GENOMICS-ASSISTED BREEDING OF MALTING BARLEY FOR NEW YORK STATE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Daniel William Sweeney

May 2021

GENOMICS-ASSISTED BREEDING OF MALTING BARLEY FOR NEW YORK STATE

Daniel William Sweeney, Ph. D.

Cornell University 2021

Barley (*Hordeum vulgare* L.) has been the primary ingredient in brewing for millennia. Recent increases in demand for local malting barley production outside of western North America for craft brewing and distilling have created a need for new varieties adapted to humid continental climates. Malting quality can be negatively affected by several seed pathogens and preharvest sprouting (PHS) in regions with wet conditions during the growing season. Genomics-assisted breeding methods, including genomic prediction and genome-wide association, have the potential to improve selection accuracy and reduce breeding cycle time to increase rates of genetic gain.

My goals were to apply genomics-assisted breeding methods to address malting barley production challenges in the northeastern United States. Specific objectives included 1) examining the genetic architecture and environmental stability of the seed dormancy 2 (*SD2*) locus to determine if tight linkage or pleiotropy is responsible for negative correlations between PHS and malting quality, 2) studying the genetic relationships between PHS, seed dormancy, and seed germination rate to identify germplasm with PHS resistance and good malting quality, 3) exploring strategies to enhance genomic prediction accuracy in a newly established spring barley breeding program, 4) evaluating the merits of a genomic selection approach using a

selection index in a newly established breeding program, and 5) releasing a two-row spring malting barley variety with disease resistance, PHS resistance, and good malting quality for local growers.

The key findings of this research are 1) three functional alleles of *HvMKK3* are the primary determinants of PHS resistance and seed dormancy at *SD2* and *HvGA20ox1* does not have an effect on these traits, 2) *HvMKK3* and *HvAlaAT1* are sensitive to environmental conditions and together explain a substantial portion of annual PHS variation, 3) a connected half-sib training population increases within and across family prediction accuracy and is highly powered for genome-wide association studies, 4) genomic selection using a selection index resulted in significant gain for highly weighted traits but retained genetic variance for lightly weighted traits, and 5) a new two-row spring barley variety, ‘Excelsior Gold’, was released as a result of these combined efforts.

BIOGRAPHICAL SKETCH

Daniel William Sweeney was born to Patrick and Alice Sweeney on September 28, 1992 in Speedway, IN. His fascination with the natural world began at an early age and was nurtured by family hiking trips and gardening with his parents, with interests transitioning from dinosaurs to marine life to insects and finally plants. Daniel first discovered agricultural science in the form of entomology and forestry through his ten years in 4-H. He graduated from Cardinal Ritter High School in Indianapolis, IN in May 2011 with honors. Daniel graduated magna cum laude from Purdue University in May 2015 with a Bachelor of Science in Plant Genetics, Breeding, and Biotechnology and a minor in Food and Agribusiness Management. At Purdue, Daniel participated in undergraduate research with Dr. Karen Hudson studying soybean fatty acid mutants and studied abroad at the University of Adelaide for five months. He began graduate studies at Cornell University in August 2015. Daniel's graduate education has been greatly enhanced by international opportunities in India, Mexico, and the United Kingdom as well as involvement in the New York state craft malting and brewing community. He has enjoyed serving in many capacities in the plant breeding and genetics graduate student organization, Synapsis, and volunteering with Tompkins County 4-H.

DEDICATION

In memory of Ralph E. Miller and Joseph P. Sweeney.

ACKNOWLEDGMENTS

I would like to thank my committee for their guidance and support. My advisor, Dr. Mark Sorrells, has given me extraordinary trust, academic freedom, and responsibility unheard of for many graduate students, Dr. Gary Bergstrom has been instrumental in promoting our barley research through grower focused extension, and Dr. Jessica Rutkoski has provided excellent technical and practical advice. I am indebted to technicians past and present: David Benscher, Amy Fox, James Tanaka and Jesse Chavez. My research would not have been possible if it were not for their support and they made coming to work every day, particularly during field season, a joy. I am grateful for the steadfast care of greenhouse experiments by Jean Koski. I would like to thank the substantial contributions of collaborators: Aaron Macleod at the Hartwick College Craft Food and Beverage Center; Kaylyn Kirkpatrick and Edmond Guidault at Cornell AgriTech; Dr. Jason Walling and Leslie Zalapa at the USDA Cereal Crops Research Unit; Jaime Cummings and Jen Starr in the Cornell small grains pathology group; New York farmers Ted Hawley, Jeff Trout, and Dave Wallace; and Dr. Julie Hansen in the Cornell forage breeding group. To labmates past and present who have enriched my education beyond measure and will be lifelong colleagues and friends, especially Drs. Margaret Krause, Shantel Martinez, Nick Santantonio, and Lynn Veenstra as well as Karl Kunze, Shitaye Megerssa, Travis Rooney, and Ellie Taagen. The support and fellowship of Synapsis has been and will continue to be a great source of ideas. I would like to thank several people who have had a far greater impact on my life and education than they know: Bob Eddleman, Dr. Tracie Egger, and Dr. Gebisa Ejeta. Most importantly, I am immensely grateful for the support, guidance, and love from my parents. Thank you for always encouraging my education, nurturing a lifelong love of learning, and constantly providing home.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABA: abscisic acid
CU: Cornell University spring barley breeding germplasm
D: dormant allele
GA: gibberellin
GBLUP: genomic best linear unbiased prediction
GBS: genotyping-by-sequencing
GE: germination energy
GI: germination index
GS: genomic selection
GWAS/GWA: genome-wide association study
HSD: honestly significant difference
JIC: John Innes Centre
LD: linkage disequilibrium
MAF: minor allele frequency
N/N*: non-dormant allele/highly non-dormant allele
NAM: nested-association mapping
PA: prediction accuracy
PHS: preharvest sprouting
PM: physiological maturity
PS: phenotypic selection
RYT: regional yield trial
QTL: quantitative trait loci
SB: AAC Synergy/Bentley
SC: AAC Synergy/Conlon
SG: AAC Synergy/ND Genesis
SN: AAC Synergy/Newdale
SNP: single nucleotide polymorphism
SP: AAC Synergy/Pinnacle
SR: AAC Synergy/Craft
ST: AAC Synergy/KWS Tinka
TP: training population
VP: validation population

CHAPTER 1

INTRODUCTION

Rationale and significance

By 2050, an estimated 9 billion people will inhabit Earth and amidst the challenges of climate change and land degradation, many of them are going to want a beer. Barley (*Hordeum vulgare* L.) is an annual diploid (2*n*=14) grass in the Triticeae tribe that has been grown on a global scale for millennia for brewing and distilling, human food, and animal feed. Most malting barley in North America is bred for large production areas in the western United States and Canada to supply industrial malting and brewing. Spring malting barley varieties from these regions tend to have an excellent industrial malting quality profile and high yield. Renewed interest in local agricultural production and value chains coupled with an explosion in craft food and beverage business, largely fueled by craft brewing, has led to opportunities for expanded malting barley production. Many of these areas of malting barley growth have significantly different climates than western North America, namely increased humidity and precipitation during the growing season. Increased moisture introduces substantial disease pressure and the threat of preharvest sprouting (PHS). Much of the available barley germplasm in North America has been specifically bred for large scale breweries who brew in the adjunct style, which involves the addition of inexpensive sources of starch like rice or corn to the barley malt mash. Adjunct brewing requires barley with high protein and high amylase enzyme activity (diastatic power) to break down the added starch because rice and corn have very low diastatic power. Craft maltsters and brewers prefer all malt brewing and have different preferences and more flexibility than large-scale malthouses and breweries, creating a need for craft malting barley breeding.

New York emerged as a center of barley production in the United States in the mid-19th century with acreage initially radiating from the arterial Erie Canal and later shifting west towards the major malthouses along the Great Lakes (Harlan et al., 1925). The McKinley Tariff of 1890 eventually destroyed the barley economy in New York by making Canadian grain imports more expensive than malt shipments from Minnesota and Wisconsin (Harlan et al., 1925). In the 21st century, New York has returned to its barley roots. New York is the fourth largest brewing state in the country with over 400 operational breweries (New York State Brewers Association) and is a malting hub with 11 craft malthouses (Cornell Field Crops, 2019). Much of this growth can be directly attributed to the 2012 New York Farm Brewery bill. This legislation provides economic incentives to brewers who brew with locally grown malt and hops and has spurred the establishment of over 150 new farm breweries in New York. Despite this encouraging growth in malting and brewing, New York barley production still lags far behind historic totals. The Farm Brewery bill mandates use of at least 90% New York grown grain in farm brewery production by 2024. Current barley acreage is insufficient to meet this goal and currently available varieties introduce a high level of grower risk due to disease and PHS susceptibility, poor agronomics, and variable malting quality. New two-row spring malting barley varieties are needed in a short amount of time to decrease grower risk, provide a stable source of high-quality grain for craft malthouses and breweries, and meet state mandated production goals.

One of the key traits for successful expansion of malting barley production into humid continental climates is PHS resistance. Grain damaged by PHS has reduced storage capacity and is unusable for malting in severe cases. Many PHS quantitative trait locus (QTL) mapping studies have been conducted in barley and two large effect loci, Seed Dormancy 1 and 2 (*SD1* and *SD2*), have been consistently identified on chromosome 5H (Oberthur et al., 1995; Han et al., 1996; Hori et al., 2007; Lin et al., 2009). *SD2* has also been mapped in a range of malting

quality QTL mapping studies (Castro et al., 2010; Mohammadi et al., 2015; Zhou et al., 2016) but PHS and malting quality are often negatively correlated (Gao et al., 2003; Li et al., 2003; Castro et al., 2010). An alanine aminotransferase (*HvAlaAT1*) has been cloned at *SD1* (Sato et al., 2016) and a mitogen activated protein kinase kinase 3 (*HvMKK3*) has been cloned at *SD2* (Nakamura et al., 2016) but a second gene in the *SD2* region, gibberellin oxidase (*HvGA20ox1*) may also be associated with PHS resistance (Li et al., 2004; Nagel et al., 2018). It is unknown if the negative linkage at *SD2* between PHS and malting quality is due to tightly linked loci or pleiotropy. Environmental conditions, particularly temperature, are known to affect PHS and seed dormancy (Rodriguez et al., 2001; Li et al., 2003; Gualano & Benech-Arnold, 2009) but the genetic causes of PHS environmental sensitivity are unknown.

Disease introduces another challenge in high moisture barley growing regions. Malting quality grain should be clean and free of visible disease. *Fusarium graminearum* ((Schwein.) Petch) is a major pathogen of wheat and barley that causes yield loss, due to reduced kernel weight, and mycotoxin accumulation. *Fusarium* head blight (FHB) is a major concern for winter and spring barley production in New York state due to favorable environmental conditions for disease and low deoxynivalenol (DON) mycotoxin thresholds (< 1 ppm) for malting and brewing. Fungicide treatments can be effective control for FHB and DON suppression, but accurate timing of fungicide application is difficult. Visual scoring of disease symptoms is often not correlated with DON accumulation, making genetic sources of FHB resistance a top breeding target. *Bipolaris sorokiniana* Sacc. causes spot blotch, a major foliar disease of barley worldwide that also infects barley grain. Spot blotch is the most serious foliar disease in spring barley in New York state. Severe spot blotch infections negatively impact yield and can reduce kernel plumpness, an important malting quality trait (Nutter et al., 1985). Diseased grain is also likely to be rejected for malting because steeping and germination conditions during malting encourage microbial growth.

Variety development is traditionally a slow process. Annual row crops typically take 10-15 years to advance from cross to screening nursery to yield trials to foundation seed production. The increased availability of affordable high-throughput genotyping platforms and the advent of genomic prediction (Bernardo, 1994; Meuwissen et al., 2001) have ushered in a new age of opportunity for accelerating and improving plant breeding and variety development. The ability to predict genetic value in early generations without allocating phenotyping resources to hundreds or thousands of lines has led to much excitement over the prospects of genomic selection (GS) (Heffner et al., 2010; Jannink et al., 2010). In recent years, optimal procedures for training population composition (Rincent et al., 2012; Lehermeier et al., 2014; Akdemir et al., 2015), marker imputation and density (Rutkoski et al., 2013; Lorenz et al., 2012), and empirical studies (Beyene et al., 2015; Rutkoski et al., 2015; Sallam and Smith, 2016; Tiede and Smith, 2018) of GS have been presented for a number of crops.

The craft malting and brewing landscape in New York presents a unique opportunity to test the merits of GS to quickly establish an effective and productive breeding program from the ground up. Nascent breeding programs must carefully manage short- and long-term objectives for success in both timelines. Long-term genetic gain relies on genetic variation, but short-term breeding goals may deplete valuable variation through high selection intensity and crosses between related elite lines. High intensity short-term solutions thus run the risk of depleting the long-term success of a breeding program. Methods to maximize initial genetic variation, consistently generate new variation, reduce phenotyping of high value but expensive traits, and increase prediction accuracy for parental and inbred selection would be particularly useful for new breeding programs. Breeding programs must select on many traits simultaneously, creating a need for selection indices (Hazel and Lush, 1942). Some of these traits will be genetically correlated which may be desirable or undesirable. Empirical examples of GS using a selection index in plants indicate gain either in overall index value (Combs and Bernardo, 2013; Massman et al., 2013) or component traits (Hernandez et al.,

2020). Heffner et al. (2011) suggested that use of selection indices in GS can predict net merit better than phenotypic selection.

High-throughput genotyping and advances in statistical modeling have also enabled efficient use of breeding germplasm for genetic mapping that does not involve costly development of biparental mapping populations that may be ultimately of little value for applied breeding. Improved understanding of the genetic architecture of complex traits in breeding programs can facilitate marker development for high value, large effect QTL and appropriately manage genetic variation for long-term genetic gain. These methods can also help elucidate complex genetic interactions such as epistasis and pleiotropy.

Objectives

This work encompasses four approaches for improving spring malting barley germplasm using genomics-assisted breeding, with emphasis on adaptation to humid continental climates, as well as an overall strategy of how genomics-assisted breeding can be practically applied in a newly founded breeding program.

1. Explore methods to improve prediction accuracy in a connected half-sib genomic prediction training population for parental and inbred selection.
2. Elucidate the genetic control of *SD2* and examining genetic relationships between PHS, seed dormancy, and germination rate using genome-wide association mapping.
3. Model environmental sensitivity of genes associated with PHS using QTL by environment modeling in historic winter and spring barley datasets.
4. Assess realized gain from GS for correlated and uncorrelated traits using a selection index.
5. Examine two-row spring malting barley as an empirical model of genomics-assisted variety development.

References

Akdemir, D., J.I. Sanchez, and J.L. Jannink. 2015. Optimization of genomic selection training populations with a genetic algorithm. Genet. Sel. Evol. 47(1): 1–10. doi: 10.1186/s12711-015-0116-6.

Bernardo, R. 1994. Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci. 34(1): 20–25. doi: 10.2135/cropsci1994.0011183X003400010003x.

Beyene, Y., K. Semagn, S. Mugo, A. Tarekegne, R. Babu, et al. 2015. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci. 55(1): 154. doi: 10.2135/cropsci2014.07.0460.

Castro, A.J., A. Benitez, P.M. Hayes, L. Viega, and L. Wright. 2010. Coincident quantitative trait loci effects for dormancy, water sensitivity and malting quality traits in the BCD47×Baronesse barley mapping population. Crop Pasture Sci. 61(9): 691. doi: 10.1071/CP10085.

Combs, E., and R. Bernardo. 2013. Genomewide selection to introgress semidwarf maize germplasm into U.S. corn belt inbreds. Crop Sci. 53(4): 1427–1436. doi: 10.2135/cropsci2012.11.0666.

Gualano, N.A., and R.L. Benech-Arnold. 2009. The effect of water and nitrogen availability during grain filling on the timing of dormancy release in malting barley crops. Euphytica 168(3): 291–301. doi: 10.1007/s10681-009-9948-x.

Harlan, H.., M.L. Martini, and M.N. Pope. 1925. Tests of barley varieties in America. United States Department of Agriculture. Department bulletin No. 1334.

Hazel, L.N., and J.L. Lush. 1942. The efficiency of three methods of selection. J. Hered. 33(11): 393–399. doi: 10.1093/oxfordjournals.jhered.a105102.

Heffner, E.L., J.-L. Jannink, and M.E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4(1): 65. doi: 10.3835/plantgenome2010.12.0029.

Heffner, E.L., A.J. Lorenz, J.-L. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: Gain per unit time and cost. Crop Sci. 50(5): 1681. doi: 10.2135/cropsci2009.11.0662.

Hernandez, C.O., L.E. Wyatt, and M.R. Mazourek. 2020. Genomic prediction and selection for fruit traits in winter squash. G3: Genes|Genomes|Genetics 10(10): 3601–3610. doi: 10.1534/g3.120.401215.

Hori, K., K. Sato, and K. Takeda. 2007. Detection of seed dormancy QTL in multiple mapping populations derived from crosses involving novel barley germplasm. Theor. Appl. Genet. 115(6): 869–876. doi: 10.1007/s00122-007-0620-3.

Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics 9(2): 166–177. doi: 10.1093/bfgp/elq001.

Li, C., P. Ni, M. Francki, A. Hunter, Y. Zhang, et al. 2004. Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. Funct. Integr. Genomics 4(2): 84–93. doi: 10.1007/s10142-004-0104-3.

Lin, R., R.D. Horsley, N.L. V. Lapitan, Z. Ma, and P.B. Schwarz. 2009. QTL mapping of dormancy in barley using the Harrington/Morex and Chevron/Stander mapping populations. Crop Sci. 49(3): 841. doi: 10.2135/cropsci2008.05.0269.

Lorenz, A.J., K.P. Smith, and J.-L. Jannink. 2012. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. Crop Sci. 52(4): 1609. doi: 10.2135/cropsci2011.09.0503.

Massman, J.M., H.G. Jung, and R. Bernardo. 2013. Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Sci. 53(1): 58–66. doi: 10.2135/cropsci2012.02.0112.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4): 1819–1829. http://www.genetics.org/content/genetics/157/4/1819.full.pdf.

Mohammadi, M., T.K. Blake, A.D. Budde, S. Chao, P.M. Hayes, et al. 2015. A genome-wide association study of malting quality across eight U.S. barley

breeding programs. Theor. Appl. Genet. 128(4): 705–721. doi: 10.1007/s00122-015-2465-5.

Nagel, M., A.M. Alqudah, M. Bailly, L. Rajjou, S. Pistrick, et al. 2019. Novel loci and a role for nitric oxide for seed dormancy and preharvest sprouting in barley. Plant. Cell Environ. doi: 10.1111/pce.13483.

Nakamura, S., M. Pourkheirandish, H. Morishige, Y. Kubo, M. Nakamura, et al. 2016. Mitogen-activated protein kinase kinase 3 regulates seed dormancy in barley. Curr. Biol. 26(6): 775–781. doi: 10.1016/J.CUB.2016.01.024.

Oberthur, L., T.K. Blake, W.E. Dyer, and S.E. Ullrich. 1995. Genetic analysis of seed dormancy in barley (Hordeum vulgare L.). J. Agric. Genomics 1: 1–10. https://www.cabdirect.org/cabdirect/abstract/20063161122.

Rodríguez, M.V., M. Margineda, J.F. González-Martín, P. Insausti, and R.L. Benech-Arnold. 2001. Predicting preharvest sprouting susceptibility in Barley. Agron. J. 93(5): 1071. doi: 10.2134/agronj2001.9351071x.

Sallam, A.H., and K.P. Smith. 2016. Genomic selection performs similarly to phenotypic selection in barley. Crop Sci. 56(6): 2871. doi: 10.2135/cropsci2015.09.0557.

Sato, K., M. Yamane, N. Yamaji, H. Kanamori, A. Tagiri, et al. 2016. Alanine aminotransferase controls seed dormancy in barley. Nat. Commun. 7(1): 11625. doi: 10.1038/ncomms11625.

Tiede, T., and K.P. Smith. 2018. Evaluation and retrospective optimization of genomic selection for yield and disease resistance in spring barley. Mol. Breed. 38(5): 55. doi: 10.1007/s11032-018-0820-3.

Zhou, G., J. Panozzo, X. Zhang, M. Cakir, S. Harasymow, et al. 2016. QTL mapping reveals genetic architectures of malting quality between Australian and Canadian malting barley (Hordeum vulgare L.). Mol. Breed. 36(6): 70. doi: 10.1007/s11032-016-0492-9.

CHAPTER 2

A CONNECTED HALF-SIB FAMILY TRAINING POPULATION FOR GENOMIC PREDICTION IN BARLEY[1]

Abstract

Genomic prediction accuracy is affected by population size, trait heritability, relatedness of training and validation populations, marker density, and genetic architecture. Nested-association mapping (NAM) populations have advantages in many of these features compared to biparental families and may be an effective strategy for increasing prediction accuracy. The classic NAM design was modified to create a two-row spring malting barley population of 1341 F_3:F_4 lines in seven families that was phenotyped for heading date, plant height, leaf rust, spot blotch, preharvest sprouting, and grain protein. Quantitative trait loci (QTL) were detected for plant height, leaf rust, preharvest sprouting, and spot blotch with genome-wide association analyses. Prediction accuracies were assessed in validation populations consisting of a single family or multiple families. Across-family prediction accuracy (0.607-0.811) generally surpassed within-family prediction accuracy, particularly for traits with high across-family variance. Reductions in marker density (70-80%) and training population size (25-50%) did not cause significant loss of prediction accuracy. Addition of fixed marker effects from genome-wide association had minimal impact on prediction accuracy in the full training population but improved accuracy in reduced training populations. Within-family prediction for traits highly influenced by family structure was improved by adding half-sibs to the training population. Connected half-sib training populations could be useful for new and established

breeding programs looking to implement genomic selection due to benefits of family structure on prediction accuracy, genotyping, genetic diversity, and genetic mapping.

[1]Originally published as Sweeney, D.W, Rutkoski, J.R., Bergstrom, G.C., and Sorrells, M.E. (2020). "A connected half-sib family training population for genomic prediction in barley." Crop Science 262-281, 60(1).

Introduction

Population improvement is typically a lengthy process, with breeding cycle times in annual crops of 5 to 12 years depending on crop species, breeding program resources, and testing methods. Reduction in breeding cycle time could increase rates of gain from selection and improve response rate to new variety needs created by climate change or shifting consumer preferences. Off-season nurseries or greenhouses and more recent approaches like speed breeding (Watson and Ghosh et al., 2018) reduce breeding cycle time up to a certain limit by increasing the number of generations per year but they do little to improve selection accuracy or relieve the phenotyping bottleneck. The adoption of genomic selection in plant breeding introduced the potential to substantially reduce breeding cycle time by enabling selection of quantitative traits prior to phenotyping (Heffner et al., 2009). The use of the term genomic selection here refers to a genomics-assisted breeding program that bases selection on genomic estimated breeding values rather than estimated breeding values based solely on phenotype whereas genomic prediction refers simply to the use of genome-wide marker data to predict breeding values. In a genomic selection breeding scheme, genetic values are estimated with genomic prediction in a training population (TP) that has been phenotyped and genotyped with genome-wide molecular markers. This population is then used to predict genetic values of related germplasm based on marker-based relationship, allowing selection to occur amongst unphenotyped lines. Genomic prediction accuracy (PA) (Heffner et al., 2011; Asoro et al., 2013; Reidelsheimer et al., 2013) and realized gain (Beyene et al., 2015; Rutkoski et al., 2015; Sallam and Smith, 2016) are not consistently better than phenotypic selection

methods in short-term breeding programs but reduced cycle time and cost through prediction of early-generation breeding material are expected to lead to higher genetic gain per unit time and cost in the long term, assuming sufficient PA (Heffner et al., 2010). The accuracy of genomic prediction largely depends on properties of the population used for model training. Training population size (Lorenz et al., 2012; Lorenz, 2013), marker density (Lorenzana and Bernardo, 2009), phenotypic variance (Marulanda et al., 2015), and relatedness to validation population (VP) (Windhausen et al., 2012; Lehermeier et al., 2014) are important factors determining the performance of a prediction model.

Another contributor to TP performance is population structure. Within plant breeding programs, populations are frequently structured by family and selection whereas more diverse plant breeding populations are additionally stratified by geographical origin (Muñoz-Amatriaín et al., 2014) or morphology (Hamblin et al., 2010). Genomic selection within biparental families has been a popular approach because relatively few individuals and markers are needed for model training due to low rates of linkage disequilibrium (LD) decay with physical distance and no population structure (Lorenzana and Bernardo, 2009; Heffner et al., 2011; Crossa et al., 2014). However, model training and prediction within a biparental requires phenotyping of new families before genomic selection can occur and PA may be limited by family size or phenotypic variance (Marulanda et al., 2015). More complex TPs can improve these shortcomings but population structure must be considered for several reasons: 1) structure may bias PA estimates (Riedelsheimer et al., 2012; Guo et al., 2014), 2) allele frequencies differ across sub-populations and can inflate genetic

effect estimates (Guo et al., 2014), 3) differences in LD phase and persistence across sub-populations can negatively impact PA (Goddard and Hayes, 2007; Windhausen et al., 2012; Jacobson et al., 2014), and 4) unrelated groups tend to have limited ability to predict each other (Lehermeier et al., 2014).

Training populations can be carefully designed to leverage population structure in attempts to improve genomic prediction accuracy. One particularly powerful family structured TP for genomic prediction is a nested association mapping (NAM) population, which is composed of multiple recombinant inbred line (RIL) families with a common parent and diverse founder lines (Yu et al., 2008). This structure enables linkage analysis and association mapping in an efficient, highly powered, and diverse germplasm resource. Lehermeier et al. (2014) showed that combinations of half-sib families from a European maize NAM could produce equivalent or better PA than a single biparental but only at increased TP size. Prediction accuracy for height in the original maize NAM was higher across all families than in biparentals at large and small TP sizes (Peiffer et al., 2014). The ability of a single half-sib family to predict another half-sib family is generally poor (Lehermeier et al., 2014; Peiffer et al., 2014), which makes use of connected half-sib families as a TP an attractive option due to the potential to improve prediction across and within families, increased allelic diversity, and elimination of the problem of unrelated material in a TP. This strategy also avoids reduced PA from unequal linkage phases. Half-sibs that share a common parent are expected to have equal linkage states and combinations of multiple half-sib families should continue to share the same linkage state as long as high LD within family prevents the creation of negative LD due to admixture (Lehermeier et al., 2014).

A NAM population has high power for genome-wide association studies (GWAS) (Yu et al., 2008). Prediction accuracy may be improved by including large effect markers as fixed effects in the genomic prediction model (Rutkoski et al., 2012; Bernardo, 2014), including markers discovered through *de novo* GWAS (Spindel et al., 2016; Moore et al., 2017). Bian and Holland (2016) proposed genomic prediction and GWAS in a nested marker effect model in the original maize NAM to allow for family-specific allelic effects. Addition of significant marker effects from the nested model increased PA for two complex disease traits, but not for plant height, a highly polygenic trait. Despite these benefits, full NAMs are not practical for applied genomic selection. Classic NAMs are very large, time consuming to create and phenotype, and often include highly diverse, unadapted founders. Founder lines in a NAM are densely genotyped or fully sequenced to provide dense genome-wide markers for association mapping but high PA in genomic prediction can be obtained with low marker density (Lorenzana and Bernardo, 2009; Lorenz et al., 2012; Abed et al., 2018).

A number of genomic prediction studies have been published in barley (*Hordeum vulgare* L.) and have assessed PA in breeding populations with complex pedigrees (Lorenz et al., 2012; Sallam et al., 2015; Abed et al., 2018; Tiede and Smith, 2018), advanced breeding lines (Nielsen et al., 2016; Schmidt et al., 2016), or biparentals (Lorenzana and Bernardo, 2009). These studies have all used malting barley germplasm, which is grown for brewing and distilling. Malting barley production in the United States is mostly located in western states such as Colorado, Idaho, Montana, and North Dakota with much of the grain produced for large-scale

domestic brewing. Craft brewing has helped create a demand for locally grown, high quality malting barley in several new regions including the northeastern United States. This demand is best illustrated by the enactment of the New York Farm Brewery bill in 2012, which incentivized the use of locally grown grain for New York brewers (New York State Brewers Association, 2018). To qualify, farm breweries must use a quota of New York grown grain that increases from 20% in 2013 to 90% in 2023. Many current North American varieties, bred for drier western environments, perform poorly in the wetter climate of the Northeast. Selection for high malting quality in barley can lead to varieties with low seed dormancy. When these varieties are exposed to excessive moisture late in the growing season, preharvest sprouting (PHS) may occur before the grain is harvested. This reduces storage time and germination capability of malting barley and leads to rejection of grain lots for malting in severe cases. In humid conditions, fungal pathogens can reduce yield or decrease malting quality through mycotoxin production and kernel discoloration. To meet increasing demand for locally adapted varieties, new malting barley varieties adapted to wetter climates are needed within a short period of time to produce high quality grain for craft brewing.

Our goal was to use some of the unique attributes of a NAM population to create a TP with higher PA than biparental populations. We developed a population of connected half-sib families in two-row spring malting barley that was used to train genomic prediction models and conduct GWAS for agronomic and quality traits. This population can be used to initiate a genomic selection program to address demand for locally adapted malting barley. The specific objectives were to 1) assess the ability of

connected half-sib family TPs to predict across and within families, 2) test the ability of models incorporating fixed marker effects from GWAS to increase PA, 3) determine if reduced TP size or marker density changed PA, and 4) evaluate broader predictive power of the TP by predicting subsets of the TP grown in regional yield trials in a different year.

Materials and methods

Population development

Eight spring two-row malting barley varieties were selected as founder lines. 'Conlon' (PI 597789), 'Pinnacle' (PI 643354), and 'ND Genesis' (PI 677345) are varieties from the North Dakota State University breeding program, 'Craft' (PI 646158) from the University of Montana, 'Newdale' (Legge et al., 2008) and 'AAC Synergy' (Legge et al., 2014) from Agriculture and Agri-food Canada, and 'Bentley' (Juskiw et al., 2009) from the Field Crop Development Centre. 'KWS Tinka' (PI 681721) is a European adapted variety from KWS SAAT SE. AAC Synergy was used as a female in all crosses. Crosses were made in spring 2016 in greenhouse facilities at Cornell University in Ithaca, NY. Generations were advanced by single seed descent in the greenhouse with a target population of 200 individuals per biparental family. Some F_3 plants did not germinate or produce sufficient seed and a total of 1341 F_3 plants were harvested. $F_{3:4}$ seed was bulked from each F_3 plant for field planting.

Phenotypic data

The training population (TP17) was grown in one-meter long headrows in an augmented design at the Helfer and Snyder fields in Ithaca, NY on 12 May 2017 with the eight founder lines included as replicated checks. Phenotypes were recorded for plant height, heading date, leaf rust (*Puccinia hordei* G. H. Otth), PHS, grain protein, and spot blotch *(Bipolaris sorokiniana* Sacc). Plant height was measured as the average height from the ground to spike tip in a headrow. Heading date was recorded as Julian date of 50% spike emergence. Leaf rust severity was measured on a modified Cobb scale as a percentage (0-90) on five flag leaves in each headrow after full head emergence (Peterson et al., 1948). Scoring relied on natural infection which was consistent throughout both trials. PHS was measured on five spikes harvested at physiological maturity, after-ripened for three days, and misted in a greenhouse for three days. PHS severity was assessed on a zero to nine scale according to Anderson et al. (1993). Grain protein was measured by NIR using a prediction model with an R^2 of 0.9711 and standard error of 0.28 (NIRsystems 6500, Foss). *Bipolaris sorokiniana* isolate Bs233NY15, derived from infected barley in central New York and the most aggressive of four isolates tested in a preliminary inoculation of some commonly grown spring barley varieties, was selected for this research. The fungus was cultured on potato dextrose agar (39 g BD Difco™ Potato Dextrose Media, 1 L DI H_2O) for 8 days and conidia were harvested and suspended in water. Spore concentrations were assessed by microscopic examination with a hemacytometer and adjusted with water to achieve an inoculum suspension of 35,000 conidia/mL. Inoculum was applied to foliage at both locations on 14 June 2017 with a pressurized backpack sprayer.

Disease severity was measured on a one to nine scale according to Fetch and Steffenson (1999) on 5 flag leaves in a headrow after full head emergence. A portion of TP17 was not phenotyped for spot blotch due to heat stress and premature senescence. PHS and grain protein were partially phenotyped in each location due to time sensitive sampling and labor.

A subset of 100 lines from TP17 were grown in replicated regional yield trials (RYT) in five locations in central and western New York in 2018. The subset was selected for leaf rust, PHS, and spot blotch resistance. All locations were phenotyped for height, yield, grain protein, and natural spot blotch and leaf rust infection. Two local RYT18 were phenotyped for heading date and PHS.

Genotypic data

Seedling leaf tissue samples were collected in the field at a single location from a single plant in each headrow. DNA was extracted from lyophilized leaf tissue using cetyl trimethylammonium bromide extraction and genotyped with genotyping-by-sequencing (GBS) (Poland et al., 2012). Four 384-plex libraries were assembled. Each of the eight founder lines were included on each 96 well plate so that they were present four times in each library (96-plex) and 16 times overall. A total of 82,561 single nucleotide polymorphisms (SNP) were called using the TASSEL v2 GBS pipeline (Bradbury et al., 2007) and aligned to the barley reference genome (Mascher et al., 2017). Sites with sequence read depth below two were removed. A total of 13,663 markers remained after the founder lines were filtered to remove any missing, monomorphic, heterozygous, or indel sites. Markers with >50% missing data in the

full population were removed, leaving 12,784 markers. Sites with equal allele calls between founder lines were forced to be monomorphic within family. Missing polymorphic markers were imputed within family using the expectation-maximization (EM) algorithm described in Poland et al. (2012) in the *rrBLUP* R package (Endelman, 2011). Markers with minor allele frequency between 0.01 and 0.25 were removed within family to meet expectations of a biparental population and markers with heterozygosity values over 13% were removed to meet expectations of F_4 lines. A total of 9962 markers remained after filtering. This marker set was additionally filtered by LD pruning markers with $r^2 > 0.8$ in the *SNPRelate* R package (Zheng et al., 2012). A final set of 6009 markers was used for analysis. Identity-by-state (IBS) relationship between founders was calculated according to Rincent et al. (2012).

Statistical modeling

Phenotypic observations were adjusted with the model

$$y_{ijk} = \mu + l_i + b(l)_{ij} + g_k + e_{ijk} \quad (2.1)$$

where y_{ijk} are phenotypic observations, l_i are fixed location effects, and $b(l)_{ij}$, g_k, and e_{ijk} are random effects for block in location, entry, and error with $b(l)_{ij} \sim N(0, \sigma^2_{b(l)})$, $g_k \sim N(0, \sigma^2_g)$, and $e_{ijk} \sim N(0, \sigma^2_e)$ where $\sigma^2_{b(l)}$ is the block variance, σ^2_g is the genetic variance, and σ^2_e is the residual variance. For leaf rust and spot blotch, heading date was added to the model as a fixed covariate and for PHS scoring date was added as a fixed covariate. Variance components were estimated using the *lme4* R package for TP17 (Bates et al., 2015). Generalized measures of heritability (H^2) on an entry mean basis for TP17 were calculated across and within family by $H_g^2 = 1 - \frac{\bar{v}_{BLUP}}{2\sigma_g^2}$ where σ^2_g

is the genetic variance in the non-inbred base population and $\bar{v}_{BLUP}$ is the average variance of the difference between two BLUPs (Cullis et al., 2006). Genetic variance across (σ^2_{ga}) and within (σ^2_{gw}) families was estimated by omitting the genetic effect g_k and adding a family effect f_l and an entry within family effect $g(f)_{kl}$ as in Windhausen et al. (2012). A random effect, $g(l)_{ik}$, was included for RYT18 analysis *with* $g(l)_{ik} \sim N(0, \sigma^2_{g(l)})$, to account for genotype-by-environment interactions. Broad-sense H^2 for RYT18 on an entry mean basis was calculated as $\sigma^2_g/(\sigma^2_g + \sigma^2_{gl}/l + \sigma^2_e/nl)$ where σ^2_{gl} is the genotype-by-location interaction variance, σ^2_e is the error variance, l is the number of locations, and n is the number of replications in each location. Variance components for RYT18 were estimated using *ASReml* R (Gilmour et al., 2009).

Genome-wide association

The *rrBLUP* package (Endelman, 2011) was used to fit GWA models of the form

$$y = \boldsymbol{X}\beta + \boldsymbol{Z}u + \boldsymbol{S}\tau + e \quad (2.2)$$

where y is a vector of unadjusted phenotypes, β is a vector of fixed effects accounting for location, additional environmental covariates, and population structure, u is a vector of random genetic effects with $u \sim N(0, \mathbf{K}\sigma^2_g)$ where **K** is the additive relationship matrix calculated from marker data, and τ is a vector of fixed additive marker effects. **X** is an incidence matrix relating fixed effects to phenotypes, **Z** is an incidence matrix relating genetic values from the additive relationship matrix to phenotypes and **S** is an incidence matrix relating fixed marker effects to phenotypes. Additional fixed environmental covariates included heading date for leaf rust and spot blotch and scoring date for PHS. GWAS were conducted for each trait separately. The

full population and the LD pruned marker set of 6009 markers were used for the analyses. Principal component analysis was used to assess population structure. One principal component was added as a fixed effect in the GWA model to account for population structure. The genomic additive relationship (**K**) matrix was created using the A.mat function in *rrBLUP*. A false discovery rate of 0.05 was used to determine marker significance.

Genomic prediction

Single trait single step genomic best-linear unbiased prediction (GBLUP) models were fit in *ASReml* R with the model

$$y = \boldsymbol{X}\beta + \boldsymbol{Z_1}u_1 + \boldsymbol{Z_2}u_2 + e \quad (2.3)$$

where y is a vector of unadjusted phenotypes, β is a vector of fixed location and environmental effects, u_1 is a vector of random block effects, and u_2 is a vector of random entry effects. Block effects were nested within location. Additional fixed covariates included heading date for leaf rust and spot blotch and scoring date for PHS. For models incorporating fixed marker effects from GWAS, markers were added as single, categorical fixed effects in β. Random effects u_1 and u_2 were assumed to be normally distributed with $u_1 \sim \mathrm{N}(0, \sigma_b^2)$ where σ^2_b is the block variance and $u_2 \sim \mathrm{N}(0, \mathbf{G}\sigma_g^2)$ where **G** is the genomic relationship matrix calculated from marker data with the first method in VanRaden (2008), and σ^2_g is the additive genetic variance. **X** is an incidence matrix relating fixed effects to phenotypes, $\mathbf{Z}_1$ is an incidence matrix relating random block effects to phenotypes, and $\mathbf{Z}_2$ is an incidence matrix relating genetic effects from the genomic relationship matrix to phenotypes.

The vector of residuals *e* is assumed to be normally distributed with $e \sim N(0, \mathbf{I}\sigma_e^2)$ where **I** is an identity matrix and σ^2_e is the error variance. The same model was fit for pedigree-based relationship with $u_2 \sim N(0, \mathbf{P}\sigma_g^2)$ where **P** is the pedigree relationship matrix constructed using the *pedigreemm* R package (Vazquez et al., 2010).

Cross-validation

We compared two general methods of cross-validation, across-family (CV_A) and within-family (CV_W), to compare effects of family structure in the TP (Table 2.1). Across and within-family refer to the makeup of the VP; CV_A methods predict a multi-family VP and CV_W methods predict a single-family VP. Prediction accuracy from cross-validation was defined as the Pearson correlation between observed BLUP values from Equation 2.1 and predicted BLUP values from Equation 2.3 divided by the square root of heritability. Within each method we tested four different scenarios:

$\underline{CV_{A1}}$: A structured TP was used to predict a structured VP. Training populations were sampled five-fold across each family twenty times. 80% of each family was used in the TP and the remaining 20% from each family was assigned to the VP. Pedigree BLUP models were only evaluated in this scenario. The ability of large effect markers detected through GWAS to improve PA was also tested in CV_{A1}. Imputed markers with decimal scores were rounded to 0, 1, or 2 to reduce the number of effect levels and improve effect estimation. For height, leaf rust, PHS, and spot blotch, GWA models were fit with the TP in each cross-validation fold and the marker with the largest $-\log_{10}(p)$ value was added as a fixed effect in the GBLUP model. Fixed marker models were not evaluated for heading date and protein because no

significant marker-trait associations were discovered. The same TP and VP were used in each fold across GBLUP and pedigree models.

Table 2.1: Description of methods for across-family (CV_A) and within-family (CV_W) cross-validation

CV method	Across or within family prediction	Training population	Validation population	Additional factors affecting prediction accuracy
CV_{A1}	Across	Sample each full-sib family	Remaining lines in each full-sib family	With and without fixed markers in TP
CV_{A2}	Across	Sample each full-sib family	Remaining lines in each full-sib family	Reduced marker density in TP
CV_{A3}	Across	Sample each full-sib family	Fixed sample size of remaining lines in each full-sib family	Reduced TP sample size; with and without fixed markers in TP
CV_{A4}	Across	All TP17	RYT18 trials	With and without fixed markers in TP
CV_{W1}	Within	Sample within one full-sib family	Remaining lines within full-sib family	
CV_{W2}	Within	All of six full-sib families	Sample within seventh full-sib family	
CV_{W3}	Within	All of six full-sib families and sample within seventh full-sib family	Sample of remaining lines within seventh full-sib family	
CV_{W4}	Within	All of one full-sib family	Sample within each half-sib family	

$\underline{CV_{A2}}$: Impacts of reduced marker density were tested by sampling subsets of 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, and 90% of the full marker set evenly and randomly

across each chromosome. Sampling was repeated 25 times at each subset size and each marker sample was used in a single five-fold cross-validation. Sampling of TP and VP was the same as $CV_{A1.}$ The same TP and VP were used in each fold across marker densities.

CV_{A3}: Reduced TP sizes were tested by evenly and randomly sampling across families to form TPs of 5, 10, 25, 50, and 75% of the full TP. A VP was sampled from the remaining lines not used for the TP. To keep comparisons across TP sizes consistent, the VP sample size was set to 20% of the full TP (a single fold of a five-fold population sample). Sampling was repeated 100 times at each subset size and each TP/VP combination was evaluated once. CV_{A3} was additionally evaluated with fixed marker models with the same TP/VP combinations for height, leaf rust, PHS, and spot blotch as described in CV_{A1}.

$CV_{A4:}$ The ability of TP17 to predict subsets in unobserved environments was tested. The full TP17 was used to predict 100 samples of each RYT18 location. CV_{A4} was also evaluated with fixed marker training models for height, leaf rust, PHS, and spot blotch with the same sampling method. The most significant marker detected in GWAS with the full TP was included as a fixed effect. The square root of the RYT18 heritability was used to calculate PA.

CV_{W1}: Within-family PA was assessed in a similar manner to CV_{A1}, but here samples of TP and VP were taken from a single biparental family twenty times. Five-fold sampling was used so that 80% of the family was used to predict the remaining 20%.

$\underline{CV_{W2}}$: Ability of related material to predict within a biparental family with leave-one-family-out cross-validation was evaluated, similar to Lehermeier et al. (2014). A TP of six half-sib families was used to predict a VP of the seventh biparental family. For each biparental family, the VP was composed of 20% of the family and all available half-sib lines were used in the TP. Sampling of the VP was repeated 100 times. The square root of the heritability of the VP was used to calculate PA.

$\underline{CV_{W3}}$: Ability of related material to predict within a biparental family was evaluated as in CV_{W2} but full-sib lines that were not included in the VP were added to the TP. The TP was composed of all half-sib and 20% of the full-sib lines not used in the VP. A disjoint set of 20% of each family was used as the VP, which was sampled 100 times.

$\underline{CV_{W4}}$: Ability of a biparental family to predict half-sib families was assessed. All lines in each biparental family were used as the TP to predict a subset of 20% of each half-sib family individually. Sampling for VP was repeated 100 times.

Results

Phenotypic data

Phenotypic summaries of TP17 and RYT18 are presented in Table 2.2. Heritabilities were moderate to high for all traits and all families in TP17 and were highly correlated with genetic variance estimates within each family, which varied greatly. There was no consistent relationship between identity-by-state (IBS) relationship of family founders (Supplemental Table 2.1) and within-family genetic variance. Patterns of

genetic variance ratios across and within families varied across traits (Supplemental Table 2.2). For leaf rust, spot blotch, and grain protein, genetic variance largely was attributed to across-family differences while heading date, height, and PHS showed greater proportions of variance within families.

Table 2.2: Means, genetic variance, and broad-sense heritability for families, the full 2017 training population, and the 2018 regional yield trials. Missing data within families is shown as a percentage when necessary. Family abbreviations are listed in the footnote. Heritabilities with an asterisk indicate family-trait combinations with negligible genetic variance but a non-zero generalized heritability estimate.

Trait	Statistic	SB	SC	SG	SN	SP	SR	ST	All 2017	All 2018
	n	191	181	196	200	185	195	193	1341	100
Heading date	μ	196	193	194	196	193	195	195	195	178
	σ^2_g	1.81	8.90	5.25	0.42	7.93	2.08	2.66	4.32	3.7
	H^2	0.81	0.88	0.87	0.70	0.90	0.75	0.78	0.87	0.88
Height	μ	66.6	65.5	67.4	63.7	64.4	67.6	67	66	74.7
	σ^2_g	1.42	5.14	6.80	3.30	4.71	1.30	21.42	6.08	10.03
	H^2	0.56	0.69	0.73	0.64	0.65	0.56	0.85	0.71	0.42
PHS	μ	3.4	4.7	4.1	4	4.4	4.7	5.3	3.5	3.3
	σ^2_g	0.38	3.08	2.66	0.22	3.14	3.86	1.78	2.47	1.58
	H^2	0.66	0.90	0.91	0.62	0.95	0.96	0.92	0.93	0.58
	% missing	62.8	56.9	68.9	58	67	66.7	70.5	64	0
Leaf rust	μ	4.8	3.1	3.1	4.8	2.6	2.3	3.1	3	0.14
	σ^2_g	0.21	0.93	0.34	0.02	0.43	0.48	0.35	0.47	0.03
	H^2	0.70	0.79	0.72	0.51*	0.72	0.69	0.74	0.70	0.44
Spot blotch	μ	2.5	3.6	2.7	2.2	3	5.1	2.3	4.3	1.7
	σ^2_g	0.17	0.11	0	0.06	0.2	0.72	0.44	0.29	0.51
	H^2	0.54	0.52	0.55*	0.53	0.55	0.70	0.63	0.65	0.72
	% missing	1.6	27.6	3.6	9	1.1	16.9	0.5	8.5	0
Grain protein	μ	13	11.8	11.1	12.8	12.3	12.6	12.3	12.3	10.8
	σ^2_g	0.19	0.27	0.27	0.3	0.32	0.2	0.21	0.62	0.12
	H^2	0.67	0.75	0.78	0.77	0.77	0.69	0.68	0.84	0.48
	% missing	72.8	69.1	66.8	66	69.2	70.8	76.2	69.7	0

n: population size; μ: trait mean; σ^2_g: genotypic variance; H^2: broad-sense heritability; SB: AAC Synergy/Bentley; SC: AAC Synergy/Conlon; SG: AAC Synergy/ND Genesis; SN: AAC Synergy/Newdale; SP: AAC Synergy/Pinnacle; SR: AAC Synergy/Craft; ST: AAC Synergy/KWS Tinka

Genotypic data

Genotyping-by-sequencing markers provided full genome coverage. Average distance between markers was less than 1 Mbp on each chromosome except 1H and 4H. Principal component analysis (PCA) showed clear clustering by family (Figure 2.1). The left-hand side of the PCA plot contains families derived from North American founders. North American two-row spring malting barley germplasm has a narrow genetic base due to high selection intensity for malting quality (Martin et al., 1991). Newdale is a parent of AAC Synergy, and so the SN family (yellow) resulted from a backcross and thus contained low genetic diversity. The ST family was a cross from a European and North American founder, which are rarely intermated, and as a result showed greater diversity than North American by North American crosses.

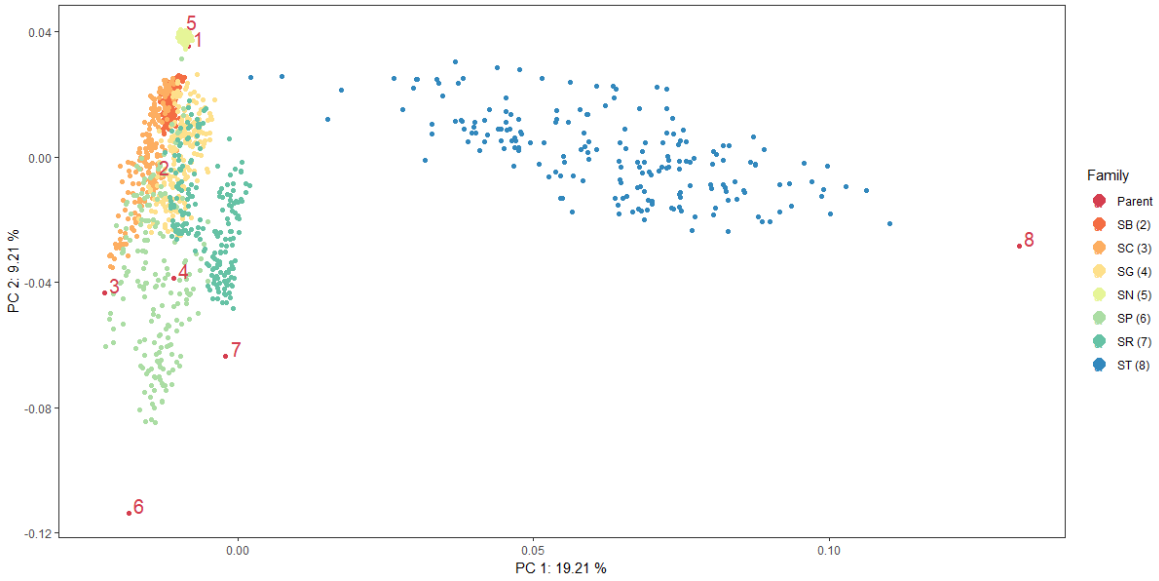


Figure 2.1: Principal component analysis of 6009 GBS markers. Red numbers signify founder lines (numbered in order of family with AAC Synergy as 1) and colored dots represent families.

Genome-wide association

Significant marker-trait associations (MTA) were found with GWAS for height, leaf rust, PHS, and spot blotch at a false-discovery rate of 0.05 (Table 2.3). No MTA were discovered for heading date or grain protein. The MTA for height on 3H may be linked to the *sdw1* locus, a commonly used semi-dwarfing gene in European malting barley. This locus encodes a gibberellin 20-oxidase-2 gene (*HvGA20ox2*) and has been cloned (Xu et al., 2017). The marker identified on 3H lies within 0.58 Mb of *HvGA20ox2* (634077598-634081600 bp). A single MTA was found for leaf rust on 5H that is not close to any previously described loci. This locus was also significant in a recent study of New York environments in similar germplasm over two years (Blachez, 2017). The PHS MTA may be linked to *SD2*, a mitogen-activated protein kinase kinase 3 that has shown involvement in abscisic acid and jasmonate signaling pathways (Nakamura et al., 2016). *SD2* has been described in many QTL mapping populations for dormancy (Prada et al., 2004; Hori et al., 1997) and has been cloned (Nakamura et al., 2016). The marker identified on 5H lies within 4 Mb of *SD2* (668430191-668438677 bp). The spot blotch MTAs on 3H and 7H may be linked to loci previously described by Zhou and Steffenson (2013), who found three loci on 1H, 3H, and 7H and referred to this haplotype as the "Midwest six-rowed durable resistant haplotype" (MSDRH). The MSDRH is also found in two-row germplasm, but disease severity tends to be higher in two-row than six-row varieties (Zhou and Steffenson, 2013). Physical positions of markers linked to the MSDRH were obtained from the barley 50K SNP chip (Bayer et al., 2017). The MTA identified on 3H lies within 0.2

MB of the MSDRH 3H locus (17409777 bp) and the 7H MTA lies within 4 Mb of the MSDRH 7H locus (30545668 bp).

Table 2.3: Significant (false-discovery rate 0.05) marker-trait associations for genome-wide association using the full 2017 training population. Segregating families indicates which families are polymorphic for the marker-trait association.

Trait	Chr	Marker position (bp)	Alleles	Segregating families	-log$_{10}$(p)	MAF	Percent σ_G^2 explained	Candidate locus
Height	3H	634581658	G/A	ST	11.17	0.055	17.6	*sdw1*
Leaf rust	5H	529384466	G/A	SC, SR	14.30	0.199	22.4	NA
PHS	5H	665601626	A/C	SC, SG, SP, SR, ST	26.69	0.336	55.7	*SD2*
Spot blotch	3H	17632511	A/G	SC, SR	5.75	0.119	11.3	*Rcs-qtl-3H-11_10565*
Spot blotch	7H	26379411	C/G	SR	5.36	0.069	12.3	*Rcs-qtl-7H-11_20162*

Prediction accuracy in CV_A

CV_A scenarios assessed across-family PA with structured TPs and structured VPs. CV_{A1} was designed to test PA across the full TP. Mean PA in CV_{A1} was moderate for height (0.610) and spot blotch (0.607) and high for heading date (0.736), leaf rust (0.811), PHS (0.774), and protein (0.711) (Figure 2.2). GBLUP models performed better than pedigree models for all traits. Differences in mean PA between GBLUP and pedigree models ranged from 0.045 for protein to 0.319 for PHS. Traits with a high proportion of within-family genetic variance showed a larger difference between GBLUP and pedigree models than traits with high across-family genetic variance. Fixed marker models did not significantly improve CV_{A1} PA except for PHS and decreased PA for leaf rust.

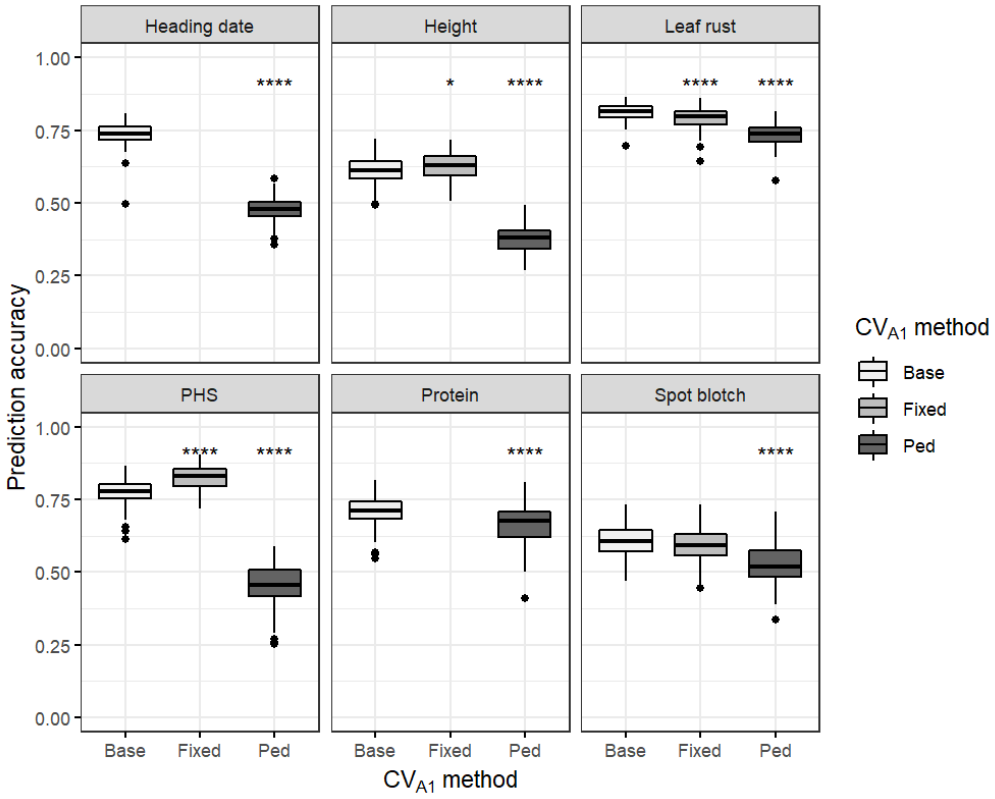


Figure 2.2: Prediction accuracy across traits in CV_{A1} with base GBLUP (Base), pedigree (Ped), and GBLUP with fixed marker effects (Fixed). Prediction accuracies were calculated using five-fold cross-validation 20 times. Boxplots with stars indicate significant difference in prediction accuracy from the base GBLUP model. Heading date and protein do not show prediction accuracy for "Fixed" models since no MTA were observed for those traits. Significance levels indicated by ***($p<0.01$), ****($p<.001$).

Reduced marker sets were evaluated in CV_{A2} using the same cross-validation strategy as CV_{A1}. Prediction accuracy significantly decreased ($p<0.05$) compared to full marker density at reduced marker densities of 1% for protein, 5% for spot blotch, 10% for heading date and height, and 20% for leaf rust and PHS and increased slightly at 80 and 90% for protein (Supplemental Figure 2.1). CV_{A3} tested performance of reduced TP size with a constant VP size. Prediction accuracy significantly decreased ($p<0.05$) compared to the full TP at 75% for leaf rust, 50% for height, PHS, and spot

blotch, and 25% for heading date and protein (Supplemental Figure 2.2). The change in PA between largest and smallest TP was smallest for leaf rust and largest for PHS. Increases in PA with models incorporating fixed marker effects in CV_{A3} were significant at reduced TP sizes for height, PHS, and spot blotch whereas leaf rust was unaffected by fixed marker models (Supplemental Figure 2.3).

CV_{A4} tested forward prediction of a TP subset in unobserved environments using a structured TP. This method was not a true cross-validation because all TP17 lines were used to predict subsets of RYT18. Each of five RYT18 was predicted with TP17. Two locations, RYT-1 and RYT-2, were grown locally in Ithaca, NY and were fully phenotyped. The three remaining RYT18 were not phenotyped for heading date or PHS. Leaf rust and spot blotch scores were based on natural infection. Leaf rust disease incidence was low in general and symptoms were not present in RYT-4 or RYT -5. Prediction accuracies were high for heading date (0.745-0.786), height (0.623-0.917), and PHS (0.918-0.969), and moderate for leaf rust (0.405-0.720), protein (0.199-0.755), and spot blotch (0.079-0.538) using TP17 to predict RYT18 (Table 2.4). Fixed marker models for TP17 did not improve PA for any trait when averaged across environments.

Table 2.4: Mean CV_{A4} prediction accuracy in each RYT18 location (validation population, VP) using TP17 as the training population. Prediction accuracies were averaged across 100 samples of 20 lines in each RYT18 location. Trait + GWAS models incorporated fixed effects not the training model with the same validation population samples.

VP	Heading date	Height	Height + GWAS	Leaf rust	Leaf rust + GWAS	PHS	PHS + GWAS	Protein	Spot blotch	Spot blotch + GWAS
RYT-1	0.786	0.804	0.758	0.515	0.495	0.969	0.847	0.755	0.538	0.414
RYT-2	0.745	0.917	0.877	0.405	0.388	0.918	0.596	0.456	0.370	0.339
RYT-3	NA	0.877	0.842	0.720	0.718	NA	NA	0.580	0.132	-0.042
RYT-4	NA	0.790	0.751	NA	NA	NA	NA	0.199	0.144	0.154
RYT-5	NA	0.623	0.625	NA	NA	NA	NA	0.452	0.079	0.148
Mean	0.766	0.802	0.771	0.547	0.534	0.943	0.721	0.489	0.253	0.203

Prediction accuracy in CV_W

CV_W methods used structured or unstructured TPs to predict unstructured (biparental) VPs. The simplest CV_W method, CV_{W1}, used a similar sampling method as CV_{A1} but sampled TP and VP from a single biparental family instead of across several families. Mean PA for CV_{W1} varied widely across families for each trait, ranging from 0.310 (SN) to 0.715 (SG) for heading date, 0.071 (SB) to 0.689 (ST) for height, 0.108 (SP) to 0.460 (SR) for leaf rust, 0.096 (SN) to 0.714 (SC) for spot blotch, 0.328 (ST) to 0.804 (SR) for PHS, and -0.013 (ST) to 0.541 (SC) for protein (Figure 2.3). CV_{W1} surpassed CV_{A1} twice for height (SG, ST), once for PHS (SR), and once for spot blotch (SC). CV_{W2} used a TP composed of all half-sibs to predict full-sib lines. CV_{W2} performed better ($p<0.05$) than CV_{W1} in 19% of the trait-family combinations and worse in 62%, and on average, performed worse than CV_{W1} on average across all traits and families (Table 2.5). CV_{W3} was similar to CV_{W2} but augmented the half-sib TP with full-sib lines not used in the VP. This approach had higher PA ($p<0.05$) than

CV_{W1} in 36% of the trait-family combinations and lower PA in 12%. CV_{W3} produced equivalent or higher PA than CV_{A1} once for height (ST) and twice for PHS (SC, SR).

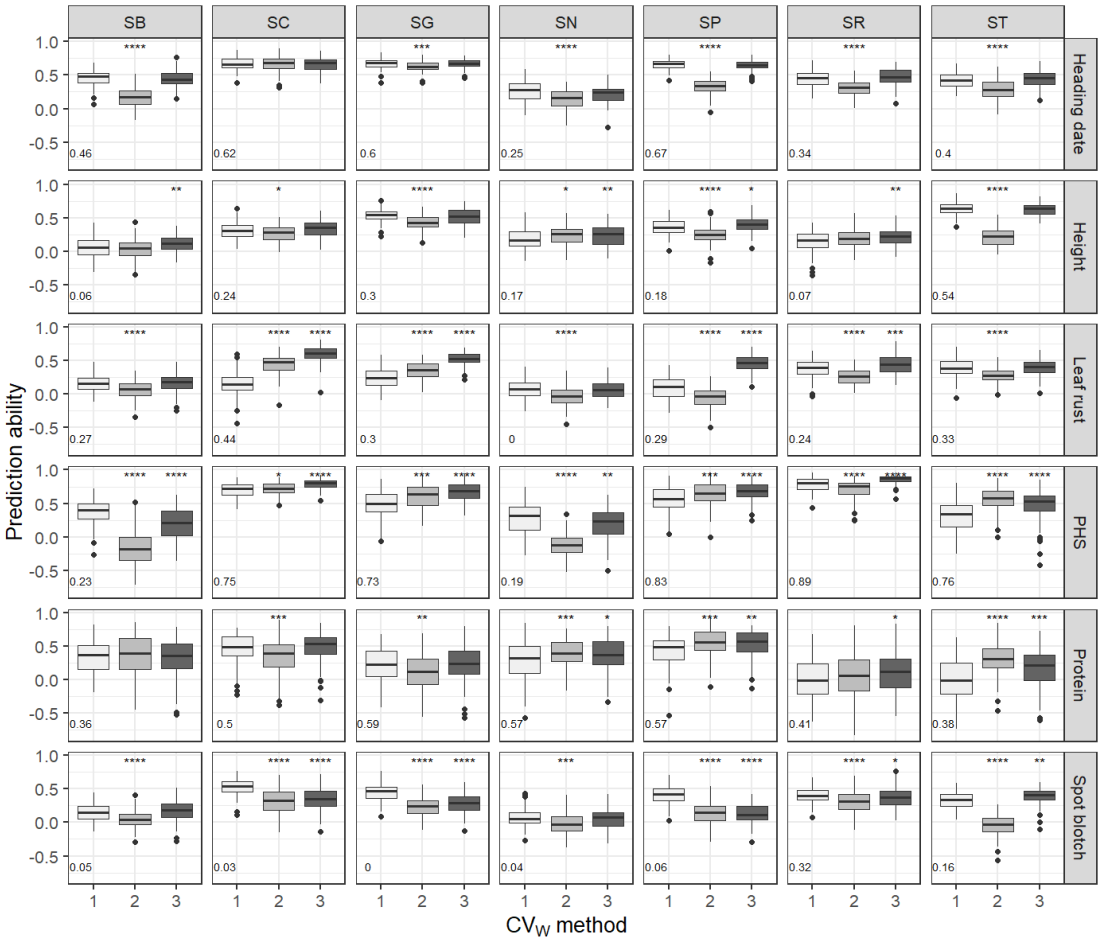


Figure 2.3: Mean prediction accuracies across seven families and six traits in CV_{W1}, CV_{W2}, and CV_{W3}. CV_{W1} sampled training population and validation population with five-fold cross-validation 20 times and CV_{W2} / CV_{W3} sampled validation population with five-fold cross-validation 20 times. Boxplots with stars indicate significant difference in prediction accuracy from CV_{W1}. Broad-sense heritabilities for each trait-family combination are printed in the lower left of each panel. Significance levels indicated by *(p<0.1), **(p<0.05), ***(p<0.01), ****(p<.001).

CV_{W3} performed as well or better on average than CV_{W1} and CV_{W2} in all families except SB and SN and all traits except spot blotch (Table 2.5). The number of half-

sibs from each family in the TP in CV_{W3} could be reduced by 50% and still perform better on average than CV_{W1}.

Single half-sib training families produced variable PA for biparental families in CV_{W4}. At least one half-sib training family performed better than CV_{W1} in each trait and in 60% of the trait-family combinations and produced higher PA than CV_{W3} in 16% of the family-trait combinations (Figure 2.4). Families with higher genetic variance tended to be able to predict and be predicted better than families with low genetic variance such as SB and SN. Methods to predict within families that incorporated related material in some form ($CV_{W2,}$ $CV_{W3,}$ or CV_{W4}) produced PA greater than or equal to standard within-family prediction (CV_{W1}) in 52% of family-trait combinations.

Discussion

Genomic selection is a promising method to quickly and accurately select superior progeny in a newly established breeding program, such as the Cornell two-row spring malting barley breeding program. We sought to increase PA by creating a connected half-sib TP, which had a similar structure to a traditional NAM. Creating a population with many diverse founder lines as in a NAM is impractical for applied breeding programs that largely rely on elite by elite crosses. Including feed or food barley

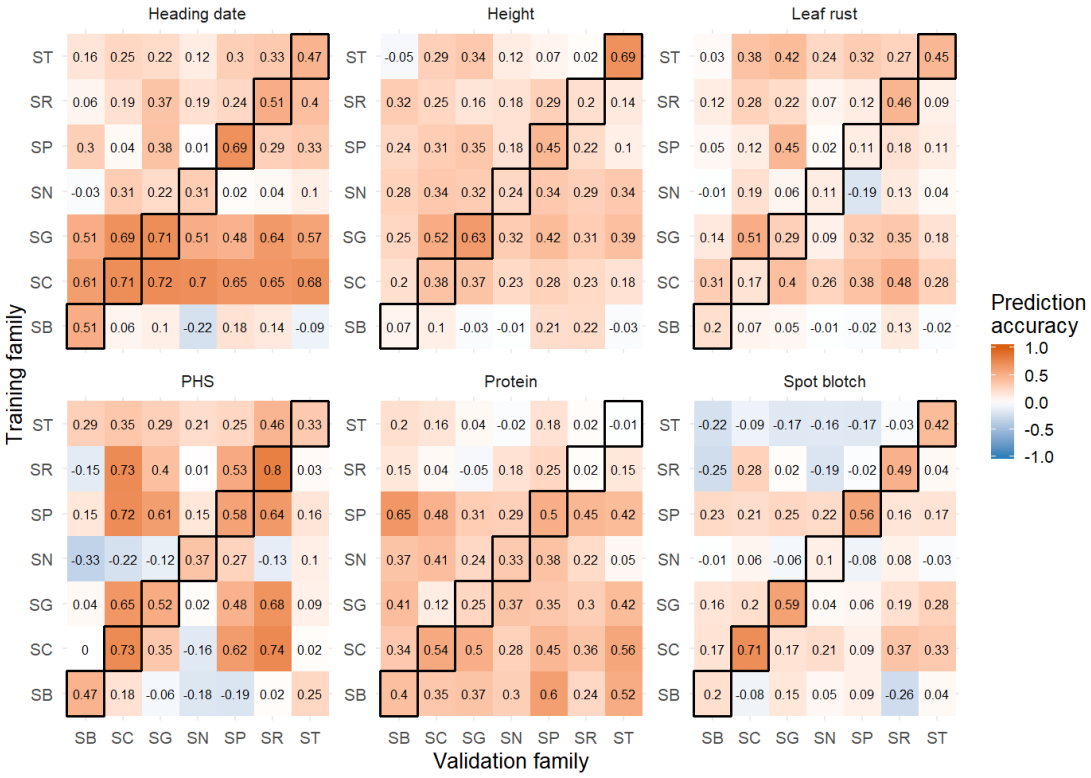


Figure 2.4: Average cross-validated prediction accuracy for each trait in CV_{W4}. Facets within each trait show prediction accuracy using a training population of a single half-sib family (y-axis) to predict a biparental family (x-axis). Each column represents all half-sib training families used to predict a single biparental family and each row represents prediction accuracy of a single biparental family across all half-sibs training families. Highlighted diagonals are prediction accuracies within each family (CV_{W1}).

Table 2.5: CV_W prediction accuracy averaged across families (a) and traits (b). Results for CV_{W3} used all available half-sibs in the training population and CV_{W3} 2-fold used 2-fold sampling to include 50% of the half-sibs from each family for the training population in each cross-validation fold. Asterisks indicate significant difference in PA compared to $CV_{W1.}$ Significance levels indicated by *($p<0.1$), **($p<0.05$).

Table 2.5a

CV method	SB	SC	SG	SN	SP	SR	ST	All families
CV_{W1}	0.307	0.541	0.498	0.241	0.480	0.415	0.391	0.411
CV_{W2}	0.100**	0.526	0.444**	0.123**	0.346**	0.343**	0.295**	0.311**
CV_{W3}	0.288	0.612**	0.550**	0.238	0.534**	0.469**	0.465**	0.451**
CV_{W3} 2-fold	0.318	0.599**	0.532**	0.208**	0.524**	0.453**	0.480**	0.445**
CV_{W4}	0.158**	0.262**	0.232**	0.128**	0.237**	0.262**	0.205**	0.212**

Table 2.5b

CV method	Heading date	Height	Leaf rust	PHS	Protein	Spot blotch	All traits
CV_{W1}	0.561	0.379	0.255	0.542	0.289	0.438	0.411
CV_{W2}	0.398**	0.280**	0.217**	0.427**	0.358**	0.185**	0.311**
CV_{W3}	0.560	0.417**	0.443**	0.584**	0.371**	0.330**	0.451**
CV_{W3} 2-fold	0.555	0.397	0.434**	0.590**	0.367**	0.325**	0.445**
CV_{W4}	0.296**	0.229**	0.181**	0.213**	0.295	0.059**	0.212**

founders could introduce useful alleles for disease resistance and yield but would hinder breeding progress for the more important malting quality traits with linkage drag or deleterious alleles. We only selected malting varieties as founders based on RYT performance over several years and added genetic diversity by including one European founder. Founder diversity should be carefully considered. Two of the founders used in this study, Bentley (IBS: 0.825) and Newdale (IBS: 0.948), were closely related to AAC Synergy. These two families showed low degrees of polymorphism and consequently had low genetic variance, heritability, and PA for several traits. Conversely, KWS Tinka was the most genetically distant founder (IBS: 0.423) to AAC Synergy, but this did not translate into a consistent increase in PA in

the ST family. Traditional NAMs are composed of RILs but inbreeding to the F_6 or later generation is time consuming and impractical for applied breeding. In a genomic prediction framework, precise estimates of QTL positions are not as important as in association mapping, so fully inbred lines are not required for model training. We were able to advance the TP from cross to field in a single year by single-seed descent in a greenhouse producing F_3-derived F_4 families, whereas RIL creation would have taken at least an extra year. Phenotypic measurements of the traits in this study were collected from headrows, which reduced field space and removed the need for an extra generation to increase seed for yield plots saving at least another year.

Family structure in the connected half-sib TP improved genotyping efficiency and imputation accuracy. The 25 founder lines of the original maize NAM were densely sequenced and high-density markers were projected onto 5000 RILs genotyped at lower density (Yu et al., 2008). Our genotyping strategy mirrored that of the maize NAM. Founders were genotyped at 96-plex and progeny were genotyped at 384-plex, which provided greater sequencing depth and coverage for founders. Studies incorporating GBS markers in genomic prediction have been published in wheat (Poland et al., 2012b; Rutkoski et al., 2014), but most barley genomic prediction studies have used genotyping arrays (Sallam et al., 2015; Schmidt et al., 2016; Nielsen et al., 2016). Abed et al. found that GBS marker density in an advanced barley breeding population could be reduced from ~30,000 markers to 2000 without loss of PA but also noted a large tradeoff between high multiplexing and missing data (2018). The TP described here is twice as large as the largest published barley training population (Sallam et al., 2015), making low cost GBS more attractive than

genotyping arrays. One disadvantage of genotyping F_3:F_4 lines is accuracy of imputing heterozygous sites but GBS marker imputation with EM has been shown to have low bias and produce good PA (Poland et al., 2012b; Rutkoski et al., 2013). Imputation based on founder marker scores allowed monomorphic markers within family to be accurately called and reduced the number of sites that needed to be imputed by EM.

The population size and structure in this study enabled detection of major QTL with GWAS. Low marker density and high LD limited the ability to highly resolve loci, but the identified MTAs were reasonably close to previously mapped or cloned loci. Observed $-\log_{10}(p)$ values met expectations at low values, validating the detected genetic signal. Power for PHS, protein, and spot blotch GWAS was limited by the number of lines that could be phenotyped but a highly significant MTA for PHS was still discovered on the distal end of chromosome 5H. Despite partial confounding of spot blotch phenotypes by environmental factors, two MTA were discovered that correspond to two-thirds of the previously described MSDRH. AAC Synergy, Bentley, Newdale, and ND Genesis were bred for regions with high spot blotch pressure. At least one of these founder lines could contain the full MSDRH but the third locus could not be resolved due to phenotype quality and quantity. Limited phenotypic data for grain protein contributed to a lack of clear QTL discovery, suggesting a more highly polygenic inheritance than PHS or spot blotch. No QTL were discovered for heading date despite high heritability and substantial genetic variance in several families. Spring barley can be susceptible to incomplete spike emergence when exposed to heat or drought stress, making heading date difficult to accurately score and contributing to error. This population was planted three to four

weeks later than the ideal planting time in central New York and was exposed to early heat stress, which may have confounded heading date scores.

Prediction accuracies from across-family prediction in CV_{A1} were high and resembled previously published results for heading date (0.55) (Abed et al., 2018), plant height (0.36 to 0.86) (Lorenz and Smith, 2015; Lorenzana and Bernardo, 2009), and grain protein (0.4 to 0.82) (Nielsen et al., 2016; Lorenzana and Bernardo, 2009). Lorenzana and Bernardo (2009) reported PA in biparental families for height (0.90) and grain protein (0.83) that surpass results from CV_{A1} and CV_{W1} in this study, but their datasets include 9 to 16 environments. No reports of PA for leaf rust, PHS, or spot blotch in barley have been previously published. GBLUP models had higher PA than pedigree models in CV_{A1} but differences were small for leaf rust, protein, and spot blotch. Pedigree relationship models effectively captured large across-family variances and were comparable to genomic relationship models for traits with large across-family variance. Across-family (CV_{A1}) PA was significantly larger than all within-family (CV_{W1}) PA for leaf rust (CV_{A1}: 0.811; CV_{W1}: 0.108-0.460) and protein (CV_{A1}: 0.671; CV_{W1}: -0.013-0.541). This indicates a particularly large effect of family structure for these traits.

Markers must be in sufficient LD with causal QTL to attain good PA (Meuwissen et al., 2001). Populations with high LD require fewer markers to adequately capture genome-wide QTL effects. Linkage disequilibrium in malting barley breeding populations tends to decay slowly due to self-pollination, limited founding material, and high selection intensity for malting quality (Kraakman et al., 2004; Hamblin et al., 2010; Schmidt et al., 2016). Linkage disequilibrium estimates in

this population were extensive (average LD decay below 0.2 r^2: 143 Mb). Several GP studies in barley have reported stable PA as marker density decreases (Lorenz et al., 2012; Nielsen et al., 2016; Abed et al., 2018), which support similar results in other crops (Lorenzana and Bernardo, 2009; Heffner et al., 2011; Spindel et al., 2016). Our results similarly support previous findings that marker density can be substantially reduced without a decrease in PA in malting barley. Marker sets at 20% density (~1200 markers) generally provided the same PA as 6009 markers. Results from CV_{A2} are supported by similar trends of PA stability at reduced marker sets within families in the CV_{W1} framework (data not shown). Marker sets at 1% density within families had high PA in some samples, particularly for PHS. A recurrent genomic selection program starting with a connected half-sib TP could continue to use GBS genotyping for future cycles. Using GBS data from the base population to create a population specific array or set of competitive allele-specific markers could be more practical since several hundred markers may be adequate for good PA with this design. Each cycle of selection would then be connected to the initial TP, allowing for imputation if necessary and decreasing the need to re-genotype the TP. This strategy would be of special interest if specific segregating genomic regions identified through GWAS were targeted for selection.

Training population size is often evaluated in genomic prediction studies. Malting barley studies (Lorenz et al., 2012; Nielsen et al., 2016) support general findings that PA generally increases as TP size increases, but reduced TP sizes can produce high PA if heritability and relatedness are sufficiently high. We found TP sizes could be reduced up to 50% in a structured population without a decrease in PA

in CV_{A3}. Increased PA in a multi-family TP can be partially due to increased population size (Schulz-Streeck et al., 2012) but population structure still has a large impact on PA when the difference in PA between the largest and smallest TPs is small (Windhausen et al., 2012; Zhao et al., 2012). The difference in PA between smallest and largest TP for leaf rust ($PA_{100\%}$-$PA_{5\%}$: 0.149) and protein ($PA_{100\%}$-$PA_{5\%}$: 0.198) was small in CV_{A3} (Supplemental Figure 2.2) and is further evidence for the large impact of family structure on these traits. Traits with high within-family genetic variance like height ($PA_{100\%}$-$PA_{5\%}$: 0.324) and PHS ($PA_{100\%}$-$PA_{5\%}$: 0.400) are less impacted by family structure and are more sensitive to reduced TP sizes, but these traits also benefit more from fixed marker models in smaller TPs (Supplemental Figure 2.3).

Ability to predict RYTs would be a useful TP feature. The number of training environments and the impact of genotype-by-environment (GxE) interaction effects impact PA across environments. Increasing the number of environments in which a TP is evaluated may increase PA (Endelman et al, 2014; Wang et al., 2014; He et al., 2016) or have little effect (Windhausen et al., 2012). Limited training environments (one or two) increase risk from poor phenotyping conditions and limit capture of GxE effects but simulations suggest that evenly or unevenly spreading lines across environments at low replication can still effectively capture environmental effects (Riedelsheimer and Melchinger, 2013; Endelman et al, 2014). Biparental simulations indicate that a single replication of a large TP produces PA equivalent to smaller TPs replicated four times (Lorenz, 2013). We evaluated single replications of TP17 lines in two locations in a single year due to seed and phenotyping constraints. Forward

prediction of RYTs (CV_{A4}) produced moderate to high PA for a variety of traits despite limited ability to capture GxE effects, likely due to the fact that RYT18 was a subset of TP17 and low GxE interaction in two-row spring barley for heading date, height, and grain protein (Neyhart et al., 2018). Little data exits for GxE interactions for PHS, spot blotch, or leaf rust. Selection for spot blotch and leaf rust resistance and uneven disease incidence due to reliance on natural infection in RYT18 likely reduced predictive ability of TP17 for those traits. Heading date and PHS are time sensitive to phenotype and require repeated visits to fully record all data. Genomic prediction in RYT of such traits is particularly attractive since this data would not otherwise be available. Prediction of quality traits like protein could decrease time intensive post-harvest phenotyping methods.

Addition of fixed marker effects detected through GWAS had variable impacts on PA in this population. Percent genetic variance explained by the most significant marker was over 10% for height, leaf rust, PHS, and spot blotch. Trait heritabilities were high for leaf rust and PHS and moderate for height and spot blotch. Fixed marker models showed a small PA increase for PHS in CV_{A1}, significantly improved PA for height, PHS, and spot blotch in CV_{A3}, and decreased PA for all traits in CV_{A4}. Prediction accuracy improved for PHS, a high heritability trait with a large effect locus, in two of the three CV_A methods evaluated with fixed marker models. Leaf rust had high heritability with a large effect locus but fixed marker models had minimal to negative impact on PA. Height was a low heritability trait with a moderate effect locus and fixed marker models had a significant positive impact on PA in CV_{A3} and a slight but non-significant improvement in CV_{A1}. Spot blotch had low heritability with two

small effect loci but fixed marker models improved PA in CV_{A3}. In the presence of substantial family structure, the allocation of variance among and within families may help explain the surprisingly improved PA of fixed marker models for height and unexpectedly worse PA for leaf rust. Prediction accuracies for leaf rust CV_A methods are largely dependent on among-family genetic variance. The effect of family structure in a trait with low within-family variance may be too high for a large-effect QTL to make a difference in PA. In contrast, height exhibited high within-family variance and the contribution of family structure to PA is thus reduced, which may increase the contribution of fixed marker effect estimates to PA. Smaller TPs benefited the most from fixed marker models in CV_{A3}, but in practice very small TPs may not have enough statistical power for QTL discovery. Smaller TPs in this dataset were consistently able to detect the same QTL as the full TP at 50% TP size (Supplemental Table 2.3). Results from CV_{A3} suggest that fixed marker models could be beneficial in maintaining PA at reduced TP size for height, PHS, and spot blotch. Reduced PA in CV_{A4} with fixed marker models was unexpected and may be due to large GxE interactions in RYT18 at the QTL detected by GWAS.

In theory, accounting for large effect markers as fixed effects in genomic prediction should improve PA and relative efficiency of selection when heritability is high and the marker explains over 10% of the genetic variance (Rutkoski et al., 2012; Bernardo, 2014). Empirical studies support incorporation of fixed marker effects from diagnostic markers (Daetwyler et al., 2014) or *de novo* GWAS (Spindel et al., 2016; Bian and Holland, 2017; Moore et al., 2017, Herter et al., 2018) in the TP to improve PA. In contrast, Rice and Lipka (2018) reported minimal to negative gains in PA from

combined genomic prediction and *de novo* GWAS across the majority of 216 genetic architecture simulations, possibly due to differences in marker effect across TP and VP. McElroy et al. (2018) similarly found little to no improvement in genomic prediction and *de novo* GWAS in cacao (*Theobroma cacao* L.). Our results are more similar to the findings reporting negligible benefit from fixed marker models than studies reporting significant improvements in PA. In this population, structure seems to have a greater impact on PA than fixed markers and addition of fixed marker effects only improved reduced TPs. The exception was PHS. Prediction of PHS with a mixed model including a single fixed effect for the PHS marker and no genomic relationship matrix produced equivalent PA to the GBLUP model (data not shown). These results suggest that genomic prediction may not be needed for PHS in this population. Considering the resources needed to collect PHS phenotypes and the susceptibility of PHS data quality to weather, genomics-based breeding approaches for PHS may still be valuable. Marker-assisted selection to target two consistently observed large effect QTL in malting barley, *SD1* and *SD2* (Ullrich et al., 2009), may be a more cost-effective method if PHS is the only trait under selection.

We evaluated PA in unstructured populations, specifically within biparental families, with CV_W methods. Despite the benefits of biparentals in genomic prediction, use of biparentals to predict other material is limited by allelic diversity and differences in LD and linkage phase (Windhausen et al., 2012). Several cross-validation methods that include structured material have been tested to increase prediction of full-sib lines (Albrecht et al., 2011; Schulz-Streeck et al., 2012; Windhausen et al., 2012; Zhao et al., 2012; Reidelsheimer et al., 2013; Guo et al.,

2014; Lehermeier et al., 2014). The two primary methods are TPs composed of structured groups, or more explicitly in this case, several half-sib families (CV_{W2}), or structured and unstructured groups, specifically several half-sib families with some full-sibs here (CV_{W3}). We also predicted full-sibs with single half-sib families (CV_{W4}). Reported performance of CV_{W2} and CV_{W3} methods varies depending on the type of population structure present. Our approach was similar to Reidelsheimer et al. (2013), Jacobson et al. (2014), Lehermeier et al. (2014), Wang et al. (2014), and Herter et al. (2018) since CV_{W2} and CV_{W3} tested TPs composed of highly related germplasm instead of germplasm from diverse breeding programs (Windhausen et al., 2012) or species-wide sub-populations (Guo et al., 2014).

Within-family prediction was enhanced when a full-sib TP was augmented with half-sib families (CV_{W3}) (Table 2.5). These results support the findings of Lehermeier et al. (2014) and Wang et al. (2014). Half-sib TPs with no full-sibs (CV_{W2}) performed poorly in comparison to CV_{W1} as in Reidelsheimer et al. (2013) and Jacobson et al. (2014). Lehermeier et al. (2014) found CV_{W2} generally matched or exceeded CV_{W1} but used twice as many families. Single half-sib family TPs (CV_{W4}) did not consistently exceed CV_{W1}, making CV_{W3} the most robust method we tested for improving within-family PA. CV_{A1} enables prediction across and within-families, but for traits with a high proportion of across-family genetic variance that are mostly predicted by family structure, actual within-family PA may be quite low (Zhao et al., 2012). CV_W, specifically CV_{W3}, complements CV_{A1} as an approach to improve within-family PA. Across all families, leaf rust and protein CV_{W3} PA was always as good or better than CV_{W1}. Sampling both across half-sib families and within a full-sib family

is a viable alternative to CV_{A1} for traits with a high proportion of across-family variance. Spot blotch PA was generally not improved by CV_{W3} and was lower than CV_{W1} for several families. Overall low heritability, low PA, and poor performance of CV_{W3} for spot blotch suggest that phenotyping error may have negatively impacted results. Improved PA for highly inbred families or low heritability family-trait combinations would be a useful attribute of CV_W methods but we did not observe any consistent improvements in either of these situations.

A connected half-sib TP provides the breeder with an opportunity to practically exploit family structure to increase PA. This approach may be beneficial for breeding programs with large sets of connected half-sib families (see Jacobson et al., 2014) that want to perform genomic selection on newly generated biparental families. The structure of a connected half-sib TP is also desirable for initiating a recurrent genomic selection program (Bernardo and Yu, 2007; Rutkoski et al., 2011). If parents for the first cycle of genomic selection are selected across and within families and intermated, significant population structure would still be present and contribute to PA. In a recurrent selection scheme, structure should persist as long as no new parents are introduced and marker effect estimates should also persist across multiple cycles of selection if a small number of parents are selected due to high LD (Windhausen et al., 2012). Both of these factors reduce phenotyping costs and the need for model updating, but this strategy has two drawbacks. Genetic variance must be carefully managed because genomic selection can deplete genetic variance more quickly than phenotypic selection (Rutkoski et al., 2014) which is particularly relevant when all selected crossing lines are half-sibs with a shared parent. In traits highly impacted by

family structure, reliance on prediction across families may strongly skew selections towards superior families, causing a more rapid increase in inbreeding and loss of potentially useful alleles (Crossa et al., 2013; Guo et al., 2014). True random mating is difficult in self-pollinated crops but making all pairwise crosses between selected lines, constraining coancestry of selected individuals while maximizing genetic value through optimum contribution selection (Meuwissen, 1997), or augmenting population wide selections with selections from within-family prediction models may help reduce inbreeding and maintain genetic variance over time. Programs with long-term breeding goals may opt to solely implement within-family selection to avoid strong bias towards families but could still benefit from a connected half-sib TP through CV_{W3} methods.

In this highly structured TP, the proportion of within-family genetic variance had a large impact on PA with CV_A and CV_W methods. Ideally, breeders choose parents that will maximize mean and genetic variance in their progeny but in reality, often make elite by elite crosses with highly related material that may have low genetic variance for some traits. Our results from CV_W suggested that adding half-sibs to the TP had limited ability to improve PA within families or traits with low genetic variance. A better option for increasing PA with CV_A or CV_W would be selection of parents that maximize within-family genetic variance for all traits under selection. Predicting genetic variance of a cross has historically been a difficult prediction problem, but recent studies have described several methods to improve prediction of progeny variance based on genome-wide molecular markers (Mohammadi et al., 2015; Lehermeier et al., 2017; Osthushenrich et al., 2018) although variable results have

been reported (Lado et al., 2017; Adeyemo and Bernardo, 2019; Neyhart and Smith, 2019). These methods require estimation of marker effects to predict progeny variance, meaning potential parents must be genotyped and phenotyped in a sufficiently large field experiment to accurately estimate marker effects. Parental selection for a connected half-sib TP based on predicted progeny variance is therefore unrealistic for new breeding programs with minimal germplasm resources and phenotypic data unless genomic estimated breeding values of potential parents are available from large, geographically diverse, and publicly available TPs. Collaborative TPs that span years and mega-environments such as that described by Neyhart et al. (2019) may provide datasets that enable accurate parental selection for new breeding programs based on expected progeny genetic variance. Initial empirical results from Neyhart and Smith (2019) suggest this approach may work for high heritability traits but not for those of low heritability.

We described the development and evaluation of genomic prediction performance of a connected half-sib TP using two-row spring malting barley. This strategy produced high PA across families and performed better than within-family prediction when half-sibs were added to the TP. Major QTL were detected in the full and reduced TPs using GWAS. Fixed marker models had little impact on PA in the full TP due to the impact of family structure, but they increased PA in reduced TPs. The connected half-sib TP showed good forward prediction of TP subsets in yield trials across environments but GxE effects may have negatively impacted PA across environments with fixed marker models. Connected half-sib TPs have potential benefits for applied genomic selection such as balancing across and within family

selection, improving genotyping efficiency, reducing TP size while maintaining power for GWAS, and generally high PA for well-established and newly created breeding programs.

Acknowledgments
We thank David Benscher, Jesse Chavez, Amy Fox, and James Tanaka for planting, phenotyping, harvesting, and greenhouse assistance. We are grateful for the contributions of Jamie Cummings for *Bipolaris sorokiniana* inoculum preparation and Julie Hansen for NIR grain protein phenotyping assistance. We also thank the breeders and companies for permission to use their varieties in crosses for this research. Funding for this project was provided by New York State Agriculture and Markets, the Genesee Valley Regional Market Authority, Hatch Project 149-447, and by the Agriculture and Food Research Initiative Competitive Grants 2011-68002-30029 (Triticeae-CAP) and 2017-67007-25939 (Wheat-CAP) from the USDA National Institute of Food and Agriculture.

References

Abed, A., P. Pérez-Rodríguez, J. Crossa, and F. Belzile. 2018. When less can be better: How can we make genomic selection more cost-effective and accurate in barley? Theor. Appl. Genet. 131(9): 1873–1890. doi: 10.1007/s00122-018-3120-8.

Adeyemo, E., and R. Bernardo. 2019. Predicting genetic variance from genomewide marker effects estimated from a diverse panel of maize inbreds. Crop Sci. 0(0): 0. doi: 10.2135/cropsci2018.08.0525.

Albrecht, T., H.-J. Auinger, V. Wimmer, J.O. Ogutu, C. Knaak, et al. 2014. Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. Theor. Appl. Genet. 127(6): 1375–1386. doi: 10.1007/s00122-014-2305-z.

Albrecht, T., V. Wimmer, H.-J. Auinger, M. Erbe, C. Knaak, et al. 2011. Genome-based prediction of testcross values in maize. Theor. Appl. Genet. 123(2): 339–350. doi: 10.1007/s00122-011-1587-7.

Anderson, J.A., M.E. Sorrells, and S.D. Tanksley. 1993. RFLP analysis of genomic regions associated with resistance to preharvest sprouting in wheat. Crop Sci. 33(3): 453. doi: 10.2135/cropsci1993.0011183X003300030008x.

Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, N.A. Tinker, et al. 2013. Genomic, marker-assisted, and pedigree-BLUP selection methods for β-glucan concentration in elite oat. Crop Sci. 53(5): 1894. doi: 10.2135/cropsci2012.09.0526.

Bastiaansen, J.W., A. Coster, M.P. Calus, J.A. van Arendonk, and H. Bovenhuis. 2012. Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures. Genet. Sel. Evol. 44(1): 3. doi: 10.1186/1297-9686-44-3.

Bates, D., M. Maechler, B. Bolker, and S. Walker 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Bates, D., A.I. Vazquez, M. Ana, and I. Vazquez. 2015. pedigreemm: Pedigree-based mixed-effects models. https://CRAN.R-project.org/package=pedigreemm (accessed 1 Dec. 2017).

Bayer, M.M., P. Rapazote-Flores, M. Ganal, P.E. Hedley, M. Macaulay, et al. 2017. Development and evaluation of a barley 50k iSelect SNP array. Front. Plant Sci. 8: 1792. doi: 10.3389/fpls.2017.01792.

Bernardo, R. 2014. Genomewide selection when major genes are known. Crop Sci. 54(1): 68. doi: 10.2135/cropsci2013.05.0315.

Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47(3): 1082. doi: 10.2135/cropsci2006.11.0690.

Beyene, Y., K. Semagn, S. Mugo, A. Tarekegne, R. Babu, et al. 2015. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci. 55(1): 154. doi: 10.2135/cropsci2014.07.0460.

Bian, Y., and J.B. Holland. 2017. Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. Heredity. 118(6): 585–593. doi: 10.1038/hdy.2017.4.

Blachez, A.F. 2017. Assessment of malting barley varieties for resistance to important barley diseases in New York [Cornell University Library]. https://cornell.worldcat.org/title/assessment-of-malting-barley-varieties-for-resistance-to-important-barley-diseases-in-new-york/oclc/1066118130&referer=brief_results (accessed 25 February 2019).

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, et al. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinforma. Appl. 23: 2633–2635. doi: 10.1093/bioinformatics/btm308.

Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, et al. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity. 112(1): 48–60. doi: 10.1038/hdy.2013.16.

Cullis, B.R., A.B. Smith, and N.E. Coombes. 2006. On the design of early generation variety trials with correlated data. doi: 10.1198/108571106X154443.

Daetwyler, H.D., U.K. Bansal, H.S. Bariana, M.J. Hayden, and B.J. Hayes. 2014. Genomic prediction for rust resistance in diverse wheat landraces. Theor. Appl. Genet. 127(8): 1795–1803. doi: 10.1007/s00122-014-2341-8.

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome J. 4(3): 250. doi: 10.3835/plantgenome2011.08.0024.

Endelman, J.B., G.N. Atlin, Y. Beyene, K. Semagn, X. Zhang, et al. 2014. Optimal design of preliminary yield trials with genome-wide markers. Crop Sci. 54(1): 48. doi: 10.2135/cropsci2013.03.0154.

Fetch, T.G., and B.J. Steffenson. 1999. Rating scales for assessing infection responses of barley infected with Cochliobolus sativus. Plant Dis. 83(3): 213–217. doi: 10.1094/PDIS.1999.83.3.213.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2009. ASReml User Guide. Hemel Hempstead, HP1 1ES, UK.

Goddard, M.E., and B.J. Hayes. 2007. Genomic selection. J. Anim. Breed. Genet. 124(6): 323–330. doi: 10.1111/j.1439-0388.2007.00702.x.

Guo, Z., D.M. Tucker, C.J. Basten, H. Gandhi, E. Ersoz, et al. 2014. The impact of population structure on genomic prediction in stratified populations. Theor. Appl. Genet. 127(3): 749–762. doi: 10.1007/s00122-013-2255-x.

Guo, Z., D.M. Tucker, J. Lu, V. Kishore, and G. Gay. 2012. Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor. Appl. Genet. 124(2): 261–275. doi: 10.1007/s00122-011-1702-9.

Hamblin, M.T., T.J. Close, P.R. Bhat, S. Chao, J.G. Kling, et al. 2010. Population structure and linkage disequilibrium in u.s. barley germplasm: implications for association mapping. Crop Sci. 50(2): 556. doi: 10.2135/cropsci2009.04.0198.

He, S., A.W. Schulthess, V. Mirdita, Y. Zhao, V. Korzun, et al. 2016. Genomic selection in a commercial winter wheat population. Theor. Appl. Genet. 129(3): 641–651. doi: 10.1007/s00122-015-2655-1.

Heffner, E.L., A.J. Lorenz, J.-L. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: Gain per unit time and cost. Crop Sci. 50(5): 1681. doi: 10.2135/cropsci2009.11.0662.

Heffner, E.L., M.E. Sorrells, and J. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49: 1–12. doi: 10.2135/cropsci2008.08.0512.

Hori, K., K. Sato, and K. Takeda. 2007. Detection of seed dormancy QTL in multiple mapping populations derived from crosses involving novel barley germplasm. Theor. Appl. Genet. 115(6): 869–876. doi: 10.1007/s00122-007-0620-3.

Hori, K., K. Sato, and K. Takeda. 2007. Detection of seed dormancy QTL in multiple mapping populations derived from crosses involving novel barley germplasm. Theor Appl Genet 115: 869–876. doi: 10.1007/s00122-007-0620-3.

Iwata, H., and J.-L. Jannink. 2011. Accuracy of genomic selection prediction in barley breeding programs: a simulation study based on the real single nucleotide polymorphism data of barley breeding lines. Crop Sci. 51(5): 1915. doi: 10.2135/cropsci2010.12.0732.

Jacobson, A., L. Lian, S. Zhong, and R. Bernardo. 2014. General combining ability model for genomewide selection in a biparental cross. Crop Sci. 54(3): 895. doi: 10.2135/cropsci2013.11.0774.

Juskiw, P.E., J.H. Helm, M. Oro, J.M. Nyachiro, and D.F. Salmon. 2009. Registration of “Bentley” barley. J. Plant Regist. 3(2): 119. doi: 10.3198/jpr2008.10.0631crc.

Kraakman, A.T.W., R.E. Niks, P.M.M.M. Van den Berg, P. Stam, and F.A. Van Eeuwijk. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168(1): 435–46. doi: 10.1534/genetics.104.026831.

Lado, B., S. Battenfield, C. Guzmán, M. Quincke, R.P. Singh, et al. 2017. Strategies for selecting crosses using genomic prediction in two wheat breeding programs. Plant Genome 10(2): 0. doi: 10.3835/plantgenome2016.12.0128.

Legge, W.G., S. Haber, D.E. Harder, J.G. Menzies, J.S. Noll, et al. 2008. Newdale barley. Can. J. Plant Sci. 88(4): 717–723. doi: 10.4141/CJPS07194.

Legge, W.G., J.R. Tucker, T.G. Fetch, S. Haber, J.G. Menzies, et al. 2014. AAC Synergy barley. Can. J. Plant Sci. 94(4): 797–803. doi: 10.4141/cjps2013-307.

Lehermeier, C., N. Krämer, E. Bauer, C. Bauland, C. Camisan, et al. 2014. Usefulness of multiparental populations of maize (Zea mays L.) for genome-based prediction. doi: 10.1534/genetics.114.161943.

Lehermeier, C., S. Teyssèdre, and C.-C. Schön. 2017. Genetic gain increases by applying the usefulness criterion with improved variance prediction in selection of crosses. Genetics 207(4): 1651–1661. doi: 10.1534/genetics.117.300403.

Li, C.D., A. Tarr, R.C.M. Lance, S. Harasymow, J. Uhlmann, et al. 2003. A major QTL controlling seed dormancy and pre-harvest sprouting a-amylase in two-rowed barley. Aust. J. Agric. Res. 54(12): 1303. doi: 10.1071/AR02210.

Lorenz, A.J., K.P. Smith, and J.-L. Jannink. 2012. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. Crop Sci. 52(4): 1609. doi: 10.2135/cropsci2011.09.0503.

Lorenz, A.J. 2013. Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. G3. 3(3): 481–91. doi: 10.1534/g3.112.004911.

Lorenz, A.J., and K.P. Smith. 2015. Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci. 55(6): 2657. doi: 10.2135/cropsci2014.12.0827.

Lorenzana, R.E., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120(1): 151–161. doi: 10.1007/s00122-009-1166-3.

Martin, J.M., T.K. Blake, and E.A. Hockett. 1991. Diversity among North American spring barley cultivars based on coefficients of parentage. Crop Sci. 31(5): 1131. doi: 10.2135/cropsci1991.0011183X003100050009x.

Marulanda, J.J., A.E. Melchinger, and T. Würschum. 2015. Genomic selection in biparental populations: assessment of parameters for optimum estimation set design. Plant Breed. 134(6): 623–630. doi: 10.1111/pbr.12317.

Mascher, M., H. Gundlach, A. Himmelbach, S. Beier, S.O. Twardziok, et al. 2017. A chromosome conformation capture ordered sequence of the barley genome. Nature 544(7651): 427–433. doi: 10.1038/nature22043.

Maurer, A., V. Draba, Y. Jiang, F. Schnaithmann, R. Sharma, et al. 2015. Modelling the genetic architecture of flowering time control in barley through nested association mapping. BMC Genomics 16(1): 290. doi: 10.1186/s12864-015-1459-7.

McElroy, M.S., A.J.R. Navarro, G. Mustiga, C. Stack, S. Gezan, et al. 2018. Prediction of cacao (Theobroma cacao) resistance to Moniliophthora spp. diseases via genome-wide association analysis and genomic selection. Front. Plant Sci. 9: 343. doi: 10.3389/fpls.2018.00343.

Meuwissen, T.H. 1997. Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75(4): 934. doi: 10.2527/1997.754934x.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157(4): 1819-1829.

Mohammadi, M., T. Tiede, and K.P. Smith. 2015. PopVar: A genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations. Crop Sci. 55(5): 2068. doi: 10.2135/cropsci2015.01.0030.

Moore, J.K., H.K. Manmathan, V.A. Anderson, J.A. Poland, C.F. Morris, et al. 2017. Improving genomic prediction for pre-harvest sprouting tolerance in wheat by weighting large-effect quantitative trait loci. Crop Sci. 57(3): 1315. doi: 10.2135/cropsci2016.06.0453.

Muñoz-Amatriaín, M., A. Cuesta-Marcos, J.B. Endelman, J. Comadran, J.M. Bonman, et al. 2014. The USDA barley core collection: Genetic diversity, population structure, and potential for genome-wide association studies. PLoS One 9(4): e94688. doi: 10.1371/journal.pone.0094688.

Nakamura, S., M. Pourkheirandish, H. Morishige, Y. Kubo, M. Nakamura, et al. 2016. Mitogen-activated protein kinase kinase 3 regulates seed dormancy in barley. Curr. Biol. 26(6): 775–781. doi: 10.1016/J.CUB.2016.01.024.

New York State Brewer's Association. 2018. Farm Brewery. https://newyorkcraftbeer.com/farm-brewery/ (accessed 1 May 2017).

Neyhart, J., D. Sweeney, M. Sorrells, C. Kapp, K. D. Kephart, J. Sherman, E. J. Stockinger, S. Fisk, P. Hayes, S. Daba, M. Mohammadi, N. Hughes,

L. Lukens, P.G. Barrios, L. Gutiérrez, and K. P. Smith. 2019. Registration of the S2MET barley mapping population for multi-environment genomewide selection. J. Plant Reg. doi:10.3198/jpr2018.06.0037crmp.

Neyhart, J., and K. P. Smith. 2019. Validating genomewide predictions of genetic variance in a contemporary breeding program. Crop. Sci. doi: 10.2135/cropsci2018.11.0716.

Nielsen, N.H., A. Jahoor, J.D. Jensen, J. Orabi, F. Cericola, et al. 2016. Genomic prediction of seed quality traits using advanced barley breeding lines. PLoS One 11(10): e0164494. doi: 10.1371/journal.pone.0164494.

Osthushenrich, T., M. Frisch, C. Zenke-Philippi, H. Jaiser, M. Spiller, et al. 2018. Prediction of means and variances of crosses with genome-wide marker effects in barley. Front. Plant Sci. 9: 1899. doi: 10.3389/fpls.2018.01899.

Peiffer, J.A., M.C. Romay, M.A. Gore, S.A. Flint-Garcia, Z. Zhang, et al. The genetic architecture of maize height. http://www.genetics.org/content/genetics/196/4/1337.full.pdf (accessed 25 January 2018).

Peterson, R.F., A. Campbelll, and A.E. Hannah. 1948. A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. Cereal Div. Exp. Farms Serv. 147. http://www.nrcresearchpress.com/doi/pdf/10.1139/cjr48c-033 (accessed 12 February 2018).

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7(2): e32253. doi: 10.1371/journal.pone.0032253.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, et al. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome J. 5(3): 103. doi: 10.3835/plantgenome2012.06.0006.

Prada, D., S.E. Ullrich, J.L. Molina-Cano, L. Cistue, J.A. Clancy, et al. 2004. Genetic control of dormancy in a Triumph/Morex cross in barley. Theor. Appl. Genet. 109(1): 62–70. doi: 10.1007/s00122-004-1608-x.

Rice, B. and A. Lipka. 2018. Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. Plant Genome. doi: 10.3835/plantgenome2018.07.0052

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, et al. 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat. Genet. 44(2): 217–220. doi: 10.1038/ng.1033.

Riedelsheimer, C., J.B. Endelman, M. Stange, M.E. Sorrells, J.-L. Jannink, et al. 2013. Genomic predictability of interconnected biparental maize populations. Genetics 194(2): 493–503. doi: 10.1534/genetics.113.150227.

Riedelsheimer, C., and A.E. Melchinger. 2013. Optimizing the allocation of resources for genomic selection in one breeding cycle. Theor. Appl. Genet. 126(11): 2835–2848. doi: 10.1007/s00122-013-2175-9.

Rincent, R., D. Laloe, S. Nicolas, T. Altmann, D. Brunel, et al. 2012. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: Comparison of methods in two diverse groups of maize inbreds (Zea mays L.). Genetics 192(2): 715–728. doi: 10.1534/genetics.112.141473.

Rutkoski, J.E., E.L. Heffner, and M.E. Sorrells. 2011. Genomic selection for durable stem rust resistance in wheat. Euphytica 179(1): 161–173. doi: 10.1007/s10681-010-0301-1.

Rutkoski, J.E., J. Poland, J-L. Jannink, M.E. Sorrells. 2013. Imputation of unordered markers and the impact on genomic selection accuracy. G3: Genes, Genomes, Genetics. 3:427-439.

Rutkoski, J.E., J.A. Poland, R.P. Singh, J. Huerta-Espino, S. Bhavani, et al. 2014. Genomic selection for quantitative adult plant stem rust resistance in wheat. Plant Genome 7(3): 0. doi: 10.3835/plantgenome2014.02.0006.

Rutkoski, J., J. Benson, Y. Jia, G. Brown-Guedira, J.-L. Jannink, et al. 2012. Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. Plant Genome J. 5(2): 51. doi: 10.3835/plantgenome2012.02.0001.

Sallam, A.H., J.B. Endelman, J.-L. Jannink, and K.P. Smith. 2015. Assessing genomic selection prediction accuracy in a dynamic barley breeding population. Plant Genome 8(1): 0. doi: 10.3835/plantgenome2014.05.0020.

Sallam, A.H., and K.P. Smith. 2016. Genomic selection performs similarly to phenotypic selection in barley. Crop Sci. 56(6): 2871. doi: 10.2135/cropsci2015.09.0557.

Schmidt, M., S. Kollers, A. Maasberg-Prelle, J. Großer, B. Schinkel, et al. 2016. Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection. Theor. Appl. Genet. 129(2): 203–213. doi: 10.1007/s00122-015-2639-1.

Schulz-Streeck, T., J.O. Ogutu, Z. Karaman, C. Knaak, and H.P. Piepho. 2012. Genomic selection using multiple populations. Crop Sci. 52(6): 2453. doi: 10.2135/cropsci2012.03.0160.

Spindel, J.E., H. Begum, D. Akdemir, B. Collard, E. Redoña, et al. 2016. Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. Heredity (Edinb). 116(4): 395–408. doi: 10.1038/hdy.2015.113.

Tiede, T., and K.P. Smith. 2018. Evaluation and retrospective optimization of genomic selection for yield and disease resistance in spring barley. Mol. Breed. 38(5): 55. doi: 10.1007/s11032-018-0820-3.

Ullrich, S.E., H. Lee, J.A. Clancy, I.A. del Blanco, V.A. Jitkov, et al. 2009. Genetic relationships between preharvest sprouting and dormancy in barley. Euphytica 168(3): 331–345. doi: 10.1007/s10681-009-9936-1.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91(11): 4414–4423. doi: 10.3168/jds.2007-0980.

Vazquez, A.I., D.M. Bates, G.J.M. Rosa, D. Gianola, and K.A. Weigel. 2010. Technical note: An R package for fitting generalized linear mixed models in animal breeding. J. Anim. Sci. 88(2): 497–504. doi: 10.2527/jas.2009-1952.

Wang, Y., M. Mette, T. Miedaner, M. Gottwald, P. Wilde, et al. 2014. The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the

accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. BMC Genomics 15(1): 556. doi: 10.1186/1471-2164-15-556.

Watson, A., S. Ghosh, M.J. Williams, W.S. Cuddy, J. Simmonds, et al. 2018. Speed breeding is a powerful tool to accelerate crop research and breeding. Nat. Plants 4(1): 23–29. doi: 10.1038/s41477-017-0083-8.

Windhausen, V.S., G.N. Atlin, J.M. Hickey, J. Crossa, J.-L. Jannink, et al. 2012. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. doi: 10.1534/g3.112.003699.

Würschum, T., J.C. Reif, T. Kraft, G. Janssen, and Y. Zhao. 2013. Genomic selection in sugar beet breeding populations. BMC Genet. 14(1): 85. doi: 10.1186/1471-2156-14-85.

Xu, Y., Q. Jia, G. Zhou, X.-Q. Zhang, T. Angessa, et al. 2017. Characterization of the sdw1 semi-dwarf gene in barley. BMC Plant Biol. 17(1): 11. doi: 10.1186/s12870-016-0964-4.

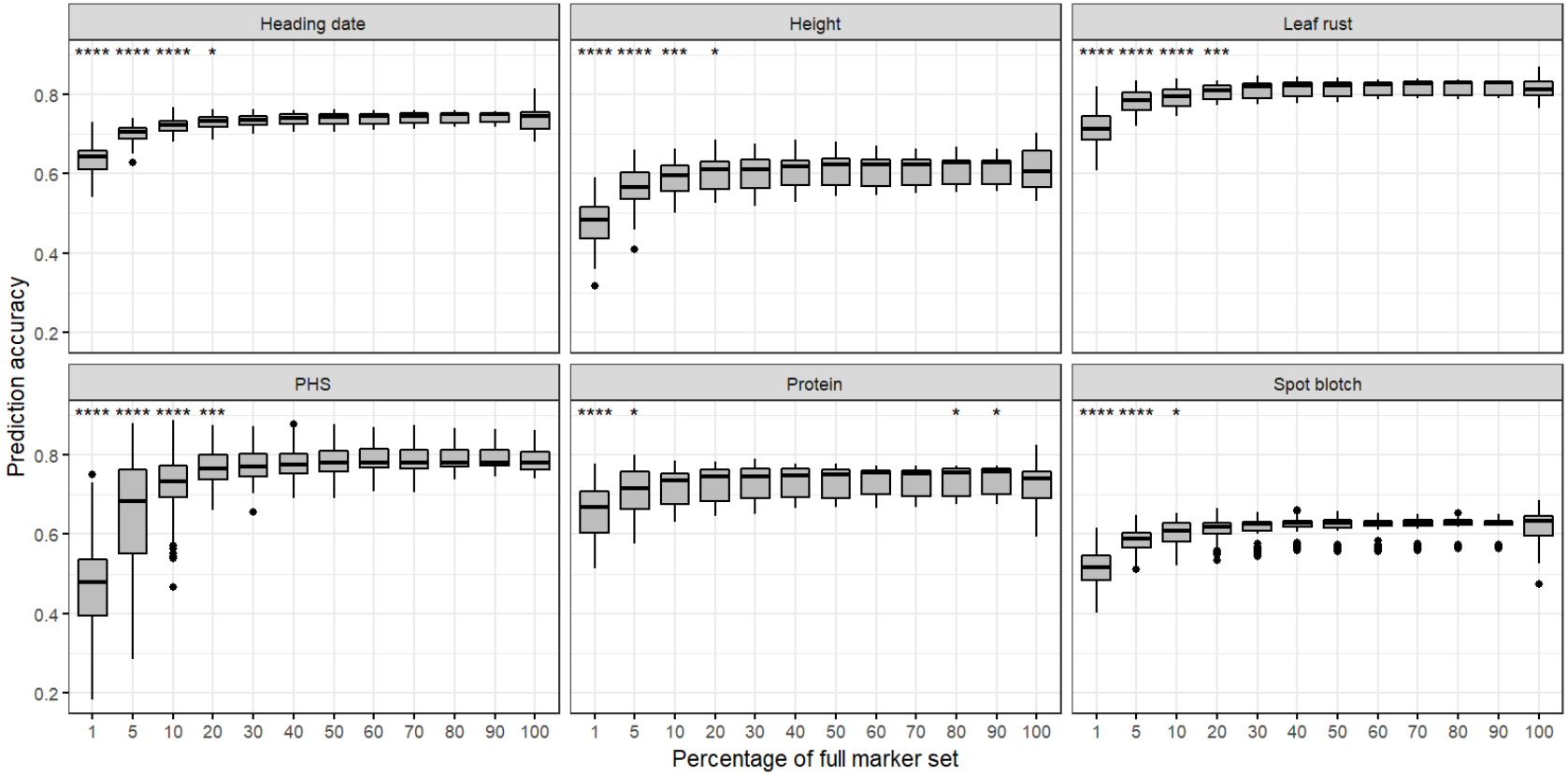
Yu, J., J.B. Holland, M.D. McMullen, and E.S. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178(1): 539–551. doi: 10.1534/GENETICS.107.074245.

Zhao, Y., M. Gowda, W. Liu, T. Würschum, H.P. Maurer, et al. 2012. Accuracy of genomic selection in European maize elite breeding populations. Theor. Appl. Genet. 124(4): 769–776. doi: 10.1007/s00122-011-1745-y.

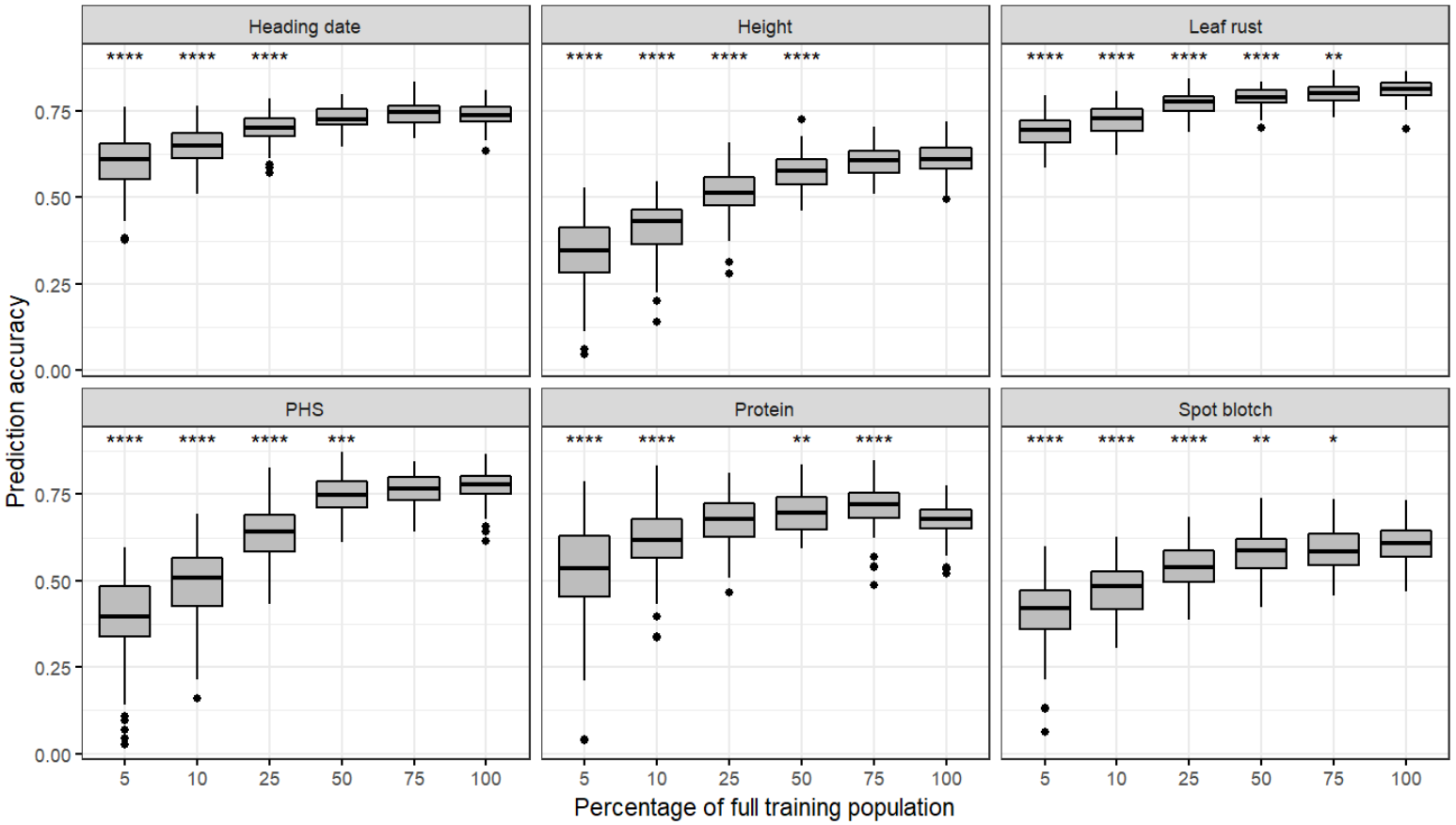
Zheng, X., D. Levine, J. Shen, S.M. Gogarten, C. Laurie, et al. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28(24): 3326–3328. doi: 10.1093/bioinformatics/bts606.

Zhou, H., and B. Steffenson. 2013. Genome-wide association mapping reveals genetic architecture of durable spot blotch resistance in US barley breeding germplasm. Mol Breed. 32: 139–54. https://link.springer.com/content/pdf/10.1007/s11032-013-9858-4.pdf (accessed 25 February 2018).

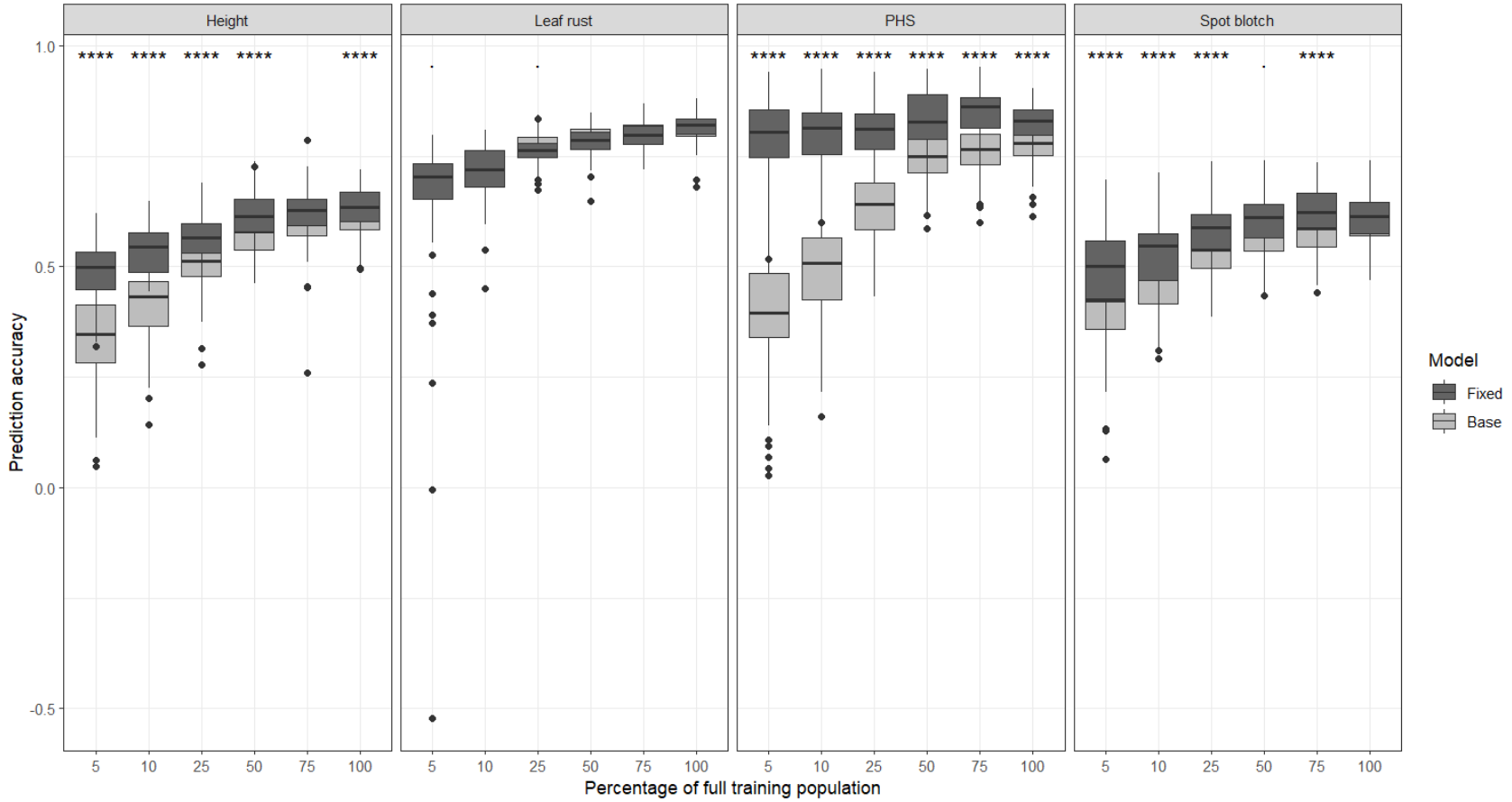
Supplementary material



Supplemental Figure 2.1: Mean CV_{A2} prediction accuracies at reduced marker density. At each marker set size, 25 marker sets were evenly sampled across chromosomes. Each marker set was used for a single five-fold cross validation. Boxplots with stars indicate significant difference in prediction accuracy from 100% of the initial marker set. Significance levels indicated by *(p<.1), ***(p<.01), ****(p<.001).



Supplemental Figure 2.2: Mean CV_{A3} prediction accuracies for training population subsets across traits. At each subset size, 100 training population samples were taken across families and validation populations of fixed size were sampled from remaining lines. Prediction accuracy with each training/validation population set was evaluated once. Boxplots with stars indicate significant difference in prediction accuracy from the full training population (100). Significance levels indicated by *(p<.1), **(p<.05), ***(p<.01), ****(p<.001).



Supplemental Figure 2.3: Effect of fixed marker models at reduced training population size in CV_{A3}. Sampling and cross validation methods are the same as in Supplemental Figure 4. Asterisks indicate significant difference between base model and fixed marker model at each training population size. Significance levels indicated by **(p<.05), ****(p<.001).

Supplemental Table 2.1: Identity-by-state relationships between founder lines calculated according to Rincent et al. (2012)

	AAC Synergy	Bentley	Conlon	Craft	ND Genesis	Newdale	Pinnacle	KWS Tinka
AAC Synergy	1							
Bentley	0.825	1						
Conlon	0.740	0.742	1					
Craft	0.767	0.761	0.734	1				
ND Genesis	0.673	0.706	0.713	0.667	1			
Newdale	0.948	0.801	0.729	0.763	0.665	1		
Pinnacle	0.670	0.740	0.720	0.737	0.672	0.663	1	
KWS Tinka	0.423	0.383	0.317	0.417	0.373	0.405	0.310	1

Supplemental Table 2.2: Genetic variance within family (σ^2_{gw}), across family (σ^2_{ga}), and percent variance explained by within family ($\sigma^2_{gw}/(\sigma^2_{gw} + \sigma^2_{ga})$).

Variance component	Heading date	Height	Leaf rust	Spot blotch	PHS	Protein
σ^2_{gw}	4.072	6.071	0.390	0.167	2.104	0.278
σ^2_{ga}	1.897	2.259	0.975	0.336	0.807	0.389
% variance within family	68.2	72.9	28.6	33.2	72.3	41.7

Supplemental Table 2.3: Power for genome-wide association studies at reduced training population size. Lines were randomly sampled 100 times at each training population size.

Trait	Training population size (% of full training population)	Training population samples with significant marker-trait association	Training population samples with same significant marker-trait association as full set
Height	25	54	54
	50	95	94
	75	100	100
Leaf rust	25	80	66
	50	100	99
	75	100	100
Preharvest sprouting	25	100	100
	50	100	100
	75	100	100
Spot blotch	25	24	14
	50	54	39
	75	72	65

CHAPTER 3

ASSOCIATION MAPPING FOR PREHARVEST SPROUTING, SEED DORMANCY, AND GERMINATION RATE IN SPRING MALTING BARLEY

Abstract

Preharvest sprouting (PHS) can severely damage barley malting quality and is of particular concern in locations with a high frequency of precipitation around harvest. Malting quality and PHS are often negatively correlated and the *SD2* locus on chromosome 5H has been associated with both traits. Malting quality phenotyping is expensive, but germination rate has been correlated with some malting quality traits. Using three spring barley populations with different structure, we mapped PHS, seed dormancy, and germination rate over multiple time points to identify changes in genetic control of these traits with after-ripening. *HvAlaAT1* at the *SD1* locus was associated with long-term dormancy and reduced germination rate. Two marker trait associations at *SD2*, likely identifying three alleles of *HvMKK3*, were repeatedly associated with primary dormancy and increased germination rate. Sixty-five lines were Sanger sequenced for *HvGA20ox1* and single base pair polymorphisms in the 5' UTR and intron were identified in non-dormant lines but *HvGA20ox1* had no effect on PHS, germination energy, or germination rate. The allelic state of *HvMKK3* determined genetic PHS susceptibility in North American spring two-row barley germplasm and interactions between *HvAlaAT1* and *HvMKK3* were associated with changes in seed dormancy and germination rate over time. KASP markers were developed for the causal mutations in *HvAlaAT* and *HvMKK3* and a diagnostic

mutation in *HvGA20ox1*. Haplotypes with PHS resistance, short primary dormancy, and a high germination rate were identified that could be useful for breeding for PHS resistance and malting quality.

Introduction

Seed dormancy is the inability of a viable seed to germinate under favorable conditions (DePauw and McCaig, 1991; Bewley et al., 2013). Dormancy is induced during seed development and dormancy maintenance is influenced by genetic and environmental factors such as temperature, water content, oxygen availability, and light (Gubler et al., 2005). The length of time required for a seed to lose dormancy is the after-ripening period. Preharvest sprouting occurs when wet or humid conditions induce starch degradation and germination in non-dormant seeds before harvest. In cereal crops, symptoms can be invisible to the naked eye (pre-germination) or visible in the form of radicle and coleoptile emergence (PHS). Storage time can be significantly reduced in pre-germinated grain and severely preharvest sprouted grain is often unusable for milling or malting.

Barley (*Hordeum vulgare* L.) is grown worldwide for food and animal feed, but malting quality grain used for brewing and distilling has the highest market value. Rapid, uniform germination of barley grain during malting enables more complete modification of the endosperm and reduces production time in the malthouse. Prolonged dormancy in malting barley is undesirable as it can result in slow, uneven germination and increased malthouse storage costs. Germination rate has been positively correlated to malting quality traits related to endosperm modification such

as malt extract, alpha-amylase, and soluble/total protein ratio (Woonton et al., 2005). Malting quality is expensive to phenotype and requires a large amount of grain, making alternative phenotyping methods like germination rate an attractive option for screening breeding material at higher volumes and at lower cost. Many modern North American malting barley varieties have been selected for high amylase activity for use in adjunct brewing. Selection for this malting quality profile has indirectly increased PHS susceptibility and reduced seed dormancy in many North American malting barley varieties. Some level of seed dormancy is required for PHS resistance but unique quantitative trait loci (QTL) for each trait have been described in the same population (Ullrich et al., 1999). In areas where PHS is a threat, PHS resistance and a short dormancy period are desired to protect and enable good malting quality.

Induction and release of seed dormancy in barley is a complex process. Abscisic acid (ABA) and gibberellin (GA) are the two major plant hormone regulators of seed dormancy and germination. ABA promotes seed dormancy induction and maintenance while GA promotes seed germination and dormancy release. Dormancy is not determined by absolute levels of ABA or GA, but rather the ratio of the two hormones and embryo sensitivity to them. GA activates alpha-amylase production in the aleurone which in turn initiates endosperm starch hydrolysis during germination in barley (Bewley et al., 2013). Reduction in ABA content and differences in ABA content between varieties are not related to loss of dormancy in barley. Rather, ABA content decreases upon imbibition in after-ripened non-dormant seeds while ABA content briefly decreases then increases to maintain dormancy in dormant seeds (Jacobsen et al., 2002). This reduction in ABA in after-ripened seed is the result of

increased expression of an ABA 8'-hydroxylase gene, *HvCYP707A* (Millar et al., 2006) but other changes in ABA sensitivity and signaling can also result in loss of dormancy. Environmental factors, particularly temperature, can influence barley seed dormancy with high temperatures during grainfill generally decreasing dormancy and increasing PHS risk (Rodriguez et al., 2001; Li et al., 2003; Gualano & Benech-Arnold, 2009).

Genetic sources of PHS resistance and seed dormancy have been well described in barley but several key questions remain unanswered. Two large effect QTL on chromosome 5H, *SD1* and *SD2*, have been consistently mapped for PHS and dormancy in a wide range of barley germplasm (Oberthur et al., 1995; Han et al., 1996; Hori et al., 2007; Lin et al., 2009). A gene has been cloned at each locus: at *SD1*, an alanine aminotransferase (*HvAlaAT1*, HORVU.MOREX.r2.5HG0398940) (Sato et al., 2016) and at *SD2*, a mitogen activated protein kinase kinase (*HvMKK3*, HORVU.MOREX.r2.5HG0447180) (Nakamura et al., 2016). Vetch et al. (2020) identified four *HvMKK3* alleles in North American malting barley germplasm and associated the E165Q mutation with PHS. The *SD2* locus has also been identified in multiple malting quality QTL mapping studies for malt extract, alpha-amylase, soluble/total protein ratio, and beta-glucan (Castro et al., 2010; Mohammadi et al., 2015; Zhou et al., 2016), but PHS resistance is typically negatively correlated with malting quality (Gao et al., 2003; Li et al., 2003; Castro et al., 2010). Several instances of double QTL peaks in the *SD2* region have been described (Li et al., 2003; Gong et al., 2014) but the genetic nature of the negative correlation between PHS and malting quality at *SD2* is still unclear, with some studies suggesting a tightly linked gene

cluster (Gong et al., 2014) and others pleiotropy (Li et al., 2003; Li et al., 2004). Edney et al. (2013) described several Baudin/TR253 progeny with a dormant *SD2* allele, PHS resistance, and good malting quality as well as one line with a non-dormant *SD2* allele, PHS resistance, and good malting quality, suggesting a tightly linked gene cluster. Hickey et al. (2012) identified a novel QTL *qSDND*, subsequently named *Qsd3*, ~2 cM proximal to *SD2* in a Flagship/ND24260 biparental population and confirmed that the *SD2* region was monomorphic in the population, but the exact nature of *Qsd3* is unclear. Gong et al. (2014) also identified a large effect seed dormancy QTL proximal to *SD2* in a Stirling/Harrington doubled haploid population that was significant across multiple years and locations.

A GA 20-oxidase (*HvGA20ox1*, HORVU.MOREX.r2.5HG0446540) (Spielmeyer et al., 2004), is 1.7 Mb proximal to *HvMKK3* in the *SD2* region and has been a suggested cause of PHS (Li et al., 2004; Nagel et al., 2018) and malting quality variation (Zhou et al, 2016). Barley has four GA 20-oxidase genes that are part of the active GA biosynthesis pathway (Spielmeyer et al., 2004). *HvGA20ox1* is expressed in the embryo and scutellum during germination as soon as 24 hours after imbibition begins and is active in the early stages of *de novo* GA synthesis converting inactive GA_{53} to active GA_1 in the scutellum (Betts et al., 2019). A closely related GA 20-oxidase in rice, *OsGA20ox2*, has been directly implicated in PHS resistance (Ye et al., 2015) and GA20ox transcripts increase during imbibition of non-dormant sorghum (*Sorghum bicolor* L.) compared to dormant sorghum lines (Perez-Flores et al., 2003). Despite an understanding of the physiological role of GA20ox1 in seed germination and evidence to support its role in PHS and seed germination in other monocot crops,

variants of *HvGA20ox1* have not been well described or definitively associated with PHS in barley. Detection of variants at both *HvGA20ox1* and *HvMKK3* associated with PHS or seed dormancy would support the hypothesis that QTL controlling PHS and malting quality at *SD2* are tightly linked, in which case selection for specific *SD2* haplotypes conferring PHS resistance and good malting quality may be possible.

Genome-wide association (GWA) has become a standard method for mapping complex traits in a range of plant population types. Improved marker density, computationally efficient models that account for population structure and kinship, and higher statistical power enable more precise mapping of QTL with current GWA methodologies (Tibbs Cortes et al., 2021). One such model is the multi-locus mixed model (MLMM), which includes large effect markers as cofactors in a forward-backward stepwise approach to increase power in structured populations and is particularly powerful for traits controlled by several medium to large effect loci (Segura et al., 2012). This approach also enables the detection of variants in tight repulsion linkage (Li et al., 2015). Association mapping studies for PHS and seed dormancy are limited in barley. Sweeney et al. (2020) found the *SD2* region to be associated with PHS in a connected half-sib population derived from eight commercially relevant, but closely related varieties, and Nagel et al. (2019) identified a number of QTL for PHS and seed dormancy in a diverse panel of landrace barley accessions. Populations composed of modern malting breeding germplasm as well as more diverse ancestral lines could strike a balance between adequate sampling of allelic diversity and relevance to current breeding programs. Recent advances in barley genotyping density with the 50k SNP chip (Bayer et al., 2017), an improved

reference genome assembly (Monat et al., 2019), and a twenty-accession pan-genome (Jayakodi et al., 2020) offer new tools for fine-scale mapping of commercially important loci like *SD2* in barley.

The objectives of this research were to 1) identify QTL for PHS, seed dormancy, and after-ripened germination rate in spring malting barley with GWA, 2) characterize and develop diagnostic high-throughput markers for *HvAlaAT1, HvGA20ox1* and *HvMKK3* variants, and 3) assess interactions between *HvGA20ox1, HvAlaAT1,* and *HvMKK3* for PHS and germination traits over an after-ripening period of three months. A detailed understanding of the changes in interactions between loci controlling genetic sources of PHS, seed dormancy, and germination rate over time may enable breeders to select for PHS resistance, rapid release of primary dormancy shortly after harvest, and high malting quality. Phenotyping PHS and seed dormancy is labor intensive and requires precise sampling, making diagnostic markers and improved phenotyping methods an attractive option to improve breeding efficiency.

Materials and methods

Plant materials

Three sets of barley germplasm were used for PHS and seed dormancy phenotyping. The first set, CU, included 424 spring two-row malting barley breeding lines consisting of three generations of a genomic selection experiment (Cycle 0/C0, Cycle 1/C1, and Cycle 2/C2; Chapter 5) and the eight founder lines: ‘Conlon’ (PI 597789), ‘Pinnacle’ (PI 643354), ‘ND Genesis’ (PI 677345), ‘Craft’ (PI 646158), ‘Newdale’ (Legge et al., 2008), ‘AAC Synergy’ (Legge et al., 2014), ‘Bentley’ (Juskiw et al.,

2009), and ‘KWS Tinka’ (PI 681721). CU contains half-sib lines from the initial Sweeney et al. (2020) study as well as recombined progeny that were selected for PHS. A detailed description of this population can be found in Chapter 5. For PHS phenotyping all 424 lines were sampled in 2019 and a subset was sampled in 2020. For germination traits the founder lines, C0, and C1 were sampled (n=308) in 2019 and all lines were sampled in 2020. The panel was grown in Ithaca, NY at two locations each in 2019 and 2020. Trials were planted in an augmented design in single one-meter rows in 2019 and twin rows in 2020 with checks arranged in randomized complete block design with two blocks per location.

The second set, JIC, included 86 predominantly European two-and six-row spring landrace varieties. This panel was initially assembled by the John Innes Centre (JIC, Norwich, UK) for disease resistance screening. JIC was planted in two replicates of single one-meter rows at one location in Ithaca, NY in 2019 and was phenotyped for PHS. In 2020, JIC along with an additional 20 spring hulless, or naked, barley lines and the six-row non-dormant ‘Morex’ (CI 15773; Rasmusson and Wilcoxson, 1979) and dormant ‘Steptoe’ (CI 15229; Muir and Nilan, 1973) checks were added to CU in the twin one-meter row augmented design in two locations in Ithaca, NY and phenotyped for PHS and germination traits. Data for JIC were included in all summary statistics and haplotype analysis. The frequency of six-row types was 5.7% (30/523 lines) in the combined CU/JIC population and to avoid potential spurious associations from population structure due to the minor allele frequency (MAF) cutoff of 0.05, six-row lines were removed for GWA.

The third set, S2MET, was composed of 223 two-row North American spring malting barley breeding lines that were part of a collaborative multi-environment trial in 2015 and 2016. The full description of this population can be found in Neyhart et al. (2019). This trial was planted in an augmented complete block design in one 2015 location and two 2016 locations in Ithaca, NY. S2MET was phenotyped for PHS in each location. Phenotypic and genotypic data for S2MET is publicly available in the T3 database (Blake et al., 2015; https://triticeaetoolbox.org/barley/).

Phenotyping

Heading date was recorded as the Julian date of 50% full spike emergence from the flag leaf within plot and maturity date was recorded as the Julian date of 50% spike physiological maturity (PM), defined as the loss of green color from the spike. For PHS measurement, five spikes were harvested at PM, after-ripened for three days at ambient room temperature and misted in a greenhouse for three days. PHS severity was assessed on a zero to nine scale according to Anderson et al. (1993) with zero indicating no PHS and nine indicating rootlet and coleoptile emergence along the entire spike. At two days post-PM, 12 to 25 spikes were collected for seed dormancy assays, dried for two days in a forced air dryer at 35 C, hand threshed, and frozen at -20 C. The number of phenotypic observations was not consistent across timepoints in 2019 and 2020. Some 2019 CU and 2020 JIC material had poor field germination and seed was limited or not available for all lines. If the seed sample was only large enough for two timepoints, the plot was phenotyped at 6 and 111 days post-PM.

Germination assays were based on the American Society of Brewing Chemists method for simultaneous determination of germination energy in barley (ASBC, 1997). Germination traits were evaluated at three time points for CU in 2019: 6 days post-PM (PM_6), 48 days post-PM (PM_{48}), and 111 days post-PM (PM_{111}). Germination traits were evaluated at six time points for CU and JIC in 2020: PM_6, 20 days post-PM (PM_{20}), 34 days post-PM (PM_{34}), PM_{48}, 69 days post-PM (PM_{69}), and PM_{111}. Samples were randomized into four batches for each time point in 2019 and a single batch in 2020. All samples were removed from the freezer one day before PM_6 to thaw. 90 mm petri dishes were filled with two Whatman No. 1 filter papers and wetted with 4 mL room temperature tap water. Thirty kernels were placed in each dish and germinated kernels, defined as radicle emergence of > 2 mm, were counted and removed after 24, 48, and 72 hours of steeping. Plates were placed in Styrofoam or hard-sided coolers at room temperature during the 72-hour period to maintain high humidity and darkness. The total number of viable kernels was determined by adding 2 mL 0.75% H_2O_2 to each plate with ungerminated kernels at 72 hours and counting germinated kernels after an additional 48 hours. In 2020, the same procedure was followed but the addition of H_2O_2 at 72 hours was omitted and germination without H_2O_2 was recorded at 96 and 120 hours instead. Tests were replicated for every time point except for 2019 PM_6.

Germination energy (GE3) is a measure of seed dormancy and was calculated as $GE3 = n_{24} + n_{48} + n_{72} / n_{total}$ where n_{24} is the number of kernels germinated at 24 hours, n_{48} is the number of kernels germinated at 48 hours, n_{72} is the number of kernels germinated at 72 hours, and n_{total} is the overall number of kernels. Germination index

(GI3) is a measure of germination rate and was calculated as GI3 = 10*(Σ(n_{24} + n_{48} + n_{72})/ (n_{24} + 2n_{48} + 3n_{72})) (Frančáková et al., 2012). Germination index does not account for dormancy, as a dormant sample with a single germinated kernel may have a germination index of 5.00 if the single kernel germinates the second day, so GI3 was scaled by multiplying by GE3 before data processing. In 2020, five-day measurements of GE (GE5) and GI (GI5) were also calculated by adding n_{96} and n_{120} terms to the equations.

Genotyping

CU and JIC were genotyped with the 50k Illumina SNP array (Bayer et al., 2017) at the USDA Small Grains Genotyping Laboratory in Fargo, ND. Marker positions were based on the Morex v2 genome assembly (Monat et al., 2019). After filtering for heterozygosity over 10% and minor allele frequency less than 0.01, 32,185 polymorphic markers remained. The *SNPrelate* R package (Zheng et al., 2012) was used to LD prune the marker set using a 25,000 bp sliding window and an LD threshold of 0.95, resulting in 21,429 markers that were used for principal component (PC) analysis with *SNPrelate*. For GWA analysis in *GAPIT* (Wang and Zhang, 2020), an additional MAF filter of 0.05 was imposed and all six-row JIC lines (5.7% of the total CU/JIC lines) were removed to avoid spurious associations, resulting in 14,015 markers for GWA. S2MET was genotyped with genotyping-by-sequencing (GBS). Full details on genotyping and quality control methods can be found in Neyhart et al. (2019). GBS reads were aligned to the Morex v1 reference genome (Mascher et al., 2017). GWA in S2MET used 11,039 markers.

Physical positions and sequence for *HvAlaAT1*, *HvGA20ox1*, and *HvMKK3* were initially referenced using the BARLEX database (https://apex.ipk-gatersleben.de/apex/f?p=284:10::::::). Sequences of *HvAlaAT1*, *HvGA20ox1*, and *HvMKK3* were BLASTed to the barley pan-genome (Jayakodi et al., 2020) assemblies of cv. Morex, Hockett, RGT Planet, Barke, Golden Promise, Igri, and Akashinriki using GrainGenes (https://wheat.pw.usda.gov/GG3/). Homoeologous copies of *TaGA20ox1* sequence on chromosomes 4AL, 5BL, and 5DL were downloaded from the 10+ Wheat Genome project (Walkowiak et al., 2020) for comparison.

Kompetitive Allele Specific Primers (KASP) markers were developed for polymorphisms located within each of the target genes (Supplementary Table 3.6). Primers were developed to amplify the causal L214F polymorphism in *HvAlaAT1* described by Sato et al. (2016), which was confirmed to be present in North American germplasm by Vetch et al. (2020). This marker was named AlaAT_L214F. Primer development for *HvGA20ox1* targeted a SNP at 331 bp in the 5' UTR of *HvGA20ox1* that was present in PHS susceptible lines (detailed in the Results section) and was named GA20ox1_331_5UTR. Two markers for *HvMKK3* were developed. The first targeted the E165Q mutation in exon 3 identified by Vetch et al. (2020) and was named MKK3_E165Q. The second, JHI-367342-KASP, was converted from a significant 50k marker trait association (MTA) in 2019 CU GWA (Table 3.3), JHI-Hv50k-2016-367342, which is highly correlated (r=0.966) with the R350G *HvMKK3* mutation from Vetch et al. (2020) in JIC and CU parent lines (data not shown). JHI-Hv50k-2016-367342 is 152 bp proximal to *HvMKK3.* KASP polymerase chain reactions (PCR) were performed in 384 well plates. Forward primer candidates were

developed using OligoCalc (http://biotools.nubic.northwestern.edu/OligoCalc.html) and reverse primer candidates were developed using NCBI Primer BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primers were diluted to 100 uM and primer master mix was composed of 12 uL allele 1 specific primer, 12 uL allele 2 specific primer, 30 uL common reverse primer, and 46 uL dd H_2O. Master mix was prepared with 315 uL PCR Allele Competitive Extension (PACE) master mix (3CR Bioscience, Essex, UK) and 8.7 uL primer mix for each 384 well plate. 2.5 uL master mix and 2.5 uL DNA were added to each well for PCR. The following amplification protocol was used: initial denaturation at 94 C for 15 minutes, 10 cycles each of 94 C for 20 seconds followed by a stepdown to 65 C by 0.8 C per minute, and 36 cycles each of 94 C for 20 seconds followed by 57 C for 1 minute. Fluorescence was measured on a ViiA 7 or Quantstudio 7 real time PCR system (Applied Biosystems, ThermoFisher Scientific). The complete CU and JIC sets were genotyped for the four KASP markers and the three new markers (AlaAT_L214F, GA20ox1_331_5UTR, and MKK3_E165Q) were added to the 50k genotype matrix. 91/223 S2MET lines were genotyped with the four KASP markers. AlaAT_L214F was not highly correlated with any GBS markers and was not added to the GBS marker set. Correlations between KASP and *SD2* region GBS markers were high, with correlations of 0.86, 0.81, and 0.71 between GA20ox1_331_5UTR/S5_665601626, JHI-367342-KASP/ S5_669551252, and MKK3_E165Q/ S5_667068417, respectively. KASP genotyped lines were used to impute heterozygous GBS calls in S2MET lines that were not KASP genotyped.

Sanger sequencing for *HvAlaAT1*, *HvGA20ox1*, and *HvMKK3* was conducted by the USDA Cereal Crops Research Unit (CCRU). JIC lines were sequenced for *HvAlaAT1* and *HvMKK3* and CU founder lines were sequenced for *HvGA20ox1* and *HvMKK3*. An additional set of spring and winter malting and naked varieties were sequenced for all three genes. All amplification reactions were performed in either polymerase chain reaction (PCR) plates or individual PCR tubes. A 25-ul reaction per sample was prepared according to manufacturer's instructions (Takara Ex Taq; Takara). All products were separated in a 1% agarose gel and stained with SyBr Safe DNA Stain (Thermo Fisher Scientific) and visualized on a gel documentation system. *HvAlaAT1* amplification reactions were performed using the following PCR cycle: initial denaturation for 2 min at 98°C, followed by 30 cycles each with 15 sec at 98°C, 30 sec at 59°C and 2 min at 72°C, with a final extension of 5 min at 72°C. *HvGA20ox1* amplicons were generated using the following PCR cycle: initial denaturation for 2 min at 98°C, followed by 30 cycles each with 15 sec at 98°C, 30 sec at 61°C and 45 sec at 72°C, with a final extension of 5 min at 72°C. For *HvMKK3* a series of PCRs consisting of a standard PCR followed by a nested PCR were used to obtain the desired amplicon as described by Nakamura et al., 2016. See Supplementary Table 3.7 for amplification reaction primers. Diluted amplicons of each gene were then used to sequence exonic regions of interest using primers detailed in Supplementary Table 3.8. Amplicons were labeled with BigDye (Thermo Fisher Scientific) according to manufacturer's instructions and submitted to the University of Wisconsin Biotechnology Center for DNA cleanup and direct amplicon sequencing. Sequencing reads were manually edited and trimmed for quality and assembled

using DNAstar Sequence annotation and alignment software (DNASTAR, 2014). Sequence variation within each gene and among lines was assessed using the SeqMan Pro Sequence alignment software (DNASTAR).

Statistical analysis

All statistical analyses were performed with R version 3.5.1 (R Core Team, 2018). CU and JIC datasets were combined for analysis and S2MET was analyzed separately. Linear mixed models were fit for each trait in each dataset using the R package *lme4* (Bates et al., 2015). Environments were defined as year/location combinations and were considered fixed effects for all models. Genetic effects (*g*) for lines were specified as fixed to extract best linear unbiased estimates (BLUEs) for GWA and as random with $g \sim (N(0, \sigma_g^2))$ where σ_g^2 is the genetic variance for heritability estimation. For CU and JIC germination trait models, fixed effects also included germination assay replication and for PHS models a random effect for PHS sampling date nested within environment (*s*) was added with $s \sim (N(0, \sigma_s^2))$ where σ_s^2 is the sampling date variance. For S2MET, *s* was also added for PHS sampling date nested within environment. Heritability estimates for augmented designs in CU, JIC, and S2MET were calculated according to Cullis et al. (2006). Fixed effect interaction models were fit to estimate additive by additive epistatic interactions between markers (M) associated with *HvAlaAT1*, *HvGA20ox1*, and *HvMKK3* as

$$y \sim \mu + M_{AlaAT1} + M_{GA20ox1} + M_{MKK3} + M_{AlaAT1}: M_{GA20ox1} + M_{AlaAT1}: M_{MKK3} + M_{GA20ox1}: M_{MKK3} + e \quad (3.1)$$

where M_{MKK3} had three levels defined by combination of MKK3_E165Q and JHI-367342-KASP alleles and *e* was the random error with $e \sim N(0, \sigma_e^2)$ where σ_e^2 is the residual error variance. Tukey-Kramer tests were performed to determine differences between unbalanced haplotype means using the *HSD.test*() function in the R package *agricolae* (de Mendiburu and Yaseen, 2020). Bivariate models were fit in *ASReml-R* version 3 (Gilmour et al., 2009) for CU 2019/2020 to determine genetic correlations (r_g) at PM_6, PM_{47}, and PM_{111} between PHS, GE3, and GI3 using a *corgh()* variance structure for entry genetic effects. Phenotypic correlations (r_p) were estimated from line means across environments.

Multi-locus mixed models were fit for each trait at each timepoint separately in GAPIT using entry BLUEs. Two principal components were added in each model to account for population structure and a MAF filter of 0.05 was applied. Analyses for CU/JIC and S2MET were conducted separately. For the CU/JIC datasets, models were fit for the 2019 dataset, the 2020 dataset, and the combined 2019/2020 dataset. A Bonferroni threshold of 0.05 was used to determine marker-trait association (MTA) significance.

Pleiotropic effects

Gibberellins are involved in a number of traits related to plant growth and development (Spielmeyer et al., 2004), notably reduced height and increased yield in cereals (Peng et al., 1999; Oikawa et al., 2004; Jia et al., 2011). Pleiotropic effects of marker GA20ox1_331_5UTR were assessed for height (cm from ground to spike) and heading date in CU and height, heading date (days after planting), and yield (kg/ha) in

S2MET. Entry means for height, heading date, and yield for the S2MET were downloaded from the T3 barley database (Blake et al., 2015) from 17 sites in 2016, 5 sites in 2015, and 1 site in 2014. The 2016 S2MET data included 231 lines and the 2014/2015 S2MET included 186 lines. Pleotropic effects of *HvGA20ox1* were tested using ANOVA with fixed effects for environment, the first two PCs to control for population structure, and GA20ox1_331_5UTR.

Results

Preharvest sprouting was highly heritable (H^2= 0.75-0.96) with similar ranges of resistant and susceptible lines observed across years in each dataset (Table 3.1). Average GE3 increased at each early time point and plateaued at PM_{33}- PM_{48} within CU and JIC, indicating most lines lost primary dormancy after one to one and a half months of after-ripening. Within CU and JIC in 2020, several lines with under 90% germination were still observed at PM_{111}. Kernels were not surface sterilized before plating due to population size and poor germination at later time points was due to diseased kernels. Heritability for GE3 was initially high (0.93-0.99) and decreased with time as genetic variance for primary dormancy decreased. GI3 heritabilities were high (0.88-0.99) in each dataset. GI3 increased with after-ripening and plateaued at PM_{33}- PM_{48} in CU and JIC. JIC showed a greater GI3 range than CU, with scores greater than 9.0 in several naked barley lines, indicating rapid germination on the first day of the assay. Within CU and JIC, Steptoe had the lowest GE3 and GI3 and did not exceed 80% germination until PM_{111}. GE5 and GI5 data were very similar to GI3 and GE3 data at each timepoint and will not be discussed in detail.

Table 3.1: Phenotypic means, genetic variances (σ_g^2), and heritability (H^2) estimates for all datasets.

Dataset	Year	Trait	Mean	Range	σ_g^2	H^2	Time point
CU	2019	PHS	2.7	0.0-8.2	2.519	0.91	PM_6
	2020	PHS	3.76	0.0-8.8	4.96	0.96	PM_6
	2019	GE3	0.66	0.0-1.0	0.06	0.93	PM_6
	2019	GE3	0.98	0.35-1.0	0.002	0.93	PM_{48}
	2019	GE3	0.99	0.93-1.0	1.9E-6	0.62	PM_{111}
	2019	GI3	3.10	0.13-6.9	2.58	0.96	PM_6
	2019	GI3	5.11	1.25-8.1	0.563	0.96	PM_{48}
	2019	GI3	5.78	4.03-8.75	0.197	0.90	PM_{111}
	2020	GE3	0.64	0.0-1.0	0.084	0.99	PM_6
	2020	GE3	0.86	0.06-1.0	0.032	0.97	PM_{20}
	2020	GE3	0.95	0.27-1.0	0.006	0.94	PM_{34}
	2020	GE3	0.95	0.23-1.0	0.005	0.94	PM_{48}
	2020	GE3	0.99	0.63-1.0	0.0001	0.73	PM_{69}
	2020	GE3	0.98	0.8-1.0	8.58E-5	0.68	PM_{111}
	2020	GI3	4.43	0.0-7.44	3.35	0.99	PM_6
	2020	GI3	4.99	3.33-7.63	1.99	0.98	PM_{20}
	2020	GI3	5.45	3.33-8.33	0.86	0.97	PM_{34}
	2020	GI3	5.75	3.51-8.82	0.94	0.96	PM_{48}
	2020	GI3	5.76	3.73-8.11	0.19	0.90	PM_{69}
	2020	GI3	5.82	4.33-8.33	0.13	0.88	PM_{111}
S2MET	2015	PHS	2.57	0.0-8.4	5.1	0.89	PM_6
	2016	PHS	5.04	0.0-9.0	2.45	0.87	PM_6
JIC	2019	PHS	1.52	0.0-7.2	1.39	0.75	PM_6
	2020	PHS	2.17	0.0-8.8	3.08	0.87	PM_6
	2020	GE3	0.73	0.0-1.0	0.07	0.97	PM_6
	2020	GE3	0.90	0.0-1.0	0.04	0.98	PM_{20}
	2020	GE3	0.94	0.08-1.0	0.02	0.96	PM_{34}
	2020	GE3	0.95	0.23-1.0	0.01	0.96	PM_{48}
	2020	GE3	0.98	0.74-1.0	9E-5	0.90	PM_{69}
	2020	GE3	0.98	0.83-1.0	2E-4	0.76	PM_{111}
	2020	GI3	3.55	0.0-8.61	3.39	0.98	PM_6
	2020	GI3	5.15	0.0-10.0	2.24	0.98	PM_{20}
	2020	GI3	5.83	0.32-10.0	1.75	0.97	PM_{34}
	2020	GI3	6.13	0.95-10.0	1.74	0.96	PM_{48}
	2020	GI3	6.17	3.16-10.0	0.57	0.94	PM_{69}
	2020	GI3	6.15	4.18-9.68	0.37	0.90	PM_{111}

Genetic correlations between PHS, GE3, and GI3 at PM_6 in CU were high, indicating shared genetic control of these traits (Table 3.2). PHS and GE3 r_g at PM_{48} was lower, suggesting a different genetic cause of short- and long-term dormancy in

this population. r_g between PHS and GI3 decreased over time. PHS had negligible r_g with height and flowering time (data not shown). GI3 at PM_{48} and PM_{111} were highly correlated (r_g=0.95) indicating phenotyping germination rate after a month and a half in two-row spring malting barley may not be necessary using the 30-kernel assay. The phenotypic correlation between the number of kernels germinated after 24 hours and scaled GI3 at 72 hours was calculated at PM_6, PM_{48}, and PM_{111} for the 2019 and 2020 datasets to assess potential to improving phenotyping efficiency. Correlations between 24-hour germination and scaled GI were high at PM_6 (r_{2019} = 0.79, r_{2020} = 0.78) but increased with after-ripening (PM_{45}: r_{2019} = 0.86, r_{2020} = 0.91; PM_{111}: r_{2019} = 0.96, r_{2020} = 0.95).

Table 3.2: Trait correlations in CU 2019/2020. Genetic correlations are shown in the lower triangle and phenotypic correlations in the upper triangle.

	PHS	GE3, PM_6	GE3, PM_{48}	GE3, PM_{111}	GI3, PM_6	GI3, PM_{48}	GI3, PM_{111}
PHS		0.77	0.32	0.05	0.83	0.61	0.55
GE3, PM_6	0.81		0.43	-0.01	0.97	0.75	0.65
GE3, PM_{48}	0.35	0.52		0.20	0.39	0.72	0.55
GE3, PM_{111}	0.05	0.05	0.26		-0.04	-0.04	0.02
GI3, PM_6	0.89	0.98	0.45	-0.01		0.76	0.67
GI3, PM_{48}	0.63	0.85	0.82	0.11	0.82		0.84
GI3, PM_{111}	0.62	0.81	0.67	0.04	0.80	0.95	

Population structure

CU and JIC separated along the first two principal components of the genotype matrix (Figure 3.1). A small cluster of CU lines appeared between the upper left quadrant of the plot and the right-hand side where the majority of the CU lines tightly clustered. These are progeny of Cycle 1 CU lines and DH130910, a facultative two-row barley

included in JIC with European two-row winter and spring six-row background that clusters in the upper left-hand quadrant with other spring six-row material. One CU founder line with European two-row background, KWS Tinka, clustered with a large portion of JIC lines in the lower middle of the plot, showing relatedness between historic and modern European two-row spring malting barley. Several naked barley lines developed by Canadian breeding programs with malting barley parentage cluster with CU. The first four principal components explained 13.9, 5.1, 3.9, and 2.8% of the genotypic variation. Within the full 32,185 marker set for CU/JIC, LD decayed below $r^2 = 0.2$ within 1.34 Mb and below $r^2 = 0.1$ in 6.2 Mb on average. Within the 11,039 marker set for S2MET, LD decayed below $r^2 = 0.2$ within 0.9 Mb and below $r^2 = 0.1$ in 4.1 Mb on average. The first four principal components of the S2MET genotypes explained 10.4, 4.5, 3.7, and 3.3% of the genotypic variation.

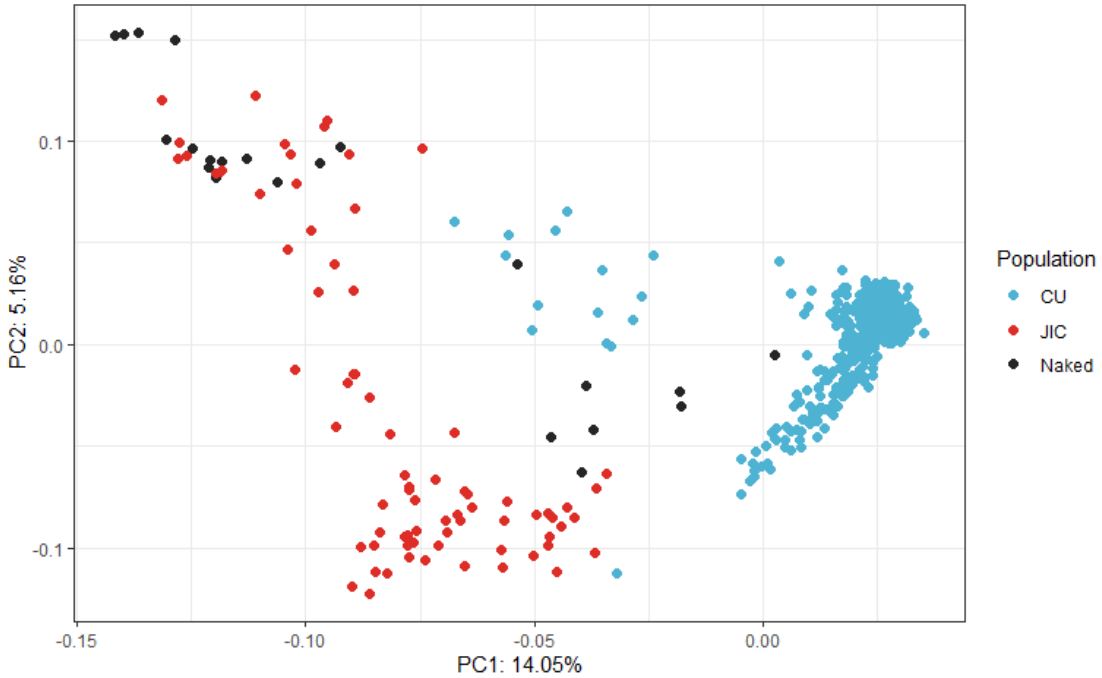


Figure 3.1: Principal component plot with all CU/JIC lines using 21,429 50k markers.

Genome-wide association

Using a Bonferroni cutoff of 0.05, one PHS MTA was detected for marker MKK3_E165Q on chromosome 5H in the combined 2019/2020 CU/JIC dataset with six-row lines omitted (Table 3.3). Three PHS MTAs were detected in S2MET, two of which were in the *SD2* region of 5H. The third MTA, S2_861435, was on chromosome 2H. The two MTA in the S2MET *SD2* region, S5_667068417 and S5_669160186, had MAF of 0.428 and 0.269, respectively, and LD (r^2) of 0.174. S5_667068417 was highly correlated and thus imputed with MKK3_E165Q KASP data and S5_669160186 was highly correlated (0.819) with S5_669551252, the marker imputed with JHI-367342-KASP. The non-dormant S5_667068417 allele was only observed with the non-dormant S5_669160186 allele, similar to the *HvMKK3* E165Q mutation only occurring with the R350G mutation (Vetch et al., 2020). S5_667068417 and S5_669160186 in combination identified three distinct alleles of *HvMKK3* in S2MET.

For germination traits in CU/JIC, GWA models were fit for 2019 and 2020 separately at all phenotyped time points within year and for the combined 2019/2020 dataset for PM_6, PM_{48}, and PM_{111}. Across these three datasets, 14 unique markers were detected in 43 MTAs across 21 timepoint/trait combinations, an average of 2.1 MTAs per timepoint/trait combination. All of the MTAs detected in 2020 and the combined 2019/2020 dataset were on chromosome 5H. Single MTAs on chromosomes 2H, and 4H were detected in 2019 CU for GI3 at PM_{111}, and GI3 at PM_{48}, respectively. These loci were designated as QSD1_1H, QSD1_2H, and QSD1_4H. Markers AlaAT_L214F, MKK3_E165Q, and JHI-367342-KASP were observed in 10,

11, and 11 unique timepoint/trait combinations, respectively. Two MTAs in the region of marker AlaAT_L214F, JHI-Hv50k-2016-308584 and JHI-Hv50k-2016-309388, were in LD of 0.83 and 0.60 $r^2$with AlaAT_L214F, respectively. This locus was designated QSD1_5H and identified *HvAlaAT1*. MTAs were detected for markers JHI-Hv50k-2016-366603, JHI-Hv50k-2016-366688, JHI-Hv50k-2016-367738, JHI-Hv50k-2016-366325, SCRI_RS_10702, and SCRI_RS_193456, which had intramarker LD (r^2) between 0.73 and 0.96 in a LD block surrounding JHI-367342-KASP. This locus was designated QSD2_5H and identified *HvMKK3.* No additional MTAs in LD with MKK3_E165Q were detected.

Seed dormancy haplotype effects

Seed dormancy haplotypes in CU and JIC were defined as the combination of *HvAlaAT1*, *HvGA20ox1*, and *HvMKK3* alleles based on KASP markers and S2MET haplotypes were defined as the combination of *HvGA20ox1* and *HvMKK3* alleles based on KASP markers. AlaAT_L214F amplified the causal mutation in *HvAlaAT1* so the dormant allele was known. The GA20ox1_331_5UTR allele most commonly present in PHS susceptible lines was labeled as non-dormant. *HvMKK3* alleles were defined based on the combination of MKK3_E165Q and JHI-Hv50k-2016-367342-KASP and lines with non-dormant alleles at both markers were labeled as N* (very non-dormant) for *HvMKK3*. Dormant alleles were labeled “D” and non-dormant alleles “N” for the other markers and haplotypes were ordered by physical position: a DDN haplotype indicates a dormant *HvAlaAT1* allele, dormant *HvGA20ox1*, and non-dormant *HvMKK3*.

Table 3.3: Marker trait associations for CU/JIC in 2019, 2020, and 2019/2020 and S2MET.

Dataset	Year	Time point	Trait	SNP	Chr	Position	MAF	Locus	Gene candidate
CU/JIC	19	PM_{111}	GI3	JHI-Hv50k-2016-58707	2H	512368	0.374	QSD1_2H	
CU/JIC	19	PM_{48}	GI3	JHI-Hv50k-2016-271675	4H	615260232	0.478	QSD1_4H	
CU/JIC	19	PM_{48}	GI3	JHI-Hv50k-2016-308584	5H	442158883	0.3	QSD1_5H	*HvAlaAT1*
CU/JIC	19/20, 20	PM_6, PM_{20}, PM_{34}, PM_{48}	GE3, GI3	AlaAT_L214F	5H	442160000	0.305-0.327	QSD1_5H	*HvAlaAT1*
CU/JIC	19	PM_{111}	GI3	JHI-Hv50k-2016-309388	5H	452787194	0.312	QSD1_5H	*HvAlaAT1*
CU/JIC	19	PM_{111}	GI3	JHI-Hv50k-2016-366380	5H	595767289	0.297	QSD2_5H	*HvMKK3*
CU/JIC	19	PM_{48}	GE3	JHI-Hv50k-2016-366603	5H	596064952	0.359	QSD2_5H	*HvMKK3*
CU/JIC	19/20	PM_{111}	GI3	JHI-Hv50k-2016-366688	5H	596189802	0.341	QSD2_5H	*HvMKK3*
CU/JIC	19	PM_{69}	GI3	SCRI_RS_193456	5H	596540049	0.378	QSD2_5H	*HvMKK3*
CU/JIC	19/20, 20	PM_6, PM_{20}, PM_{34}, PM_{48}	GE3, GI3	JHI-367342-KASP	5H	596729543	0.36-0.38	QSD2_5H	*HvMKK3*
CU/JIC	20	PM_{34}	GE3	SCRI_RS_10702	5H	596742751	0.388	QSD2_5H	*HvMKK3*
CU/JIC	19	PM_{48}	GI3	JHI-Hv50k-2016-367738	5H	597955901	0.357	QSD2_5H	*HvMKK3*
CU/JIC	20	PM_{111}	GI3	JHI-Hv50k-2016-366325	5H	598980001	0.388	QSD2_5H	*HvMKK3*
CU/JIC	19, 19/20, 20	PM_6, PM_{20}, PM_{34}, PM_{48}	GE3, GI3, PHS	MKK3_E165Q	5H	596732030	0.191-0.296	QSD3_5H	*HvMKK3*
S2MET	15/16	PM_6	PHS	S2_861435	2H	861435	0.11	QSD1_2H	
S2MET	15/16	PM_6	PHS	S5_667068417	5H	667068417	0.428	QSD2_5H	*HvMKK3*
S2MET	15/16	PM_6	PHS	S5_669160186	5H	669160186	0.269	QSD2_5H	*HvMKK3*

Six *HvGA20ox1/HvMKK3* haplotypes were identified in S2MET (Figure 3.2; NN*, DN*, NN, DN, ND, DD) and eleven *HvAlaAT1*/*HvGA20ox1*/*HvMKK3* haplotypes (Figure 3.3) were identified in CU/JIC. Seven haplotypes (DNN*, NNN*, DDN*, DDD, NDD, DDN, NDN) were identified in CU and four (DNN, NNN, DND, NND) were identified in JIC. Haplotype frequency was highly unbalanced within each dataset. PHS resistant (<3) and susceptible (>3) lines were identified within every haplotype group except NNN*, DDN*, DND, and NND in CU/JIC and NN*, DN*, and NN in S2MET. PHS was primarily determined by *HvMKK3* allele. Lines with the N* *HvMKK3* allele had moderate to high PHS, low or no primary seed dormancy, and a high germination rate at all timepoints (Figures 3.4 and 3.5). The effect of the dormant *HvAlaAT1* allele was dependent on *HvMKK3* allele. Dormant *HvAlaAT1* decreased PHS and had no effect on GE3 or GI3 with N* *HvMKK3* but had no effect on PHS and decreased initial GE3 and GI3 across all timepoints with N or D *HvMKK3* alleles. Haplotypes with non-dormant *HvAlaAT1* and either N or D *HvMKK3* alleles (NNN, NND, NDN, NDD) generally had PHS resistance, moderate primary seed dormancy that was quickly released, and germination rates similar to DNN* and NNN* with a month and a half of after-ripening. The 2020 dataset enabled a finer scale estimation of after-ripening time needed for dormancy release and increase in germination rate (Figure 3.5). Primary seed dormancy was released by PM_{34} for all haplotypes except haplotypes with dormant *HvAlaAT1* and dormant *HvMKK3* alleles (DND, DDD) which remained dormant up to PM_{48} with a few lines still showing partial dormancy at PM_{69}. NNN, NND, NDN, NDD, and DNN haplotypes showed a clear difference in GI3 compared to dormant *HvAlaAT1* haplotypes at all time points and approached the GI3 of DNN* and NNN* lines as soon as PM_{20}.

SD2 haplotype	S5_665601626	S5_667068417	S5_669551252	n	TukeyHSD	TukeyHSD combined
NN*	A	G	C	94	a	a
DN*	C	G	C	2	ab	
NN	A	C	C	3	bc	b
DN	C	C	C	70	bc	
ND	A	C	T	9	c	c
DD	C	C	T	37	c	

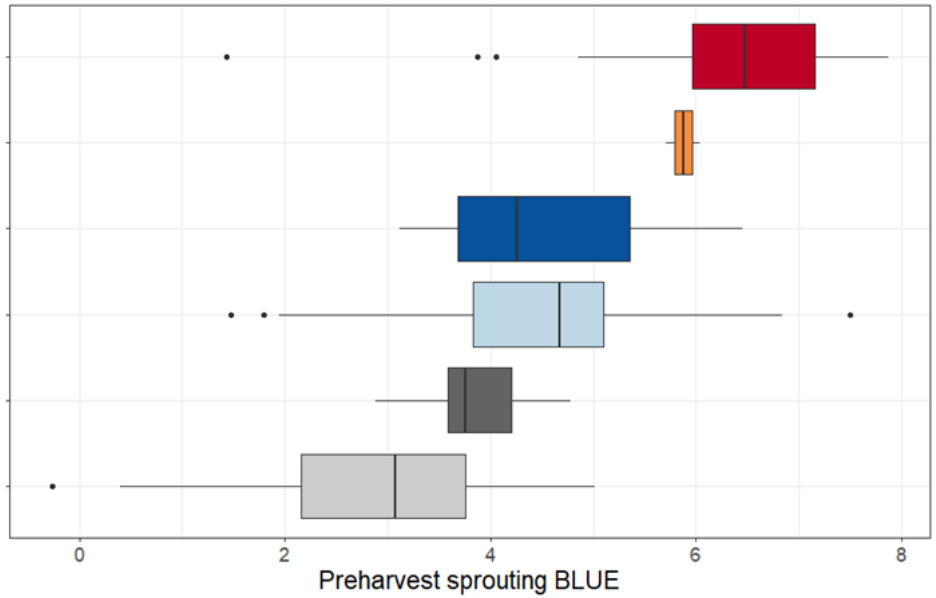


Figure 3.2: S2MET preharvest sprouting BLUEs separated by *SD2* seed dormancy haplotype. Minor alleles are shaded gray. n indicates the number of S2MET lines with each haplotype. Haplotypes that share the same TukeyHSD letter do not have significantly different means ($p < 0.05$). TukeyHSD combined indicates Tukey HSD results when the *HvGA20ox1* allele at S5_665601262 was removed.

Haplotype	AlaAT1_L214F	GA20ox1_331_5UTR	JHI-Hv50k-2016-367342-KASP	MKK3_E165Q	n	TukeyHSD	TukeyHSD combined
NNN*	G	G	T	C	21	a	a
DNN*	C	G	T	C	71	ab	b
DDN*	C	T	T	C	2	b	
NNN	G	G	T	G	9	cd	c
NDN	G	T	T	G	49	c	
DNN	C	G	T	G	7	c	c
DDN	C	T	T	G	152	c	
NND	G	G	C	G	17	cd	c
NDD	G	T	C	G	79	cd	
DND	C	G	C	G	8	d	c
DDD	C	T	C	G	98	c	

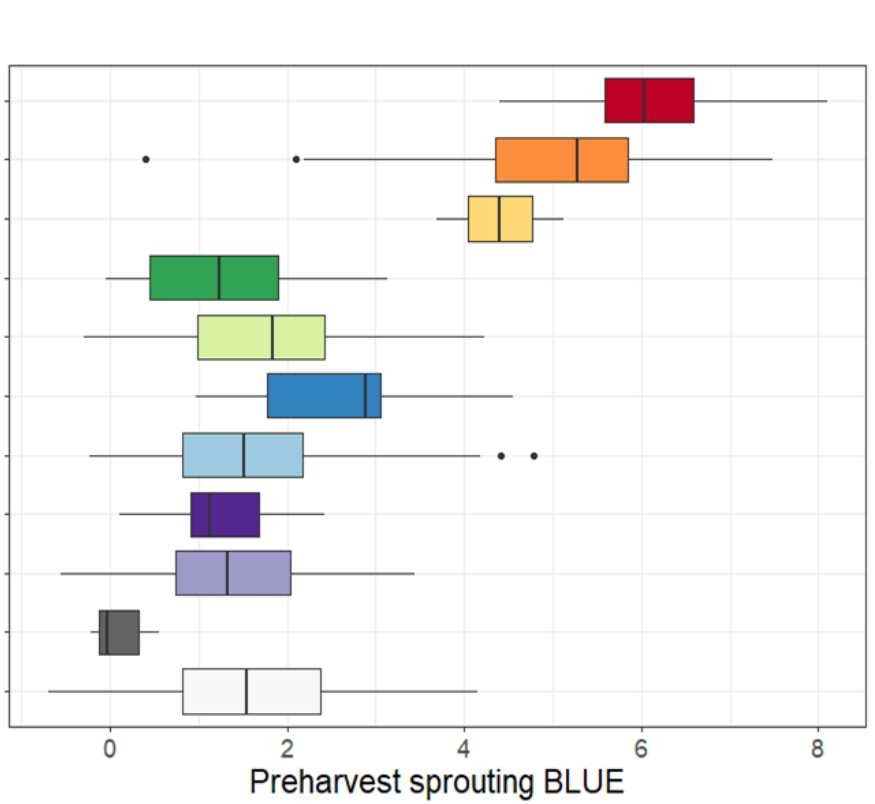


Figure 3.3: CU and JIC 2019/2020 preharvest sprouting BLUEs separated by *SD1* and *SD2* seed dormancy haplotype. Shading indicates the minor allele for each marker. The number of lines in each haplotype is indicated by n. A Tukey-Kramer test was used to determine significant differences ($p < 0.05$) between means of haplotype groups of unequal size. Haplotypes that share the same letter in the TukeyHSD column are not significantly different from each other. TukeyHSD combined indicates Tukey HSD results when GA20ox1_331_5UTR was removed.

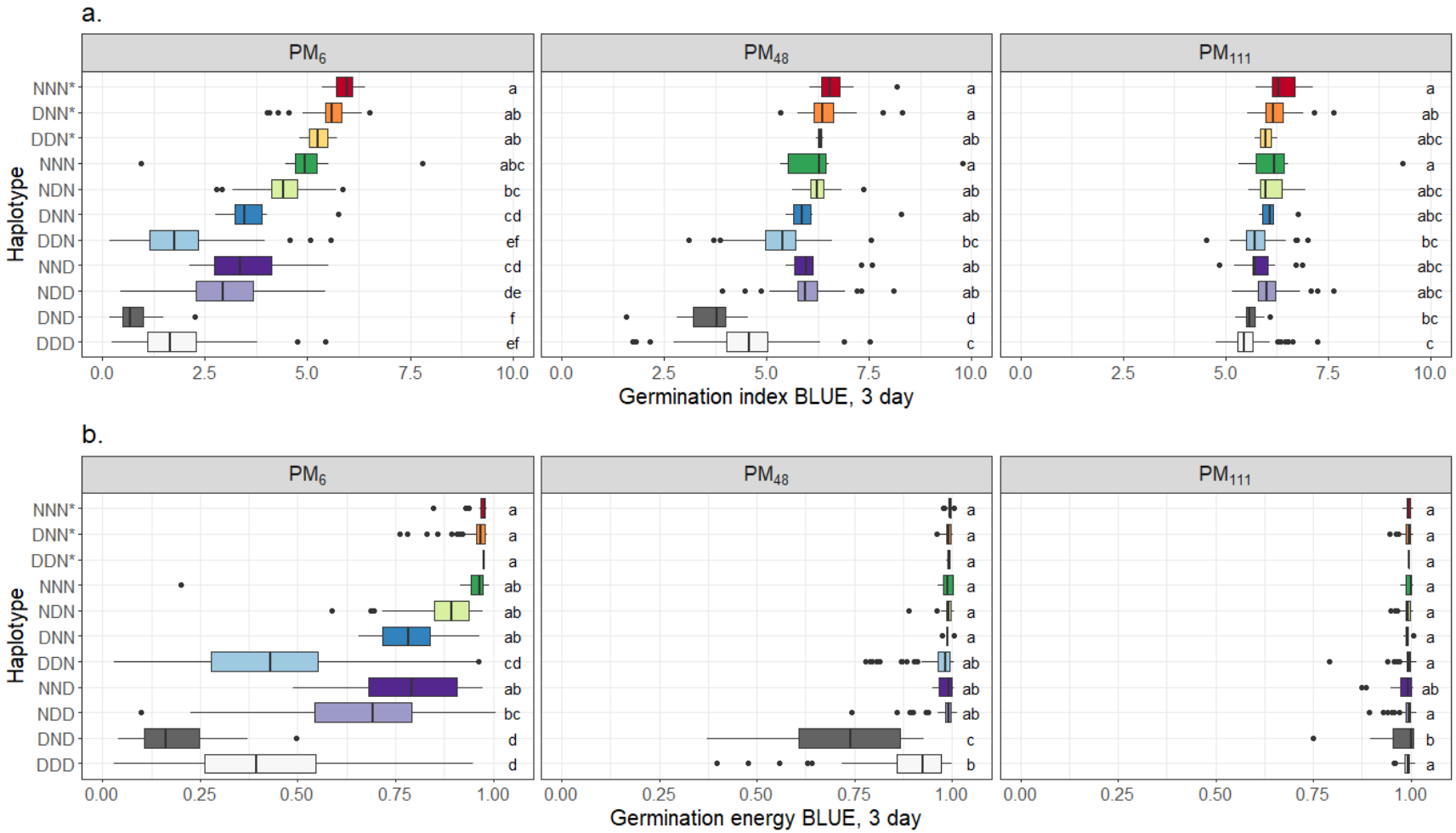


Figure 3.4: CU and JIC 2019/2020 three-day germination index (GI, facet a), and germination energy (GE, facet b) BLUEs at 6, 48, and 111 days after physiological maturity. A Tukey-Kramer test was used to determine significant differences ($p < 0.05$) between means of haplotype groups of unequal size and haplotypes that share the same letter in the TukeyHSD column are not significantly different from each other.

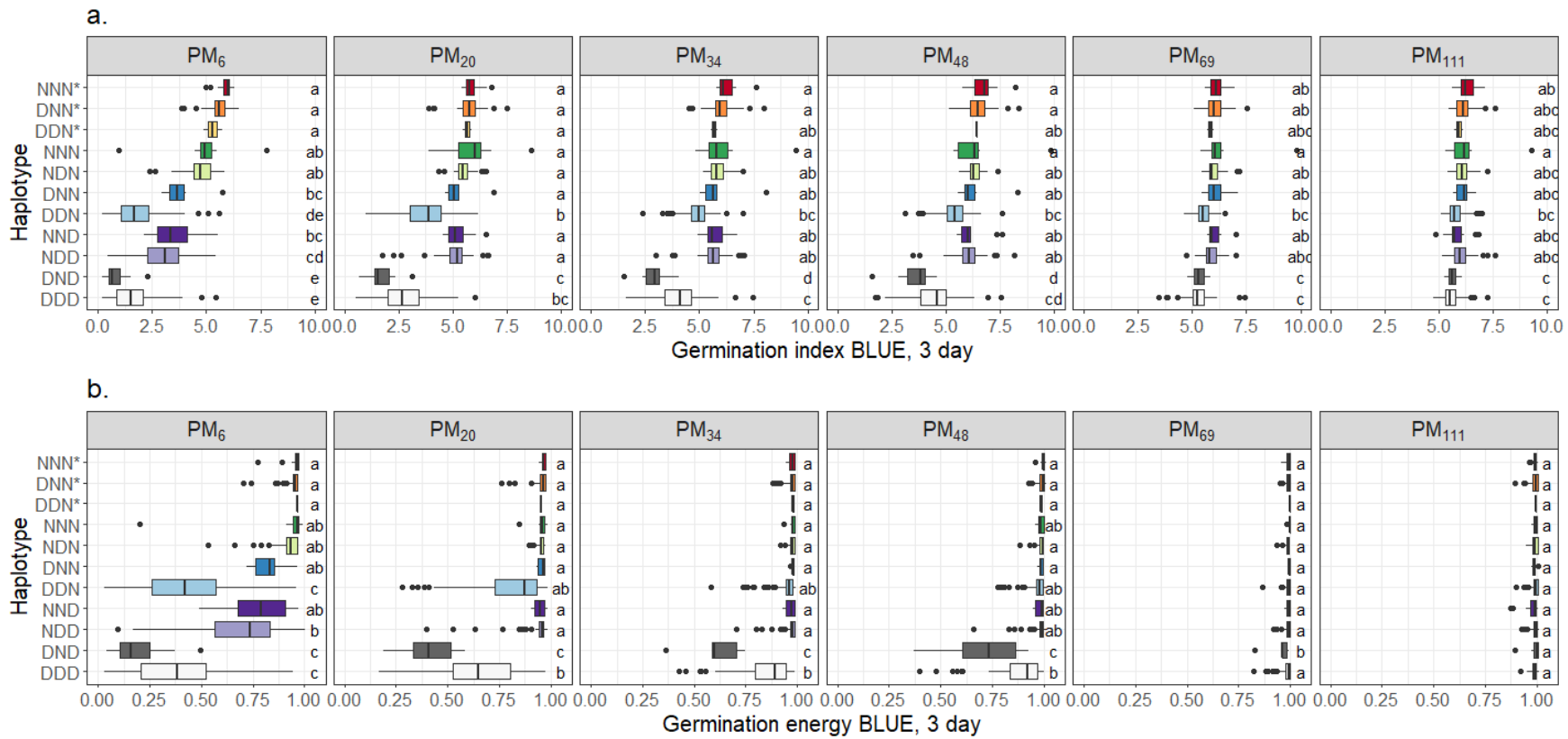


Figure 3.5: CU and JIC 2020 germination index (GI, facet a) and germination energy (GE, facet b) BLUEs at six time points after physiological maturity. Facet headers indicate the number of days post-physiological maturity (PM). A Tukey-Kramer test was used to determine significant differences ($p < .0.05$) between means of haplotype groups of unequal size. Haplotypes that share the same letter above the boxes are not significantly different from each other.

Table 3.4: Estimates (standard error in parentheses) for single and additive by additive interaction fixed marker effects for *HvAlaAT1*, *HvGA20ox1*, and *HvMKK3 for* CU/JIC 2020. Asterisks indicate marker effect significantly different from zero ($p < 0.05$). Marker effects indicate dormant allele effects except for $MKK3_N$ and $MKK3_D$, which indicate non-dormant and dormant allele effects respectively in reference to the very non-dormant MKK3 allele.

Dataset	Trait	Model term	PM_6	PM_{20}	PM_{34}	PM_{48}	PM_{69}	PM_{111}
CU/JIC	GE3	Intercept	0.952 (0.04)*	0.964 (0.031)*	0.973 (0.017)*	0.998 (0.016)*	0.987 (0.005)*	0.991 (0.004)*
CU/JIC	GE3	AlaAT1	-0.009 (0.045)	-0.011 (0.035)	-0.005 (0.019)	-0.009 (0.019)	0.002 (0.005)	-0.001 (0.004)
CU/JIC	GE3	GA20ox1	-0.001 (0.144)	-0.162 (0.114)	-0.169 (0.063)*	-0.139 (0.061)*	-0.025 (0.017)	0.01 (0.014)
CU/JIC	GE3	$MKK3_N$	0.069 (0.068)	0.056 (0.054)	0.038 (0.031)	0.01 (0.03)	0.007 (0.008)	-0.001 (0.007)
CU/JIC	GE3	$MKK3_D$	-0.241 (0.056)*	-0.065 (0.048)	-0.028 (0.027)	-0.031 (0.027)	0.004 (0.008)	-0.015 (0.006)*
CU/JIC	GE3	AlaAT1: $MKK3_N$	-0.473 (0.078)*	-0.274 (0.062)*	-0.192 (0.035)*	-0.142 (0.034)*	-0.035 (0.01)*	0.005 (0.008)
CU/JIC	GE3	AlaAT1: $MKK3_D$	-0.346 (0.076)*	-0.459 (0.061)*	-0.29 (0.035)*	-0.226 (0.034)*	-0.044 (0.01)*	0.003 (0.008)
CU/JIC	GE3	AlaAT1:GA20ox1	0.023 (0.063)	0.16 (0.051)*	0.179 (0.029)*	0.137 (0.029)*	0.035 (0.008)*	-0.002 (0.006)
CU/JIC	GE3	GA20ox1: $MKK3_N$	-0.131 (0.143)	0.081 (0.112)	0.125 (0.062)*	0.114 (0.06)	0.018 (0.017)	-0.009 (0.014)
CU/JIC	GE3	GA20ox1: $MKK3_D$	-0.005 (0.142)	0.204 (0.111)	0.193 (0.062)*	0.152 (0.06)*	0.023 (0.017)	0.003 (0.014)
CU/JIC	GI3	Intercept	5.866 (0.205)*	5.834 (0.203)*	6.245 (0.162)*	6.724 (0.171)*	6.106 (0.104)*	6.309 (0.096)*
CU/JIC	GI3	AlaAT1	-0.3 (0.233)	-0.081 (0.231)	-0.258 (0.185)	-0.219 (0.194)	-0.084 (0.118)	-0.186 (0.109)
CU/JIC	GI3	GA20ox1	-0.421 (0.748)	-0.734 (0.744)	-1.107 (0.601)	-0.711 (0.634)	-0.443 (0.381)	-0.147 (0.353)
CU/JIC	GI3	$MKK3_N$	-0.371 (0.353)	0.558 (0.351)	0.279 (0.292)	0.086 (0.308)	0.371 (0.186)*	0.007 (0.166)
CU/JIC	GI3	$MKK3_D$	-2.585 (0.292)*	-0.856 (0.312)*	-0.705 (0.261)*	-0.696 (0.282)*	0.005 (0.18)	-0.415 (0.141)*
CU/JIC	GI3	AlaAT1: $MKK3_N$	-2.603 (0.405)*	-2.259 (0.407)*	-1.434 (0.336)*	-1.284 (0.356)*	-0.663 (0.217)*	-0.114 (0.193)
CU/JIC	GI3	AlaAT1: $MKK3_D$	-1.436 (0.395)*	-2.946 (0.4)*	-2.148 (0.332)*	-2.01 (0.352)*	-0.754 (0.216)*	-0.173 (0.19)
CU/JIC	GI3	AlaAT1:GA20ox1	0.152 (0.326)	0.644 (0.332)	0.796 (0.279)*	0.623 (0.297)*	0.261 (0.182)	-0.049 (0.158)
CU/JIC	GI3	GA20ox1: $MKK3_N$	-0.522 (0.739)	-0.221 (0.732)	0.395 (0.587)	0.178 (0.618)	-0.017 (0.37)	-0.058 (0.347)
CU/JIC	GI3	GA20ox1: $MKK3_D$	0.286 (0.735)	0.872 (0.728)	1.193 (0.587)*	0.732 (0.619)	0.176 (0.37)	0.267 (0.347)
CU/JIC	PHS	Intercept	6.124 (0.234)*					
CU/JIC	PHS	AlaAT1	-1.127 (0.267)*					
CU/JIC	PHS	GA20ox1	-1.004 (0.858)					

CU/JIC	PHS	$MKK3_N$	-4.179 (0.406)*					
CU/JIC	PHS	$MKK3_D$	-5.175 (0.335)*					
CU/JIC	PHS	AlaAT1: $MKK3_N$	0.648 (0.469)					
CU/JIC	PHS	AlaAT1: $MKK3_D$	0.863 (0.456)					
CU/JIC	PHS	AlaAT1:GA20ox1	0.409 (0.379)					
CU/JIC	PHS	GA20ox1: $MKK3_N$	0.794 (0.853)					
CU/JIC	PHS	GA20ox1: $MKK3_D$	1.493 (0.842)					
S2MET	PHS	Intercept	5.876 (0.766)*					
S2MET	PHS	GA20ox1	0.279 (0.387)*					
S2MET	PHS	$MKK3_N$	-1.451 (0.777)					
S2MET	PHS	$MKK3_D$	-2.961 (0.786)*					
S2MET	PHS	GA20ox1: $MKK3_N$	0.189 (0.436)					
S2MET	PHS	GA20ox1: $MKK3_D$	-0.187 (0.502)					

By PM_{111}, differences in GI3 between haplotypes were still present but were considerably smaller than earlier time points. *HvGA20ox1* main and interaction effects were mostly not significant. For several time point/trait combinations, dormant *HvGA20ox1* caused a significant reduction in GE3 or GI3, but this was typically countered by a positive interaction with *HvAlaAT1* or *HvMKK3* (Table 3.4). The only haplotype pair where *HvGA20ox1* may have had a consistent effect was DNN/DDN. DDN lines had longer seed dormancy and slower GI3 than DNN lines across time points, but DNN only contained 7 lines so comparisons between the two haplotypes may not be accurate. Additive by additive epistatic interactions between *HvAlaAT1* and *HvMKK3* alleles were significant ($p < 0.05$) at most time points for GE3 and GI3, but epistatic effects were generally smaller than main marker effects. All epistatic interaction effects decreased as after-ripening progressed.

Sanger sequencing results

Eight polymorphisms were detected in the coding sequence of *HvGA20ox1* which differentiated four unique alleles (Table 3.5.2). Three SNPs in the 5' UTR, two synonymous SNPs in exon 1, a single base pair deletion in the intron, and two synonymous SNPs in exon 2 were detected. Alignment of the barley pan-genome sequences revealed an additional three to six base pair deletion at the beginning of the 5' UTR as well as a number of SNPs and indels in the putative promoter region 1500 bp upstream of the 5' UTR transcription start site. Most PHS susceptible phenotypes had the $A003_{GA20ox1}$ allele but this allele was also present in PHS resistant lines, including Steptoe and winter six-row lines with Steptoe in their pedigree. Most PHS susceptible two-row spring lines in CU and S2MET with $A003_{GA20ox1}$ also carried the E165Q *HvMKK3* allele, $A001_{MKK3}$. $A004_{GA20ox1}$ was almost exclusively found in two-row spring lines while $A001_{GA20ox1}$ was more common in winter barley lines.

In the JIC, variants at the four *HvMKK3* and the single *HvAlaAT1* SNPs described by Vetch et al. (2020) were identified as well as a novel variant in a single line in exon 10 of *HvAlaAT1* (Table 3.5.1 and 3.5.3). Three JIC lines, Kagelkorn, Tamparkorn, and St. Davids, carried $A001_{MKK3}$. *HvMKK3* sequences from Harrington, Steptoe, and 36 unique alleles detected in cultivated barley by Nakamura et al. (2016) were aligned with *HvMKK3* sequence of the pan-genome lines. $A001_{MKK3}$ was only observed in Hockett, Harrington, and a single accession from Nakamura et al. with Hap_033, but Hap_033 had intron length variants and a unique SNP in exon 9 compared to Hockett and Harrington. The CU parents AAC Synergy, Newdale, and Bentley carried $A001_{MKK3,}$ Pinnacle and ND Genesis carried $A003_{MKK3}$, (Vetch et al., 2020), and Craft, Conlon, and KWS Tinka carried $A002_{MKK3}$. The only naked barley to carry $A001_{MKK3}$ was

CDC Clear, which has Canadian malting germplasm in its pedigree. In CU $A001_{MKK3}$ was almost completely linked with $A003_{GA20ox1}$, confounding the effects of the two loci. 94/96 S2MET lines with $A001_{MKK3}$ also carried $A003_{GA20ox1}$. No crossovers between *HvGA20ox1* and *HvMKK3* were detected in CU based on 50k markers, but two lines from the same family with a putative DN* *HvGA20ox1*/*HvMKK3* haplotype were identified from KASP markers.

Pleiotropy

CU and S2MET were used to assess potential pleiotropic effects of *HvGA20ox1* on agronomic traits. CU and S2MET were structured populations and *HvGA20ox1* was correlated with the first (r=-0.577) and second (0.234) PCs in S2MET but only showed a small correlation (0.33) with PC2 in CU. After accounting for population structure, *HvGA20ox1* had no effect on grain yield or plant height across diverse environments in S2MET but had a small, significant ($p < 0.05$) effect for heading date accounting for the first and second principal components of S2MET. When each environment was analyzed separately, *HvGA20ox1* was not significant in any environment for plant height. *HvGA20ox1* did not show any effect on height in CU but had a small effect on heading date ($p < 0.05$). Lines with the non-dormant *HvGA20ox1* allele headed one day later on average than dormant *HvGA20ox1* lines.

Table 3.5.1: Summary of *HvAlaAT1* alleles in winter and spring breeding lines and the JIC panel

		Gene	***HvAlaAT1***	
		Exon/Intron	**Exon 9**	**Exon 10**
		Mutation type	*missense*	*missense*
Allele	**n**	**bp position**	**222**	**467**
$A001_{AlaAT1}$	33		C	C
$A002_{AlaAT1}$	109		G	C
$A003_{AlaAT1}$	1		C	A

Table 3.5.2: Summary of *HvGA20ox1* alleles in winter and spring breeding lines

		Gene	***HvGA20ox1***								
		Exon/Intron	**5' UTR**			**Exon 1**		**Intron 1**	**Exon 2**		
		Mutation type	*n/a*	*n/a*	*n/a*	*silent*	*silent*	*n/a*	*silent*	*silent*	*missense*
Allele	**n**	**bp position**	**284**	**316**	**331**	**458**	**764**	**954**	**1216**	**1381**	**1480**
$A001_{GA20ox1}$	21		T	T	C	C	G	-	C	G	A
$A002_{GA20ox1}$	3		C	A	C	C	G	-	C	G	A
$A003_{GA20ox1}$	16		C	A	G	A	C	G	T	C	A
$A004_{GA20ox1}$	25		C	A	C	C	G	-	T	G	A

Table 3.5.3: Summary of *HvMKK3* alleles in winter and spring breeding lines and the JIC panel

		Gene	***HvMKK3***				
		Exon/Intron	**Exon 3**	**Exon 4**	**Exon 7**	**Exon 7**	**Exon 8**
		Mutation type	*missense*	*missense*	*missense*	*silent*	*missense*
Allele	**n**	**bp position**	**189**	**498**	**476**	**512**	**32**
$A001_{MKK3}$	5		C	G	G	T	G
$A002_{MKK3}$	53		G	G	C	T	A
$A003_{MKK3}$	15		G	G	G	T	G
$A004_{MKK3}$	2		G	G	G	G	
$A005_{MKK3}$	8		G	T	C	T	A

Discussion

Much effort has gone towards identifying sources of PHS resistance and seed dormancy in barley that do not compromise malting quality or malthouse performance. Mapping studies have repeatedly detected two large effect PHS and seed dormancy loci, *SD1* and *SD2*, on chromosome 5H but fine scale mapping of the critical *SD2* locus has not been achieved due to insufficient marker density and limited mapping population diversity. Three sets of spring malting barley germplasm, highly structured breeding material from a single program, more diverse breeding material from multiple programs, and a landrace panel, were mapped for PHS, seed dormancy (GE3), and germination rate (GI3) with high density markers and a high power GWA model to determine change in genetic control of germination traits over time. Marker-trait associations in the *SD1* and *SD2* regions were identified from six to eleven days after physiological maturity (PM_6 to PM_{111}). Epistatic interactions between *SD1* and *SD2* changed over time as dormancy was released.

Dormancy did not extend beyond four months for any CU line, with the majority of lines fully breaking dormancy in less than two months. Germination rate increased then plateaued at different points post-PM for each seed dormancy haplotype in CU, indicating that even for lines with low primary dormancy, after-ripening improves germination rate. A small amount of ABA sensitivity that decreases the germination rate for a short time may still be present even for lines with very low primary dormancy. Correlations between 24-hour germination and GI3 were high for CU and improved with after-ripening, similar to Woonton et al. (2005). A one-day germination test anywhere between one and three months post-harvest is a simple, inexpensive phenotype that may be of use to malting barley breeders if high germination rate has even a moderate correlation with malting quality.

High r_g between PHS, GE3, and GI3 at PM_6 indicated similar genetic control of the three traits immediately after PM. GWA results for PHS, GE3, GI3 at PM_6 detected large effects of MKK3_E165Q in 2019, 2020, and 2019/2020 but large effects from AlaAT_L214F and JHI-367342-KASP were only detected for GE3 and GI3 in 2020 and 2019/2020. Although PHS, GE3, and GI3 measure similar phenotypes, the scales are not in perfect correspondence. The PHS assay used here is more a measure of GI3 than GE3 since coleoptile emergence factors into the scoring scale. A GE3 score of 1.0 indicating complete germination may not score more than a 4 or 5 out of 9 on the PHS scale if no coleoptile emergence is observed. Spikes used in PHS tests were sampled at PM and phenotypes were recorded six days later. Spikes for GE and GI assays were purposefully sampled two days after 50% of the headrow had reached PM to facilitate efficient harvesting. The addition of two to three days of after-ripening in a heated (35 C) dryer may have partially released primary dormancy in lines with non-dormant *HvAlaAT1* and *HvMKK3* alleles, leading to the detection of *HvAlaAT1* and N or D *HvMKK3* alleles in GWA. The precise timing of primary dormancy release from PM_0 to PM_9 in these haplotypes should be further examined to improve harvest time decisions for growers.

Phenotypic variation for PHS, GE3, and GI3 was observed within seed dormancy haplotype in CU, JIC, and S2MET. A number of ABA metabolism genes like 9-cis-epoxycarotenoid dioxygenase (NCED) (Leymarie et al., 2008) and CYP707A (Millar et al., 2006; Chono et al., 2006) have been connected to regulation of seed dormancy in barley. Small effect variants or expression differences caused by environmental factors at such loci could alter the ABA/GA balance in developing seeds and introduce additional variation within seed dormancy haplotype. Sampling error cannot be discounted because of the difficulty of precisely identifying PM when spike sampling. The effects of *HvAlaAT1* and *HvMKK3* are large enough

that further mapping should be conducted within haplotype. ND Genesis and Pinnacle have the same haplotype (NDN) but ND Genesis had higher PHS (BLUE = 3.32) than Pinnacle (BLUE = 1.86) and several *HvMKK3* N* lines in CU and S2MET unexpectedly had very low PHS.

HvAlaAT1

SD1 was initially identified as a seed dormancy QTL in barley by Oberthur et al. (1993) and its role in long-term seed dormancy has been further examined by many authors, including Romagosa et al. (1999), Han et al. (1999), Sato et al. (2016), and Vetch et al. (2020). Sato et al. (2016) cloned *HvAlaAT1* and showed it was the causative gene at *SD1*. *TaAlaAT* has also been associated with seed dormancy and is potentially involved in ABA signaling in wheat, as the dormant allele was associated with increased expression and signaling of NCED (Wei et al., 2019). *HvAlaAT1/SD1* has not been observed to affect PHS when phenotyped within 7 days of PM (Ullrich et al., 2009; Vetch et al., 2020) but was associated with GE3 at PM_7 in Prada et al. (2004) and Bonnardeaux et al. (2008). Similar results were observed in the CU/JIC population. *HvAlaAT1* did not have a significant effect on PHS but had a large effect on primary seed dormancy with persisting effects on seed dormancy and germination rate. DND lines exhibited lower PHS than NND lines but the sample sizes for both haplotypes were very small. *HvAlaAT1* effects were conditioned by allelic state of *HvMKK3*. Epistatic interactions between *SD1* and *SD2* have been previously described across an after-ripening regime **(**Romagosa et al. ,1999) and at single time points (Han et al., 1996; Bonnardeaux et al., 2008; Gong et al., 2014). The AlaAT_L214F marker designed to capture the causal SNP in *HvAlaAT1* identified by Sato et al. (2016) was the most frequently detected MTA in QSD1_5H, validating the effect of this polymorphism in historic European and modern North American germplasm.

The dormant *HvAlaAT1* allele is likely undesirable for malting barley breeding since a slow germination rate may be correlated with poor modification and malting quality. Sato et al. (2016) and Vetch et al. (2020) found the non-dormant *HvAlaAT1* allele to be enriched in malting barley varieties and reached similar conclusions. We observed similar trends in all three populations in this study but noted several high malting quality varieties with dormant *HvAlaAT1* and non-dormant N* *HvMKK3*, further supporting evidence of epistatic interactions in this haplotype. Previous seed dormancy mapping efforts in barley have focused on spring germplasm. Winter malting barley varieties with good malting quality were also sequenced for *HvAlaAT1* and *HvMKK3* and some carried dormant alleles of both genes, indicating either additional malting quality loci in winter germplasm or the sensitivity of seed dormancy to environmental factors.

HvGA20ox1

HvGA20ox1 has been suggested as a candidate gene for seed dormancy and PHS (Li et al., 2003; Nagel et al., 2019) but direct evidence of *HvGA20ox1* association with these traits and interactions of *HvGA20ox1* with *HvAlaAT1* and *HvMKK3* has not been previously described. Several SNPs and indels in *HvGA20ox1* were identified in the 5' UTR and intron but exonic mutations were non-synonymous. $A003_{GA20ox1}$, which was carried by most PHS susceptible lines, was also present in PHS resistant winter six-row naked barley lines as well as the dormant Steptoe. Toora (2018) sequenced the coding region of *HvGA20ox1* in ~20 spring barley lines, including AC Metcalfe, AAC Synergy, Morex, and Steptoe, and identified the same polymorphisms in the coding region as this study. *HvGA20ox1* has six alternate transcripts but none of the polymorphisms become non-synonymous in the alternate transcripts, although

several still contain polymorphisms in the 5' UTR. The GA20ox1_331_5UTR marker did not have a significant effect on PHS, seed dormancy, and germination rate in CU/JIC or PHS in S2MET.

The sequencing results suggest that any change in phenotype due to *HvGA20ox1* would be caused by changes in gene expression and not loss-of-function mutations. Polymorphisms located in non-exonic regions as well as large deletions in GA 20 oxidases have been connected to seed dormancy, height, and yield in other grass species. *HVGA20ox2* (HORVU.MOREX.r2.3HG0256590) has multiple alleles conferring a semi-dwarf phenotype in barley (Xu et al., 2017). *sdw1.a* (formerly *sdw1*) has 60-fold less gene expression than wild type and is only suitable for feed barleys due to slow germination while *sdw1.c* (formerly *denso*) and *sdw1.d* have been widely used in malting barley breeding programs worldwide (Jia et al. 2011; Xu et al., 2017). The *sdw1.a* allele is a result of the deletion of *HvGA20ox2* but *sdw1.c* and *sdw1.d* contain indels in the 5' UTR and exon 1, respectively (Xu et al., 2017). A 383 bp deletion in *OsGA20ox2* is responsible for the "Green Revolution" semi-dwarf phenotype in rice as well as enhanced seed dormancy (Ye et al., 2015). Several *OsGA20ox1* promoter region variants have been detected and associated with changes in grain number per panicle (Wang et al., 2020) and seedling vigor (Abe et al., 2012) in rice. Large deletions in *HvGA20ox1* were not detected in the pan-genome or any line sequenced in this study but a 3-6 bp deletion was observed at the beginning of the 5' UTR that was perfectly correlated with $A003_{GA20ox1}$. The primers used to sequence *HvGA20ox1* did not extend far enough to capture the entire 5' UTR so the effect of that deletion remains unknown. Potential promoter sites up to 1500 bp proximal to the beginning of the 5' UTR were identified using web-based promoter sequence prediction tool, PromPredict (Morey et al., 2011), using barley pan-genome sequence. A single ~50 bp putative promoter

region was identified that contained an indel in Morex, Hockett, Igri, and Akashinriki. This promoter sequence was perfectly correlated with $A003_{GA20ox1}$ and the 5' UTR deletion. Given the lack of phenotypic and mapping evidence in these populations, any promoter variants seem unlikely to contribute to *HvGA20ox1* influence on PHS and seed dormancy.

Pleiotropic effects of *HvGA20ox1* for height and heading date were not detected in CU and a very small positive effect of the non-dormant allele on heading date was detected in S2MET. KWS Tinka was a donor of the semi-dwarf allele at *HvGA20ox2* in CU but there was no observed impact of the semi-dwarf *HvGA20ox2* allele on PHS, seed dormancy, or germination rate in CU nor were there additional effects on height in phenotypically semi-dwarf CU lines with dormant *HvGA20ox1* (data not shown). Pleiotropic effects of *HvGA20ox1* on agronomic traits seem to be of minimal concern to breeders. Potential pleiotropy for malting quality traits is of greater consequence and is under further study in CU.

Wheat carries three copies of GA20ox1 on chromosomes 4AL, 5BL, and 5DL (Spielmeyer et al., 2004). The telomeric end of chromosome 5HL in barley is syntenic with a region on chromosome 4AL in wheat (Li et al., 2004) that is known to be part of an ancestral 5AL/4AL translocation (Naranjo et al., 1987). This region has repeatedly been associated with PHS and seed dormancy in wheat (Torada et al., 2008; Kulwal et al., 2012; Shorinola et al., 2017) and *TaMKK3-A* was cloned and identified as the causal gene at *Phs1* by Torada et al. (2016). Given the high synteny between wheat 4AL and barley 5HL, variation in *TaGA20ox1* could affect PHS resistance and seed dormancy in wheat and may be an additional source of variation at *Phs1* but the 4AL copy of *TaGA20ox1* did not contain any variants in the 10+ genome lines in the coding region or 5' UTR.

HvMKK3

MKK3 has been associated with PHS and seed dormancy in wheat and rice. *OsMKK3* is part of a pathway regulating expression of mother of flowering time (*OsMFT*) in rice. Knock outs of *OsMKK3* increased expression of *OsMFT* and increased seed dormancy (Mao et al., 2019). A single non-synonymous SNP mutation in *TaMKK3* was identified and proposed to be the causal polymorphism at the *Phs-A1* locus on chromosome 4A by Torada et al. (2016), which was later confirmed in more diverse germplasm by Shorinola et al. (2017). Nakamura et al. (2016) cloned *HvMKK3,* demonstrated its role in seed dormancy in barley, and identified a remarkable 92 *HvMKK3* alleles in wild and cultivated barley. Vetch et al. (2020) did not find the N260T mutation identified by Nakamura et al. (2016) as the causal polymorphism in cv. Azumamugi in North American spring malting barley germplasm but described four additional alleles, identifying E165Q ($A001_{MKK3}$ or very non-dormant/N* in this study) as the primary non-dormant *HvMKK3* allele. Repeated association of the MKK3_E165Q KASP marker with PHS, GE3, and GI3 validated these findings. Additional results from this study also strongly suggested that two other *HvMKK3* alleles, $A002_{MKK3}$ (dormant/D) and $A003_{MKK3}$ (non-dormant/N), have large effects on seed dormancy and germination rate in spring malting barley germplasm. $A001_{MKK3}$ is carried by Klages, which is in the pedigree of many current varieties with $A001_{MKK3}$ in this study and Vetch et al. (2020). Klages was derived from Betzes/Domen. Domen, derived from OpalB/Maskin, also carries $A001_{MKK3}$ and Betzes carries $A002_{MKK3}$**.** OpalB is a selection from the Danish line Opal, which was sequenced by Nakamura et al. (Hap_001) and does not carry $A001_{MKK3}$. No *HvMKK3* sequence for Maskin, a Norwegian six-row spring landrace (Aikasalo, 1988), is available but $A001_{MKK3}$ was identified in two other Norwegian six-row spring landraces in the JIC, Kagelkorn and Tamparkorn. Seed dormancy tends to increase in

cooler growing conditions (Chapter 4) and genetic sources of very low seed dormancy may have been historically favored for malting in very cool climates such as Scandinavia. $A001_{MKK3}$ was primarily found in varieties from breeding programs in western Canada, Washington, Montana, and Idaho while $A002_{MKK3}$ and $A003_{MKK3}$ were present in a wider range of modern and historic germplasm from North America and Europe. $A001_{MKK3}$ seems to be relatively rare globally but is enriched in western North American malting germplasm that has 'Canadian-style' malting quality, characterized by high enzyme activity, free amino nitrogen (FAN), and low beta-glucan (Edney et al., 2013). This malt quality profile is highly preferred by North American adjunct brewers. Varieties with 'Canadian-style' malting quality frequently have PHS susceptibility. Malting quality and PHS QTL co-localize at *SD2* in this germplasm (Mohammadi et al., 2015; Zhou et al., 2016). European and North American craft brewing styles rarely use adjuncts and have different preferences for malting quality profiles, generally favoring lower enzyme levels, FAN, and protein than adjunct brewing. Combined with favorable malting quality QTL at loci besides *SD2* and non-dormant *HvAlaAT1*, the $A002_{MKK3}$ and $A003_{MKK3}$ alleles may help provide effective protection against PHS without compromising malting quality for craft malting and brewing markets.

Genetics of the SD2 locus

HvGA20ox1 and *HvMKK3* are tightly linked and recombinants in CU were very rare (only 2 lines out of 424). Association mapping takes advantage of ancestral recombination to identify loci (Ersoz et al., 2009). QTL in tight repulsion linkage can be identified using MLMM and increased population size and heritability improve the chances of detecting two separate QTL (Li et al., 2015). The large population sizes used in this study, high trait heritability, presence of

ancestral recombination between *HvGA20ox1* and *HvMKK3* in JIC and S2MET, and the ability of MLMM to detect tightly linked variants by adding significant QTL as covariates before testing other SNPs (Segura et al., 2012) likely could have identified effects of each gene if present. Genotyping the *HvMKK3* E165Q mutation was crucial in differentiating the effects of *HvMKK3* and *HvGA20ox1*. GWA without MKK3_E165Q in CU/JIC and S2MET resulted in two MTAs in the *SD2* region, with one MTA in an LD block around *HvGA20ox1* and the second in an LD block around *HvMKK3* (data not shown). Differentiating $A001_{MKK3}$ from $A002_{MKK3}$ and $A003_{MKK3}$ is dependent on two SNPs with different MAF, as observed for MKK3_E165Q (CU/JIC: 0.191, S2MET: 0.428) and JHI-367342-KASP (CU/JIC: 0.380, S2MET: 0.269). This difference in MAF enabled two MTAs to be identified. In mapping populations with extended LD, limited or no recombination between *HvMKK3* and *HvGA20ox1*, and low-density genotyping, *HvMKK3* alleles may appear to be two separate loci.

Sequencing results for *HvGA20ox1* and *HvMKK3* in this study complement previous work by Nakamura et al. (2016) in *HvMKK3* to elucidate results from earlier PHS and seed dormancy mapping studies that detected *SD2*, specifically Steptoe/Morex (Oberthur et al., 1993), Triumph/Morex (Prada et al., 2004), Haruna Nijo/H602 (Hori et al., 2007), and Harrington/Morex and Chevron/Stander (Lin et al., 2009). Harrington carries the $A001_{MKK3}$ allele, Steptoe, Morex, Stander, and Haruna Nijo carry $A003_{MKK3}$, and Triumph, Chevron, and H602 carry $A002_{MKK3}$. In Harrington/Morex, Morex donated the dormant allele at *SD2* but in Triumph/Morex, Morex donated the non-dormant allele. Lin et al. (2009) initially proposed multiple alleles of a single major gene at *SD2* and our findings of three significantly different *HvMKK3* alleles with $A001_{MKK3}$ and $A002_{MKK3}$ as the least and most dormant alleles support this hypothesis. Stander and Haruna Nijo were the non-dormant donor in their respective populations,

supporting our findings of increased dormancy of $A002_{MKK3}$ compared to $A003_{MKK3}$. Steptoe was the dormant allele donor at *SD2* in Steptoe/Morex, indicating another gene at *SD2*, splicing or promoter variation in *HvMKK3*, or potentially undetected variation in the 5' UTR or promoter region of *HvGA20ox1* in Steptoe.

Other marker trait associations

A locus at the proximal end of chromosome 2H was identified in S2MET for PHS and in CU 2019 for GI3 at PM_{111}. This QTL may be coincident with a malt extract QTL from a Harrington/Morex DH population (Marquez-Cedillo et al., 1999). Candidate genes in this region include an ethylene receptor, cysteine synthase, cytochrome P450 family, protein phosphatase 2C (PP2C), and polycomb group proteins similar to VERNALIZATION 2 and EMBRYONIC FLOWER2 in Arabidopsis. PP2C is a negative regulator of ABA signaling (Arc et al., 2013). Cyanoalanine/cysteine synthase is activated by ethylene during germination to catalyze hydrogen cyanide to b-cyanoalanine (Hasegawa et al., 1994). Expression data from Betts et al. (2019) indicated decreased expression of PP2C in the embryo and scutellum 24 hours after imbibition and increased expression of cyanoalanine/cysteine synthase in the embryo and scutellum. These data are consistent with the roles of the respective genes in germination. Down-regulation of a negative ABA regulator like PP2C could be involved in inducing ABA insensitivity and reducing dormancy. Genes of interest in the QSD1_4H region include a jasmonate-zim-domain protein and an F-box transcription factor. Methyl jasmonate has been implicated in seed dormancy in wheat (Xu et al., 2016).

At PM_6 in the combined 2019/2020 dataset, a MTA from marker JHI-Hv50k-2016-311615 for GE3 at 475 Mb on chromosome 5H was detected that fell just below the Bonferroni

cutoff (p-value = 7.0e-6). Isoamylase 3 (*HvISA3*, HORVU.MOREX.r2.5HG0404420) is a starch-debranching enzyme located at 475796690-475807295 bp, 93619 bp distal to marker JHI-Hv50k-2016-311615. *HvISA3* was the strongest candidate gene in the region. Isoamylases hydrolyze α-(1→6) glycosidic linkages and debranch amylopectin during grain filling (Gous and Fox, 2017). Three isoforms of isoamylase have been identified in barley and *HvISA3* is expressed in the pericarp and endosperm during early grain filling (Radchuk et al., 2009). Limit dextrinase is the other starch debranching enzyme to hydrolyze amylopectin in barley but it is more active during germination, making it an important malting quality trait. Shu and Rasmussen (2014) identified a MTA for amylose content highly associated with the contig MLOC_10776, which includes *HvISA3* (https://ics.hutton.ac.uk/morexGenes/). Starch with higher amylose content is hydrolyzed more slowly by amylolytic enzymes and higher amylose content has been hypothesized to be a contributing factor to grain dormancy (Chu et al., 2014). The dormant allele of JHI-Hv50k-2016-311615 had a small effect on decreasing PHS and GE at early time points as well as a small but consistent reduction in GI across after-ripening timepoints. This locus may be useful for PHS resistance but the consistent decrease in GI is not desirable. Negative impacts on starch related malting quality traits like malt extract may also be present. Although the role of amylose content in seed dormancy has been hypothesized, prior evidence for the role of *HvISA3* is limited. Hickey et al. (2013) identified the *qFlag* seed dormancy QTL in a Flagship/2ND24640 cross flanked by the DaRT markers bPb-46127 and bPb-7763, which map to 446906000 and 448737500 bp on chromosome 5H, respectively (https://www.diversityarrays.com/technology-and-resources/sequences/). These markers are physically closer to *HvAlaAT1* but there may have been extended LD with the *HvISA3* region in that population. The same Flagship/2ND24640 population showed variation for grain amylose

content in a staygreen experiment looking at impacts of heat and drought stress on starch content (Gous et al., 2017).

Conclusion

Results from GWA supported by KASP markers and Sanger sequencing indicate that three previously known *HvMKK3* alleles present in North American and European malting barley have distinct effects on PHS and seed dormancy. Length of primary seed dormancy and germination rate from physiological maturity to full after-ripening are largely determined by interactions of *HvMKK3* alleles with *HvAlaAT1*. GWA results from three diverse populations and Sanger sequencing of *HvGA20ox1* do not suggest that *HvGA20ox1* is associated with PHS, seed dormancy, or germination rate. The *SD2* tight linkage hypothesis for PHS and seed dormancy seems unlikely based on these results, but the effects of *HvMKK3* alleles and *HvGA20ox1* on malting quality traits remain unknown. Given the prevalence of $A001_{MKK3}$ in lines used to make high enzyme malt for adjunct brewing, $A001_{MKK3}$ likely has pleiotropic effects on malting quality. Ongoing work in the CU population is focused on determining the effects of *HvAlaAT1*, *HvGA20ox1*, and *HvMKK3* on malting quality. The KASP markers developed for *HvAlaAT1*, *HvGA20ox1*, and multiple alleles of *HvMKK3* are a resource for high-throughput screening of desirable dormancy haplotypes for breeding efforts. We identified several haplotypes of *HvAlaAT1* and *HvMKK3* with PHS resistance, short primary dormancy, and rapid germination after a short after-ripening period. Different seed dormancy haplotypes may be preferred depending on climate and target brewing market. Due to variation within haplotype, phenotyping for PHS and seed dormancy should still be conducted within breeding programs at high risk for

PHS. One-day GI assays may be a better throughput method for screening dormancy and malting quality at early stages of the breeding cycle than current methods.

References

Abe, A., H. Takagi, T. Fujibe, K. Aya, M. Kojima, et al. 2012. OsGA20ox1, a candidate gene for a major QTL controlling seedling vigor in rice. Theor. Appl. Genet. 125(4): 647–657. doi: 10.1007/s00122-012-1857-z.

Anderson, J.A., M.E. Sorrells, and S.D. Tanksley. 1993. RFLP analysis of genomic regions associated with resistance to preharvest sprouting in wheat. Crop Sci. 33(3): 453. doi: 10.2135/cropsci1993.0011183X003300030008x.

Arc, E., J. Sechet, F. Corbineau, L. Rajjou, and A. Marion-Poll. 2013. ABA crosstalk with ethylene and nitric oxide in seed dormancy and germination. Front. Plant Sci. 4(MAR): 63. doi: 10.3389/fpls.2013.00063.

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, **67**(1), 1–48. doi: 10.18637/jss.v067.i01.

Bayer, M.M., P. Rapazote-Flores, M. Ganal, P.E. Hedley, M. Macaulay, et al. 2017. Development and evaluation of a barley 50k iSelect SNP array. Front. Plant Sci. 8: 1792. doi: 10.3389/fpls.2017.01792.

Bewley, J.D., K.J. Bradford, H.W.M. Hilhorst, and H. Nonogaki. 2013. Seeds: Physiology of development, germination and dormancy, 3rd edition. Springer New York.

Blake, V.C., C. Birkett, D.E. Matthews, D.L. Hane, P. Bradbury, et al. 2016. The Triticeae Toolbox: Combining phenotype and genotype data to advance small-grains breeding. Plant Genome 9(2): plantgenome2014.12.0099. doi: 10.3835/plantgenome2014.12.0099.

Bonnardeaux, Y., C. Li, R. Lance, X.Q. Zhang B,C, K. Sivasithamparam, et al. Seed dormancy in barley: identifying superior genotypes through incorporating epistatic interactions. doi: 10.1071/AR07345.

Castro, A.J., A. Benitez, P.M. Hayes, L. Viega, and L. Wright. 2010. Coincident quantitative trait loci effects for dormancy, water sensitivity and malting quality traits in the BCD47×Baronesse barley mapping population. Crop Pasture Sci. 61(9): 691. doi: 10.1071/CP10085.

Chono, M., I. Honda, S. Shinoda, T. Kushiro, Y. Kamiya, E. Nambara, N. Kawakami, S. Kaneko, Y. Watanabe. 2006. Field studies on the regulation of abscisic acid content and

germinability during grain development of barley: molecular and chemical analysis of pre-harvest sprouting | Journal of Experimental Botany | Oxford Academic. J. Exp. Bot. Vol. 57, Issue 10. doi: https://doi.org/10.1093/jxb/erj215.

Chu, S., J. Hasjim, L.T. Hickey, G. Fox, and R.G. Gilbert. 2014. Structural changes of starch molecules in barley grains during germination. Cereal Chem. J. 91(5): 431–437. doi: 10.1094/CCHEM-09-13-0174-R.

Cullis, B.R., A.B. Smith, and N.E. Coombes. 2006. On the design of early generation variety trials with correlated data. Journal of Agricultural, Biological, and Environmental Statistics. 11(4): 381-393. doi: 10.1198/108571106X154443.

de Mendiburu, F. and Yaseen, M. 2020. agricolae: Statistical procedures for agricultural research. R package version 1.4.0 , https://myaseen208.github.io/agricolae/https://cran.r-project.org/package=agricolae.

Depauw, R.M., and T.N. Mccaig. 1991. Components of variation, heritabilities and correlations for indices of sprouting tolerance and seed dormancy in Triticum spp. Euphytica 52: 221-229.

Edney, M.J., W.G. Legge, M.S. Izydorczyk, T. Demeke, and B.G. Rossnagel. 2013. Identification of barley breeding lines combining preharvest sprouting resistance with “Canadian-type” malting quality. Crop Sci. 53(4): 1447. doi: 10.2135/cropsci2012.11.0649.

Ersoz, E.S., J. Yu, and E.S. Buckler. 2009. Applications of linkage disequilibrium and association mapping in maize. Biotechnology in Agriculture and Forestry. Springer International Publishing. p. 173–195

Frančáková, H., M. Líšková, T. Bojňanská, and J. Mareček. 2012. Germination index as an indicator of malting potential. Czech J. Food Sci. 377-384, 30(4).

Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2009. ASReml User Guide. Hemel Hempstead, HP1 1ES, UK.

Gong, X., C. Li, M. Zhou, Y. Bonnardeaux, and G. Yan. 2014. Seed dormancy in barley is dictated by genetics, environments and their interactions. Euphytica 197: 355:368. doi: 10.1007/s10681-014-1072-x.

Gous, P.W., and G.P. Fox. 2017. Review: Amylopectin synthesis and hydrolysis – Understanding isoamylase and limit dextrinase and their impact on starch structure on barley (Hordeum vulgare) quality. Trends Food Sci. Technol. 62: 23–32. doi: 10.1016/j.tifs.2016.11.013.

Gous, P.W., F. Warren, R. Gilbert, and G.P. Fox. 2017. Drought-proofing barley (*Hordeum vulgare*): The effects of stay green on starch and amylose structure. Cereal Chem. J. 94(5): 873–880. doi: 10.1094/CCHEM-02-17-0028-R.

Gualano, N.A., and R.L. Benech-Arnold. 2009. The effect of water and nitrogen availability during grain filling on the timing of dormancy release in malting barley crops. Euphytica 168(3): 291–301. doi: 10.1007/s10681-009-9948-x.

Gubler, F., A.A. Millar, and J. V. Jacobsen. 2005. Dormancy release, ABA and pre-harvest sprouting. Curr. Opin. Plant Biol. 8(2): 183–187. doi: 10.1016/j.pbi.2005.01.011.

Han, F., S.E. Ullrich, J.A. Clancy, and I. Romagosa. 1999. Inheritance and fine mapping of a major barley seed dormancy QTL. Plant Sci. 143(1): 113–118. doi: 10.1016/S0168-9452(99)00028-X.

Hasegawa, R., T. Tada, Y. Torii, and Y. Esashi. 1994. Presence of beta-cyanoalanine synthase in unimbibed dry seeds and its activation by ethylene during pre-germination. Physiol. Plant. 91(2): 141–146. doi: 10.1111/j.1399-3054.1994.tb00411.x.

Hickey, L.T., W. Lawson, V.N. Arief, G. Fox, J. Franckowiak, et al. 2012. Grain dormancy QTL identified in a doubled haploid barley population derived from two non-dormant parents. Euphytica 188(1): 113–122. doi: 10.1007/s10681-011-0577-9.

Hori, K., K. Sato, and K. Takeda. 2007. Detection of seed dormancy QTL in multiple mapping populations derived from crosses involving novel barley germplasm. Theor. Appl. Genet. 115(6): 869–876. doi: 10.1007/s00122-007-0620-3.

Ignacio, R., F. Han, J.A. Clancy, and S.E. Ullrich. 1999. Individual locus effects on dormancy during seed development and after ripening in barley. Crop Sci. 39(1): 74. doi: 10.2135/cropsci1999.0011183X003900010012x.

Jacobsen, J. V., D.W. Pearce, A.T. Poole, R.P. Pharis, and L.N. Mander. 2002. Abscisic acid, phaseic acid and gibberellin contents associated with dormancy and germination in barley. Physiol. Plant. 115(3): 428–441. doi: 10.1034/j.1399-3054.2002.1150313.x.

Jayakodi, M., S. Padmarasu, G. Haberer, V.S. Bonthala, H. Gundlach, et al. 2020. The barley pan-genome reveals the hidden legacy of mutation breeding. Nature 588(7837): 284–289. doi: 10.1038/s41586-020-2947-8.

Jia, Q., X.-Q. Zhang, S. Westcott, S. Broughton, M. Cakir, et al. 2011. Expression level of a gibberellin 20-oxidase gene is associated with multiple agronomic and quality traits in barley. Theor. Appl. Genet. 122(8): 1451–1460. doi: 10.1007/s00122-011-1544-5.

Juskiw, P.E., J.H. Helm, M. Oro, J.M. Nyachiro, and D.F. Salmon. 2009. Registration of 'Bentley' Barley. J. Plant Regist. 3(2): 119. doi: 10.3198/jpr2008.10.0631crc.

King, R.W., and P. von Wettstein-Knowles. 2000. Epicuticular waxes and regulation of ear wetting and pre-harvest sprouting in barley and wheat. Euphytica 112(2): 157–166. doi: 10.1023/A:1003832031695.

Kulwal, P., G. Ishikawa, D. Benscher, Z. Feng, L.X. Yu, et al. 2012. Association mapping for pre-harvest sprouting resistance in white winter wheat. Theor. Appl. Genet. 125(4): 793–805. doi: 10.1007/s00122-012-1872-0.

Legge, W.G., S. Haber, D.E. Harder, J.G. Menzies, J.S. Noll, et al. 2008. Newdale barley. Can. J. Plant Sci. 88(4): 717–723. doi: 10.4141/CJPS07194.

Legge, W.G., J.R. Tucker, T.G. Fetch, S. Haber, J.G. Menzies, et al. 2014. AAC Synergy barley. Can. J. Plant Sci. 94(4): 797–803. doi: 10.4141/cjps2013-307.

Leymarie, J., M.E. Robayo-Romero, E. Gendreau, R.L. Benech-Arnold, and F. Corbineau. 2008. Involvement of ABA in induction of secondary dormancy in barley (Hordeum vulgare L.) Seeds. Plant Cell Physiol. 49(12): 1830–1838. doi: 10.1093/pcp/pcn164.

Li, C., P. Ni, M. Francki, A. Hunter, Y. Zhang, et al. 2004. Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. Funct. Integr. Genomics 4(2): 84–93. doi: 10.1007/s10142-004-0104-3.

Li, X., X. Li, E. Fridman, T.T. Tesso, J. Yu, et al. 2015. Dissecting repulsion linkage in the dwarfing gene Dw3 region for sorghum plant height provides insights into heterosis. Proc. Natl. Acad. Sci. U. S. A. 112(38): 11823–11828. doi: 10.1073/pnas.1509229112.

Mao, X., J. Zhang, W. Liu, S. Yan, Q. Liu, et al. 2019. The MKKK62-MKK3-MAPK7/14 module negatively regulates seed dormancy in rice. Rice 12(1): 2. doi: 10.1186/s12284-018-0260-z.

Marquez-Cedillo, L.A., P.M. Hayes, B.L. Jones, A. Kleinhofs, W.G. Legge, et al. 2000. QTL analysis of malting quality in barley based on the doubled-haploid progeny of two elite North American varieties representing different germplasm groups. Theor. Appl. Genet. 101(1–2): 173–184. doi: 10.1007/s001220051466.

Millar, A.A., J. V. Jacobsen, J.J. Ross, C.A. Helliwell, A.T. Poole, et al. 2006. Seed dormancy and ABA metabolism in Arabidopsis and barley: the role of ABA 8′-hydroxylase. Plant J. 45(6): 942–954. doi: 10.1111/j.1365-313X.2006.02659.x.

Mohammadi, M., T.K. Blake, A.D. Budde, S. Chao, P.M. Hayes, et al. 2015. A genome-wide association study of malting quality across eight U.S. barley breeding programs. Theor. Appl. Genet. 128(4): 705–721. doi: 10.1007/s00122-015-2465-5.

Monat, C., S. Padmarasu, T. Lux, T. Wicker, H. Gundlach, et al. 2019. TRITEX: Chromosome-scale sequence assembly of Triticeae genomes with open-source tools. Genome Biol. 20(1): 284. doi: 10.1186/s13059-019-1899-5.

Morey, C., S. Mookherjee, G. Rajasekaran, and M. Bansal. 2011. DNA free energy-based promoter prediction and comparative analysis of Arabidopsis and rice genomes. Plant Physiol. 156(3): 1300–1315. doi: 10.1104/pp.110.167809.

Muir, C.E., and R.A. Nilan. 1973. Registration of Steptoe Barley 1 (Reg. No. 134). Crop Sci. 13(6): 770–770. doi: 10.2135/cropsci1973.0011183x001300060063x.

Nagel, M., A.M. Alqudah, M. Bailly, L. Rajjou, S. Pistrick, et al. 2019. Novel loci and a role for nitric oxide for seed dormancy and preharvest sprouting in barley. Plant. Cell Environ. doi: 10.1111/pce.13483.

Nakamura, S., M. Pourkheirandish, H. Morishige, Y. Kubo, M. Nakamura, et al. 2016. Mitogen-activated protein kinase kinase 3 regulates seed dormancy in barley. Curr. Biol. 26(6): 775–781. doi: 10.1016/J.CUB.2016.01.024.

Naranjo, T., A Roca, P.G. Goicoechea, and D.R. Giraldez. 1987. Arm homoeology of wheat and rye chromosomes. Genome, 29: 873-882.

Neyhart, J.L., D. Sweeney, M. Sorrells, C. Kapp, K.D. Kephart, et al. 2019. Registration of the S2MET barley mapping population for multi-environment genomewide selection. J. Plant Regist. 13(2). doi: 10.3198/jpr2018.06.0037crmp.

Oberthur, L., T.K. Blake, W.E. Dyer, and S.E. Ullrich. 1995. Genetic analysis of seed dormancy in barley (Hordeum vulgare L.). J. Agric. Genomics 1: 1–10. https://www.cabdirect.org/cabdirect/abstract/20063161122.

Peng, J., D.E. Richards, N.M. Hartley, G.P. Murphy, K.M. Devos, et al. 1999. "Green revolution" genes encode mutant gibberellin response modulators. Nature 400(6741): 256–261. doi: 10.1038/22307.

Prada, D., S.E. Ullrich, J.L. Molina-Cano, L. Cistué, J.A. Clancy, et al. 2004. Genetic control of dormancy in a Triumph/Morex cross in barley. Theor. Appl. Genet. 109(1): 62–70. doi: 10.1007/s00122-004-1608-x.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Radchuk, V. V, L. Borisjuk, N. Sreenivasulu, K. Merx, H.-P. Mock, et al. 2009. Spatiotemporal profiling of starch biosynthesis and degradation in the developing barley grain. doi: 10.1104/pp.108.133520.

Rasmusson, D.C., and R.W. Wilcoxson. 1979. Registration of Morex barley (Reg. No. 158). Crop Sci. 19(2): 293–293. doi: 10.2135/cropsci1979.0011183x001900020032x.

Rodríguez, M.V., M. Margineda, J.F. González-Martín, P. Insausti, and R.L. Benech-Arnold. 2001. Predicting preharvest sprouting susceptibility in barley. Agron. J. 93(5): 1071–1079. doi: 10.2134/agronj2001.9351071x.

Sato, K., M. Yamane, N. Yamaji, H. Kanamori, A. Tagiri, et al. 2016. Alanine aminotransferase controls seed dormancy in barley. Nat. Commun. 7(1): 11625. doi: 10.1038/ncomms11625.

Segura, V., B.J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, et al. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. 44(7): 825–830. doi: 10.1038/ng.2314.

Shorinola, O., B. Balcárková, J. Hyles, J.F.G. Tibbits, M.J. Hayden, et al. 2017. Haplotype analysis of the pre-harvest sprouting resistance locus Phs-A1 reveals a causal role of TaMKK3-A in global germplasm. Front. Plant Sci. 8: 1555. doi: 10.3389/fpls.2017.01555.

Shu, X., and S.K. Rasmussen. 2014. Quantification of amylose, amylopectin, and B-glucan in search for genes controlling the three major quality traits in barley by genome-wide association studies. Front. Plant Sci. 5: 197. doi: 10.3389/fpls.2014.00197.

Spielmeyer, W., M. Ellis, M. Robertson, S. Ali, J.R. Lenton, et al. 2004. Isolation of gibberellin metabolic pathway genes from barley and comparative mapping in barley, wheat and rice. Theor. Appl. Genet. 109(4): 847–855. doi: 10.1007/s00122-004-1689-6.

Sweeney, D.W., J. Rutkoski, G.C. Bergstrom, and M.E. Sorrells. 2020. A connected half-sib family training population for genomic prediction in barley. Crop Sci. 60(1): 262–281. doi: 10.1002/csc2.20104.

Tibbs Cortes, L., Z. Zhang, and J. Yu. 2021. Status and prospects of genome-wide association studies in plants. Plant Genome. doi: 10.1002/tpg2.20077.

Toora, P.K. 2018. Genetic and molecular bases of abscisic acid and gibberellin metabolism in regulating seed dormancy in malting barley.

Torada, A., M. Koike, S. Ikeguchi, and I. Tsutsui. 2008. Mapping of a major locus controlling seed dormancy using backcrossed progenies in wheat (*Triticum aestivum* L.) (D. Somers, editor). Genome 51(6): 426–432. doi: 10.1139/G08-007.

Torada, A., M. Koike, T. Ogawa, Y. Takenouchi, K. Tadamura, et al. 2016. A causal gene for seed dormancy on wheat chromosome 4A encodes a MAP kinase kinase. Curr. Biol. 26: 782–787. doi: 10.1016/j.cub.2016.01.063.

Ullrich, S.E., H. Lee, J.A. Clancy, I.A. del Blanco, V.A. Jitkov, et al. 2009. Genetic relationships between preharvest sprouting and dormancy in barley. Euphytica 168(3): 331–345. doi: 10.1007/s10681-009-9936-1.

Vetch, J.M., J.G. Walling, J. Sherman, J.M. Martin, and M.J. Giroux. 2020. Mutations in the HvMKK3 and HvAlaAT1 genes affect barley preharvest sprouting and after-ripened seed dormancy. Crop Sci. doi: 10.1002/csc2.20178.

Walkowiak, S., L. Gao, C. Monat, G. Haberer, M.T. Kassa, et al. 2020. Multiple wheat genomes reveal global variation in modern breeding. Nature 588(7837): 277–283. doi: 10.1038/s41586-020-2961-x.

Wang, Y., L. Zhai, K. Chen, C. Shen, Y. Liang, et al. 2020. Natural sequence variations and combinations of GNP1 and NAL1 determine the grain number per panicle in rice. Rice 13(1): 14. doi: 10.1186/s12284-020-00374-8.

Wei, W., X. Min, S. Shan, H. Jiang, J. Cao, et al. 2019. Isolation and characterization of TaQsd1 genes for period of dormancy in common wheat (Triticum aestivum L.). doi: 10.1007/s11032-019-1060-x.

Woonton, B.W., J. V Jacobsen, F. Sherkat, and I.M. Stuart. 2005. Changes in germination and malting quality during storage of barley. J. Inst. Brew 33-41, 111(1)

Xu, Q., T.T. Truong, J.M. Barrero, J. V. Jacobsen, C.H. Hocart, et al. 2016. A role for jasmonates in the release of dormancy by cold stratification in wheat. J. Exp. Bot. 67(11): 3497–3508. doi: 10.1093/jxb/erw172.

Xu, Y., Q. Jia, G. Zhou, X.-Q. Zhang, T. Angessa, et al. 2017. Characterization of the sdw1 semi-dwarf gene in barley. BMC Plant Biol. 17(1): 11. doi: 10.1186/s12870-016-0964-4.

Ye, H., J. Feng, L. Zhang, J. Zhang, M.S. Mispan, et al. 2015. Map-based cloning of Seed Dormancy1-2 identified a gibberellin synthesis gene regulating the development of endosperm-imposed dormancy in rice. Plant Physiol. 169: 2152–2165. doi: 10.1104/pp.15.01202.

Zheng, X., D. Levine, J. Shen, S.M. Gogarten, C. Laurie, et al. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28(24): 3326–3328. doi: 10.1093/bioinformatics/bts606.

Supplementary material

Table 3.6: Primer sequences for KASP markers

Marker	Chr	Position	Gene	Allele-1 forward primer	Allele-2 forward primer	Universal reverse primer	**Non-dormant**/dormant allele
AlaAT_L214F	5H	442161820	*HvAlaAT1*	GATTTTCGAAGTAAAGAGGTGCTT**G**	GATTTTCGAAGTAAAGAGGTGCTT**C**	CACGAACAGTCAAACCTGCG	**G**/C
GA20ox1_331_5UTR	5H	595056546	*HvGA20ox1*	TCGAGTGTGGGCGTTTGC**G**	TCGAGTGTGGGCGTTTGC**C**	ACCATGATCGCCTCCACCAA	**G**/C
JHI-Hv50k-2016-367342-KASP	5H	596729543	*HvMKK3*	GTGATTCCTCGCTGCTTGGT**A**	GTGATTCCTCGCTGCTTGGT**G**	AGTGAGTAATAATGAGCCCAGCC	**A**/G
MKK3_E165Q	5H	596732030	*HvMKK3*	GGACAAATAAGCATTGCCCTT**G**	GGACAAATAAGCATTGCCCTT**C**	GCAAGAACCGGCTCTGGTAT	**A**/T

Table 3.7: Sanger sequencing template synthesis primers

Gene	**Exon**	**Primer name**	**Nucleotide sequence (5' → 3')**	**Primer length (bp)**	**Tm (°C)**	**Amplicon size (bp)**
HvAlaAT1	1 - 15	Qsd1 F	AGGTCCCAAGAAGATGGCATT	21	56.5	~ 2,271
		Qsd1 R	CATGAAGGACTCGTGGAAGGC	21	58.0	
HvGA20ox1	1, 2	V2 GAox1 F1	GCACACCACACACAGAGAGAAA	22	61.0	1605
		V2 GAox1 R1	TCCATGCATCGCTACCACTCTA	22	61.0	
HvMKK3	1 - 9	MKK_UTRexon2 F1	CAACCGAAGAGACCGTGATT	20	57.0	> 8,181
		MKK_UTRexon12 R1	TGTGATCTGAAGCCTGTCCA	20	58.0	
	1 - 9	MKK_UTRexon2 Nested F2	GAAGAGACCGTGATTCCTCG	20	57.0	8,181
		MKK_UTRexon12 Nested R2	GGCGTGGAATCATTTAAGGA	20	55.0	

Table 3.8: Sanger sequencing primers

Gene	Target	Direction	Nucleotide sequence (5' → 3')	Primer length (bp)	Tm (°C)
HvAlaAT1	Exons 9, 10	F	TCCATTATACTCGGCGTCCATT	22	55.9
		R	ATCTCCATGTAGCCTCCCCTT	21	57.6
HvGA20ox1	Exons 1, 2	F	CGGAGCTCGATCGTCTGTTG	20	60.9
		R	TTCTCACTCCCTGCACTGCA	20	61.4
	Exons 1, 2	F	GCACACCACACACAGAGAGAAA	22	61.0
		R	TCCATGCATCGCTACCACTCTA	22	61.0
HvMKK3	Exons 3, 4	F	GTGTTTGGTATTGCACGGAGTAA	23	59.0
		R	AGTCCACGCATTTTACATCAGACT	24	60.0
	Exon 7	F	TCTTTGGTCCCTTTCCATTTT	21	59.8
		R	GTTGGCAGAGGCAAAAGAAG	21	60.0
	Exon 8	F	TCATTAAAGATGGGCAGGGTGC	22	61.0
		R	TGATCGACGGCATTAAGAGCA	21	59.0

CHAPTER 4

QTL X ENVIRONMENT MODELING OF SPRING AND WINTER TWO-ROW MALTING BARLEY PREHARVEST SPROUTING

Abstract

Preharvest sprouting (PHS) can severely damage barley (*Hordeum vulgare* L.) malting quality but selection for PHS resistance often results in poor malting quality. Seed dormancy is closely related to PHS and increased temperature during grain fill can result in decreased seed dormancy in barley. Several large effect seed dormancy quantitative trait loci (QTL) have been identified in barley, but genetic components of seed dormancy temperature sensitivity are poorly understood. Six years of PHS data from Cornell University field trials were used to fit QTL x environment mixed models incorporating haplotypes of seed dormancy genes *HvAlaAT1*, *HvGA20ox1,* and *HvMKK3* and weather covariates in spring and winter two-row malting barley. Variation in winter barley PHS was best modeled by average temperature range during grain fill and spring barley PHS by total precipitation during grain fill. Average high temperature during grain fill also accurately modeled PHS for both datasets. A highly non-dormant *HvMKK3* allele determined baseline PHS susceptibility and *HvAlaAT1* interactions with multiple *HvMKK3* alleles conferred environmental sensitivity. Polygenic variation for PHS within haplotype was detected. Haplotype responses to environmental covariates were similar across winter and spring datasets but winter entries showed more extreme high and low average PHS. Residual genotype and QTL by environment interaction variance indicated additional environmental and genetic factors involved in PHS control. Entries with dormant *HvMKK3* were most stable

across environments. These models provide insight into genotype and environmental regulation of barley seed dormancy, a method for PHS forecasting, and a tool for breeders to improve PHS resistance.

Introduction

Seed dormancy, the inability of a seed to germinate under conducive environmental conditions (DePauw and McCaig, 1991; Bewley et al., 2013), has many physiological determinants but the ratio of and sensitivity to abscisic acid (ABA) and gibberellin (GA) hormones are particularly important (Finch-Savage & Leubner-Metzger, 2006). ABA induces and maintains seed dormancy and GA promotes germination. Seed dormancy in barley is primarily imposed by seed covering structures, namely the glumellae, pericarp, and endosperm (Lenoir et al., 1986; Benech-Arnold et al., 1999). Excised embryos readily germinate as soon as 22 days after pollination but retain sensitivity to exogenous ABA through maturity (Benech-Arnold et al., 1999). Dormancy is not determined by absolute ABA content or GA content (Jacobsen et al., 2002); rather, embryo sensitivity to ABA is correlated with dormancy imposition and maintenance (Van Beckum et al., 1993; Benech-Arnold et al., 1999; 2006; Bradford et al., 2008). Physiological regulation of seed dormancy is also sensitive to environmental factors, with temperature playing an especially important role. Increased temperature during grain fill has been linked to reduced seed dormancy in wheat (*Triticum aestivum* L.) (Biddulph et al., 2005; 2007; Nakamura et al., 2011) and barley (Rodriguez et al., 2001; Li et al., 2004; Gualano & Benech-Arnold, 2009; Gong et al., 2014). In barley, maximum ABA content is reached in the middle of seed

development and decreases as physiological maturity approaches (Goldbach & Michael, 1976; Benech-Arnold et al., 1999) with similar results observed in wheat (King, 1976). Timing of peak ABA accumulation and total ABA content at maturity are known to be temperature sensitive in barley (Goldbach & Michael, 1976; Walker-Simmons and Sesing, 1990; Chono et al., 2006). In rice (*Oryza sativa* L.), GA accumulation peaks before ABA and also decreases as physiological maturity approaches (Liu et al., 2014) but little is known about GA accumulation or temperature sensitivity during grain fill in barley.

Preharvest sprouting occurs when excess moisture prematurely initiates seed germination, either visibly or non-visibly, in the field before harvest and can severely reduce barley malting quality. This phenomenon is related to seed dormancy as varieties that quickly lose primary dormancy are at high risk for PHS damage. A number of QTL have been mapped for PHS in barley but the *SD1* and *SD2* seed dormancy loci on chromosome 5H are the most consistent and largest effect QTL (Oberthur et al., 1995; Han et al., 1996; Hori et al., 2007; Lin et al., 2009). An alanine aminotransferase, *HvAlaAT1*, has been cloned at *SD1* (Sato et al., 2016) and a mitogen associated protein kinase kinase (*HvMKK3*) has been cloned in the *SD2* region (Nakamura et al., 2016). AlaAT is likely involved in ABA signaling but the exact mechanism is unknown (Wei et al., 2019) and in rice, MKK3 is part of an ABA signaling pathway regulating expression of *MOTHER OF FLOWERING TIME* (*OsMFT*) (Mao et al., 2019). In modern North American spring malting barley, three *HvMKK3* variants are associated with PHS (Vetch et al., 2020; Chapter 3). Gibberellin 20-oxidases (GA20ox) are important enzymes in bioactive gibberellin

production (Spielmeyer et al., 2004). GA20ox1 is a key part of active gibberellin synthesis during seed germination and is highly expressed in late stages of grain development and during germination (Sreenivasulu et al., 2008; Liu et al., 2014; Betts et al., 2019). *HvGA20ox1* is in the *SD2* region 1.7 Mb proximal to *HvMKK3* and its role in PHS has been suggested in barley (Li et al., 2004; Nagel et al., 2018) and sorghum (Perez-Flores et al., 2003; Rodriguez et al., 2012) and germination rate in rice (Abe et al., 2012) but association analysis in three spring barley populations indicated that *HvGA20ox1* does not have main or epistatic effects on PHS or seed dormancy (Chapter 3). Genetic evidence supporting the correlation between increased temperature and PHS susceptibility has been reported in barley. *SD2*-associated dormancy in Steptoe/Morex increased in environments with lower average temperatures during grain fill (Gao et al., 2003). Germination percentage QTL in the *SD1* and *SD2* regions were detected in a Stirling/Harrington doubled haploid population with variable effects across environments differentiated by temperature (Gong et al., 2014). Increased temperature during embryo development in the twenty days after flowering was correlated with reduced seed dormancy.

Phenotypic stability is one of the main objectives in crop genotype by environment interaction studies. Often, breeders are interested in identifying lines that perform similarly across diverse environments rather than optimizing performance in a subset of environments. QTL by environment interaction (QEI) models are a special class of genotype by environment interaction models that incorporate known QTL information to estimate change in QTL effects across environment. Genetic effects are partitioned into main QTL effects and residual polygenic effects and genotype by

environment interaction (GEI) effects into QEI effects and residual polygenic GEI effects (Malosetti et al., 2004). Variation in QTL effects can be further defined by introducing environmental covariates such as temperature or precipitation into a QEI mixed model framework (van Eeuwijk et al., 2005). Malosetti et al. (2004) estimated a QTL main effect, environment specific effects, and QTL sensitivity to temperature during heading for a grain yield QTL on chromosome 2H in barley. PHS susceptibility in South American barley varieties has been modeled as a function of temperature within grain fill intervals defined by thermal time, revealing strong linear relationships between increased temperature in late grain fill stages and PHS (Rodriguez et al., 2001; Gaulano and Benech-Arnold, 2009), but a QEI framework has not been implemented for modeling barley PHS. Temperature sensitivity of *HvAlaAT1* and *HvMKK3* have not been formally described and temperature sensitivity models are not available for European and North American malting barley germplasm. Although *HvGA20ox1* is not associated with overall PHS, expression could be sensitive to environmental factors. Spring malting barley tends to be more susceptible to PHS than winter malting barley but is it unknown if this difference is genetic, environmental, or a combination of both. Preharvest sprouting is often negatively correlated with malting quality so additional genetic sources of PHS resistance could be useful for combining high malting quality potential with PHS resistance. Modeling the response of *HvAlaAT1*, *HvMKK3,* and *HvGA20ox1* to environmental covariates could identify climate stable PHS resistance and reveal residual polygenic variation for PHS resistance or environmental sensitivity. Improved understanding of genetic and environmental causes of PHS variability for commonly grown varieties would also

help growers adjust management strategies to minimize risk from PHS-inducing weather.

The objective of this study was to use seed dormancy haplotype data and weather covariates to model genotypic and environmental influences on annual variation in PHS using historical data from 61 two-row spring malting barley entries in twenty-five environments over six years and 21 two-row winter malting barley entries in nine environments over five years. Three facultative entries planted in both winter and spring were also analyzed. The seed dormancy haplotypes of these entries were classified using Kompetitive Allele Specific Primers (KASP) assays for *HvAlaAT1*, *HvGA20ox1,* and *HvMKK3.* Five temperature and precipitation covariates were tested in QEI models to estimate seed dormancy QTL by environmental covariate interaction effects, residual polygenic effects, and residual QEI and GEI effects.

Materials and methods

Data curation and phenotyping

Two datasets, 1) seven spring barley experiments planted across six years and 2) three winter barley experiments planted across five years, were used for analysis. Spring and winter datasets were always analyzed separately. These experiments were highly unbalanced across years and included both commercial variety yield trials, cooperative regional nurseries, and preliminary breeding trials (Table 4.1). A total of 61 spring entries and 21 winter entries that represented the range of PHS scores in each dataset and were present in multiple experiments across years either as checks or experimental entries were used for analysis (Table 4.2). The entries were all modern European or

North American two-row malting barleys, with the exception of the NakedReg experiment composed of spring hulless, or naked, barley entries and a hulled check. Winter 2015 data was removed due to collection and sample preparation discrepancies. Environments were defined as year-location combinations. Within year, different experiments planted at the same location on the same day with shared checks were grouped in the same environment. For each experiment, heading date was recorded as the Julian date that 50% of the spikes in a plot had fully emerged from the flag leaf. Heat, drought, and semi-dwarfism can cause heading date to be delayed in spring barley, making accurate measurement of the grain fill period difficult. To improve estimates of the grain fill period, in 2019 and 2020 spring barley experiments, tipping date was recorded as the date that 50% of the plot showed awn protrusion of 2 mm. Anthesis in spring barley occurs at this stage (Zadoks scale Z49), not at heading date as in winter barley (Alqudah & Schnurbusch, 2017). The period between heading date and spike sampling for winter barley and the period between tipping date and spike sampling for spring barley were defined

Table 4.1: Summary of preharvest sprouting experiments. Experiments consisted of replicated yield trial plots and early generation meter long headrow nurseries.

Dataset	Experiment	Years	Type
Spring	SMBReg	2015-2020	yield
Spring	ESBN	2015-2020	yield
Spring	S2MET	2015-2016	yield
Spring	CU_TP	2017	headrow
Spring	CUReg	2018-2019	yield
Spring	CU1	2019-2020	headrow
Spring	NakedReg	2018-2020	yield
Winter	WMBReg	2016-2020	yield
Winter	WMBCoop	2016-2020	yield
Winter	OSUmalt	2016-2018	yield

as the grain fill period. The average difference between tipping date and heading date in 2019 and 2020 spring experiments, 9 days, was added to the heading date of spring samples collected from 2015-2018 to estimate tipping date. In all experiments, five spikes per plot were sampled at physiological maturity (defined as the loss of green color from the peduncle), after-ripened at ambient temperature for three days, misted in an artificial greenhouse mist chamber, and scored on a 0-9 scale as described by Anderson et al. (1993) with 0 indicating no visible PHS and 9 indicating advanced radicle and coleoptile development on all caryopses on the spike. Winter barley spikes were after-ripened for four days in 2019 and 2020 experiments to increase variation for selection. Preharvest sprouting scores of 0-2 can be classified as resistant, 3-5 as susceptible, and 6-9 as highly susceptible.

Daily high, low, and average temperature and daily precipitation were downloaded from the Northeast Regional Climate Center station (http://www.nrcc.cornell.edu/wxstation/ithaca/ithaca.html) at the Game Farm Road station, which is centrally located between the testing locations for all trials (average distance of 1100 m). For each plot in each environment, the environmental covariates of average high temperature during grain fill (T_{max}), average low temperature during grain fill (T_{min}), average temperature during grain fill (T_{avg}), average difference between high and low temperature during grain fill (T_r), and total precipitation during grain fill (P_{sum}) were calculated (Table 4.3).

Table 4.2: Summary of genotyped entries in spring and winter datasets. CU entries signify Cornell University experimental entries; OSU, Oregon State University; MSU, Montana State University; and NDSU, North Dakota State University. All entries are two-row malting barleys except for CDC_CLEAR, DH133529, and DH133535, which are two-row spring naked barleys. Three entries, LIGHTNING, DH130935, and DH131055, exhibit a facultative growth habit (labeled as F) and were evaluated in spring and winter experiments. The number of preharvest sprouting (PHS) observations on a per plot basis for each entry is indicated by n.

Dataset	Entry	Source	Haplotype	n	PHS μ	PHS σ_p^2
Spring	AC_METCALFE	commercial	NNN*	15	5.43	2.08
Spring	BENTLEY	commercial	NNN*	48	5.19	2.39
Spring	CDC_CLEAR	commercial	NNN*	14	6.84	1.35
Spring	CDC_COPELAND	commercial	NNN*	12	5.93	2.73
Spring	FULL_PINT	commercial	NNN*	20	5.71	2.34
Spring	SG293-3	CU	NNN*	8	6.96	0.75
Spring	SG542-1	CU	NNN*	8	6.58	0.83
Spring	SP362R-2	CU	NNN*	8	6.63	0.58
Spring	08MT-03	MSU	NDD	14	2.03	2.04
Spring	09N2-65	NDSU	NDD	14	3.14	0.90
Spring	09N2-96	NDSU	NDN	14	5.22	3.46
Spring	2ND32529	NDSU	NDD	15	5.44	1.89
Spring	ACCORDINE	commercial	NDD	15	1.99	0.85
Spring	CONLON	commercial	NDD	83	2.44	2.19
Spring	CRAFT	commercial	NDD	70	1.10	1.07
Spring (F)	DH130935	OSU	NDD	12	2.14	1.28
Spring (F)	DH131055	OSU	NDD	6	3.34	0.52
Spring	DH133529	OSU	NDD	14	1.46	2.53
Spring	DH133535	OSU	NDD	14	1.39	3.10
Spring	ESMA	commercial	NDD	29	0.91	0.46
Spring	EXPLORER	commercial	NDD	32	0.32	0.44
Spring	SC132-11	commercial	NDD	15	0.84	0.76
Spring	KWS_CHRISSIE	commercial	NDD	15	2.04	1.52
Spring	KWS_JESSIE	commercial	NDD	18	1.55	1.31
Spring	KWS_TINKA	commercial	NDD	102	2.41	1.22
Spring	SANGRIA	commercial	NDD	26	1.09	1.37
Spring	ST1453-4	CU	NDD	16	2.63	1.95
Spring	09N2-16	NDSU	NDN	9	4.68	3.04
Spring	09N2-84	NDSU	NDN	8	4.51	4.34
Spring	EIFEL	commercial	NDN	18	1.76	1.00
Spring	LCS_GENIE	commercial	NDN	29	1.43	1.67

Spring	ND_GENESIS	commercial	NDN	116	3.43	2.07
Spring	PINNACLE	commercial	NDN	113	1.55	2.93
Spring	SG5123-1	CU	NDN	17	2.59	3.43
Spring	SP333R-1	CU	NDN	14	3.18	2.86
Spring	SP572-3	CU	NDN	13	2.01	1.28
Spring	SP575-1	CU	NDN	10	2.80	1.13
Spring	09N2-67	NDSU	NND	33	2.64	2.95
Spring	09N2-68	NDSU	NND	14	3.31	1.27
Spring	AAC_CONNECT	commercial	DNN*	9	6.44	1.69
Spring	AAC_SYNERGY	commercial	DNN*	217	5.10	1.41
Spring	NEWDALE	commercial	DNN*	94	4.61	2.45
Spring	SB182R-3	CU	DNN*	8	5.89	2.29
Spring	SN391R-1	CU	DNN*	14	4.91	2.76
Spring	SR556-3	CU	DNN*	10	4.62	2.96
Spring	SR591R-3	CU	DNN*	14	4.70	2.45
Spring (F)	LIGHTNING	commercial	DDD	26	1.44	0.97
Spring	SC662R-1	CU	DDD	12	3.22	3.63
Spring	SR575-2	CU	DDD	10	1.86	2.98
Spring	SR6122R-1	CU	DDD	14	1.93	0.92
Spring	SR632R-3	CU	DDD	12	0.82	1.01
Spring	ST1442R-2	CU	DDD	12	1.54	1.33
Spring	ST1453-2	CU	DDD	14	1.65	0.60
Spring	ST1481R-1	CU	DDD	15	2.91	1.35
Spring	SC742R-2	CU	DDN	15	1.60	1.70
Spring	SC795-4	CU	DDN	16	2.03	2.34
Spring	SG223R-1	CU	DDN	12	3.40	3.27
Spring	SG231R-1	CU	DDN	12	2.81	1.52
Spring	SP323R-2	CU	DDN	12	2.79	3.00
Spring	SP382R-3	CU	DDN	10	2.24	3.49
Spring	SP585-2	CU	DDN	13	1.45	1.41
Winter	DH130004	OSU	NNN*	12	6.47	3.62
Winter	DH130718	OSU	NNN*	21	6.46	2.47
Winter	DH130939	OSU	NNN*	24	6.61	2.90
Winter	DH140088	OSU	NNN*	30	5.84	3.64
Winter	THUNDER	commercial	NNN*	16	6.78	3.74
Winter (F)	DH131055	OSU	NDD	18	0.00	0.00
Winter	DH131738	OSU	NDD	20	0.28	0.15
Winter (F)	DH130935	OSU	NDD	21	1.07	0.81
Winter	KWS_DONAU	commercial	NDD	30	1.40	1.60
Winter	KWS_SCALA	commercial	NDD	30	0.4	0.34

Winter	KWS_SOMERSET	commercial	NDD	30	0.28	0.18
Winter	NECTARIA	commercial	NDD	21	1.29	0.87
Winter	SY_TEPEE	commercial	NDD	27	0.19	0.08
Winter	VIOLETTA	commercial	NDD	18	0.08	0.02
Winter	WINTMALT	commercial	NDD	26	1.46	1.29
Winter	CHARLES	commercial	DNN*	51	4.66	7.1
Winter	ENDEAVOR	commercial	DNN*	48	6.94	2.39
Winter	CALYPSO	commercial	DDD	24	0.28	0.25
Winter (F)	LIGHTNING	commercial	DDD	42	0.00	0.00
Winter	FLAVIA	commercial	DDD	39	0.88	0.88
Winter	SU_MATEO	experimental	DDD	24	0.77	0.92

Table 4.3: Environmental summaries with planting date, average heading date, average daily high temperature (T_{max}) (C), average daily low temperature (T_{min}), average daily temperature (T_{avg}), and total precipitation (P_{sum}) (cm) during the grain fill period.

Dataset	Environment	Planting date	Average heading date	T_{max}	T_{avg}	T_{min}	P_{sum}
Spring	Ketola2015	5/4/15	6/30/15	25.3	19.1	13.4	12.4
Spring	Snyder2015	4/27/15	6/19/15	24.8	19.0	13.1	16.5
Spring	S2MET_KT2015	5/5/15	6/29/15	25.0	18.9	13.2	12.2
Spring	Helfer2016	4/27/16	7/2/16	29.1	21.2	14.4	3.8
Spring	Ketola2016	4/22/16	6/25/16	29.5	20.4	14.2	2.8
Spring	S2MET_Hel2016	4//16	6/29/16	28.4	20.9	14.1	3.6
Spring	S2MET_Ket2016	4/28/16	6/25/16	28.3	20.6	14.0	2.8
Spring	Helfer2017	4/27/17	6/23/17	27.2	20.2	14.2	14.0
Spring	Ketola2017	4/23/17	6/24/17	27.5	20.3	14.4	14.5
Spring	CU_TP_Hel2017	5/12/17	7/9/17	25.9	20.3	15.0	15.0
Spring	CU_TP_Sny2017	5/12/17	7/12/17	26.1	20.4	14.7	13.2
Spring	Caldwell2018	4/30/18	6/22/18	27.2	19.6	13.1	4.6
Spring	Helfer2018	5/7/18	6/29/18	27.3	20.3	13.4	9.1
Spring	Ketola2018	5/6/18	6/28/18	27.3	20.3	13.4	7.4
Spring	Naked_Freeville2018	5/8/18	7/1/18	27.3	20.6	13.8	10.9
Spring	Ketola2019	4/17/19	6/22/19	27.3	20.2	14.7	8.1
Spring	Snyder2019	4/9/19	6/17/19	27.4	19.9	14.3	12.9
Spring	CU1_Caldwell2019	4/12/19	6/24/19	27.2	20.1	14.4	5.6
Spring	CU1_Ketola2019	4/12/19	6/28/19	27.6	20.6	14.8	3.8
Spring	Naked_Caldwell2019	4/23/19	6/25/19	26.9	20.8	14.7	12.4
Spring	Naked_Freeville2019	4/22/19	6/22/19	26.8	20.7	14.6	13.2

Spring	Helfer2020	4/6/20	6/20/20	27.4	20.7	14.0	11.2
Spring	CU1_Caldwell2020	4/22/20	7/1/20	28.4	22.1	15.7	12.2
Spring	CU1_Helfer2020	4/15/20	6/28/20	28.2	21.7	15.2	12.2
Spring	Naked_Freeville2020	5/4/20	6/24/20	27.9	21.4	14.8	11.4
Winter	Ketola2016	10/5/15	5/23/16	25.2	17.8	10.5	3.6
Winter	Snyder2016	9/23/15	5/20/16	24.7	17.2	10.0	3.6
Winter	Ketola2017	10/6/16	5/24/17	23.3	17.4	11.4	12.4
Winter	McGowan2017	9/28/16	5/20/17	23.2	17.1	11.2	10.9
Winter	Ketola2018	10/3/17	5/25/18	24.4	17.8	11.4	4.8
Winter	Snyder2018	10/2/17	5/27/18	24.5	18.1	11.7	6.1
Winter	Snyder2019	10/10/19	5/31/19	23.7	17.6	11.4	15.7
Winter	Ketola2020	9/27/19	5/25/19	25.2	18.4	11.7	3.6
Winter	Snyder2020	9/26/19	5/26/19	25.4	18.7	11.9	6.4

Statistical analysis

All statistical analyses were performed with R version 3.5.1 (R Core Team, 2018). Within each experiment/environment combination, single spike scores within entry were aggregated and outlier spike scores with a z-score greater than two were removed. Spike scores were then averaged per plot within experiment/environment. All experiments were randomized complete block designs except S2MET and CU1, which were planted in an augmented design with replicated checks. Replication was not significant for either dataset and was not included in the models. Single step QTL by environment models were fit with *ASReml-R* (Butler et al., 2009) for each environmental covariate within each dataset as

$$y_{ijk} \sim \mu + E_j + Q_i + Q_i{:}Z_j + G_k + G_k{:}Z_j + G_k{:}E_j + Q_i{:}E_j + E_j(sample_l) + \varepsilon_{ijk} \quad (4.1)$$

where y_{ijk} is the phenotypic observation for PHS, μ is the overall mean, E_j is the fixed effect for environment, Q_i is the fixed effect for seed dormancy haplotype with seven

levels for spring entries and four levels for winter entries (Table 4.2), Z_j is the fixed environmental covariate, and G_k is the random effect of entry with $G_k \sim N(0, \sigma_g^2)$ where σ_g^2 is the genetic variance. QEI was partitioned into fixed QTL by environmental covariate interaction (Q_iZ_j) and residual QTL by environment interaction (Q_iE_j) effects with $Q_iE_j \sim N(0, \sigma_{QE}^2)$ where σ_{QE}^2 is the QTL by environment interaction variance. GEI was similarly partitioned as entry by environmental covariate interaction, $G_kZ_j \sim N(0, \sigma_{GZ}^2)$ where σ_{GZ}^2 is the entry by environmental covariate interaction variance, and entry by environment interaction, $G_kE_j \sim N(0, \sigma_{GE}^2)$ where σ_{GE}^2 is the entry by environment interaction variance (Malosetti et al., 2004). Spike sampling date was nested within environment and specified as $E_j(sample_l) \sim N(0, \sigma_{E(s)}^2)$ where $\sigma_{E(s)}^2$ is the spike sampling date variance. Heterogeneous residuals were fit to estimate a unique residual variance in each environment with $e_{ijk} \sim N(0, \sigma_e^2)$ where σ_e^2 is the residual error variance. All phenotypic observations were included in the models to accurately estimate environmental effects. A dummy variable was added to differentiate observations with and without haplotype data. For each G_k main or interaction effect, the *ASReml-R* function *at()* was used to estimate separate variance components for entries with and without haplotype data ('at(dummy, c(1,2)):Entry' where 1 indicates lines without a haplotype and 2 indicates lines with a haplotype). Variance components for entries without haplotype data were not included in the results. A likelihood ratio test was used to test if variance components were significantly different (Pr(Chisq) < 0.05) from zero. Fixed effect significance for E_j, Q_i, and Q_iZ_j terms was tested with Wald tests. To test if individual Q_iZ_j slopes were significant (different from zero), the

Kenward-Roger degrees of freedom were extracted from the wald() object and used to determine a p-value using the t-distribution. T values were calculated as the fixed effect estimate divided by the standard error. Best model fit was determined by Akaike information criterion (AIC) estimated using icREML to compare models with different fixed environmental covariates (Verbyla, 2019). Briefly, the icREML method decomposes full likelihood into residual and conditional likelihood using restricted maximum likelihood (REML) and derives AIC from the full log-likelihood.

Kompetitive Allele Specific Primer (KASP) markers were used to genotype *HvAlaAT1, HvGA20ox1,* and *HvMKK3*. Primer design and KASP methods are detailed in the Methods section of Chapter 3. In brief, KASP marker AlaAT_L214F amplified the causal L214F polymorphism in *HvAlaAT1* described by Sato et al. (2016), GA20ox1_331_5UTR amplified a SNP in the 5' UTR of *HvGA20ox1* (Chapter 3), MKK3_E165Q amplified the E165Q mutation in *HvMKK3* exon 3 identified by Vetch et al. (2020), and JHI-367342-KASP is highly correlated with the R350G *HvMKK3* mutation from Vetch et al. (2020). For AC Metcalfe and CDC Copeland, AlaAT_L214F, MKK3_E165Q, and JHI-367342-KASP data were based on Vetch et al. (2020) and GA20ox1_331_5UTR data was taken from Sanger sequencing results (Chapter 3). Haplotypes were defined as the combination of alleles at *HvAlaAT1, HvGA20ox1,* and *HvMKK3*, respectively, with "N" indicating a non-dormant allele and "D" indicating a dormant allele. Haplotype order was based on the physical position of the three genes on chromosome 5H. The dormant *HvMKK3* allele is dormant at both MKK3_E165Q and JHI-367342-KASP, the non-dormant allele is dormant at MKK3_E165Q and non-dormant at JHI-367342-KASP, and a third, highly

non-dormant allele is non-dormant at MKK3_E165Q and JHI-367342-KASP. This third *HvMKK3* allele was labeled "N*". Thus, an NDN haplotype indicates non-dormant *HvAlaAT1*, putatively dormant *HvGA20ox1*, and non-dormant *HvMKK3*. NNN*, DNN*, NND, NDN, NDD, DDD, and DDN haplotypes were included in this analysis. NNN, DNN, and DND haplotypes were omitted due to limited data only observed in CU1_Caldwell2020 and CU1_Helfer2020 environments.

The S2MET population was grown in 2015 (S2MET_KT2015) and 2016 (S2MET_Hel2016, S2MET_Ket2016) and experienced a substantial change in T_{max} (3.3 C) and P_{sum} (-9.01 cm) between 2015 and 2016. The full S2MET (n=223) was genotyped with GA20ox1_331_5UTR, MKK3_E165Q, and JHI-367342-KASP (Chapter 3) and 181 lines were homozygous at each marker that also had observations in 2015 and 2016. Because the S2MET was planted in an augmented design, each line only had 2 or 3 observations. PHS BLUEs were estimated within year using the model

$$y_{ijk} = \mu + E_i + G_j + s_{k(i)} + \epsilon_{ijk} \quad (4.2)$$

where μ is the overall mean, E_i is the fixed effect for environment, G_j is the fixed effect of entry, $s_{k(i)}$ is the random effect of PHS scoring date nested within E_j with $s_{k(i)} \sim N(0, \sigma_s^2)$ where σ_s^2 is the variance due to scoring date, and ε_{ijk} is the random error with $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ where σ_ε^2 is the residual error variance. The change in PHS, *ΔPHS*, was calculated as the absolute value of the 2015 PHS BLUE subtracted from the 2016 PHS BLUE for each line. *ΔPHS* between *HvGA20ox1/HvMKK3* haplotypes and *HvMKK3* alleles was tested with a Tukey-Kramer test was used to determine significant differences ($p < 0.05$) in haplotype PHS stability between groups of unequal size using the *agricolae* R package (de Mendiburu and Yaseen, 2020).

Results

The spring dataset contained seven seed dormancy haplotypes but the number of observations for each haplotype was unbalanced (NNN*=133, DNN*=366, DDD=115, NND=47, NDD=510, NDN=361, DDN=90). Lines with the N* *HvMKK3* allele are typically selected against in Cornell spring regional testing due to high PHS susceptibility but the PHS susceptible DNN* lines AAC Synergy and Newdale were used as malting quality checks in the majority of the experiments, which accounted for the high number of DNN* observations. The DDN and DDD haplotypes were not observed in any spring commercial lines and were only present in Cornell experimental lines or the facultative variety Lightning (DDD). Experimental and commercial lines containing NDD and NDN haplotypes were common in spring regional and experimental testing. The winter dataset contained four seed dormancy haplotypes which were more balanced (NNN*=103, DNN*=99, DDD=129, NDD=241) but only two winter lines, Charles and Endeavor, contained the DNN* haplotype. These lines were used as malting quality checks in all winter experiments. Cornell winter barley breeding lines were not available for analysis, resulting in a smaller number of entries for the winter dataset. The mean (standard deviation) observed PHS scores for the spring dataset were μ_{NNN*} = 5.81 (1.52), μ_{DNN*} = 4.99 (1.39), μ_{DDD} = 1.88 (1.37), μ_{NDD} = 1.91 (1.51), μ_{NDN} = 2.6 (1.88), μ_{DDN} = 2.29 (1.61), μ_{NND} = 2.84 (1.58) and for the winter dataset were μ_{NNN*} = 6.36 (1.8), μ_{DNN*} = 5.76 (2.47), μ_{DDD} = 0.46 (0.79), μ_{NDD} = 0.67 (0.93). Observed average PHS scores for winter N* *HvMKK3* haplotypes were significantly higher ($p < 0.05$) than equivalent

spring haplotype means but winter D *HvMKK3* haplotype means were significantly lower ($p < 0.01$) than equivalent spring haplotype means.

Models were fit for five environmental covariates: T_{max}, T_{min}, T_{avg}, T_r, and P_{sum}. Average temperature and precipitation during the months of May to August from 2015 to 2020 were representative of weather in those months in the past 30 years (Figure 4.1). For both winter and spring datasets, T_{max} and P_{sum} were negatively correlated (Table 4.4), partially confounding the effects of increased temperature and decreased precipitation. T_{max} and T_{min} had low correlations in both datasets. Environment and QTL fixed effects were highly significant ($Pr(Chisq) < 0.001$) for all spring and winter models (Table 4.5). QZI was significant ($Pr(Chisq) < 0.05$) for spring $T_{max,}$ T_{min}, T_{avg} and winter T_{avg}, T_{min}, and P_{sum}. Variance components for G, GEI, and QEI were significant for the majority of environmental covariates in spring and winter models but GZI variance was only significant for spring T_{max} and T_R and winter T_{avg}. Spring P_{sum} and winter T_r had the best model fit as determined by AIC estimated with icREML. For the spring P_{sum} model, polygenic background variation due to entry (G) accounted for 30% of the within-environment variance. Entry by environment variance (GE), QTL by environment (QE), and entry by precipitation (GZ) explained 17%, 21%, and <0.01% of the within-environment variance, respectively. The winter T_r model had a lower proportion of within-environment variance explained by G (16%), higher proportion by QE (51%), and similar proportions by GZ (<0.01%), and GE (20%). Variance component estimates for other models within dataset were similar (Table 4.5). The percentage of variance explained by QE was consistently higher in winter than spring models. QZ interactions and

deviations due to G and GZ were plotted for spring P_{sum} (Figure 4.2), spring T_{max} (Figure 4.3), winter T_r (Figure 4.4), and winter T_{max} (Figure 4.5).

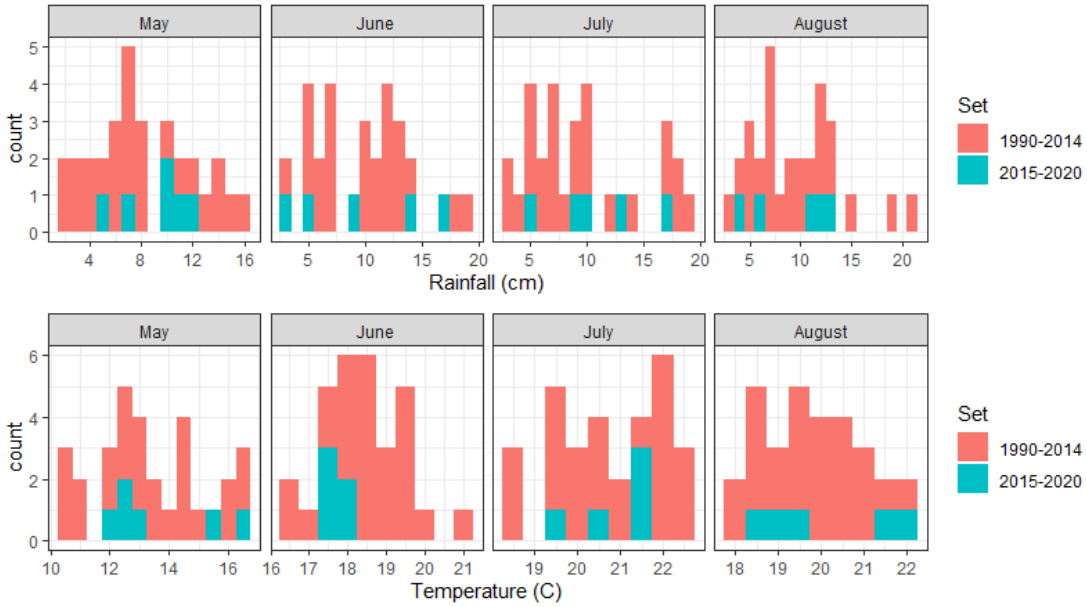


Figure 4.1: Average monthly May-August temperature and rainfall for Ithaca, NY from 1990-2020. Blue bars indicate years used in this study.

Table 4.4: Correlations between environmental covariates during the grain fill period for winter and spring (lower diagonal) datasets.

	T_{max}	T_{avg}	T_{min}	T_R	P_{sum}
T_{max}	1	0.81	0.21	0.70	-0.67
T_{avg}	0.83	1	0.74	0.16	-0.29
T_{min}	0.23	0.73	1	-0.79	0.29
T_R	0.54	0.02	-0.61	1	-0.79
P_{sum}	-0.68	-0.20	0.46	-0.82	1

Table 4.5: QTL x environment interaction model summaries for spring and winter models fitting environmental covariates (Z) as average daily high temperature (T_{max}), average daily low temperature (T_{min}), average daily temperature (T_{avg}), average daily temperature range (T_R), and the sum of precipitation (P_{sum}) during the grain fill period. Wald tests were conducted for environment (E), QTL (Q), and QTL by environmental covariate (QZI) fixed effects. Wald test significance is designated as Pr(Chisq) < 0.001 (***), Pr(Chisq) < 0.01 (**), Pr(Chisq) < 0.05 (*), or not significant (ns). Error variance (e) is presented as the average of heterogeneous error variance estimates at each environment. Variance component estimates are for entries with haplotype data and asterisks indicate significant difference from zero (Pr(Chisq) < 0.05).

Spring	Fixed	Random	T_{max}	T_{avg}	T_{min}	T_R	P_{sum}
Wald	E		***	***	***	***	***
	Q		***	***	***	***	***
	QZI		*	**	*	ns	ns
Variance		G	0.575*	0.632*	0.653*	0.532*	0.647*
		GZI	2.36e-2*	3.28e-2	1.46e-2	2.99e-2*	6.34e-4
		GEI	0.416*	0.424*	0.419*	0.368*	0.377*
		QEI	0.390*	0.374*	0.393*	0.472*	0.448*
		e	0.706	0.708	0.726	0.722	0.689
AIC			8920.57	8912.93	8899.95	8924.39	8888.13
Winter	Fixed	Random	T_{max}	T_{avg}	T_{min}	T_R	P_{sum}
Wald	E		***	***	***	***	***
	Q		***	***	***	***	***
	QZI		ns	*	*	ns	***
Variance		G	0.349*	0.398*	0.307	0.362*	0.367*
		GZI	2.68e-8	2.68e-8*	1.05e-3	2.68e-8	1.19e-7
		GEI	0.454*	0.447*	0.442*	0.457*	0.455*
		QEI	1.03*	1.102*	1.271*	1.14*	1.55*
		e	0.276	0.276	0.275	0.275	0.274
AIC			853.36	854.61	855.65	834.0	850.28

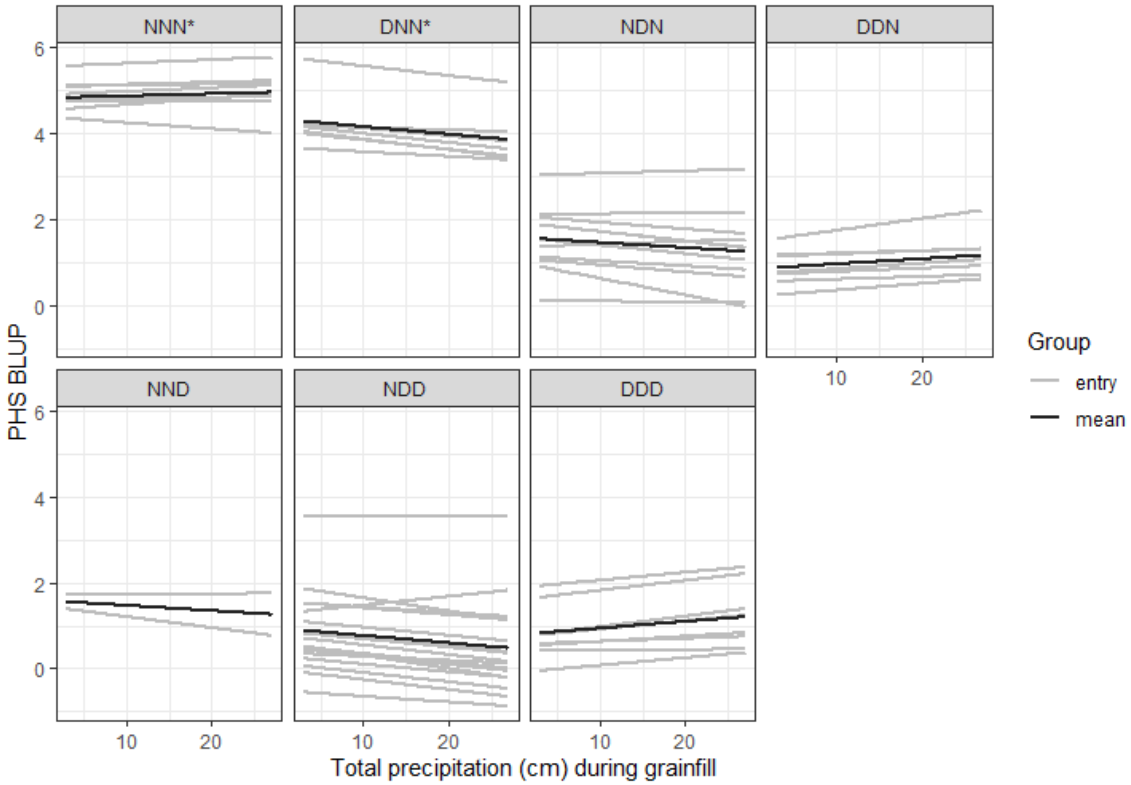


Figure 4.2: Spring barley QTL by precipitation interactions during grain fill and deviations due to background polygenic entry effects. Black lines indicate the main haplotype intercept and QTL by precipitation interaction and gray lines indicate entry deviations from the main haplotype intercept and slope.

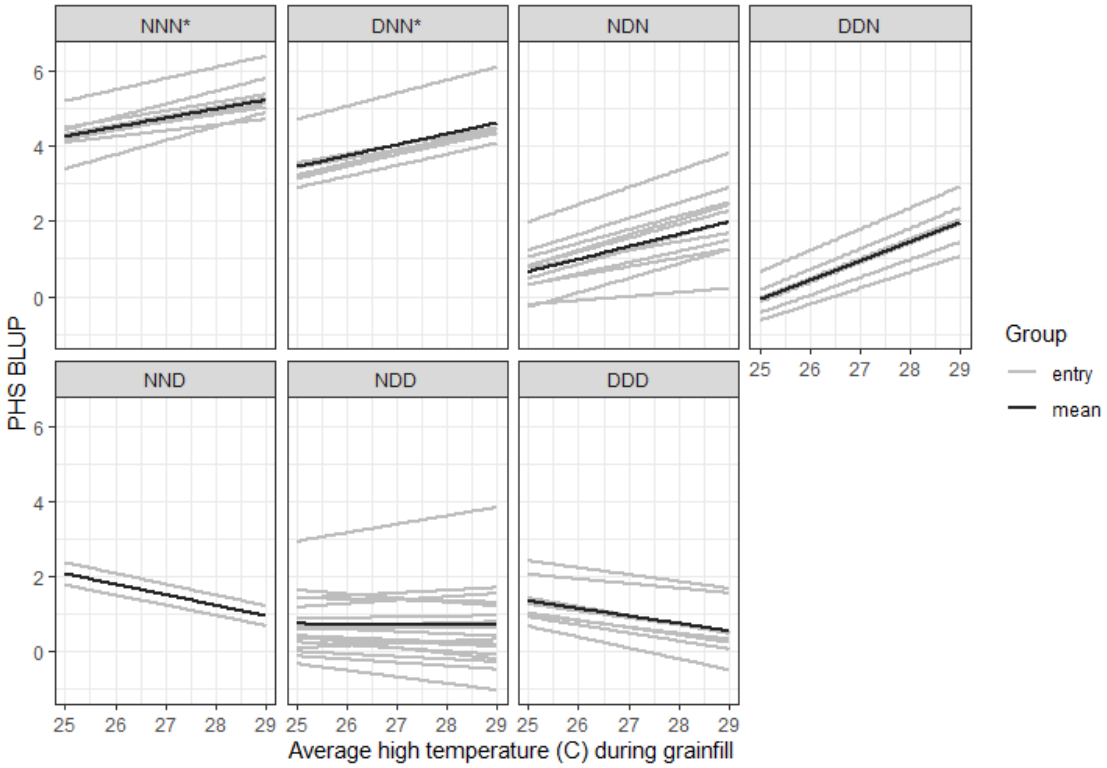


Figure 4.3: Spring barley QTL by average high temperature interactions during grain fill and deviations due to background polygenic entry effects. Black lines indicate the main haplotype intercept and QTL by average high temperature interaction and gray lines indicate entry deviations from the main haplotype intercept and slope.

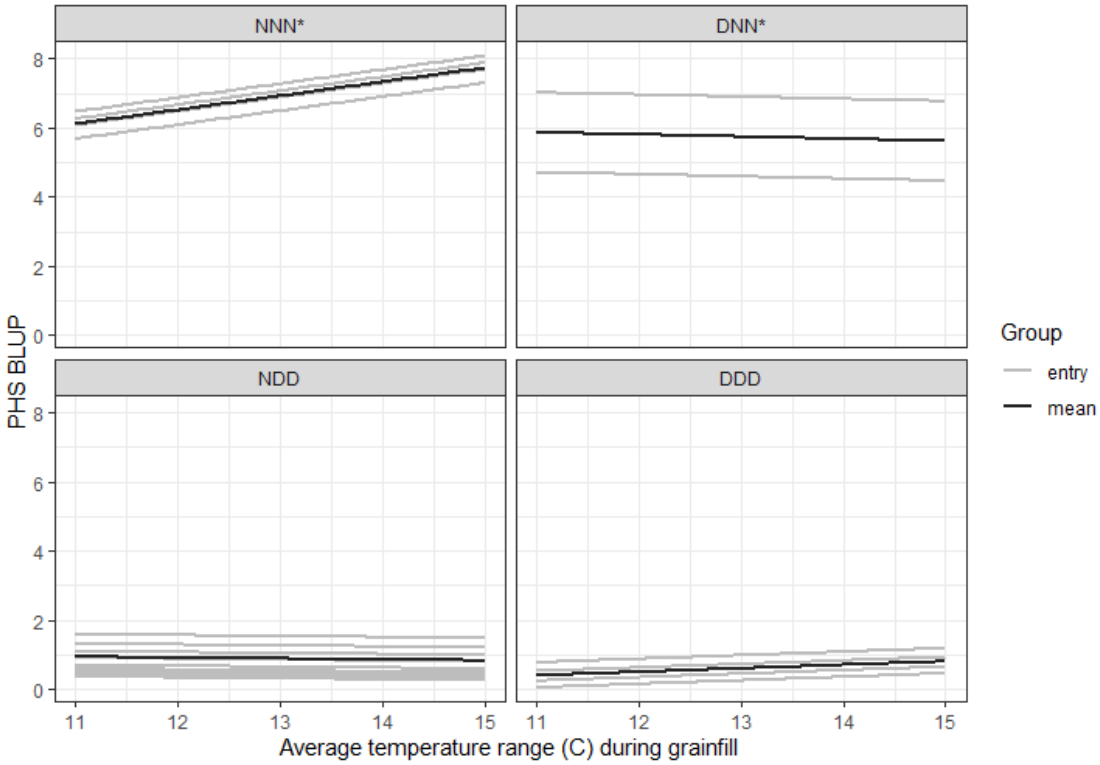


Figure 4.4: Winter barley QTL by temperature range interactions during grain fill and deviations due to background polygenic entry effects. Black lines indicate the main haplotype intercept and QTL by temperature range interaction and gray lines indicate entry deviations from the main haplotype intercept and slope.

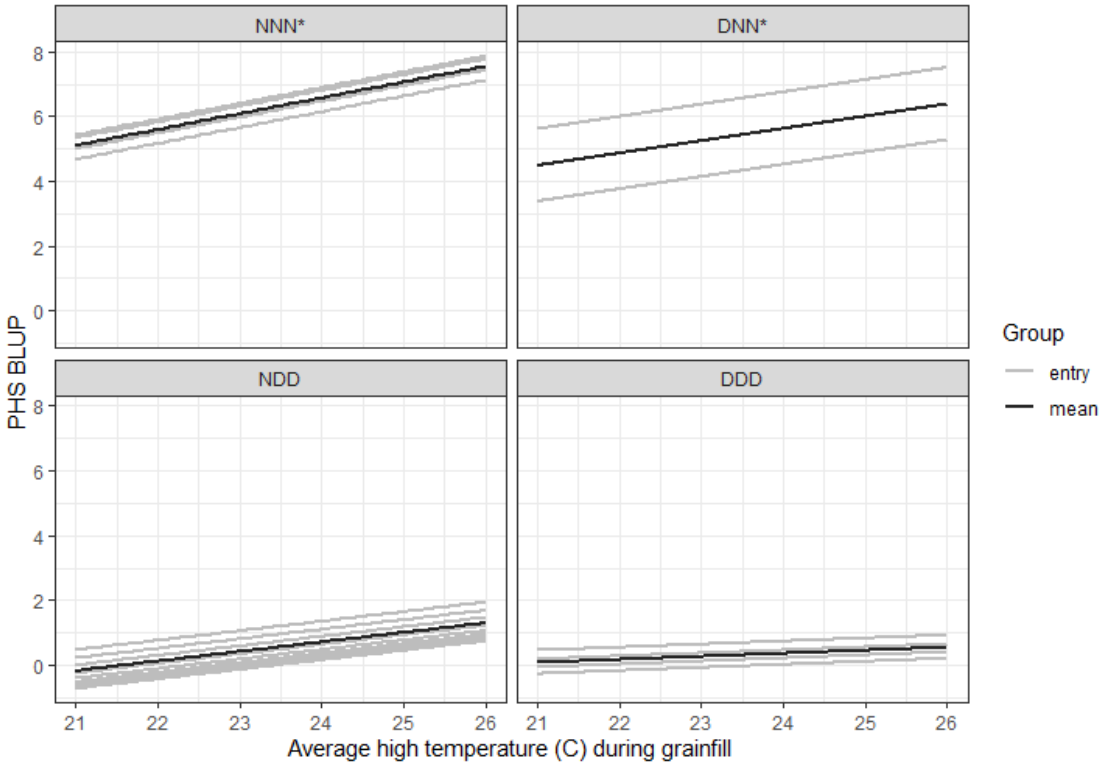


Figure 4.5: Winter barley QTL by average high temperature interactions during grain fill and deviations due to background polygenic entry effects. Black lines indicate the main haplotype intercept and QTL by average high temperature interaction and gray lines indicate entry deviations from the main haplotype intercept and slope.

Table 4.6: Fixed QTL by environmental covariate interaction slopes and standard errors for spring and winter datasets. Asterisks indicate slope is significantly different ($p<0.05$) from zero.

Dataset	Haplotype	T_{max}	T_{avg}	T_{min}	T_R	P_{sum}
Spring	NNN*	0.241 (0.17)	0.367 (0.2)	0.356 (0.183)	-0.161 (0.171)	0.005 (0.034)
Spring	DNN*	0.293* (0.142)	0.328 (0.166)	0.187 (0.144)	0.285* (0.132)	-0.018 (0.028)
Spring	DDD	-0.199 (0.22)	-0.263 (0.257)	-0.176 (0.211)	-0.068 (0.209)	0.015 (0.045)
Spring	NDD	-0.01 (0.105)	-0.132 (0.124)	-0.229* (0.111)	0.12 (0.105)	-0.012 (0.024)
Spring	NDN	0.331* (0.127)	0.339* (0.151)	0.158 (0.127)	0.220 (0.126)	-0.017 (0.02)
Spring	DDN	0.511 (0.277)	0.493 (0.342)	0.053 (0.256)	0.255 (0.218)	0.012 (0.055)
Spring	NND	-0.28 (0.27)	-0.748 (0.415)	-0.7 (0.363)	-0.054 (0.241)	-0.012 (0.046)
Winter	NNN*	0.488* (0.233)	0.503 (0.264)	0.367 (0.275)	0.404 (0.367)	-0.093 (0.081)
Winter	DNN*	0.38 (0.24)	0.624* (0.296)	0.615* (0.296)	-0.062 (0.323)	0.152* (0.049)
Winter	DDD	0.095 (0.203)	0.12 (0.237)	0.103 (0.247)	0.102 (0.284)	-0.083 (0.046)
Winter	NDD	0.295 (0.183)	0.449* (0.208)	0.415* (0.202)	-0.027 (0.267)	0.098* (0.04)

Haplotypes with a N* *HvMKK3* allele (NNN*, DNN*) exhibited higher baseline PHS susceptibility than N or D *HvMKK3* haplotypes (Figures 4.2-4.5). QZI effects generally had high standard errors and only a few haplotype specific QZI effects were significantly different than zero in both datasets (Table 4.6). Both spring and winter N* *HvMKK3* haplotypes were sensitive to T_{max}, T_{avg}, and T_{min}, but winter haplotypes were generally more sensitive than springs (Table 4.6, Figure 4.3 and 4.5). N* *HvMKK3* spring and winter haplotypes showed contrasting sensitivities to T_r and P_{sum}. Spring NNN* and DNN* haplotypes were sensitive to T_r but not $P_{sum,}$ winter NNN* haplotypes were sensitive to T_r and P_{sum} (negative slope), and winter DNN*

haplotypes were sensitive to P_{sum} but not T_r. Haplotypes with D *HvMKK3* alleles (DDD, NDD, NND) in spring and winter datasets were generally more temperature insensitive than N* *HvMKK3* haplotypes and had baseline PHS resistance. Winter NDD haplotypes showed some sensitivity to T_{max}, T_{avg}, T_{min}, and P_{sum} but winter DDD haplotypes were only negatively sensitive to P_{sum}. Spring DDD haplotypes had lower PHS as T_{max}, T_{avg}, and T_{min} increased, NDD haplotypes were only moderately sensitive to T_{avg} and T_{min}, and NND haplotypes were sensitive to T_{max} and highly sensitive to T_{avg} and $T_{min.}$ All spring haplotypes had very low sensitivity to precipitation. Spring N *HvMKK3* haplotypes (NDN, DDN) were highly sensitive to temperature and had baseline PHS resistance similar to dormant *HvMKK3* haplotypes. Within haplotype for both winter and spring datasets, there was considerable variation for baseline PHS resistance due to entry. Although the variance due to GZ was small for spring datasets, the change in haplotype temperature sensitivity due to entry was large enough to change PHS resistance ranks for entries within spring NNN*, NDD, and NDN haplotypes as T_{max} increased (Figure 4.3). Winter GZ variance was negligible for all models and did not result in rank changes across environmental covariates.

Most S2MET lines showed PHS instability between the cool wet 2015 environment to the hot and dry 2016 environments, with *ΔPHS* greater than 1.5 for 130/181 lines and increased PHS in 154/181 lines. Haplotype observations were highly unbalanced when *HvGA20ox1* was used to define haplotypes, but N* *HvMKK3* lines had the lowest *ΔPHS* (Table 4.7). When *HvGA20ox1* was omitted from haplotypes, significant differences were observed between *HvMKK3* alleles for *ΔPHS*

and PHS. N* *HvMKK3* lines had high baseline PHS with low *ΔPHS*, N *HvMKK3* lines had low to moderate baseline PHS with high *ΔPHS*, and D *HvMKK3* lines had low baseline PHS with moderate *ΔPHS.*

Table 4.7: Change in preharvest sprouting (ΔPHS) BLUEs for S2MET between 2015 and 2016. Haplotypes are defined at the combination of alleles at *HvGA20ox1* and *HvMKK3* or only *HvMKK3* where N* specifies a highly non-dormant allele, N specifies a non-dormant allele, and D specifies a dormant allele. Mean PHS BLUEs within year and ΔPHS BLUEs with standard deviations in parentheses are shown for each haplotype. A Tukey-Kramer test was used to determine significant differences ($p < 0.05$) between ΔPHS means of haplotype groups of unequal size. Haplotypes that share the same letter in the TukeyHSD column are not significantly different from each other.

Genes	Haplotype	n	PHS_{2015}	PHS_{2016}	ΔPHS (sd)	TukeyHSD
HvGA20ox1/HvMKK3	NN*	60	5.38 (1.45)	5.98 (0.73)	1.28 (0.81)	b
HvGA20ox1/HvMKK3	DN	62	1.16 (1.11)	4.85 (1.29)	3.77 (1.21)	a
HvGA20ox1/HvMKK3	NN	3	1.43 (1.92)	5.01 (1.55)	3.58 (0.88)	a
HvGA20ox1/HvMKK3	ND	5	0.66 (0.52)	4.23 (0.56)	3.55 (0.92)	a
HvGA20ox1/HvMKK3	DD	31	0.61 (0.7)	2.92 (1.73)	2.51 (1.4)	ab
HvMKK3	N*	60	5.38 (1.45)	5.98 (0.73)	1.28 (0.81)	c
HvMKK3	N	65	1.18 (1.13)	4.85 (1.29)	3.76 (1.19)	a
HvMKK3	D	36	0.62 (0.67)	3.1 (1.67)	2.66 (1.38)	b

Discussion

Annual variation in two-row winter and spring malting barley PHS across six years was modeled with environmental covariates related to temperature and precipitation. The years used in this analysis, 2015-2020, were representative of the past 30 years in Ithaca, NY and captured a wide range of temperature and precipitation during grain

fill. Annual PHS variation was partitioned into genetic, environmental, and GEI effects. Genetic effects were partitioned into seed dormancy QTL and residual polygenic entry effects and environmental effects were partitioned into year-location and environmental covariate effects. Although the testing sites sampled in this dataset are physically close, there was substantial environmental variation between and within sites for soil type, management history, topography, and disease pressure. These factors were not quantified and contributed to GEI, QEI, and error. Uniform sampling of physiologically mature spikes for PHS assays is affected when disease pressure is high or temperature sharply increases just before sampling, leading to discoloration of the peduncle or premature senescence. Heading date phenotyping accuracy was affected by incomplete spike emergence in spring barley. Severe winter injury can substantially delay heading date and cause irregular maturities within plot in winter barley, biasing sampling date and grain fill intervals. The grain fill period for both winter and spring datasets varied substantially across years and within entry, making accurate comparisons of developmental stages across years and experiments difficult. Increased modeling precision of seed dormancy haplotype sensitivity to environmental conditions would benefit from measurement of environmental covariates within defined seed development intervals. Despite these phenotyping challenges, PHS response to temperature was similar to previously reported results. Temperature has a highly positive correlation with PHS during late grain fill stages defined by thermal time (Gualano and Benech-Arnold, 2009) and increased temperature in the first several weeks of grain fill during embryo development has been correlated with a reduction in seed dormancy (Gong et al., 2014). Environmental sensitivity of ABA

and GA biosynthesis and catabolism genes at critical developmental timepoints may play a role in determining PHS susceptibility. Expression of the ABA biosynthesis gene *HvNCED1* and the ABA catabolism gene *HvCYP707A1* vary during grain fill and change ABA content and germination percentage, indicating sensitivity to environmental factors (Chono et al., 2006). Change in GA content during grain fill and the impact of environment on GA content, ABA/GA ratio, and sensitivity to ABA and GA is poorly understood in barley. Change in expression of large effect seed dormancy genes during grain fill across varieties and environmental conditions, particularly *HvAlaAT1* and *HvMKK3*, is also not understood.

Environmental sources of PHS variation

Average observed PHS for winter N* *HvMKK3* haplotypes was higher than spring N* *HvMKK3* haplotypes but winter D *HvMKK3* haplotypes had consistently lower observed PHS than comparable spring D *HvMKK3* haplotypes. These differences between spring and winter datasets may have a genetic component, but Endeavor and a number of the winter DH entries have spring germplasm in their pedigree, reducing the probability that winter germplasm specific loci alone introduced additional sources of PHS variation. Environmental components, particularly temperature, are a more likely cause of the more extreme observed PHS values in the winter dataset. Average T_{max}, T_{avg}, and T_{min} were each about 2.5 C higher in the spring dataset than the winter dataset. T_{max} never exceeded 26 C for the winter dataset but 76% of the spring T_{max} observations exceeded 26 C. Limited data from three facultative lines included in spring and winter datasets also supports the role of temperature in observed

differences between winter and spring PHS. Average observed PHS in winter and spring, respectively, was 0 and 1.44 for Lightning (DDD), 0 and 3.34 for DH131055 (NDD), and 1.07 and 2.14 for DH130935 (NDD) (Table 4.2). Cooler temperatures during grain fill may have increased ABA content, decreased ABA decay rate, or increased ABA sensitivity during grain fill in the winter dataset, all of which could have induced stronger primary dormancy in D *HvMKK3* haplotypes. Unless variants at ABA biosynthesis, catabolism, or sensitivity loci are present, increased PHS in winter N* *HvMKK3* haplotypes may be due to temperature mediated increases in GA content or GA sensitivity that counter any increase in ABA.

The environmental covariates that produced the best model fit for spring (P_{sum}) and winter (T_r) datasets were different than the expected best fit of T_{max} or T_{avg}. The winter T_{max}, T_{avg}, and T_{min} models indicated substantial QZ interaction for NNN*, DNN*, and NDD haplotypes but the T_r model only indicated QZ interaction for NNN* (Table 4.6). This result was not observed for the spring T_r model. The two winter experiments grown in 2016 had significantly ($p < 2.2$ e-16) higher average T_r (μ=14.7 C) than experiments in other years (μ=12.6 C) and T_r was less correlated with T_{max} in 2016 (r=0.584) than in other years (r=0.933). The smaller size of the winter dataset may have been more sensitive to years with large differences in temperature during grain fill than the spring dataset, as this relationship was not observed for spring. The ability of T_r to capture variation in T_{min} may have been more informative under the cooler growth conditions of winter barley. The spring P_{sum} model indicated very low sensitivity to precipitation for all haplotypes but a small increase in genetic variance was observed compared to other spring models, indicating greater baseline

variation in PHS resistance in drought conditions. These results were in stark contrast to the winter P_{sum} model which had highly significant QZ interactions (Pr(Chisq<0.01)) but inconsistent precipitation sensitivity across haplotypes. The two non-dormant winter haplotypes showed contrasting QZ slopes (NNN*= -0.093, DNN*=0.152) as did the dormant winter haplotypes (DDD = -0.083, NDD = 0.098). These results might also be partially explained by the smaller winter dataset and the presence of only two entries for DNN*. Differences between baseline haplotype effect may be exacerbated with smaller sample size, especially if the entries within haplotype are highly variable in their baseline PHS in drought conditions. Haplotypes were partially confounded with genetic background which limited the resolution to define haplotype and residual genetic effects.

Variance component estimates for QEI were consistently large compared to G and GEI variance components in the winter dataset. Winter GEI variance components were also consistently larger than G variance components, unlike the spring dataset. Several environmental factors may have increased QEI and GEI variance in the winter dataset. The limited number of DNN* entries may have contributed to larger QEI variance, as both DNN* lines, Charles and Endeavor, are susceptible to the foliar disease scald, caused by *Rhynchosporium commune* (Zaffarano, B.A. McDonald & Linde), which can infect seeds and may affect seed viability. Charles was a check for all winter experiments, due to its success in western growing regions, but it is poorly adapted to New York and had the largest phenotypic variance (σ_p^2=7.1) across both datasets. In 2019 and 2020, an additional after-ripening day was added for winter barley in an attempt to increase the variation for PHS for selection purposes. Adding

this information as a covariate in the winter model had no effect but still may have biased environmental effects upwards for those years, particularly for N* *HvMKK3* haplotypes. Bias in heading and sampling date due to winter injury may have also contributed.

Spring two-row malting barley experiments typically have less PHS stability across years than winter malting barley. Spring two-row barley had more seed dormancy haplotype combinations, greater polygenic variation, and more variation in temperature sensitivity than winter two-row malting barley. However, almost three times more spring than winter entries were evaluated in this study. Winter germplasm was limited due to the recent addition of winter malting barley breeding at Cornell. More genetic variation may be present in other winter germplasm, although malting barley generally has less genetic variation than other barley market classes due to strong selection for malting quality traits (Martin et al., 1991). Additional spring polygenic sensitivity to temperature may be due to ABA or GA synthesis, catabolism, and signaling loci. The increased G and GEI variation within haplotype in the spring dataset is of particular interest. ND Genesis and Pinnacle are NDN entries from the same breeding program (NDSU) but had observed PHS means of 3.43 and 1.55, respectively. Despite overall haplotype T_{max} instability for NDN, some NDN entries showed low baseline PHS and little change as temperature increased. Within the DDD, DNN*, and NDD spring haplotype groups, several entries had noticeably higher baseline PHS BLUPs than average. This may be due to genotyping errors, modifiers of *HvAlaAT1* and *HvMKK3*, or other unknown large effect loci. Further mapping is needed in two-row spring malting barley germplasm within *HvMKK3* allele to better

understand the basis of quantitative variation for PHS and PHS temperature sensitivity.

Physiological temperature sensitivity of seed dormancy genes

Preharvest sprouting increased in N* and N *HvMKK3* spring and winter haplotypes and NDD winter haplotypes as grain fill temperature increased. Dormant spring *HvMKK3* and DDD winter haplotypes had little change, or in some cases, a small decrease in PHS as grain fill temperature increased. *HvMKK3* alleles were observed to have different *ΔPHS* in S2MET across two years with very different weather patterns. These observations support differential temperature sensitivity of *HvMKK3* alleles. Although the physiological function of *HvMKK3* is still unknown, results from wheat and rice provide a potential model for temperature sensitivity of *HvMKK3*. *TaMKK3-A* has been identified as the causal gene at the *Phs-A1* locus for PHS resistance on chromosome 4A in wheat (Torada et al., 2016; Shorinola et al., 2017). The ABA hypersensitive *ENHANCED RESPONSE TO ABA8* (*ERA8*) mutant in wheat is likely a novel allele of *TaMKK3-A* (Martinez et al., 2020) that results in increased ABA sensitivity at physiological maturity through the after-ripening period but does not increase ABA content (Martinez et al., 2016). In rice (*Oryza sativa* L.), *OsMKK3* is part of a MAPK cascade system composed of MKKK62-MKK3-MAPK7/14 (Mao et al., 2019). Overexpression of MKKK62 reduced seed dormancy, reduced ABA sensitivity, and reduced expression of *OsMFT* while knockouts of MKK3 and MAPK7/14 increased seed dormancy and expression of *OsMFT*. *MOTHER OF FT AND TFL1* (*MFT*) is a highly conserved regulator of seed

germination in the phosphatidylethanolamine-binding protein family, which also includes *FLOWERING LOCUS T* (*FT*) and *TERMINAL FLOWER1* (*TFL1*) (Xi et al., 2010). *OsMFT2* was recently identified as a positive regulator of ABA signaling, and thus seed dormancy, in rice through interactions with three basic leucine zipper transcription factors (Song et al., 2020). A wheat *MFT* homolog is the underlying causal gene for the *TaPHS1* locus on chromosome 3AS (Nakamura et al., 2011; Liu et al., 2013) and positively regulates seed dormancy. *TaMFT* displays different temperature sensitivity before and after seed physiological maturity. Low grain fill temperatures (13 C) increased *TaMFT* expression and embryo dormancy while high grain fill temperatures (25 C) reduced *TaMFT* expression and embryo dormancy (Nakamura et al., 2011). *TaMFT* expression is reduced by low germination temperatures (4 C overnight treatment) in after-ripened seed, leading to increased germination (Lei et al., 2013). Vetch et al. (2020) sequenced *HvMFT* in North American spring malting germplasm but did not detect sequence variants. In wheat, *TaMKK3* is likely involved in ABA signaling but has not been directly associated with *TaMFT*. It is unknown if *TaMFT* is directly temperature sensitive or regulated by upstream temperature sensitive factors but several ABA-regulated MAPK cascades are known to be induced by abiotic stress (Danquah et al., 2015; Colcombet & Hirt, 2008). *HvMKK3* may be involved in ABA signaling and regulation of *HvMFT*. The MAPK cascade including *HvMKK3* may not be directly temperature sensitive, but it is likely ABA sensitive and barley seed ABA content is temperature sensitive. Reduced ABA sensitivity in an ABA-mediated MAPK cascade conferred by *HvMKK3* could lead to a baseline reduction in *HvMFT* expression, and therefore seed dormancy, that

would further be reduced by decreased ABA content resulting from high temperature during grain fill. Our results indicate N* *HvMKK3* allele confers high baseline PHS and temperature sensitivity, the N allele confers low baseline PHS and high temperature sensitivity, and the D allele confers low baseline PHS and low to moderate temperature sensitivity. These results are supported by the S2MET *ΔPHS* analysis. This suggests low, high, and low ABA sensitivity of the N*, N, and D *HvMKK3* alleles, respectively. Temperature sensitivity of *HvGA20ox1* could not accurately be estimated in these datasets due to incomplete haplotype observations. Although the NND haplotype exhibited greater temperature sensitivity than NDD haplotypes in the spring dataset, only two NND lines had enough observations to be included in the models. As previously mentioned, temperature induced changes in GA content, decay, and sensitivity are unknown in malting barley and are an area for future research.

HvAlaAT1 also showed evidence of temperature sensitivity, with differences in slope for haplotype pairs differing by allelic state at *HvAlaAT1* (NNN*/DNN*, DDD/NDD, NDN/DDN) for some environmental covariates in winter and spring datasets. The physiological mechanism of *HvAlaAT1* imposed seed dormancy is also unknown but has also been speculated to be connected to ABA signaling (Sato et al., 2016). This hypothesis is consistent with the temperature sensitivity results of this study. *HvAlaAT1* has been observed to have a larger effect on dormancy duration than primary dormancy level (Sato et al., 2016; Vetch et al., 2020; Chapter 3). *TaQsd1* was identified as a dormancy period QTL in wheat and was found to encode AlaAT, with variants in *TaQsd1-5B* (Wei et al., 2019). *TaQsd1-5B* variants associated with

prolonged dormancy had higher expression of the ABA biosynthesis genes, *TaNCED1* and *TaNCED2*, decreased expression of the ABA catabolism gene *TaCYP707A1*, and increased ABA content compared with short dormancy *TaQsd1-5B* variants. Temperature sensitivity of *HvAlaAT1* may be similarly related to the change in *HvNCED1* and *HvCYP707A1* expression across environments noted by Chono et al. (2006). The *SD1* locus has rarely been detected in PHS mapping studies with two-row by two-row parentage (Li et al., 2003; Ullrich et al., 2009; Hickey et al., 2012; Chapter 2 and 3). The exceptions are Bonnardeaux et al. (2008) and Gong et al. (2014) who both detected *SD1* as a minor effect QTL in a Stirling/Harrington doubled haploid population grown in three and ten environments, respectively. However, *SD1* was not detected in all environments and had variable effects across environment. Gong et al. (2014) speculated that growth conditions did not favor *SD1* expression or that germination test conditions favored rapid germination, masking the dormancy release rate effects of *SD1*. This supports the findings of this study that *HvAlaAT1* in combination with *HvMKK3* alleles has variable effects across temperature regimes.

Conclusion

Our results demonstrate different seed dormancy haplotype response to temperature and precipitation covariates in winter and spring malting barley. These results provide valuable information for barley breeders looking to select stable PHS resistance and may be useful for wheat breeding as well since MKK3 and AlaAT are known to affect PHS in both species. Marker assisted selection is a promising tool to use in breeding for PHS resistance due to the presence of several large effect seed

dormancy QTL but background polygenic effects and additional polygenic temperature sensitivity suggest the potential for a genomic selection approach. This quantitative variation indicates that further breeding progress may be possible for combining and fine tuning PHS resistance with good malting quality. The impact of temperature on malting quality and its relationship with PHS susceptibility in the spring and winter datasets is unknown. In regions with high annual PHS risk and large variation in spring and summer temperatures, the dormant *HvMKK3* allele is the most stable source of PHS resistance for winter and spring barley. Triple dormant lines may exhibit excessive primary dormancy in spring germplasm (Chapter 3), which is undesirable for malting and may explain why no commercial spring DDD lines were observed in this study. Several of the winter DDD entries are known to have good malting quality, suggesting possible differences in dormancy release rate or ABA sensitivity between winter and spring DDD entries. For areas with cooler average summer temperatures, the NDN haplotype may provide adequate PHS resistance for spring germplasm but the increased polygenic variation in this haplotype group requires further local testing. Winter and spring entries with the highly non-dormant *HvMKK3* allele have consistently high risk for PHS in all environments and must be carefully managed in high moisture environments. These models provide a practical model for forecasting PHS risk by variety with simple weather data to prioritize harvest in high-risk years.

References

Abe, A., H. Takagi, T. Fujibe, K. Aya, M. Kojima, et al. 2012. OsGA20ox1, a candidate gene for a major QTL controlling seedling vigor in rice. Theor. Appl. Genet. 125(4): 647–657. doi: 10.1007/s00122-012-1857-z.

Alqudah, A.M., and T. Schnurbusch. 2017. Heading date is not flowering time in spring barley. Front. Plant Sci. 8: 896. doi: 10.3389/fpls.2017.00896.

Anderson, J.A., M.E. Sorrells, and S.D. Tanksley. 1993. RFLP analysis of genomic regions associated with resistance to preharvest sprouting in wheat. Crop Sci. 33(3): 453. doi: 10.2135/cropsci1993.0011183X003300030008x.

Benech-Arnold, R.L., M.C. Giallorenzi, J. Frank, and V. Rodriguez. 1999. Termination of hull-imposed dormancy in developing barley grains is correlated with changes in embryonic ABA levels and sensitivity. Seed Sci. Res. 9(1): 39–47. doi: 10.1017/S0960258599000045.

Benech-Arnold, R.L., N. Gualano, J. Leymarie, D. Côme, and F. Corbineau. 2006. Hypoxia interferes with ABA metabolism and increases ABA sensitivity in embryos of dormant barley grains. J. Exp. Bot. 57(6): 1423–1430. doi: 10.1093/jxb/erj122.

Betts, N.S., C. Dockter, O. Berkowitz, H.M. Collins, M. Hooi, et al. 2020. Transcriptional and biochemical analyses of gibberellin expression and content in germinated barley grain. J. Exp. Bot. 71(6): 1870–1884. doi: 10.1093/jxb/erz546.

Bewley, J.D., K.J. Bradford, H.W.M. Hilhorst, and H. Nonogaki. 2013. Seeds: Physiology of development, germination and dormancy, 3rd edition. Springer New York.

Biddulph, T.B., D.J. Mares, J.A. Plummer, and T.L. Setter. 2005. Drought and high temperature increases preharvest sprouting tolerance in a genotype without grain dormancy. Euphytica. Springer. p. 277–283

Biddulph, T.B., J.A. Plummer, T.L. Setter, and D.J. Mares. 2007. Influence of high temperature and terminal moisture stress on dormancy in wheat (Triticum aestivum L.). F. Crop. Res. 103(2): 139–153. doi: 10.1016/j.fcr.2007.05.005.

Blake, V.C., C. Birkett, D.E. Matthews, D.L. Hane, P. Bradbury, et al. 2016. The Triticeae toolbox: Combining phenotype and genotype data to advance small-grains breeding. Plant Genome 9(2): plantgenome2014.12.0099. doi: 10.3835/plantgenome2014.12.0099.

Bonnardeaux, Y., C. Li, R. Lance, X.Q. Zhang B,C, K. Sivasithamparam, et al. Seed dormancy in barley: identifying superior genotypes through incorporating epistatic interactions. doi: 10.1071/AR07345.

Bradford, K.J., R.L. Benech-Arnold, D. Côme, and F. Corbineau. 2008. Quantifying the sensitivity of barley seed germination to oxygen, abscisic acid, and gibberellin using a population-based threshold model. J. Exp. Bot. 59(2): 335–347. doi: 10.1093/jxb/erm315.

Chono, M., I. Honda, S. Shinoda, T. Kushiro, Y. Kamiya, E. Nambara, N. Kawakami, S. Kaneko, Y. Watanabe. 2006. Field studies on the regulation of abscisic acid content and germinability during grain development of barley: molecular and chemical analysis of pre-harvest sprouting | Journal of Experimental Botany | Oxford Academic. J. Exp. Bot. Vol. 57, Issue 10. doi: https://doi.org/10.1093/jxb/erj215.

Colcombet, J., and H. Hirt. 2008. Arabidopsis MAPKs: a complex signalling network involved in multiple biological processes. Biochem. J 413: 217–226. doi: 10.1042/BJ20080625.

Danquah, A., A. de Zélicourt, M. Boudsocq, J. Neubauer, N. Frei dit Frey, et al. 2015. Identification and characterization of an ABA-activated MAP kinase cascade in *Arabidopsis thaliana*. Plant J. 82(2): 232–244. doi: 10.1111/tpj.12808.

de Mendiburu, F. and Yaseen, M. 2020. agricolae: Statistical Procedures for Agricultural Research.R package version 1.4.0 , https://myaseen208.github.io/agricolae/https://cran.r-project.org/package=agricolae.

Depauw, R.M., and T.N. Mccaig. 1991. Components of variation, heritabilities and correlations for indices of sprouting tolerance and seed dormancy in Triticum spp. Euphytica 52: 221-229.

Finch-Savage, W.E., and G. Leubner-Metzger. 2006. Seed dormancy and the control of germination. New Phytol. 171(3): 501–523. doi: 10.1111/j.1469-8137.2006.01787.x.

Gao, W., J.A. Clancy, F. Han, D. Prada, A. Kleinhofs, et al. 2003. Molecular dissection of a dormancy QTL region near the chromosome 7 (5H) L telomere in barley. Theor. Appl. Genet. 107(3): 552–559. doi: 10.1007/s00122-003-1281-5.

Goldbach, H., and G. Michael. 1976. Abscisic acid content of barley grains during ripening as affected by temperature and variety. Crop Sci. 16(6): 797–799. doi: 10.2135/cropsci1976.0011183x001600060015x.

Gong, X., C. Li, M. Zhou, Y. Bonnardeaux, and G. Yan. 2014. Seed dormancy in barley is dictated by genetics, environments and their interactions. Euphytica 197: 355:368. doi: 10.1007/s10681-014-1072-x.

Gualano, N.A., and R.L. Benech-Arnold. 2009. The effect of water and nitrogen availability during grain filling on the timing of dormancy release in malting barley crops. Euphytica 168(3): 291–301. doi: 10.1007/s10681-009-9948-x.

Han, F., S.E. Ullrich, J.A. Clancy, and I. Romagosa. 1999. Inheritance and fine mapping of a major barley seed dormancy QTL. Plant Sci. 143(1): 113–118. doi: 10.1016/S0168-9452(99)00028-X.

Hickey, L.T., W. Lawson, V.N. Arief, G. Fox, J. Franckowiak, et al. 2012. Grain dormancy QTL identified in a doubled haploid barley population derived from two non-dormant parents. Euphytica 188(1): 113–122. doi: 10.1007/s10681-011-0577-9.

Hori, K., K. Sato, and K. Takeda. 2007. Detection of seed dormancy QTL in multiple mapping populations derived from crosses involving novel barley germplasm. Theor. Appl. Genet. 115(6): 869–876. doi: 10.1007/s00122-007-0620-3.

Jacobsen, J. V., D.W. Pearce, A.T. Poole, R.P. Pharis, and L.N. Mander. 2002. Abscisic acid, phaseic acid and gibberellin contents associated with dormancy and germination in barley. Physiol. Plant. 115(3): 428–441. doi: 10.1034/j.1399-3054.2002.1150313.x.

King, R.W. 1976. Abscisic acid in developing wheat grains and its relationship to grain growth and maturation. Planta 132(1): 43–51. doi: 10.1007/BF00390329.

Lenoir, C., F. Corbineau, and D. Come. 1986. Barley (Hordeum vulgare) seed dormancy as related to glumella characteristics. Physiol. Plant. 68(2): 301–307. doi: 10.1111/j.1399-3054.1986.tb01930.x.

Li, C., P. Ni, M. Francki, A. Hunter, Y. Zhang, et al. 2004. Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. Funct. Integr. Genomics 4(2): 84–93. doi: 10.1007/s10142-004-0104-3.

Lin, R., R.D. Horsley, N.L. V. Lapitan, Z. Ma, and P.B. Schwarz. 2009. QTL mapping of dormancy in barley using the Harrington/Morex and Chevron/Stander mapping populations. Crop Sci. 49(3): 841. doi: 10.2135/cropsci2008.05.0269.

Liu, Y., J. Fang, F. Xu, J. Chu, C. Yan, et al. 2014. Expression Patterns of ABA and GA metabolism genes and hormone levels during rice seed development and imbibition: A comparison of dormant and non-dormant rice cultivars. J. Genet. Genomics 41(6): 327–338. doi: 10.1016/j.jgg.2014.04.004.

Malosetti, M., J. Voltas, I. Romagosa, S.E. Ullrich, and F.A. Van Eeuwijk. 2004. Mixed models including environmental covariables for studying QTL by environment interaction. Kluwer Academic Publishers.

Mao, X., J. Zhang, W. Liu, S. Yan, Q. Liu, et al. 2019. The MKKK62-MKK3-MAPK7/14 module negatively regulates seed dormancy in rice. Rice 12(1): 2. doi: 10.1186/s12284-018-0260-z.

Martin, J.M., T.K. Blake, and E.A. Hockett. 1991. Diversity among North American spring barley cultivars based on coefficients of parentage. Crop Sci. 31(5): 1131. doi: 10.2135/cropsci1991.0011183X003100050009x.

Martinez, S.A., O. Shorinola, S. Conselman, D. See, D.Z. Skinner, et al. 2020. Exome sequencing of bulked segregants identified a novel TaMKK3-A allele linked to the wheat ERA8 ABA-hypersensitive germination phenotype. Theor. Appl. Genet. 133(3): 719–736. doi: 10.1007/s00122-019-03503-0.

Martinez, S.A., K.M. Tuttle, Y. Takebayashi, M. Seo, K.G. Campbell, et al. 2016. The wheat ABA hypersensitive ERA8 mutant is associated with increased preharvest

sprouting tolerance and altered hormone accumulation. Euphytica 212(2): 229–245. doi: 10.1007/s10681-016-1763-6.

Nagel, M., A.M. Alqudah, M. Bailly, L. Rajjou, S. Pistrick, et al. 2019. Novel loci and a role for nitric oxide for seed dormancy and preharvest sprouting in barley. Plant. Cell Environ. doi: 10.1111/pce.13483.

Nakamura, S., F. Abe, H. Kawahigashi, K. Nakazono, A. Tagiri, et al. 2011. A wheat homolog of MOTHER of FT and TFL1 acts in the regulation of germination. Plant Cell 23(9): 3215–3129. doi: 10.1105/tpc.111.088492.

Nakamura, S., M. Pourkheirandish, H. Morishige, Y. Kubo, M. Nakamura, et al. 2016. Mitogen-activated protein kinase kinase 3 regulates seed dormancy in barley. Curr. Biol. 26(6): 775–781. doi: 10.1016/J.CUB.2016.01.024.

Oberthur, L., T.K. Blake, W.E. Dyer, and S.E. Ullrich. 1995. Genetic analysis of seed dormancy in barley (Hordeum vulgare L.). J. Agric. Genomics 1: 1–10. https://www.cabdirect.org/cabdirect/abstract/20063161122 (accessed 7 May 2019).

Pérez-Flores, L., F. Carrari, R. Osuna-Fernández, M.V. Rodríguez, S. Enciso, et al. 2003. Expression analysis of a GA 20-oxidase in embryos from two sorghum lines with contrasting dormancy: Possible participation of this gene in the hormonal control of germination. J. Exp. Bot. 54(390): 2071–2079. doi: 10.1093/jxb/erg233.

Rodríguez, M.V., M. Margineda, J.F. González-Martín, P. Insausti, and R.L. Benech-Arnold. 2001. Predicting preharvest sprouting susceptibility in barley. Agron. J. 93(5): 1071. doi: 10.2134/agronj2001.9351071x.

Rodríguez, M.V., G.M. Mendiondo, R. Cantoro, G.A. Auge, V. Luna, et al. 2012. Expression of seed dormancy in grain sorghum lines with contrasting pre-harvest sprouting behavior involves differential regulation of gibberellin metabolism genes. Plant Cell Physiol. 53(1): 64–80. doi: 10.1093/pcp/pcr154.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Sato, K., M. Yamane, N. Yamaji, H. Kanamori, A. Tagiri, et al. 2016. Alanine aminotransferase controls seed dormancy in barley. Nat. Commun. 7(1): 11625. doi: 10.1038/ncomms11625.

Shorinola, O., B. Balcárková, J. Hyles, J.F.G. Tibbits, M.J. Hayden, et al. 2017. Haplotype analysis of the pre-harvest sprouting resistance locus Phs-A1 reveals a causal role of TaMKK3-A in global germplasm. Front. Plant Sci. 8: 1555. doi: 10.3389/fpls.2017.01555.

Song, S., G. Wang, H. Wu, X. Fan, L. Liang, et al. 2020. OsMFT2 is involved in the regulation of ABA signaling-mediated seed germination through interacting with OsbZIP23/66/72 in rice. Plant J. 103(2): 532–546. doi: 10.1111/tpj.14748.

Spielmeyer, W., M. Ellis, M. Robertson, S. Ali, J.R. Lenton, et al. 2004. Isolation of gibberellin metabolic pathway genes from barley and comparative mapping in barley, wheat and rice. Theor. Appl. Genet. 109(4): 847–855. doi: 10.1007/s00122-004-1689-6.

Sreenivasulu, N., B. Usadel, A. Winter, V. Radchuk, U. Scholz, et al. 2008. Barley grain maturation and germination: Metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. Plant Physiol. 146(4): 1738–1758. doi: 10.1104/pp.107.111781.

Van Eeuwijk, F.A., M. Malosetti, X. Yin, P.C. Struik, and P. Stam. 2005. Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. Aust. J. Agric. Res. 56: 883–894. doi: 10.1071/AR05153.

Vetch, J.M., J.G. Walling, J. Sherman, J.M. Martin, and M.J. Giroux. 2020. Mutations in the HvMKK3 and HvAlaAT1 genes affect barley preharvest sprouting and after-ripened seed dormancy. Crop Sci. doi: 10.1002/csc2.20178.

Walker-Simmons, M., and J. Sesing. 1990. Temperature effects on embryonic abscisic acid levels during development of wheat grain dormancy. J. Plant Growth Regul. 9(1): 51–56. doi: 10.1007/BF02041941.

Wei, W., X. Min, S. Shan, H. Jiang, J. Cao, et al. 2019. Isolation and characterization of TaQsd1 genes for period of dormancy in common wheat (Triticum aestivum L.). doi: 10.1007/s11032-019-1060-x.

Xi, W., C. Liu, X. Hou, and H. Yu. 2010. MOTHER OF FT and TFL1 regulates seed germination through a negative feedback loop modulating ABA signaling in Arabidopsis. Plant Cell 22(6): 1733–1748. doi: 10.1105/tpc.109.073072.

CHAPTER 5

EVALUATION OF GENETIC GAIN USING A SELECTION INDEX IN GENOMIC SELECTION OF TWO-ROW SPRING BARLEY

Abstract

New breeding programs are faced with many challenges, including evaluation of unknown germplasm, initiation of breeding populations that will satisfy short- and long-term breeding goals, and implementation of efficient phenotyping strategies for multiple traits. Genomic selection (GS) is a potentially valuable tool for recently established breeding programs to quickly accelerate genetic gain. Genomic selection using selection index values may increase gain over phenotypic selection but empirical studies remain limited. We compared gain in overall selection index value for height, heading date, preharvest sprouting (PHS) resistance, and spot blotch resistance and component traits in two cycles of GS with one round of phenotypic selection (PS) in two-row spring malting barley. Genomic selection increased gain for selection index value, height, and PHS but produced no gain for heading date and spot blotch. Genetic variance for lightly weighted traits was not reduced with GS but was substantially reduced for heavily weighted PHS and correlated seed germination traits. Inbreeding was increased by GS compared to PS but restricted mating of high breeding value individuals limited potential inbreeding. Our results indicate GS can result in gain for component traits and overall selection index value.

Introduction

Applied plant breeding programs must simultaneously select multiple traits to competitively produce superior varieties. A breeder can select multiple traits in a stepwise fashion, such as in tandem selection or independent culling levels, but simultaneous selection on multiple traits using an index of net merit is the most efficient method (Lush and Hazel, 1942). In addition to the necessity of multi-trait selection, plant breeders are sometimes faced with the need to rapidly produce improved germplasm. Introduction or breakdown of biotic stressors, new target populations of environment, and shifting consumer preferences can make successful varieties obsolete within the course of a few growing seasons. New breeding programs are also faced with the challenge of efficiently producing competitive material, generating genetic variance for long term gain, and purging highly deleterious alleles from nascent breeding populations.

Marker assisted selection (Lande and Thompson, 1990) and genomic prediction (Bernardo, 1994; Meuwissen et al., 2001; Dekkers 2007) are extensions of selection index theory that estimate breeding value based on the combination of phenotypes and marker-derived genotypes. Genomic prediction of genomic estimated breeding values (GEBVs) in unobserved lines or environments and GS on GEBVs are theoretically able to increase genetic gain compared to PS by increasing selection accuracy of Mendelian sampling (Daetwlyer et al., 2011) and reducing cycle time (Heffner et al., 2010). Empirical gain from GS experiments in plants suggest GS and PS may perform similarly (Rutkoski et al., 2015; Sallam & Smith, 2016) but reduced cycle time in GS leads to greater gain per unit time. Use of a selection index in GS

may result in gain for component traits (Combs and Bernardo, 2013), index value (Massman et al., 2013), or both for positively (Hernandez et al., 2020) and negatively genetic correlated traits (Tiede and Smith, 2018). Prediction accuracy for selection index value may be higher with GS than PS even if single trait accuracies are similar between the two methods (Heffner et al., 2011).

Malting barley has a long production history in New York (Harlan et al., 1925) but acreage in modern times was negligible until the New York Farm Brewery bill initiated demand for locally produced malting barley starting in 2012. The humid continental climate of New York state makes production of high-quality malting grade barley a challenge due to disease pressure and PHS. Malting quality and PHS are often negatively correlated (Gao et al., 2003; Li et al., 2004; Castro et al., 2010) but many malting quality QTL are independent of PHS QTL (Mohammadi et al., 2015). Seed germination rate can be used as a proxy for some malting quality traits (Woonton et al., 2005). The correlated response of malting quality and seed germination traits from selection on PHS is unknown. Currently available spring malting barleys are adapted to drier climates of the western United States and Canada and have inconsistent performance in New York. To meet the new demand for New York malting barley, a two-row spring barley breeding program was initiated at Cornell University in 2016.

The need for improvement of multiple traits for local adaptation in a short time frame suggested a GS strategy. The objectives of this study were to compare genetic gain using a base selection index across two cycles of recurrent GS with a single cycle of phenotypic selection (PS), evaluate correlated response to selection of seed

germination traits, and examine the consequences of GS on inbreeding and genetic variance.

Materials and methods

Population development and selection scheme

The development and phenotyping of the training population, referred to here as TP17, is described in detail in Chapter 2. In brief, eight spring two-row malting barley varieties were selected as founder lines for C0: 'Conlon' (PI 597789), 'Pinnacle' (PI 643354), 'ND Genesis' (PI 677345), 'Craft' (PI 646158), 'Newdale' (Legge et al., 2008), 'AAC Synergy' (Legge et al., 2014), 'Bentley' (Juskiw et al., 2009), and 'KWS Tinka' (PI 681721). AAC Synergy was used as a female in all crosses to create a seven-family connected half-sib population. TP17 was composed of 1341 $F_{3:4}$ lines planted in two locations in Ithaca, NY in 2017. TP17 was phenotyped for plant height, heading date, leaf rust (*Puccinia hordei* G. H. Otth), PHS, grain protein, and spot blotch *(Bipolaris sorokiniana* Sacc) and genotyped with 6009 genotyping-by-sequencing (GBS) markers as described in Chapter 2. Genotypic (GEBV) and phenotypic estimated breeding values (EBV) for TP17 lines were predicted with GBLUP models (Equation 5.1 below) incorporating data from all TP17 lines. A base selection index (see 'Selection Index' below) was calculated using either GEBVs or EBVs. The AAC Synergy/Bentley and AAC Synergy/Newdale families were omitted from selection due to close parental relatedness and subsequent low genetic variation. The two lines within each of the remaining five families (0.7% selection intensity) with the highest GEBV and EBV selection index value were selected as parents for

Cycle 1 genomic (C1G) and Cycle 1 phenotypic (C1P), respectively. Within selection cohort, each selected C0 line was intermated with the eight half-sib lines to avoid full-sib crosses, resulting in 39 C1G crosses and 37 C1P crosses. C1G and C1P F1 seed was planted in the greenhouse. Six F2 seeds per C1P cross (n=211) and 9-10 F2 seeds per C1G cross (n=368) were planted in the greenhouse. The C1G plants were tissue sampled and genotyped using a subset of the GBS markers (see 'Genotyping' section). Prior to prediction of C1G lines, the GBLUP model was updated using 2018 replicated regional trial data from 100 TP17 and five founder lines evaluated in two to four locations. The top 15 F2 lines (4% selection intensity) were selected as Cycle 2 genomic (C2G) parents and intermated in the F2 generation to make 22 crosses. Base selection index weights were modified when selecting parents from C1G to emphasize greater disease resistance. The predicted top 25% C1G lines (n=92) were advanced to the F3 stage and the remaining plants were discarded. Six F2 seeds per C2G cross were planted and F2:F3 seed, along with F4:F5 seed of C1G and C1P, was harvested in spring 2019. A subset of TP17 labeled C0 (15 lines each from seven families, n=105), C1G (n=92), a random subset of C1P (n=106), and C2G (n=100) were planted in two locations in Ithaca, NY in 2019 and 2020 in an augmented block design with the eight founder lines arranged in a randomized complete block design as checks. This population will be referred to as $Cycle_{all}$ (C_{all}, Figure 5.1). Trials in 2020 received irrigation in June due to drought conditions.

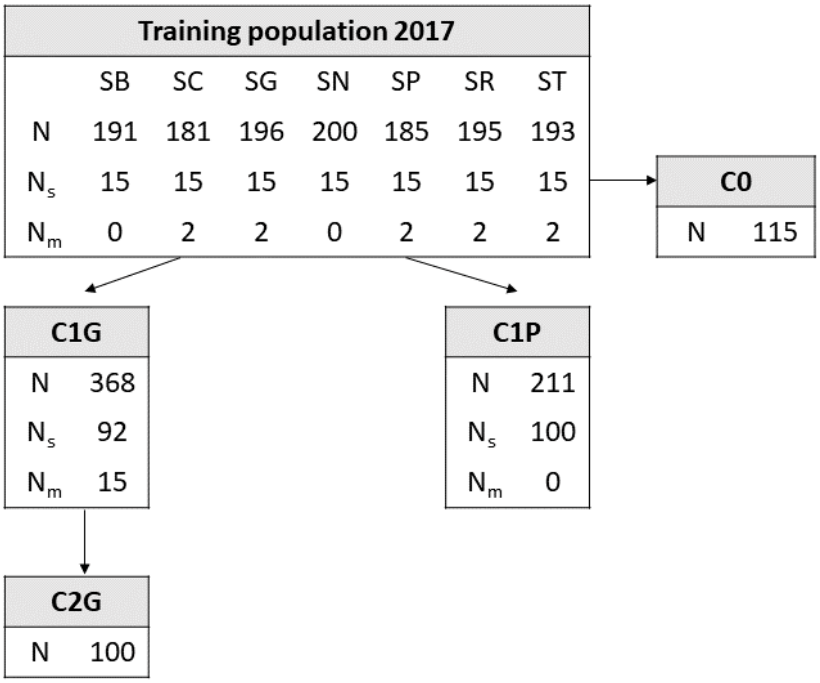


Figure 5.1: Selection scheme and composition of $C_{all.}$ N indicates the total number of lines in each population, N_s indicates the number of lines selected for inclusion in C_{all}, and N_m indicates the number of lines selected to intermate to create the next cycle. Family names in training population are provided above N.

Phenotyping

Heading date was recorded as the Julian date that 50% of the spikes in a plot had fully emerged from the flag leaf. Height was recorded as the average plot height from the ground to the tip of the spike, excluding awns. Preharvest sprouting was measured on a 0 (resistant) to 9 (susceptible) scale according to Anderson et al. (1993) on 5 spikes sampled at physiological maturity (loss of green color from the peduncle), after-ripened for three days, and misted in a greenhouse chamber for three days. 2020 PHS C_{all} phenotypes were incomplete due to COVID pandemic induced labor restraints. Germination energy (GE) and germination index (GI) were measured in 2019 and 2020 on samples collected two days after physiological maturity, dried for one day,

hand threshed, and frozen. 4 mL water and 30 kernels were added to petri plates with two Whatman filter papers. Germination, defined as the emergence of 2 mm of the radicle, was counted each day for three days at 6 (PM_6), 48 (PM_{48}) , and 111 (PM_{111}) days after physiological maturity. Germination energy, a measure of seed dormancy, was calculated as the percentage of kernels that germinated after three days. Germination index, a measure of germination rate, was calculated as $10*(\Sigma(n_{24} + n_{48} + n_{72})/ (n_{24} + 2n_{48} + 3n_{72}))$ where n_{24} is the number of kernels germinated at 24 hours, n_{48} is the number of kernels germinated at 48 hours, and n_{72} is the number of kernels germinated at 72 hours. See Chapter 3 for more details on germination trait phenotyping. Leaf rust was not observed in 2019 or 2020 and was not phenotyped. Grain protein was not phenotyped in 2019 or 2020 due to labor constraints.

Spot blotch isolate Bs233NY15 was used to evaluate seedling stage reaction. This isolate was collected in central New York in 2015 and is an aggressive isolate that produces symptoms characteristic of field symptoms found across New York state. Freezer stock of Bs233NY15 was removed from -80°C and plated to potato dextrose agar with antibiotics (neomycin and streptomycin). Plates were incubated for seven days at room temperature (20-22°C) with 12 hour black light, 12 hour darkness. After one week of growth, plates were flooded with sterile deionized water. 10-20 µl of spore suspension was removed from the flooded plates and sub-cultured to V8 agar. A lawn of spore suspension was made with a sterile cell spreader and sub-plates were incubated at room temperature (20-22°C) with 12 hour white light, 12 hour darkness. After ten days of growth the V8 sub-plates were flooded with sterile deionized water and the spore suspension was collected in a sterile container. Once enough suspension

liquid was collected, a haemocytometer was used to calculate the initial concentration. The final concentration was adjusted to 1 x 10^4 (or 10,000) conidia per ml. 1 mL Tween (Croda International PLC) was added as a surfactant.

Three replicates of C_{all} and founder lines were planted one day apart in a randomized complete block design in greenhouse facilities at Cornell University. Three seeds per cell were sown in 24-well flats. Before inoculation each cell was thinned to two plants. Fourteen-day old seedlings were inoculated using a Preval sprayer (Chicago Aerosol, Coal City, IL). Approximately 120-130 ml was used for every four 24-well flats. Flats were lightly sprayed with pure water once inoculum had dried and covered in plastic bags for one day to maintain high local humidity. The second full leaf was scored according to Fetch and Steffenson (1999) at ten days post-inoculation. The cell mean of two plants was recorded and used for further analysis.

Genotyping

The initial 6009 GBS marker set used to train genomic prediction models in TP17 was detailed in Chapter 2. This marker set will be referenced as G_1. To genotype C1G lines, a subset of the GBS markers were converted into SNPSeq assays (LGC Genomic, Berlin, Germany). Markers at least 100 bp apart that were polymorphic in the selected C1G parents were randomly sampled across each chromosome and 768 markers were converted for SNPSeq genotyping. Missing markers were imputed using the "EM" imputation method in the A.mat function in *rrblup* (Endelman et al., 2011) and monomorphic marker calls were corrected within C1G family based on the parents. This marker set will be referenced as G_2 and was used to select among C1G

lines and to select parents to intermate for C2G. C_{all} was genotyped with the 50k Illumina SNP array (Bayer et al., 2017) at the USDA Small Grains Genotyping Laboratory in Fargo, ND. Marker positions were based on the Morex v2 genome assembly (Monat et al., 2019). 20,929 polymorphic markers were obtained for C_{all} and founders after filtering for heterozygosity over 10% and minor allele frequency less than 0.05. This marker set will be referenced as G_3.

Statistical analysis

All statistical analyses were completed using R version 3.5.1 (R Core Team, 2018). All statistical models were fit with *ASReml-R* version 3 (Gilmour et al., 2009) unless otherwise noted.

Genomic prediction

Genomic best linear unbiased prediction (GBLUP) models were fit as

$$y = \boldsymbol{X}\beta + \boldsymbol{Z_1}u_1 + \boldsymbol{Z_2}u_2 + \epsilon \qquad (5.1)$$

where y is a vector of unadjusted phenotypes, β is a vector of fixed location and environmental effects, u_1 is a vector of random block effects, and u_2 is a vector of random entry effects. Random effects u_1 and u_2 were assumed to be normally distributed with $u_1 \sim \mathrm{N}(0, \sigma_b^2)$ where σ^2_b is the block variance and $u_2 \sim \mathrm{N}(0, \mathbf{G}\sigma_a^2)$ where **G** is the genomic relationship matrix calculated from marker data using the first method in VanRaden (2008), and σ^2_a is the additive genetic variance. **X** is an incidence matrix relating fixed effects to phenotypes, $\mathbf{Z_1}$ is an incidence matrix relating random block effects to phenotypes, and $\mathbf{Z}_2$ is an incidence matrix relating

genetic effects from the genomic relationship matrix to phenotypes. The vector of residuals ε is assumed to be normally distributed with $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ where σ^2_ε is the error variance. Equation 5.1 was used to select TP lines to initiate C1G and C1G lines to initiate C2G. A similar model was fit to select TP lines to initiate C1P where $u_2 \sim N(0, \sigma_g^2)$. Equation 5.1 was also used to assess GS accuracy, the Pearson correlation of GEBVs and phenotypic EBVs, within C0 and C1G using C_{all} data. PS accuracy was calculated as $h = \sigma_a/\sigma_p$ where σ_a is the square root of the additive genetic variance and σ_p is the square root of the phenotypic variance with $\sigma_p = \sigma_a + \sigma_e/l$ where l=4 is the number of locations.

Selection index

A base selection index (Williams, 1962) was used to select multiple traits instead of a Smith-Hazel index because phenotypic records were incomplete in the initial TP and phenotypic covariances could not be accurately calculated. The general form of the index, *I*, is given as

$$I = \Sigma a_i z_i \quad (5.2)$$

where *a* is an economic weight, *z* is a matrix of GEBVs for each trait, and *i* indicates the number of traits under selection. Economic weights were arbitrarily assigned based on breeding priorities. In TP17, selection index 1 (SI1) was calculated with weights of 1 for heading date, -2 for height, -6 for PHS, -4 for spot blotch, -2 for leaf rust, and -2 for grain protein. Index weights were modified in selection index 2 (SI2) for selection of C2G parents with weights of 1 for heading date, -1 for height, -4 for PHS, -6 for spot blotch, -4 for leaf rust, and -3 for grain protein This index reflected updated

breeding program priorities of improved disease resistance and maintenance of malting quality variation by reducing the weight on PHS.

Realized gain and correlated response

Adjusted cycle means for each trait in TP, C0, C1P, C1G, and C2G were calculated as $p_l + \mu$ from the model

$$y_{ijkl} = \mu + \beta_i + g_j + r_k + p_l + \epsilon_{ijkl} \quad (5.3)$$

where y_{ijkl} is a vector of unadjusted phenotypes, μ is the overall mean, β_i is the fixed environmental effect, g_j is the random genotype effect with $g_j \sim \mathrm{N}(0, \sigma_\mathrm{g}^2)$ where σ_g^2 is as before, r_k is the heterogeneous random block within environment effect with $r_k \sim \mathrm{N}(0, \sigma_\mathrm{r}^2)$ where σ_r^2 is the block variance, p_l is the fixed population effect, and ε_{ijkl} is the heterogeneous random error with $\epsilon_{ijkl} \sim \mathrm{N}(0, \sigma_\epsilon^2)$ where σ_ε^2 is the residual error variance. To calculate adjusted cycle means for selection indices, Equation 5.3 was fit with the p_l term omitted to extract entry BLUPs for each trait because phenotypic records were incomplete in several cycles. Selection index cycle means were then calculated with

$$y_{ij} = \mu + p_i + \epsilon_{ij} \quad (5.4)$$

where y_{ij} is a vector of selection index values calculated from adjusted phenotypes from Equation 5.3, μ is the overall mean, p_i is the fixed cycle effect, and ε_{ij} is the heterogeneous random error as before. Correlated response to selection for GE and GI at PM_6, PM_{47}, and PM_{111} was calculated using Equation 5.3 with r_k omitted and a single error variance estimate.

Realized gain from selection was calculated by subtracting cycle means from either the TP17 or C0 mean. Percentage gain per cycle was calculated by dividing the realized gain per cycle by TP17 or C0 mean. To test if fixed cycle effects were significant (different from zero), a Wald test was conducted in *ASReml-R*. The Kenward-Roger degrees of freedom were extracted from the wald() object and used to determine a p-value using the t-distribution. T values were calculated as the fixed effect estimate divided by the standard error.

Variance component estimation

Genetic variance, $\sigma^2_{g,}$ was estimated by first removing nongenetic effects using Equation 5.5 and then fitting residuals from Equation 5.5 as the response variable in Equation 5.6. Equation 5.6 was used to estimate σ^2_g in cycle specific models for each trait.

$$y_{ij} = \mu + \beta_i + r_j + \epsilon_{ij} \quad (5.5)$$

where y_{ij} is a vector of unadjusted phenotypes, β_i is the fixed environmental effect, r_j is the heterogeneous random block within environment effect with $r_j \sim \mathrm{N}(0, \sigma^2_\mathrm{r})$ within each environment where σ^2_r is the block variance, and ε_{ij} is the heterogeneous random error within environment with $\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2_\epsilon)$ where σ^2_ϵ is the residual error variance.

$$y'_i = \mu + g_i + \epsilon_i \quad (5.6)$$

where y'_i is a vector of residuals from Equation 5.5, g_j is the random heterogeneous genotype effect within cycle with $g_j \sim \mathrm{N}\left(0, \sigma^2_\mathrm{g}\right)$, and ε_i is the random error within cycle with $\epsilon_i \sim \mathrm{N}(0, \sigma^2_\epsilon)$. Additive genetic variance, σ^2_a , was calculated with the same

models except in Equation 5.6 $g_j \sim \mathrm{N}(0, \mathbf{G}\sigma_a^2)$ where σ_a^2 and **G** are as before. $\sigma_a^2 + \sigma_\epsilon^2$ was used to calculate σ_p^2. A likelihood ratio test was used in *ASReml-R* to test for equal variance between cycles by comparing a model with heterogeneous cycle variances and a model with a single cycle variance component. A Pr(Chisq) < 0.05 indicated variance components were significantly different.

Genetic correlation

Bivariate mixed models were fit to determine genetic correlations (r_g) between PHS, height, heading date, and germination traits at different time points within C_{all} using a *corgh()* variance structure for genetic effects. Seedling spot blotch data from the greenhouse was not included. This model was compared to a model with a *diag()* variance structure for genetic effects (no correlation between genetic effects) using a likelihood ratio test. Although not a formal significance test, this was used to determine if estimating genetic correlations improved model fit. Environmental genetic correlations for each trait were assessed with the Pearson correlation (r_p) of genetic values estimated in each environment. Phenotypic correlations were estimated from C_{all} line means across environments and significance was tested using the rcorr() function in the *Hmisc* R package (Harrell, 2021).

Expected gain from selection

Expected response to selection for multiple traits in a base selection index ($\boldsymbol{E_B}$) was estimated by

$$\boldsymbol{E_B} = i\frac{\boldsymbol{Gw}}{\sqrt{\boldsymbol{w'Pw}}} \quad (5.7)$$

where *i* is the selection intensity, **G** is the covariance matrix of breeding values, **w** is a vector of trait weights, and **P** is the phenotypic covariance matrix (Ceron-Rojas & Crossa, 2018, p. 28). The heritability of the base selection index was calculated as $h^2_{I_B} = \frac{\boldsymbol{w'Gw}}{\boldsymbol{w'Pw}}$.

Expected and realized inbreeding

The expected pedigree coefficient of relationship (r_A) was calculated within each cycle using the additive relationship matrix **A** in the *pedigreemm* R package (Bates et al., 2015) and average r_A was calculated as the mean of the off-diagonal elements within cycle. Realized inbreeding was measured in two ways. First, the number of monomorphic markers in each cycle was calculated. Second, the average genomic coefficient of relationship (r_G) within cycle was calculated as the mean of the off-diagonals of the scaled genomic relationship matrix $\mathbf{G_s}$. The G_1 and G_3 derived **G** matrices contained negative off-diagonal elements, indicating that some pairs of individuals were less molecularly related in the sense of identity by state than average pairs (Yu et al., 2017). To accurately compare **A** and **G** and set off-diagonal elements to zero, $\mathbf{G_s}$ was calculated as

$$G_{s_{ij}} = \frac{(G_{s_{max}} - G_{s_{min}})(G_{ij} - G_{min})}{G_{max} - G_{min}} \quad (5.8)$$

where $\mathbf{G_{ij}}$ is the i^{th}, j^{th} element of **G**, and $\mathbf{G_{max}}$ and $\mathbf{G_{min}}$ are the maximum and minimum values of **G**. $\mathbf{G_{smax}}$ and $\mathbf{G_{smin}}$ specify the maximum and minimum values of $\mathbf{G_s}$, set here to 2 and 0 (Momen et al., 2017). Two-sided t-tests were used to test differences between r_A and r_G within cycle.

Expected r_A was simulated for C2G in a scenario where full-sib mating was not restricted. The top ten TP17 lines ranked by selection index value were selected as parents for $C1G_{sim}$. All possible crosses (n=45) were made with 8 lines per family. $C2G_{sim}$ was then simulated 500 times. In each replication, one third of the $C1G_{sim}$ families were randomly dropped to represent selection and 15 lines from the remaining families were selected as parents. 22 crosses, excluding selfs, with 6 lines per family were simulated. r_A was calculated within $C1G_{sim}$ and $C2G_{sim}$ as described above. The number of crosses and lines within each family were similar to observed structures of C1G and C2G.

Results

Realized genetic gain trial

Within C_{all} cycle, broad sense heritabilities (H^2) were generally high (0.62-0.92, Table 5.1). The heritabilities, $h_{I_B}^2$, of SI1 estimated using TP17 and C0 data were 0.37 and 0.55, respectively. $h_{I_B}^2$ for SI2 estimated using C1G data was 0.74. Average environmental correlations across the four C_{all} environments were high for PHS (r=0.795), moderate for HD (r=0.56), and low for height (r=0.364). C0 was used as a proxy for TP17 in C_{all} to make phenotyping manageable. Cycle means were similar between C0 and TP17 except for TP17 which had later average HD due to a later planting date than C_{all} locations. Genetic variance estimates were also similar between the two populations for most traits, although C0 had higher height variance. Height and spot blotch H^2 was lower in TP17 than C0. Overall, C0 effectively captured the variation present in the larger TP17. Genotypic and phenotypic correlations between

PHS and germination traits were generally high and significant (Chapter 3, Table 5.2) except for GE at PM_{111} when dormancy had fully broken and almost all lines had GE > 0.95. Heading date had small but significant genetic correlations with height and GE at PM_{48} and height had small but significant genetic correlations with GE at PM_6 and PM_{48} and GI at all PM dates.

Gain from selection

Genetic gain in C1P and C2G was significantly different from zero for HD, height, and PHS compared to TP17 and for height and PHS compared to C0 (Table 5.3). Gain in SI1 and SI2 was significant for C2G compared to TP17 and C0, but C1P only showed significant gain for SI1 compared to TP17. Significant gain for spot blotch was not observed for any cycle. The greatest percentage gain in selection index traits after two cycles of GS was observed for PHS (-73.9%). C1P gains were significantly different than C2G for PHS, SI1, and SI2 but not HD ($p=0.2$), height ($p=0.8$), or spot blotch ($p=0.14$). Realized GS accuracy was high for C0 (heading date=0.849, height=0.861, PHS=0.785, spot blotch=0.931, SI1=0.809, SI2=0.724) and variable for C1G (heading date=0.757, height=0.672, PHS=0.706, spot blotch=0.287, SI1=0.631, SI2=0.536). Realized PS for C0 was generally lower than realized GS (heading date=0.743, height=0.632, PHS=0.929, spot blotch=0.615, SI1=0.742, SI2=0.86). Average difference between expected and observed gain in C1G was not significantly different ($p<0.05$) when calculated in reference to TP17 or C0. Expected gain was generally similar to observed gain for C1G and C1P with the exception of C1P PHS (Table 5.4). Expected gain was not similar to observed gain for any trait in C2G.

Table 5.1: Summary of realized genetic gain cycles and the initial training population (TP17). r_G is mean genomic coefficient of relationship and r_A is mean pedigree coefficient of relationship with significant difference ($p<0.05$) indicated by asterisk.

Cycle	**n**	**Parameter***	**Heading date (Julian days)**	**Height (cm)**	**PHS (0-9)**	**Spot blotch (1-9)**	r_G	r_A	**Monomorphic marker %****
TP17	1341	μ	179.4	65.9	3.29	3.49	0.253*	0.321	0.03%
		σ_g^2	5.64	8.05	2.77	0.59			
		H^2	0.72	0.43	0.84	0.36			
C0	105	μ	177.6	65.8	3.22	3.22	0.304*	0.317	2.21%
		σ_g^2	6.56	16.44	4.13	1.3			
		H^2	0.76	0.7	0.92	0.86			
C1G	92	μ	177.6	61.0	0.75	3.28	0.368	0.369	17.03%
		σ_g^2	10.12	18.99	1.3	1.22			
		H^2	0.87	0.74	0.84	0.84			
C1P	106	μ	177.3	63.4	2.38	3.35	0.379*	0.349	9.62%
		σ_g^2	6.99	14.56	3.79	1.44			
		H^2	0.76	0.62	0.91	0.86			
C2G	100	μ	177.8	63.6	0.86	3.12	0.413*	0.401	21.72%
		σ_g^2	7.85	12.84	0.93	1.19			
		H^2	0.84	0.66	0.8	0.85			

*μ=cycle mean, σ_g^2 = genetic variance, H^2 = broad-sense heritability
**: Percent monomorphic markers out of 20,929

Table 5.2: Genetic (lower triangle) and phenotypic (upper triangle) correlations in C_{all}

	HD	Height	PHS	GE_6	GE_48	GE_111	GI_6	GI_48	GI_111
HD		-0.37*	0.09	0.11*	-0.06	-0.11	0.13*	0.09	0.07
Height	-0.113*		-0.01	0.03	0.15*	0.1	0.01	0.11	0.14*
PHS	0.108*	0.105		0.77*	0.32*	-0.05	0.84*	0.62*	0.56*
GE_6	-0.045	0.128*	0.836*		0.42*	-0.02	0.97*	0.75*	0.65*
GE_48	-0.14*	0.12*	0.375*	0.551*		0.19*	0.39*	0.72*	0.57*
GE_111	0.16	0.204	0.03	0.02	0.177		-0.05	-0.06	0.03
GI_6	0.017	0.118*	0.908*	0.979*	0.474*	-0.07		0.76*	0.67*
GI_48	-0.078	0.234*	0.659*	0.868*	0.831*	-8.4e-5	0.83*		0.85*
GI_111	-0.056	0.333*	0.661*	0.835*	0.715*	-0.118	0.816*	0.963*	

Inbreeding and genetic variance

Expected r_A in each cycle was significantly different than the observed r_G for all cycles except C1G (Table 5.1). Average r_G was highest among C2G lines (0.413) and lowest among TP17 (0.252). Relationship significantly increased in C1P and C2G ($p<0.001$) compared to C0. Relatedness to the TP17 common parent, AAC Synergy, significantly increased from C0 (r_G =0.308) to C2G (0.414) and C1P (0.383). Monomorphic marker percentage increased from C0 (2.21%) to C1P (9.62%), C1G (17.03%), and C2G (21.72%). C2G had significantly different genetic variance than C0 and C1P for PHS but not for heading date, height, or spot blotch.

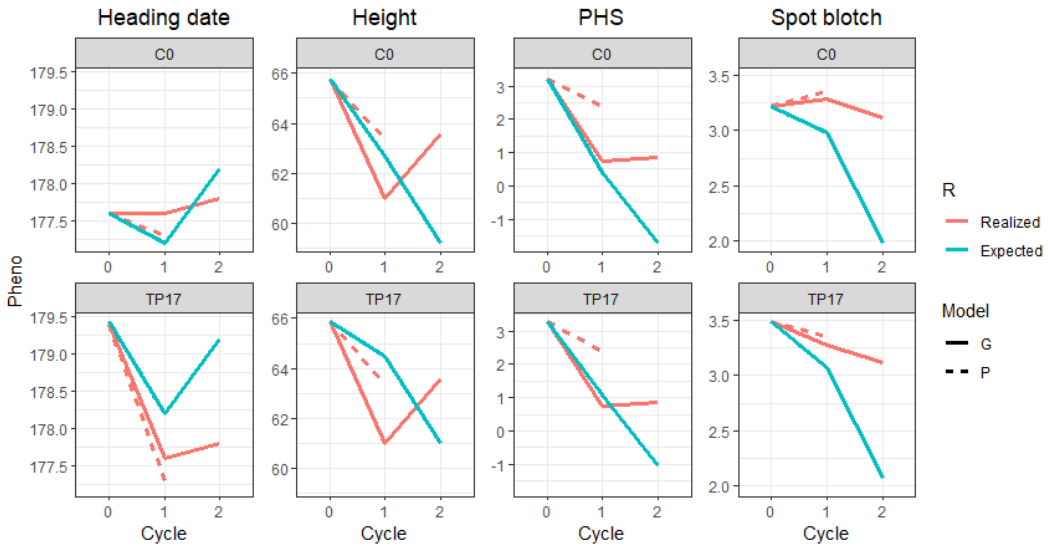


Figure 5.2: Realized and expected genetic gain for selection index traits. Solid lines indicate genomic selection cycles and dashed lines indicate phenotypic selection cycles.

Table 5.3: Realized gain and percent gain per cycle in reference to TP17 and C0. Wald tests were used to test overall significance of 'Cycle' as a fixed effect (Pr(Chisq) < 0.05), indicated by an asterisk in the column header. Specific cycle fixed effects were tested using a t-test with Kenward-Roger denominator degrees of freedom estimation and cycles with a fixed effect significantly different from zero ($p < 0.05$) are indicated by an asterisk within rows.

Cycle	Reference cycle	Heading date*	Height*	PHS*	Spot blotch	SI1	SI2
TP17 μ		179.44	65.88	3.29	3.49	-3.34	-3.6
C0 μ		177.63	65.77	3.22	3.22	-2.46	-1.32
C1P	TP17	-2.11 (-1.17%)*	-2.43 (-3.69%)*	-0.9 (-27.35%)*	-0.14 (-4.01%)	2.86 (85.7%)*	2.79 (77.4%)
	C0	-0.3 (-0.17%)	-2.32 (-3.53%)*	-0.83 (-25.78%)*	0.13 (4.04%)	1.98 (80.6%)	0.5 (38.2%)
C1G	TP17	-1.88 (-1.05%)*	-4.88 (-7.41%)*	-2.54 (-77.2%)*	-0.21 (6.02%)	14.4 (431%)*	10.2 (284.4%)*
	C0	-0.07 (-0.04%)	-4.77 (-7.25%)*	-2.47 (-76.71%)*	0.06 (1.86%)	13.52 (549.5%)*	6.04 (603.7%)*
C2G	TP17	-1.63 (-0.91%)*	-2.29 (-3.48%)*	-2.43 (-73.86%)*	-0.37 (10.6%)	11.12 (332.9%)*	9.05 (251.3%)*
	C0	0.18 (0.1%)	-2.18 (-3.31%)*	-2.36 (-73.29%)*	-0.1 (3.1%)	10.24 (416.3%)*	6.77 (513.5%)*

Table 5.4: Expected gain for individual traits in a selection index. E_{S1} is the expected gain from selection index 1 for each individual trait parameterized from TP17 or C0 data. E_{S2} is expected gain from selection index 2 parameterized with C1G data.

Cycle	Heading date	Height	PHS	Spot blotch
E_{S1} TP17	0.596	-1.291	-2.156	-0.149
E_{S1} C0	-0.419	-3.089	-2.81	-0.24
E_{S2}	1.049	-3.48	-2.13	-1.0

Correlated response of germination traits

Selection on an index heavily weighted towards PHS resistance resulted in significant reductions in GE at PM_6 and PM_{48} and GI at all timepoints in C1G and C2G (Table 5.6, Figure 5.2). C1P means were not significantly different than C0 and were

significantly higher ($p<0.01$) than C2G for all germination traits. C2G had significantly different genetic variance than C0 and C1P for GE at PM_{48} and GI at PM_6 and PM_{48}.

Table 5.5: Correlated response from selection for germination traits and percent gain per cycle from C0. Wald tests were used to test overall significance of 'Cycle' as a fixed effect (Pr(Chisq) < 0.05), indicated by an asterisk in the column header. Specific cycle fixed effects were tested using a t-test with Kenward-Roger denominator degrees of freedom estimation and cycles with a fixed effect significantly different from zero ($p < 0.05$) are indicated by an asterisk within rows.

Cycle	**GE_6**	**GE_{48}**	**GI_6**	**GI_{48}**	**GI_{111}**
C0 μ	0.757	0.989	3.735	5.353	5.822
C1P	-0.029 (-3.83%)	-0.002 (-0.2%)	-0.237 (-6.3%)	-0.096 (-1.8%)	-0.07 (-1.2%)
C1G	-0.322 (-42.54%)*	-0.045 (-0.45%)*	-2.106 (-56.4%)*	-1.025 (-19.2%)*	-0.46 (-7.9%)*
C2G	-0.312 (-41.21)*	-0.029 (-0.29)*	-2.089 (-55.9%)*	-0.714 (-13.3%)*	-0.28 (-4.9%)*

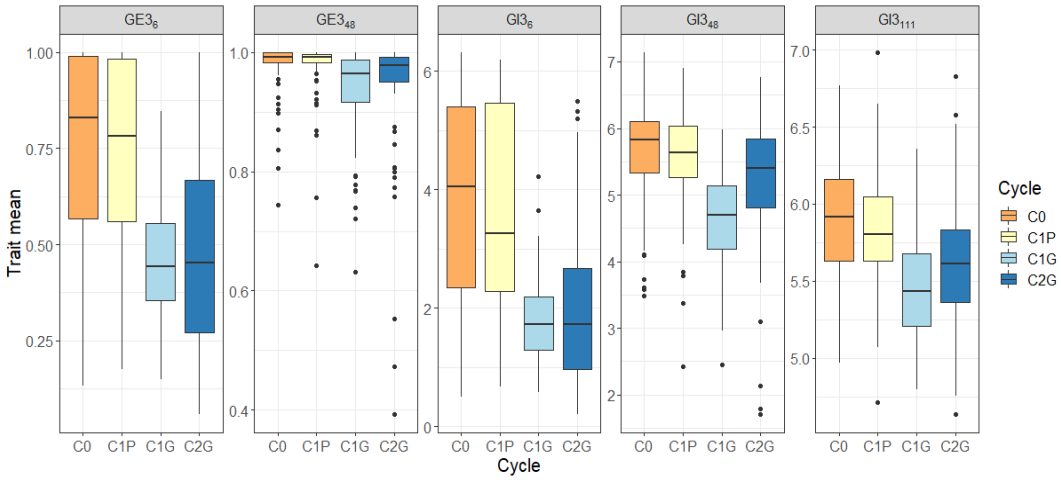


Figure 5.3: Distributions of correlated germination traits per selection cycle.

Discussion

Genetic gain from two cycles of recurrent GS was compared with one cycle of PS in four environments for four uncorrelated traits in a selection index. In addition to examining gain from GS, this population was the primary germplasm in a newly established two-row spring malting barley breeding program. As a result, several breeding program decisions impacted the design of the genetic gain experiment. Selection cycles were not replicated, which made the impacts of drift unknown and did not allow for statistical comparison of gain between GS and PS. The selection index weights were updated for C2G parental selection based on updated breeding priorities. This complicated the estimation of expected gain on a selection index, particularly from C1G to C2G. The C1G population used in C_{all} was selected and represents a quarter of the initial C1G. Estimates of expected gain from C1G to C2G may be biased because of this.

The base selection indices included weights for leaf rust resistance and grain protein but these traits were not phenotyped in C_{all}. These phenotyping omissions affect the expected and observed gain for selection indices. Additional sources of error in selection index calculation include incomplete PHS phenotyping in C_{all} 2020 and subsequent error in **P** calculation. Although the increased selection index mean in C2G compared with C1P must be qualified, percentage gain in selection index value was far higher than component traits. Gain in selection index value and some component traits after calculating selection index with GEBVs has been previously reported (Tiede and Smith, 2018; Hernandez et al., 2020).

Discrepancies in expected and observed genetic gain

Previous empirical gain from GS studies that estimated expected gain have reported similar observed and expected gain (Combs and Bernardo, 2013; Rutkoski et al., 2015) or greater observed than expected gain (Veenstra et al., 2020). We found inconsistent expected and observed gain for several trait/cycle combinations. Phenotyping and genotyping discrepancies likely contributed to these results. Due to the time sensitive and labor-intensive nature of PHS sampling, only 36% of TP17 was phenotyped for PHS (Chapter 2). Four of the ten C1P parents were not phenotyped for PHS, resulting in an EBV equal to the population mean. One or more of these parents likely carried PHS susceptible alleles, resulting in reduced gain for PHS and subsequently selection index value in C1P. One benefit of GS is the ability to improve prediction of unphenotyped material based on genetic relationships. This is particularly attractive for traits like PHS, germination, and malting quality that require time sensitive phenotyping, are labor intensive, or are cost-prohibitive at early generations. Three of the ten C1G parents were not phenotyped for PHS. The prediction accuracy of unobserved PHS GEBVs was high enough in TP17 to correctly select PHS resistant parents and recombined progeny, leading to gain in PHS and selection index value.

Realized gain for spot blotch resistance was substantially lower than expected gain. C_{all} spot blotch reaction was phenotyped with seedling assays in greenhouse facilities to minimize environmental variation with field inoculation and avoid seed infection since accurate phenotyping of PHS and germination traits require clean seed. There was no correlation (r_p=0.05) between spot blotch reaction for lines evaluated in

both TP17 adult field and C_{all} seedling greenhouse experiments. Genome-wide association analysis results for spot blotch reaction in TP17 (Chapter 2) and C_{all} (Appendix) were very similar and other studies indicate partial shared genetic control of seedling and adult plant spot blotch resistance (Bovill et al., 2010; Visioni et al., 2020), indicating the small set of lines evaluated in TP17 and C_{all} may have been insufficient to accurately calculate correlation. The GS model was updated with 2018 regional trial data before C1G selection. Spot blotch was phenotyped in uninoculated plots in four locations and in some locations, spot blotch symptoms were potentially confused with bacterial leaf streak (*Xanthomonas translucens* pv. *translucens).* Spot blotch reaction H^2 was low in TP17 and this phenotyping error may have further compromised the predictive ability of the model. Expected gain for spot blotch reaction estimated from highly heritable greenhouse data was likely overestimated in C_{all}. The expected and observed gains in height from C1G to C2G were of opposite sign. Relaxation of the index weight, reduced prediction accuracy, and increased genetic variance for height in C1G may have allowed drift to impact gain. Expected gain in PHS from C1G to C2G was below the biological minimum value of zero, reflecting error in the calculation of **G** or **P**. Although genetic correlations between heading date, height, and PHS were low, some correlated response to selection may have also occurred.

Another potential factor contributing to discrepancies between expected and observed gain between C1G and C2G was genotyping error in C1G. Realized GS accuracy for C1G using the largely imputed G_2 marker set (HD= 0.776, height=0.392, PHS=0.557, spot blotch = 0.337) was generally lower compared to the G_3 set. The

difference in prediction accuracy between G_2 and G_3 may be due to imputation error in G_2 or the 3.5-fold greater marker coverage in G_3.

Inbreeding and genetic variance

Two cycles of GS led to a substantial increase in average marker-based relationship compared to a single cycle of phenotypic selection, similar to results in Rutksoki et al. (2015) and Veenstra et al. (2020). Inbreeding in C2G was higher than expected based on pedigree. Despite increase in inbreeding from GS, the purposeful avoidance of full-sib mating at each stage of selection limited the possible extent of inbreeding. In TP17, eight out of the top ten lines with the highest selection index calculated using GEBVs were full sibs. In C1G, two families had four lines each in the top ten selection index ranking. Expected $C1G_{sim}$ r_A if the top ten TP17 selection index lines irrespective of family had been intermated was 0.464, significantly higher than observed C1G. $C2G_{sim}$ r_A was 0.528 ± .03, significantly higher than observed r_G in observed C2G (0.413). This mating control strategy was simple but prudent given the structure of the original training population. Continuation of closed recurrent selection within C_{all} would result in further inbreeding regardless of selection method given that a single line, AAC Synergy, is in the background of every line. Average relationship to AAC Synergy increased more in two cycles of GS than one cycle of PS, but 13% of the possible mating pairs in C2G maintained r_G below 0.2 compared with 15% in C1P. Strategies like optimum contribution selection (Meuwissen 1997) may be effective at decreasing inbreeding in GS (Veenstra et al., 2020) and would be of particular value in populations with small effective population size such as C_{all}. Relaxing the selection

intensity in future generations may also be helpful. North American malting barley is known to have limited genetic diversity due to strong selection pressure on malting quality traits (Martin et al., 1991) and genomic inbreeding estimates were expected to be higher than expected from pedigree. One of the TP17 founder lines, KWS Tinka, is a European spring two-row variety and the impact of this more diverse germplasm was shown by respective average r_G in TP17 and C0 of 0.309 and 0.458 without the Tinka family and 0.253 and 0.304 with the Tinka family.

C2G had significantly reduced genetic variance for PHS and some correlated germination traits compared to C0 and C1P. Reduction in PHS genetic variance was favorable because non-adapted PHS susceptible alleles were purged but correlated reduction in germination trait variance was unfavorable. Although a small amount of PHS genetic variance remained in C1G and C2G, the minimal gain between the two cycles indicate further gain for PHS would be small as the mean value is close to the biological minimum. C2G variance estimates for heading date and height were similar to C1P and C0, indicating C2G may have retained variance for other important non-selected traits such as yield and malting quality components.

Correlated response of germination traits

Correlated response for GE and GI was expected because three alleles of a single gene, *HvMKK3*, are known to have pleiotropic effects on PHS, GE, and GI (Chapter 3). Percentage genetic gain for PHS in C2G (-73%) was higher than for germination traits (-5 to -41%). The decrease in genetic correlation between PHS and germination traits as after-ripening time increased was mirrored by a decrease in percentage gain in

germination traits. C2G average GI at PM_{111} was only 5% lower than C0 with similar genetic variance. A selection index for PHS and GI at later after-ripening stages may be beneficial for maintaining PHS resistance without further decreasing GI or impacting malting quality variance. The impact of selection for PHS on malting quality remains unknown but is a topic of continuing research. Height had significant genetic correlations of 0.118, 0.234, and 0.333 with GI3 at PM_6, PM_{48}, and PM_{111}, respectively. Pleiotropic effects could explain this relationship since gibberellin is a positive regulator of seed germination and plant height. Taller plants may also have escaped seed infection, which negatively impacts germination rate.

Conclusion

Two cycles of GS with a selection index resulted in gain in overall selection index value, low to moderate gain with maintenance of genetic variance for lightly weighted component traits, and variable results for gain and genetic variance in highly weighted traits compared to PS. Expected and observed gains were dissimilar, especially from C1G to C2G, possibly as a result of selection on C1G, incomplete index phenotypes, and phenotyping inconsistencies. The importance of good phenotyping in the TP was highlighted as gain for spot blotch was negligible in both GS and PS. Although PS is likely to be equally effective in selection for monogenic or oligogenic traits such as PHS, GS has an advantage if the trait phenotyping is high-cost or time sensitive and other traits are under selection. Our GS strategy with high weights for quality and disease traits and low weights for agronomic traits might be useful to enrich early generations for high-cost traits before advanced yield trialing.

References

Anderson, J.A., M.E. Sorrells, and S.D. Tanksley. 1993. RFLP analysis of genomic regions associated with resistance to preharvest sprouting in wheat. Crop Sci. 33(3): 453. doi: 10.2135/cropsci1993.0011183X003300030008x.

Bates, D., A.I. Vazquez, M. Ana, and I. Vazquez. 2015. Package "pedigreemm" Title Pedigree-based mixed-effects models.

Bayer, M.M., P. Rapazote-Flores, M. Ganal, P.E. Hedley, M. Macaulay, et al. 2017. Development and evaluation of a barley 50k iSelect SNP array. Front. Plant Sci. 8: 1792. doi: 10.3389/fpls.2017.01792.

Bernardo, R. 1994. Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci. 34(1): 20–25. doi: 10.2135/cropsci1994.0011183X003400010003x.

Bovill, J., A. Lehmensiek, M.W. Sutherland, G.J. Platz, T. Usher, et al. 2010. Mapping spot blotch resistance genes in four barley populations. Mol. Breed. 26(4): 653–666. doi: 10.1007/s11032-010-9401-9.

Castro, A.J., A. Benitez, P.M. Hayes, L. Viega, and L. Wright. 2010. Coincident quantitative trait loci effects for dormancy, water sensitivity and malting quality traits in the BCD47×Baronesse barley mapping population. Crop Pasture Sci. 61(9): 691. doi: 10.1071/CP10085.

Céron-Rojas, J.J., and J. Crossa. 2018. Linear selection indices in modern plant breeding. SpringerOpen, Switzerland. https://doi.org/10.1007/978-3-319-91223-3

Combs, E., and R. Bernardo. 2013. Genomewide selection to introgress semidwarf maize germplasm into U.S. corn belt inbreds. Crop Sci. 53(4): 1427–1436. doi: 10.2135/cropsci2012.11.0666.

Daetwyler, H.D., B. Villanueva, P. Bijma, and J.A. Woolliams. 2007. Inbreeding in genome-wide selection. J. Anim. Breed. Genet. 124(6): 369–376. doi: 10.1111/j.1439-0388.2007.00693.x.

Dekkers, J.C.M. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124(6): 331–341. doi: 10.1111/j.1439-0388.2007.00701.x.

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome J. 4(3): 250. doi: 10.3835/plantgenome2011.08.0024.

Fetch, T.G., and B.J. Steffenson. 1999. Rating scales for assessing infection responses of barley infected with *Cochliobolus sativus*. Plant Dis. 83(3): 213–217. doi: 10.1094/PDIS.1999.83.3.213.

Gao, W., J.A. Clancy, F. Han, D. Prada, A. Kleinhofs, et al. 2003. Molecular dissection of a dormancy QTL region near the chromosome 7 (5H) L telomere in barley. Theor. Appl. Genet. 107(3): 552–559. doi: 10.1007/s00122-003-1281-5.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2009. ASReml User Guide. Hemel Hempstead, HP1 1ES, UK.

Harlan, H.., M.L. Martini, and M.N. Pope. 1925. Tests of barley varieties in America. United States Department of Agriculture. Department bulletin No. 1334.

Harrell, F. 2021. Package “Hmisc”. Harrell miscellaneous.

Hazel, L.N., and J.L. Lush. 1942. The efficiency of three methods of selection. J. Hered. 33(11): 393–399. doi: 10.1093/oxfordjournals.jhered.a105102.

Heffner, E.L., J.-L. Jannink, H. Iwata, E. Souza, and M.E. Sorrells. 2011. Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci. 51(6): 2597. doi: 10.2135/cropsci2011.05.0253.

Heffner, E.L., A.J. Lorenz, J.-L. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: Gain per unit time and cost. Crop Sci. 50(5): 1681. doi: 10.2135/cropsci2009.11.0662.

Hernandez, C.O., L.E. Wyatt, and M.R. Mazourek. 2020. Genomic prediction and selection for fruit traits in winter squash. G3; Genes|Genomes|Genetics 10(10): 3601–3610. doi: 10.1534/g3.120.401215.

Juskiw, P.E., J.H. Helm, M. Oro, J.M. Nyachiro, and D.F. Salmon. 2009. Registration of 'Bentley' barley. J. Plant Regist. 3(2): 119. doi: 10.3198/jpr2008.10.0631crc.

Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124(3).

Legge, W.G., S. Haber, D.E. Harder, J.G. Menzies, J.S. Noll, et al. 2008. Newdale barley. Can. J. Plant Sci. 88(4): 717–723. doi: 10.4141/CJPS07194.

Legge, W.G., J.R. Tucker, T.G. Fetch, S. Haber, J.G. Menzies, et al. 2014. AAC Synergy barley. Can. J. Plant Sci. 94(4): 797–803. doi: 10.4141/cjps2013-307.

Li, C., P. Ni, M. Francki, A. Hunter, Y. Zhang, et al. 2004. Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. Funct. Integr. Genomics 4(2): 84–93. doi: 10.1007/s10142-004-0104-3.

Martin, J.M., T.K. Blake, and E.A. Hockett. 1991. Diversity among North American Spring barley cultivars based on coefficients of parentage. Crop Sci. 31(5): 1131. doi: 10.2135/cropsci1991.0011183X003100050009x.

Massman, J.M., H.G. Jung, and R. Bernardo. 2013. Genomewide Selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Sci. 53(1): 58–66. doi: 10.2135/cropsci2012.02.0112.

Meuwissen, T.H. 1997. Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75(4): 934. doi: 10.2527/1997.754934x.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4): 1819–1829. http://www.genetics.org/content/genetics/157/4/1819.full.pdf (accessed 25 February 2018).

Mohammadi, M., T.K. Blake, A.D. Budde, S. Chao, P.M. Hayes, et al. 2015. A genome-wide association study of malting quality across eight U.S. barley breeding programs. Theor. Appl. Genet. 128(4): 705–721. doi: 10.1007/s00122-015-2465-5.

Momen, M., A.A. Mehrgardi, A. Sheikhy, A. Esmailizadeh, M.A. Fozi, et al. 2017. A predictive assessment of genetic correlations between traits in chickens using markers. Genet. Sel. Evol. 49(1): 1–14. doi: 10.1186/s12711-017-0290-9.

Monat, C., S. Padmarasu, T. Lux, T. Wicker, H. Gundlach, et al. 2019. TRITEX: Chromosome-scale sequence assembly of Triticeae genomes with open-source tools. Genome Biol. 20(1): 284. doi: 10.1186/s13059-019-1899-5.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Rutkoski, J., R.P. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, et al. 2015. Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. Plant Genome 8(2): plantgenome2014.10.0074. doi: 10.3835/plantgenome2014.10.0074.

Sallam, A.H., and K.P. Smith. 2016. Genomic selection performs similarly to phenotypic selection in barley. Crop Sci. 56(6): 2871. doi: 10.2135/cropsci2015.09.0557.

Tiede, T., and K.P. Smith. 2018. Evaluation and retrospective optimization of genomic selection for yield and disease resistance in spring barley. Mol. Breed. 38(5): 1–16. doi: 10.1007/s11032-018-0820-3.

Veenstra, L.D., J. Poland, J.L. Jannink, and M.E. Sorrells. 2020. Recurrent genomic selection for wheat grain fructans. Crop Sci. 60(3): 1499–1512. doi: 10.1002/csc2.20130.

Visioni, A., S. Rehman, S.S. Viash, S.P. Singh, R. Vishwakarma, et al. 2020. Genome wide association mapping of spot blotch resistance at seedling and adult plant stages in barley. Front. Plant Sci. 11: 642. doi: 10.3389/fpls.2020.00642.

Williams, J.S. 1962. The evaluation of a selection index. Biometrics 18(3): 375. doi: 10.2307/2527479.

Woonton, B.W., J. V Jacobsen, F. Sherkat, and I.M. Stuart. 2005. Changes in germination and malting quality during storage of barley. J. Inst. Brew. 111(1): 33-41.

Yu, H., M.L. Spangler, R.M. Lewis, and G. Morota. 2017. Genomic relatedness strengthens genetic connectedness across management units. G3 Genes, Genomes, Genet. 7(10): 3543–3556. doi: 10.1534/g3.117.300151.

CHAPTER 6

VARIETY DESCRIPTIONS OF 'EXCELSIOR GOLD' AND CU198 TWO-ROW SPRING MALTING BARLEY

Abstract

Excelsior Gold and CU198 are two-row spring malting barleys (*Hordeum vulgare* L.) developed and released by the Cornell Agricultural Experiment Station in 2020 with spot blotch (*Bipolaris sorokiniana* Sacc) and preharvest sprouting (PHS) resistance. The New York Farm Brewery bill was passed in 2012, creating a special designation for "farm" breweries and distilleries that used a quota of New York grown malt and grain in their products. This legislation spurred rapid growth in the number of breweries in the state, increasing from 92 in 2012 to 460 in 2020 (Brewers Association, 2020). After several years of commercial variety trials did not identify well-adapted germplasm, Cornell University initiated a two-row spring malting barley breeding program in 2016. The humid continental climate of upstate New York makes PHS damage an annual risk for malting barley production. Currently available varieties with PHS resistance do not have acceptable malting quality in New York state for craft brewing and high malting quality varieties are prone to severe PHS damage. Excelsior Gold and CU198 combine good foliar disease resistance, increased PHS resistance, and acceptable malting quality to help support the unique requirements of malting barley production in New York state.

Methods

Pedigree

Excelsior Gold was derived from the cross Craft/AAC Synergy and was initially tested as SR564R-4 and CU31. Craft (PI 646158) was developed by Montana State University and is derived from the cross Klages/Baronesse. AAC Synergy was developed by Agriculture and Agri-Food Canada and is derived from the cross TR02267/Newdale (Legge et al., 2014). AAC Synergy was selected as a high malting quality and foliar disease resistance donor and Craft was selected for malting quality and PHS resistance. A public naming contest was held to solicit a name for CU31, and attendees of the 2020 New York Barley and Malt Summit selected 'Excelsior Gold'. CU198 is derived from the cross KWS Tinka/AAC Synergy and was initially tested as ST563R-3. KWS Tinka (PI 681721) is a semi-dwarf European spring two-row malting variety with good malting quality and moderate PHS resistance.

Selection and evaluation

Initial crosses were made in greenhouse facilities in Ithaca, NY in April 2016 as part of a connected half-sib population with AAC Synergy as the common female parent (Chapter 2). F_1, F_2, and F_3 generations were advanced by single seed descent in the greenhouse and 1341 $F_{3:4}$ lines were planted in single 1M rows at two locations in Ithaca, NY in 2017. Phenotypic data were collected for plant height, heading date, leaf rust (*Puccinia hordei* G. H. Otth), PHS, grain protein, and spot blotch and 250 lines were selected. One spike per headrow was sampled and the F_5 generation of the 250 selected lines was grown in the greenhouse in the winter of 2017 for pure line

advancement. At the same time, F_3:F_5 bulk seed from the field was sent to New Zealand for winter seed increase. Leaf tissue was sampled from a single F_4 plant in each headrow for genotyping-by-sequencing and genomic prediction model training as described in Chapter 2. Genomic prediction GBLUP models for disease resistance, PHS, and agronomic performance were used to select 100 of the 250 lines sent to winter nursery and these 100 lines were planted in preliminary yield trials with two replications in a randomized complete block design in five locations in spring 2018. Genomic prediction GBLUP models were again used to advance 60 of the 100 lines to preliminary yield trials with two replicates in four locations in 2019. These 60 lines were grown in strip plots in a New Zealand winter nursery to increase seed for strip plot planting in New York in 2019. CU198 was also entered into advanced yield trial testing in 2019 and was observed in the same four locations with three replicates. Excelsior Gold and CU198 were evaluated in advanced yield trials with three replications in three locations in 2020. Excelsior Gold and CU198 were entered into the Eastern Spring Barley Nursery (ESBN) and the American Malting Barley Midwest pilot quality trial in 2021.

Malting quality phenotyping was conducted at the USDA Cereal Crops Research Unit (CCRU) in Madison, WI using a small-scale “teaball” method (Schmitt and Budde, 2011) and at the Hartwick College Craft Food and Beverage Center in Oneonta, NY. Preharvest sprouting was measured on a 0 (resistant) to 9 (susceptible) scale according to Anderson et al. (1993) on five spikes sampled at physiological maturity (loss of green color from the peduncle), after-ripened for three days, and misted in a greenhouse chamber for three days. Germination energy and germination

intensity were measured in 2019 and 2020 on samples collected two days after physiological maturity, dried for one day, and hand threshed. For germination tests, 4 mL water and 30 kernels were added to petri plates with two Whatman filter papers at 45 days after physiological maturity. Germination, defined as radicle emergence of > 2 mm, was counted each day for three days. Germination energy (GE) was calculated as the percentage of kernels that germinated after three days and was used as a measure of seed dormancy. Germination index (GI) was calculated as $10*(\Sigma(n_{24} + n_{48} + n_{72})/(n_{24} + 2n_{48} + 3n_{72}))$ where n_{24} is the number of kernels germinated at 24 hours, n_{48} is the number of kernels germinated at 48 hours, and n_{72} is the number of kernels germinated at 72 hours.

Excelsior Gold and CU198 were evaluated for Fusarium head blight (FHB, *Fusarium graminearum* Schwabe) and spot blotch resistance. FHB was evaluated in misted, inoculated nurseries in 2018 and 2020. The FHB headrow nurseries were spray inoculated with a 1×10^5 conidia concentration according to Dill-Macky (2003). Twenty spikes were evaluated in each headrow and severity was assessed on a 0 (no infection) to 5 (more than 75% of kernels infected) scale. The headrows were hand harvested and threshed grain was sent for deoxynivalenol (DON) analysis. Spot blotch was evaluated in yield trials and in inoculated seedling greenhouse assays replicated three times (Chapter 5). Spot blotch resistance was measured according to Fetch and Steffenson (1999) on a 1 (resistant) to 9 (susceptible) scale.

2020 foundation seed increase fields of Excelsior Gold were grown by Francis Domoy in Orleans County, NY and CU198 was grown on the Cornell University Helfer farm in Ithaca, NY by the Cornell Small Grains breeding project.

Statistical analysis

All statistical analyses were conducted with R version 3.5.1 (R Core Team, 2018). Excelsior Gold and CU198 were compared to the two-row malting varieties AAC Synergy, Newdale (Legge et al., 2008), KWS Tinka, and ND Genesis (PI 677345) for agronomic traits. AAC Synergy, Newdale, and ND Genesis are American Malting Barley Association approved varieties and KWS Tinka was included for yield and disease resistance comparison. Honestly significant differences (HSD) between means ($p<0.05$) were calculated using the *agricolae* R package (de Mendiburu and Yaseen, 2020). AAC Synergy was used for malting quality comparisons. Preliminary and advanced yield trials were planted with 68 g of seed per plot in 4 by 1.26 m plots and trimmed to 3 by 1.26 m plots for harvest. Irrigation was not used for any trials. Environments were defined as year-location combinations. Entry means were used for analysis.

Characteristics

General description

Excelsior Gold spikes are smooth-awned with an erect spike, dense head, and awn length equal to or longer than the spike. Spikes fully emerge from the flag leaf. Kernels have an adhering hull, short rachilla hairs, white aleurone, medium glume awns (same length as kernel), slightly wrinkled hull, and horseshoe depression at the attachment point. Leaf sheathes are green.

CU198 spikes are rough-awned with a lax head, semi-nodding spike, and *deficiens* type lateral florets. Some spikes with full lateral florets can be observed, but

other spike characteristics are the same for this type. Spikes sometimes exhibit incomplete emergence from the flag leaf. Kernels have an adhering hull, long rachilla hairs, horseshoe depression at attachment point, white aleurone, long glume awns, and a slightly wrinkled hull. Leaf sheathes have some purple color.

Agronomic performance

In statewide yield trials conducted by Cornell University, CU198 yielded better than several check varieties and Excelsior Gold performed similarly to AAC Synergy and Newdale (Table 6.1). Both Excelsior Gold and CU198 had good test weights, earlier heading dates, higher PHS resistance, and were taller than AAC Synergy and Newdale. After 45 days of after-ripening, Excelsior Gold and CU198 were fully non-dormant (GE > 0.95) but germination rate (GI) was slightly lower than check varieties.

Table 6.1: Comparison of agronomic performance of Excelsior Gold and CU198 with check varieties. n specifies the number of year-location combinations observed for each trait.

Entry	Yield	Test weight	Height	Heading date	Boot	PHS	GE	GI
	kg/ha	kg/hl	cm	Julian days	1-5	0-9	%	0-10
AAC Synergy	3520	59.8	73.4	176	2.2	4.7	1.0	6.2
KWS Tinka	3425	59.4	71.1	175	2.4	2.4	0.97	6.1
ND Genesis	3226	61.3	78.7	170	1.6	3.2	0.98	6.1
Newdale	3676	62.5	69.0	174	2.9	3.5	0.99	6.2
Excelsior Gold	3688	61.9	80.8	172	1.5	2.1	1.0	5.4
CU198	3880	63.4	78.5	171	1.8	2.0	1.0	5.5
HSD (a=0.05)	450	1.1	5.2	2	0.5	1.3	ns	0.6
n	14	14	12	7	11	7	4	4

Disease resistance

Excelsior Gold and CU198 have good seedling and adult plant spot blotch resistance, improved FHB index compared to AAC Synergy and Newdale, and lower DON than AAC Synergy (Table 6.2). Excelsior Gold and CU198 carry the resistant allele for spot blotch at the *Rcs5* locus on chromosome 7H (Steffenson et al., 1996; Ameen et al., 2020).

Table 6.2: Comparison of Excelsior Gold and CU198 disease resistance to check varieties from 2018 to 2020. n specifies the number of year-location combinations observed for each trait.

Entry	Spot blotch greenhouse	Spot blotch field	FHB incidence	FHB severity	FHB index	DON
	1-9	1-9	%	1-5		ppm
AAC Synergy	2.5	1.7	50.2	1.75	8.3	8.4
KWS Tinka	4.5	4.1	51.5	1.59	6.6	9.1
ND Genesis	3.2	2.9	57.9	1.73	8.3	4.4
Newdale	3.0	3.0	49.0	1.89	9.9	4.5
Excelsior Gold	2.8	1.0	34.2	1.32	3.5	5.2
CU198	3.5	2.7	46.8	1.55	5.7	4.6
HSD (a=0.05)	1.44	1.1	ns	0.45	ns	4.5
n	1	9	4	4	4	2

Quality screening

Excelsior Gold and CU198 have good malting quality profiles compared to AAC Synergy based on data from two malt quality labs using slightly different protocols (Tables 6.3 and 6.4). Excelsior Gold is similar to AAC Synergy for most traits but has elevated beta-glucan levels and lower free amino nitrogen (FAN). CU198 has beta-

glucan levels more comparable to AAC Synergy but lower FAN and soluble protein, giving it a favorable craft malt profile.

Table 6.3: Comparison of malting quality of Excelsior Gold and CU198 with AAC Synergy. Malting quality data is from five locations across two years. Malting was completed at Hartwick College.

Entry	Protein	Malt extract	Beta-glucan	Soluble protein	Soluble/total protein	FAN	Diastatic power	Alpha-amylase
	%	%	ppm	%	%	mg/L	°L.	D.U.
AAC Synergy	10.0	82.4	74.8	5.31	53.8	241	114	67
Excelsior Gold	11.0	81.8	247.0	5.19	47.6	218	107	58
CU198	10.5	82.4	147.4	4.60	44.6	193	116	60
HSD (a=0.05)	ns	ns	97.4	0.67	ns	32	ns	ns
n	5	5	5	5	5	5	5	5

Table 6.4: Comparison of malting quality of Excelsior Gold and CU198 with AAC Synergy. Malting quality data is from four locations in 2019. Malting was completed at the USDA Cereal Crops Research Unit using small-scale "teaball" malting.

Entry	Malt extract	Beta-glucan	Soluble/total protein	FAN	Diastatic power	Alpha-amylase
	%	ppm	%	mg/L	°L.	D.U.
AAC Synergy	80.3	93.5	4.84	235	87	81
Excelsior Gold	80.2	220.0	4.66	198	103	66
CU198	81.7	67.1	4.12	181	100	65
HSD (a=0.05)	1.2	ns	0.44	36	ns	16
n	4	4	4	4	4	4

Genotypic information

Excelsior Gold and CU198 were genotyped using KASP markers for genes involved in PHS in barley (Chapter 3). Both Excelsior Gold and CU198 have the non-dormant allele at *HvAlaAT1* and a dormant allele at *HvMKK3.* This haplotype confers PHS resistance but does not induce long-term seed dormancy. CU198 was not explicitly

screened for the *sdw* semi-dwarf gene on chromosome 3H but based on genome wide association for plant height (Chapter 1) CU198 does not carry the semi-dwarf allele at *sdw*. This genetic evidence is supported by the height comparison between CU198 and KWS Tinka (Table 6.1).

Discussion

Plant breeders are charged with releasing improved varieties with increased yield, resistance to insects and diseases, increased resiliency to abiotic stressors, and acceptable industrial and consumer quality. The "breeder's equation" (Lush, 1943) is a useful model for breeding programs that defines genetic gain (ΔG) as the product of additive genetic variation within a breeding population (σ_a), selection intensity (i), and selection accuracy (r). Genetic gain per unit time can be expressed by dividing ΔG by years per cycle (L). In many cases, new breeding goals and targets appear more quickly than the breeding process can introduce an acceptable alternative and genetic variation for the new trait may be minimal or completely lacking. Achieving high selection accuracy in plant breeding is a resource intensive process that requires extensive testing of experimental germplasm to accurately separate genetic effects from environmental effects. Cycle time reductions often have definite ceilings based on crop biology.

To meet the needs of the New York craft malting and brewing industry, we addressed all four aspects of the breeder's equation to accelerate genetic gain using traditional and modern plant breeding technologies. We initiated our spring barley breeding population using a connected half-sib design. Malting barley breeding

typically involves elite by elite crosses due to stringent industry malting quality standards. Our strategy used elite malting parents from different breeding programs to generate good genetic variation for non-malting quality traits without jeopardizing average malting quality. The population structure imposed with such a design also enabled us to improve prediction accuracy and understand the genetic architecture of key traits (Chapter 2). Smaller family sizes (50-100) with more elite malting founder lines (>10) would have increased genetic diversity and potentially prediction accuracy while maintaining a similar population size. We imposed a high selection intensity using a combination of phenotypic and genomic selection, selecting 100 lines for yield trials out of almost 1400 lines in the initial population. Genomic selection was then used to further advance lines into multiple years of yield evaluation. High genetic variation and selection intensity will translate to genetic gain if selection accuracy is good. The initial ~1400 line population was fully phenotyped for basic agronomic and disease traits, but important quality traits like PHS had to be subsampled due to the high labor requirement. We decided to select from one year of training population phenotypes in two locations and could not measure yield or malting quality on small quantities of seed in a large population. Marker data was not available in time to use genomic prediction to select lines for the first winter nursery so phenotypic selection was used to first select 250 lines to send to New Zealand and genomic prediction was used to select 100 of those 250 lines to enter into yield trials the following year. Lines without PHS phenotypes were therefore omitted in the initial selection, potentially limiting selection accuracy and genetic gain. Collecting malting quality data continued to be a challenge in the yield trials due to weather and cost constraints. Our

selection accuracy for agronomic performance was high because we phenotyped those traits in a number of replicated locations over several years but malting quality accuracy may have been lower. Traditional plant breeding strategies like shuttle breeding with a winter nursery can accelerate the breeding process by enabling multiple generations per year. Although we did not use winter nursery to select multiple times in one year, we used it to increase seed of selected material to quickly move from headrow plots to multiple locations of yield trials and from yield trials to large-scale seed increase.

Two additional cycles of breeding material have been generated using a selection index in a genomic selection framework (Chapter 5). These lines were evaluated for two years in meter-long single or double rows to estimate genetic gain and are not formally entering the breeding program pipeline until 2021. In a purely applied breeding framework, cycle time could have been further reduced by selection in the F2 or F3 stage, one season of headrows for seed increase, and preliminary yield trials the following year. In this framework, Cycle 1 material could have entered preliminary yield trials in 2020 and Cycle 2 material in 2021. The correlated response of yield and malting quality from strong selection on PHS is yet to be determined for this material.

Plant breeders must accept tradeoffs in the breeder's equation imposed by resource limitations, crop biology, and available technology. Our spring malting barley strategy prioritized genetic variation, selection intensity, and cycle time to rapidly develop New York adapted germplasm. Both Excelsior Gold and CU198 provide improved agronomics and competitive malting quality compared to the best

currently available two-row spring malting quality varieties in New York. Enrichment for disease resistance QTL, definition of ideal ranges for basic agronomic characteristics such as heading date and plant height, and improved understanding of the relationship between PHS, seed dormancy, and malting quality have already begun to shape the Cornell two-row spring malting barley breeding program and will help drive genetic gain for New York craft malting and brewing in the future.

References

Ameen, G., S. Solanki, T. Drader, L. Sager-Bittara, B. Steffenson, et al. 2020. rcs5-mediated spot blotch resistance in barley is conferred by wall-associated kinases that resist pathogen manipulation. bioRxiv: 2020.04.13.040238. doi: 10.1101/2020.04.13.040238.

de Mendiburu, F. and Yaseen, M. 2020. agricolae: Statistical Procedures for Agricultural Research.R package version 1.4.0 , https://myaseen208.github.io/agricolae/https://cran.r-project.org/package=agricolae.

Dill-Macky, R. 2003. Inoculation methods and evaluation of Fusarium head blight resistance in wheat. In: Leonard, K.J., Bushnell, W.R., editors, Fusarium head blight wheat barley. APS Press, St. Paul. p.184–210

Fetch, T.G., and B.J. Steffenson. 1999. Rating scales for assessing infection responses of barley infected with *Cochliobolus sativus*. Plant Dis. 83(3): 213–217. doi: 10.1094/PDIS.1999.83.3.213.

Legge, W.G., S. Haber, D.E. Harder, J.G. Menzies, J.S. Noll, et al. 2008. Newdale barley. Can. J. Plant Sci. 88(4): 717–723. doi: 10.4141/CJPS07194.

Lush, J.L. 1943. Animal breeding plans. (Edn 2). The Iowa State College Press, Ames.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Schmitt, M.R., and A.D. Budde. 2011. Malting extremely small quantities of barley 1. J. Am. Soc. Brew. Chem 69(4): 191–199. doi: 10.1094/ASBCJ-2011-0728-01.

Wenzel, G., B.J. Steffenson, P.M. Hayes, and A. Kleinhofs. 1996. Genetics of seedling and adult plant resistance to net blotch and spot blotch in barley. Theor. Appl. Genet. 92: 552–558.

APPENDIX

GENOME-WIDE ASSOCIATION FOR BARLEY SEEDLING RESISTANCE TO SPOT BLOTCH

Introduction

Spot blotch (*Cochliobolus sativus*, anamorph *Bipolaris sorokiniana* Sacc.) is a major foliar disease of barley worldwide but can also infect barley grain. Severe spot blotch infections negatively impact yield and can reduce kernel plumpness, an important malting quality trait (Nutter et al., 1985). Diseased grain is also likely to be rejected for malting because steeping and germination conditions during malting encourage microbial growth. This disease is common in warm humid growing environments across the world and in the Northeast is primarily a spring barley disease. Spot blotch resistance has been an important breeding target for many years in North America but the majority of resistance screening have used isolates from the Upper Midwest (Valjavec-Gratian and Steffenson, 1997). A number of genome-wide association analyses for spot blotch have been published with North American (Zhou and Steffenson, 2013), global (Wang et al., 2017; Gyawali et al., 2018; Novakazi et al., 2019), and wild (Roy et al., 2010) germplasm and resistance loci have been discovered on all seven barley chromosomes. Several large effect QTL have been fine mapped for spot blotch resistance, including *Rcs5* (Steffenson et al., 1996; Ameen et al., 2020), *Rcs6* (Bilgic et al., 2006), and *Rbs7* (Wang et al., 2019). Zhou and Steffenson identified three QTL, the Midwest Six-rowed Durable Resistant Haplotype (MSDRH), that were consistently found in resistant North American six-row barleys coming from Upper Midwest breeding programs (2013). This haplotype was present in two-row

germplasm but did not confer the same level of resistance as in six-row germplasm. Genomic selection results for quantitative disease resistance in cereal grains are limited to Rutkoski et al. (2015) for stem rust (*Puccinia graminis*) in wheat and Tiede and Smith (2018) for *Fusarium graminearum* head blight related deoxynivalenol accumulation in spring six-row barley but both studies reported significant gain from genomic selection for disease resistance. Previous work using a New York *Bipolaris* isolate identified two QTL that may be part of the MSDRH (Blachez 2017). Our objectives were to assess genetic gain for spot blotch resistance across two cycles of genomic selection and map spot blotch resistance QTL using association mapping to enhance future breeding efforts.

Materials and methods

Phenotyping and genotyping

The C_{all} population described in Chapter 5 was used to evaluate seedling spot blotch resistance. Inoculum preparation and application methods, phenotyping scale, and genotyping are described in detail in Chapter 5. Marker set G_3, containing 20,929 Illumina 50k markers, was used for association analysis.

Statistical analysis

All statistical analyses were performed with R version 3.5.1 (R Core Team, 2018). Spot blotch best linear unbiased predictors (BLUP) were obtained with the *lme4* package (Bates et al., 2015) for scores at seven and ten days after inoculation using the model:

$$y_{ijk} \sim \mu + r_i + g_j + t_{k(i)} + \epsilon_{ijk}$$

where μ is the overall mean, r_i is the fixed effect of replicate, g_j is the random effect of genotype where $g_j \sim N(0, \sigma_g^2)$ with genetic variance σ_g^2, and $t_{k(i)}$ is the random effect of tray nested within replicate where $t_{k(i)} \sim N(0, \sigma_t^2)$ where σ_t^2 is the variance due to tray, and ε_{ijk} is the random error with $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ where σ_ε^2 is the residual error variance. Broad-sense heritability (H^2) was estimated by $\sigma_g^2/(\sigma_g^2 + \sigma_\varepsilon^2/r)$ where r is the number of replicates.

Genome wide association (GWA) models were fit using a multi-locus mixed model (MLMM) (Segura et al., 2012) in the R package *GAPIT* (Wang and Zhang, 2020). Population structure was accounted for by adding the first two principal components of the eigendecomposition of the marker matrix. A Bonferroni correction ($p<0.05$) was used to control false positives marker trait associations (MTA).

Results

Infection response at ten days had higher genetic variance and lower tray variance than seven-day scores and ten-day scores were used for further analysis. Broad-sense heritability for ten-day spot blotch infection response was 0.87. The susceptible C_{all} founder line Craft had a mean ten-day infection response of 7.6 and the resistant founder ACC Synergy had an infection response of 2.5 (Figure A.1). 207 lines showed a very low infection response with a mean of 3.0 or lower, 171 lines showed a low infection response with a mean between 5.0 and 3.0, and 31 lines showed a high infection response with a mean greater than 5.0. Phenotypic distributions were similar for each selection cycle.

LD decayed below 0.2 r^2 at around 8 Mb. Two spot blotch seedling resistance QTL with a single MTA each were identified using MLMM for association analysis (Table A.1). The most significant MTA was marker JHI-Hv50k-2016-457496 on the short arm of chromosome 7H and explained 90.2% of the phenotypic variation. A MTA on chromosome 3H for marker JHI-Hv50k-2016-156842 was also detected that explained 8.1% of the phenotypic variation.

Table A.1: Significant genome-wide association marker trait associations for spot blotch resistance.

SNP	Chr	Position	MAF	n	p-value	Alleles*	Susceptible parent
JHI-Hv50k-2016-457496	7	30096226	0.067	407	9.61E-84	A/**G**	Craft
JHI-Hv50k-2016-156842	3	14114956	0.185	407	5.53E-12	**A**/G	Conlon, KWS Tinka

*Bold allele is resistant

Discussion

Greenhouse inoculation of spot blotch conidia in flats at 1 x 10^4 conidia per ml successfully induced phenotypic variation for seedling response in a two-row spring barley population. Phenotypic variation and ease of phenotyping were greater at ten days after inoculation than seven days. Plastic bags were effective at maintaining high humidity around inoculated plants. A significant replication effect was not observed but at least two replications should still be used for future studies.

The large effect MTA on 7H is most likely *rcs5*. This locus has been associated with spot blotch in a number of mapping experiments in diverse germplasm (Steffenson et al., 1996; Zhou and Steffenson, 2013; Wang et al., 2017; Gyawali et al., 2018; Novakazi et al., 2019). Two wall-associated kinase genes, *Sbs1* and *Sbs2*, have been

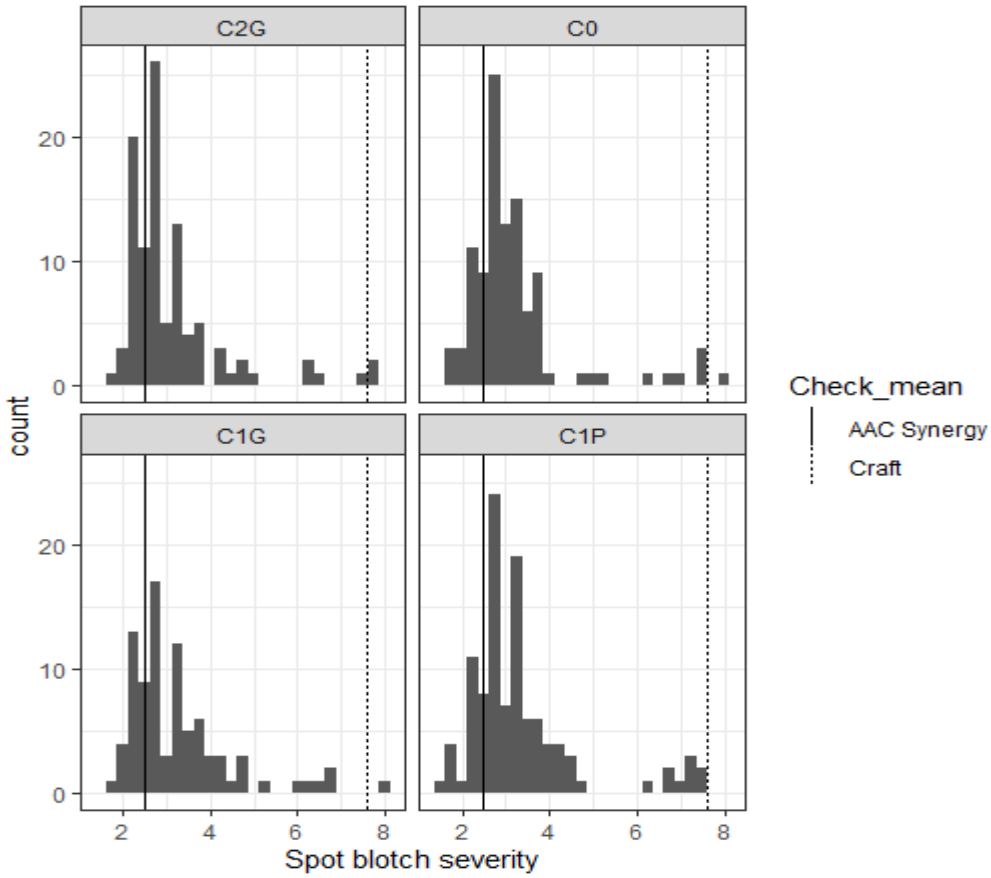


Figure A.1: Histogram of average spot blotch scores taken ten days after inoculation in C_{all}. C0 indicates the base population, C1P Cycle 1 phenotypic selection, C1G Cycle 1 genomic selection, and C2G Cycle 2 genomic selection. The solid vertical line indicates the mean of the resistant founder AAC Synergy and the dashed vertical line indicates the mean of the susceptible founder Craft.

cloned at *rcs5* and are pathogen susceptibility targets that are hijacked into inducing programmed cell death after pathogen infection (Ameen et al., 2020).The physical location of *Sbs1* (HORVU.MOREX.r2.7HG0539110) on the Morex v2 genome is 7H:26054957-26055559. A more diverse germplasm set may be able to identify a marker closer to the physical position of *Sbs1* (see Novokazi et al., 2019). JHI-Hv50k-2016-156842 is associated with Rcs-qtl-3H-11_10565 identified by Zhou and Steffenson (2013) and Rcs-qtl-3H-25.27 identified by Gyawali et al. (2018). Both

studies identified an association with the marker 11_10565/ BOPA1_3906-558, which is 44,396 bp from JHI-Hv50k-2016-156842. In this region, there are several gene families that have been associated with plant disease resistance in previous studies including an ABC transporter, glutathione-S-transferase, and an ankyrin repeat protein (Table A.2).

The 1H MSDRH locus was not identified in this population. Blachez (2017) and mapping in Chapter 2 did not identify the 1H locus in GWA in similar germplasm using adult plant phenotypes induced with the same NY isolate, indicating this locus may be fixed in the C_{all} panel or the isolate may not induce a susceptible response at the 1H locus.

Table A.2: Genes within 250 kb of JHI-Hv50k-2016-156842. Gene annotations collected from GrainGenes JBrowse for Morex v2 genome assembly.

Gene name	Description
HORVU.MOREX.r2.3HG0187360	ABC transporter G family member
HORVU.MOREX.r2.3HG0187370	Hexose transporter
HORVU.MOREX.r2.3HG0187380	DNA-binding storekeeper protein-related transcriptional regulator
HORVU.MOREX.r2.3HG0187390	Hydroxyacylglutathione hydrolase
HORVU.MOREX.r2.3HG0187400	Core-2/I-branching beta-1-6-N-acetylglucosaminyltransferase family protein
HORVU.MOREX.r2.3HG0187420	Glutathione S-transferase T3
HORVU.MOREX.r2.3HG0187430	Core-2/I-branching beta-1-6-N-acetylglucosaminyltransferase family protein
HORVU.MOREX.r2.3HG0187440	Core-2/I-branching beta-1-6-N-acetylglucosaminyltransferase family protein
HORVU.MOREX.r2.3HG0187460	Eukaryotic translation initiation factor 4E
HORVU.MOREX.r2.3HG0187470	Receptor-like kinase
HORVU.MOREX.r2.3HG0187510	Ankyrin repeat family protein-like
HORVU.MOREX.r2.3HG0187520	Forkhead-associated (FHA) domain-containing protein
HORVU.MOREX.r2.3HG0187530	Delta(7)-sterol-C5(6)-desaturase
HORVU.MOREX.r2.3HG0187560	Calmodulin-binding protein-like
HORVU.MOREX.r2.3HG0187620	Endo-1-4-beta-xylanase
HORVU.MOREX.r2.3HG0187630	Endo-1-4-beta-xylanase

302 lines carried the resistant marker at *rcs5* and Rcs-qtl-3H, 61 lines were resistant at *rcs5* and susceptible at Rcs-qtl-3H, and 20 lines were susceptible at both loci. No susceptible *rcs5*/resistant Rcs-qtl-3H lines were identified, likely because susceptible alleles from the parental donors Craft (*rcs5*) and Conlon and KWS Tinka (Rcs-qtl-3H Rcs-qtl-3H) would have only been combined in Cycles 1 and 2, at which point selection for spot blotch resistance had been imposed.

The Cornell spring malting barley breeding program currently has limited genetic diversity. Although spot blotch resistance loci in the breeding program currently confer good resistance against a prevalent New York state isolate, new sources of spot blotch resistance are needed due to the importance of spot blotch as a foliar and seed pathogen in malting barley in New York. Detection of the 3H QTL supports previous mapping efforts in more diverse germplasm. Neither MTA detected in this study is likely a good candidate for marker assisted selection outside of this narrow population. Greenhouse screening was labor intensive and breeding efforts may benefit more from seedling evaluation in inoculated field nurseries. Field based evaluation of seedling spot blotch resistance has not been evaluated at Cornell.

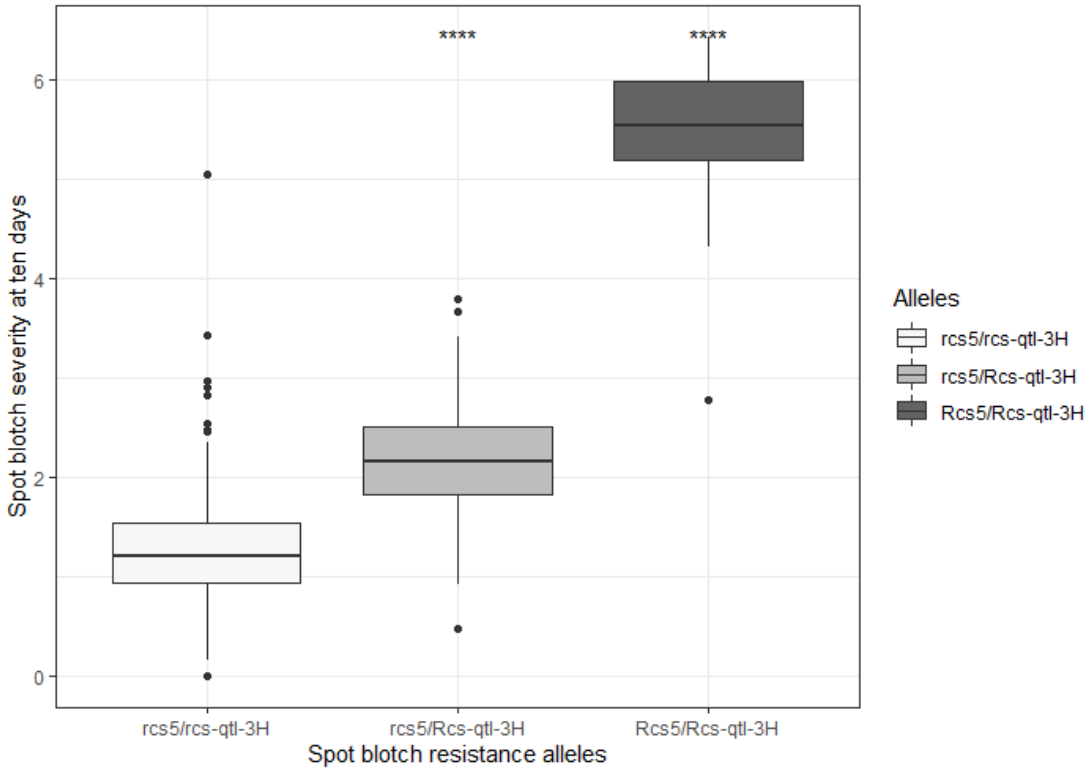


Figure A.2: Distribution of spot blotch severity BLUPs measured at ten days post-inoculation. Resistant/susceptible alleles for marker JHI-Hv50k-2016-457496 are indicated by rcs5/Rcs5, respectively, and from JHI-Hv50k-2016-156842 by rcs-qtl-3H/ Rcs-qtl-3H, respectively. Asterisks indicate haplotype means are significantly different ($p < 0.05$).

References

Ameen, G., S. Solanki, T. Drader, L. Sager-Bittara, B. Steffenson, et al. 2020. rcs5-mediated spot blotch resistance in barley is conferred by wall-associated kinases that resist pathogen manipulation. bioRxiv: 2020.04.13.040238. doi: 10.1101/2020.04.13.040238.

Bates D., Mächler M., Bolker B., Walker S. (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, **67**(1), 1–48. doi: 10.18637/jss.v067.i01.

Bilgic, H., B.J. Steffenson, and P.M. Hayes. 2006. Genetics and resistance molecular mapping of loci conferring resistance to different pathotypes of the spot blotch pathogen in barley. Phytopathology. 96(7): 699. doi: 10.1094/PHYTO-96-0699.

Blachez, A.F. 2017. Assessment of malting barley varieties for resistance to important barley diseases in New York. Cornell University Library. https://cornell.worldcat.org/title/assessment-of-malting-barley-varieties-for-resistance-to-important-barley-diseases-in-new-york/oclc/1066118130&referer=brief_results (accessed 25 February 2019).

Gyawali, S., S. Chao, S.S. Vaish, S.P. Singh, S. Rehman, et al. 2018. Genome wide association studies (GWAS) of spot blotch resistance at the seedling and the adult plant stages in a collection of spring barley. Mol. Breed. 38(5): 1–14. doi: 10.1007/s11032-018-0815-0.

Novakazi, F., O. Afanasenko, N. Lashina, G.J. Platz, R. Snowdon, et al. 2019. Genome-wide association studies in a barley (Hordeum vulgare) diversity set reveal a limited number of loci for resistance to spot blotch (Bipolaris sorokiniana) (T. Miedaner, editor). Plant Breed.: pbr.12792. doi: 10.1111/pbr.12792.

Nutter, F.W., V.D. Pederson, and A.E. Foster. 1985. Effect of inoculations with Cochliobolus sativus at specific growth stages on grain yield and quality of malting barley. Crop Sci. 25(6): 933–938. doi: 10.2135/cropsci1985.0011183x002500060008x.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Roy, J.K., K.P. Smith, G.J. Muehlbauer, S. Chao, T.J. Close, et al. 2010. Association mapping of spot blotch resistance in wild barley. Mol. Breed. 26(2): 243–256. doi: 10.1007/s11032-010-9402-8.

Rutkoski, J., R.P. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, et al. 2015. Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. Plant Genome 8(2): plantgenome2014.10.0074. doi: 10.3835/plantgenome2014.10.0074.

Steffenson, B.J., P.M. Hayes, and A. Kleinhofs. 1996. Genetics of seedling and adult plant resistance to net blotch (Pyrenophora teres f. teres) and spot blotch (Cochliobolus sativus) in barley. Theor. Appl. Genet. 92(5): 552–558. doi: 10.1007/BF00224557.

Tiede, T., and K.P. Smith. 2018. Evaluation and retrospective optimization of genomic selection for yield and disease resistance in spring barley. Mol. Breed. 38(5): 1–16. doi: 10.1007/s11032-018-0820-3.

Valjavec-Gratian, M., and B.J. Steffenson. 1997. Pathotypes of Cochliobolus sativus on barley in North Dakota. Plant Dis. 81:1275-1278.

Wang, J., and Z. Zhang. 2020. GAPIT version 3: Boosting power and accuracy for genomic association and prediction. bioRxiv: 2020.11.29.403170. doi: 10.1101/2020.11.29.403170.

Wang, R., Y. Leng, S. Ali, M. Wang, and S. Zhong. 2017. Genome-wide association mapping of spot blotch resistance to three different pathotypes of Cochliobolus sativus in the USDA barley core collection. Mol. Breed. 37(4): 1–14. doi: 10.1007/s11032-017-0626-8.

Wang, R., Y. Leng, M. Zhao, and S. Zhong. 2019. Fine mapping of a dominant gene conferring resistance to spot blotch caused by a new pathotype of Bipolaris sorokiniana in barley. Theor. Appl. Genet. 132(1): 41–51. doi: 10.1007/s00122-018-3192-5.

Zhou, H., and B. Steffenson. 2013. Genome-wide association mapping reveals genetic architecture of durable spot blotch resistance in US barley breeding germplasm. Mol Breed. 32: 139–54. https://link.springer.com/content/pdf/10.1007/s11032-013-9858-4.pdf