Inference, exploration, and prediction in hard red winter wheat: multi-environment genomic-to-phenotype framework

by

Lucas Alexandre Batista

B.S., Universidade Federal De Lavras, 2019
M.S., North Dakota State University, 2023

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agronomy – Interdepartmental Genetics Program
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Abstract

Wheat breeding must deliver genetic gain while coping with strong environmental variability, polygenic trait architectures, and rapidly expanding data streams. The framework for breeding analytics organizes methods around inference, exploration, and prediction, emphasizing that model usefulness depends on alignment between the biological question, the available data, and realistic validation. Using hard red winter wheat as the primary system, this dissertation first quantifies genotype-by-environment interaction and genetic progress from long-term multi-environment trials in Kansas, showing how models and stability analyses can separate genetic signal from environmental and design noise to support cultivar selection and deployment. It then investigates the genetic architecture of key developmental, floral, and disease traits in a D-genome–enriched population derived from *Aegilops tauschii*, integrating controlled greenhouse and field phenotyping. Finally, it evaluates multi-omics prediction across locations, years, and new genotypes by combining genomics with environmental and phenotyping features under breeding-relevant validation schemes, demonstrating that the value of added data layers and model complexity is scenario- and trait-dependent. Collectively, these results support an integrated, goal-driven approach to wheat improvement that couples robust multi-environment inference with hypothesis-guided discovery and deployment-oriented prediction, providing actionable guidance for selection, parent choice, and cultivar placement.

Inference, exploration, and prediction in hard red winter wheat: multi-environment genomic-to-phenotype framework

by

Lucas Alexandre Batista

B.S., Universidade Federal De Lavras, 2019
M.S., North Dakota State University, 2023

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agronomy – Interdepartmental Genetics Program
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Approved by:

Major Professor
Dr. Allan K. Fritz

Copyright

Abstract

Wheat breeding must deliver genetic gain while coping with strong environmental variability, polygenic trait architectures, and rapidly expanding data streams. The framework for breeding analytics organizes methods around inference, exploration, and prediction, emphasizing that model usefulness depends on alignment between the biological question, the available data, and realistic validation. Using hard red winter wheat as the primary system, this dissertation first quantifies genotype-by-environment interaction and genetic progress from long-term multi-environment trials in Kansas, showing how models and stability analyses can separate genetic signal from environmental and design noise to support cultivar selection and deployment. It then investigates the genetic architecture of key developmental, floral, and disease traits in a D-genome–enriched population derived from *Aegilops tauschii*, integrating controlled greenhouse and field phenotyping. Finally, it evaluates multi-omics prediction across locations, years, and new genotypes by combining genomics with environmental and phenotyping features under breeding-relevant validation schemes, demonstrating that the value of added data layers and model complexity is scenario- and trait-dependent. Collectively, these results support an integrated, goal-driven approach to wheat improvement that couples robust multi-environment inference with hypothesis-guided discovery and deployment-oriented prediction, providing actionable guidance for selection, parent choice, and cultivar placement.

Table of Contents

List of Figures

List of Tables

Acknowledgements

It is with deep humility and gratitude that I acknowledge the privilege of pursuing my doctoral studies at Kansas State University, particularly within the Department of Agronomy and the College of Agriculture. This journey has been made possible through the support, mentorship, and encouragement of many individuals and institutions, and I am sincerely grateful to all who contributed along the way.

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Allan Fritz, for his unwavering guidance, mentorship, and support throughout my doctoral journey. The days spent walking breeding populations in the field, traveling across Kansas during harvest, and the conversations in his office were invaluable learning experiences. These moments shaped my understanding of wheat breeding and greatly influenced my development as a scientist.

I would also like to extend my sincere appreciation to my committee members, Dr. David Hessel, Dr. Alencar Xavier, and Dr. Eduard Akhunov, for their insightful discussions, thoughtful feedback, and expertise. Their rigorous scientific perspectives greatly strengthened this work and deepened my understanding of quantitative genetics and plant breeding.

Special thanks are extended to the Kansas State Wheat Breeding Team, including Shaun Winnie, Angela Matthews, Andrew Auld, Carter Westerhold, Addy Carroll, and Eduardo Melgar, as well as former graduate students Augusto Tessele and Enzo Unzain. Their dedication and efforts in supporting the breeding program were essential to the success of this research, and the many days working together in the field made this experience both productive and enjoyable.

I would also like to acknowledge Corteva Agriscience, especially the Wamego Station Team, including Miranda Smith, Mike Legleiter, James Martin, Jennifer Glaum, Erin Boot, and Sean Wallace, for their contributions to the generation of phenotypic and genotypic data through

the collaborative agreement with Corteva Agriscience (TRAC #2023-0635). Their support and collaboration were instrumental in advancing this work, and the many harvest days shared together were both productive and memorable.

To my friends, whose support, encouragement, and companionship made this journey more enjoyable and meaningful, I extend my sincere thanks.

Finally, and most importantly, I express my deepest gratitude to my family. I thank my parents, José Antonio and Marcelia, my brother Guilherme, my sister Isabela, and my grandparents, Ivanete, José Maria, Maria das Graças, and José Geraldo, for their unconditional love, encouragement, and unwavering belief in me. Their support has been the foundation of my achievements, and this thesis stands as a testament to their faith and sacrifices. I would also like to express my sincere gratitude to my girlfriend Alyse Laporte, whose support, encouragement, and companionship throughout this journey have meant a great deal and have been fundamental during this important stage of my life.

Dedication

I dedicate this work to all who devote their lives to advancing science and improving agriculture. To the breeders who envision better crops, the scientists who pursue discovery, the agronomists who translate knowledge into practice, and the farmers who cultivate the land.

Chapter 1 - Literature Review - Inference, Exploration, and Prediction in Crop Breeding

Abstract

This literature review develops a goal-oriented framework for modeling in crop breeding by organizing analytical approaches around inference, exploration, and prediction, following the principle that model usefulness depends on alignment between the scientific question and the available data. The chapter begins by establishing the global importance of wheat (*Triticum aestivum* L.) and the regional significance of Hard Red Winter wheat in Kansas, emphasizing the crop's central role in food security and agricultural economies. It then introduces George Box's "all models are wrong, but some are useful" perspective as a conceptual foundation for treating models as tools and datasets as fit-for-purpose parts within breeding pipelines. The review synthesizes literature on genotype-by-environment interaction (G×E) as an inferential problem, covering classical stability and adaptability methods for multi-environment trials. It next examines trait–marker association analysis as an exploratory framework, highlighting the progression from biparental QTL mapping to GWAS, the importance of controlling population structure and kinship, and the value of multi-trait, multi-environment, and multi-population designs for robust hypothesis generation. Finally, the chapter reviews prediction-focused breeding methods, including genomic selection, multi-environment genomic prediction, high-throughput phenotyping, and multi-omics integration, with attention to model validation, scalability, and deployment realism. Across these themes, the chapter argues that robust crop-breeding analytics require explicit goal definition, appropriate model selection, and continuous validation as germplasm, environments, and data modalities evolve. This framework provides the

conceptual foundation for the subsequent dissertation chapters on G×E inference, GWAS-based exploration in a D-genome–enriched population, and multi-omics prediction in breeding datasets.

Crop importance

Common wheat (*Triticum aestivum* L.) is one of the world's most important food-security crops, contributing roughly one-fifth of global food calories and protein and providing unique processing functionality for bread and other staple foods; consequently, continued genetic improvement remains essential for sustaining yield stability and grain quality under changing climates (Johansson et al., 2024; Shewry, 2024). In Kansas, this global importance is expressed most strongly through Hard Red Winter (HRW) wheat, which dominates production and underpins the regional grain economy: Kansas State University reports that Kansas is the leading U.S. producer of HRW wheat, with approximately 7–8 million acres planted annually (Kansas State University Department of Plant Pathology, 2025). Kansas Wheat similarly notes that about 95% of Kansas wheat is HRW, a class valued for high protein and strong gluten that supports bread flour and a broad range of wheat-based foods (Kansas Wheat, n.d.). At the national level, USDA-ERS indicates that HRW accounts for about 40% of total U.S. wheat production and is produced primarily in the Great Plains, reinforcing Kansas's strategic role in the U.S. wheat sector (U.S. Department of Agriculture, Economic Research Service [USDA-ERS], 2025). USDA-NASS further reports 7.3 million planted acres, 346.8 million bushels of wheat production, and approximately $1.595 billion in wheat production value for Kansas in 2025, emphasizing wheat's continued agricultural and economic significance in the state (U.S. Department of Agriculture, National Agricultural Statistics Service [USDA-NASS], 2025).

Models and their application in breeding

George Box's famous reminder that "all models are wrong, but some are useful" (Box, 1979) is a fitting epigraph for modern plant breeding because breeding is, at its core, a sequence of decisions made under uncertainty and complexity. The practical question is rarely whether a model is "true," but whether it is useful for a purpose: to learn (inference), to discover patterns and generate hypotheses (exploration), or to make accurate decisions about future performance (prediction). Tredennick et al., (2021) sharpen this idea by emphasizing that model choice should be governed by the scientific goal —exploration, inference, or prediction—because the same dataset can support very different questions, and the "best" model for one goal may be inappropriate for another. Although their paper is rooted in ecology, its framing maps directly onto crop improvement, where the same phenotypic, environmental, and genomic measurements can be used to estimate variance components and genetic gain (inference), to scan genomes for trait associations (exploration), or to predict untested genotypes in untested environments (prediction) (Box, 1976, 1979; Tredennick et al., 2021).

These goals overlap in practice, but in breeding they correspond to distinct operational decisions. Inference supports questions such as whether G×E is large enough to justify regional targeting, whether heritability is sufficient for early-generation selection, and which management factors interact with genotype (Cooper & DeLacy, 1994; van Eeuwijk et al., 2016). Exploration supports questions such as which genomic regions warrant fine mapping or candidate-gene follow-up and which correlated trait sets might share genetic drivers (Brachi et al., 2011; Zhu et al., 2008). Prediction supports decisions about which lines to advance, which parents to cross, and how to allocate testing resources under budget and cycle-time constraints (Bernardo, 2008; Crossa et al., 2017; Meuwissen et al., 2001). Modern breeding increasingly integrates all three

modes, but the logic and evaluation standards differ by goal, making explicit goal alignment central to robust modeling practice (Box, 1976; Tredennick et al., 2021).

G×E as an inferential problem

Genotype-by-environment interaction (G×E) is a defining feature of crop deployment because the same genotype can rank differently across sites and years as temperature, rainfall, soils, disease pressure, and management change. This is true in wheat and widely observed across maize, rice, soybean, and other crops, with the magnitude and structure of G×E depending on trait architecture, the target population of environments (TPE), and the breeding program's sampling strategy (Cooper & DeLacy, 1994; van Eeuwijk et al., 2016). From an inferential perspective, G×E modeling asks what portion of phenotypic variance is attributable to genotype, environment, and their interaction; whether the interaction is mainly "scale" (changing variance without reshuffling ranks) or "crossover" (rank changes); and which genotypes are broadly adapted versus specifically adapted (Eberhart & Russell, 1966; Finlay & Wilkinson, 1963; Shukla, 1972). In HRW wheat, G×E is often intensified by diverse production conditions and episodic stresses (heat events, drought timing, disease outbreaks), which can obscure genetic differences in any single environment and make multi-environment trials (MET) essential for stable selection decisions (Cooper & DeLacy, 1994; van Eeuwijk et al., 2016).

Some of the earliest tools for G×E inference treat genotype performance as a function of an environmental index, producing interpretable stability and adaptability parameters. Finlay and Wilkinson's regression approach estimates genotype-specific slopes to summarize responsiveness to environmental quality (Finlay & Wilkinson, 1963). Eberhart and Russell extend this approach by pairing slope (responsiveness) with deviation from regression (stability), distinguishing predictable from unpredictable responses (Eberhart & Russell, 1966). These

approaches remain widely used because of their interpretability and computational simplicity, and they can provide useful first-pass summaries before more complex modeling (Eberhart & Russell, 1966; Finlay & Wilkinson, 1963). However, they also reflect Box's warning that usefulness depends on fit-for-purpose assumptions: regression-on-index approaches compress multidimensional environmental variation into a single axis and can mask mechanistic drivers when stress patterns differ in timing or intensity (Box, 1976; Cooper & DeLacy, 1994). When environments are not representative samples of the TPE or management is confounded with location, the interpretability of slopes and stability deviations becomes less defensible, and uncertainty can be underappreciated (Cooper & DeLacy, 1994; van Eeuwijk et al., 2016).

A major shift came from viewing G×E as a structured matrix problem (genotypes × environments) with interaction patterns that can be decomposed. The additive main effects and multiplicative interaction (AMMI) model combines ANOVA for main effects with principal components for the interaction, often revealing low-rank patterns that correspond to latent environmental gradients or genotype groups (Gauch, 1988). GGE biplots similarly emphasize genotype main effects plus G×E and are often used to visualize "which-won-where" patterns and mega-environment structure (Yan et al., 2007). These tools are valuable for both exploration and inference when used appropriately, but their usefulness depends strongly on experimental design and data balance. Because breeding datasets are often unbalanced (not all genotypes tested in all environments, changing trial designs across years), naive AMMI/GGE on raw means can be misleading. Improved practice is to apply decomposition to adjusted estimates (e.g., BLUPs or design-corrected means) derived from mixed models (Piepho, 1998; Smith et al., 2005).

Linear mixed models provide the modern backbone for MET inference because they accommodate complex trial designs, unbalanced data, spatial field trends, heterogeneous error

variances, and flexible representations of G×E (Piepho et al., 2008; Smith et al., 2005). In this framework, genotypes can be treated as random (supporting BLUPs and variance component estimation) or fixed (supporting BLUEs and contrasts), while environments can be modeled as random or fixed depending on inferential aims and the intended generalization (Piepho et al., 2008). Mixed models help separate design effects (blocks, replicates, spatial heterogeneity) from genetic signal, improving inferential validity and producing cleaner inputs for downstream genomic prediction and association analyses (Piepho et al., 2008; Smith et al., 2005). Even when the primary goal is inference, modern MET networks and large genotype sets increasingly motivate scalable multivariate approaches that preserve key interaction information while remaining computationally feasible, including latent-representation methods designed to handle many environments (Xavier et al., 2025). The practical implication is that breeders often rely on hybrid workflows: mixed models for inference and data correction, coupled with cross-validation or forward checks to ensure that inferential summaries remain useful under deployment (Box, 1979; Tredennick et al., 2021).

Trait–marker association as exploration

Trait–marker association analysis is fundamentally exploratory: it aims to identify genomic regions associated with phenotypic variation, generate biological hypotheses, and provide markers for selection or deeper functional studies. Classical biparental QTL mapping offers high power for segregating loci within a cross but is constrained by limited allelic diversity and mapping resolution (Brachi et al., 2011). GWAS extends discovery to diverse panels, leveraging historical recombination and broader allelic diversity, often yielding higher resolution but increased vulnerability to confounding from population structure and relatedness (Myles et al., 2009; Yu et al., 2006). This issue is particularly acute in self-pollinated crops such

as wheat, where breeding history can induce long haplotypes and heterogeneous linkage disequilibrium patterns across subgenomes (Flint-Garcia et al., 2003; Zhu et al., 2008). Across crops, the central exploratory challenge is therefore to detect signals while controlling spurious associations arising from shared ancestry rather than causal loci (Price et al., 2006; Yu et al., 2006).

A major methodological advance was the unified mixed-model framework that incorporates population structure (often via principal components or Q matrices) and kinship (K matrices) as covariates or random effects, improving control of confounding effects and increasing credibility of association signals (Yu et al., 2006; Zhang et al., 2010). Subsequent methods improved computational scalability for large marker sets and sample sizes, including efficient variance-component models and widely adopted breeding-oriented software frameworks (Kang et al., 2010; Lipka et al., 2012). In parallel, multi-locus methods such as FarmCPU and BLINK were developed to increase power and manage confounding under complex, polygenic architectures (Huang et al., 2018; Liu et al., 2016). Recent benchmarking in wheat continues to compare GWAS models and threshold strategies, reinforcing that model choice is part of the discovery process and that robustness across modeling choices can matter more than any single p-value (Sandhu et al., 2024).

Exploratory power increases when GWAS leverages additional structure: multiple traits, multiple environments, or multiple related populations. Multi-trait mixed models can detect pleiotropy and increase power when traits share genetic basis (Korte et al., 2012). Multi-environment association analyses can identify loci with environment-dependent effects, linking marker discovery back to G×E and helping distinguish broadly stable loci from those that contribute to specific adaptation (Brachi et al., 2011). Multi-population GWAS can increase

power and robustness by borrowing information across populations while still allowing population-specific effects, a strategy that is increasingly relevant for breeding programs with limited depth in any single population (Skovbjerg et al., 2025). At the same time, dense marker datasets and many significant associations can create a false sense of causal certainty: even with mixed-model correction, GWAS results are statistical associations that may reflect linkage, residual confounding, or environment dependence and typically require follow-up validation (Visscher et al., 2017; Zhu et al., 2008). In practical breeding, GWAS outputs are therefore most defensibly treated as decision-support signals and hypothesis generators, with continuous validation needed as LD patterns, genotyping platforms, and breeding germplasm evolve over time (Box, 1979; Desta & Ortiz, 2014).

Beyond frequentist mixed-model GWAS, Bayesian GWAS reframes association mapping as a probabilistic inference problem in which SNP effects are modeled jointly with explicit prior distributions (often mixture or "spike-and-slab" priors that allow many effects to be near zero while a minority have nonzero effects), yielding posterior summaries that are often more directly interpretable than single-marker p-values for polygenic traits. In practice, Bayesian models can simultaneously account for a polygenic background while allowing a sparse subset of larger effects, and they commonly report evidence as posterior inclusion probabilities (PIPs) or region-level credible sets, which is useful when linkage disequilibrium spreads signal across correlated markers (Clauw et al., 2025; Guan & Stephens, 2011; Samaddar et al., 2024; Zhou et al., 2013). Bayesian approaches can also replace (or complement) classical significance thresholds by ranking associations with Bayes factors, making the role of effect-size priors and sample size more explicit and improving comparability of evidence across studies (Wakefield, 2009). In plant GWAS, where complex trait architectures, population structure, and heterogeneous LD are

common, Bayesian and other advanced GWAS frameworks are increasingly emphasized as valuable complements to standard mixed-model GWAS, especially when the goal is to quantify uncertainty in multi-locus signals rather than to declare isolated “significant SNPs” (Clauw et al., 2025). At the same time, Bayesian GWAS requires careful practice: results can be sensitive to prior specification, computationally demanding (often relying on MCMC or scalable approximations), and in high-LD regions single-SNP PIPs may be diluted—making robust reporting (e.g., credible sets, sensitivity analyses, and replication) central to defensible biological interpretation (Guan & Stephens, 2011; Samaddar et al., 2024).

Prediction as decision-making

Genomic selection (GS) reframed breeding as a prediction problem: rather than focusing on a small number of significant markers, GS uses genome-wide marker information to predict breeding values for complex traits controlled by many small-effect loci (Meuwissen et al., 2001). This predictive framing aligns closely with the prediction goal emphasized by Tredennick et al. (2021): model utility is judged by out-of-sample accuracy for selection candidates, not by interpretability of individual parameters (Tredennick et al., 2021). In practice, GS has been adopted across major crops because it can accelerate genetic gain by shortening selection cycles and enabling earlier advancement decisions (Crossa et al., 2017; Desta & Ortiz, 2014).

Recent syntheses emphasize that GS performance depends on training population design, relatedness between training and target sets, trait heritability, and the degree to which models account for G×E and other complexities (Crossa et al., 2017; Crossa et al., 2025). The most operationally valuable prediction problems in breeding are often the hardest: predicting untested genotypes in untested environments, predicting across years with different stress patterns, and predicting performance in target environments that have not yet been observed. Models that

ignore G×E can perform well within narrow environmental domains, but their accuracy often degrades when genotype performance is environment-sensitive (Jarquín et al., 2014; van Eeuwijk et al., 2016). This has motivated extensive work in multi-environment genomic prediction, including reaction norm models that integrate high-dimensional genomic and environmental data (Jarquín et al., 2014) and deep learning approaches designed for multi-environment settings (Montesinos-López et al., 2018). More recently, scalable multivariate approaches have been proposed to better capture complex G×E patterns under sparse testing, which is increasingly relevant as breeding networks expand in size and heterogeneity (Xavier et al., 2025). Explainable ensemble frameworks have also been developed to improve prediction while providing feature attribution, supporting breeder trust and interpretation when models incorporate many environment variables (Yu et al., 2025). Because predictive performance can be inflated by data leakage or unrealistic validation, robust evaluation requires validation schemes that mimic real breeding decisions, such as leave-one-environment-out and forward-in-time validation (Crossa et al., 2017; Tredennick et al., 2021).

High-throughput phenotyping and phenomics

Prediction accuracy often improves not only through better algorithms, but through better phenotyping. High-throughput phenotyping (HTP) platforms (including UAV multispectral imaging and ground-based sensors) provide time-series traits that capture plant status and stress responses more directly than end-of-season yield alone. These phenomic traits can be used as covariates, auxiliary traits in multi-trait models, or inputs to machine learning architectures (Araus & Cairns, 2014; Rutkoski et al., 2016). In wheat, incorporating HTP-derived traits into genomic prediction has repeatedly improved predictive ability for yield and related traits, particularly when phenomic features provide complementary information about stress responses

and developmental dynamics (Kaushal et al., 2024; Rutkoski et al., 2016). In this sense, phenomics expands the “parts” available for prediction, potentially absorbing some G×E complexity into measurable physiological trajectories (Araus & Cairns, 2014; Kaushal et al., 2024).

Multi-omics prediction

Multi-omics prediction extends the same logic further: although genomics captures inherited variation, phenotypes emerge through regulatory, transcriptional, epigenetic, proteomic, and metabolic processes. For complex traits—especially those shaped by stress responses and developmental timing—integrating additional omics layers can provide complementary signals and improve prediction. However, multi-omics integration also introduces major challenges, including extreme dimensionality, differences in measurement scales, noise, missingness, and context dependence across tissues, developmental stages, and environments, making integration strategy central to usefulness (Montesinos-López et al., 2025; Wang et al., 2024; Zhang et al., 2025). Recent work highlights both promise and caution: integrating genomic, transcriptomic, and methylomic information can improve predictive performance and reveal biologically meaningful signals when carefully validated (Wang et al., 2024), while broader evaluations emphasize that naive concatenation can underperform and that more principled fusion strategies are often required to realize consistent gains (Montesinos-López et al., 2025). Multi-omics atlases in wheat further illustrate the biological richness of integrated data layers and their relevance for breeding-relevant processes, but also underscore tissue and stage dependence that must be respected in modeling and sampling decisions (Zhang et al., 2025). In practice, because many breeding programs cannot collect multi-omics data on all candidates at scale, the most useful strategies often leverage omics on subsets to improve

training, to identify robust intermediate phenotypes, or to support targeted biological interpretation rather than assuming full omics coverage every cycle (Desta & Ortiz, 2014; Montesinos-López et al., 2025).

Connecting the modeling frame to the dissertation chapters

Together, these themes support a coherent goal-driven narrative across the dissertation. Chapter 2 emphasizes inference by characterizing G×E patterns across older and newer wheat varieties and estimating genetic gain through time, using MET-based modeling to separate genetic signal from environmental and design noise. Chapter 3 emphasizes exploration by using GWAS in a D-genome–enriched introgression population to identify genomic regions associated with modern breeding traits, treating association results as robust hypotheses that motivate biological interpretation and follow-up. Chapter 4 emphasizes prediction by modeling diverse breeding "omics" data types with multiple analytical frameworks to evaluate their value for inference, exploration, and especially out-of-sample prediction under realistic deployment assumptions. Across chapters, the unifying principle is consistent: models are evaluated by their usefulness for the stated goal, and data are collected and processed to serve as fit-for-purpose parts for the chosen analytical tool.

References

Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. Trends in Plant Science, 19(1), 52–61. https://doi.org/10.1016/j.tplants.2013.09.008

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Science, 48(5), 1649–1664.

Box, G. E. P. (1976). Science and statistics. Journal of the American Statistical Association, 71(356), 791–799.

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), Robustness in statistics (pp. 201–236). Academic Press.

Brachi, B., Morris, G. P., & Borevitz, J. O. (2011). Genome-wide association studies in plants: The missing heritability is in the field. Genome Biology, 12, 232.

Clauw, P., Ellis, T. J., Liu, H.-J., & Sasaki, E. (2025). Beyond the standard GWAS—A guide for plant biologists. Plant and Cell Physiology, 66(4), 431–443.

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(5), 561–572.

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.

Crossa, J., Martini, J. W. R., Vitale, P., Pérez-Rodríguez, P., Costa-Neto, G., Fritsche-Neto, R., Runcie, D., Cuevas, J., Toledo, F., Li, H., De Vita, P., Gerard, G., Dreisigacker, S., Crespo-Herrera, L., Saint Pierre, C., Bentley, A., Lillemo, M., Ortiz, R., Montesinos-López, O. A., & Montesinos-López, A. (2025). Expanding genomic prediction in plant breeding: Harnessing big data, machine learning, and advanced software. Trends in Plant Science, 30(7), 756–774.

Desta, Z. A., & Ortiz, R. (2014). Genomic selection: Genome-wide prediction in plant improvement. Trends in Plant Science, 19(9), 592–601.

Eberhart, S. A., & Russell, W. A. (1966). Stability parameters for comparing varieties. Crop Science, 6(1), 36–40.

Finlay, K. W., & Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. Australian Journal of Agricultural Research, 14(6), 742–754.

Flint-Garcia, S. A., Thornsberry, J. M., & Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. Annual Review of Plant Biology, 54, 357–374.

Gauch, H. G. (1988). Model selection and validation for yield trials with interaction. Biometrics, 44(3), 705–715.

Guan, Y., & Stephens, M. (2011). Bayesian variable selection regression for genome-wide association studies and other large-scale problems. The Annals of Applied Statistics, 5(3), 1780–1815.

Huang, M., Liu, X., Zhou, Y., Summers, R. M., & Zhang, Z. (2018). BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience, 8(2), giy154.

Jarquín, D., Crossa, J., Lacaze, X., du Cheyron, P., Daucourt, J., Lorgeou, J., Piraux, F., Guerreiro, L., Pérez, P., Calus, M. P. L., Burgueño, J., & de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theoretical and Applied Genetics, 127(3), 595–607.

Johansson, E., Lan, Y., Olalekan, O., Kuktaite, R., Chawade, A., & Rahmatov, M. (2024). Alien introgression to wheat for food security: Functional and nutritional quality for novel products under climate change. Frontiers in Nutrition, 11, 1393357.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-Y., Freimer, N. B., Sabatti, C., & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nature Genetics, 42(4), 348–354.

Kansas State University Department of Plant Pathology. (2025, September 23). Wheat. https://www.plantpath.k-state.edu/extension/wheat-and-forages/wheat.html

Kansas Wheat. (n.d.). Which wheat for what? https://kswheat.com/news/which-wheat-for-what

Kaushal, S., Gill, H. S., Billah, M. A., Khan, N., Halder, J., Bernardo, A. E., St Amand, P. C., Bai, G., Glover, K., Maimaitijiang, M., & Sehgal, S. K. (2024). Enhancing the potential of phenomic and genomic prediction in winter wheat breeding using high-throughput phenotyping and deep learning. Frontiers in Plant Science, 15, 1410249.

Korte, A., Vilhjálmsson, B. J., Segura, V., Platt, A., Long, Q., & Nordborg, M. (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nature Genetics, 44(9), 1066–1071.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. Bioinformatics, 28(18), 2397–2399.

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genetics, 12(2), e1005767.

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.

Montesinos-López, A., Montesinos-López, O. A., Gianola, D., Crossa, J., & Hernández-Suárez, C. M. (2018). Multi-environment genomic prediction of plant traits using deep learners with dense architecture. G3: Genes, Genomes, Genetics, 8(12), 3813–3828.

Montesinos-López, O. A., Montesinos-López, A., Mosqueda-González, B. A., Delgado-Enciso, I., Chavira-Flores, M., Crossa, J., Dreisigacker, S., Sun, J., & Ortiz, R. (2025). Genomic prediction powered by multi-omics data. Frontiers in Genetics, 16, 1636438.

Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., & Buckler, E. S. (2009). Association mapping: Critical considerations shift from genotyping to experimental design. The Plant Cell, 21(8), 2194–2202.

Piepho, H.-P. (1998). Methods for comparing the yield stability of cropping systems—A review. Journal of Agronomy and Crop Science, 180(4), 193–213.

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

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8), 904–909.

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., González Pérez, L., Crossa, J., Reynolds, M., & Singh, R. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. G3: Genes, Genomes, Genetics, 6(9), 2799–2808.

Samaddar, A., Maiti, T., & de los Campos, G. (2024). Bayesian hierarchical hypothesis testing in large-scale genome-wide association analysis. Genetics, 228(4), iyae164.

Sandhu, K. S., Burke, A. B., Merrick, L. F., Pumphrey, M. O., & Carter, A. H. (2024). Comparing performances of different statistical models and multiple threshold methods in a nested association mapping population of wheat. Frontiers in Plant Science, 15, 1460353.

Shewry, P. R. (2024). Can we increase the use of wheat and other cereals as sources of protein? Journal of Cereal Science, 117, 103899.

Shukla, G. K. (1972). Some statistical aspects of partitioning genotype–environmental components of variability. Heredity, 29, 237–245.

Skovbjerg, C. K., Sarup, P., Wahlström, E., Jensen, J. D., Orabi, J., Olesen, L., Jensen, J., Jahoor, A., & Ramstein, G. (2025). Multi-population GWAS detects robust marker associations in a newly established six-rowed winter barley breeding program. Heredity, 134, 33–48.

Smith, A. B., Cullis, B. R., & Thompson, R. (2005). The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. Journal of Agricultural Science, 143(6), 449–462.

Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. Ecology, 102(6), e03336.

U.S. Department of Agriculture, Economic Research Service. (2025, March 6). Wheat sector at a glance. https://www.ers.usda.gov/topics/crops/wheat/wheat-sector-at-a-glance

U.S. Department of Agriculture, National Agricultural Statistics Service. (2025). 2025 state agriculture overview: Kansas. https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=KANSAS

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.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. The American Journal of Human Genetics, 101(1), 5–22.

Wakefield, J. (2009). Bayes factors for genome-wide association studies: Comparison with P-values. Genetic Epidemiology, 33(1), 79–86.

Wang, P., Lehti-Shiu, M. D., & Shiu, S.-H. (2024). Prediction of plant complex traits via integration of multi-omics data. Nature Communications, 15, 6856.

Xavier, A., Runcie, D. E., & Habier, D. (2025). Megavariate methods capture complex genotype-by-environment interactions. Genetics, 229(4), iyae179.

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), 643–655.

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, 38(2), 203–208.

Yu, T., Zhang, H., Chen, S., Gao, S., Liu, Z., Wang, J., Crossa, J., Montesinos-López, O. A., Hearne, S., & Li, H. (2025). EXGEP: A framework for predicting genotype-by-environment interactions using ensembles of explainable machine-learning models. Briefings in Bioinformatics, 26(4), bbaf414.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., & Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. Nature Genetics, 42(4), 355–360.

Zhang, N., Tang, L., Li, S., Liu, L., Gao, M., Wang, S., Chen, D., Zhao, Y., Zheng, R., Soleymaniniya, A., Zhang, L., Wang, W., Yang, X., Ren, Y., Sun, C., Wilhelm, M., Wang, D., Li, M., & Chen, F. (2025). Integration of multi-omics data accelerates molecular analysis of common wheat traits. Nature Communications, 16(1), 2200.

Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. PLoS Genetics, 9(2), e1003264.

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. The Plant Genome, 1(1), 5–20.

Chapter 2 - Yield Stability, Genotype x Environment Interaction, and Genetic Gain in Kansas Hard Red Winter Wheat: Investigation from Multi-Environment Trials

Abstract

Kansas hard red winter wheat (*Triticum aestivum* L.) is grown under strong spatial and temporal variability, making genotype x environment interaction (G×E) a recurring challenge for cultivar development and a key driver of selection risk. We analyzed the Kansas State University long term Environment Evaluation (ENEV) trials from 2017-2025 across 41 environments (10 counties), evaluating 61 entries (54 cultivars and seven elite lines; release years 1992-2025). Trials followed an α-lattice design with two replications; yield data were spatially adjusted and summarized as genotype BLUEs/BLUPs using mixed models. Dynamic stability was assessed with Eberhart-Russell mixed-model reaction norms and parameters (slope b_i and deviation variance S^2d_i), while rank-based G×E was summarized with a ranking-genotypes GGE biplot and mean rank ± SD. Relative genetic gain was estimated by regressing yield BLUPs on year of release. Across environments, the environment main effect dominated yield variance (78.73% of total; $P < 0.001$), while genotype (7.39%; $P < 0.001$) and G×E (6.04%; $P < 0.001$) were also significant; heritability across environments was 0.35 with strong site-year heterogeneity (H^2 = 0.11-0.94). The environmental index covered ≈ −2500 to +2000 kg ha^{-1}, and reaction norms were largely parallel with fan-shaped divergence in favorable environments, indicating predominantly scale-type G×E and increased discrimination among cultivars at high productivity. Most genotypes clustered near $b_i \approx 1$ with low S^2d_i (broad adaptation), whereas

smaller subsets showed either enhanced responsiveness ($b_i > 1$) or better relative performance in poor environments ($b_i < 1$); a few entries showed high S^2d_i, indicating yield plasticity. Relative genetic trend analysis showed continued yield improvement of 24.67 kg ha^{-1} yr^{-1} across all entries (24.85% total; 0.64% yr^{-1}). The ENEV trial provides an evidence-based "thermometer" for monitoring adaptation and documenting measurable yield gains under Kansas production conditions, guiding cultivar deployment and breeding decisions statewide.

Introduction

Kansas, often recognized as "The Wheat State", plays a fundamental role in winter wheat production in the United States, contributing approximately 26% of the nation's total wheat (Holman, J. D., et al., 2024; NASS, 2022; Obembe, O. S., et al., 2021). Meeting the increasing demand for food crops such as wheat, presents a significant challenge for breeding programs, driven by the dual pressures of a rapidly growing global population and the impacts of annual environmental differences, both of which are projected to intensify in the coming decades (Zhang-Biehn et al., 2021; Godfray et al., 2010). Developing wheat varieties that demonstrate both high and stable yields across diverse climates is crucial for meeting humanity's growing needs and maintaining economic viability at the producer level.

Stability, as a concept rooted in physics and mathematics, describes a system's tendency to maintain or return to equilibrium when perturbed, indicating resistance to disturbance and predictability in behavior (Holling, C. S., 1973; Mayar, K., Carmichael, D. G., & Shen, X., 2022). In statistics, stability often denotes robustness and consistency of outcomes, meaning results or models remain largely unchanged under small variations in data or assumptions (Breiman, 1996; Huber, P. J., & Ronchetti, E. M., 2011; Maronna, R. A., Martin, R. D., Yohai,

V. J., & Salibián-Barrera, M., 2019). In biological and agricultural sciences - particularly in breeding - *yield stability* refers to the consistency of a genotype's performance across diverse environments, defined either in a static sense (yield remains nearly constant across environments) or a dynamic sense (yield responds to environments but with a predictable relative performance advantage) (Becker & Léon, 1988; Annicchiarico, 2002, Pour-Aboughadareh, Alireza, et al., 2022), yet due to the differences in the environment types, we should expect a predominance of dynamic stability for agricultural science, especially for developed countries. Breeders quantify yield stability using methods such as Finlay-Wilkinson regression (assessing adaptation via the regression slope of yield on environmental index) and the Eberhart-Russell model (a regression approach incorporating a stability parameter for deviation from prediction) (Finlay & Wilkinson, 1963; Eberhart & Russell, 1966). They also employ multivariate techniques like the AMMI (additive main effects and multiplicative interaction) model and GGE (genotype + genotype-by-environment) biplots to analyze and visualize genotype-environment interactions, facilitating identification of high-yielding and stable genotypes (Zobel et al., 1988; Yan & Kang, 2002, Olivoto, T., & Lúcio, A. D. C., 2020).

For any breeding program, G×E is a challenge for selecting genotypes that perform well across the target environments. G×E interaction is an essential concept in plant breeding because it explains how genetic backgrounds can perform differently across a range of environmental conditions. In wheat, G×E interactions significantly influence yield rank performance, disease response, and quality traits (Crespo-Herrera et al., 2021). Understanding these interactions helps breeders identify stable and high-performing wheat varieties that can adapt to diverse climates and management practices. Managing G×E is key to enhancing global food security, especially

under increasingly unpredictable climatic conditions, as it influences the ability to maintain relatively consistent production.

Over the past two centuries, US wheat breeding programs have continually developed varieties that help farmers stay ahead of new and emerging pests and pathogens. Early on, during the late 1800s and early 1900s, efforts focused on the selection of landraces and their variants to identify cultivars adapted to previously unfamiliar growing conditions. The release of semi-dwarf, higher-yielding lines in the 1970s further propelled nationwide wheat yields upward, due to what is known as the “Green Revolution”. These accomplishments have profoundly shaped the trajectory of US wheat production, fortifying its resilience and overall productivity (Kusunose, Y., 2023; Graybosch & Peterson, 2012).

The positive impacts of breeding efforts continue to manifest in measurable gains. For instance, genetic gain for cultivars tested in Kansas from different breeding programs has ranged from 0.37% to 1.92% per year, reflecting the ongoing effectiveness of targeted breeding strategies (Rife, T. W. et al., 2019). Through sustained dedication to both traditional methods and emerging technologies, US wheat breeding programs remain poised to meet the ever-evolving demands of agriculture.

The Kansas State University-Manhattan program, which targets development of hard winter wheat varieties adapted to central Kansas, established the ENEV experiment as a long-term research initiative to evaluate wheat cultivar performance across diverse environmental conditions. Given the increasing challenges posed by weather variation, this initiative serves as a “thermometer” for newly released varieties in Kansas, offering annual data on how they adapt and respond to shifting environment types. The purpose is to generate robust, evidence-based information that benefits the Kansas wheat breeding program, the broader scientific community,

and local producers. In this manuscript, we present data from 2017 to 2025 regarding G×E and relative genetic gain. We also take the opportunity to describe and discuss the stability terminology for plant breeding.

Materials and methods

Plant Materials

A total of 54 hard winter wheat cultivars and 7 elite breeding lines are presented in the ENEV experiment. The foundation of this panel is varieties released by the KSU-Manhattan breeding program since the release of Karl 92 in 1992, along with commercially relevant checks that changed over time. The cultivars selected represent a range of genetic backgrounds and were released between 1992 and 2025 (Table 1). These cultivars are/were widely grown in Kansas and include both public and private materials, as well as experimental lines that were, or are being, considered for release. The complete list, along with the year each cultivar was first made commercially available, and the commercializing institution, is summarized in Table 1.

Table 2.1. Hard red winter wheat genotypes included in the ENEV network, with year of release and breeding institution.

Genotype, Year of Release, and Institution					
Karl 92	1992	KSU-M	KS Silverado	2019	KSU-H
Jagger	1994	KSU-M	KS090387K-20	2019	KSU-M
2137	1995	KSU-M	LCS Valiant	2019	LCS
2145	2001	KSU-M	SY Wolverine	2019	SY
Overley	2003	KSU-M	WB 4595	2019	WB
Fuller	2006	KSU-M	WB 4699	2019	WB
Billings	2009	OSU	WB 4792	2019	WB
Everest	2009	KSU-M	AM Cartwright	2020	AM
WB Cedar	2011	WB	AP EverRock	2020	SY
Gallagher	2012	OSU	KS Hatchett	2020	KSU-M
LCS Mint	2012	LCS	Kivari AX	2020	CSU
WB 4458	2012	WB	WB 4401	2020	WB
WB Grainfield	2012	WB	KS Ahearn	2021	KSU-M

Doublestop CL Plus	2013	OSU	AP Prolific	2022	SY
LCS Wizard	2013	LCS	KS Providence	2022	KSU-M
KanMark	2014	KSU-M	LCS Steel AX	2022	LCS
SY Monument	2014	SY	WB 4422	2022	WB
Hot Rod	2015	AGSECO	High Cotton	2023	OSU
SY Flint	2015	SY	KS Mako	2023	KSU-M
LCS Chrome	2016	LCS	KS160786S-6	2023	KSU-M
Langin	2016	CSU	KS170013D-19	2023	KSU-M
Larry	2016	KSU-M	KS170025D-11	2023	KSU-M
Zenda	2016	KSU-M	Paradox	2023	OSU
Bob Dole	2017	KSU-M*	Golden Hawk	2024	PS
Rock Star	2017	PS	KS Bill Snyder	2024	KSU-H
Smith's Gold	2017	OSU	AP24AX	2025	SY
WB 4269	2017	WB	KS150167-17	2025	KSU-M
WB 4515	2017	WB	KS150167-18	2025	KSU-M
Green Hammer	2018	OSU	KS150167-19	2025	KSU-M
Showdown	2018	OSU	KS Flintlock	2025	KSU-H
AG Icon	2019	AGSECO			

Year of release corresponds to the first year the cultivar/line was made available (or first entered as an advanced, when applicable). Institution codes denote the originating breeding program or brand: KSU-M = Kansas State University, Manhattan Program; KSU-H = Kansas State University, Hays Program, OSU = Oklahoma State University; WB = WestBred (Bayer); LCS = Limagrain Cereal Seeds; SY/AP = AgriPro/Syngenta; CSU = Colorado State University; AM = AgriMaxx; PS = Polansky Seed; AG = AGSECO, Inc.; IG = InterGrain. *Bob Dole was developed by Kansas State University - Manhattan program and commercialized by Syngenta.

Experimental Design and Region Characterization

This is a recurrent experiment initiated in the 2016/2017 growing season. The experiment follows an α-lattice design with two replications for each genotype. Not all entries were planted every year, as some cultivars were introduced in subsequent years after the experiment began (annexed data). In total, the study encompasses 41 environments, defined as combinations of locations and years. The entries were evaluated in six-row plots measuring 1.5 meters × 2.44 meters. The genotypes were tested in 10 counties of the state of Kansas that represent the target population of environments for the K-State Manhattan breeding program (Figure 1) and the

experiment site-years varied (annexed data). The soil properties across the experimental locations were extracted from ‘Xpolaris’ R package (Moro Rosso et al., 2021); it shows variability, reflecting diverse soil compositions arrays. Clay content varies between 12% and 30%, with higher concentrations observed in the northeastern sites, whereas sand content follows an inverse pattern, with higher sand percentages found in the central and southwestern areas (Figure 1). Silt content is distributed evenly, with some regions exhibiting values up to 60% (Figure 1). Organic matter (OM) content ranges from 1.25% to 2.0%, with higher concentrations observed in central northeastern sites (Figure 1).

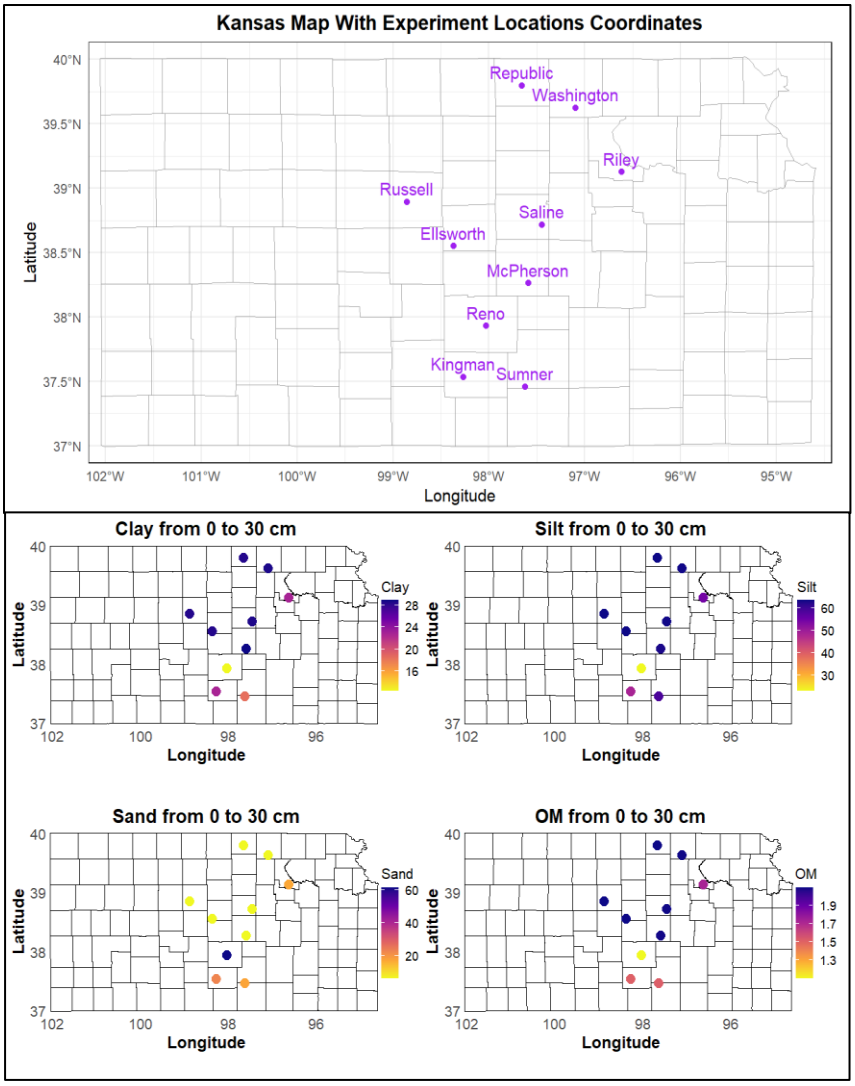

Figure 2.1. Kansas map with experiment locations coordinates and Soil characterization of the experimental environments plotted in Kansas map. The maps depict key soil attributes, including clay, sand, silt content, and organic matter (OM), for average 0–30 cm soil layer. These variations illustrate the diverse soil.

Statistical Analysis

To quantify dynamic yield stability and adaptability across the experiments, we modeled each genotype's response to environment with the regression framework of Eberhart and Russell (1966), an extension of the adaptation analysis originally proposed by Finlay and Wilkinson (1963). For each genotype i grown in environment j, the observed entry-mean yield y_{ij} was expressed as

$$\mathbf{y_{ij} = \mu + g_i + \beta_i I_j + \delta_{ij}},$$

where μ is the overall mean, g_i is the average deviation of genotype i from μ, I_j is the environmental index (the deviation of environment j from μ), β_i is the regression (slope) parameter that measures the environmental sensitivity of genotype i, and δ_{ij} is a genotype-specific deviation term that captures lack of fit to the linear response. After fitting the model by weighted least squares, two complementary stability statistics were extracted: (i) the slope $\hat{\beta}_i$ (with $\hat{\beta}_i = 1$ indicating average responsiveness, $\hat{\beta}_i < 1$ broad adaptation to poor environments, and $\hat{\beta}_i > 1$ specific adaptation to favorable environments), and (ii) the deviation variance $s_{di}^2 = \text{Var}\,(\hat{\delta}_{ij})$, where smaller s_{di}^2 values denote greater predictability (stability) across environments.

To quantify dynamic yield stability within the experiments, statistical analysis was conducted using the data collected from the 41 environments and the 'R software' (R Core Team., 2021). Package 'MrBean' (Aparicio J., et al., 2022) was used to calculate best linear unbiased estimates (BLUEs) and best linear unbiased predictions BLUPs (Piepho, H. P., et al., 2008). Potential outliers were identified using studentized residuals with threshold of 3 standard deviation (Aparicio J., et al., 2022). An observation with studentized residual greater than 3 (in absolute value) was considered to be an outlier. Spatial analysis was then performed to compute BLUEs of genotypes (Batista, L. A., et al., 2024; Lell, M., et al., 2021) following the model:

$$\mathbf{Y_{ij} = \mu + g_i + e_j + \varepsilon}$$

where Y_{ij} is the observed phenotype, μ is the overall mean, g_i represents the random effect of the i^{th} genotype (i.e., wheat line), e_j represents the random effect of the j^{th} environment (location-year) and ε_{ij} represents the residual term capturing the non-explained variability and it is

assumed to be an independent and identically distributed (IID) random outcome following a normal density.

For the purpose of estimating the broad-sense heritability (H^2) of each phenotype, variance components were estimated using the restricted maximum likelihood. All effects were assumed to be random. Broad-sense heritability on an entry-mean basis was calculated across and within experiments as

$$\mathbf{H^2} = \frac{\boldsymbol{\sigma^2_G}}{(\boldsymbol{\sigma^2_G} + \boldsymbol{\sigma^2_{GxE}} + \boldsymbol{\sigma^2_e})}$$

Where σ_G^2 is the genotypic variance, σ_{GxE}^2 is the genotype-by-environment interaction, $\sigma_e^2$is the residual variance. Heritability analysis was implemented with the R package 'MrBean' (Aparicio J., et al., 2022).

To look into genotype by environment interactions at the yield and rank levels, we used the R package 'Metan' (Olivoto, T., & Lúcio, A. D. C., 2020) to model ranking genotype by environment interactions. In the analysis, the mean yield for genotype i in environment j is modeled using a general linear framework. Specifically, the observed mean yield, $\hat{y}_{ij}$, is expressed as:

$$\mathbf{\hat{y}_{ij}} = \boldsymbol{\mu} + \boldsymbol{\alpha_i} + \boldsymbol{\beta_j} + \boldsymbol{\varphi_{ij}}$$

Here, $\hat{y}_{ij}$ is the yield for genotype i in environment j, μ is the overall grand mean, α_i captures the main effect of genotype i, β_j reflects the effect of environment j, and φ_{ij} represents the genotype-by-environment interaction. Decomposing the interaction term φ_{ij}, using Singular Value Decomposition (SVD), yields the additive main effects and multiplicative interaction (AMMI) model (Olivoto, T., & Lúcio, A. D. C., 2020; W. Yan et al., 2007; Weikai Yan and

Kang., 2002). In this formulation, the genotype main effect α_i is removed so that its variation is absorbed into φ_{ij}. The data matrix is centered by environment, meaning that the adjusted interaction is computed as:

$$\boldsymbol{\varphi_{ij} = \hat{y}_{ij} - \mu - \beta_j}$$

The centered data is then subjected to SVD, the partition factor for the k^{th} principal component, i_k along with j_k are the principal component scores for genotype i and environment j, respectively.

We also calculated the non-parametric genotype rank stability based on Nassar, R., & Huehn, M. (1987) by averaging the number of times the genotypes placed a ranking ranging from $r_1, r_2 \ldots r_n$ for the j environments as:

$$\boldsymbol{\hat{r} = \frac{1}{n} \sum_{i=1}^{n} r_j}$$

And calculated the standard deviation of the average rank for each genotype $r_1, r_2, \ldots, r_N$ with mean μ, as:

$$\boldsymbol{\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (r\_i - \mu)^2}}$$

Genetic Gain Calculations

We calculated a relative genetic gain for the genotypes used in the experiment. Some genotypes were released during the course of the experiment and we used best linear unbiased prediction (BLUP) (Piepho, H. P., et al., 2008). To obtain the yield BLUP values, we used 'lme4' R package (Bates, D., et. all., 2015) in a second step analysis, with genotypes as random effects and environments as fixed effects. The yield BLUPs obtained and the release year of each

genotype to be used in a linear regression model. The linear regression model (Montgomery, D. C., et al., 2021; Weisberg, S., 2005) assumed relationship between the release year and yield BLUP and the equation can be expressed as:

$$\textbf{Yield BLUP} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} \times \textbf{Release Year}$$

Where β_0 is the intercept (expected yield when Release Year is zero). β_1 is the slope (the change in yield per unit change in release year). The slope (β_1) indicates the average increase in yield for each additional year. If β_1 is positive, it suggests that yield has been improving over the years. Then, total genetic gain over a specified number of years was calculated as:

$$\textbf{Total Gain} = \boldsymbol{\beta_1} \times \textbf{Number of Years}$$

To express this gain as a percentage and compare the total gain to the average yield at the start of the observation period:

$$\%\,\textbf{Gain} = \left(\frac{\textbf{Total Gain}}{\textbf{Average Yield at Start}}\right) \times \textbf{100}$$

And to express in Percentage Gain per Year we divided the percentage gain for the total year in comparison as:

$$\%\,\textbf{Gain/Year} = \frac{\left(\frac{\textbf{Total Gain}}{\textbf{Average Yield at Start}}\right) \times \textbf{100}}{\textbf{Total Years of Improviment}}$$

Note that the percentage of genetic gain is relative to the set of cultivars here represented.

Results

As part of this long-term, ongoing multi-environment trial ENEV network, we present initial results and insights into G×E interactions for hard red winter wheat across Kansas.

Because cultivar performance reflects both genetic potential and environmental opportunity, this dataset provides a practical framework to define and interpret G×E as the extent to which genotypes respond differently to changes in location-year conditions- capturing both large differences in overall productivity among environments and differences in cultivar ranking and sensitivity across those environments. In the sections below, we first quantify how phenotypic variation is partitioned among genotype, environment, and G×E effects, and then connect these components to broader patterns of adaptability and stability across the testing network. Finally, because the panel includes widely adopted cultivars spanning multiple decades of release, we evaluate evidence of genetic gain over time and discuss how these historical trends relate to on-farm progress in the region.

Table 2 summarizes the variance partitioning from the across-location model. The model explained a large proportion of the phenotypic variation, with a conditional r^2 of 0.92, indicating that the full model captured most of the observed variability. The environment effect dominated the variance structure, accounting for 73.56% of the total variance and was highly significant ($P < 0.001$), highlighting strong differences among location-year environments. The genotypic variance was smaller but still significant (6.94%; $P < 0.001$), indicating detectable genetic differences across the full testing network. The genotype x environment interaction was also significant (5.70%; $P < 0.001$), confirming that cultivar performance was not perfectly consistent across environments, although G×E contributed substantially less than the environment main effect. Spatial variation within trials was captured by Row within environment (3.59%) and Range within environment (3.25%)**,** both significant ($P < 0.001$), supporting the inclusion of spatial terms to account for within-field heterogeneity. Despite the strong environmental

component, the entry-mean broad-sense heritability was high ($H^2 = 0.88$)**,** suggesting that genotype means were estimated with good precision across the multi-environment network.

Table 2.2. Variance Components Results Across Locations.

Variance Components (Across Locations)				
Model Conditional $R^2 = 0.93$ (92.7% of variance explained)				
Component	**Variance σ^2**	**SD**	**% Total**	**P-value**
Genotype (σ^2)	**101,043.25**	**317.87**	**7.39**	**<0.001**
Environment (σ^2)	**1,076,190.13**	**1,037.40**	**78.73**	**<0.001**
G x E (σ^2)	**82,523.05**	**287.27**	**6.04**	**<0.001**
Residual (Error)	**107,197.02**	**327.41**	**7.84**	**-**
Broad-sense Heritability (H^2): 0.35				
Significance based on Likelihood Ratio Test: * P < 0.001, ** P < 0.01, * P < 0.05				

Variance components (σ^2) were estimated using a linear mixed model with Genotype, Environment, and Genotype × Environment effects, with Row and Range modeled as spatial terms nested within Environment. Conditional R^2 reflects the variance explained by the full model (fixed + random effectsP-values were obtained by likelihood ratio tests (LRT); Significance codes: *** P < 0.001, ** P < 0.01, * P < 0.05.

Figure 2 summarizes cultivar yield responses across the full range of environmental productivity captured by the ENEV network. Importantly, not all genotypes were evaluated in all environments, because new cultivars were released and entered the testing network over time, and some entries were added or removed across years. Therefore, the reaction norms depict predicted grain yield from the mixed model, where genotype responses are inferred from the deviation patterns observed in the subset of environments in which each genotype was actually tested, and then expressed along the common environmental index (EI). The EI spanned a wide gradient (≈ −2500 to +2000 kg ha^{-1} deviation from the overall mean), and predicted genotype yields increased nearly linearly with EI, ranging from roughly 1500 to 7500 Kg ha^{-1}. Figure 2 summarizes rank-type G×E by plotting cultivar rank trajectories across environments ordered by environmental effect (from strongly unfavorable to highly favorable). This pattern highlights

strong differences in overall productivity among location-year environments and shows that genotypic differences become more expressed as environments become more favorable.

Across the EI gradient, most reaction norms were broadly parallel, with a “fan-shaped” widening in high-yielding environments, indicating differential responsiveness among genotypes and predominantly scale-type G×E (differences in slope/response magnitude) rather than pervasive rank re-ordering. Several cultivars show trajectories with lower slopes, indicating comparatively stable ranking across the gradient, other varieties display pronounced rank shifts and line crossings, evidencing re-ranking among genotypes as environments improve. Overall, this pattern indicates that yield differences are not purely scale changes; rather, genotypes differ in how their relative performance is expressed across contrasting productivity levels. Nonetheless, some line crossing was evident at intermediate EI values, consistent with rank-type changes for specific cultivar comparisons. Several genotypes maintained above-average predicted yields across both low- and high-EI environments (high intercept with near-average slopes), reflecting broad adaptation, whereas others showed steeper slopes (greater responsiveness to favorable conditions) or flatter slopes (greater relative stability). High-performing and stable responses were observed across multiple breeding institutions (colors in Figure 2), indicating that broadly adapted materials are represented among entries from several programs. This is expected especially as they are mostly released cultivars and are/were broadly planted in the region.

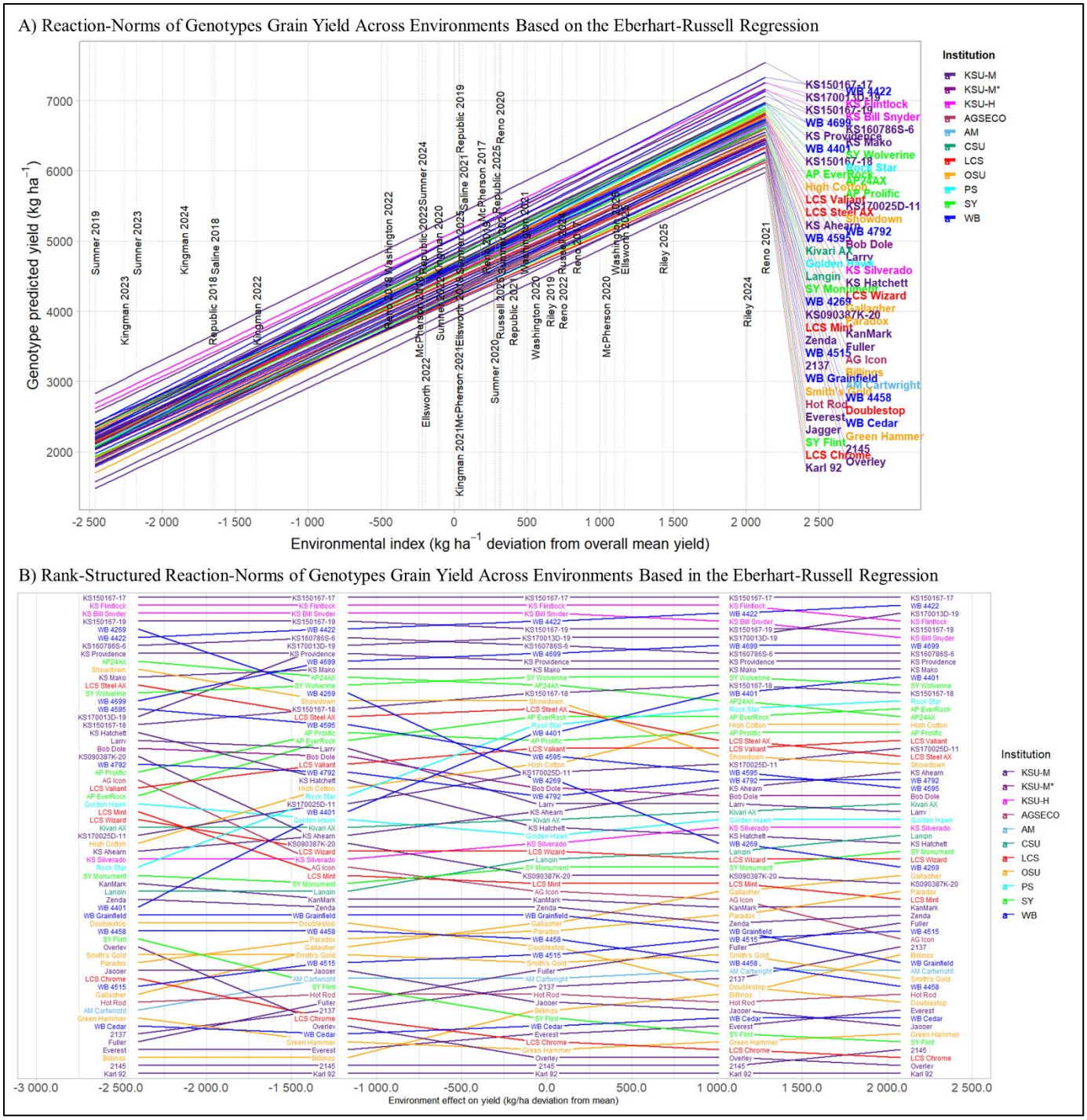

Figure 2.2. Eberhart–Russell reaction norms for grain yield across environments. (A) Model-predicted grain yield reaction norms for each genotype across the environmental index (EI), where EI is the environment mean yield deviation from the overall mean across all environments (kg ha⁻¹). Each line represents a genotype and is colored by breeding institution of origin (see legend); environments are labeled at their EI positions along the vertical dotted lines. (B) Rank-structured Eberhart–Russell reaction norms showing model-predicted genotype yield responses across EI, emphasizing changes in genotype ranking along the environmental gradient.

Lines represent genotypes and are colored by breeding institution of origin (see legend). KSU-M = Kansas State University, Manhattan Program; KSU-H = Kansas State University, Hays Program, OSU = Oklahoma State University; WB = WestBred (Bayer); LCS = Limagrain Cereal Seeds; SY/AP = AgriPro/Syngenta; CSU = Colorado State University; AM = AgriMaxx; PS = Polansky Seed; AG = AGSECO, Inc.; IG = InterGrain. *Bob Dole was developed by Kansas State University - Manhattan program and commercialized by Syngenta.

Figure 3 highlights differences among genotypes in both responsiveness (b_i) and stability (S^2d_i) across the ENEV network. Most genotypes clustered near $b_i \approx 1$ with below-average S^2d_i, indicating broadly adapted performance with relatively stable yield responses across environments. A subset of entries showed $b_i > 1$ combined with low S^2d_i (e.g., KS170025D-11, KS160786S-6, KS150167-17), suggesting strong responsiveness to favorable environments while maintaining stable performance, though these lines have been testing in one single year with a positive yield impact, it was also tested in low and high yield location within the year. Conversely, several genotypes exhibited $b_i < 1$ with low S^2d_i (e.g., KS090387K-20, KS150167-19, Jagger), consistent with more conservative responsiveness and better relative adaptation to low-yielding conditions. In contrast, a small group showed high S^2d_i values- most notably WB 4458, along with KS Providence, Billings, and LCS Chrome- indicating more plasticity performance and larger unexplained deviations from the linear response across environments. Overall, genotypes from multiple breeding institutions occurred in all quadrants, as it might be reflecting diverse breeding strategies for either broad adaptation or specific adaptation to low or high-yielding conditions. Some varieties included in the analysis were are also not specifically bred with central Kansas adaptation in mind.

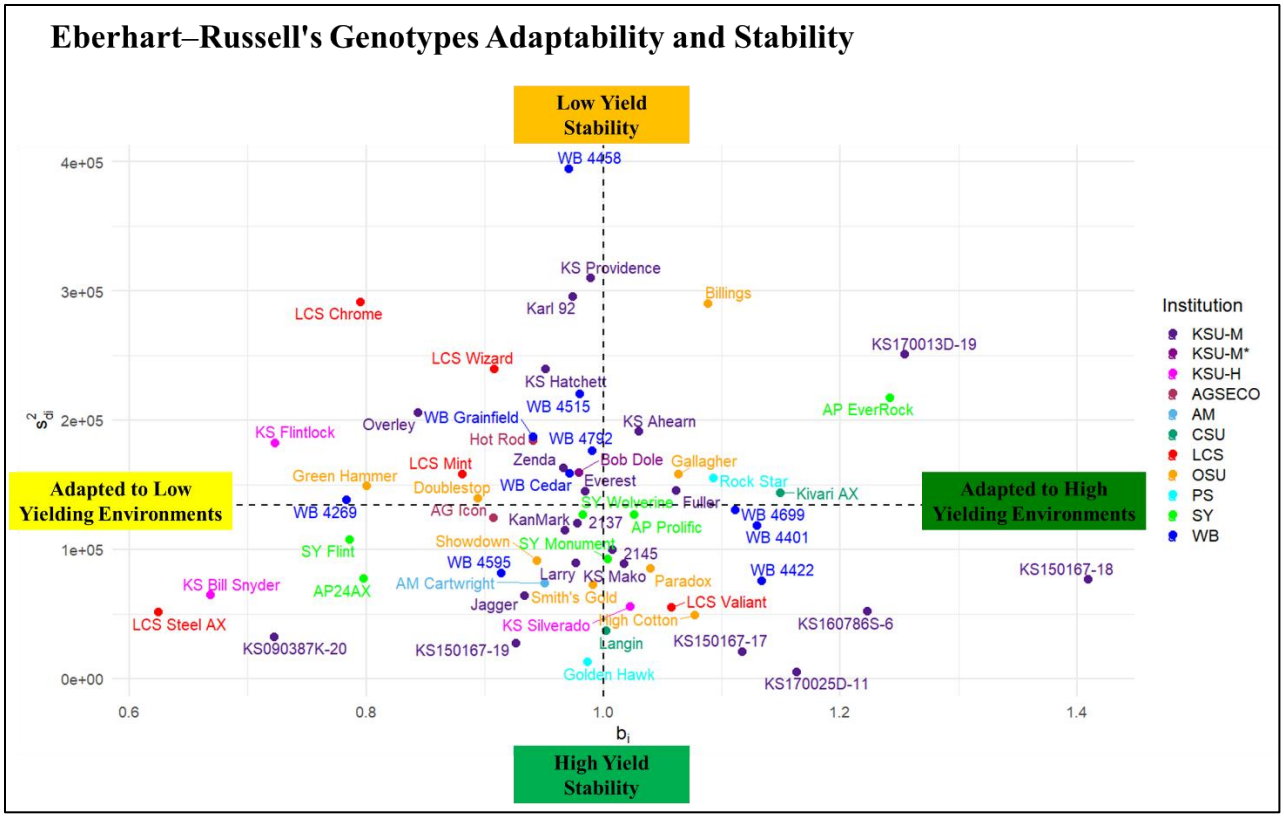

Figure 2.3. Eberhart–Russell adaptability and stability of hard red winter wheat grain yield across the ENEV multi-environment network. Each point represents a genotype. Adaptability is summarized by the Eberhart-Russell regression coefficient (b_i), estimated from regressing genotype mean yield (predicted from the mixed model) on the environmental index; $b_i \approx 1$ indicates average responsiveness (broad adaptation), $b_i < 1$ indicates better relative performance in low-yielding environments, and $b_i > 1$ indicates greater responsiveness and adaptation to high-yielding environments. Stability is summarized by the variance of deviations from regression ($\boldsymbol{S^2d_i}$); lower $\boldsymbol{S^2d_i}$ values indicate greater yield stability across environments. Dashed reference lines mark $b_i = 1$ (vertical) and the overall mean $\boldsymbol{S^2d_i}$ (horizontal), defining quadrants for interpretation. Colors indicate breeding institutions. KSU-M = Kansas State University, Manhattan Program; KSU-H = Kansas State University, Hays Program, OSU = Oklahoma State University; WB = WestBred (Bayer); LCS = Limagrain Cereal Seeds; SY/AP = AgriPro/Syngenta; CSU = Colorado State University; AM = AgriMaxx; PS = Polansky Seed; AG = AGSECO, Inc.; IG = InterGrain. *Bob Dole was developed by Kansas State University - Manhattan program and commercialized by Syngenta.

Variance components estimated separately for each environment (site-year) revealed substantial heterogeneity in the relative contribution of genetic and residual sources of variation across the ENEV network (Table 3). Genotypic variance (varG) ranged from 31,133.27 in Saline_2018 to 428,417.26 in Russell_2025, whereas residual (error) variance (varE) ranged

from 7.83 in Republic_2018 to 529,530.09 in Ellsworth_2022. Consistent with these differences in signal-to-noise ratio, entry-mean broad-sense heritability (H^2) varied widely, from 0.11 in Ellsworth_2022 to 0.94 in Republic_2018, indicating that the precision of genotype mean estimates was strongly environment-dependent. Overall model fit was generally high, with r^2 values ranging from 0.48 to 1.00 (most environments ≥ 0.80), supporting the adequacy of the within-environment mixed models to capture genotype mean patterns (Table 3). Residual variation was typically low to moderate, with CV values spanning 0.00% to 14.32%; the largest residual variability occurred in Ellsworth_2022 (CV = 14.32%), while several environments showed very low residual variation (Table 3).

Table 2.3. Variance Components, Heritability, and Model Fit.

Env	var^G	var^E	h^2	r^2	cv
Reno_2017	375558.33	49413.86	0.87	0.98	1.98
McPherson_2017	274529.99	145156.80	0.71	0.89	4.91
Ellsworth_2018	93343.68	13313.74	0.87	0.98	1.28
McPherson_2018	90797.54	91078.36	0.56	0.83	4.58
Reno_2018	210035.10	67005.09	0.80	0.94	3.69
Republic_2018	124058.60	7.83	0.94	1.00	0.00
Saline_2018	31133.27	54026.22	0.43	0.77	5.49
Riley_2019	66657.88	62226.94	0.60	0.89	3.32
Reno_2019	105037.25	81111.13	0.62	0.87	4.15
Republic_2019	327703.26	153728.49	0.74	0.91	5.52
Sumner_2019	63928.64	11949.42	0.78	1.00	2.31
Kingman_2020	75234.68	12899.25	0.85	0.98	1.26

McPherson_2020	214938.16	75081.39	0.78	0.92	2.99
Reno_2020	161471.07	229154.45	0.54	0.68	7.57
Sumner_2020	75874.12	19218.40	0.82	0.96	1.66
Washington_2020	153185.55	29719.46	0.87	0.95	2.03
Kingman_2021	54023.95	37904.10	0.67	0.90	2.89
McPherson_2021	101433.53	53463.47	0.77	0.84	3.69
Reno_2021	211483.25	102123.05	0.77	0.92	3.33
Republic_2021	238818.62	77964.76	0.84	0.88	3.90
Saline_2021	141527.41	24969.43	0.90	0.94	2.20
Sumner_2021	91462.92	52408.04	0.75	0.87	3.40
Washington_2021	215869.45	57700.55	0.87	0.90	3.37
Ellsworth_2022	42993.68	529530.09	0.11	0.48	14.32
Kingman_2022	50734.86	48105.13	0.58	0.87	4.76
Reno_2022	118631.71	82734.71	0.58	0.91	3.37
Republic_2022	175732.59	51938.91	0.83	0.93	3.42
Sumner_2022	84253.58	46353.37	0.76	0.80	3.57
Washington_2022	64454.49	103095.30	0.51	0.80	6.16
Kingman_2023	33763.04	11934.22	0.76	0.94	2.90
Sumner_2023	77657.19	22610.86	0.84	0.92	4.31
Kingman_2024	71851.18	49527.39	0.70	0.83	6.03
Riley_2024	247367.08	156998.94	0.72	0.82	4.29
Russell_2024	286394.62	37448.01	0.90	0.96	2.06
Sumner_2024	186324.23	48067.27	0.87	0.92	3.45
Ellsworth_2025	38944.52	122978.78	0.32	0.78	4.72

Riley_2025	161883.60	149144.10	0.61	0.84	4.55
Republic_2025	242928.06	21585.35	0.93	0.98	1.83
Russell_2025	428417.26	123015.80	0.85	0.91	4.97
Sumner_2025	369019.90	31336.75	0.92	0.98	1.99
Washington_2025	299441.61	44152.98	0.91	0.97	2.32

var^G = genotypic variance; var^E = residual (error) variance; h^2 = entry-mean broad-sense heritability; outliers = proportion of phenotypic records removed after outlier detection; r^2 = coefficient of determination between observed and model-predicted genotype means; cv = residual coefficient of variation (%). Variance components were obtained from linear mixed models fitted separately for each environment.

Figure 4 compares genotypes for overall yield performance across the ENEV network using the Eberhart-Russell model intercept (μ_g) as an estimate of mean yield effect. Genotypes were ranked from lowest to highest μ_g, showing a broad distribution of mean performance around the network average. Mean yield effects ranged from −682.6 kg ha^{-1} for the lowest-performing entry to +849.3 kg ha^{-1} for the highest-performing entry, indicating substantial genetic variation in overall yield potential across the panel. Genotypes on the left side of the ranking had negative μ_g values, reflecting below-average performance across environments, whereas those on the right showed positive μ_g values and above-average performance. Confidence interval widths varied among genotypes, consistent with differences in estimation precision expected in a long-term, unbalanced multi-environment dataset where entries were not present in all site-years.

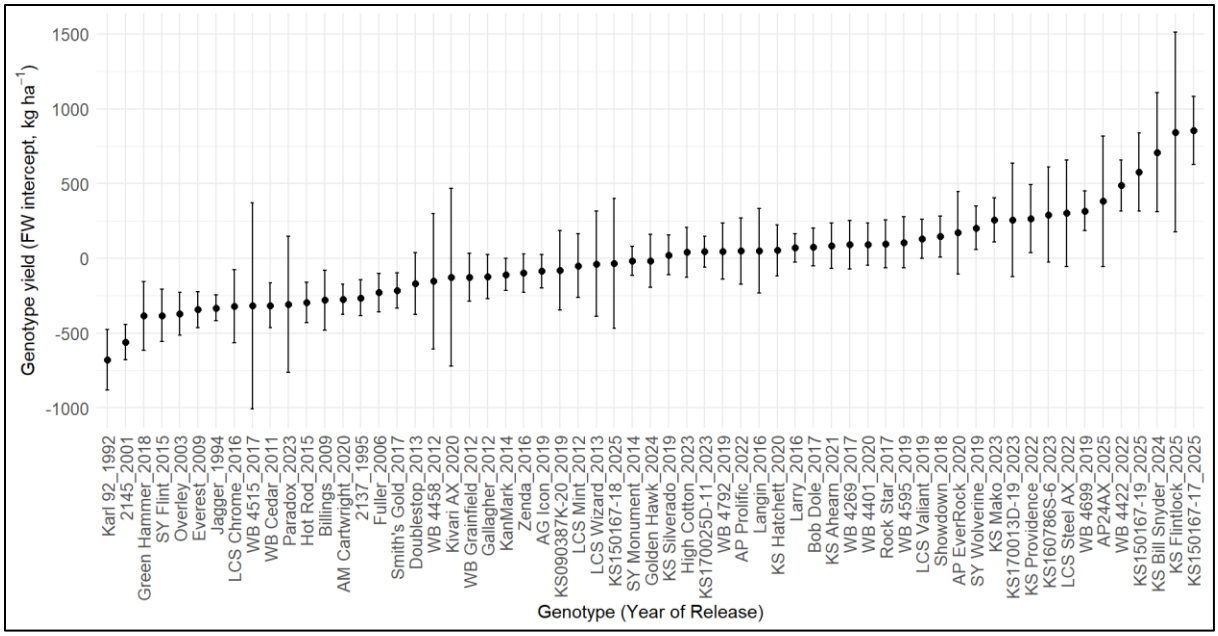

Figure 2.4. Genotypes mean yield effects across the ENEV network based on the Eberhart–Russell mixed model. Points represent genotype mean yield effects (FW intercept, μg) expressed as kg ha⁻¹ relative to the overall mean (0 = network average). Error bars show the uncertainty of μg (± 1.96 SE, i.e., 95% confidence intervals). Genotypes are ordered from lowest to highest μg and labeled with year of release. Because the trial set is unbalanced (not all genotypes were tested in all environments due to staggered entry and release over time), μg values are model-based estimates derived from the environments in which each genotype was evaluated.

Figure 5A summarizes genotype rank performance and stability across the ENEV network using a "ranking genotypes" GGE biplot based on the environment-centered (G+GE) matrix. The first two principal components captured the major patterns of G×E structure for yield. Genotypes positioned closer to the "ideal genotype" target (center of the concentric circles) combine better average rank performance with greater stability across environments, whereas genotypes located farther from the target show either lower rank performance, greater instability, or both. Environments that cluster together indicate more similar genotype discrimination patterns, while environments positioned farther apart reflect stronger differences in genotype ranking responses.

Figure 5B provides a complementary summary of rank stability by displaying each genotype's mean rank across environments together with its rank variability (± SD). Genotypes are ordered from best to worst mean rank, and the dashed vertical line marks the threshold of mean rank ≤ 10, highlighting entries that fall, on average, within the top ~20% of the tested genotypes. Narrower SD bars indicate more stable ranks across environments, whereas wider bars suggest greater rank variability. Point size reflects the number of environments in which each genotype was evaluated, emphasizing that rank stability estimates are supported by more evidence for genotypes tested in more site-years. Several genotypes met the mean-rank ≤ 10 criteria, including WB 4422, KS Flintlock, KS Bill Snyder, KS Providence, WB 4699, KS Mako, and LCS Wizard.

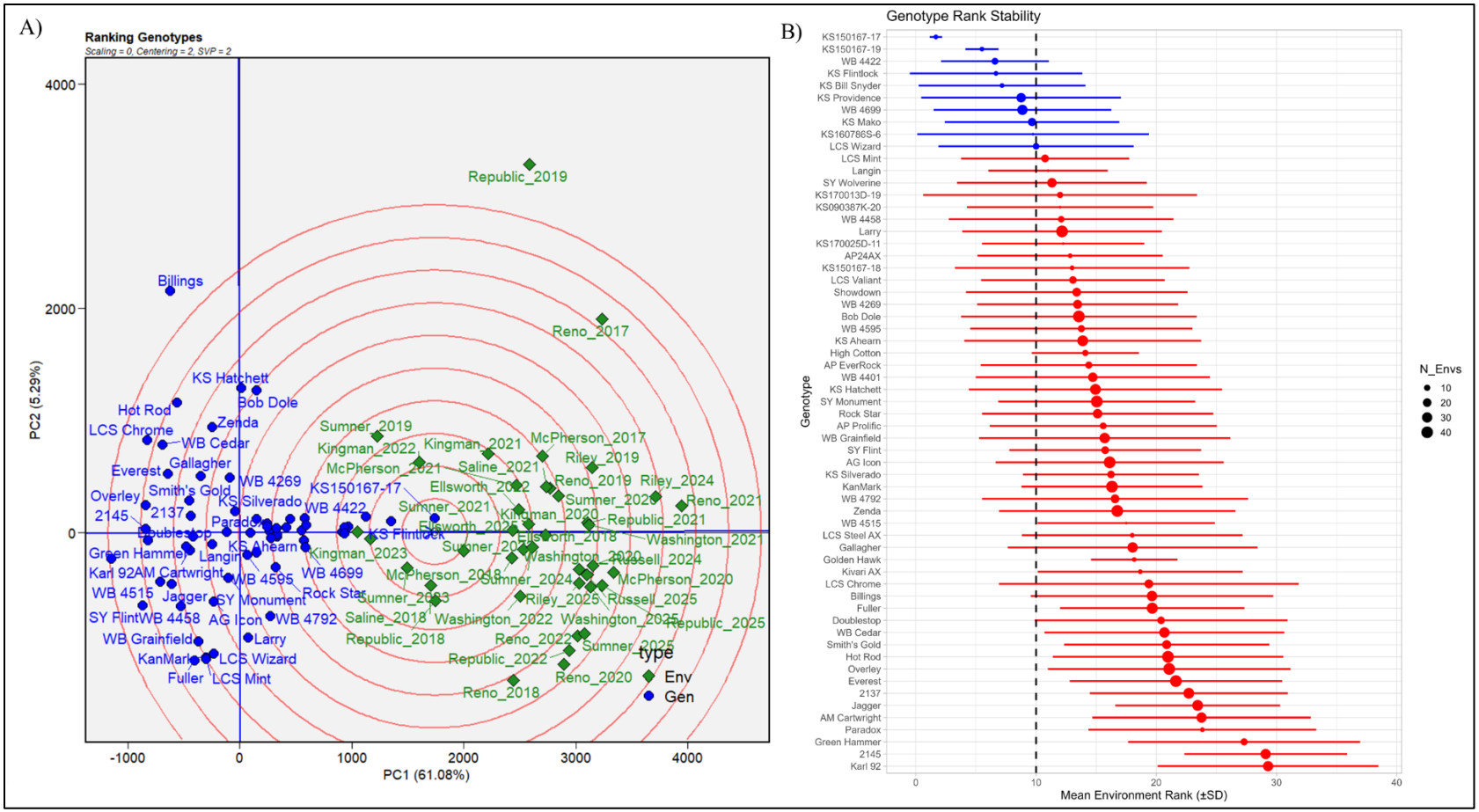

Figure 2.5. Genotype rank performance and stability across environments using a GGE biplot and mean-rank variability. A) "Ranking genotypes" GGE biplot for grain yield based on the environment-centered (G+GE) matrix (centering = 2) with singular value partitioning to genotype scores (SVP = 1). Genotypes are shown as blue points and environments as green points. The "ideal genotype" target is indicated at the center of the concentric circles; genotypes closer to this target combine higher mean performance with greater stability in ranking across environments, whereas genotypes farther away show poorer rank and/or greater instability. B) Mean genotype rank across environments (x-axis) with ± SD bars indicating rank stability (smaller SD = more stable ranks). Point size is proportional to the number of environments in which each genotype was evaluated. The vertical dashed line indicates the mean-rank threshold used to highlight entries with average rank ≤ 10 (blue) versus > 10 (red).

Linear genetic gain trend analysis indicated clear evidence of yield improvement over time in the ENEV cultivar panel (Table 4; Figure 6). Across all entries released from 1992–2025, regressing grain yield BLUPs on year of release returned a positive slope of 24.95 kg ha^{-1} yr^{-1} (Y = 24.95 * x + -45851.43), corresponding to an estimated cumulative gain of 823.19 kg ha^{-1} over 33 years (21.44 % total; 0.65% yr^{-1}). When restricted to KSU-M entries, the long-term rate of gain was similar (24.48 kg ha^{-1} yr^{-1}; Y = 24.48 × Year − 44,865.67), translating to a total gain of 807.95 kg ha^{-1} (20.69% total; 0.63% yr^{-1}) over the same period. Notably, the more recent release window (2009–2025) showed steeper gains, with slopes of 36.71 kg ha^{-1} yr^{-1} overall (Y = 37.71 × Year – 71,626.60; 603.36 kg ha^{-1}, 14.60% total; 0.91% yr^{-1}) and 42.79 kg ha^{-1} yr^{-1} for KSU-M (Y = 42.79 × Year − 81,849.85; 684.60 kg ha^{-1}, 16.66% total; 1.04% yr^{-1}). Consistent with these estimates, Figure 6 shows a clear upward shift in cultivar BLUPs with release year, with modern cultivars clustering toward the upper range of predicted yield while older cultivars largely occupy the lower portion of the distribution, supporting continued genetic progress for grain yield across decades of cultivar development.

Table 2.4. Estimated genetic gain over time based on regression of yield BLUPs on year of release.

Period	Equation	Total Gain (kg ha^{-1})	Gain (kg ha^{-1} year^{-1})	Total Gain (%)	Annual Gain (%)
KSU-M (1992–2025)	Y = 24.48 * x + -44865.67	807.95	24.48	20.69	0.63
Overall (1992–2025)	Y = 24.95 * x + -45851.43	823.19	24.95	21.44	0.65
KSU-M (2009–2025)	Y = 42.79 * x + -81849.85	684.60	42.79	16.66	1.04
Overall (2009–2025)	Y = 37.71 * x + -71626.60	603.36	36.71	14.60	0.91

Linear regression slopes from the model Yield BLUP ~ Year of release, corresponding to the fitted lines shown in Figure 6. Total gain (kg ha^{-1}) calculated as slope × length of the period (1992-2025 = 33 years) and (2009-2025 = 16 years). Total gain (%) and annual gain (%) were calculated relative to the predicted yield at the earliest year of the period (i.e., baseline defined from the fitted regression).

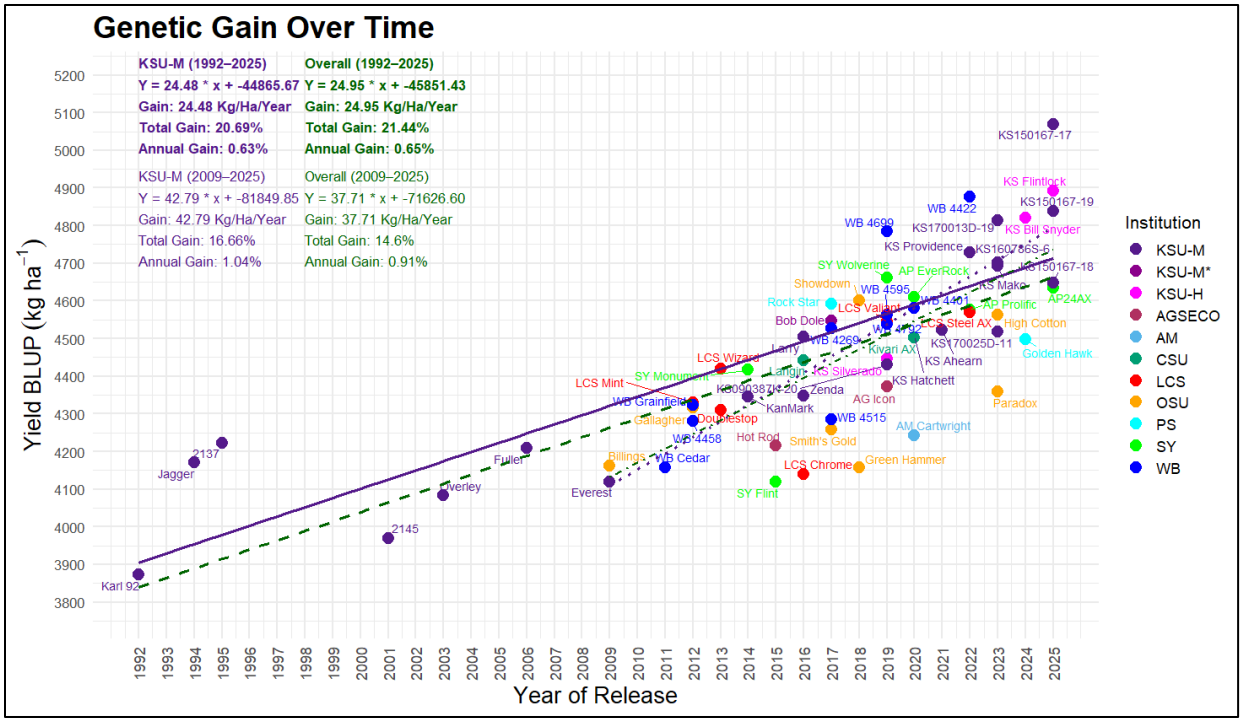

Figure 2.6. Genetic gain in grain yield across hard red winter wheat cultivars as a function of year of release. Each point represents a cultivar's grain yield BLUP (kg ha⁻¹) averaged across environments, plotted by year of release and colored by breeding institution (cultivar names are annotated). Genetic gain was quantified as the slope of linear regressions of yield BLUP on release year for KSU-M entries and for all entries combined. Solid lines summarize long-term trends across the full release window (1992-2025), while dotted lines highlight more recent trends for modern releases (2009-2025); corresponding regression equations and gain metrics (kg ha⁻¹ yr⁻¹ and percent gain) are reported on the plot. KSU-M = Kansas State University, Manhattan Program; KSU-H = Kansas State University, Hays Program, OSU = Oklahoma State University; WB = WestBred (Bayer); LCS = Limagrain Cereal Seeds; SY/AP = AgriPro/Syngenta; CSU = Colorado State University; AM = AgriMaxx; PS = Polansky Seed; AG = AGSECO, Inc.; IG = InterGrain. *Bob Dole was developed by Kansas State University - Manhattan program and commercialized by Syngenta.

Discussion and Conclusions

The ENEV serves as a practical "thermometer" for monitoring wheat performance under the dynamic and often unpredictable conditions of central Kansas. Consistent with the nature of Kansas production systems, the across-environment variance component analysis (Table 2) showed that environmental differences accounted for most of the phenotypic variation when modeling all experiments together, while genotype and G×E effects were also significant,

confirming that both overall productivity differences among site-years and differential cultivar responses contribute to observed yield outcomes.

Environmental heterogeneity across Kansas remains a primary driver of yield variability and helps explain why genotypes differ in mean performance and stability across the network. Soil attributes such as pH, organic matter, and texture are known to influence root development, nutrient availability, and water capture, with downstream impacts on yield formation (Hobson et al., 2023; Gámez-Arjona et al., 2022). In parallel, spatial and temporal variation in precipitation and temperature imposes changing stress profiles across regions and years- particularly drought and episodic heat during critical developmental stages- which can strongly reduce wheat yield potential (Reynolds et al., 2007; Reynolds et al., 2021; Asseng et al., 2011; Singh Sangha et al., 2025). These environmental contrasts underscore the importance of selecting adapted cultivars that either maintain stable performance across variable conditions or express adaptive plasticity, where responsiveness translates into agronomic advantage rather than unpredictable yield fluctuations (Whitman & Agrawal, 2009).

Reaction norm analyses (Figure 2) revealed that genotype yield responses increased along the environmental index gradient, with most genotypes showing broadly similar direction of response but differing in magnitude. Our findings indicate substantial phenotypic plasticity for yield. Most cultivars showed “fan-shaped” response patterns, producing much higher yields than others in the best environments. This plasticity is adaptive under Kansas conditions, as genotypes with higher responsiveness achieved greater yield in high-input, and relatively high-yield without a proportional penalty in poor conditions. Such patterns agree with evidence that breeding and management intensification have favored genotypes capable of exploiting favorable conditions (Jaenisch & Lollato, 2019). For example, enhanced fertility and integrated management

increased average yields by ~14% (from ~63 to 72 bu/ac) in Kansas wheat trials (Jaenisch & Lollato, 2019). This implies that genotypes with high plasticity will respond strongly to fertilizer, fungicides, and other inputs, widening yield gaps between the most and least responsive cultivars. Conversely, less plastic (more "stable") genotypes may underperform under intensive management. Our results therefore reinforce the view that phenotypic plasticity is agronomically beneficial, yielding more when conditions improve (Sadras et al., 2016; Sultan, 1987). From a breeding standpoint, these G×E patterns suggest strategic opportunities. In high-yield environments, genotype rankings diverge and plasticity differences become more pronounced, which increases the accuracy of selection (Cooper & Messina, 2023) evidenced in figure 5B. In moderately yielding trials (e.g., with limited inputs or mild stress), the total environmental variance can be lower or more uniform across plots (Renaud et al., 2014). With less "noise" from micro-environmental fluctuations, the signal from genetic differences is relatively stronger, boosting heritability. By contrast, high-yield environments can occasionally introduce new variability (e.g., pest outbreaks or lodging in lush plots) that masks genetic effects. In practice, this means that selecting parents and breeding lines in sites or seasons with above-average productivity may yield higher genetic gains in yield as we may select cultivar with positive plasticity. It also argues for exploiting genotype-specific management, deploying highly plastic cultivars in regions or years with assured high inputs, while using more stable cultivars where conditions are marginal. Such an approach mirrors the "genotype-by-management" paradigm (Lollato et al., 2019). In summary, our data and others' (Jaenisch & Lollato, 2019; Raj et al., 2023; Giordano et al., 2025) suggest that yield plasticity and management interact synergistically, highlighting the need to consider agronomy and genetics together when aiming to raise wheat yields. In other words, as breeding has increased the interdependence of traits,

yield becomes more influenced by environmental noise (Sadras, V. O., 2024). Notably, yield and its stability also appear to be under partially independent genetic control. The classic theory of Bradshaw (1965) proposed that a trait baseline and its plastic responsiveness could map to different genes, and empirical studies (Kusmec et al., 2017; Reymond et al., 2003) support this view.

The within-environment analyses (Table 3) further demonstrate that the reliability of selection varies substantially among site-years. This variation is expected in long-term ENEV networks and has practical implications for resource allocation. Environments that both represent relevant production conditions and effectively discriminate among genotypes are particularly valuable for breeding progress (Holland et al., 2003; Costa-Neto et al., 2021). Together, these results support continued emphasis on robust trial design and spatial adjustment, as well as maintaining broad environmental representation to capture meaningful G×E patterns.

The rank-based summaries (Figure 5A–B) provide complementary evidence for identifying broadly adapted, high-ranking cultivars and separating them from entries that perform well only in specific contexts. In Figure 5A, genotypes positioned closer to the “ideal genotype” target represent combinations of favorable average performance and rank stability across environments, while more distant genotypes reflect poorer rank performance and/or instability. Figure 5B reinforces these patterns by displaying mean rank and rank variability (± SD), while also accounting for differences in the number of environments tested (point size). Genotypes that achieve strong ranks in a limited set of environments may not generalize across the full range of Kansas conditions. Therefore, combining parametric reaction-norm approaches with rank-based summaries provides a more complete view of performance, stability, and selection risk (Pour-Aboughadareh et al., 2022; Khare et al., 2024). In practice, breeders can thus

use a mean-rank vs rank-SD plot to identify “specialists” (highly ranked only in some environments) versus “generalists” (consistently well-ranked). Further genetic studies could confirm if these are related or unique genetic underpinnings driving the yield stability response.

The genetic trend analysis (Table 4; Figure 6) indicates sustained progress in grain yield over time within this cultivar panel, especially the modern era from ~2009 to 2025. These trends align with prior reports of long-term genetic gain in Kansas and Great Plains wheat, which demonstrate roughly ~1% annual yield improvements historically. For instance, extensive trials in the southern Great Plains (1959–2022) show significant genetic gain (~24 kg ha^{-1} yr^{-1}) over decades, with continued improvement even in recent years (~47 kg ha^{-1} yr^{-1} in the last 15 years) (Boehm et al., 2023). This supports the conclusion that consistent field-based evaluation, combined with modern breeding strategies, continues to deliver incremental improvement (Cox et al., 1989; Donmez et al., 2001; Graybosch and Peterson, 2010; Rife et al., 2019; Crespo-Herrera et al., 2021; Boehm et al., 2023). In the context of rising climatic uncertainty and intensifying abiotic stress risk, maintaining this “thermometer” dataset is especially valuable for detecting shifts in cultivar performance, identifying resilient materials, and guiding breeding and deployment strategies.

A complementary analysis of cultivar BLUPs (data not shown) indicated that heading date (HD) and year of release jointly explained a substantial portion of yield differences in this panel (multiple regression $R^2 = 0.55$), with higher yields associated with both more recent releases and a tendency toward later-heading phenology. This pattern is notable because Kansas variety guidance has historically emphasized that the “successful” agronomic profile differs by region (e.g., early to medium-early maturity often fitting central Kansas, versus more medium to medium-late types in the west), while also stressing that disease and pest pressure - and thus

realized yield - depends on year-to-year pathogen dynamics and management decisions (Onofre, K. A., et al., 2025). In that context, the positive HD-yield association is consistent with a varietal shift toward longer-season cultivars that can capitalize on favorable springs (potentially via greater tiller survival and sink development), alongside a management landscape in which rust risk is increasingly mitigated. For example, the Kansas variety guide explicitly notes that western-adapted cultivars such as “Avery” is “later” with improved yield potential but may “need a fungicide” in rust years - illustrating how yield gains can be tied to phenology choices that are viable only when disease is actively managed (Onofre, K. A., et al., 2023). Long-term Kansas fungicide research summarized by De Wolf, E. D. (2020) further supports this framing: a single foliar application between flag-leaf emergence and anthesis often increases yield by ~4-14% (average ~10%) under disease-favorable scenarios, effectively reducing the historical reliance on “maturity escape” as the primary way to avoid rust-driven yield loss. Mechanistically, the phenology signal also aligns with Kansas wheat development principles - tillers can contribute a large share of final yield under Kansas conditions - so management that preserves canopy health later into the season can amplify the payoff of later phenology (Thiry, D. E., et all., 2002). The regression context helps explain why an early-maturing cultivar like Everest can appear less competitive in this specific analysis despite periods of strong producer adoption: Everest is described as early and moderately resistant to barley yellow dwarf (BYDV), and BYDV itself can explain meaningful yield variation across years (with models showing sizable yield penalties at high incidence, while moderate resistance in Everest was estimated to reduce BYDV-related yield loss by ~73%) (Gaunce, G. M., et. al., 2015). Because BYDV pressure is episodic and strongly shaped by aphid dynamics and management (e.g., delayed planting relative to fly-free dates, seed treatments, and use of moderately resistant varieties),

shifts in disease/pest pressure and management can change which maturity strategies “win” over time - supporting a coherent interpretation that recent genetic gain is partly realized through later phenology enabled by improved disease control rather than by earlier maturity alone (De Wolf, E. D., 2018).

Looking forward, the integration of genomic selection and enviromics-informed selection frameworks will increase the rate of gain while improving targeting of genotypes to specific stress patterns (Costa-Neto et al., 2021; Zhang-Biehn et al., 2021; Kusunose et al., 2023). Stability in crop performance should be interpreted in a dynamic biological sense, because living organisms are not static systems but continuously respond to changing internal and external conditions. In plant breeding, yield stability does not imply that a genotype produces identical yields across all environments; rather, it reflects the ability to maintain predictable performance and avoid disproportionate yield loss when environmental conditions fluctuate (Becker & Léon, 1988; Annicchiarico, 2002). Therefore, stability is inherently tied to adaptation, but it represents the portion of adaptation that remains consistent across relevant environmental gradients. The extent to which a genotype expresses stability depends on both its genetic background and the environmental context in which it is evaluated, because the same cultivar may appear stable in one target region but unstable in another due to genotype-by-environment interaction (Finlay & Wilkinson, 1963; Eberhart & Russell, 1966; Yan & Kang, 2002). For this reason, stability must be treated as a context-dependent property, emerging from the interaction between genetic potential and environmental variability rather than as an intrinsic fixed trait of the genotype. In summary, our multi-environment analysis of Kansas hard red wheat reveals that environmental variability is the dominant determinant of yield, but modern breeding has shifted genotypes toward greater yield plasticity. High-plasticity cultivars capitalize on favorable conditions,

especially under intensive management, but their superiority hinges on inputs and good weather. We also acknowledge that the genetic architecture of yield and its plasticity can be at least partly decoupled (Bradshaw, 1965), exemplified by major loci like the $2N^{v}S$ translocation (Gao et al., 2021, Giordano, et al., 2025). Rising phenotypic similarity appears to have lowered broad-sense heritability of yield (Giordano et al., 2025) even as genetic gain continues, suggesting that future progress will depend on combining high selection intensity with genomics and environmics tools (Costa-Neto et al., 2021). These insights carry direct implications for wheat improvement. Breeders should exploit favorable environments to detect genetic differences but also test lines in marginal sites to ensure stability. Agronomists should leverage genotype-specific management, recognizing that inputs will unevenly benefit cultivars. Together, these strategies, informed by continued research on G×E and plasticity, can help sustain wheat yield gains and adaptability in the face of evolving challenges.

References

Annicchiarico, P. (2002). Genotype x environment interactions: challenges and opportunities for plant breeding and cultivar recommendations.

Aparicio J (2022). MrBean: Web application for analyzing field experiments. R package version 2.0.9, https://github.com/AparicioJohan/MrBeanApp.

Asseng, S., Foster, I. A. N., & Turner, N. C. (2011). The impact of temperature variability on wheat yields. Global change biology, 17(2), 997-1012.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. convergence, 12(1), 2.

Batista, L. A., Bandillo, N., Friskop, A., & Green, A. (2024). Accelerating genetic gain through strategic speed breeding in spring wheat. Crop Science, 64(6), 3311-3322.

Becker, H. C., & Leon, J. I. (1988). Stability analysis in plant breeding. Plant breeding, 101(1).

Bradshaw, A. D. (1965). Evolutionary significance of phenotypic plasticity in plants. Advances in genetics, 13, 115-155.

Breiman, L. (1996). Bagging predictors. Machine learning, 24, 123-140.

Boehm Jr., J. D., Masterson, S., Palmer, N., Cai, X., & Miguez, F. (2023). Genetic improvement of winter wheat (Triticum aestivum L.) grain yield in the Northern Great Plains of North America, 1959–2021. Crop Science, 63(6), 3236–3249.

Cooper, M., & Messina, C. D. (2023). Breeding crops for drought-affected environments and improved climate resilience. The Plant Cell, 35(1), 162-186.

Costa-Neto, G., Galli, G., Carvalho, H. F., Crossa, J., & Fritsche-Neto, R. (2021). EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. G3, 11(4), jkab040.

Crespo-Herrera, L. A., Crossa, J., Huerta-Espino, J., Mondal, S., Velu, G., Juliana, P., ... & Singh, R. P. (2021). Target population of environments for wheat breeding in India: definition, prediction and genetic gains. Frontiers in Plant Science, 12, 638520.

Cox, T. S., Shogren, M. D., Sears, R. G., Martin, T. J., & Bolte, L. C. (1989). Genetic improvement in milling and baking quality of hard red winter wheat cultivars, 1919 to 1988. Crop Science, 29(3), 626-631.

De Wolf, E. D. (2018). Barley yellow dwarf (EP165). Kansas State University Agricultural Experiment Station and Cooperative Extension Service.

De Wolf, E. D. (2020). Evaluating the need for foliar fungicides on wheat (Agronomy eUpdate, Issue 797). Kansas State University Department of Agronomy.

Donmez, E., Sears, R. G., Shroyer, J. P., & Paulsen, G. M. (2001). Genetic gain in yield attributes of winter wheat in the Great Plains. Crop science, 41(5), 1412-1419.

Gámez-Arjona, F. M., Sánchez-Rodríguez, C., & Montesinos, J. C. (2022). The root apoplastic pH as an integrator of plant signaling. Frontiers in Plant Science, 13, 931979.

Gao, Liangliang, Dal-Hoe Koo, Philomin Juliana, Trevor Rife, Daljit Singh, Cristiano Lemes da Silva, Thomas Lux et al. "The Aegilops ventricosa 2NvS segment in bread wheat: cytology, genomics and breeding." Theoretical and Applied Genetics 134, no. 2 (2021): 529-542.

Gaunce, G. M., & Bockus, W. W. (2015). Estimating yield losses due to barley yellow dwarf on winter wheat in Kansas using disease phenotypic data. Plant Health Progress, 16(1), 1–6. doi:10.1094/PHP-RS-14-0039

Giordano, N., Sadras, V. O., Guttieri, M. J., Cruppe, G., Bai, G., Fritz, A., ... & Lollato, R. (2025). Genetic modulation of yield and phenotypic plasticity of yield in winter wheat. Journal of Experimental Botany, eraf478.

Graybosch, R. A., & Peterson, C. J. (2010). Genetic improvement in winter wheat yields in the Great Plains of North America, 1959–2008. Crop Science, 50(5), 1882-1890.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., ... & Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. science, 327(5967), 812-818.

Graybosch, R. A., & Peterson, C. J. (2012). Specific adaptation and genetic progress for grain yield in Great Plains hard winter wheats from 1987 to 2010. Crop Science, 52(2), 631-643.

Finlay, K. W., & Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. Australian journal of agricultural research, 14(6), 742-754.

Eberhart, S. T., & Russell, W. (1966). Stability parameters for comparing varieties 1. Crop science, 6(1), 36-40.

Hobson, D. J., Harty, M. A., Langton, D., McDonnell, K., & Tracy, S. R. (2023). The establishment of winter wheat root system architecture in field soils: The effect of soil type on root development in a temperate climate. Soil use and Management, 39(1), 198-208.

Holman, J. D., Obour, A. K., O'Brien, D., & Assefa, Y. (2024). Historic winter wheat yield, production, and economic value trends in Kansas, the “Wheat State”. Crop Science, 64(2), 925-941.

Holland, J. B., Nyquist, W. E., Cervantes-Martínez, C. T., & Janick, J. (2003). Estimating and interpreting heritability for plant breeding: an update. Plant breeding reviews, 22(9-112).

Holling, C. S. (1973, November). Resilience and stability of ecological systems.

Huber, P. J., & Ronchetti, E. M. (2011). Robust statistics. John Wiley & Sons.

Jaenisch, B. R., & Lollato, R. P. (2019). Integrated wheat management for improved wheat yield and protein in Kansas. Kansas Agricultural Experiment Station Research Reports, 5(6), 14**.**

Khare, V., Shukla, R. S., Pandey, S., Singh, S. K., & Singh, C. (2024). Exploring the genotype-environment interaction of bread wheat in ambient and high-temperature planting conditions: a rigorous investigation. Scientific Reports, 14(1), 2402.

Kusmec, A., Srinivasan, S., Nettleton, D., & Schnable, P. S. (2017). Distinct genetic architectures for phenotype means and plasticities in Zea mays. Nature plants, 3(9), 715-723.

Kusunose, Y., Rossi, J. J., Van Sanford, D. A., Alderman, P. D., Anderson, J. A., Chai, Y., ... & Wright, B. D. (2023). Sustaining productivity gains in the face of climate change: A research agenda for US wheat. Global change biology, 29(4), 926-934.

Lell, M., Reif, J., & Zhao, Y. (2021). Optimizing the setup of multienvironmental hybrid wheat yield trials for boosting the selection capability. The Plant Genome, 14(3), e20150.

Lollato, R. P., Ruiz Diaz, D. A., DeWolf, E., Knapp, M., Peterson, D. E., & Fritz, A. K. (2019). Agronomic practices for reducing wheat yield gaps: a quantitative appraisal of progressive producers. Crop Science, 59(1), 333-350.

Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibián-Barrera, M. (2019). Robust statistics: theory and methods (with R). John Wiley & Sons.

Mayar, K., Carmichael, D. G., & Shen, X. (2022). Stability and resilience—A systematic approach. Buildings, 12(8), 1242.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. John Wiley & Sons.

Moro Rosso, L. H., de Borja Reis, A. F., Correndo, A. A., & Ciampitti, I. A. (2021). XPolaris: an R-package to retrieve United States soil data at 30-meter resolution. BMC Research Notes, 14(1), 327.

NASS. (2022). Kansas wheat history. USDA, NASS. https://www.nass.usda.gov/Statistics_by_State/Kansas/Publications/Cooperative_Projects/KS-wheat-history22.pdf

Nassar, R., & Huehn, M. (1987). Studies on estimation of phenotypic stability: Tests of significance for nonparametric measures of phenotypic stability. Biometrics, 45-53.

Obembe, O. S., Hendricks, N. P., & Tack, J. (2021). Decreased wheat production in the USA from climate change driven by yield losses rather than crop abandonment. Plos one, 16(6), e0252067.

Olivoto, T., & Lúcio, A. D. C. (2020). metan: An R package for multi-environment trial analysis. Methods in Ecology and Evolution, 11(6), 783-789.

Onofre, K. A., De Wolf, E. D., Lollato, R., & Whitworth, R. J. (2023). Kansas Wheat Variety Guide 2023 (MF991). Kansas State University Research and Extension.

Onofre, K. A., De Wolf, E. D., Lollato, R., & Whitworth, R. J. (2025, August). Kansas Wheat Variety Guide 2025: Wheat ratings (MF991). Kansas State University Agricultural Experiment Station and Cooperative Extension Service.

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

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.

Raj, A. S., Siliveru, K., McLean, R., Prasad, P. V. V., & Lollato, R. P. (2023). Intensive management simultaneously reduces yield gaps and improves milling and baking properties of bread wheat. Crop Science, 63(2), 936-955.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rife, T. W., Graybosch, R. A., & Poland, J. A. (2019). A field-based analysis of genetic improvement for grain yield in winter wheat cultivars developed in the US central plains from 1992 to 2014. Crop Science, 59(3), 905-910.

Renaud, E. N., Lammerts van Bueren, E. T., Paulo, M. J., van Eeuwijk, F. A., Juvik, J. A., Hutton, M. G., & Myers, J. R. (2014). Broccoli cultivar performance under organic and conventional management systems and implications for crop improvement. Crop Science, 54(4), 1539-1554.

Reymond, M., Muller, B., Leonardi, A., Charcosset, A., & Tardieu, F. (2003). Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. Plant physiology, 131(2), 664-675.

Reynolds, M., Dreccer, F., & Trethowan, R. (2007). Drought-adaptive traits derived from wheat wild relatives and landraces. Journal of experimental botany, 58(2), 177-186.

Reynolds, M. P., Lewis, J. M., Ammar, K., Basnet, B. R., Crespo-Herrera, L., Crossa, J., ... & Braun, H. J. (2021). Harnessing translational research in wheat for climate resilience. Journal of Experimental Botany, 72(14), 5134-5157.

Sadras, V. O. (2024). Bread and hummus: trait connectance and correlation pleiades in grain crops. Journal of Experimental Botany, 75(22), 6903-6908.

Sadras, V. O., Hayman, P. T., Rodriguez, D., Monjardino, M., Bielich, M., Unkovich, M., ... & Wang, E. (2016). Interactions between water and nitrogen in Australian cropping systems: physiological, agronomic, economic, breeding and modelling perspectives. Crop and Pasture Science, 67(10), 1019-1053.

Singh Sangha, J., Wang, W., Knox, R., Ruan, Y., Cuthbert, R. D., Isidro-Sánchez, J., ... & Selvaraj, G. (2025). Phenotypic plasticity of bread wheat contributes to yield reliability under heat and drought stress. PloS one, 20(3), e0312122.

Sultan, S. E. (1987). Evolutionary implications of phenotypic plasticity in plants. In Evolutionary biology: volume 21 (pp. 127-178). Boston, MA: Springer US.

Thiry, D. E., Sears, R. G., Shroyer, J. P., & Paulsen, G. M. (2002). Tillering in wheat (SRL134). Kansas State University Research and Extension.

Weisberg, S. (2005). Applied linear regression. https://www.stat.cmu.edu/~brian/valerie/617-2022/0%20-%20books/weissberg%20ALR%203rd%20ed.pdf

Whitman, D. W., & Agrawal, A. A. (2009). What is phenotypic plasticity and why is it important. Phenotypic plasticity of insects: Mechanisms and consequences, 1-63.

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

Zhang-Biehn, S., Fritz, A. K., Zhang, G., Evers, B., Regan, R., & Poland, J. (2021). Accelerating wheat breeding for end-use quality through association mapping and multivariate genomic prediction. The Plant Genome, 14(3), e20164.

Zobel, R. W., Wright, M. J., & Gauch Jr, H. G. (1988). Statistical analysis of a yield trial. Agronomy journal, 80(3), 388-393.

Chapter 3 - Genetic Architecture of Floral Traits and Fusarium Head Blight in D-Genome–Enriched Wheat Panel

Abstract

Floral and phenological traits can influence both hybrid wheat outcrossing potential and Fusarium head blight (FHB) infection dynamics at anthesis, yet their genetic architecture in winter wheat remains incompletely resolved. We evaluated 101 *Aegilops tauschii*–derived hard winter wheat introgression lines (34 pedigree families) across two greenhouse environments (Floral traits) and two inoculated, mist-irrigated field nurseries in Kansas (Floral + FHB traits). Traits included heading date (HD), plant height (PH), anther extrusion (AE; timing and 0–9 intensity scores), FHB composite indices (7, 14, and 21 days post-anthesis), Fusarium-damaged kernels (FDK), and deoxynivalenol (DON). Genotypes were generated by skim sequencing followed by pedigree-informed hierarchical phasing and imputation, producing a genome-wide SNP dataset (≈75k markers). Within each environment, linear mixed models were used to estimate genotype BLUEs and variance components for heritability, correlations, and principal component analysis; genomic kinship summarized population structure. Bayesian whole-genome regression was applied for GWAS and genomic prediction. Developmental timing and AE traits showed consistent genetic control, with moderate heritability for HD (h^2 = 0.50–0.58) and AE timing (h^2 = 0.43–0.69) and high heritability for AE intensity scores (h^2 = 0.67–0.78). FHB traits were more environment dependent but retained meaningful genetic variation (FHB1–FHB3 h^2 = 0.52–0.81; FDK h^2 = 0.46–0.51; DON h^2 = 0.45–0.77). Multivariate analyses separated phenology/timing, AE-intensity, and disease/toxin domains, with AE intensity generally inversely aligned with FHB outcomes. Together, these results demonstrate that *Ae. tauschii*-

derived diversity captures heritable variation for AE and FHB components and provides a foundation for D-genome-informed selection targeting both hybrid breeding efficiency and reduced FHB risk.

Introduction

Floral and phenological traits such as anther extrusion (AE), heading date (HD), and flowering time (FT) influence pollen dispersal, floral opening dynamics, and exposure of susceptible floral tissues to infection, and have been repeatedly associated with passive or field components of Fusarium head blight (FHB) resistance in wheat (*Triticum aestivum* L.) (Xu et al., 2019; Buerstmayr et al., 2020). FHB, primarily caused by *Fusarium graminearum*, is among the most economically important diseases of wheat worldwide because it reduces grain yield and test weight and contaminates grain with mycotoxins such as deoxynivalenol (DON), impacting food and feed safety (Buerstmayr et al., 2020). A consistent observation across mapping populations and diversity panels is that greater AE is often linked to reduced initial infection (Type I resistance), presumably by reducing the retention of senescing anthers that can serve as nutrient sources for pathogen establishment and by altering the microenvironment around florets (Xu et al., 2019; Miedaner et al., 2024).

Beyond disease resistance, AE is also a key male trait for hybrid wheat seed production. High and consistent anther extrusion and pollen shed increase outcrossing ability and seed set efficiency, making AE a primary target trait in hybrid breeding pipelines (Boeven et al., 2016; El Hanafi et al., 2022; Garst et al., 2023). However, AE and related floral traits are typically quantitative and strongly influenced by environment, requiring careful phenotyping across contrasting conditions to estimate genetic parameters and identify stable genotypes (Adhikari et al., 2020; Muqaddasi et al., 2017a).

The D genome wild progenitor *Aegilops tauschii* (syn. *Triticum tauschii*) is a major source of novel allelic diversity for bread wheat improvement. Introgression of *Ae. tauschii* segments into elite winter wheat has expanded haplotype diversity and can contribute to agronomic and adaptive traits (Nyine et al., 2021). Because many floral traits and disease responses are polygenic and may be constrained by limited diversity in elite germplasm (especially in the D subgenome), *Ae. tauschii* derivative lines may offer a strategic resource to uncover genetic variation affecting AE, flowering, and FHB-related traits (Dubcovsky, J., & Dvorak, J., 2007; Akhunov, E. D., et al., 2010; Gaurav, K., et al., 2022; Cavalet-Giorsa, E., et al., 2024).

Here, we evaluate *Ae. tauschii* derived winter wheat lines to quantify phenotypic variation and broad-sense heritability for key floral, and developmental traits, as well as FHB infection and toxin accumulation, characterize their relationships, and provide a foundation for downstream genetic dissection and selection strategies relevant to both FHB resistance and hybrid wheat breeding.

Materials and Methods

A panel of 101 *Ae. tauschii* derived wheat lines, including 30 *Ae. tauschii* accessions and representing 34 pedigree families, was evaluated across greenhouse and field experiments (Table 1). All introgression lines were obtained from the Wheat Genetics Resource Center (WGRC; Kansas State University) and were developed by one backcross to elite hard winter wheat recurrent parents, followed by inbreeding/line development within each family. The panel was dominated by recurrent winter wheat backgrounds – Danby, Everest, and KanMark used in several families. Zenda was included as check along with the recurrent-parent set. Danby was released in 2005 (Kansas Agricultural Experiment Station, 2007), Everest in 2009 (Kansas State

University Research Foundation, 2010), KanMark in 2014 (Kansas State University Research Foundation, 2015), and Zenda in 2016 (Kansas State University Research Foundation, 2017). Across the checks, plant height ranged from 73 to 93 cm. For height-related alleles, Danby and KanMark carries Rht1-B1b / Rht2-D1a, whereas Everest and Zenda carried Rht1-B1a / Rht2-D1b. For Fusarium head blight, Danby and KanMark are classified as susceptible, whereas Everest and Zenda were classified as moderately resistant–intermediate, and none of the recurrent parents carried the Fhb1 resistance QTL. Heading date ranged from 105 to 111 days, and anther extrusion score ranges from 2 to 8. For photoperiod response, Danby and Zenda carried Ppd-D1b (photoperiod-sensitive), whereas Everest and KanMark carried Ppd-D1a (photoperiod-insensitive). The recurrent parents were included as checks in addition to Zenda in both greenhouse and field trials to provide within-environment benchmarks for phenology, plant architecture, floral traits, and disease response.

Because the population consisted of backcross-derived families sharing common recurrent parents, entries were not independent and were expected to show family-level relatedness and subgrouping driven largely by the recurrent parent background and shared introgression segments. This pedigree structure was considered in downstream analyses through the use of mixed models for phenotypic adjustment and, for association analyses, marker-derived relatedness (kinship) to account for population structure and shared ancestry.

Table 3.1. Pedigree families and number of genotypes per family (panel composition).

Pedigrees	# Genotypes
Danby/TA 10106//Danby	7
Danby/TA 2401//Danby	7
Danby/ TA 10210//KanMark	5
Danby/TA 2508//KanMark	5

Everest/TA 1680//Everest	5
Danby/ TA 2458//Danby	4
Danby/ TA 2461//Danby	4
Danby/ TA 2545//KanMark	4
Danby/TA 2374//Danby	4
Everest/TA 10210//Everest	4
Everest/TA 2395//Everest	4
TA 10141/Danby//Danby	4
Danby/ TA 1645//Danby	3
Danby/ TA 2388//Danby	3
Danby/ TA 2536//Danby	3
Danby/ TA 2538//KanMark	3
Danby/TA 1707//Danby	3
Everest/TA 10113//Everest	3
Everest/TA 1697//Everest	3
Danby/TA 1585//Danby	2
Danby/TA 1605//KanMark	2
Danby/TA 1694//KanMark	2
Danby/TA 2374//KanMark	2
Danby/TA 2468//KanMark	2
Everest/TA 2482//Everest	2
Everest/TA 2512//Everest	2
TA 2458/Danby//Danby	2
Danby/ TA 10330//Danby	1
Danby/ TA 10918//Danby	1
Danby/ TA 1662//Danby	1
Danby/TA 2395//KanMark	1
Danby/TA 2556//Danby	1
Everest/TA 10197//Everest	1
Everest/TA 2433//Everest	1

Greenhouse Experiment

A greenhouse study was conducted at the Kansas State University (KSU) Throckmorton Greenhouses using a randomized complete block design (RCBD) in 2023 and 2024. In 2023 and 2024 greenhouse experiments, we planted the genotypes in a tray and vernalized them for six weeks in a cold room with temperature near 6 °C. After the vernalization period, each genotype transplanted into pots, with four pots per genotype, and five plants per pot (20 plants per genotype), including recurrent parents as checks and Zenda. Greenhouse conditions were maintained near optimal for wheat development, with day/night temperatures of approximately 24/18 °C and a 16-h photoperiod; natural light was supplemented as needed to maintain consistent daily light availability. Standard greenhouse management practices (irrigation and fertilization) were applied to minimize water and nutrient limitations.

Greenhouse Trait Definitions and Scoring

For greenhouse-recorded traits, we use the prefix GH. Heading date (HD) was recorded as the number of days after the vernalization period until 50% of spikes in a pot had emerged from the boot. Plant height (PH) was measured at maturity from the soil surface to the tip of the spike. For floral traits, we recorded: the first date anther extrusion was observed (AE1_D), the visual score at first anther extrusion (AE1_S), the date of complete anther extrusion (AE2_D), and the visual score at complete anther extrusion (AE2_S). Anther extrusion score (AE_S) was visually rated on a 0–9 scale, where 0 indicates no visible extrusion and 9 indicates near-complete extrusion across spikes (Skinnes et al., 2010; Xu et al., 2019) (Figure 1). Trait definitions followed established hybrid wheat floral phenotyping protocols (Boeven et al., 2016; Langer et al., 2014) and prior anther extrusion–FHB studies (Xu et al., 2019).

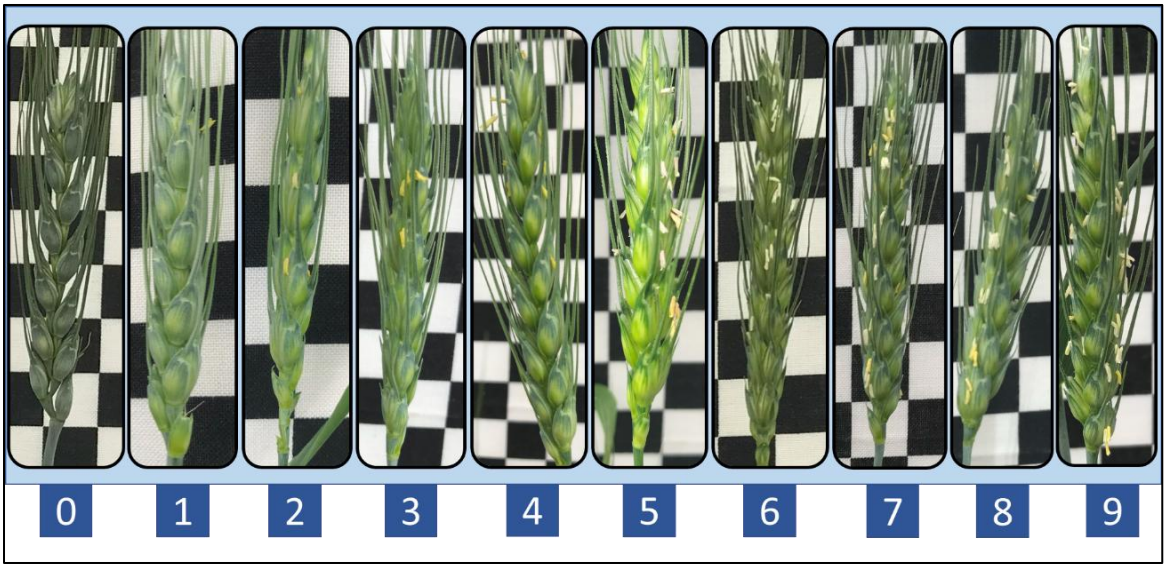

Figure 3.1. Figure 1 Anther Extrusion Scoring note. AE was rated on a 0-9 scale (0 = no extrusion; 9 = near-complete extrusion).

Field Nursery Experiment (FHB Epidemic + Agronomic/Floral Traits)

Field trials were conducted during the 2023–2024 and 2024–2025 seasons at the KSU Rocky Ford Research Station (Manhattan, KS; 39.23°N, 96.58°W). Trials were planted in October and harvested in July of the following year. Both seasons used a partially replicated (p-rep) design (Hoefler et al., 2020). All 101 lines were grown in unreplicated plots per environment, with a subset (~10%) designated as repeated check cultivars (including recurrent parents) randomly distributed to support spatial adjustment.

Plots consisted of a single-row nursery configuration (1.0 m × 0.35 m) planted with ~2g of seed per plot. Standard agronomic management was applied. To ensure uniform Fusarium head blight (FHB) pressure, artificial epidemics were promoted by distributing *F. graminearum* - colonized maize kernels throughout rows before flowering and using nocturnal mist irrigation to maintain prolonged humidity and canopy wetness during anthesis - conditions favorable for FHB infection and symptom development (Cowger et al., 2013; Šarčević et al., 2023; Xu, Y., et al.,

2023). Overhead sprinklers were run overnight (approximately 18:00–07:00) in short cycles of 3 minutes per hour daily between flowering and milky ripe to sustain head wetness through flowering, consistent with standard FHB nursery practice (Xu, Y., et al., 2023).

Field Trait Definitions and Scoring

For field-nursery-recorded traits, we use the prefix FN. Heading date (HD) was recorded when ≥50% spike emergence was observed in a plot. Plant height (PH) was measured near maturity using a measuring stick. Floral traits (AE1_D, AE1_S, AE2_D, AE2_S, AE_S) were recorded using the same definitions and the same 0–9 scoring scale described for the greenhouse (Skinnes et al., 2010; Xu et al., 2019) (Figure 1).

FHB severity was visually assessed at 3 time points post-anthesis beginning 7 days after anthesis and continuing after 7 days periods. Disease was summarized as the FHB index:

$$\mathbf{FBHindex} = \frac{\mathbf{Incidence}(\%) \times \mathbf{Severity}(\%)}{\mathbf{100}}$$

following standard disease quantification practices used in breeding and genetic studies (Buerstmayr et al., 2020). For Fusarium damaged kernels (FDK), ten randomly sampled spikes per plot were harvested and threshed, the grains were used for visual FDK scoring on a 1-9 scale (low to high damage) (Gaikpa, D. S., et al., 2020). After scoring FDK, grain samples were ground with a coffee grinder, and 10 grams per samples were sent to University of Minnesota DON Testing Lab. DON concentration was quantified using a gas chromatography-mass spectrometry from harvested grain samples (ppm), following standard mycotoxin quantification procedures from USWBSI-funded lab (Xu, Y., et al., 2023).

Table 2 summarizes all phenotypic traits evaluated in this study, including phenology and floral traits measured in both the greenhouse (GH) and field nursery (FN). Check marks indicate where each trait was recorded, and the final column reports the scale or units used for scoring or quantification.

Table 3.2. Trait acronyms and where they were measured (greenhouse vs. field)

Acronym	Trait (definition)	Greenhouse (GH)	Field Nursery (FN)	Scale / units
HD	Heading date (50% spike emergence)	✓	✓	Days After vernalization
PH	Plant height (soil/pot surface to spike tip)	✓	✓	Centimeters (cm)
AE1_D	Date of first observed anther extrusion	✓	✓	Days After vernalization
AE1_S	Anther extrusion score at first extrusion	✓	✓	Visual scale 0–9; 0 = no anther, 9 = full anther extrusion
AE2_D	Date of complete anther extrusion	✓	✓	Days After vernalization
AE2_S	Anther extrusion score at complete extrusion	✓	✓	Visual scale 0–9; 0 = no anther, 9 = full anther extrusion
FHB1	Composite FHB index = (Incidence × Severity)/100 (7 days after anthesis)	—	✓	Visual index (0–100) 0 = no FHB, 100 = High FHB
FHB3	Composite FHB index = (Incidence × Severity)/100 (14 days after anthesis)	—	✓	Visual index (0–100) 0 = no FHB, 100 = High FHB
FHB3	Composite FHB index = (Incidence × Severity)/100 (21 days after anthesis)	—	✓	Visual index (0–100) 0 = no FHB, 100 = High FHB
FDK	Fusarium-damaged kernels (visual rating)	—	✓	Visual scale 1–9; 1 = no damage, 9 = full damage
DON	Deoxynivalenol concentration in grain	—	✓	Part Per Million (ppm)

Genotyping

Skim sequencing libraries were prepared following the WGRC DNA library preparation workflow, which uses a single-tube tagmentation reaction to fragment genomic DNA and add adapters for multiplexed single-read sequencing. Genomic DNA was standardized to 1 ng μL^{-1} prior to library preparation. Libraries were generated by tagmenting 1 µL of DNA (1 ng μL^{-1}) in a 5 µL reaction (including TDE1 enzyme and TD buffer) and incubating at 55 °C for 15 min,

followed by PCR amplification using NextSeq i7/i5 indices (2.5 µM) and NEB 2× Taq Master Mix (25 µL final volume). PCR cycling included an initial extension and denaturation, followed by 18 cycles of amplification (95 °C denaturation, 55 °C annealing, 72 °C extension). Amplicons were quantified using Quant-iT PicoGreen (Synergy H1 microplate reader), normalized (15 µL at 6 ng μL^{-1}), and pooled by combining 6 µL (36 ng) of normalized PCR product per sample. Libraries were then cleaned (0.8×), size-selected to 600–800 bp (Pippin Prep), cleaned again (0.7×), quantified using the Qubit dsDNA HS assay, and fragment size was verified with an Agilent Bioanalyzer or 4200 TapeStation (Adhikari et al., 2022). Sequencing reads were processed to generate genotype likelihoods/calls.

For variant calling, raw sequencing reads were trimmed and quality filtered using fastp (Chen et al., 2018) and aligned to the Chinese Spring wheat reference genome IWGSC RefSeq v2. (IWGSC, 2018; Zhu et al., 2021) using BWA-MEM (Li, 2013). BAM files were processed by marking and removal of duplicates with Picard (Broad Institute, 2019). SNPs were called using BCFtools (Danecek et al., 2021), and a reference SNP set was constructed from all parental lines used for developing the introgression population, with SNPs from these parents treated as high-quality variants. These positions were applied to skim-sequenced progeny to recover polymorphic sites, even when they appeared monomorphic in the introgression population because of the high missingness inherent to skim sequencing. Variant calling was performed using a panel of 50 diverse *Aegilops tauschii* accessions. Because the A and B subgenomes were contributed primarily by a single homozygous recurrent parent, polymorphism in A/B was limited, whereas the diverse *Ae. tauschii*–derived D genome resulted in a much higher SNP density. The high SNP density in the D genome increased power to detect rare alleles segregating within the introgression lines. The resulting VCF file was used for downstream processing. To

maximize imputation accuracy, genotype phasing and imputation were performed hierarchically using pedigree structure and parental information. First, all parental lines were phased to obtain high-confidence, pedigree-consistent haplotypes. These phased parental haplotypes were then used as reference panels for downstream imputation. Phasing was performed independently of progeny to ensure stable parental haplotype structure (Adhikari et al., 2022).

Imputation was conducted family-by-family, conditioning progeny genotypes on the corresponding phased parental haplotypes to constrain inference to biologically plausible haplotypes and recombination patterns within each biparental family. After family-level imputation, all families were merged into a single dataset and a second round of population-level imputation was performed, leveraging shared haplotypes across families to further refine genotype calls and harmonize marker states (Sargolzaei et al., 2014; Browning & Browning, 2016; Whalen et al., 2020). The resulting imputed dataset (~75k SNPs) was used for downstream analyses. SNP density across the wheat genome was summarized to assess marker distribution along chromosomes (Figure 2). Marker physical positions (bp) were extracted from the SNP map and chromosomes were standardized to the 21 bread wheat chromosomes (1A–7D). For each chromosome, SNPs were binned into consecutive, non-overlapping 1-Mb windows and the number of markers per window was counted. Windows with no markers were retained and displayed as zero to preserve chromosomal continuity. Counts were visualized as a heatmap (one row per chromosome; x-axis in Mb), where color intensity represents the number of SNPs per window using discrete classes (0, 1, 5, 10, and ≥20 SNPs per window) (Figure 2). For each chromosome, the total number of SNPs and its percentage of the genome-wide total (N = 75,861) are shown adjacent to the chromosome label (Figure 2).

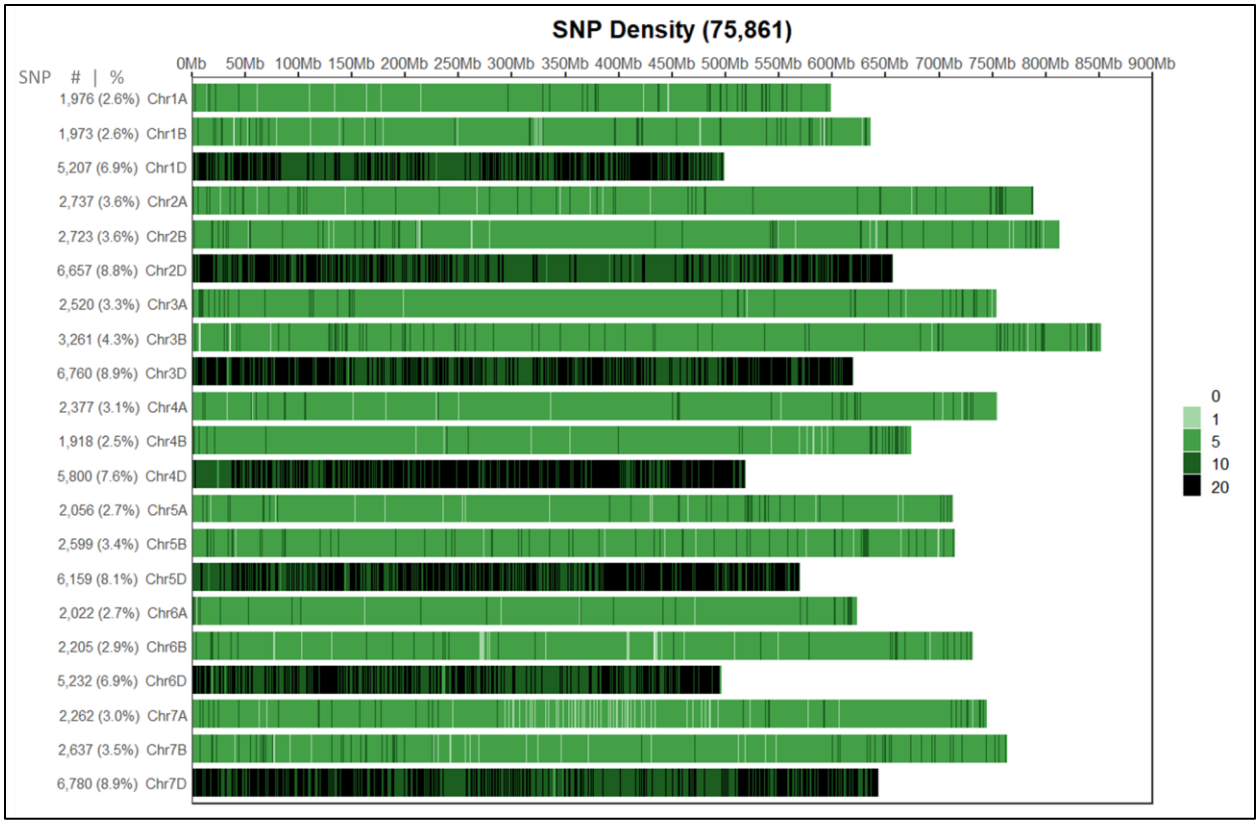

Figure 3.2. Genome-wide SNP density across the 21 wheat chromosomes. Each horizontal track represents a chromosome (Chr1A–Chr7D). SNP density is summarized in consecutive physical windows along each chromosome (x-axis in Mb). Color intensity indicates the number of SNPs per window (white = 0; increasing green = low to high density; darkest = highest class shown in the legend). Chromosome track lengths correspond to the maximum mapped physical position in the dataset.

Statistical Analysis

BLUE estimation and variance components

For each environment, phenotypes were analyzed using linear mixed models to obtain genotype adjusted means (best linear unbiased estimates, BLUEs) and to partition variance into genetic and non-genetic components. Spatial heterogeneity was accounted for using row and column effects. For a given environment, the mixed model can be written as:

$$\mathbf{Y = \mu + g_i + r_j + row_k + col_l + e}$$

where y is the observed phenotype, μ is the overall mean, g_i is the effect of genotype i, r_j is the replicate effect when available, row_k and col_l represent spatial effects, and e is the residual. For

BLUE estimation, genotype was modeled as a fixed effect and nuisance terms (replicate and spatial terms) were modeled as random effects. Models were fit by REML using lmer function in lme4 package in R software (Bates et al., 2015).

To evaluate relationships among traits and the consistency of trait expression across trial types (Field Nursery vs Greenhouse), genotypic adjusted means were obtained for each trait within each experiment using the mixed-model framework described above. Pairwise correlations among traits were then calculated across genotypes within each experiment. Pearson's correlation coefficient (r) was used (Pearson, K., 1895). Correlation significance was assessed using two-sided tests, and when multiple correlations were evaluated simultaneously, P-values were adjusted using a false discovery rate procedure (Benjamini, Y., & Hochberg, Y,1995).

To summarize multivariate relationships among phenology, floral, and FHB-related traits across the four site-year environments, we conducted a principal component analysis (PCA) on a genotype-by-trait matrix constructed from genotype-level trait estimates, after standardizing each trait to mean 0 and variance 1 (z-scores) to place measurements on a common scale (Jolliffe, 2002; Abdi & Williams, 2010). PCA was implemented in R using the prcomp function (singular value decomposition of the centered/scaled matrix), and the first two PCs were visualized using a biplot in which trait vectors represent loadings: vector length reflects the magnitude of contribution to the PCs, and angles between vectors approximate correlations (small angles = positive association; opposite directions = negative association) (Gabriel, 1971; Lever et al., 2017; R Core Team, 2025). Traits were color-coded by functional group (phenology, floral, disease) to facilitate interpretation of how trait domains align with the major axes of variation.

Genomic Kinship Analysis

Genomic relationships among lines were quantified using high-density SNP genotype data by coding marker dosages as 0, 1, and 2, corresponding to homozygous reference, heterozygous, and homozygous alternate alleles, respectively (VanRaden, 2008). Pairwise genetic dissimilarities were computed as squared Euclidean distances across all markers, and these distances were transformed into a kernel-based genomic kinship (similarity) matrix in genotype space (Endelman, 2011; Gianola & van Kaam, 2008). The kernel bandwidth was defined as the inverse of the mean squared distance among individuals to provide a data-informed scaling of genomic similarity consistent with RKHS/gaussian-kernel approaches (Gianola & van Kaam, 2008; Endelman, 2011). The resulting kinship matrix was visualized as a heatmap with hierarchical clustering, and individuals were grouped/annotated by recurrent parent background to facilitate interpretation of genetic structure. Heatmaps were generated in R using the ComplexHeatmap package (Gu et al., 2016).

. Narrow-sense heritability was computed using the bWGR function zsemf (Xavier et al., 2020; VanRaden, 2008) as:

$$\mathbf{h^2} = \frac{\mathbf{V_g}}{\mathbf{V_g} + \mathbf{V_e}}$$

where V_g is the additive genotype variance, V_e is the residual variance fitting.

Significant markers were used to nominate candidate genes by querying annotated loci within a ±1 Mb window surrounding each marker position on the IWGSC RefSeq v2.1 bread wheat reference genome (≈14.5 Gb) for Chinese Spring (*Triticum aestivum* L.) (Zhu, Tingting, et al. 2021).

Results

Across experiments, phenology and flowering-related traits showed consistently strong additive genetic signal relative to residual variation (Table 3). Heading date displayed moderate phenotypic variance (Vp = 8.13-14.93) and stable narrow-sense heritability across environments (h^2 = 0.50-0.58), with comparable contributions of additive genetic variance (Vg = 4.09-8.65) and residual variance (Ve = 4.03-6.49) in both field nursery (FN) and greenhouse (GH) trials. Anther extrusion (AE) timing traits followed a similar pattern, with moderate-to-high heritability overall (AE1_D: h^2 = 0.48-0.68; AE2_D: h^2 = 0.43-0.69). Notably, heritability for AE timing was generally higher in GH (AE1_D h^2 = 0.68; AE2_D h^2 = 0.69) than in FN (AE1_D h^2 = 0.48–0.59; AE2_D h^2 = 0.43-0.56), consistent with reduced environmental heterogeneity under controlled conditions. In contrast, ordinal AE score traits were strongly repeatable and predominantly genetically driven across both settings, with consistently high narrow-sense heritability for AE1_S (h^2 = 0.69-0.72) and AE2_S (h^2 = 0.67-0.78), reflecting large Vg relative to Ve across experiments.

FHB traits expressed substantial phenotypic variance and a stronger environment-dependent component, but still retained moderate-to-high additive genetic control depending on assessment timing (Table 3). The earliest composite index (FHB1) was moderately heritable (h^2 = 0.52-0.72), whereas later indices showed higher and more stable heritability, particularly in 24_FN (FHB2 h^2 = 0.79; FHB3 h^2 = 0.81) compared with 25_FN (FHB2 h^2 = 0.60; FHB3 h^2 = 0.65). Kernel damage (FDK) exhibited moderate heritability (h^2 = 0.46-0.51). Deoxynivalenol (DON) showed high heritability in 24_FN (h^2 = 0.77) but a reduced signal in 25_FN (h^2 = 0.45), indicating a larger proportional contribution of residual variance in that environment-year. PH was moderately heritable overall (h^2 = 0.45-0.69), with higher heritability in GH (0.69) than FN

(0.45-0.49), again suggesting that controlled conditions improved the ratio of genetic to non-genetic variance. Collectively, Table 3 indicates that developmental timing traits and AE scores are under consistent additive genetic control across environments, while FHB-related traits show greater sensitivity to environment and sampling time, yet maintain meaningful heritable variation for selection in favorable testing conditions.

Table 3.3. Variance components, and narrow-sense heritability by experiment and trait across Field Nursery and Greenhouse Experiments.

Environment	Trait	n_GE	V_p	V_g	V_e	h^2
24_FN	HD	105	10.45	5.75	4.70	0.55
24_GH	HD	105	14.93	8.65	6.28	0.58
25_FN	HD	105	13.81	7.33	6.49	0.53
25_GH	HD	105	8.13	4.09	4.03	0.50
24_FN	AE1_D	104	9.45	4.54	4.91	0.48
24_GH	AE1_D	105	19.81	13.44	6.37	0.68
25_FN	AE1_D	105	14.33	8.47	5.85	0.59
25_GH	AE1_D	105	10.18	5.09	5.09	0.50
24_FN	AE1_S	105	2.50	1.79	0.71	0.72
24_GH	AE1_S	105	1.05	0.72	0.33	0.69
25_FN	AE1_S	105	2.21	1.54	0.68	0.69
25_GH	AE1_S	105	1.67	1.17	0.50	0.70
24_FN	AE2_D	104	11.71	5.03	6.68	0.43
24_GH	AE2_D	105	21.97	15.15	6.82	0.69
25_FN	AE2_D	105	13.28	7.41	5.88	0.56
25_GH	AE2_D	105	9.42	4.52	4.91	0.48
24_FN	AE2_S	105	3.13	2.14	0.99	0.68
24_GH	AE2_S	105	3.36	2.25	1.12	0.67
25_FN	AE2_S	105	4.73	3.67	1.06	0.78
25_GH	AE2_S	105	3.78	2.84	0.94	0.75
24_FN	FHB1	105	55.44	40.09	15.35	0.72
25_FN	FHB1	105	84.91	44.23	40.68	0.52
24_FN	FHB2	105	144.74	114.98	29.76	0.79
25_FN	FHB2	105	143.55	85.45	58.10	0.60
24_FN	FHB3	105	584.70	472.18	112.52	0.81
25_FN	FHB3	105	170.75	111.46	59.30	0.65
24_FN	PH	105	80.45	36.50	43.95	0.45

24_GH	PH	105	81.24	56.00	25.24	0.69
25_FN	PH	105	62.66	30.93	31.73	0.49
24_FN	FDK	105	2.41	1.24	1.17	0.51
25_FN	FDK	105	1.45	0.66	0.79	0.46
24_FN	DON	105	0.68	0.52	0.16	0.77
25_FN	DON	102	0.38	0.17	0.21	0.45

FN = Field Nursery; GH = Greenhouse; HD = Heading date; AE1_D = date of first observed anther extrusion; AE1_S = anther extrusion score at first extrusion; AE2_D = date of complete anther extrusion; AE2_S = anther extrusion score at complete extrusion; FHB1 = Fusarium head blight composite index 7 days after anthesis (Incidence × Severity / 100); FHB2 = Fusarium head blight composite index 14 days after anthesis (Incidence × Severity / 100); FHB3 = Fusarium head blight composite index 21 days after anthesis (Incidence × Severity / 100); PH = plant height; FDK = Fusarium-damaged kernels; DON = deoxynivalenol concentration; n_GE = number of genotypes evaluated; Vp = phenotypic variance; Vg = additive genetic variance; Ve = residual variance; h^2 = narrow-sense heritability.

In Figure 3, the distributions of genotype-level values for each trait highlight substantial phenotypic diversity across the panel, while also revealing clear differences in the degree and shape of variation among trait classes. HD and AE_D (AE1_D, AE2_D) show largely unimodal, approximately symmetric distributions with comparatively narrow spread, indicating that most lines cluster around the population center with fewer extreme early/late genotypes. In contrast, ordinal score traits related to anther extrusion (AE1_S, AE2_S) display broader dispersion and noticeable skewness, with a larger fraction of lines concentrated at lower scores and an extended tail toward higher values. This pattern is consistent with traits scored on bounded scales where many genotypes express limited extrusion under some conditions, while a subset shows strong expression. Disease and grain quality traits associated with FHB exhibit the greatest heterogeneity across genotypes. The composite indices (FHB1-FHB3) show wider distributions and stronger right-skew (particularly at later assessments), reflecting that while many entries express low-to-moderate symptom levels, a smaller subset accumulates high disease values. Similarly, DON is markedly right-skewed, indicating that extreme toxin accumulation is concentrated in relatively few genotypes even when most lines remain near the lower-to-

intermediate range. FDK shows intermediate dispersion relative to the FHB indices, consistent with partial overlap between visual kernel damage and toxin/disease expression. Plant height (PH) exhibits a comparatively smooth, near-normal distribution, suggesting a quantitative architecture with broad but continuous variation across the panel. Across panels, the check cultivars (colored vertical lines) span a meaningful portion of each trait's range, providing fixed benchmarks and confirming that the population captures both favorable and unfavorable phenotypic space relative to commonly used standards.

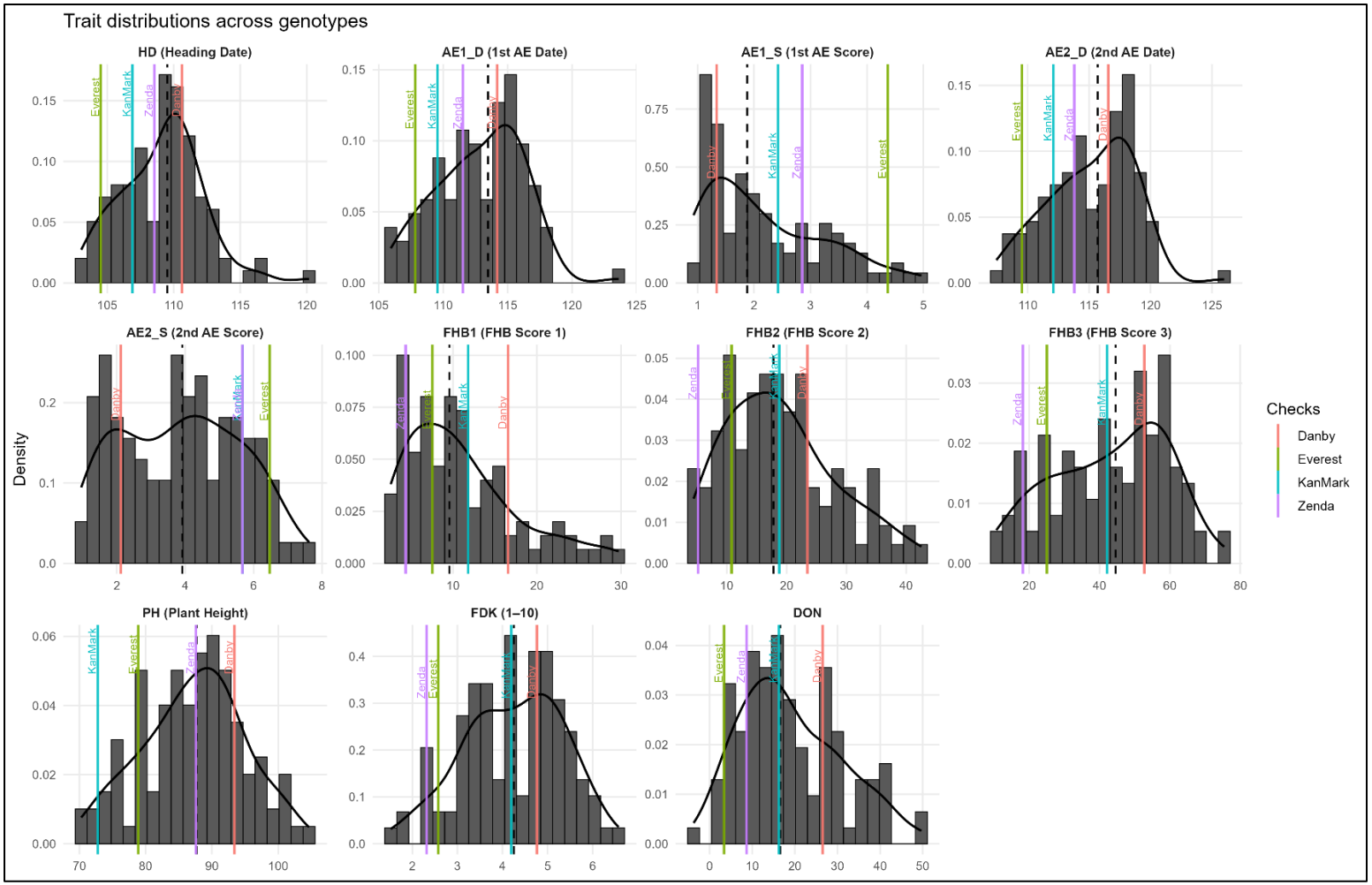

Figure 3.3. Distribution of traits across the genotype panel. Panels show genotype-level trait distributions and recurrent parents. The dashed vertical line indicates the trait median across genotypes. Trait acronyms: HD = heading date; AE1_D, AE2_D, AE3_D = date of the 1st, 2nd, and 3rd anther extrusion observation; AE1_S, AE2_S, AE3_S = anther extrusion score at the 1st, 2nd, and 3rd observation; FHB1, FHB2, FHB3 = Fusarium head blight severity scores at scoring events 1–3; PH = plant height; FDK = Fusarium-damaged kernels rating (1–10); DON = deoxynivalenol (vomitoxin) concentration. Axes are trait-specific (free scales) to preserve the natural units and ranges of each trait.

Pearson correlation heatmap (Figure 4) resolves the trait network into visually distinct color “quadrants/blocks” that summarize both the direction and strength of associations. In this figure, blue quadrants/cells represent positive correlations (traits increase together), whereas red quadrants/cells represent negative correlations (trade-offs; one trait increases as the other decreases). In addition, the heatmap is partitioned by the clustering into major modules highlighted by the large colored rectangles: the yellow quadrant delineates negative correlations patterns while the purple quadrant delineates positive correlations patterns. Within traits, though in different environments, yellow quadrants show consistence in scoring. Between traits, some preformed negative correlation blocks showed in purple quadrants.

Three major correlation modules are apparent as dense within-module blocks. The yellow phenology/timing module (HD and AE date traits) shows the strongest and most consistent positive correlations across environments, indicating a shared developmental axis. Within each environment, AE timing measures are highly redundant, and HD closely tracks the same timing structure, producing a prominent positive (blue) block for the phenology cluster (yellow quadrant) (Figure 4). The purple AE score module forms a second internally coherent block: AE score traits are strongly correlated with each other across timepoints, indicating that extrusion intensity behaves as a coordinated trait group that is partially distinct from timing (Figure 4).

A key feature is that the cross-module region between the yellow (timing) and purple (AE score) quadrants shifts from near-neutral to strongly negative depending on environment, showing that the coupling between when extrusion occurs and how much extrusion is expressed is environment-dependent. For example, in FN during 2024, the first AE score was essentially unrelated to the first AE date (24_FN_AE1_D vs 24_FN_AE1_S: $r = 0.01$; near-neutral), whereas in GH the same comparison was strongly negative (24_GH_AE1_D vs 24_GH_AE1_S:

r = -0.73; negative), and FN in 2025 was intermediate (25_FN_AE1_D vs 25_FN_AE1_S: r = −0.44) (Figure 4). This indicates that developmental timing remains tightly coordinated, but the relationship between timing and extrusion intensity can decouple or invert across settings.

Plant height (PH) occupies an intermediate position: PH aligned moderately with phenology (e.g., 24_GH_PH vs 24_GH_HD: r = 0.65; 25_FN_PH vs 25_FN_HD: r = 0.41; positive), but its association with AE scores changed sign across environments (e.g., 24_FN_PH vs 24_FN_AE1_S: r = 0.22 vs 24_GH_PH vs 24_GH_AE1_S: r = −0.40), indicating PH is not simply a proxy for heading or extrusion intensity (Figure 4).

For disease/quality module (FHB indices, DON, FDK) forms another internally positive block, with year-dependent differences in strength. FHB scores were tightly correlated with each other (24_FN_FHB1 vs 24_FN_FHB2: r = 0.90; 25_FN_FHB1 vs 25_FN_FHB2: r = 0.80) and linked to DON/FDK more strongly in 2024 than 2025 (24_FN_DON vs 24_FN_FDK: r = 0.51 vs 25_FN_DON vs 25_FN_FDK: r = 0.22; 24_FN_DON vs 24_FN_FHB3: r = 0.52 vs 25_FN_DON vs 25_FN_FHB1: r = 0.31) (Figure 4). Importantly, the off-diagonal region comparing AE scores to disease shows a stronger negative pattern in 2024 (e.g., 24_FN_AE2_S vs 24_FN_FHB1: r = −0.56; 24_FN_AE2_S vs 24_FN_DON: r = −0.51) that weakens in 2025 (25_FN_AE2_S vs 25_FN_FHB1: r = −0.20; 25_FN_AE2_S vs 25_FN_DON: r = −0.22), indicating that the AE score–disease trade-off was more pronounced under the 2024 field conditions (Figure 4).

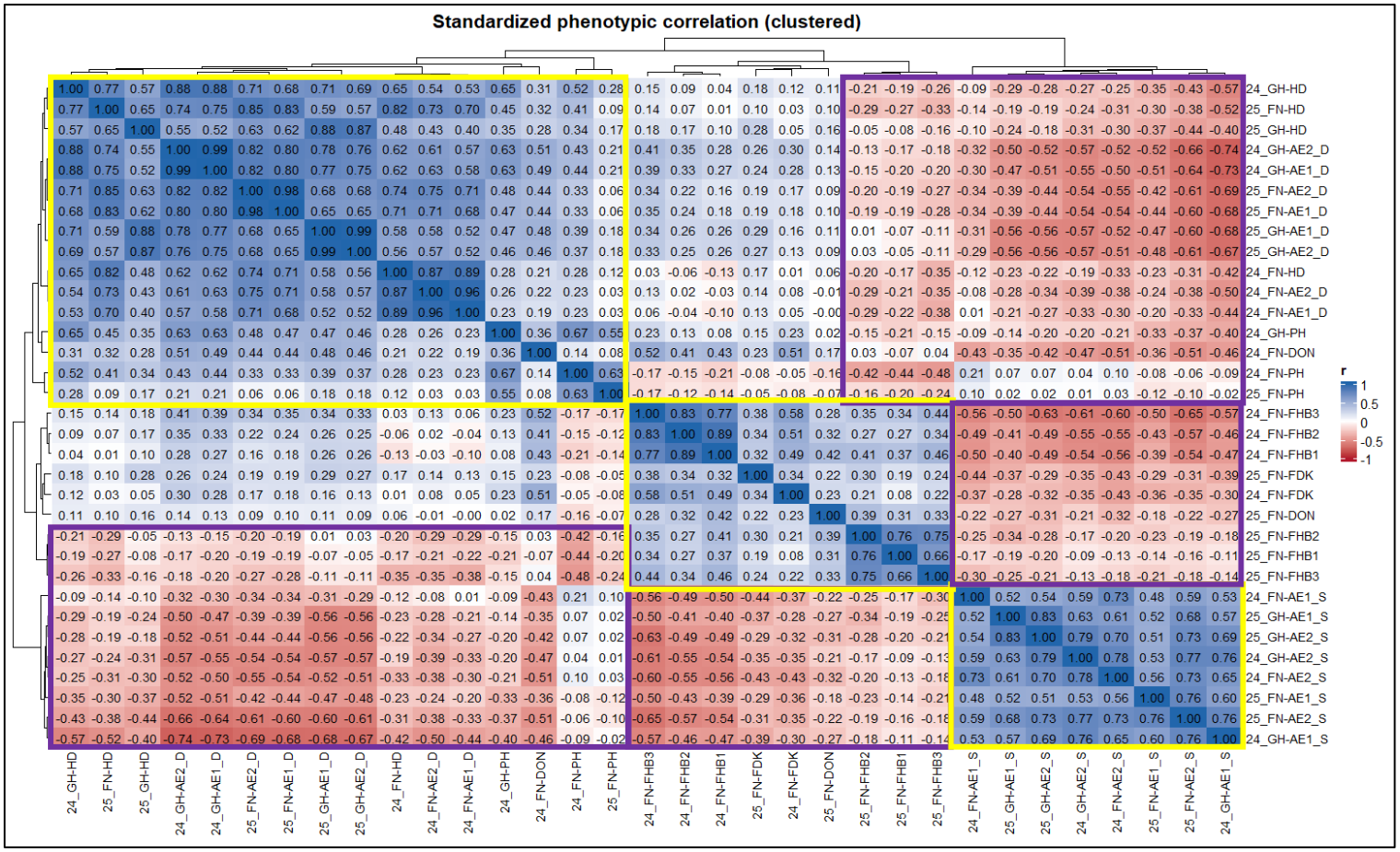

Figure 3.4. Multi-Trait phenotypic correlations across environments. Heatmap values are Pearson correlation coefficients (r) computed using standardized trait values within each year × site dataset; blue = positive and red = negative correlations. Dendrograms show hierarchical clustering of traits based on similarity in correlation patterns. Trait labels follow YY_SITE_TRAIT, where YY = 24 (2024) or 25 (2025), and SITE = FN or GH (site codes). Acronyms: HD = heading date; AE = anther extrusion; AE1_D and AE2_D = date of the 1st and 2nd anther extrusion observation, respectively; AE1_S, AE2_S = anther extrusion score at the 1st, and 2nd observation, respectively; FHB1, FHB2, FHB3 = Fusarium head blight severity scores at scoring events 1-3; PH = plant height; FDK = Fusarium-damaged kernels rating (1–10); DON = deoxynivalenol concentration. Yellow quadrants show positive correlations; Purple quadrants shows negative correlations.

Figure 5 summarizes the structure of phenology, floral, and Fusarium-related traits across environments and confirms a strong separation among trait groups in the first two principal components. The biplot shows that PC1 captured the primary axis of variation and was largely defined by opposing loadings of disease traits (FHB1-FHB3, FDK, DON) versus floral extrusion scores (AE1_S, AE2_S). Disease traits loaded strongly in the positive PC1 direction, with the three FHB indices clustering tightly and showing similar vector orientation, consistent with their strong positive intercorrelations. In contrast, AE score vectors projected in the opposite direction along PC1, indicating an overall negative association between extrusion intensity and FHB-related outcomes in the combined standardized dataset.

PC2 further differentiated components of plant development and flowering time from disease severity. Phenology and timing traits - particularly HD and PH, along with AE date traits (AE1_D and AE2_D) - contributed substantially to PC2 and tended to align together, reflecting shared covariation among developmental timing/architecture traits. The relatively small angles between HD and AE date vectors indicate that later heading generally coincided with later anther extrusion timing across environments, while their broader separation from AE score vectors indicates that extrusion intensity captures a partially distinct dimension of floral variation. Notably, DON aligned more closely with PC1 and showed weaker separation along PC2 compared with FHB indices, suggesting that toxin accumulation tracked overall disease pressure more strongly than the phenology/timing axis in this ordination.

Overall, the biplot reinforces three coordinated but distinguishable trait domains: (i) a disease severity/quality domain dominated by FHB indices, FDK, and DON; (ii) a floral extrusion intensity domain (AE scores) that is oriented opposite to disease loadings; and (iii) a developmental timing/plant architecture domain (HD, PH, and AE dates) that contributes more

strongly to the second axis. Together, these patterns indicate that while phenology and AE timing covary as a shared developmental process, extrusion intensity represents a separate floral dimension that is inversely aligned with Fusarium-related outcomes at the multivariate level (Figure 5).

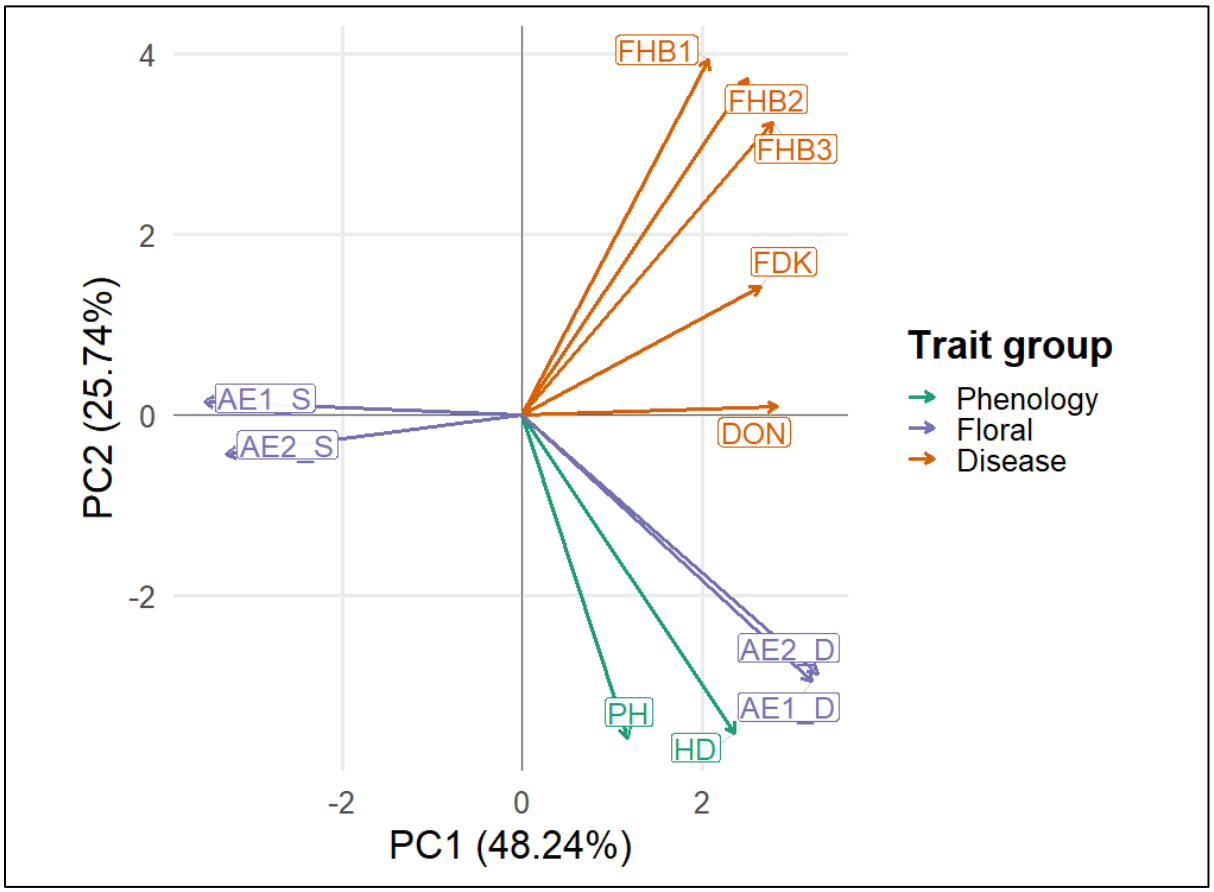

Figure 3.5. Principal component biplot summarizing phenology, floral, and disease traits across environments. Biplot of the first two principal components (PC1 = 39.48% and PC2 = 19.18% of total variation) computed from standardized trait values across the four site-year environments. Vectors (arrows) represent trait loadings; longer arrows indicate stronger contributions to the PCs, and smaller angles between arrows indicate positive correlations (opposite directions indicate negative correlations). Traits are color-coded by group: phenology (green), floral (purple), and disease (orange). Trait acronyms are: HD (heading date), PH (plant height), AE1/AE2 (first/second anther extrusion) with D = date and S = score, FHB1-FHB3 (Fusarium head blight scores across rating times), FDK (Fusarium-damaged kernels), and DON (deoxynivalenol).

The genomic kinship matrix exposed population structure consistent with recurrent parent backgrounds (Figure 6). Clear block patterns along the diagonal indicated high relatedness, particularly for lines derived from the same recurrent parents. In contrast, reduced kinship values were observed between families, reflecting genetic differentiation across recurrent parent groups. The pronounced separation among Danby, KanMark, Everest, and Zenda backgrounds confirms that the breeding design preserved distinct genetic pools. These results demonstrate that structure is captured by genome-wide markers and supports the use of genomic information for downstream analyses.

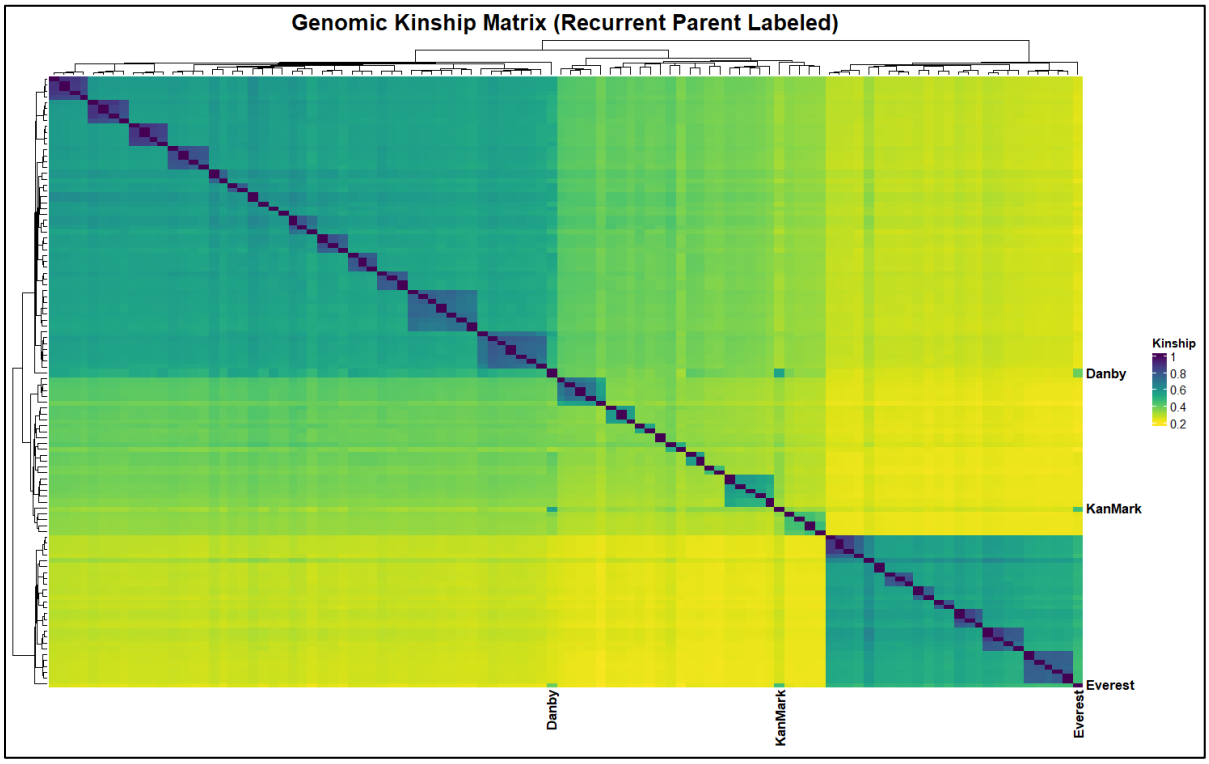

Figure 3.6. Genomic kinship matrix of introgression lines recurrent parents' background. Genomic kinship matrix estimated from genome-wide SNP markers for all evaluated lines. Lines are hierarchically clustered and grouped with recurrent parent labeled (Danby, KanMark, and Everest). The diagonal represents self-relatedness, while off-diagonal blocks illustrate genetic relationships, highlighting clear pedigree structure and reduced relatedness across recurrent parent groups.

Discussion and Conclusion

This study was designed as a screening and prioritization effort to identify phenotypic floral and Fusarium-related traits that are most relevant for both hybrid wheat utility and line breeding, while providing practical direction on how *Ae. tauschii*–derived materials could be deployed as potential parental sources. By evaluating phenology (e.g., heading date), key floral attributes (especially anther extrusion), and FHB-related responses in a D-genome–enriched panel, our goal was to highlight trait relationships that matter for selection under variable environments and to flag germplasm that combines favorable reproductive biology with disease tolerance.

A key feature of this study is the genetic background of the panel; the germplasm is *Aegilops tauschii*–derived, meaning it carries an enriched contribution from the D-genome progenitor of bread wheat. This matters for two reasons. First, the modern bread wheat D genome is widely recognized as having experienced a strong bottleneck relative to the A and B genomes, so "standard" elite panels can be comparatively underpowered for discovering D-genome associations unless D diversity has been deliberately broadened (Akhunov et al., 2010). Second, *Ae. tauschii*–derived panels reintroduce polymorphism and haplotypes that were rare or absent in mainstream hexaploid breeding pools, expanding the adaptive landscape available to breeding for stress responsiveness and resilience (He et al., 2019; Gaurav et al., 2021; Nyine et al., 2021). Therefore, the observation that many of the most significant SNPs detected here fall on chromosomes D is not only biologically plausible -it is also consistent with what you would predict from the panel design, where D-genome haplotypes are more polymorphic and thus more "discoverable" than in many elite-only panels (Akhunov et al., 2010; He et al., 2019).

Heading date (HD) remains one of the most repeatable developmental traits in wheat, often showing high heritability because core variation is anchored in large-effect photoperiod and vernalization pathways (e.g., Ppd and Vrn), with additional tuning from earliness-per-se and other regulatory layers (Grogan et al., 2016; Okada et al., 2019). The D-genome enrichment adds an interpretive angle: D-genome variants can contribute disproportionately to phenological plasticity and the alignment of flowering with local stress windows when *Ae. tauschii* haplotypes are present (He et al., 2019; Gaurav et al., 2021; Nyine et al., 2021). The recurrent parents differ in photoperiod response at the Ppd-D1 locus, with Everest and KanMark reported to carry the photoperiod-insensitive allele Ppd-D1a, whereas Danby is photoperiod sensitive (USDA-ARS., 2004; USDA-ARS., 2018; USDA-ARS; 2022). Consistent with this, loci originating from the recurrent parent were detected in similar genomic regions in our analysis.

The relationship between phenology and Fusarium head blight (FHB) was clearly context dependent, which aligns with a large body of field evidence. In some environments, earlier heading can coincide with peak inoculum and favorable infection weather, producing an apparent penalty for early lines; in other environments, early flowering can create partial "escape" from infection windows, reversing the sign of the association (Clark et al., 2016; Prat et al., 2017). This duality is a major reason phenology alone is not a reliable proxy for physiological resistance, especially when environments differ in epidemic timing, humidity, and temperature during anthesis (Buerstmayr & Buerstmayr, 2015; Miedaner et al., 2017). Importantly, multi-environment QTL work has shown that some loci near major developmental regions can look "FHB-associated" until heading date (and plant height) are accounted for, supporting the view that part of the signal can be indirect and mediated by exposure rather than defense (He et al., 2016; Xu et al., 2019). In practical breeding terms, this underscores the need

to treat HD as a covariate (or to use multi-trait models) when interpreting FHB associations, particularly in panels where D-genome diversity may broaden phenological responses and thus broaden exposure differences among lines (Arruda et al., 2016; Steiner et al., 2019; He et al., 2019). Everest has been reported to carry Fusarium head blight (FHB) resistance quantitative trait loci (QTL) on chromosomes 2D, 4B, and 4D. In our analysis, several loci with allelic origin from the recurrent parent were identified in similar genomic regions (Supplementary Table S1), suggesting that part of the observed variation in FHB-related traits may be associated with native resistance alleles contributed by Everest (Clinesmith, M. A., et al., 2019).

AE emerged as a central trait linking reproductive biology to both hybrid-wheat utility and disease epidemiology. AE is a quantitatively inherited trait with substantial polygenic background, and GWAS across spring and winter wheat panels supports a many-locus architecture with moderate repeatability across environments (Muqaddasi et al., 2017a, 2017b; Okada et al., 2019). A particularly relevant breeding interaction is the coupling between major agronomic loci and floral traits: alleles selected for plant stature and adaptation can shift AE as a correlated response, which matters directly for outcrossing efficiency in hybrid seed production and can also reshape disease exposure at flowering (He et al., 2016; Okada et al., 2019; Longin et al., 2012).

The AE–FHB relationship is most compelling when framed as an infection-biology problem. Retained or partially extruded anthers provide nutrient-rich tissue that can facilitate early pathogen establishment, so even modest differences in extrusion can translate into meaningful differences in initial infection success (Kubo et al., 2013). This principle has been demonstrated in wheat studies where “minor differences” in anther extrusion measurably affected FHB outcomes, supporting floral morphology as a contributor to Type I (initial

infection) resistance rather than a purely incidental correlation. Comparative mapping in winter wheat likewise reports overlap between QTL for FHB resistance and anther retention/extrusion, reinforcing the interpretation of floral morphology as a meaningful infection-modulating trait (Buerstmayr & Buerstmayr, 2015; Lu et al., 2013; Steiner et al., 2019). Together, these studies support interpreting our negative AE–FHB association as biologically grounded: improving AE should improve pollination potential for hybrid wheat systems while also reducing a key infection opportunity during flowering (Longin et al., 2012; Buerstmayr & Buerstmayr, 2015).

There were differences in Fusarium head blight (FHB) pressure between the two growing seasons. The 2023–2024 season showed the expected disease development pattern, whereas the 2024–2025 season experienced relatively lower disease pressure due to irrigation issues. This contrast is reflected in the heritability estimates for FHB-related traits reported in Table 3. For FHB severity, Fusarium damaged kernels (FDK), and DON accumulation, the broader pattern fits the "QTLome" view of FHB: resistance is highly polygenic, with some repeatedly observed regions embedded in a large landscape of smaller effects, and different resistance components can share partial overlap while also retaining component-specific signals (Venske et al., 2019; Buerstmayr et al., 2020). In an *Ae. tauschii*-derived panel, these insights are especially actionable because D-genome haplotypes can introduce new allelic variation in upstream signaling and stress responsiveness, potentially altering disease progress and toxin outcomes even when visual symptoms appear similar (He et al., 2019; Nyine et al., 2021).

From a breeding perspective, the most important implication is integrative rather than trait-by-trait. Heading date should be tuned to target environments, but phenology must be explicitly controlled in FHB analyses to avoid confounding and spurious "resistance" signals driven by exposure (He et al., 2016; Xu et al., 2019; Arruda et al., 2016). AE appears to be a

high-leverage trait because it simultaneously supports hybrid seed production through improved pollen shedding/outcrossing and reduces a biologically meaningful infection niche during flowering (Longin et al., 2012; Muqaddasi et al., 2017b; Buerstmayr & Buerstmayr, 2015). Durable improvement for disease and grain-safety endpoints should combine pyramiding of stable loci with approaches that capture the remaining polygenic variance, while recognizing that severity, FDK, and DON are related but not identical biological outcomes (Venske et al., 2019; Buerstmayr et al., 2020). Finally, the *Ae. tauschii*-derived nature of the panel strengthens this strategy by offering a clear route to deploy D-genome alleles that enhance plasticity, reproductive performance, and defense -precisely the combination needed under variable and increasingly unpredictable environments (Akhunov et al., 2010; He et al., 2019; Gaurav et al., 2021; Nyine et al., 2021).

This study demonstrates that a *Ae. tauschii*-derived wheat panel is especially powerful for uncovering biologically meaningful D-genome associations across phenology, floral morphology, and Fusarium-related traits. The enrichment of significant SNPs on D-genome chromosomes is consistent with the panel's origin and with the broader understanding that expanding D-genome diversity can reveal adaptive and resilience-related variation that is otherwise difficult to detect in conventional elite panels (Akhunov et al., 2010). Within this framework, heading date loci reinforce the role of multi-layer developmental regulation, while AE emerges as a particularly strategic target that links hybrid-wheat functionality with reduced infection opportunity at anthesis.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433–459.

Adhikari, A., Basnet, B. R., Crossa, J., Dreisigacker, S., Camarillo, F., Bhati, P. K., Jarquin, D., Manes, Y., & Ibrahim, A. M. H. (2020). Genome-wide association mapping and genomic prediction of anther extrusion in CIMMYT hybrid wheat breeding program via modeling pedigree, genomic relationship, and interaction with the environment. Frontiers in Genetics, 11, 586687.

Adhikari, L., Shrestha, S., Wu, S., Crain, J., Gao, L., Evers, B., Wilson, D., Ju, Y., Koo, D.-H., Hucl, P., Pozniak, C., Walkowiak, S., Wang, X., Wu, J., Glaubitz, J. C., DeHaan, L., Friebe, B., & Poland, J. (2022). A high-throughput skim-sequencing approach for genotyping, dosage estimation and identifying translocations. Scientific Reports, 12, 17583.

Akhunov, E. D., Akhunova, A. R., Anderson, O. D., Anderson, J. A., Blake, N., Clegg, M. T., ... & Dvorak, J. (2010). Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC genomics, 11(1), 702.

Arruda, M. P., Lipka, A. E., Brown, P. J., Krill, A. M., Thurber, C., Brown, P. J., Brown-Guedira, G., … & Kolb, F. L. (2016). Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (Triticum aestivum L.). Molecular Breeding, 36(3), 84.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of statistical software, 67, 1-48.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300.

Boeven, P. H., Longin, C. F. H., Leiser, W. L., Kollers, S., Ebmeyer, E., & Würschum, T. (2016). Genetic architecture of male floral traits required for hybrid wheat breeding. Theoretical and Applied Genetics, 129, 2343–2357.

Browning, B. L., & Browning, S. R. (2016). Genotype imputation with millions of reference samples. The American Journal of Human Genetics, 98(1), 116–126.

Buerstmayr, M., & Buerstmayr, H. (2015). Comparative mapping of quantitative trait loci for Fusarium head blight resistance and anther retention in the winter wheat population Capo × Arina. Theoretical and Applied Genetics, 128(8), 1519–1530.

Buerstmayr, M., Steiner, B., & Buerstmayr, H. (2020). Breeding for Fusarium head blight resistance in wheat—progress and challenges. Plant Breeding.

Cavalet-Giorsa, E., Gonzalez-Munoz, A., Athiyannan, N., Holden, S., Salhi, A., Gardener, C., ... & Krattinger, S. G. (2024). Origin and evolution of the bread wheat D genome. Nature, 633(8031), 848-855.

Clark, A. J., Sarti-Dvorjak, D., Brown-Guedira, G., Dong, Y., Baik, B. K., & Van Sanford, D. A. (2016). Identifying rare FHB-resistant segregants in intransigent backcross and F2 winter wheat populations. Frontiers in Microbiology, 7, 277.

Clinesmith, M. A., Fritz, A. K., Lemes da Silva, C., Bockus, W. W., Poland, J. A., Dowell, F. E., & Peiris, K. H. S. (2019). QTL mapping of Fusarium head blight resistance in winter wheat cultivars ‘Art’ and ‘Everest’. Crop Science, 59(3), 911–924.

Cowger, C., & Arellano, C. (2013). Fusarium graminearum infection and deoxynivalenol concentrations during development of wheat spikes. Phytopathology, 103(5), 460-471.

Dubcovsky, J., & Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. Science, 316(5833), 1862-1866.

El Hanafi, S., et al. (2022). Hybrid seed set in relation with male floral traits in wheat. Plants, 11(4), 508.

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. The plant genome, 4(3).

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

Gaikpa, D. S., Lieberherr, B., Maurer, H. P., Longin, C. F. H., & Miedaner, T. (2020). Comparison of rye, triticale, durum wheat and bread wheat genotypes for Fusarium head blight resistance and deoxynivalenol contamination. Plant Breeding, 139(2), 251-262.

Garst, N., et al. (2023). Evaluation of pollination traits important for hybrid wheat breeding. Crop Science.

Gaurav, K., Arora, S., Silva, P., Sánchez-Martín, J., Horsnell, R., Gao, L., ... & Wulff, B. B. (2022). Population genomic analysis of Aegilops tauschii identifies targets for bread wheat improvement. Nature biotechnology, 40(3), 422-431.

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. Journal of the American Statistical Association, 88(423), 881–889.

Gianola, D., & van Kaam, J. B. C. H. M. (2008). Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics, 178(4), 2289–2303.

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics, 32(18), 2847–2849.

He, X., Singh, P. K., Dreisigacker, S., Singh, S., Lillemo, M., & Duveiller, E. (2016). Dwarfing genes Rht-B1b and Rht-D1b are associated with both type I FHB susceptibility and low anther extrusion in two bread wheat populations. PLOS ONE, 11(9), e0162499.

Jolliffe, I. T. (2002). Principal component analysis (2nd ed.). Springer.

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., & Eskin, E. (2008). Efficient control of population structure in model organism association mapping. Genetics, 178(3), 1709–1723.

Kansas Agricultural Experiment Station. (2007). Plant variety protection application for wheat cultivar “Danby” (Triticum aestivum L.) (PVP No. 200700177). U.S. Department of Agriculture, Plant Variety Protection Office.

Kansas State University Research Foundation. (2010). Plant variety protection application for wheat cultivar “Everest” (Triticum aestivum L.) (PVP No. 201000387). U.S. Department of Agriculture, Plant Variety Protection Office.

Kansas State University Research Foundation. (2015). Plant variety protection application for wheat cultivar “KanMark” (Triticum aestivum L.) (PVP No. 201500379). U.S. Department of Agriculture, Plant Variety Protection Office.

Kansas State University Research Foundation. (2017). Plant variety protection application for wheat cultivar “Zenda” (Triticum aestivum L.) (PVP No. 201700282). U.S. Department of Agriculture, Plant Variety Protection Office.

Kubo, K., Fujita, M., Kawada, N., Nakajima, T., Nakamura, K., Maejima, H., … & Matsunaka, H. (2013). Minor differences in anther extrusion affect resistance to Fusarium head blight in wheat. Journal of Phytopathology, 161(5), 308–314.

Langer, S. M., Longin, C. F. H., & Würschum, T. (2014). Phenotypic evaluation of floral and flowering traits with relevance for hybrid breeding in wheat (Triticum aestivum L.). Plant Breeding, 133(4), 433–441.

Lever, J., Krzywinski, M., & Altman, N. (2017). Principal component analysis. Nature Methods, 14(7), 641–642.

Longin, C. F. H., Mühleisen, J., Maurer, H. P., Zhang, H., Gowda, M., & Reif, J. C. (2012). Hybrid breeding in autogamous cereals. Theoretical and applied genetics, 125(6), 1087-1096.

Lu, Q., Lillemo, M., Skinnes, H., He, X., Shi, J., Ji, F., … & Bjørnstad, Å. (2013). Anther extrusion and plant height are associated with type I resistance to Fusarium head blight in bread wheat line “Shanghai-3/Catbird”. Theoretical and Applied Genetics, 126(2), 317–334.

Miedaner, T., Flamm, C., & Oberforster, M. (2024). The importance of Fusarium head blight resistance in the cereal breeding industry: Case studies from Germany and Austria. Plant Breeding, 143(1), 44-58.

Miedaner, T., Sieber, A. N., Desaint, H., Buerstmayr, H., Longin, C. F. H., & Würschum, T. (2017). The potential of genomic-assisted breeding to improve Fusarium head blight resistance in winter durum wheat. Plant Breeding, 136(5), 610-619.

Muqaddasi, Q. H., Brassac, J., Börner, A., Pillen, K., & Röder, M. S. (2017a). Genetic architecture of anther extrusion in spring and winter wheat. Frontiers in Plant Science, 8, 754.

Muqaddasi, Q. H., Pillen, K., Plieske, J., Ganal, M. W., & Röder, M. S. (2017b). Genetic and physical mapping of anther extrusion in elite European winter wheat. PLOS ONE, 12(11), e0187744.

Nyine, M., Adhikari, E., Clinesmith, M., Aiken, R., Betzen, B., Wang, W., … & Akhunov, E. (2021). The haplotype-based analysis of Aegilops tauschii introgression into hard red winter wheat and its impact on productivity traits. Frontiers in Plant Science, 12, 716955.

Okada, T., Jayasinghe, J. E. A. R. M., Eckermann, P., Watson-Haigh, N. S., Warner, P., Hendrikse, Y., … & Whitford, R. (2019). Effects of Rht-B1 and Ppd-D1 loci on pollinator traits in wheat. Theoretical and Applied Genetics, 132(7), 1965–1979.

Pearson, K. (1895). VII. Note on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London, 58(347–352), 240–242.

Prat, N., Guilbert, C., Prah, U., Wachter, E., Steiner, B., Langin, T., ... & Buerstmayr, H. (2017). QTL mapping of Fusarium head blight resistance in three related durum wheat populations. Theoretical and applied genetics, 130(1), 13-27.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8), 904–909.

R Core Team. (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. BMC Genomics, 15, 478.

Šarčević, H., et al. (2023). Evaluation of inoculation methods for determination of FHB resistance in wheat (including maize-stalk/maize-based inoculation approaches). Agronomy, 13(4), 1175.

Skinnes, H., Semagn, K., Tarkegne, Y., Marøy, A. G., & Bjørnstad, Å. (2010). The inheritance of anther extrusion in hexaploid wheat and its relationship to Fusarium head blight resistance and deoxynivalenol content. Plant Breeding, 129(2), 149–155.

Steiner, B., Buerstmayr, M., Wagner, C., Danler, A., Eshonkulov, B., Ehn, M., … & Buerstmayr, H. (2019). Fine-mapping of the Fusarium head blight resistance QTL Qfhs.ifa-5A identifies two resistance QTL associated with anther extrusion. Theoretical and Applied Genetics, 132(7), 2039–2053.

USDA-ARS. (2004). Southern Regional Performance Nursery (SRPN) table index [Dataset]. U.S. Department of Agriculture.

USDA-ARS. (2018). Southern Regional Performance Nursery (SRPN) table index [Dataset]. U.S. Department of Agriculture.

USDA-ARS. (2022). Southern Regional Performance Nursery (SRPN) table index [Dataset]. U.S. Department of Agriculture.

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

Venske, E., dos Santos, R. S., Farias, D. R., Rother, V., da Maia, L. C., Pegoraro, C., & de Oliveira, A. C. (2019). Meta-analysis of the QTLome of Fusarium head blight resistance in bread wheat: refining the current puzzle. Frontiers in Plant Science, 10, 727.

Whalen, A., Gorjanc, G., & Hickey, J. M. (2020). AlphaFamImpute: High-accuracy imputation in full-sib families from genotype-by-sequencing data. Bioinformatics, 36(15), 4369–4371.

Xavier, A., Muir, W. M., & Rainey, K. M. (2020). bWGR: A Bayesian whole-genome regression package for advanced genomic prediction. Bioinformatics, 36(8), 1466–1468.

Xu, K., He, X., Dreisigacker, S., He, Z., & Singh, P. K. (2019). Anther extrusion and its association with Fusarium head blight in CIMMYT wheat germplasm. Agronomy, 10(1), 47.

Xu, Y., Li, Y., Bian, R., Zhang, G., Fritz, A. K., Dong, Y., ... & Bai, G. (2023). Genetic architecture of quantitative trait loci (QTL) for FHB resistance and agronomic traits in a hard winter wheat population. The Crop Journal, 11(6), 1836-1845.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, 38(2), 203–208.

Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. PLOS Genetics, 9(2), e1003264.

Zhu, T., Wang, L., Rimbert, H., Rodriguez, J. C., Deal, K. R., De Oliveira, R., ... & Luo, M. C. (2021). Optical maps refine the bread wheat Triticum aestivum cv. Chinese Spring genome assembly. The Plant Journal, 107(1), 303-314.

Chapter 4 - Exploring Omics and Methods to Improve Wheat Prediction Across Locations, Years, and New Genotypes

Abstract

Genotype-by-environment interaction limits the environment-specific accuracy of genomic prediction for complex traits such as grain yield, motivating parametrizing models with a combination of genomics with environmental and high-throughput phenotyping data. We evaluated multi-omics prediction in a winter wheat (*Triticum aestivum* L.) panel of 399 lines tested across 11 site-year environments in Kansas and Nebraska over two seasons (2023–24 and 2024–25), spanning irrigated and dryland conditions. Four traits (grain yield, days to heading, manual plant height, sensor-based plant height) were modeled using baseline BLUP and GBLUP and a suite of regularized regression and machine-learning methods. Features included SNP markers, soil and weather covariates, seed NIR spectra, and phenomic covariates (WithPheno) versus models excluding these phenotypes (WithoutPheno). Predictive ability, measured as the Pearson (r) correlation between predicted and observed values, were computed within and across environments under five validation schemes: CV00 (75/25 random split), CV01 (25/75), LOLO (leave-one-location-out), LGO (leave-20%-of-genotypes-out), and LOYO (leave-one-year-out). For grain yield within environments, the prediction improvement over the baseline (WithoutPheno/WithPheno) were CV00 -0.04/0.13, CV01 0.01/0.13, LOLO 0.00/0.11, LGO 0.05/0.28, and LOYO 0.01/0.11. Overall, phenomic augmentation delivered trait- and scenario-dependent gains, whereas predictive contribution from soil, weather, and NIR were negligible. Results support selecting models and predictor sets based on the intended breeding deployment

scenario, prioritizing phenomics-enabled pipelines for product placement and genomics-centered approaches for estimation of breeding values.

Introduction

Genotype-by-environment (G×E) interactions pose a central challenge in crop breeding, as genotypes that excel in one environment may perform poorly in another (Jubair & Domaratzki, 2023). Plant breeders traditionally conduct multi-environment trials (MET) to evaluate new germplasm across diverse locations and years, aiming to select genotypes with broad adaptation or specific adaptation (Barreto et al., 2024, Resende et al., 2025). However, METs are resource-intensive, often constrained by limited seed, large entry numbers, and high costs, leading to unbalanced experimental designs and delaying identification of stable high-yielding cultivars (Jarquin et al., 2020). Consequently, promising genotypes can be missed if their potential is not expressed under the tested conditions (Krause et al., 2020). These challenges motivate the development of predictive approaches to evaluate performance in untested environments and accelerate breeding progress (Heffner et al., 2009; Watson et al., 2019), highly quantitative traits like yield can be hard to accurately predict in wheat using traditional genomics-only approaches, therefore, improving prediction accuracies for these types of traits can lead to improved genetic gains in breeding programs. Genomic selection (GS) is one such approach that has transformed plant breeding by using genome-wide DNA markers to predict breeding values, thus enabling selection prior to phenotypic evaluation (Meuwissen et al., 2001; Crossa et al., 2017). By capturing small-effect quantitative trait loci across the genome, GS can increase genetic gain per unit time - shortening breeding cycles and improving selection intensity for complex traits like yield (Crossa et al., 2017; Wang et al., 2024). In wheat, genomic prediction has been implemented to enhance grain yield and other traits (Battenfield et al., 2016).

Yet, prediction accuracy often declines when candidate genotypes or target environments differ from the training data, in large part due to G×E interactions that are not accounted for by baseline models (Jarquín et al., 2014; Heslot et al., 2014). To address this limitation, researchers have developed models that explicitly incorporate G×E, including reaction-norm (genotype-specific environmental response) models and the use of environmental covariates in genomic prediction (Heslot et al., 2014; Jarquín et al., 2014, Costa-Neto et al., 2021). Early studies demonstrated that integrating specific environment descriptors (e.g., weather and soil variables) into genomic prediction models can improve the ability to predict yield in new seasons or sites not represented in the training set (Heslot et al., 2014; Millet et al., 2019). For instance, using historical climate data as covariates in genomic models enabled more accurate yield prediction for untested environments, effectively capturing G×E-driven yield potential and helping breeders cope with climate variability (Gillberg et al., 2019; Roorkiwal et al., 2018). These findings underscore that incorporating "enviromics" - high-dimensional environmental data - into GS has potential for improving prediction across variable conditions by modeling genetic and non-genetic factor (Costa-Neto et al., 2021; Jarquin et al., 2020, Wang et al., 2025).

In parallel with advances in envirotypic modeling, machine learning (ML) has emerged as a powerful complement to traditional linear models for genomic prediction (Montesinos-López et al., 2019; Azodi et al., 2019). Unlike linear mixed models (e.g., GBLUP) that assume additive genetic effects and mainly capture linear patterns, some ML methods can learn complex non-linear associations among genotypic markers, environmental factors, and management variables without the explicit parameterization of these interactions (González-Recio & Forni, 2011; Jubair & Domaratzki, 2023). For example, tree-based ensemble methods such as random forests and gradient boosting partition the predictor space in a data-driven manner and can model

high-order interactions in the parameter spaces; these methods may yield prediction accuracies for complex traits on par with or superior to GBLUP, especially when epistasis or G×E effects are pronounced (González-Recio & Forni, 2011; Spindel et al., 2015; González-Recio et al., 2013). As a result, there is growing interest in deploying state-of-the-art model approaches to exploit the increasing volumes of genomic and environmental data ("big data") now available in plant breeding (Desta & Ortiz, 2014; González-Camacho et al., 2018).

In addition to weather and climate information, modern breeding programs are generating other high-dimensional datasets - often termed *omics* - that can be attached for prediction. For example, near-infrared (NIR) spectral profiles of seeds or canopy reflectance from high-throughput screenings can serve as phenotypic predictors correlated with complex traits like yield (Montesinos-López et al., 2023). Integrating such multi-omics data with genomic markers has been shown to markedly improve predictive power for complex traits, as each data layer captures different aspects of the phenotype continuum (Wang et al., 2024). These highlight that no single data layer captures all heritable variation for a complex trait; instead, an integrative approach that fuses genomic, environmental, and other omics predictors is needed to maximize predictive accuracy and gain biological insight.

GS in the era of big data is moving toward a more integrative, predictive paradigm. By coupling genome-wide markers with environmental covariates and high-throughput phenotypic/omics data in robust ML frameworks, breeders can better model G×E and gene-network effects to improve the reliability of predictions (Crossa et al., 2017; Jarquin et al., 2020). This study embraces that paradigm. We investigate the integration of multi-omics features into genomic prediction for wheat traits. Specifically, we compare a traditional phenomic best linear

unbiased prediction (BLUP) and genomic best linear unbiased prediction (GBLUP) against several ML methods. Through this comprehensive exploration, we aim to demonstrate how coupling GS with omics features and ML can enhance predictive ability in wheat breeding, thereby helping to select superior genotypes more efficiently for variable target environments and improve genetic gain. We also investigate the contributions of the omics layers into the predictive ability.

Materials and Methods

Plant Materials and Field Trials

The study was conducted on a panel of 399 winter wheat lines developed at Kansas State University, representing a broad germplasm base of elite breeding lines, commercial cultivars, and historical varieties. These lines originate from 203 distinct crosses and encompass substantial variation in traits such as maturity and yield potential. Field trials were carried out over two growing seasons (Crop years 2024 and 2025) at multiple test sites in Kansas and Nebraska (Figure 1). In total, 11 distinct environments (location-year combinations) were evaluated, four in 2023/2024 season and seven 2024/2025 season, capturing a range of climatic and soil conditions typical of the U.S. central Great Plains. The four environments in 2023/2024 season were under pivot irrigation compensating water necessity during the growth stages, and from the seven environments in 2024/2025 season, four were irrigated and three dryland. Irrigated locations control water availability and reduce a major source of environmental variation, thereby minimizing error variance. The inclusion of dryland locations, however, expanded the range of environmental conditions evaluated and enabled the study to better represent the most common growth conditions encountered in hard red winter wheat production.

The experimental design in each location followed a unreplicated optimized arrangement (Murillo, D. A. et al., 2021). All 399 lines were grown in unreplicated plots per environment, with a subset (~5%) of entries designated as repeated check cultivars randomly interspersed to enable spatial adjustment. Each plot consisted of seven rows (2.15 m length x 1.2 m width, with 0.15 m inter-row spacing). To ensure uniform stand establishment, approximately 1350 seeds per plot were planted using standard planting practices. Sowing was done between late September to early November of each year at each site (Data Annex 1), and standard agronomic practices (fertilization, weed and pest control) were applied uniformly. Harvest dates ranged from late June to early July, varying slightly by location and year depending on maturity and weather (Data Annex 1). Plots were harvested with a small-plot combine harvester (Wintersteiger/Zurn) to measure moisture and harvest weight, which were used to calculate grain yield.

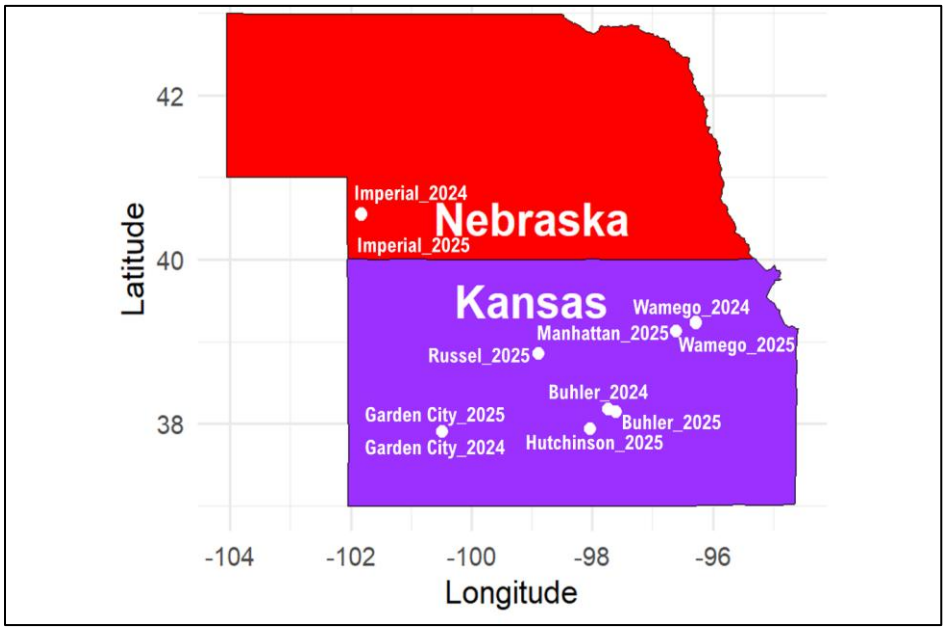

Figure 4.1. Geographic distribution of field trial environments. Map showing the locations of the 11 site-year environments in Kansas and Nebraska, including irrigated and dryland trials in the 2023–24 and 2024–25 growing seasons. Site names and years are indicated for each point.

Phenotypic Data Collection

Throughout the growing season, multiple phenotypic traits were measured for each plot. Grain yield (kg ha^{-1}, adjusted to 12% moisture) was computed from the plot harvest weight. Plant height (cm) was measured at maturity using a meter stick from the soil surface to the tip of the spikes (excluding awns) for representative plants in each plot. An RGB remote sensing platform was also provided by Corteva Agriscience to collected plant height data, with height values calibrated for each location using ground collected data. Days to heading was recorded as the number of days after January 1st at which 50% of the spikes in a plot had emerged from the boot as recorded in Julian days. In addition to field observations, high-throughput phenomic data were collected. A UAV-based remote sensing platform was deployed at approximately from spring green-up (weather permitting) to capture multispectral data. The UAV (DJI Matrice 600) carried an integrated multi-sensor payload. Flights were conducted during the spring, between Feekes growth stages 6 and 10, with two to three flights performed per field at the irrigated sites. No flights were conducted at the dryland sites. Each flight produced an orthomosaic image of the trial from which individual plot images were extracted. Corteva proprietary applications automated image processing, reflectance spectra for each plot were calculated by interpolating and calibrating raw data, yielding reflectance values for the vegetation in each plot as we refer to as spring greenness features. Spring greenness describes the amount and intensity of green vegetation present in the plot during spring growth, providing an indicator of canopy development and early-season vigor. These high-throughput imaging data provided temporal phenotypic information on crop growth and canopy status, complementing ground measurements.

To characterize seed composition, we obtained NIR spectral profiles on seed samples for each genotype. In the first year, ~300 g of foundation seed per line was used for both planting

and initial NIR analysis. From each line's seed lot, a 50g sample was scanned with a FOSS DS2500 Analyzer prior to planting. The spectrometer collected absorbance spectra from 850 to 2500nm at 8nm resolution, yielding 206 wavelength features. This "baseline" NIR scan (on the seed that was subsequently planted) captures inherent genotype differences in grain composition. After harvest of the first-year trials, a second 50g grain sample was taken from the harvested grain of each line in each environment and analyzed by NIR in the same manner. This provided a post-harvest spectral profile of the grain grown under field conditions. In the second year, we extended this approach. Seeds harvested from each specific environment of year one was retained, and a representative subset was scanned to obtain environment-specific NIR signatures. The NIR data were integrated as predictive features in the prediction models with data scams prior to planting, representing an additional phenomic layer. Because NIR spectra are highly collinear, NIR wavelength features were averaged across the first year harvested locations and compressed using Principal Component Analysis (PCA) and replaced by the minimum number of nirPCs components explaining ≥99% of NIR variance.

Genotyping and Genomic Data

All lines were genotyped with genome-wide DNA markers to enable genomic prediction. Two seeds per line were collected and submitted for genotyping. Genotyping was performed by Corteva Agriscience's genotyping laboratory using the Illumina Infinium SNP array platform. Each sample was assayed with a 10k SNP array (custom wheat SNP array) according to the Corteva's protocol (DNA amplification, fragmentation, array hybridization, and single-base extension with fluorescent detection). Quality control filters were applied to remove markers with poor clustering, low call rates (<95%), or minor allele frequency < 0.01. The final dataset

included approximately 5k high-quality SNP markers distributed across the wheat genome. These SNPs were encoded as 0/1/2 allele dosage (for homozygous reference, heterozygous, homozygous alternate) for use in analyses. Prior to model training, missing SNP genotypes (<1% overall) were imputed using marker imputation function IMP from Xavier, A., 2020, Package 'bWGR', imputed with the expected frequency.

Environmental Data (Enviromics)

A comprehensive set of environmental covariates was gathered to characterize each trial environment. We leveraged the NASA POWER database, which provides gridded daily climate data globally (Sparks, 2018). Using the EnvRtype R package (Costa-Neto et al., 2021), we extracted daily weather variables for each field location over the growing season of each year. The variables included solar radiation metrics (e.g., top-of-atmosphere insolation and surface shortwave radiation), temperature indices (daily minimum, maximum, and dew point temperatures), precipitation (rainfall mm), relative humidity, wind speed, and derived ag-climatic indices. Among the derived indices were evapotranspiration (ET), water balance (precipitation minus ET), vapor pressure deficit and its diurnal slope, an effect of temperature on radiation use efficiency (cumulative thermal time weighted by radiation, denoted FRUE), and grow degree days (GDD) accumulated daily. To reduce dimensionality and align with crop developmental stages meaningfully, the daily weather time-series were aggregated into temporal windows. We partitioned the growing season into discrete 30-day intervals from planting until maturity (approximately 240 days, resulting in ~8 intervals), corresponding roughly to establishment, vegetative, heading/flowering, and grain fill periods. For each variable, the mean within each 30-day window was calculated to represent the environmental conditions during that growth phase.

This yielded a set of environmental features for each location-year such as “mean temperature in heading period” or “total radiation in grain fill period”. The compilation of these climate and site variables forms the enviromics dataset for our study, enabling models to leverage G×E by associating genotype performance with specific environmental conditions (Costa-Neto et al., 2021). Additionally, to provide soil environmental context, data were retrieved using the ‘Xpolaris’ R package (Moro Rosso et al., 2021). Soil data composed of mean values of organic matter, sand, clay and silt content of soil layers from 0 to 30 centimeter.

Statistical Analysis

Statistical analyses were conducted in R (R Core Team, 2021). Field-trial data were first adjusted for spatial field effects, and genotype performance within each environment was summarized using BLUEs (Best Linear Unbiased Estimation). Potential outliers were screened using studentized residuals, flagged prior to downstream analyses (Aparicio, 2022) (Table 1). Narrow-sense heritability was calculated as using additive and residual variance components estimated from genomic models (VanRaden, 2008) (Table 1). Predictive models spanning linear genomic prediction and ML approaches were trained on designated training sets and evaluated on held-out testing sets under the cross-validation schemes described below; software packages, functions, and citations are provided in Table 1. Variance components were estimated with bWGR package function ‘zsemf’ for genetic variance and residual (Xavier et al., 2019).

Table 4.1. Summary of statistical and genomic prediction methods, R implementations, and citations.

Method	R package	Function(s) | implementation	Citation
Spatial adjustment + BLUE/BLUP + Outlier	MrBean	(R app) | $\mathbf{Y_{ij} = \mu + g_i + e_j + \varepsilon}$	Aparicio, 2022

Mixed-model/GBLUP	bWGR	mm(…) | $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$	VanRaden, 2008; Xavier, 2019
Ridge Regression (RR)	glmnet	cv.glmnet(alpha = 0) | $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$	Hoerl & Kennard, 1970; Friedman et al., 2010
Bayesian Ridge Regression (BRR)	bWGR	wgr(...) | $\mathbf{Y} = \boldsymbol{\mu}\mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$,	de los Campos et al., 2013; Xavier, 2020
Elastic Net (EN)	glmnet	cv.glmnet(alpha = 0.5) | $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$	Zou & Hastie, 2005; Tibshirani, 1996; Friedman et al., 2010
Principal Components Regression (PCR)	pls	pcr(…) | $\mathbf{Y} = \boldsymbol{\beta}_0 + \mathbf{T}_k\boldsymbol{\gamma} + \mathbf{e}$	Mevik & Wehrens, 2007
Partial Least Squares (PLS) regression	pls	plsr(...) | $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$	Wold et al., 2001; Mevik & Wehrens, 2007
Support Vector Regression (SVR)	e1071	svm(...) | $\mathbf{Y}_{\mathbf{SVR(G,E)}} = \mathbf{w} \cdot \emptyset\,(\mathbf{G}, \mathbf{E}) + \mathbf{e}$	Cortes & Vapnik, 1995; Dimitriadou et al., 2006; Ma et al., 2018
Random Forest (RF)	ranger	ranger(…) | $\mathbf{Y}(\mathbf{x}) = \frac{\mathbf{1}}{\mathbf{T}}\sum_{\mathbf{t=1}}^{\mathbf{T}} \mathbf{f_t}(\mathbf{x})$	Breiman, 2001; Wright & Ziegler, 2017
Extreme Gradient Boosting (XGBoost)	xgboost	xgboost(…) | $\mathbf{Y}(\mathbf{x}) = \sum_{\mathbf{m=1}}^{\mathbf{M}} \mathbf{f_m}(\mathbf{x})$	Friedman, 2001; Chen & Guestrin, 2016
Multivariate SNP-BLUP (MSB)	bWGR	zsemf (…) | $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$	Xavier et al., 2019; Xavier et al., 2025
Narrow-sense heritability	bWGR	zsemf (…) | $\mathbf{h^2} = \frac{\boldsymbol{\sigma}_\mathbf{a}^\mathbf{2}}{\boldsymbol{\sigma}_\mathbf{a}^\mathbf{2} + \boldsymbol{\sigma}_\mathbf{e}^\mathbf{2}}$	VanRaden, 2008

Methods are listed with the corresponding R package/tool and primary function(s) used in this study. Equations are provided in simplified form to indicate the general model structure; MrBean refers to the MrBean R application used for spatial adjustment, BLUE estimation, and outlier screening; bWGR implementations include mixed-model genomic prediction (GBLUP) and multivariate SNP-BLUP (ZSEMF). Narrow-sense heritability (h^2) was computed from additive genetic and residual variance components estimated from the genomic model.

Cross-validation

To evaluate predictive performance under realistic breeding scenarios, we implemented five cross-validation schemes representing different degrees of extrapolation in genotype and environment space (Table 2). CV00 and CV01 assess interpolation accuracy under high- and low-training-data conditions using random observation holdouts (25% and 75% test fractions, respectively). Leave-one-location-out (LOLO) evaluates spatial transferability by predicting an unobserved location using all remaining locations for training, whereas Leave-Genotype-Out (LGO) assesses prediction of entirely new genotypes by withholding 20% of lines across all

environments (Jarquin et al., 2020). Finally, Leave-One-Year-Out (LOYO) tests temporal transferability by predicting one year from the other (2024→2025 and 2025→2024). All cross-validations schemes were evaluated within environment to avoid bias and average location prediction, and across environment as well.

Table 4.2. Cross-validation schemes used for genomic prediction model evaluation

Scheme	Name	Training set	Test set	Repeats	Breeding scenario
CV00	Random 25% holdout	75% of observations	Random 25% of observations	10	Interpolation within environments; predicts missing line performance where training data are abundant
CV01	Random 75% holdout	25% of observations	Random 75% of observations	10	Data-scarce setting (early-stage trials); highlights model efficiency with limited phenotypic information
LOLO	Leave-one-location-out	All locations excluding one	All observations from one location (across years)	11 (1/location)	Spatial extrapolation to a new, untested site; relies on transferability/enviromic signal
LGO	20% genotypes out	Remaining 80% of genotypes	20% of genotypes (across environments)	10 repeats	GS scenario: predict new, unphenotyped lines absent from training (Jarquin et al., 2020)
LOYO	Leave-one-year-out	The other year	All observations from one year	2 (2024→2025, 2025→2024)	Temporal extrapolation; predicts a future year using historical data

CV00 = random 25% observation holdout cross-validation; CV01 = random 75% observation holdout cross-validation; LOLO = leave-one-location-out cross-validation (one location withheld across years); LGO = leave-genotypes-out cross-validation (20% of genotypes withheld across all environments); LOYO = leave-one-year-out cross-validation (train on one year and test on the other).

Predictive Traits and Models Evaluation

In this study we modeled four target traits, grain yield (kg ha^{-1}) (YIELD), plant height (cm) (PLTHT), remotely sensed plant height (RS; cm) (PLTHT_RS), and days to heading

(HDDAT). For each trait, we evaluated two predictor configurations: WithPheno, which included within-season phenotypic covariates (e.g., UAV or field-derived measurements) as well additional visual scores predictors (fall stand score 1-9 scale), and WithoutPheno, which excluded within-season phenotypes and relied on pre-season or fixed data sources (e.g., genomics, soil, weather, and seed NIR before planting). As baseline references, we implemented a marker-based GBLUP model and a phenotypic BLUP model using phenotype-only information. The phenotypic BLUP baseline was used where relevant as a non-genomic benchmark; however, in the LGO scenario (prediction of new, unphenotyped genotypes), only genomic-based prediction (GBLUP) is feasible by design because no phenotype-derived information is available for the target lines. Predictive performance was assessed using the Pearson correlation coefficient (r) between observed and predicted values in the held-out (test) set within and across locations. Correlations were computed separately for each trait, model, and cross-validation replicate, and summarized to compare model ranking and quantify the benefit of adding multi-omics and environmental predictors. Model performance is reported in the results as predictive correlation within and across each validation scenario, along with the relative contribution of multi-omics and environmental predictors to overall prediction performance. Within-environment prediction evaluates the ability of a model to rank genotypes within a specific environment, such as a single location-year, whereas across-environment prediction uses information from multiple environments to predict performance across the broader multi-environment trial structure (Guo et al., 2013; Jarquín et al., 2014). Across-environment predictive ability is often higher under random cross-validation because training and testing sets may share environmental structure and relatedness, but this can inflate accuracy relative to more

realistic leave-one-environment-out, forward, or future-environment prediction scenarios (Werner et al., 2020; Sandhu et al., 2021).

Variable Importance

Prediction ability was evaluated using repeated cross-validation under multiple breeding-relevant validation schemes (Table 2). Distributions of predictive ability (r) were summarized using boxplots across all trait × model × schemes × repeat combinations for each predictor set (Figure 3) and across all trait × predictor-set × scheme × repeat combinations for each model (Figure 4). Differences among predictor sets and models were assessed using Tukey's HSD multiple-comparison procedure ($\alpha = 0.05$), and mean-separation groupings were displayed as letters and color-coded categories in the figures**.**

We quantified model-based variable importance to interpret which data layers drove prediction in the best pipelines for each trait and validation scenario. For every trait x cross-validation scheme, we first selected the top-ranked pipeline separately for the WithPheno and WithoutPheno configurations. Using the exact predictor set and algorithm of that top pipeline, we refit the model on the trait dataset after standardization. To make importance comparable within each fitted pipeline, we normalized them to sum to 1 and retained the smallest set of variables accounting for 99% of cumulative importance; these "top variables" were then mapped to predictor groups (Genomic, Soil, Weather, NIR, Pheno), and group contributions were computed as the sum of normalized importance within each group.

Results

We assembled a multi-environment dataset that combines multi-omics information for 399 winter wheat lines evaluated across 11 site-year environments. Below, we first summarize

the phenotypic patterns and trait heritability across environments, and then describe the predictive performance of the genomic and ML methods under different cross-validation schemes for the top best combinations for each model subdivided in “WithPheno” and “WithoutPheno”. As summarized in Figure 2, we present a comprehensive comparison of all prediction models relative to the baseline BLUP and GBLUP across the five validation schemes (CV00, CV01, LOLO, LGO, and LOYO), evaluating performance using both within-environment (within site-year) and across-environment (across site-years) correlations. For each scheme and evaluation metric, we further contrast predictor configurations WithPheno versus WithoutPheno to quantify the incremental value of phenomic information across traits and deployment scenarios. All model × predictor-set × scheme combinations and their summarized results are reported in Supplementary Tables 3 (within-environment) and 4 (across-environment), with complete combination-level outputs provided.

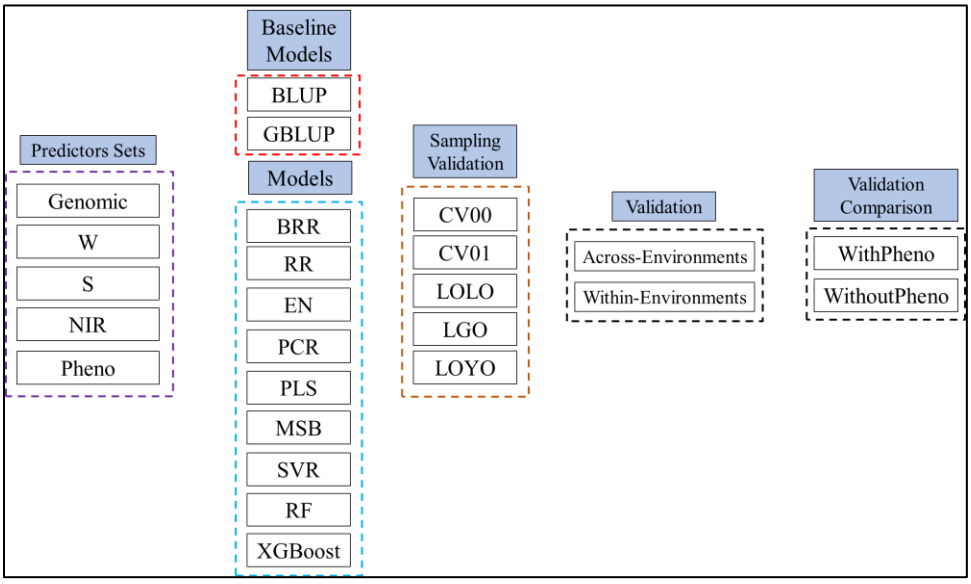

Figure 4.2. Multi-omics prediction workflow and validation framework. Schematic overview of the analysis pipeline used in this study. Predictor sets (left) are assembled into candidate feature configurations and evaluated using baseline models and a panel of statistical and ML models. Baseline models include BLUP (phenotype-only mixed model using pedigree-free phenotypic information) and GBLUP (genomic BLUP using a genomic relationship matrix

from SNP markers). Additional models span linear/shrinkage and nonlinear learners, including BRR (Bayesian ridge regression), RR (ridge regression), EN (elastic net), PCR (principal component regression), PLS (partial least squares), MSB (multivariate SNP-BLUP), SVR (support vector regression), RF (random forest), and XGBoost (gradient-boosted decision trees).

Descriptive statistics for grain yield, plant height, remote-sensed height, and days to heading are summarized for each site-year in Table 3. Grain yield means ranged from approximately 4500 to 11800 Kg ha^{-1} across environments, indicating substantial variation in productivity among locations and years. For example, Manhattan_2025 and Russell_2025 had the lowest average yields (4538 and 4643 Kg ha^{-1}), while Wamego_2025 and Buhler_2024 showed the highest yields (11804 and 9775 kg ha^{-1}, respectively). Coefficients of variation (CV) for yield were generally between ~10-20%, reflecting moderate to high within-environment variability typical of elite breeding trials.

Plant height measured with a meter stick varied from roughly 78 to 102 cm among environments, with slightly taller plants in some irrigated western locations and shorter plants in drier or more stressed environments. CV values for plant height were mostly around 6-9%, indicating relatively consistent stature within each site-year compared with yield. Remote-sensed plant height (lidar-based measurements) was available for a subset of environments (e.g., Wamego_2024, Imperial_2025), where mean heights were close to those obtained from manual measurements, supporting their use as complementary phenotypic predictors (Table 3).

Days to heading, showed clear differences among environments, with means typically between ~118 and 138 days. For instance, Imperial_2024 exhibited one of the latest heading averages, whereas Buhler_2024 and Russel_2025 headed earlier on average. CV values for heading time were generally low (≈2-3%), consistent with the relatively narrow spread of maturity within this elite panel (Table 3).

Narrow-sense heritability (h^2) estimates differed among traits and environments. For grain yield, h^2 ranged from low to moderate, with values as low as 0.06 (Manhattan_2025) and up to 0.48 (Imperial_2025), reflecting differences in environmental noise and the strength of genetic signal across site-years. Plant height tended to show moderate to high heritability, with many environments having h^2 around 0.4-0.7. For days to heading, h^2 was generally moderate (≈0.3 - 0.6), indicating that genetic differences in maturity were consistently expressed across most environments. Overall, Table 3 demonstrates that the trial network captured substantial environmental and phenotypic diversity, while still providing traits with sufficient heritability to support genomic and ML prediction.

Table 4.3. Trait descriptive statistics and narrow-sense heritability estimates for 11 site-year environments in 2023-24 and 2024-25 growing seasons.

Environments	Yield (kg ha^{-1})				Plant height (cm)				Plant height RS (cm)				Days to heading			
	GM	CV	Ve (%)	h^2	GM	CV	Ve (%)	h^2	GM	CV	Ve (%)	h^2	GM	CV	Ve (%)	h^2
Buhler 2024	9776.94	11.55	0.54	0.46	89.13	8.49	0.48	0.52	-	-	-	-	117.88	2.39	0.53	0.47
Garden City 2024	7801.13	17.46	0.70	0.30	-	-	-	-	94.08	7.57	0.64	0.36	131.33	1.85	0.64	0.36
Imperial 2024	7685.57	17.61	0.70	0.30	-	-	-	-	94.64	7.82	0.78	0.22	138.46	0.96	0.73	0.27
Wamego 2024	7477.30	13.70	0.75	0.25	82.75	7.93	0.51	0.49	85.00	5.89	0.27	0.73	123.42	2.16	0.50	0.50
Buhler 2025	-	-	-	-	94.53	8.92	0.54	0.46	101.88	7.77	0.53	0.47	119.72	1.76	0.49	0.51
Garden City 2025	7977.50	19.90	0.76	0.24	-	-	-	-	96.15	8.46	0.68	0.32	128.47	2.44	0.47	0.53
Hutchinson 2025	-	-	-	-	82.76	8.08	0.60	0.40	-	-	-	-	117.27	1.77	0.57	0.43
Imperial 2025	5936.52	9.65	0.52	0.48	-	-	-	-	83.36	7.57	0.64	0.36	135.51	1.14	0.70	0.30
Manhattan 2025	4537.75	12.98	0.94	0.06	78.28	8.01	0.55	0.45	-	-	-	-	127.40	3.03	0.39	0.61
Russel 2025	4643.42	15.92	0.88	0.12	87.90	6.50	0.56	0.44	-	-	-	-	122.68	10.36	0.67	0.33
Wamego 2025	11802.63	14.03	0.67	0.33	95.26	7.06	0.57	0.43	102.16	5.86	0.78	0.22	123.18	2.47	0.51	0.49

General mean (GM), coefficient of variation (CV), residual variance (Ve %), and narrow-sense heritability (h^2) for grain yield, plant height, remote-sensed plant height (RS), and days to heading across each location-year environment.

CV00 scheme: Random 25% Holdout

Within environments (Figure 3; Table S1) CV00, predictive ability was highest overall across traits. For HDDAT, the strongest pipeline was PCR with Genomic + S + W + Pheno (r = 0.61), which exceeded both BLUP and GBLUP (r = 0.56) (Figure 3; Table S1). The best NoPheno configuration also cleared the baseline, led by XGBoost with Genomic + S + W (r = 0.60) (Figure 2; Table S1). For PLTHT, the best baseline-beating pipeline was XGBoost with Genomic + S + Pheno (r = 0.52), exceeding BLUP/GBLUP (r = 0.48); however, the best NoPheno option (BRR with Genomic + W + NIR, r = 0.45) did not surpass the baseline (Figure 2; Table S1). For PLTHT_RS, the top performer was RF with S + Pheno (r = 0.58), strongly exceeding BLUP/GBLUP (r ~ 0.41–0.40), while the best NoPheno result (RF with Genomic, r = 0.37) stayed below baseline (Figure 3; Table S1). For YIELD, the strongest WithPheno pipeline was BRR with Genomic + S + W + NIR + Pheno (r = 0.36), exceeding both BLUP (0.27) and GBLUP (0.18) (Figure 3; Table S1). In contrast, the best NoPheno yield pipeline (EN with Genomic + S + W, r = 0.23) exceeded GBLUP but not BLUP, reinforcing that secondary phenotypes were needed to consistently beat the full phenotypic baseline (Figure 3; Table S1).

Across environments (Figure 4; Table S2), several pipelines still beat baseline, but HDDAT baselines were already very strong: BLUP/GBLUP reached r = 0.83, and no model exceeded that (best WithPheno: BRR with W + NIR + Pheno, r = 0.74) (Figure 4; Table S2). For YIELD, the best baseline-beating model was BRR with S + W + Pheno (r = 0.90), exceeding BLUP (0.89) and GBLUP (0.84); the best NoPheno yield model (EN with Genomic + W + NIR, r = 0.89) exceeded GBLUP but not BLUP (Figure 4; Table S2). For PLTHT and PLTHT_RS, RF was the top baseline-beating method (PLTHT: S + NIR + Pheno, r = 0.77; PLTHT_RS: S +

Pheno, r = 0.83), while the best NoPheno configurations did not clear the corresponding BLUP/GBLUP baselines (Figure 4; Table S2).

CV01 scheme: Random 75% Holdout

Within environments (Figure 3; Table S1), with reduced training data, several models still beat baseline. For HDDAT, the strongest baseline-winning pipeline was RF with S + Pheno (r = 0.55), exceeding BLUP (0.36) and GBLUP (0.48) (Figure 3; Table S1). The best NoPheno option (MSB with Genomic + W + NIR, r = 0.47) exceeded BLUP but not GBLUP (Figure 3; Table S1). For PLTHT, the best pipeline was XGBoost with Genomic + S + W + Pheno (r = 0.42), exceeding BLUP/GBLUP (0.34–0.38), while the best NoPheno (BRR with Genomic + S + W, r = 0.36) again exceeded BLUP but not GBLUP (Figure 3; Table S1). For PLTHT_RS, the top baseline-beating result was RF with S + Pheno (r = 0.53) vs. BLUP/GBLUP (0.29–0.31); the best NoPheno (BRR with Genomic + W, r = 0.30) exceeded BLUP but not GBLUP (Figure 3; Table S1). For YIELD, the best baseline-winning pipeline was PLS with Genomic + S + W + Pheno (r = 0.31) vs. BLUP/GBLUP (0.18–0.11). Notably, the best NoPheno yield model (MSB with Genomic + S + W + NIR, r = 0.19) also exceeded both baselines, though the WithPheno advantage remained clear (Figure 3; Table S1).

Across environments (Figure 4; Table S2), HDDAT remained baseline-dominated: GBLUP reached r = 0.84, and the best WithPheno (BRR with S + NIR + Pheno, r = 0.81) exceeded BLUP but not GBLUP. For PLTHT, BRR with Genomic + S + W + Pheno (r = 0.71) was the top baseline-winning method (Figure 4; Table S2). For PLTHT_RS, RF with Pheno alone reached r = 0.80, exceeding BLUP/GBLUP (0.73) (Figure 4; Table S2). For YIELD, BRR with W + NIR + Pheno (r = 0.89) exceeded BLUP/GBLUP (0.88–0.83), while the best NoPheno

yield model (BRR with S + W + NIR, r = 0.88) again exceeded GBLUP but not BLUP (Figure 4; Table S2).

LOLO scheme: Random Location Holdout

Within environments (Figure 3; Table S1), LOLO reduced performance relative to random-split schemes, but several pipelines still won against baseline, especially with richer inputs. For HDDAT, PCR with Genomic + S + W + Pheno reached r = 0.61, exceeding BLUP/GBLUP (0.56–0.58), and the best NoPheno model (SVR with Genomic + S + W) reached r = 0.65, also exceeding both baselines (Figure 3; Table S1). For PLTHT, the best WithPheno pipeline was RF with Genomic + S + NIR + Pheno (r = 0.52), exceeding BLUP/GBLUP (0.50–0.46); the best NoPheno option (PLS with Genomic + S, r = 0.50) matched BLUP and exceeded GBLUP but did not exceed BLUP (Figure 3; Table S1). For PLTHT_RS, BRR with Genomic + W + Pheno reached r = 0.52, exceeding BLUP/GBLUP (0.43–0.41), and the best NoPheno (BRR with Genomic + W, r = 0.43) matched BLUP and exceeded GBLUP (Figure 3; Table S1). For YIELD, BRR with Genomic + S + W + NIR + Pheno reached r = 0.40, exceeding BLUP/GBLUP (0.29–0.19); the best NoPheno (BRR with Genomic + W + NIR, r = 0.29) matched BLUP and exceeded GBLUP (Figure 3; Table S1).

The LOLO across-environment summary showed the same “baseline-winning” leaders: SVR (NoPheno) for HDDAT and BRR/RF for PLTHT/PLTHT_RS/YIELD, with the same best-performing feature sets and r values reported above (Figure 4; Table S2).

LGO scheme: Random 25% of Genotypes Holdout

Within environments (Figure 3; Table S1), LGO was the most challenging design, and BLUP is not applicable here, so the baseline comparison is against GBLUP only. In this setting,

RF dominated baseline wins across traits when phenotypic covariates were included. For HDDAT, RF with S + NIR + Pheno reached r = 0.61, exceeding GBLUP (0.27) (Figure 3; Table S1). For PLTHT_RS, RF with S + Pheno reached r = 0.56, exceeding GBLUP (0.27) (Figure 3; Table S1). For PLTHT, RF with Genomic + S + W + Pheno reached r = 0.50, exceeding GBLUP (0.37) (Figure 3; Table S1). For YIELD, RF with S + Pheno reached r = 0.37, exceeding GBLUP (0.09) (Figure 3; Table S1). In contrast, best NoPheno results were often at or below GBLUP for height traits (e.g., PLTHT_RS best NoPheno = 0.27, matching GBLUP), while yield still showed a small baseline win without Pheno (EN with Genomic + S + W + NIR, r = 0.14 > 0.09) (Figure 3; Table S1).

Across environments (Figure 4; Table S2), LGO baseline wins became trait-dependent. For HDDAT, GBLUP remained strongest (r = 0.87) and no model exceeded it (best WithPheno: BRR with S + W + NIR + Pheno, r = 0.80). For PLTHT and PLTHT_RS, RF with S + Pheno produced the best baseline wins (PLTHT: r = 0.77 > 0.72; PLTHT_RS: r = 0.80 > 0.72). For YIELD, the best baseline-winning pipeline was again RF with S + Pheno (r = 0.90 > 0.84), and even the best NoPheno yield model (BRR with S + W + NIR, r = 0.89) exceeded GBLUP (Figure 4; Table S2).

LOYO scheme: Year Holdout

Within environments (Figure 3; Table S1) under LOYO, several pipelines outperformed baseline, with one clear exception in HDDAT where NoPheno could be superior. For HDDAT, the best WithPheno model (BRR with Genomic + W + Pheno, r = 0.59) exceeded BLUP (0.57) but did not exceed GBLUP (0.59), whereas the best NoPheno model (SVR with Genomic + S + W) reached r = 0.62, exceeding both baselines (Figure 3; Table S1). For PLTHT and

PLTHT_RS, WithPheno pipelines led by BRR were the strongest baseline winners (PLTHT: Genomic + S + W + Pheno, r = 0.47; PLTHT_RS: Genomic + S + Pheno, r = 0.49), both exceeding BLUP/GBLUP (Figure 2; Table S1). For YIELD, the top baseline-winning pipeline was MSB with Genomic + S + W + NIR + Pheno (r = 0.34), exceeding BLUP/GBLUP (0.23–0.14), and the best NoPheno yield model (MSB with Genomic + NIR, r = 0.24) also exceeded both baselines (Figure 3; Table S1).

Across environments (Figure 4; Table S2), transfer was challenging for the baselines (notably for HDDAT), and multi-omics models strongly outperformed them. For HDDAT, MSB with S + Pheno reached r = 0.78 vs BLUP/GBLUP at 0.17, and the best NoPheno (XGBoost with S) reached r = 0.65. For YIELD, the strongest baseline win was RR with Genomic + Pheno (r = 0.51) vs BLUP/GBLUP (0.14–0.08), and the best NoPheno (XGBoost with W) reached r = 0.39 (Figure 4; Table S2). Similar baseline wins were observed for PLTHT (BRR with S + NIR + Pheno, r = 0.61) and PLTHT_RS (BRR with S + W + NIR + Pheno, r = 0.58) (Figure 4; Table S2).

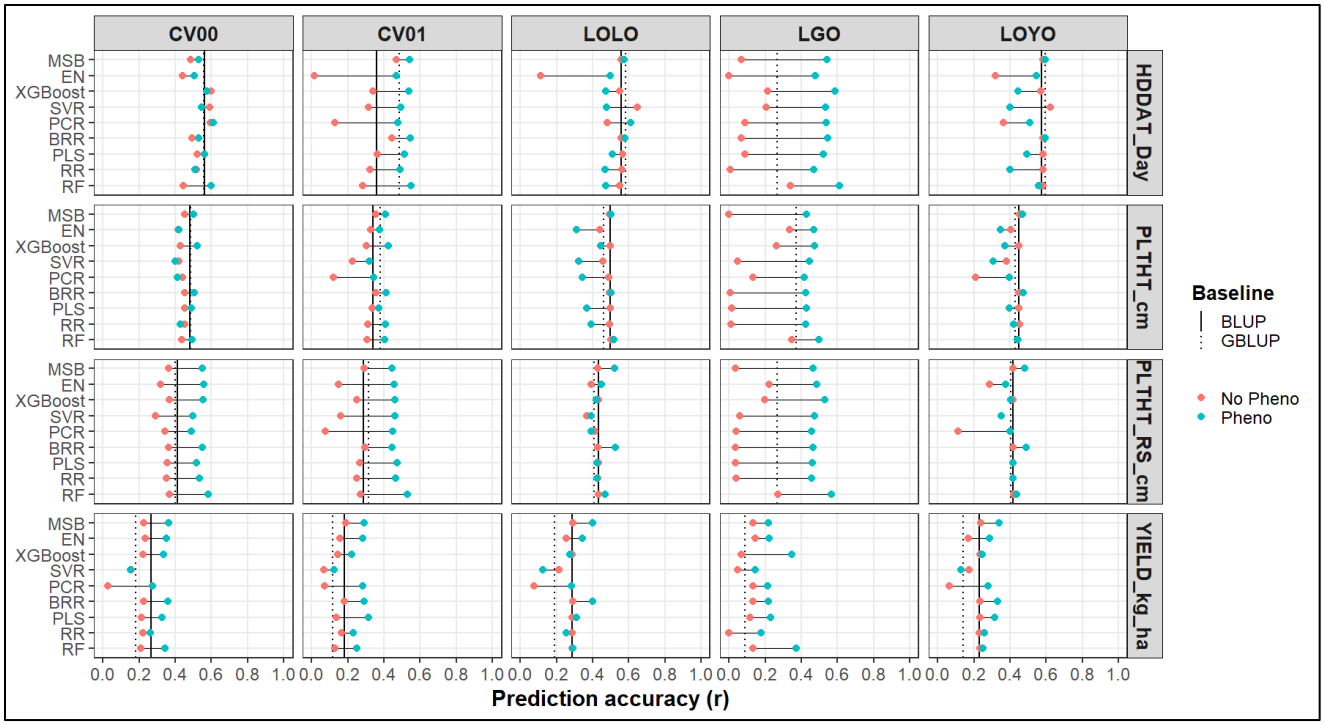

Figure 4.3. Top Prediction ability (Pearson's *r*) for each model within-environments cross validation schemes for four traits, with and without phenotypic covariates. Prediction accuracy is reported as Pearson correlation (r) between observed and predicted values in the test set (higher is better). No Pheno = models trained without phenotypic covariates; Pheno = models trained with phenotypic covariates (secondary/HTP traits) included as predictors. Validation schemes: CV00 = 75% training / 25% testing; CV01 = 25% training / 75% testing; LOLO = leave-one-location-out (train on all locations except one; test on the held-out location); LGO = leave-genotypes-out (hold out a 20% of genotypes across environments); LOYO = leave-one-year-out (train year except one; test on the held-out year). Traits: HDDAT = heading date (days); PLTHT = plant height (cm); PLTHT_RS = remote-sensing–derived plant height (cm); YIELD = grain yield (kg ha^{-1}). Baselines: BLUP = best linear unbiased prediction (phenotype-only mixed model); GBLUP = genomic best linear unbiased prediction (marker-based mixed model). Models: RF = random forest; RR = ridge regression; PLS = partial least squares regression; BRR = Bayesian ridge regression; PCR = principal component regression; SVR = support vector regression; XGBoost = extreme gradient boosting; EN = elastic net regression; MSB = Multivariate SNP-BLUP ensemble. Vertical lines mark baseline performance (BLUP = solid; GBLUP = dashed); paired points are connected within each model to show the effect of including phenotypic covariates.

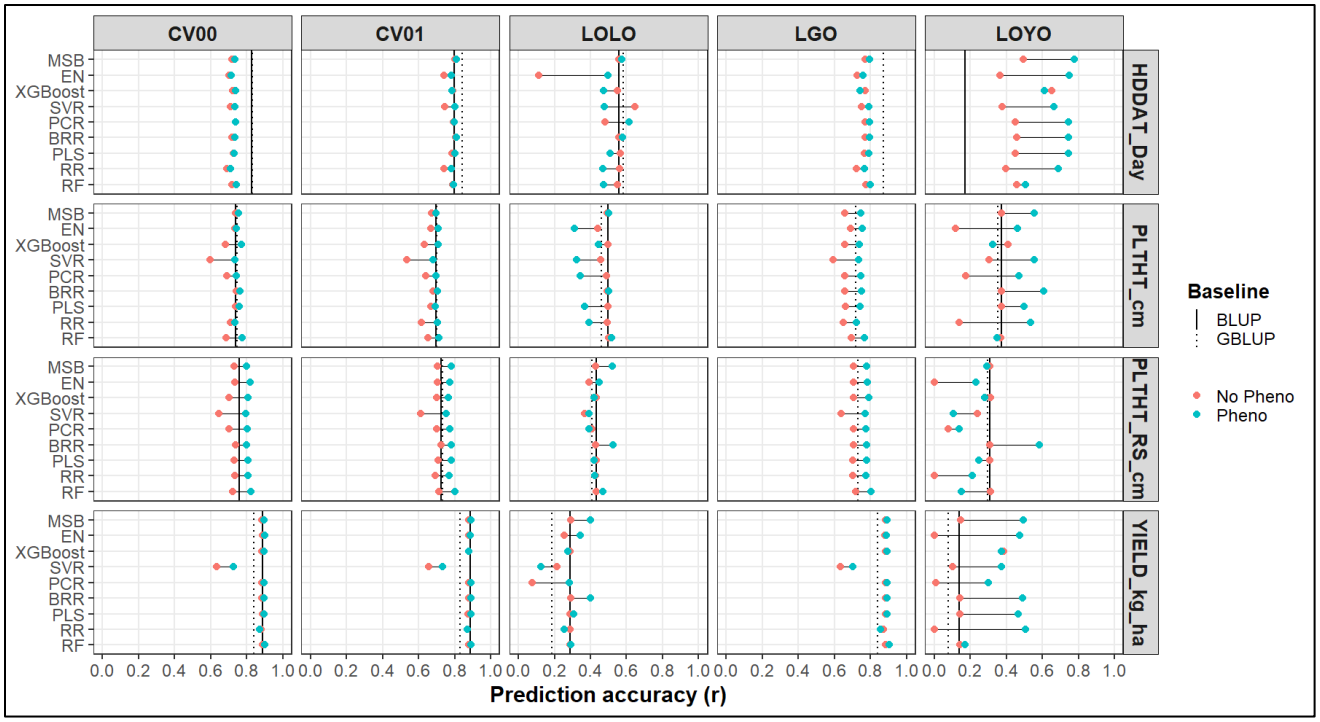

Figure 4.4. Top Prediction ability (Pearson's *r*) for each model across-environments cross-validation schemes for four traits, with and without phenotypic covariates. Prediction accuracy is reported as Pearson correlation (r) between observed and predicted values in the test set (higher is better). No Pheno = models trained without phenotypic covariates; Pheno = models trained with phenotypic covariates (secondary/HTP traits) included as predictors. Validation schemes: CV00 = 75% training / 25% testing; CV01 = 25% training / 75% testing; LOLO = leave-one-location-out (train on all locations except one; test on the held-out location); LGO = leave-genotypes-out (hold out a 20% of genotypes across environments); LOYO = leave-one-year-out (train year except one; test on the held-out year). Traits: HDDAT = heading date (days); PLTHT = plant height (cm); PLTHT_RS = remote-sensing–derived plant height (cm); YIELD = grain yield (kg ha^{-1}). Baselines: BLUP = best linear unbiased prediction (phenotype-only mixed model); GBLUP = genomic best linear unbiased prediction (marker-based mixed model). Models: RF = random forest; RR = ridge regression; PLS = partial least squares regression; BRR = Bayesian ridge regression; PCR = principal component regression; SVR = support vector regression; XGBoost = extreme gradient boosting; EN = elastic net regression; MSB = Multivariate SNP-BLUP ensemble. Vertical lines mark baseline performance (BLUP = solid; GBLUP = dashed); paired points are connected within each model to show the effect of including phenotypic covariates.

Across predictor sets, prediction ability showed a strong dependence on whether correlations were computed within or across locations (Figure 5A). Within-location performance (Panel A) was lowest for single-layer environmental inputs (S, W, and S_W; medians near zero), and only modestly improved by NIR alone or simple NIR-environment combinations. In

contrast, genomic information increased within-location accuracy, and multi-layer combinations that included Genomic plus environmental covariates generally shifted the distribution upward. The clearest gains occurred when secondary phenotypic covariates (WithPheno) were included; predictor sets containing Pheno consistently produced the highest within-location correlations, with the top-ranked combinations forming the best-performing Tukey groups and showing higher medians and upper quartiles than non-Pheno sets (Figure 5A).

Across-location correlations (Figure 5B) were generally higher overall, with many multi-layer predictor sets reaching moderate-to-high median correlations (~0.6-0.75). As in the within-location analysis, the weakest performance was observed for NIR alone and Genomic_NIR, which clustered in the lowest significance groups. However, once Pheno was incorporated - especially in combination with soil and/or weather (e.g., W_Pheno, S_W_Pheno, and multi-layer Genomic_S/W/NIR_Pheno sets) - predictive ability increased markedly and many of the best-performing predictor sets became statistically indistinguishable (shared Tukey letter groups), suggesting a plateau where multiple Pheno-containing combinations provide similarly strong cross-environment prediction.

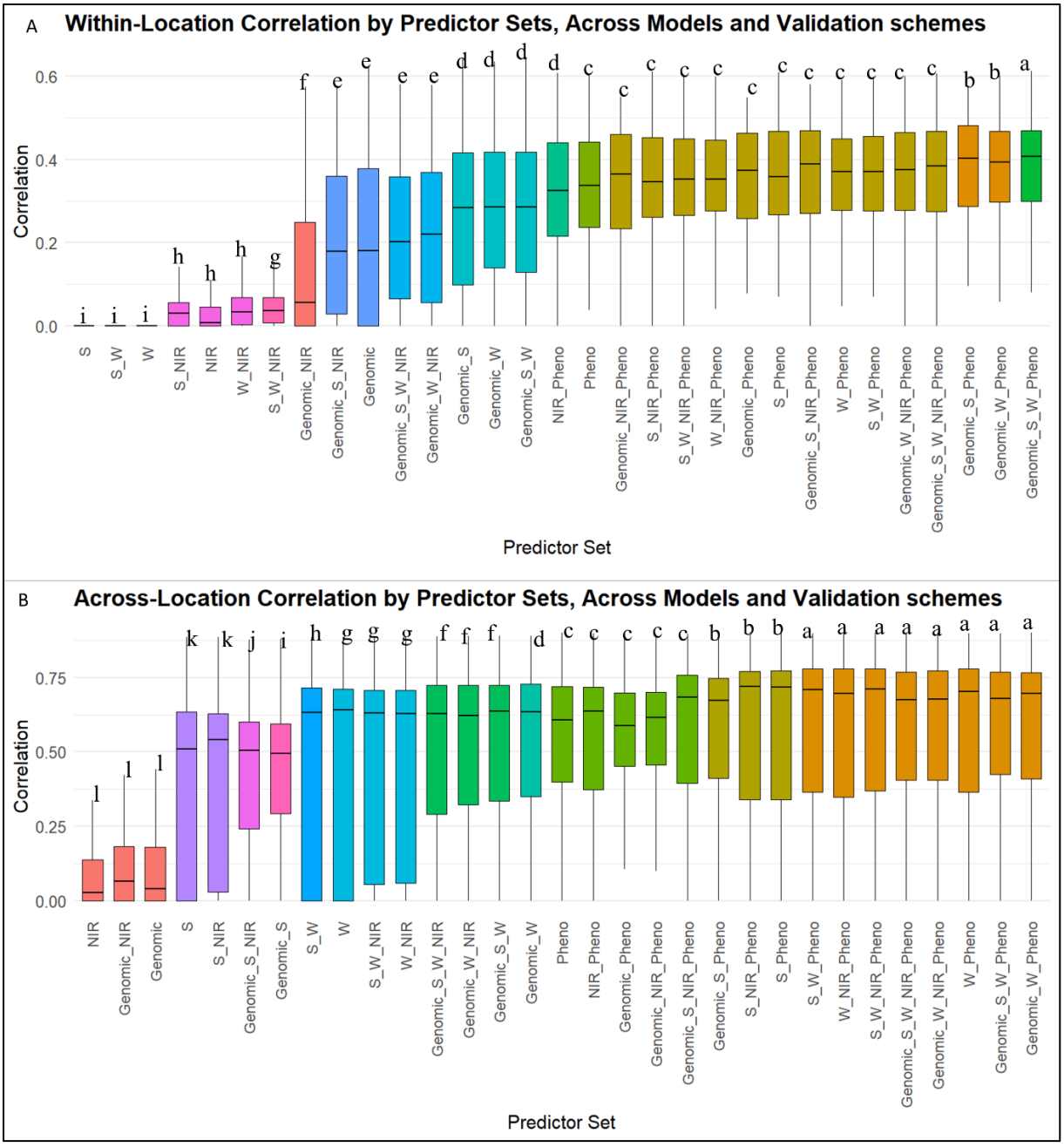

Figure 4.5. Predictor-set effects on prediction ability (r) within vs. across locations across all validation schemes. Panels summarize model prediction ability (Pearson's correlation, r) aggregated across all cross-validation schemes for each predictor set. (A) Within-location correlations were computed within each location (and summarized across repeats/schemes), reflecting model performance for ranking genotypes within environments. (B) Across-location correlations were computed by combining predictions across locations, reflecting model performance for predicting performance across environments. Boxplots show the distribution of correlations across model × trait × scheme combinations for each predictor set; center lines indicate average performance, boxes represent interquartile ranges, whiskers extend to 1.5×IQR,

and points are outliers. Boxes show the interquartile range (Q1–Q3) with the center line indicating the median; whiskers extend to 1.5×IQR and points denote outliers. Predictor sets are ordered by mean r. Letters (and matching box colors) indicate Tukey HSD mean-separation groups at $\alpha = 0.05$; categories sharing a letter are not significantly different. Acronyms / predictor-set codes: Genomic: genome-wide marker data (SNP genotypes); S: soil covariates; W: weather (environmental) covariates; NIR: near-infrared spectroscopy covariates; Pheno: secondary phenotypic covariates (auxiliary phenotypes used as predictors); Combined labels (e.g., S_W, Genomic_W_NIR_Pheno) indicate the concatenation of those data layers in the predictor set.

Model comparisons (Figure 6) indicated clearer separation within locations than across locations. Within-location correlations differed significantly among modeling approaches (Panel A). BLUP and GBLUP produced the highest median performance and formed the top mean-separation groups, followed by RF, XGBoost, MSB, BRR, and PLS, while SVR, PCR, RR and EN showed the lowest medians. In contrast, across-location correlations (Panel B) showed no significant differences among models under Tukey HSD (all models shared the same grouping), although GBLUP/BLUP tended to have slightly higher medians.

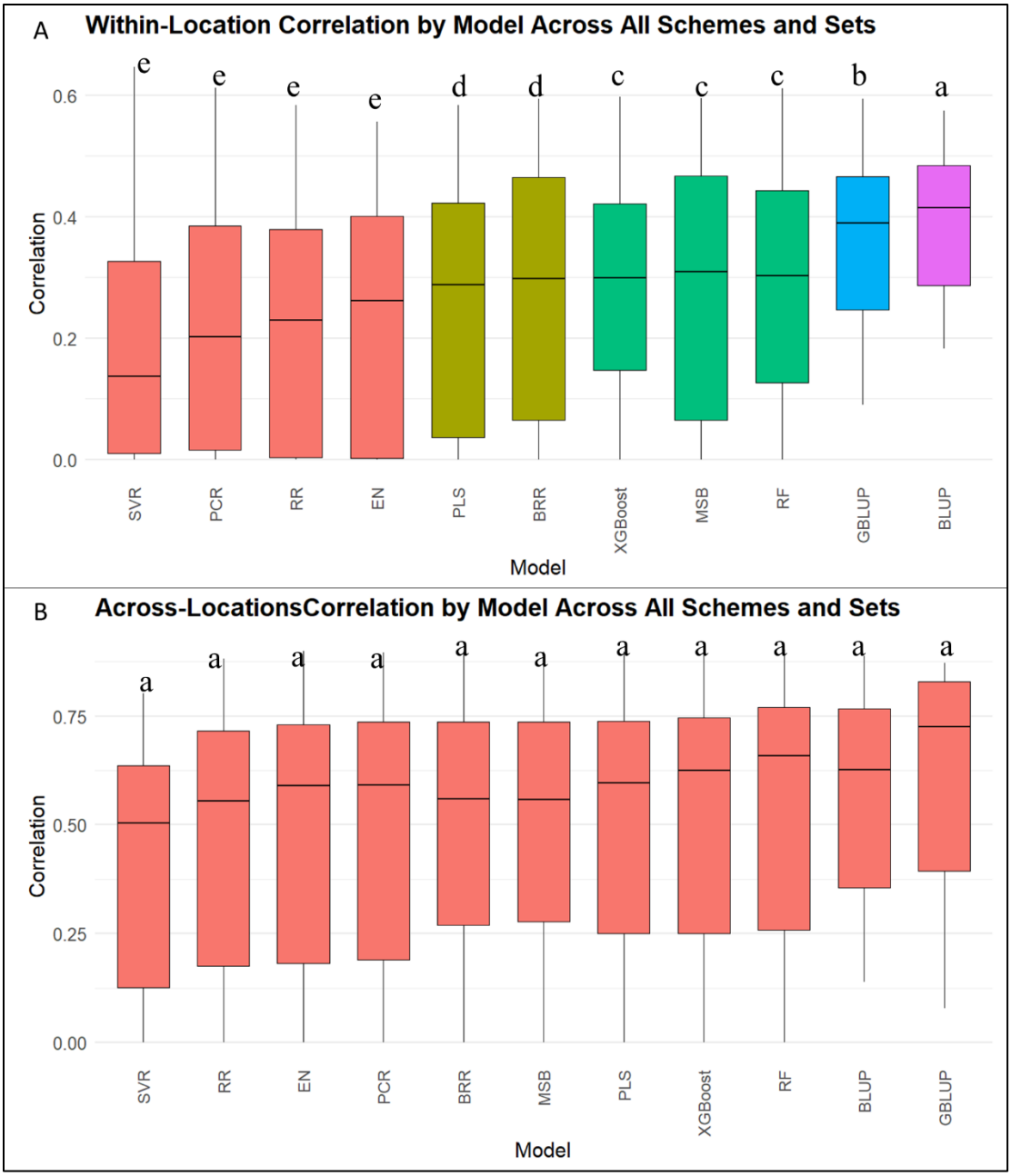

Figure 4.6. Model-specific prediction predictive ability (*r*) within locations and across locations across all validation schemes, predictors sets and traits. Boxplots summarize prediction accuracy (Pearson's correlation, r) across all traits, predictor sets, and validation schemes for each modeling approach. (A) Within-location correlations were computed within each location (environment) and summarized across all runs, reflecting the ability to rank genotypes within environments. (B) Across-location correlations were computed by pooling predictions across locations, reflecting performance for predicting and ranking genotypes across environments. Boxes represent the interquartile range (Q1–Q3) with the center line indicating the median; whiskers extend to 1.5×IQR. Box colors indicate Tukey HSD mean-separation groups ($\alpha = 0.05$); models sharing a color are not significantly different. Model acronyms: SVR: Support

Vector Regression; PCR: Principal Component Regression; RR: Ridge Regression EN: Elastic Net regression; PLS: Partial Least Squares regression; BRR: Bayesian Ridge Regression; XGBoost: Extreme Gradient Boosting; MSB: Multivariate SNP-Based; RF: Random Forest

Model-based variable-importance profiles confirmed that the top pipelines relied on different information sources depending on both the target trait and the cross-validation scheme (Figure 7). When secondary phenotypes were available, phenotypic covariates captured a large share of the total importance for most trait × scheme combinations, indicating that repeated plot-level scores provided highly informative for genetic and environmental effects. Genomic markers, soil, and weather covariates contributed smaller but non-negligible fractions, with their relative weight varying among traits and between the more lenient (CV00, CV01) and more stringent (LOLO, LGO, LOYO) schemes.

In contrast, when phenotypic predictors were omitted ("WithoutPheno", figure 7B), genomic markers became the dominant information source across traits and schemes, with weather (and to a lesser extent soil and NIR) covariates absorbing most of the remaining importance. This shift highlights that, in the absence of secondary phenotypes, the models rely primarily on marker-based relatedness and broad-scale environmental descriptors to recover predictive signal. Overall, Figure 5 shows that no single predictor group is universally optimal, the composition of the most influential variables is strongly trait and scheme-dependent, and adding secondary phenotypes chiefly reallocates importance away from genomic and environmental covariates toward directly measured plant performance traits.

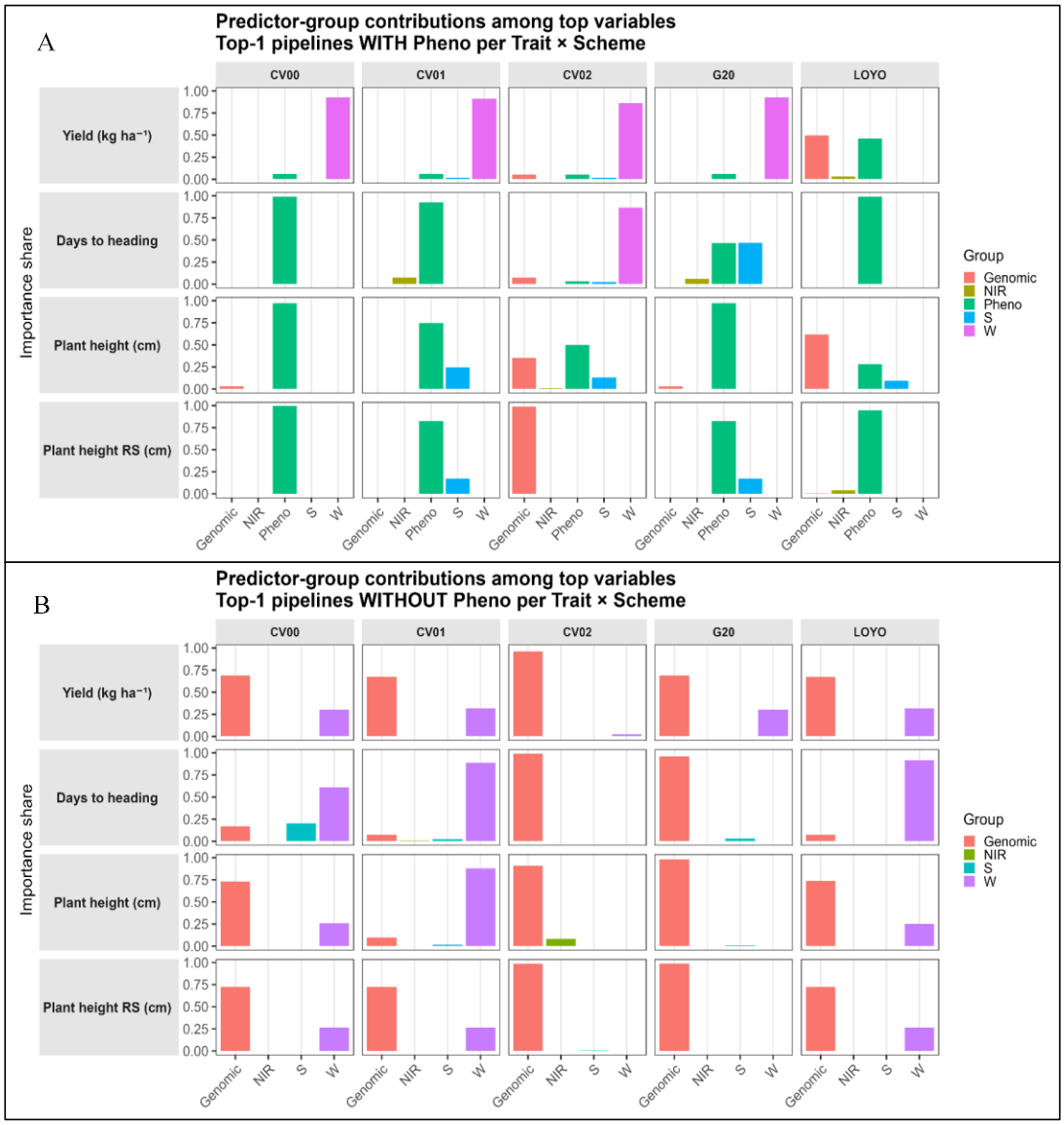

Figure 4.7. Relative contribution of genomic, environmental, NIR, and phenotypic predictor groups in the best pipelines across traits and cross-validation schemes. Share of total model-based variable importance attributed to each predictor group in the top-ranked pipeline for every trait × cross-validation combination. Panel A shows pipelines that include secondary phenotypes ("WithPheno"), whereas panel B shows pipelines that omit phenotypic predictors ("WithoutPheno"). Bars represent the proportion of summed importance associated with genomic markers, NIR features, phenotypic covariates, soil covariates (S), and weather covariates (W) among the most influential variables in each pipeline (see Methods for the variable-importance metric). Traits are grain yield, days to heading, plant height, and remote-sensed plant height (RS); columns correspond to CV00, CV01, LOLO, LGO, and LOYO.

Discussion and Conclusion

Reference to the heritability estimates in Table 3 helps explain part of the prediction pattern observed here, because narrow-sense heritability reflects the proportion of phenotypic variance attributable to additive genetic effects, which are the effects most directly captured by marker-based genomic models (Visscher et al., 2008). Nevertheless, predictive performance is not determined by heritability alone, as factors such as training population size, composition, and relatedness to the prediction set also influence model accuracy (Zhang et al., 2017). Thus, the improved performance of models incorporating additional omics or phenotypic layers likely reflects their ability to capture environmental effects and genotype-by-environment interactions that are not fully explained by additive genomic signal alone (Sandhu et al., 2021).

One of the clearest findings of this study is the contrast between within-environment and across-environment prediction performance. Across-site predictions yielded much higher accuracies than within-site predictions, indicating that models can capture broad genotype rank differences across locations more easily than fine-scale variation within a single environment. We also observed that the value of additional data layers depended on the scenario. In before-planting predictions (no in-season phenotypes), a simple genomic baseline model was often sufficient for most traits. However, grain yield was a notable exception, the baseline lost considerable accuracy on yield, which benefited greatly from integrating multi-layer data. In general, including in-season phenotypic covariates ("WithPheno" predictors) dramatically improved within-location accuracy, shifting correlations from near zero (with only soil or weather inputs) to significantly positive values. By contrast, across-location predictions were

overall higher (many multi-layer models achieved moderate to high median correlations, ~0.6–0.75) and showed a plateauing effect once phenotypic covariates were included. In fact, any predictor set containing in-season phenotypes tended to perform similarly well across sites, suggesting diminishing returns beyond a certain point. These trends highlight that multi-environment (across-site) predictions can leverage large environmental variance, whereas within-site predictions are inherently more challenging and rely on capturing indirect genetic signals – which is only feasible when phenotypes are available.

Yield exemplified a complex, low-heritability trait with strong environmental modulation, and thus showed the greatest gains from data integration and flexible modeling. In the CV00 scenario, pipelines that combined genomic markers with envirotyping and high-throughput phenotypic indices achieved very high predictive correlations outperforming the marker-only GBLUP baseline. This supports broader evidence that fusing genomics with environmental data and secondary phenotypes can markedly improve prediction accuracy for polygenic traits (Togninalli et al., 2023; Montesinos-López et al., 2023). The added value of UAV-derived vegetation indices is consistent with the concept of phenomic selection, as shown in prior wheat studies where spectral indices proxy growth and stress responses not captured by markers alone (Crain et al., 2018; Rutkoski et al., 2016). Likewise, the strong contribution of weather covariates aligns with previous reports that incorporating environmental descriptors improves genomic prediction under G×E (Heslot et al., 2014; Gillberg et al., 2019; Costa-Neto et al., 2021). Notably, in our yield predictions the best-performing models were those that included in-season phenotype measurements (e.g., mid-season canopy traits) along with weather - these consistently formed the top Tukey groups and surpassed genomic-only models. This finding

resonates with recent multi-modal wheat prediction studies that successfully integrated genomic and high-throughput phenotypic data for yield improvement (Togninalli et al., 2023), as well as with results in other crops demonstrating better prediction in untested environments when envirotyping data are included (Costa-Neto et al., 2021).

When the training set size was reduced (the CV01 scenario, only ~25% training data), predictive correlations dropped across all methods, as expected. Nonetheless, multi-source pipelines retained a notable advantage over single-source models, indicating that auxiliary data can compensate for limited training examples. In particular, regularized learning approaches that leveraged weather, NIR, and/or correlated traits remained among the top performers under CV01. This is consistent with empirical evidence that including correlated secondary information can offset sparse phenotypic training data (Crain et al., 2018; Montesinos-López et al., 2023). It also reflects the classic bias–variance tradeoff, in low-data regimes, simpler well-regularized models tend to excel by controlling overfitting, whereas overly complex models suffer (Friedman et al., 2010; Azodi et al., 2019). Indeed, under CV01 our simpler pipelines (e.g. ridge regression or Bayesian ridge with weather/NIR) performed comparably to more complex algorithms, underscoring that controlling model complexity was decisive when data were limited.

Extrapolating predictions beyond the training domain proved especially challenging, with accuracy generally declining in leave-one-out scenarios. When predicting into completely new locations (LOLO), models that explicitly incorporated environmental covariates outperformed genotype-only baselines. This reinforces that envirotyping can improve across-site prediction by capturing environmental similarities and genotype-environment interaction structure (Jarquín et

al., 2014; Gillberg et al., 2019; Millet et al., 2019). In LOYO, genomic-only models had very limited ability to anticipate year-specific yield shifts, whereas providing the model with target-year weather data improved predictions substantially. This finding highlights that temporal transferability of predictions depends on having environmental information describing the target growing season (Heslot et al., 2014; Roorkiwal et al., 2018). It aligns with emerging "forecasting" approaches that integrate climate or crop model outputs into genomic prediction to better anticipate performance under future or abnormal conditions (Watson et al., 2019; Zhang et al., 2022). The gains observed with including actual weather for the target year suggest that in practice, a hybrid approach (combining genomics with seasonal climate forecasts or real-time weather measurements) could enhance predictions for upcoming growing seasons - a strategy that will become increasingly relevant under climate variation (Watson et al., 2019).

Interestingly, when predicting entirely new genotypes that were unphenotyped in training (the LGO scenario), predictive ability was moderate and driven primarily by relatedness captured through the DNA markers. This outcome is expected from quantitative genetics theory - the accuracy of genomic prediction for new individuals relies on kinship linkage disequilibrium with the training set (Daetwyler et al., 2013). In this context, additional omics layers provided relatively smaller gains, since in-season phenotypes cannot be obtained for brand-new lines that have never been grown. Nonetheless, we did observe modest improvements from including seed NIR spectral data (which can be available pre-field), suggesting that certain compositional or seed-quality signals are correlated with yield potential, especially the fact that NIR is used to accurately predict grain protein content and protein content is negative correlated with yield (Ye, D., et al, 2028; Eichi, V. R. et al, 2020). This hints that "pre-field" omics can add value for

selection of new candidates, in line with recent efforts using other omics (e.g. transcriptomic or metabolomic profiles) to augment genomic predictions in the absence of phenotypes (Westhues et al., 2017). Ultimately, the limited accuracy in LGO emphasizes the importance of training population design and genetic relatedness for complex trait prediction (Norman et al., 2018). It also suggests that while multi-omics integration yields big gains when phenotypes are available, its benefit for completely new genotypes is constrained - reinforcing that genomic relationships remain the cornerstone of breeding value prediction for novel lines.

Another notable result was the comparison among different modeling approaches. Within a given scenario, we found significant performance differences, classic linear methods like BLUP and GBLUP achieved the highest median correlations for most traits, forming the top statistical group, whereas several more flexible ML models (e.g. SVR, PCR, regularized regression) showed lower medians. In contrast, across locations there were no statistically significant differences among models (all approaches fell into one shared Tukey group), although GBLUP/BLUP still tended to have slightly higher medians than others. In practical terms, this means that our baseline linear models performed as well as or better than advanced ML models for this dataset - especially for within-environment prediction. This agrees with prior benchmarking studies which often find that no single algorithm consistently outperforms GBLUP for genomic prediction of agronomic traits, likely due to the limited size of typical breeding data and the effectiveness of shrinkage in linear models (Azodi et al., 2019; Montesinos-López et al., 2023). The strong showing of BLUP/GBLUP also underscores the earlier bias-variance point, simpler models with fewer tuning parameters were robust in capturing the genetic signal, whereas more complex methods did not yield additional benefit

given the training data available and mostly captured environmental noise. Notably, even in the multi-layer data settings, the linear models were able to capitalize on the extra predictors effectively, narrowing the gap with black-box methods. Therefore, while incorporating diverse data layers can be advantageous, it does not necessarily require highly complex algorithms - well-regularized linear or kernel models can often integrate this information optimally.

From a breeding perspective, these findings illustrate when and how to exploit phenotypic data versus relying on baseline genomic predictions. In scenarios where in-season phenotypes can be collected - for example, using mid-season UAV imagery or early growth measurements in target environments - our results show that incorporating those phenotypic covariates yields substantial accuracy gains. This could be valuable for product placement decisions or late-stage selection, where breeders evaluate which lines are best suited for specific environments using real-time phenotypic indicators (Crain et al., 2018; Rutkoski et al., 2016). On the other hand, when making selections before field phenotyping (e.g. advancing lines based purely on genetic merit), one must rely on genomic estimated breeding values. In those early selection contexts, our study suggests that the simple GBLUP marker-based model (baseline) often suffices for many traits, with little added benefit from extra data layers - except in the case of highly complex traits like yield, where any additional informative data (environment descriptors, seed properties, etc.) can still improve predictions. This differentiation implies that breeding programs should tailor their prediction pipelines to the decision stage, leveraging rich multi-layer models with phenotypes for environment-specific recommendations and cultivar placement, versus using parsimonious genomic models for routine selection of untested lines (Norman et al., 2018). Overall, our study demonstrates that integrating genomics with high-

throughput phenotyping and envirotyping leads to the greatest predictive gains when genotype–environment interactions are strong, while also affirming the enduring strength of baseline genomic prediction models for capturing the core genetic signal. These insights can help breeders balance investment in phenotyping versus genotyping depending on the trait and use-case, ultimately improving genetic gain under variable environmental conditions (Montesinos-López et al., 2023). Traits derived from UAV flight using multispectral or RGB sensors provide an inexpensive, highly high-throughput phenotypic supplement for prediction, enabling breeders to focus more intensive phenotyping on traits and stages where the return on investment is greatest.

It is noteworthy that with phenotypic covariates models most of the times outperformed genomic only based models in our results, an outcome that has been reported in recent phenomic selection studies (Rutkoski et al., 2016; Adak et al., 2023; Roscher-Ehrig et al., 2024). We observed a similar trend; for certain traits and environments, a model using secondary phenotypic inputs achieved nearly the same predictive correlation as the full multi-omics model. This underscores that phenomic data capture critical environmental and developmental signals that pure genomic models lack. However, we caution that this does not mean genomic data are obsolete - rather, it highlights the context-dependent value of each data type. Genomic predictions target stable genetic merit (breeding values), whereas phenomic predictions capture current trait expression, including G×E and transient effects. Plus, although the sets of genotypes presenting in this study is very diverse, we recognize it is relatively small and with few years of training, which in a scenario with increase of diversity and years for training, GBLUP could have performed better, as well as the other models. Our findings support the view of Wang et al.

(2025) that one should not directly benchmark phenomic prediction against genomic prediction in terms of accuracy, because they are fundamentally predicting different quantities. As Wang, Feldmann, and Runcie argue, a high cross-validation r for a phenomic model reflects its ability to predict phenotypic performance (often exploiting environment-specific variation), whereas a genomic model's r reflects predicting genetic potential across environments. Thus, a phenomic model can appear "more accurate" in a naive comparison if the validation scheme allows it to leverage instantaneous environmental signals, even though that may not translate to stable genetic gain across cycles. In our study, random CV (CV00) likely overestimates the advantage of phenomics because training and test sets share the same environments - under those conditions, models including rich phenotypic data (e.g. remote sensing) can essentially "memorize" environmental variation to boost correlation. But in LOLO, LOYO, or GLO scenarios where we do not use Pheno, the genomic component becomes more important for maintaining accuracy, since phenomic data from one environment may not directly transfer to another. The key lesson is that genomic and phenomic approaches are complementary. Rather than asking which is superior in accuracy, breeders should ask how to optimally combine them to improve selection decisions (Wang et al., 2025). Our results show that an integrated model leveraging both provides the most robust predictions overall - genomic markers ensure the model captures stable genetic effects, while phenotypic covariates inject additional information about genotype performance under specific conditions, and together they often produce the highest accuracies.

Across traits and validation scenarios, our results demonstrate that the payoff of multi-omics and ML is strongly trait- and context-dependent favoring more product placement. For

grain yield, integrating genomics, envirotyping, and high-throughput phenotypes substantially improved prediction - particularly in extrapolation and data-sparse regimes - supporting the "integrative genomic prediction" paradigm (Crossa et al., 2017; Montesinos-López et al., 2023). These patterns are consistent with recent work showing that multi-modal methods can exploit heterogeneous inputs to boost complex-trait prediction in wheat (Togninalli et al., 2023; McBreen et al., 2025). Conceptually, our approach parallels reaction-norm modeling of G×E (Jarquín et al., 2014), and the practical implication is clear, environment-informed prediction can reduce "surprises" when conditions shift and can support sparse testing strategies in METs (Jarquín et al., 2020). The LOYO results suggest that incorporating climate descriptors - and ultimately forecast- or simulation-informed covariates - may help breeding decisions remain robust under changing climate (Watson et al., 2019; Zhang et al., 2022). In contrast, traits with simpler architecture (PLTHT) achieved high predictive ability with standard genomic models, and the marginal gains from additional layers were small. HDDAT occupied an intermediate position, genomic prediction was already strong, with multi-omics primarily improving robustness to unusual environments/years. This supports a resource-allocation strategy where multi-omics investment is prioritized for traits with strong G×E, lower heritability, or high value for early indirect selection (Crossa et al., 2025).

Finally, variable-importance patterns provided interpretable, biologically sensible signals, strengthening confidence that improved performance came from meaningful predictors rather than artifacts - an important requirement for breeder adoption (Azodi et al., 2020). Looking forward, more flexible data-fusion approaches and ensemble stacking may further improve performance while preserving interpretability (Ma et al., 2018; Wang et al., 2024; Montesinos-

López et al., 2023; Sagi & Rokach, 2021). Extending this framework to additional omics and to genotype-by-management predictors (including crop-model outputs) is a promising direction to further narrow prediction gaps (Heslot et al., 2014; Cooper et al., 2021). Our findings reinforce that multi-omics + ML can meaningfully "climatize" GS for complex traits like yield, enabling faster and more resilient breeding decisions as environmental variability increases and product placement is necessary (Watson et al., 2019; Zhang et al., 2022). Our results show that the biggest gains come from matching model and data complexity to trait biology and deployment context, using integrated, environment-aware prediction where it is needed most to make earlier, faster, and more reliable breeding decisions.

References

Adak, A., Murray, S. C., & Anderson, S. L. (2023). Temporal phenomic predictions from unoccupied aerial systems can outperform genomic predictions. G3: Genes, Genomes, Genetics, 13(1), jkac294.

Aparicio, J. (2022). MrBean: Web application for analyzing field experiments. R package version, 2(9).

Azodi, C. B., Bolger, E., McCarren, A., Roantree, M., de los Campos, G., & Shiu, S.-H. (2019). Benchmarking parametric and machine learning models for genomic prediction of complex traits. G3: Genes, Genomes, Genetics, 9(11), 3691–3702.

Azodi, C. B., Tang, J., & Shiu, S.-H. (2020). Opening the black box: interpretable machine learning for geneticists. Trends in Genetics, 36(6), 442–455.

Barreto, C. A. V., das Graças Dias, K. O., de Sousa, I. C., Azevedo, C. F., Nascimento, A. C. C., Guimarães, L. J. M., ... & Nascimento, M. (2024). Genomic prediction in multi-environment trials in maize using statistical and machine learning methods. Scientific Reports, 14(1), 1062.

Battenfield, S. D., Guzmán, C., Gaynor, R. C., Singh, R. P., Peña, R. J., Dreisigacker, S., ... & Poland, J. (2016). Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. The Plant Genome, 9(2), 1–12.

Ben Hassen, M., Cao, T. V., Bartholome, J., Orasen, G., Colombi, C., Rakotomalala, J., ... & Ahmadi, N. (2018). Rice diversity panel provides accurate genomic predictions for complex traits in the progenies of biparental crosses involving members of the panel. Theoretical and Applied Genetics, 131(2), 417-435.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) (pp. 785–794).

Cooper, M., Voss-Fels, K. P., Messina, C. D., Tang, T., & Hammer, G. L. (2021). Tackling G× E× M interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. Theoretical and Applied Genetics, 134(6), 1625-1644.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

Costa-Neto, G., Galli, G., Carvalho, H. F., Crossa, J., & Fritsche-Neto, R. (2021). EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. G3: Genes, Genomes, Genetics, 11(4), jkab040.

Crain, J., Mondal, S., Rutkoski, J., Singh, R. P., & Poland, J. (2018). Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. The plant genome, 11(1), 170043.

Crossa, J., Martini, J. W. R., Vitale, P., Pérez-Rodríguez, P., Costa-Neto, G., et al. (2025). Expanding genomic prediction in plant breeding: harnessing big data, machine learning, and advanced software. Trends in Plant Science, 30(7), 756–774.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., ... & Burgueño, J. (2017). Genomic selection in plant breeding: methods, models, and perspectives. Trends in Plant Science, 22(11), 961–975.

Daetwyler, H. D., Calus, M. P., Pong-Wong, R., de Los Campos, G., & Hickey, J. M. (2013). Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics, 193(2), 347-365.

de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., and Calus, M. P. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics, 193(2), 327-345.

Desta, Z. A., & Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. Trends in Plant Science, 19(9), 592–601.

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., Weingessel, A., & Leisch, M. F. (2006). The e1071 package. Misc Functions of Department of Statistics (e1071), TU Wien, 297.

Eichi, V. R., Okamoto, M., Garnett, T., Eckermann, P., Darrier, B., Riboni, M., & Langridge, P. (2020). Strengths and Weaknesses of National Variety Trial Data for Multi-Environment Analysis: A Case Study on Grain Yield and Protein Content. Agronomy, 10(5), 753.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33(1), 1–22.

Gillberg, J., Marttinen, P., Mamitsuka, H., & Kaski, S. (2019). Modeling G×E with historical weather information improves genomic prediction in new environments. BioRxiv, 213231 (preprint).

González-Camacho, J. M., de Los Campos, G., Pérez-Rodríguez, P., Ozdemir, D., Dreisigacker, S., & Crossa, J. (2018). Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. The Plant Genome, 11(2), 170104.

González-Recio, O., & Forni, S. (2011). Genome-wide prediction of discrete traits using Bayesian regressions and machine learning. Genetics Selection Evolution, 43(1), 7.

González-Recio, O., Rosa, G. J. M., & Gianola, D. (2013). Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. Livestock Science, 166, 217–231.

Guo, Z., Tucker, D. M., Wang, D., Basten, C. J., Ersoz, E., Briggs, W. H., Lu, J., Li, M., & Gay, G. (2013). Accuracy of across-environment genome-wide prediction in maize nested association mapping populations. G3: Genes, Genomes, Genetics, 3(2), 263–272.

Heffner, E. L., Sorrells, M. E., & Jannink, J. L. (2009). Genomic selection for crop improvement. Crop Science, 49(1), 1–12.

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(2), 463–480.

Heslot, N., Yang, H.-P., Sorrells, M. E., & Jannink, J.-L. (2012). Genomic selection in plant breeding: a comparison of models. Crop Science, 52(1), 146–160.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.

Jarquin, D., Howard, R., Crossa, J., Beyene, Y., Gowda, M., Martini, J. W., ... & Prasanna, B. M. (2020). Genomic prediction enhanced sparse testing for multi-environment trials. G3: Genes, Genomes, Genetics, 10(8), 2725–2739.

Jarquín, D., Crossa, J., Lacaze, X., Duffaud, M., Tourrette, Y., Lorgeou, J., ... & de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theoretical and Applied Genetics, 127(3), 595–607.

Jonas, E., & de Koning, D.-J. (2013). Does genomic selection have a future in plant breeding? Trends in Biotechnology, 31(9), 497–504.

Jubair, S., & Domaratzki, M. (2023). Crop genomic selection with deep learning and environmental data: A survey. Frontiers in Artificial Intelligence, 5, 1040295.

Krause, M. D., Dias, K. O. G., de Santos, J. P. R., Oliveira, A. A., Guimarães, L. J. M., Pastina, M. M., ... & Garcia, A. A. F. (2020). Boosting predictive ability of tropical maize hybrids in sparse testing networks via genotype × environment interaction models. Crop Science, 60(6), 3049–3065.

Krause, M. R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R. P., ... & Mondal, S. (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. G3: Genes, Genomes, Genetics, 9(4), 1231-1247.

Lell, M., Reif, J., & Zhao, Y. (2021). Optimizing the setup of multi-environmental hybrid wheat yield trials for boosting the selection capability. The Plant Genome, 14(3), e20150.

Ma, W., Qiu, Z., Song, J., Li, J., Cheng, Q., Zhai, J., ... & Zhang, W. (2018). A deep convolutional neural network approach for predicting phenotypes from genotypes. * Planta*, 248(5), 1307–1318.

McBreen, J., Babar, M. A., Jarquin, D., Ampatzidis, Y., Khan, N., Kunwar, S., ... & Brown-Guedira, G. (2025). Leveraging Multi-Omics Data with Machine Learning to Predict Grain Yield in Small vs. Big Plot Wheat Trials. Agronomy, 15(6), 1315.

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.

Mevik, B.-H., & Wehrens, R. (2007). The pls package: principal component and partial least squares regression in R. Journal of Statistical Software, 18(2), 1–24.

Millet, E. J., Kruijer, W., Blanc, L.,Journet, E.-P., Reinhold, M., Pletcher, L., ... & Vile, D. (2019). Genomic prediction of maize yield across European environmental conditions. Nature Genetics, 51(6), 952–956.

Montesinos-López, A., Rivera, C., Pinto, F., Piñera, F., Gonzalez, D., Reynolds, M., ... & Crossa, J. (2023). Multimodal deep learning methods enhance genomic prediction of wheat breeding. G3: Genes, Genomes, Genetics, 13(5), jkad045.

Montesinos-López, O. A., Montesinos-López, A., Luna-Vázquez, F. J., Toledo, F. H., Pérez-Hernández, O., Eskridge, K. M., & Crossa, J. (2019). An R package for Bayesian analysis of multi-environment and multi-trait multi-output genomic models. G3: Genes, Genomes, Genetics, 9(5), 1355–1369.

Moro Rosso, L. H., de Borja Reis, A. F., Correndo, A. A., & Ciampitti, I. A. (2021). XPolaris: an R-package to retrieve United States soil data at 30-meter resolution. BMC Research Notes, 14(1), 327.

Murillo, D. A., Gezan, S. A., Heilman, A. M., Walk, T. C., Aparicio, J. S., & Horsley, R. D. (2021). FielDHub: A shiny app for design of experiments in life sciences.

Norman, A., Taylor, J., Edwards, J., & Kuchel, H. (2018). Optimising genomic selection in wheat: Effect of marker density, population size and population structure on prediction accuracy. G3: Genes, Genomes, Genetics, 8(9), 2889–2899.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Resende, R. T., Xavier, A., Silva, P. I. T., Resende, M. P., Jarquin, D., & Marcatti, G. E. (2025). GIS-based G× E modeling of maize hybrids through enviromic markers engineering. New Phytologist, 245(1), 102-116.

Roorkiwal, M., Jarquin, D., Singh, M. K., Gaur, P. M., Bharadwaj, C., Rathore, A., ... & Varshney, R. K. (2018). Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotype × environment interaction on prediction accuracy in chickpea. Scientific Reports, 8(1), 11701.

Roscher-Ehrig, L., Weber, S. E., Abbadi, A., Malenica, M., Abel, S., Hemker, R., Snowdon, R. J., Wittkop, B., & Stahl, A. (2024). Phenomic selection for hybrid rapeseed breeding. Plant Phenomics, 6, 0215.

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., Crossa, J., ... & Singh, R. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. G3: Genes, Genomes, Genetics, 6(9), 2799-2808.

Sagi, O., & Rokach, L. (2021). Approximating XGBoost with an interpretable decision tree. Information sciences, 572, 522-542.

Sandhu, K. S., Mihalyov, P. D., Lewien, M. J., Pumphrey, M. O., & Carter, A. H. (2021). Combining genomic and phenomic information for predicting grain protein content and grain yield in spring wheat. Frontiers in Plant Science, 12, 613300.

Sandhu, K. S., Aoun, M., Morris, C. F., & Carter, A. H. (2021). Genomic selection for end-use quality and processing traits in soft white winter wheat breeding program with machine and deep learning models. Biology, 10(7), 689.

Sparks, A. H. (2018). nasapower: A NASA POWER global meteorology, surface solar energy and climatology data client for R. Journal of Open Source Software, 3(30), 1035.

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., ... & McCouch, S. (2015). Genomic selection and association mapping in rice (Oryza sativa): a comparison of models including genotype × environment interaction. The Plant Genome, 8(1), eplantgenome2014.04.0018.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288.

Togninalli, M., Wang, X., Kucera, T., Shrestha, S., Juliana, P., Mondal, S., ... & Poland, J. (2023). Multi-modal deep learning improves grain yield prediction in wheat breeding by fusing genomics and phenomics. Bioinformatics, 39(6), btad336.

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

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era: Concepts and misconceptions. Nature Reviews Genetics, 9(4), 255–266.

Wang, F., Feldmann, M. J., & Runcie, D. E. (2025). Do not benchmark phenomic prediction against genomic prediction accuracy. The Plant Phenome Journal, 8, e70029.

Wang, P., Lehti-Shiu, M. D., Lotreck, S., Segura Abá, K., Krysan, P. J., & Shiu, S. H. (2024). Prediction of plant complex traits via integration of multi-omics data. Nature Communications, 15(1), 6856.Watson, A., Hickey, L. T., Christopher, J., Rutkoski, J., Poland, J., & Hayes, B. J. (2019). Multivariate genomic selection and potential of rapid indirect selection with speed breeding in spring wheat. Crop Science, 59(5), 1945-1959.

Werner, C. R., Gaynor, R. C., Gorjanc, G., Hickey, J. M., Kox, T., Abbadi, A., Leckband, G., Snowdon, R. J., & Stahl, A. (2020). How population structure impacts genomic selection accuracy in cross-validation: Implications for practical breeding. Frontiers in Plant Science, 11, 592977.

Westhues, M., Schrag, T. A., Heuer, C., Thaller, G., Utz, H. F., Schön, C.-C., & Peter Maurer, H. (2017). Omics-based hybrid prediction in maize. Theoretical and Applied Genetics, 130(9), 1927–1939.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58(2), 109–130.

Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software, 77(1), 1–17.

Xavier, A. (2019). Efficient estimation of marker effects in plant breeding. G3: Genes, Genomes, Genetics, 9(11), 3855-3866.

Xavier, A., Muir, W. M., & Rainey, K. M. (2020). bWGR: A Bayesian whole-genome regression package for advanced genomic prediction. Bioinformatics, 36(8), 1466–1468.

Xavier, A., Runcie, D., & Habier, D. (2025). Megavariate methods capture complex genotype-by-environment interactions. Genetics, 229(4), iyae179.

Ye, D., Sun, L., Zou, B., Zhang, Q., Tan, W., & Che, W. (2018). Non-destructive prediction of protein content in wheat using NIRS. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 189, 463-472.

Zhang, T., He, Y., DePauw, R., Jin, Z., Garvin, D., Yue, X., ... & Yang, X. (2022). Climate change may outpace current wheat breeding yield improvements in North America. Nature Communications, 13(1), 5591.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320.