of the grain yield.

Season	Trial	Location	Plots	Lines	Planted	Harvested	Mean	SD
2015/2016	AYN	MP	40	75			5.050	0.9
		MANH	240	75			3.206	0.651
		HUTCH	320	109			4.139	0.92
		GYP	40	109			5.739	0.726
	PYN	MP	504	439			5.113	1.21
		MANH	504	379			3.720	0.657
		HUTCH	594	516			3.770	0.886
		GYP	593	506			5.421	0.702
2016/2017	AYN	MP	315	91	10/11/16	20/06/17	5.933	1.00
		MANH	320	91	10/18/16	22/06/17	5.822	0.768
		HUTCH	168	97	10/12/16	21/06/17	6.242	0.845
		BEL	48	36	10/19/16	28/06/17	5.110	0.776
	PYN	MP	126	120	10/11/16	20/06/17	4.805	1.015
		HUTCH	108	102	10/12/16	21/06/17	6.319	0.925
		BEL	8	8	10/19/16	28/06/17	4.5	0.385
2017/2018	AYN	MP	199	90	10/19/17	29/06/18	3.203	0.412
		MANH	200	90	10/20/17	23/06/18	2.644	0.402
		HUTCH	280	125	10/18/17	28/06/18	3.345	0.489
		BEL	280	125	10/17/17	30/06/18	2.561	0.430
		GYP	279	125	10/18/17	18/06/18	2.343	0.318
	PYN	MP	323	282	10/19/17	29/06/18	3.048	0.490

		MANH	414	366	10/20/17	23/06/18	2.880	0.431
		HUTCH	414	366	10/18/17	28/06/18	3.236	0435
		BEL	287	251	10/17/17	30/06/18	2.198	0.412
		GYP	286	250	10/18/17	18/06/18	1.973	0.278
2018/2019	AYN	MANH	80	40	11/01/18		5.798	0.535
		HUTCH	212	99	10/24/18	01/07/19	4.925	0.563
		BEL	130	63	10/24/18	17/07/19	4.965	0.897
	PYN	HUTCH	504	454	10/24/18	01/07/19	4.623	0.794

Table 4.2. Summary of vegetation indices used in the KSU breeding program across 5 years. For each index the formula and reference are provided.

Abbr.	Index	Equation	Reference
GNDVI	Green Normalized Difference Vegetation Index	$GNDVI = \frac{NIR - Green}{NIR + Green}$	(Blackmer et al. 1996)
NDRE	Red Edge Normalized Difference Vegetation Index	$NDRE = \frac{NIR - RedEdge}{NIR + RedEdge}$	(Elliott and Regan 1993)
NDVI	Normalized Difference Vegetation Index	$NDVI = \frac{NIR - Red}{NIR + Red}$	(Curran et al. 1983)

Table 4.3. Genomic Prediction Accuracies as the correlation between BLUPs and GEBVs. The Pearson correlation coefficients is given with a 95% CI and the rank correlation between the GEBVs and the BLUPs.

Training Population	Testing Population	Correlation	95% CI	Rank Correlation
Excluding 2015/2016	2015/2016	0.09	[-0.004, 0.182]	0.093
Excluding 2015/2016 PYN	2015/2016 PYN	0.078	[-0.026, 0.180]	0.093
Excluding 2015/2016 MP MANH GYP	2015/2016 MP MANH GYP	0.024	[-0.08, 0.127]	0.019
Excluding 2016/2017	2016/2017	-0.164	[-0.336, 0.018]	-0.125
Excluding 2016/2017 PYN	2016/2017 PYN	-0.055	[-0.249, 0.144]	0.010
Excluding 2016/2017 MP MANH BEL	2016/2017 MP MANH BEL	-0.009	[-0.206, 0.189]	0.059
Excluding 2017/2018	2017/2018	0.048	[-0.049, 0.145]	0.030
Excluding 2017/2018 PYN	2017/2018 PYN	0.061	[-0.046, 0.165]	0.019
Excluding 2017/2018 MP MANH BEL	2017/2018 MP MANH BEL	0.009	[-0.097, 0.114]	-0.026
Excluding 2018/2019	2018/2019	-0.091	[-0.186, 0.005]	-0.13
Excluding 2018/2019 PYN	2018/2019 PYN	-0.039	[-0.135, 0.058]	-0.079
Excluding 2018/2019 MANH BEL	2018/2019 MANH BEL	0.572	[0.502, 0.634]	0.537

References

Álvarez, María F., Teresa Mosquera, and Matthew W. Blair. 2014. “The Use of Association Genetics Approaches in Plant Breeding.” In *Plant Breeding Reviews: Volume 38*, 17–68. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118916865.ch02.

Ammiraju, J. S. S., B. B. Dholakia, D. K. Santra, H. Singh, M. D. Lagu, S. A. Tamhankar, H. S. Dhaliwal, V. S. Rao, V. S. Gupta, and P. K. Ranjekar. 2001. “Identification of Inter Simple Sequence Repeat (ISSR) Markers Associated with Seed Size in Wheat:” *Theoretical and Applied Genetics* 102 (5): 726–32. https://doi.org/10.1007/s001220051703.

Anand, Shivani. 2019. “India Smartphone Market Registers Highest Second-Quarter Shipments in 2Q19, IDC India Reports.” IDC: The Premier Global Market Intelligence Company. August 13, 2019. https://www.idc.com/getdoc.jsp?containerId=prAP45442419.

Andrade-Sanchez, Pedro, Michael A. Gore, John T. Heun, Kelly R. Thorp, A. Elizabete Carmo-Silva, Andrew N. French, Michael E. Salvucci, and Jeffrey W. White. 2014. “Development and Evaluation of a Field-Based High-Throughput Phenotyping Platform.” *Functional Plant Biology* 41 (1): 68. https://doi.org/10.1071/FP13126.

Araus, José Luis, and Jill E. Cairns. 2014. “Field High-Throughput Phenotyping: The New Crop Breeding Frontier.” *Trends in Plant Science* 19 (1): 52–61. https://doi.org/10.1016/j.tplants.2013.09.008.

Bajarin, Ben. 2014. “Why India Will Be the World’s Second Biggest Smartphone Market.” Fortune. December 2, 2014. https://fortune.com/2014/12/02/why-india-will-be-the-worlds-second-biggest-smartphone-market/.

Barker, Jared, Naiqian Zhang, Joshua Sharon, Ryan Steeves, Xu Wang, Yong Wei, and Jesse Poland. 2016. “Development of a Field-Based High-Throughput Mobile Phenotyping Platform.” *Computers and Electronics in Agriculture* 122 (March): 74–85. https://doi.org/10.1016/j.compag.2016.01.017.

Bassi, Filippo M., Alison R. Bentley, Gilles Charmet, Rodomiro Ortiz, and Jose Crossa. 2016. “Breeding Schemes for the Implementation of Genomic Selection in Wheat (Triticum Spp .).” *Plant Science* 242 (January): 23–36. https://doi.org/10.1016/j.plantsci.2015.08.021.

Bean, S. R., O. K. Chung, M. R. Tuinstra, J. F. Pedersen, and J. Erpelding. 2006. “Evaluation of the Single Kernel Characterization System (SKCS) for Measurement of Sorghum Grain Attributes.” *Cereal Chemistry Journal* 83 (1): 108–13. https://doi.org/10.1094/CC-83-0108.

Beavis, William D. 1998. “QTL Analyses: Power, Precision, and Accuracy.” *Molecular Dissection of Complex Traits*, 145–62.

Bernardo, Rex. 2008. “Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years.” *Crop Science* 48 (5): 1649. https://doi.org/10.2135/cropsci2008.03.0131.

Blackmer, Tracy M., James S. Schepers, Gary E. Varvel, and Elizabeth A. Walter-Shea. 1996. “Nitrogen Deficiency Detection Using Reflected Shortwave Radiation from Irrigated Corn Canopies.” *Agronomy Journal* 88 (1): 1–5. https://doi.org/10.2134/agronj1996.00021962008800010001x.

Blake, Victoria C., Clay Birkett, David E. Matthews, David L. Hane, Peter Bradbury, and Jean-Luc Jannink. 2016. “The Triticeae Toolbox: Combining Phenotype and Genotype Data to Advance Small-Grains Breeding.” *The Plant Genome* 9 (2): 0. https://doi.org/10.3835/plantgenome2014.12.0099.

Bond, Jennifer K. 2020. “USDA ERS - Wheat Sector at a Glance.” June 26, 2020. https://www.ers.usda.gov/topics/crops/wheat/wheat-sector-at-a-glance/.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. “TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples.” *Bioinformatics* 23 (19): 2633–35. https://doi.org/10.1093/bioinformatics/btm308.

Bradski, Gary. 2000. *The OpenCV Library*. https://www.drdobbs.com/open-source/the-opencv-library/184404319.

Breseghello, Flavio, and Mark E. Sorrells. 2006. “Association Mapping of Kernel Size and Milling Quality in Wheat (Triticum Aestivum L.) Cultivars.” *Genetics* 172 (2): 1165–77. https://doi.org/10.1534/genetics.105.044586.

Breseghello, Flavio, and Mark E. Sorrells. 2007. “QTL Analysis of Kernel Size and Shape in Two Hexaploid Wheat Mapping Populations.” *Field Crops Research* 101 (2): 172–79. https://doi.org/10.1016/j.fcr.2006.11.008.

Broman, Karl W., Hao Wu, Saunak Sen, and Gary A. Churchill. 2003. “R/Qtl: QTL Mapping in Experimental Crosses.” *Bioinformatics (Oxford, England)* 19 (7): 889–90. https://doi.org/10.1093/bioinformatics/btg112.

Browning, Brian L., Ying Zhou, and Sharon R. Browning. 2018. “A One-Penny Imputed Genome from Next-Generation Reference Panels.” *The American Journal of Human Genetics* 103 (3): 338–48. https://doi.org/10.1016/j.ajhg.2018.07.015.

Browning, Brian L., and Sharon R. Browning. 2016. “Genotype Imputation with Millions of Reference Samples.” *The American Journal of Human Genetics* 98 (1): 116–26. https://doi.org/10.1016/j.ajhg.2015.11.020.

Burgueño, Juan, Gustavo de los Campos, Kent Weigel, and José Crossa. 2012. “Genomic Prediction of Breeding Values When Modeling Genotype × Environment Interaction Using Pedigree and Dense Molecular Markers.” *Crop Science* 52 (2): 707–19. https://doi.org/10.2135/cropsci2011.06.0299.

Busemeyer, Lucas, Daniel Mentrup, Kim Möller, Erik Wunder, Katharina Alheit, Volker Hahn, Hans Peter Maurer, et al. 2013. “BreedVision — A Multi-Sensor Platform for Non-Destructive Field-Based Phenotyping in Plant Breeding.” *Sensors* 13 (3): 2830–47. https://doi.org/10.3390/s130302830.

Butler, David. 2020. *Asreml: Fits the Linear Mixed Model*. www.vsni.co.uk.

Cabrera-Bosquet, Llorenç, José Crossa, Jarislav von Zitzewitz, María Dolors Serret, and José Luis Araus. 2012. “High-Throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding Converge: High-Throughput Phenotyping and Genomic Selection.” *Journal of Integrative Plant Biology* 54 (5): 312–20. https://doi.org/10.1111/j.1744-7909.2012.01116.x.

Campbell, Kim Garland, Christine J. Bergman, Daisy G. Gualberto, James A. Anderson, Micheal J. Giroux, Gary Hareland, R. Gary Fulcher, Mark E. Sorrells, and Patrick L. Finney. 1999. “Quantitative Trait Loci Associated with Kernel Traits in a Soft × Hard Wheat Cross.” *Crop Science* 39 (4): 1184. https://doi.org/10.2135/cropsci1999.0011183X003900040039x.

Chaerle, L., and D. Van Der Straeten. 2001. “Seeing Is Believing: Imaging Techniques to Monitor Plant Health.” *Biochimica Et Biophysica Acta* 1519 (3): 153–66. https://doi.org/10.1016/s0167-4781(01)00238-x.

Core Team, R. 2020. *R: A Language and Environment for Statistical Computing* (version 4.0.0). Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Crain, Jared L., Yong Wei, Jared Barker, Sean M. Thompson, Phillip D. Alderman, Matthew Reynolds, Naiqian Zhang, and Jesse Poland. 2016. “Development and Deployment of a Portable Field Phenotyping Platform.” *Crop Science* 56 (3): 965–75. https://doi.org/10.2135/cropsci2015.05.0290.

Crain, Jared, Suchismita Mondal, Jessica Rutkoski, Ravi P. Singh, and Jesse Poland. 2018. “Combining High-Throughput Phenotyping and Genomic Information to Increase Prediction and Selection Accuracy in Wheat Breeding.” *The Plant Genome* 11 (1): 0. https://doi.org/10.3835/plantgenome2017.05.0043.

Curran, P. J., M. D. Steven, John Theodore Houghton, Arthur Herbert Cook, and Henry Charnock. 1983. “Multispectral Remote Sensing for the Estimation of Green Leaf Area Index.” *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 309 (1508): 257–70. https://doi.org/10.1098/rsta.1983.0039.

Dawson, Julie C., Jeffrey B. Endelman, Nicolas Heslot, Jose Crossa, Jesse Poland, Susanne Dreisigacker, Yann Manès, Mark E. Sorrells, and Jean-Luc Jannink. 2013. “The Use of Unbalanced Historical Data for Genomic Selection in an International Wheat Breeding Program.” *Field Crops Research* 154 (December): 12–22. https://doi.org/10.1016/j.fcr.2013.07.020.

Dekkers, Jack C. M., and Frédéric Hospital. 2002. “The Use of Molecular Genetics in the Improvement of Agricultural Populations.” *Nature Reviews Genetics* 3 (1): 22–32. https://doi.org/10.1038/nrg701.

Dholakia, B. B., J. S. S. Ammiraju, H. Singh, M. D. Lagu, M. S. Roder, V. S. Rao, H. S. Dhaliwal, P. K. Ranjekar, and V. S. Gupta. 2003. “Molecular Marker Analysis of Kernel Size and Shape in Bread Wheat.” *Plant Breeding* 122 (5): 392–95. https://doi.org/10.1046/j.1439-0523.2003.00896.x.

Dhondt, Stijn, Nathalie Wuyts, and Dirk Inzé. 2013. “Cell to Whole-Plant Phenotyping: The Best Is yet to Come.” *Trends in Plant Science* 18 (8): 428–39. https://doi.org/10.1016/j.tplants.2013.04.008.

Dunckel, Sandra M. 2015. “Whole Genome Approaches for Characterizing and Utilizing Synthetic Wheat.” https://krex.k-state.edu/dspace/handle/2097/20518.

Eathington, Sam R., Theodore M. Crosbie, Marlin D. Edwards, Robert S. Reiter, and Jason K. Bull. 2007. “Molecular Markers in a Commercial Breeding Program.” *Crop Science* 47 (S3): S-154-S-163. https://doi.org/10.2135/cropsci2007.04.0015IPBS.

Elliott, G. A., and K. L. Regan. 1993. “Use of Reflectance Measurements to Estimate Early Cereal Biomass Production on Sandplain Soils.” *Australian Journal of Experimental Agriculture* 33 (2): 179–83. https://doi.org/10.1071/ea9930179.

Elshire, Robert J., Jeffrey C. Glaubitz, Qi Sun, Jesse A. Poland, Ken Kawamoto, Edward S. Buckler, and Sharon E. Mitchell. 2011. “A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species.” Edited by Laszlo Orban. *PLoS ONE* 6 (5): e19379. https://doi.org/10.1371/journal.pone.0019379.

Endelman, Jeffrey B. 2011. “Ridge Regression and Other Kernels for Genomic Selection with R Package RrBLUP.” *The Plant Genome Journal* 4 (3): 250. https://doi.org/10.3835/plantgenome2011.08.0024.

Endelman, Jeffrey B., and Jean-Luc Jannink. 2012. “Shrinkage Estimation of the Realized Relationship Matrix.” *G3: Genes|Genomes|Genetics* 2 (11): 1405–13. https://doi.org/10.1534/g3.112.004259.

Falconer, Douglas S., and Trudy F. C. Mackay. 2009. *Introduction to Quantitative Genetics*. 4. ed., [16. print.]. Harlow: Pearson, Prentice Hall.

Federer, W. T., and D. Raghavarao. 1975. “On Augmented Designs.” *Biometrics* 31 (1): 29–35. https://doi.org/10.2307/2529707.

Furbank, Robert T., and Mark Tester. 2011. “Phenomics – Technologies to Relieve the Phenotyping Bottleneck.” *Trends in Plant Science* 16 (12): 635–44. https://doi.org/10.1016/j.tplants.2011.09.005.

Georges, Michel, Carole Charlier, and Ben Hayes. 2019. “Harnessing Genomic Information for Livestock Improvement.” *Nature Reviews Genetics* 20 (3): 135–56. https://doi.org/10.1038/s41576-018-0082-2.

Gitelson, Anatoly A., Yoram J. Kaufman, and Mark N. Merzlyak. 1996. “Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS.” *Remote Sensing of Environment* 58 (3): 289–98. https://doi.org/10.1016/S0034-4257(96)00072-7.

Giura, A., and N. N. Saulescu. 1996. “Chromosomal Location of Genes Controlling Grain Size in a Large Grained Selection of Wheat (Triticum Aestivum L.).” *Euphytica* 89 (1): 77–80. https://doi.org/10.1007/BF00015722.

Glaubitz, Jeffrey C., Terry M. Casstevens, Fei Lu, James Harriman, Robert J. Elshire, Qi Sun, and Edward S. Buckler. 2014. “TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline.” Edited by Nicholas A. Tinker. *PLoS ONE* 9 (2): e90346. https://doi.org/10.1371/journal.pone.0090346.

Haghighattalab, Atena, Lorena González Pérez, Suchismita Mondal, Daljit Singh, Dale Schinstock, Jessica Rutkoski, Ivan Ortiz-Monasterio, Ravi Prakash Singh, Douglas Goodin, and Jesse Poland. 2016. “Application of Unmanned Aerial Systems for High Throughput Phenotyping of Large Wheat Breeding Nurseries.” *Plant Methods* 12 (1): 35. https://doi.org/10.1186/s13007-016-0134-6.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. “Invited Review: Genomic Selection in Dairy Cattle: Progress and Challenges.” *Journal of Dairy Science* 92 (2): 433–43. https://doi.org/10.3168/jds.2008-1646.

Heslot, Nicolas, Hsiao-Pei Yang, Mark E. Sorrells, and Jean-Luc Jannink. 2012. “Genomic Selection in Plant Breeding: A Comparison of Models.” *Crop Science* 52 (1): 146–60. https://doi.org/10.2135/cropsci2011.06.0297.

Holland, James B., Wyman E. Nyquist, and Cuauhtemoc T. Cervantes-Martínez. 2010. “Estimating and Interpreting Heritability for Plant Breeding: An Update.” In *Plant Breeding Reviews*, 9–112. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470650202.ch2.

Houle, David, Diddahally R. Govindaraju, and Stig Omholt. 2010. “Phenomics: The next Challenge.” *Nature Reviews Genetics* 11 (12): 855–66. https://doi.org/10.1038/nrg2897.

Huang, Ning, Arnold Parco, Teresita Mew, Gerard Magpantay, Susan McCouch, Emmanuel Guiderdoni, Jichen Xu, Prasanta Subudhi, Enrique R. Angeles, and Gurdev S. Khush. 1997. “RFLP Mapping of Isozymes, RAPD and QTLs for Grain Shape, Brown Planthopper Resistance in a Doubled Haploid Rice Population.” *Molecular Breeding* 3 (2): 105–13. https://doi.org/10.1023/A:1009683603862.

International Wheat Genome Sequencing Consortium, Michael Alaux, Jane Rogers, Thomas Letellier, Raphaël Flores, Françoise Alfama, Cyril Pommier, et al. 2018. “Linking the International Wheat Genome Sequencing Consortium Bread Wheat Reference Genome Sequence to Wheat Genetic and Phenomic Data.” *Genome Biology* 19 (1): 111. https://doi.org/10.1186/s13059-018-1491-4.

Jackson, Christopher. 2011. “Multi-State Models for Panel Data: The Msm Package for R.” *Journal of Statistical Software* 38 (1): 1–28. https://doi.org/10.18637/jss.v038.i08.

Jarquín, Diego, Cristiano Lemes da Silva, R. Chris Gaynor, Jesse Poland, Allan Fritz, Reka Howard, Sarah Battenfield, and Jose Crossa. 2017. “Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype × Environment Interactions in Kansas Wheat.” *The Plant Genome* 10 (2): plantgenome2016.12.0130. https://doi.org/10.3835/plantgenome2016.12.0130.

Kang, Hyun Min, Noah A. Zaitlen, Claire M. Wade, Andrew Kirby, David Heckerman, Mark J. Daly, and Eleazar Eskin. 2008. “Efficient Control of Population Structure in Model Organism Association Mapping.” *Genetics* 178 (3): 1709–23. https://doi.org/10.1534/genetics.107.080101.

Korkmaz, Selçuk, Dincer Goksuluk, and Gokmen Zararsiz. 2014. “MVN: An R Package for Assessing Multivariate Normality.” https://doi.org/10.32614/RJ-2014-031.

Lenaerts, Bert, Bertrand C.Y. Collard, and Matty Demont. 2019. “Review: Improving Global Food Security through Accelerated Plant Breeding.” *Plant Science* 287 (October): 110207. https://doi.org/10.1016/j.plantsci.2019.110207.

Li, Jiming, Jinhua Xiao, Silvana Grandillo, Longying Jiang, Yizhen Wan, Qiyun Deng, Longping Yuan, and Susan R. McCouch. 2004. “QTL Detection for Rice Grain Quality Traits Using an Interspecific Backcross Population Derived from Cultivated Asian (O. Sativa L.) and African (O. Glaberrima S.) Rice.” *Genome* 47 (4): 697–704. https://doi.org/10.1139/g04-029.

Liu, Ying, Liwei Wang, Chuanlong Sun, Zuxin Zhang, Yonglian Zheng, and Fazhan Qiu. 2014. “Genetic Analysis and Major QTL Detection for Maize Kernel Size and Weight in Multi-Environments.” *Theoretical and Applied Genetics* 127 (5): 1019–37. https://doi.org/10.1007/s00122-014-2276-0.

Marshall, D. R., D. J. Mares, H. J. Moss, and F. W. Ellison. 1986. “Effects of Grain Shape and Size on Milling Yields in Wheat. II. Experimental Studies.” *Australian Journal of Agricultural Research* 37 (4): 331–42. https://doi.org/10.1071/ar9860331.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. “Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps.” *Genetics* 157 (4): 1819–29.

Moles, A. T. 2005. “A Brief History of Seed Size.” *Science* 307 (5709): 576–80. https://doi.org/10.1126/science.1104863.

Montesinos-López, Osval A., Abelardo Montesinos-López, José Crossa, Fernando H. Toledo, Oscar Pérez-Hernández, Kent M. Eskridge, and Jessica Rutkoski. 2016. “A Genomic Bayesian Multi-Trait and Multi-Environment Model.” *G3: Genes|Genomes|Genetics* 6 (9): 2725–44. https://doi.org/10.1534/g3.116.032359.

Moose, Stephen P., and Rita H. Mumm. 2008. “Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement.” *Plant Physiology* 147 (3): 969–77. https://doi.org/10.1104/pp.108.118232.

Neilsen, Mitchell, Shravan D Gangadhara, and Trevor W Rife. 2016. “Extending Watershed Segmentation Algorithms for High-Throughput Phenotyping.” In . Denver USA. https://www.researchgate.net/publication/322369315_Extending_Watershed_Segmentation_Algorithms_for_High-Throughput_Phenotyping.

Ortiz, Rodomiro, Richard Trethowan, Guillermo Ortiz Ferrara, Masa Iwanaga, John H. Dodds, Jonathan H. Crouch, Jose Crossa, and Hans-Joachim Braun. 2007. “High Yield Potential, Shuttle Breeding, Genetic Diversity, and a New International Wheat Improvement Strategy.” *Euphytica* 157 (3): 365–84. https://doi.org/10.1007/s10681-007-9375-9.

Pask, A., J. Pietragalla, D. Mullan, and M. P. Reynolds. 2012. “Physiological Breeding II: A Field Guide to Wheat Phenotyping.” https://repository.cimmyt.org/handle/10883/1288.

Patterson, H. D., and E. R. Williams. 1976. “A New Class of Resolvable Incomplete Block Designs.” *Biometrika* 63 (1): 83–92. https://doi.org/10.2307/2335087.

Peng, Jinrong, Donald E. Richards, Nigel M. Hartley, George P. Murphy, Katrien M. Devos, John E. Flintham, James Beales, et al. 1999. “‘Green Revolution’ Genes Encode Mutant Gibberellin Response Modulators.” *Nature* 400 (6741): 256–61. https://doi.org/10.1038/22307.

Pérez, Paulino, and Gustavo de los Campos. 2014. “Genome-Wide Regression and Prediction with the BGLR Statistical Package.” *Genetics* 198 (2): 483–95. https://doi.org/10.1534/genetics.114.164442.

Pérez-Rodríguez, Paulino, José Crossa, Jessica Rutkoski, Jesse Poland, Ravi Singh, Andrés Legarra, Enrique Autrique, Gustavo de los Campos, Juan Burgueño, and Susanne Dreisigacker. 2017. “Single-Step Genomic and Pedigree Genotype × Environment Interaction Models for Predicting Wheat Lines in International Environments.” *The Plant Genome* 10 (2): 0. https://doi.org/10.3835/plantgenome2016.09.0089.

Poland, Jesse, Jeffrey Endelman, Julie Dawson, Jessica Rutkoski, Shuangye Wu, Yann Manes, Susanne Dreisigacker, et al. 2012. “Genomic Selection in Wheat Breeding Using Genotyping-by-Sequencing.” *The Plant Genome Journal* 5 (3): 103. https://doi.org/10.3835/plantgenome2012.06.0006.

Ramya, P., A. Chaubal, K. Kulkarni, L. Gupta, N. Kadoo, H. S. Dhaliwal, P. Chhuneja, M. Lagu, and V. Gupt. 2010. “QTL Mapping of 1000-Kernel Weight, Kernel Length, and Kernel Width in Bread Wheat (Triticum Aestivum L.).” *Journal of Applied Genetics* 51 (4): 421–29. https://doi.org/10.1007/BF03208872.

Rasheed, Awais, Xianchun Xia, Francis Ogbonnaya, Tariq Mahmood, Zongwen Zhang, Abdul Mujeeb-Kazi, and Zhonghu He. 2014. “Genome-Wide Association for Grain

Morphology in Synthetic Hexaploid Wheats Using Digital Imaging Analysis." *BMC Plant Biology* 14 (1): 128. https://doi.org/10.1186/1471-2229-14-128.

Ray, Deepak K., Nathaniel D. Mueller, Paul C. West, and Jonathan A. Foley. 2013. "Yield Trends Are Insufficient to Double Global Crop Production by 2050." *PLoS ONE* 8 (6): e66428. https://doi.org/10.1371/journal.pone.0066428.

Revelle, William. 2019. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. https://CRAN.R-project.org/package=psych.

Reynolds, Matthew, Scott Chapman, Leonardo Crespo-Herrera, Gemma Molero, Suchismita Mondal, Diego N. L. Pequeno, Francisco Pinto, et al. 2020. "Breeder Friendly Phenotyping." *Plant Science*, Food Security under Climate Change, 295 (June): 110396. https://doi.org/10.1016/j.plantsci.2019.110396.

Rife, Trevor W., and Jesse A. Poland. 2014. "Field Book: An Open-Source Application for Field Data Collection on Android." *Crop Science* 54 (4): 1624–27. https://doi.org/10.2135/cropsci2013.08.0579.

Rutkoski, Jessica, Jesse Poland, Suchismita Mondal, Enrique Autrique, Lorena González Pérez, José Crossa, Matthew Reynolds, and Ravi Singh. 2016. "Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat." *G3: Genes|Genomes|Genetics* 6 (9): 2799–2808. https://doi.org/10.1534/g3.116.032888.

Sankaran, Sindhuja, Lav R. Khot, and Arron H. Carter. 2015. "Field-Based Crop Phenotyping: Multispectral Aerial Imaging for Evaluation of Winter Wheat Emergence and Spring Stand." *Computers and Electronics in Agriculture* 118 (October): 372–79. https://doi.org/10.1016/j.compag.2015.09.001.

Sankaran, Sindhuja, Lav R. Khot, Carlos Zúñiga Espinoza, Sanaz Jarolmasjed, Vidyasagar R. Sathuvalli, George J. Vandemark, Phillip N. Miklas, et al. 2015. "Low-Altitude, High-Resolution Aerial Imaging Systems for Row and Field Crop Phenotyping: A Review." *European Journal of Agronomy* 70 (October): 112–23. https://doi.org/10.1016/j.eja.2015.07.004.

Schwieters, T, and H Saleem. 2013. "Smartphone Uptake Gaining Pace in Africa as IDC Tips Shipments to Double over next Four Years." October 25, 2013. http://itersnews.com/?p=55430.

Sorrells, Mark E., J. Perry Gustafson, Daryl Somers, Shiaoman Chao, David Benscher, Gina Guedira-Brown, Eric Huttner, et al. 2011. "Reconstruction of the Synthetic W7984 × Opata M85 Wheat Reference Population." Edited by A. Van Deynze. *Genome* 54 (11): 875–82. https://doi.org/10.1139/g11-054.

Stacklies, W., H. Redestig, M. Scholz, D. Walther, and J. Selbig. 2007. "PcaMethods a Bioconductor Package Providing PCA Methods for Incomplete Data." *Bioinformatics* 23 (9): 1164–67. https://doi.org/10.1093/bioinformatics/btm069.

Subramanian, Ram, Edgar P. Spalding, and Nicola J. Ferrier. 2013. “A High Throughput Robot System for Machine Vision Based Plant Phenotype Studies.” *Machine Vision and Applications* 24 (3): 619–36. https://doi.org/10.1007/s00138-012-0434-4.

Tanabata, Takanari, Taeko Shibaya, Kiyosumi Hori, Kaworu Ebana, and Masahiro Yano. 2012. “SmartGrain: High-Throughput Phenotyping Software for Measuring Seed Shape through Image Analysis.” *Plant Physiology* 160 (4): 1871–80. https://doi.org/10.1104/pp.112.205120.

Tester, M., and P. Langridge. 2010. “Breeding Technologies to Increase Crop Production in a Changing World.” *Science* 327 (5967): 818–22. https://doi.org/10.1126/science.1183700.

United Nations, Department of Economic and Social Affairs, and Population Division. 2019. *World Population Prospects 2019: Highlights*. Vol. (ST/ESA/SER.A/423).

Wang, Shichen, Debbie Wong, Kerrie Forrest, Alexandra Allen, Shiaoman Chao, Bevan E. Huang, Marco Maccaferri, et al. 2014. “Characterization of Polyploid Wheat Genomic Diversity Using a High-Density 90 000 Single Nucleotide Polymorphism Array.” *Plant Biotechnology Journal* 12 (6): 787–96. https://doi.org/10.1111/pbi.12183.

Wang, Xu, Daljit Singh, Sandeep Marla, Geoffrey Morris, and Jesse Poland. 2018. “Field-Based High-Throughput Phenotyping of Plant Height in Sorghum Using Different Sensing Technologies.” *Plant Methods* 14 (1): 53. https://doi.org/10.1186/s13007-018-0324-5.

Watson, Amy, Sreya Ghosh, Matthew J. Williams, William S. Cuddy, James Simmonds, María-Dolores Rey, M. Asyraf Md Hatta, et al. 2018. “Speed Breeding Is a Powerful Tool to Accelerate Crop Research and Breeding.” *Nature Plants* 4 (1): 23–29. https://doi.org/10.1038/s41477-017-0083-8.

Whan, Alex P, Alison B Smith, Colin R Cavanagh, Jean-Philippe F Ral, Lindsay M Shaw, Crispin A Howitt, and Leanne Bischof. 2014. “GrainScan: A Low Cost, Fast Method for Grain Size and Colour Measurements.” *Plant Methods* 10 (1): 23. https://doi.org/10.1186/1746-4811-10-23.

White, Jeffrey W., Pedro Andrade-Sanchez, Michael A. Gore, Kevin F. Bronson, Terry A. Coffelt, Matthew M. Conley, Kenneth A. Feldmann, et al. 2012. “Field-Based Phenomics for Plant Genetics Research.” *Field Crops Research* 133 (July): 101–12. https://doi.org/10.1016/j.fcr.2012.04.003.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the Tidyverse.” *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wilhelm, Edward P., Ian J. Mackay, Robert J. Saville, Andrey V. Korolev, Francois Balfourier, Andy J. Greenland, Margaret I. Boulton, and Wayne Powell. 2013. "Haplotype Dictionary for the Rht-1 Loci in Wheat." *Theoretical and Applied Genetics* 126 (7): 1733–47. https://doi.org/10.1007/s00122-013-2088-7.

Williams, Keith, and Mark E. Sorrells. 2014. "Three-Dimensional Seed Size and Shape QTL in Hexaploid Wheat (L.) Populations." *Crop Science* 54 (1): 98. https://doi.org/10.2135/cropsci2012.10.0609.

Xu, Yunbi, and Jonathan H. Crouch. 2008. "Marker-Assisted Selection in Plant Breeding: From Publications to Practice." *Crop Science* 48 (2): 391–407. https://doi.org/10.2135/cropsci2007.04.0191.

Yang, Wanneng, Lingfeng Duan, Guoxing Chen, Lizhong Xiong, and Qian Liu. 2013. "Plant Phenomics and High-Throughput Phenotyping: Accelerating Rice Functional Genomics Using Multidisciplinary Technologies." *Current Opinion in Plant Biology* 16 (2): 180–87. https://doi.org/10.1016/j.pbi.2013.03.005.

Yu, Jianming, Gael Pressoir, William H Briggs, Irie Vroh Bi, Masanori Yamasaki, John F Doebley, Michael D McMullen, et al. 2006. "A Unified Mixed-Model Method for Association Mapping That Accounts for Multiple Levels of Relatedness." *Nature Genetics* 38 (2): 203–8. https://doi.org/10.1038/ng1702.

Zhang, Zhanhui, Zonghua Liu, Yanmin Hu, Weihua Li, Zhiyuan Fu, Dong Ding, Haochuan Li, Mengmeng Qiao, and Jihua Tang. 2014. "QTL Analysis of Kernel-Related Traits in Maize Using an Immortalized F2 Population." Edited by James C. Nelson. *PLoS ONE* 9 (2): e89645. https://doi.org/10.1371/journal.pone.0089645.

Appendix A - Supplementary Material Chapter 2

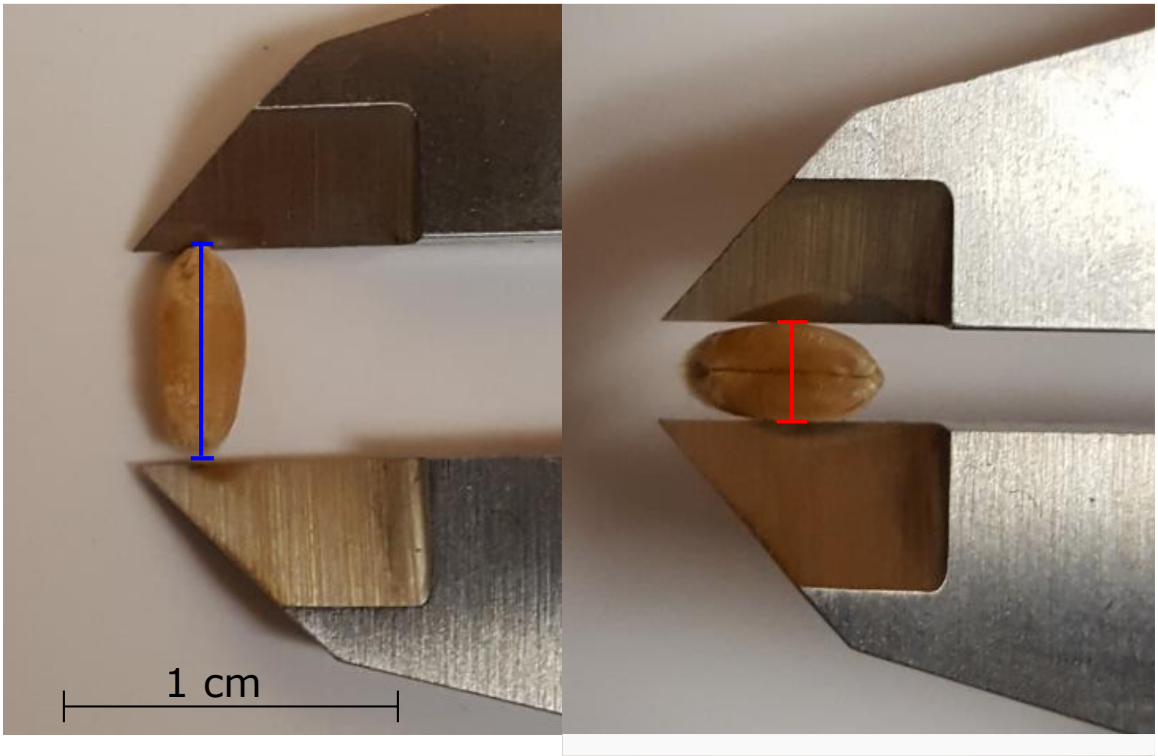


Figure A.1. Parameters measured manually in determining length and width.

The blue line represents the length measurement and the red line represents the width measurement. These measurements are subject to error as the width measured may not be exactly perpendicular to the length.

Table A.1. Average length and width for manual hand measurements and OneKK measurements. All measurements in mm.

Sample	Hand Length	Hand Width	App Length	App Width
14GHS04251S01	6.43777778	3.615	6.26096879	3.76906366
14GHS00164S01	5.77954546	3.10977273	6.2087206	3.65213782
14GHS03953S02	5.51125	2.15625	5.26645648	2.30834333
14GHS03054S01	5.89892857	3.19107143	5.85752571	3.23742261
14GHS04251S02	5.65428571	3.19306122	5.69855144	3.31528373
14GHS03540S02	5.90727273	2.94045455	5.87227804	3.18275993
14GHS03324S01	5.72155172	2.88206897	6.05950642	3.33717766
14GHS02903S01	5.44268293	2.42634146	5.81081639	2.88826545
14GHS02180S01	5.58595238	3.17	5.99118973	3.64690862
14GHS02903S02	5.053125	2.093125	5.28768382	2.48833034
14GHS02458S01	5.38529412	2.38294118	5.36991206	2.63306162
14GHS03577S02	4.97891892	1.96297297	5.27129495	2.41278834
14GHS03623S02	5.47454546	2.27454546	5.59096664	2.50760646
14GHS04164S02	5.06566038	2.08377359	5.54534093	2.67624029
14GHS03567S02	5.44822222	2.42444444	5.33786195	2.65282837
14GHS00902S01	6.35771429	3.53142857	6.42374339	3.60678977
14GHS02737S02	5.03894737	2.27236842	5.01367965	2.37154742
14GHS03818S02	5.34333333	2.2025	5.38451237	2.54281646
14GHS01966S01	4.80034483	2.49965517	4.73041035	2.41817426
14GHS04279S01	5.8984	2.6608	5.76476395	2.64990789
14GHS04279S02	5.64632653	2.33693878	5.92539745	2.58315039
14GHS02033S01	5.34441177	2.90058824	5.7327286	3.17623443
14GHS01778S01	6.06358974	2.70076923	6.30506918	2.89254194
14GHS05159S01	5.0192	2.9612	4.93484719	2.87272517
14GHS03359S02	5.764	3.01	5.98562422	3.25329523
14GHS04483S02	5.94733333	3.393	5.82530921	3.40632934
14GHS00226S01	6.02729167	3.44145833	5.91707823	3.40774982
14GHS01684S01	5.652	3.638	5.62153973	3.52957908

14GHS04629S02	6.05794118	3.33117647	6.02668853	3.29901579
14GHS04376S01	6.03861111	3.77888889	5.93277378	3.72724586
14GHS01443S01	5.99380952	3.1552381	6.06582842	3.3878297
14GHS02992S01	5.94093023	2.62511628	6.0426781	2.67776044
14GHS02963S01	6.31156863	3.51607843	6.06032232	3.22555936
14GHS02749S01	6.06533333	2.642	6.06318835	2.63147881
14GHS03857S01	5.8082	2.8738	5.76545291	2.87529293
14GHS03676S01	6.30368421	2.9181579	6.26727283	2.82582928
14GHS00393S01	5.39125	3.1953125	5.43952967	3.01139845
14GHS02920S01	5.57476923	2.79092308	5.66413988	2.80991114
14GHS00291S01	5.7778125	3.3353125	5.85094546	3.2433813
14GHS03624S01	5.675	2.90304348	5.9273068	2.92643268
14GHS00645S01	5.90932203	2.42271186	6.04362544	2.73299152
14GHS01635S01	5.205	2.6375	5.20602813	2.80031369
14GHS03901S01	5.92	2.60333333	6.03893336	2.57159479
14GHS03773S02	3.31142857	1.30285714	3.57019937	1.48869435
14GHS02544S02	5.14357143	2.32107143	4.97573465	2.44168333
14GHS02669S02	5.44435897	2.30769231	5.36784822	2.37051624
14GHS03952S02	6.64	3.24	6.07810985	2.9458523
14GHS03763S02	5.8927907	2.85255814	6.09616807	3.04237151
14GHS03763S01	5.44066667	3.03366667	5.20635195	2.95145684
14GHS03703S02	6.35105263	2.75578947	6.56503009	2.86392407
14GHS03579S02	5.49666667	1.92	5.28795279	1.99918268
14GHS03559S02	4.89	2.372	4.89487765	2.39746833
14GHS04425S01	5.687	3.0835	6.14217425	3.31775967
14GHS04443S02	4.72863014	2.66630137	5.00004705	2.78388291
14GHS04451S02	5.68	2.7	5.72220689	2.48246727
14GHS04826S02	5.32375	3.23375	5.40586739	3.19825706
14GHS04530S01	5.82633333	4.32633333	5.67733467	3.18634123
14GHS04704S02	6.01833333	2.9425	6.03324191	2.83801015

14GHS04910S02	5.29344828	2.63344828	5.62393782	2.79487638
14GHS04819S02	5.32571429	2.52785714	5.45203249	2.62535683
14GHS04528S02	5.626	3.126	5.45008988	3.01720983
14GHS04872S02	5.47394737	2.77	5.21452489	2.55616414
14GHS02667S02	5.48214286	2.86142857	4.86034039	2.53892381
14GHS02538S01	5.39216216	2.96567568	5.64908272	2.7722535
14GHS02667S01	6.06733333	3.07633333	6.24389091	3.01820723
14GHS02695S02	5.37642857	2.95904762	5.44978886	2.62346171
14GHS02639S02	5.12	2.53222222	5.47107807	2.66838394
14GHS03254S02	5.46590909	3.21	5.94936747	3.15758188
14GHS04622S02	5.22263158	2.31421053	5.34622532	2.36379769
14GHS05064S02	5.43820513	2.6574359	5.15487892	2.41838558
14GHS05064S01	5.88631579	2.98947368	5.78438057	2.68972497
14GHS04817S01	5.41969697	3.02333333	5.47566581	2.921961
14GHS04817S02	5.04166667	2.70166667	5.30764784	2.99112479
14GHS05134S01	5.47666667	2.57	5.60052247	2.73465667
14GHS05062S01	6.07625	3.210625	6.27808052	3.29533726
14GHS05062S02	5.535	2.67395833	6.03986006	2.82064181
14GHS05216S01	5.6845	3.4165	6.30117628	3.43240263
14GHS04621S02	5.67192308	3.05730769	6.08331331	3.16588155
14GHS04621S01	5.843	3.08133333	6.24649597	3.10742757
14GHS02704S01	5.72125	2.573	6.2656136	2.86686013
14GHS03722S02	5.32	2.318	6.09862674	2.74858454
14GHS03930S02	4.8823913	1.945	5.38536321	2.20487746

Appendix B - Supplementary Material Chapter 3

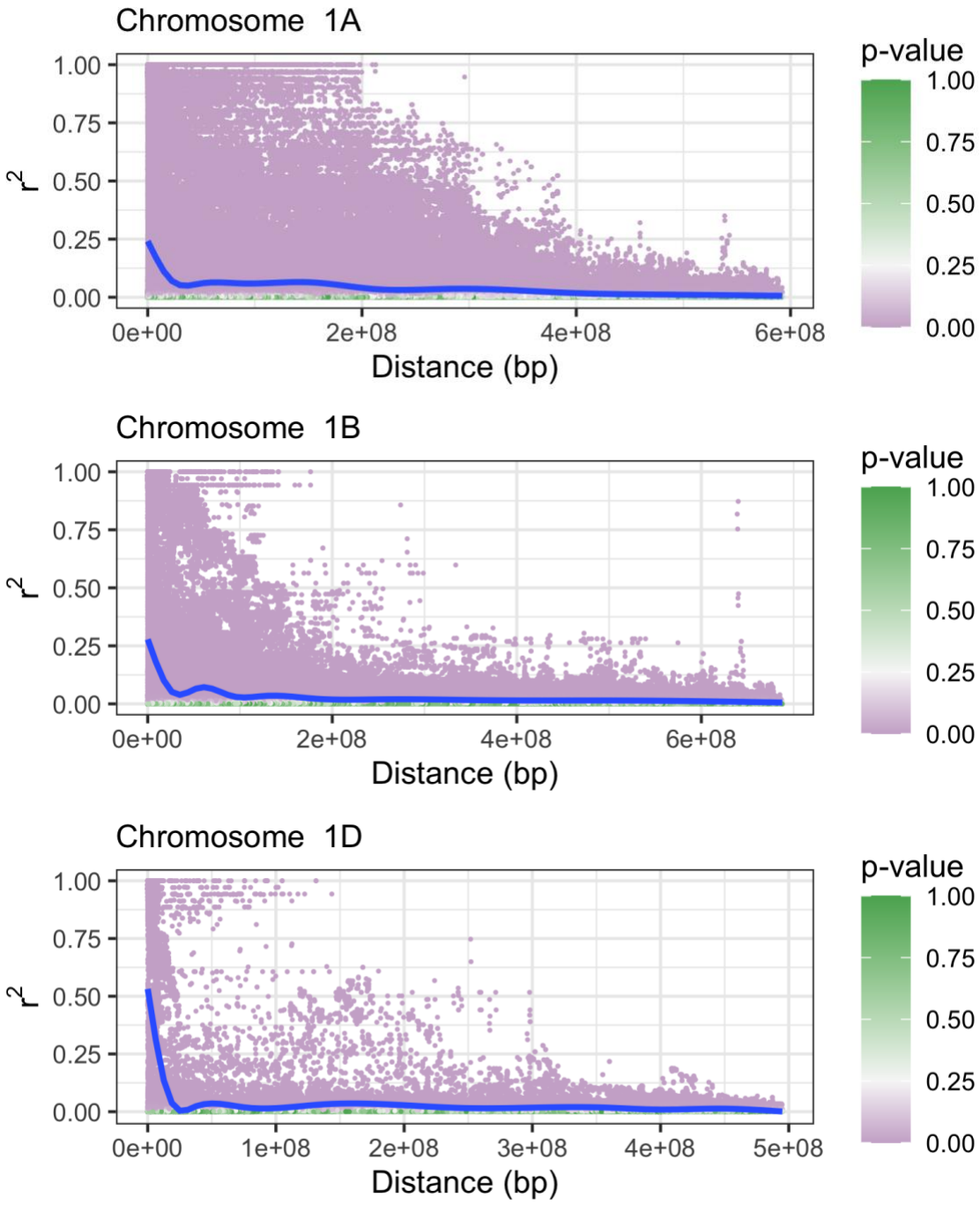


Figure B.1. Linkage disequilibrium decay by chromosome.

The distance between markers is given on the x-axis in base pairs and the squared correlation coefficient is given on the y-axis. The blue line represents the trend line.

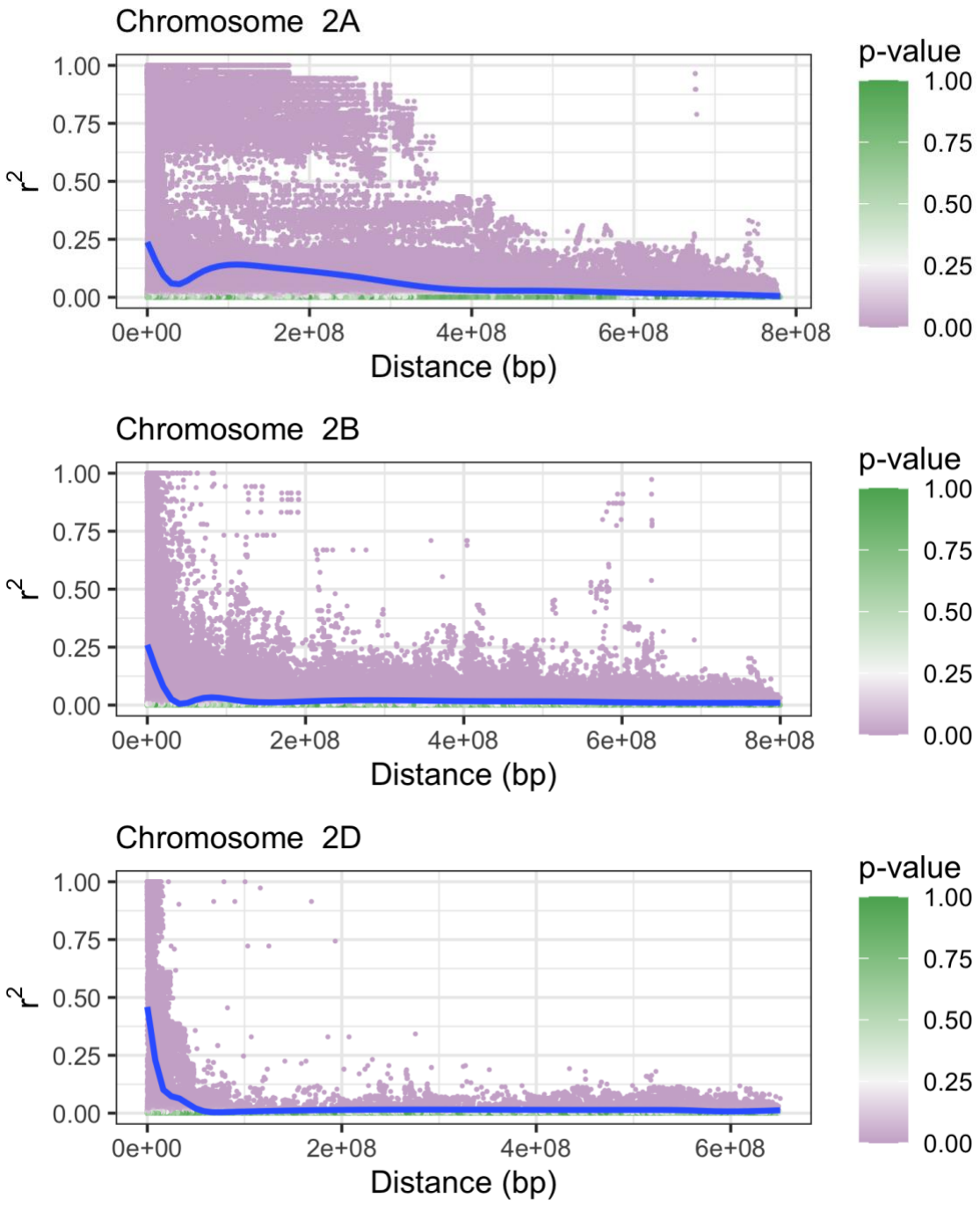


Figure B.1. Linkage disequilibrium decay by chromosome (continued).

The distance between markers is given on the x-axis in base pairs and the squared correlation coefficient is given on the y-axis. The blue line represents the trend line.

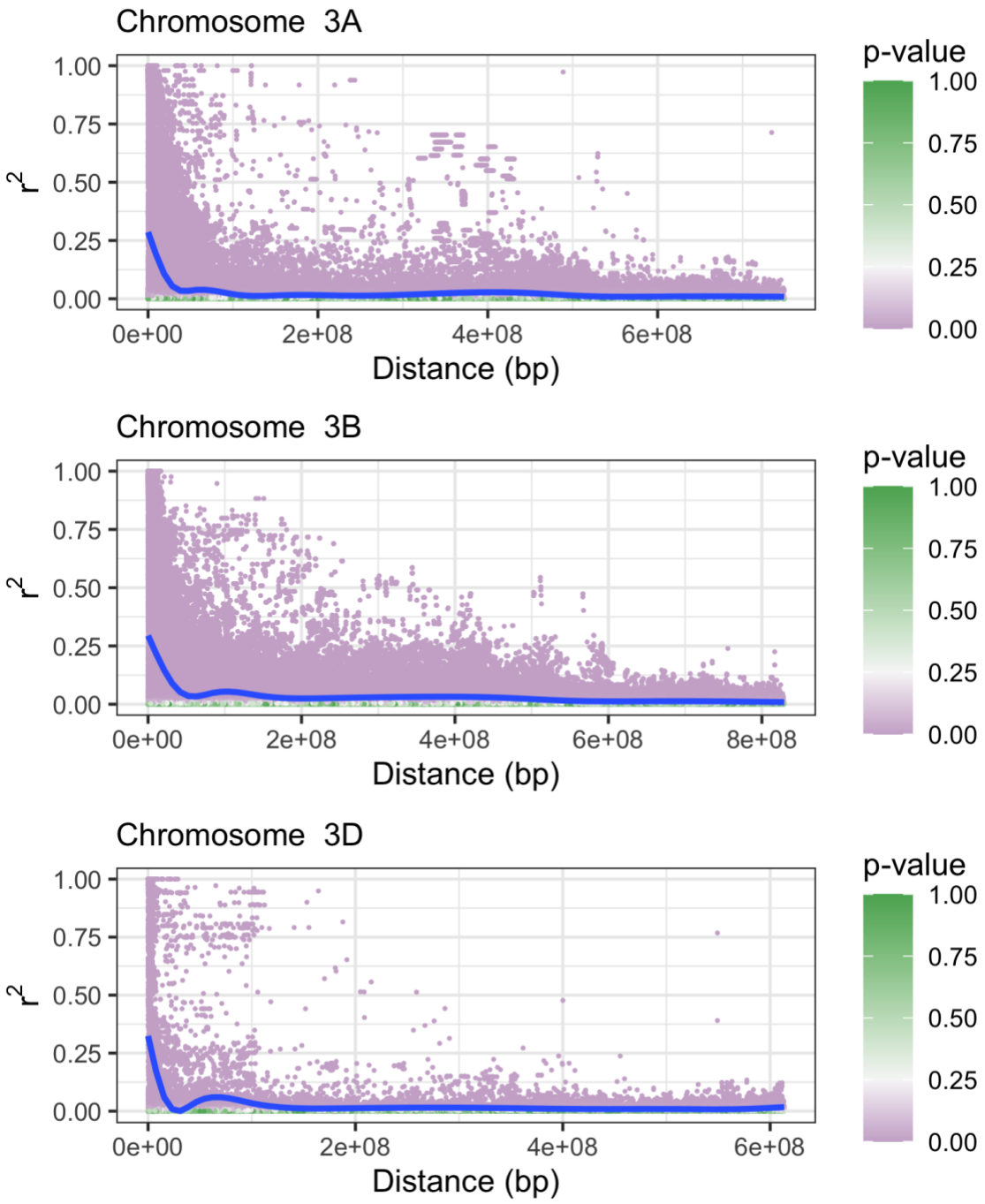


Figure B.1. Linkage disequilibrium decay by chromosome (continued).

The distance between markers is given on the x-axis in base pairs and the squared correlation coefficient is given on the y-axis. The blue line represents the trend line.

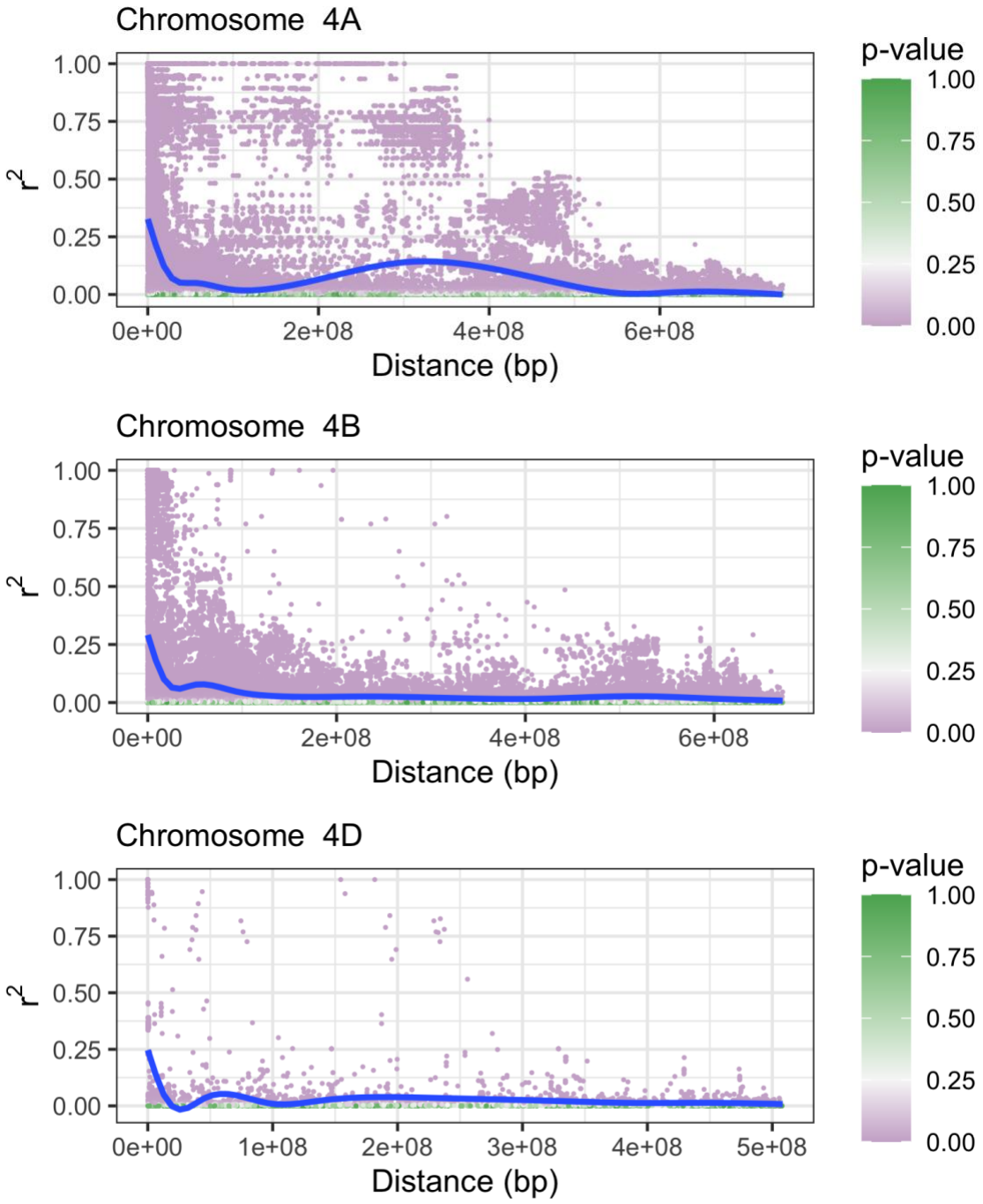


Figure B.1. Linkage disequilibrium decay by chromosome (continued).

The distance between markers is given on the x-axis in base pairs and the squared correlation coefficient is given on the y-axis. The blue line represents the trend line.

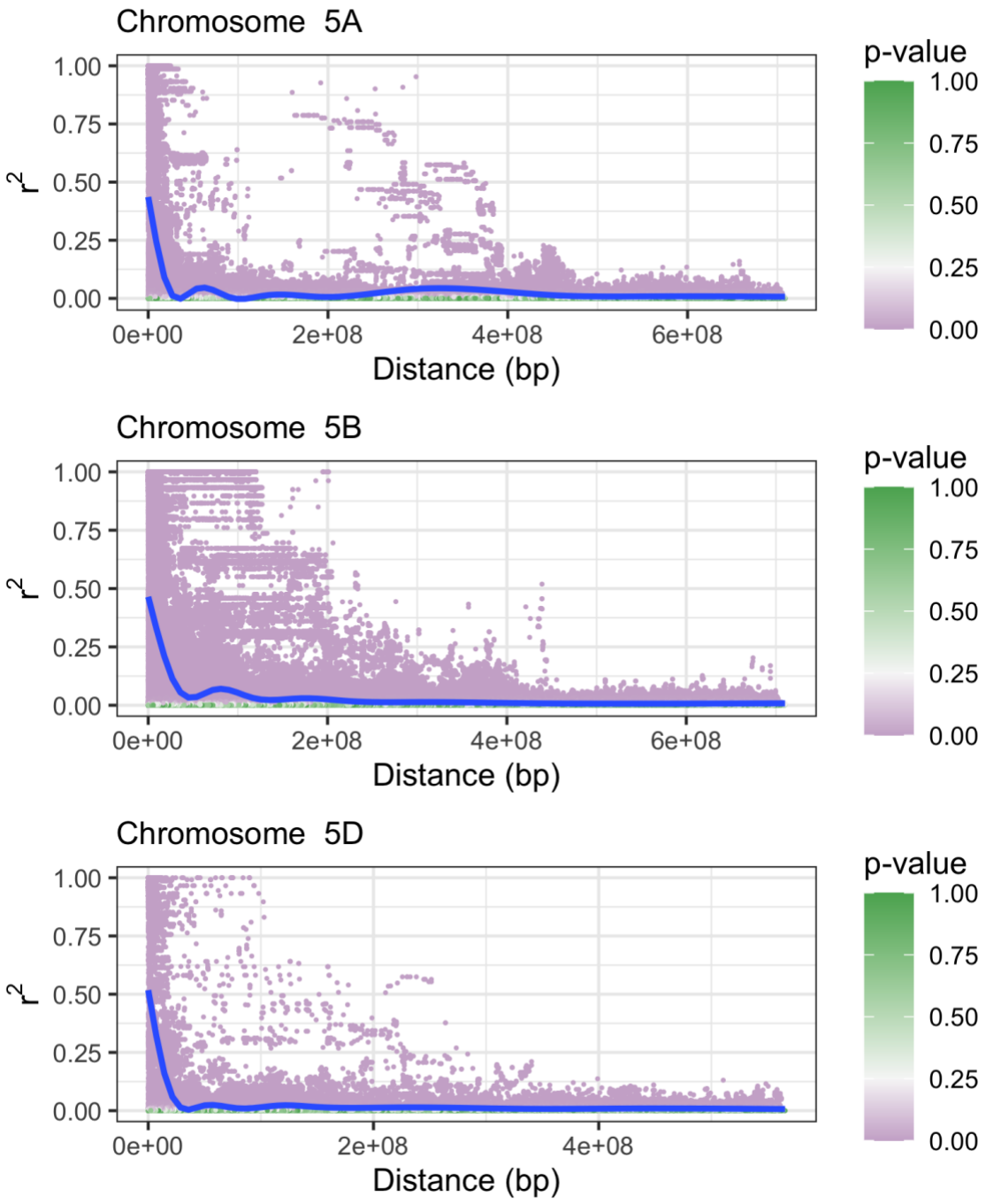


Figure B.1. Linkage disequilibrium decay by chromosome (continued).

The distance between markers is given on the x-axis in base pairs and the squared correlation coefficient is given on the y-axis. The blue line represents the trend line.

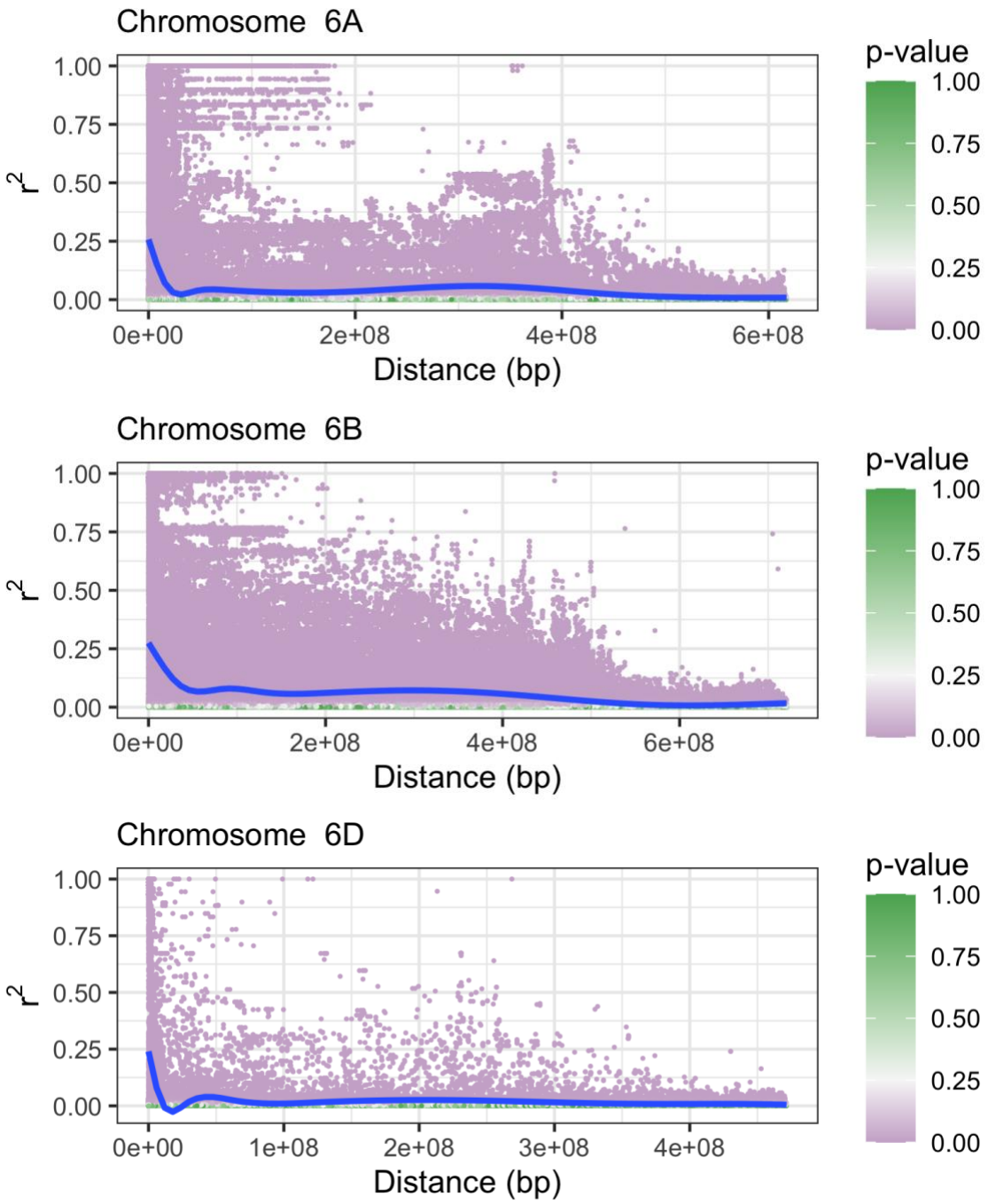


Figure B.1. Linkage disequilibrium decay by chromosome (continued).

The distance between markers is given on the x-axis in base pairs and the squared correlation coefficient is given on the y-axis. The blue line represents the trend line.

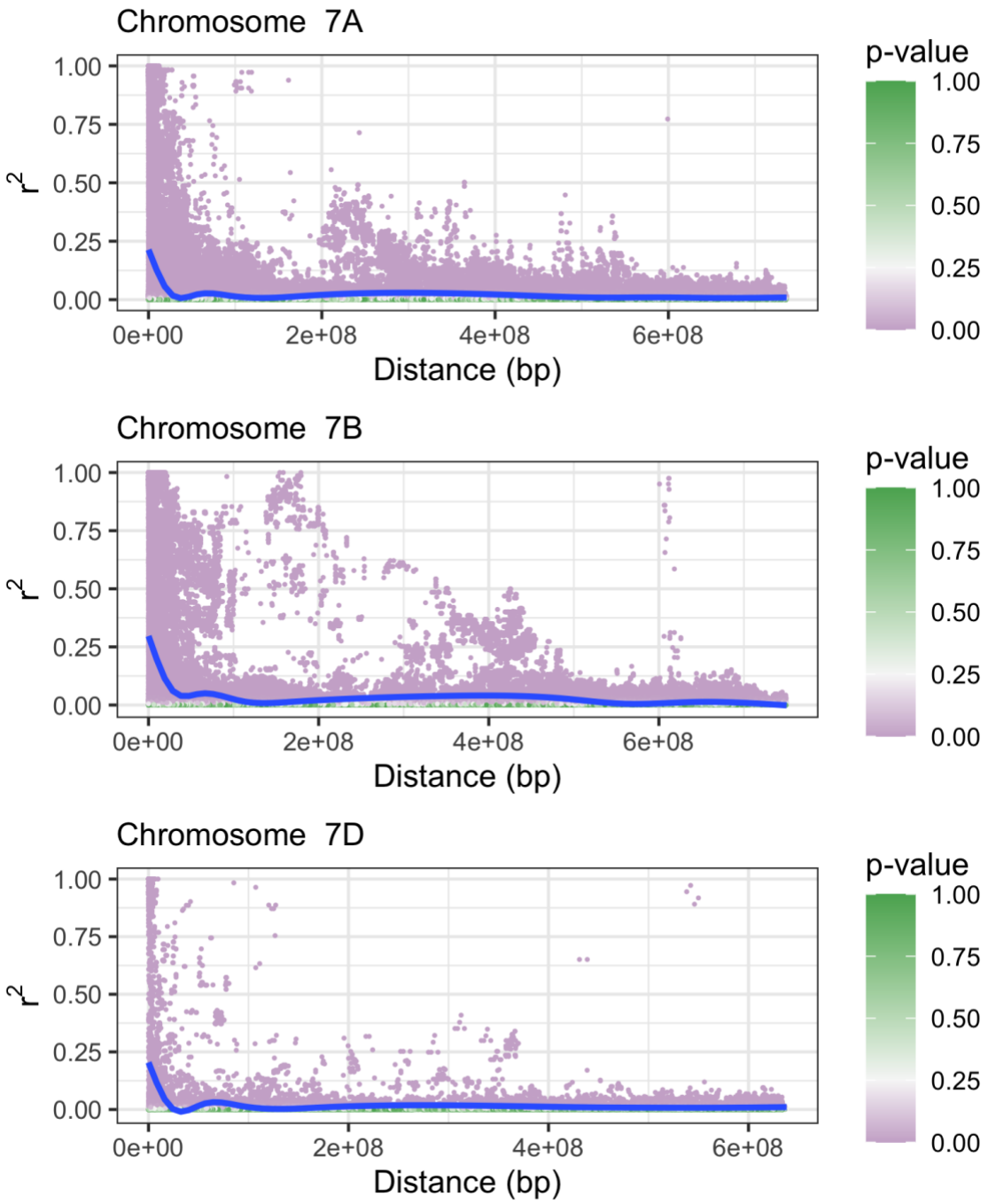


Figure B.1. Linkage disequilibrium decay by chromosome (continued).

The distance between markers is given on the x-axis in base pairs and the squared correlation coefficient is given on the y-axis. The blue line represents the trend line.

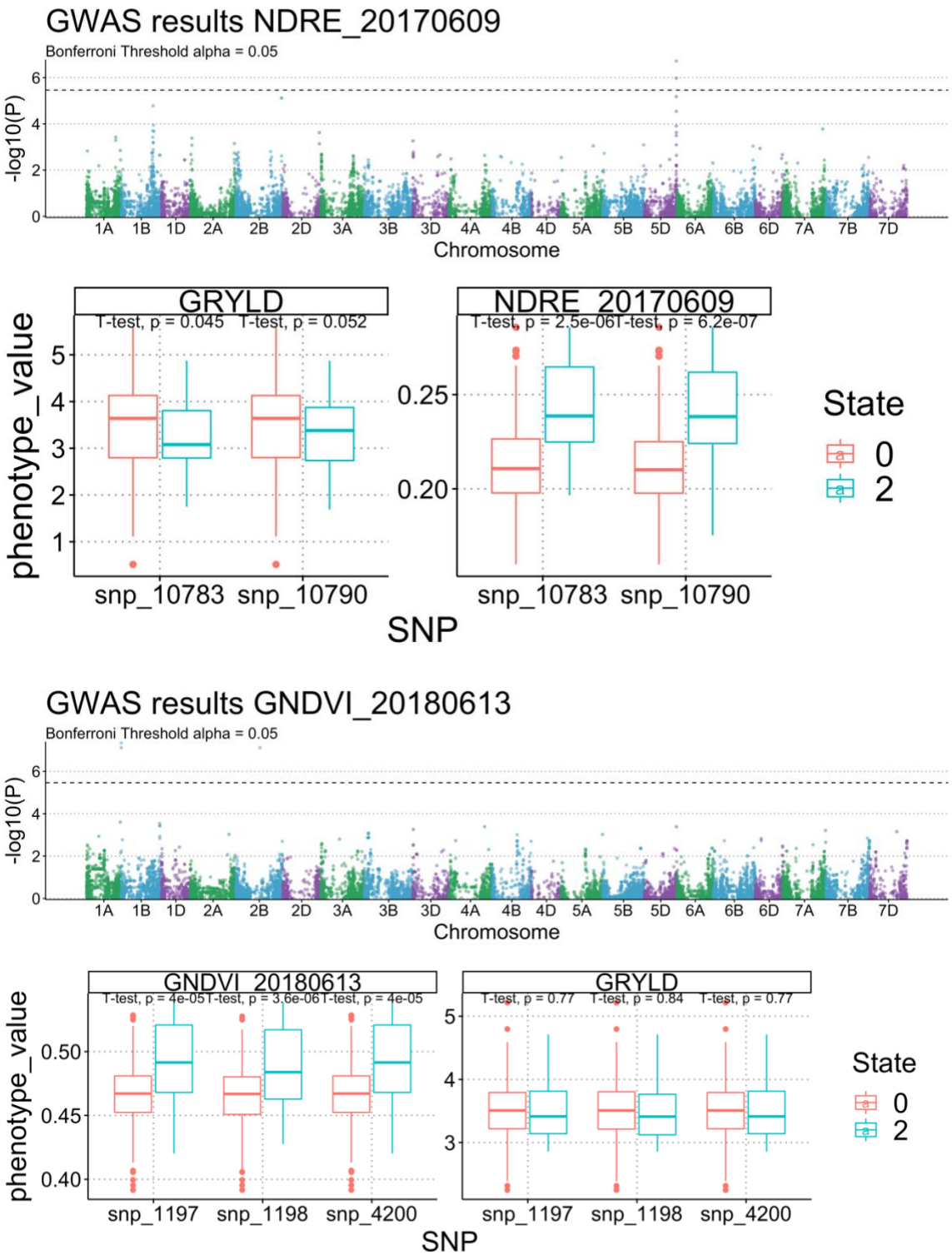


Figure B.2. Results of GWAS Analysis.

The effect of the significant SNPs identified in the GWAS analysis for the specific phenotype and grain yield are given in the boxplot below the GWAS results. The points are colored by chromosome.

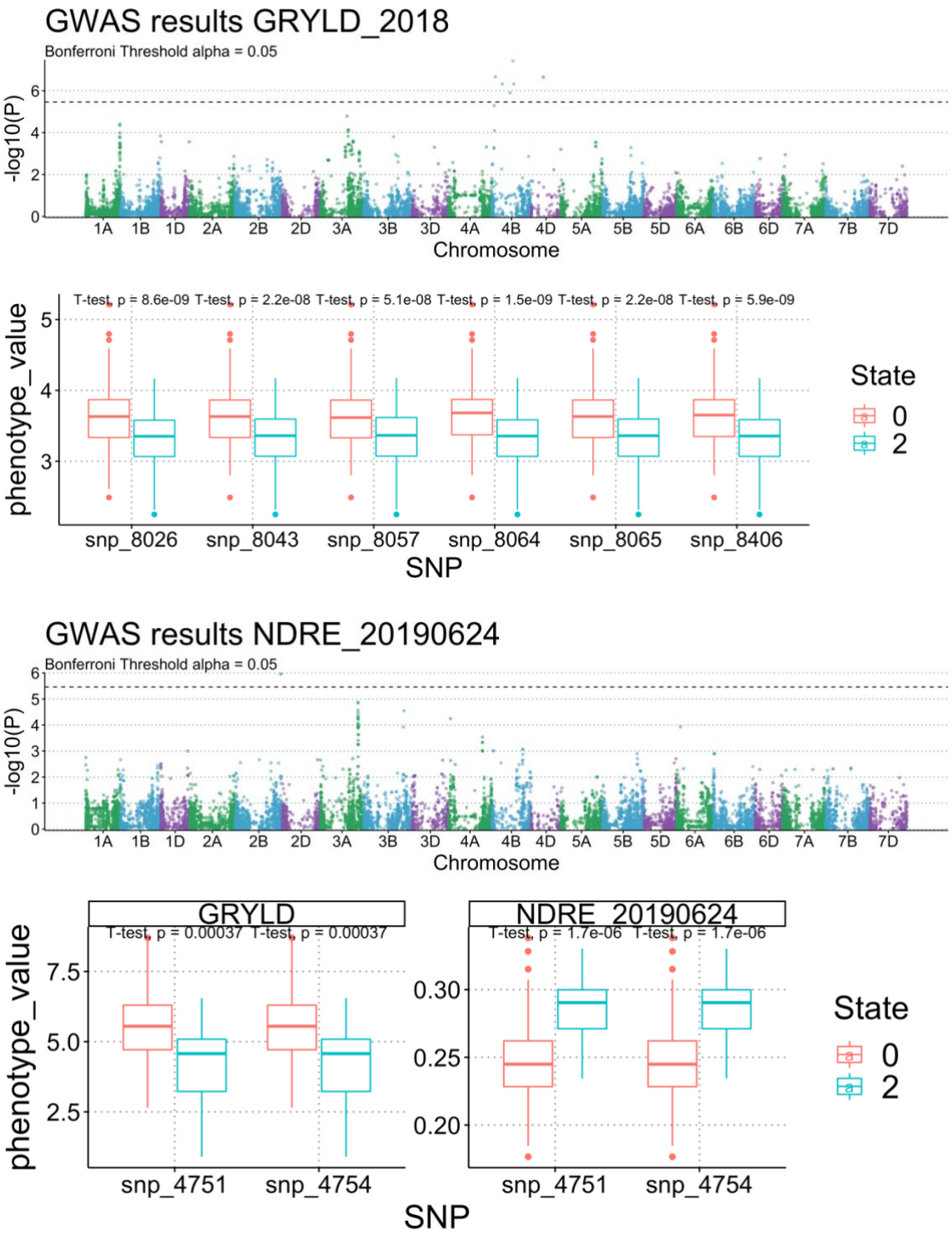


Figure B.2. Results of GWAS Analysis Continued.

The effect of the significant SNPs identified in the GWAS analysis for the specific phenotype and grain yield are given in the boxplot below the GWAS results. The points are colored by chromosome.

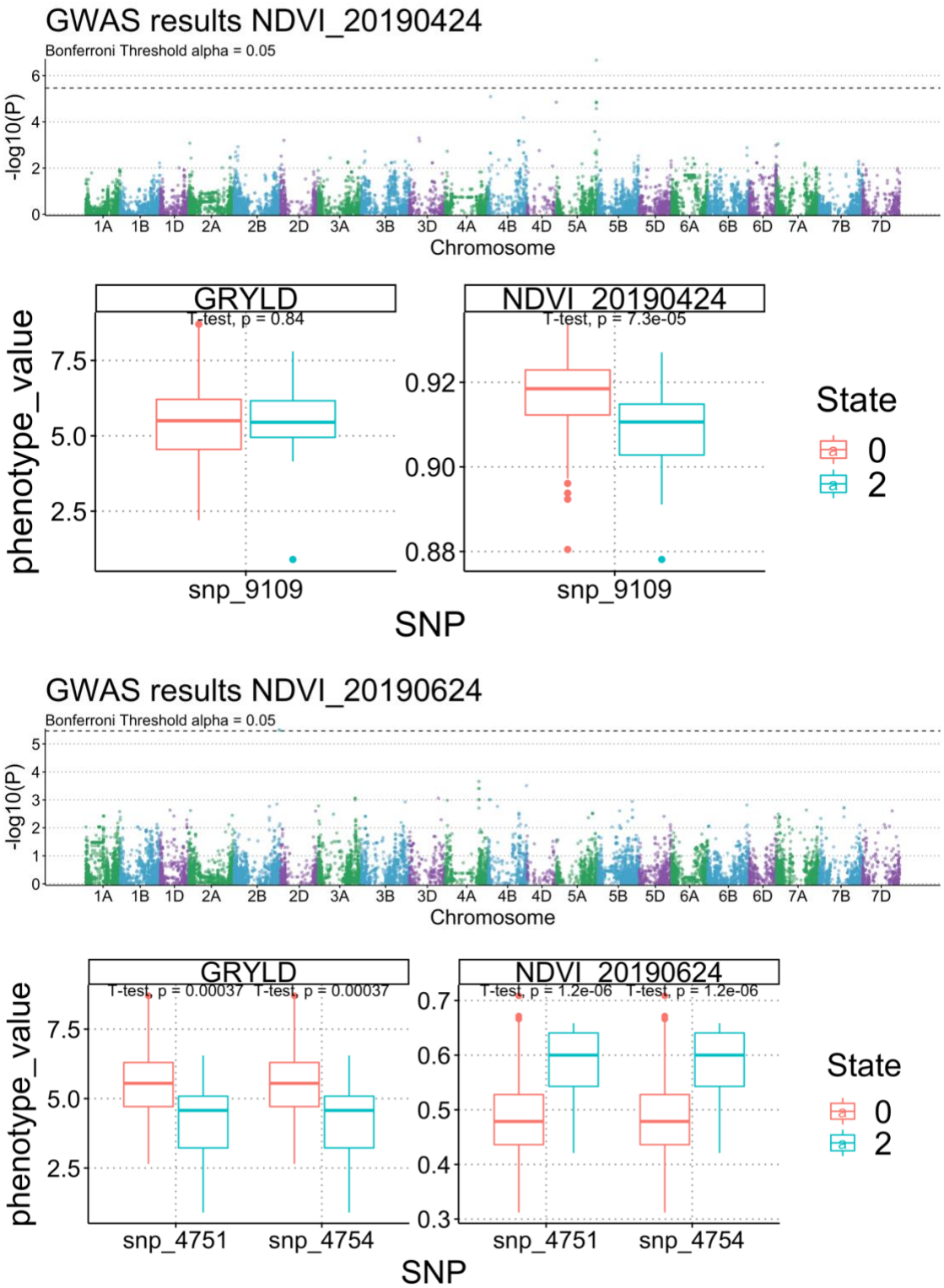


Figure B.2. Results of GWAS Analysis Continued.

The effect of the significant SNPs identified in the GWAS analysis for the specific phenotype and grain yield are given in the boxplot below the GWAS results. The points are colored by chromosome.

Table B.1. Planting and yield information for AM panel between 2016/2017-2018/2019. The location of the field in Riley County Kansas, the planting date, harvest date, mean yield for the entire field and the standard deviation are given.

Season	Location	Date Planted	Date Harvested	Mean Yield (t/ha)	SD Yield (t/ha)
2016/2017	Ashland Bottoms	19/10/2016	23/06/2017	3.469	1.023
2017/2018	Ashland Bottoms	26/10/2017	27/06/2017	3.496	0.621
2018/2019	Rocky Ford	24/10/2018	01/07/2019	5.424	1.402

Table B.2. Vegetation Indices used with references

Abbr.	Index	Equation	Reference
GNDVI	Green Normalized Difference Vegetation Index	$GNDVI = \frac{NIR - Green}{NIR + Green}$	(Blackmer et al. 1996)
NDRE	Red Edge Normalized Difference Vegetation Index	$NDRE = \frac{NIR - RedEdge}{NIR + RedEdge}$	(Elliott and Regan 1993)
NDVI	Normalized Difference Vegetation Index	$NDVI = \frac{NIR - Red}{NIR + Red}$	(Curran et al. 1983)

Appendix C - Supplementary Material Chapter 4

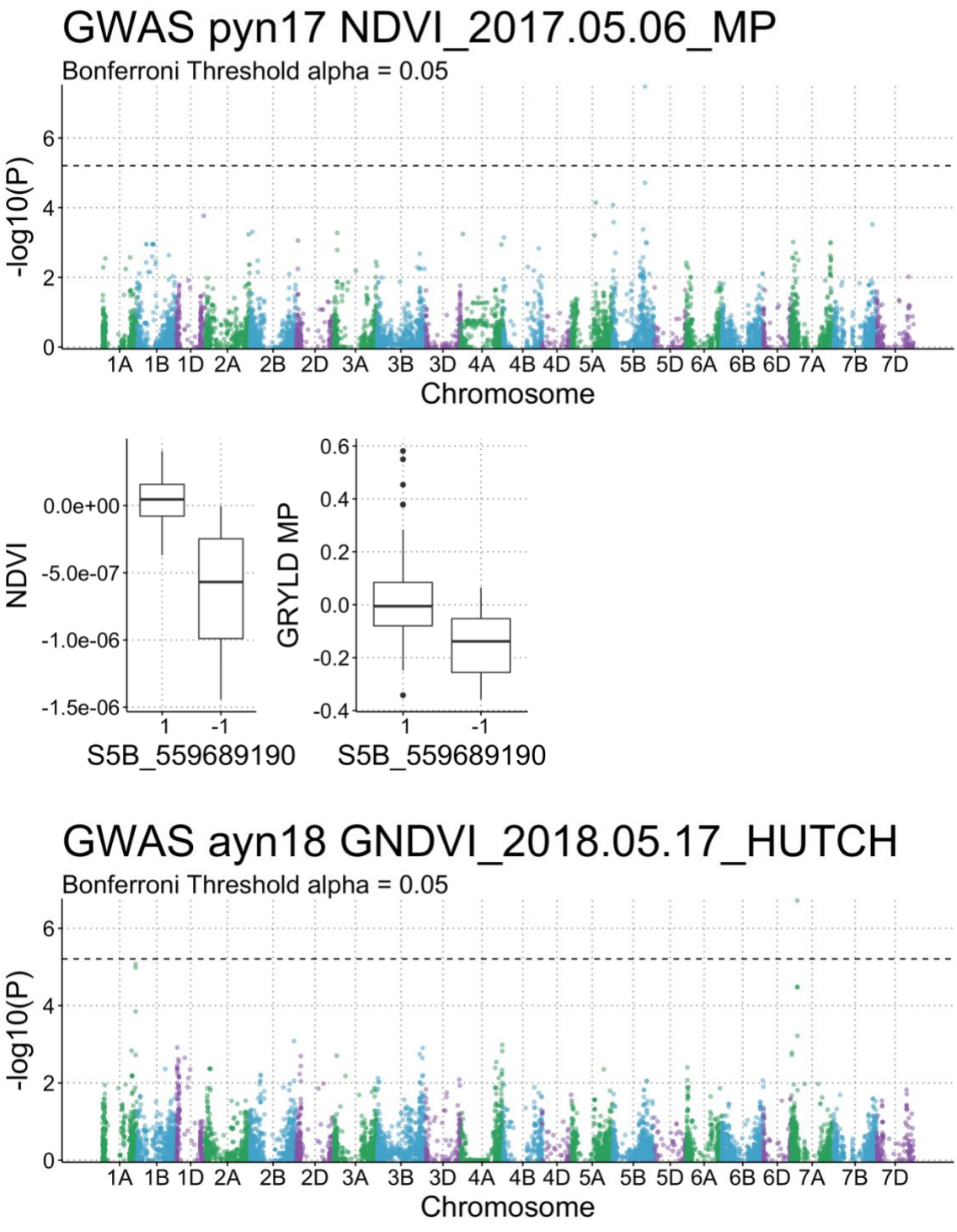


Figure C.1. Results of GWAS Analysis.

The effect of the significant SNPs identified in the GWAS analysis for the specific phenotype and grain yield are given in the boxplot below the GWAS results. The points are colored by chromosome.

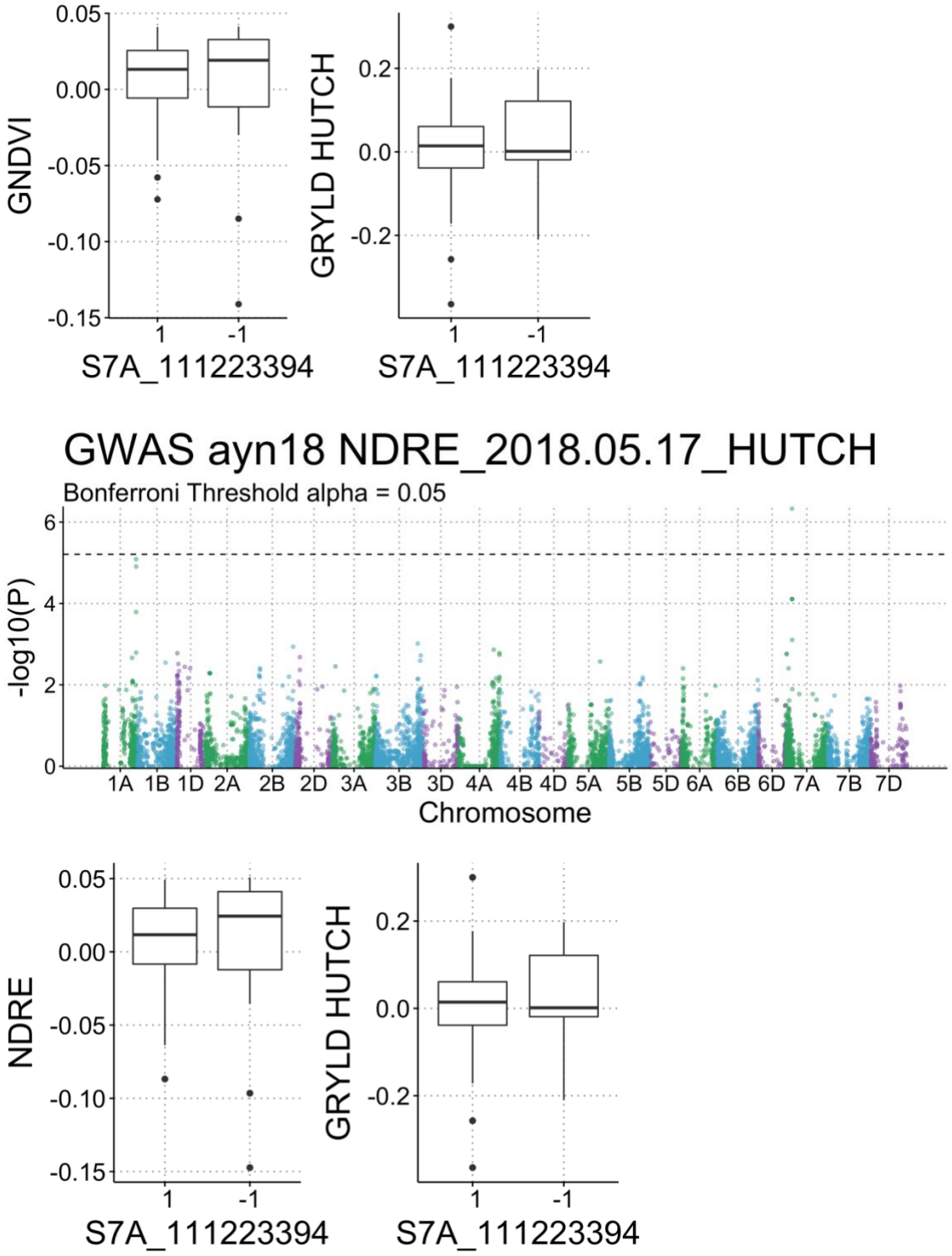
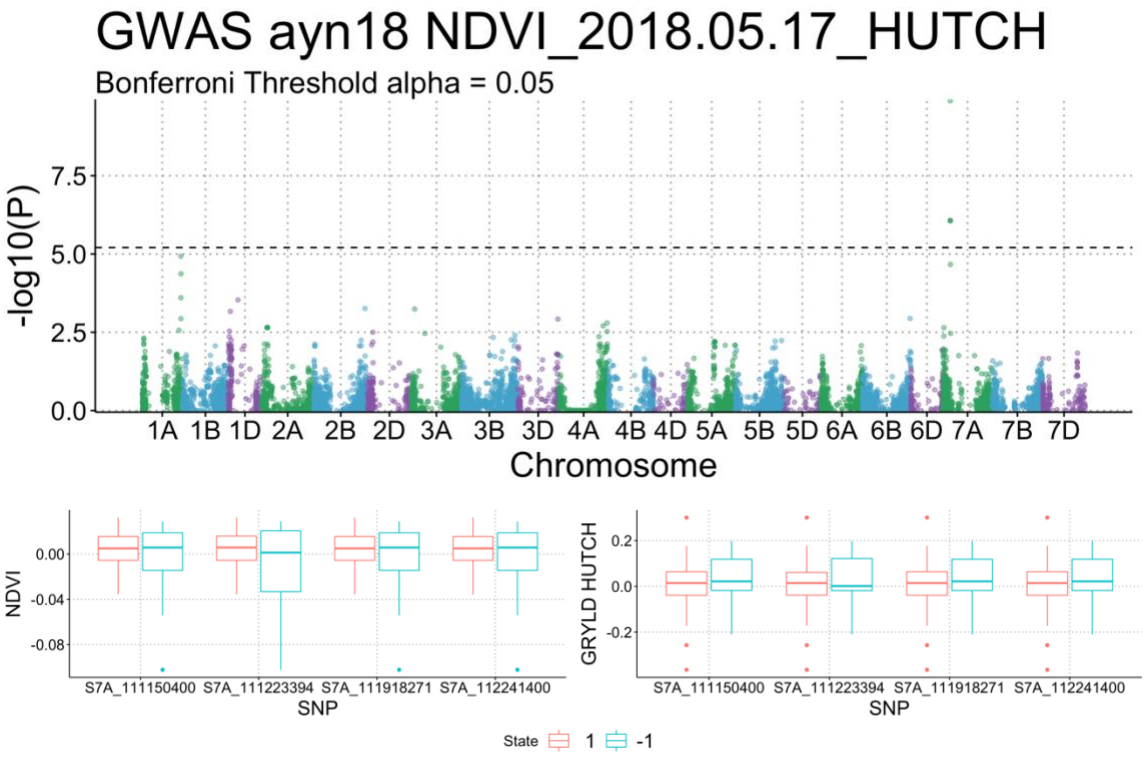


Figure C.1. Results of GWAS Analysis Continued.

The effect of the significant SNPs identified in the GWAS analysis for the specific phenotype and grain yield are given in the boxplot below the GWAS results. The points are colored by chromosome.



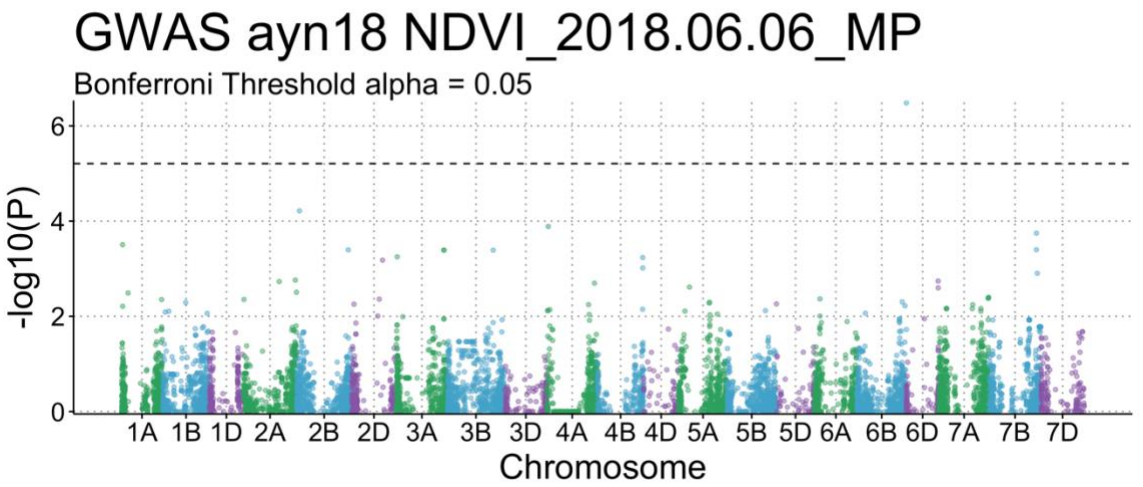


Figure C.1. Results of GWAS Analysis Continued.

The effect of the significant SNPs identified in the GWAS analysis for the specific phenotype and grain yield are given in the boxplot below the GWAS results. The points are colored by chromosome.

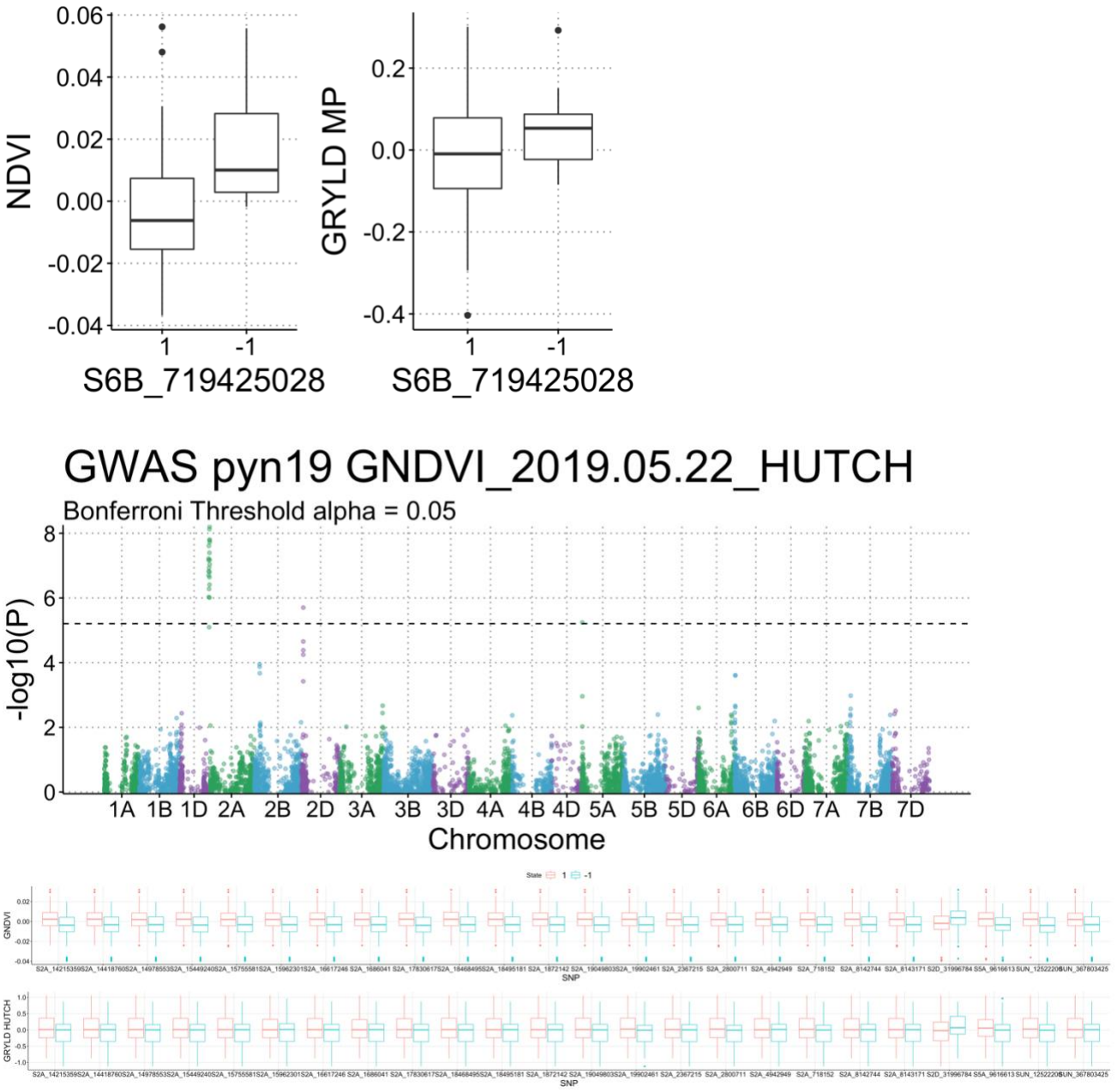


Figure C.1. Results of GWAS Analysis Continued.

The effect of the significant SNPs identified in the GWAS analysis for the specific phenotype and grain yield are given in the boxplot below the GWAS results. The points are colored by chromosome.

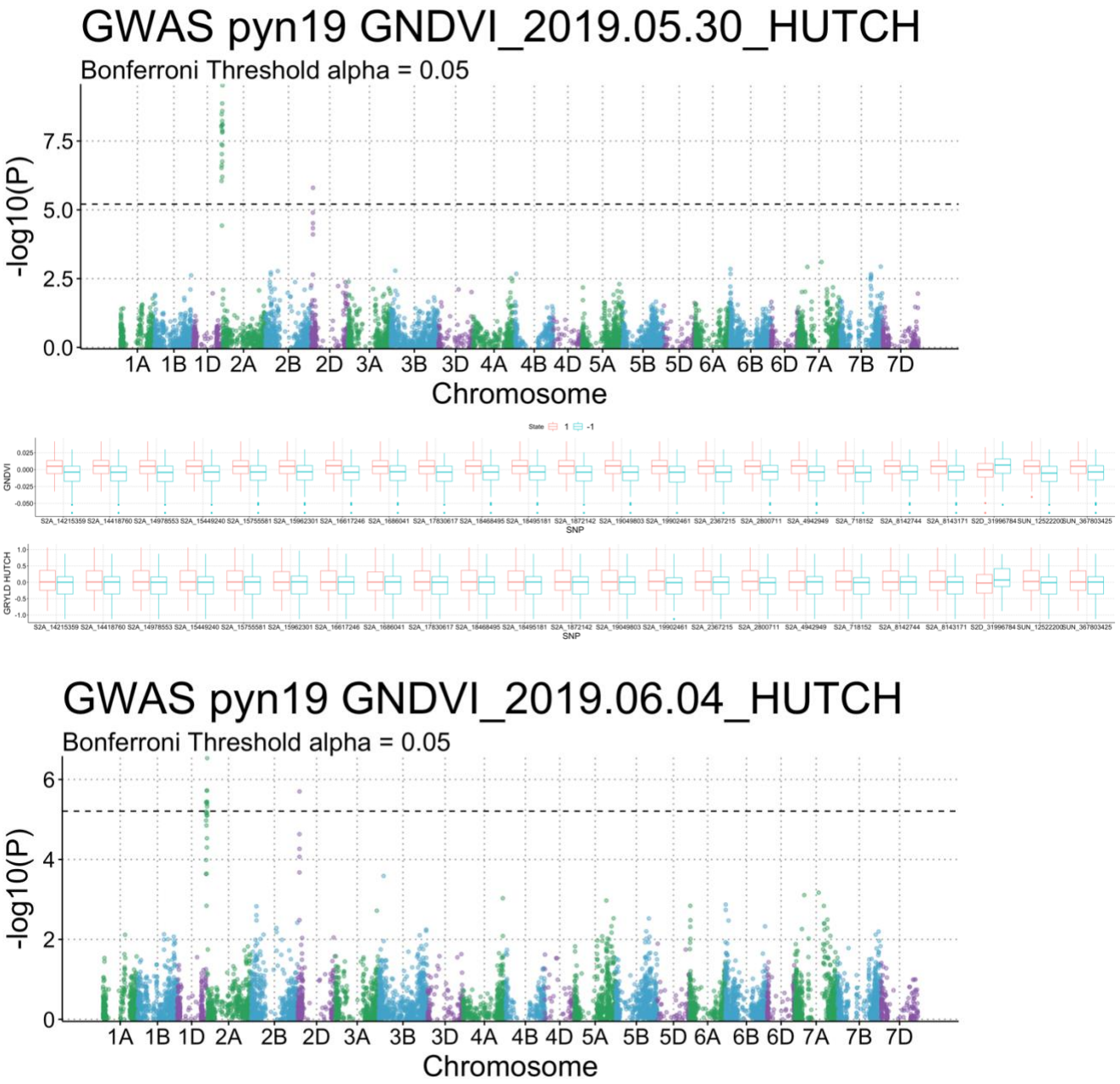


Figure C.1. Results of GWAS Analysis Continued.

The effect of the significant SNPs identified in the GWAS analysis for the specific phenotype and grain yield are given in the boxplot below the GWAS results. The points are colored by chromosome.

Table C.1. Summary of simulation costs. The range over which the parameter was tested as well as the step change for each are given.

Parameter	Range	Step
Line Development	$4-40	$1
Genotyping Costs	$1-20	$1
Phenotyping Costs	$35-70	$5
Phenotyping Replications	6	N/A
Heritability of Primary Trait	0-1	0.1
Heritability of Correlated Trait	0.95	N/A
Prediction Accuracy	0-1	0.1
AYN Selected	100	N/A

Table C.2. Accuracy of different GP models. The average correlation between the observed and GEBVs for the different genomic prediction models for 100 CV's is given with the standard deviation in brackets.

Season	Ridge Regression	Reproducing Kernel Hilbert Space	BayesC
All	0.3274 (0.0494)	0.3279 (0.0492)	0.3274 (0.0496)
2015/2016	0.4505 (0.0700)	0.4495 (0.0700)	0.4500 (0.0703)
2016/2017	0.4834 (0.1075)	0.4843 (0.1075)	0.4840 (0.1067)
2017/2018	0.2963 (0.0702)	0.2962 (0.0704)	0.2971 (0.0702)
2018/2019	0.3700 (0.0802)	0.3704 (0.0793)	0.3704 (0.0801)
Excluding_2016	0.3197 (0.0516)	0.3201 (0.0518)	0.3201 (0.0526)
Excluding_2017	0.3342 (0.0551)	0.3345 (0.0551)	0.3341 (0.0555)
Excluding_2018	0.3619 (0.0533)	0.3634 (0.0539)	0.3623 (0.0534)
Excluding_2019	0.3682 (0.0610)	0.3683 (0.0611)	0.3678 (0.0612)