Phenomics enabled genetic dissection of complex traits in wheat breeding

by

Daljit Singh

B.Sc., Punjab Agricultural University, 2010
M.S., Virginia Tech, 2013

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Abstract

A central question in modern biology is to understand the genotype-to-phenotype (G2P) link, that is, how the genetics of an organism results in specific characteristics. However, prediction of phenotypes from genotypes is a difficult problem due to the complex nature of genomes, the environment, and their interactions. While the recent advancements in genome sequencing technologies have provided almost unlimited access to high-density genetic markers, large-scale rapid and accurate phenotyping of complex plant traits remains a major bottleneck. Here, we demonstrate field-based complex trait assessment approaches using a commercially available light-weight Unmanned Aerial Systems (UAS). By deploying novel data acquisition and processing pipelines, we quantified lodging, ground cover, and crop growth rate of 1745 advanced spring wheat lines at multiple time-points over the course of three field seasons at three field sites in South Asia. High correlations of digital measures to visual estimates and superior broad-sense heritability demonstrate these approaches are amenable for reproducible assessment of complex plant traits in large breeding nurseries. Using these validated high-throughput measurements, we applied genome-wide association and prediction models to assess the underlying genetic architecture and genetic control. Our results suggest a diffuse genetic architecture for lodging and ground cover in wheat, but heritable genetic variation for prediction and selection in breeding programs. The logistic regression-derived parameters of dynamic plant height exhibited strong physiological linkages with several developmental and agronomic traits, suggesting the potential targets of selection and the associated tradeoffs. Taken together, our highly reproducible approaches provide a proof-of-concept application of UAS-based phenomics that is scalable to tens-of-thousands of plots in breeding and genetic studies as will be needed to understand the G2P and increase the rate of gain for complex traits in crop breeding.

Phenomics enabled genetic dissection of complex traits in wheat breeding

by

Daljit Singh

B.Sc., Punjab Agricultural University, 2010
M.S., Virginia Tech, 2013

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Approved by:

Major Professor
Jesse A. Poland

Copyright

Abstract

A central question in modern biology is to understand the genotype-to-phenotype (G2P) link, that is, how the genetics of an organism results in specific characteristics. However, prediction of phenotypes from genotypes is a difficult problem due to the complex nature of genomes, the environment, and their interactions. While the recent advancements in genome sequencing technologies have provided almost unlimited access to high-density genetic markers, large-scale rapid and accurate phenotyping of complex plant traits remains a major bottleneck. Here, we demonstrate field-based complex trait assessment approaches using a commercially available light-weight Unmanned Aerial Systems (UAS). By deploying novel data acquisition and processing pipelines, we quantified lodging, ground cover, and crop growth rate of 1745 advanced spring wheat lines at multiple time-points over the course of three field seasons at three field sites in South Asia. High correlations of digital measures to visual estimates and superior broad-sense heritability demonstrate these approaches are amenable for reproducible assessment of complex plant traits in large breeding nurseries. Using these validated high-throughput measurements, we applied genome-wide association and prediction models to assess the underlying genetic architecture and genetic control. Our results suggest a diffuse genetic architecture for lodging and ground cover in wheat, but heritable genetic variation for prediction and selection in breeding programs. The logistic regression-derived parameters of dynamic plant height exhibited strong physiological linkages with several developmental and agronomic traits, suggesting the potential targets of selection and the associated tradeoffs. Taken together, our highly reproducible approaches provide a proof-of-concept application of UAS-based phenomics that is scalable to tens-of-thousands of plots in breeding and genetic studies as will be needed to understand the G2P and increase the rate of gain for complex traits in crop breeding.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to extend my sincere gratitude to my advisor Dr. Jesse Poland for his continued support during this scientific journey. It has been truly an honor to have worked alongside a great scientist, leader, and a human being. Jesse, thank you for patiently teaching me the basics of science, leadership, and collaborative work. Your continued encouragement has not only allowed me to expand knowledge, it has inculcated a strong sense of responsibility towards science and global food security. I would also like to thank Dr. Geoff Morris for bringing the "broader picture" and "hypothesis driven" focus into my academic training. I appreciate my committee members Dr. Shawn Hutchinson and Dr. Ignacio Ciampitti for their guidance and support.

Dr. Xu "Kevin" Wang has been more than a colleague during this interdisciplinary journey. Kevin, thanks for bringing the engineering temperament into my thought process. Haley Ahlers, your creative inputs into this research work were invaluable. Bo Winters, thank you for putting up with my million pages long travel documentation. Thank you, Jared and Narinder, for stimulating discussions on a range of topics.

Dr. Uttam Kumar (BISA) for being a great collaborator and travelling companion during long train journeys through the Indian wheat belt. This research work would not have been possible without the support of field staff at Borlaug Institute for South Asia, CIMMYT, Poland Lab members as well as our US and International collaborators. A big thanks also goes to friends abroad and home, especially Gurinderpal and Vipin, for their support.

Dedication

This work is dedicated to my parents Karamjit Kaur and Harjinder Singh, wife Komal, brother Taran, and mentors.

Chapter 1 - Connecting Genotype-to-Phenotype (G2P)

A central question in modern biology is to understand the genotype-to-phenotype (G2P) link, that is, how the genetics of an organism results in specific characteristics. However, prediction of phenotypes from genotypes is a challenging problem due to the complex nature of genomes, the environment, and their interactions (Turelli and Barton, 2004; Hammer et al., 2006; van Eeuwijk et al., 2016). The complex interactions between the G2P landscape components are dictated by the underlying trait architecture that shapes the variation within and between populations (Cooper et al., 2002; Mitchell-Olds et al., 2007). Many traits of agronomic importance such as yield are under polygenic control and are highly influenced by the environments (Fridman et al., 2004; Laurie et al., 2004; Sehgal et al., 2017). Furthermore, the complex intra- and inter-locus gene action makes the dissection of genetic architecture of polygenic traits a challenging task (Mei et al., 2005; Morgante et al., 2018). To this end, genetic studies comprising larger population sizes and robust statistical modeling framework that takes into account complex genetic architecture are highly desired (Segura et al., 2012; Sehgal et al., 2017; Morgante et al., 2018).

To facilitate the dissection of quantitative variation underlying complex traits, quality phenotypic measurements are highly desired. While the recent advancements in genome sequencing technologies have provided greater access to high-density genetic markers, large-scale rapid and accurate phenotyping of complex plant traits remains a major constraint in field studies (White et al., 2012; Poland, 2015). This phenotyping constraint is often termed as phenotyping bottleneck in literature and is currently an active area of research in plant sciences (Furbank and Tester, 2011). Rapid and high-throughput phenotyping (HTP) strategies, also known as phenomics, have made significant progress in the past decade, largely spurred by the interest in controlled environment experiments (Chen et al., 2014; Honsdorf et al., 2014;

Campbell et al., 2015; Neilson et al., 2015). Initial studies provided proof-of-concept of novel applications of HTP in controlled conditions for model plants (Samuel et al., 2011; Moore et al., 2013). Through a combination of genetic and temporal analysis, QTL underlying temporal phenotypic responses have been identified (Chen et al., 2014; Bac-Molenaar et al., 2015; Campbell et al., 2015). While indoor HTP systems have played a hugely disruptive role in the area of phenomics, the high cost, limited space, and inability to replicate natural field conditions were considered to be the major constraints (White et al., 2012).

For in-field breeding and genetics evaluations, the outdoor ground-based HTP platforms have been proposed (Araus and Cairns, 2014). The first generation of field-based HTP platforms were typically high-clearance customized vehicles (Montes et al., 2011; Andrade-Sanchez et al., 2014; Barker et al., 2016). The utility of these platforms in genetics studies was explored through QTL mapping (Busemeyer et al., 2013). Although these platforms offered a significant advantage in terms of data throughput, the higher equipment costs led to limited spread of this technology. Consequently, lightweight and low-cost HTP solutions such as pushcarts were developed (Crain et al., 2016).

But despite such technological progress the scalability and throughput remain a constraint in the context of large breeding and genetics experiments involving thousands of small plots(Reynolds et al., 2018). Particularly, the low throughput of ground-based vehicles can be a drawback for abiotic stress traits, the expression of which is highly time-sensitive (Baret et al., 2018).

For targeting the scope of field experiments, unmanned aerial systems (UAS) offer a flexible alternative to ground-based phenotyping platforms, particularly for large-breeding nurseries and genetic studies with thousands or tens-of-thousands of plots (Poland, 2015). Recently, UAS have been deployed in HTP of wheat breeding nurseries (Sankaran et al., 2015;

Haghighattalab et al., 2016), expanding previous work using multi-rotor UAS of varying sizes and payload capacity to phenotype small-sized test plots (Bendig et al., 2014; Chapman et al., 2014; Shi et al., 2016). In regard to the applications of UAS-phenotyping in large genetics experiments, there are very few published reports on genetic dissection of plant traits (Watanabe et al., 2017; Xavier et al., 2017; Condorelli et al., 2018). With rapid development of low-cost consumer-grade sensors and platforms, UAS phenotyping holds great potential to be an integral part of plant genomics and breeding for precise, quantitative assessment of otherwise low-throughput and complex traits on large populations (Baret et al., 2018). However, significant developments in processing, methodology and analysis of UAS-derived data are needed to realize its full potential. This dissertation seeks to address the core question of G2P through the application of UAS in large wheat breeding and genetics populations. A better understanding of G2P is expected to accelerate the genetic gains in plant breeding to meet the global demand for food, fiber and fuel. The specific research problems investigated in this dissertation are: 1.) application of a novel field-based lodging assessment approach. 2.) Genetic dissection of ground cover. 3) genetic and physiological dissection of dynamic plant height.

Application of Novel Field-Based Lodging Assessment Approaches

Lodging, the permanent displacement of the plant stem from vertical position, is an example of a complex trait that is difficult to quantify in the field (Pinthus, 1974). Conventional phenotyping methods for lodging are based on visual ratings of incidence and severity scores, and can be associated with stem or root lodging (Berry et al., 2003; Piñera-Chavez et al., 2016). The physiology of lodging in wheat is associated with complex genetic architecture involving numerous small effect loci (Verma et al., 2005; Liu et al., 2015; Miller et al., 2016). Therefore, the complex trait architecture of lodging is a good target for field-based high-throughput phenotyping to enable precise measurement of very large populations needed for genomics

studies and breeding progress. Proof-of-concept studies for lodging assessment have been demonstrated on a limited-scale in crops (Chapman et al., 2014; Chu et al., 2017; Yang et al., 2017). Chapter 2 investigates the methods of image-based quantitative assessment of lodging, and demonstrates the scalable application of this assessment methodology in large wheat breeding and genetic studies. This comprehensive evaluation of lodging assessment methods provides an analytical framework to further investigate the functional underpinnings of lodging in wheat and other crops.

Genetic Dissection of Ground Cover

Ground cover, measured as the fraction of plant canopy covering soil, is an important breeding target in many crops due to its key role in stress adaptation and grain yield. However, the dissection of complex physiological and genetic components of ground cover is hindered by the subjective and tedious nature of the phenotypic measurements in the field. Digital ground cover measurements enabled by UAS can improve the scale (temporal and spatial) and throughput of field experiments to uncover the complex physiological and genetic mechanisms underlying ground cover trait (Duan et al., 2016; Xavier et al., 2017). Chapter 3 investigates the potential of image-derived digital ground cover as an alternative to visual scores of ground cover in wheat. Physiological and genotype-by-environment interactions are also discussed. An advanced understanding of the complex biological components through phenomics enabled breeding will accelerate the rate of genetic gains in wheat and related crops.

Genetic and Physiological Dissection of Dynamic Plant Height

Optimization of morphological and developmental characteristics is critical to maximize the yield potential in crop varieties. Plant height is a key agronomic trait that is an important breeding target in many crops due to its association with biomass and grain yield (Allan, 1986; Peng et al., 1999; Hedden, 2003). However, the slow and static nature of the conventional

measurements is a major bottleneck to elucidate the genetic basis of the temporal dynamics of plant height in field experiments. In chapter 4, we investigate the genetic and physiological underpinnings of growth rate using high quality plant height measurements collected on 2400 plots at 34 time-points from early and normal planted field experiments in wheat. We envision that the full utility of these phenomics-based approaches will be realized by integration with whole-genome prediction, crop growth models, and other molecular biology tools (Cooper et al., 2014; van Eeuwijk et al., 2018).

References

Allan, R.E., 1986. Agronomic Comparisons among Wheat Lines Nearly Isogenic for Three Reduced-Height Genes1. Crop Science 26, 707-710.

Andrade-Sanchez, P., Gore, M.A., Heun, J.T., Thorp, K.R., Carmo-Silva, A.E., French, A.N., Salvucci, M.E., White, J.W., 2014. Development and evaluation of a field-based high-throughput phenotyping platform. Functional Plant Biology 41, 68-79.

Araus, J.L., Cairns, J.E., 2014. Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci 19, 52-61.

Bac-Molenaar, J.A., Vreugdenhil, D., Granier, C., Keurentjes, J.J., 2015. Genome-wide association mapping of growth dynamics detects time-specific and general quantitative trait loci. J Exp Bot 66, 5567-5580.

Baret, F., Madec, S., Irfan, K., Lopez, J., Comar, A., Hemmerle, M., Dutartre, D., Praud, S., Tixier, M.H., 2018. Leaf-rolling in maize crops: from leaf scoring to canopy-level measurements for phenotyping. J Exp Bot 69, 2705-2716.

Barker, J., Zhang, N.Q., Sharon, J., Steeves, R., Wang, X., Wei, Y., Poland, J., 2016. Development of a field-based high-throughput mobile phenotyping platform. Computers and Electronics in Agriculture 122, 74-85.

Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., Bareth, G., 2014. Estimating Biomass of Barley Using Crop Surface Models (CSMs) Derived from UAV-Based RGB Imaging. Remote Sensing 6, 10395-10412.

Berry, P.M., Spink, J., Sterling, M., Pickett, A.A., 2003. Methods for rapidly measuring the lodging resistance of wheat cultivars. Journal of Agronomy and Crop Science 189, 390-401.

Busemeyer, L., Ruckelshausen, A., Moller, K., Melchinger, A.E., Alheit, K.V., Maurer, H.P., Hahn, V., Weissmann, E.A., Reif, J.C., Wurschum, T., 2013. Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. Sci Rep 3, 2442.

Campbell, M.T., Knecht, A.C., Berger, B., Brien, C.J., Wang, D., Walia, H., 2015. Integrating Image-Based Phenomics and Association Analysis to Dissect the Genetic Architecture of Temporal Salinity Responses in Rice. Plant Physiol 168, 1476-1489.

Chapman, S., Merz, T., Chan, A., Jackway, P., Hrabar, S., Dreccer, M., Holland, E., Zheng, B., Ling, T., Jimenez-Berni, J., 2014. Pheno-Copter: A Low-Altitude, Autonomous Remote-Sensing Robotic Helicopter for High-Throughput Field-Based Phenotyping. Agronomy 4, 279-301.

Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., Klukas, C., 2014. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. Plant Cell 26, 4636-4655.

Chu, T.X., Starek, M.J., Brewer, M.J., Murray, S.C., Pruter, L.S., 2017. Assessing Lodging Severity over an Experimental Maize (Zea mays L.) Field Using UAS Images. Remote Sensing 9, 923.

Condorelli, G.E., Maccaferri, M., Newcomb, M., Andrade-Sanchez, P., White, J.W., French, A.N., Sciara, G., Ward, R., Tuberosa, R., 2018. Comparative Aerial and Ground Based High Throughput Phenotyping for the Genetic Dissection of NDVI as a Proxy for Drought Adaptive Traits in Durum Wheat. Front Plant Sci 9, 893.

Cooper, M., Chapman, S.C., Podlich, D.W., Hammer, G.L., 2002. The GP problem: quantifying gene-to-phenotype relationships. In silico biology 2, 151-164.

Cooper, M., Messina, C.D., Podlich, D., Totir, L.R., Baumgarten, A., Hausmann, N.J., Wright, D., Graham, G., 2014. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop & Pasture Science 65, 311-336.

Crain, J.L., Wei, Y., Barker, J., Thompson, S.M., Alderman, P.D., Reynolds, M., Zhang, N.Q., Poland, J., 2016. Development and Deployment of a Portable Field Phenotyping Platform. Crop Science 56, 965-975.

Duan, T., Chapman, S.C., Holland, E., Rebetzke, G.J., Guo, Y., Zheng, B., 2016. Dynamic quantification of canopy structure to characterize early plant vigour in wheat genotypes. J Exp Bot 67, 4523-4534.

Fridman, E., Carrari, F., Liu, Y.S., Fernie, A.R., Zamir, D., 2004. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305, 1786-1789.

Furbank, R.T., Tester, M., 2011. Phenomics--technologies to relieve the phenotyping bottleneck. Trends Plant Sci 16, 635-644.

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

Hammer, G., Cooper, M., Tardieu, F., Welch, S., Walsh, B., van Eeuwijk, F., Chapman, S., Podlich, D., 2006. Models for navigating biological complexity in breeding improved crop plants. Trends Plant Sci 11, 587-593.

Hedden, P., 2003. The genes of the Green Revolution. Trends Genet 19, 5-9.

Honsdorf, N., March, T.J., Berger, B., Tester, M., Pillen, K., 2014. High-throughput phenotyping to detect drought tolerance QTL in wild barley introgression lines. PLoS One 9, e97047.

Laurie, C.C., Chasalow, S.D., LeDeaux, J.R., McCarroll, R., Bush, D., Hauge, B., Lai, C., Clark, D., Rocheford, T.R., Dudley, J.W., 2004. The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168, 2141-2155.

Liu, W.X., Leiser, W.L., Maurer, H.P., Li, J.H., Weissmann, S., Hahn, V., Wurschum, T., 2015. Evaluation of genomic approaches for marker-based improvement of lodging tolerance in triticale. Plant Breeding 134, 416-422.

Mei, H.W., Li, Z.K., Shu, Q.Y., Guo, L.B., Wang, Y.P., Yu, X.Q., Ying, C.S., Luo, L.J., 2005. Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. Theor Appl Genet 110, 649-659.

Miller, C.N., Harper, A.L., Trick, M., Werner, P., Waldron, K., Bancroft, I., 2016. Elucidation of the genetic basis of variation for stem strength characteristics in bread wheat by Associative Transcriptomics. BMC Genomics 17, 500.

Mitchell-Olds, T., Willis, J.H., Goldstein, D.B., 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? Nature Reviews Genetics 8, 845.

Montes, J.M., Technow, F., Dhillon, B.S., Mauch, F., Melchinger, A.E., 2011. High-throughput non-destructive biomass determination during early plant development in maize under field conditions. Field Crops Research 121, 268-273.

Moore, C.R., Johnson, L.S., Kwak, I.Y., Livny, M., Broman, K.W., Spalding, E.P., 2013. High-throughput computer vision introduces the time axis to a quantitative trait map of a plant growth response. Genetics 195, 1077-1086.

Morgante, F., Huang, W., Maltecca, C., Mackay, T.F.C., 2018. Effect of genetic architecture on the prediction accuracy of quantitative traits in samples of unrelated individuals. Heredity 120, 500-514.

Neilson, E.H., Edwards, A.M., Blomstedt, C.K., Berger, B., Moller, B.L., Gleadow, R.M., 2015. Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. J Exp Bot 66, 1817-1832.

Peng, J.R., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., Sudhakar, D., Christou, P., Snape, J.W., Gale, M.D., Harberd, N.P., 1999. 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400, 256-261.

Piñera-Chavez, F.J., Berry, P.M., Foulkes, M.J., Molero, G., Reynolds, M.P., 2016. Avoiding lodging in irrigated spring wheat. II. Genetic variation of stem and root structural properties. Field Crops Research 196, 64-74.

Pinthus, M.J., 1974. Lodging in Wheat, Barley, and Oats: The Phenomenon, its Causes, and Preventive Measures. Advances in Agronomy 25, 209-263.

Poland, J., 2015. Breeding-assisted genomics. Curr Opin Plant Biol 24, 119-124.

Reynolds, D., Baret, F., Welcker, C., Bostrom, A., Ball, J., Cellini, F., Lorence, A., Chawade, A., Khafif, M., Noshita, K., Mueller-Linow, M., Zhou, J., Tardieu, F., 2018. What is cost-efficient phenotyping? Optimizing costs for different scenarios. Plant Science.

Samuel, A., Paulino, P.R., Bernd, M.R., 2011. A growth phenotyping pipeline for Arabidopsis thaliana integrating image analysis and rosette area modeling for robust quantification of genotype effects. New Phytologist 191, 895-907.

Sankaran, S., Khot, L.R., Carter, A.H., 2015. Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. Computers and Electronics in Agriculture 118, 372-379.

Segura, V., Vilhjalmsson, B.J., Platt, A., Korte, A., Seren, U., Long, Q., Nordborg, M., 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet 44, 825-830.

Sehgal, D., Autrique, E., Singh, R., Ellis, M., Singh, S., Dreisigacker, S., 2017. Identification of genomic regions for grain yield and yield stability and their epistatic interactions. Scientific Reports 7, 41578.

Shi, Y., Thomasson, J.A., Murray, S.C., Pugh, N.A., Rooney, W.L., Shafian, S., Rajan, N., Rouze, G., Morgan, C.L., Neely, H.L., Rana, A., Bagavathiannan, M.V., Henrickson, J., Bowden, E., Valasek, J., Olsenholler, J., Bishop, M.P., Sheridan, R., Putman, E.B., Popescu, S., Burks, T., Cope, D., Ibrahim, A., McCutchen, B.F., Baltensperger, D.D., Avant, R.V., Jr., Vidrine, M., Yang, C., 2016. Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research. PLoS One 11, e0159781.

Turelli, M., Barton, N., 2004. Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G× E interactions. Genetics 166, 1053-1079.

van Eeuwijk, F., Bustos-Korts, D., Millet, E.J., Boer, M., Kruijer, W., Thompson, A., Malosetti, M., Iwata, H., Quiroz, R., Kuppe, C., Muller, O., Blazakis, K.N., Yu, K., Tardieu, F., Chapman, S., 2018. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. Plant Science.

van Eeuwijk, F.A., Bustos-Korts, D.V., Malosetti, M., 2016. What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype × Environment Interactions? Crop Science 56, 2119-2140.

Verma, V., Worland, A.J., Sayers, E.J., Fish, L., Caligari, P.D.S., Snape, J.W., 2005. Identification and characterization of quantitative trait loci related to lodging resistance and associated traits in bread wheat. Plant Breeding 124, 234-241.

Watanabe, K., Guo, W., Arai, K., Takanashi, H., Kajiya-Kanegae, H., Kobayashi, M., Yano, K., Tokunaga, T., Fujiwara, T., Tsutsumi, N., Iwata, H., 2017. High-Throughput Phenotyping of Sorghum Plant Height Using an Unmanned Aerial Vehicle and Its Application to Genomic Prediction Modeling. Frontiers in Plant Science 8.

White, J.W., Andrade-Sanchez, P., Gore, M.A., Bronson, K.F., Coffelt, T.A., Conley, M.M., Feldmann, K.A., French, A.N., Heun, J.T., Hunsaker, D.J., Jenks, M.A., Kimball, B.A., Roth, R.L., Strand, R.J., Thorp, K.R., Wall, G.W., Wang, G.Y., 2012. Field-based phenomics for plant genetics research. Field Crops Research 133, 101-112.

Xavier, A., Hall, B., Hearst, A.A., Cherkauer, K.A., Rainey, K.M., 2017. Genetic Architecture of Phenomic-Enabled Canopy Coverage in Glycine max. Genetics 206, 1081-1089.

Yang, M.D., Huang, K.S., Kuo, Y.H., Tsai, H.P., Lin, L.M., 2017. Spatial and Spectral Hybrid Image Classification for Rice Lodging Assessment through UAV Imagery. Remote Sensing 9, 583.

Chapter 2 - Phenomics Enabled Genetic Dissection of Crop Lodging in Wheat

This chapter is under review as following journal article:

Singh, D., Wang, X., Kumar, U., Gao, L., Noor, M., Imtiaz, M., Singh, R.P., Poland, J. High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. Frontiers in Plant Science.

Abbreviations

BLUE, best linear unbiased estimate; CIMMYT, International Wheat and Maize Improvement Center; DEM; digital elevation model; DLmean, digital lodging mean; DLmix, digital lodging mixture; FAS, Faisalabad; GBS, genotyping-by-sequencing; GCP, ground control point; GNSS, global navigation satellite systems; GWAS, genome-wide association studies; HTP, high-throughput phenotyping; LDH, Ludhiana; QTL, quantitative trait loci; RR-BLUP, ridge regression best linear unbiased predictors; RKHS, reproducing kernel Hilbert space; SNP, single nucleotide polymorphism ; UAS, unmanned aerial systems.

Abstract

Novel high-throughput phenotyping (HTP) approaches are needed to advance the understanding of genotype-to-phenotype and accelerate plant breeding. The first generation of HTP has examined simple spectral reflectance traits from images and sensors but is limited in advancing our understanding of crop development and plant architecture. Lodging is a complex trait that significantly impacts yield and quality in many crops including wheat. Conventional visual assessment methods for lodging are time-consuming, relatively low-throughput, and subjective, limiting phenotyping accuracy and population sizes in breeding and genetics studies. Here we demonstrate the considerable power of unmanned aerial systems (UAS) or drone-based

phenotyping as a high-throughput alternative to visual assessments for the complex phenological trait of lodging, which significantly impacts yield and quality in many crops including wheat. We tested and validated quantitative assessment of lodging on 2,640 wheat breeding plots over the course of two years using differential digital elevation models from high-throughput imaging from UAS. The wheat lines were part of the International Wheat and Maize Improvement Center's (CIMMYT) advanced elite yield nurseries in South Asia. High correlations of digital measures of lodging to visual estimates (r = 0.76-0.93) and equivalent broad-sense heritability demonstrate this approach is amenable for reproducible assessment of lodging in large breeding nurseries. Using these high-throughput measures to assess the underlying genetic architecture of lodging in wheat, we applied genome-wide association analysis and identified a key genomic region on chromosome 2A, consistent across digital and visual scores of lodging. However, these associations accounted for a very minor portion of the total phenotypic variance. We therefore investigated whole genome prediction models and found high prediction accuracies across populations and environments. This adequately accounted for the diffuse architecture of numerous small effect loci, consistent with the previously described complex genetic architecture of lodging in wheat. Our study provides a proof-of-concept application of UAS-based phenomics that is scalable to tens-of-thousands of plots in breeding and genetic studies as will be needed to uncover the genetic factors and increase the rate of gain for complex traits in crop breeding.

Key words: *Triticum aestivum*, GWAS, genomic selection, high-throughput phenotyping, lodging, UAS, unmanned aerial systems, wheat breeding

Introduction

A deeper understanding of the biological processes mediated by plant genomes is needed to develop crops with improved stress resilience and yield potential. Connecting genotype to phenotype for quantitative plant traits on a genome level necessitates high-density genetic markers and large population sizes to gain sufficient power and resolution. While the recent advancements in sequencing technologies have provided almost unlimited access to high-density genetic markers, large-scale rapid and accurate phenotyping of complex traits remains a major constraint (Furbank and Tester, 2011). High-throughput phenotyping (HTP) tools with improved spatial and temporal resolution can help address this phenotyping bottleneck (Furbank and Tester, 2011; White et al., 2012).

Several HTP platforms including greenhouse, ground-based, and aerial systems have been demonstrated for crops (Andrade-Sanchez et al., 2014; Honsdorf et al., 2014; Crain et al., 2016), such as enabling the dissection of stress and growth traits in controlled conditions (Chen et al., 2014; Campbell et al., 2015; McCormick et al., 2016). For targeting the scope of field experiments, UAS offer a flexible alternative to ground-based phenotyping platforms, particularly for large-breeding nurseries and genetic studies with thousands or tens of thousands of plots (Poland, 2015). Recently, UAS have been deployed in HTP of wheat breeding nurseries (Sankaran et al., 2015; Haghighattalab et al., 2016), expanding previous work using multi-rotor UAS of varying sizes and payload capacity to phenotype small-sized test plots (Bendig et al., 2014; Chapman et al., 2014; Shi et al., 2016). With rapid development of low-cost consumer-grade sensors and platforms, UAS phenotyping holds great potential to be an integral part of plant genomics and breeding for precise, quantitative assessment of otherwise low-throughput and complex traits on large populations. However, significant developments in processing, methodology and analysis of UAS-derived data are needed to realize its full potential.

Lodging, the permanent displacement of the plant stem from vertical position, is an example of a complex trait that is difficult to quantify in the field (Pinthus, 1974). Conventional phenotyping methods for lodging are based on visual ratings of incidence and severity scores, and can be associated with stem or root lodging (Berry et al., 2003; Piñera-Chavez et al., 2016). A lodging-resistant ideotype for wheat has been described as a strong root system, wider root plates, larger stem diameter, and moderate to short height (Berry et al., 2007). The physiology of lodging in wheat is associated with a complex genetic architecture (Verma et al., 2005; Liu et al., 2015; Miller et al., 2016). Only a few small to moderate effect quantitative trait loci (QTL) explaining 2-27% variation for lodging and stem strength have been identified (Keller et al., 1999; Hai et al., 2005; Berry and Berry, 2015). As such, the complex genetic architecture of lodging is a good target for field-based high-throughput phenotyping to enable precise measurement of very large populations needed for genomics studies and breeding progress.

Image-derived lodging assessments have been proposed in wheat, maize, and rice (Chapman et al., 2014; Chu et al., 2017; Yang et al., 2017). Albeit on a limited number of plots and without quantitative assessment or ground-truth validations, a proof-of-concept study by Chapman et al. (2014) demonstrated the possibility to assess the presence of lodging with UAS. An image-based lodging assessment was validated relative to visual scores of lodging on 288 maize plots (Chu et al., 2017). Thus, with strong proof of concept and scalable potential, UAS assessment of lodging phenotypes in large wheat breeding nurseries has potential to transform throughput, and hence the power, for genetic studies and breeding programs.

Here, we demonstrate novel field-based lodging assessment approaches using a commercially available light-weight UAS. By developing multiple time-points of three-dimensional digital elevation models (DEM) from UAS-acquired stereo imaging, we quantified lodging in 2640 wheat breeding plots with high correlation to visual scores and comparable

repeatability. Using these precise phenotypic measurements, we identified genomic regions associated with lodging in wheat from a genome-wide association analysis. The limited genetic variation explained by the genome-wide associations led us to test whole-genome prediction models which accounted for a much larger portion of the heritable variation and supported the need for large, precisely measured populations to understand the functional genomics of lodging. Here we report an original application of UAS for large-scale, high-throughput assessment of complex plant architecture and physiology in breeding and genomic studies with evaluation of lodging in wheat. This highly reproducible approach is scalable to tens-of-thousands of plots or even individual plants of different crops to rapidly quantify plant height, lodging, and could potentially be extended to traits like growth rate.

Methods

Plant Material and Field Layout

Advanced spring wheat (*Triticum aestivum* L.) breeding lines from CIMMYT's South Asia Bread Wheat Genomic Prediction Yield Trials were sown in the first week of November (4 Nov 2015 and 7 Nov 2016) at the Borlaug Institute for South Asia's Ludhiana (LDH), Punjab, India (30°59' N and 75°44' E) location during seasons 2016 and 2017. A total of 590 and 595 unique wheat entries along with the check varieties (consisting of early and late flowering, locally adaptive high yielding and disease resistance varieties) were planted in alpha-lattice field design during seasons 2016 and 2017, respectively. Entries in each year were divided into 11 trials with each trial containing 53 entries and 7 checks laid out in two complete replicate blocks of 120 plots per trial. Each replicate block was divided into six subblocks of 10 plots each. The experimental unit was an individual six-row plot with dimensions 1.3m × 3.8m. Plot-to-plot spacing was 80 cm and 52 cm between rows and columns, respectively. Sixty best entries from season 2016 were repeated in season 2017 as an additional trial. A similar experiment was

replicated in Faisalabad (FAS) in Pakistan (31°24' N and 73°02' E), except the 11th trial was excluded in this location. The experimental location in LDH is situated in the north-western wheat growing belt of India. LDH and FAS environments are classified as irrigated mega-environments (ME1) according to CIMMYT's wheat breeding mega-environment classification system (Rajaram et al., 1995). Field trials were managed following the established standard agronomic practices at each location.

UAS and Sensor Specifications

Two different UAS quadcopters were deployed for data acquisition during two seasons at LDH. In 2016 season, an IRIS+ quadcopter (3DR Robotics Inc., CA) equipped with a 3-channel Canon S100 digital camera (Canon, USA) was used to collect data over the wheat plots. In 2017, the UAS platform was upgraded to a high payload carrying capacity quadcopter DJI Matrice100 (DJI, USA) carrying a 5-channel multispectral RedEdge camera (MicaSense Inc., USA). A detailed list of UAS and sensor specifications is provided (Table 1).

UAS-based Image Acquisition

Each year the semi-autonomous UAS flights were conducted between 11AM to 2PM. Data acquisition followed the standard operating procedures developed within the Poland Lab at Kansas State University (Wang et al., 2018). In each of the years, the field trials experienced natural lodging from the combination of heavy rain and wind during the grain-filling stage. A total of four UAS flights were made on dates 01-Mar-2016, 16-Mar-2016, 02-Mar-2017, and 15-Mar-2017. The flights were conducted pre- and post-lodging events in each year and the flight dates corresponded to pre- and post-lodging events, respectively. The flight plans were created using Mission Planner desktop application for Windows (http://ardupilot.org/planner/) for IRIS+ UAS, and Litchi Android App (VC Technology Ltd.) and CSIRO mission planner application (https://croptsrv-cdc.it.csiro.au/shiny/users/zhe00a/missionplanner/; accessed Oct 2, 2018) for

DJI Matrice100. All flights were made at a ground altitude of 25 meters in 2016 and 2017. In both years, the image overlap rate between two geospatially adjacent images was set to 80% sequentially and 78% laterally to ensure optimal orthomosaic photo stitching quality. Accordingly, the flight speed, the flight elevation above the ground, and the width between two parallel flight paths were adjusted based on the overlap rate and the camera field of view. Both cameras were automatically triggered with the onboard GPS following a constant interval of distance travelled.

To ensure highly accurate digital elevation maps, the UAS images were geo-referenced and geo-rectified using 12 white colored ground control points (GCPs) that were uniformly distributed over the 1.5 ha field area. These GCPs were surveyed using a SXBlue III-L differential Global Navigation Satellite System (GNSS) unit (Geneq Inc., Montreal, Canada) and Precis BX305 Real Time Kinematics GNSS unit (Tersus GNSS Inc., Shanghai, China) in 2016 and 2017, respectively. To preserve the image pixel intensity, the Canon S100 camera was set to capture raw images, while the MicaSense RedEdge camera was set to capture uncompressed TIFF images.

Digital Elevation Model (DEM) Generation

Raw images captured by Canon S100 camera were imported to Canon Digital Photo Professional Software (Canon, USA) for lens distortion correction and converted to 16-bit TIFF images. Lens distortion corrections were not required for images captured by MicaSense RedEdge camera (https://micasense.github.io/imageprocessing/). After preprocessing, images of both cameras were processed in Agisoft PhotoScan Pro (Version 1.3.1, Agisoft LLC, Russia) following the internally established protocols in Poland Lab. In the first step of image alignment, the settings were: key-point limit 15000 (MicaSense) and 80000 (Canon) points, reference pre-selection, accuracy high, tie-point limit 0, and adaptive camera model. A sparse point cloud of

the entire field area was stitched through the process of image alignment. In the subsequent step, the GCP coordinates were assigned to the individual images where white-colored, square-shaped GCPs were visible on the ground. Our tests with Photoscan program suggested that at least three images per GCP are required to accurately geo-rectify and geo-reference the orthomosaics. The GCP assignment step was followed by the camera optimization step that adjusted the estimated point coordinates and camera parameters in the model. Based on the optimized camera positions, a dense point cloud model of the entire field was generated by setting the parameters to high quality and moderate depth. Finally, the DEM was built from the dense point cloud model. The detailed processing reports are available on the project data repository (Singh et al., 2018a). Each pixel in this DEM had three attributes namely latitude, longitude and height. These three attributes corresponded to geo-position and height of DEM points, and were used to calculate plot-level height/lodging information. A total of four DEM with two DEM each season corresponding to pre- and post-lodging, respectively, were processed.

Lodging Assessment

In the present study, we describe two crop lodging assessment algorithms based on image-derived DEM. Both approaches used a differential DEM model that was generated by subtracting the post-lodging DEM from pre-lodging DEM each year (**Figure 2.1- 2.2**). In the first method, a simple arithmetic mean (DLmean; Digital Lodging mean) of the differential digital elevation pixels for each plot was calculated. In the second method, a two-component normal mixture distribution with parameters μ_2 and λ_2 was estimated through an iterative Expectation Maximization algorithm in *R* package *mixtools* (Benaglia et al., 2009) by constraining the first mean parameter μ_1 to zero. The parameters μ_2 and λ_2 correspond to the mean and proportion of the lodged DEM pixels, respectively, and were combined to create a mixture lodging index of the digital lodging (henceforth, Digital Lodging Mixture, DLmix; $\mu_2 \times$

λ_2). Additionally, visual assessment of lodging was carried out post-lodging events in both seasons. Visual scores included the lodging intensity (LOI; percent plot area lodged; 0-100%), severity (LOS; angle of plant lodging; 0-10), and a combined lodging index (LI; LOI × LOS). Additional supporting phenotypic data included agronomic and phenological traits. A detailed trait ontological description is provided (**Supplementary Table B.1**).

Statistical Data Analysis

The phenotypic data on lodging included: i) three visual scores of lodging namely intensity (LOI), severity (LOS) and lodging index (LI) per plot and additional supporting agronomic measurements per plot; ii) two digital lodging scores obtained by taking overall summary mean per plot (DLmean) or combined lodging index of normal mixture parameters (DLmix). The variance components for broad-sense line mean heritability or repeatability for each trait and trial were calculated using *lme4* package (Bates et al., 2014) in R with the following model:

$$y_{ikl} = \mu + G_i + M_{k(l)} + e_{ikl} \quad (1)$$

where y_{ikl} is the phenotypic response variable, μ is the fixed overall mean, G_i is the random genotype effect, $M_{k(l)}$ is the random effect of sub-blocks nested within a replicate, and e_{ikl} is the residual. The variance components derived from the model were used to calculate broad-sense heritability on entry-mean basis for each trait:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_e^2/r} \quad (2)$$

where σ^2_G is the genotypic variance, σ^2_e is the residual variance, and r is the number of replicates. Genotypic best linear unbiased estimates (BLUEs) were calculated as follows:

$$y_{ijkl} = \mu + G_i + Z_j + R_{k(j)} + M_{kl(j)} + e_{ijkl} \quad (3)$$

where y_{ijkl} is the phenotypic response variable, μ is the fixed overall mean, G_i is the fixed genotype effect, Z_j is the random trial effect, $R_{k(j)}$ is the random effect of replicate nested within a trial, $M_{kl(j)}$ is the random effect of sub-blocks and replications nested within a trial, and e_{ijkl} is the residual. The marker based genetic correlations between each pair of traits were calculated with *sommer* (Covarrubias-Pazaran, 2016) package in R as follows:

$$r_{g(x,y)} = \frac{cov_g(x,y)}{\sqrt{var_g(x) * var_g(y)}} \quad (4)$$

where $cov_g(x,y)$ is the covariance of the trait pairs; $var_g(x)$ and $var_g(y)$ is the variance of traits.

Genotyping

All 1185 lines from both seasons were profiled using the genotyping-by-sequencing protocol of Poland et al. (2012b) and sequenced on an Illumina Hi Seq2000 or HISeq2500. Single nucleotide polymorphism (SNP) markers were called with TASSEL *v*5 pipeline (Glaubitz et al., 2014b) and aligned to the reference Chinese Spring Wheat Assembly v1.0 (Consortium, 2014). Genotyping calls were extracted and filtered so that the percent missing data per marker was less than 40% and percent heterozygosity was less than 10%. Lines with more than 50% missing data were removed. After filtering, a total of 10,878 SNP markers were retained and missing data were imputed with Beagle *v*4.1 (Browning and Browning, 2016). Another filtration step was applied on imputed SNPs to remove heterozygous calls. In addition, we built a bioinformatics pipeline to predict the presence or absence of the 2NS segment based on genotyping-by-sequencing. Briefly, wheat and alien specific tags were identified using a training set of cultivars or lines that are known to be 2NS positive and negative. The presence or absence of the 2NS segment was predicted based on relative counts of wheat or alien specific tags. A custom R function that takes input of alien or wheat specific tags and tags by taxa file through TASSEL pipeline was used to predict the presence or absence of the 2NS segments. The method

was validated using a wet lab method (Ventriup-LN2) and proved to be highly accurate (>99%) (Liangliang Gao, personal communication).

Genome-wide Association Study (GWAS)

A GWAS and genomic prediction analyses were performed on 10,166 SNPs scored on 590 and 595 lines from cropping seasons 2016 and 2017, respectively. A combined GWAS analysis on 1035 genotypes from both years was also performed with year as a fixed effect. A two-step adjusted means model with genotypes and year as fixed terms was used to generate BLUEs for the 1035 lines for all lodging measurements as following:

$$y_{ij} = G_i + E_j + e_{ij} \quad (5)$$

where y_{ij} is the phenotypic value, G_i is the fixed genotype effect, E_j is the environment fixed effect, and e_{ij} is the residual error. The resulting adjusted means were used as response variable in the combined association analysis. The following mixed model genome-wide association analysis (Yu et al., 2006) was implemented in *rrBLUP* package in *R* (Endelman, 2011):

$$y = X\beta + Zu + S\tau + e \quad (6)$$

where y is the vector of adjusted means (BLUEs) of phenotypes, β is the vector of fixed terms (intercept and principal component-based population structure covariates), u is vector of genetic marker effects, and $\boldsymbol{\tau}$ is the random effect of genetic background of each line; and X, Z, and S are the respective design matrices. For each trait, a genome-wide false discovery rate threshold was calculated based on the QVALUE function in *R* (Storey and Tibshirani, 2003).

Genomic Prediction and Cross-validation

To test for the diffused genetic architecture of lodging in wheat, we generated *k*-fold based genomic predictions. To reduce the prediction bias resulting from training and testing sets similarities, 11-fold training-testing composition was chosen based on total number of trials in

the experiment. Two linear parametric methods of genomic predictions, ridge regression BLUP (RR-BLUP (Endelman, 2011)) and Bayes Cπ (Perez and de los Campos, 2014), and a non-linear method, Reproducing Kernel Hilbert Space (RKHS (Gianola et al., 2006)), were used to calculate genomic estimated breeding values for each trait. While RR-BLUP assumes an infinite number of loci with infinitesimally small effects, the Bayes Cπ is a variable selection method that allows for a proportion of marker effects to be set to zero, assuming a common non-zero variance for rest of the marker effects (Legarra et al., 2008; Habier et al., 2013). RKHS is a kernel-based regression method. Two different empirical cross-validation schemes were tested for prediction models: (1) predictions across location on same genetic material (16LDH-16FAS, 17LDH-17FAS); (2) predictions across years on independent set of lines (16LDH-17LDH, 17LDH-16LDH, 16LDH-17FAS, 17LDH-16FAS). In the case of LDH-FAS training-testing combinations, each of the five lodging measures (LOI, LOS, LI, DLmean and DLmix) was used in the training set at LDH to predict the LOI at FAS. The genomic prediction models were implemented in *R* packages BGLR and rrBLUP (Endelman, 2011; Perez and de los Campos, 2014). We used 10000 iterations and 3000 burn-ins for Bayes Cπ and RKHS models in BGLR.

Data and Code Availability

All data associated with the experiments including raw images, orthomosaics, polygons, etc. can be accessed at the public repository (https://doi.org/10.6084/m9.figshare.6151127). Analysis scripts are available at github/singhdj2/digital_lodging.

Results and Discussion

High Throughput Phenotyping of Wheat Breeding Trials

To assess yield potential and agronomic performance of elite breeding lines as part of the International Wheat and Maize Research Center's (CIMMYT) breeding efforts, wheat trials were

established at Ludhiana (LDH) in NW India and Faisalabad (FAS) in central Pakistan in 2016 and 2017 (**Table 2.A**). Throughout the growing season, autonomous phenotyping operations were conducted at LDH with a GPS-guided UAS equipped with modified-RGB and multi-spectral digital cameras. Mid-day flight missions covered an area of 1.5 ha containing the entire trial of over 1,320 plots in ~25 minutes and obtained a ground spatial resolution of 1.5 to 3.5 cm per pixel. In each of the years, the trials experienced natural lodging from the combination of heavy rain and wind during the grain-filling stage. Following these lodging events, lodging incidence and severity was visually scored as a 'ground-truth' for subsequent validation of the image-derived lodging values (**Figure 2.2 and Table 1.A**). The breeding lines in the trials showed considerable phenotypic variation with 0 to 100% lodging severity and incidence, and moderate broad-sense heritability (H^2) of 0.50 to 0.66 for the lodging incidence (**Figure 2.3**).

Extraction of Image-derived Digital Lodging

To quantitatively assess the amount of lodging from UAS collected images across 1,320 field plots in each of the two years, we built digital elevation models (DEM) for the crop before and after the lodging events. A differential DEM was generated by subtracting the post- from pre-lodging DEM giving the overall elevation change between the two time-points (**Figure 2.2**). We observed large elevation changes that were commensurate with severely lodged plots. To derive a quantitative measurement of lodging, we first calculated a simple arithmetic mean (henceforth, Digital Lodging mean: DLmean) of all differential DEM height points falling under the area of each plot polygon. This measure of lodging was phenotypically and genetically well-correlated to the visual scores of incidence, severity, and lodging index (r_{pheno} = 0.77-0.93; r_{geno} = 0.93-0.96; $P < 0.001$; **Figure 2.3** and **Supplementary Table B.3**).

Following on the simple mean difference we applied a more informed normal mixture model of the differential DEM pixel distributions. A combined mixture lodging index (DLmix)

of digital lodging was derived from the mixture model parameters and compared with the visual scores. The ground-truth validation of the mixture model again showed high phenotypic and genetic correlations to the visual scores (r_{pheno} = 0.76-0.91; r_{geno} = 0.93-0.97; $P < 0.001$; **Figure 2.3** and **Supplementary Table B.3**).

As an additional measure of accuracy of the visual scores and digital image-based estimations of lodging, we calculated the broad-sense heritability (H^2), or repeatability, on an entry-mean basis for each of eleven trials in both years (**Figure 2.3**). The repeatability of visual scores, DLmean, and DLmix was in the range of 0.5 to 0.7, consistent with previous studies on wheat and sorghum that reported similar heritability for lodging (Liu et al., 2015; Piñera-Chavez et al., 2016; Yu et al., 2016). Digital lodging outperformed visual scores in terms of heritability for 8 out of 11 trials in year 2017 (**Supplementary Table B.2**). Digital and visual measures of lodging were genetically highly correlated in both years ($r_{geno} > 0.93$; **Supplementary Table B.3**), suggesting they are capturing the same variance and supporting the effectiveness of image-based lodging assessment for HTP within large wheat breeding and genetic studies.

Genome-wide Association Analysis of Lodging

To assess the genetic architecture of lodging using the validated digital image-based estimations, we conducted a genome-wide association analysis on 1,185 (590 in 2016 and 595 in 2017) elite wheat breeding lines. Genome profiling was performed with genotyping-by-sequencing and markers were fitted in a linear mixed model with terms to account for population structure and cryptic relationships (Equation 4). Despite having a relatively large population size (n=590), and moderate to high heritability, no significant SNPs were identified for any lodging measure in 2016 (**Supplementary Fig. A.1**). For the 2017 field trial, an association peak on chromosome 2A was observed for visual and digital scores of lodging (**Supplementary Fig. A.2**). To leverage the power of a larger population size, association analysis on combined data

from two years was performed. The association test showed a highly significant and consistent peak at chromosome 2A (**Figure 2.4**). The markers on 2A coincided with a region corresponding to the 2NS *Aegilops ventricosa-Triticum aestivum* translocation (Doussinault et al., 1983; Gao, 2018). We investigated lines positive for 2NS translocation, which showed reduction in lodging incidence for both visual and digital lodging measures (**Figure 2.5**; P = 0.049-0.002, n = 1010*; t-test*). A survey for the 2NS fragment in our material suggested that more than 75% lines carry this translocation fragment, which is known to harbor multiple disease resistance genes in wheat (Jahier et al., 2001; Helguera et al., 2003; Williamson et al., 2013; Cruz et al., 2016). However, reports on its impact on lodging are lacking. Interestingly, the significant markers from association analysis only explained up to 2% of genetic variation for lodging and the majority of the markers were below the significance threshold. Overall, these results point towards a complex genetic architecture for lodging in wheat and 'hidden heritability' like the classic example of human height (Gudbjartsson et al., 2008), where a large number of significant loci identified via GWAS explained only a small proportion of the total additive genetic variation for height.

Genome-wide Predictions and Cross-validations

To address the postulate of hidden heritability due to a diffuse genetic architecture of numerous small-effect QTL, we generated genome-wide predictions for digital and visual measures of lodging. Three different genomic prediction models (BayesCπ, RKHS, and RR-BLUP) were compared to account for complex genetic architecture of lodging in wheat. As all three models yielded comparable results (**Supplementary Table A.4**), only RR-BLUP based predictions are discussed. To assess the proportion of genetic variance captured using the whole-genome models, we calculated k-fold cross validation prediction accuracy within each year. Cross-validations within the environment were able to explain up to 53% of the heritable genetic

effects (**Table 3.C**). Finding that the whole genome models accounted for more than half of the heritable genetic effects for lodging, we further validated this observation by comparing the phenotypic and genetic correlation to the prediction accuracy for lodging measured in a second environment location of Faisalabad, Pakistan (FAS) (**Supplementary Table A.5**). In majority of the cases (7 of 10) the genomic prediction accuracies were equal or higher than the phenotypic correlation. In 2017 the whole-genome prediction model had a predictive correlation of 0.45, accounting for 55% of the heritable genetic effects for DLmix. Commensurate with application of genomic prediction in a breeding program with confounding environmental effects, the prediction accuracies across environments were lower but still captured heritable variance with predictions in the range of 0.19 to 0.55 (**Table 2.D**). In contrast to the lack of power to find individual genetic associations, we were able to capture a substantial portion of the heritable genetic effects using whole-genome models that account for many small effect QTL and support the hypothesis of a diffuse genetic architecture for lodging in elite wheat germplasm (Kooke et al., 2016). Furthermore, we support the observation that much larger populations must be evaluated to uncover the genetic basis and identify causative variants for lodging.

Relationship of Lodging to Phenology and Agronomic Traits

Finally, to investigate the relationship of different measures of lodging with plant developmental and agronomic traits, we calculated pairwise correlations of lodging with different traits within each environment. Consistent with previous reports in wheat (Keller et al., 1999; Berry and Berry, 2015), we found that taller plants ($r = 0.12$, $P < 0.01$) with early heading ($r = -0.15$, $P < 0.001$) tend to have more lodging while thousand grain weight was negatively associated with lodging (**Figure 2.6**). A positive relationship of lodging and plant biomass traits such as ground coverage, plant stand, and spike length also highlights the vulnerability of high yielding, high biomass cultivars to the crop lodging. Inconsistent trends, however, in the

relationship between grain yield and lodging were observed with a negative correlation in 2016 and a positive correlation in 2017. This suggests that on occasion higher grain weight on the maturing spike can weigh the plants down and increase lodging, while in other conditions the lodging can occur at a stage that will prevent completion of grain filling leading to yield loss. Under preferred mechanical harvesting operations in farmers' fields, either of these situations will lead to economic loss of harvestable yield and decreased quality. These trends suggest the need for developing cultivars with better stem strength characteristics to mitigate lodging-associated losses.

Implementation of the Proposed Methodology in Field Experiments

The present study implemented highly scalable approach to measure complex agronomic trait of lodging in the field experiments. Data collection component of this phenotyping approach includes UAS setup (i.e. mission generation and upload, calibration info collection) and aerial image acquisition. The flight time depends on the field scale, UAS flying elevation, moving velocity, and the overlapping rate between the successive aerial images. Data processing component includes DEM generation by photogrammetry and plot-level data extraction, and the processing time depend on the data volume and the computer hardware settings. For the implementation stand-point, this work can be replicated with an initial investment of USD 12000, which will cover the cost of sensor hardware (USD 2000), UAS platform (USD 5000), a high-precision GNSS (USD 2000), and the computer software and hardware (USD 3000). A practical implementation of our lodging assessment approach in the field would require a careful monitoring of weather and crop growth conditions. Furthermore, as the operational costs and scale of breeding programs grow in future, the cost-effective and high-throughput tools that can provide multiple layers of data at a fraction of cost would be highly desired. Therefore, a full benefit of our proposed methodology can be realized by integrating it with the routine

application of UAS-based sensor measurements in research programs. As such, this lodging assessment approach can provide an additional data layer on top of the routine phenotypic measurements (e.g., spectral, morphological, physiological) without incurring any extra cost and time effort to the researchers.

Conclusions

UAS-enabled phenotyping allowed us to quantify lodging on 2,640 wheat plots. Using validated digital lodging measurements along with association and genomic prediction analyses, we provide evidence in support of a diffuse genetic architecture of lodging in wheat. Our findings have diverse applications in plant breeding and genetics. First, our highly reproducible UAS based digital lodging methods can be easily scaled and also applied to different crops to rapidly quantify plant height, lodging, and should be extensible to traits like growth rate on large populations. Second, for a complex and quantitatively controlled trait like lodging, whole-genome predictions can account for large portion of variation not captured by regular GWAS. Undoubtedly, accurate phenotypic assessment is a critical prerequisite for breeding for lodging resilience, and as shown here, UAS-enabled large-scale quantitative assessment of lodging can be a powerful approach to identify genetic variants for lodging. This comprehensive evaluation of lodging assessment methods lays the foundation for improving our understanding of biological underpinnings of lodging in wheat and other crops. Overall this highlights the future of modern breeding where, in conjunction with powerful genomics and informatics tools, UAS-enabled phenotyping can accelerate the genetic gains in plant breeding to meet the global demand for food, fiber and fuel.

Author Contributions

JP and DS conceived and designed the study; DS collected and analyzed UAS and ground-truth data; DS and XW performed image analysis; LG contributed alien-fragment data;

UK supervised field experiments and collected data in India; MN and MI supervised field experiments and collected ground-truth data in Pakistan; RS provided experimental lines; DS and JP wrote the manuscript; JP directed the overall project; all authors edited and reviewed the manuscript.

Acknowledgements

We thank the support staff at Borlaug Institute for South Asia in India including Yogesh Gautam, Manish Kumar, Avadhesh Kumar; CIMMYT Pakistan Majid Nadeem and Dr. Makhdoom Hussain; and CIMMYT Mexico Dr. Suchismita Mondal. We appreciate the assistance of colleagues at Kansas State University including Shuangye Wu, Mark Lucas, Richard Brown, Haley Ahlers, Jared Crain, and Byron Evers; Scott Chapman, CSIRO, Australia. This work was supported by the National Science Foundation (NSF) Plant Genome Research Program (PGRP) (Grant No. IOS-1238187), the Kansas Wheat Commission and Kansas Wheat Alliance, the US Agency for International Development (USAID) Feed the Future Innovation Lab for Applied Wheat Genomics (Cooperative Agreement No. AID-OAA-A-13-00051), and by the NIFA International Wheat Yield Partnership (Grant No. 2017-67007-25933/project accession no. 1011391) from the USDA National Institute of Food and Agriculture. The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of the U.S. Agency for International Development, the U.S. National Science Foundation, or the U.S. Department of Agriculture.

References

Andrade-Sanchez, P., Gore, M.A., Heun, J.T., Thorp, K.R., Carmo-Silva, A.E., French, A.N., Salvucci, M.E., White, J.W., 2014. Development and evaluation of a field-based high-throughput phenotyping platform. Functional Plant Biology 41, 68-79.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1, 1-23.

Benaglia, T., Chauveau, D., Hunter, D.R., Young, D.S., 2009. mixtools: An R Package for Analyzing Finite Mixture Models. Journal of Statistical Software 32, 1-29.

Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., Bareth, G., 2014. Estimating Biomass of Barley Using Crop Surface Models (CSMs) Derived from UAV-Based RGB Imaging. Remote Sensing 6, 10395-10412.

Berry, P.M., Berry, S.T., 2015. Understanding the genetic control of lodging-associated plant characters in winter wheat (Triticum aestivum L.). Euphytica 205, 671-689.

Berry, P.M., Spink, J., Sterling, M., Pickett, A.A., 2003. Methods for rapidly measuring the lodging resistance of wheat cultivars. Journal of Agronomy and Crop Science 189, 390-401.

Berry, P.M., Sylvester-Bradley, R., Berry, S., 2007. Ideotype design for lodging-resistant wheat. Euphytica 154, 165-179.

Browning, B.L., Browning, S.R., 2016. Genotype Imputation with Millions of Reference Samples. Am J Hum Genet 98, 116-126.

Campbell, M.T., Knecht, A.C., Berger, B., Brien, C.J., Wang, D., Walia, H., 2015. Integrating Image-Based Phenomics and Association Analysis to Dissect the Genetic Architecture of Temporal Salinity Responses in Rice. Plant Physiol 168, 1476-1489.

Chapman, S., Merz, T., Chan, A., Jackway, P., Hrabar, S., Dreccer, M., Holland, E., Zheng, B., Ling, T., Jimenez-Berni, J., 2014. Pheno-Copter: A Low-Altitude, Autonomous Remote-Sensing Robotic Helicopter for High-Throughput Field-Based Phenotyping. Agronomy 4, 279-301.

Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., Klukas, C., 2014. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. Plant Cell 26, 4636-4655.

Chu, T.X., Starek, M.J., Brewer, M.J., Murray, S.C., Pruter, L.S., 2017. Assessing Lodging Severity over an Experimental Maize (Zea mays L.) Field Using UAS Images. Remote Sensing 9, 923.

Consortium, I.W.G.S., 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345, 1251788.

Covarrubias-Pazaran, G., 2016. Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. PLoS One 11, e0156744.

Crain, J.L., Wei, Y., Barker, J., Thompson, S.M., Alderman, P.D., Reynolds, M., Zhang, N.Q., Poland, J., 2016. Development and Deployment of a Portable Field Phenotyping Platform. Crop Science 56, 965-975.

Cruz, C.D., Peterson, G.L., Bockus, W.W., Kankanala, P., Dubcovsky, J., Jordan, K.W., Akhunov, E., Chumley, F., Baldelomar, F.D., Valent, B., 2016. The 2NS Translocation from Aegilops ventricosa Confers Resistance to the Triticum Pathotype of Magnaporthe oryzae. Crop Sci 56, 990-1000.

Doussinault, G., Delibes, A., Sanchezmonge, R., Garciaolmedo, F., 1983. Transfer of a Dominant Gene for Resistance to Eyespot Disease from a Wild Grass to Hexaploid Wheat. Nature 303, 698-700.

Endelman, J.B., 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome 4, 250-255.

Furbank, R.T., Tester, M., 2011. Phenomics--technologies to relieve the phenotyping bottleneck. Trends Plant Sci 16, 635-644.

Gao, L., Dorn, K., Rife, T. W., Wang, X., Lemes, C., Clinesmith, M., Silva, P., Fritz, A., Stein, N., Mascher, M. and Poland, J., 2018. Completion of the 'Jagger' Wheat Genome Leads to Identification of Aegilops ventricosa 2NS Translocation and Its Impact in Wheat Breeding. . Plant and Animal Genome Conference XXVI, San Diego, CA.

Gianola, D., Fernando, R.L., Stella, A., 2006. Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173, 1761-1776.

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

Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G., Stefansson, H., Halldorsson, B.V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S., Helgadottir, A., Ingason, A., Steinthorsdottir, V., Olafsdottir, E.J., Olafsdottir, G.H., Jonsson, T., Borch-Johnsen, K., Hansen, T., Andersen, G., Jorgensen, T., Pedersen, O., Aben, K.K., Witjes, J.A., Swinkels, D.W., den Heijer, M., Franke, B., Verbeek, A.L., Becker, D.M., Yanek, L.R., Becker, L.C., Tryggvadottir, L., Rafnar, T., Gulcher, J., Kiemeney, L.A., Kong, A.,

Thorsteinsdottir, U., Stefansson, K., 2008. Many sequence variants affecting diversity of adult human height. Nat Genet 40, 609-615.

Habier, D., Fernando, R.L., Garrick, D.J., 2013. Genomic BLUP decoded: a look into the black box of genomic prediction. Genetics 194, 597-607.

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

Hai, L., Guo, H.H., Xiao, S.H., Jiang, G.L., Zhang, X.Y., Yan, C.S., Xin, Z.Y., Jia, J.Z., 2005. Quantitative trait loci (QTL) of stem strength and related traits in a doubled-haploid population of wheat (Triticum aestivum L.). Euphytica 141, 1-9.

Helguera, M., Khan, I.A., Kolmer, J., Lijavetzky, D., Zhong-qi, L., Dubcovsky, J., 2003. PCR assays for the Lr37-Yr17-Sr38 cluster of rust resistance genes and their use to develop isogenic hard red spring wheat lines. Crop Science 43, 1839-1847.

Honsdorf, N., March, T.J., Berger, B., Tester, M., Pillen, K., 2014. High-throughput phenotyping to detect drought tolerance QTL in wild barley introgression lines. PLoS One 9, e97047.

Jahier, J., Abelard, P., Tanguy, A.M., Dedryver, F., Rivoal, R., Khatkar, S., Bariana, H.S., 2001. The Aegilops ventricosa segment on chromosome 2AS of the wheat cultivar 'VPM1' carries the cereal cyst nematode resistance gene Cre5. Plant Breeding 120, 125-128.

Keller, M., Karutz, C., Schmid, J.E., Stamp, P., Winzeler, M., Keller, B., Messmer, M.M., 1999. Quantitative trait loci for lodging resistance in a segregating wheat x spelt population. Theor Appl Genet 98, 1171-1182.

Kooke, R., Kruijer, W., Bours, R., Becker, F., Kuhn, A., van de Geest, H., Buntjer, J., Doeswijk, T., Guerra, J., Bouwmeester, H., Vreugdenhil, D., Keurentjes, J.J., 2016. Genome-Wide Association Mapping and Genomic Prediction Elucidate the Genetic Architecture of Morphological Traits in Arabidopsis. Plant Physiol 170, 2187-2203.

Legarra, A., Robert-Granie, C., Manfredi, E., Elsen, J.M., 2008. Performance of genomic selection in mice. Genetics 180, 611-618.

Liu, W.X., Leiser, W.L., Maurer, H.P., Li, J.H., Weissmann, S., Hahn, V., Wurschum, T., 2015. Evaluation of genomic approaches for marker-based improvement of lodging tolerance in triticale. Plant Breeding 134, 416-422.

McCormick, R.F., Truong, S.K., Mullet, J.E., 2016. 3D Sorghum Reconstructions from Depth Images Identify QTL Regulating Shoot Architecture. Plant Physiol 172, 823-834.

Miller, C.N., Harper, A.L., Trick, M., Werner, P., Waldron, K., Bancroft, I., 2016. Elucidation of the genetic basis of variation for stem strength characteristics in bread wheat by Associative Transcriptomics. BMC Genomics 17, 500.

Perez, P., de los Campos, G., 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics 198, 483-495.

Piñera-Chavez, F.J., Berry, P.M., Foulkes, M.J., Molero, G., Reynolds, M.P., 2016. Avoiding lodging in irrigated spring wheat. II. Genetic variation of stem and root structural properties. Field Crops Research 196, 64-74.

Pinthus, M.J., 1974. Lodging in Wheat, Barley, and Oats: The Phenomenon, its Causes, and Preventive Measures. Advances in Agronomy 25, 209-263.

Poland, J., 2015. Breeding-assisted genomics. Curr Opin Plant Biol 24, 119-124.

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

Rajaram, S., Van Ginkel, M., Fischer, R.A., 1995. CIMMYT's wheat breeding mega-environments (ME). Proceedings of the 8th International Wheat Genetics Symposium, Beijing, China, 1-10.

Sankaran, S., Khot, L.R., Carter, A.H., 2015. Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. Computers and Electronics in Agriculture 118, 372-379.

Shi, Y., Thomasson, J.A., Murray, S.C., Pugh, N.A., Rooney, W.L., Shafian, S., Rajan, N., Rouze, G., Morgan, C.L., Neely, H.L., Rana, A., Bagavathiannan, M.V., Henrickson, J., Bowden, E., Valasek, J., Olsenholler, J., Bishop, M.P., Sheridan, R., Putman, E.B., Popescu, S., Burks, T., Cope, D., Ibrahim, A., McCutchen, B.F., Baltensperger, D.D., Avant, R.V., Jr., Vidrine, M., Yang, C., 2016. Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research. PLoS One 11, e0159781.

Singh, D., Wang, X., Kumar, U., Gao, L., Noor, M., Imtiaz, M., Singh, R.P., Poland, J., 2018. Full dataset for high-throughput phenotyping-enabled genetic dissection of crop lodging in wheat. Figshare.

Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100, 9440-9445.

Verma, V., Worland, A.J., Sayers, E.J., Fish, L., Caligari, P.D.S., Snape, J.W., 2005. Identification and characterization of quantitative trait loci related to lodging resistance and associated traits in bread wheat. Plant Breeding 124, 234-241.

Wang, X., Singh, D., Marla, S., Morris, G., Poland, J., 2018. Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. Plant Methods 14, 53.

White, J.W., Andrade-Sanchez, P., Gore, M.A., Bronson, K.F., Coffelt, T.A., Conley, M.M., Feldmann, K.A., French, A.N., Heun, J.T., Hunsaker, D.J., Jenks, M.A., Kimball, B.A., Roth, R.L., Strand, R.J., Thorp, K.R., Wall, G.W., Wang, G.Y., 2012. Field-based phenomics for plant genetics research. Field Crops Research 133, 101-112.

Williamson, V.M., Thomas, V., Ferris, H., Dubcovsky, J., 2013. An Aegilops ventricosa Translocation Confers Resistance Against Root-knot Nematodes to Common Wheat. Crop Sci 53, 1412-1418.

Yang, M.D., Huang, K.S., Kuo, Y.H., Tsai, H.P., Lin, L.M., 2017. Spatial and Spectral Hybrid Image Classification for Rice Lodging Assessment through UAV Imagery. Remote Sensing 9, 583.

Yu, J.M., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S., 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38, 203-208.

Yu, X., Li, X., Guo, T., Zhu, C., Wu, Y., Mitchell, S.E., Roozeboom, K.L., Wang, D., Wang, M.L., Pederson, G.A., Tesso, T.T., Schnable, P.S., Bernardo, R., Yu, J., 2016. Genomic prediction contributing to a promising global strategy to turbocharge gene banks. Nat Plants 2, 16150.

Table 2.A. Experimental details of the study.

	Season 2016	Season 2017
No. unique entries	590	595
No. of plots	1320	1320
Plot size	1.3 × 3.8 m^2	1.3 × 3.8 m^2
Field design	α-lattice	α-lattice
Pre-lodging flight	March 01	March 02
Post-lodging flight	March 16	March 15
Ground-truth date	March 18	March 15
UAS platform	3DR IRIS+	DJI M100
In-air flight duration	25-30 min	20-25 min
Flight speed	2 m/s	2 m/s
Flight altitude	25 m	25 m
DEM resolution	1.5 cm/pixel	3.5 cm/pixel
Camera Sensor	Modified Canon S100	MicaSense RedEdge
Camera channels (nm)	Blue (460), Green (525), Near Infrared (710)	Blue (475), Green (560), Red (668), RedEdge (717), Near Infrared (840)

UAS: Unmanned Aerial System

DEM: Digital Elevation Model

Table 2.B. Approaches to assess lodging using digital images derived from UAS and ground-based assessment.

Approach	Trait	Description
Ground-truth	Lodging Incidence (LOI; 0-100%) Lodging Severity (LOS; 0-10) Lodging Index (LI; LOI × LOS)	Visual scores of lodging
Image-based digital lodging	Differential Mean (DLmean)	Plot summary mean
	Digital Lodging Mixture (DLmix)	Normal mixture-based lodging index

Table 2.C. 11-fold cross-validation predictive ability (r_{pv}), broad-sense heritability (H), and prediction accuracy (r_{pa}) of visual and digital lodging measures in years 2016 and 2017 at LDH.

	2016			*2017*		
	r_{pv}	H	r_{pa}	r_{pv}	H	r_{pa}
LOI	0.30	0.59	0.39	0.41	0.67	0.50
LOS	0.31	0.50	0.44	0.41	0.63	0.52
LI	0.32	0.55	0.43	0.40	0.63	0.50
DLmean	0.35	0.56	0.47	0.40	0.71	0.47
DLmix	0.31	0.52	0.43	0.42	0.63	0.53

r_{pv}: correlation between genomic estimated breeding values and phenotypic values

r_{pa}: predictive ability scaled to the squared root of heritability

Table 2.D. Prediction accuracies of lodging measures generated from different training-testing combinations on untested genotypes at LDH and FAS locations, e.g., 16LDH-17FAS is 16LDH training set predicting 17FAS.

	16LDH-17LDH	17LDH-16LDH	16LDH-17FAS	17LDH-16FAS	Average
LOI	0.37	0.45	0.28	0.27	0.34
LOS	0.42	0.54	0.32	0.23	0.38
LI	0.39	0.49	0.27	0.21	0.34
DLmean	0.37	0.51	0.30	0.20	0.35
DLmix	0.38	0.55	0.29	0.19	0.35

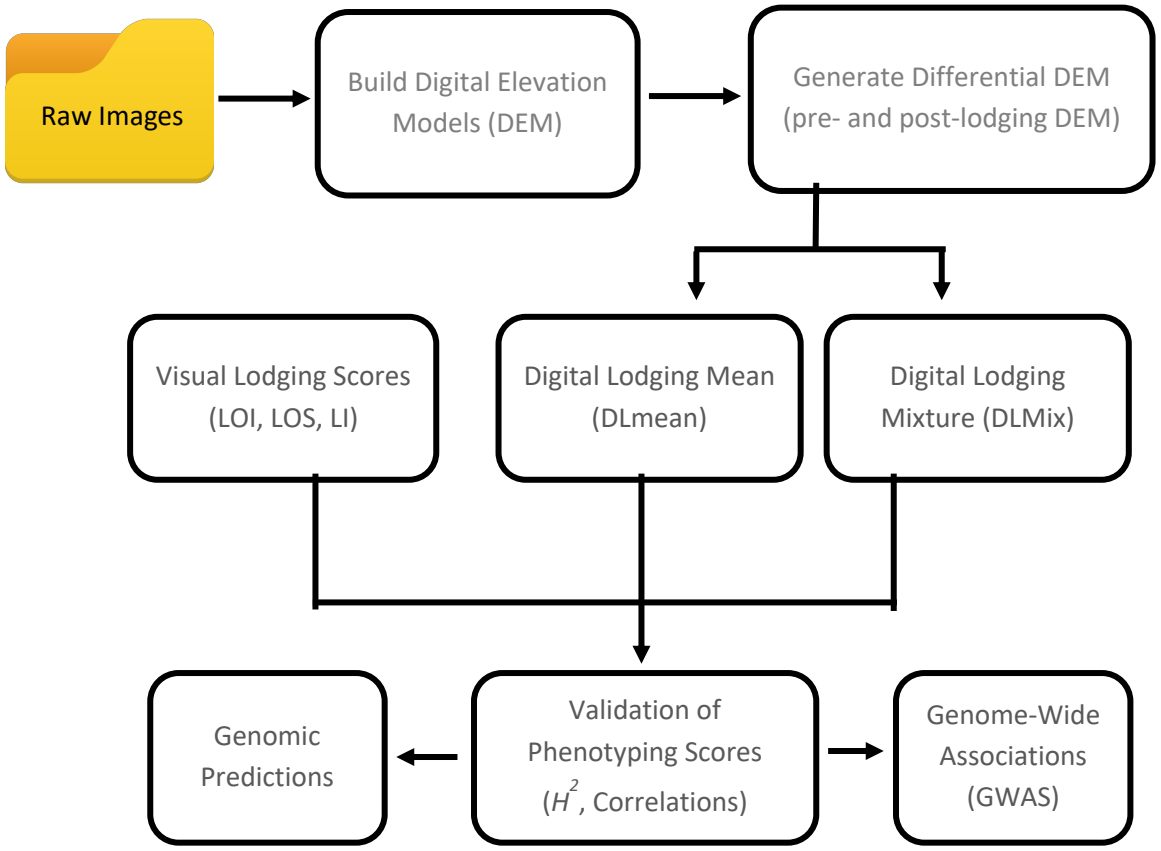


Figure 2.1. Workflow of digital and visual phenotypic analysis approaches used to assess the crop lodging in wheat.

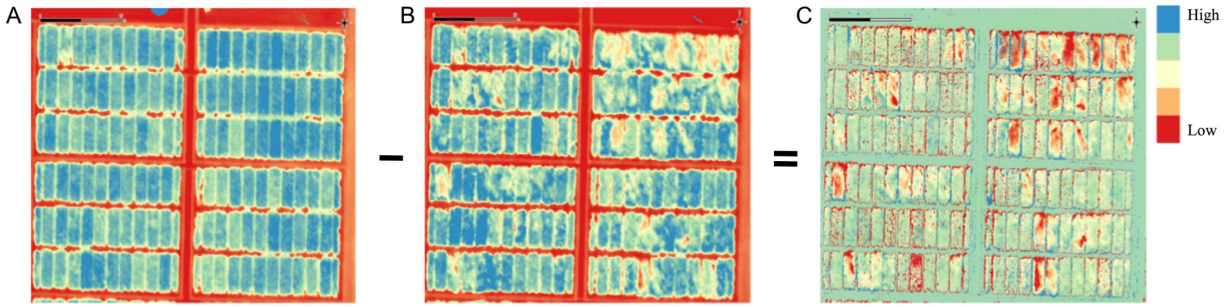


Figure 2.2. Processing of pre- and post-lodging digital elevation models (DEM) to obtain differential DEM of lodging. Post-lodging DEM is subtracted from pre-lodging DEM to generate a differential DEM of lodging. Panels are (A) pre-lodging, (B) post-lodging, and (C) differential DEM. Elevation differences are color coded with red corresponding to low elevation in (A and B) or high differences in (C), blue is areas of high elevation (A and B) or low differences (C).

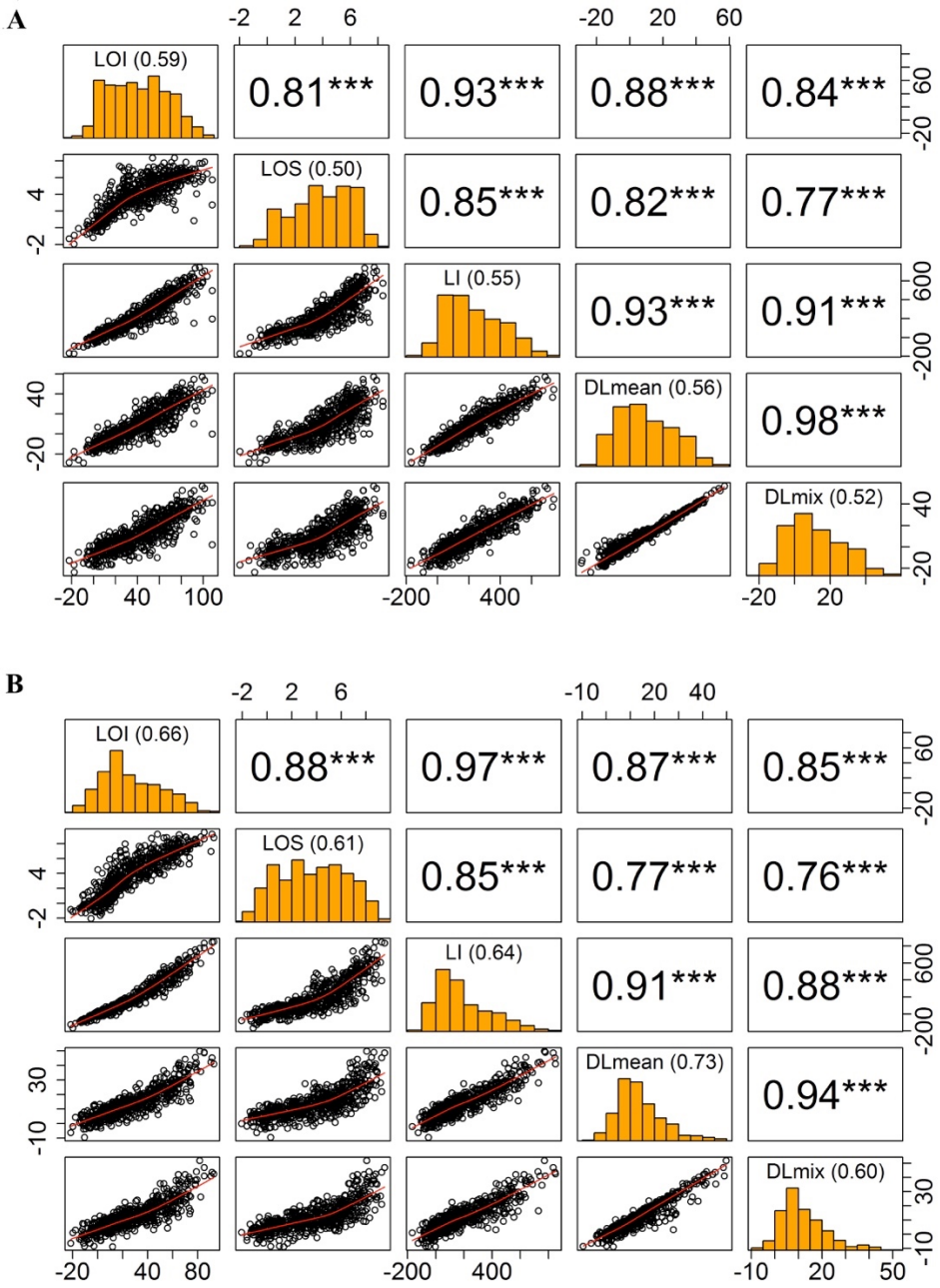


Figure 2.3. Relationship of visual and digital lodging scores. Pairwise correlation matrix of visual and digital measures of lodging in (A) year 2016, (B) year 2017. Diagonal panels show the trait distributions and broad-sense entry mean heritability; upper triangle is the Pearson's correlation coefficient values with significance levels as superscript ($^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$); lower triangle is the scatter plot.

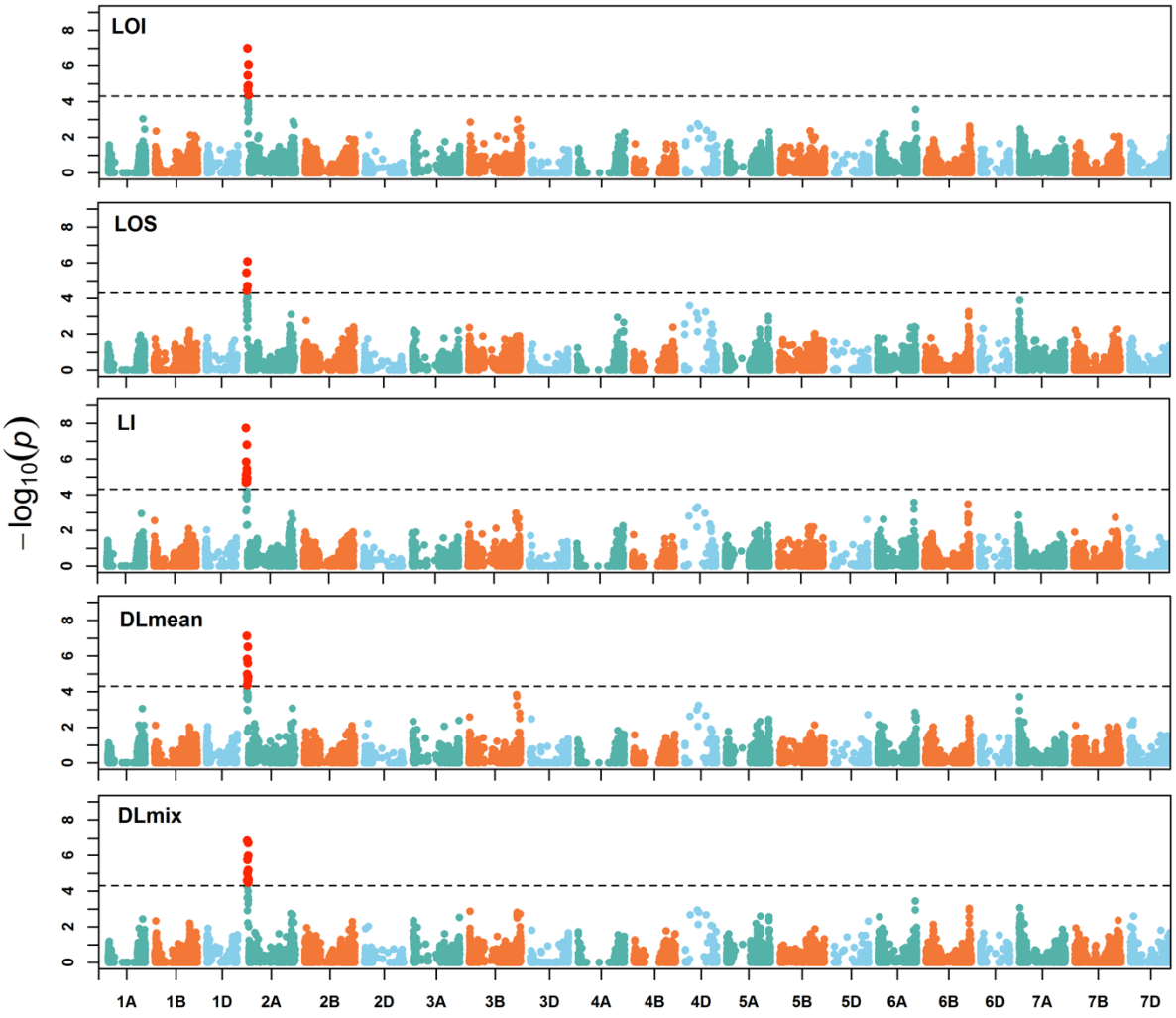


Figure 2.4. Manhattan plot of genome-wide associations. Manhattan plots of visual and digital lodging scores from combined analysis of genotypes from 2016 and 2017 (no. of genotypes=1035). The dashed lines on y-axis correspond to the genome-wide false discovery rate (FDR) threshold.

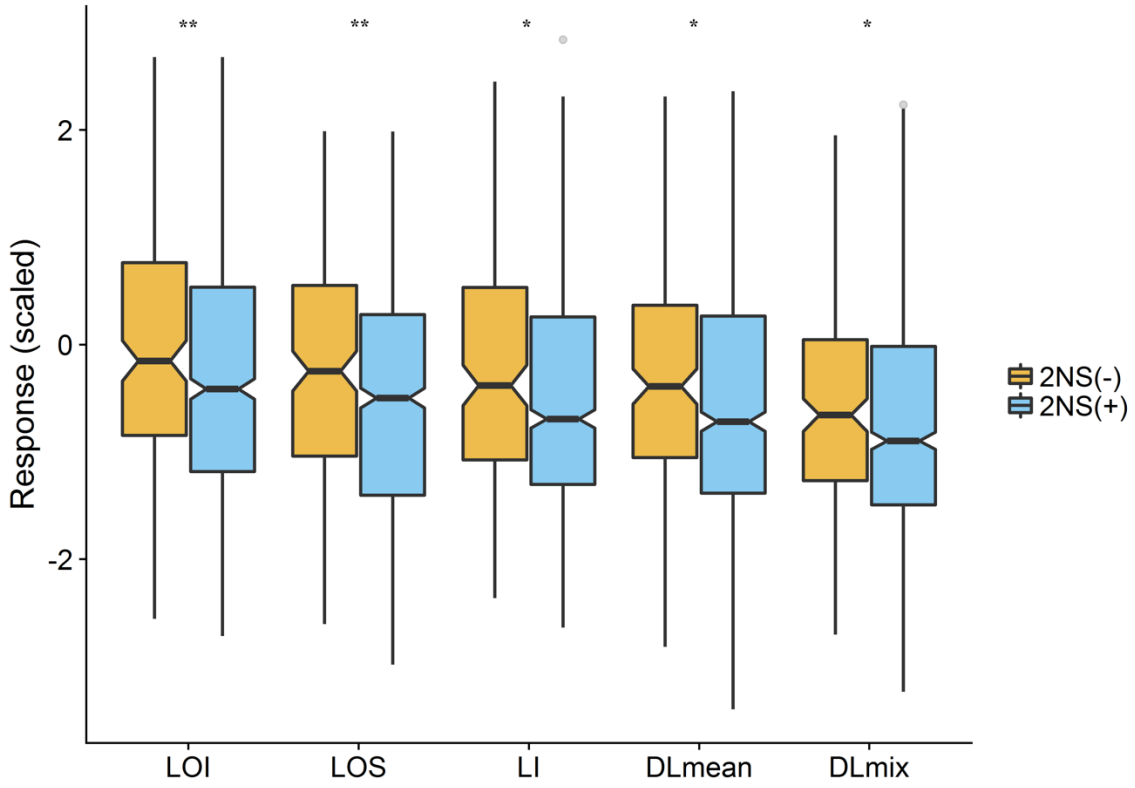


Figure 2.5. ***Effect of 2NS fragment on lodging.*** **The boxplot of scaled phenotypic values of lodging measures for 2NS positive (2NS+) and negative (2NS-) genotypes. The asterisks show the significant p-value for each trait (*t-test; n*=1010*; *P* < 0.05, ***P* < 0.01). Lodging incidence (LOI), Lodging Severity (LOS), Lodging Index (LI), Digital Lodging Mean (DLmean), Digital Lodging Mixture (DLmix).**

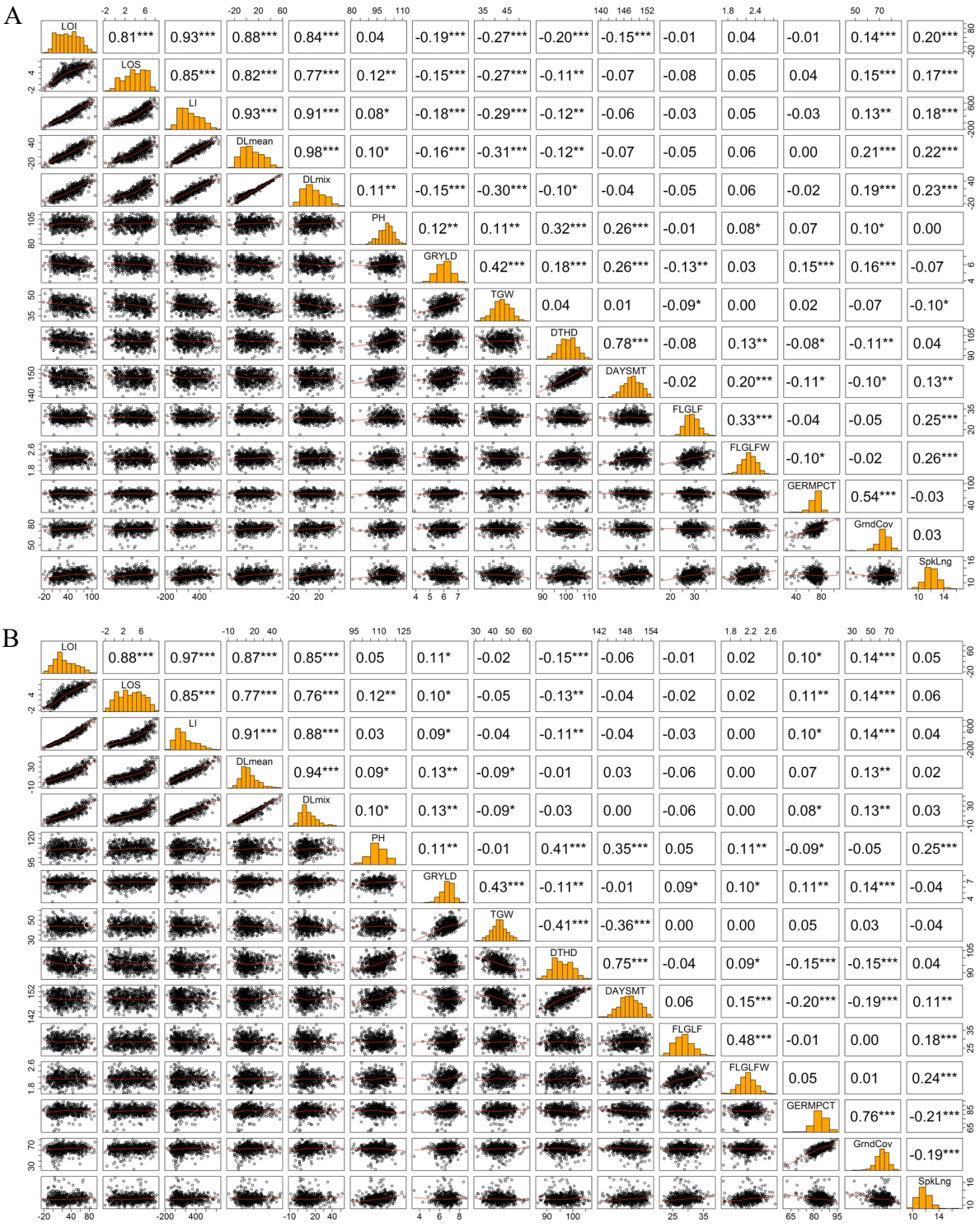


Figure 2.6. ***Relationship of lodging and agronomic traits.*** **Pairwise relationship of lodging and agronomic traits in (A) year 2016, (B) year 2017. Diagonal panels show trait distributions; upper triangle is the Pearson's correlation coefficient with significance levels as superscript ($^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$); lower triangle is the scatter plot.**

Chapter 3 - Phenomics-enabled Genetic and Physiological Dissection of Ground Cover in Wheat

Abbreviations

BLUP, best linear unbiased predictor; CIMMYT, International Maize and Wheat Improvement Center; DAS, days after sowing; GEI, genotype by environment interactions; GBS, genotyping-by-sequencing; GCP, ground control point; GWAS, genome-wide association studies; HTP, high-throughput phenotyping; JBL, Jabalpur; LDH, Ludhiana; MAF, minor allele frequency; PUS, PUSA; QTL, quantitative trait loci; QEI, QTL by environment interactions; SNP, single nucleotide polymorphism ; UAS, unmanned aerial systems.

Abstract

Ground cover, measured as the fraction of plant canopy covering soil, is an important breeding target in many crops due to its key role in stress adaptation and grain yield. However, the dissection of complex physiological and genetic components of ground cover is hindered by the subjective and tedious nature of the phenotypic measurements in the field. Digital ground cover measurements enabled by Unmanned Aerial Systems (UAS) can increase the scale (temporal and spatial) and throughput of field experiments to uncover the complex physiological and genetic mechanisms underlying ground cover trait. We tested this hypothesis by deploying an image-based classification procedure to quantify the temporal dynamics of UAS-derived digital ground cover on 8760 spring wheat plots over the course of two years in seven field environments (year × site × planting time). Accurate measurements were reflected by the high broad-sense heritability of image-based digital ground cover. High genetic correlation between

digital and visual ground cover as well as the high correlation to grain yield of visual ground cover demonstrated the potential of digital measurements as an alternative to visual ground cover ratings in wheat. To get a deeper insight into the genetic architecture of ground cover, we applied genome-wide association analysis at individual time-points and identified association signals at multiple genomic regions on wheat genome. The strength of the association signals modulated over the course of growing season and across environments highlighting the complex interplay of environmental and genetic factors in shaping the temporal phenotypic trait responses. To further investigate the impact of environmental factors on the physiology of ground cover, we assessed the sensitivity of the GWAS-based significant markers to quantitative environmental covariates across environments. The QTL-by-environment (QEI) analysis showed the significant contribution of mean and maximum temperatures during first 40 days after sowing. Collectively, by integrating the physiological and genetic components of ground cover trait, we demonstrate the power of phenomics enabled breeding to get insights into the genotype-to-phenotype link in wheat and related crops.

Keywords: *Triticum aestivum*, high-throughput phenotyping, phenomics, ground cover, early vigor, unmanned aerial systems UAS, vegetation index, genotyping-by-sequencing, breeding, genotype-by-environment interactions, genome-wide association study GWAS.

Introduction

A major question in biology is to understand the genotype-to-phenotype (G2P) link. In crop breeding, efforts are undergoing to leverage the scale and throughput of next-generation

phenomics and genomics to improve our understanding of G2P and predict the most promising candidate varieties in the breeding pipeline. Recent developments, especially in the field of phenomics, have shown great promise in accelerating the rate of genetic gains in breeding through data-driven insights into the G2P challenge.

Among the key breeding target traits, ground cover has a complex genetic and physiological architecture due to its association with multiple plant developmental processes and environmental conditions, as well as a complex association with grain yield at the end of the season. Ground cover is measured as the fraction of plot area covered by the plant canopy. Rapid early ground cover, also known as early vigor, contributes to the above ground biomass and final crop yield through efficient radiation capture, water-use efficiency, nutrient uptake, and competitiveness against weeds (Regan et al., 1997; Coleman et al., 2001; Rebetzke et al., 2002; Liao et al., 2004; Mullan and Reynolds, 2010). Previously, several key component traits of ground cover in wheat were identified, namely: embryo size, coleoptile length, width of third expanded leaf, leaf length, and specific leaf area (Richards and Lukacs, 2002; Moore and Rebetzke, 2015). Several physiological and genetic trade-offs associated with faster early growth and other developmental traits exist (Rebetzke et al., 2004; Botwright et al., 2005). The genetic variation in component traits of ground cover was successfully leveraged to develop drought resilient high-yielding wheat lines by increasing early season ground cover (Richards and Lukacs, 2002; Wilson et al., 2015). As an important component trait for overall productivity, early-season ground cover in wheat will continue to remain important breeding target especially in non-optimal environments.

Traditionally, ground cover is estimated by visual ratings of the plot area by a human rater in the field (Richards and Lukacs, 2002). These ratings are labor intensive, slow, and subjective. Therefore, it may confound the important biological signals in field experiments and

greatly limits time-series assessment. Image-based measurements enable a quick scan of the large breeding nurseries in relatively shorter time while providing increased temporal and spatial resolution of measurements (Haghighattalab et al., 2016; Shi et al., 2016). Therefore, image-based measurements can potentially address the shortcomings in traditional assessment methods of ground cover. A variety of digital phenotyping sensors have been deployed for dynamic measurements of ground cover (Mullan and Reynolds, 2010; Duan et al., 2016; Yu et al., 2017; Jimenez-Berni et al., 2018). Recently, the UAS-derived RGB imagery was used to study the genetic architecture of canopy coverage trait in soybean (Xavier et al., 2017). Thus, with the scalable and dynamic potential of UAS-based phenotyping, it is possible to increase the power of field experiments to dissect the genetic and physiological underpinnings of the ground cover trait in crops.

To evaluate the potential of UAS-derived digital ground cover and examine the genetic association of ground cover on large-scale breeding and genetic studies, we conducted multi-environment spring wheat trials in South Asia in 2017 and 2018 seasons. Multiple digital measurements and a visual assessment of 8760 advanced yield plots were performed for agronomic and phenological traits in wheat under seven field environments. Our primary objectives of this research were to: 1) evaluate the potential of UAS-derived digital ground cover as a high-throughput alternative to visual ground cover ratings in large breeding experiments, 2) investigate the quantitative relationship of genetic and environmental factors for ground-cover and grain yield in spring wheat.

Materials and Methods

Experimental Design

Seven field experiments (year × location × planting-date) were established at the Borlaug Institute for South Asia's research sites in India in wheat seasons of 2016-17 and 2017-18.

There were three identical normal-planting-date (hereafter, denoted as Normal) trials at each Ludhiana (LDH), PUSA (PUS), and Jabalpur (JBL) sites in both seasons, in addition to an early-planted trial (denoted as Early) at LDH in 2018. For simplification, hereafter, the experiments are designated in *yyLoc.Planting* format, where the *yy* refers to the year (2017, 2018), *Loc* to location (LDH, JBL, PUS), and *Planting* to planting date (Early and Normal). These field sites represent a diverse set of environments and management regimes from critical wheat growing areas of North-West (LDH), North-East (PUS), and Central India (JBL). While LDH and PUS sites have cooler and foggy early season, JBL has long sunny days throughout the wheat season. In addition, PUS and JBL have relatively shorter growth season due to high average daily temperatures compared to LDH. A total of 546 and 595 advanced CIMMYT spring wheat lines were planted in two replications in an alpha-lattice design with checks at each location in 2017 and 2018. These lines represent the advanced breeding material from Global Wheat Program from CIMMYT based in Mexico. Among the main selection targets of the program are grain yield, multiple rust resistance, grain quality, and drought and heat stress tolerance. The field experiment consisted of 1200 and 1320 six-row plots of size 5 m^2 at each site in two seasons. The field plots were further divided into smaller blocks of 120 plots called trials at each year/location. Each trial consisted of two replicate blocks each divided into 6 subblocks, where each replicate block contained 53 entries and 7 check varieties. The supplemental irrigation was applied via flood irrigation at LDH and PUS, and with sprinkler irrigation system at JBL sites. Recommended crop management practices (irrigation, fertilizer application etc.) at each site were followed throughout the field season.

Data Collection

Ground and aerial imaging data was collected at regular intervals throughout the field season at each environment (Table 3.1). In terms of phenotypic data, some field sites were

covered less frequently than others due to logistical constraints. The aerial imagery was captured using a DJI Matrice100 (DJI, USA) quadcopter Unmanned Aerial System (UAS) equipped with a 5-band multi-spectral RedEdge camera (Micasense Inc. USA) (Wang et al., 2018). Images were collected on clear, sunny days between 10:30 AM -2:00 PM with a total of 25 UAS flights covering vegetative, reproductive and maturity growth stages of wheat in two seasons (Table 3.1). To normalize the calendar days, the crop growth progression at each site was characterized in terms of thermal time (°Cd) and days after sowing (DAS). The thermal time calculated as aggregated mean daily temperature with 0°C as the base temperature.

Visual Measurements

In addition to the UAS-based imagery, visual ground cover ratings were collected between 35-45 DAS to "ground-truth" the digital measurements. The ground cover ratings showed considerable variation from 0 to 100 % at each location across time-points (Fig. 3.1). To further assess the relationship of ground cover with yield, individual plots were harvested after the physiological maturity to estimate the final grain yield. The details trait ontological information is provided (Table S2).

HTP Data Processing

Raw multi-spectral image pixels were converted to reflectance values using an in-house automated Python image processing pipeline (https://github.com/xwangksu/uavBatch; accessed Sept 24, 2018). After initial pre-processing, the experimental metadata along with the original calibrated images were stored on the remote database. Subsequently, the image stitching operation was performed via photogrammetry software Photoscan Pro (Agisoft Inc., Russia). The detailed description of image processing parameters can be found in Wang et al. (2018) and Singh et al. (2018b).

Digital Ground Cover Estimation

Digital ground cover was estimated via supervised image classification procedure on the orthomosaic geoTiff. Image pixels were classified belonging to each of soil, green-vegetation, and shadow via semi-automatic classification (SCP) plugin in QGIS (Congedo, 2016) (Supplementary Fig. B.1). Subsequent to orthomosaic classification, the plot-level ground cover was extracted using the custom R functions. The functions take classified raster image (i.e. output from SCP plugin) and plot boundary shapefile as input to extract the plot-level digital ground cover as the fraction of vegetation pixels falling under the plot boundary shapefile layer. The data distributions from each time-point were plotted and investigated for outliers. Some early time-points from 2018 optimal-planting-date experiment at Ludhiana (18LDH.Normal) had trials with areas of missing data due to incomplete flight plans that were removed.

Spatial Corrections and BLUP

The raw phenotypic data for ground cover and grain yield showed prominent spatial trends for field variability. A 2D spatial p-spline method was applied to each trait separately to correct for the spatial trends (Rodríguez-Álvarez et al., 2018). The spatial correction procedure implemented in R package SpATS can incorporate local and global spatial trends in a single-step mixed model avoiding the model selection procedure often required in commonly used spatial adjustment procedures. The SpATS spatial model is described as following:

$$y = f(u, v) + X_d b_d + Z_d h_d + \varepsilon \qquad [1]$$

where y is the phenotypic response variable (e.g. yield or ground-cover); $f(u, v)$ is the p-spline smooth bivariate function that takes row (u) and column (v) positions as input; X_d and Z_d are the design matrices associated with fixed (intercept) and random blocking factors (trial, trial:rep, trial:rep:sub-block), respectively; b_d and h_d are fixed and random effect coefficients of design matrix; ε is the random residual error. It is further assumed $h_d \sim N(0, G)$ and $\varepsilon \sim N(0,$

I), where G and I are the block diagonal and identity matrices. The detailed theoretical description of 2D p-spline models, including *f (u, v)* component above is presented in Rodríguez-Álvarez et al. (2018). The Best Linear Unbiased Predictors (BLUP) obtained from equation 1 were used as trait input values in subsequent steps.

Broad-sense Heritability

The broad-sense line-mean heritability or repeatability was used to evaluate the quality of phenotypic measurements (equation. 2). The repeatability on entry-mean basis for each trait was calculated following:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_e^2/r} \quad [2]$$

where σ^2_G is the genotypic variance, σ^2_e is the residual variance, and r is the number of replicates. The variance components for H^2 were calculated using *lme4* package in R (Bates et al., 2014). Genotypic best linear unbiased predictors (BLUPs) were calculated as follows:

$$y_{ijkl} = \mu + G_i + Z_j + R_{k(j)} + P_{c(j)} + Q_{r(j)} + M_{kl(j)} + e_{ijkl} \quad [3]$$

where y_{ijkl} is the phenotypic response variable, μ is the fixed overall mean, G_i is the random genotype effect, Z_j is the random trial effect, $R_{k(j)}$ is the random effect of replicate nested within a trial, $P_{c(j)}$ is the random effect of columns nested within a trial, $Q_{r(j)}$ is the random effect of rows nested within a trial, $M_{kl(j)}$ is the random effect of sub-blocks and replications nested within a trial, and e_{ijkl} is the residual effect. The accuracy of the phenotypic data measured in terms of repeatability fluctuated significantly throughout the season in each environment. As a quality control step, the time-points with repeatability below 20% were removed from subsequent analyses. This led to the exclusion of a several digital ground cover time-points and one visual ground cover measurement at 18PUS.Normal environment.

Genetic Correlations

Marker based genetic correlations between a pair of traits were calculated with *sommer (Covarrubias-Pazaran, 2016)* package in R as follows:

$$r_{g(x,y)} = \frac{cov_g(x,y)}{\sqrt{var_g(x) * var_g(y)}} \quad [4]$$

where $cov_g\,(x,y)$ is the covariance of the traits *x* and *y*; $var_g\,(x)$ and $var_g\,(y)$ is the genetic variance of traits x and y, respectively. The Pearson correlation coefficients of pair of traits were estimated via *cor* function in R. The raw data associated with the experiment can be accessed at the Figshare public repository (DOI: 10.6084/m9.figshare.7180625). The analysis scripts can be found at github/singhdj2/ind-GrndCov-2018.

Genotypic Data Processing

Two-enzyme genotyping-by-sequencing (GBS) method was used for reduced representation library preparation and DNA sequencing (Poland et al., 2012b). GBS based single nucleotide polymorphic (SNP) markers with minor allele frequency (MAF) of above 0.01 were called using TASSEL5 v2 pipeline (Glaubitz et al., 2014a) from a large cohort of ~50,000 advanced CIMMYT wheat breeding lines. This cohort represent the F_6:F_7 derived lines developed using the selected bulk method within the Global Wheat Program at CIMMYT, Mexico. The unique tags were aligned to the Chinese Spring RefSeq v1.0 using bowtie2 (Langmead and Salzberg, 2012; IWGSC, 2018). The tags with mapping quality greater than or equal to 20 were selected to call SNPs. All SNPs from the discovery step were recovered and filtered based on three criteria: inbreeding coefficient of at least 80, Chi-square test to test expected inbreeding of 96% at $F_{5:6}$ generation, and Fisher test to confirm bi-allelic single locus SNPs (Poland et al., 2012a). The SNPs that passed at least one filter were selected and further filtered to remove SNPs with at least 70% missing data. The missing data in the remaining SNPs

were imputed with Beagle v4.1 (Browning and Browning 2016). A final round of quality filtering was performed on subset of lines belonging to this study. After the quality filtering, a set of 19,217 high-quality SNPs were retained for subsequent genetic analyses.

Genome-wide Association Analysis (GWAS)

A mixed linear model was fit via rrBLUP package in R (Endelman, 2011) as follows:

$$y = X\beta + Zu + S\tau + e \quad [5]$$

where y is the vector of adjusted means (BLUPs) of phenotypic values, β is the vector of fixed terms (intercept and principal component-based population structure covariates), u is vector of genetic marker effects, and $\boldsymbol{\tau}$ is the random effect of genetic background of each line; and X, Z, and S are the respective design matrices. An exploratory genome-wide threshold of *-log10(p-value)* 3 was selected.

Genotype × Environment Analysis

Two groups of digital ground cover time-points were chosen for Genotype × Environment (GE) analysis based on the criteria of at least three environments falling within the intersectional thermal time window of 100 °Cd in each season 2017 and 2018. The first group comprised the 17PUS.Normal, 17LDH.Normal, and 17JBL.Normal environments from season 2017 covering the thermal time window of 575-666 °Cd, which corresponded to approximately 5-6 days in calendar day units. The second group belonged to advanced stage time-points from season 2018 18LDH.Early, 18JBL.Normal, 18PUS.Normal, and 18LDH.Normal environments covering the thermal window of 881-925 °Cd. To estimate the effect of environmental factors on marker sensitivity for ground cover across the selected environments, a linear model with main effect term for environment (Z; treated as a qualitative covariate), marker main effect (Q) and an interaction term for QTL-by-environment (QE) covariate (e.g., temperature, thermal time) was fit for the most significant GWAS markers separately. The environmental covariates used were:

thermal time; mean, minimum, and maximum temperature, where each temperature covariate was further divided into 20, 40, 60, and 80 Day aggregated values for each environment. Briefly, a regression model that incorporates explicit genotypic and environment factors was fit for each of the marker and environment covariate combinations following Malosetti et al. (2013). The significance of the main effect and QE interaction terms were used to assess the relative sensitivity of the markers in relation to the environmental covariates. The QE interaction term was declared significant above the *p*-value threshold of 0.01. A second linear model was fit by dropping the interaction term from the above model to estimate the average additive effect for each marker across qualitative environments

Results and Discussion

UAS-based Estimation of Digital Ground Cover

The advanced wheat yield trials were setup at three diverse wheat growing environments in India namely, JBL, PUS and LDH in 2017 and 2018. Aerial UAS phenotyping operations were performed throughout the growing season at regular intervals to cover the key wheat growth stages starting from two-leaf stage up to the physiological maturity. A total of 34 datasets covering multiple timepoints were assessed to evaluate the dynamic nature of the ground cover trait in wheat across seven environments (Table 3.1). A quantitative assessment of ground cover was carried out using classified orthomosaics of multi-spectral UAS imagery (Fig. C.1). The plot-level data from each time-point was extracted by overlaying classified orthomosaics with plot-boundary shapefiles. A total of 41,880 plot level ground cover measurements were generated over the course of two seasons along with the early-season visual scores of ground cover at each location. Collectively, within the scope of large breeding and genetics studies, this temporal and spatial extent is reflective of the scalable potential of the phenomics based approaches.

Phenotypic Variation of Ground Cover

The image-derived quantitative digital ground cover measurements were compared to the single time-point visual ground cover scores in each experiment. Both digital and visual ground cover showed considerable variation ranging from 0-100% across environments and growth stages (Fig. 3.1). The average daily temperature at each site varied considerably in both seasons; therefore, the thermal time (°Cd; measured as cumulative daily mean temperature) was used to normalize the growth stages for comparisons across environments. Typically, the JBL site experiences faster thermal period and has a much shorter growing period compared to other two locations of PUS and LDH. The LDH location is located in cooler NW region where the shorter days and lower temperatures stretch the growing season longer compared to PUS and JBL. We observed that the majority of plots at LDH and JBL achieved 75% ground cover in 700 °Cd (Fig. 3.1). In 2018, the plots at PUS location showed slow growth in ground cover where it took longer than 700 °Cd to cover 75% of the ground. Though the dense temporal coverage of data-points across environments is required to draw more informed inferences, the general trends indicate the plants at LDH and JBL sites went through exponential growth in ground cover to achieve 90% coverage during the 500-1000 °Cd period. In contrast, the PUS site showed a slow and more protracted growth period that spanned 500-1400 °Cd to achieve 90% ground coverage. The slow canopy growth at PUS can be attributed to low germination rates at this location during the study years. These results show that the multi-time-point ground cover measurements can provide insights into the temporal growth dynamics across environments.

Spatial Adjustment of Field Effects

The large-scale field studies necessitate an adequate control of non-random spatial effects caused by the soil, environment, and management practices within the field experiments (Gilmour et al., 1997). We observed significant spatial trends in phenotypic data in our

experiments (Fig. 3.2). The spatial trends were more prominent in some parts of the field than the other. To account for the spatial heterogeneity, we applied a p-spline based spatial corrections following Rodríguez-Álvarez et al. (2018). By fitting a single-step mixed model procedure, we were able to account for the continuous global and local spatial trends in our experiments as suggested by the plots of the residuals and fitted BLUPs (Fig. 3.2). Spatial adjustments of breeding experiments were shown to improve the breeding selection and genomic predictions (Qiao et al., 2004; Lado et al., 2013; Velazco et al., 2017). Cooper et al. (2014) highlighted the positive impact of spatial adjustments on genetic correlations for 1126 multi-environment trials conducted in US corn-belt from 2002 and 2009. Collectively, the previous reports and our current results support the need for spatial corrections of field and high-throughput phenotyping experiments through spatial models.

Comparative Accuracy of Phenotypic Measurements

We assessed the accuracy of phenotypic measurements via repeatability or broad-sense heritability on entry-mean basis. The repeatability of visual ground cover showed higher fluctuations (0.04-0.61) compared to digital ground cover (0.05-0.70) (Table 3.A). The moderate to high overall repeatability of digital and visual ground cover reflected good accuracy of the assessment methodologies. We compared the repeatability of visual assessment to the digital measure from the closest timepoint and found the digital traits outperform visual ground cover assessment in five out of seven environments. To further assess the influence of genetic and environmental factors on visual and digital traits, the genetic correlations between the pairs of closest measurement time-points were compared (Table 3.B). The visual and digital ground cover ratings showed strong genetic correlations ($r > 0.68$-0.99). Thus, the accurate digital ground cover measurements, a finding supported by strong genetic correlations and moderate repeatability, can potentially be an alternative to the conventional visual ground cover scores.

Genome-wide Association Study (GWAS) of Ground Cover

The GWAS was conducted to dissect the genetic underpinnings of ground cover in our populations. Each individual time-point was fit as a response variable in a mixed linear model to explore the marker-trait associations. A total of ca. 70 unique association signals was observed in years 2017 and 2018 (Fig. 3.3, Supplementary Figs. B.2-B.5). Many association signals coincided at more than one time-points in 17LDH.Normal environment. In 2018, three SNPs at chromosomes 3A, 3D, and 5A featured twice within and across environments. The association peak at 1BL was coincident at DAS 46 and 57 in 18JBL.Normal and 18LDH.Early, with both time-points belonging to the closer thermal window of 881-893 °Cd. A similar peak at 6DL was also observed for 18JBL.Normal and 18LDH.Early at thermal window of 830-893 °Cd, suggesting a growth stage specific signal. QTL on chromosomes 1B, 2B, 6B, and 7A were earlier reported to be associated with early vigor related traits in wheat (Coleman et al., 2001; Moore and Rebetzke, 2015). Overall, the significant markers each explained less than 1% of the total phenotypic variation for ground cover across time-points and environments (data not shown), suggesting a complex genetic control of ground cover and its component traits influenced by the environmental conditions (Coleman et al., 2001; Rebetzke et al., 2001; ter Steege et al., 2005; Xavier et al., 2017).

QTL Effects Across Environments

To assess the effect of environments on the QTL, two groups of time-points, one each for early season and advanced season time-points were selected. The first group consisted of three environments from 575-666 °Cd thermal window in 2017 and the second group consisted four environments from 881-925 °Cd thermal window from 2018. After adjusting for the main environmental effect, a total of 27 and 21 QTL with significant main effects (p-value < 0.01) were detected in first and second groups, respectively. Notably, the QTL in both groups did not

overlap. The allelic effect of the individual QTL ranged from -0.3 to 0.6 (in % ground cover units) in both groups (Fig. 3.4). While 89% of the QTL in the early season group had significantly positive allelic effects for ground cover, only 62% of the SNPs during advanced season group had significantly positive allelic effects. Consistent with the previous reports of significant GEI for early growth traits (Moore and Rebetzke, 2015), these results highlight the complex interplay of genotypic and environmental factors influencing the trait physiology in wheat.

Environmental Determinants of Ground Cover

The varied environmental conditions at our research sites afforded the opportunity to explore the relationship of temperature-based environmental covariates and ground cover. The QTL × Environment Interactions (QEI) were modeled by including quantitative environmental covariates in the linear model (Mathews et al., 2008). A total of 18 unique QTL in each early (Group I) and advanced (Group II) stages were found to have significant QEI effects for environmental covariates (p-value < 0.01; Fig. 3.5). While mean and maximum temperature during the first 40 DAS had the highest number of QTL with significant QEI in group I, the minimum and mean temperature had most significant contribution to the QEI for group II. Overall, these results highlight the significant role of temperature during the early developmental stages that can have profound effect on the genotype-environment sensitivity.

Relationship of Ground Cover and Grain Yield

Following on the genetic and physiological dissection of ground cover, we explored the relationship of ground cover with the grain yield. Digital ground cover and grain yield were phenotypically positively correlated although the relationship showed significant temporal fluctuations (Table 3.C). The peak trends in phenotypic correlations with grain yield were observed ca. 80-90 DAS at 18PUS, consistent with the slow increase in canopy cover at this site

(Fig. 3.1). Some recent studies have reported high genetic correlations between grain yield and digital canopy cover in maize and soybean (Xavier et al., 2017; Makanza et al., 2018). However, in our experiments, the genetic correlations between grain yield and ground cover were inconsistent likely due to varying heritability and/or high sampling error associated with the markers (Visscher et al., 2014). Therefore, to get a better estimate of comparative selection accuracy in the absence of genetic correlations, we derived a trait selection efficiency index (SEI) by combining the correlation coefficient of grain yield and ground cover with the broad-sense heritability for ground cover (SEI = Correlation × H^2; Table 3.C). A high index value was interpreted as the high comparative selection efficiency for the trait or time-point being compared. In five out of seven environments, the digital ground cover showed higher index scores than visual ratings, suggesting the relative superiority of UAS-derived digital measurements over visual ratings as a potential selection target in wheat.

Genetic Tradeoffs of Grain Yield and Ground Cover

To further investigate the grain yield-ground cover genetic tradeoffs, we quantified the genetic effects of the colocalized QTLs of the two traits in early and late season groups. In the early season group, all six QTL had positive effect on both ground cover and grain yield, where the major alleles on average increased the grain yield by 25 kg ha^{-1} (Table 3.D). In contrast, the majority of the QTL in late season group showed opposite allelic effects for two traits, suggesting a significant genetic tradeoff between two traits. Similar tradeoffs at different developmental stages and environments have been attributed to antagonistic pleiotropy in plants (El-Soda et al., 2014; Li et al., 2018). Further work is needed to investigate the molecular basis of plant adaptation and its genetic tradeoffs for ground cover in wheat. We envision that the described phenomics based approaches will provide a powerful framework to test these hypotheses.

Conclusion

The study of complex traits is often hindered by the underpowered genetic experiments due to the phenomic bottleneck. We show that the phenomics based approaches provide an accurate and high-throughput framework to assess large number of genotypes at a significantly high temporal and spatial resolution. By integrating validated multi-time-point measurements with the genomics and environmental modelling, we provide a glimpse into the complex genetic architecture of ground cover in wheat. The environmentally sensitive genetic architecture of ground cover in wheat makes for a good case for whole-genome prediction models that incorporate explicit environmental factors in the genomic models (Heslot et al., 2014; Jarquin et al., 2014). The growth-stage and environment conditioned changes in the size and magnitude of the QTL effects point towards a significant genetic tradeoffs of grain yield and ground cover in wheat.

Further research is required to get a deeper understanding of the complex pleiotropic control of the two traits. For crop breeding perspective, the future modelling approaches that can leverage the crop growth models, G × E, and phenomics are needed to increase the genetic gains in crops. To that end, significant progress has been made recently in terms of theoretical framework development to implement these models in breeding programs (Messina et al., 2018; van Eeuwijk et al., 2018). Arguably, next generation crop improvement programs would greatly benefit from the emerging technologies via improved understanding of genotype-to-phenotype linkage.

References

Bates, D., Maechler, M., Bolker, B., Walker, S., 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1, 1-23.

Botwright, T.L., Rebetzke, G.J., Condon, A.G., Richards, R.A., 2005. Influence of the Gibberellin-sensitive Rht8 Dwarfing Gene on Leaf Epidermal Cell Dimensions and Early Vigour in Wheat (Triticum aestivum L.). Annals of Botany 95, 631-639.

Coleman, R.K., Gill, G.S., Rebetzke, G.J., 2001. Identification of quantitative trait loci for traits conferring weed competitiveness in wheat Triticum aestivum. Australian Journal of Agricultural Research 52, 1235-1246.

Congedo, L., 2016. Semi-automatic classification plugin documentation. Release 4, 29.

Cooper, M., Messina, C.D., Podlich, D., Totir, L.R., Baumgarten, A., Hausmann, N.J., Wright, D., Graham, G., 2014. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop & Pasture Science 65, 311-336.

Covarrubias-Pazaran, G., 2016. Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. PLoS One 11, e0156744.

Duan, T., Chapman, S.C., Holland, E., Rebetzke, G.J., Guo, Y., Zheng, B., 2016. Dynamic quantification of canopy structure to characterize early plant vigour in wheat genotypes. J Exp Bot 67, 4523-4534.

El-Soda, M., Malosetti, M., Zwaan, B.J., Koornneef, M., Aarts, M.G.M., 2014. Genotype x environment interaction QTL mapping in plants: lessons from Arabidopsis. Trends in Plant Science 19, 390-398.

Endelman, J.B., 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome 4, 250-255.

Gilmour, A.R., Cullis, B.R., Verbyla, A.P., 1997. Accounting for natural and extraneous variation in the analysis of field experiments. JABES, 269-293.

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

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

Heslot, N., Akdemir, D., Sorrells, M.E., Jannink, J.-L., 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor Appl Genet 127, 463-480.

IWGSC, 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361.

Jarquin, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., Piraux, F., Guerreiro, L., Perez, P., Calus, M., Burgueno, J., de los Campos, G., 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127, 595-607.

Jimenez-Berni, J.A., Deery, D.M., Rozas-Larraondo, P., Condon, A.G., Rebetzke, G.J., James, R.A., Bovill, W.D., Furbank, R.T., Sirault, X.R.R., 2018. High Throughput Determination of Plant Height, Ground Cover, and Above-Ground Biomass in Wheat with LiDAR. Frontiers in Plant Science 9.

Lado, B., Matus, I., Rodriguez, A., Inostroza, L., Poland, J., Belzile, F., del Pozo, A., Quincke, M., Castro, M., von Zitzewitz, J., 2013. Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. G3 (Bethesda) 3, 2105-2114.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9, 357.

Li, X., Guo, T., Mu, Q., Li, X., Yu, J., 2018. Genomic and environmental determinants and their interplay underlying phenotypic plasticity. Proceedings of the National Academy of Sciences.

Liao, M., Fillery, I.R.P., Palta, J.A., 2004. Early vigorous growth is a major factor influencing nitrogen uptake in wheat. Functional Plant Biology 31, 121-129.

Makanza, R., Zaman-Allah, M., Cairns, J., Magorokosho, C., Tarekegne, A., Olsen, M., Prasanna, B., 2018. High-Throughput Phenotyping of Canopy Cover and Senescence in Maize Field Trials Using Aerial Digital Canopy Imaging. Remote Sensing 10, 330.

Malosetti, M., Ribaut, J.-M., van Eeuwijk, F.A., 2013. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Frontiers in Physiology 4.

Mathews, K.L., Malosetti, M., Chapman, S., McIntyre, L., Reynolds, M., Shorter, R., van Eeuwijk, F., 2008. Multi-environment QTL mixed models for drought stress adaptation in wheat. Theor Appl Genet 117, 1077-1091.

Messina, C.D., Technow, F., Tang, T., Totir, R., Gho, C., Cooper, M., 2018. Leveraging biological insight and environmental variation to improve phenotypic prediction:

Integrating crop growth models (CGM) with whole genome prediction (WGP). European Journal of Agronomy.

Moore, C., Rebetzke, G., 2015. Genomic Regions for Embryo Size and Early Vigour in Multiple Wheat (Triticum aestivum L.) Populations. Agronomy 5, 152.

Mullan, D.J., Reynolds, M.P., 2010. Quantifying genetic effects of ground cover on soil water evaporation using digital imaging. Functional Plant Biology 37, 703-712.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S.Y., Manes, Y., Dreisigacker, S., Crossa, J., Sanchez-Villeda, H., Sorrells, M., Jannink, J.L., 2012a. Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome 5, 103-113.

Poland, J.A., Brown, P.J., Sorrells, M.E., Jannink, J.L., 2012b. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7, e32253.

Qiao, C.G., Basford, K.E., DeLacy, I.H., Cooper, M., 2004. Advantage of single-trial models for response to selection in wheat breeding multi-environment trials. Theor Appl Genet 108, 1256-1264.

Rebetzke, G., Appels, R., Morrison, A., Richards, R., McDonald, G., Ellis, M., Spielmeyer, W., Bonnett, D., 2001. Quantitative trait loci on chromosome 4B for coleoptile length and early vigour in wheat (Triticum aestivum L.). Australian Journal of Agricultural Research 52, 1221-1234.

Rebetzke, G.J., Botwright, T.L., Moore, C.S., Richards, R.A., Condon, A.G., 2004. Genotypic variation in specific leaf area for genetic improvement of early vigour in wheat. Field Crops Research 88, 179-189.

Rebetzke, G.J., Condon, A.G., Richards, R.A., Farquhar, G.D., 2002. Selection for Reduced Carbon Isotope Discrimination Increases Aerial Biomass and Grain Yield of Rainfed Bread Wheat. Crop Science 42, 739-745.

Regan, K.L., Siddique, K.H.M., Tennant, D., Abrecht, D.G., 1997. Grain yield and water use efficiency of early maturing wheat in low rainfall Mediterranean environments. Australian Journal of Agricultural Research 48, 595-604.

Richards, R., Lukacs, Z., 2002. Seedling vigour in wheat-sources of variation for genetic and agronomic improvement. Australian Journal of Agricultural Research 53, 41-50.

Rodríguez-Álvarez, M.X., Boer, M.P., van Eeuwijk, F.A., Eilers, P.H.C., 2018. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spatial Statistics 23, 52-71.

Shi, Y., Thomasson, J.A., Murray, S.C., Pugh, N.A., Rooney, W.L., Shafian, S., Rajan, N., Rouze, G., Morgan, C.L., Neely, H.L., Rana, A., Bagavathiannan, M.V., Henrickson, J., Bowden, E., Valasek, J., Olsenholler, J., Bishop, M.P., Sheridan, R., Putman, E.B., Popescu, S., Burks, T., Cope, D., Ibrahim, A., McCutchen, B.F., Baltensperger, D.D., Avant, R.V., Jr., Vidrine, M., Yang, C., 2016. Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research. PLoS One 11, e0159781.

Singh, D., Wang, X., Kumar, U., Gao, L., Noor, M., Imtiaz, M., Singh, R.P., Poland, J., 2018. High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. Frontiers in Plant Science (Accepted).

ter Steege, M.W., den Ouden, F.M., Lambers, H., Stam, P., Peeters, A.J., 2005. Genetic and physiological architecture of early vigor in Aegilops tauschii, the D-genome donor of hexaploid wheat. A quantitative trait loci analysis. Plant Physiol 139, 1078-1094.

van Eeuwijk, F., Bustos-Korts, D., Millet, E.J., Boer, M., Kruijer, W., Thompson, A., Malosetti, M., Iwata, H., Quiroz, R., Kuppe, C., Muller, O., Blazakis, K.N., Yu, K., Tardieu, F., Chapman, S., 2018. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. Plant Science.

Velazco, J.G., Rodriguez-Alvarez, M.X., Boer, M.P., Jordan, D.R., Eilers, P.H.C., Malosetti, M., van Eeuwijk, F.A., 2017. Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. Theor Appl Genet 130, 1375-1392.

Visscher, P.M., Hemani, G., Vinkhuyzen, A.A.E., Chen, G.-B., Lee, S.H., Wray, N.R., Goddard, M.E., Yang, J., 2014. Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. PLOS Genetics 10, e1004269.

Wang, X., Singh, D., Marla, S., Morris, G., Poland, J., 2018. Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. Plant Methods 14, 53.

Wilson, P.B., Rebetzke, G.J., Condon, A.G., 2015. Pyramiding greater early vigour and integrated transpiration efficiency in bread wheat; trade-offs and benefits. Field Crops Research 183, 102-110.

Xavier, A., Hall, B., Hearst, A.A., Cherkauer, K.A., Rainey, K.M., 2017. Genetic Architecture of Phenomic-Enabled Canopy Coverage in Glycine max. Genetics 206, 1081-1089.

Yu, K., Kirchgessner, N., Grieder, C., Walter, A., Hund, A., 2017. An image analysis pipeline for automated classification of imaging light conditions and for quantification of wheat canopy cover time series in field phenotyping. Plant Methods 13, 15.

Table 3.A. Repeatability (H^2) of visual and digital ground cover (GrndCov) across seven environments (year × site × planting).

Environment	DAS†	GrndCov	H^2
17LDH.Normal	37	Digital	0.61
	38	Visual	0.54
17JBL.Normal	38	Digital	0.24
	45	Visual	0.32
17PUS.Normal	34	Digital	0.27
	41	Visual	0.27
18LDH.Normal	29	Visual	0.59
	37	Digital	0.61
18LDH.Early	33	Digital	0.70
	44	Visual	0.61
18JBL.Normal	37	Visual	0.57
	46	Digital	0.58
18PUS.Normal	44	Visual	0.04
	59	Digital	0.32

†DAS- Days after sowing

Table 3.B. Genetic correlations between visual and UAS-based digital ground cover recorded at six environments (year × site × planting).

Environment	Visual DAS†	Digital DAS†	r_{geno} ‡
17LDH.Normal	38	37	0.68 ± 0.11
17JBL.Normal	45	38	0.91 ± 0.06
17PUS.Normal	41	34	0.85 ± 0.08
18LDH.Normal	29	33	0.99 ± 0.09
18LDH.Early	44	33	0.92 ± 0.06
18JBL.Normal	37	46	0.76 ± 0.15

†DAS- Days after sowing

‡ r_{geno}- Genetic correlation ± Standard Error

Table 3.C. Repeatability (H^2) and phenotypic correlation ($r_{GY\text{-}GC}$, ground cover and grain yield) of the visual and digital ground cover time-point pairs across environments.

Environment	DAS	GrndCov	H^2	$r_{GY\text{-}GC}$	$H^2 \times r_{GY\text{-}GC}$
17LDH.Normal	37	Digital	0.61	0.21	0.13
	38	Visual	0.54	0.20	0.11
17JBL.Normal	38	Digital	0.24	0.28	0.07
	45	Visual	0.32	0.32	0.10
17PUS.Normal	34	Digital	0.27	0.30	0.08
	41	Visual	0.27	0.27	0.07
18LDH.Normal	29	Visual	0.59	0.03	0.02
	37	Digital	0.61	-0.01	-0.01
18LDH.Early	40	Digital	0.31	0.16	0.05
	44	Visual	0.61	0.01	0.01
18JBL.Normal	37	Visual	0.57	0.43	0.25
	46	Digital	0.58	0.19	0.11
18PUS.Normal	44	Visual	0.04	0.30	0.01
	59	Digital	0.32	0.31	0.10

Table 3.D. Additive allelic effect of the significant markers for grain yield and ground cover. The allelic effect is interpreted as the average substitution effect of the allele at QTL across environments.

Group	SNP	GRYLD (kg ha^{-1})	GrndCov (%)	MAF
575-666 °Cd	S2B_559049321	25.1*	0.24**	0.42
	S2B_560269926	25.5*	0.24**	0.41
	S2B_564924279	23.1*	0.27**	0.45
	S2B_565068485	23.4*	0.27**	0.45
	S5B_49665997	28.0*	0.16*	0.43
	S5B_50528895	25.5*	0.21**	0.37
881-925 °Cd	S2B_799538648	-23.1*	0.35**	0.15
	S2B_800464700	-31.3**	0.43**	0.13
	S3A_21408488	19.5*	-0.19**	0.46
	S6A_39465299	-34.7**	-0.15**	0.44
	S7A_448594243	31.8*	0.21*	0.09
	S7B_83987519	-21.6**	0.12*	0.31

p-value significance: * 0.01; ** 0.001

MAF- Minor allele frequency

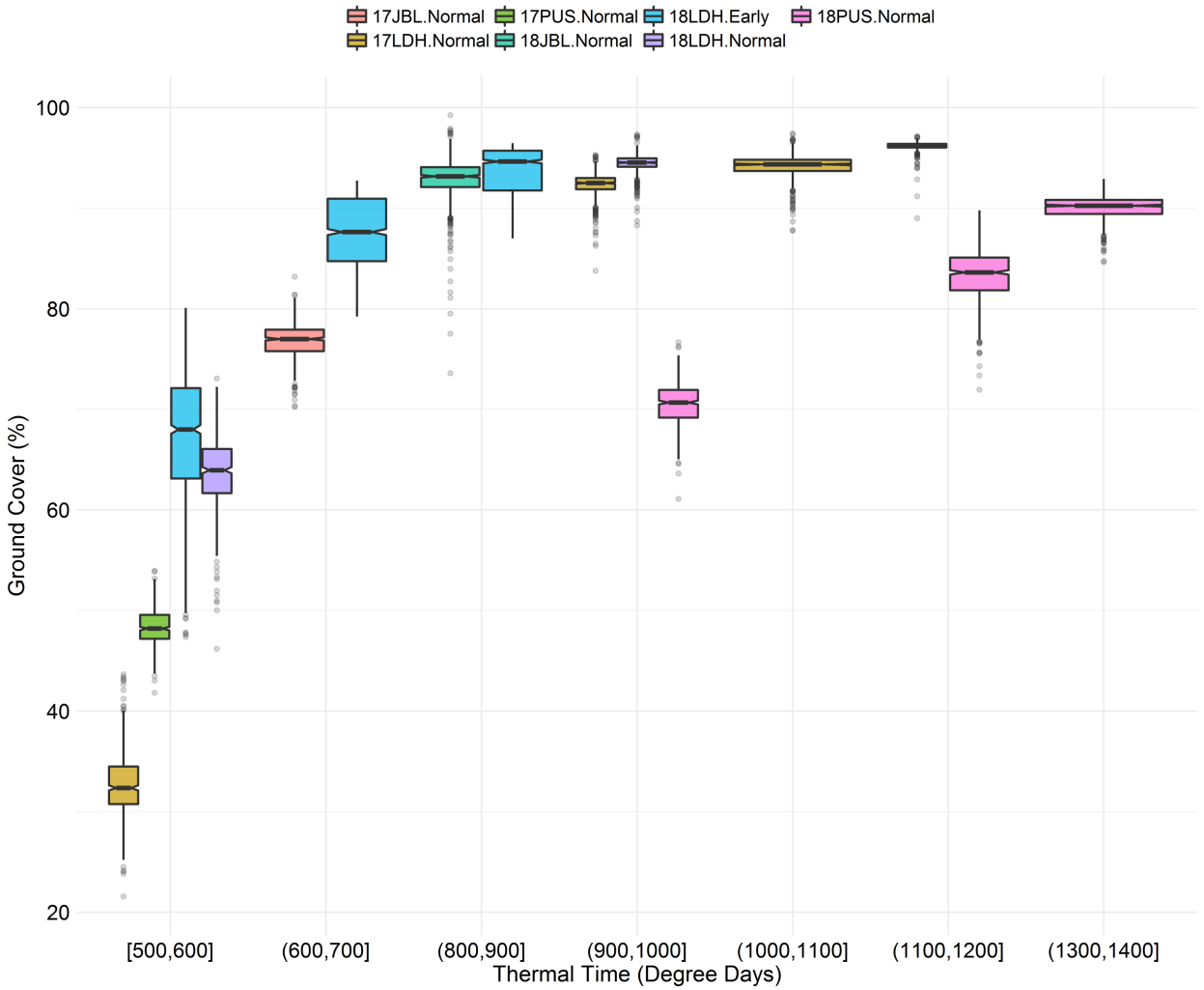


Figure 3.1. Temporal trend of digital ground cover measurements collected from seven environments and expressed as a function of thermal time.

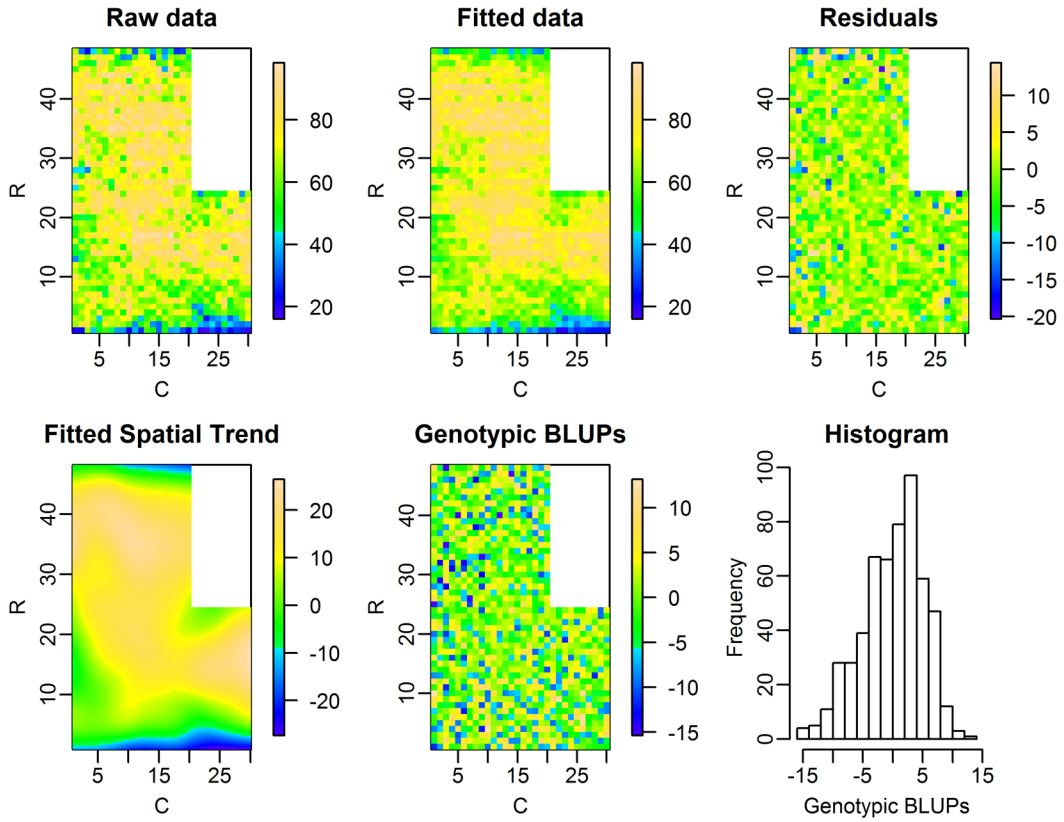


Figure 3.2. An example illustrating the modeling of field spatial variation with 2D splines at 18LDH.Early environment for digital ground cover. The terms C and R refer to the column and row layout factors, respectively.

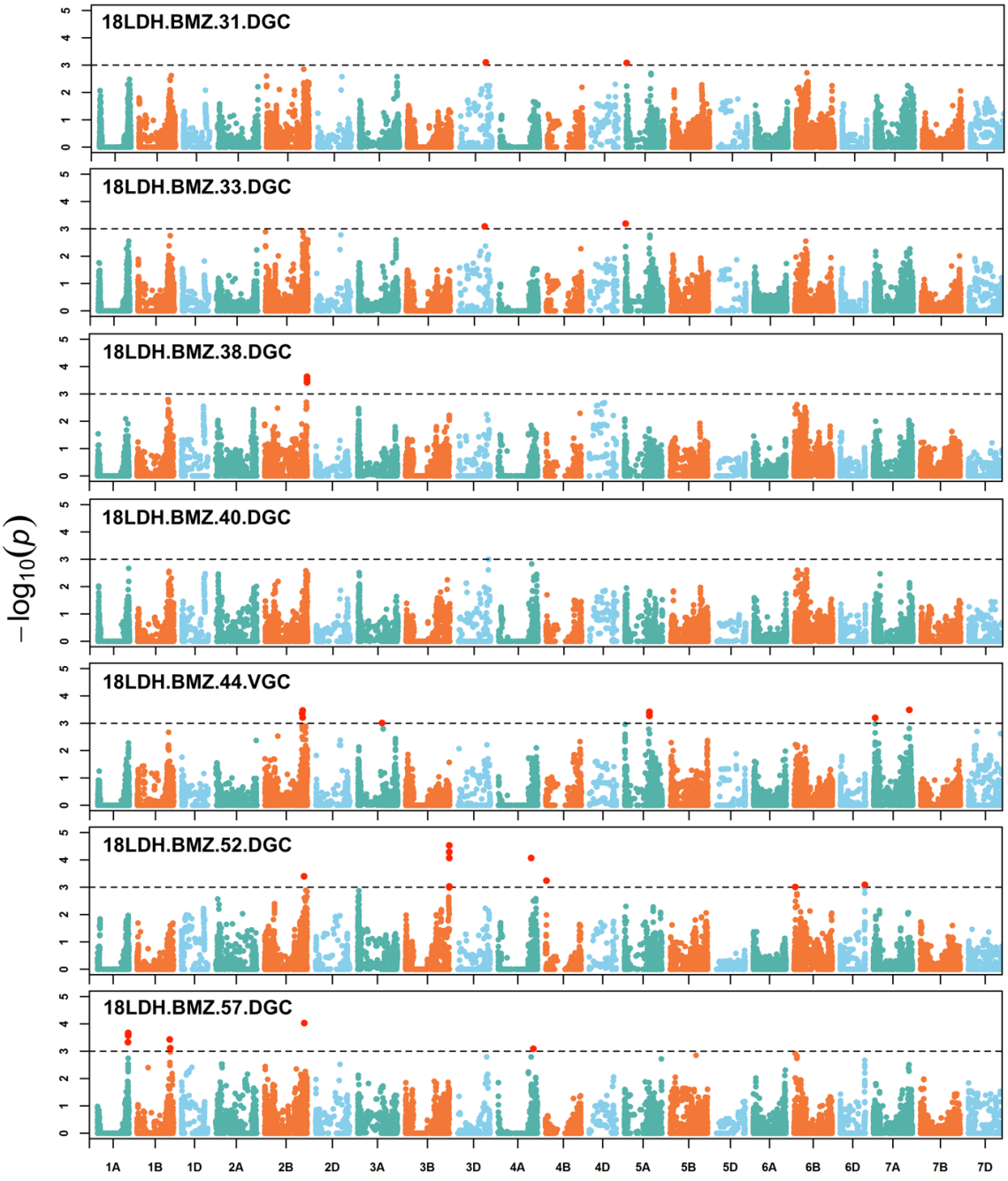


Figure 3.3. GWAS Manhattan plots of ground cover measurements from 18LDH.Early environment. Each panel corresponds to days of measurements shown as Days after Sowing (DAS). The hexaploid wheat sub-genomes are marked by three distinct colors. Abbreviations: DGC, Digital Ground Cover; VGC, Visual Ground Cover. The dashed lines indicate the genome-wide significance threshold.

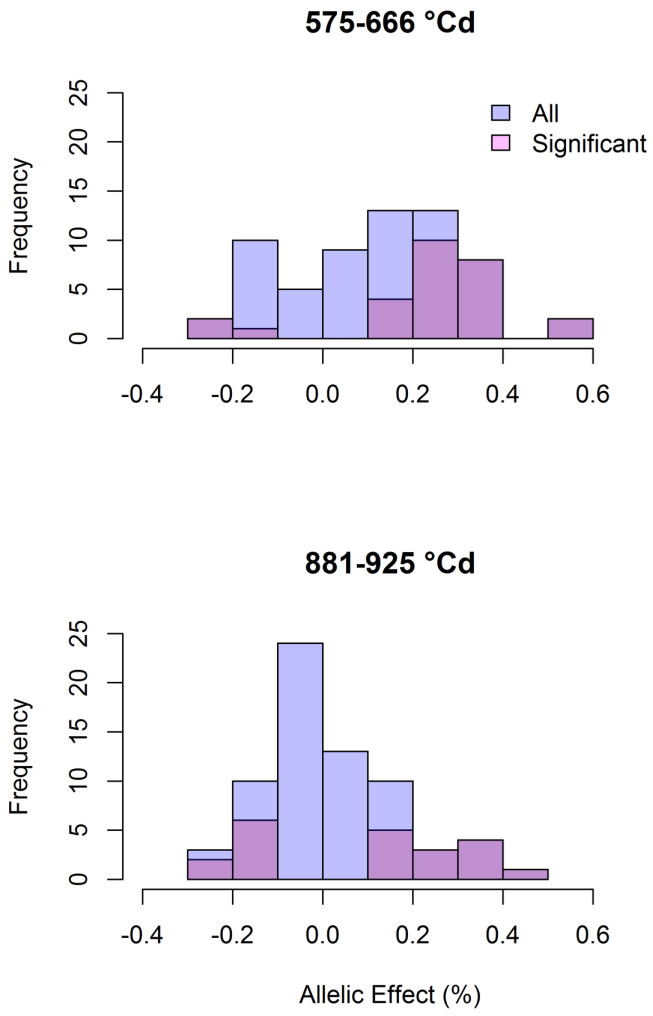


Figure 3.4. Environment-adjusted allelic substitution effect of GWAS significant markers for ground cover at early and advanced wheat growth stages. a) Group I: early growth stage data representing three environments (17LDH.Normal, 17JBL.Normal, and 17PUS.Normal) from 575-666 thermal days (°Cd); b) Group II: advanced growth stage representing four environments (18LDH.Early, 18LDH.Normal, 18JBL.Normal, and 18PUS.Normal) from 881-925 °Cd thermal days window.

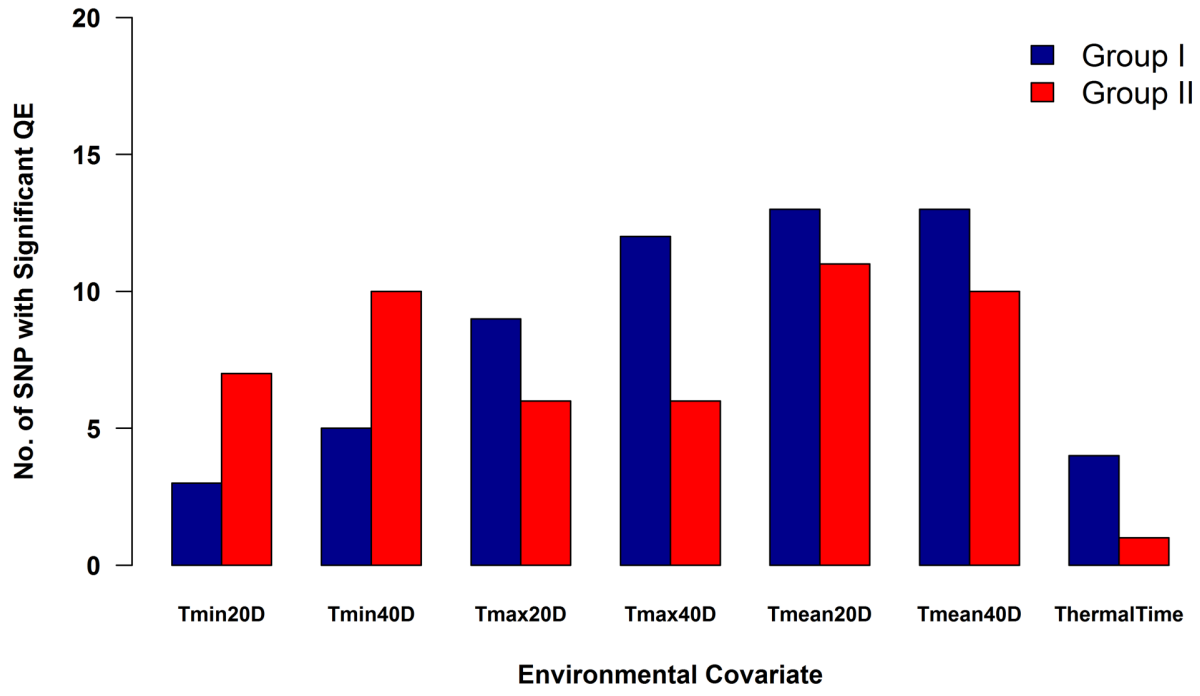


Figure 3.5. The number of SNPs with significant QTL-by-Environment interactions (QEI) for temperature-based environmental covariates. Each group is divided into two groups of thermal time windows: Group I (575-666°Cd) and Group II (881-925°Cd).

Chapter 4 - Large-scale In-Field Digital Phenotyping Reveals Genomic Regions Regulating Crop Growth Dynamics in Wheat

Abbreviations

BLUE, best linear unbiased estimates; CIMMYT, International Maize and Wheat Improvement Center; DAS, days after sowing; GBS, genotyping-by-sequencing; GCP, ground control point; GWAS, genome-wide association studies; HTP, high-throughput phenotyping; JBL, Jabalpur; LDH, Ludhiana; MAF, minor allele frequency; PUS, PUSA; SNP, single nucleotide polymorphism; UAS, unmanned aerial systems.

Abstract

Optimization of morphological and developmental characteristics is critical to maximize the yield potential in crop varieties. Plant height is a key agronomic trait that is an important breeding target in many crops due to its association with biomass and grain yield. However, the slow and static nature of the conventional measurements is a major bottleneck to elucidate the genetic basis of the temporal dynamics of plant height in field experiments. By using the unmanned aerial systems (UAS) estimated canopy height measurements, we demonstrate the considerable power of field phenomics to dissect the genetic and physiological underpinnings of growth components in 546 elite CIMMYT spring wheat breeding lines grown in normal and early planting field experiments in South Asia. Plant height estimates were extracted for 2400 plots at 34 timepoints using digital elevation models from UAS imaging. A logistic growth function was fit to derive genotype specific growth parameters, namely: upper asymptote (final height), slope (growth rate), and inflection point (time of maximum growth rate). These growth

parameters exhibited very high heritability (0.82-0.93) in the two studied environments. By leveraging these highly heritable phenotypic measurements, we found 35 association signals. Multiple coincident signals at known developmental genes were observed for growth rate, heading and maturity dates, as well as agronomic traits, suggesting a significant genetic interdependence of morphological and developmental processes in wheat. Significant physiological tradeoffs associated with faster crop growth in two environments were also uncovered. Through integration of dynamic height measurements with physiology and genomics, our study provides strong evidence in support of large-scale, field-based phenomics to uncover the biological underpinnings of complex plant developmental traits.

Keywords: Triticum aestivum, Logistic regression, plant height, genome-wide association studies GWAS, wheat breeding, high-throughput phenotyping, unmanned aerial systems UAS, genomics

Introduction

Crop growth is a highly dynamic biological process which is an outcome of an intricate interplay of genetics, environment, and their interaction (Angus et al., 1981; Hammer et al., 2006). Plant height is a key component of crop growth in plants. Although the trait has been extensively studied, it remains an important breeding target due to its role in plant architecture and consequently biomass and grain yield and quality. Thus, better understanding of the genetic components, and discovery of alternate sources of variation for height in wheat (*Triticum aestivum* L.) is highly desired for crop improvement (Mo et al., 2018; Würschum et al., 2018).

Height measurements in the field are typically recorded manually with a measuring stick and is relatively low-throughput, particularly for early growth stages with small plant heights. Consequently, due to the low-throughput nature, these measurements are often taken once or twice during the season in large experimental plots. This process however, fails to capture the dynamic temporal changes of height during the season and may miss important biological signals. By leveraging the temporal growth measurements in controlled environments, novel insights into the dynamic genetic control of plant growth have been illustrated (Chen et al., 2014; Honsdorf et al., 2014; Bac-Molenaar et al., 2015; Campbell et al., 2017). However, not much progress has been achieved in terms of large-scale dynamic monitoring of crop growth in field breeding and genetics experiments (Xavier et al., 2017). Thus, scalable in-field high-throughput phenotyping (HTP) techniques are highly desired to unravel the biological underpinnings of the dynamic developmental trait such as height.

Unmanned aerial systems (UAS)-based HTP can facilitate the scanning of tens-of-thousands of field plots at an impressive throughput (Haghighattalab et al., 2016; Duan et al., 2017; Singh et al., 2018b). The improved temporal resolution of UAS-based measurements led to important insights into the genetic components of ground cover and drought stress in soybean and wheat (Xavier et al., 2017; Condorelli et al., 2018). Although at a relatively smaller spatial and temporal scale, the field studies have demonstrated the estimation of crop height via UAS-derived three-dimensional stereo images (Madec et al., 2017; Malambo et al., 2018; Wang et al., 2018). Therefore, UAS-based HTP can be a suitable platform to study the temporal dynamics of height and other developmental traits in large breeding and genetic experiments.

The identification and introduction of reduced height (*Rht*) loci in wheat has been an important hallmark of the green revolution (Hedden, 2003). Among the major *Rht* genes, gibberellin (GA)-insensitive *Rht-B1* and *Rht-D1* homeoloci are the most extensively utilized in

the breeding programs worldwide. However, *Rht-B1b* and *Rht-D1b* alleles were found to be associated with reduced anther extrusion, increased susceptibility to fusarium head blight, reduced drought tolerance due to shorter coleoptile length, and in certain environments reduced grain yield (Richards, 1992; Rebetzke et al., 2001; Miedaner and Voss, 2008). Therefore, alternate sources of *Rht* alleles are suggested to avoid the negative pleiotropic effects of major *Rht-1* loci. More recently a novel loci *Rht24* on chromosome 6A has been found to have comparable effect on height as that of major *Rht-1* loci but without any negative impact on anther extrusion (Tian et al., 2017; Würschum et al., 2018). Studies are also underway to find other alternative dwarfing loci with relatively milder pleiotropic effects on other agronomic traits in wheat (Mo et al., 2018). Thus, a better understanding of alternative height loci regulating plant height is critical to fine-tune the harvest index and consequently yield and stress adaptability in wheat.

By taking advantage of the UAS-acquired three-dimensional digital elevation models (DEM) and dense genomic markers, we sought to dissect the physiological and genetic components of dynamic crop height trait in wheat. By using the multi-temporal plot-level measurements, we estimated three key crop growth parameters of plant height, namely: final height (θ_1), growth rate (θ_2), the time of maximal growth rate (θ_3), in 546 elite CIMMYT spring wheat breeding lines grown in early and normal planting dates in field experiments. The genome-wide association analysis revealed multiple coincident genomic signals for known and unknown developmental genes in crop growth, phenology, and morphology in wheat. Thermal time dependent physiological and genetic relationships between early and normal planting dates were established, which could be useful in making appropriate breeding selection decisions. This highly scalable approach can be applied to large number of experimental plots for dynamic

quantification of plant height, crop growth rate, as well as lodging and ground cover in different crops.

Materials and Methods

Plant material and field layout

Advanced spring wheat (*Triticum aestivum* L.) breeding lines from CIMMYT's South Asia Bread Wheat Genomic Prediction Yield Trials were planted in two sowing dates (early planting, 23 Oct 2017; normal planting, 17 Nov 2017) at the Borlaug Institute for South Asia's Ludhiana (LDH), Punjab, India (30°59' N and 75°44' E) site in wheat season 2018. Briefly, a total of 546 lines were planted in an alpha-lattice design in six-row plots of size 5 m^2 in each experiment. The field plots were further divided into smaller blocks of 120 plots called trials. Each trial consisted of two replicate blocks each divided into 6 subblocks, where each replicate block contained 53 entries and 7 check varieties. Further layout details can be found in chapter 2.

UAS Data Acquisition and Processing

The UAS flights were made at regular intervals during the course of the field season in both experiments. Raw images captured by the RedEdge camera (Micasense Inc., USA) were processed in Agisoft PhotoScan Pro (Version 1.3.1, Agisoft LLC, Russia) following the internally established protocols in Poland Lab. The UAS data acquisition and pre-processing methods are described in detail in chapter 2.

Digital Canopy Height

A two-step strategy was implemented for automated plot-level digital height extraction in *python* (Fig. 4.1). First, an expanded shapefile that included the surrounding bare soil areas was used to estimate the base elevation for each plot. Base height was estimated by taking the arithmetic mean of the lowest 1-2 percentile from the overall distribution of DEM points within

the expanded shapefile. Second, a buffered (shrunk) version of the original shapefile was used to estimate the mean of the digital points falling within the top 85% and 95% percentile distribution. Subsequently the plot-level absolute height was then calculated by subtracting the top and bottom height per each plot. This memory efficient *python*-based implementation allowed us to extract the plot-level height measurements from the entire data set of 34 DEM rasters and ~15Gb within a few minutes on a single computing node.

Logistic Growth Model

To estimate the genotype-specific crop growth parameters, the following logistic growth model was fit in *nlme* package in *R* (Pinheiro et al., 2014):

$$\mu\,(t) = \frac{\theta_1}{1+\exp^{(-\theta_2(t-\theta_3))}} \qquad [1]$$

where, $\mu(t)$ is the genotypic mean at the time-point t, θ_1 is the upper asymptote (*i.e.,* final height), θ_2 is the growth rate (*i.e.,* change in height over the time-course), θ_3 is the inflection point (*i.e.,* time of maximum growth rate), t is the accumulated thermal time (Table 4.A).

Visual Measurements

In addition to the UAS-based imagery, visual plant height ratings were collected after heading stage to "ground-truth" the digital measurements. To further assess the relationship of height and growth parameters with agronomic measurements, secondary phenotypic data was also collected during the growing season for each experiment. The description of traits and their ontologies are provided in chapter 1 and 2.

Best Linear Unbiased Estimates (BLUE)

The adjusted genotypic means (BLUE) were calculated using the following linear mixed model in *lme4* package in R (Bates et al., 2014) as follows:

$$y_{ijkl} = \mu + G_i + Z_j + R_{k(j)} + P_{c(j)} + Q_{r(j)} + M_{kl(j)} + e_{ijkl} \qquad [2]$$

where y_{ijkl} is the phenotypic response variable, μ is the fixed overall mean, G_i is the fixed genotype effect, Z_j is the random trial effect, $R_{k(j)}$ is the random effect of replicate nested within a trial, $P_{c(j)}$ is the random effect of columns nested within a trial, $Q_{r(j)}$ is the random effect of rows nested within a trial, $M_{kl(j)}$ is the random effect of sub-blocks and replications nested within a trial, and e_{ijkl} is the residual effect. The random terms were assumed to be *IID Normal*.

Broad-sense Heritability / Repeatability

The matric of broad-sense line-mean heritability or repeatability was used to evaluate the quality of phenotypic measurements. The repeatability on entry-mean basis for each trait was calculated following:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_e^2/r} \quad [3]$$

where σ^2_G is the genotypic variance, σ^2_e is the residual variance, and r is the number of replicates. The variance components in *equation 3* were calculated by setting the genotype term as random in *equation 2*.

Genetic Correlations

Marker based genetic correlations between a pair of traits were calculated with *sommer (Covarrubias-Pazaran, 2016)* package in R as follows:

$$r_{g(x,y)} = \frac{cov_g(x,y)}{\sqrt{var_g(x) * var_g(y)}} \quad [4]$$

where $cov_g\,(x,y)$ is the covariance of the traits x and y; $var_g\,(x)$ and $var_g\,(y)$ is the variance of traits x and y, respectively.

Genome-wide Association Analysis (GWAS)

Following the quality filtering, a set of 19217 high-quality genotyping-by-sequencing based SNPs were retained for subsequent genetic analyses (refer to chapter 2 for processing details). A mixed linear model algorithm called *FarmCPU* was implemented via *GAPIT*

package in *R* (Liu et al., 2016). The algorithm adjusts the genetic background effect and population structure in an iterative and efficient fashion. A false discovery rate (FDR) of 0.05 was used to test the statistical significance of individual markers.

Results and Discussion

High-throughput phenotyping of Wheat Breeding Trials

To assess the yield potential and agronomic performance of elite breeding lines, early-planting and normal planting wheat trials were established at Ludhiana (LDH) in NW India in 2018 season. Throughout the growing season, 34 autonomous phenotyping flights were conducted with a UAS equipped with multi-spectral digital camera (Fig. 4.2). Mid-day flight missions covered over 2,400 plots in both trials in ~50 minutes and obtained a DEM spatial resolution of ~3 cm. Plot-level digital plant height was subsequently extracted using an automated image processing pipeline in *python*. In each of the experiments, the plant height (PH) was visually scored as a 'ground-truth' for subsequent validation of the image-derived crop growth parameters. The breeding lines in the trials showed considerable phenotypic variation, and repeatability of 0.85 to 0.87 for the PH (Table 4.B).

The growth trajectory of the plots at each environment closely followed a logistic growth pattern (Fig. 4.2). The logistic growth can be mathematically defined by three parameters θ_1, θ_2, θ_3 (Table 4.A). Due to its resemblance to many developmental processes, the logistic growth parameters have been extensively used to describe key plant growth processes (Malosetti et al., 2006; Chen et al., 2014). In our experiments, all growth parameters could be estimated with very high accuracy as reflected by the high broad-sense heritability (H^2 = 0.82-0.93; Table 4.B). The maximum height parameter (θ_1) was genetically highly correlated with "ground truth" plant

height (PH) in both experiments (r_{geno} = 0.90-0.99, data not shown), supporting the effectiveness of digital crop growth measurements in capturing the genetic variation for plant height in wheat.

Temporal Dynamics of Crop Growth

The genotypes exhibited considerable temporal variation for growth in both environments (Fig. 4.2 and 4.3). The genotypes in normal-planting trial registered 17% faster growth rate (θ_2) and a 200 thermal units shorter inflection point (θ_3) compared to the early-planted trial (Fig. 4.2). A faster and shorter growth cycle in normal-planted trial underscores the physiological adaptation in response to the constrained thermal period of a normal growing season. Similar adaptive mechanisms in response to varying environmental conditions have been previously described in wheat (Sadras and Monzon, 2006; Xie et al., 2016). To ascertain if this adaptive response had a genetic basis, we calculated broad-sense heritability (H^2) across two trials by including a genotype-by-environment (G × E) component in the linear model. The H^2 of θ_2 and θ_3 was 80% and 87%, suggesting a strong genetic component underlying the differential growth rate.

Genetic Basis of Crop Growth

To elucidate the genetic controls of plant growth using the digital measurements, we conducted a genome-wide association scan on 546 elite breeding lines. We found significant associations near major wheat developmental genes *PpD-D1* (chr 2BS), *PpD-B1* (chr 2DS) as well as associations on 5BL and 7DS (Fig. 4.4). Many of the association signals for growth parameters coincided with these major developmental genes (Supplementary Fig. D.1-D.2). Previous reports in wheat have highlighted the close relationship of developmental and growth traits (Addisu et al., 2008; Wilhelm et al., 2013). In addition, we found growth rate specific signals at 1DL and 4BL in both trials. The signals on 2DL, 6AS, 6BS, and 6BL were trial specific, suggesting the environment specific differential expression of some genes. We also

observed an increased number of signals for inflection point trait in normal planted trial compared to the early-planted trial. These observations suggest a strong genetic component of crop growth, the effect of which is highly dependent on environment. The ability to dissect these genetic components in the field grown crop populations can greatly increase our understanding of the interplay of the genes that shape the phenotypes of plants under varying environmental conditions.

Close Relationship of Crop Growth and Phenology

Early but faster growth rate (i.e., low θ_3 and high θ_2) was associated with early heading and maturity in both trials (Fig. 4.5), though the strength of association was higher in early planted trials compared to the normal planted trials. The temperature dependent changes in phenology have been previously described in wheat (Asseng et al., 2004; Sadras and Monzon, 2006; Xie et al., 2016). Together with the observation of several coincident QTL in the GWAS analysis, and consistent with the typical coincidence of time of heading and onset of slow growth rate in determinate cereals, our results suggest a shared genetic control between growth and phenological processes in wheat.

Crop Growth and Morphology

Faster growth (θ_2) was associated with shorter final height (θ_1) (r = -0.17 to -0.38; P < 0.01) and early onset of decreasing growth rate (θ_3) (r = - 0.21 to -0.41, P < 0.01) in both planting dates (Fig. 4.5). Furthermore, delayed but faster growth rate (i.e., increased θ_3 and θ_2) was associated with reduced ground cover (r = -0.07 to -0.26; P < 0.05). The growth rate had a positive relationship with manually measured height in early planting trials and negative relationship in normal planted trial. Interestingly, higher lodging incidence was associated with faster growth, early onset of decreasing growth, and early heading and maturity in both trials. The absolute strength of this association was higher in early planted trials, indicating the fast but

shorter growth in early planted trials may increase the vulnerability of plants to lodging. Earlier studies and our own experiments (*see* chapter 1) have described the increased sensitivity of taller, high yielding and early flowering genotypes to lodging in wheat (Keller et al., 1999; Berry and Berry, 2015).

Crop Growth and Agronomic Yield

Early heading, late maturity, and early onset of slow growth was associated with higher grain yield in general and more significantly in normal planting trial compared to late planted trial (Fig. 4.5). Furthermore, a positive, although insignificant, association of growth rate and grain yield in normal planting trials suggest that the quick plant growth is more favorable for grain yield in short cycle environments. Shorter growth cycles were shown to be associated with stress avoidance in wheat cultivars (Lopes and Reynolds, 2011; Kumar, 2018). The early planted trials had higher grain yield compared to the normal planted trials, suggesting the extended growth period and favorable weather conditions positively impact the grain yields in these environments (Fig. 4.6). Physiological trade-offs associated with grain filling duration and yield traits have been reported previously in wheat (Lizana and Calderini, 2013; Xie et al., 2016). Therefore, a better understanding of the physiological tradeoffs under differing environments can facilitate better decision-making in regard to the breeding selection in target population of environments. To this end, the phenomics based approaches with its scalable and high-throughput potential can easily offset the increased burden of more frequent sampling times for growth analysis.

Conclusions

Due to strong molecular and physiological interlinking of crop growth processes, a finer understanding of key growth components is required to maximize the genetic gains in crops. The dynamic nature of the growth traits requires a season-long monitoring, which is particularly

challenging in natural field environments. By using the high-throughput UAS-enabled phenotyping, we extracted high quality plant height measurements on 2400 plots at 34 time-points from early and normal planted field experiments in wheat. By integrating highly accurate digital growth phenotypes with whole-genome association analysis, we show a considerable power of phenomics based approaches in dissecting the physiological and genetic underpinnings of dynamic height trait in field experiments. A better understanding of physiological tradeoffs as well as the ability to overcome negative pleiotropic effects through alternative height genes can allow breeders to maximize yield potential of wheat. This reproducible and highly scalable methodology has potential to assay thousands to tens-of-thousands of experimental plots across entire breeding programs. We envision that the full utility of these phenomics-based approaches will be realized by integration with whole-genome prediction, ecophysiological crop growth models, and other molecular biology tools.

References

Addisu, M., Snape, J.W., Simmonds, J.R., Gooding, M.J., 2008. Reduced height (Rht) and photoperiod insensitivity (Ppd) allele associations with establishment and early growth of wheat in contrasting production systems. Euphytica 166, 249.

Angus, J.F., Mackenzie, D.H., Morton, R., Schafer, C.A., 1981. Phasic development in field crops II. Thermal and photoperiodic responses of spring wheat. Field Crops Research 4, 269-283.

Asseng, S., Jamieson, P.D., Kimball, B., Pinter, P., Sayre, K., Bowden, J.W., Howden, S.M., 2004. Simulated wheat growth affected by rising temperature, increased water deficit and elevated atmospheric CO2. Field Crops Research 85, 85-102.

Bac-Molenaar, J.A., Vreugdenhil, D., Granier, C., Keurentjes, J.J., 2015. Genome-wide association mapping of growth dynamics detects time-specific and general quantitative trait loci. J Exp Bot 66, 5567-5580.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1, 1-23.

Berry, P.M., Berry, S.T., 2015. Understanding the genetic control of lodging-associated plant characters in winter wheat (Triticum aestivum L.). Euphytica 205, 671-689.

Campbell, M.T., Du, Q., Liu, K., Brien, C.J., Berger, B., Zhang, C., Walia, H., 2017. A Comprehensive Image-based Phenomic Analysis Reveals the Complex Genetic Architecture of Shoot Growth Dynamics in Rice (Oryza sativa). The Plant Genome 10.

Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., Klukas, C., 2014. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. Plant Cell 26, 4636-4655.

Condorelli, G.E., Maccaferri, M., Newcomb, M., Andrade-Sanchez, P., White, J.W., French, A.N., Sciara, G., Ward, R., Tuberosa, R., 2018. Comparative Aerial and Ground Based High Throughput Phenotyping for the Genetic Dissection of NDVI as a Proxy for Drought Adaptive Traits in Durum Wheat. Front Plant Sci 9, 893.

Covarrubias-Pazaran, G., 2016. Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. PLoS One 11, e0156744.

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

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

Hammer, G., Cooper, M., Tardieu, F., Welch, S., Walsh, B., van Eeuwijk, F., Chapman, S., Podlich, D., 2006. Models for navigating biological complexity in breeding improved crop plants. Trends Plant Sci 11, 587-593.

Hedden, P., 2003. The genes of the Green Revolution. Trends Genet 19, 5-9.

Honsdorf, N., March, T.J., Berger, B., Tester, M., Pillen, K., 2014. High-throughput phenotyping to detect drought tolerance QTL in wild barley introgression lines. PLoS One 9, e97047.

Keller, M., Karutz, C., Schmid, J.E., Stamp, P., Winzeler, M., Keller, B., Messmer, M.M., 1999. Quantitative trait loci for lodging resistance in a segregating wheat x spelt population. Theor Appl Genet 98, 1171-1182.

Kumar, S., 2018. Impact of Extreme Weather Events on Wheat Yield in Different Agro-Ecological Zones of Middle Indo-Gangetic Plain. Agricultural Research.

Liu, X., Huang, M., Fan, B., Buckler, E.S., Zhang, Z., 2016. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLOS Genetics 12, e1005767.

Lizana, X.C., Calderini, D.F., 2013. Yield and grain quality of wheat in response to increased temperatures at key periods for grain number and grain weight determination: considerations for the climatic change scenarios of Chile. The Journal of Agricultural Science 151, 209-221.

Lopes, M.S., Reynolds, M.P., 2011. Drought Adaptive Traits and Wide Adaptation in Elite Lines Derived from Resynthesized Hexaploid Wheat. Crop Science 51, 1617-1626.

Madec, S., Baret, F., de Solan, B., Thomas, S., Dutartre, D., Jezequel, S., Hemmerle, M., Colombeau, G., Comar, A., 2017. High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates. Front Plant Sci 8, 2002.

Malambo, L., Popescu, S.C., Murray, S.C., Putman, E., Pugh, N.A., Horne, D.W., Richardson, G., Sheridan, R., Rooney, W.L., Avant, R., Vidrinec, M., McCutchen, B., Baltensperger, D., Bishop, M., 2018. Multitemporal field-based plant height estimation using 3D point clouds generated from small unmanned aerial systems high-resolution imagery. International Journal of Applied Earth Observation and Geoinformation 64, 31-42.

Malosetti, M., Visser, R.G.F., Celis-Gamboa, C., van Eeuwijk, F.A., 2006. QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato. Theor Appl Genet 113, 288-300.

Miedaner, T., Voss, H.-H., 2008. Effect of Dwarfing Rht Genes on Fusarium Head Blight Resistance in Two Sets of Near-Isogenic Lines of Wheat and Check Cultivars. Crop Science 48, 2115-2122.

Mo, Y., Vanzetti, L.S., Hale, I., Spagnolo, E.J., Guidobaldi, F., Al-Oboudi, J., Odle, N., Pearce, S., Helguera, M., Dubcovsky, J., 2018. Identification and characterization of Rht25, a locus on chromosome arm 6AS affecting wheat plant height, heading time, and spike development. Theor Appl Genet.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., 2014. R Core Team (2014) nlme: linear and nonlinear mixed effects models. . CRAN.

Rebetzke, G., Appels, R., Morrison, A., Richards, R., McDonald, G., Ellis, M., Spielmeyer, W., Bonnett, D., 2001. Quantitative trait loci on chromosome 4B for coleoptile length and early vigour in wheat (Triticum aestivum L.). Australian Journal of Agricultural Research 52, 1221-1234.

Richards, R., 1992. The effect of dwarfing genes in spring wheat in dry environments. I. Agronomic characteristics. Australian Journal of Agricultural Research 43, 517-527.

Sadras, V.O., Monzon, J.P., 2006. Modelled wheat phenology captures rising temperature trends: Shortened time to flowering and maturity in Australia and Argentina. Field Crops Research 99, 136-146.

Singh, D., Wang, X., Kumar, U., Gao, L., Noor, M., Imtiaz, M., Singh, R.P., Poland, J., 2018. High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. Frontiers in Plant Science (Accepted).

Tian, X., Wen, W., Xie, L., Fu, L., Xu, D., Fu, C., Wang, D., Chen, X., Xia, X., Chen, Q., He, Z., Cao, S., 2017. Molecular Mapping of Reduced Plant Height Gene Rht24 in Bread Wheat. Front Plant Sci 8, 1379.

Wang, X., Singh, D., Marla, S., Morris, G., Poland, J., 2018. Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. Plant Methods 14, 53.

Wilhelm, E.P., Boulton, M.I., Al-Kaff, N., Balfourier, F., Bordes, J., Greenland, A.J., Powell, W., Mackay, I.J., 2013. Rht-1 and Ppd-D1 associations with height, GA sensitivity, and days to heading in a worldwide bread wheat collection. Theor Appl Genet 126, 2233-2243.

Würschum, T., Liu, G., Boeven, P.H.G., Longin, C.F.H., Mirdita, V., Kazman, E., Zhao, Y., Reif, J.C., 2018. Exploiting the Rht portfolio for hybrid wheat breeding. Theor Appl Genet.

Xavier, A., Hall, B., Hearst, A.A., Cherkauer, K.A., Rainey, K.M., 2017. Genetic Architecture of Phenomic-Enabled Canopy Coverage in Glycine max. Genetics 206, 1081-1089.

Xie, Q., Mayes, S., Sparkes, D.L., 2016. Early anthesis and delayed but fast leaf senescence contribute to individual grain dry matter and water accumulation in wheat. Field Crops Research 187, 24-34.

Table 4.A. Description of logistic regression-based crop growth parameters.

Growth Parameter	Description
Upper Asymptote (θ_1)	Final/maximum height
Growth Rate (θ_2)	Rate of growth during the exponential phase
Inflection Point (θ_3)	The time of the maximal growth rate

Table 4.B. The repeatability (within environment) and heritability (across environments) of elite wheat breeding lines measured in two planting date experiments: early and normal planting. Logistic growth parameters (θ_1, maximum height; θ_2, growth rate; θ_3, thermal time at maximum growth rate); PH, plant height; AgrScr, agronomic score; DTB, days to booting; DTHD, days to heading; DAYSMT, days to maturity; FLGL, flag leaf length; FLGLW, flag leaf width; GrndCov, ground cover; GRYLD, grain yield; LOI, lodging incidence; SpkLng, spike length; TGW, thousand grain weight.

Trait	Repeatability		Heritability
	Early	Normal	
θ_1	0.93	0.89	0.89
θ_2	0.88	0.82	0.80
θ_3	0.89	0.87	0.87
PH	0.85	0.87	0.87
AgrScr	0.60	0.47	0.60
DTB	0.95	0.85	0.63
DTHD	0.98	0.97	0.79
DAYSMT	0.89	0.91	0.91
FLGL	0.72	0.80	0.81
FLGLW	0.80	0.83	0.86
GrndCov	0.61	0.59	0.40
GRYLD	0.63	0.65	0.63
LOI	0.70	0.56	0.42
SpkLng	0.72	0.67	0.76
TGW	0.91	0.89	0.91

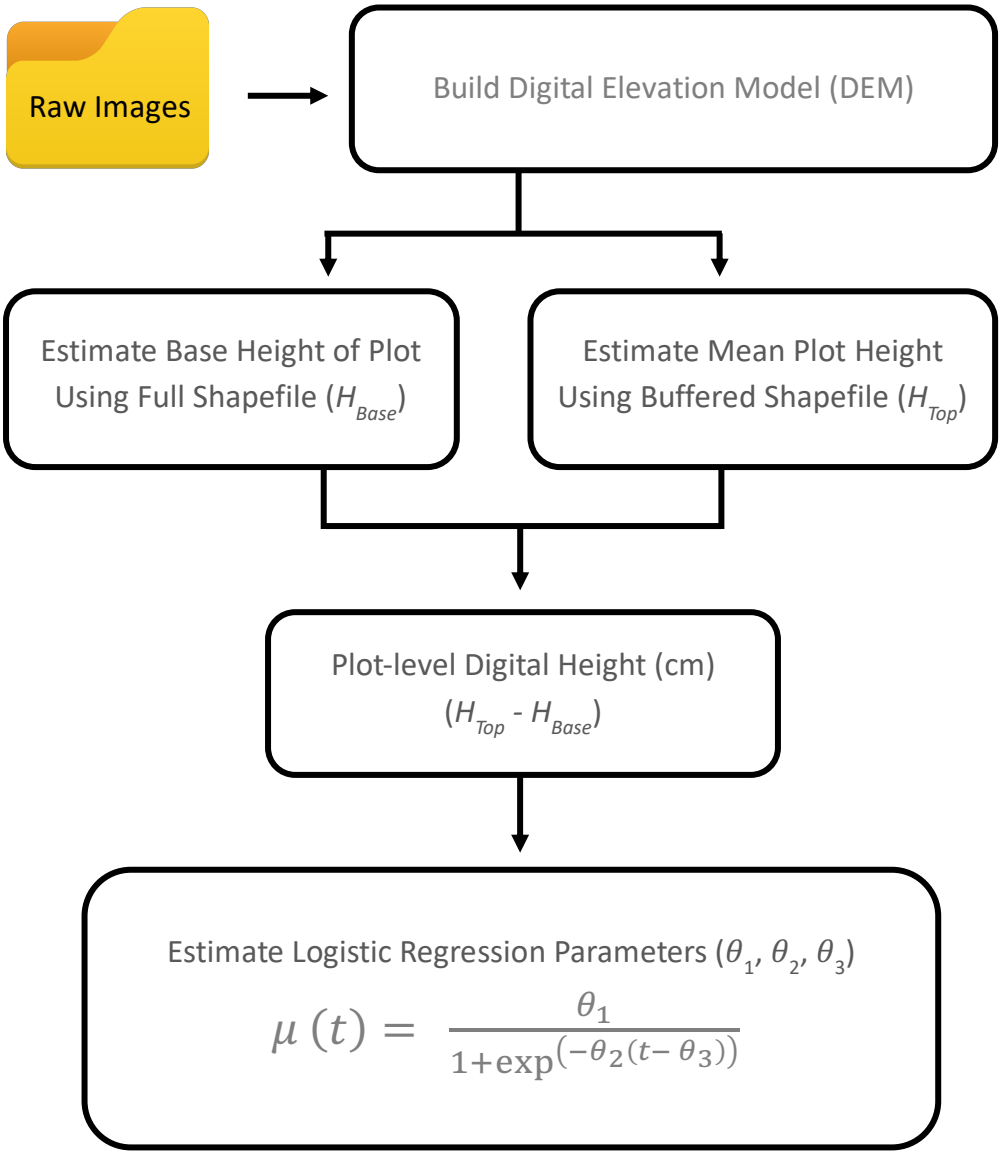


Figure 4.1. Workflow of phenotypic analysis approaches used to assess the digital crop growth in wheat.

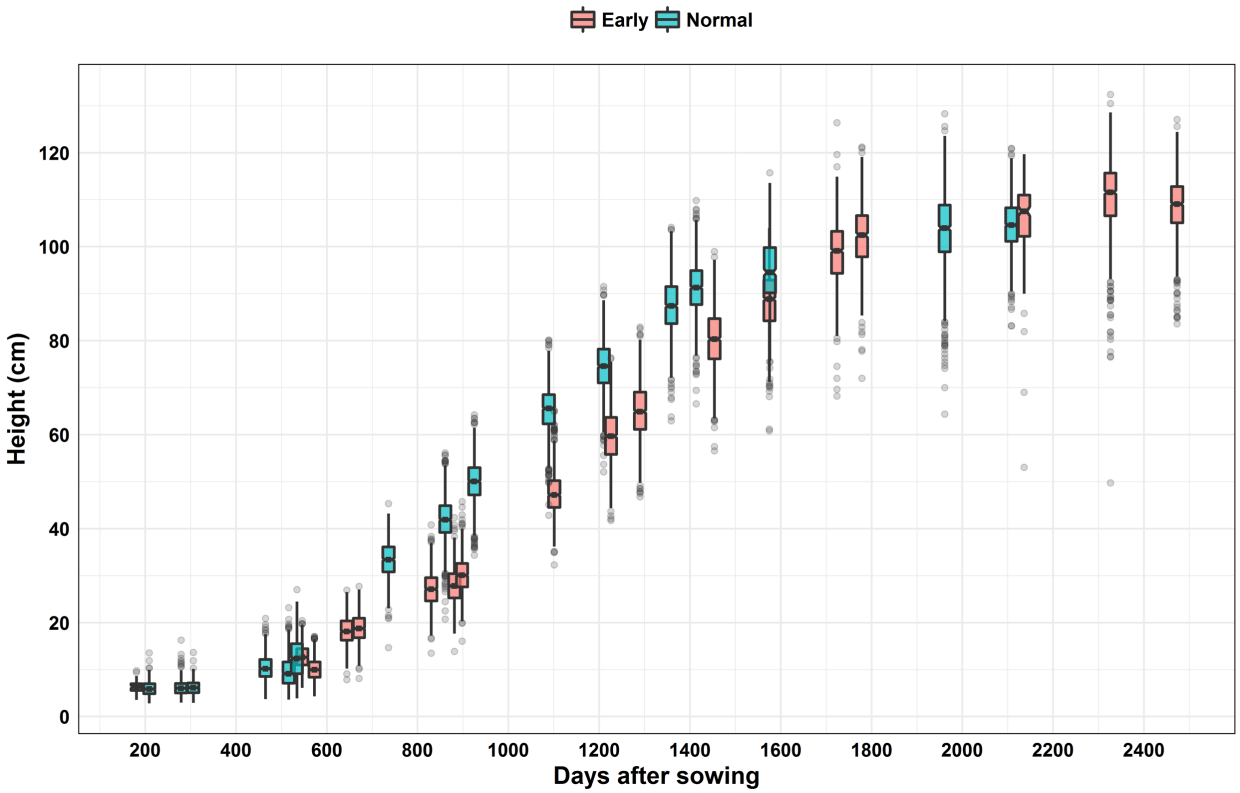


Figure 4.2. Temporal dynamics of UAS-estimated plant height at early and normal planting dates in 2018 in wheat.

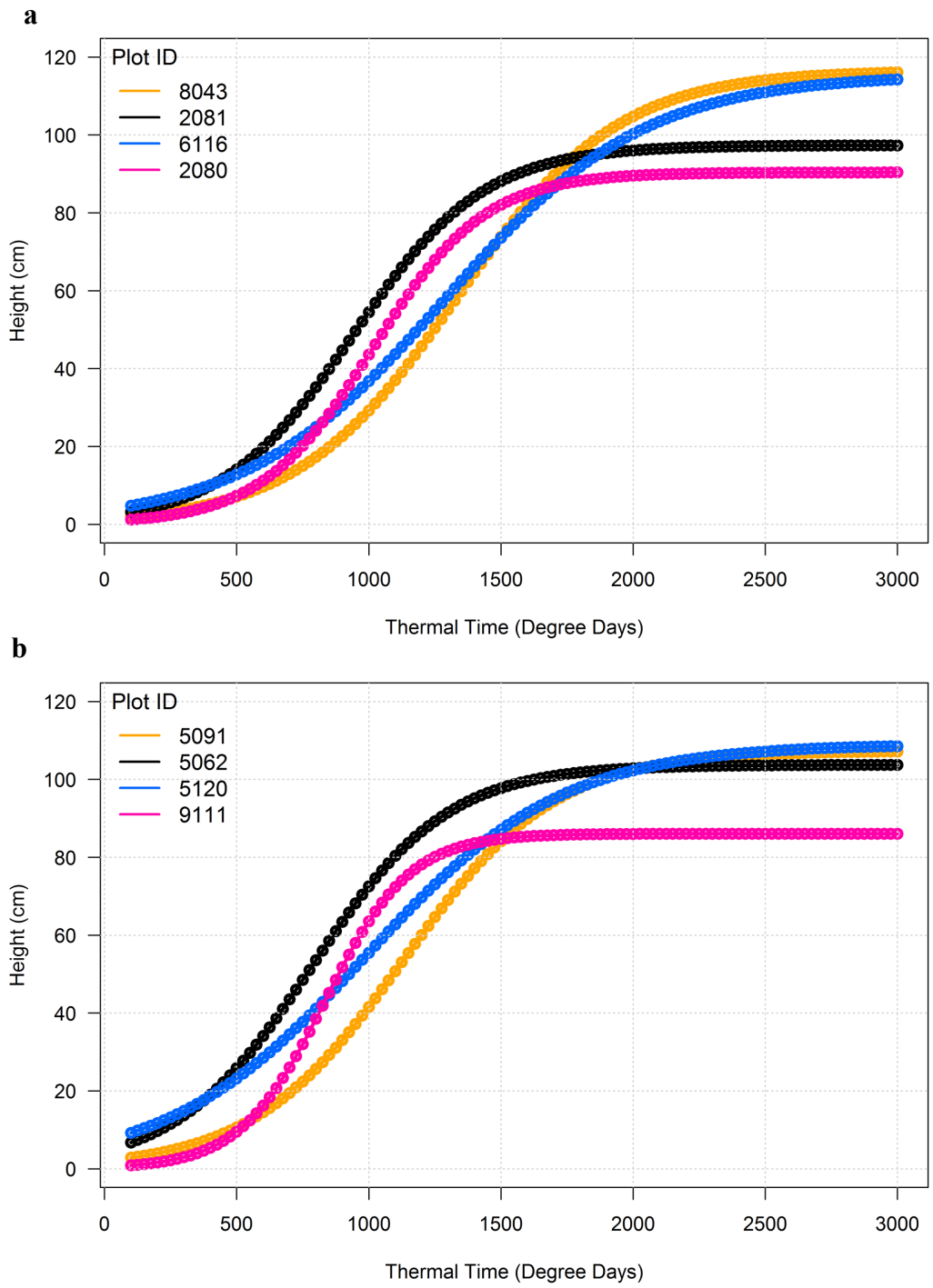
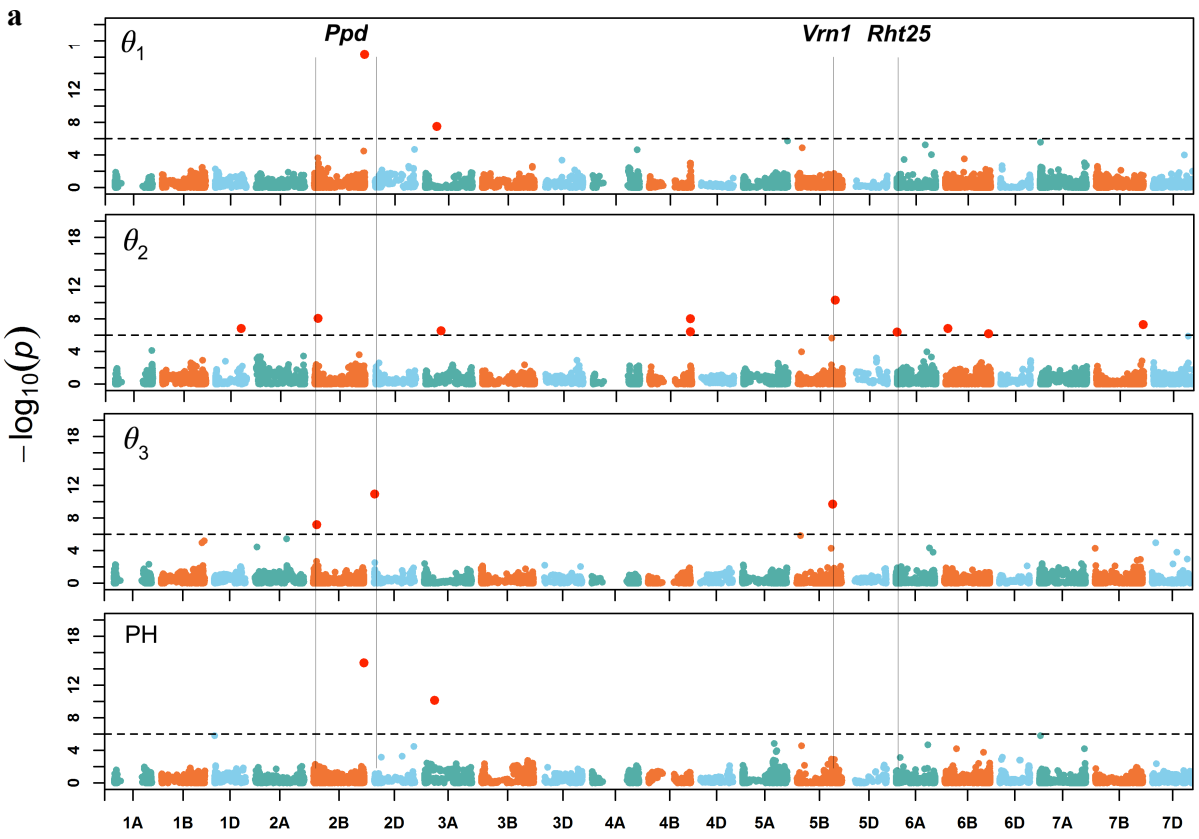


Figure 4.3. Growth trajectories of wheat lines showing variation in digital logistic growth at: a) early, b) normal planting dates.


a
Ppd
Vrn1
Rht25
θ1
θ2
θ3
PH
−log10(p)
1A
1B
1D
2A
2B
2D
3A
3B
3D
4A
4B
4D
5A
5B
5D
6A
6B
6D
7A
7B
7D

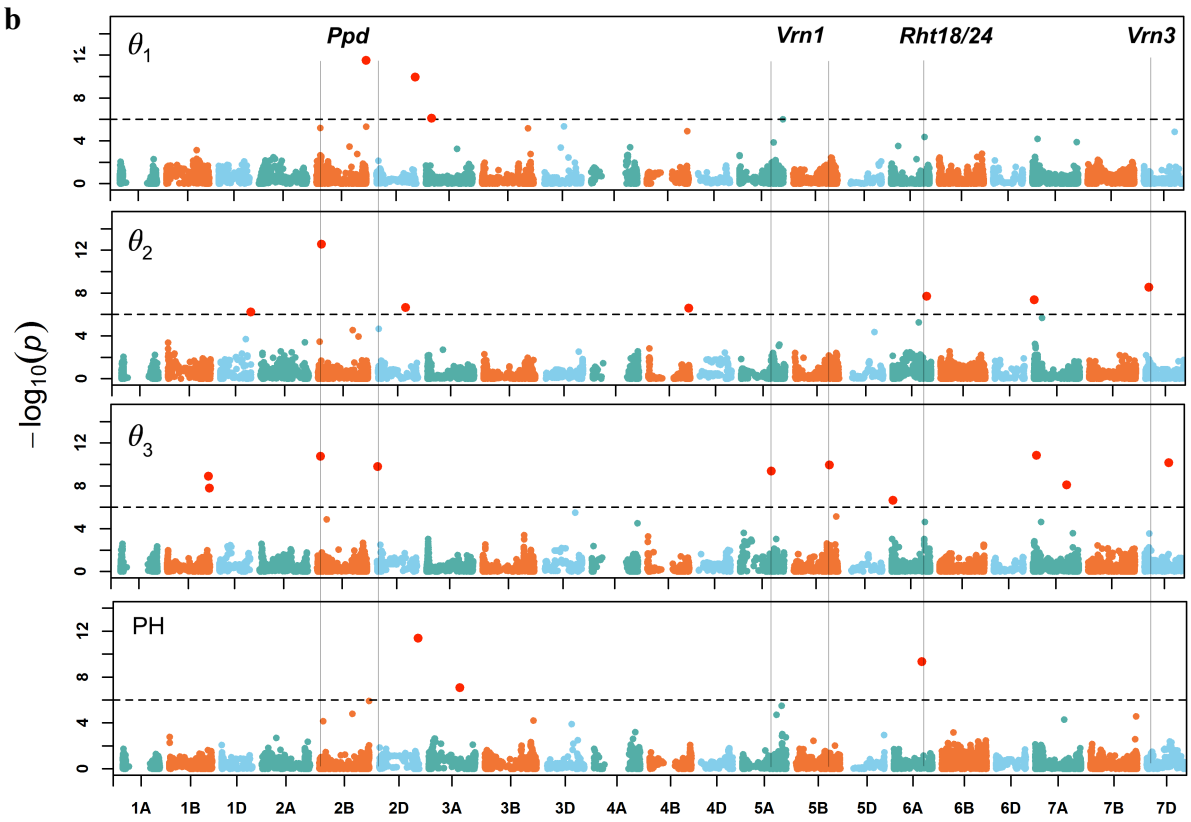
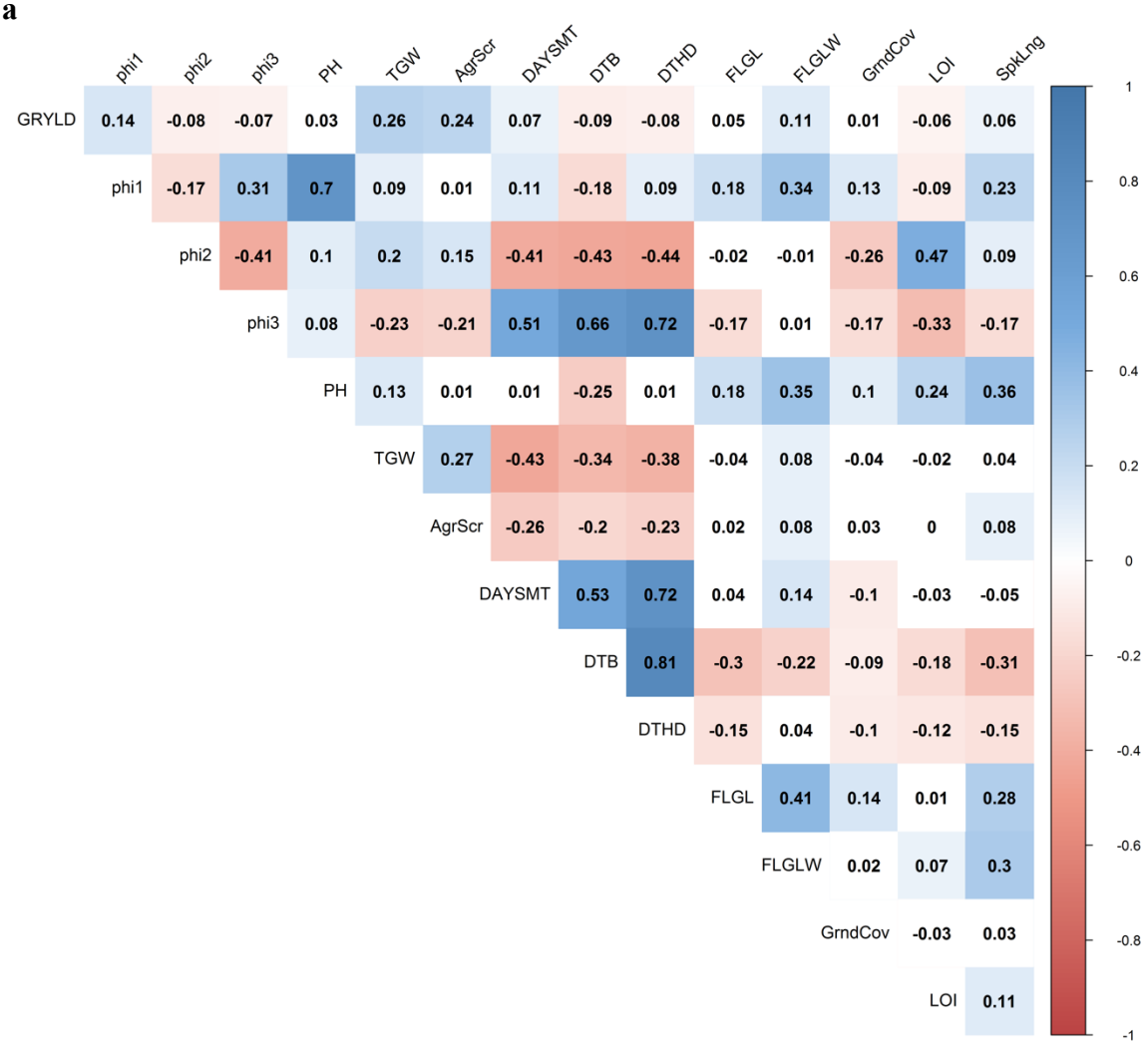


Figure 4.4. Manhattan plot of genome-wide associations of 542 wheat breeding lines. Digital crop growth parameters (θ_1, θ_2, θ_3) and plant height (PH) during: a) early planting trial, b) normal planting trials. The dashed lines on *y*-axis correspond to the genome-wide false discovery rate (FDR) threshold of 0.05.

a


phi1
phi2
phi3
PH
TGW
AgrScr
DAYSMT
DTB
DTHD
FLGL
FLGLW
GrndCov
LOI
SpkLng
GRYLD 0.14 -0.08 -0.07 0.03 0.26 0.24 0.07 -0.09 -0.08 0.05 0.11 0.01 -0.06 0.06
phi1 -0.17 0.31 0.7 0.09 0.01 0.11 -0.18 0.09 0.18 0.34 0.13 -0.09 0.23
phi2 -0.41 0.1 0.2 0.15 -0.41 -0.43 -0.44 -0.02 -0.01 -0.26 0.47 0.09
phi3 0.08 -0.23 -0.21 0.51 0.66 0.72 -0.17 0.01 -0.17 -0.33 -0.17
PH 0.13 0.01 0.01 -0.25 0.01 0.18 0.35 0.1 0.24 0.36
TGW 0.27 -0.43 -0.34 -0.38 -0.04 0.08 -0.04 -0.02 0.04
AgrScr -0.26 -0.2 -0.23 0.02 0.08 0.03 0 0.08
DAYSMT 0.53 0.72 0.04 0.14 -0.1 -0.03 -0.05
DTB 0.81 -0.3 -0.22 -0.09 -0.18 -0.31
DTHD -0.15 0.04 -0.1 -0.12 -0.15
FLGL 0.41 0.14 0.01 0.28
FLGLW 0.02 0.07 0.3
GrndCov -0.03 0.03
LOI 0.11
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

b

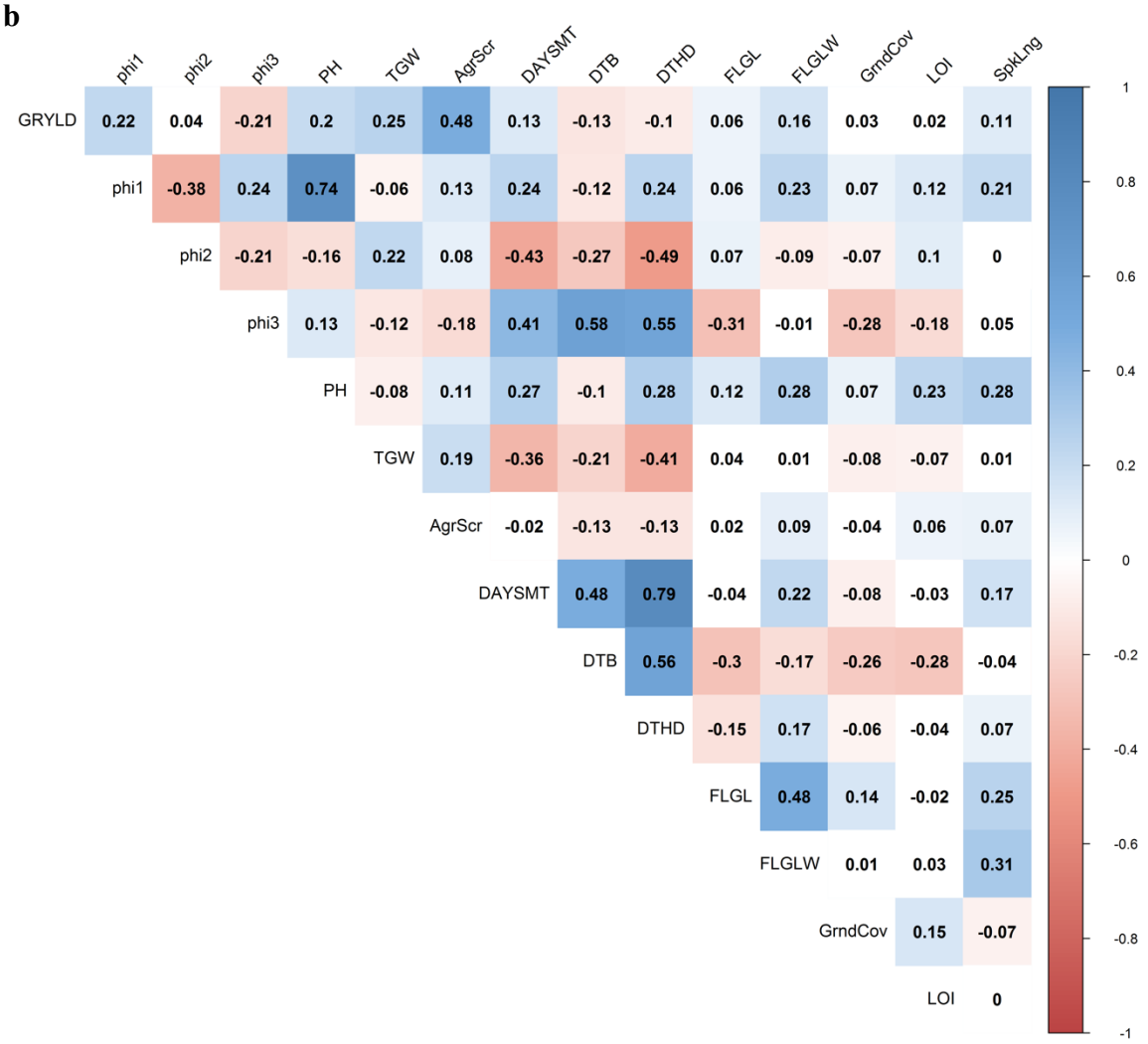


Figure 4.5. Phenotypic correlation matrix of crop growth and agronomic traits measured during the field experiments: a) early planting, b) normal planting. The absolute correlation coefficient values above 0.06 are significant at $\alpha < 0.05$.

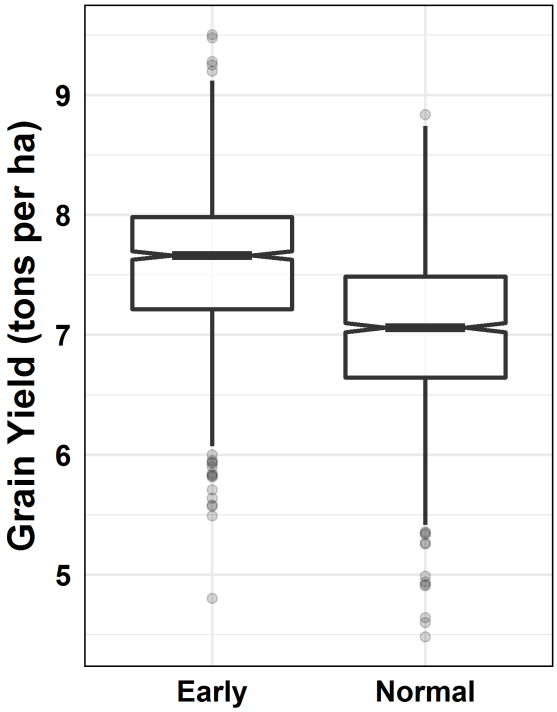


Figure 4.6. Distribution of grain yield of early and normal planted wheat trials in 2018 field season at Ludhiana.

Appendix A - Supplementary Material Chapter 2

This appendix contains supplementary figures and tables for Chapter 2.

Table A.1. Ontological description of phenotypic traits assessed during the study.

Trait	Abbreviation	Units	Ontology	Description
Ground Cover	GrndCov	%	CO_321:0000014	Crop ground cover, or the percentage of soil surface covered by plant foliage
Grain Yield	GRYLD	t ha^{-1}	CO_321:0000013	Amount (weight) of grains that was harvested
Plant Height	PH	cm	CO_321:0000020	Height of plant from ground to top of spike, excluding awns
Thousand Grain Weight	TGW	g	CO_321:0000025	Grain weight expressed as weight of a thousand grains
Heading Date	DTHD	-	CO_321:0000007	Heading time extends from the time of emergence of the tip of the spike from the flag leaf sheath to when the spike has completely emerged but has not yet started to flower
Days to Maturity	DAYSMT	-	CO_321:0000022	Maturity time starts at hard dough stage (ds87) often called physiological maturity
Flag Leaf	FLGLF	cm	CO_321:0001524	Length of flag leaf

Length				lamina
Flag Leaf Width	FLGLFW	cm	CO_321:0001525	Width of flag leaf lamina
Germination Percentage	GERMPCT	%	CO_321:0000011	Proportion of seed that germinate under proper condition
Spike Length	SpkLng	cm	CO_321:0000056	Length of spike.
Lodging Incidence	LOI	%	CO_321:0000167	Indicates incidence of lodged plants
Lodging Severity	LOS	-	-	Severity of the lodged plants on 0 to 10 scale

Table A.2. Broad-sense heritability of lodging measurements by trial at Ludhiana in 2016 and 2017. The visual (LOI, LOS, LI) and digital lodging measurements were estimated for each trial separately in two seasons. LOI, lodging incidence; LOS, lodging severity; LI, lodging index (LOI × LOS); DLmean, digital lodging mean; DLmix, digital lodging mixture

Year	Trial #	LOI	LOS	LI	DLmean	DLmix
2016	1	0.58	0.64	0.71	0.73	0.73
	2	0.65	0.50	0.61	0.45	0.4
	3	0.87	0.74	0.87	0.88	0.84
	4	0.74	0.68	0.74	0.62	0.58
	5	0.59	0.68	0.60	0.70	0.64
	6	0.70	0.40	0.66	0.64	0.63
	7	0.59	0.12	0.72	0.67	0.67
	8	0.35	0.42	0.40	0.35	0.36
	9	0.64	0.65	0.50	0.44	0.48
	10	0.57	0.61	0.43	0.48	0.52
	11	0.45	0.26	0.36	0.30	0.30
2017	1	0.50	0.35	0.52	0.59	0.54
	2	0.67	0.50	0.59	0.64	0.54
	3	0.69	0.77	0.61	0.64	0.42
	4	0.57	0.57	0.59	0.78	0.59
	5	0.64	0.65	0.64	0.85	0.76
	6	0.75	0.57	0.71	0.82	0.73
	7	0.76	0.71	0.67	0.77	0.79
	8	0.46	0.44	0.52	0.56	0.49
	9	0.67	0.63	0.65	0.80	0.21
	10	0.70	0.65	0.67	0.78	0.76
	11	0.85	0.85	0.87	0.76	0.79

Table A.3. Genetic correlations of lodging measures in years 2016 and 2017. Lower triangle (2016); upper triangle (2017). LOI, lodging incidence; LOS, lodging severity; LI, lodging index (LOI × LOS); DLmean, digital lodging mean; DLmix, digital lodging mixture.

Phenotype	LOI	LOS	LI	DLmean	DLmix
LOI	-	0.98	0.99	0.95	0.96
LOS	0.93	-	0.96	0.93	0.93
LI	0.93	0.99	-	0.96	0.97
DLmean	0.94	0.96	0.96	-	1.00
DLmix	0.93	0.96	0.97	1.00	-

Table A.4. Comparison of predictive abilities of three genomic prediction models (RR: RRBLUP; BCπ: Bayesian Cπ; RKHS: Reproducing Kernel Hilbert Space) of visual and digital lodging measures. Tr: Training set; Pr: Prediction set.

	Tr16-Pr17			Tr17-Pr16		
	RR	BCπ	RKHS	RR	BCπ	RKHS
LOI	0.30	0.30	0.30	0.32	0.32	0.33
LOS	0.35	0.35	0.35	0.38	0.37	0.38
LI	0.32	0.32	0.32	0.35	0.35	0.35
DLmean	0.32	0.31	0.32	0.37	0.37	0.37
DLmix	0.32	0.31	0.32	0.39	0.39	0.39

Table A.5. Predictive ability (r_{pv}), phenotypic correlation (r_{ph}) and genetic correlation (r_g) of lodging between LDH and FAS in years 2016 and 2017.

	2016			2017		
	r_{pv}	r_{ph}	r_g	r_{pv}	r_{ph}	r_g
LOI	0.43	0.39	0.65	0.45	0.46	0.75
LOS	0.32	0.29	0.45	0.46	0.44	0.88
LI	0.36	0.33	0.54	0.44	0.44	0.77
DLmean	0.33	0.32	0.54	0.41	0.42	0.66
DLmix	0.30	0.29	0.51	0.40	0.41	0.67

r_{pv}: correlation between genomic estimated breeding values and phenotypic values across environments

r_{ph}: correlation of phenotypic values at two environments

r_g: the marker-based genetic correlation between two environments

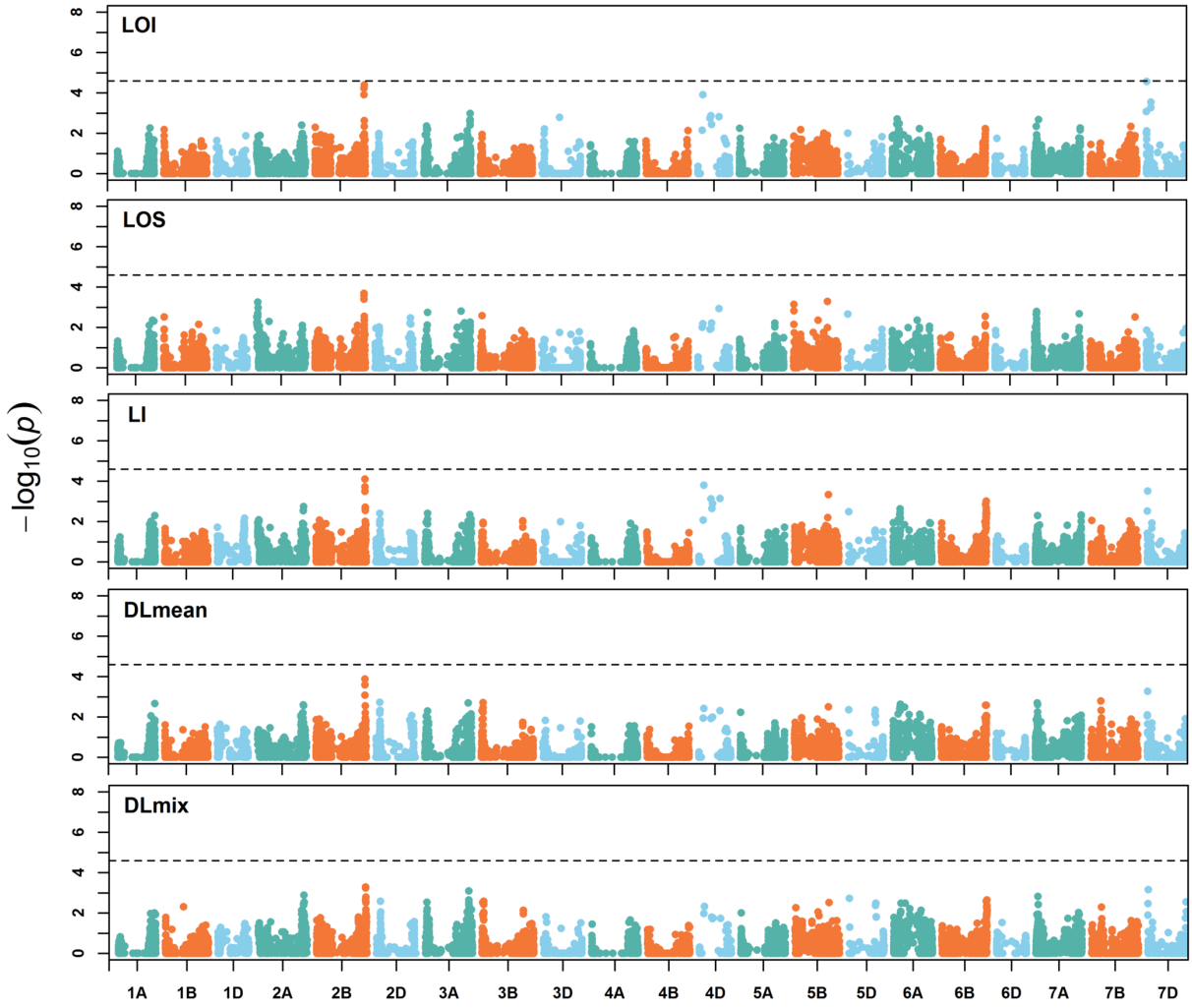


Figure A.1. Manhattan plot of genome-wide associations in 2016. The dashed lines on y-axis correspond to the genome-wide false discovery rate (FDR) threshold.

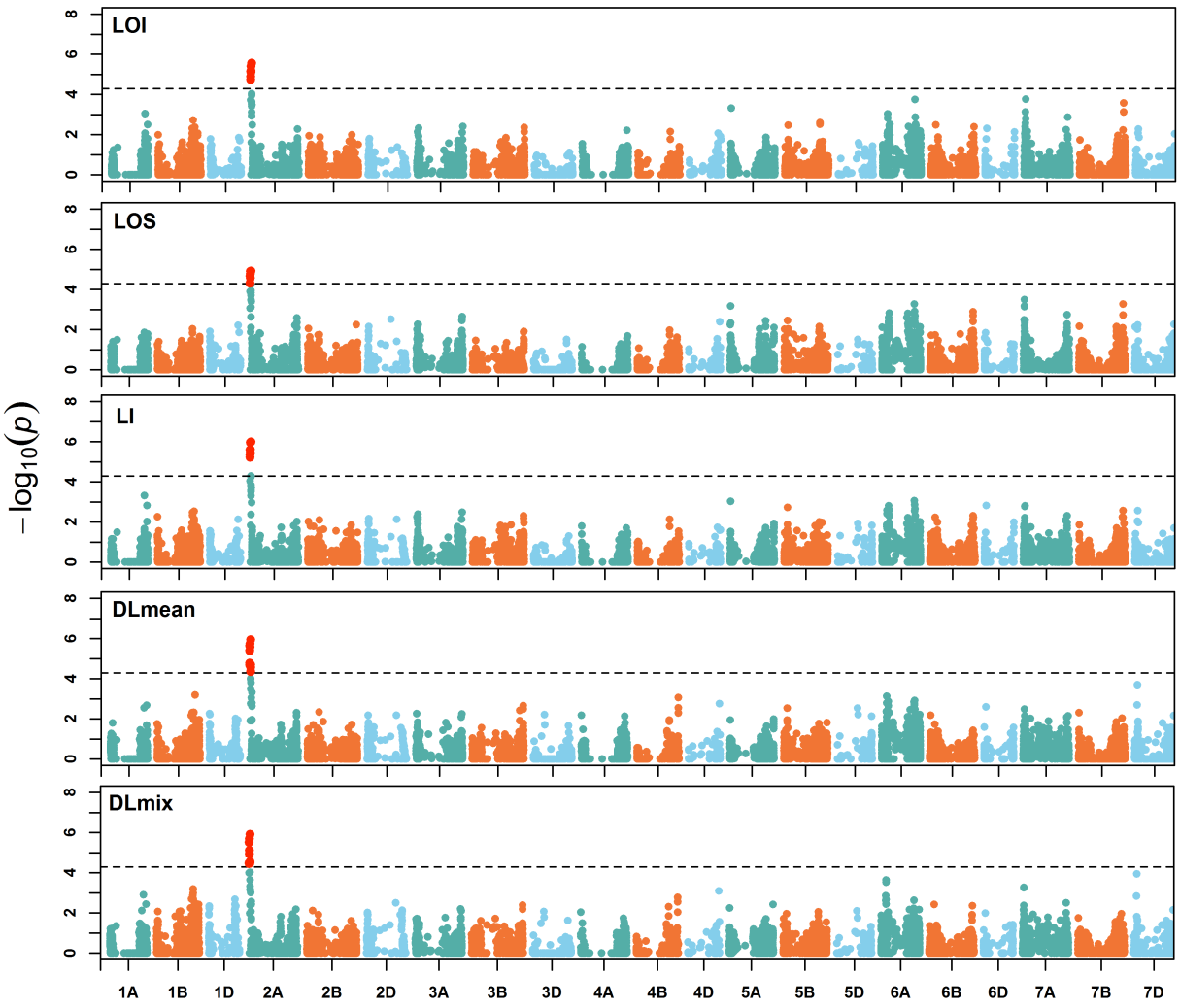


Figure A.2. Manhattan plot of genome-wide associations in 2017. The dashed lines on y-axis correspond to the genome-wide false discovery rate (FDR) threshold. The markers crossing FDR threshold are highlighted in red.

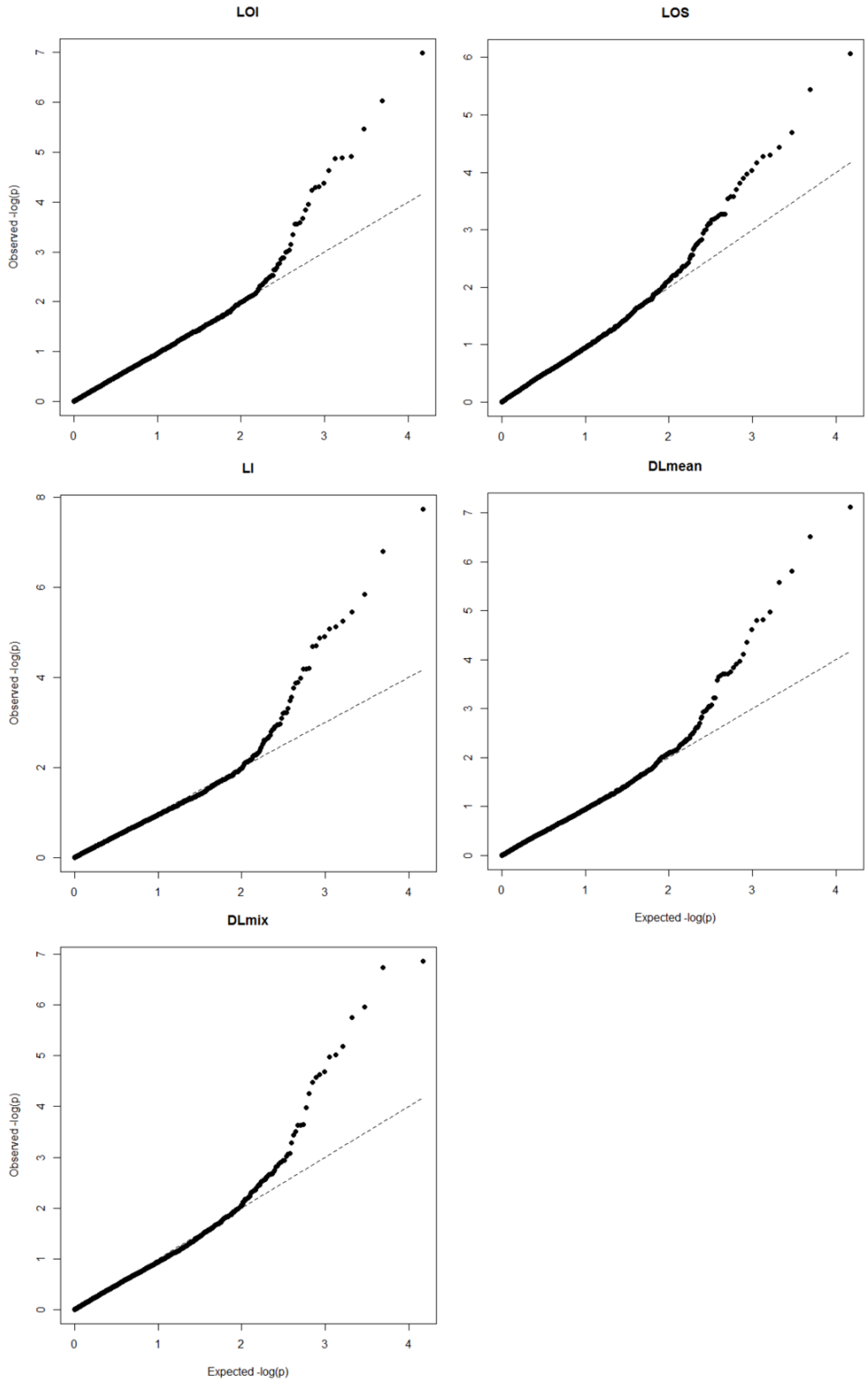


Figure A.3. Quantile-Quantile (QQ) plots of the genome-wide associations. The QQ plots of SNP markers of the measured lodging traits.

Appendix B - Supplementary material Chapter 3

This appendix contains supplementary figures and tables for Chapter 3.

Table B.1. The UAS flight dates at three locations in wheat seasons 2017 and 2018.

	Ludhiana (LDH)	Jabalpur (JBL)	PUSA (PUS)
2017	Dec-13	Dec-16	Dec-19
	Jan-15	Dec-18	
	Jan-20		
	Jan-25		
	Jan-31		
	Feb-07		
2018	Nov-22	Dec-26	Dec-28
	Nov-24	Jan-14	Jan-12
	Nov-29		Feb-02
	Dec-01		Feb-12
	Dec-13		Feb-22
	Dec-18		
	Jan-09		
	Jan-20		
	Jan-26		

Table B.2. Ontological description of phenotypic traits assessed during field experiments.

Trait	Abbreviation	Units	Ontology	Description
Ground Cover	GrndCov	%	CO_321:0000014	Crop ground cover, or the percentage of soil surface covered by plant foliage
Grain Yield	GRYLD	t ha^{-1}	CO_321:0000013	Amount (weight) of grains that was harvested
Normalized Difference Vegetation Index	NDVI	-	CO_321:0000301	Wheat canopy normalized difference vegetation index trait

Table B.3. Repeatability (H^2), phenotypic correlation ($r_{GY\text{-}GrndCov}$, ground cover and grain yield) of visual and digital ground cover across environments.

Environment	DAS	GrndCov	H^2	$r_{GY\text{-}GrndCov}$	$H^2 \times r_{GY\text{-}GrndCov}$
17PUS.Normal	34	Digital	0.27	0.30	0.08
	41	Visual	0.27	0.27	0.07
17JBL.Normal	38	Digital	0.24	0.28	0.07
	45	Visual	0.32	0.32	0.10
17LDH.Normal	37	Digital	0.61	0.21	0.13
	38	Visual	0.54	0.20	0.11
	70	Digital	0.42	0.32	0.13
	80	Digital	0.46	0.28	0.13
	86	Digital	0.32	0.44	0.14
18PUS.Normal	44	Visual	0.04	0.30	0.01
	59	Digital	0.32	0.31	0.10
	80	Digital	0.46	0.42	0.19
	90	Digital	0.34	0.44	0.15
18JBL.Normal	37	Visual	0.57	0.43	0.25
	46	Digital	0.58	0.19	0.11
18LDH.Early	31	Digital	0.61	0.12	0.07
	33	Digital	0.70	0.12	0.08
	38	Digital	0.33	0.15	0.05
	40	Digital	0.31	0.16	0.05
	44	Visual	0.61	0.01	0.01
	52	Digital	0.24	0.13	0.03
	57	Digital	0.58	0.01	0.01
18LDH.Normal	29	Visual	0.59	0.03	0.02
	37	Digital	0.61	-0.01	-0.01
	38	Digital	0.64	0.12	0.08
	76	Digital	0.45	-0.05	-0.02

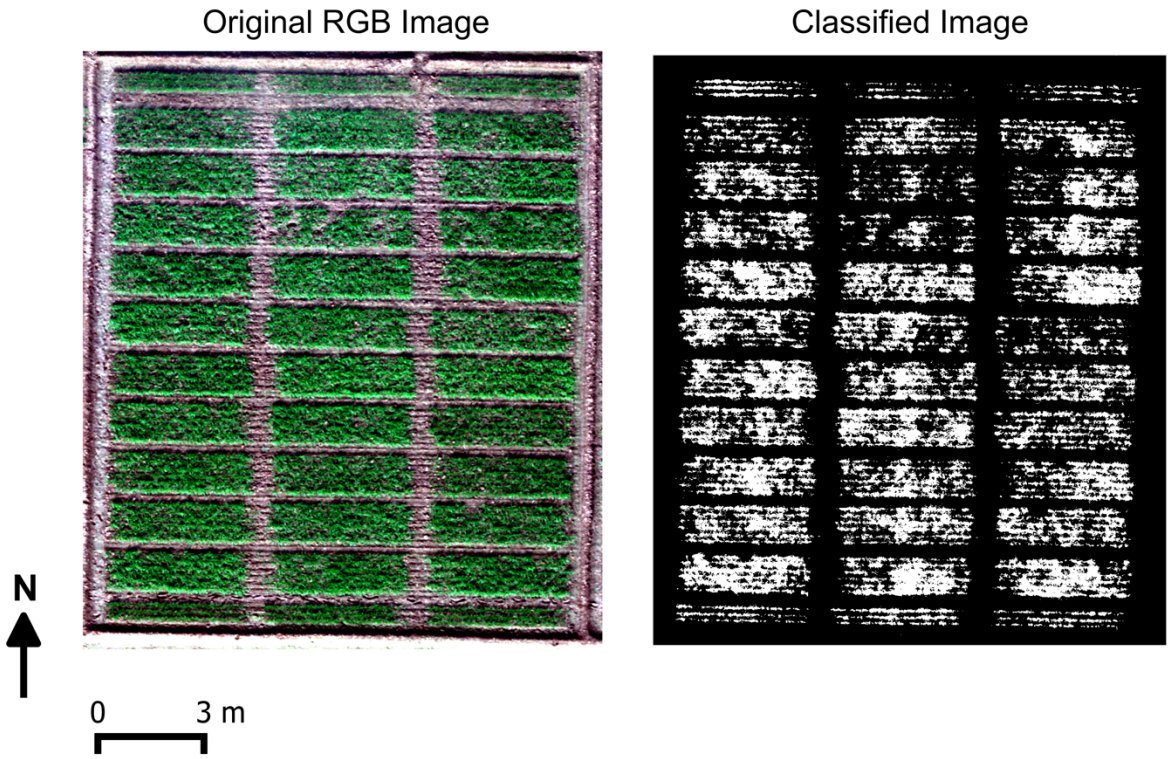


Figure B.1. Digital extraction of ground cover using supervised image classification procedure.

Figure B.2. GWAS Manhattan plots of ground cover measurements from 17LDH.Normal environment. Each panel corresponds to days of measurements shown as Days after Sowing (DAS). The hexaploid wheat sub-genomes are marked by three distinct colors. Abbreviations: DGC, Digital Ground Cover; VGC, Visual Ground Cover.

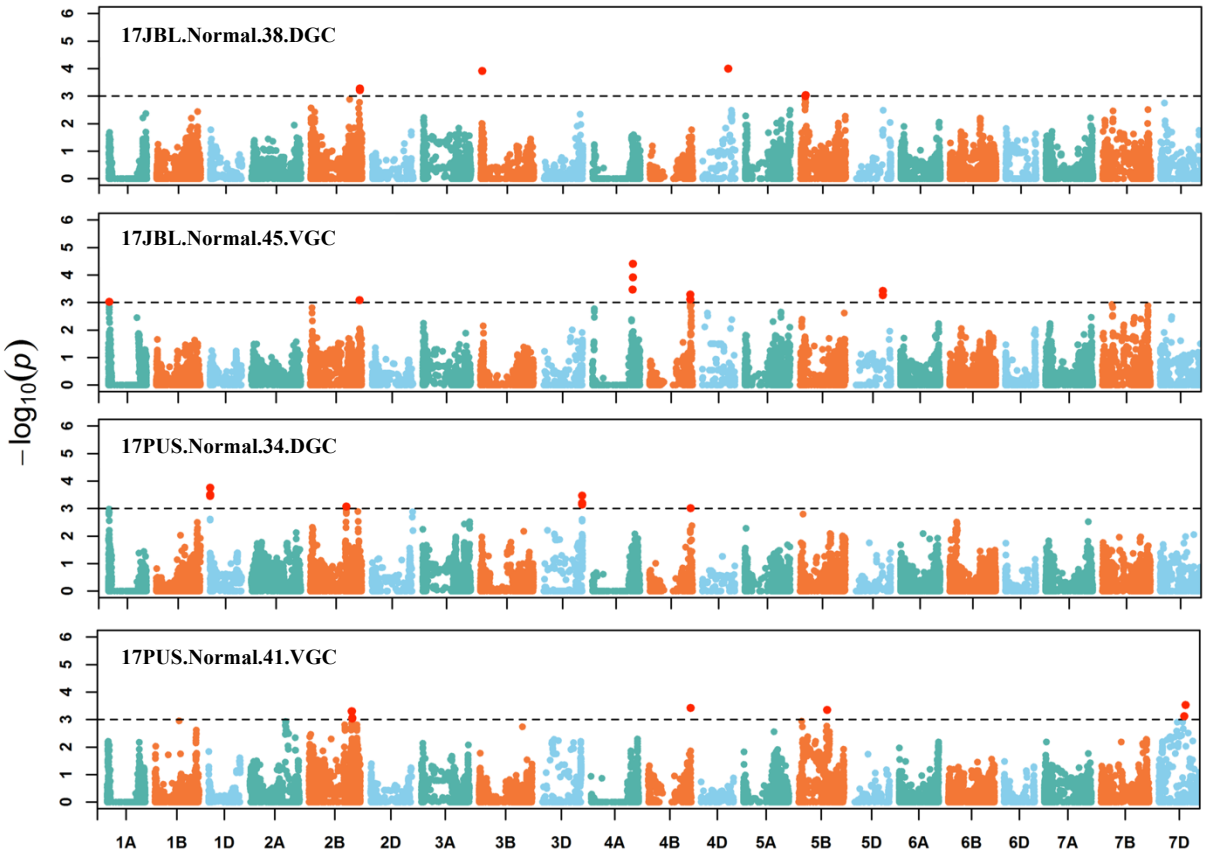


Figure B.3. GWAS Manhattan plots of ground cover measurements from 17JBL.Normal and 17PUS.Normal environments.

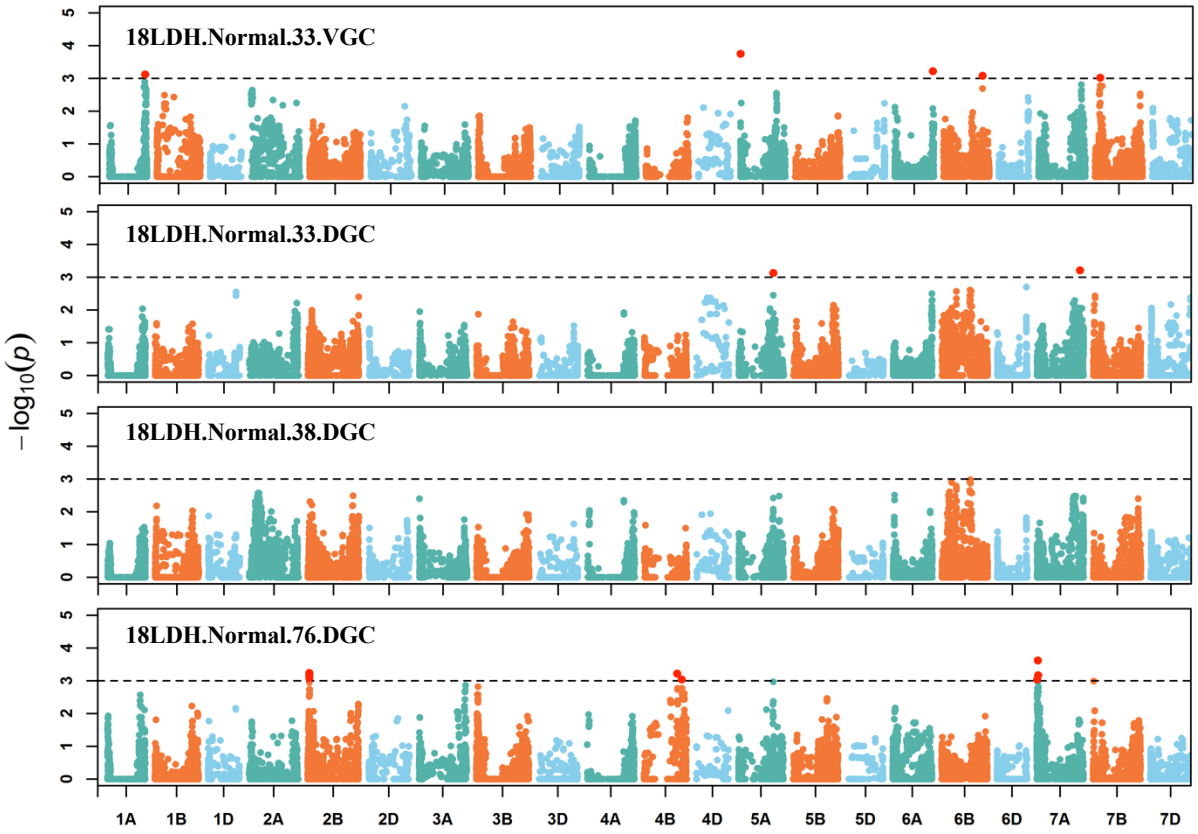


Figure B.4. GWAS Manhattan plots of ground cover measurements from 18LDH.Normal environments.

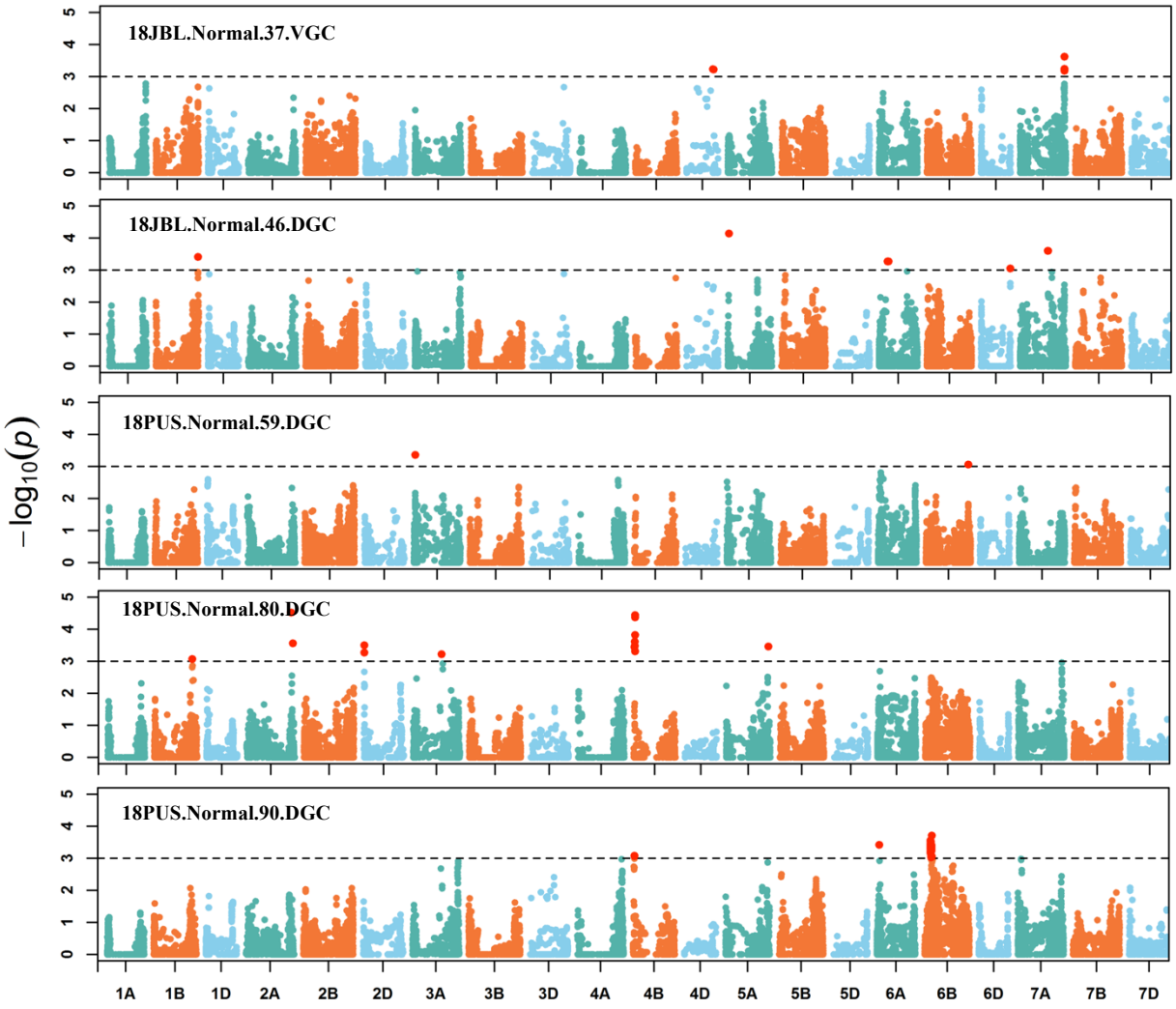


Figure B.5. GWAS Manhattan plots of ground cover measurements from 18JBL.Normal and 18PUS.Normal environments.

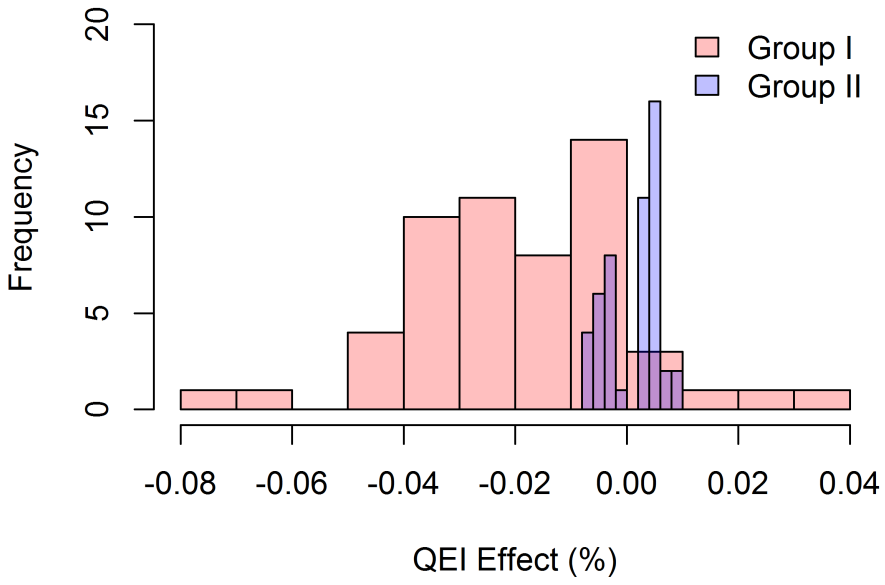
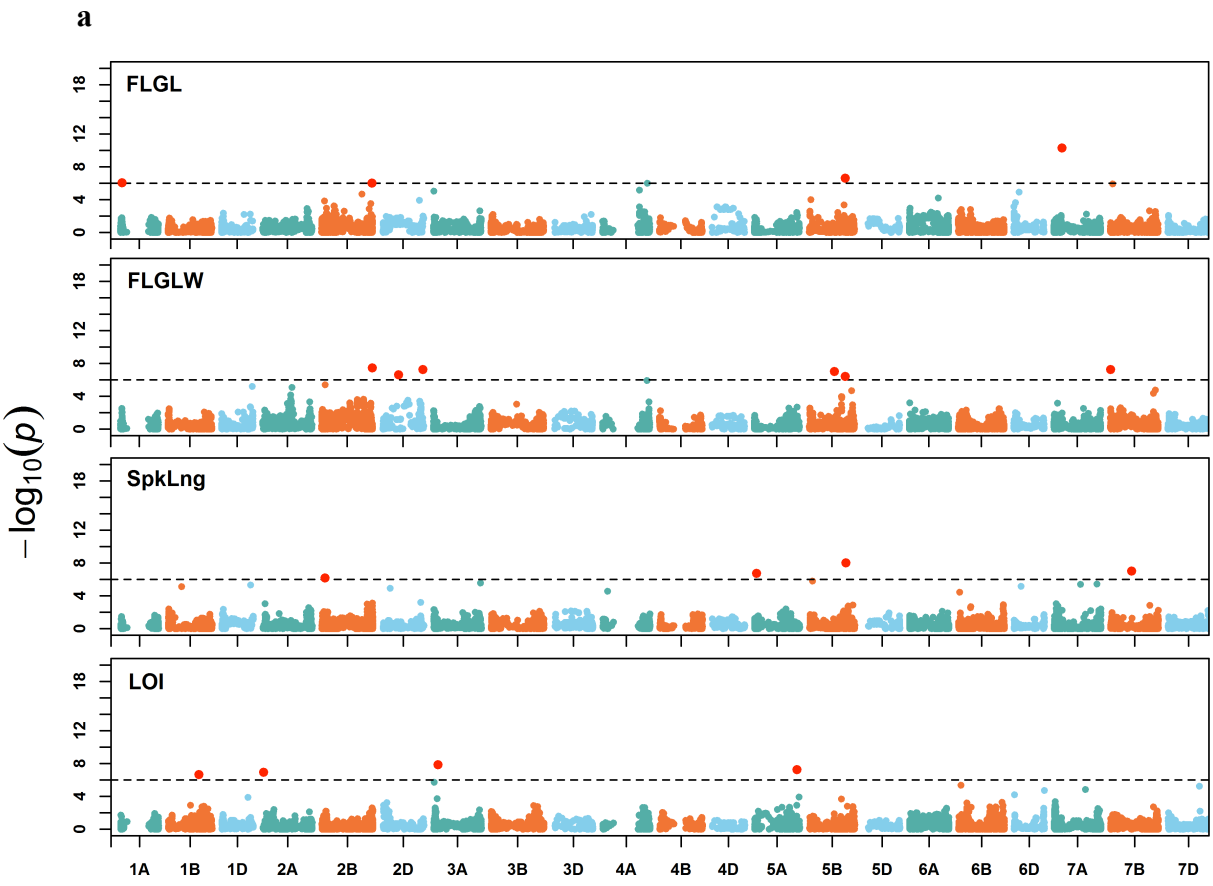


Figure B.6. The QTL × Environment interaction (QEI) effect of GWAS significant markers for ground cover at early and advanced wheat growth stages. a) Group I: early growth stage data representing three environments (17LDH.Normal, 17JBL.Normal, and 17PUS.Normal) from 575-666 thermal days (°Cd); b) Group II: advanced growth stage representing four environments (18LDH.Early, 18LDH.Normal, 18JBL.Normal, and 18PUS.Normal) from 881-925 °Cd thermal days window.

Appendix C - Supplementary material Chapter 4

This appendix contains supplementary figures and tables for Chapter 4.



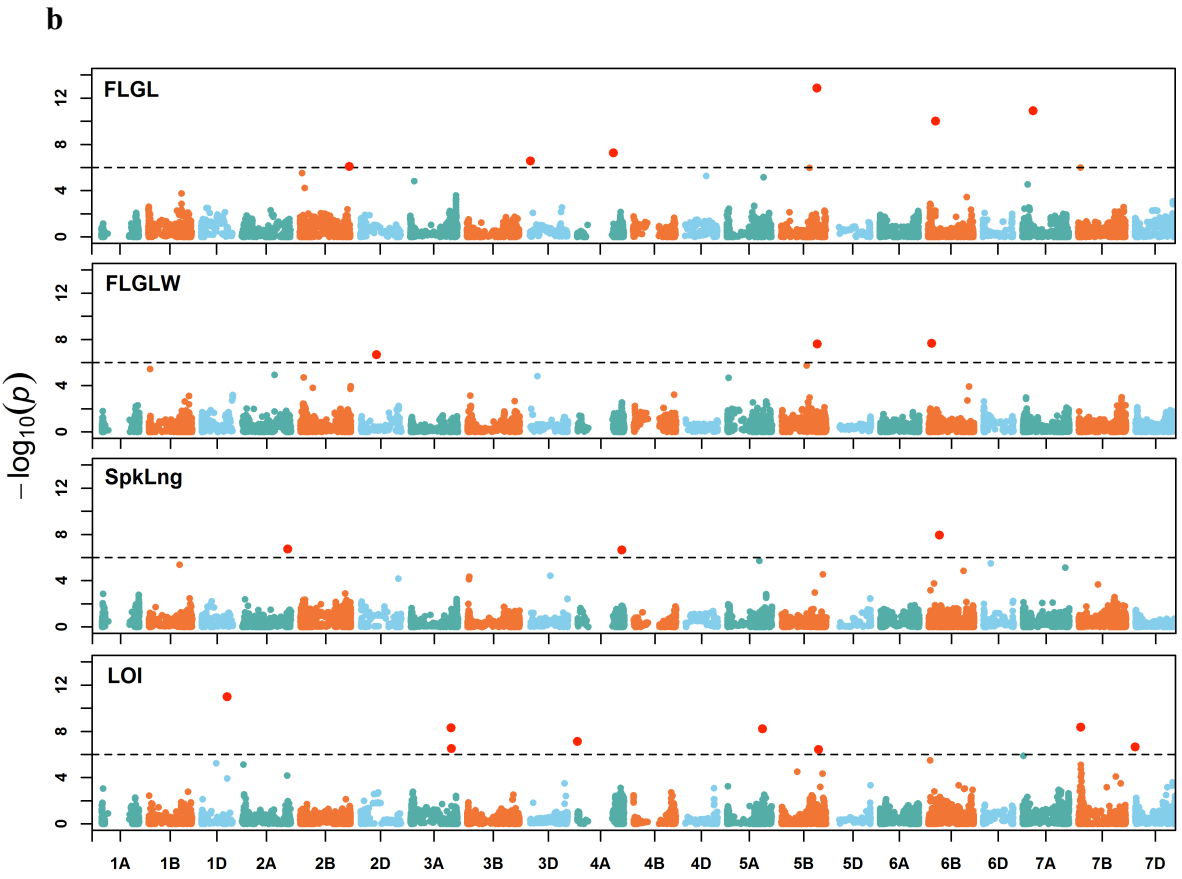


Figure C.1. Manhattan plot of genome-wide associations of morphological traits in wheat. Flag leaf length (FLGL), Flag leaf width (FLGLW), Spike length (SpkLng), and lodging index (LOI) during: a) early planting trial, b) normal planting trials. The dashed lines on *y*-axis correspond to the genome-wide false discovery rate (FDR) threshold of 0.05.

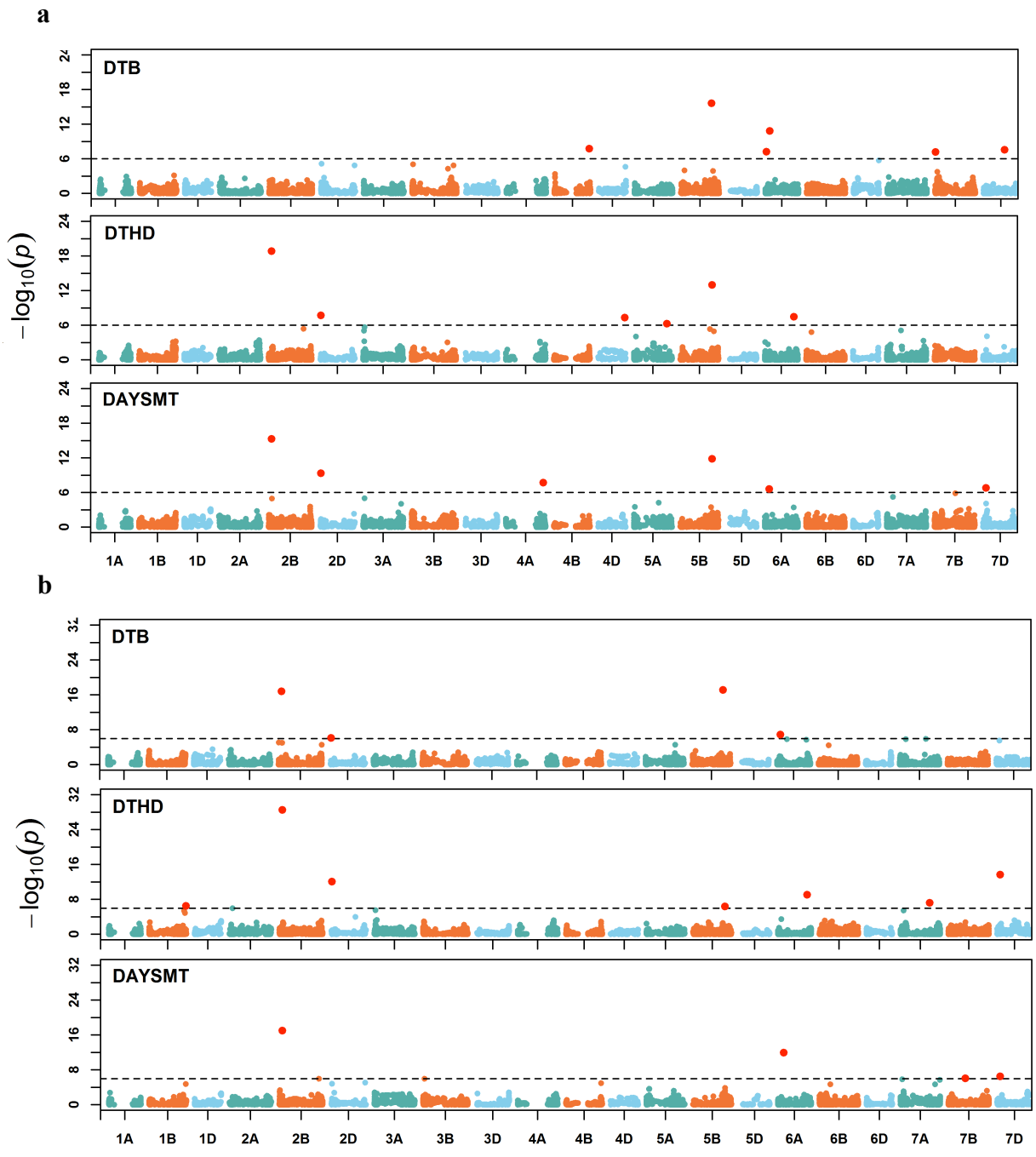


Figure C.2. Manhattan plot of genome-wide associations of phenological traits in wheat. Days to booting (DTB), Days to heading (DTHD), and Days to maturity (DAYSMT) during: a) early planting trial, b) normal planting trials. The dashed lines on *y*-axis correspond to the genome-wide false discovery rate (FDR) threshold of 0.05.