Farmer Driven Breeding Strategies

By

Pablo Andrés Sandro García

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

Plant Breeding and Plant Genetics

at the

UNIVERSITY OF WISCONSIN – MADISON

2024

Date of final oral examination: November 1st, 2024

The dissertation is approved by the following members of the Final Oral Committee:
Lucia Gutierrez, Associate Professor, Plant and Agroecosystems Sciences
Julie Dawson, Professor, Plant and Agroecosystems Sciences
Valentin, Picasso, Associate Professor, Plant and Agroecosystems Sciences
Jeffrey Endelman, Professor, Plant and Agroecosystems Sciences
Francisco Peñagaricano, Associate Professor, Animal & Dairy Sciences

DEDICATION

To my family and friends whom, with their support, made this dissertation possible.

Acknowledgement

I would like to express my heartfelt gratitude to Dr. Lucia Gutiérrez for the opportunity to join her lab and for her unwavering dedication to mentoring and guiding me throughout my academic journey. I also extend my thanks to both past and present members of the Cereals Breeding and Quantitative Genetics lab for their invaluable support. Special thanks to my friends Ines, German, Gina, and Venancio for always being there with support and valuable feedback. Lastly, I am deeply grateful to my family and friends for their encouragement and support.

TABLE OF CONTENT

List of Figures

List of Tables

List of Supplementary Figures

List of supplentatry Tables

Abstract

Cereals are the principal source of food in modern society. Adapted and stable cultivars are important to ensure food security, especially in the U.S. Midwest where incorporating oat, barley, and wheat in crop rotations is beneficial to sustainability. This dissertation uses plant breeding tools to deliver cultivars and information to help farmers incorporate small grains in crop rotations. Specific objectives include developing a breeding program for organic bread wheat, characterizing genotype by environment interaction in oats, and predicting genotypic responses across environments using multi-environment datasets. To achieve the first objective, in conjunction with farmers and bakers we selected a hard red winter wheat population for disease tolerance and protein content, testing its baking quality and stability in organic farms. For the second objective, we analyzed a 25-year, 67-location oats database to predict genotype rankings in mega-environments and we made this information accessible to farmers and researchers with a variety selection tool. For our final objective, we used Finlay-Wilkinson random regressions across oat, wheat, and barley datasets to assess grain yield stability, noting the impacts of the number of environment and variance on estimation. We are releasing a stable winter wheat cultivar that meets farmers' and bakers' needs. We successfully identified high-yielding oat cultivars and improved information availability. We predicted Finlay-Wilkinson regression for stability with high accuracy for genotypes lacking phenotypic data. In conclusion, diverse plant breeding methodologies tailored responses to production system limitations.

Chapter1. Introduction

Background

Humans have historically been connected to plants in many ways, especially as a source of food (Hardy et al. 2016), but around 10,000 years ago, we started to manage the surrounding landscape to favor the growth of species with nutritional value (Harlan 1986; Smith 2001; Diamond 2002; Purugganan and Fuller 2009). The shift from natural selection to human selection (Spencer and Cross 2007), triggered the process of domestication and marked the beginning of agriculture (Harlan 1986; Smith 2001; Diamond 2002). This change lead to the development of farming societies and today's modern societies (Diamond 2002). For cereals, the differences between cultivated species and their wild relatives are given by a group of characteristics that allow domesticated species to thrive in an anthropically modified environment and is known as the domestication syndrome (De Wet and Harlan 1975; Allaby 2014). Some of these characteristics are for example, low toxin content, photoperiod response/insensibility, increase in grain/fruit size, absence of shattering, synchronized maturity, and other physiological traits that help the propagation in a farming context (Allaby, 2014; De Wet and Harlan, 1975). The accumulation of domestication syndrome traits occurred during a prolonged period (Allaby, 2014). It is suggested that favorable alleles of genes responsible for domestication traits were combined and their frequency increased by selection. For example, the loss of shattering in wheat (*Triticum aestivum* L.; Fuller and Allaby, 2009; Tanno and Willcox, 2006), barley (*Hordeum vulgare* L.; Fuller and Allaby, 2009), and rice (*Oryza sativa* L.; Fuller and Allaby, 2009) have an accumulation and spreading period between 3000 and 4000 years (Fuller and Allaby, 2009). Another example can be found in soybean (*Glycine max* L.) where gene modifiers of the response to photoperiod were sequentially accumulated during an approximate period of 3000 years (Lu et al., 2020).

Farmers drove the selection process in their fields until the rise of scientific-driven plant breeding in the first part of the 20 century (Cox 2009). With the theoretical developments in genetics and statistics (Darwin 1859, 1868; Fisher 1922, 1925, 1938; Wright 1922, 1931, 1939, 1949; Lush 1943), scientific plant breeding matured, shifting breeding efforts from farmers to centralized public and private breeding programs. Centralized scientific plant breeding changed who is doing the selection and proposing the selection goals (Ashby 2009). Centralized plant breeding objectives are now meant to fill a regional need with a focus on the main production system (conventional farming) and industry requirements (Weltzien and Christinck 2009). Since the beginning of the green revolution farm specialization has risen (Evenson and Gollin 2003), establishing the concept of conventional farming which is composed of a well-defined uniform group of practices (i.e. crop fertilization and pesticide package), leaving outside a universe of alternative farming options such as organic or regenerative farming (Shennan et al. 2017). Consequently, cultivars selected aiming at conventional farming systems do not perform well or are not the top performers in alternative farming environments (Przystalski et al., 2008; Wolfe et al., 2008; Hildermann et al., 2009; Reid, et al. 2009). The low correlation between farming environments makes the use of indirect selection ineffective (Murphy et al. 2007; Przystalski et al. 2008). Furthermore, there is also a well-known gap in performance between experimental and conventional farmers' fields (Simmonds 1980, 1991; Annicchiarico 2002).

Modern agricultural practices are diverse, and are based on crop rotations, which is a cultural practice that increases the diversity of the agricultural field, helps to cut disease cycles (Schwartz 2020), nutrients cycles (Gaudin, et al., 2015), ecosystem services (Tamburini et al 2020) and boosts the overall productivity of the system (Janovicek et al. 2021). Diversity in crop rotations help to protect against climatic variability and helps to maintain yield stability over time (Gaudin, et al.,

2015) In the Midwest of the United States, the most common crop is corn (*Zea maiz* L.) grown in monoculture. An alternative when there is a crop rotation, it involves corn and soybean (*Glicine max* L. Burchfield et al. 2024). This crop rotations have reduced diversity, they have negative ecological impact (Dougherty et al. 2020), and have limited crop yield (Burchfield et al. 2024). Increasing diversity in crop rotations carries environmental benefits, such as increase in ecosystem diversity, soil quality and nutrient cycling. It can also increase the system productivity and grain yield (Tamburini et al 2020). On the diversity of options to incorporate in more diverse rotations, small grains promise to have a positive impact in the system productivity and sustainability (Janovicek et al. 2021).

Participatory plant breeding can be used to reduce the gap between research and farms in two ways, first, by reducing the differences between experimental stations and farm fields, and second by involving farmers, consumers and stakeholders in the selection process and decision-making (Ceccarelli & Grando, 2007). Participatory plant breeding implies an interaction between breeders and farmers and the participation of end-users in setting selection goals that will better meet their needs with a highly local focus (Ashby 2009). Participatory breeding can include different levels of involvement in the breeding process (Lammerts van Bueren and Myers 2012). In regions that are beyond subsistence agriculture, another layer of participatory breeding can be added including other stakeholders in the value-added chain (e.i. millers, bakers, chefs, the general public, etc. Dawson et al. 2011).

The high cost of a traditional breeding program forces breeders to aim to produce cultivars that can be marketed at the regional level. This breeding for a regional level creates the need to address the genotype by environment imteractiom (GEI). Understanding GEI well enough to develop adapted varieties can define the success or failure of plant breeding (Becker and Leon

1988). GEI is observed when genotypes show variation in relative performance across a set of environments (Bernardo 2002). GEI can be exploited by selecting for local adaptation or overcome by selecting genotypes with wide adaptation and low GEI (Bernardo 2002).

History

An important awakening occurred in biological sciences after the publication of Darwin's theory of natural selection (Darwin 1859) and the incorporation of selection as components of change in species (Darwin 1868; Allard 1961a). In addition, the works of Gregor Mendel incorporated the concept of segregation over the previous theory of hybrid blending (Bernardo 2014). These first approaches were expanded by Fisher (Fisher 1918), establishing a clear model for the inheritance of multiple mendelian 'factors'. Combined with theoretical developments in selection (Wright 1921, 1932; Haldane 1926, 1927), and population genetics (Wright 1922, 1949), constitute the basis for the foundations and development of quantitative genetics.

The next big advance in methods was the understanding of the structure of the DNA, the development of molecular markers, and the theory associated with their use. Genetic markers had limited use until the development of second-generation sequencing and the development of high throughput genomic information (Ekblom and Galindo 2011, Davey et al. 2011, Kumar et al. 201). These techniques need specialized personnel for plant breeding, which shifted the breeders' work from farmers' fields to research stations where plant breeding is carried out by scientists.

Breeder's equation

Traditional plant breeding works by recurrent cycles of evaluation, selection, and recombination to increase the genetic value of the population based on phenotypic information (Bernardo 2002; Rutkoski 2019; Cobb et al. 2019). The change in the mean between the parental population and the population generated from recombining the selected individuals is known as the realized repose to selection (Falconer 1966). The expected response to selection in a population can be estimated using the Breeder's equation (Lush 1937, Walsh and Lynch 2008).

$$R = ih\sigma_a$$

where R is the expected response to selection, i is the standardized selection intensity (i.e. standardized difference between the mean of the population and the mean of the selected individuals), h is the square root of the heritability (proportion of the total variance attributed to additive genetic variance) and represents the precision of the observed phenotype, and σ_a is the additive genetic deviation (square root of the additive genetic variance) and reflects the genetic diversity of the population. The equation can be weighted by the length of the breeding cycle to compare breeding cycles with different lengths (Eberhart 1970). The difficulty in observing the value of the genotype directly and the need to select indirectly by the observed phenotype make the phenotypic selection in characters with lower heritability less efficient (Bernardo, 2010).

Use of molecular markers in plant breeding

Since the structure of DNA was understood, a series of different attempts to generate molecular markers to identify and select genotypes have been developed. The first molecular markers were developed by using the detection of variations in the presence and absence of restriction sites Amplified Fragment Length Polymorphism (AFLP, Grover and Sharma 2016). The development of the polymerase chain reaction (PCR) allowed the use of markers based on the enrichment of DNA fragments, Restriction Fragment Length Polymorphism RFLP, Inter-simple sequence repeat (ISSR), Simple sequence repeat (SSR), etc. (Grover and Sharma 2016). With the development of the first generation of sequencing technology (Sanger and Coulson 1975; Sanger et al. 1977), other markers such as single nucleotide polymorphism (SNPs) were developed and applied in plant breeding (Rafalski 2002). As the molecular markers technology and availability evolved, diverse methodologies and an extensive theory have been developed around the identification of genotypic variants and their effect on the phenotype (Simko et al. 2021). The development of the second and third generation of sequencing technologies allowed the creation of reference genomes and the development of genetic maps with high density and coverage, with the possibility of easily genotyping hundreds or thousands of individuals; this enabled the extensive widespread use of molecular markers in plant breeding, marker-assisted selection, and genomic prediction.

The availability of whole-genome marker information and the possibility of genotyping a large number of individuals and population structure opens the possibilities for genomic selection (GS) (Meuwissen et al. 2001; Heffner et al. 2009). GS consists of the use of a population with phenotypic and genotypic information which is used to train a model and to predict/select individuals that only have genotypic information (Meuwissen et al. 2001). Genomic selection and the incorporation of wide genome marker data have better precision than the second-best technic available (marker-assisted selection; Bernardo and Yu 2007; Gaynor et al. 2017). Additionally, incorporating genomic selection in breeding programs can increase the rates of genetic gain per unit of time by shortening breeding cycles (Gaynor et al. 2017), reducing the amount of phenotyping needed by genotyped line optimizing the use of resources and increasing the genetic gain by the unit of time and unit of resource (phenotyped plot; Endelman et al. 2014), even when using phenotypic information with low heritability (Endelman 2014). GS opens the possibility to predict the performance of untested genotypes in both known (Burgeno 2008, Dawson 2013, Lado et al 2016), and unknown environments (Heslot et al 2014). GS can be beneficial for farmers in a direct way through the release of better cultivars and can be used to improve the quality of the information that farmers are receiving.

Genotype by environment interaction in plant breeding

Genotype-by-environment interaction (GEI) happens when a group of genotypes reacts in different ways to the changes in the environment (Lynch and Walsh 1998). Modeling the GEI is useful in plant breeding to identify differences in GEI and shape the allocation of resources for cultivar evaluation and selection strategies. Modeling GEI is particularly of interest when combined with genomic information, improving the precision of estimations in tested genotypes (Lopez-Cruz et al. 2015), and also improving predictions in untested genotypes (Dawson 2013; Lado et al. 2016). Studying GEI using methods such as AMMI (additive main effects and multiplicative interaction) and GGE biplot analysis allows for defining mega environments with common genotypic rankings (Yan et al. 2007; Gauch et al. 2008).

Another dimension to observe GEI is at the genotype level, analyzing how the genotypes respond to changes in the environment, which is known as stability (Becker and Leon 1988). Stability can be classified as static or dynamic based on the function of the response to changes in the environment (Becker and Leon 1988). A genotype with dynamic stability responds to improvements in the environment showing increases equal to the average environment in a linear form. This can be tested in different ways, for example, using environmental regressions, Wricke's ecovalence, etc. (Finlay and Wilkinson 1963, Wricke, 1962). Obtaining a reliable yield stability characterization for a genotype requires at least a sample of 150 environments (Wang et al. 2023). A genotype with static genetic response is useful in plant breeding when breeders are looking for a genotype that does not show changes in performance in a specific group of target environments (Annicchiarico 2002; van Eeuwijk et al. 2016).

Participatory plant breeding

Participatory breeding has been used as a strategy to overcome the lack of relevance of traditional breeding, for example, in subsistence agriculture, small-scale farms, or organic farms (Ceccarelli and Grando 2007; Ashby 2009; Kucek 2017). The main idea behind participatory plant breeding is to involve the beneficiary of the final product in the process of decision-making, selection, and breeding (Ceccarelli and Grando 2007). Where subsistence agriculture is practiced the main beneficiary is the farmer (Ceccarelli and Grando 2007; Ashby 2009), in places where economic resources are abundant the production chain is more extensive and farmers, processors, and end-users can be involved (Dawson et al. 2011). Objectives of selection and priorities can be set using tools such as surveys (Kucek 2017). Participatory plant breeding represents a strategy for improving the delivery of genetic material to non-traditional farmers filling market niches that are not attractive to larger seed companies.

Cereals

Wheat is the most important food grain produced in the United States, with a production of over 51 million metric tons (FAOSTATS 2020).The majority of the U.S. commercial wheat varieties are developed by public plant breeding programs (Wheat CAP 2018), contributing billions in production value (USDA/NASS 2021). Most of the wheat is grown for the commodity grain system and industrial baking markets, with most breeders developing varieties adapted to conventional systems (Tilley et al. 2012; Kiszonas and Morris 2018). Wheat is a good example of a crop where the seed market is highly influenced by varieties developed for conventional agriculture. Still, land in organic wheat production increased by 37% from 2019 to 2024 (USDA 2019, USDA 2024). Specifically, organic production is one of the fastest-growing sectors in agriculture (Matlock 2021) and production is expected to continue growing since revenue and the number of organic farms have been growing in recent years in the U.S. (57% from 2019 to 2023) and particularly in the U.S. Upper Midwest (119% 2019 to 2024) (USDA 2024). However, there has been a very limited effort in developing wheat varieties adapted to organic systems in the U.S. with few breeding programs dedicated to developing varieties for whole-grain end-uses or sourdough, resulting in a lack of evaluation and selection strategies as well as limited availability of varieties with suitable end-use quality (Ross 2018; Krill-Brown et al. 2019). Therefore, breeding programs developing hard red wheat varieties adapted to regional climates and organic products with good quality for artisanal products are necessary and will provide farmers with greater opportunities to reach high-value markets.

The bread wheat (*Triticum aestivum L.*) is an allohexaploid (2n=6x= 42) with an autogamous reproductive system. Bread wheat is part of the wheat group which includes, diploids (2x), allotetraploids (4x), and allohexaploid (6x) generated by interspecific hybridization and genome duplication (Feldman and Levy 2012). Bread wheat is an amphidiploid, which means that its genome is composed of three different sub-genomes (AABBDD) but the segregation and meiotic behavior is equal to a diploid (Breiman and Graur 1995). This has evolutive advantages and in the case of wheat is closely related to the domestication process (Feldman and Levy 2012). Bread wheat has been divided by physiological functionality into two groups winter wheat (which requires vernalization before flowering) and spring wheat which does not have cold tolerance and does not require vernalization. Also, each of these groups has been divided into four market categories (soft white, soft red, hard white, and hard red) which are grouped based on color, starch and protein content and their end-use. Wheat is among the first species with science-guided breeding efforts (Gayon and Zallen 1998); because of its importance in food it has an extended breeding history.

Oats (*Avena sativa*) is another cereal where public breeding efforts are important in the release of cultivars. Among cereals, oat ranks 7th at the world level with average production in the U.S. of 16 million tons which represent 7% of the total oat produced in the world (Webster 2016). Oats is a multipurpose crop, being used as forage, and its grain as feed or food depending on the quality levels. Also, it is the cereal with the highest level of protein and other several nutritional and health benefits (Mushtaq et al. 2014; Hu et al. 2021). Oats have a particular interest because of their plasticity in grain production used as either feed or human food. There is exhaustive testing of oat cultivars at the state level and across states (Gutierrez 2021, Smith 2021, Kleinjan 2021). This network of trials has been repeated over many years and locations. Still, this information is not easily accessible to interpret due to the volume, unbalance, and complexity of the data make it difficult for farmers and/or researchers to access.

Oat is part of an polyploid complex with ploidies ranging from diploid to hexaploid with a basic chromosome number of 7 and an autogamous reproductive system (Thomas 1992). Oats breeding in the U.S. is done in allohexaploid oats (*Avena sativa L.;* Thomas 1992) and started with the selection from a few widespread landraces with high diversity. The main selection target traits in oats are grain yield, grain quality, and disease resistance, and since oats is a multipurpose crop, forage quality is also important. Line selection and development are performed using classical breeding approaches in the principal oat breeding programs in the U.S. There is a reference genome available for oats (Yao et al. 2022), and efforts to incorporate next generation techniques genomics and phenomics in oat have been done successfully (Asoro et al. 2011; Haikka et al. 2020b, a; Campbell et al. 2021; Berro et al. 2023) and other technologies to accelerate the breeding efficiency have been tested (Heuschele et al. 2019; González-Barrios et al. 2020).

Thesis overview

This thesis presents the results of three projects that advance farmer driven plant breeding and sustainable cereal breeding in the Upper Midwest. Chapter 2 describes a participatory wheat selection process that resulted in a new cultivar for organic systems and artisan bread quality. Chapter 3 presents the development of a decision support tool for farmers. The tool makes oats variety recommendations to farmers based on their location and a genomic prediction model that draws on historical trial data. Chapter 4 addresses the issue of cultivar stability in addition to prediction of performance in the face of GEI. This chapter used mixed models to make predictions of cultivar performance and stability. The analysis, decision support tools and germplasm development through this project should be useful to breeders and farmers working to increase the performance and stability of cereal cultivars.

Objectives:

The objective of this dissertation is to help increase the sustainability of agricultural systems in the Upper Midwest using plant breeding as the means to help incorporate small grains in crop rotations. This will be achieved using plant breeding tools through different avenues. First, we will describe the development of a breeding program for bread wheat varieties targeted to organic systems and artisan baking in the U.S. Upper Midwest. We have involved farmers, millers, and bakers in the decision process for advancing breeding lines. Specifically, our goals include 1) analysis of the genotype-by-environment interaction (GEI) for agronomic and quality performance, including the study of stability indicators for transient year-to-year GEI, 2) a comparison of on-station and on-farm performance for advanced breeding lines, and the incorporation of on-farm evaluation and farmers input in the decision process, and 3) development of baking trials with bakers input to evaluate artisan whole grain baking quality in advanced breeding lines. This chapter was published as an invited paper to a special edition of Theoretical Applied and Genetics themed on organic wheat breeding.

Second, we will characterize the GEI and identify oat mega-environments in the Midwestern regions of the United States using a large-scale historical variety trial dataset from the Midwestern oat breeding programs. We will combine sources of historical information and genomic selection to predict the lines and cultivars that are better adapted to specific farmers' regions, and to support the development of a "Variety decision tool" (https://decide.practicalfarmers.org/) where farmers can get an individualized prediction of the performance of available varieties for their local environment.

Third, we will predict genotypic reaction norms and sensitivity to environmental changes across environments using a data set in barley, one in oats, one data set in wheat and mixed models for genomic prediction. We will use available information from relatives tested in other environments by exploiting genetic relationships to leverage information.

References

Allaby RG (2014) Domestication syndrome in plants. In: Encyclopedia of global archaeology. Springer New York, pp 2182–2184

Allard RW (1961a) Principles of plant breeding. Soil Sci 91:414

Annicchiarico P (2002) Genotype x environment interactions: challenges and opportunities for plant breeding and cultivar recommendations. Food & Agriculture Org., Rome, Italy.

Ashby JA (2009) The impact of participatory plant breeding. In: Plant breeding and farmer participation. FAO, Rome, Italy, pp 649–671

Becker HC, Leon J (1988) Stability Analysis in Plant Breeding. Plant Breed 101:1–23. https://doi.org/10.1111/j.1439-0523.1988.tb00261.x

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma press Woodbury

Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Sci 47:1082–1090. https://doi.org/10.2135/cropsci2006.11.0690

Bernardo, R. (2010). Breeding for quantitative traits in plants 2nd ed. Stemma Press, Woodbury, MN.

Berro I, Varela J, Gutiérrez L (2023) An image-based methodology to evaluate oat panicle architecture. Crop Sci 63:648–661. https://doi.org/10.1002/csc2.20884

Breiman A, Graur D (1995) WHEAT EVOLUTION. Isr J Plant Sci 43:85–98. https://doi.org/10.1080/07929978.1995.10676595

Burchfield, E. K., Crossley, M. S., & Nelson, K. S. (2024). Rotational complexity across US counties is currently insufficient to observe yield gains in major crops. Environmental Research Letters, 19(4), 044024. https://doi.org/10.1088/1748-9326/ad300b

Burgueño, J., Crossa, J., Cornelius, P. L., & Yang, R.-C. (2008). Using factor analytic models for joining environments and genotypes without crossover genotype× environment interaction. Crop Science, 48(4), 1291–1305.

Campbell MT, Hu H, Yeats TH, et al (2021) Improving Genomic Prediction for Seed Quality Traits in Oat (Avena sativa L.) Using Trait-Specific Relationship Matrices. Front Genet 12:643733. https://doi.org/10.3389/fgene.2021.643733

Ceccarelli S, Grando S (2007) Decentralized-participatory plant breeding: an example of demand driven research. Euphytica 155:349–360. https://doi.org/10.1007/s10681-006-9336-8

Cobb JN, Juma RU, Biswas PS, et al (2019) Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder’s equation. Theor Appl Genet 132:627–645. https://doi.org/10.1007/s00122-019-03317-0

Cox S (2009) Crop domestication and the first plant breeders. Plant Breed Farmer Particip 1–26

Darwin C (1859) On the Origin of Species, 1859. Routledge, London

Darwin CR (1868) Variation of plants and animals under domestication

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics, 12(7), 499–510. https://doi.org/10.1038/nrg3012

Dawson JC, Rivière P, Berthellot J-F, et al (2011) Collaborative Plant Breeding for Organic Agricultural Systems in Developed Countries. Sustainability 3:1206–1223. https://doi.org/10.3390/su3081206

Dawson, J. C., Endelman, J. B., Heslot, N., Crossa, J., Poland, J., Dreisigacker, S., Manès, Y., Sorrells, M. E., & Jannink, J.-L. (2013). The use of unbalanced historical data for genomic selection in an international wheat breeding program. Field Crops Research, 154, 12–22.

De Wet JMJ, Harlan JR (1975) Weeds and Domesticates: Evolution in the man-made habitat. Econ Bot 29:99–108. https://doi.org/10.1007/BF02863309

Diamond J (2002) Evolution, consequences and future of plant and animal domestication. Nature 418:700–707. https://doi.org/10.1038/nature01019

Dougherty, B. W., Pederson, C. H., Mallarino, A. P., Andersen, D. S., Soupir, M. L., Kanwar, R. S., & Helmers, M. J. (2020). Midwestern cropping system effects on drainage water quality and crop yields. Journal of Environmental Quality, 49(1), 38–49. https://doi.org/10.1002/jeq2.20007

Eberhart SA (1970) Factors affecting efficiencies of breeding methods. Afr Soils 15:655–680

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity, 107(1), 1–15. https://doi.org/10.1038/hdy.2010.152

Endelman, J. B., Atlin, G. N., Beyene, Y., Semagn, K., Zhang, X., Sorrells, M. E., & Jannink, J.-L. (2014). Optimal design of preliminary yield trials with genome-wide markers. Crop Science, 54(1), 48–59.

Evenson RE, Gollin D (2003) Assessing the Impact of the Green Revolution, 1960 to 2000. Science 300:758–762. https://doi.org/10.1126/science.1078710

FAOSTATS (2020) World Food and Agriculture - Statistical Yearbook 2020. FAO, Rome, Italy

Feldman M, Levy AA (2012) Genome Evolution Due to Allopolyploidization in Wheat. Genetics 192:763–774. https://doi.org/10.1534/genetics.112.146316

Finlay K, Wilkinson G (1963) The analysis of adaptation in a plant-breeding programme. Aust J Agric Res 14:742. https://doi.org/10.1071/AR9630742

Fisher RA (1922) On the mathematical foundations of theoretical statistics. Philos Trans R Soc Lond Ser Contain Pap Math Phys Character 222:309–368

Fisher RA (1938) The statistical utilization of multiple measurements. Ann Eugen 8:376–386

Fisher RA (1925) Theory of statistical estimation. In: Mathematical proceedings of the Cambridge philosophical society. Cambridge University Press, pp 700–725

Fisher RA (1918) 009: The Correlation Between Relatives on the Supposition of Mendelian Inheritance.

Fuller, D. Q., & Allaby, R. (2009). Seed Dispersal and Crop Domestication: Shattering, Germination and Seasonality in Evolution under Cultivation. In L. Stergaard (Ed.), Fruit Development and Seed Dispersal (pp. 238–295). Wiley-Blackwell. https://doi.org/10.1002/9781444314557.ch7

Gauch HG, Piepho H-P, Annicchiarico P (2008) Statistical Analysis of Yield Trials by AMMI and GGE: Further Considerations. Crop Sci 48:866. https://doi.org/10.2135/cropsci2007.09.0513

Gaudin, A. C. M., Janovicek, K., Deen, B., & Hooker, D. C. (2015). Wheat improves nitrogen use efficiency of maize and soybean-based cropping systems. Agriculture, Ecosystems & Environment, 210, 1–10. https://doi.org/10.1016/j.agee.2015.04.034

Gaynor RC, Gorjanc G, Bentley AR, et al (2017) A Two-Part Strategy for Using Genomic Selection to Develop Inbred Lines. Crop Sci 57:2372–2386. https://doi.org/10.2135/cropsci2016.09.0742

Gayon J, Zallen DT (1998) The Role of the Vilmorin Company in the Promotion and Diffusion of the Experimental Science of Heredity in France, 1840-1920. J Hist Biol 31:241–262

González-Barrios P, Bhatta M, Halley M, et al (2020) Speed breeding and early panicle harvest accelerates oat (Avena sativa L.) breeding cycles. Crop Sci csc2.20269. https://doi.org/10.1002/csc2.20269

Grover A, Sharma PC (2016) Development and use of molecular markers: past and present. Crit Rev Biotechnol 36:290–302. https://doi.org/10.3109/07388551.2014.959891

Haikka H, Knürr T, Manninen O, et al (2020a) Genomic prediction of grain yield in commercial Finnish oat (Avena sativa) and barley (Hordeum vulgare) breeding programmes. Plant Breed 139:550–561. https://doi.org/10.1111/pbr.12807

Haikka H, Manninen O, Hautsalo J, et al (2020b) Genome-wide Association Study and Genomic Prediction for Fusarium graminearum Resistance Traits in Nordic Oat (Avena sativa L.). Agronomy 10:174. https://doi.org/10.3390/agronomy10020174

Haldane JBS (1926) A mathematical theory of natural and artificial selection. Math Proc Camb Philos Soc 23:363–372. https://doi.org/10.1017/S0305004100015176

Haldane JBS (1927) A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. Math Proc Camb Philos Soc 23:838–844. https://doi.org/10.1017/S0305004100015644

Hardy K, Radini A, Buckley S, et al (2016) Dental calculus reveals potential respiratory irritants and ingestion of essential plant-based nutrients at Lower Palaeolithic Qesem Cave Israel. Quat Int 398:129–135. https://doi.org/10.1016/j.quaint.2015.04.033

Harlan, J. R. (1986). Plant Domestication: Diffuse Origins and Diffusions. In Developments in Agricultural and Managed Forest Ecology (Vol. 16, pp. 21–34). Elsevier. https://doi.org/10.1016/B978-0-444-42703-8.50007-5

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic Selection for Crop Improvement. Crop Sci 49:1–12. https://doi.org/10.2135/cropsci2008.08.0512

Heslot, N., Akdemir, D., Sorrells, M. E., & Jannink, J.-L. (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theoretical and Applied Genetics, 127, 463–480.

Heuschele DJ, Case A, Smith KP (2019) Evaluation of Fast Generation Cycling in Oat (Avena sativa). Cereal Res Commun 47:626–635. https://doi.org/10.1556/0806.47.2019.40

Hildermann I, Thommen A, Dubois D, et al (2009) Yield and baking quality of winter wheat cultivars in different farming systems of the DOK long-term trial. J Sci Food Agric 89:2477–2491. https://doi.org/10.1002/jsfa.3750

Hu H, Campbell MT, Yeats TH, et al (2021) Multi-omics prediction of oat agronomic and seed nutritional traits across environments and in distantly related populations. Theor Appl Genet 134:4043–4054. https://doi.org/10.1007/s00122-021-03946-4

Janovicek, K., Hooker, D., Weersink, A., Vyn, R., & Deen, B. (2021). Corn and soybean yields and returns are greater in rotations with wheat. Agronomy Journal, 113(2), 1691–1711. https://doi.org/10.1002/agj2.20605

Kiszonas AM, Morris CF (2018) Wheat breeding for quality: A historical review. Cereal Chem 95:17–34

Krill-Brown A, Vang T, Carter C (2019) Developing Wheat Varieties for Whole Grain Artisanal Baking. Cereal Foods World 64

Kucek L (2017) Participatory Breeding of Wheat for Organic Production. Dissertation, Cornell University

Kumar, R., Ichihashi, Y., Kimura, S., Chitwood, D. H., Headland, L. R., Peng, J., Maloof, J. N., & Sinha, N. R. (2012). A High-Throughput Method for Illumina RNA-Seq Library Preparation. Frontiers in Plant Science, 3. https://doi.org/10.3389/fpls.2012.00202

Lado, B., Barrios, P. G., Quincke, M., Silva, P., & Gutiérrez, L. (2016). Modeling genotype× environment interaction for genomic selection with unbalanced data from a wheat breeding program. Crop Science, 56(5), 2165–2179.

Lopez-Cruz M, Crossa J, Bonnett D, et al (2015) Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker × Environment Interaction Genomic Selection Model. G3amp58 GenesGenomesGenetics 5:569–582. https://doi.org/10.1534/g3.114.016097

Lu, S., Fang, C., Abe, J., Kong, F., & Liu, B. (2022). Current overview on the genetic basis of key genes involved in soybean domestication. aBIOTECH, 3(2), 126–139. https://doi.org/10.1007/s42994-022-00074-5

Lush JL (1943) Animal breeding plans. Anim Breed Plans

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 154:1819–1829

Murphy KM, Campbell KG, Lyon SR, Jones SS (2007) Evidence of varietal adaptation to organic farming systems. Field Crops Res 102:172–177. https://doi.org/10.1016/j.fcr.2007.03.011

Mushtaq A, Gul-Zaffar ., Z. AD, Mehfuza H (2014) A review on Oat (Avena sativa L.) as a dual-purpose crop. Sci Res Essays 9:52–59. https://doi.org/10.5897/SRE2014.5820

Przystalski M, Osman A, Thiemt EM, et al (2008) Comparing the performance of cereal varieties in organic and non-organic cropping systems in different European countries. Euphytica 163:417–433. https://doi.org/10.1007/s10681-008-9715-4

Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. Nature 457:843–848. https://doi.org/10.1038/nature07895

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100. https://doi.org/10.1016/S1369-5266(02)00240-6

Reid TA, Yang R-C, Salmon DF, Spaner D (2009) Should spring wheat breeding for organically managed systems be conducted on organically managed land? Euphytica 169:239–252. https://doi.org/10.1007/s10681-009-9949-9

Ross AS (2018) Flour Quality and Artisan Bread. Cereals Foods World 63

Rutkoski JE (2019) A practical guide to genetic gain. In: Advances in Agronomy. Elsevier, Cambridge, MA, pp 217–249

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94:441–448

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74:5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Schwartz T (2020) Making Diverse Rotations Work - Small Grains in the Corn Belt

Shennan C, Krupnik TJ, Baird G, et al (2017) Organic and Conventional Agriculture: A Useful Framing? Annu Rev Environ Resour 42:317–346. https://doi.org/10.1146/annurev-environ-110615-085750

Simko, I., Jia, M., Venkatesh, J., Kang, B.-C., Weng, Y., Barcaccia, G., Lanteri, S., Bhattarai, G., & Foolad, M. R. (2021). Genomics and Marker-Assisted Improvement of Vegetable Crops. Critical Reviews in Plant Sciences, 40(4), 303–365. https://doi.org/10.1080/07352689.2021.1941605

Simmonds NW (1980) Comparisons of the Yields of Four Potato Varieties in Trials and in Agriculture. Exp Agric 16:393–398. https://doi.org/10.1017/S0014479700012059

Simmonds NW (1991) Selection for local adaptation in a plant breeding programme. Theor Appl Genet 82:363–367. https://doi.org/10.1007/BF02190624

Smith BD (2001) Documenting plant domestication: The consilience of biological and archaeological approaches. Proc Natl Acad Sci 98:1324–1326. https://doi.org/10.1073/pnas.98.4.1324

Spencer RD, Cross RG (2007) The International Code of Botanical Nomenclature (ICBN), the International Code of Nomenclature for Cultivated Plants (ICNCP), and the cultigen. TAXON 56:938–940. https://doi.org/10.2307/25065875

Tamburini, G., Bommarco, R., Wanger, T. C., Kremen, C., Van Der Heijden, M. G. A., Liebman, M., & Hallin, S. (2020). Agricultural diversification promotes multiple ecosystem services without compromising yield. Science Advances, 6

Thomas, Hugh. (1992). Cytogenetics of Avena. In H. G. Marshall & M. E. Sorrells (Eds.), Agronomy Monographs (pp. 473–507). American Society of Agronomy, Crop Science Society of America. https://doi.org/10.2134/agronmonogr33.c14

Tanno K, Willcox G. How fast was wild wheat domesticated? Science. 2006 Mar 31;311(5769):1886. doi: 10.1126/science.1124635. PMID: 16574859.

Tilley M, Chen YR, Miller RA (2012) Wheat breeding and quality evaluation in the US. In: Breadmaking. Elsevier, Cambridge, UK, pp 216–236

USDA/NASS (2021) USDA/NASS QuickStats Ad-hoc Query Tool. In: QuickStats Ad-Hoc Query Tool. https://quickstats.nass.usda.gov/results/DF990781-9A03-30B0-BA7A-AE0D627F1F0B?pivot=short_desc. Accessed 14 Oct 2021

USDA/NASS (2019) 2019 Organic Survey. https://www.nass.usda.gov/Publications/AgCensus/2017/Online_Resources/Organics/index.php. Accessed 14 Oct 2021

USWBSI S-U (2019) Home Page | USWBSI. https://scabusa.org/. Accessed 14 Oct 2021 van Eeuwijk FA, Bustos-Korts DV, Malosetti M (2016) What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype × Environment Interactions? Crop Sci 56:2119–2140. https://doi.org/10.2135/cropsci2015.06.0375

Walsh, B., & Lynch, M. (2018). Evolution and selection of quantitative traits. Oxford University Press.

Wang, T.-C., Casadebaig, P., & Chen, T.-W. (2023). More than 1000 genotypes are required to derive robust relationships between yield, yield stability, and physiological parameters: A computational study on wheat crop. Theoretical and Applied Genetics, 136(3), 34. https://doi.org/10.1007/s00122-023-04264-7

Webster F (2016) Oats: Chemistry and Technology. Academic Press

Weltzien E, Christinck A (2009) Methodologies for priority setting

Wolfe MS, Baresel JP, Desclaux D, et al (2008) Developments in breeding cereals for organic agriculture. Euphytica 163:323–346. https://doi.org/10.1007/s10681-008-9690-9

Wricke, G. (1962). Uber eine methode zur erfassung der okologischen streubreite in feldversucen. Z. Pflanzenzuchtung, 47, 92–96.

Wright S (1922) Coefficients of Inbreeding and Relationship. Am Nat 56:330–338. https://doi.org/10.1086/279872

Wright S (1931) Evolution in Mendelian Populations. Genetics 16:97–159

Wright S (1939) Genetic principles governing the rate of progress of livestock breeding. J Anim Sci 1939:18–26

Wright S (1949) the genetical structure of populations. Ann Eugen 15:323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Wright S (1921) Systems of mating. IV. The effects of selection. Genetics 6:162

Wright S (1932) The roles of mutation, inbreeding, crossbreeding, and selection in evolution

Yan W, Kang MS, Ma B, et al (2007) GGE Biplot vs. AMMI Analysis of Genotype-by-Environment Data. Crop Sci 47:643–653. https://doi.org/10.2135/cropsci2006.06.0374

Yao, E., Blake, V. C., Cooper, L., Wight, C. P., Michel, S., Cagirici, H. B., Lazo, G. R., Birkett, C. L., Waring, D. J., Jannink, J.-L., Holmes, I., Waters, A. J., Eickholt, D. P., & Sen, T. Z. (2022). GrainGenes: A data-rich repository for small grains genetics and genomics. Database, 2022, baac034. https://doi.org/10.1093/database/baac034

Chapter 2. Developing high-quality value-added cereals for organic systems in the U.S. Upper Midwest: hard red winter wheat (*Triticum aestivum* L.) breeding

This article was previously published in Theorical and Applied Genetics on 28 May 2022.

Sandro, P., Kucek, L.K., Sorrells, M.E., Dawson, J.C., and Gutierrez, L. Developing high-quality value-added cereals for organic systems in the US Upper Midwest: hard red winter wheat (*Triticum aestivum* L.) breeding. Theor Appl Genet 135, 4005–4027 (2022). https://doi.org/10.1007/s00122-022-04112-0

Author contributions

JD and MES performed the original crosses. JD and LKK performed evaluation and selection of segregating populations. PS, JD, MES, and LG performed evaluation and selection of advanced breeding lines. PS and LG designed phenotyping experiments. PS, LG, and JD performed the statistical analyses and wrote the manuscript. JD and LG designed the study and hypothesis. All authors read and approved of the final manuscript.

*Corresponding authors: *dawson@wisc.edu*, *gutierrezcha@wisc.edu*

ORCID ID: https://orcid.org/0000-0002-2957-308

Keywords. wheat, organic breeding, participatory plant breeding, genotype by environment interaction, stability, farm to table

Funding. USDA/NIFA, Organic Agriculture Research & Extension Initiative, Proposal Number: 2020-02136 and 2011–51300-30697; United States Department of Agriculture; USDA/SARE Sustainable Agriculture Research and Education grants LNE12-318 and GNE15-107; Hatch project 149-945; and WARF/WCIA/CALS MSN234477, MSN244376, and MSN256676. This project was also funded by the UW-Madison Center for Integrated Agricultural Systems and the College of Agricultural and Life Sciences (CIAS).

Availability of data and material. All data will be deposited in the T3 database (*https://wheat.triticeaetoolbox.org/*). Researchers interested in plant materials should contact corresponding authors.

Ethics approval

The sensory evaluation component of the project was determined to be exempt by the UW Madison Education and Social/Behavioral Science IRB, Submission ID Number 2014-1008 under category 45 CFR 46.101(b)(6) Taste and food quality evaluation.

Key Message

We describe a hard-red winter wheat breeding program focused on developing genotypes adapted to organic systems in the U.S. Upper Midwest with high end artisan baking quality using participatory approaches.

Abbreviations

BLUEs: best linear unbiased estimates

BYDV: barley yellow dwarf virus

DON: deoxynivalenol

FHB: Fusarium head blight

FW: Finlay-Wilkinson regression

GEI: genotype by environment interaction

RCBD: randomized complete block design

W: Wricke's ecovalence stability coefficient

Acknowledgments

We thank David Benscher and James Tanaka of the Cornell Small Grains program for early generation trial management, David Benscher, Phil Holman, Allison Krill, Olivia Handal and Jessica Rutkoski for advance generation trial management in NY, Spooner and Illinois, and Janet Hedtcke for organic field management in Madison. We thank participating farmers and bakers: Thor Oeschner of Oeschner Farms, Harold Wilken of Janie's Farm, Paul Bickford, John Wepking and Halee Wepking of Meadowlark Organics, bakers Melina Kelson of Bootleg Batard in Chicago, IL, Solveig Tofte of Sun Street Breads in Minneapolis, MN, Greg Wade of Publican Quality Bread in Chicago, Kirk Smock of ORIGIN Breads in Madison, WI, Matt Kronschnabel of Bard Bread in Viroqua, WI, and Andrew Hutchison of Madison Sourdough. We would also like to thank Madison Sourdough staff for hosting baking and sensory evaluations and Alyssa Hartman and the Artisan Grain Collaborative for facilitating baking and sensory evaluations, and all participants in the tasting events.

Abstract

There is an increased demand for food grade grains grown sustainably. Hard-red winter wheat has comparative advantages for organic farm rotations due to fall soil cover, weed competition, and grain yields. However, limitations of currently available cultivars such as poor disease resistance, winter hardiness, and baking quality, challenges its adoption and use. Our goal was to develop a participatory hard-red winter wheat breeding program for the U.S. Upper Midwest involving farmers, millers, and bakers. Specifically, our goals include an evaluation of genotype by environment interaction (GEI) and genotypic stability for both agronomic and quality traits, and the development of on-farm trials and baking and sensory evaluations of genotypes to include farmers, millers, and bakers' perspectives in the breeding process. Selection in early generations for diseases and protein content was followed by multi-environment evaluations for agronomic, disease, and quality traits in three locations during five years, on-farm evaluations, baking trials, and sensory evaluations. GEI was substantial for most traits, but no repeatable environmental conditions were significant contributors to GEI making selection for stability a critical trait. Breeding lines had similar performance in on-station and on-farm trials compared to commercial checks, but some breeding lines were more stable than the checks for agronomic, quality traits, and baking performance. These results suggest that stable lines can be developed using a participatory breeding approach under organic management. Crop improvement explicitly targeting sustainable agriculture practices for selection with farm-to-table participatory perspectives are critical to achieve long term sustainable crop production.

Introduction

Plant breeding has been highly successful in improving crops for serving humans food, fiber, and fuel by focusing on traits such as yield, quality, and disease resistance among others (Bernardo, 2010; Duvick, 2003). There is currently a need to achieve sustainable intensification by expanding the food, feed, and fuel goals to include soil, water, and biodiversity targets for agriculture (Heaton et al. 2013; Runck et al. 2014). Because of their role in creating new crops, plant breeders play a crucial role in the development of sustainable agricultural systems for the future (Brummer et al. 2011). Some general breeding strategies have included selection for lower input requirements that may decrease fossil fuel use, water contamination, and production cost (Tillman, 1999; Robertson and Swinton, 2005; Dawson et al. 2008); adaptation to stress (Araus et al. 2008; Cattivelli et al. 2008; Jacobs, 2007; Santini et al. 2007) and highly diverse cropping systems (Cecarelli et al. 2010; Liebman et al. 2012). Regional breeding strategies to withstand variable and or changing environments; selection for local adaptation (Cecarelli and Grando, 2007) and cropping systems (Cook, 2006) that improve performance under specific environments while providing ecosystem services have also been implemented. Finally, breeding for strategic spatio-temporal utilization of resources (Runck et al. 2014) that produce winter-annual or perennial crops in areas that would not otherwise be used at that time (Schulte et al. 2006; Heaton et al. 2013); and selection for completely new systems such as natural ecosystems (Glover et al. 2010) or organic agriculture (Dawson et al 2008(Wolfe et al. 2008; Dawson et al. 2011; Lammerts van Bueren and Myers 2012) have been successful.

Wheat (*Triticum aestivum* L.) is the most important food grain produced in the United States, with production over 51 million metric tons (FAOSTATS 2020), and the U.S. is among the top three exporters of wheat in the world. However, wheat differs from other major crops grown in the U.S.

because the majority of the U.S. commercial wheat varieties are developed by public plant breeding programs (Wheat CAP 2018), contributing billions in production value (USDA/NASS 2021). Furthermore, in the predominant wheat producing regions of the U.S., growers are supported by a commodity grain system and industrial baking markets, with most breeders developing varieties adapted to conventional systems (Tilley et al. 2012; Kiszonas and Morris 2018).

In many regions of the U.S., the demand for local and organic products is growing (Green et al. 2017; Rana and Paul 2017) while organic production is one of the fastest growing sectors in agriculture (Matlock 2021). Specifically, land in organic wheat production increased 15% from 2015 to 2019 (Matlock 2021) and production is expected to continue growing since revenue and the number of organic farms has been growing in recent years in the U.S. (15% 2017 to 2019) and particularly in the U.S. Upper Midwest (10% 2017 to 2019) (USDA/NASS 2019). Although most cereal grains in the U.S. Upper Midwest are grown for feed, food-grade grains present a high value and more consistent market outside of the commodity wheat system. Local grain markets are also growing, and both organic and local markets are responding to strong consumer demand for artisanal breads such as whole grain naturally leavened (sourdough). Expanding the production of food-grade grains for artisanal products also supports an increase in whole grain consumption with its associated health benefits (Mellen et al. 2008; Jonnalagadda et al. 2011; Reynolds et al. 2019). However, few breeding programs are dedicated to developing varieties for whole grain end-uses or sourdough, resulting in a lack of evaluation and selection strategies as well as limited availability of varieties with suitable end-use quality (Ross 2018; Krill-Brown et al. 2019). Therefore, breeding programs developing hard red wheat varieties adapted to regional climates and organic production with good quality for artisanal products will provide farmers with greater opportunities to reach high-value markets.

Breeding targets for hard red bread wheat generally include grain yield, winter hardiness, protein concentration, resistance to pre-harvest sprouting (for baking quality) and resistance to Fusarium head blight (*Fusarium graminearum* L., FHB) and foliar diseases. However, cultivars selected under conventional systems may not be suitable for organic production because genotype by management performance rank changes have been observed between organic and conventional systems for grain yield and protein content in cereals in general (Przystalski et al. 2008; Wolfe et al. 2008) and in wheat in particular (Reid et al. 2009; Hildermann et al. 2009). Moreover, this strong interaction between genotype and management system, and high heritability for these traits in organic trials, makes direct selection under organic conditions more effective (Murphy et al. 2007; Przystalski et al. 2008). A survey and interviews with organic wheat farmers identified traits where more research and development are needed for organic bread wheat production (Kucek 2017). High protein with artisanal baking quality and flavor, disease resistance including resistance to FHB and other leaf diseases, winter hardiness, weed competitive ability and overall good agronomic performance were identified as the key target traits for farmers (Kucek 2017). Discussions with bakers and millers have identified stability for quality parameters and good technical performance in whole grain and sourdough products as key traits. Stability is particularly important in regional grains systems as less blending is possible for grain with different quality parameters from different production regions to buffer yearly fluctuations in protein levels, falling number or disease incidence.

In addition to meeting the target traits for organic systems, it is important to consistently obtain high-quality grain across locations and years. There are two levels where stability is important:

temporal and spatial stability. Organic farms tend to be more diverse from farm-to-farm and field-to-field than conventional farms (Shennan et al. 2017; Knapp and van der Heijden 2018) with additional changes in genotypic rankings from year to year (Kucek et al., 2019) and would therefore benefit from both spatial and temporal stability. Furthermore, crop yield losses resulting from extreme temperatures, recurring droughts, and erratic rainfall patterns due to climate change are expected to have global impact (Wassman et al. 2009). The goal of breeding for climate-resilient crops is to breed for increasing crop yield production that is not limited by multiple challenges related to abiotic stress, such as drought or heat stress (Bhatta et al. 2018; Picasso et al. 2019), and biotic stresses, such as diseases, pests, and weeds (Ceccarelli et al. 2010; Malosetti et al. 2013; Wani et al. 2018). Therefore, breeding efforts should focus on continuing to identify genotypes with high yield potential while reducing the yield gap under sub-optimal growing conditions due to extreme weather events and climate change (Araus et al. 2008; Fischer and Edmeades 2010; Pennacchi et al. 2018). The increasing frequency of extreme weather events means that the environmental conditions faced by new genotypes may be outside of the range of historic variation in a target region. This makes it critical to select genotypes for spatial stability at a larger scale such as a regional level. Longer term stability may also be favored by higher levels of genetic diversity that can be accomplished by higher within variety diversity or higher field diversity by growing multiple varieties in a field (Weedon and Finckh 2019; Wolfe and Ceccarelli 2020).

One way for breeding programs to address the heterogeneity of target environments and emerging markets is to work directly with farmers, bakers, and other expert professionals. Participatory plant breeding, where farmers and other stakeholders such as bakers, millers, seed producers, and consumers, are involved in the breeding process, can take many shapes and may have different levels of involvement (Lammerts van Bueren and Myers 2012). One of the aspects that participatory plant breeding can address is the gap between performance at the experimental stations and in farmers' fields (Simmonds 1980, 1991; Annicchiarico 2002a). There are multiple reasons why this gap occurs, but a primary cause is that in some cases selection environments are not representative of production environments (Ceccarelli and Grando 2007).

Involving farmers and growers in the decision-making process for the selection and advancement of candidate lines would increase the adoption of new variety releases (Annicchiarico 2002a), provide helpful information for breeding decisions (Ashby 2009), have industry-ready varieties for release, and provide a more diverse set of environmental evaluations (Ceccarelli and Grando 2007). Therefore, participatory plant breeding could include other stakeholders involved in the wheat production from the farm to the table such as millers, bakers, and consumers in order to increase the success of varieties (Dawson et al. 2011). Participatory bread wheat breeding programs for organic systems and artisanal baking exist in Europe (France and Italy), Canada, and in Washington State (U.S.). Many of these programs share methods and follow similar selection protocols. This includes using organic trials and farmer input to select parents for crosses, decentralizing selection and testing starting as early as possible, often in the F_3 or F_4 generation, and involving farmers and bakers in the evaluation of breeding lines on research stations, farms, and bakeries. These programs were reviewed by (Colley et al. 2021).

The objective of this paper was to illustrate the development of a small breeding program for bread wheat varieties targeted to organic systems and artisan baking in the U.S. Upper Midwest. We have involved farmers, millers, and bakers in the decision process for advancing breeding lines. Specifically, our goals include: 1) description of the breeding process from initial crosses to end-product, 2) analysis of the genotype-by-environment interaction (GEI) for agronomic and quality performance, including the study of stability indicators for transient year-to-year GEI, 3) a comparison of on-station and on-farm performance for advanced breeding lines, and the incorporation of on-farm evaluation and farmers input in the decision process, and 4) development of baking trials with bakers input to evaluate artisan whole grain baking quality in advanced breeding lines.

Materials and methods

Population development

Sixteen parents of hard red winter wheat tested in organic systems in the U.S. and Europe were selected based on their bread quality attributes, adaptation to organic conditions, and resistance to FHB and other diseases and were used as progenitors of the breeding program (see Figure 1 for a description of the breeding program). Eight parents were crossed in a partial diallel in 2012 in a greenhouse at Cornell University (Ithaca, NY). A second crossing block was conducted in 2013 to add eight additional parents to the partial diallel based on results from multi-location organic winter wheat trials. Between one and 20 successful F_1 progeny per cross were self-pollinated. F_2 plants were grown in a high tunnel from October 2013 to June 2014 and up to 30 individual plants per original cross were selected. A total of 300 $F_{2:3}$ derived families were evaluated with 12 parents used as checks from October 2014 to July 2015 in 1m head-rows in two trials: an agronomic trial at the Cornell University Homer C. Thompson Vegetable Research Farm in Freeville, New York (42.5˚ N, 76.3˚ W) under certified organic conditions and an FHB nursery at the Caldwell Research Farm in Ithaca New York (42.5˚ N, 76.3˚ W). Head-rows in the certified organic location were evaluated for lodging, winter survival, grain yield, and test weight. In the FHB nursery, the lines were inoculated with FHB on June 2, 5, 9, 11 and 16, 2015 (Zadoks 61, 65, and 69). Head-rows were scored for FHB incidence and severity at the soft dough stage (Zadock 83), about 24 days after flowering following (Fulcher et al. 2021). Briefly, twenty spikes in each head-row were scored from 1-5 for FHB where 1 = one spikelet infected, 2 = two spikelets infected, 3 = up to half of the spikelets in the spike infected, 4 = more than half of the spikelets in the spike infected, 5 = the entire spike is dead. The incidence of FHB was calculated as the percent of non-zero scores among the 20 spikes scored, while severity was calculated as the mean FHB of the non-zero scores. From 115 $F_{2:3}$ families that scored a FHB index less than 15, ten individual plants were selected at Freeville, NY. and eight seeds from each plant were evaluated for protein content using a near infrared spectroscopy single-seed analyzer following (Carlson et al. 2019). $F_{2:3}$ families with mean protein content above the 50% percentile were selected. These 465 $F_{3:4}$ families were evaluated from October 2015 to July 2016 in 1m head-rows in two trials: an agronomic trial at the Freeville location, using three check lines replicated 10 times throughout the nursery to assess spatial variability, and where families were evaluated for lodging, winter survival, grain yield, and test weight. Another trial was conducted at the FHB nursery in Caldwell, but due to seed limitations for some families, only 406 $F_{3:4}$ families were grown in the FHB nursery. Two check lines were replicated 10 times throughout the nursery to assess spatial variability. $F_{3:4}$ families were selected based on an index incorporating grain yield, leaf disease, and FHB index and 98 families were selected. Eight seeds from each selected family were evaluated for protein content using a near infrared spectroscopy single-seed analyzer following Carson et al. (2019) and the 50 families with the highest mean protein content were advanced to F_5.

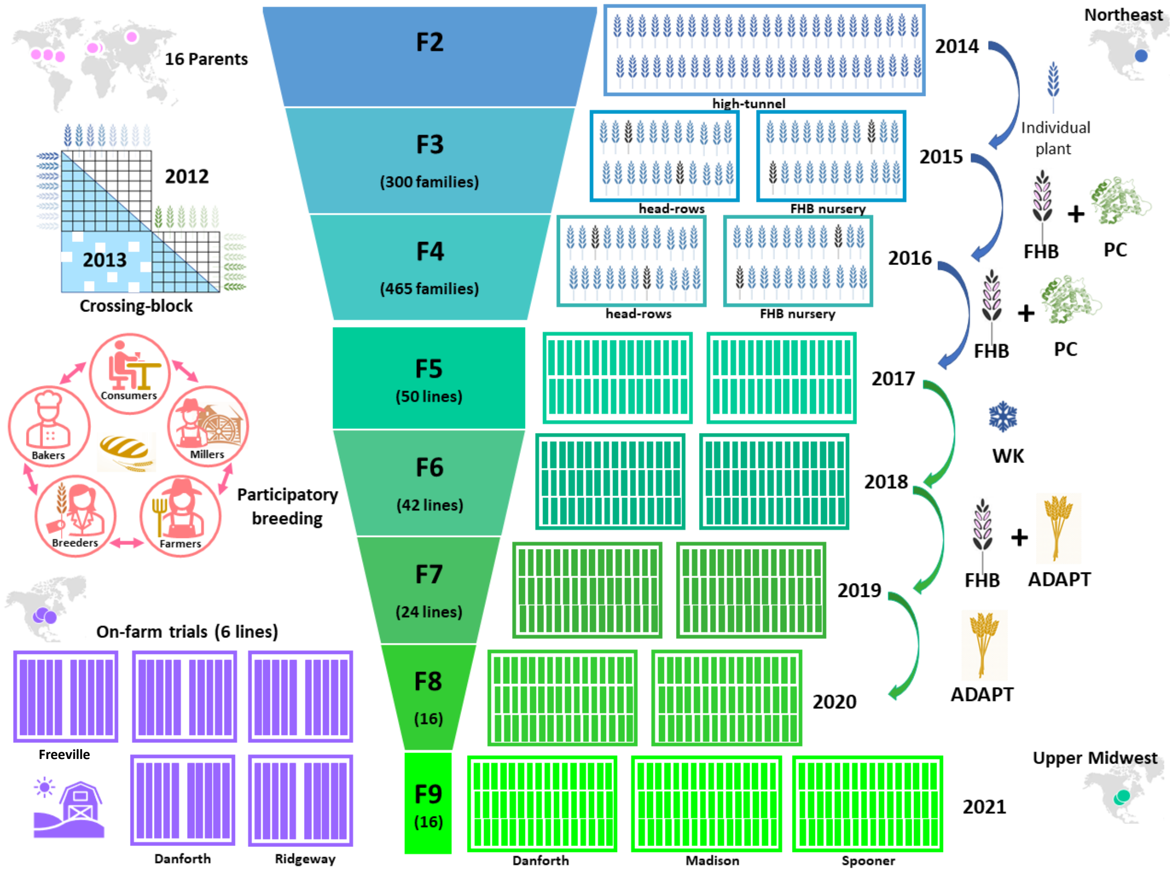


Figure 2. 1. Characterization of the organic hard red winter wheat breeding program. Sixteen parents were used in a partial diallel crossing block to generate the initial breeding population and families were selected initially based on Fusarium head blight and protein content in the Northeast. At the F5 generation, lines were evaluated in the U.S. Upper Midwest in three on-station locations (Madison, WI; Spooner, WI; and Danforth, IL) in five years (2017-2021). Advanced lines were evaluated in three on-farm locations (Freeville, NY; Danforth, IL; and Ridgeway, WI). Participatory breeding strategies were used to include farmers, millers, bakers, and consumers in the decision to advance and release breeding lines.

On-station phenotypic evaluation

The 50 selected advanced inbred lines (i.e. F_5 families derived from $F_{3:4}$ families) and three organic artisan milling and baking industry checks (*Red Fife*, *Arapahoe*, and *Warthog*) were evaluated between 2017 and 2021 at the West Madison Agricultural Research Station in Verona, Wisconsin (Madison, 43°04'22.6"N 89°32'46.8"W), and at Janie's Farm in Danforth, Illinois (Danforth, 40°50'30.2"N 87°58'10.1"W). A third location was used during the 2021 year at Spooner Agricultural Research Station in Spooner, Wisconsin (Spooner, 45°49'21.0"N 91°52'35.3"W). Trials were planted as close as possible at ideal planting dates for each location ranging from September 25th to October 22nd. Planting date is restricted by aphid feeding and their potential to transmit barley yellow dwarf virus until late September, and by the growing degrees days remaining to support seedling establishment and growth required for winter survival in early October (Conley et al. 2015). All trials were managed following organic practices on certified organic land, except for the Spooner location, which has been managed according to organic standards since 2018 but is not certified. *Red Fife* is an historic facultative variety originally selected in Ontario, Canada, that is popular for baking because of its flavor and its inclusion in the Slow Food Ark of Taste (Ark of Taste 2014). *Red Fife* was a standard for the Canadian baking and milling industries in the mid-late 1900s. *Arapahoe* is a modern hard red winter wheat variety with very good winter hardiness and baking quality released by the University of Nebraska in 1989 (Baenziger et al. 1989). *Arapahoe* has high protein levels and test weight and is resistant to stem rust and moderately resistant to leaf rust. *Warthog* is a modern hard red winter wheat variety released in Canada in 2001 (Registration #5390, Thompsons Limited). *Warthog* has been widely grown by farmers in the U.S. Upper Midwest and has acceptable baking quality.

Experiments at Madison followed a four-year rotation: 3-years of alfalfa followed by one year of cereals. Management of the experiments included mechanical weeding using mowers in alleys and a cultivator between plots (2019 and 2021), hand weeding inside plots (intense weeding in 2017 and 2018, but minimal weeding in 2019 and 2021), hand rogueing off-types to maintain genetic purity and to avoid cross-contamination (i.e. one or two plants per plot), and frost-seeded red clover planted in the winter to maintain soil cover. The Madison 2017 experiment was a randomized complete block design (RCBD) with two replications. The Madison 2018, 2019, and 2020 experiments were augmented alpha designs with three complete replications. The Madison 2021 experiment was a RCBD with three replications. The Madison 2017 experiment was evaluated in 1.2 x 1.85 m (2.22 m^2) plots while the 2018-2021 experiments were evaluated in 1.52 x 3.20 m (4.86 m^2) plots. Planting was on 3 October 2016, 28 September 2017, 28 September 2018, 25 September 2019, and 25 September 2020. Planting density was 134 kg ha^{-1} for the 2017 experiments and 157 kg ha^{-1} for the 2018-2021 experiments. Clover was frost seeded on 3 March 2017, 16 February 2018, 3 February 2019, 16 February 2020, and 26 February 2021. Harvest dates were 17 July 2017, 18 July 2018, 18 July 2019, 17 July 2020, and 10 July 2021. All experiments were planted with an inter-column spacing of 30 cm and an inter-row spacing of 2.4 m.The 2021 experiment at Spooner was a RCBD with three replications evaluated in 1.52 x 3.65 m (5.54 m^2) plots. Planting was on 29 September 2020 at a planting density of 157 kg ha^{-1}, while harvest date was 21 July 2021. Experiments at Danforth consisted of a RCBD with two (2017) or three (2018-2021) replications evaluated in 1.4 x 2.4 m (3.36 m^2) plots with a planting density of 117 kg ha^{-1}, an inter-column spacing of 23.3 cm and an inter-row spacing of 30 cm. Planting was on 22 October 2016, 3 October 2017, 22 September 2018, 3 October 2019, and 9 September 2020 for the 2017-2021 growing seasons. Harvest date was 18 July 2017, 26 June 2018, 11 July 2019, July 2020, and 8 July 2021.

All genotypes were evaluated for agronomic traits including winter survival, plant height, lodging, heading date and grain yield; disease traits including FHB, leaf rust, barley yellow dwarf virus, and powdery mildew; and for grain quality traits including test weight, falling number, and deoxynivalenol (DON), ash, and protein content. Winter survival was estimated as the percent of plants that were still green after the snow melts in the spring in early April. Plant height was measured during the grain filling stage (Zadok's 73 to 83) as the height (in cm) of an average plant in the plot from the ground to the tip of the spike not including awns. Lodging was estimated during the grain filling stage (Zadok's 73 to 83) as the percent of plants in the plot that were falling over. Heading date was recorded on the date when 50% of the tillers in a plot have spikes completely emerged from the flag leaf sheath (Zadok's 60). Grain yield was evaluated as the weight of each plot harvested at maturity (Zadok 92) and expressed in kg ha^{-1} obtained by harvesting and threshing with a self-propelled Wintersteiger Masterplot experimental plot combine harvester, drying the grain, and cleaning the grain with a Pfeuffer Sample MLN grain cleaner. Grain yield was corrected to a kernel moisture level of 12%. FHB was evaluated on a 1-9 scale using a combined score between incidence and severity after flowering (Zadok's 83) using natural infection. Leaf rust and powdery mildew were scored on a 1-9 scale similar to FHB based on natural infection in years where the disease was detected. The incidence of barley yellow dwarf virus was evaluated based on natural infection as the percentage of plants affected by the disease in years where the disease was present. All agronomic and disease traits were recorded in Madison. Grain yield, plant height, and heading date were recorded in Danforth. Grain yield, plant height, and heading date were recorded in Spooner.

Test weight was measured as the weight of 500 mL of dry and clean grains using a Cox funnel and following the Canadian Grain Commission official grain grading guide (Candadian Grain Comission). A subsample of 500 grams of grain from each plot from the 2019, 2020, and 2021 experiments were sent to the Integrated Bioprocessing Research Laboratory of the College of Agricultural, Consumer and Environmental Sciences at the University of Illinois at Urbana-

Champaign for quality processing. Grain samples were analyzed for grain protein content, falling number, and ash content. Protein and ash content were evaluated with a Perten Inframatic 9500 NIR (Near Infra-Red) grain analyzer and expressed in percent values. Falling number was evaluated with a Perten Falling Number® system that measures alpha-amylase enzyme activity and structural integrity of the starch and is expressed in seconds. Grain quality traits were evaluated on samples from Madison and Danforth.

Statistical analysis of on-station experiments

On-station genotypic means. Empirical best linear unbiased estimates (BLUEs) of each agronomic and quality trait for each genotype in each location and year (i.e. environment) were estimated using plot-level information from all genotypes planted at a given location and year (i.e. environment) and corrected for experimental design and spatial variation with the following linear mixed model:

[1] $$\underline{y_{ijkl}} = \mu + G_i + \underline{\beta_j} + \underline{R_k} + \underline{C_l} + \underline{\varepsilon_{ijkl}}$$

where y_{ijkl} is the plot level observation, μ is the overall mean, G_i is the effect of the ith genotype, β_j is the effect of the jth block with $\beta_j \sim N(0,\sigma^2_\beta)$, R_k is the effect of the kth row with $R_k \sim N(0,\sigma^2_R)$, C_l is the effect of the lth column with $C_l \sim N(0,\sigma^2_C)$, and ε_{ijkl} is the residual term with $\varepsilon_{ijkl} \sim N(0,\sigma^2_\varepsilon)$, with the covariance among random effects equal to zero and σ^2_β, σ^2_R, σ^2_C, and σ^2_ε, being the variance components of blocks, row, column, and residual error respectively. Row and column effects were used as a post-blocking control of spatial variation and were only considered when their inclusion affected the model (see supplementary files 1-5). An alpha design was used in some trials with incomplete blocks following rows, and therefore, the same model was used for those trials. This analysis was performed in R statistical software (R. Core Team 2013) fitting the *lmer* function of the *lme4* package (Bates et al. 2007). Genotypic BLUEs were obtained using the *emmeans* function of the *emmeans* package (Lenth 2021) in R statistical software This model was used for all agronomic and quality traits in each environment (i.e. combination of location and year).

Variance components estimation. The following random effects model was used to estimate variance components for genotype, location, year, their interactions (i.e. location by year, genotype by year, genotype by location, and genotype by location by year) and the residual error for grain yield, protein content, falling number, and ash content:

$$[2] \qquad \underline{y_{ijkl}} = \mu + \underline{G_i} + \underline{L_j} + \underline{A_k} + \underline{LA_{jk}} + \underline{\beta_{l(jk)}} + \underline{GL_{ij}} + \underline{GA_{ik}} + \underline{GLA_{ijk}} + \underline{\varepsilon_{ijkl}}$$

where y_{ijkl} is the plot level observation, μ is the overall mean, G_i is the effect of the ith genotype with $G_i \sim N(0,\sigma^2_G)$, L_j is the effect of the jth location with $L_j \sim N(0,\sigma^2_L)$, A_k is the effect of the kth year with $A_k \sim N(0,\sigma^2_A)$, LA_{jk} is the location by year interaction with $LA_{jk} \sim N(0,\sigma^2_{LA})$, GL_{ij} is the genotype by location interaction with $GL_{ij} \sim N(0,\sigma^2_{GL})$, GA_{ik} is the genotype by year interaction with $GA_{ik} \sim N(0,\sigma^2_{GA})$, GLA_{ijk} is the genotype by location by year interaction with $GLA_{ijk} \sim N(0,\sigma^2_{GLA})$, and ε_{ijkl} is the residual term with $\varepsilon_{ijkl} \sim N(0,\sigma^2_\varepsilon)$, with the covariance among random effects equal to zero and σ^2_G, σ^2_L, σ^2_A, σ^2_{LA}, σ^2_β, σ^2_{GL}, σ^2_{GA}, σ^2_{GLA}, and σ^2_ε, being the genotypic, location, year, location by year, block, genotype by location, genotype by year, genotype by location by year, and residual error variance components. The variance components were then expressed as a proportion of the total genetic variance ($\sigma^2_G + \sigma^2_{GL} + \sigma^2_{GA} + \sigma^2_{GLA}$).

Heritability. Heritability for each trait in each environment was calculated *ad hoc* following Piepho (2019) based on Holland et al. (2003) using harmonic means:

$$[3] \qquad H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ga}^2}{\bar{n}_a} + \frac{\sigma_{gl}^2}{\bar{n}_l} + \frac{\sigma_{gal}^2}{\bar{n}_{al}} + \frac{\sigma_e^2}{\bar{n}_{alr}}}$$

where H^2 is the estimate of the heritability, $\sigma_g^2, \sigma_{ga}^2, \sigma_{gl}^2, \sigma_{gal}^2$, σ_e^2 are the variance component estimates of genotype, genotype by year, genotype by location, genotype by year by location, and the residual error respectively from model [2], and $\bar{n}_a$, $\bar{n}_l$, $\bar{n}_{al}$, $\bar{n}_{alr}$ the harmonic means of the number of years, locations, location-year, and location-year-replications. Variance components and heritabilities were also estimated from models [2] and [3] for each location removing the location components, and for each environment removing the location-year components.

GGE biplots. BLUEs from model [1] were used to graphically represent the genotype and genotype by environment effects through a biplot (GGE biplot, Wickham 2016) using the package *gge* (Laffont et al. 2013) from the R statistical program (R. Core Team 2013).

Finlay-Wilkinson Regression. Stability for grain yield was estimated with the Finlay and Wilkinson (1963) analysis (hereafter FW) using the following model:

[4] $$y_{ij} = \mu + G_i + E_j + \beta_i E_j + \varepsilon_{ij}$$

where y_{ij} is the BLUE of the ith genotype in the jth environment (i.e. combination of location and year) estimated from equation [1], μ is the overall mean, G_i is the effect of the ith genotype, E_j is the effect of the jth environment, β_i is the FW regression coefficient of the ith genotypic performance over the environmental mean, also called sensitivity, and ε_{ij} is the residual genotype by environment interaction not explained by the FW regression model, where $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. Genotypes with sensitivity values closer to zero have static stability, while genotypes with sensitivity close to one have dynamic stability, and genotypes with higher sensitivity have low stability. The R^2 coefficient representing the lack of fit of the FW regression for each genotype was estimated with the same model. The FW regression models were run on R statistical program with basic regression functions (R. Core Team 2013).

Wricke's ecovalence stability coefficient. Wricke's ecovalence stability was estimated for grain quality, protein content, falling number, and ash content using the following model (Wricke, 1962):

$$W_i = \sum_j (\hat{G}_{ij} - \bar{G}_i - \bar{E}_j - \hat{\mu})^2 \tag{5}$$

where W_i is Wricke's ecovalence stability coefficient $\hat{G}_{ij}$ is the BLUE of the ith genotype in the jth environment (combination of location and year) estimated in equation [1], $\bar{G}_i$is the genotypic mean of the ith genotype across environments, $\bar{E}_j$ is the environmental mean of the jth environment across all genotypes, and $\hat{\mu}$ it is the overall mean. Wricke's ecovalence stability coefficient represents the volatility of genotypes to changes in the environment; a genotype with no genotype by environment interaction will show a W value of zero. Wricke's ecovalence was estimated using the *StatGxE* package (van Rossum et al. 2021) in R statistical program (R. Core Team 2013).

Selection of advanced breeding lines

Advanced breeding lines were selected for grain yield, test weight, FHB and other, diseases, winter survival, and protein content. The selection was applied over the population reducing the number of genotypes each year from 50 to 42 in 2018, to 24 in 2019, and finally, to 16 in 2020 (Figure 2). The Red Fife check was also discarded after the first year of evaluation because it did not survive the first winter and was not included in the multi-year analysis. The BLUEs for all traits, field notes, and input from farmers and bakers, described below, were used in the process of selection conducted each year based on a the data from current and previous years. Winter survival was the trait with the highest selection pressure during the years 2017 and 2018. Lines that did not survive or showed more than 70% winterkill were discontinued. High selection intensity for absence of FHB symptoms was applied in all years with high natural infection. Other traits such as lodging, and disease prevalence were criteria used to discard additional lines each year (Figure 2; supplementary files 1-5). Finally, pedigree information was considered in the selection process to avoid shrinking the genetic diversity early in the process. Among and within original F_1 cross selection was performed through the years, but at least one genotype per cross was always advanced unless it consistently underperformed.

Participatory on-farm trials

On-farm experiments. Six breeding lines and one commercial check were evaluated in on-farm trials in three farms in three states in the U.S.: Harold Wilken, Janie's Farm, Danforth, IL (40°50'30.2"N 87°58'10.1"W), John and Halee Wepking and Paul Bickford, Meadowlark Organics, Ridgeway, WI (42°59'25.2"N 90°00'43.0"W), and Freeville, New York (Freeville, 42°31'06.7"N 76°20'04.6"W). The two farms in the Upper Midwest have been leaders in growing organic food grade grains, and both have commercial mills as part of their operations to supply regional bakeries and consumers. The following lines were evaluated: *47.04* (*Maxine* X *Gua*), *140.05* (*Warthog* X *Gua*), *174.01* (*Clarks Cream* X *Arapahoe*), *202.07* (*Rouge de Bordeaux* X *Arapahoe*), *212.08* (*Red Fife* X *Arapahoe*), *260.06* (*Rouge de Bordeaux* X *Warthog*), and *Warthog* was used as the commercial check on all farms. Lines to be evaluated at the farm were selected in a participatory manner after field-days and meetings with farmers, bakers, millers, and other stakeholders using the on-station data based on winter survival, grain yield, agronomic performance, resistance to FHB and other diseases evaluated on-station. On-farm experiments were planted in a RCBD with two replications in each location with a plot size of 18.5m^2 and 97.4

m^2 for 2020 and 2021. Planting was on 26 September 2019, and 11 October 2020 at Ridgeway, 3 October 2019 at Danforth, and 21 October 2019 at Freeville. Harvest date was 16 July 2020 and 17 July 2021 at Ridgeway, on 7 July 2020 at Danforth, and 22 July 2020 at Freeville. Experiments were managed by the participating farmers according to their standard practices for certified organic winter wheat at each farm, except that the Freeville location was managed by the Cornell University Small Grains Team. The Freeville trial was still included as part of the on-farm trials due to the nature of trials including plot size and number of breeding lines evaluated. Traits evaluated were grain yield and DON content. Grain yield was harvested with an experimental plot combine (Zurn 150 plot harvester) at Ridgeway and a Winterstiger plot master at Danforth. A bulked subsample of 100 g of grain from both replications was sent to the Wisconsin Crop Improvement association for DON level testing.

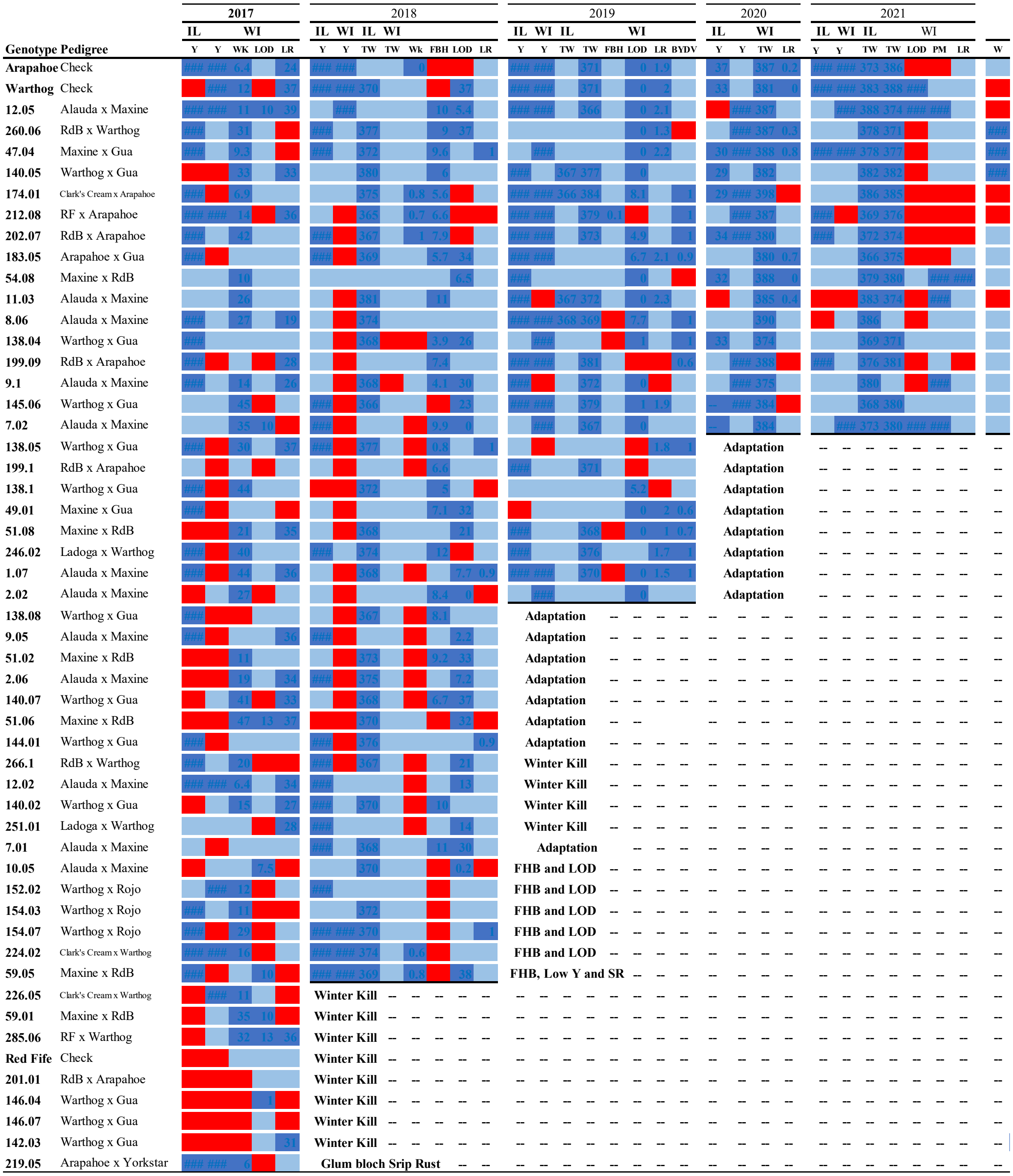


Figure 2.2 Characterization of the advanced breeding line selection process in the hard red winter wheat breeding program for artisan baked whole grain. The heatmap is a qualitative characterization of trait performance categories with blue being good performer, light blue an average, and red a bad performer for each trait in each location (IL, Danforth, Illinois; WI,

Madison, Wisconsin) and year (2017-2021). Grain yield (Y) and Wricke's ecovalence stability coefficient (W) relative performance was color-coded in relation to the overall mean performance across all environments. Winter survival (WK), lodging (LOD), leaf rust severity (LR), test weight (TW), fusarium head blight severity (FHB), powdery mildew (PM), and barley yellow dwarf virus severity (BYDV) were color-coded in absolute values. The main driver for discontinuation of early lines was indicated at the right of the heatmap of their last evaluation in the program.

On-farm genotypic statistical analysis. Empirical best linear unbiased estimates (BLUEs) for grain yield of each genotype for each location and year (i.e. environment) were estimated using plot-level information from all genotypes planted at a given environment and correcting for experimental design following the linear mixed model described in [1] but without using spatial corrections. A Dunnet test (α=0.05) was used to compare the performance of the experimental lines to that were statistically different from the *Warthog* check.

Participatory baking trials and sensory evaluations

Participatory baking trials and sensory evaluations were conducted in 2020 and 2021 to evaluate the performance of lines for artisan baking. Grain was obtained from large seed increase plots grown on-station in the experiments described above. Grain from Madison was used in 2020, while a proportional mix of grain from Madison and Freeville was used in 2021 (70% NY and 30% WI). Lines were selected based on on-station trial performance, DON levels, and grain availability. All grain was tested for DON levels prior to milling to ensure food safety. Grain was also tested for protein content and falling number to better interpret the results of the baking trials. However, the values of protein content and falling number were not provided to bakers prior to the baking tests to avoid biasing expectations. We used the process described in Kucek et al (2017) to evaluate the breeding lines and checks in naturally leavened (sourdough) bread. Briefly, this involves creating a levain out of the flour of each genotype and then scoring the mixing, proofing, and shaping stages of dough development, followed by scores for the exterior, interior and flavor of the final bread. Scores are given on a 10 point scale from poor to optimal, with notes to indicate why a variety may not have scored optimally since there can be multiple reasons for non-optimal performance.

2020 Baking trial. Four breeding lines *140.05* (*Warthog* X *Gua*), *174.01* (*Clarks Cream* X *Arapahoe*), *212.08* (*Red Fife* x *Arapahoe*), *260.06* (*Rouge de Bordeaux* X *Warthog*), and *Warthog* as a commercial check were milled by Madison Sourdough Bakery on an Osstiroller 700 MSM Combi Mill with 28-inch stones. The flour was slightly sifted to achieve approximately 95% extraction for each genotype. The sifting process removes bran flakes above a certain particle size, and this was kept constant across all genotypes. The breeding lines and check were evaluated in a baking test conducted on 11 and 12 March 2020. Six professional artisan bakers traveled to Madison for the evaluation and participated over the two days. The bakers who participated in the evaluation included Melina Kelson of Bootleg Batard in Chicago, IL, Solveig Tofte of Sun Street Breads in Minneapolis, MN, Greg Wade of Publican Quality Bread in Chicago, Kirk Smock of ORIGIN Breads in Madison, WI, Matt Kronschnabel of Bard Bread in Viroqua, WI, and Andrew Hutchison of Madison Sourdough. The six bakers collectively worked on each variety during the baking process and adjustments were made by consensus for each variety to optimize the level of hydration, rest time and mix time without changing the base formula (other than hydration). Each baker then individually assigned scores to each variety during each step of the process.

2021 Baking trial. Two breeding lines *47.04* (*Maxine* x *Gua*), *260.06* (*Rouge de Bordeaux* X *Warthog*), and *Warthog* as commercial check were milled and evaluated in the remote baking tests conducted during the week of 5 May 2021. Remote tests were done in 2021 due to COVID-19 restrictions. The grain was milled by Meadowlark Organics in Ridgeway, Wisconsin, on a Meadows 30” stone mill. The flour was slightly sifted to achieve approximately 85% extraction to follow the commercial standard protocol for Meadowlark Organic. This was determined to be the best point of comparison by the participating bakers as they are familiar with working with Meadowlark’s flour. The baking evaluation also included Meadowlark’s current commercial bread flour (12.4% protein blend of *Warthog* winter wheat and a spring wheat variety) as a second commercial flour check. Flour was then shipped to each baker and were evaluated in a baking test done remotely by five bakers in their own bakeries, all of whom participated in the 2020 bake test. The same protocol was used from the 2020 trials but done individually by each baker. Data collection was reduced to one score for each of the major phases of the baking process rather than each individual step, with additional room for bakers to write descriptive evaluations of each step in lieu of more detailed scoring.

Sensory evaluation. Bread from the baking evaluation was used the following day in a hedonic sensory evaluation by members of the public and the research team in 2020 and 2021. Each bread was rated for appearance, texture, and flavor on a 1 to 5 scale, with 5 being the most preferred. The sensory evaluation in 2020 was done at Madison Sourdough on March 11 with 55 participants. The sensory evaluation in 2021 was done remotely by 40 participants with kits assembled including all samples from Madison Sourdough and Origin Breads in Madison, WI, and then tasted by individuals off-site due to COVID-19 restrictions.

Statistical Analysis. Empirical best linear unbiased estimates (BLUEs) of each baking and sensory trait for each genotype for each year (i.e. environment) were estimated using baker’s and taster’s scores information from all genotypes tested at a given year (i.e. environment) with the following linear mixed model:

[6] $$\underline{y_{ij}} = \mu + G_i + \underline{\beta_j} + \underline{\varepsilon_{ij}}$$

where y_{ij} is the score, μ is the overall mean, β_j is either the effect of the jth baker or taster with $\beta_j \sim N(0,\sigma^2{}_\beta)$, and ε_{ijkl} is the residual term with $\varepsilon_{ijkl} \sim N(0,\sigma^2{}_\varepsilon)$, with the covariance among random effects equal to zero and $\sigma^2{}_\beta$, and $\sigma^2{}_\varepsilon$ the evaluator and residual error variances respectively. When dough, bread and bake summaries were analyzed, the term Υ_k was included as the effect of the kth baking stage with $\Upsilon_k \sim N(0,\sigma^2{}_\Upsilon)$.The ε_{ijk} is the residual term with $\varepsilon_{ijk} \sim N(0,\sigma^2{}_\varepsilon)$, with the covariance among random effects equal to zero and $\sigma^2{}_\beta$, and $\sigma^2{}_\Upsilon$, and $\sigma^2{}_\varepsilon$ the evaluator, baking stage, and residual error variances, respectively.

Results

Genotype by environment interaction (GEI) and variance components. Approximately 54% of the total genetic variance (i.e. G+GL+GA+GLA) of grain yield was explained by the genotypic main effects (G) while the remaining 46% was explained by GEI (GL+GA+GLA), with genotype by location by year (GLA=25%) being the largest effect (Figure 3). Protein content also had a strong genotypic main effect, with 52% of the genetic variance component explained by genotypic main effects. On the other hand, GEI explained a large proportion of the total genetic variance of falling number and ash content (78% for falling number, and 66% for ash content; Figure 3). Finally, although three mega-environments were identified in the GGE biplots (supplementary file 6), no-repeatable pattern of GEI could be identified as drivers of those groups with mega-environments grouping environments across years and locations.

Finlay-Wilkinson Stability. Only small differences in FW slopes for grain yield were observed among genotypes, with regression slopes between 0.79 (*212.08*) and 1.20 (*Warthog*) for all the environments (Table 1, Figure 4). This indicates that similar dynamic stability exists among genotypes. On the other hand, R^2 values were higher for most of the breeding lines than for the commercial checks, with values ranging from 0.91 (*212.08*) to 0.99 (*47.04*), and both checks, *Arapahoe* and *Warthog*, having values of 0.92 indicating higher predictability from the FW model for breeding lines than commercial checks (Table 1). Finally, because the years 2020 and 2021 in Madison were extremely high yielding environments and regressions can be affected by extreme values, a regression model excluding the years 2020 and 2021 for Madison was evaluated. The variance in the regression coefficients and R^2 values were larger after excluding the extreme values. Some genotypes, including the checks, had lower R^2 values in the remaining environments (0.19 and 0.30 for *Arapahoe* and *Warthog* respectively).

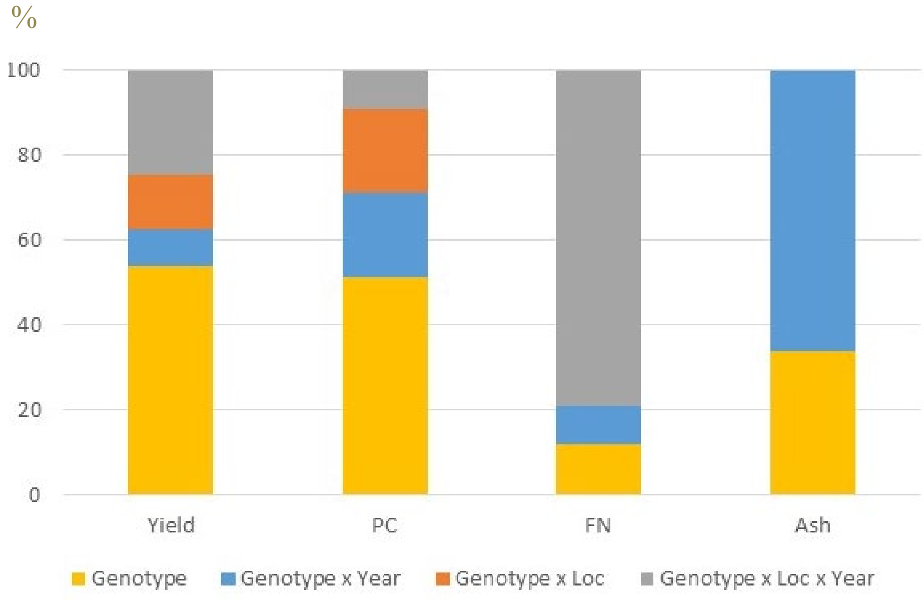


Figure 2.3. Relative proportion of all genetic variance components (G, GA, GL, GAL) for grain yield (Yield, 2017-2021), protein content (PC, 2019-2021), falling number (FN, 2019-2021) and ash content (Ash, 2020-2021). All the traits were evaluated in two locations (Madison and Danforth).

Table 2.1. Grain yield stability parameters of hard red winter wheat breeding lines and commercial checks evaluated in ten environments in two locations (Danforth and Madison) and five years (2017-2021). Finlay-Wilkinson regression coefficients (b1, values closer to zero indicate static stability) and non-linear stability (R^2 of FW, values near to 1 show more predictable performance), as well as Wricke's ecovalence stability coefficient (W, lower values have for more dynamically stable performance) are shown. Because the years 2020 and 2021 were extremely high-yielding years in Madison (see supplemental file 1), values are also shown for the stability analysis without the years 2020 and 2021 in Madison (stability no extremes). Mean grain yield performance is also shown for comparison purposes.

	Overall Stability				Stability (no extremes)		
Genotype	**Mean**	**β_1**	**R^2**	**W (10^{+6})**	**β_1**	**R^2**	**W (10^{+6})**
7.02	2766	1.06	0.98	0.86	1.00	0.85	0.40
8.06	2587	0.97	0.98	0.30	0.97	0.91	0.25
9.1	2750	1.07	0.98	0.98	0.76	0.55	0.76
11.03	2519	0.89	0.97	1.24	0.71	0.66	0.81
12.05	3082	1.11	0.95	1.56	0.93	0.41	0.95
47.04	2997	1.08	0.99	0.39	1.00	0.73	0.20
54.08	2655	0.95	0.98	0.71	0.58	0.37	0.63
138.04	2604	0.99	0.93	1.06	0.91	0.44	0.76
140.05	2689	1.00	0.96	0.97	1.00	0.43	0.66
145.06	2642	1.03	0.99	0.34	1.04	0.99	0.11
174.01	2769	0.97	0.96	0.93	1.30	0.68	0.47
183.05	2698	0.93	0.95	0.87	1.20	0.75	0.68
199.09	2750	0.93	0.96	1.27	0.93	0.80	0.93
202.07	2836	0.96	0.95	0.73	0.92	0.93	0.30
212.08	2639	0.79	0.91	2.17	0.41	0.34	0.68
260.06	2863	1.01	0.99	0.26	0.94	0.77	0.16
Arapahoe	3206	1.07	0.92	1.87	0.97	0.19	0.94
Warthog	3234	1.20	0.92	3.10	1.05	0.30	1.05

Wricke's ecovalence stability coefficient. There were orders of magnitude of differences in Wricke's ecovalence stability coefficients (W) for grain yield among genotypes, with W values between 0.26 x 10^6 (*260.06*) and 3.10 x 10^6 (*Warthog*) (Table 1). A similar rank-order of genotypes was found when the extremely high yielding environments were removed from the analysis (Table 1). The performance of genotypes with the highest and lowest W coefficients are shown in Figure 4 for all the environments (excluding the extremes) to illustrate predictability across environments.

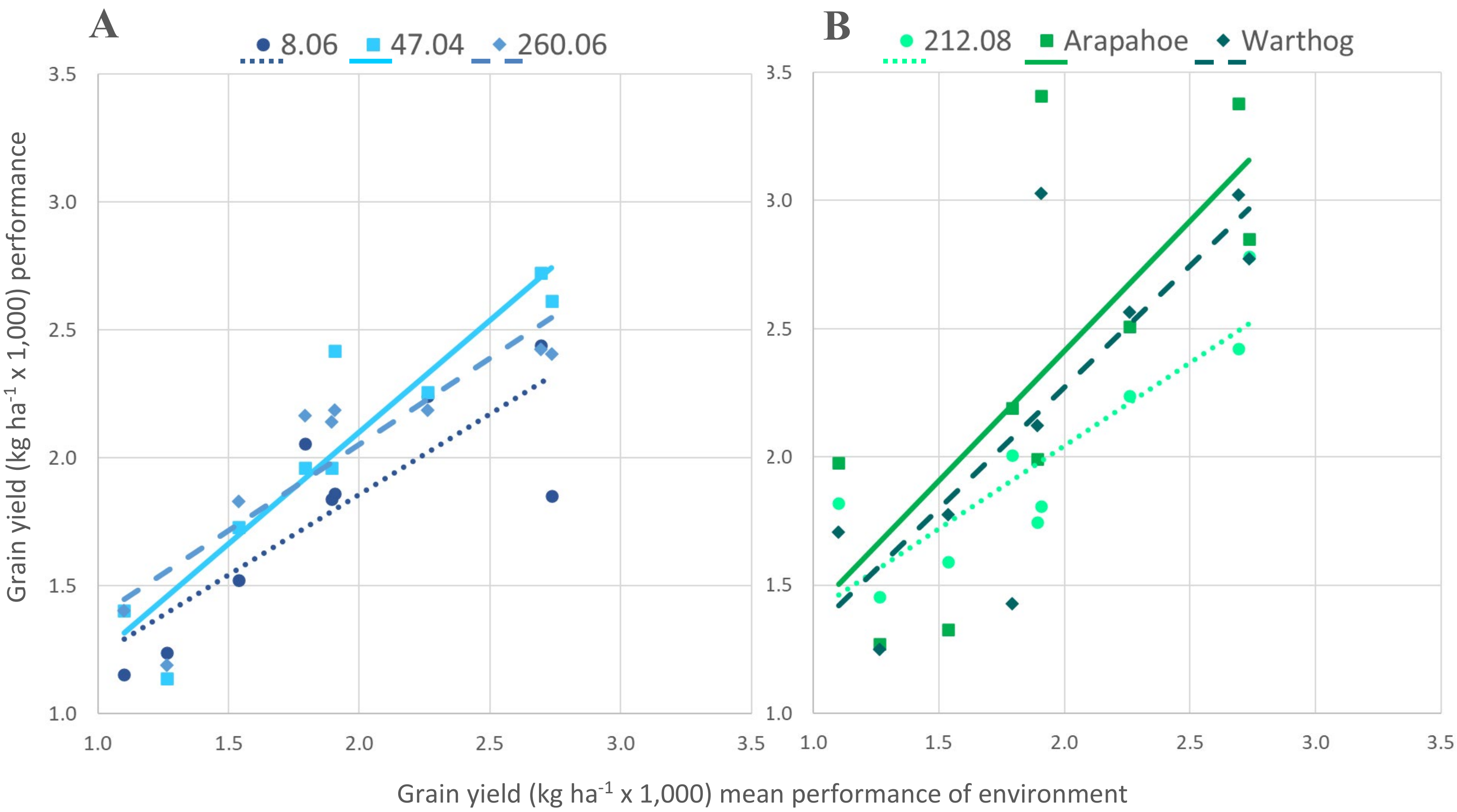


Figure 2.4. Finlay - Wilkinson regression for grain yield of hard red winter wheat breeding lines and commercial checks evaluated in seven environments in three locations (Madison, Danforth and Spooner) and five years (2017-2021). Only genotypes that showed the A) lowest (i.e. more stable, *8.06*, *47.04*, and 260.06) and B) highest (i.e. less stable, *212.08, Arapahoe*, and *Warthog*) values for the Wricke's ecovalence stability coefficient (from Table 1, no extremes) are shown.

Mean protein content across years and locations was 10.7, in general, higher for the breeding lines than for the commercial checks *Arapahoe* (10.5%) and *Warthog* (10.1%). Some breeding lines *260.06* (10.1%) and *47.05* (10.4%) were similar to the *Arapahoe* and *Warthog*. Breeding lines *11.03* (11.5%), *140.05* (11.9%) had significantly higher protein than *Arapahoe* and *Warthog*. Breeding lines *11.03*, and *140.05* were the least stable genotypes for protein content with W values of 1.92 and 2.23 respectively (Table 2). On the other hand, some of the breeding lines such as *8.06* have high protein (11.0%) contents with low values of W (W=0.58). Falling number was between 300 and 400 seconds for all the breeding lines with a range of values for W (Table 2, supplementary file 4). Both *Arapahoe* and *Warthog* were among the most volatile genotypes for falling number, with W values of 12,431 and 13,183 respectively. Ash content was similar between breeding lines ranging between 1.64% and 1.88% and commercial checks *Arapahoe* (1.79%) and *Warthog* (1.90%) with small overall variance (Table 2, supplementary file 2).

Table 2.2. Protein concentration (PC), falling number (FN), and ash content (Ash), mean and stability parameters of hard red winter wheat breeding lines and commercial checks evaluated in two locations (i.e. Danforth and Madison) and three years (i.e. 2019-2021). Wricke's ecovalence stability coefficient (W) are shown for each trait. Ash content was evaluated in two years (2020-2021) in two locations (Danforth and Madison).

Genotype	PC (%)	W_{PC}	FN (s)	W_{FN}	Ash (%)	W_{Ash} 10^{+2}
7.02	10.4	1.95	322	1026	1.70	0.26
8.06	11.0	0.58	335	1699	1.66	2.50
9.1	10.7	0.30	317	2889	1.71	3.48
11.03	11.4	1.92	346	485	1.64	0.43
12.05	11.0	1.14	333	6386	1.69	0.26
47.04	10.4	0.86	323	3694	1.71	0.92
54.08	10.7	0.13	303	2254	1.72	2.02
138.04	10.7	0.13	342	4017	1.84	1.06
140.05	11.9	2.23	345	4762	1.88	1.52
145.06	10.7	2.00	361	7336	1.77	0.14
174.01	10.7	0.61	356	1435	1.66	0.87
183.05	10.6	0.44	344	6175	1.87	3.30
199.09	10.6	1.52	346	11706	1.73	0.71
202.07	10.0	0.33	334	842	1.73	0.10
212.08	11.0	0.84	328	4204	1.77	0.24
260.06	10.1	0.20	332	6181	1.79	1.14
Arapahoe	10.5	3.01	336	13183	1.79	4.33
Warthog	10.1	0.73	382	12431	1.90	1.04

Participatory on-farm evaluations

There was a change in genotype ranking between on-station and on-farm for grain yield, even within the same year (Table 3). The on-station experiments showed that the selected breeding lines were not different from *Warthog* in most of the environments (supplementary file 1, Table 3). The breeding lines *140.05* (4525 kg ha^{-1}) , *47.04* (6074kg ha^{-1}), and *260.06* (5196 kg ha^{-1}) were superior in yield to the commercial check *Warthog* (2711 kg ha^{-1}) in Ridgeway in 2020. Breeding lines

212.08 (2783 ha^{-1}) and *140.05* (4525 ha^{-1}) were inferior to *Warthog* (3824 kg ha^{-1}) in Ridgeway in 2021, and no differences were found among genotypes in Danforth, 2020 (Table 3). In Ridgeway on-farm trials in 2020, *140.05* was the only line with DON level over 1 ppm, while none of the lines had high DON in Freeville. *260.06* was not different from *Warthog* in Freeville 2020, while the lines *47.04* and *174.01* were significatively inferiors than *Warthog*. In on-farm trials in Ridgeway in 2021 lines *140.05* and *212.08* had values of DON over 1 ppm (supplementary file 7).

Table 2.2 Grain yield (kg ha^{-1}) performance of breeding lines and commercial checks in on-farm trials during the 2020 and 2021 years, and summary information for the same years of on-station performance.

	Wisconsin						Illinois				New York	Overall
	On-Station				On-Farm		On-Station			On-Farm	On-Farm	
	Madison			Spooner	Ridgeway		Danforth			Danforth	Freeville	
Genotype	2020	2021	2017-2021	2021	2020	2021	2020	2021	2017-2021	2020	2020	2017-2021
212.08	5054	4555^{B}	3429	1618	-	2783^{B}	1795	2781	2001	-	-	2639
140.05	4473	5994^{B}	3653	1597	4525^{A}	3293^{B}	1838	2261	1940	1425	-	2689
174.01	4991	5216^{B}	3715	1309	2157	-	1964	2387	2033	1484	3706^{B}	2769
202.07	5223	5176^{B}	3652	1467	-	3439	2420	2647	2219	-	-	2836
47.04	5123	6235^{B}	3917	1759	6074^{A}	3516	2017	2614	2225	1642	3807^{B}	2997
260.06	5059	5920^{B}	3762	1861	5196^{A}	3713	1795	2406	2093	1805	4195	2863
Warthog	4865	7443	4387	1805	2711	3824	2238	2773	2193	1805	5260	3234
S.E.	471	463	449	232	195	175	253	295	324	175	563	434
Dunnet	1115	1096	1064	549	461	414	600	699	768	414	1333	1028

Genotypes were compared to the check Warthog with a Dunnet test within trial. A represents mean performance superior to the check, B represents mean performance inferior to the check, all other mean performances are not different from the check performance.

Baking test and sensory evaluations

All the breeding lines evaluated in the 2020 bake tests had suitable baking properties compared to the commercial check (Figure 5). The *140.05* breeding line was unstable in the short fermentation test with an overall poor performance (Figure 5) but performed better in the long fermentation test and some bakers identified it as their favorite breeding line in the long fermentation evaluation (data not shown). The *174.01* breeding line was stable and the best performing genotype at most stages during the baking process (Figure 5). The bakers described the *212.18* breeding line as insufficient for extensibility during mixing compared to other lines but was close to optimal for the rest of the process (Figure 5). Line 260.06 was close to the check throughout the baking process (Figure 5). The check, Arapahoe, was rated slightly insufficient for extensibility by the bakers during mixing and for proofing strength, and then close to optimal for the rest of the process (Figure 5). All breeding lines were rated similarly to the check for flavor which is promising for commercialization. In the 2021 bake tests, the commercial Meadowlark bread flour (a blend including high-protein spring wheat) performed the best while the commercial check *Warthog* performed the worst (Figure 5). The two breeding lines had suitable baking properties for high quality artisanal bread at lower protein concentrations than the Meadowlark bread flour. The *260.06* breeding line was stable throughout the baking process, similar to the Meadowlark bread flour, and better than *Warthog*. This line was present in both years and performed well compared to other breeding lines and the checks. Although some breeding lines were different from the commercial checks at some stages in the baking trials, there were no statistical differences between breeding lines and checks for sensory traits in either year (data not shown).

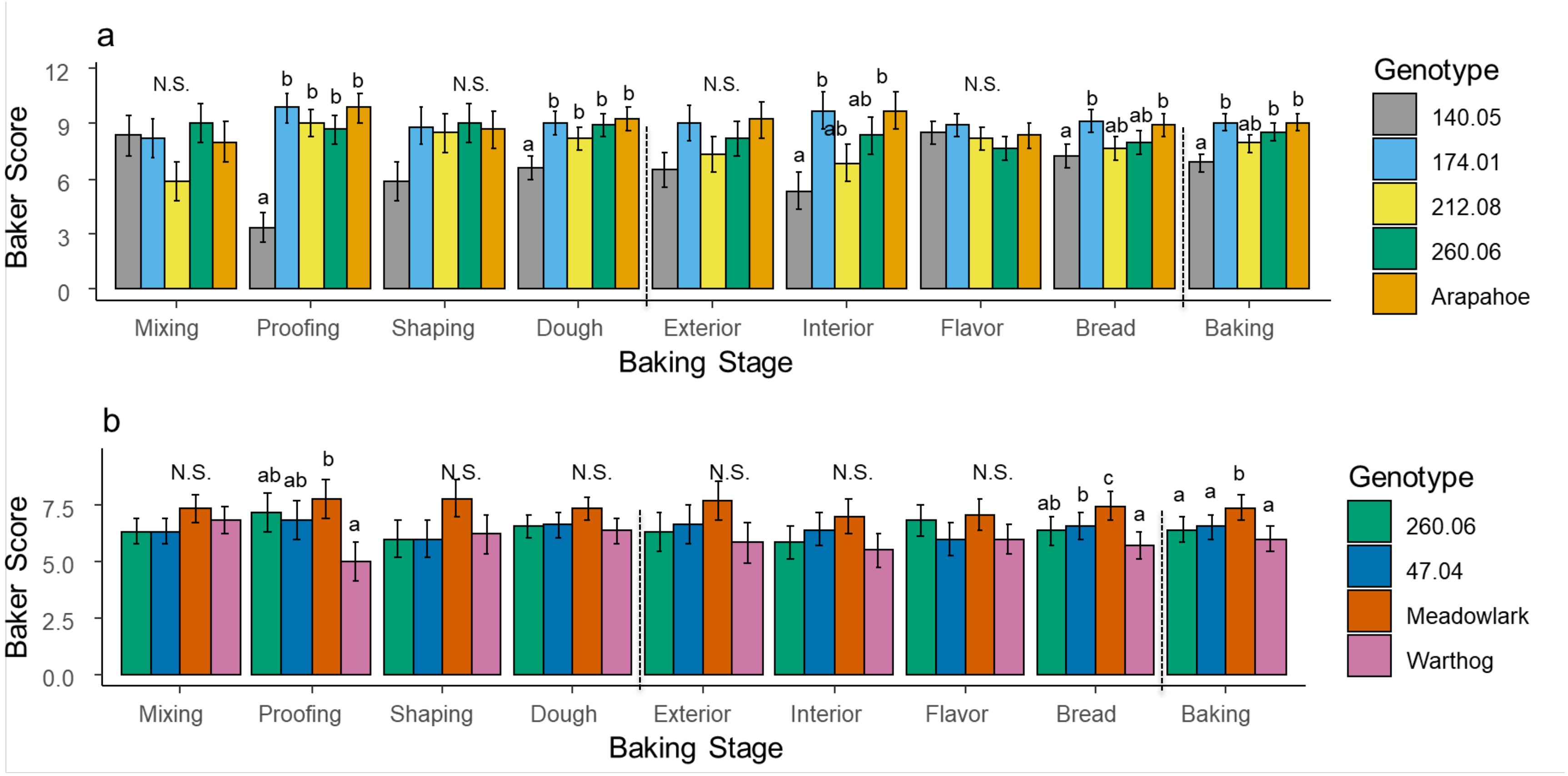


Figure 2.5. Baking trial test of hard red winter wheat breeding lines and commercial checks evaluated in naturally leavened artisan bread trials in 2020 and 2021. **a**. 2020 Madison Sourdough baking test. **b**. 2021 Remote baking test.

Discussion

Breeding strategies and characterization of selection

The breeding lines advanced and selected in our program have better grain quality than the checks and are more stable for grain quality and grain yield without compromising grain yield. Even though this project started and had its initial stages in the U.S. Northeast, the genetic diversity available was still relevant in the U.S. Upper Midwest probably because of similar breeding goals and the focus on stability. There are currently no cultivars of hard red winter wheat developed in the U.S. Upper Midwest available to farmers, and until this work started in 2012, there were no breeding programs with the explicit goal of breeding and selecting under organic conditions in the U.S. Upper Midwest. Furthermore, these plant breeding efforts are unique in the U.S. by targeting high-end artisan baking quality, although programs with similar goals and methods exist in Europe (Löschenberger et al. 2008; Wolfe et al. 2008; Dawson et al. 2011; Osman et al. 2012) and in the U.S. for spring wheat (Hills 2012; Kucek et al. 2017). Seven cycles of selection under certified organic conditions including two cycles of selection for FHB and protein in the Northeast (New York) and five cycles of selection for adaptation in the U.S. Upper Midwest (Illinois and Wisconsin, Figure 2) resulted in a set of 16 high performing breeding lines that perform well in terms of agronomic, disease, and grain quality performance. Importantly, some of these lines have equivalent mean performance and are more stable than existing check cultivars. Strong selection pressure was used initially in the breeding program to discard 98% of the lines based on tolerance to FHB from an FHB inoculated nursery and a tandem selection for protein content. Later, the largest driver of selection was adaptation to the U.S. Upper Midwest and disease resistance. We discarded eight breeding lines in 2017 that did not survive the winter in either location. Additionally, one line was discarded because it was the only breeding line showing a severe infection of glum blotch (*Phaeosphaeria nodorum*) which can severely reduce the quality of the grain. In 2018, we discarded seven breeding lines with a high level of natural infection to FHB, eight lines with low grain yield in either location, and four lines due to poor winter survival in either location in combination with a low-grain yield in 2017 or 2018. A strong selection for grain yield performance occurred in 2019 where overall performance across six environments in three years was used to discard eight lines based on general poor or inconsistent performance. Stability for grain yield was not explicitly estimated until the fourth year when genotypes had sufficiently sampled a range of environmental conditions. The core population of selected breeding lines therefore consisted of 16 $F_{3:8}$ breeding lines representing 64% of the $F_{2:3}$ families. This intensity of selection is relatively high (Rutkoski 2019). The 16 breeding lines represent nine of the original crosses combining eight parents. The core breeding population has a high level of genetic diversity (supplementary files 1-5) and has the potential for at least one commercial release as well as continuing additional breeding cycles to combine complementary traits.

Trait performance and GEI

Disease resistance. FHB can be caused by several species of the *Fusarium* and *Microdochium* genera, but is dominated by *Fusarium graminearum* Schwabe, teleomorph: *Gibberella zeae* in the U.S. Upper Midwest. It is one of the most damaging diseases in wheat because it has a direct effect on grain yield, and because grain contaminated with deoxynivalenol (DON) is a health hazard for people and animals (Su et al. 2019). Therefore, grain with high levels of DON cannot be sold to food markets and needs to be sold with a discounted price (Su et al. 2019). The development of FHB is favored by wet weather during flowering, which often occurs in the U.S. Upper Midwest and Eastern U.S. We had a strong selection for FHB using inoculated nurseries early in our breeding cycle and natural infection later. The commercial check Warthog did not meet food grade levels during our first year of baking trials and could therefore not be used for baking. Some of our selected breeding lines were able to meet food safety parameters consistently. Our work in collaboration with multi-year national efforts to breed cereal grain varieties with higher tolerance to FHB including the U.S. wheat and barley scab initiative will be of great benefit to both organic and conventional breeding programs (USWBSI 2019).

Agronomics and grain yield. Although GEI was relevant in our study, over 50% of the total genetic variance was due to genotypic main effect for grain yield and protein content. The remaining GEI did not represent repeatable GEI effects such as genotype by location or management practices and is therefore more difficult to predict. The strong effect of the genotype by location and genotype by location by year for grain yield is commonly found in other species (Gutierrez et al., 2015; Monteverde et al., 2018; Monteverde et al., 2019; Gonzalez-Barrios et al. 2019; Bhatta et al., 2020; Neyhart et al., 2021a,b) and is concordant with wheat results (Lado et al., 2016; Kucek et al. 2019), with the difference that mega-environments do not show repeatable sources of GEI in our study (supplementary file 6). Our study included two locations evaluated during five years, with only a third location incorporated during the last year. A small genotype by location variance could be the result of fewer locations being tested but could also mean that organic management systems in the U.S. Upper Midwest may represent a coherent target population of environments, within one mega-environment, with existing GEI due largely to unrepeatable causes.

Organic farmers identified weed competitive ability as a trait of interest (Kucek, 2017), likely due to limited weed control methods in organic systems. Genotypes with good weed competitive ability often have higher plant biomass and moderate to tall plant height (Kissing Kucek et al. (in press)). We found that the breeding lines that advanced to the later stages of evaluation are tall or have intermediate to high plant height. Although we did not evaluate weed competitive ability directly in our study, we believe that weed competitive ability might be an underlying reason for the superior performance of taller genotypes in our program. In some cases, tall plants might exhibit higher lodging incidence, however, our tall breeding lines exhibited a low incidence of lodging that might have contributed to their advantage.

Baking quality. Protein content is one of the most frequently used proxy traits for baking quality even though it does not fully explain baking quality properties (Borghi 1999; Gabriel et al. 2017). Baking quality is measured as the volume of bread by unit of flour. The relationship between baking quality and protein is not linear as quality increases with protein until 12%, and it might decrease with higher values of protein (Timms et al. 1981; Gabriel et al. 2017). This relationship is dependent on the baking method (Færgestad et al. 2000; Tronsmo et al. 2003). Therefore, protein content standards developed for industrial baking (Færgestad et al. 2000; Tronsmo et al. 2003) might be completely disconnected from requirements for artisanal baking where longer fermentation methods might be used (Ross 2018). The ratios between protein components are as important for baking quality as the total amount of protein (He et al. 2005). The protein quality is often more strongly related to genotype than overall protein content. For example, in French seed markets, varieties obtain an overall quality rating when released, which is independent of their protein content in any particular year. Breeding for stable protein levels and evaluating breeding lines for baking quality directly can result in better recommendations for farmers about which varieties have consistently good quality.

There is a long tradition of grain quality standards for bread wheat quality under conventional production that meet the standards of industrial milling and baking (Hills 2012; Sanchez-Garcia et al. 2015). However, artisanal baking quality is a more elusive trait to evaluate as no standards have been developed. For example, protein content and protein quality are still relevant traits for baking quality in artisanal baking (Hills 2012; Kucek et al. 2017), but quality parameters for artisan baked whole grain bread are not the same as for industrial bread baking that is based on ultra-refined flour (Ross 2018). Due to the longer sourdough fermentation and typically less intensive mixing process used by artisanal bakers, genotypes that do not meet the 12.5% of protein content required by industrial milling may still perform well in artisanal processes. This suggests that it may be possible to develop varieties of winter wheat that perform well in artisanal baking at lower protein concentrations if they have good protein quality. Different levels of product development input will be needed to develop standards for artisanal bread products, including baker and consumer testing. Although more challenging to implement, these evaluations will facilitate the incorporation of a broader set of users in the decision making process of end products for variety development (Dawson et al. 2011). High grain quality and flavor for artisanal baking are especially important traits in organic agriculture (Hills et al. 2013).

A strong genotypic main effect for protein content was observed in our study and has also been observed in other studies (Williams et al. 2008). High levels of diversity are still present for protein content among our breeding lines. This result was surprising because strong selection intensity for protein content in early generations was applied, and although no further direct selection for protein content was done at later stages of selection, it is still possible that indirect had an effect in our population in later generations given that genetic correlation between protein content and other traits such as grain yield (Kramer 1979; Marinciu 2009; Montesinos-López et al. 2016), and positively correlated to plant height (Montesinos-López et al. 2016).

Some of our breeding lines showed a higher level of stability for protein content through Wricke's ecovalence stability coefficient (i.e *54.08* W=0.13*, 138.04* W=0.13*, 260.06* W=0.2) than others (i.e. *11.03* W=1.92, *7.02* W= 1.95). Furthermore, there is no apparent correlation between the protein content and the stability for protein content of the breeding lines, so these might be able to be improved independently. Based on the artisanal baking tests, even though breeding lines were tested at lower than typical protein contents for industrial processes, they performed well compared to commercial checks. Farmers in the U.S. Upper Midwest frequently use the check variety *Warthog* because it typically has more consistent quality and performance than other commercially available varieties (participating farmers, pers. com.), yet it was one of the most volatile lines in our study for protein content.

Our results show a high level of environmental influence on falling number and large differences among genotypes. Falling number is a complex trait that depends on several factors and has a strong environmental influence (Johansson 2002; Kucek et al. 2017; Sjoberg et al. 2020). Sjoberg et al. (2020) report that genotypes with higher values of falling number tend to be more stable than genotypes with low levels of falling number. However, *Warthog* had the highest falling number (382 s) in our trial and was also the least stable genotype for falling number with the highest Wricke's ecovalence stability coefficient (W=1.24), indicating that the response of this check is more volatile to changes in the environment than the breeding lines. On the other hand, experimental lines had medium levels of falling number with a high Wricke's ecovalance stability coefficient (i.e. 202.07, FN= 334, W=0.08). Falling number was evaluated in six environments including two locations and three years. To have a better evaluation of the relationship between falling number mean values and the stability, a larger dataset would be required.

Two thirds of the genetic variation associated with ash content was explained by the genotype by year interaction, while the remaining third was explained by genotypic effects. These high levels of GEI are similar to results reported in other studies (Morris et al. 2009; Ficco et al. 2020). Even though one third of the genetic variation is associated to genotypic main effects, a very small range of values was observed in our study. Based on the Wricke's ecovalence stability coefficient, different genotypes are the most stable for falling number compared to grain yield. There are orders of magnitude for differences among genotypes for the Wricke's ecovalence stability coefficient and the breeding lines had low values (i.e. *212.08*, W=0.24) while *Arapahoe* had values twenty times higher (W=4.33). Grain size was correlated with ash content in some studies, including a durum wheat study (Ficco et al. 2020). This would be expected because the ash measures the bran-endosperm ratio. As an incidental observation, breeding lines with the lowest ash content seem to have plumper grain (i.e. *11.03*), unfortunately, we did not measure size and shape of the grain in this study.

On-farm trials

We found crossover GEI between on-farm and on-station performance for grain yield. This difference in performance between research stations and on-farm trials has been documented in the literature (Simmonds 1980, 1991), and given the prevalence of genotype by location interactions in multi-environment trials, this is not remarkable (Annicchiarico 2002a; Annicchiarico et al. 2006). On the other hand, the variation among on-station trials was higher than the differences between farms/years and research stations. This observed variation can be associated with uncontrolled factors that are not related to the commercial or research objective of the land management. However, given that our breeding lines are more stable overall than commercial checks, some of our high yielding breeding lines such as *260.06* and *47.04* could have consistently good agronomic performance, without a yield penalty, and with high baking quality. Another important consideration in the interpretation of the on-farm results is that they are based on two-years of data in two locations, and given that genotype by year and genotype by year by location interactions were substantial in our study, and are commonly reported as relevant in other studies (Kucek et al. 2019), caution should be exercised with the interpretation of these results.

Baking test

Our baking tests represent an exploratory approach to incorporate bakers in the process of choosing promising lines. Because of the large amount of grain needed for a baking trial and the small number of lines that can be tested at once, we were not able to test all lines in both years. The baking tests are time intensive and logistically difficult but worth the effort to obtain feedback from professional bakers prior to release of varieties specifically targeted at that market. The results from the baking trials, even though preliminary, are promising, and we will continue to test these lines prior to release. The results in the baking trials do not necessarily match the laboratory baking quality evaluations (supplementary file 5) and reports in the literature (Timms et al. 1981; Færgestad et al. 2000; Tronsmo et al. 2003; Sanchez-Garcia et al. 2015; Gabriel et al. 2017). Further work should identify or develop better laboratory tests that will allow selection of lines in early stages and strategic use of baking tests in the final stages on the process. The development of these laboratory resources for conventional programs aided selection for quality traits needed by industrial bakers. Reliable tests for artisanal baking quality that can be conducted with small grain samples will help in selection for the quality parameters needed by this market.

Breeding for grain yield stability

All breeding lines and the checks showed a dynamic stability for grain yield with different levels of performance of the FW linear model. The checks used in our experiments, *Warthog* and *Arapahoe*, were less stable than all of the breeding lines, showing a larger lack of fit of the FW regression (i.e. smaller R^2) and larger values for the Wricke's ecovalence stability coefficient indicating a more volatile response to the environment. Furthermore, some of the breeding lines combine high grain yield performance similar to the checks with more stable performance and a dynamic response to the environment. This combination of agronomic performance with stability is valuable in general, but especially under organic production systems where more diverse management practices are used (Wolfe et al. 2008; Lammerts van Bueren et al. 2011). A possible explanation for the high stability and performance could be the level of remnant genetic diversity present in some of these lines. The breeding lines were bulked at the F_4 level, where many loci are still segregating, maintaining some level of genetic diversity within the lines. In winter wheat populations with very high levels of genetic diversity (mixtures or composite crosses), a higher level of phenotypic plasticity has been reported (Weedon and Finckh 2019). The advantages of retaining more genetic diversity in breeding populations and farm fields has been extensively documented in cereals (Allard 1961b; Wolfe and Ceccarelli 2020; Bocci et al. 2020a). While our breeding lines are not as genetically diverse as the populations studied by those authors, there is some remaining phenotypic and genetic variability that may contribute to their stability. Another possible explanation is the highly exposure to multiple locations in early stages of line development (before reach F6 these lines were planted in 3 different locations). In many conventional breeding programs later generation lines are advanced to elite trials in multiple locations.

Our breeding lines have been selected during seven cycles under organic management practices, while the checks were selected under conventional management. Therefore, our breeding lines might also be more adapted to the specific environments of organic production. While selection in the target environment is key to achieving the best response to selection, this principle has not always been applied for breeding for organic production as this often represents a small market for conventional breeding programs. As a result, many lines used by organic farmers were developed under conventional management and then tested in organic systems. Unsurprisingly, studies have documented better adaptation to organic management when selecting under organic conditions

(Reid et al. 2009; Bocci et al. 2020a). Because of the interrelated nature of organic, local, and artisanal grain markets as a growing high-value option for farmers, we have developed a breeding program focused on this target environment, without a parallel conventional program. While the program is smaller than many conventional programs, it is focused on regional adaptation to organic systems and artisanal food end uses, which allows us to prioritize traits that may be unique to organic systems such as stable production and quality. Stability is important in all systems, but may be more relevant in organic production systems than in conventional systems (Knapp and van der Heijden 2018). It is therefore necessary to take stability into consideration to select varieties that can perform well across the diverse landscape of management systems that goes under the umbrella of organic production. There are multiple definitions of stability in plant breeding, and therefore, the interpretation of stability results should be performed carefully. We tried to focus on both spatial (i.e. location) as well as temporal (i.e. years) stability in our study, and we used both dynamic and static stability indicators. In this context, static stability refers to genotypes that perform similar across environmental conditions (Becker and Leon 1988), while dynamic stability refers to genotypes that maintain a relative performance according to the environmental mean and therefore respond to changes in environmental conditions (Finlay and Wilkinson 1963a). Dynamic stability is usually sought for traits such as grain yield while static stability is usually desirable for quality traits (Becker and Leon 1988), where a predictable and consistent product is advantageous for the industry and to sustain long-term profitability of regional grain systems.

Conclusions

The results presented here show that a small-scale breeding program with a distinct focus can produce relevant lines for an important emerging market. We have used the results of these trials to continue the process of crossing and selection to build a breeding pipeline that can continue to deliver improved germplasm for farmers in the U.S. Upper Midwest. While our lines have similar average performance to existing checks, the average performance does not capture the critical aspect of stability for farmers and bakers in regional grain systems. By choosing parents based on both high artisanal breadmaking quality and reliable performance in organic trials we have been able to select for improved stability of performance and quality. The fact that we can select more stable lines than existing commercial checks for both production and quality traits at the beginning phases of a breeding program is a promising indicator of the potential to develop high performing, stable varieties for the U.S. Upper Midwest. These results suggest that stable lines can be developed using a participatory breeding approach under organic management. Crop improvement explicitly targeting sustainable agriculture practices for selection with farm to table participatory perspectives are critical to achieve long term sustainable crop production.

References

Allard RW (1961) Relationship Between Genetic Diversity and Consistency of Performance in Different Environments 1. Crop Sci 1:127–133. https://doi.org/10.2135/cropsci1961.0011183X000100020012x

Annicchiarico P (2002) Genotype x environment interactions: challenges and opportunities for plant breeding and cultivar recommendations. Food & Agriculture Org., Rome, Italy.

Annicchiarico P, Bellah F, Chiari T (2006) Repeatable genotype×location interaction and its exploitation by conventional and GIS-based cultivar recommendation for durum wheat in Algeria. European Journal of Agronomy 24:70–81. https://doi.org/10.1016/j.eja.2005.05.003

Araus J, Slafer G, Royo C, and Serret MD (2008) Breeding for yield potential and stress adaptation in cereals. Crit Rev Plant Sci 27:377–412

Ark of Taste (2014) Red Fife Wheat. In: Ark of Taste. https://arkoftaste.slowfood.ca/red-fife-wheat/. Accessed 15 Oct 2021

Ashby JA (2009) The impact of participatory plant breeding. In: Plant breeding and farmer participation. FAO, Rome, Italy, pp 649–671

Baenziger PS, Schmidt JW, Peterson CJ, et al (1989) Registration of 'Arapahoe' Wheat. Crop Sci 29:832–832. https://doi.org/10.2135/cropsci1989.0011183X002900030070x

Bates D, Sarkar D, Bates MD, Matrix L (2007) The lme4 package. R package version 2:74

Becker HC, Leon J (1988) Stability Analysis in Plant Breeding. Plant Breeding 101:1–23. https://doi.org/10.1111/j.1439-0523.1988.tb00261.x

Bernardo, R (2010) Breeding for quantitative traits in plants 2nd ed

Bhatta M, Gutierrez L, Cammarota L, Cardozo F, Germán S, Gómez-Guerrero B, Pardo MF, Lanaro V, Sayas M, Castro AJ (2020) Multi-trait genomic prediction model increased the predictive ability for agronomic and malting quality traits in barley (Hordeum vulgare L.) G3: Genes, Genomes, Genetics Mar 1;10(3):1113-24 https://doiorg/101534/g3119400968

Bocci R, Bussi B, Petitti M, et al (2020) Yield, yield stability and farmers' preferences of evolutionary populations of bread wheat: A dynamic solution to climate change. European Journal of Agronomy 121:126156. https://doi.org/10.1016/j.eja.2020.126156

Borghi B (1999) Nitrogen As Determinant. In: Wheat: Ecology and physiology of yield determination. CRC Press, Binghamton, NY, p 67

Brummer EC, Barber WT, Collier SM, Cox TS, Johnson R, Murray SC, Olsen RT, Pratt RC, Thro AM (2011) Plant breeding for harmony between agriculture and the environment. Frontiers in Ecology and the Environment Dec;9(10):561-8

Canadian Grain Commission Determining test weight. https://www.grainscanada.gc.ca/en/grain-quality/official-grain-grading-guide/01-determining-test-weight/determining-test-weight.html. Accessed 6 Oct 2021

Carlson MO, Montilla-Bascon G, Hoekenga OA, et al (2019) Multivariate genome-wide association analyses reveal the genetic basis of seed fatty acid composition in oat (Avena sativa L.). G3: Genes, Genomes, Genetics 9:2963–2975

Cattivelli L, Rizza F, Badeck FW, et al (2008) Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. Field Crop Res 105: 1–14

Ceccarelli S, Grando S (2007) Decentralized-participatory plant breeding: an example of demand driven research. Euphytica 155:349–360. https://doi.org/10.1007/s10681-006-9336-8

Ceccarelli S, Grando S, Maatougui M, et al (2010) Plant breeding and climate changes. J Agr Sci 148 : 627–37

Colley MC, Dawson JC, McCluskey JR, et al (2021) Exploring the emergence of participatory plant breeding in countries of the global North – a review. J Agric Sci. https://doi.org/10.1017/S0021859621000782

Conley S, John G, Madburger D, Smith D (2015) Top 8 Recommendations for Winter Wheat Establishment in 2015 – Integrated Pest and Crop Management – UW–Madison. https://ipcm.wisc.edu/blog/2015/08/top-8-recommendations-for-winter-wheat-establishment-in-2015/. Accessed 6 Oct 2021

Cook RJ (2006) Toward cropping systems that enhance productivity and sustainability. P Natl Acad Sci USA 103: 18389–94

Dawson JC, Huggins DR, Jones SS (2008) Characterizing nitrogen use efficiency in natural and agricultural ecosystems to improve the performance of cereal crops in low-input and organic agricultural systems. Field Crops Research 107(2): 89-101

Dawson JC, Rivière P, Berthellot J-F, et al (2011) Collaborative Plant Breeding for Organic Agricultural Systems in Developed Countries. Sustainability 3:1206–1223. https://doi.org/10.3390/su3081206

Duvick DN (2003) The current state of plant breeding: how did we get here? In: Sligh M and Lauffer L (Eds). Seeds and breeds for 21st century agriculture. www.leopold.iastate.edu/sites/default/

Færgestad EM, Molteberg EL, Magnus EM (2000) Interrelationships of Protein Composition, Protein Level, Baking Process and the Characteristics of Hearth Bread and Pan Bread. Journal of Cereal Science 31:309–320. https://doi.org/10.1006/jcrs.1999.0304

FAOSTATS (2020) World Food and Agriculture - Statistical Yearbook 2020. FAO, Rome, Italy

Ficco DBM, Beleggia R, Pecorella I, et al (2020) Relationship between Seed Morphological Traits and Ash and Mineral Distribution along the Kernel Using Debranning in Durum Wheats from Different Geographic Sites. Foods 9:1523. https://doi.org/10.3390/foods9111523

Finlay K, Wilkinson G (1963) The analysis of adaptation in a plant-breeding programme. Aust J Agric Res 14:742. https://doi.org/10.1071/AR9630742

Fischer RA, Edmeades GO (2010). Breeding and cereal yield progress. Crop Science 5:S-85.

Fulcher, MR, & Bergstrom, GC (2020) First report of Fusarium armeniacum causing Fusarium head blight of wheat in New York. Plant Dis 104:3080. https://doi.org/10.1094/PDIS-04-20-0799-PDN

Fulcher MR, Winans JB, Benscher D, et al (2021) Triticum varieties grown as 'ancient grains' in New York differ in susceptibility to Fusarium head blight and harbor diverse Fusarium flora. Eur J Plant Pathol 159:693–699. https://doi.org/10.1007/s10658-020-02183-7

Gabriel D, Pfitzner C, Haase NU, et al (2017) New strategies for a reliable assessment of baking quality of wheat – Rethinking the current indicator protein content. J Cereal Sci 77:126–134. https://doi.org/10.1016/j.jcs.2017.08.002

Glover JD, Reganold JP, Bell LW, et al (2010) Increased food and ecosystem security via perennial grains. Science 328: 1638–39

González-Barrios P, Diaz-Garcia L, Gutierrez L (2019) Mega-environmental design: using genotype by environment interaction to optimize resources for cultivar testing. Crop Science 59:1–17 doi: 102135/cropsci2018110692

Green C, Ferreira G, Carlson A, et al (2017) USDA ERS - Growing Organic Demand Provides High-Value Opportunities for Many Types of Producers. In: Amber Waves Magazine. https://www.ers.usda.gov/amber-waves/2017/januaryfebruary/growing-organic-demand-provides-high-value-opportunities-for-many-types-of-producers. Accessed 5 Oct 2021

Gutierrez L, Germán S, Pereyra S, Hayes PM, Pérez CA, Capettini F, Locatelli A, Berberian NM, Falcioni E, Estrada R, Fros D, Gonza V, Altamirano H, Huerta-Espino J, Neyra E, Orjeda G, Sandoval-Islas S, Singh R, Turkington K, Castro AJ (2015) Multi-environment multi-QTL association mapping identifies disease resistance QTLs in barley germplasm from Latin America. Theoretical and Applied Genetics 128(3): 501-516 doi: 101007/s00122-014-2448-y

Heaton EA, Schulte LA, Berti M, Langeveld H, Zegada-Lizarazu W, Parrish D, Monti A (2013) Managing a second-generation crop portfolio through sustainable intensification: Examples from the USA and the EU. Biofuels, Bioproducts and Biorefining Nov;7(6):702-14

He ZH, Liu L, Xia XC, et al (2005) Composition of HMW and LMW Glutenin Subunits and Their Effects on Dough Properties, Pan Bread, and Noodle Quality of Chinese Bread Wheats. Cereal Chem 82:345–350. https://doi.org/10.1094/CC-82-0345

Hildermann I, Thommen A, Dubois D, et al (2009) Yield and baking quality of winter wheat cultivars in different farming systems of the DOK long-term trial. J Sci Food Agric 89:2477–2491. https://doi.org/10.1002/jsfa.3750

Hills K (2012) Improving baking quality and local markets for wheat grown in western Washington. Dissertation, Washington State University

Hills K, Goldberger J, Jones S (2013) Commercial Bakers' View on the Meaning of "Local" Wheat and Flour in Western Washington State. JAFSCD 1–20. https://doi.org/10.5304/jafscd.2013.034.022

Holland JB, Nyquist WE, Cervantes-Martínez CT (2003) Estimating and interpreting heritability for plant breeding: an update. Plant breeding reviews 22:9–112

Jacobs DF (2007) Toward development of silvical strategies for forest restoration of American chestnut (Castanea dentate) using blight-resistant hybrids. Biol Conserv 137: 497–506

Jonnalagadda SS, Harnack L, Hai Liu R, et al (2011) Putting the whole grain puzzle together: health benefits associated with whole grains—summary of American Society for Nutrition 2010 Satellite Symposium. The Journal of nutrition 141:1011S-1022S

Johansson E (2002) Effect of two wheat genotypes and Swedish environment on falling number, amylase activities, and protein concentration and composition. 8

Kissing Kucek L, Mallory E, Darby H, et al. (2021) Breeding wheat for weed-competitive ability: I. Correlated traits. Euphytica 217: 202. https://doi.org/10.1007/s10681-021-02930-9

Kiszonas AM, Morris CF (2018) Wheat breeding for quality: A historical review. Cereal Chemistry 95:17–34

Knapp S, van der Heijden MGA (2018) A global meta-analysis of yield stability in organic and conservation agriculture. Nat Commun 9:3632. https://doi.org/10.1038/s41467-018-05956-1

Kramer Th (1979) Environmental and genetic variation for protein content in winter wheat (Triticum aestivum L.). Euphytica 28:209–218. https://doi.org/10.1007/BF00056577

Krill-Brown A, Vang T, Carter C (2019) Developing Wheat Varieties for Whole Grain Artisanal Baking. Cereal Foods World 64

Kucek L (2017) Participatory Breeding of Wheat for Organic Production. Dissertation, Cornell University

Kucek L, Dyck E, Russell J, et al (2017) Evaluation of wheat and emmer varieties for artisanal baking, pasta making, and sensory quality. J of Cereal Sci 74:19e27. https://doi.org/10.1016/j.jcs.2016.12.010

Kucek L, Santantonio N, Gauch HG, et al (2019) Genotype × Environment Interactions and Stability in Organic Wheat. Crop Science 59:25–32. https://doi.org/10.2135/cropsci2018.02.0147

Lado B, Gonzalez-Barrios P, Quincke M, Silva P, Gutierrez L (2016) Modelling genotype by environment interaction for genomic selection with unbalanced data from a wheat (Triticum aestivum L.) breeding program. Crop Science doi: 102135/cropsci2015040207

Laffont J-L, Wright K, Hanafi M (2013) Genotype Plus Genotype × Block of Environments Biplots. Crop Science 53:2332–2341. https://doi.org/10.2135/cropsci2013.03.0178

Lammerts van Bueren ET, Jones SS, Tamm L, et al (2011) The need to breed crop varieties suitable for organic farming, using wheat, tomato and broccoli as examples: A review. NJAS - Wagening J Life Sci 58:193–205. https://doi.org/10.1016/j.njas.2010.04.001

Lammerts van Bueren ET, Myers JR (eds) (2012) Organic crop breeding. Wiley-Blackwell, Chichester, West Sussex, UK ; Ames, Iowa

Lenth RV (2021) emmeans: Estimated Marginal Means, aka Least-Squares Means. Version 1.5.4. CRAN, CRAN. URL https://CRAN.R-project.org/package=emmeans

Liebman M (2012) Increasing cropping system diversity balances productivity, profitability and environmental health. PloS one. Oct 10;7(10):e47149

Löschenberger F, Fleck A, Grausgruber H, et al (2008) Breeding for organic agriculture: the example of winter wheat in Austria. Euphytica 163:469. https://doi.org/10.1007/s10681-008-9709-2

Malosetti M, Ribaut J-M, van Eeuwijk FA (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front. Physiol. 4:44. doi: 10.3389/fphys.2013.00044

Marinciu C (2009) Grain yield and protein concentration in winter wheat cultivars tested with and without nitrogen fertilizer. Romanian Agrc Resch 7

Matlock T (2021) Organic: A Thriving Agriculture Segment. https://www.usda.gov/media/blog/2020/10/28/organic-thriving-agriculture-segment. Accessed 14 Oct 2021

Mellen PB, Walsh TF, Herrington DM (2008) Whole grain intake and cardiovascular disease: a meta-analysis. Nutrition, Metabolism and Cardiovascular Diseases 18:283–290

Montesinos-López OA, Montesinos-López A, Crossa J, et al (2016) A Genomic Bayesian Multi-trait and Multi-environment Model. G3 6:2725–2744. https://doi.org/10.1534/g3.116.032359

Monteverde E, Gutierrez L, Blanco P, Pérez de Vida F, Rosas JE, Bonnecarrère V, Quero G, McCouch S (2019) Integrating molecular markers and environmental covariates to interpret genotype by environment interaction in rice (Oryza sativa L.) grown in temperate areas G3: Genes, Genomes, Genetics 9(5): 1519-1531; https://doiorg/101534/g3119400064

Monteverde E, Rosas JE, Blanco P, Perez de Vida F, Bonnecarrere V, Quero G, Gutiérrez L, McCouch S (2018) Multi-Environment models increase prediction accuracy of complex traits in rice advanced breeding lines of rice (O. sativa). Crop Science 58:1519–1530 doi: 102135/cropsci2017090564

Morris CF, Li S, King GE, et al (2009) A Comprehensive Genotype and Environment Assessment of Wheat Grain Ash Content in Oregon and Washington: Analysis of Variation. Cereal Chemistry Journal 86:307–312. https://doi.org/10.1094/CCHEM-86-3-0307

Murphy KM, Campbell KG, Lyon SR, Jones SS (2007) Evidence of varietal adaptation to organic farming systems. Field Crops Research 102:172–177. https://doi.org/10.1016/j.fcr.2007.03.011

Neyhart JL, Silverstein K, Gutierrez L, Smith KP (2021)a Optimizing the choice of test locations for multi-trait genotypic evaluation. Crop Science (in press)

Neyhart JL, Gutiérrez L, Smith KP (2021)b Using environmental similarities to design training sets for genomewide selection. Crop Science 61(1): 396-409 https://doiorg/101002/csc220303

Osman AM, Struik PC, Lammerts van Bueren ET (2012) Perspectives to breed for improved baking quality wheat varieties adapted to organic growing conditions: Perspectives to breed for improved baking quality of organic wheat. J Sci Food Agric 92:207–215. https://doi.org/10.1002/jsfa.4710

Pennacchi JP, Carmo-Silva E, Andralojc PJ, Lawson T, Allen AM, Raines CA, Parry MA (2019) Stability of wheat grain yields over three field seasons in the UK. Food and energy security, 8(2), p.e00147

Piepho H-P (2019) A coefficient of determination (R2) for generalized linear mixed models. Biometrical Journal 61:860–872

Przystalski M, Osman A, Thiemt EM, et al (2008) Comparing the performance of cereal varieties in organic and non-organic cropping systems in different European countries. Euphytica 163:417–433. https://doi.org/10.1007/s10681-008-9715-4

R. Core Team (2013) R: A language and environment for statistical computing. Vienna, Austria

Rana J, Paul J (2017) Consumer behavior and purchase intention for organic food: A review and research agenda. Journal of Retailing and Consumer Services 38:157–165. https://doi.org/10.1016/j.jretconser.2017.06.004

Reid TA, Yang R-C, Salmon DF, Spaner D (2009) Should spring wheat breeding for organically managed systems be conducted on organically managed land? Euphytica 169:239–252. https://doi.org/10.1007/s10681-009-9949-9

Reynolds A, Mann J, Cummings J, et al (2019) Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. The Lancet 393:434–445

Robertson GP and Swinton SM (2005) Reconciling agricultural productivity and environmental integrity: a grand challenge for agriculture. Front Ecol Environ 3: 38–46

Ross AS (2018) Flour Quality and Artisan Bread. Cereals Foods World 63

Runck BC, Kantar MB, Jordan NR, Anderson JA, Wyse DL, Eckberg JO, Barnes RJ, Lehman CL, DeHaan LR, Stupar RM, Sheaffer CC (2014) The reflective plant breeding paradigm: A robust system of germplasm development to support strategic diversification of agroecosystems. Crop Science. 54(5):1939-48

Rutkoski JE (2019) A practical guide to genetic gain. In: Advances in Agronomy. Elsevier, Cambridge, MA, pp 217–249Shennan C, Krupnik TJ, Baird G, et al (2017) Organic and Conventional Agriculture: A Useful Framing? Annu Rev Environ Resour 42:317–346. https://doi.org/10.1146/annurev-environ-110615-085750

Tilley M, Chen YR, Miller RA (2012) Wheat breeding and quality evaluation in the US. In: Breadmaking. Elsevier, Cambridge, UK, pp 216–236

Sanchez-Garcia M, Álvaro F, Peremarti A, et al (2015) Changes in bread-making quality attributes of bread wheat varieties cultivated in Spain during the 20th century. European Journal of Agronomy 63:79–88. https://doi.org/10.1016/j.eja.2014.11.006

Santini A, La Porta N, Ghelardini L, and Mittempergher L (2007) Breeding against Dutch elm disease adapted to the Mediterranean climate. Euphytica 163: 45–56

Schulte LA, Asbjornsen H, Liebman M, Crow TR (2006) Agroecosystem restoration through strategic integration of perennials. Journal of Soil and Water Conservation Nov 1;61(6):164A-9A

Simmonds NW (1980) Comparisons of the Yields of Four Potato Varieties in Trials and in Agriculture. Ex Agric 16:393–398. https://doi.org/10.1017/S0014479700012059

Simmonds NW (1991) Selection for local adaptation in a plant breeding programme. Theoret Appl Genetics 82:363–367. https://doi.org/10.1007/BF02190624

Sjoberg SM, Carter AH, Steber CM, Garland Campbell KA (2020) Application of the factor analytic model to assess wheat falling number performance and stability in multienvironment trials. Crop Sci 61:372–382. https://doi.org/10.1002/csc2.20293

Shennan C, Krupnik TJ, Baird G, et al (2017) Organic and Conventional Agriculture: A Useful Framing? Annu Rev Environ Resour 42:317–346. https://doi.org/10.1146/annurev-environ-110615-085750

Su Z, Bernardo A, Tian B, et al (2019) A deletion mutation in TaHRC confers Fhb1 resistance to Fusarium head blight in wheat. Nature genetics 51:1099–1105

Tilman D (1999) Global environmental impacts of agricultural expansion: the need for sustainable and efficient practices. P Natl Acad Sci USA 96: 5995–6000.Ross AS (2018) Flour Quality and Artisan Bread. Cereals Foods World 63

Tilley M, Chen YR, Miller RA (2012) Wheat breeding and quality evaluation in the US. In: Breadmaking. Elsevier, Cambridge, UK, pp 216–236

Timms MF, Bottomley RC, Ellis JRS, Schofield JD (1981) The baking quality and protein characteristics of a winter wheat grown at different levels of nitrogen fertilisation. J Sci Food Agric 32:684–698. https://doi.org/10.1002/jsfa.2740320709

Tronsmo KM, Færgestad EM, Schofield JD, Magnus EM (2003) Wheat protein quality in relation to baking performance evaluated by the Chorleywood bread process and a hearth bread baking test. Journal of Cereal Science 38:205–215. https://doi.org/10.1016/S0733-5210(03)00027-4

UPOV (2017) Guidelines for the conduct of tests for distinctness, uniformity and stabiliy. Geneva. https://www.upov.int/edocs/tgdocs/en/tg003.pdf

USDA/NASS (2021) USDA/NASS QuickStats Ad-hoc Query Tool. In: QuickStats Ad-hoc Query Tool. https://quickstats.nass.usda.gov/results/DF990781-9A03-30B0-BA7A-AE0D627F1F0B?pivot=short_desc. Accessed 14 Oct 2021

USDA/NASS (2019) 2019 Organic Survey. https://www.nass.usda.gov/Publications/AgCensus/2017/Online_Resources/Organics/index.php. Accessed 14 Oct 2021

USWBSI S-U (2019) Home Page | USWBSI. https://scabusa.org/. Accessed 14 Oct 2021

van Rossum B-J, van Eeuwijk F, Boer M, Malosetti M (2021) Package ‘statgenGxE.’ In: CRAN. https://cran.r-project.org/web/packages/statgenGxE/index.html

Wani SH, Choudhary M, Kumar P, Akram NA, Surekha C, Ahmad P, Gosal SS (2018) Marker-assisted breeding for abiotic stress tolerance in crop plants. In: Biotechnologies of crop improvement, vol 3. Springer, Cham, pp 1–23

Wassman R, Jagadish SVK, Sumfleth K, Pathak H, Howell G, Ismail A, Serraj R, Redona E, Singh RK, Heuer S (2009) Regional vulnerability of climate change impacts on Asian rice production and scope for adaptation. Adv Agron 102:91–133

Weedon OD, Finckh MR (2019) Heterogeneous Winter Wheat Populations Differ in Yield Stability Depending on their Genetic Background and Management System. Sustainability 11:6172. https://doi.org/10.3390/su11216172

Wheat CAP (2018) APPENDIX 5 Economic impact report of WheatCAP, TCAP and USDA-ARS genotyping labs. In: Wheat CAP. https://www.triticeaecap.org/wp-content/uploads/2018_Wheat_CAP_economic_report.pdf

Wickham H (2016) ggplot2: elegant graphics for data analysis. springer

Williams RM, O'Brien L, Eagles HA, et al (2008) The influences of genotype, environment, and genotype×environment interaction on wheat quality. Aust J Agric Res 59:95. https://doi.org/10.1071/AR07185

Wolfe MS, Baresel JP, Desclaux D, et al (2008) Developments in breeding cereals for organic agriculture. Euphytica 163:323–346. https://doi.org/10.1007/s10681-008-9690-9

Wolfe MS, Ceccarelli S (2020) The increased use of diversity in cereal cropping requires more descriptive precision. J Sci Food Agric 100:4119–4123. https://doi.org/10.1002/jsfa.9906

Wricke G (1962) Evaluation method for recording ecological differences in field trials. Z Pflanzenzücht 47: 92–96

Annex 1

Supplementary Table 2.1 Best linear unbiased estimates (BLUEs) of grain yield (kg ha^{-1}) of hard red winter wheat breeding lines and commercial checks evaluated in eleven environments in three locations (Madison, Danforth and Spooner) and five years (i.e. 2017-2021). Trial mean, standard error (S.E.), Fisher's least square differences (LSD) for α=0.05, heritability (H^2), and whether a post-blocking row and/or column effect was used to obtain BLUEs is reported.

	Madison						Spooner	Danforth						
Genotype	2017	2018	2019	2020	2021	5-yr	2021	2017	2018	2019	2020	2021	5-yr	Overall
7.02	2836	1870	4736	4763	6209	3776	1517	1678	1995*	1182	--	2118	1964	2766
8.06	2399	1859	4368	4875	5497	3612	1521	2056	1838	1244*	1805	1852	1810	2587
9.1	2398	1964	3397	5377*	5761	3610	1504	2166*	1879	1209	1658	2403	2089*	2750
11.03	2453	2076	4358	4842	4832	3376	1389	1662	1858	1208	1552	1961	1807	2518
12.05	3643	2907*	5530*	5597*	6318	3938	1414	1877	1885	1421*	1552	2333	1996	3082*
47.04	2920	2417	4495	5123	6235	3917	1730*	1960	1961*	1142	2017	2614*	2225*	2997*
54.08	3041	2348	5053*	4773	5545	3460	1528	1633	1800	1343*	2133	2218	2024	2655
138.04	2306	1103	4824	4460	5744	3285	1514	2056	1734	1062	2217*	2152	2081*	2604
140.05	2195	2725*	3721	4473	5994	3653	1571	1517	1748	--	--	2261	1940	2689
145.06	2362	1898	3997	5052	5731	3658	1581	1784	1891	1250*	--	2181	1844	2642
174.01	2117	2650	4364	4991	5216	3715	1287	2380*	1746	1375*	1964	2387	2033	2769
183.05	1833	1610	3493	4741	5359	3497	1724	2265*	2051*	1251*	1784	2216	2101*	2698
199.09	2153	1548	4378	5074	5245	3515	1275	1962	1790	1342*	1784	2751*	2130*	2750
202.07	2626	1759	4561	5223	5176	3652	1443	1863	1898	1233	2281*	2647*	2219*	2836*
212.08	3793	1808	4460	5054	4555	3429	1591	2008	1747	1463*	1795	2781*	2001	2639
260.06	2922	2185	4515	5059	5920	3762	1830*	2165*	2140*	1195	1795	2406	2093*	2863*
Arapahoe	4119	3408*	4867	4871	6699	4159*	1328	2191*	1991*	1276*	2502*	2850*	2355*	3206*
Warthog	3557	3028*	4221	4865	7443	4387*	1775*	1430	2124*	1256*	2238*	2773*	2193*	3234*
Trial mean	2293	2155	4357	5001	5748	3719	1529	1793	1886	1255	1958	2383	2050	2793
Trial S.E.	35.5	80.0	52.4	70.8	44.0	18.2	14.0	35.5	24.3	31.6	45.0	23.3	13.3	12.5
LSD	363.5	733.9	453.6	530.4	439.6	426.0	100.4	224.1	225.2	236.7	331.4	280.0	302.0	410.4
Row	Yes	Yes	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No
Col	No	No	Yes	No	No	Yes	No	No	No	No	No	Yes	No	No
H^2	0.73	0.65	0.84	0.44	0.94	0.76	0.86	0.92	0.75	0.37	0.85	0.81	0.53	0.80

Genotypes that are not significantly different (α=0.05) from the highest yielding genotype in the trial are marked with a star. Danforth 5-year mean, and overall mean did not show significant differences among genotypes for grain yield.

Supplementary Table 2.2. Best linear unbiased estimates (BLUEs) of test weight (g 500 ml^{-1}) of hard red winter wheat breeding lines and commercial checks evaluated in seven environments in two locations (Madison and Danforth) and four years (i.e. 2018-2021). Trial mean, standard error (S.E.), heritability (H^2), Fisher's least square differences (LSD) for α=0.05, and whether a post-blocking row and/or column effect was used to obtain BLUEs is reported.

	Madison					Danforth				
Genotype	2018	2019	2020	2021	4-yr	2018	2019	2021	3-yr	Overall
7.02	330	367	384	380	368	363	344	373	361	364
8.06	332	369	391	364	368	374*	368*	386*	376*	371*
9.1	314	372	375	357	356	368	358	380	368	359
11.03	336	372	385	374	368	381*	367*	383	377*	369*
12.05	342*	366	387	374	370	361	352	388*	367	369*
47.04	334	362	388	377	369	372	358	378	369	368
54.08	338	364	388	380	368	365	347	379	362	366
138.04	310	355	374	371	357	368	359	369	366	358
140.05	348*	377*	382	382	376*	380*	367*	382	376*	374*
145.06	331	379*	384	380	370	366	363	368	365	365
174.01	351*	384*	398*	385*	379*	375*	366*	386*	375*	376*
183.05	327	357	380	375	360	369	356	366	364	358
199.09	327	381	388	381	372*	362	362	376	365	368
202.07	341*	373*	380	374	368	367	359	372	366	365
212.08	347*	379*	387	376	374*	365	360	369	366	368
260.06	332	357	387	371	368	377	359	378	372	367
Arapahoe	352*	371	387	386*	376*	356	351	373	360	368
Warthog	355*	371	381	388*	377*	354	347	383	370	372*
Trial mean	340	368	384	376	369	368	357	377	368	367
Trial S.E.	1.4	0.9	0.5	0.4	0.4	0.7	0.4	0.2	0.2	0.3
Trial LSD	13.9	8.2	4.4	3.7	7.5	7.3	3.5	2.4	3.8	7.4
Row	Yes	Yes	No	No	Yes	No	No	Yes	No	No
Col	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	No
H^2	0.62	0.81	0.97	0.91	0.70	0.72	0.94	0.99	0.89	0.87

Genotypes that are not significantly different (α=0.05) from the highest test weight genotype in the trial are marked with a star.

Supplementary Table 2.3. Best linear unbiased estimates (BLUEs) of plant height (cm) and heading date (in Julian days) of hard red winter wheat breeding lines and commercial checks evaluated in multiple environments in three locations (Madison, Spooner, and Danforth) and five years (2017-2021). Trial mean, standard error (S.E.), heritability (H^2), Fisher's least square differences (LSD) for α=0.05, and whether a post-blocking row and/or column effect was used to obtain BLUEs is reported.

	Plant height (cm)											Heading date (day)						
	Wisconsin						Spooner	Illinois				Wisconsin						Spooner
Genotype	2017	2018	2019	2020	2021	5-yr	2021	2018	2019	2021	3-yr	2017	2018	2019	2020	2021	5-yr	2021
7.02	99	87	118	101	116	106	90	81	118	95	93	154	157	158	155	145	153	157
8.06	114	105	136	115	123	122	95	0	136	114	104*	154	160	158	159*	147	156	158
9.1	108	111	127	112	121	117	90	94	127	116	105*	154	158	151	159*	148	154	158
11.03	126*	128	144	130	123	131*	98	100	144	111	104*	158	162*	151	162	153	157	159
12.05	114	97	113	101	117	108	88	86	113	104	95	150	153	157	156	144	151	155
47.04	110	102	112	99	100	103	78	83	112	96	89	154	158	156	158	145	153	158
54.08	99	96	115	110	117	108	102*	81	136	119*	100	152	156	158	158	145	154	155
138.04	110	101	114	102	114	109	87	82	114	103	92	160*	163*	151	160*	147	156	159*
140.05	97	103	126	102	123*	111	90	86	126	107	95	153	157	158	158	147	155	157
145.06	116	115	125	95	119*	117	90	90	125	108	99	157	155	162	158	148	156	158
174.01	117	108	126	110	117	117	93	95	126	113	104*	147	160	156	156	145	151	154
183.05	105	109	130	105	111	111	85	85	130	100	94	156	158	151	160*	154*	157	159*
199.09	108	104	126	110	112	112	88	91	126	108	99	157	156	154	159	148	154	157
202.07	118	112	134	114	116	119	87	94	134	110	102*	157	162*	162*	160*	153	158*	159
212.08	123*	114	128	103	111	114	92	91	128	109	101	153	161	162*	158	147	156	157
260.06	110	110	129	107	119*	116	90	92	129	108	98	157	160	151	159*	152	156	158
Arapahoe	104	110	112	108	104	110	77	77	112	98	88	147	154	158	156	146	152	155
Warthog	100	98	116	96	106	104	88	81	116	92	87	157	156	160	158	145	153	157
Trial mean	109	106	124	107	115	113	89	89	123	106	97	154	157	156	158	148	155	157
Trial S.E.	0.49	1.2	0.6	1.5	0.6	0.2	0.4	0.6	1	0.3	0.2	1	0.2	0.4	0.3	0.1	0.04	0.1
Trial LSD	4.85	11.4	5.3	11.6	5.9	5.5	3	5.3	8.5	3.1	4.3	0.8	1.4	3.5	2.6	0.9	1.0	0.5
H^2	0.85	0.84	0.96	0.87	0.85	0.87	0.87	0.87	0.79	0.97	0.93	0.99	0.95	0.99	0.84	0.96	0.47	0.94
Row	No	Yes	No	No	Yes	Yes	No	No	No	No	No	No	Yes	No	No	No	Yes	No
Col	No	No	No	Yes	Yes	Yes	No	No	No	No	No	No	No	No	Yes	No	Yes	No

Genotypes that are not significantly different (α=0.05) from the highest plat height or earliest plant height genotype in the trial are marked with a star.

Supplementary Table 2.4. Best linear unbiased estimates (BLUEs) of ash content (%) and falling number (in seconds) of hard red winter wheat breeding lines and commercial checks evaluated in multiple environments in two locations (Danforth and Madison) and three years (2019-2021). Trial mean, standard error (S.E.), heritability (H^2), Fisher's least square differences (LSD) for α=0.05, and whether a post-blocking row and/or column effect was used to obtain BLUEs is reported.

	Ash (%)							Falling Number (s)								
	Danforth			Madison				Danforth				Madison				
Genotype	2020	20211	2-yr	2020	2021	2-yr	Overall	2019	2020	2021	3-yr	2019	2020	2021	3-yr	Overall
7.02	-	1.7*	1.7	1.7	1.7	1.7	1.7	361	-	225	312	369	354	281	334	322
8.06	1.8	1.6*	1.7	1.6*	1.7	1.6	1.7	351	346	271	322	370	362	315	348	335
9.1	1.7	1.6*	1.7	1.7	1.9*	1.8	1.7	363	316	201	294	364	328	312	341	317
11.03	1.6*	1.6*	1.6	1.7	1.7	1.7	1.6	363	374	247	328	399	358	322*	365	346
12.05	1.7	1.7*	1.7	1.7	1.6	1.7	1.7	311	359	285	318	356	381	305	348	333
47.04	1.8	1.7*	1.7	1.7	1.8*	1.7	1.7	376*	365	202	313	365	332	275	323	323
54.08	1.8	1.6*	1.7	1.8	1.7	1.7	1.7	332	357	214	301	352	335	242	305	303
138.04	1.8	1.9	1.8	1.9	1.8*	1.9*	1.8	-	369	273	351*	360	359	304	342	342
140.05	1.8	1.9	1.9*	2.0	1.8*	1.9*	1.9*	400*	365	289	351*	350	379*	307	338	345
145.06	-	1.7*	1.7	1.8	1.8*	1.8	1.8	390*	-	275	350*	485*	319	310	359	361*
174.01	1.6*	1.7*	1.6	1.7	1.7	1.7	1.7	403*	389	276	356	396	377*	299	356	356
183.05	1.7	1.9	1.8	2.0	1.9*	1.9*	1.9*	334	349	318	334	397	337	335	377	344
199.09	1.7	1.8	1.7	1.7	1.8*	1.7	1.7	364	373	206	314	495	335	301	438*	346
202.07	1.7	1.8	1.7	1.7	1.7	1.7	1.7	360	363	254	326	370	347	311	343	334
212.08	1.8	1.8	1.8	1.8	1.7	1.8	1.8	378*	375	199	317	396	354	267	337	328
260.06	1.8	1.8	1.8	1.8	1.9*	1.8	1.8	354	334	302	331	340	350	297	331	332
Arapahoe	1.7	1.9	1.8	1.8	1.6	1.8	1.8	362	364	164	295	461	337	335*	377	336
Warthog	2.0	1.9	1.9*	1.9	1.8*	1.9*	1.9*	364	382	383*	368**	491*	390*	324*	401	382*
Trial mean	1.8	1.8	1.8	1.8	1.7	1.8	1.8	363	361	255	327	395	352	302	354	338
Trial S.E.	0.004	0.01	0.006	0.01	0.008	0.007	0.005	3.3	1.6	2.2	1.7	1.9	1.7	2.7	1.7	1.8
LSD	0.039	0.050	0.100	0.050	0.052	0.056	0.050	29.1	14.2	17.0	20.0	19.0	16.8	18.2	20.0	21.0
Row	No	Yes	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Col	No	No	No	Yes	Yes	No	No	No	No	No	No	No	No	No	No	No
H^2	0.94	0.91	0.70	0.87	0.82	0.71	0.87	0.47	0.85	0.96	0.16	0.94	0.93	0.83	0.47	0.57

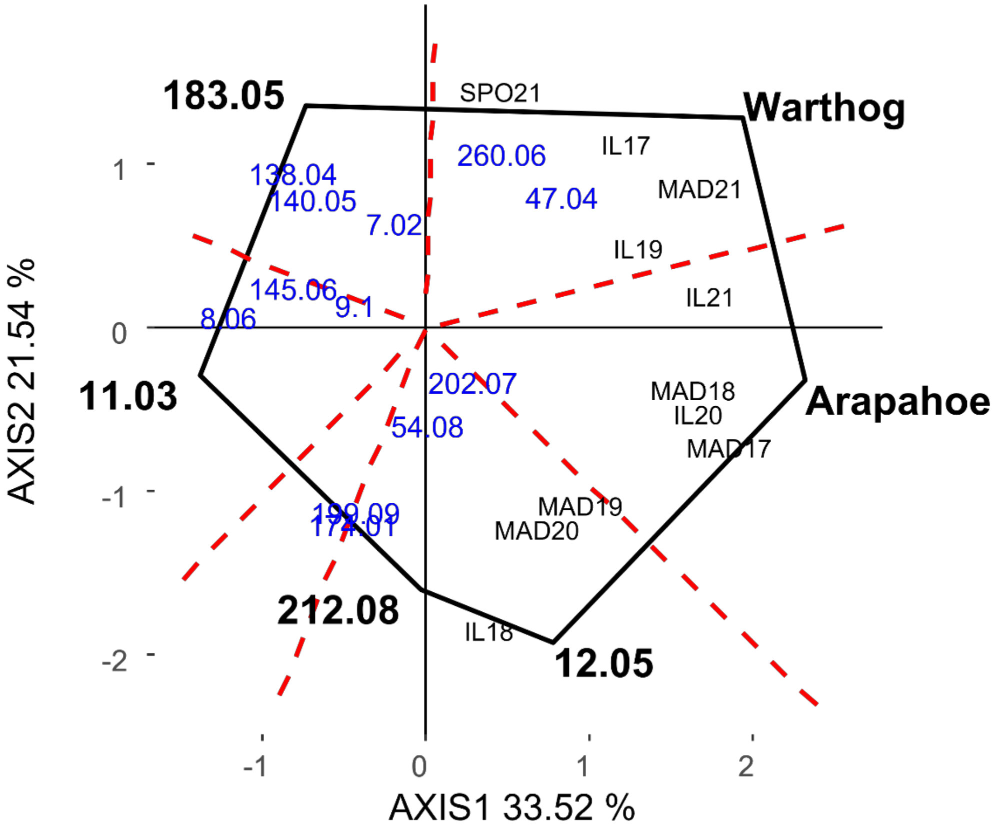
Genotypes that are not significantly different (α=0.05) from the highest ash content or falling number value genotype in the trial are marked with a star.

Supplementary Table 2.5. Best linear unbiased estimates (BLUEs) of protein content (%) of hard red winter wheat breeding lines and commercial checks evaluated in multiple environments in two locations (Danforth and Madison) and three years (i.e. 2019-2021). Trial mean, standard error (S.E.), heritability (H^2), Fisher's least square differences (LSD) for α=0.05, and whether a post-blocking row and/or column effect was used to obtain BLUEs is reported.

	Danforth				Madison				
Genotype	2019	2020	2021	3-yr	2019	2020	2021	3-yr	Overall
7.02	9.6	-	9.8	9.4	11.9	10.5	12.2	11.5	10.4
8.06	9.5	8.9	10.6	9.8	12.8	10.8	13.2	12.0	11.0
9.1	10.2	8.8	9.7	9.4	12.4	10.7	12.6	12.0	10.7
11.03	9.3	9.1	10.1	9.8	14.0	12.2*	13.1	13.0*	11.4
12.05	9.7	10*	10.6	10.0	12.3	11.5	12.5	12.0	11.0
47.04	9.2	8.7	10.0	9.5	12.1	10.4	11.6	11.2	10.4
54.08	9.3	9.0	10.1	9.4	12.4	11.0	12.3	11.9	10.7
138.04	9.1	9.1	10	9.4	12.7	10.9	12.5	11.9	10.7
140.05	10.7*	9.9*	11.0*	10.5*	14.8*	12.3*	12.6	13.3*	11.9*
145.06	9.0	-	9.6	9.8	11.9	10.4	12.4	11.7	10.7
174.01	9.5	9.3	10.5	9.6	12.2	10.7	12.3	11.7	10.7
183.05	8.8	8.8	9.7	9.2	12.4	10.7	12.8*	12.0	10.6
199.09	9.0	8.6	9.8	9.0	13.4	10.7	12.8	12.3	10.6
202.07	8.5	8.7	9.0	8.8	11.7	10.0	11.8	11.1	10.0
212.08	9.2	8.9	10.1	9.5	12.9	11.2	13.3	12.5	11.0
260.06	8.9	8.7	9.3	8.9	11.7	10.3	12.0	11.4	10.1
Arapahoe	8.6	8.9	10.9*	9.5	12.1	10.0	10.9	11.4	10.5
Warthog	9.0	8.8	9.5	8.9	12.5	10.0	11.3	11.2	10.1
Trial mean	9.3	9.0	10.0	9.5	12.6	10.8	12.3	11.9	10.7
Trial S.E.	0.0004	0.02	0.03	0.0	0.0003	0.02	0.08	0.02	0.02
LSD	0.0034	0.17	0.24	0.4	0.0032	0.22	0.52	0.39	0.38
Row	No	No	No	No	No	No	No	No	No
Col	No	No	No	No	Yes	Yes	Yes	No	No
H^2	0.76	0.96	0.93	0.78	0.94	0.96	0.86	0.90	0.86

Genotypes that are not significantly different (α=0.05) from the highest protein content genotype in the trial are marked with a star.

Supplementary Figure 2.1. GGE biplot representation of 16 hard red winter wheat genotypes evaluated in multi-environment trials for grain yield (kg ha^{-1}) for Danforth (2017-2021), Madison (2017-2021), and Spooner (2021). Wining genotypes on each mega environment are shown in bold, environments are represented in black, and genotypes in blue.



Supplementary Table 2.6. DON content (ppm) performance of hard red winter wheat breeding lines and commercial check (i.e. Warthog) evaluated in multiple in two on-farm locations in 2020 (i.e. Illinois and Wisconsin) and 2021 (i.e. Wisconsin).

	Ridgeway, WI		Freeville, NY
Genotype	2020	2021	2020
212.08	-	<0.3	-
140.05	1.2	1.1	-
174.01	0.6	-	0.8
202.07	-	<0.3	-
47.04	0.6	0.6	0.5
260.06	0.6	0.4	0.7
Warthog	0.7	1.3	0.8

Chapter 3. Genomic prediction for targeted populations of environments in oat (*Avena sativa* L.)

This chapter was published in Crop & Pasture Science on 30 April 2024

Sandro Pablo, Bhatta Madhav, Bower Alisha, Carlson Sarah, Jannink Jean-Luc, Waring David J., Birkett Clay, Smith Kevin, Wiersma Jochum, Caffe Melanie, Kleinjan Jonathan, McMullen Michael S., English Lydia, Gutierrez Lucia (2024) Genomic prediction for targeted populations of environments in oat (Avena sativa). Crop & Pasture Science 75, CP23126. DOI: https://doi.org/10.1071/CP23126

Author contributions

All authors, edited and approved the final manuscript and contributed to the proposal of new hypotheses and discussed results. PS Primary contributor to the written manuscript, including the figures, curated database, perform statistical analysis. MB, Preliminary data set curation and statistical analysis. AB, Iowa Multi environment data contribution, webpage set up, coordination and project execution. SC, Iowa Multi environment data contribution, webpage set up, coordination and project execution. JLJ, T3 database multi environment data contribution, t3 database curation and data intake. DJW, T3 data base multi environment data contribution, t3 database curation and data intake. CB, T3 database multi environment data contribution, t3 database curation and data intake. KS, Minnesota multi environment data contribution. JW, Minnesota multi environment data contribution. MC, South Dakota multi environment data contribution. JK, South Dakota multi environment data contribution. MSMM, North Dakota multi environment data contribution. LE, Iowa Multi environment data contribution, coordination, webpage set up, and project execution. LG Wisconsin Multi environment data contribution, designed the study and hypothesis, and wrote the manuscript.

Summary

The rankings of genotypic performance may change from one environment to the next creating what is known as genotype by environment interaction (GEI). Genomic prediction is not exempt from the impact of GEI, and therefore, the identification of superior genotypes becomes non-trivial. We compared strategies for predicting the performance of individuals in the presence of GEI. Our results provide useful tools for the optimization of training populations and testing environments for the prediction of individuals in the targeted populations of environments.

Abstract

Context. Long-term multi environment trials (METs) could improve genomic prediction models for plant breeding programs by better representing the target population of environments (TPE). However, METs are generally highly unbalanced as genotypes are routinely dropped from trials after a few years. Furthermore, in the presence of genotype by environment interaction (GEI), the selection of the environments to include in a prediction set becomes critical to represent specific TPEs.

Aims. The goal of this study was to compare strategies for modeling GEI in genomic prediction using large METs from oat breeding programs in the Midwest United States and to develop a variety decision tool for farmers and plant breeders.

Methods. The performance of genotypes in TPEs was predicted using different strategies for handling GEI in genomic prediction models including systematic and/or random GEI components. These strategies were also used to build the variety-decision tool for farmers.

Results. Genomic prediction for unknown genotypes, locations, and years within TPEs had moderate to high predictive ability, accuracy, and reliability. Modeling GEI was beneficial in small but not in large mega-environments. The latest three years were highly predictive of performance in an upcoming year for most years but not for years with unusual weather patterns.

Implications. We have shown the performance of a simple modeling strategy for handling prediction for TPEs deploying large historical datasets.

Key results. High predictive ability, accuracy, and reliability were obtained when large datasets were used in TPEs.

Conclusions. Deploying historical datasets can be accomplished through meaningful delineation and prediction for TPEs.

Keywords: genomic prediction, genomic selection, genotype by environment interaction (GEI), multi-environment trials (METs), genotypic performance, genomic best linear unbiased predictions (GBLUP), unbalanced dataset, targeted populations of environments (TPE).

Introduction

The phenotype of an individual plant is determined by its genotype, the environment, and its genotype by environment interaction (GEI; van Eeuwijk 2006). Multi-environment trials (METs) are therefore routinely conducted in plant breeding to capture GEI (Smith *et al.* 2001; Piepho *et al.* 2008) and to estimate (or predict) genotypic performance. Furthermore, plant breeders aim to produce varieties that are best suited for their specific target population of environments (TPE). i.e., any set of fields and future seasons in which the varieties may grow (Atlin *et al.* 2011). Therefore, TPE and genotypic characterization in those environments are critical for performance evaluation and prediction (Yan 2015). The traditional approach to characterize GEI uses a simplification by grouping locations with similar rankings of genotypes, reducing the variability within a group (i.e. no ranking crossover GEI; Yan *et al.* 2007, Crossa and Cornelius 1997; Burgueño *et al.* 2008). The subdivision of environments based on GEI is often referred to as mega-environments (ME, Braun 1997; Gauch and Zobel 1997; Yan 2015). The ME is useful in guiding decision-making for TPE, helping on the management of breeding goals and environments by leveraging relevant information to improve the prediction of the performance of genotypes.

Different strategies have been used to characterize the GEI in general, including mixed models (Piepho 2000; Verbyla *et al.* 2003; Malosetti *et al.* 2004; van Eeuwijk *et al.* 2005; Boer *et al.* 2007; Mathews *et al.* 2008; Lado *et al.* 2016) and the use of environmental covariates in factorial and partial least squares regression (Vargas *et al.* 1999; Rebollo *et al.*, 2023). Mixed models can also deal with the unbalanced structure of most METs where new genotypes are added every year and only the highest performing genotypes, relevant varieties, and long-term checks are retained (Dawson *et al.* 2013). With the widespread availability of genomic data, genomic prediction models (Meuwissen et al. 2001) have become an important tool for predicting the genomic estimated breeding values or genotypic values of individuals for economically relevant traits. It is only natural then that mixed models were extended to account for GEI in genomic prediction by modeling a variance-covariance matrix between environments based on the genotypic correlation among environments (Piepho 1998; Burgueño *et al.* 2012; Malosetti *et al.* 2016; Lado *et al.* 2016; Bhatta *et al*., 2020) or based on environmental covariates (Jarquín *et al.* 2014). Some of these modeling approaches include the use of mixed models such as genomic best-linear unbiased prediction, or GBLUP (Bernardo 1996). In this strategy, a set of genotypes with genomic and phenotypic information is used to train a model that can predict un-phenotyped individuals by borrowing information from relatives through a genomic-estimated relationship matrix (Meuwissen *et al*., 2001; Heffner *et al.* 2009). However, a model is only as good as the data used to generate it, and multiple years and locations within a TPE are required to obtain accurate predictions.

Several studies have tapped into the large historical MET datasets available for many crops (Dawson *et al.* 2013; Lado *et al.* 2016; Ward *et al.* 2019; Bhatta *et al.* 2020), however, a clear strategy for predicting a specific TPE, locations, and year to year variation remains largely needed. The goal of this study was to compare strategies for modeling GEI in genomic prediction using large historical MET data from oat breeding programs in the Midwest of the United States. Specifically, our goals were to evaluate strategies for predicting the performance of genotypes in TPEs; evaluate the predictive ability for random GEI such as genotype by year interactions; and develop the concept for a variety decision tool that farmers and plant breeders can use to aid in the selection of candidate genotypes for their area.

Materials and Methods

Plant material, multi-environment trial data, and genomic information

This study used grain yield from MET historical data of 5,220 oat genotypes that were grown from 1997 to 2021 in 59 locations (Table 1) across 9 states (Illinois, Indiana, Iowa, Michigan, Minnesota, New York, North Dakota, South Dakota, and Wisconsin) in the United States (Figure 1, Figure S1, Table S1). This dataset was compiled from different types of plant breeding trials including preliminary yield trials (PYT, F7) or stage 1, advanced yield trials (AYT, F8) or stage 2, elite yield trials (EYT, F9+) or stage 3, early maturity trials (ET), late maturity trials (LT), several collaborative nurseries including the uniform early oat performance nursery (UEOPN), uniform mid-season oat performance nursery (UMPON), Midwest collaborative nursery (MCON), and a few large experiments including the mega-environmental design (MED), and Metabolomics (META) studies. Genotypes were evaluated either on partially replicated or augmented experimental designs (MCON and META), or with different number of replications in alpha-latices (AYT, PYT, MED), or randomized complete block design (EYT, ET, LT, UEOPN, and UMOPN), depending on the experiment type (available at T3/Oat database, https://oat.triticeaetoolbox.org). This dataset, typical of plant breeding evaluations, is highly unbalanced as not all genotypes were evaluated across all combinations of locations and years.

This study also used the genotypic data from the "GBS POGI" genotype protocol (available at https://oat.triticeaetoolbox.org/breeders_toolbox/protocol/8) from T3/Oat database (available at https://oat.triticeaetoolbox.org) that had 1,189 genotypes with genotyping-by-sequencing derived SNPs that after filtering for a minor allele frequency >0.05 and missing data <20% contained 3,034 high quality SNPs.

Phenotypic data analyses for GGE biplot and ME delineation

Given the high level of sparsity in the dataset and the fact that proper two-step models tend to perform similar to single-step models but are less computationally intensive and do not have as many convergence problems (Piepho *et al.* 2012), a two-step approach was used for obtaining best linear unbiased estimates (BLUEs) of grain yield for each genotype at each location for GEI characterization. In the first step, genotypic BLUEs for grain yield were obtained after controlling for multiple experiments within a location-year combination using linear mixed models with experiment within location-year combination modeled as random effects. In the second step, genotypic BLUEs for grain yield in each location-year combination were modeled using the following linear mixed model in the *asreml-r* (Butler *et al.* 2021) package of the R statistical software (R Development Core Team, 2020):

$$y_{ij} = \mu + g_i + e_j + ge_{ij} + \varepsilon_{ij} \qquad [1]$$

y_{ij} are the genotypic BLUEs for grain yield of the i^{th} genotype in the j^{th} environment (i.e. location-year combination), μ is the overall mean or intercept, g_i is the effect of the ith genotype, e_j is the effect of the jth environment, ge_{ij} is the effect of the interaction of the ith genotype evaluated in the jth environment with $ge \sim \mathrm{N}(0, \boldsymbol{\Sigma})$, and ε_{ij} is the residual error with $\varepsilon \sim \mathrm{N}(0, \boldsymbol{R})$. Where $\boldsymbol{\Sigma}$ is the variance-covariance matrix of the genotype by environment interaction, with $\boldsymbol{\Sigma}_{([\boldsymbol{gxe}]\boldsymbol{x}[\boldsymbol{gxe}])} = \boldsymbol{\Sigma_G} \otimes \boldsymbol{\Sigma_E}$ where $\boldsymbol{\Sigma}_{\boldsymbol{G}(gxg)}$ is the realized additive relationship matrix estimated with the molecular markers using the VanRaden *et al.* (2008) method in the package *AGHmatrix* (Amadeu *et al.* 2016) in R statistical software (R Development Core Team, 2020), and $\boldsymbol{\Sigma}_{\boldsymbol{E}(exe)}$ is the genotypic variance–covariance matrix among environments modeled as a factor analytic of order 1 (FA1). Where $\mathbf{R}_{([\mathrm{gxe}]\mathrm{x}[\mathrm{gxe}])}$ is the variance-covariance matrix of residuals and was modeled as a diagonal matrix with the reciprocals of the standard error of the BLUEs estimates for each genotype in each environment to account for the heterogeneity in the precision of means estimates in the two-step model following Piepho et al. (2012). However, not all genotypes were used in this second step because of the high level of sparsity. Therefore, a subset of 2,000 genotypes that had been evaluated in at least four environments were included in the second step. This ended up accounting for approximately 82% of all data information. Of these, 401 genotypes had phenotypic and genotypic information. Variance components were estimated from a modification of model [1] where genotype by environment was modeled fully by genotype by location by year interaction (see Table S2 for a full description of the model).

Mega-environments were then delineated using GGE biplot analysis from the grain yield BLUEs for each location using the *GGEplot* (Dumble 2022) in the R statistical software (R Development Core Team, 2020). Basically, a singular value decomposition of the genotypic main effect and GEI matrix was conducted (Yan *et al.* 2000; Yan and Tinker 2006) and then it was graphically represented in a biplot (Gabriel 1971; Yan and Tinker 2006). A GGE2 biplot was used for this representation. Winning genotypes were then used to delineate MEs from the GGE2 model (Yan and Tinker 2006). For this analysis a modification of model [1] was used (see Table S2 for details).

Genomic prediction models

Genomic predictions were accomplished using small modifications from the general GBLUP model described in model [1] with a full description of all the prediction strategies fully described in Table S2. When no explicit GEI is modeled, the general model was denoted $\mathrm{GBLUP_M}$ following the nomenclature of Lado et al. (2016). On the other hand, when genotype by environment interaction (ge_{ij}) was modeled with a with $\boldsymbol{\Sigma}_{([\boldsymbol{gxe}]\boldsymbol{x}[\boldsymbol{gxe}])} = \boldsymbol{\Sigma}_{\boldsymbol{G}} \otimes \boldsymbol{\Sigma}_{\boldsymbol{E}}$ with some type of variance-covariance structure for $\boldsymbol{\Sigma}_{\boldsymbol{E}}$, the model was denoted as $\mathrm{GBLUP_{GxE}}$ following the Burgueño *et al.*

(2012) model and the nomenclature of Lado *et al.* (2016). The $\text{GBLUP}_{\text{GxE}}$ assumes known environments where the variance-covariance relationship matrix can be estimated.

Genomic prediction cross-validation schemes

Several cross-validation schemes were implemented to address different prediction goals and scenarios. In all the cross-validation scenarios, a portion of the phenotypic dataset is masked, the remaining data is used to predict their performance, and the predictive ability, or correlation between the predicted and observed values are estimated. One of the interesting prediction scenarios is the prediction of the performance of genotypes that have not yet been evaluated in any environment, Burgueño *et al.* (2012) called this strategy a CV1. We used the CV1 strategy with a 10-fold cross-validation scheme. Each iteration of the 10-fold cross-validation consists of 1) randomly splitting the dataset in 10 subsets, 2) using 9 out of 10 of the subsets to predict the remaining subset, 3) repeating step 2 multiple times for predicting the remaining 9 subsets, and 4) estimating the correlation between observed and predicted values for the full set. The CV1 strategy was used for an overall prediction of the dataset and for predicting the performance in each ME using 100 iterations.

Another prediction scenario consists of the leave-one-environment-out prediction strategy that Jarquin *et al.* (2017) called CV0. In this case, all the phenotypic data for an environment at a time is masked and data from the remaining environments is used for the prediction of the new environment. We evaluated six different scenarios with the CV0 approach. First, within MEs the GBLUP_{M} model was used to predict the new or masked location ($\text{CV0}_{\text{LOC|M}}$). With this strategy, we assume that GEI within MEs is very low and that locations can be properly grouped in MEs even before new performance data is available. Second, within MEs the $\text{GBLUP}_{\text{GxE}}$ model was used to predict the new or masked locations ($\text{CV0}_{\text{LOC|GEI}}$). This situation assumes that the GEI within MEs, although small, is known and can be used to model the variance-covariance structure. Our next scenarios consisted in the leave-one-year-out strategy within each ME. For these scenarios, we used the GBLUP$_M$ model assuming unknown genotype by year relationships. We then either predicted each year within a location (CV0$_{Y|LOC}$), or each year within a ME (CV0$_{Y|ME}$). The predictions by location were performed only for the three largest locations in each ME (i.e., Minot ND, Wayland NY, and West Lafayette IN for ME1; Crookston MN, Ithaca NY, and Morris MN for ME2; and Brookings SD, Madison WI, and Urbana IL for ME3; see Table S1). Using only the largest locations may create an upward bias in predictive ability due to the population size effect. On the other hand, in the predictions for the whole ME, both genotype by year and genotype by location will affect the predictive ability. Although year-to-year variability is random and years cannot be reproduced, an understanding of the stability of the predictions from year-to-year can be valuable in the decision process of choosing genotypes. Finally, we used overall predictions (instead of by ME) for the CV0$_{LOC|M}$ and CV0$_{Y|ME}$ scenarios. All these models and strategies are fully described in Table S2 and in all cases, predictive ability was evaluated as the correlation between predicted and observed genotypic values.

A third prediction scenario was the use of different number of previous years to predict the current year. We use a sliding window approach where we consider periods of three, five, ten, fifteen, and twenty-year periods to predict the genotypic performance, moving one year at a time in the range between 1997 to 2021. (i.e. We use 1997, 1998, and 1999 to predict the year 2000). We used a modification of model [1] to account for the year, location, and interactions. A full description of this model is presented in Table S2. The predictive ability was evaluated as the correlation between predicted and observed genotypic values for each year.

For all the genomic predictions with cross-validations, a subset of 401 oat genotypes out of the 5,220 with 3,034 high-quality GBS-derived SNP markers (minor allele frequency >0.05 and missing data <20%) were used. Additionally, predictions for specific MEs were only evaluated for ME1-ME3. Because ME4 consists by a single location and therefore no cross-validation schemes were evaluated for this ME.

Genomic prediction model for variety decision tool

We also predicted the performance of all 1,180 genotypes for which marker information was available for the variety decision tool. Genotypes had a large range of *per se* phenotypic information available in the dataset ranging from no phenotypic evaluation to evaluation in 53 environments. The GBLUP_M model was used for this purpose in the *asreml-r* (Butler *et al.* 2021) package of the R statistical software (R Development Core Team, 2020). The prediction error variance (PEV), accuracy, and reliability were estimated for each genotype following Gezan (2019). The PEV was estimated as the squared standard error of the genotypic effects ($\widehat{g_i}$) as follows:

$$PEV(\widehat{g_i}) = SE(\widehat{g_i})^2 \qquad [2]$$

The accuracy (h) was estimated as the correlation between true and predicted ($COR(g, \hat{g})$) effects as:

$$h = COR(g_i, \widehat{g_i}) = \sqrt{1 - \frac{PEV(\widehat{g_i})}{\widehat{\sigma_g^2}}} \qquad [3]$$

Finally, the reliability (r) was estimated as the squared accuracy as:

$$r = COR(g_i, \widehat{g_i})^2 = 1 - \frac{PEV(\widehat{g_i})}{\widehat{\sigma_g^2}} \qquad [5]$$

The variety decision tool has two goals, to serve farmers and to serve oat breeders in the Midwest US. The identification of superior oat genotypes for each TPE can aid oat breeders in the identification of release candidates for a specific region or in the identification of parents for future crosses. On the other hand, by filtering the dataset to genotypes that have been released (i.e. varieties) and that are available for sale in the Midwest (surveyed in November 2020), the decision tool can aid farmers in the identification of superior varieties for their local environment. Therefore, both the top ten genotypes, as well as the top ten varieties were identified for each TPE. The variety decision tool is hosted in a web server by the Practical Farmers of Iowa at (https://decide.practicalfarmers.org/).

Results

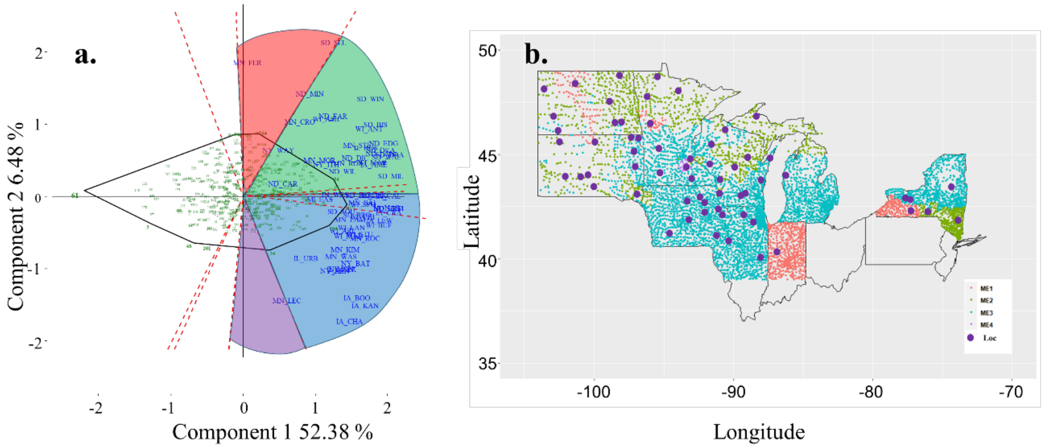
Phenotypic data summary

This study evaluated 5,220 genotypes for grain yield across 59 locations in 25 years resulting in 665 location-year (environment) combinations. Best linear unbiased estimates for grain yield ranged from 74.9 to 798 g m^{-2} with a median of 416 g m^{-2} across locations (Table S1) and substantial GEI was found (Figure 1).

GEI characterization

Mega-environment delineation using GGE biplot analysis of grain yield BLUEs for each one of the 59 locations explained 58.7% of the observed phenotypic variation (Figure 1A). Four MEs were identified based on the winning genotypes on the genotypic polygon (Figure 1A and 1B). ME3a and ME3b had different winning genotypes but similar overall ranking of the genotypes and were merged into a single ME3 for further analysis. The percentage of the variance explained by the genotype by environment interaction was 1.85 times higher than the genotypic variance (Figure 1C and 1D).

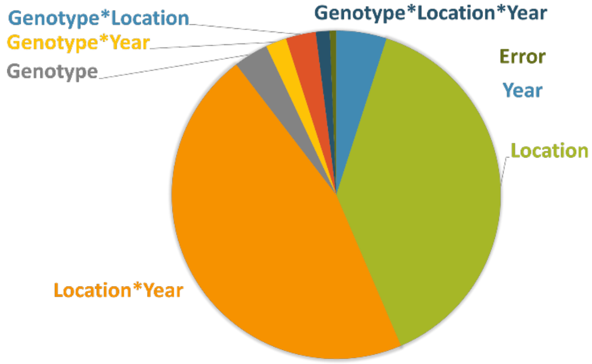
Figure 3.1. Genotype by environment interaction characterization and mega-environment (ME) delineation for a historical dataset for grain yield of oat genotypes evaluated in 59 locations in 9 states in the Midwest United States from 1997-2021. a. GGE biplot for grain yield showing grouping of the 59 locations into four MEs. b. Smoothed map of the four oat MEs for grain yield obtained with the historical dataset of the 59 locations for oat. c. Variance decomposition of the phenotypic variation for grain yield. d. Pie chart of the variance decomposition of the phenotypic variation for grain yield.



c.

Variance component	%
Year	5.0
Location	38.6
Location*Year	46.0
Genotype	3.4
Genotype*Year	2.0
Genotype*Location	2.9
Genotype*Location*Year	1.4
Error	0.6

d.

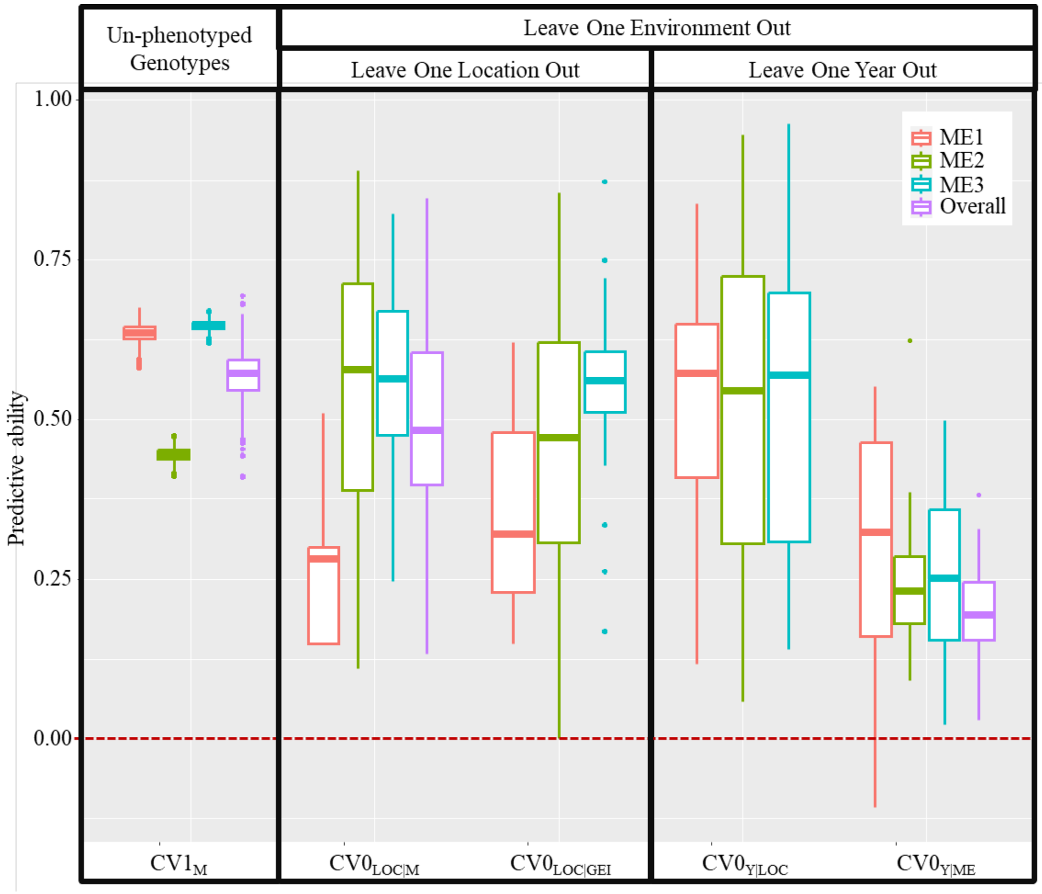


Genomic prediction with different cross-validation strategies

The predictive ability for grain yield for un-phenotyped genotypes in the 10-fold CV1 strategy within MEs ranged from 0.45 (ME1) to 0.65 (ME3) with an overall (all MEs combined) predictive ability of 0.57 (Figure 2, Table S3). When using a CV0 for the leave-one-location-out within ME with the $\mathrm{GBLUP_M}$ modeling approach ($\mathrm{CV0_{LOC|ME}}$), the ME2 and ME3 had the highest predictive ability with an average of 0.55 (range: 0.20 to 0.89 for ME2, and range: 0.16 to 0.82 for ME3, Figure 2), followed by ME1 with an average of 0.28 (range: 0.15 to 0.51). When using the $\mathrm{GBLUP_{GxE}}$ modeling approach on a CV0 for the leave-one-location-out ($\mathrm{CV0_{LOC|GEI}}$), ME3 had the highest predictive ability with an average of 0.55 (range: 0.14 to 0.87, Figure 2), followed by ME2 with an average of 0.47 (range: 0.23 to 0.85) and ME1 with an average of 0.36 (range: 0.19 to 0.32).

The predictive ability for the CV0 for the leave-one-year-out model was 0.29 for ME1 (range: 0.20 to 0.55, Figure 2), 0.29 for ME2 (range: 0.10 to 0.62), and 0.25 for ME3 (range: 0.12 to 0.50) when predicting years within ME ($\mathrm{CV0_{Y|ME}}$, Figure 2). When predicting for locations within MEs ($\mathrm{CV0_{Y|LOC}}$) we obtained higher predictive values with a wide range of variation i.e. 0.52 (0.10 to 0.83 for ME1, 0.51 (0.05 to 0.95) for ME2, and 0.53 (0.20 to 0.96 for ME3).

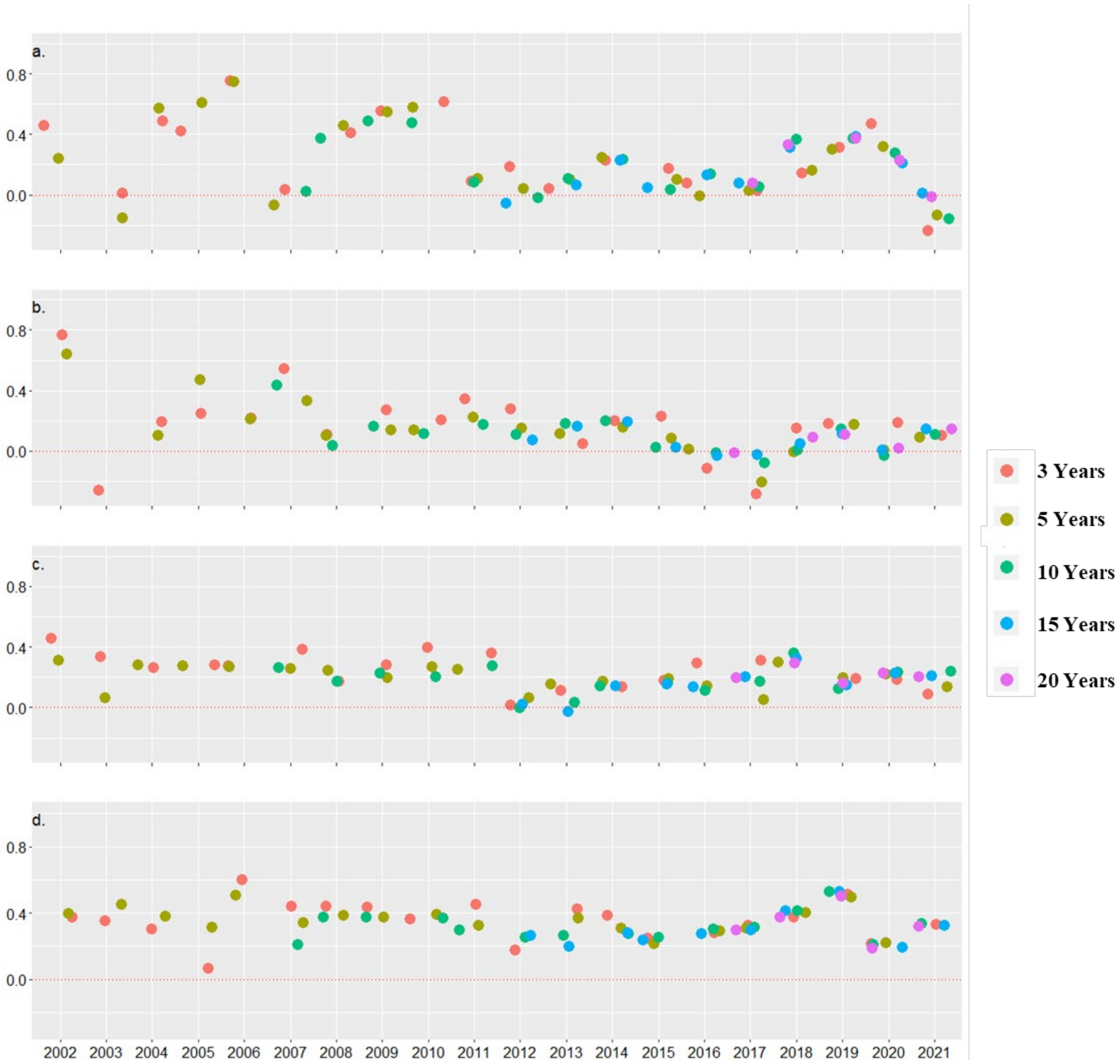
Figure 3.2. Predictive ability within and across mega-environments (ME) for different cross validation strategies for an oat dataset consisting of genotypes evaluated in 59 locations in the Midwest United States over 25 years. The **CV1** used a 10-fold cross-validation strategy for predicting the performance of genotypes that have not been evaluated in any environment within a ME (ME1-ME3) or overall. The **$CV0_{LOC|M}$** used a leave-one-location-out approach with an average GBLUP modeling approach ($GBLUP_M$) within each ME. The **$CV0_{LOC|GEI}$** used a leave-one-location-out approach with a GEI GBLUP modeling approach ($GBLUP_{GxE}$) within each ME. The **$CV0_{Y|LOC}$** used a leave-one-year-out approach with a $GBLUP_M$ modeling approach within each location. The **$CV0_{Y|ME}$** used a leave-one-year-out approach with a $GBLUP_M$ modeling approach within each ME. The predictive ability was estimated as the correlation between predicted and observed values.



Non-repeatable GEI and years

When using the previous years to predict the next year, predictive abilities for using 3, 5, 10, 15, and 20 years were similar in most cases. The overall predictive ability using the most recent three years was highest in 45% of the years and within 20% of the highest predictive ability in 20% of the years (Figure 3). The three most recent years were highest or within the highest predictive ability in 75% (65% highest, 10% not-different) of the years for ME1, 75% (50% highest, 25% not-different) of the years for ME2, and in 80% (40% highest, 40% not-different) of the years for ME3. The predictive ability using more than three years tends to be better when the general predictive ability is low such as in 2005, 2013, and 2021, possibly associated with an unusual weather or stress year.

Figure 3.3. Predictive ability using 3, 5, 10, 15, or 20 previous years to predict the current year across (overall) or within mega-environments (ME1, ME2, or ME3) for an oat dataset where genotypes were evaluated in 59 locations in the Midwest United States over 25 years. a. Overall predictions b. predictions for ME1, c. predictions for ME2, d. predictions for ME3.



Genomic prediction for the variety decision tool

While the accuracy and reliability of released varieties was higher than the advanced breeding lines or general genotypes, the actual predicted values were lower (Table 1 and Table 2). Neither of the 10 genotypes differs more than 2% from the top yielding genotype in the overall and ME2 predictions of genotypes (Table 1). While the six or four first genotypes are within the 2% of the top performing genotypes in ME1 and ME2 respectively. On the other hand, there were six varieties with less than 2% percent difference from the top performing variety in the overall predictions, while there were four in ME1, and two in ME2 and ME3 (Table 2).

Table 3.1. List of top ten ranked genotypes from a genomic prediction model for grain yield (g m^{-2}) using all available information from an oat dataset consisting of genotypes evaluated in 59 locations in 9 states of the Midwest United States over 25 years. The genotype name, ranking, predicted value for grain yield (g m^{-2}), predictor error variance (PEV), accuracy (h), and reliability (r) are shown for each genotype. **a.** Predictions across all environments or Overall model, **b.** Predictions for ME1, **c.** Predictions for ME2, **d.** Predictions for ME3.

a.	Genotype	Ranking	BLUP ($g*m^{-2}$)	P.E.V.	Accuracy (h)	reliability (r)
Overall	X10627.3	1	531	491	0.69	0.47
	SD150012	2	531	200	0.89	0.78
	SD150081	3	528	263	0.85	0.72
	MN11139	4	526	146	0.92	0.84
	MN11140	5	526	152	0.91	0.84
	SD120640	6	525	170	0.90	0.82
	X10679.1	7	524	336	0.80	0.64
	SD150112	8	524	308	0.82	0.67
	MN11231	9	523	165	0.91	0.82
	MNBT1021.1	10	523	156	0.91	0.83

b.	Genotype	Ranking	BLUP ($g*m^{-2}$)	P.E.V.	Accuracy (h)	reliability (r)
ME1	SD111922	1	554	178	0.90	0.81
	SD111946	2	553	178	0.90	0.81
	SD111931	3	552	178	0.90	0.81
	SD111939	4	552	176	0.90	0.81
	SD081108	5	550	172	0.90	0.81
	HAYDEN	6	549	175	0.90	0.81
	SD111753	7	535	245	0.86	0.74
	SD111736	8	534	245	0.86	0.74
	NATTY	9	530	246	0.86	0.74
	SD081107	10	530	230	0.87	0.75

c.	Genotype	Ranking	BLUP ($g*m^{-2}$)	P.E.V.	Accuracy (h)	reliability (r)
ME2	SD081577	1	512	226	0.87	0.76
	SD081563	2	509	225	0.87	0.76
	MN05119	3	508	224	0.87	0.76
	MN11211	4	507	381	0.77	0.59
	MN10209	5	506	289	0.83	0.69
	OA1250.2	6	506	352	0.79	0.62
	MN06108	7	503	442	0.72	0.52
	MN07204	8	503	292	0.83	0.69
	X10199.6	9	502	454	0.72	0.51
	SD081949	10	502	243	0.86	0.74

d.	Genotype	Ranking	BLUP ($g*m^{-2}$)	P.E.V.	Accuracy (h)	reliability (r)
ME3	SD120638	1	488	221	0.87	0.76
	SD120640	2	482	223	0.87	0.76
	IL10.9867	3	480	228	0.87	0.75
	SD150112	4	478	323	0.81	0.65
	SD150012	5	477	247	0.86	0.73
	SD150081	6	476	331	0.80	0.64
	X10627.3	7	474	603	0.59	0.35
	SD141130	8	474	296	0.83	0.68
	MN11140	9	474	185	0.89	0.80
	X10679.1	10	474	371	0.78	0.60

Table 3.2. List of top ten ranked commercially available varieties (as of November 2020) from a genomic prediction model for grain yield (g m^{-2}) using all available information from an oat dataset consisting of genotypes evaluated in 59 locations in 9 states of the Midwest United States over 25 years. The genotype name, ranking, predicted value for grain yield (g m^{-2}), predictor error variance (PEV), accuracy (h), and reliability (r) are shown for each genotype. **a.** Predictions across all environments or Overall model, **b.** Predictions for ME1, **c.** Predictions for ME2, **d.** Predictions for ME3.

a.	Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
Overall	DEON	1	488	68	0.96	0.93
	CDC MINSTREL	2	486	177	0.90	0.81
	ESKER2020	3	483	89	0.95	0.90
	AC PINNACLE	4	481	167	0.91	0.82
	HAYDEN	5	479	62	0.97	0.93
	NATTY	6	479	78	0.96	0.92
	AAC ROSKENS	7	477	219	0.87	0.76
	RON	8	469	83	0.95	0.91
	GOLIATH	9	468	79	0.96	0.91
	NEWBURG	10	467	61	0.97	0.93

b.	Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
ME1	AAC.ROSKENS	1	501	313	0.81	0.66
	CDC.MINSTREL	2	499	187	0.89	0.80
	AC PINNACLE	3	499	181	0.90	0.81
	AAC.BULLET	4	498	333	0.80	0.64
	NEWBURG	5	486	121	0.93	0.87
	SOURIS	6	483	120	0.93	0.87
	DEON	7	482	168	0.90	0.82
	AAC.OAKLIN	8	482	306	0.82	0.67
	AAC.ALMONTE	9	481	305	0.82	0.67
	HAYDEN	10	480	117	0.93	0.87

c.	Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
ME2	HAYDEN	1	549	175	0.90	0.81
	NATTY	2	530	246	0.86	0.74
	EXCEL	3	515	253	0.85	0.73
	HORSEPOWER	4	512	222	0.87	0.76
	SOURIS	5	507	152	0.91	0.84
	NEWBURG	6	506	183	0.90	0.80
	GOLIATH	7	504	201	0.89	0.78
	SHELBY427	8	502	228	0.87	0.75
	ROCKFORD	9	502	258	0.85	0.72
	CORRAL	10	501	361	0.78	0.61

d.	Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
ME3	ESKER2020	1	436	116	0.94	0.88
	DEON	2	430	82	0.96	0.91
	NATTY	3	425	98	0.95	0.89
	HAYDEN	4	415	76	0.96	0.92
	GOLIATH	5	415	97	0.95	0.90
	RON	6	413	106	0.94	0.89
	CDC MINSTREL	7	408	201	0.89	0.78
	BETAGENE	8	407	96	0.95	0.90
	SHELBY427	9	404	66	0.96	0.93
	EXCEL	10	404	102	0.94	0.89

Discussion

Genomic prediction success and strategies

In oat, genomic prediction models have been used to assess prediction accuracy of genomic prediction models with the potential to improve disease resistance (Haikka *et al.* 2020a), seed quality (Campbell *et al.* 2021; Hu *et al.* 2021), grain yield (Haikka *et al.* 2020b), nutritional compounds and other metabolites (Brzozowski *et al*., 2021; 2022 2023), as well as other quantitative traits (Asoro *et al.* 2011) and testing optimization (Gonzalez-Barrios *et al.* 2019). Our study has looked at other aspects of genomic prediction including the prediction for specific TPE, and we have achieved relatively high predictive abilities of around 0.45 to 0.65 for predicting new genotypes and predictive abilities of 0.15 to 0.89 for new environments for a trait as complex as grain yield.

Deploying big data in genomic selection for predicting for the TPE

Genotype by environment interactions create all sorts of challenges for breeding programs, some of which are easier to deal with than others (Piepho *et al*. 2008). Under GEI, the efficiency of the prediction relies on the representativity of the tested environments and how they are used for the construction of the training population (Haile *et al.* 2020; Atanda *et al.* 2022; Jarquin *et al*. 2022). Genomic selection can take advantage of large historical datasets by connecting the phenotypic information from different genotypes through borrowing information from relatives and thus controlling the sparsity (Bernardo 2010; Dawson *et al*. 2013; Lado *et al*. 2016; Monteverde *et al*. 2018) and producing higher predictive ability (Endelman *et al*. 2014; Gonzalez-Barrios *et al*. 2019; Hoefler *et al*. 2020; Atanda *et al*. 2021). Therefore, the use of big data consisting of large historical datasets where individual genotypes, or their relatives have been evaluated, can substantially improve the identification of superior individuals that will perform better in the TPE.

Large MET historical datasets have their share of constraints (i.e., sparsity and breeding structure), that need to be addressed in order to make efficient use of the available databases. How to best handle the level of sparsity is still an ongoing area of research given that genotypes are not missing at random in the dataset because poorer-performing genotypes are dropped every year (Aguate *et al.* 2019; Hartung and Piepho 2021; Wang *et al*. 2023). Other relevant limitations for the use of large datasets are due to the complexity of the data, the statistical models, and the requirement of high computational power (Atanda *et al*. 2021). This study utilized a highly unbalanced dataset from several different MET across 25 years to investigate the use of a mixed model approach in characterizing GEI, identifying MEs associated with the TPEs, and determining the potential of genomic selection of grain yield utilizing this dataset. Overall, the results obtained in this study show the high predictive abilities for large-scale historical data in genomic prediction.

The management of the GEI and information leverage is determined by the size and overlap of the ME and TPE areas. MEs can be defined based on the breeding program necessities adopting different spatial scales, (i.e. at a regional scale, Gonzales Barrios *et al*. 2019, country-wide scale, Crespo-Herrera *et al*. 2021, or worldwide, Braun *et al*. 1997). We used defined MEs by providing a broad idea of the GEI in the Midwest region of the United States. Even though TPE might be smaller or larger than the MEs, using the ME definition can be beneficial. For example, the use of information from other locations can still be beneficial in the prediction for TPEs that are smaller than the MEs by increasing the number of evaluations in environments with low GEI to the TPE. On the other hand, in the case that the TPEs are larger than the MEs or go beyond the limits of a single ME, the use ME units will allow for example to tailor the MET evaluations to account for the variation that exist between ME in the TPE and the breeder will need to decide on the release strategies for this case. Furthermore, this work is focused on the use of the groups of locations defined as a TPE instead of predicting locations per se. The shift from a specific location to a TPE is grounded in the fact that farmer fields performance will rarely coincide with the evaluations in research stations that breeders carry out, and therefore, a focus on predicting individual locations might not be useful for the goal of the breeding program unless it is used as an intermediate step in the prediction process. We were able to predict all MEs with a high level of predictive ability. Furthermore, we tested a situation where GEI within MEs was not modeled ($CV0_{LOC|M}$) and a situation where GEI within ME was modeled using a variance-covariance matrix that accounts for genetic correlations between locations ($CV0_{LOC|GEI}$). The use of the known correlation structure did not improve the predictive ability reported in other studies (Lado *et al*. 2016). There could be multiple reasons for this, including the level of GEI within the ME, an artifact due to the population size, or the level of unbalance in the population (Lado *et al.* 2018; Berro *et al.* 2019; Bassi *et al.* 2016), or just the fact that specific location to location genotype by environment interaction become less relevant when the number of locations is very large. Even though the mean predictive ability did not improve with modeling the GEI, the variability in the predictive ability for each location decreased with the $GBLUP_{GxE}$ model. Arguments can be made that the use of the correlation between environments cannot be known for new locations or for farmers' fields, and therefore they should not be used in the prediction models. However, if the focus remains on the overall predictions within MEs, modeling the correlation is just another modeling tool that can be deployed, especially when large historical datasets are used where the estimation of those correlations might be more relevant. Another opportunity to improve predictions when no previous information is available is by incorporating environmental covariates explaining the genotypic performance in the ME (van Eeuwijk, *et al.* 2016; Bustos-Korts *et al.* 2016, Monteverde *et al.*

2019; Neyhart *et al*. 2022; Rebollo *et al*. 2023). An advantage of the use of environmental covariates is that they could help to fine-tune the limits and boundaries of the ME by the definition of growing zones and modeling the crop growth (Heslot *et al.* 2013; Bustos-Korts *et al.* 2016; Monteverde *et al.* 2019; Bustos *et al*. 2021). Although we did not use environmental covariates in this study, high to medium predictive abilities were obtained. However, we believe that the future use of environmental covariates could improve the predictive ability of these models and help in the delineation of MEs.

Even though we obtained high levels of predictive ability for all MEs, the performance of genotypes in some MEs was easier to predict than in others. For example, ME1 had consistently low predictive ability (i.e. lower than 0.4), this was expected because the number of locations within the ME was only five. This ME was the only one that benefited from modeling genetic correlations among locations using the $\mathrm{GBLUP}_{\mathrm{GxE}}$ model, where predictive ability increased from 0.27 to 0.36. This is also an artifact of the low number of locations used in the study, where the specific pairwise correlations have a larger impact than in the case of the other MEs with dozens of locations where specific location to location interactions are overall less relevant.

Non-repeatable GEI and years

Year to year variation is known to explain an important part of the GEI variation in MET trials, and more importantly for being nonreproducible. Furthermore, because many plant breeding programs rely on the evaluation of individuals in a small sample of years, genotypes with positive genotype by year interactions in testing years can be advanced when they wouldn’t necessarily have a superior performance on average years for the region. Therefore, datasets with more years may reduce the bias associated with the unbalance in MET, improve the variance component estimation (Aguate *et al.* 2019; Hartung and Piepho 2021), and improve the actual genotypic predictions (Oakley *et al.* 2016). Our results suggest that increasing the number of years beyond the last three years will not improve the predictive ability of the upcoming unknown year. This could be partially explained by the higher connectivity in the genotypic pool between consecutive number years. Some years had a low predictive ability regardless of the set of years used in the prediction model. These are years with unusual weather patterns and some abiotic stress. The odd performance in those years will be very difficult to predict regardless of the amount of previous information we use or have available. The ME1 had 20% of years with a predictive ability smaller than 0.1, while ME3 had only 5% of years with low predictive ability. This can be partially explained by the populational size and the frequency of climatic stresses present in the ME.

We also evaluated the predictive ability for specific years within the ME ($CV0_{Y|ME}$) and for specific years within locations ($CV0_{Y|LOC}$). The predictions for specific locations were higher (0.20 to 0.90) than the predictions for the whole ME (0.10 to 0.60). Multiple factors can explain these findings. On the one hand, the locations used to test the year-to-year variability had more data points and specifically more years than the average location in any given ME. Additionally, predictions for specific locations might not be as relevant as predictions for the ME as previously discussed. Finally, predictions for the ME include both genotype by year and genotype by location interactions, and although in many cases genotype by year interactions are larger than genotype by location, in our case, they were of the same magnitude. Although in a breeding program we are not able to replicate a year nor predict the type of year that will come next, the evaluation of the predictions are relevant for understanding the stability in the predictions that can be expected from year to year in real future evaluations.

Variety decision tools and prospects

To the best of our knowledge, ours is the first study to deliver a variety decision tool for farmers in the Midwestern United States based on the farms location. There is a limited list of genotypes that are relevant to farmers and those are genotypes that have been released as a cultivar and that are currently commercially available. These genotypes are usually connected in METs because they were tested as advanced breeding lines and later most probably continued to be tested in regional trials. However, because there is not a unified variety testing system in the United States, the list of varieties to be tested in each State is tailored to each State and therefore different. The reports are then usually provided on a location-by-location basis or in some cases a State-wise recommendation is provided. How years are dealt in those reports is also State-specific but most of the States do not report any information beyond three years of data. This reporting system creates difficulties at several levels. For example, it is hard for farmers to decide on which data to rely on for deciding on the best performing varieties for their specific farm. This is especially challenging for farmers close to State borders or far from the specific sparse testing locations for their State. Furthermore, there is no reason for limiting the flow of information across State borders. The use of MEs to delineate TPEs and the grouping of locations improves the accuracy of predictions and allows the distribution of resources in a more efficient way than by location State evaluation sheets. This helps to create a more efficient delivery of higher quality information to farmers. Finally, being able to include the information from advanced breeding lines is also key to improving the prediction of the future released varieties. This can be seen for example in the higher accuracy and reliability values for varieties than for advanced breeding genotypes that may not have as many relatives in the dataset. Additionally, this is also reflected in the mean differences between experimental genotypes and varieties, where varieties are more than two percent apart. This decision tool is being created as a first exploratory strategy. It represents an improvement over State reports, but further development would improve the predictions. For example, a very simple strategy was used to delineate the borders of the MEs, but environmental covariates could be used to further refine these delineations following the work of Rebollo *et al*. (2023) and Bustos *et al*. (2021). Additionally, environmental covariates could be used to improve the actual predictions. In summary, the strategy used in this study will assist in developing a tool for selecting the top-ranking genotypes in specific locations using a zip code-based system in the Midwest region from the deployment of the genomic selection model in each TPE and similar strategies could be expanded to other crops in any region around the world.

Data availability.

All data are publicly available in T3/oats database and data used in this study were from May 1st, 2023.

Conflict of Interest

The authors declare no conflict of interest.

Declaration of funding

This project was partially funded through multiple competitive Federal awards from the from the United States Department of Agriculture, National Institute of Food and Agriculture, including the USDA-SARE 2018-LNC18-400, by the Hatch Act Formula Fund WISO1984 and WIS03002, the USDA-AFRI-NIFA-OREI 2020-51300-32379 and 2023-51300-40916, and the USDA-AFRI-NIFA 2022-68013-36439 (WheatCAP).

Acknowledgements

We would like to thank the USDA-ARS small grains genotyping lab at Fargo, North Dakota for its genotyping service. Additionally, the authors would like to thank two anonymous reviewers and the editor for thoughtful comments that improved the manuscript.

Annex 2

Supplemental Figures and Tables

Supplementary Table 3.1. Summary statistics and characterization of the 59 locations where oat performance was evaluated in 9 states of the Midwest of the United States from 1997 to 2021. Location, State, environmental code (Env), latitude (Lat), longitude (Lon), grain yield mean (g m^{-2}), the number of data-points in the dataset (N), mega-environment classification (ME, four MEs were obtained: me1 to me4), number of years (#year), first (first year), and last (last year) year in the dataset.

State	Location	Env	Lat	Lon	grain yield (g m^{-2})	N	ME	#years	first year	last year
IA	Ames	IA_AME	42.03	-93.61	501	905	ME2	12	1997	2008
IA	Boone	IA_BON	42.06	93.89	331	86	ME3	3	2018	2020
IA	Charles City	IA_CHA	43.07	-92.67	261	62	ME3	4	2016	2019
IA	Crawfordsville	IA_CRA	41.21	-91.53	437	235	ME3	7	1997	2003
IA	Kanawha	IA_KAN	42.93	-93.81	316	132	ME3	6	2015	2020
IA	Lewis	IA_LEW	41.31	-95.08	418	236	ME3	7	1997	2003
IA	Nashua	IA_NAS	42.95	-92.53	446	369	ME3	13	1997	2020
IL	Dekalb	IL_DKL	41.93	-88.75	391	242	ME3	13	2002	2015
IL	Monmouth	IL_MON	40.92	-90.64	452	434	ME3	17	2004	2021
IL	Urbana	IL_URB	40.11	-88.21	421	1698	ME3	25	1997	2021
IN	West Lafayette	IN_WES	40.46	-86.94	496	995	ME1	17	1997	2013
MI	Calumet	IA_CAL	42.96	-95.57	397	203	ME2	6	1997	2003
MI	East Lansing	MI_EAS	42.75	-84.47	360	796	ME3	16	1997	2013
MN	Crookston	MN_CRO	47.76	-96.58	579	1583	ME2	8	2014	2021

MN	Fergus Falls	MN_FER	46.28	-96.07	479	163	ME1	5	2021	2016
MN	Kimball	MN_KIM	45.40	-94.32	543	51	ME3	2	2016	2019
MN	Lamberton	MN_LAM	44.23	-95.28	438	1635	ME3	6	2016	2021
MN	Lecenter	MN_LEC	44.39	-93.73	567	188	ME4	6	2016	2021
MN	Morris	MN_MOR	45.58	-95.91	448	1152	ME2	19	1997	2020
MN	Rochester	MN_ROCH	44.01	-92.47	530	90	ME3	2	2020	2021
MN	Roseau	MN_ROS	48.82	-95.75	617	244	ME2	6	2016	2021
MN	Rosemount	MN_RSM	44.73	-93.13	309	838	ME2	15	1997	2011
MN	Saint Paul	MN_SAI	44.96	-93.08	395	256	ME3	3	2012	2020
MN	Stephen	MN_STE	48.48	-96.83	562	188	ME2	6	2016	2021
MN	Waseca	MN_WAS	44.07	-93.51	355	632	ME3	20	1999	2021
ND	Carrington	ND_CAR	47.45	-99.09	426	708	ME2	20	1998	2020
ND	Dickinson	ND_DIC	46.82	-102.66	432	282	ME2	15	2005	2020
ND	Edgeley	ND_EDG	46.36	-98.72	477	86	ME2	5	2014	2018
ND	Fargo	ND_FAR	46.90	-96.80	551	1066	ME2	22	1997	2020
ND	Hettinger	ND_HET	46.04	-102.60	380	282	ME3	15	2005	2020
ND	Lamoure	ND_LAM	46.36	98.29	425	28	ME3	2	2012	2013
ND	Langdon	ND_LNG	48.81	-98.31	575	274	ME2	15	2005	2020
ND	Minot	ND_MIN	48.23	-101.27	454	958	ME1	23	1997	2020
ND	Williston	ND_WIL	48.17	-103.61	285	223	ME2	12	2005	2020
NY	Batavia	NY_BAT	43.00	-78.19	253	80	ME3	4	2015	2018
NY	Henrietta	NY_HEN	43.06	77.61	199	53	ME3	2	2009	2011
NY	Ithaca	NY_ITH	42.45	-76.50	322	1257	ME2	23	1997	2020
NY	Sprout Brook	NY_SPR	42.85	74.69	255	38	ME3	2	2013	2014

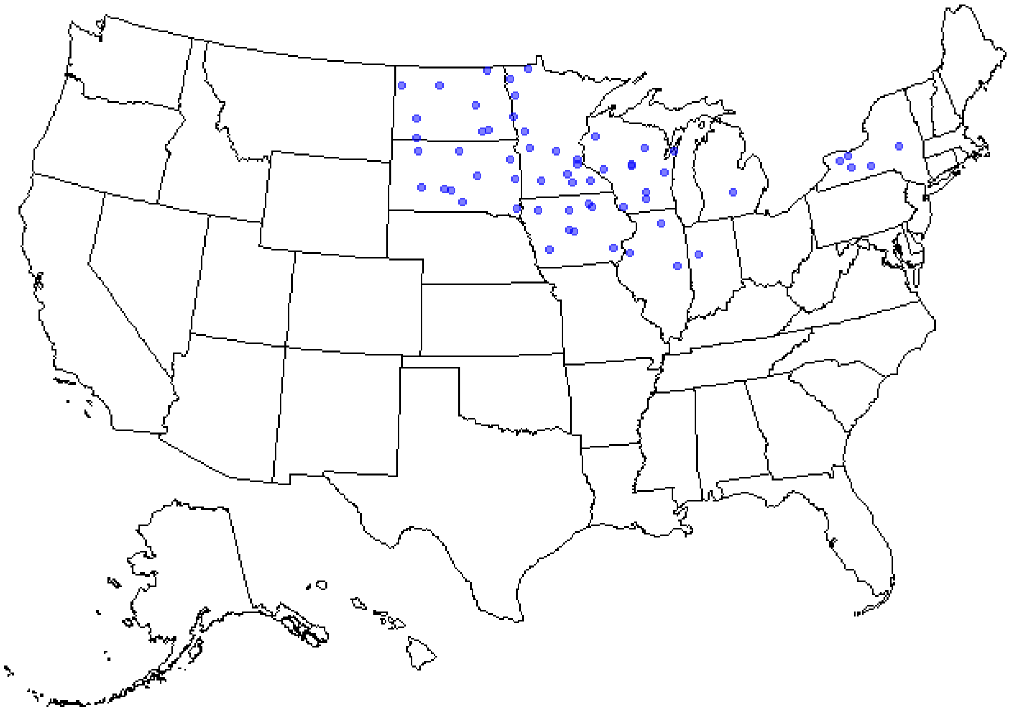
NY	Wayland	NY_WAY	42.56	-77.60	248	151	ME1	7	2011	2018
SD	Beresford	SD_BER	43.07	-96.79	379	910	ME3	21	1999	2020
SD	Bison	SD_BIS	45.52	-102.46	196	28	ME2	2	2014	2015
SD	Brookings	SD_BRO	44.31	-96.79	415	1667	ME3	23	1997	2020
SD	Draper	SD_DRP	43.86	-100.53	308	29	ME2	2	2016	2017
SD	Miller	SD_MIL	44.52	-98.99	375	86	ME2	6	2014	2020
SD	Okaton	SD_OKAT	43.96	-100.92	317	42	ME2	3	2014	2020
SD	Selby	SD_SELBY	45.51	-100.03	503	71	ME1	6	2014	2021
SD	South Shore	SD_SOU	45.13	-97.03	483	1037	ME3	21	1997	2020
SD	Wall	SD_WAL	43.99	-102.24	299	72	ME2	5	2014	2020
SD	Winner	SD_WIN	43.38	-99.86	297	85	ME2	6	2015	2020
WI	Antigo	WI_ANT	45.12	-89.13	267	84	ME2	4	2018	2021
WI	Arlington	WI_ARL	43.32	-89.37	456	1340	ME3	22	1998	2021
WI	Ashfield	WI_ASH	44.50	-90.00	366	106	ME2	4	1998	2001
WI	Buffalo County	WI_BUF	44.47	-91.68	346	78	ME3	4	2018	2020
WI	Chilton	WI_CHI	44.03	-88.18	322	503	ME3	18	1997	2019
WI	Lancaster	WI_LAN	42.85	-90.70	362	541	ME3	19	1997	2021
WI	Madison	WI_MAD	43.07	-89.42	403	3242	ME3	24	1997	2021
WI	Marshfield	WI_MAR	44.47	-90.02	358	536	ME2	18	1998	2017
WI	Spooner	WI_SPO	45.85	-91.94	268	607	ME3	21	1997	2020
WI	Sturgeon Bay	WI_STU	44.84	-87.38	383	610	ME3	21	1997	2018

Supplementary Table 3.2. General description of genomic prediction models and strategies used in the characterization of a large oat population evaluated in 59 locations over a period of 25 years in the Midwest United States. Models are described as modification from model [1] $y_{ij}=\mu+g_i+e_j+ge_{ij}+\varepsilon_{ij}$, where genotype by environment interaction was modeled either as genotype (G) by location (LOC), year (YEAR), or mega-environment (ME). The variance-covariance matrix for the genotype by environment interaction was modeled in general as the Kronecker product of the realized additive relationship matrix estimated with marker data ($\mathbf{K}_{(gxg)}$) and the genetic variance-covariance among environments ($\mathbf{\Sigma}_{\mathbf{E(exe)}}$), with $\mathbf{\Sigma}_{\mathbf{E}}$ either an independent variance-covariance structure, a factor-analytic (FA(1)), a diagonal matrix (**D**), or a known unstructured (UN*) variance-covariance. The residuals were modeled with a variance-covariance matrix, $\mathbf{R}_{([gxe]x[gxe])}$, as a diagonal matrix with the reciprocals of the square standard error of the BLUEs estimates for each genotype in each environment to account for the heterogeneity in the precision of means estimates in the two-step model following Piepho et al. (2012).

Strategy	**y**	**Fixed**	**GEI**	**Other Components**	**Res**	**Var(ge_{ij})**	$\mathbf{\Sigma_G}$	$\mathbf{\Sigma_E}$	**Var(ε_{ij})**
GEI Characterization									
VC	gxlxa	mu	ge_{ij}=G·LOC·YEAR		ε_{ij}	$\mathbf{I}$			$\mathbf{R}=\mathbf{I}(1/SE)$
GGE	gxl	e_j=LOC	ge_{ij}=**G·LOC**		ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}$=FA(1)	$\mathbf{R}=\mathbf{I}(1/SE)$
Un-phenotyped individuals									
CV1 by ME_m	gxl_m	e_j=LOC	ge_{ij}=G·LOC	**G**	ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}=\mathbf{D}$	$\mathbf{R}=\mathbf{I}(1/SE)$
CV1 Overall	gxl	e_j=ME	ge_{ij}=G·ME	**G**	ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}=\mathbf{D}$	$\mathbf{R}=\mathbf{I}(1/SE)$
Leave one location out									
$CV0_{LOC\|M}$ by ME_m	gxl_m	e_j=LOC	ge_{ij}=**G·LOC**		ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}$=FA(1)	$\mathbf{R}=\mathbf{I}(1/SE)$
$CV0_{LOC\|M}$ Overall	gxl	e_j=ME	ge_{ij}=G·ME	ME·LOC + **G**	ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}$=FA(1)	$\mathbf{R}=\mathbf{I}(1/SE)$
$CV0_{LOC\|GEI}$ by ME_m	gxl_m	e_j=LOC	ge_{ij}=**G·LOC**		ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}$=UN*	$\mathbf{R}$=I(1/SE)
Leave one year out									
$CV0_{Y\|ME}$ by ME_m	gxl_mxa	e_j=ME	ge_{ij}=**G·ME**	YEAR|LOC|ME	ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}$=FA(1)	$\mathbf{R}=\mathbf{I}(1/SE)$
$CV0_{Y\|ME}$ Overall	gxlxa	e_j=ME	ge_{ij}=G·ME	YEAR|LOC|ME + **G**	ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}$=FA(1)	$\mathbf{R}=\mathbf{I}(1/SE)$
$CV0_{Y\|LOC}$ by LOC_{lm}	gxa_{lm}	e_j=YEAR	ge_{ij}=**G·YEAR**		ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}$=FA(1)	$\mathbf{R}=\mathbf{I}(1/SE)$
Predicting years									
YEAR by ME_m	gxl_mxa	e_j=LOC	ge_{ij}=**G·LOC**	YEAR	ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}$=FA(1)	$\mathbf{R}=\mathbf{I}(1/SE)$
YEAR Overall	gxlxa	e_j=ME	ge_{ij}=**G·ME**	ME·LOC + YEAR	ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}=\mathbf{I}$	$\mathbf{R}=\mathbf{I}(1/SE)$
Variety decision tool									
VDT by ME_m	gxl_mxa	e_j=ME	ge_{ij}=**G·ME**	YEAR|LOC|ME	ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}$=FA(1)	$\mathbf{R}=\mathbf{I}(1/SE)$
VDT Overall	gxlxa	e_j=ME	ge_{ij}=G·ME	YEAR|LOC|ME + **G**	ε_{ij}	$\mathbf{\Sigma_G}\otimes\mathbf{\Sigma_E}$	$\mathbf{\Sigma_G}=\sigma_g^2\mathbf{K}$	$\mathbf{\Sigma_E}$=FA(1)	$\mathbf{R}=\mathbf{I}(1/SE)$

Supplementary Table 3.3. Predictive ability within and across mega-environments (ME) for different cross validation strategies. The CV0-loc-GEI used a leave-one-location-out approach with a GEI GBLUP modeling approach ($GBLUP_{GxE}$) within each ME. The CV0-loc-MEAN uses a leave-one-location-out approach with an average GBLUP modeling approach ($GBLUP_{M}$) within each ME. The CV0-year-by-LOC used a leave-one-year-out approach with a $GBLUP_{M}$ modeling approach within each location. The CV0-year-by-ME used a leave-one-year-out approach with a $GBLUP_{M}$ modeling approach within each ME. Finally, CV2 used a 10-fold cross-validation strategy for predicting the performance of genotypes that have been evaluated in some environments but not others within a ME (ME1-ME3) or overall.

	ME1	ME2	ME3	Overall
CV0-Loc-GEI	0.36 (0.19-0.62)	0.47 (0.24-0.86)	0.55 (0.14-0.87)	--
CV0-Loc-Mean	0.27 (0.15-0.51)	0.55 (0.27-0.89)	0.55 (0.15-0.82)	--
CV0-Year-by-Loc	0.52 (0.10-0.83)	0.51 (0.05-0.95)	0.53 (0.20-0.96)	--
CV0-Year-by-ME	0.30 (0.55-0.10)	0.24 (0.62-0.10)	0.25 (0.50-0.02)	--
CV2-Gen-Mean	0.64 (0.02-0.67)	0.45 (0.01-0.47)	0.65 (0.01-0.67)	0.57 (0.41-0.69)
#genotypes	304	393	401	401
#locations	5	23	30	59



Supplementary Figure 3.1. Map of the United States showing the 59 locations where oat performance was evaluated in 9 states of the Midwest of the United States from 1997 to 2021.

Chapter 3 Errata:

After publishing the paper we identified a a series of mistakes in the reporting of the materials and method, Supplementary Table 3.2, and Results Table 3.2.

Materials and Methods:

Line 155 to 158

Erroneous sentence: "Where R([gxe]x[gxe]) is the variance-covariance matrix of residuals and was modeled as a diagonal matrix with the reciprocals of the standard error of the BLUEs estimates for each genotype in each environment to account for the heterogeneity in the precision of means estimates in the two-step model **following Piepho et al. (2012)**."

Corrected sentence *"*Where R([gxe]x[gxe]) is the variance-covariance matrix of residuals and was modeled as a diagonal matrix with the reciprocals of the standard error of the BLUEs estimates for each genotype in each environment to account for the heterogeneity in the precision of means estimates in the two-step model."

Supplementary table Table 3.2.

Erroneous sentence: "The residuals were modeled with a variance-covariance matrix, R([gxe]x[gxe]), as a diagonal matrix with the reciprocals of the standard error of the BLUEs estimates for each genotype in each environment to account for the heterogeneity in the precision of means estimates in the two-step model **following Piepho et al. (2012)**."

Corrected sentence: "The residuals were modeled with a variance-covariance matrix, R([gxe]x[gxe]), as a diagonal matrix with the reciprocals of the standard error of the BLUEs estimates for each genotype in each environment to account for the heterogeneity in the precision of means estimates in the two-step model."

Table 3.2.Tables b and c are mislabel, where ME1 appears ME2, and ME2 appears as Me1

Errouneous: Table 3.2. List of top ten ranked commercially available varieties (as of November 2020) from a genomic prediction model for grain yield (g m-2) using all available information from an oat dataset consisting of genotypes evaluated in 59 locations in 9 states of the Midwest United States over 25 years. The genotype name, ranking, predicted value for grain yield (g m-2), predictor error variance (PEV), accuracy (h), and reliability (r) are shown for each genotype. a. Predictions across all environments or Overall models, b. Predictions for ME1, c. Predictions for ME2, d. Predictions for ME3.

a.

Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
DEON	1	488	68	0.96	0.93
CDC MINSTREL	2	486	177	0.90	0.81
ESKER2020	3	483	89	0.95	0.90
AC PINNACLE	4	481	167	0.91	0.82
HAYDEN	5	479	62	0.97	0.93
NATTY	6	479	78	0.96	0.92
AAC ROSKENS	7	477	219	0.87	0.76
RON	8	469	83	0.95	0.91
GOLIATH	9	468	79	0.96	0.91
NEWBURG	10	467	61	0.97	0.93

Overall

b.

Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
AAC.ROSKENS	1	501	313	0.81	0.66
CDC.MINSTREL	2	499	187	0.89	0.80
AC PINNACLE	3	499	181	0.90	0.81
AAC.BULLET	4	498	333	0.80	0.64
NEWBURG	5	486	121	0.93	0.87
SOURIS	6	483	120	0.93	0.87
DEON	7	482	168	0.90	0.82
AAC.OAKLIN	8	482	306	0.82	0.67
AAC.ALMONTE	9	481	305	0.82	0.67
HAYDEN	10	480	117	0.93	0.87

ME1

c.

Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
HAYDEN	1	549	175	0.90	0.81
NATTY	2	530	246	0.86	0.74
EXCEL	3	515	253	0.85	0.73
HORSEPOWER	4	512	222	0.87	0.76
SOURIS	5	507	152	0.91	0.84
NEWBURG	6	506	183	0.90	0.80
GOLIATH	7	504	201	0.89	0.78
SHELBY427	8	502	228	0.87	0.75
ROCKFORD	9	502	258	0.85	0.72
CORRAL	10	501	361	0.78	0.61

ME2

d.

Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
ESKER2020	1	436	116	0.94	0.88
DEON	2	430	82	0.96	0.91
NATTY	3	425	98	0.95	0.89
HAYDEN	4	415	76	0.96	0.92
GOLIATH	5	415	97	0.95	0.90
RON	6	413	106	0.94	0.89
CDC MINSTREL	7	408	201	0.89	0.78
BETAGENE	8	407	96	0.95	0.90
SHELBY427	9	404	66	0.96	0.93
EXCEL	10	404	102	0.94	0.89

ME3

Corrected: Table 3.2. List of top ten ranked commercially available varieties (as of November 2020) from a genomic prediction model for grain yield (g m^{-2}) using all available information from an oat dataset consisting of genotypes evaluated in 59 locations in 9 states of the Midwest United States over 25 years. The genotype name, ranking, predicted value for grain yield (g m^{-2}), predictor error variance (PEV), accuracy (h), and reliability (r) are shown for each genotype. **a.** Predictions across all environments or Overall model, **b.** Predictions for ME1, **c.** Predictions for ME2, **d.** Predictions for ME3.

a. Overall

Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
DEON	1	488	68	0.96	0.93
CDC MINSTREL	2	486	177	0.90	0.81
ESKER2020	3	483	89	0.95	0.90
AC PINNACLE	4	481	167	0.91	0.82
HAYDEN	5	479	62	0.97	0.93
NATTY	6	479	78	0.96	0.92
AAC ROSKENS	7	477	219	0.87	0.76
RON	8	469	83	0.95	0.91
GOLIATH	9	468	79	0.96	0.91
NEWBURG	10	467	61	0.97	0.93

b. ME1

Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
HAYDEN	1	549	175	0.90	0.81
NATTY	2	530	246	0.86	0.74
EXCEL	3	515	253	0.85	0.73
HORSEPOWER	4	512	222	0.87	0.76
SOURIS	5	507	152	0.91	0.84
NEWBURG	6	506	183	0.90	0.80
GOLIATH	7	504	201	0.89	0.78
SHELBY427	8	502	228	0.87	0.75
ROCKFORD	9	502	258	0.85	0.72
CORRAL	10	501	361	0.78	0.61

c. ME2

Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
AAC.ROSKENS	1	501	313	0.81	0.66
CDC.MINSTREL	2	499	187	0.89	0.80
AC PINNACLE	3	499	181	0.90	0.81
AAC.BULLET	4	498	333	0.80	0.64
NEWBURG	5	486	121	0.93	0.87
SOURIS	6	483	120	0.93	0.87
DEON	7	482	168	0.90	0.82
AAC.OAKLIN	8	482	306	0.82	0.67
AAC.ALMONTE	9	481	305	0.82	0.67
HAYDEN	10	480	117	0.93	0.87

d. ME3

Genotype	Ranking	BLUP (g*m-2)	P.E.V.	Accuracy (h)	reliability (r)
ESKER2020	1	436	116	0.94	0.88
DEON	2	430	82	0.96	0.91
NATTY	3	425	98	0.95	0.89
HAYDEN	4	415	76	0.96	0.92
GOLIATH	5	415	97	0.95	0.90
RON	6	413	106	0.94	0.89
CDC MINSTREL	7	408	201	0.89	0.78
BETAGENE	8	407	96	0.95	0.90
SHELBY427	9	404	66	0.96	0.93
EXCEL	10	404	102	0.94	0.89

References

Aguate, F., Crossa, J., & Balzarini, M. (2019). Effect of Missing Values on Variance Component Estimates in Multienvironment Trials. Crop Science, 59(2), 508–517. https://doi.org/10.2135/cropsci2018.03.0209

Amadeu, R. R., Cellon, C., Olmstead, J. W., Garcia, A. A., Resende Jr, M. F., & Muñoz, P. R. (2016). AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: A blueberry example. The Plant Genome, 9(3), plantgenome2016.01.0009.

Asoro, F. G., Newell, M. A., Beavis, W. D., Scott, M. P., & Jannink, J. (2011). Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. The Plant Genome, 4(2), plantgenome2011.02.0007. https://doi.org/10.3835/plantgenome2011.02.0007

Atanda, S. A., Govindan, V., Singh, R., Robbins, K. R., Crossa, J., & Bentley, A. R. (2022). Sparse testing using genomic prediction improves selection for breeding targets in elite spring wheat. Theoretical and Applied Genetics, 135(6), 1939–1950. https://doi.org/10.1007/s00122-022-04085-0

Atanda, S. A., Olsen, M., Burgueño, J., Crossa, J., Dzidzienyo, D., Beyene, Y., Gowda, M., Dreher, K., Zhang, X., Prasanna, B. M., Tongoona, P., Danquah, E. Y., Olaoye, G., & Robbins, K. R. (2021). Maximizing efficiency of genomic selection in CIMMYT's tropical maize breeding program. Theoretical and Applied Genetics, 134(1), 279–294. https://doi.org/10.1007/s00122-020-03696-9

Atanda, S. A., Olsen, M., Crossa, J., Burgueño, J., Rincent, R., Dzidzienyo, D., Beyene, Y., Gowda, M., Dreher, K., Boddupalli, P. M., Tongoona, P., Danquah, E. Y., Olaoye, G., & Robbins, K. R. (2021). Scalable Sparse Testing Genomic Selection Strategy for Early Yield Testing Stage. Frontiers in Plant Science, 12, 658978. https://doi.org/10.3389/fpls.2021.658978

Atlin, G. N., Kleinknecht, K., Singh, K. P., & Piepho, H. P. (2011). Managing genotype× environment interaction in plant breeding programs: A selection theory approach. Journal of the Indian Society of Agricultural Statistics, 65(2), 237–247.

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., & Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). Plant Science, 242, 23–36.

Bernardo, R. (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Science, 34(1), 20–25.

Bernardo, R. (2010). Breeding for quantitative traits in plants 2nd ed. Stemma Press, Woodbury, MN.

Berro, I., Lado, B., Nalin, R. S., Quincke, M., & Gutiérrez, L. (2019). Training population optimization for genomic selection. The Plant Genome, 12(3), 190028.

Bhatta, M., Gutierrez, L., Cammarota, L., Cardozo, F., Germán, S., Gómez-Guerrero, B., Pardo, M. F., Lanaro, V., Sayas, M., & Castro, A. J. (2020). Multi-trait genomic prediction model increased the predictive ability for agronomic and malting quality traits in barley (Hordeum vulgare L.). G3: Genes, Genomes, Genetics, 10(3), 1113–1124.

Bhatta, M., Sandro, P., Smith, MR., Delaney, O., Voss-Fels, K-P., Gutierrez, L., & Hickey, L. T. (2021). Need for speed: manipulating plant growth to accelerate breeding cycles. Current Opinion in Plant Biology 60, 101986.

Boer, M. P., Wright, D., Feng, L., Podlich, D. W., Luo, L., Cooper, M., & van Eeuwijk, F. A. (2007). A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics, 177(3), 1801–1813.

Braun, H.-J., Rajaram, S., & Ginkel, M. (1997). CIMMYT's approach to breeding for wide adaptation. Adaptation in Plant Breeding: Selected Papers from the XIV EUCARPIA Congress on Adaptation in Plant Breeding Held at Jyväskylä, Sweden from July 31 to August 4, 1995, 197–205.

Budhlakoti, N., Kushwaha, A. K., Rai, A., Chaturvedi, K. K., Kumar, A., Pradhan, A. K., Kumar, U., Kumar, R. R., Juliana, P., Mishra, D. C., & Kumar, S. (2022). Genomic Selection: A Tool for Accelerating the Efficiency of Molecular Breeding for Development of Climate-Resilient Crops. Frontiers in Genetics, 13, 832153. https://doi.org/10.3389/fgene.2022.832153

Buntaran, H., Forkman, J., & Piepho, H.-P. (2021). Projecting results of zoned multi-environment trials to new locations using environmental covariates with random coefficient models: Accuracy and precision. Theoretical and Applied Genetics, 134, 1513–1530.

Burgueño, J., Crossa, J., Cornelius, P. L., & Yang, R.-C. (2008). Using factor analytic models for joining environments and genotypes without crossover genotype× environment interaction. Crop Science, 48(4), 1291–1305.

Burgueño, J., de los Campos, G., Weigel, K., & Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype× environment interaction using pedigree and dense molecular markers. Crop Science, 52(2), 707–719.

Bustos-Korts, D., Malosetti, M., Chapman, S., & van Eeuwijk, F. (2016). Modelling of genotype by environment interaction and prediction of complex traits across multiple environments as a synthesis of crop growth modelling, genetics and statistics. Crop Systems Biology: Narrowing the Gaps between Crop Modelling and Genetics, 55–82.

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2020). ASReml-R reference manual. Version 4.1. 0.130. VSN International Ltd.

Campbell, M. T., Hu, H., Yeats, T. H., Brzozowski, L. J., Caffe-Treml, M., Gutiérrez, L., Smith, K. P., Sorrells, M. E., Gore, M. A., & Jannink, J.-L. (2021). Improving Genomic Prediction for Seed Quality Traits in Oat (Avena sativa L.) Using Trait-Specific Relationship Matrices. Frontiers in Genetics, 12, 643733. https://doi.org/10.3389/fgene.2021.643733

Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., Hagen, T., Quinn, M., & Ng, E. H. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: Lessons from the breeder's equation. Theoretical and Applied Genetics, 132(3), 627–645. https://doi.org/10.1007/s00122-019-03317-0

Cooper, M., & DeLacy, I. H. (1994). Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. Theoretical and Applied Genetics, 88, 561–572.

Costa-Neto, G., Fritsche-Neto, R., & Crossa, J. (2021). Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. Heredity, 126(1), 92–106.

Crossa, J., & Cornelius, P. L. (1997). Sites regression and shifted multiplicative model clustering of cultivar trial sites under heterogeneity of error variances. Crop Science, 37(2), 406–415.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., & Varshney, R. K. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends in Plant Science, 22(11), 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Crossa, J., Vargas, M., Van Eeuwijk, F. A., Jiang, C., Edmeades, G. O., & Hoisington, D. (1999). Interpreting genotype× environment interaction in tropical maize using linked molecular markers and environmental covariables. Theoretical and Applied Genetics, 99, 611–625.

Cullis, B. R., Smith, A. B., Beeck, C. P., & Cowling, W. A. (2010). Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. Genome, 53(11), 1002–1016.

Dawson, J. C., Endelman, J. B., Heslot, N., Crossa, J., Poland, J., Dreisigacker, S., Manès, Y., Sorrells, M. E., & Jannink, J.-L. (2013). The use of unbalanced historical data for genomic selection in an international wheat breeding program. Field Crops Research, 154, 12–22.

Dumble, S., Bernal, E. F., & Villardon, P. G. (2017). GGEBiplots: GGE Biplots with ‘ggplot2.’ R Package Version 0.1, 1. https://CRAN.R-project.org/package=GGEBiplots

Endelman, J. B., Atlin, G. N., Beyene, Y., Semagn, K., Zhang, X., Sorrells, M. E., & Jannink, J.-L. (2014). Optimal design of preliminary yield trials with genome-wide markers. Crop Science, 54(1), 48–59.

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. Biometrika, 58(3), 453–467.

Gauch Jr, H. G. (1988). Model selection and validation for yield trials with interaction. Biometrics, 705–715.

Gauch Jr, H. G., & Zobel, R. W. (1997). Identifying mega-environments and targeting genotypes. Crop Science, 37(2), 311–326.

Gezan, S. (2020). Linear mixed models: Obtaining estimates of random effects with ASReml-R. Asreml-r Blog. https://vsni.co.uk/case-studies/calculating-accuracy-of-random-effect-estimates-with-ASReml-R

González-Barrios, P., Díaz-García, L., & Gutiérrez, L. (2019). Mega-environmental design: Using genotype× environment interaction to optimize resources for cultivar testing. Crop Science, 59(5), 1899–1915.

Haikka, H., Knürr, T., Manninen, O., Pietilä, L., Isolahti, M., Teperi, E., Mäntysaari, E. A., & Strandén, I. (2020). Genomic prediction of grain yield in commercial Finnish oat (Avena sativa) and barley (Hordeum vulgare) breeding programmes. Plant Breeding, 139(3), 550–561. https://doi.org/10.1111/pbr.12807

Haikka, H., Manninen, O., Hautsalo, J., Pietilä, L., Jalli, M., & Veteläinen, M. (2020). Genome-wide Association Study and Genomic Prediction for Fusarium graminearum Resistance Traits in Nordic Oat (Avena sativa L.). Agronomy, 10(2), 174. https://doi.org/10.3390/agronomy10020174

Haile, T. A., Heidecker, T., Wright, D., Neupane, S., Ramsay, L., Vandenberg, A., & Bett, K. E. (2020). Genomic selection for lentil breeding: Empirical evidence. The Plant Genome, 13(1), e20002.

Hartung, J., & Piepho, H.-P. (2021). Effect of missing values in multi-environmental trials on variance component estimates. Crop Science, 61(6), 4087–4097.

Heffner, E. L., Jannink, J.-L., & Sorrells, M. E. (2011). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. The Plant Genome, 4(1).

Heslot, N., Akdemir, D., Sorrells, M. E., & Jannink, J.-L. (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theoretical and Applied Genetics, 127, 463–480.

Heslot, N., Rutkoski, J., Poland, J., Jannink, J.-L., & Sorrells, M. E. (2013). Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. PloS One, 8(9), e74612.

Hoefler, R., Gonzalez-Barrios, P., Bhatta, M., Berro, I., Nalin, R.S., Borges, A., Covarrubias, E., Diaz- Garcia, L., & Gutierrez, L. (2020). Do spatial designs outperform classic experimental designs? Journal of Agricultural, Biological, and Environmental Statistics 25:523-552.

Hu, H., Campbell, M. T., Yeats, T. H., Zheng, X., Runcie, D. E., Covarrubias-Pazaran, G., Broeckling, C., Yao, L., Caffe-Treml, M., Gutiérrez, L., Smith, K. P., Tanaka, J., Hoekenga, O. A., Sorrells, M. E., Gore, M. A., & Jannink, J.-L. (2021). Multi-omics prediction of oat agronomic and seed nutritional traits across environments and in distantly related populations. Theoretical and Applied Genetics, 134(12), 4043–4054. https://doi.org/10.1007/s00122-021-03946-4

Jarquin, D. (2022). From Sparsity to Sparse Testing Designs to Accomplish Trait Forecasting. Plant and Animal Genome XXIX Conference (January 8-12, 2022).

Jarquín, D., Lemes da Silva, C., Gaynor, R. C., Poland, J., Fritz, A., Howard, R., Battenfield, S., & Crossa, J. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype× environment interactions in Kansas wheat. The Plant Genome, 10(2), plantgenome2016.12.0130.

Lado, B., Barrios, P. G., Quincke, M., Silva, P., & Gutiérrez, L. (2016). Modeling genotype× environment interaction for genomic selection with unbalanced data from a wheat breeding program. Crop Science, 56(5), 2165–2179.

Lado, B., Vázquez, D., Quincke, M., Silva, P., Aguilar, I., & Gutiérrez, L. (2018). Resource allocation optimization with multi-trait genomic prediction for bread wheat (Triticum aestivum L.) baking quality. Theoretical and Applied Genetics, 131, 2719–2731.

Lush, J. L. 1937. Animal breeding plans. Iowa State College Press,Ames, IA

Malosetti, M., Bustos-Korts, D., Boer, M. P., & van Eeuwijk, F. A. (2016). Predicting responses in multiple environments: Issues in relation to genotype× environment interactions. Crop Science, 56(5), 2210–2222.

Malosetti, M., Ribaut, J.-M., & van Eeuwijk, F. A. (2013). The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. Frontiers in Physiology, 4, 44.

Malosetti, M., Voltas, J., Romagosa, I., Ullrich, S. E., & Van Eeuwijk, F. A. (2004). Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica, 137, 139–145.

Mathews, K. L., Malosetti, M., Chapman, S., McIntyre, L., Reynolds, M., Shorter, R., & Van Eeuwijk, F. (2008). Multi-environment QTL mixed models for drought stress adaptation in wheat. Theoretical and Applied Genetics, 117, 1077–1091.

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics, 157(4), 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Monteverde, E., Gutierrez, L., Blanco, P., Pérez de Vida, F., Rosas, J. E., Bonnecarrère, V., Quero, G., & McCouch, S. (2019). Integrating Molecular Markers and Environmental Covariates To Interpret Genotype by Environment Interaction in Rice (Oryza sativa L.) Grown in Subtropical Areas. G3 Genes|Genomes|Genetics, 9(5), 1519–1531. https://doi.org/10.1534/g3.119.400064

Monteverde, E., Rosas, J. E., Blanco, P., Pérez de Vida, F., Bonnecarrère, V., Quero, G., Gutierrez, L., & McCouch, S. (2018). Multienvironment models increase prediction accuracy of complex traits in advanced breeding lines of rice. Crop Science, 58(4), 1519–1530.

Neyhart, J. L., Silverstein, K. A., & Smith, K. P. (2022). Accurate predictions of barley phenotypes using genomewide markers and environmental covariates. Crop Science, 62(5), 1821–1833.

Oakey, H., Cullis, B., Thompson, R., Comadran, J., Halpin, C., & Waugh, R. (2016). Genomic selection in multi-environment crop trials. G3: Genes, Genomes, Genetics, 6(5), 1313–1326.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58(3), 545–554.

Piepho, H. P., Möhring, J., Melchinger, A. E., & Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. Euphytica, 161(1–2), 209–228.

Piepho, H.-P. (1998). Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures: Theoretical and Applied Genetics, 97(1–2), 195–201. https://doi.org/10.1007/s001220050885

Piepho, H.-P. (2000). A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics, 156(4), 2043–2050.

Piepho, H.-P., Möhring, J., Schulz-Streeck, T., & Ogutu, J. O. (2012). A stage-wise approach for the analysis of multi-environment trials. Biometrical Journal, 54(6), 844–860.

Rogers, A. R., & Holland, J. B. (2021). Environment-specific genomic prediction ability in maize using environmental covariates depends on environmental similarity to training data. G3 Genes|Genomes|Genetics, jkab440. https://doi.org/10.1093/g3journal/jkab440

Rutkoski, J. E., Crain, J., Poland, J., & Sorrells, M. E. (2017). Genomic selection for small grain improvement. Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement, 99–130.

Rutkoski, J., Singh, R. P., Huerta-Espino, J., Bhavani, S., Poland, J., Jannink, J. L., & Sorrells, M. E. (2015). Efficient use of historical data for genomic selection: A case study of stem rust resistance in wheat. The Plant Genome, 8(1), plantgenome2014.09.0046.

Schaeffer, L. R. (2004). Application of random regression models in animal breeding. Livestock Production Science, 86(1–3), 35–45.

Semagn, K., Iqbal, M., Crossa, J., Jarquin, D., Howard, R., Chen, H., Bemister, D. H., Beres, B. L., Randhawa, H., N'Diaye, A., Pozniak, C., & Spaner, D. (2022). Genome-based prediction of agronomic traits in spring wheat under conventional and organic management systems. Theoretical and Applied Genetics, 135(2), 537–552. https://doi.org/10.1007/s00122-021-03982-0

Smith, A., Cullis, B., & Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics, 57(4), 1138–1147.

Spindel, J., & Iwata, H. (2018). Genomic selection in rice breeding. Rice Genomics, Genetics and Breeding, 473–496.

Team, R. D. C. (2017). R Core Team. R Foundation for Statistical Computing Austria.

van Eeuwijk, F. A. (2006). Genotype by Environment Interaction–Basics and Beyond. Plant Breeding: The Arnel R. Hallauer International Symposium, 155–170.

van Eeuwijk, F. A., Bustos-Korts, D. V., & Malosetti, M. (2016). What should students in plant breeding know about the statistical aspects of genotype× environment interactions? Crop Science, 56(5), 2119–2140.

van Eeuwijk, F. A., Malosetti, M., Yin, X., Struik, P. C., & Stam, P. (2005). Statistical models for genotype by environment data: From conventional ANOVA models to eco-physiological QTL models. Australian Journal of Agricultural Research, 56(9), 883–894.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. Journal of Dairy Science, 91(11), 4414–4423.

Vargas, M., Crossa, J., van Eeuwijk, F. A., Ramírez, M. E., & Sayre, K. (1999). Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype× environment interaction. Crop Science, 39(4), 955–967.

Varshney, R. K., Roorkiwal, M., & Sorrells, M. E. (Eds.). (2017). Genomic Selection for Crop Improvement. Springer International Publishing. https://doi.org/10.1007/978-3-319-63170-7

Verbyla, A. P., Eckermann, P. J., Thompson, R., & Cullis, B. R. (2003). The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. Australian Journal of Agricultural Research, 54(12), 1395–1408.

Wang, T.-C., Casadebaig, P., & Chen, T.-W. (2023). More than 1000 genotypes are required to derive robust relationships between yield, yield stability and physiological parameters: A computational study on wheat crop. Theoretical and Applied Genetics, 136(3), 34. https://doi.org/10.1007/s00122-023-04264-7

Ward, B. P., Brown-Guedira, G., Kolb, F. L., Van Sanford, D. A., Tyagi, P., Sneller, C. H., & Griffey, C. A. (2019). Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia. PloS One, 14(2), e0208217.

Yan, W. (2015). Mega-environment analysis and test location evaluation based on unbalanced multiyear data. Crop Science, 55(1), 113–122.

Yan, W., Cornelius, P. L., Crossa, J., & Hunt, L. A. (2001). Two types of GGE biplots for analyzing multi-environment trial data. Crop Science, 41(3), 656–663.

Yan, W., Hunt, L. A., Sheng, Q., & Szlavnics, Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Science, 40(3), 597–605.

Yan, W., & Kang, M. S. (2002). GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC press.

Yan, W., Kang, M. S., Ma, B., Woods, S., & Cornelius, P. L. (2007). GGE biplot vs. AMMI analysis of genotype-by-environment data. Crop Science, 47(2), 643–653.

Yan, W., & Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. Canadian Journal of Plant Science, 86(3), 623–645.

Chapter 4. Using Finlay-Wilkinson Random Regression to Predict Yield and Yield Stability in Cereals

Pablo Sandro[1], Justin Blancon[2], Julie Dawson[1], Lucia Gutierrez[1]

[1] Department of plant and agroecosystem sciences, University of Wisconsin, Madison, WI 53706
[2] INRAE, GDEC - DiGEN (Diversité et Génome) - Clermont-Ferrand, France

Author contributions:

All authors read and approved of the final manuscript. PS wrote the original manuscript, performed statistical analysis and figures. JB discussed methodology and model utilization and contributes wheat data set. JD read, comment and discussed methodology and model utilization, LG designed the study and hypothesis, discussed methodology and model utilization and contributed oats and barley data sets.

Core ideas

1. Standard Finaly-Wilkinson regression and Finlay-Wilkinson random regression do not show differences in estimation of stability.

2. Finlay-Wilkinson random regression to predict grain yield and grain yield stability for non-observed genotypes by leveraging their relative's phenotypic information.

3. Small numbers of environments with a high variance are as accurate in the estimation of grain yield stability as a MET with over 20 environments

Abstract

Climate change poses a challenge for agriculture by increasing climatic variability and compromising crop yields, hence there is a need to select cultivars that can withstand climatic variability. Through plant breeding, we can develop cultivars with yield stability. Cultivar yield stability can be evaluated by different methods. Most stability indices require complete datasets spanning multiple environments. A feature of breeding programs is that they typically generate sparse multi-environment datasets. Methods such as Finlay-Wilkinson random regression (RFW) allow us to use sparse datasets and incorporate genomic data to leverage phenotypic information from related lines to predict yield stability. Our objective was to compare RFW with the traditional Finlay-Wilkinson regression (FW), and GBLUP models to predict yield and yield stability and examine how variation in the number of environments used to fit the model and environmental variance impact those predictions. We used three very different datasets, one highly unbalanced dataset for oats, and two completely balanced datasets with a different number of environments in barley and wheat. We ran FW and RFW models in different scenarios where we estimated stability and predicted grain yield. We found that using whole datasets for FW or RFW, estimated stability values obtained were similar. When we used RFW to predict grain yield in unobserved genotypes we obtained higher predictive ability than using only genotypic information as in GBLUP. When we look at variations in the number of environments we found that selecting the right set of environments is more important than adding more environments.

Introduction

Increase climate variability and the impact of climate change on biotic (e.g., weeds, pests) and abiotic (e.g., drought, floods, heat) factors challenge agricultural production worldwide (Alotaibi, 2023, IIPC, 2023). Plant breeding programs need to update their goals to develop varieties with adaptation to future climatic conditions (Paniza, 2024). It is expected that climate change will affect plant breeding and selection objectives in unknown and unprecedented ways (Xiong et al., 2022). Another study by Macholdt & Honermeier, (2017) found that farmers, extension experts, and scientists expect that yield stability will become more relevant in the upcoming years due to climate change. One possible solution to this problem is to select stable genotypes that can withstand climatic stresses (Picasso et al., 2019).

Phenotypic stability is referred as the consistent performance of a genotype relative to other varieties tested on the same set of environments (Annicchiarico, 2002; Becker & Léon, 1988). An environment, in the context of multi-environment trials, is considered a combination of location and year (Bernardo, 2014). This environment concept summarizes the particular combination of biotic and abiotic factors that are affecting the phenotype (Bernardo, 2014). Genotypes under evaluation in a breeding program are tested in several locations across several years (Piepho & Möhring, 2006; Yan, 2015). The locations selected for a multi-environment trial represent a repeatable sample of the possible locations within the target population of environment for the breeding program (Yan et al., 2011). On the other hand, year and location × year interactions are inherently unpredictable, yet they account for a significant portion of the overall variance in phenotypic performance (Yan et al., 2011). The target population of the environment is defined by genotypes' performance in a set of locations over several years, given the increasing unpredictability of environmental conditions due to climate change, breeding programs will need to adapt and develop new strategies (Alotaibi, 2023). For example, the use of environments outside of the TPE, where exposure to specific or controllable abiotic stresses can be leveraged (Bernardo, 2014).

The characterization of the property of stability for a genotype or a system has a prolific literature including many indices to estimate stability (Pour-Aboughadareh et al., 2022), however, consensus about which is the best estimator has not been reached (Reckling et al., 2021). Therefore a common practice is to calculate several indices to capture different aspects of stability (i.e. Bocci et al., 2020; Mohammadi & Amri, 2013; Mühleisen et al., 2014). Most stability indices need complete balanced datasets (i.e. all genotypes need to be evaluated in all environments, Wricke 1962; Francis and Kannenberg 1978; Gauch 1992; Yan and Kang 2002; Bacsi and Hollósy 2019; Lin and Binns). Other indices such as β from Finlay-Wilkinson regression (Finlay & Wilkinson, 1963) have more flexibility and allow the use of unbalanced data for genotype to calculate stability (Pour-Aboughadareh et al., 2022). This can be favorable to leveraging historical information from plant breeding multi-environment trials.

Plant breeding evaluation trials have a hierarchical structure where genotypes are selected and discarded as they are tested, generating unbalanced datasets (Piepho & Möhring, 2006). In the final stages of evaluation, there is a tradeoff between the number of locations assigned to a genotype and the number of genotypes in evaluation (Gauch & Zobel, 1996). As a consequence, genotypes are evaluated in a subset of the total environments available for testing (Piepho & Möhring, 2006; Smith et al., 2001). There is no consensus on how many environments are needed to accurately estimate yield stability (Reckling et al., 2021). Numbers can range from 21 when characterizing GEI in oats (Yan et al., 2015) to as many as 150 environments needed to estimate stability in physiological traits (Wang et al., 2023). The number of environments will change with the heritability of traits in the study and the trials (Yan et al. 2015). In recent times the availability of genomic information at low cost allows the further optimization of resource allocation across locations without losing trial quality (Endelman et al., 2014). Genomic information also facilitates the use of historical datasets composed of multi-environment trials (METs) to understand genotype-by-environment interactions (González-Barrios et al., 2019; Sandro et al., 2024). Several genomics techniques can be used, fitting prediction models in unbalanced datasets to calculate stability indicators (Carvalho et al., 2024), using linear regressions (Neyhart et al. 2024 in review), or combining genomics with Finlay-Wilkinson random regression (Ly et al. 2018; Piepho & Blancon, 2023). Another example of the use of random regression and environment covariates can be used for environmental characterization to improve genomic predictions (Rebollo et al., 2023). In a similar way, Finlay-Wilkinson extended using random regression can be combined with genomic and sparse phenotypic data using relatives' information as leverage to predict phenotypic stability (Piepho & Blancon, 2023). There is a need to identify high-yielding and stable genotypes in plant breeding, but most breeding programs face challenges posed by unbalanced datasets. To respond and overcome these limitations, our objective was to integrate Finlay-Wilkinson random regression (RFW) as a genomic prediction tool for characterizing yield stability. We aim to compare it with the traditional Finlay-Wilkinson regression that uses a fixed effects model, and a traditional prediction model such as GBLUP. We also aim to compare how RFW and FW are affected by variations in the number of environments used to fit the model and how the environmental variance impacts the results observed in each model.

Materials and methods

Materials and Experiments

We used datasets for three crops: barley (*Hordeum vulgare* L.), initially reported in (Neyhart et al., 2019); oats (*Avena sativa* L.), reported in (Sandro et al., 2024); and wheat (*Triticum aestivum* L.), reported in (Rincent et al., 2019).

Barley dataset is balanced, and it was generated between 2016 and 2017 across 17 locations across the north of the United States and south of Canada, representing areas where Barley has been traditionally grown. The oats dataset is highly unbalanced, and it was generated between 1998 and 2021 across 59 locations across the Midwest region of the United States. The trials are part of the USDA-coordinated collaborative nursery including public breeding programs across the US and Canada where oats have been traditionally grown. We only used data corresponding to the Midwest of the United States, all data is available in the T3/Oat database (available at https://oat.triticeaetoolbox.org). The wheat dataset is balanced, and it was generated during the years 2012, 2013, and 2014, across 14 locations testing water and nitrogen use efficiency across France.

Table 4.1. Description of barley, oats, and wheat datasets including number of genotypes, number of environments (combination of location and year), number of environments for each genotype, and number of markers used to estimate the realized additive relationship matrix.

Dataset	Number of Genotypes	Number of Environments	Number of Environments per Genotype	Number of SNP Markers
barley	223	30	28-30	6,321
oat	581	673	3-480	2,699
wheat	210	16	16	34,464

Environmental means

The mean yield performance for each environment was obtained as the Best Linear Unbiased Estimator (BLUEs) of each environment using the following model.

Eq [1] $$y_{ij} = \mu + g_i + e_j + ge_{ij} + \varepsilon_{ij}$$

where y_{ij} is the BLUE for grain yield for the *ith* genotype in the *jth* environment, g_i is the effect of the *ith* genotype, e_j is the effect of the *jth* environment (i.e. year-location combination), ge_{ij} is the effect of the *ith* genotype and the *jth* environment interaction, and e_{ij} the error associated with the *ith* genotype in the *jth* environment with $\varepsilon_{ij} \sim$ N(0,$\sigma_{\varepsilon}^{2}D$) where σ_{ε}^{2} is the error variance and D= $\mathbf{I}*SE_{ij}^{2}{}^{-1}$is a parametrization of the standard error for the estimation of the BLUEs for the *ith* genotype and the *jth* environment, and **I** is an identity matrix of order n.

Finlay-Wilkinson regression

We fitted a Finlay-Wilkinson regression over the environment, grain yield mean, and where parameters for genotype and environment are fitted as fix effects. This estimate of the intercept ($\hat{G}_{\mathrm{i}}$), and slope ($\hat{\beta}_{\mathrm{i}(FFW)}$) for each genotype was considered the best approximation to the true value. We used the following model:

Eq [2] $$y_{ij} = G_i + \beta_i E_j + e_{ij}$$

where y_{ij} is the grain yield BLUEs for *ith* genotype in *jth* environment, G_i represents the grain yield mean (i.e. regression intercept) for the *ith* genotype, β_i is the regression slope for the *ith* genotype, E_j is the mean of the *jth* environment estimated by the mean of all genotypes across the environment and e_{ij} the error associated with the *ith* genotype in the *jth* environment, with $\varepsilon_{ij} \sim$ N(0,$\sigma_{\varepsilon}^{2}D$) where

σ_ε^2 is the error variance and D= I*${SE_{ij}^2}^{-1}$ is a parametrization of the standard error for the estimation of the corrected means for the *ith* genotype and the *jth* environment.

Finlay-Wilkinson Random Regression (RFW)

We estimated random regression Finlay-Wilkinson (RFW) to leverage information from relatives that were tested in other environments. We used the Eq[2] with G_i and $\beta_i E_j$ as random where G_i is the random regression intercept for the i^{th} genotype, β_i is the random regression slope to the i^{th} genotype, E_j correspond to the environmental mean for the j^{th} environment, and e_{ij} is the regression error with, g~ MVN$(0, \sigma^2\boldsymbol{K})$, β_i~ N$(0, \sigma^2\boldsymbol{K})$, and e_{ij}~ N(0,$\sigma_\varepsilon^2 D$) where σ_ε^2 is the error variance and D= I*${SE_{ij}^2}^{-1}$ is a parametrization of the standard error for the estimation of the BLUEs for the i^{th} genotype and the j^{th} environment, and $\boldsymbol{K}$ is the realized additive relationship estimated matrix. Covariance between G_i and β_i, $\sigma_{G_i\beta_i}$ was modeled $\sigma_{G_i\beta_i}\boldsymbol{K}$.

GBLUP Model

We fit a GBLUP (Bernardo, 1994) model that was modeled using the following linear mixed model in the *asreml-r* (Butler *et al.* 2020) package of the R statistical software (R Development Core Team, 2020):

Eq [2]
$$y_{ij} = \mu + g_i + \varepsilon_{ij}$$

y_{ij} are the genotypic BLUEs for grain yield of the i^{th} genotype in the j^{th} environment (i.e. location-year combination), μ is the overall mean or intercept, g_i is the effect of the ith genotype g~ MVN$(0, \sigma^2\boldsymbol{K})$, and ε_{ij} is the residual error with ε ~ N(0,$\sigma_\varepsilon^2 D$) where σ_ε^2 is the error variance and D= I*${SE_{ij}^2}^{-1}$ is a parametrization of the standard error for the estimation of the BLUEs for the i^{th} genotype and the j^{th} environment, and $\boldsymbol{K}$ is the realized additive relationship estimated matrix.

Cross validations and predictive abilities

We performed a 10- fold cross-validation (CV1, Burgueño et al., 2012) with 100 iterations to determine the ability of the RFW model to predict stability ($\tilde{\beta}_{i(RFW)}$) and grain yield. On 10-fold CV1 genotypes are assigned to one of ten groups, and in the correspondent fold all the information from the genotypes in the group is masked, this simulates the incorporation of a new genotype with only genomic information. Predictive ability (PA) for the slope of grain yield was calculated as the correlation between the predicted slope ($\tilde{\beta}_{i(RFW)}$) for each genotype using the complete data set and the CV1 predicted slope ($\tilde{\beta}_{i(cv1|RFW)}$), cor($\tilde{\beta}_{i(RFW)}, \tilde{\beta}_{i(cv1|RFW)}$). At the same time, we used the predicted slope ($\hat{\beta}_{i(cv1|RFW)}$) to solve the regression equation and predict grain yield for each environment where the genotype was tested. To avoid inflating the correlation between observed and predicted due to the environmental effects, we centered both predicted grain yield ($\tilde{y}_{ij(cv1|RFW)}$) and the observed BLUEs for grain yield ($\hat{y}_{ij}$) in each environment by subtracting the environmental mean ($\bar{E}_j$). PA for grain yield was calculated as the correlation between the environment-centered predicted grain yield and the environment-centered observed grain yield, cor(($\hat{y}_{ij} - \bar{E}_j$),($\tilde{y}_{ij(cv1|RFW)} - \bar{E}_j$)). PA was graphically represented using ggplot2 (Wickham, 2016) in R (R Core Team, 2017).

To compare the use of RFW yield predictions to other prediction methods, we used the same data set to perform a 10-fold cross-validation one (CV1 Burgueño et al., 2012) with 100 iterations to determine the predictive ability for grain yield. $\tilde{y}_{ij(cv1|gBLUP)}$ using a gBLUP model (Bernardo, 1994).

Effect of the number of environments on β

To analyze the effect of the number of environments used to fit the regression on β, we subset the available environments to form different environment data sets which we used to fit both FW and RFW models. We performed 100 iterations on each number of subset environments, each subset was formed by two, five, and ten environments in wheat, and two, five, ten, fifteen, and 20 environments in barley. On each iteration, we used only the information on the environments in the subset and masked all information on the remnant environments. For each subset of environments, we recorded the list of environments, the minimum environmental mean, the maximum environmental mean, the environmental mean range, and the variance of environmental means. PA was calculated as the Pearson correlation between the FW slope estimated with the full environmental data set ($\hat{\beta}_{i(FW)}$) and the FW slope estimated using each partitioned subset of environments ($\hat{\beta}_{i(p|FW)}$) as cor($\hat{\beta}_{i(FW)}$, $\hat{\beta}_{i(p|FW)}$). PA was calculated similarly for RFW cor($\tilde{\beta}_{i(RFW)}$, $\tilde{\beta}_{i(p|RFW)}$), where $\tilde{\beta}_{i(RFW)}$is the slope for the RFW predicted using the full environmental data set, and $\tilde{\beta}_{i(p|RFW)}$ is the slope for the RFW predicted using the environmental subset. Grain yield was predicted as described in the previous section. The predictive ability for grain yield was calculated as the Pearson correlation between predicted grain yield for FW slope fitted using the environment subset ($\hat{y}_{i(p|FFW)}$) and the observed grain yield corrected mean ($\hat{y}_{ij}$) for the environments that were not part of the subset of environments used to fit the model; cor(($\hat{y}_{ij} - \bar{E}_j$), ($\hat{y}_{ij(p|FW)} - \bar{E}_j$)). PA for grain yield using RFW was calculated following the same procedure: cor(($\hat{y}_{ij} - \bar{E}_j$), ($\tilde{y}_{ij(p|RFW)} - \bar{E}_j$)).

To further understand the effect of the number of environments in oats, we fitted FW regressions and estimated. $\hat{\beta}_{i(FW)}$ Using a progression of environment numbers between two and two hundred environments. We used oats genotypes that were tested in over 200 environments, for each number of environments (i.e. two, three, four, …, two hundred) we drew a random sample from all the available environments where the genotypes were tested. We repeated the sampling process 100 times for each number of environments in the progression. For each sample, we estimated the environment variance, range, and $\hat{\beta}_{i(p|FW)}$. We selected four representative genotypes with more than 200 environments to graph their results (i.e. Ogle, Horsepower, Deon, Clintland64).

Results

Effect of contrasting testing environments

Increases in the variance of environmental means in general increased the ability of the set of environments to estimate $\hat{\beta}_{\mathrm{i}(FW)}$. A small set of environments (i.e. two environments) with high variance between their means showed a higher predictive ability to estimate $\hat{\beta}_{\mathrm{i}(FW)}$, $\tilde{\beta}_{\mathrm{i}(RFW)}$, and grain yield than subsets with more environments and a lower environmental variance (Fig 4.2 in barley, Fig 4.1 in oats, and Fig 4.3 in wheat). Even though when we compare intermediate variance values, a set of environments with an equal number of environments and equal variance shows different ability to predict $\hat{\beta}_{\mathrm{i}(p|FW)}$. The effects of the variance and number of environments have more impact on the predictive ability in environment subsets integrated by fewer environments (i.e., two and five environments). The predictability of $\tilde{\beta}_{\mathrm{i}(RFW)}$ when we fit the RFW model, with a different number of environments and variance of the environment means follow the same pattern as FW described above.

Number of environments

The effect of the number of environments used to fit the model is similar in barley and wheat. The correlation between $\hat{\beta}_{\mathrm{i}(p|FW)}$ with $\hat{\beta}_{\mathrm{i}(FW)}$ increases as the number of environments increases for barley and wheat (Fig. 4.3 & 4.4). In barley, the maximum PA is reached with 20 environments yet the increment from using 15 environments is marginal. In oats (Fig 4.1, Supplementary Figures 4.1, 4.2, and 4.3) we observe similar patterns to those observed in barley and wheat, for the number of environments, the variance of the mean of the environments. Also, since we have a higher number of environments, we observe that the increase in the predictive ability of $\hat{\beta}_{\mathrm{i}(FW)}$ are marginal beyond 20 environments.

On the barley and wheat datasets when $\hat{\beta}_{i(p|FW)}$ and $\tilde{\beta}_{\mathrm{i}(p|RFW)}$ were used to predict grain yield in environments not included to fit the model, the average PA was 0.33 (Range: 0 and 0.62) in barley (Fig. 4.2) and 0.52 (Range: 0.05 – 0.82) in wheat (Fig. 4.3). There is variability in grain yield PA relative to the number of environments used in the prediction. Training sets with a small number of environments (i.e. two or five environments) showed a higher level of variability for estimation and predictions. As the number of environments used to fit the model increased, the variability in the results was reduced (Fig. 4.1, 4.2, and 4.3). In the oat data set, changes in the mean of environments variance, and $\hat{\beta}_{i(p|FW)}$ variability where marginal after 20 environments were included.

Even though sets with two and five environments showed more variability, some of the environment combinations have values of $\hat{\beta}_{i(p|FW)}$ on the range of the expected for the $\hat{\beta}_{i(FW)}$ estimated using the full data set. This is translated to better grain yield predictions for those sets with accurate $\hat{\beta}_{i(FW)}$ estimations. It may be due to differences in environmental variance, as measured by the differences in mean grain yield among environments included in the training set.

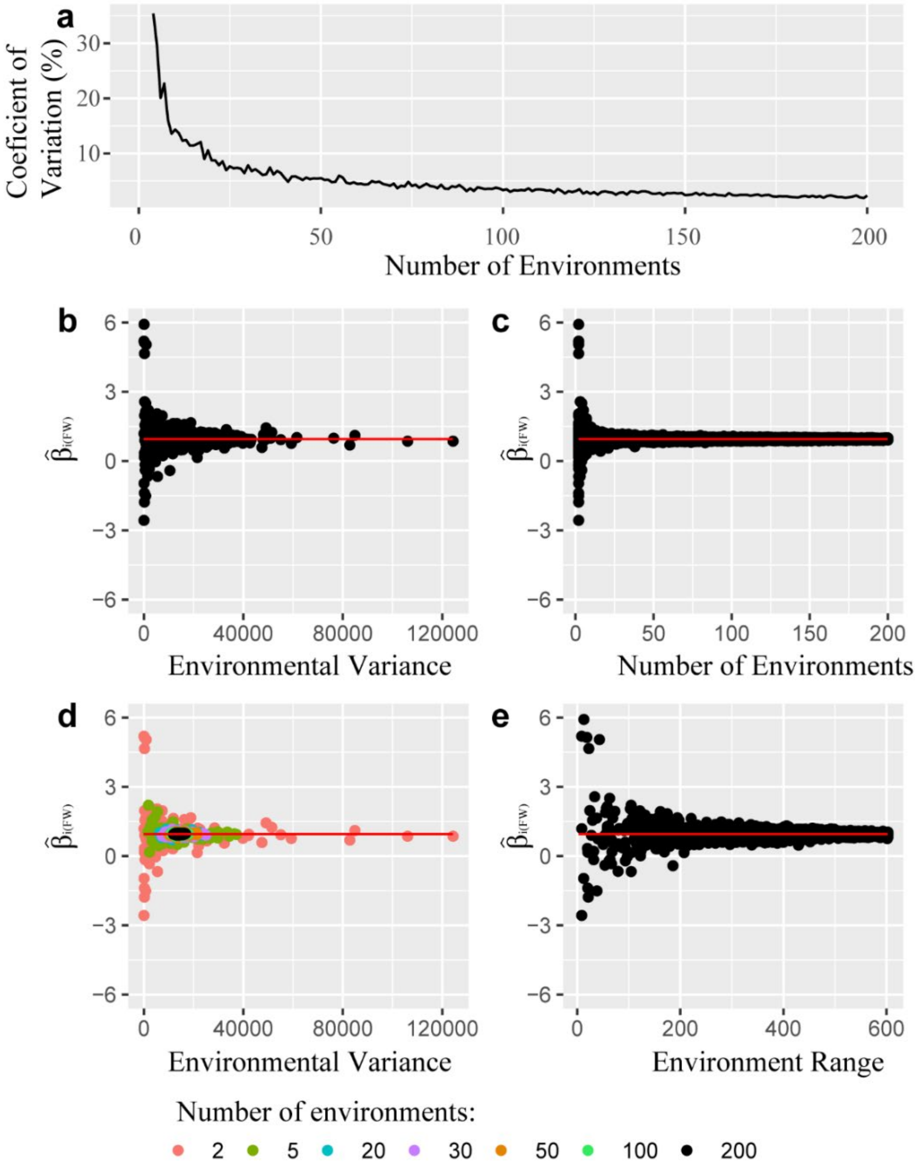


Figure 4.1 Cultivar Ogle's $\hat{\beta}_{i(p|FW)}$ variation on number, environmental variance, and of environments used to fit the model. The red lines represent the value of $\hat{\beta}_{i(p|FW)}$ estimated using all environments available for the line (480 environments). **a.** Coefficient of variation for $\hat{\beta}_{i(p|FW)}$ per number of environments used to fit the model. **b.** $\hat{\beta}_{i(p|FW)}$ against environmental variance for the environments used to fit the model. **c.** $\hat{\beta}_{i(p|FW)}$ against the number of environments used to fit the model. d. $\hat{\beta}_{i(p|FW)}$ environmental variance for 100 sets of 2, 5, 20, 30, 50, 100, and 200

environments used to fit the model. **e.** $\hat{\beta}_{i(p|FW)}$ against the range of environments used to fit the model.

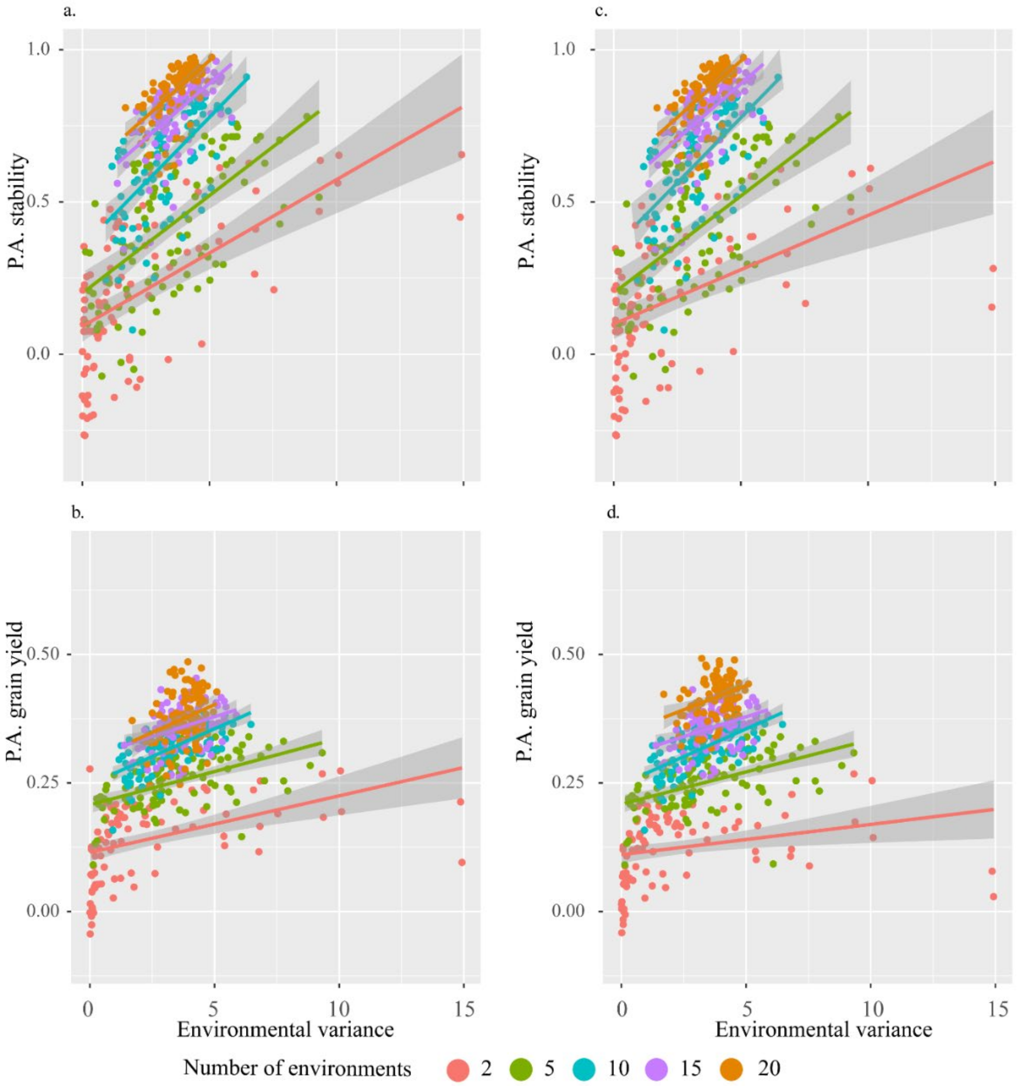


Figure 4.2 Correlation between stability for the complete data set versus stability for a partial data set of environments in barley dataset. **a.** environmental variance of the partial set of environments versus the predictive ability of the partial set of environments cor($\tilde{\beta}_{\mathrm{i}(RFW)}$ and $\tilde{\beta}_{\mathrm{i}(p|RFW)}$). **b.** the environmental variance of the partial set of environments against predictive ability as cor($\hat{\beta}_{\mathrm{i}(FW)}$ and $\hat{\beta}_{\mathrm{i(p}|RFW)}$). **c.** Environmental variance for the set of environments against predictive ability for

grain yield predicted using $\tilde{\beta}_{\mathrm{i}(p|RFW)}$ calculated as environment-centered cor($\tilde{y}_{\mathrm{i}(RFW)}$ and $\tilde{y}_{\mathrm{i}(p|RFW)}$). **d.** Environmental variance for the set of environments against predictive ability for grain yield predicted using $\hat{\beta}_{\mathrm{i(p}|FW)}$ calculated as environment-centered cor($\hat{y}_{\mathrm{i}(FW)}$ and $\hat{y}_{\mathrm{i(p}|RFW)}$).

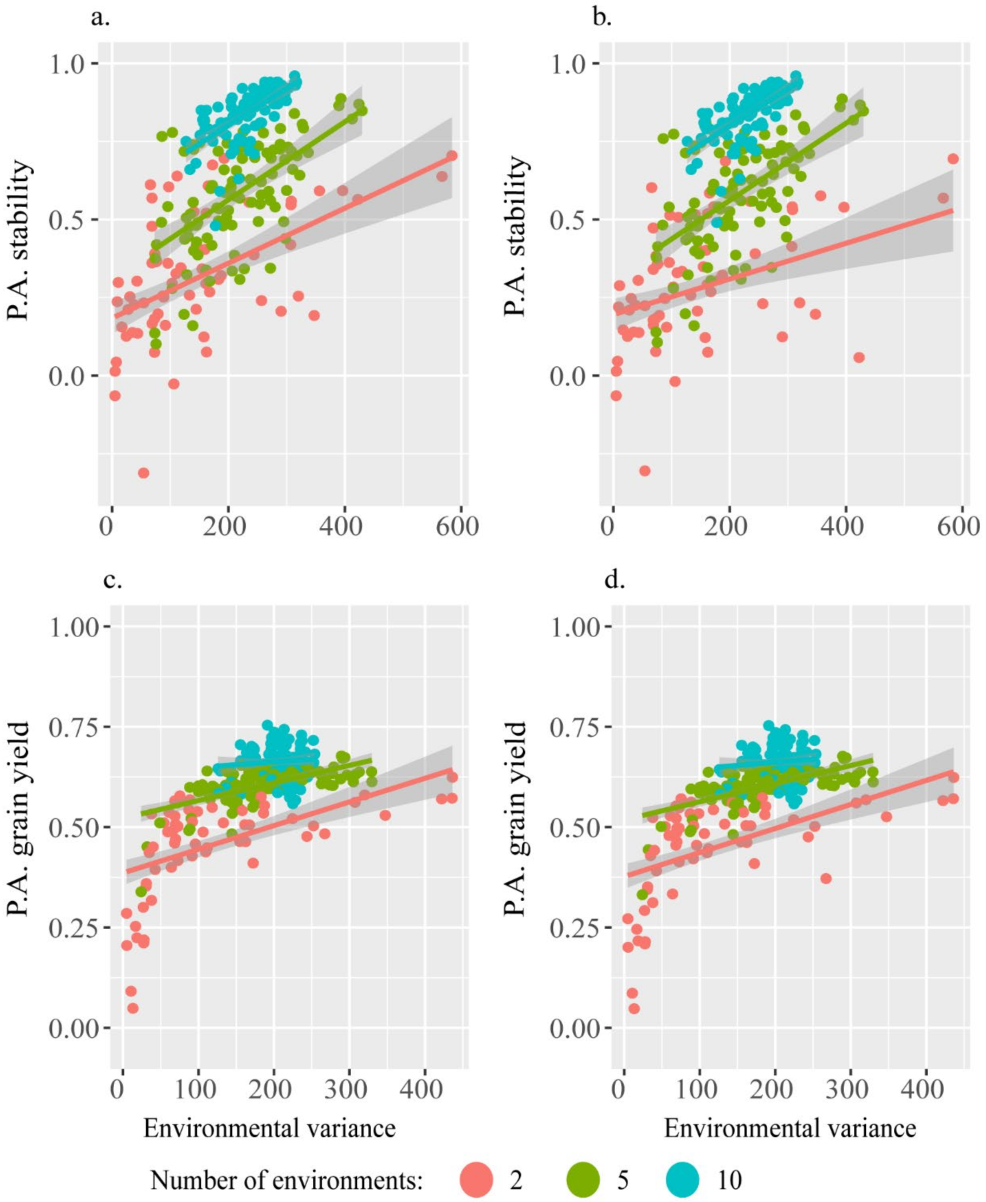


Figure 4.3. Correlation between stability for the complete data set versus stability for a partial set of environments in the wheat dataset. Number of environments in the set two (red), five (green), ten (cyan) **a.** environmental variance versus the predictive ability of the partial set of environments

cor($\tilde{\beta}_{\mathrm{i}(RFW)}$ and $\tilde{\beta}_{\mathrm{i}(p|RFW)}$). **b.** environmental variance against predictive ability as cor($\hat{\beta}_{\mathrm{i}(FW)}$ and $\hat{\beta}_{\mathrm{i}(\mathrm{p}|RFW)}$). **c.** environmental variance against the predictive ability for grain yield predicted using $\tilde{\beta}_{\mathrm{i}(p|RFW)}$ calculated as environment-centered cor($\tilde{y}_{\mathrm{i}(RFW)}$ and $\tilde{y}_{\mathrm{i}(p|RFW)}$). **d.** environmental variance against the predictive ability for grain yield predicted using $\hat{\beta}_{\mathrm{i}(\mathrm{p}|FW)}$ calculated as environment-centered cor($\hat{y}_{\mathrm{i}(FW)}$ and $\hat{y}_{\mathrm{i}(\mathrm{p}|RFW)}$).

Stability estimation and prediction

On both balanced and unbalanced datasets when we use RFW to predict the value of $\tilde{\beta}_{\mathrm{i(RFW)}}$ in new genotypes using the CV1, we were able to predict stability on unbalanced data sets genotypes with less than 20 environments showed lower predictive ability than genotypes with more phenotypic information.

Table 4.2. Range of FW estimated slope $\hat{\beta}_{\mathrm{i}(FW)}$, RFW predicted slope $\tilde{\beta}_{\mathrm{i}(RFW)}$, CV1 RFW predicted slope $\tilde{\beta}_{\mathrm{i}(cv1|RFW)}$ for barley, oats, and wheat datasets.

Dataset	$\hat{\boldsymbol{\beta}}_{\mathbf{i}(FW)}$	$\tilde{\boldsymbol{\beta}}_{\mathbf{i}(RFW)}$	$\tilde{\boldsymbol{\beta}}_{\mathbf{i}(cv1\|RFW)}$
barley	0.71 – 1.18	0.71 – 1.18	0.80 – 1.19
oats	0.01 – 1.99	0.02 – 1.98	0.73 – 1.33
wheat	0.57 – 1.22	0.57 – 1.22	0.80 – 1.13

In oats, where the dataset is highly unbalanced, and the number of environments by genotype is variable, the range of the estimated $\hat{\beta}_{\mathrm{i}(FW)}$ values and predicted $\tilde{\beta}_{\mathrm{i}(RFW)}$ values is wider than in the case of complete balanced datasets for barley and wheat (Tables 4.1 and 4.2.).

Table 4.3 Predictive ability for $\tilde{\beta}_{\mathrm{i}(cv1|RFW)}$ calculated as cor($\tilde{\beta}_{i(RFW)}$, $\tilde{\beta}_{i(cv1|RFW)}$), grain yield $\tilde{y}_{ij(cv1|RFW)}$ calculated as cor($\hat{y}_{ij}$, $\tilde{y}_{ij(cv1|RFW)}$) and predictability for $\tilde{y}_{\mathrm{i}(cv1|gBLUP)}$, calculated as cor($\hat{y}_{ij}$, $\tilde{y}_{ij(cv1|gBLUP)}$), the range for each estimator is shown between parenthesis.

Dataset	$cor(\tilde{\boldsymbol{\beta}}_{\mathrm{i(RFW)}}, \tilde{\boldsymbol{\beta}}_{\mathrm{i(CV1\|RFW)}})$	$cor(\tilde{\boldsymbol{y}}_{\mathrm{i(RFW)}}, \tilde{\boldsymbol{y}}_{\mathrm{i(CV1\|RFW)}})$	$cor(\tilde{\boldsymbol{y}}_{\mathrm{i(RFW)}}, \tilde{\boldsymbol{y}}_{\mathrm{i(CV1\|GBLUP)}})$
barley	0.37 (-0.30, 0.85)	0.24 (-0.05, 0.46)	0.07 (-0.14, 0.18)
oats	0.26 (-0.35, 0.74)	0.18 (-0.11, 0.50)	0.08 (-0.10, 0.26)
wheat	0.26 (-0.60, 0.86)	0.37 (-0.13, 0.70)	0.17 (-0.04, 0.40)

The predictive ability of the $\tilde{\beta}_{\mathrm{i(RFW)}}$ is moderate (PA 0.26 to 0.37) in the three datasets considering the highly quantitative nature of the trait. Barley had the higher PA of the three datasets for $\tilde{\beta}_{\mathrm{i(RFW)}}$. When the $\tilde{\beta}_{\mathrm{i(RFW)}}$ is used to predict the grain yield in a known environment the three data sets have better predictions than when we use GBLUP.

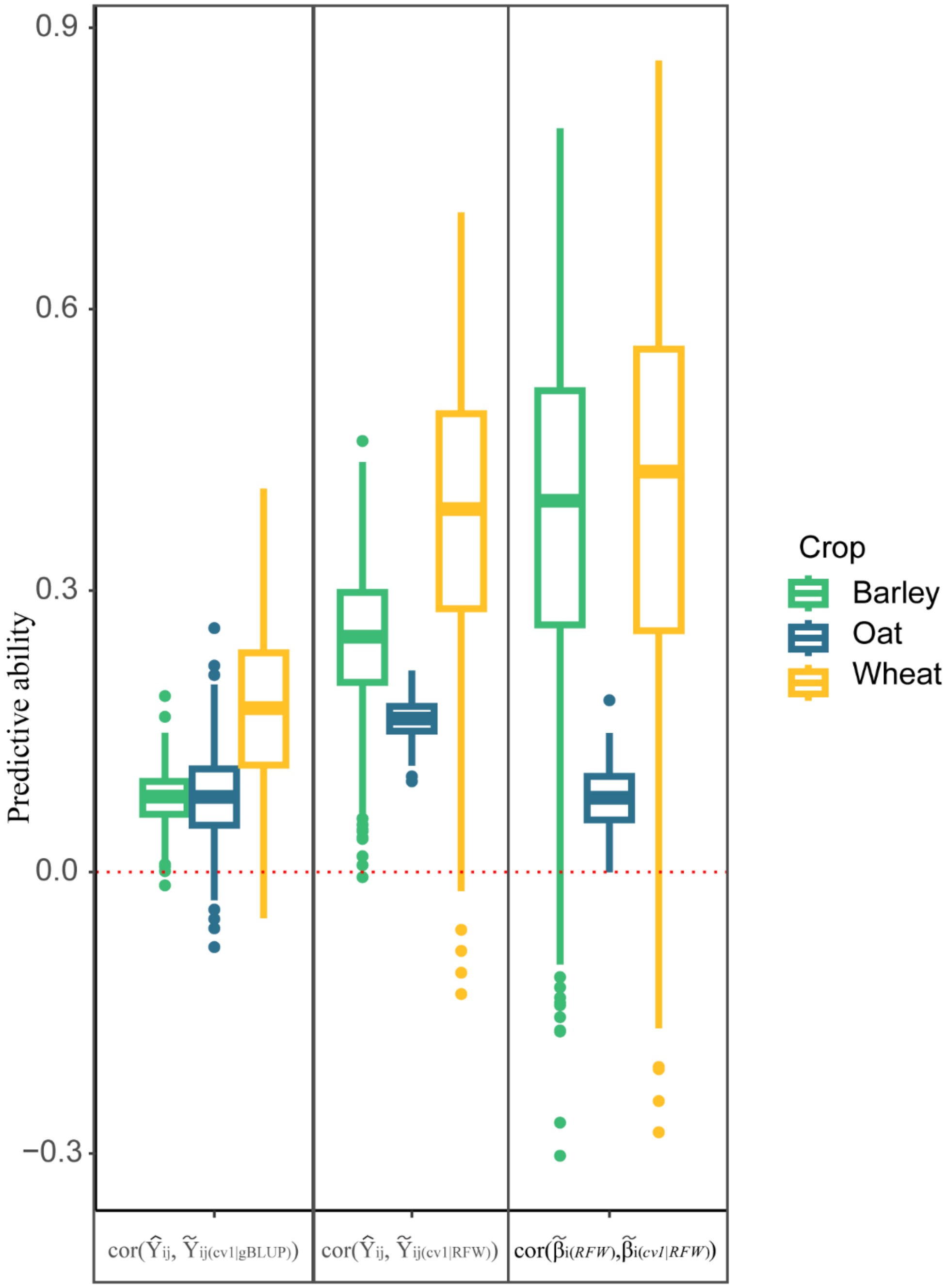


Figure 4.4 Predictive ability for $\hat{y}_{ij}$ and $\tilde{\beta}_{i(RFW)}$ in barley (green), oats (blue), and wheat (gold), for gBLUP and RFW models, with $\tilde{y}_{ij(cv1|gBLUP)}$ predictive ability as cor($\hat{y}_{ij}$, $\tilde{y}_{ij(cv1|gBLUP)}$), $\tilde{y}_{ij(cv1|RFW)}$ predictive ability as cor($\hat{y}_{ij}$, $\tilde{y}_{ij(cv1|RFW)}$), and $\tilde{\beta}_{\mathrm{i}(cv1|RFW)}$, predictive ability as cor($\tilde{\beta}_{i(RFW)}$, $\tilde{\beta}_{i(cv1|RFW)}$),.

Discussion

Our results suggest that predictions from RFW do not differ from estimations using FW when using the full data set. RFW has the advantage of allowing the prediction of grain yield and grain yield stability, in genotypes without phenotypic information. Our study found that the use of RFW opens the possibility of genomic-informed selection in the early stages of the breeding program evaluation. Early selection will allow us to select genotypes for stability during the line development stage before we observe the first grain yield results, and test only genotypes we expect to be stable, optimizing the use of resources by helping to spot stable genotypes early.

When we compare the predicted value of $\tilde{\beta}_{\mathrm{i}(cv1|RFW)}$ to $\tilde{\beta}_{\mathrm{i}(RFW)}$ we observe a shrinkage towards the mean. The predictability of $\tilde{\beta}_{\mathrm{i}(cv1|RFW)}$ values in oats 0.1 (0 – 0.2) are lower than previously reported values of highly quantitative traits like grain yield (0.25; 0.5 – 0.75, Sandro et al. 2024). The value predictability (0.37) for $\tilde{\beta}_{\mathrm{i}(cv1|RFW)}$ is in the range previously reported predictive ability for stability by Neyhart et al. 2024 in Review). When we use $\tilde{\beta}_{\mathrm{i}(cv1|RFW)}$ to predict grain yield 0.24 (-0.05, 0.46)for barley, 0.16; (0.10, 0.21) for oats, and 0.37 (-0.13, 0.70) for wheat, for each crop our result showed higher PA than using GBLUP, 0.07 (-0.14, 0.18) for barley, for oats 0.08 (-0.10, 0.26) and 0.17 (-0.04, 0.40) for wheat. The predictive abilities obtained using $\tilde{\beta}_{\mathrm{i}(cv1|RFW)}$ to model grain yield in unobserved genotypes is similar to the predictive abilities reported in the literature when environmental effects are included in the model (0.2-0.8, Hoefler et al., 2020; 0.25-0.58, Lado et al., 2016; 0.65, and Rebollo et al., 2023).

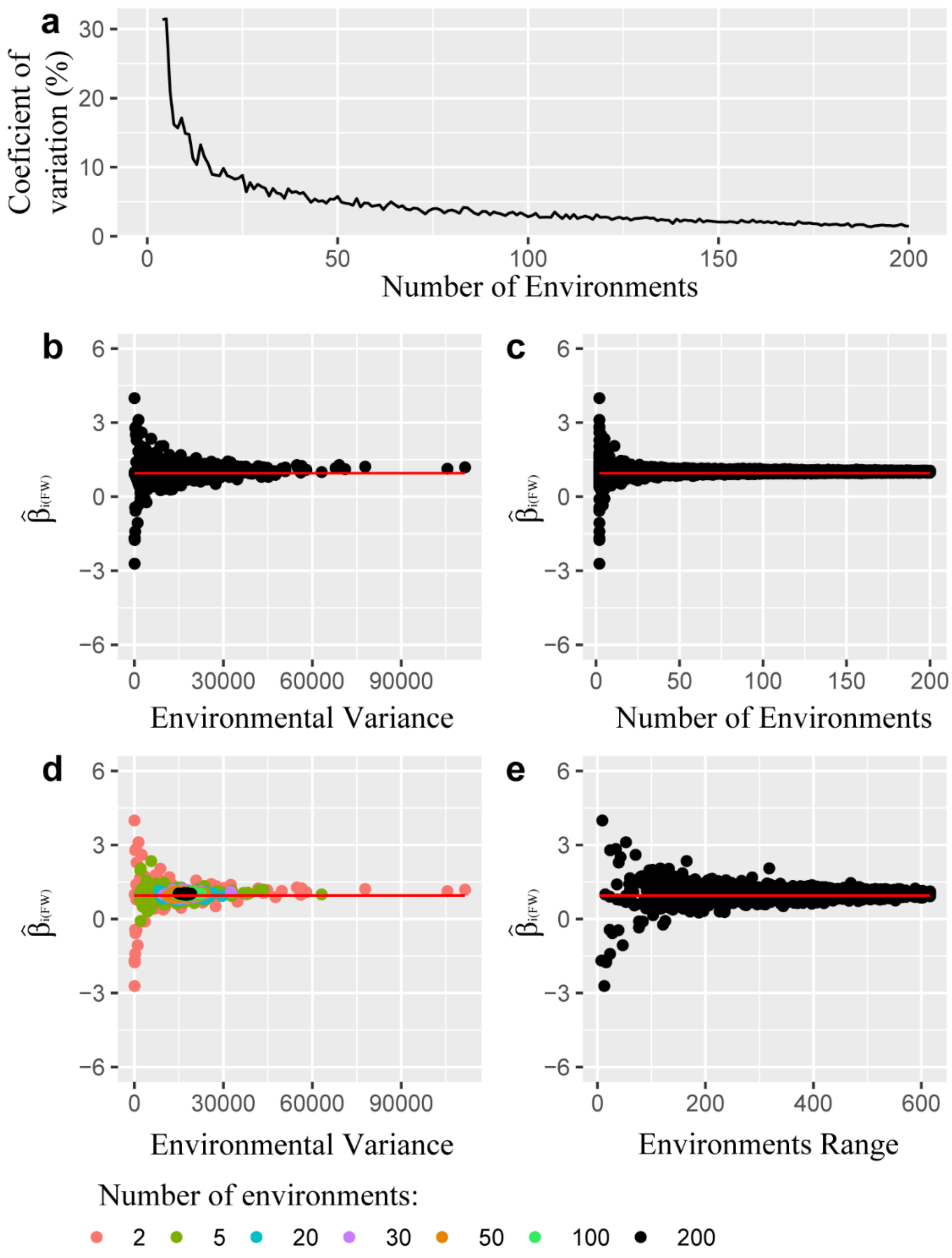
Breeding programs are in a perpetual struggle to maximize the use of resources to obtain the best possible phenotypic information in as many genotypes as possible in the shortest possible time (González-Barrios et al., 2019; Hoefler et al., 2020; Lado et al., 2018; Rebollo et al., 2023). The heritability of the trait defines how intensive phenotyping is required to differentiate cultivars that meet the requirements for the target population of environments (Hoefler et al., 2020). Quantitative traits demand the highest levels and quality in phenotyping (Bernardo, 2010). Grain yield stability is a challenging trait to estimate due to the quantitative nature of grain yield and the number of environments that are required to achieve a good characterization (Wang et al., 2023). When designing multi-environment trials (MET) to characterize GEI, repeatable and unrepeatable environmental factors should be considered in GEI characterization (Van Eeuwijk et al., 2016). It is necessary to test genotypes in several environments, including a combination of locations and years, to identify the adaptation of the genotype to the environment that is being targeted (González-Barrios et al., 2019; Sandro et al., 2022). The locations are considered repeatable, on variety trials researchers usually pick research stations in their target population-of-environments and the options to change locations are limited. On the other hand, years are not repeatable and can be accumulated for a location, and different arranges of years and locations can be used for GEI characterization (Kleinknecht et al., 2008). The year component is a very important part of the environmental variance in multi-environment trials (Laidig et al., 2008). An extreme year can greatly impact the results when the number of environments is small (i.e. ≤ six) (Westcott, 1986). Extreme environments in general have an important effect on the slope of the linear regression (Westcott, 1986). Yan et al. (2015) established a method to estimate how many environments are necessary to correctly characterize GEI. The number of environments varies based on the trait (genetic variance, heritability), the quality of the trials (error variance, heritability), the mega environment structure (environmental variance, heritability), and the precision threshold. In oats, Yan et al.,(2015) found that on average 21 (15 to 34) is the number of environments necessary to characterize GEI for grain yield in undivided mega environments (Yan et al., 2015). When mega-environments are considered, the number of environments is reduced to between 4 and 6 environments per mega-environment due to an increase in heritability within ME (Yan et al. 2015). Simulation studies for physiological traits in wheat found that at least 150 environments are needed accurately calculate stability (Wang et al. 2023). Our results are aligned with Yan et al. (2015) on the three crops we studied $\hat{\beta}_{i(FW)}$ stabilizes and converges around $\hat{\beta}_{i(FW)}$ calculated using the full data set after 20 to 30 environments are used in the estimation. A limitation of our results is that we are working with breeding populations, where stable genotypes may be overrepresented due to high selection pressure. Released cultivars went through an exhaustive process of evaluation in several years and locations and have a proven record of good performance in the intended target population of environments. Those lines are probably stable and consistently respond to the environment. On the other hand, breeding lines that did not yield at the top of ranking on all environments will show a less stable performance and therefore are discarded in the selection process. There may also be an overrepresentation of stable genotypes in historical datasets, which also have more phenotypic information and may cause a bias overestimating grain yield stability.

Our results show that both the number of environments and the environmental variance affect β estimation and by extension the quality of yield prediction. After including at least 20 environments, the differences in $\hat{\beta}_{i(FW)}$ by adding another environment is practically not significant. When the number of environments available is limited (i.e. between two and five environments), it is important to use the combination of environments that will provide the maximum variance of the means of the environments because it will give better estimation. This implies that a breeding program can capitalize on doing stability characterization by testing in highly informative environments into the target population of environments early on. For example, by using two environments with a high environmental variance, a large set of genotypes can be screened in the early stages. Also, the effect of the environmental variance suggests that selecting environments that maximize the environmental variance is the best strategy to extract the most information per environment added in the early stages. As the information is accumulated and more environments are added, the marginal improvements from single environments and their effects are reduced, and the overall quality of the estimation increases.

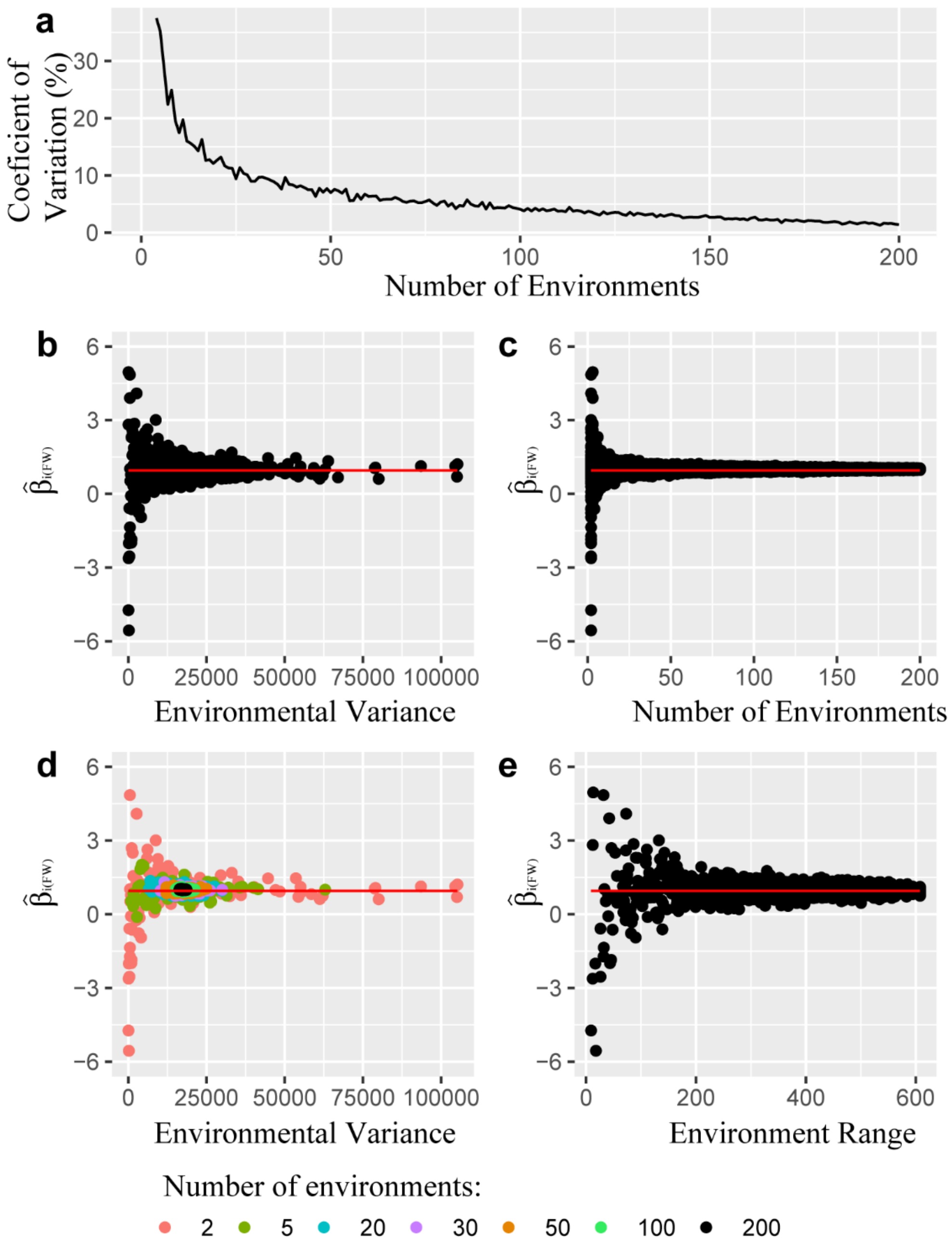
Conclusions

In conclusion, when we have phenotypic information available, FW and RFW do not differ in their values of $\hat{\beta}_{i(FW)}$ and $\tilde{\beta}_{\mathrm{i}(RFW)}$. Stability in new genotypes without grain yield information can be predicted using RFW. Grain yield predictions showed higher predictive abilities when we compared RFW to those obtained using GBLUP. Our results suggest that by adding more environments we could improve stability estimations up to 20 or 30 environments. After the 30 environments are tested the estimations converge and do not change. Moreover, when fewer than 20 environments are available, the contrast between testing environments is the most important factor affecting the estimation.

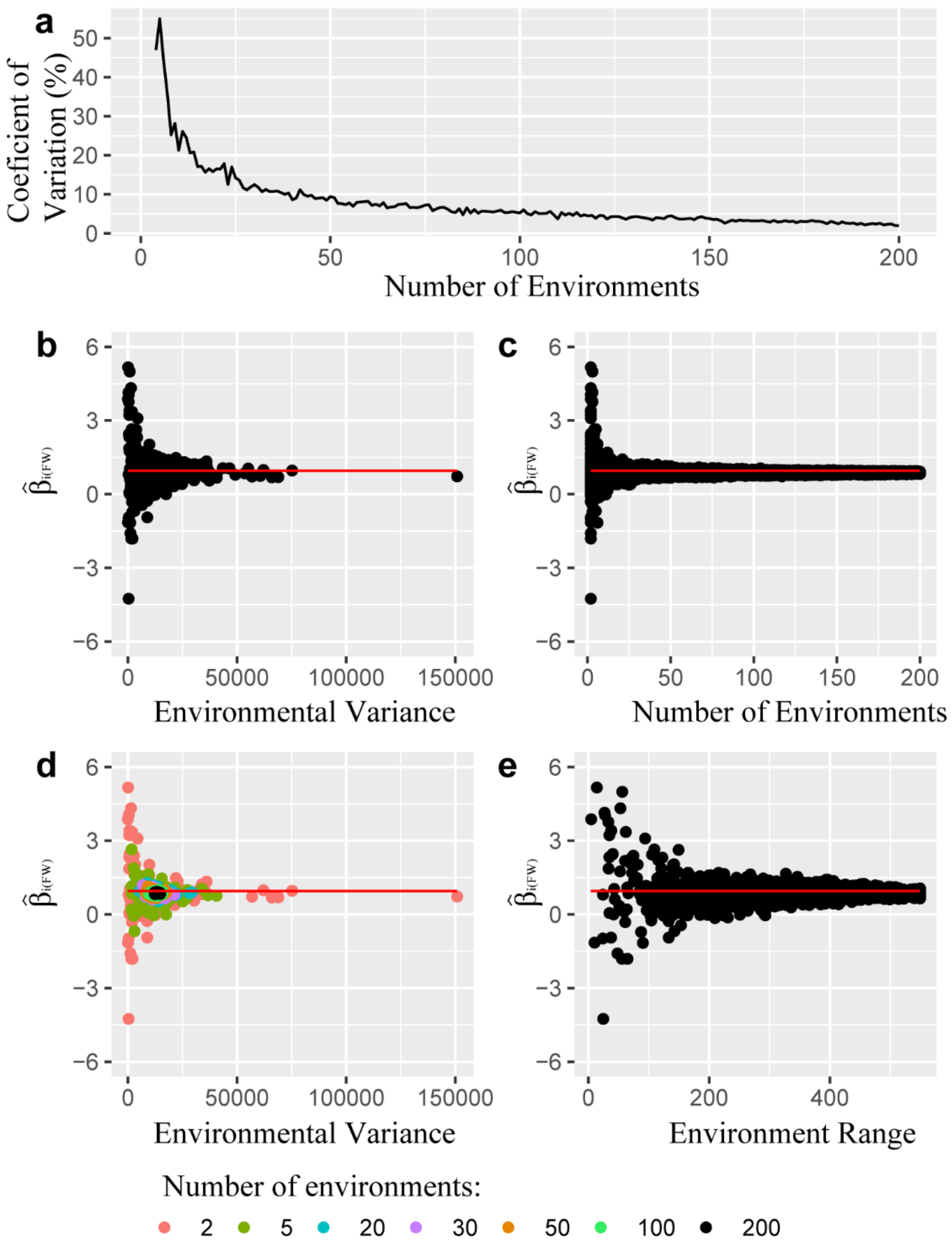
Annex 3



Supplementary Figure 4.1. Deon's $\hat{\beta}_{i(p|FFW)}$ variation on number of environments, environmental variance, and range of the set of environments used to estimate $\hat{\beta}_{i(p|FW)}$. The red lines represent the value of $\hat{\beta}_{i(p|FW)}$(1.05) estimated using all environments available for the line (235 environments). **a.** Coefficient of variation for $\hat{\beta}_{i(p|FW)}$ per number of environments used to fit the model. **b.** $\hat{\beta}_{i(p|FW)}$ against environmental variance for the environments used to fit the model. **c.** $\hat{\beta}_{i(p|FW)}$ against the number of environments used to fit the model. **d.** $\hat{\beta}_{i(p|FW)}$ environmental variance for 100 sets of 2, 5, 20, 30, 50, 100, and 200 environments used to fit the model. **e.** $\hat{\beta}_{i(p|FW)}$ against the range of environments used to fit the model.



Supplementary Figure .2. Horsepower's $\hat{\beta}_{i(p|FW)}$ variation on number of environments, environmental variance, and range of the set of environments used to fit the model. The red lines represent the value of $\hat{\beta}_{i(p|FW)}$ (1.00) estimated using all environments available for the line (232 environments). **a.** Coefficient of variation for $\hat{\beta}_{i(p|FW)}$ per number of environments used to fit the model. **b.** $\hat{\beta}_{i(p|FW)}$ against environmental variance for the environments used to fit the model. **c.** $\hat{\beta}_{i(p|FW)}$ against the number of environments used to fit the model. **d.** $\hat{\beta}_{i(p|FW)}$ environmental variance for 100 sets of 2, 5, 20, 30, 50, 100, and 200 environments used to fit the model. **e.** $\hat{\beta}_{i(p|FW)}$ against the range of environments used to fit the model.



Supplementary Figure 4.3. Clintland64's $\hat{\beta}_{i(p|FW)}$ variation on number of environments, environmental variance, and range of the set of environments used to fit the model. The red lines represent the value of $\hat{\beta}_{i(p|FW)}$ (0.87) estimated using all environments available for the line (265 environments). **a.** Coefficient of variation for $\hat{\beta}_{i(p|FW)}$ per number of environments used to fit the model. **b.** $\hat{\beta}_{i(p|FW)}$ against environmental variance for the environments used to fit the model. **c.** $\hat{\beta}_{i(p|FW)}$ against the number of environments used to fit the model. **d.** $\hat{\beta}_{i(p|FW)}$ environmental variance for 100 sets of 2, 5, 20, 30, 50, 100, and 200 environments used to fit the model. **e.** $\hat{\beta}_{i(p|FW)}$ against the range of environments used to fit the model.

References

Alotaibi, M. (2023). Climate change, its impact on crop production, challenges, and possible solutions. Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 51(1), 13020. https://doi.org/10.15835/nbha51113020

Annicchiarico, P. (2002). Genotype x environment interactions: Challenges and opportunities for plant breeding and cultivar recommendations. https://books.google.com/books?hl=en&lr=&id=so4UJSzWNbIC&oi=fnd&pg=PR8&dq=Genotype+x+environment+interactions:+Challenges+and+opportunities+for+plant+breeding+and+cultivar+recommendations&ots=dD_pjzx8S5&sig=URiP9ZDJN3XK_V_cNWIG929YYJI

Bacsi, Z., & Hollósy, Z. (2019). A yield stability index and its application for crop production. Analecta Technica Szegedinensia, 13(1), 11–20. https://doi.org/10.14232/analecta.2019.1.11-20

Becker, H. C., & Léon, J. (1988). Stability Analysis in Plant Breeding. Plant Breeding, 101(1), 1–23. https://doi.org/10.1111/j.1439-0523.1988.tb00261.x

Bernardo, R. (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Science, 34(1), 20–25.

Bernardo, R. (2014). *Essentials of plant breeding*. Stemma Press, Woodbury, MN.

Bocci, R., Bussi, B., Petitti, M., Franciolini, R., Altavilla, V., Galluzzi, G., Di Luzio, P., Migliorini, P., Spagnolo, S., Floriddia, R., Rosi, G. L., Petacciato, M., Battezzato, V., Albino, A., Faggio, G., Arcostanzo, C., & Ceccarelli, S. (2020). Yield, yield stability and farmers' preferences of evolutionary populations of bread wheat: A dynamic solution to climate change. European Journal of Agronomy, 121, 126156. https://doi.org/10.1016/j.eja.2020.126156

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2020). ASReml-R reference manual. Version 4.1. 0.130. VSN International Ltd.

Burgueño, J., de los Campos, G., Weigel, K., & Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype× environment interaction using pedigree and dense molecular markers. Crop Science, 52(2), 707–719.

Carvalho, H. F., Rio, S., García-Abadillo, J., & Isidro Y Sánchez, J. (2024). Revisiting superiority and stability metrics of cultivar performances using genomic data: Derivations of new estimators. Plant Methods, 20(1), 85. https://doi.org/10.1186/s13007-024-01207-1

Endelman, J. B., Atlin, G. N., Beyene, Y., Semagn, K., Zhang, X., Sorrells, M. E., & Jannink, J.-L. (2014). Optimal design of preliminary yield trials with genome-wide markers. Crop Science, 54(1), 48–59.

Finlay, K., & Wilkinson, G. (1963). The analysis of adaptation in a plant-breeding programme. Australian Journal of Agricultural Research, 14(6), 742. https://doi.org/10.1071/AR9630742

Francis, T. R., & Kannenberg, L. W. (1978). YIELD STABILITY STUDIES IN SHORT-SEASON MAIZE. I. A DESCRIPTIVE METHOD FOR GROUPING GENOTYPES. Canadian Journal of Plant Science, 58(4), 1029–1034. https://doi.org/10.4141/cjps78-157

Gauch, H. G., & Zobel, R. W. (1996). Optimal Replication in Selection Experiments. Crop Science, 36(4), 838–843. https://doi.org/10.2135/cropsci1996.0011183X003600040002x

Gauch, J. (1992). Statistical analysis of regional yield trials: AMMI analysis of factorial designs. https://www.cabidigitallibrary.org/doi/full/10.5555/19931643324

González-Barrios, P., Díaz-García, L., & Gutiérrez, L. (2019). Mega-environmental design: Using genotype× environment interaction to optimize resources for cultivar testing. Crop Science, 59(5), 1899–1915.

Hoefler, R., González-Barrios, P., Bhatta, M., Nunes, J. A. R., Berro, I., Nalin, R. S., Borges, A., Covarrubias, E., Diaz-Garcia, L., Quincke, M., & Gutierrez, L. (2020). Do Spatial Designs Outperform Classic Experimental Designs? Journal of Agricultural, Biological and Environmental Statistics, 25(4), 523–552. https://doi.org/10.1007/s13253-020-00406-2

IPCC. (2023). Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157896

Lado, B., Barrios, P. G., Quincke, M., Silva, P., & Gutiérrez, L. (2016). Modeling genotype× environment interaction for genomic selection with unbalanced data from a wheat breeding program. Crop Science, 56(5), 2165–2179.

Lado, B., Vázquez, D., Quincke, M., Silva, P., Aguilar, I., & Gutiérrez, L. (2018). Resource allocation optimization with multi-trait genomic prediction for bread wheat (Triticum aestivum L.) baking quality. Theoretical and Applied Genetics, 131, 2719–2731.

Laidig, F., Drobek, T., & Meyer, U. (2008). Genotypic and environmental variability of yield for cultivars from 30 different crops in German official variety trials. Plant Breeding, 127(6), 541–547. https://doi.org/10.1111/j.1439-0523.2008.01564.x

Lin, C. S., & Binns, M. R. (1988). A method of analyzing cultivar • location • year experiments: A new stability parameter.76: 425-430 Theorical Applied Genetics

Ly, D., Huet, S., Gauffreteau, A., Rincent, R., Touzy, G., Mini, A., Jannink, J.-L., Cormier, F., Paux, E., Lafarge, S., Le Gouis, J., & Charmet, G. (2018). Whole-genome prediction of reaction norms to environmental stress in bread wheat (Triticum aestivum L.) by genomic random regression. Field Crops Research, 216, 32–41. https://doi.org/10.1016/j.fcr.2017.08.020

Macholdt, J., & Honermeier, B. (2017). Yield Stability in Winter Wheat Production: A Survey on German Farmers' and Advisors' Views. Agronomy, 7(3), Article 3. https://doi.org/10.3390/agronomy7030045

Mohammadi, R., & Amri, A. (2013). Genotype × environment interaction and genetic improvement for yield and yield stability of rainfed durum wheat in Iran. Euphytica, 192(2), 227–249. https://doi.org/10.1007/s10681-012-0839-1

Mühleisen, J., Piepho, H.-P., Maurer, H. P., Zhao, Y., & Reif, J. C. (2014). Exploitation of yield stability in barley. Theoretical and Applied Genetics, 127(9), 1949–1962. https://doi.org/10.1007/s00122-014-2351-6

Neyhart, J. L., Sweeney, D., Sorrells, M., Kapp, C., Kephart, K. D., Sherman, J., Stockinger, E. J., Fisk, S., Hayes, P., Daba, S., Mohammadi, M., Hughes, N., Lukens, L., Barrios, P. G., Gutiérrez, L., & Smith, K. P. (2019). Registration of the S2MET Barley Mapping Population for Multi-Environment Genomewide Selection. Journal of Plant Registrations, 13(2), 270–280. https://doi.org/10.3198/jpr2018.06.0037crmp

Paniza, H. M. (2024). Challenges in Plant Breeding Under Climate Change: A Review. In K. A. Abd-Elsalam & S. M. Abdel-Momen (Eds.), Plant Quarantine Challenges under Climate Change Anxiety (pp. 533–556). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-56011-8_17

Picasso, V. D., Casler, M. D., & Undersander, D. (2019). Resilience, Stability, and Productivity of Alfalfa Cultivars in Rainfed Regions of North America. Crop Science, 59(2), 800–810. https://doi.org/10.2135/cropsci2018.06.0372

Piepho, H., & Möhring, J. (2006). Selection in Cultivar Trials—Is It Ignorable? Crop Science, 46(1), 192–201. https://doi.org/10.2135/cropsci2005.04-0038

Piepho, H.-P., & Blancon, J. (2023). Extending Finlay–Wilkinson regression with environmental covariates. Plant Breeding, 142(5), 621–631. https://doi.org/10.1111/pbr.13130

Pour-Aboughadareh, A., Khalili, M., Poczai, P., & Olivoto, T. (2022). Stability Indices to Deciphering the Genotype-by-Environment Interaction (GEI) Effect: An Applicable Review for Use in Plant Breeding Programs. Plants, 11(3), 414. https://doi.org/10.3390/plants11030414

Rebollo, I., Aguilar, I., Pérez De Vida, F., Molina, F., Gutiérrez, L., & Rosas, J. E. (2023). Genotype by environment interaction characterization and its modeling with random regression to climatic variables in two rice breeding populations. Crop Science, 63(4), 2220–2240. https://doi.org/10.1002/csc2.21029

Reckling, M., Ahrends, H., Chen, T.-W., Eugster, W., Hadasch, S., Knapp, S., Laidig, F., Linstädter, A., Macholdt, J., Piepho, H.-P., Schiffers, K., & Döring, T. F. (2021). Methods of yield stability analysis in long-term field experiments. A review. Agronomy for Sustainable Development, 41(2), 27. https://doi.org/10.1007/s13593-021-00681-4

Rincent, R., Malosetti, M., Ababaei, B., Touzy, G., Mini, A., Bogard, M., Martre, P., Le Gouis, J., & van Eeuwijk, F. (2019). Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. Theoretical and Applied Genetics, 132(12), 3399–3411. https://doi.org/10.1007/s00122-019-03432-y

Sandro, P., Bhatta, M., Bower, A., Carlson, S., Jannink, J.-L., Waring, D. J., Birkett, C., Smith, K., Wiersma, J., Caffe, M., Kleinjan, J., McMullen, M. S., English, L., & Gutierrez, L. (2024). Genomic prediction for targeted populations of environments in oat (Avena sativa). Crop & Pasture Science, 75(5). https://doi.org/10.1071/CP23126

Sandro, P., Kucek, L. K., Sorrells, M. E., Dawson, J. C., & Gutierrez, L. (2022). Developing high-quality value-added cereals for organic systems in the US Upper Midwest: Hard red winter wheat (Triticum aestivum L.) breeding. Theoretical and Applied Genetics, 135(11), 4005–4027. https://doi.org/10.1007/s00122-022-04112-0

Smith, A., Cullis, B., & Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics, 57(4), 1138–1147.

Team, R. D. C. (2017). R Core Team. R Foundation for Statistical Computing Austria.

Van Eeuwijk, F. A., Bustos-Korts, D. V., & Malosetti, M. (2016). What should students in plant breeding know about the statistical aspects of genotype× environment interactions? Crop Science, 56(5), 2119–2140.

Wang, T.-C., Casadebaig, P., & Chen, T.-W. (2023). More than 1000 genotypes are required to derive robust relationships between yield, yield stability and physiological parameters: A computational study on wheat crop. Theoretical and Applied Genetics, 136(3), 34. https://doi.org/10.1007/s00122-023-04264-7

Westcott, B. (1986). Some methods of analysing genotype—Environment interaction. Heredity, 56(2), 243–253. https://doi.org/10.1038/hdy.1986.37

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Oxford University Press. https://academic.oup.com/biometrics/article-abstract/67/2/678/7381027

Wricke, G. (1962). Uber eine methode zur erfassung der okologischen streubreite in feldversucen. Z. Pflanzenzuchtung, 47, 92–96.

Xiong, W., Reynolds, M., & Xu, Y. (2022). Climate change challenges plant breeding. Current Opinion in Plant Biology, 70, 102308. https://doi.org/10.1016/j.pbi.2022.102308

Yan, W., Frégeau-Reid, J., Martin, R., Pageau, D., & Mitchell-Fetch, J. (2015). How many test locations and replications are needed in crop variety trials for a target region? Euphytica, 202(3), 361–372. https://doi.org/10.1007/s10681-014-1253-7

Yan, W., Pageau, D., Frégeau-Reid, J., & Durand, J. (2011). Assessing the Representativeness and Repeatability of Test Locations for Genotype Evaluation. Crop Science, 51(4), 1603–1610. https://doi.org/10.2135/cropsci2011.01.0016

Yan, W., & Kang, M. S. (2002). GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC press.

Chapter 5. General Discussion

Breeding and new technologies have a role in delivering stable cultivars that are necessary to achieve sustainable agriculture (Zhang et al. 2018; Isidro-Sánchez et al. 2020). In the years to come, breeders will need to deliver integrative cultivars to meet farmers' needs and adapt to the agricultural landscape and climate challenges (Bassi et al. 2024). Across this dissertation, we applied participatory breeding, genomic prediction, and stability indices intended to meet the requirements of different farming systems. We aim to deliver cultivars and decision-making information, integrating genotype by management, and genotype by environment interactions. The goal was to improve cultivar stability as a step towards the sustainability of the production systems.

In Chapter 2 we used consultative participatory breeding to develop a winter wheat variety. This was done by integrating farmers and final users in the process of selection. The culmination of this collaborative breeding cycle will be the release of a new cultivar "Bickford" (experimental line 260.06). In addition, we have developed a breeding population that has promising lines to be used as parents to the next generations of breeding lines.

The release of "Bickford" represents the work of a multitude of people and several years of constant research. Still, this is just the beginning. Plant breeding is a continuous effort where we select and breed new cultivars to withstand biotic and abiotic stress every year. To advance the breeding process and genetic gain, we need to start new recombination and selection cycles every year. We need to take advantage of the genetic resources generated in this project during the selection process using promissory lines as parents of new cycles of selection. We need to add new sources of disease resistance, agronomics, and baking quality to our breeding population to increase genetic diversity and recombination. We also must shorten the breeding cycle, to reduce the time between crosses and the evaluation of the cultivars. Also, climatic uncertainty will demand an increase in the speed at which cultivars adapted to new conditions are delivered (Xiong et al., 2022). Nowadays modern breeding can accelerate the delivery of new cultivars by integrating several sources of information, genomics, phenomics, and environmental data (Crossa et al. 2021).

One important aspect of the cultivar “Bickford” is the integration of on-farm trials and artisanal baking as inputs in the selection process. In our case we evaluated using artisanal baking only a few elite lines, this evaluation happened late in the selection cycle. Artisanal baking requires a minimum of 2.5 kg of flour, per genotype tested, and a group of bakers can test around 4 or 5 experimental lines in two days. Grain volume is limited in the early stages of the breeding cycle and artisanal bakers' time is finite. We should find ways to improve the testing for artisanal baking in the early stages of the breeding process with a few kilograms of seed and increase the number of lines evaluated. There is a gap in the knowledge about how useful tests are designed to evaluate baking quality in industrial settings to inform about the performance of artisanal baking. Evaluating artisanal baking quality in the early stages of the breeding cycle will help to improve the selection process and allow for the incorporation of the baking quality lines as parents early on, which will improve baking quality genetic gain. Early testing would help us also to identify lines that have other attributes, such as flavor and aromas that could be lost by chance or negative association with other characteristics. To this end, we need to design testing schemes that allow the interaction with bakers with small flour batches. Another aspect is to identify which traditional baking tests are informative in artisanal baking, and which is the correlation structure between traits to be integrated into the selection process as an indirect selection scheme.

In Chapter 3 we used quantitative methods to facilitate access of cultivar ranking information to farmers from the Midwest of the United States. We integrated genomics, genotype by environment modeling, and historical databases to define mega environments and create a ranking of cultivars for each mega environment to recommend to farmers. We matched the zip codes to the mega environments and their cultivar rankings. We uploaded this information to the variety selection tool webpage where farmers can use their farm zip code to access the list of the 10 best cultivars for ME where the farm is located. This makes information available to any farmer with a computer or mobile device able to access clear and easy-to-read information, thus facilitating the selection of cultivars to be planted.

We demonstrated that historical data sets can be used to predict cultivar performance and help farmers access information. Further work should aim to work on the extension and validation of the farmers' fields. On-farm trials would have three purposes, extension, validation, and adding high-quality data. Demonstrating to farmers and extension agents that the prediction for their farms is valid will help with the adoption and use of the tool. Validation, testing the model, and adding the new data will help to further improve the model. Adding a network of farms will provide improved geographical coverage which would help us to better understand the mega-environments and how environmental variation is distributed across the region. To achieve this we should plan a mega environment design trial (González-Barrios et al. 2019) to integrate market-available cultivars, and experimental lines to further improve the prediction model. If we combine this information with environmental data (Jarquín et al. 2017), potentially will help to define the limits and transition zones between mega environments (Bustos-Korts et al. 2021).

We modeled one trait at a time, using a multi-trait model (Crossa and Cerón-Rojas 2011; VanRaden et al. 2014), would improve predictions for sparse traits such as disease resistance (we do not have conditions to the expression of a particular disease every year). Finally, a prediction model for all relevant traits for farmers and researchers will help to accelerate genetic gain and reduce the risk of selecting a cultivar for the farmer.

In Chapter 4 we evaluated the use of Finlay-Wilkinson random regression (RFW) and traditional Finlay-Wilkinson regression (FFW) to analyze yield stability on different databases. We compared the predictive ability of RFW for yield stability and grain yield on the CV1 scheme using phenotypic information from their relatives. We evaluate the effect of the number of environments over the stability prediction whereas a rule of thumb of over 25 environments is enough to estimate stability accurately. This will potentially allow us to perform genomic selection in the early stages of a breeding program.

We proposed to integrate genomic information into Finlay-Wilkinson random regression to predict stability and grain yield. This can be incorporated into the breeding program selection pipeline adding grain yield stability as another of the traits in the selection index (Crossa et al. 2022). How much value is assigned to grain yield in the index will be defined by the importance of the grain yield stability for the target population of environments. The methodology applied in Chapter 4 will take advantage of information generated by the continuation of Chapter 3 which will be useful in continuing working on this topic. For example, incorporating on-farm trials in the RFW model would be a valuable resource. Combining on-farm data, genomic information, and environmental covariates in a Finlay-Wilkinson model (Piepho and Blancon 2023), would allow us to predict yield at the regional or even farm level (if environmental covariates are in the range of the regression (Malosetti et al. 2013). Another aspect that may be worth studying is real-time yield prediction. Using environmental information from the first half of the growing season to predict grain yield. This yield forecast will allow the farmer to evaluate economic margins and make an early decision on harvest forage or grain. Defining other traits' stability and incorporating this into the selection indices is a new challenge that maybe is worth exploring in the future.

In conclusion, we have successfully met the objective of delivering cultivars and decision-making tools that integrate genotype-by-management and genotype-by-environment interactions. While these projects are not silver bullets, they represent important steps toward a more sustainable agricultural future. We remain committed to advancing this work, continuously improving our results, and contributing through plant breeding towards agricultural sustainability.

References

Bassi, F. M., Sanchez-Garcia, M., & Ortiz, R. (2024). What plant breeding may (and may not) look like in 2050? The Plant Genome, 17(1), e20368. https://doi.org/10.1002/tpg2.20368

Bustos-Korts, D., Boer, M. P., Chenu, K., Zheng, B., Chapman, S., & Van Eeuwijk, F. A. (2021). Genotype-specific P-spline response surfaces assist interpretation of regional wheat adaptation to climate change. In Silico Plants, 3(2), diab018.

Crossa, J., & Cerón-Rojas, J. J. (2011). Multi-trait Multi-environment Genome-wide Molecular Marker Selection Indices. Journal of the Indian Society of Agricultural Statistics.

Crossa, J., Cerón-Rojas, J. J., Martini, J. W., Covarrubias-Pazaran, G., Alvarado, G., Toledo, F. H., & Govindan, V. (2022). Theory and practice of phenotypic and genomic selection indices. Wheat Improvement, 593.

Crossa, J., Fritsche-Neto, R., Montesinos-Lopez, O. A., Costa-Neto, G., Dreisigacker, S., Montesinos-Lopez, A., & Bentley, A. R. (2021). The Modern Plant Breeding Triangle: Optimizing the Use of Genomics, Phenomics, and Enviromics Data. Frontiers in Plant Science, 12, 651480. https://doi.org/10.3389/fpls.2021.651480

González-Barrios, P., Díaz-García, L., & Gutiérrez, L. (2019). Mega-environmental design: Using genotype× environment interaction to optimize resources for cultivar testing. Crop Science, 59(5), 1899–1915.

Isidro-Sánchez, J., Prats, E., Howarth, C., Langdon, T., & Montilla-Bascón, G. (2020). Genomic Approaches for Climate Resilience Breeding in Oats. In C. Kole (Ed.), Genomic Designing of Climate-Smart Cereal Crops (pp. 133–169). Springer International Publishing. https://doi.org/10.1007/978-3-319-93381-8_4

Jarquín, D., Lemes da Silva, C., Gaynor, R. C., Poland, J., Fritz, A., Howard, R., Battenfield, S., & Crossa, J. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype× environment interactions in Kansas wheat. The Plant Genome, 10(2), plantgenome2016.12.0130.

Malosetti, M., Ribaut, J.-M., & van Eeuwijk, F. A. (2013). The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. Frontiers in Physiology, 4, 44.

Piepho, H.-P., & Blancon, J. (2023). Extending Finlay–Wilkinson regression with environmental covariates. Plant Breeding, 142(5), 621–631. https://doi.org/10.1111/pbr.13130

VanRaden, P. M., Tooker, M. E., Wright, J. R., Sun, C., & Hutchison, J. L. (2014). Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. Journal of Dairy Science, 97(12), 7952–7962. https://doi.org/10.3168/jds.2014-8489

Zhang, H., Li, Y., & Zhu, J.-K. (2018). Developing naturally stress-resistant crops for sustainable agriculture. Nature Plants, 4(12), 989–996. https://doi.org/10.1038/s41477-018-0309-4

Xiong, W., Reynolds, M., & Xu, Y. (2022). Climate change challenges plant breeding. Current Opinion in Plant Biology, 70, 102308. https://doi.org/10.1016/j.pbi.2022.102308