

Improvement of quantitative traits in elite wheat breeding germplasm through introgression from
Triticum dicoccoides via a wheat-*Aegilops tauschii* synthetic octoploid bridge

by

Addison Hope Carroll

B.S., Colorado State University, 2019

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics Program
Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Abstract

Wild relatives of wheat provide an important reservoir of allelic diversity for crop improvement, but their use in elite breeding germplasm is constrained by reduced agronomic adaptation, persistent yield disadvantages, and the difficulty of recovering favorable recombinant deviates for quantitative traits. Wild emmer (*Triticum dicoccoides*), the progenitor of domesticated tetraploid and hexaploid wheats, carries extensive diversity for disease resistance, abiotic stress response, functional quality, phenology, and yield component traits. However, systematic introgression of wild emmer diversity into adapted hexaploid wheat backgrounds remains challenging, particularly when the breeding objective is improvement of complex quantitative traits rather than transfer of major effect loci. This project evaluated the use of a wheat-*Aegilops tauschii* synthetic octoploid amphiploid bridge as a strategy for introducing allelic diversity from wild emmer into elite winter wheat germplasm while retaining regional adaptation to Kansas production environments.

Seven BC₂-derived *Triticum dicoccoides* germplasm development (TdGD) populations were generated using two adapted hexaploid recurrent parents, two *Aegilops tauschii* accessions, and seven wild emmer accessions representing geographically diverse sources of A and B genome variation. High-density sequencing data from population founders were used to perform pedigree-based imputation into progeny with low-density SNP data to characterize genome-wide introgression architecture and detect donor segments. Across populations, donor introgression was successfully recovered on all three wheat subgenomes, with introgressions enriched in distal, high recombination regions and reduced in pericentromeric intervals. Chromosome-specific patterns were influenced by founder divergence, marker density, and recombination landscape. Population structure was driven primarily by recurrent parent background and variation in donor

allele retention. These results demonstrate that the amphiploid bridge strategy can generate wild introgression germplasm and allow for quick recovery of an adapted hexaploid wheat background.

To evaluate the breeding potential of this germplasm, one population, TdGD Pop80, was phenotyped across six Kansas field environments for agronomic, yield component, functional quality, and ionomic traits. Mixed model analyses were used to estimate within- and across-environment performance, trait relationships, and repeatability. Despite the observation that wild introgression broadly reduced agronomic performance, Pop80 exhibited substantial quantitative variation and included lines with favorable combinations of grain yield, kernel traits, protein quality, and elemental composition. Trait relationships reflected both expected physiological tradeoffs and opportunities for selection, including relationships among phenology, vegetative and reproductive performance under drought, grain yield, kernel size, grain protein, functional quality, and ionomic traits. Genome-wide association analyses identified marker-trait associations across trait groups, including loci corresponding to known developmental, quality, kernel size, and ionomic regions, as well as additional regions requiring validation. Overlaying association signals with donor introgression profiles provided a framework for distinguishing candidate wild-derived regions and for prioritizing loci for future breeding and genetic studies.

A yield-weighted multi-trait selection index was constructed to identify competitive germplasm. Two Pop80 wild introgression lines, KSTdGD80-55 and KSTdGD80-175, were selected for germplasm release based on their favorable agronomic performance, yield potential, quality traits, and characterized introgression profiles. These lines represent adapted winter wheat germplasm carrying defined introgressions from wild emmer and *Aegilops tauschii*, and they provide useful material for continued evaluation of wild-derived alleles in breeding-relevant

backgrounds. Together, this research demonstrates that bridge-mediated introgression can broaden the allelic base of elite wheat germplasm while maintaining sufficient agronomic performance for selection. The resulting genomic resources, phenotypic datasets, marker-trait associations, and released germplasm provide a foundation for future efforts to refine wild introgressions, validate candidate loci, and deploy favorable wild alleles for quantitative trait improvement in wheat breeding programs.

Improvement of quantitative traits in elite wheat breeding germplasm through introgression from
Triticum dicoccoides via a wheat-*Aegilops tauschii* synthetic octoploid bridge

by

Addison Hope Carroll

B.S., Colorado State University, 2019

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics Program
Department of Agronomy
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Approved by:

Major Professor
Dr. Allan Fritz

Copyright

© Addison Carroll 2026

Abstract

Wild relatives of wheat provide an important reservoir of allelic diversity for crop improvement, but their use in elite breeding germplasm is constrained by reduced agronomic adaptation, persistent yield disadvantages, and the difficulty of recovering favorable recombinant deviates for quantitative traits. Wild emmer (*Triticum dicoccoides*), the progenitor of domesticated tetraploid and hexaploid wheats, carries extensive diversity for disease resistance, abiotic stress response, functional quality, phenology, and yield component traits. However, systematic introgression of wild emmer diversity into adapted hexaploid wheat backgrounds remains challenging, particularly when the breeding objective is improvement of complex quantitative traits rather than transfer of major effect loci. This project evaluated the use of a wheat-*Aegilops tauschii* synthetic octoploid amphiploid bridge as a strategy for introducing allelic diversity from wild emmer into elite winter wheat germplasm while retaining regional adaptation to Kansas production environments.

Seven BC₂-derived *Triticum dicoccoides* germplasm development (TdGD) populations were generated using two adapted hexaploid recurrent parents, two *Aegilops tauschii* accessions, and seven wild emmer accessions representing geographically diverse sources of A and B genome variation. High-density sequencing data from population founders were used to perform pedigree-based imputation into progeny with low-density SNP data to characterize genome-wide introgression architecture and detect donor segments. Across populations, donor introgression was successfully recovered on all three wheat subgenomes, with introgressions enriched in distal, high recombination regions and reduced in pericentromeric intervals. Chromosome-specific patterns were influenced by founder divergence, marker density, and recombination landscape. Population structure was driven primarily by recurrent parent background and variation in donor

allele retention. These results demonstrate that the amphiploid bridge strategy can generate wild introgression germplasm and allow for quick recovery of an adapted hexaploid wheat background.

To evaluate the breeding potential of this germplasm, one population, TdGD Pop80, was phenotyped across six Kansas field environments for agronomic, yield component, functional quality, and ionomic traits. Mixed model analyses were used to estimate within- and across-environment performance, trait relationships, and repeatability. Despite the observation that wild introgression broadly reduced agronomic performance, Pop80 exhibited substantial quantitative variation and included lines with favorable combinations of grain yield, kernel traits, protein quality, and elemental composition. Trait relationships reflected both expected physiological tradeoffs and opportunities for selection, including relationships among phenology, vegetative and reproductive performance under drought, grain yield, kernel size, grain protein, functional quality, and ionomic traits. Genome-wide association analyses identified marker-trait associations across trait groups, including loci corresponding to known developmental, quality, kernel size, and ionomic regions, as well as additional regions requiring validation. Overlaying association signals with donor introgression profiles provided a framework for distinguishing candidate wild-derived regions and for prioritizing loci for future breeding and genetic studies.

A yield-weighted multi-trait selection index was constructed to identify competitive germplasm. Two Pop80 wild introgression lines, KSTdGD80-55 and KSTdGD80-175, were selected for germplasm release based on their favorable agronomic performance, yield potential, quality traits, and characterized introgression profiles. These lines represent adapted winter wheat germplasm carrying defined introgressions from wild emmer and *Aegilops tauschii*, and they provide useful material for continued evaluation of wild-derived alleles in breeding-relevant

backgrounds. Together, this research demonstrates that bridge-mediated introgression can broaden the allelic base of elite wheat germplasm while maintaining sufficient agronomic performance for selection. The resulting genomic resources, phenotypic datasets, marker-trait associations, and released germplasm provide a foundation for future efforts to refine wild introgressions, validate candidate loci, and deploy favorable wild alleles for quantitative trait improvement in wheat breeding programs.

Table of Contents

List of Figures	xv
List of Tables	xix
Acknowledgements	xxii
Dedication	xxiv
Chapter 1 - Literature review	1
Wild emmer as a resource for allelic diversity	1
The origins of tetraploid and hexaploid wheats	2
The importance of increasing wheat allelic diversity	6
<i>T. dicoccoides</i> geographical and ecological adaptation	7
Wild emmer as a resource for phenotypic diversity	9
Wild emmer as a resource for improving yield and physiological components	10
Wild emmer as a resource for disease resistance variation	12
Powdery mildew	13
Leaf rust	14
Stripe rust	18
<i>Fusarium</i> head blight	22
Wild emmer as a resource for improvement of functional quality	25
Wild emmer and flowering time variation across vernalization requirements	28
Wild emmer as a reservoir of abiotic stress tolerance	30
Historic prioritization of qualitative traits in introgression breeding	33
Evaluation of approaches to introgression breeding	37
Conclusion	41
Chapter 2 - Genomic signals of introgression using a wheat- <i>Aegilops tauschii</i> amphiploid bridge to bring allelic diversity from wild emmer into adapted hexaploid wheat	43
Abstract	43
Introduction	44
Methods	48
Population development	48
Array-based genotyping	51

Quality control and population structuring analysis	52
Sequencing-based genotyping	54
Nucleotide diversity between population founders.....	57
Preparation of sequence- and array-based genotypes for imputation	58
Pedigree-based imputation of sparsely genotyped progeny.....	59
Introgression mapping in imputed progeny	62
Results.....	64
Array-based genotyping.....	64
Sequencing-based genotyping	68
Allelic differentiation between TdGD populations and founders	72
PCA of progeny genotypes	73
Nucleotide diversity between population founders.....	77
Pedigree-based imputation performance.....	81
Mapping introgression segments in imputed WILs	85
Discussion	94
Implications of SNP array-based genotyping of introgression germplasm	94
Impact of sequencing methodology on founder variant detection.....	97
Performance of pedigree-based imputation for introgression detection in BC ₂ populations	98
Determinants of population structure in BC ₂ introgression germplasm	102
Comparison between amphiploid bridge and direct crossing approaches	106
Conclusion	109
Chapter 3 - Quantitative phenotypic evaluation of wild emmer introgression lines developed	
using an amphiploid bridge for breeding germplasm improvement.....	111
Abstract	111
Introduction.....	112
Methods	115
Population development.....	115
Seed increase and population selection.....	116
Trialing and phenotypic data collection.....	117
Phenological measurements – heading date, senescence date, grain fill period.....	119
Plant height	119

Tiller count.....	119
Drone-based measurements – NDVI, green reflectance	120
Grain yield	120
Test weight.....	121
Marvin Kernel Analyzer – kernel width, length, area, TKW	121
Grain protein	121
Grain protein deviation	122
Lactic acid SDS-SRC – protein quality score, solvent retention capacity.....	122
Standardization of phenotypic data.....	123
Within-environment analysis of phenotypic data	124
Across-environment analysis of phenotypic data	124
Repeatability calculations	126
Assessment of trait relationships	128
SNP genotyping and GWAS.....	129
Results.....	131
Characterization of trialing environments.....	131
Phenotypic evaluations of TdGD Pop80.....	133
Agronomic performance – across environments	134
Agronomic performance – within environments	136
Yield and kernel components – across environments.....	139
Yield and kernel components – within environments.....	143
Functional quality – across environments.....	145
Functional quality – within environments	150
Marker-trait associations in TdGD Pop80	152
Agronomic performance	153
Yield and kernel components.....	157
Functional quality	165
Discussion.....	170
Feasibility of maintaining adaptive performance in bridge-mediated wheat introgression breeding germplasm.....	172

Candidate loci for marker-trait associations identified under an exploratory analytical framework	176
Future directions for hypothesis testing and application in breeding germplasm improvement	183
Conclusion	187
Chapter 4 - Grain ionomic profiling of wild emmer introgression lines developed using an amphiploid bridge.....	190
Abstract.....	190
Introduction.....	191
Methods	197
Population development and field trialing.....	197
Grain ionomic profiling	198
Phenotypic analysis of elemental concentrations	199
Assessment of trait relationships	201
Genome-wide association analysis and introgression overlap.....	202
Results.....	203
Phenotypic evaluation.....	203
Across environments.....	203
Within environments.....	206
Multivariate relationships among traits	210
Genome-wide association	213
Discussion.....	218
Impact of wild emmer introgression on grain ionomic variation.....	218
Integration of the grain ionome with broader phenotypic performance	220
Marker-trait associations for the grain ionome and colocalizations with related traits	222
Limitations and complications for interpretation.....	226
Breeding implications and conclusions.....	229
Chapter 5 - Multi-trait index selection of amphiploid bridge-derived wild emmer introgression lines for wheat improvement	231
Abstract.....	231
Introduction.....	232

Methods	236
Germplasm development	236
Field evaluation and selection.....	237
Phenotypic characterization	238
Statistical analysis	239
Grain yield by trait selection index	240
Genotypic characterization and introgression profiling.....	241
Results.....	242
Grain yield by trait selection index	243
Characterization of KSTdGD80-175 and KSTdGD80-55.....	245
Agronomic performance	245
Yield and kernel components.....	246
Functional quality	247
Ionomics.....	248
Genomic introgression	249
Discussion.....	254
Multi-trait selection translated introgression-derived variation into breeding utility	254
Population size influenced the opportunity to recover rare favorable recombinants.....	257
Selected introgression lines provide complementary breeding resources	258
Characterized introgressions increase the value of germplasm releases.....	259
Conclusion	261
Availability	261
References.....	262
Appendix A - Supplementary tables and figures for Chapter 2.....	305
Appendix B - Supplementary tables and figures for Chapter 3	320
Appendix C - Supplementary tables and figures for Chapter 4	359
Appendix D - Supplementary tables for Chapter 5.....	368

List of Figures

Figure 1.1. Schematic diagram of wheat evolution in terms of time, threshing ability, and ploidy level.....	5
Figure 2.1. Schematic overview of the <i>Triticum dicoccoides</i> germplasm development pipeline	49
Figure 2.2. Location of origin for accessions used in <i>Triticum dicoccoides</i> germplasm development populations	50
Figure 2.3. Distribution of MAF across markers for each TdGD population.....	67
Figure 2.4. Principal component analysis of WIL progeny across TdGD populations	74
Figure 2.5. SNP-PC correlation plots for the first two PCs derived from population structure analyses of TdGD WILs	75
Figure 2.6. Pairwise nucleotide diversity between population founders across TdGD populations, plotted by genomic position and faceted by chromosome.....	79
Figure 2.7. Distributions of donor introgression segment lengths in TdGD WILs	88
Figure 2.8. Population averaged donor introgression profiles across the genome	91
Figure 3.1. Schematic of the experimental design for TdGD Pop80 trials conducted from 2020 to 2023.....	118
Figure 3.2. Ten-day precipitation totals from March 1 through harvest across TdGD testing environments	132
Figure 3.3. Distributions of agronomic performance traits in TdGD Pop80 within and across environments.....	135
Figure 3.4. Distributions of yield and kernel component traits in TdGD Pop80 within and across environments.....	141
Figure 3.5. Distributions of functional quality traits in TdGD Pop80 within and across environments	147
Figure 3.6. Pearson correlations among agronomic performance, yield and kernel component, functional quality, and extended drone data BLUPs in TdGD Pop80 WILs.....	149
Figure 3.7. GWAS Manhattan plots for HD and SD BLUPs overlaid with mean donor proportions in TdGD Pop80.....	154
Figure 3.8. GWAS Manhattan plot for PH BLUPs overlaid with mean donor proportions in TdGD Pop80	155

Figure 3.9. GWAS Manhattan plots for GR and NDVI BLUPs overlaid with mean donor proportions in TdGD Pop80.....	157
Figure 3.10. GWAS Manhattan plots for kernel size BLUPs (KL, KW, KA, TKW) overlaid with mean donor proportions in TdGD Pop80.....	162
Figure 3.11. GWAS Manhattan plots for TW BLUPs overlaid with mean donor proportion in TdGD Pop80	163
Figure 3.12. GWAS Manhattan plot for GY BLUPs overlaid with mean donor proportion in TdGD Pop80	164
Figure 3.13. GWAS Manhattan plot for TC BLUPs overlaid with mean donor proportion in TdGD Pop80	165
Figure 3.14. GWAS Manhattan plot for GP BLUPs overlaid with mean donor proportion in TdGD Pop80	166
Figure 3.15. GWAS Manhattan plots for SRC and PQS BLUPs overlaid with mean donor proportions in TdGD Pop80.....	167
Figure 4.1. Distributions of ionomic traits in TdGD Pop80 within and across environments ...	210
Figure 4.2. Pearson correlations among BLUPs for ionomics traits in TdGD Pop80 WILs.....	211
Figure 4.3. GWAS Manhattan plots for vascular transport proxy elements, overlaid with mean donor proportions in TdGD Pop80	215
Figure 4.4. GWAS Manhattan plots for trace elements, overlaid with mean donor proportions in TdGD Pop80	215
Figure 4.5. GWAS Manhattan plots for metabolic micronutrients, overlaid with mean donor proportions in TdGD Pop80.....	216
Figure 4.6. GWAS Manhattan plot for Cd BLUPs overlaid with mean donor proportion in TdGD Pop80	217
Figure 4.7. GWAS Manhattan plots for structural macronutrients, overlaid with mean donor proportions in TdGD Pop80.....	217
Figure 5.1. Distribution of the BLUP GYT index among TdGD Pop80 WILs.....	243
Figure 5.2. Multi-trait profiles of KSTdGD80-175 and KSTdGD80-55.....	245
Figure 5.3. Genome-wide introgression profile for KSTdGD80-55.....	250
Figure 5.4. Genome-wide introgression profile for KSTdGD80-175.....	251

Appendix Figure A.1. Representative depth thresholding analyses on chromosome 3B, used to inform VCF DP filters	308
Appendix Figure A.2. Marker density of genotypic matrices used for pedigree-based imputation in TdGD populations.....	310
Appendix Figure A.3. Distribution of observed heterozygosity per WIL across markers within each TdGD population.....	311
Appendix Figure A.4. Distribution of observed heterozygosity per marker across WILs within each TdGD population.....	312
Appendix Figure A.5. Marker density of imputation genotype matrices for TdGD Pop31 and Pop41	313
Appendix Figure A.6. PCA results for TdGD Pop30	314
Appendix Figure A.7. PCA and SNP-PC correlation results for TdGD populations derived from KS061278M-4.....	315
Appendix Figure A.8. Per-individual heterozygosity in TdGD WILs following imputation.....	316
Appendix Figure A.9. Mean marker heterozygosity in TdGD Pop30 following imputation and heterozygote smoothing	317
Appendix Figure A.10. Marker MAF in TdGD WILs following imputation, summarized by population and subgenome.....	318
Appendix Figure A.11. Comparative alignment coverage profiles to assess 1BL.1RS presence in KS061278M-4.....	319
Appendix Figure B.1. Scree plot of the first ten principal components from genome-wide SNP marker data.....	350
Appendix Figure B.2. Distributions of normalized difference vegetation index (NDVI) and green reflectance performance by growth stage	352
Appendix Figure B.3. Pearson correlations for agronomic performance, yield and kernel component, functional quality, and extended drone data BLUEs in 21THD for TdGD Pop80 WILs	353
Appendix Figure B.4. Pearson correlations for agronomic performance, yield and kernel component, functional quality, and extended drone data BLUEs in 22EL for TdGD Pop80 WILs	354

Appendix Figure B.5. Pearson correlations for agronomic performance, yield and kernel component, functional quality, and extended drone data BLUEs in 22RL for TdGD Pop80 WILs	355
Appendix Figure B.6. Pearson correlations for agronomic performance, yield and kernel component, and functional quality BLUEs in 22THD for TdGD Pop80 WILs	356
Appendix Figure B.7. Pearson correlations for agronomic performance, yield and kernel component, and functional quality BLUEs in 23RL for TdGD Pop80 WILs	357
Appendix Figure B.8. Pearson correlations for agronomic performance, yield and kernel component, and functional quality BLUEs in 23THD for TdGD Pop80 WILs	358
Appendix Figure C.1. Pearson correlations among BLUEs for ionomics traits measured in 21THD for TdGD Pop80 WILs	364
Appendix Figure C.2. Pearson correlations among BLUEs for ionomics traits measured in 22THD for TdGD Pop80 WILs	365
Appendix Figure C.3. Pearson correlations among BLUEs for ionomics traits measured in 23THD for TdGD Pop80 WILs	366
Appendix Figure C.4. Hierarchical clustering of TdGD Pop80 trait BLUPs	367

List of Tables

Table 2.1. Source information for <i>Triticum dicoccoides</i> accessions used in germplasm development populations	48
Table 2.2. Number of array-based markers retained in WIL progeny and TdGD population founders.....	66
Table 2.3. Population genetic summary statistics by TdGD population.....	68
Table 2.4. Sequencing information for the founders used in TdGD population development.....	69
Table 2.5. Variant counts across successive marker thinning steps, by subgenome and TdGD population	71
Table 2.6. Pairwise nucleotide diversity between population founders across TdGD populations	81
Table 2.7. Average marker heterozygosity and minor allele frequency in TdGD WILs following imputation	84
Table 2.8. Summary of donor introgression in TdGD WILs, aggregated by population and subgenome	86
Table 2.9. Proportion of WILs with donor introgression segments detected on each chromosome arm, summarized by TdGD population.....	92
Table 3.1. <i>Triticum dicoccoides</i> germplasm development (TdGD) populations and corresponding number of wild introgression lines (WILs).....	117
Table 3.2. Summary of mixed model results across agronomic performance, yield and kernel component, and functional quality traits evaluated in TdGD Pop80.....	134
Table 3.3. Multiple linear regression of GFP on HD and SD.....	137
Table 3.4. Multiple linear regression of PQS on GP and SRC	148
Table 3.5. GWAS results summary for agronomic performance using BLUPs	156
Table 3.6. GWAS results summary for yield and kernel components using BLUPs	159
Table 3.7. Summary of donor allele direction for GR and NDVI GWAS associations	161
Table 3.8. GWAS results summary for functional quality traits using BLUPs.....	169
Table 4.1. Summary of mixed model results across ionomic traits evaluated in TdGD Pop80 .	203
Table 4.2. GWAS results summary for ionomics using BLUPs.....	213

Table 5.1. Subset of grain yield by trait (GYT) index results for RP, performance checks, and top three WILs	244
Table 5.2. High confidence marker-trait association regions containing typed donor alleles in KSTdGD80-175 or KSTdGD80-55	252
Appendix Table A.1. Subset of a pedigree input file for TdGD population 80 used for imputation with AlphaImpute2	305
Appendix Table A.2. Variant counts for TdGD population founders across successive filtering steps.....	306
Appendix Table A.3. Mean and median lengths of donor introgression segments detected in WILs within each TdGD population under moving average (MA) based segment calling parameterizations	307
Appendix Table B.1. Trial management details for TdGD testing environments	320
Appendix Table B.2. UAS platform, sensor, altitude, and image resolution specifications used for drone-based data collection in 2021 and 2022	322
Appendix Table B.3. Cumulative precipitation for TdGD field trials	323
Appendix Table B.4. Trait means and standard deviations across TdGD testing environments	324
Appendix Table B.5. Summary of mixed model results for extended drone data, split by growth stage, for TdGD Pop80	325
Appendix Table B.6. Within-environment estimates of repeatability for agronomic performance, yield and kernel component, functional quality, and extended drone data traits in TdGD Pop80	326
Appendix Table B.7. GWAS results summary for agronomic performance using BLUEs	327
Appendix Table B.8. GWAS results summary for yield and kernel components using BLUEs	331
Appendix Table B.9. GWAS results summary for functional quality using BLUEs	339
Appendix Table B.10. GWAS results summary for extended drone data using BLUPs.....	342
Appendix Table B.11. GWAS results summary for extended drone data using BLUEs	344
Appendix Table C.1. Soil testing data for Thomas County, Kansas in 2022 and 2023	359
Appendix Table C.2. Itonomic trait means and standard deviations in Thomas County across years	360

Appendix Table C.3. Within-environment estimates of repeatability for ionomic traits in TdGD Pop80	361
Appendix Table C.4. GWAS results summary for ionomics using BLUEs.....	362
Appendix Table D.1. Agronomic performance BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks	368
Appendix Table D.2. Grain yield BLUEs and BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks	369
Appendix Table D.3. Yield and kernel component BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks	370
Appendix Table D.4. Functional quality BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks	371
Appendix Table D.5. Macronutrient BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks.....	372
Appendix Table D.6. Micronutrient and trace element BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks	373
Appendix Table D.7. Toxic element BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks.....	374

Acknowledgements

None of what this research has accomplished would have been possible without the support, guidance, encouragement, and belief of many people, especially during the moments when I did not fully have confidence in myself.

My greatest thanks go to my advisor, Dr. Allan Fritz. Working with him has been, and will remain, one of the great honors of my professional career. His knowledge of wheat breeding, his philosophy toward teaching, research, and life in general, and his steady belief in the value of this work have shaped who I am both personally and professionally. Learning from him has been a joy and an experience I will never forget.

I am also deeply grateful for the mentorship I have received from Dr. Blaine Johnson. His consistent guidance, thoughtful conversations about breeding, and willingness to share perspective on many other experiences have been integral to the completion of this work. His mentorship has helped shape the way I view research and has strengthened my understanding of how my own contributions can be valuable to wheat breeding.

I would also like to thank my committee members, Drs. Mary Guttieri and Eduard Akhunov, for their time, expertise, and engagement with my research. Both are outstanding scientists whose feedback has changed the way I approach scientific inquiry and the standard to which I hold myself. I have genuinely appreciated the opportunity to grow as a scientist with the guidance of my committee.

I have also had the privilege of receiving mentorship outside of my immediate academic program, which has given me confidence in my professional abilities and supported my ambitions beyond the research itself. In particular, the professional development training and support I received through the FFAR Fellows program provided me with skills, perspectives, and

connections that will continue to shape my career. I am especially grateful to Drs. Sarah Mcneil and Rima Thapa for their continued guidance throughout my graduate career, for offering perspectives I may not otherwise have encountered, and for consistently expressing confidence in my ability to succeed.

Beyond the philosophical, scientific, and professional guidance I have received, countless individuals contributed to the technical execution of this project. I am grateful to every member of the K-State wheat breeding program with whom I have overlapped during my time here, including the many student workers and interns who spent long days in the Kansas sun and humidity all while maintaining a positive attitude and a willingness to learn.

I am especially thankful to Angie Matthews, who trained me in many aspects of lab work, including DNA extraction and marker screening, and to Shaun Winnie, whose organization, precision, and support through planting preparations, harvest, and seed processing helped ensure that this work could be carried out carefully and successfully. I also thank Dr. Yemane Belaineh for his support with genomic and bioinformatic analyses, and for providing essential training, ideas, and guidance during the development of my analysis pipeline.

One of the greatest outcomes of my time in the K-State wheat breeding program has been the friendships I made along the way. I am deeply grateful to Dr. Augusto Tessele, Enzo Unzain, Kristen Kimbrell, Carter Westerhold, Dr. Lucas Batista, Eduardo Melgar, and Dr. Marshall Clinesmith for their friendship, support, help with data collection, conversations about analysis methods, and discussions about breeding philosophy. I cherish the time we have spent together and will always treasure the conversations, laughter, and shared experiences that made this journey meaningful.

Dedication

I dedicate this work to my family. The unwavering belief my parents and siblings have always had in my aspirations gave me the confidence to believe that I could accomplish this, and, as it turns out, you were right all along.

Most importantly, I dedicate this work to the love of my life, my partner, Zach Velez. You have never doubted me for a second and have made countless sacrifices to follow me, support me, and encourage me in every sense of the word. You lifted me up when I was down and filled in where I fell short. There is no one I would rather have moved to Kansas with. We did it, my love.

Chapter 1 - Literature review

Wild emmer as a resource for allelic diversity

T. dicoccoides (wild emmer) is an invaluable genetic resource for improving the allelic diversity of wheat breeding programs. Given the central role of tetraploid wheats in transmitting ancestral alleles to modern wheat, wild emmer offers an abundance of allelic diversity that can enhance germplasm variation. Alongside other tetraploid wheats (AABB genome) and *Ae. tauschii* (DD genome), *T. dicoccoides* forms part of the secondary gene pool of hexaploid bread wheat, which allows for relatively straightforward introgressions through homologous recombination (Hao et al., 2020; Nyine et al., 2020).

Other hexaploid wheats, such as spelt, are part of the primary gene pool, allowing for unrestricted recombination due to complete genome compatibility. In contrast, introgression from the secondary gene pool – though highly valuable – can lead to reduced fertility in progeny due to unbalanced chromosome constitutions. Despite this, *T. dicoccoides* and *Ae. tauschii* are prime sources of allelic diversity due to their high nucleotide diversity relative to other species in the wheat gene pool (Dreisigacker et al., 2008; Nyine et al., 2020; Zhou et al., 2020).

There is considerable evidence of historical gene flow between tetraploid wheats, such as *T. dicoccoides*, and hexaploid bread wheat. This has shaped the genetic landscape of germplasm currently used in breeding programs worldwide (Blanco-Pastor, 2022; He et al., 2019; Zhou et al., 2020). Notably, there is a greater reduction in allelic diversity in the D subgenome of wheat compared to the A and B subgenomes. This discrepancy likely arises from the much higher level of gene flow from tetraploid wild emmer compared to the D subgenome donor, *Ae. tauschii* (Dvorak et al., 2006; He et al., 2019).

Regions of the wheat genome containing introgressions from *T. dicoccoides* are associated with higher regional nucleotide diversity, leading to less differentiation at the nucleotide level between hexaploid wheat and wild emmer (He et al., 2019; Zhou et al., 2020). However, not all diversity is beneficial. Differentiation between wheat and wild emmer occurred for a reason, and the primary focus of breeding programs remains the release of high-yielding varieties well-adapted to mechanical harvest. Introgression from wild emmer often results in linkage drag, where deleterious alleles from wild relatives reduce yield – the most significant area of phenotypic differentiation between wheat and its wild ancestors (Chandrasekhar et al., 2017; Deblieck et al., 2022).

Nonetheless, introgressions that have been retained at high frequencies in the wheat genome are often linked to traits under improvement selection and local adaptation. This indicates that selection against deleterious alleles has been effective in the wheat improvement process (Blanco-Pastor, 2022; He et al., 2019).

The origins of tetraploid and hexaploid wheats

Wheat's evolution into the hexaploid species we recognize in farmers' fields today was driven by two key allopolyploidization events. Modern hexaploid bread wheat (*Triticum aestivum*, $2n = 6x = 42$, AABBDD) evolved through spontaneous hybridizations between wild grass species in the Fertile Crescent. The first of these events involved hybridization between the wild diploid wheat species *Triticum urartu* ($2n = 2x = 14$, AA) and an unknown, likely extinct relative of *Aegilops speltoides* ($2n = 2x = 14$, BB) (Dvorak et al., 1993; IWGSC, 2014; Li et al., 2022; Peng et al., 2011; Rahman et al., 2020). This resulted in the formation of the tetraploid species *Triticum turgidum* ssp. *dicoccoides* ($2n = 4x = 28$, AABB), hereafter referred to as wild emmer wheat or *T. dicoccoides*.

Domestication of wild emmer, likely through selection for a non-shattering spike by early human societies, eventually led to the emergence of cultivated emmer (*T. turgidum* ssp. *dicoccum*). Spike fragility is considered one of the earliest targets of domestication, with brittle rachis loci on chromosomes 3A and 3B (*Btr-A1* and *Btr-B1*) controlling the disarticulation of spikelets from the spike (Nalam et al., 2006; Peng et al., 2011; Zhao et al., 2019). Although this brittleness facilitates seed dispersal in wild species, it would have been disadvantageous during intentional cultivation and harvest (Peleg et al., 2011; Peng et al., 2011; Tzarfati et al., 2014). Population mapping studies consistently indicate that the non-brittle allele is contributed by the domesticated parent, indicating that non-brittle *btr* alleles had already become somewhat fixed during early tetraploid cultivation (Nalam et al., 2006; Peleg et al., 2011). After the selection of non-brittle rachis types, cultivated emmer (*T. dicoccum*) further evolved into the free-threshing pasta wheat *T. turgidum* ssp. *durum*. This transition was facilitated, at least in part, by a gain-of-function mutation at the domestication gene *Q*, which contributes to free-threshing ability and exerts pleiotropic effects on spike architecture (Faris et al., 2003; Simons et al., 2006; Tzarfati et al., 2014). *Btr* and *Q* may participate in a shared developmental pathway, jointly influencing spike morphology, glume anatomy, and ultimately grain weight, pointing to coordinated regulatory changes during wheat domestication (Tzarfati et al., 2014; Zhao et al., 2019).

The second allopolyploidization event, crucial to the formation of hexaploid wheat, involved hybridization between tetraploid wheat and the wild diploid goatgrass *Aegilops tauschii* ($2n = 2x = 14$, DD) (Dvorak et al., 1998; McFadden & Sears, 1946; Petersen et al., 2006). The free-threshing character of hexaploid bread wheat is also governed by the *Q* gene, which is present in both tetraploid and hexaploid wheat. This indicates that the mutation giving rise to the *Q* gene gain-of-function occurred only once in evolutionary history (Dubcovsky & Dvorak,

2007; Simons et al., 2006). As a result, the A and B subgenome donor of bread wheat must have already possessed the free-threshing trait conferred by *Q* and, given the early and widespread fixation of non-brittle *btr* alleles in cultivated tetraploids, most likely displayed a tough, non-shattering rachis (Peleg et al., 2011; Peng et al., 2011; Tzarfati et al., 2014). Yet, the presence of hulled, non-free-threshing hexaploid varieties, such as spelt (*Triticum spelta*), has led to multiple hypotheses regarding the evolution of hexaploid wheat.

Extensive research by Dvorak and colleagues (1998, 2012) has led to a widely accepted model in which free-threshing tetraploid wheat hybridized with *Ae. tauschii*, producing primitive, hulled hexaploid wheat that carried the *Q* gain-of-function mutation. To achieve full free-threshing status, an additional mutation was required in the glume tenacity gene (*Tg*) from *Ae. tauschii*. The reversion to a hulled phenotype, as seen in spelt wheat, likely arose via hybridization with non-free-threshing emmer wheat, reverting back to the ancestral *q* allele in a process known as de-domestication (Dvorak et al., 2012; Jiang et al., 2019; Zhou et al., 2020).

Thus, *T. dicoccoides* occupies a pivotal role in the evolution of modern wheat, acting as a central conduit of gene flow between its wild ancestors and modern wheat species. Figure 1.1 diagrams this current model of wheat evolution. Although this widely accepted model of wheat evolution posits a linear trajectory from wild emmer to durum wheat (*T. turgidum* ssp. *durum*), with free-threshing tetraploid wheat donating the A and B subgenomes to hexaploid bread wheat (*T. aestivum*), alternative models challenge this narrative. Some researchers, citing genetic similarities between the A and B subgenomes of wild emmer and bread wheat, argue that *T. dicoccoides* itself may be the direct progenitor of *T. aestivum*'s A and B subgenomes (Cheng et al., 2019; Petersen et al., 2006). However, these similarities are more plausibly attributed to

historical gene flow between tetraploid and hexaploid wheats, which readily hybridize and shared overlapping geographic distributions (Dvorak et al., 2006; Zhou et al., 2020).

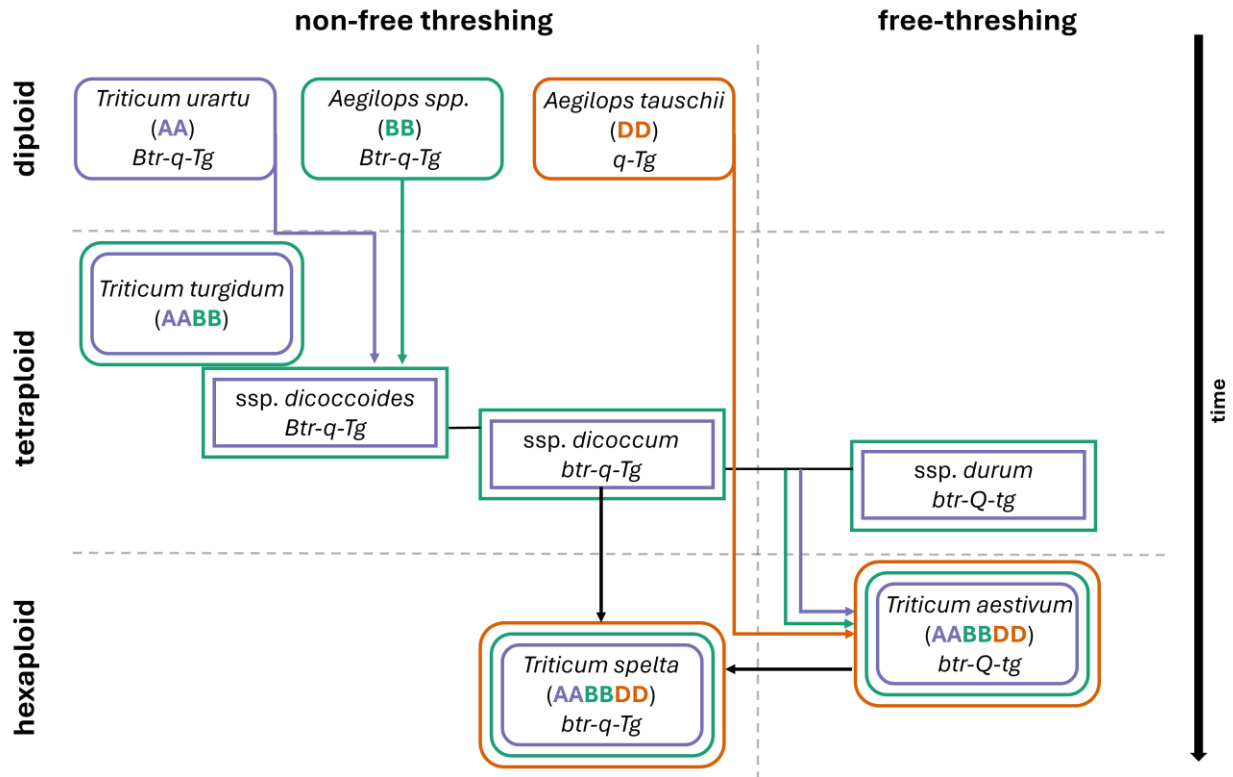


Figure 1.1. Schematic diagram of wheat evolution in terms of time, threshing ability, and ploidy level

Allelic states at the domestication loci *Btr*, *q*, and *Tg* are indicated, where functional *Btr* alleles confer a brittle rachis, the domesticated *Q* allele promotes free-threshing and compact spike architecture, and the recessive *tg* allele reduces tenacious glumes.

An alternative hypothesis suggests that durum wheat is the direct progenitor of the A and B subgenomes of bread wheat, as its free-threshing character aligns with that of *T. aestivum* (Matsuoka & Nasuda, 2004; Pont et al., 2019). However, the more pronounced reduction in allelic diversity in durum wheat compared to bread wheat resulting from domestication bottlenecks implies that the two species underwent separate domestication processes. This perspective is further supported by debates regarding the origins of durum wheat itself. While some researchers propose a linear evolution from cultivated emmer (*T. turgidum* ssp. *dicoccum*),

others hypothesize parallel domestication pathways for durum wheat and cultivated emmer, both diverging independently from wild emmer (He et al., 2019; Zhou et al., 2020).

Despite these debates, the significant reduction in allelic diversity during the transitions from wild emmer to cultivated emmer, and subsequently to durum wheat, supports the likelihood of a linear evolutionary trajectory. Tracing these historical events remains challenging due to continuous gene flow between tetraploid and hexaploid wheats, as well as genomic similarities that blur distinctions between modern and ancestral wheats. These complexities underscore the need for further genomic research to unravel the intricate history of wheat domestication and evolution.

The importance of increasing wheat allelic diversity

There is ample evidence that modern wheat has experienced significant genetic bottlenecks during polyploidization and domestication, resulting in a marked loss of nucleotide diversity (Akhunov et al., 2010; Fu & Somers, 2009; Haudry et al., 2007). The domestication bottleneck was particularly severe in wheat due to the multiple rounds of interspecies hybridization that led to the formation of hexaploid wheat. Studies have shown that haplotype diversity in *T. aestivum* is lower than in synthetic hexaploid wheat, which was created through artificial crosses between tetraploid wheat and *Ae. tauschii*. This demonstrates that the domestication process unintentionally reduced diversity in wheat (Akhunov et al., 2010; Dreisigacker et al., 2008).

The recombination landscape of wheat further highlights this reduction in diversity. Long-range linkage disequilibrium (LD) is more frequent in *T. aestivum* than in wild emmer, indicating fewer recombination events and larger LD blocks in cultivated wheat (Akhunov et al., 2010; Fu & Somers, 2009; Haudry et al., 2007). The D subgenome of wheat shows particularly

large LD blocks compared to the A and B subgenomes, reflecting lower diversity and gene flow from *Ae. tauschii* compared to tetraploid wheats (Akhunov et al., 2010).

Domestication loci, including the *Q* gene, are often associated with selective sweeps, where purifying selection reduces allelic diversity around loci of interest (Haudry et al., 2007). Enriching allelic diversity in highly conserved regions, without eliminating the effects of key domestication loci, is therefore a key strategy in breeding programs. One approach is to introduce diversity from *T. dicoccoides*, which lacks extreme long-range LD, enabling breeders to restore diversity in critical genomic regions.

It is also important to note that tetraploid durum wheat (*T. turgidum* ssp. *durum*) has an even more pronounced reduction in diversity than bread wheat. This suggests that durum wheat was not the direct tetraploid progenitor of hexaploid wheat, but that the two types of wheat were improved separately but simultaneously (Haudry et al., 2007). This reinforces the importance of using *T. dicoccoides* to increase allelic diversity in both hexaploid and tetraploid wheat.

***T. dicoccoides* geographical and ecological adaptation**

The single-origin hypothesis suggests that hexaploid bread wheat emerged through a polyploidization event approximately 8,000-10,000 years ago. This event combined the genomes of free-threshing tetraploid wheat (AABB) with *Aegilops tauschii* (DD), resulting in the hexaploid genome structure (AABBDD) found in modern bread wheat (Cheng et al., 2019; Pont et al., 2019; Z. Wang et al., 2023). Following its formation, *T. aestivum* spread widely across the Fertile Crescent via human migration and trade routes. Archaeological evidence confirms its simultaneous emergence in regions such as modern-day Turkey, Israel, and Iran (Cheng et al., 2019; Luo et al., 2007).

The wild progenitor of tetraploid wheats, wild emmer or *T. dicoccoides*, spans a broad range across the Fertile Crescent and has adapted to diverse ecological and geographical conditions. Key environmental factors – such as altitude, soil type, water availability, and temperature – have driven local adaptation and allelic divergence within *T. dicoccoides*, resulting in substantial diversity across its accessions (Li et al., 2001; Peleg et al., 2008a; Ren et al., 2013; Vuorinen et al., 2018). Genetic studies have classified *T. dicoccoides* into two primary groups: a northern population found in northeastern Syria, Turkey, Iraq, and Iran, and a southern population concentrated in Israel, Jordan, Lebanon, and southwestern Syria. Northern populations, situated in regions of overlap with domesticated wheat, have experienced greater gene flow with hexaploid and tetraploid wheats. In contrast, the southern populations, isolated by geography, represent a distinct reservoir of allelic diversity, less frequently involved in hybridization with domesticated species (Cheng et al., 2019; Luo et al., 2007; Z. Wang et al., 2023; Zhou et al., 2020).

The overlapping ranges of *Ae. tauschii* and northern *T. dicoccoides* in the Fertile Crescent support the hypothesis that the polyploidization event leading to hexaploid wheat likely occurred in this region (Zhou et al., 2021). Northern *T. dicoccoides* populations likely played a critical role in this evolutionary process due to their proximity to *Ae. tauschii*. Meanwhile, the geographic isolation of southern populations has preserved alleles adapted to unique environments, making these populations particularly valuable for traits essential to future wheat improvement, such as drought tolerance and disease resistance (Adhikari et al., 2026; Peleg et al., 2005; Vuorinen et al., 2018; Zhou et al., 2020).

Importantly, ecological conditions exert a stronger influence on allelic divergence within *T. dicoccoides* than geographic distance. Studies from the University of Haifa, Israel, reveal

significant differentiation among accessions based on environmental gradients, including altitude and seasonal growth habits (spring versus winter), which are shaped by local climatic conditions (Cheng et al., 2019; Kato et al., 1997, 1998; Ren et al., 2013). Traits such as nutrient use efficiency, drought tolerance, and resilience to extreme temperatures often reflect the specific environments in which populations evolved (Vuorinen et al., 2018). For example, accessions from arid or high-altitude regions exhibit unique adaptations to water scarcity or cold stress, underscoring their potential for breeding climate-resilient wheat varieties (Z. Wang et al., 2023).

Accessions from environmentally stressful or niche conditions often harbor valuable allelic diversity shaped by selective pressures in their habitats. By incorporating these novel alleles into hexaploid wheat, breeders can address challenges such as climate variability and resource limitations. Southern populations of *T. dicoccoides*, in particular, are underutilized but represent a unique resource for broadening the allelic pool of wheat breeding germplasm and enhancing its adaptability to diverse agricultural environments (L. Huang et al., 2019; Li et al., 2001; Peleg et al., 2008a).

Overall, *T. dicoccoides* offers an invaluable genetic reserve, with diversity driven by both geographic and ecological factors. While the southern populations are especially distinct due to their isolation, accessions from across environmentally challenging habitats hold tremendous potential for modern wheat improvement. Leveraging this diversity is critical for equipping hexaploid wheat with the resilience needed to thrive in changing climatic pressures.

Wild emmer as a resource for phenotypic diversity

Given *T. aestivum*'s advanced improvement status in terms of yield and agronomic stability, the use of *T. dicoccoides* (wild emmer) as a source of trait diversity must be well-justified. Introgressions from wild emmer often introduce linkage drag – manifesting as

reductions in yield and domestication traits such as seed size and threshability – potentially offsetting breeding gains (Bohra et al., 2022; Hao et al., 2020; Huang et al., 2003; Klindworth et al., 2013; Liu et al., 2006). Consequently, the incorporation of wild emmer alleles into modern wheat must be purposeful, targeting traits with proven potential for agronomic improvement rather than introducing alleles with unknown or deleterious functions.

Fortunately, extensive research by the Institute of Evolution at the University of Haifa, Israel, has characterized the phenotypic diversity of *T. dicoccoides*. Studies have analyzed individual accessions as well as recombinant inbred line (RIL) mapping populations derived from crosses between *T. durum* and wild emmer (Peng et al., 2013; Xie & Nevo, 2008). This work provides a critical foundation for leveraging wild emmer trait diversity in hexaploid wheat breeding. Although hybridization between *T. dicoccoides* and *T. aestivum* is complicated by their differing ploidy levels, these efforts are essential to harness the genetic potential of wild emmer, which is central to wheat's evolutionary and domestication history.

Wild emmer as a resource for improving yield and physiological components

One of the primary challenges in utilizing wild emmer for introgression breeding is its frequent association with reduced yield, driven by the co-inheritance of negative alleles and the potential disruption of domestication traits such as threshability and rachis robustness (Bohra et al., 2022; Hao et al., 2020; Huang et al., 2003; Klindworth et al., 2013; Liu et al., 2006). Wild emmer accessions also exhibit markedly greater asynchrony in tiller development, leading to a parallel asynchrony in flowering time (Rahman et al., 2023). Such variability poses challenges for field management and mechanical harvest, which rely on uniform crop maturity. In addition, wild emmer produces a higher proportion of non-fruiting tillers; although these do not contribute directly to grain number, they may still contribute positively to photosynthetic capacity by

increasing total leaf area and improving radiation interception (Rahman et al., 2023). By contrast, yield gains in modern wheat correspond with selection for an erect canopy architecture to support increased planting density, promote fewer but more productive tillers, and improve fruiting efficiency (Shan & Osborne, 2024; L. Zhao et al., 2023).

Nevertheless, wild emmer possesses valuable alleles for enhancing yield component traits, such as kernel characteristics and biomass production. For example, while domesticated wheat grains are generally plump (wide and short), wild grains are often narrower and longer. Combining alleles from both wild and domesticated grains may improve grain weight by developing kernels that are both long and wide (Golan et al., 2015). Crucially, improvements in yield potential are not driven solely by sink traits (e.g., spike architecture and grain characteristics). A balance between source and sink traits must be achieved. This balance involves sufficient tillering and biomass production to support photosynthetic capacity while maintaining the ability to fill sink organs like grains. Excessive biomass production can waste energy if the plant cannot efficiently utilize photoassimilates for grain filling (Murchie et al., 2023; Reynolds et al., 2017; Slafer et al., 2023). Historically, wheat breeding has focused on improving sink traits through maximization of harvest index, resulting in domesticated wheats with significantly reduced biomass compared to their wild relatives (Katamadze et al., 2023; Reynolds et al., 2017).

Wild emmer, in contrast, offers exceptional potential for improving wheat biomass and photosynthetic capacity through its greater leaf area, enhanced canopy development, and superior radiation interception (Merchuk-Ovnat et al., 2016b). Although *T. dicoccoides* produces lower grain yields than *T. aestivum*, the substantial variation it displays in yield component traits highlights its capacity to contribute meaningfully to modern wheat breeding. It is important to

recognize, however, that yield improvement in wheat is not governed solely by additive genetic effects. Combining high-harvest index alleles from *T. aestivum* with high-biomass alleles from *T. dicoccoides* will not necessarily produce a linear or predictable enhancement of source-sink balance. Rather, yield is shaped by a combination of additive and epistatic variation, with allelic interactions occurring both within and across subgenomes (Tessele et al., 2025a, 2025b). As a result, introducing new alleles for a specific yield component may not express the intended phenotype if those alleles disrupt the finely tuned epistatic architecture that has been optimized within a particular breeding program. Thus, while targeted introgression of wild emmer diversity holds promise for improving yield and agronomic performance, it also carries the risk of perturbing established allelic interactions that underpin adaptation and productivity in elite germplasm.

Wild emmer as a resource for disease resistance variation

While not always directly tied to yield improvements, resistance to pests and diseases is an essential component of wheat breeding efforts. Resistance traits can significantly influence the final yield outcome, often through their capacity to mitigate losses under biotic stress. Complete susceptibility to specific pathogens can drastically reduce yields, particularly in environments with high disease pressure or in varieties lacking background resistance. The extent of yield loss due to disease often depends on the level of pre-existing breeding progress for resistance in the target variety (Barkley et al., 2014; Munaro et al., 2020). Historical yield improvements in wheat have often been closely linked to advances in breeding for disease resistance, with disease-resistant varieties enabling consistent production across a range of growing conditions (Zetzsche et al., 2020).

Wheat breeding for disease resistance has traditionally relied on loci with major effects, which offer robust and easily traceable resistance. This focus has made introgression breeding for disease resistance a primary strategy, as it avoids the complexities associated with transferring and tracking numerous minor effect quantitative loci. Instead, the efficient introgression of major loci into elite wheat backgrounds reduces the risk of linkage drag and allows for faster deployment of resistant varieties (Shahin et al., 2024).

Wild emmer is an exceptional genetic resource for disease resistance due to its vast allelic diversity. This diversity stems from its adaptation to a broad range of environmental conditions and pressures across the Fertile Crescent, which has led to the evolution of resistance mechanisms against many pathogens. A multitude of resistance genes for critical wheat diseases – such as powdery mildew, leaf rust, stripe rust, and *Fusarium* head blight – have been identified in *T. dicoccoides* accessions, making it a key contributor to modern wheat improvement (Peng et al., 2013; Xie & Nevo, 2008).

Powdery mildew

Among the diseases affecting wheat, powdery mildew caused by *Blumeria graminis* f. sp. *tritici* has been a primary focus in resistance breeding efforts. *T. dicoccoides* has contributed extensively to the identification and characterization of powdery mildew resistance genes, many of which have been mapped to specific chromosomal regions and are now utilized in wheat breeding programs. Early studies, such as those by Reader and Miller (1991), established the potential of wild emmer for powdery mildew resistance. Subsequent efforts, including collaborations between the University of Haifa in Israel and the China Agricultural University in Beijing, have refined this understanding, identifying numerous genes tailored to diverse environmental pressures (Chen et al., 2005; Liu et al., 2002).

Some powdery mildew resistance genes derived from *T. dicoccoides* show region-specific effectiveness, likely reflecting the environmental conditions where the accessions evolved or where the mapping was conducted. For example, many genes mapped in China exhibit high efficacy under the country's significant powdery mildew pressure (Chen et al., 2005). However, broader-spectrum resistance genes – effective across multiple isolates – have also been identified. These genes, such as those mapped by Blanco et al. (2008) and Ben-David et al. (2010), offer greater potential for global applicability, enabling the development of cultivars with more widespread durable resistance under conditions of strategic deployment.

The ability of wild emmer to provide resistance to a wide array of powdery mildew isolates underscores its utility in wheat improvement. Genes like *Pm36*, mapped to chromosome 5BL, and *PmG16*, mapped to chromosome 7AL, exemplify the potential of wild emmer-derived alleles to enhance global wheat resilience (Ben-David et al., 2010; Blanco et al., 2008). The early mapping of resistance genes, such as *Pm16* and *Pm30*, highlights the historical and ongoing importance of wild emmer in resistance breeding (Blanco et al., 2008; Chen et al., 2005). Recent cloning of *Pm41* on chromosome 3BL has enabled functional validation within a hexaploid wheat background and revealed that this resistance gene is unique to southern wild emmer populations (M. Li et al., 2020). This finding highlights the importance of tapping into geographically isolated and genetically differentiated groups of *T. dicoccoides* as reservoirs of effective and novel disease-resistance alleles.

Leaf rust

Wheat rust diseases, including leaf rust, stripe rust, and stem rust, are among the most significant threats to global wheat production (Huerta-Espino et al., 2011). Leaf rust, caused by *Puccinia triticina*, is particularly widespread and has severe implications for yield potential,

especially in regions like Kansas. In Kansas, where leaf rust frequently exerts high disease pressure, foliar fungicides are critical for mitigating yield losses during epidemic years (Barkley et al., 2014; Cruppe et al., 2021; Munaro et al., 2020). Globally, leaf rust affects both tetraploid and hexaploid wheat production, with durum wheat being especially vulnerable due to limited historical breeding for resistance (Knott et al., 2005; Lhamo et al., 2023).

Although yield losses from leaf rust are typically less severe than those caused by stripe rust or stem rust, the disease still accounts for significant annual losses because of its widespread prevalence and frequency. Relative to stem rust, specifically, its early onset in the growing season allows it to reduce photosynthetic capacity earlier, which can compound its impact on overall plant health and productivity (Cruppe et al., 2021; Huerta-Espino et al., 2011). This underscores the importance of developing resistant wheat varieties to complement chemical control measures.

Multiple sources of genetic resistance to leaf rust have been identified in wild emmer wheat. The first major resistance gene derived from wild emmer for wheat improvement was *Lr53*, located on chromosome 6B (Marais et al., 2005). This gene has been widely used due to its broad-spectrum effectiveness against diverse *P. triticina* pathotypes and its durable protection across varied environmental conditions (Knott et al., 2005; Raghunandan et al., 2022). Additionally, *Lr53* has been successfully deployed in both *T. durum* and *T. aestivum* (Knott et al., 2005; Marais et al., 2005).

Despite its success, the emergence of pathogen virulence against *Lr53* highlights the challenges of relying on single resistance genes. Continuous pathogen evolution can overcome such major effect loci, necessitating the pyramiding of *Lr53* with other *Lr* genes. Pyramiding involves combining additive and epistatic effects of multiple resistance loci, offering broader and

more durable protection against evolving *P. triticina* pathotypes (Lhamo et al., 2023; Raghunandan et al., 2022).

While wild emmer provides invaluable genetic diversity, introgressing its resistance loci can pose challenges. Genes such as *Lr53* often exhibit stronger performance in their native background or closely related lines than in unrelated breeding populations (Knott et al., 2005; Marais et al., 2005); effectiveness of resistance loci is highly dependent on the genetic background into which they are transferred. This issue is especially pronounced when transferring resistance into hexaploid wheat, where greater background allelic diversity can alter or diminish gene expression (Peng et al., 2013). Reports of resistance suppression during interspecific crosses, particularly across ploidy levels, add further complexity. In hexaploid wheat, resistance dilution is more likely because each of the three subgenomes can carry suppressors that attenuate the effect of introgressed genes (Chen et al., 2013; Hiebert et al., 2020; Knott et al., 2005; L. Wu et al., 2015). This phenomenon is well documented in rust diseases; nevertheless, because leaf rust resistance is comparatively rare in durum wheat, the genetic diversity present in wild emmer remains an essential resource for addressing this vulnerability (Lhamo et al., 2023).

Since the identification of *Lr53*, additional sources of resistance from wild emmer have been discovered, including genes conferring both seedling resistance and durable, slow-rusting adult plant resistance. The variation in resistance reflects the diversity of *P. triticina* pathotypes and the evolutionary adaptation of wild emmer to different environmental pressures (Huerta-Espino et al., 2011; Sela et al., 2011; Xie & Nevo, 2008). Breeding programs that rely heavily on single genes, such as *Lr53*, are particularly vulnerable to new virulent strains of *P. triticina*.

Breeding programs that adopt a more diversified strategy – stacking multiple, complementary sources of resistance – are better positioned to counter rapidly evolving pathogen populations. For instance, *Lr67*, a race-nonspecific slow-rusting gene introgressed from wild emmer to hexaploid *T. aestivum*, demonstrates strong efficacy in reducing leaf rust severity when combined with other *Lr* genes (Raghunandan et al., 2022; Lhamo et al., 2023). However, interpreting its effectiveness is complicated by the expression of leaf tip necrosis (LTN), which reduces functional leaf area and is frequently associated with yield penalties (Herrera-Foessel et al., 2014; Johnston et al., 2017; Sivasamy et al., 2014). When pyramiding multiple leaf rust resistance genes, the appearance of LTN can artificially inflate perceived resistance by simply reducing the amount of susceptible leaf tissue. Thus, to effectively harness durable resistance from wild emmer, improvements in leaf rust resistance must be disentangled from the confounding physiological consequences of LTN.

Although the potential of wild emmer as a source of durable resistance for modern wheat has been recognized, its deployment has been only moderately successful. This is due in part to genomic suppressors that reduce the efficacy of transferred resistance genes (Chen et al., 2013; Hiebert et al., 2020; Knott et al., 2005; L. Wu et al., 2015) and, additionally, to the yield-reducing effects of LTN associated with some pyramided leaf rust resistance genes (Herrera-Foessel et al., 2014; Johnston et al., 2017; Sivasamy et al., 2014). Nevertheless, genes such as *Lr53* remain promising resources, particularly for enhancing resistance in durum wheat. Innovative approaches to resistance breeding remain essential, especially with the high leaf rust pathotype diversity and widespread adaptation to many wheat growing regions (Huerta-Espino et al., 2011; Llamas et al., 2023; Raghunandan et al., 2022). The discovery, characterization, and

strategic deployment of novel sources of resistance from wild emmer remains important for sustaining long-term progress in improvement of both hexaploid and tetraploid wheats.

Stripe rust

Stripe rust, caused by *Puccinia striiformis* f. sp. *tritici* (Pst), is one of the most significant threats to global wheat production. Unlike leaf rust and stem rust, stripe rust thrives under cooler temperatures (7°C–15°C) and humid conditions, making it particularly devastating in temperate and high-altitude regions (Chen, 2005; De Wolf et al., 2023). The disease impacts both seedling and adult plants, resulting in characteristic yellow-orange pustules arranged in stripes along the leaf veins. By disrupting photosynthetic processes, stripe rust reduces grain fill, kernel weight, and test weight, leading to substantial yield losses (Chen, 2005, 2013). Severe epidemics can result in near-total yield loss in susceptible varieties (De Wolf et al., 2023; Wan et al., 2007).

In the United States, stripe rust has become increasingly severe and frequent due to the emergence of aggressive pathogen races adapted to changing environment conditions. Multiple reports indicate emerging races of stripe rust that are now capable of surviving higher temperatures that would previously have been lethal, evolving the capacity to enter a dormant state and maintain disease proliferation once permissible temperatures are regained (Ali et al., 2025; Gardner et al., 2023; Milus et al., 2009; Milus & Seyran, 2006; Miñana-Posada et al., 2025; Zhang et al., 2024). In Kansas and the Central Great Plains, this thermal adaptation has contributed to increasingly early-season infections, especially when mild winters and wet springs align (Chen et al., 2002; De Wolf et al., 2023). These shifts have enabled stripe rust to spread in regions that historically were not well suited for its overwintering or early spring establishment. Early-season environmental indicators – particularly temperature and precipitation patterns – are therefore critical for predicting epidemic onset and implementing timely management strategies

(Chen, 2005; De Wolf et al., 2023). In epidemic years, yield losses can exceed 50% in susceptible cultivars, highlighting the importance of genetics-based resistance and integrated disease control (Chen, 2013). Indeed, stripe rust susceptibility has been identified as the strongest predictor of yield loss among Kansas wheat cultivars (Cruppe et al., 2021).

T. dicoccoides (wild emmer) has been a foundational source of resistance genes for combatting stripe rust. Two of the most significant resistance genes derived from wild emmer are *Yr15* and *Yr36*. Found on chromosome 1B, *Yr15* provides race-specific resistance through a hypersensitive response that limits pathogen growth. This gene is effective against a broad range of *Pst* pathotypes and has been widely used in tetraploid and hexaploid wheat breeding programs (Klymiuk et al., 2018; Sun et al., 1997). Advances in molecular tools, such as marker-assisted selection (MAS), have facilitated the introgression of *Yr15* into elite wheat cultivars without significant linkage drag (Peng et al., 2000a; Qie et al., 2019; Yaniv et al., 2015). Located on chromosome 6BS, *Yr36* confers durable, non-race-specific resistance. This gene encodes a kinase-START protein, playing a role in temperature-dependent adult-plant resistance (Fu et al., 2009). *Yr36* is particularly valuable in temperate regions where stripe rust pressure persists throughout the growing season, providing sustainable control measures (Chen, 2013). Notably, *Yr36* was absent in modern bread and pasta wheats until it was reintroduced to *T. durum* and *T. aestivum* from wild emmer germplasm, highlighting the importance of *T. dicoccoides* in improving stripe rust resistance across diverse wheat varieties (Fu et al., 2009). However, a potential yield disadvantage associated with *Yr36* has contributed to its limited deployment in breeding programs (Brevis & Dubcovsky, 2010; Hale et al., 2012). Additional resistance sources from wild emmer continue to be identified. For example, *YrH52*, also mapped to chromosome 1B, confers resistance to specific *Pst* races (Peng et al., 2000b). More recently, genome-wide

association studies (GWAS) have uncovered novel stripe rust resistance loci within wild emmer populations, offering new opportunities to leverage this diversity for durable resistance in modern wheat (Tene et al., 2022).

The rapid evolution of *Pst* populations poses a persistent challenge to resistance breeding. Historical reliance on single major-effect genes has often led to the emergence of virulent pathogen strains capable of overcoming these resistance mechanisms (Chen, 2005; Chen et al., 2002; Shahin et al., 2024). To address this, breeders are increasingly adopting pyramiding strategies, combining multiple resistance genes to provide broad-spectrum and durable protection (Klymiuk et al., 2018; Singh et al., 2005).

A key focus in recent breeding efforts has been the selection of slow rusting, quantitative disease resistance alleles. Unlike qualitative resistance, which relies on major-effect genes to completely block pathogen growth, slow rusting involves multiple minor-effect genes that collectively reduce the rate of infection and disease progression (Chen, 2005, 2013; Shahin et al., 2024; Singh et al., 2005). This form of resistance, associated with partial tolerance, is more durable because it exerts less selection pressure on the pathogen to evolve virulence. Notably, cultivars carrying only 4-5 slow rusting genes can achieve high resistance to both leaf and stripe rusts (Singh et al., 2005).

Loci such as the *Lr34/Yr18/Sr57/Pm38* complex, located on chromosome 7D, exemplify slow rusting resistance mechanisms that span multiple diseases, including leaf rust, stripe rust, stem rust, and powdery mildew (Krattinger et al., 2009). These adult-plant resistance (APR) genes confer partial, broad-spectrum resistance by slowing disease progression rather than completely halting infection. This approach makes them more durable, as the slower resistance mechanism imposes less selective pressure on pathogens, reducing the likelihood of resistance

breakdown over time. The *Lr34/Yr18/Sr57/Pm38* complex encodes an ABC transporter protein and has been a cornerstone of rust resistance breeding globally, particularly in hexaploid bread wheat (Krattinger et al., 2009; Li et al., 2023). However, its utility in tetraploid species like durum wheat is limited by the challenges of transferring D genome sequences, which has necessitated alternative strategies such as transgenics or homoeologous recombination for introgression into these genetic backgrounds (Li et al., 2023).

Similarly, the *Lr53/Yr35* complex, derived from wild emmer wheat, demonstrates another example of co-inherited loci systematically transferred together during introgression into a hexaploid background. These loci provide partial resistance to both leaf and stripe rust and highlight the value of wild emmer as a reservoir of multi-trait resistance (Marais et al., 2005). The co-inheritance of these loci simplifies breeding efforts, as LD ensures their simultaneous transfer, facilitating pyramiding strategies to enhance resistance durability. Such linked loci are crucial for breeding programs aiming to address the increasing variability of pathogen populations driven by climate change and evolving virulence (Shahin et al., 2024).

Genetic diversity for stripe rust resistance in wild emmer germplasm is closely linked to ecogeographical factors. Accessions from regions such as northern Israel, particularly the Mount Hermon area, are remarkable sources of resistance due to the high prevalence of *Pst*-favorable climatic conditions in these areas (Sela et al., 2014; L. Huang et al., 2019). This highlights the importance of incorporating geographically diverse germplasm into breeding programs to address the variability of stripe rust pathotypes.

Stripe rust remains a significant challenge for wheat production, with wild emmer providing an invaluable reservoir of resistance genes that have demonstrated effectiveness in both tetraploid and hexaploid backgrounds (Fu et al., 2009; Klymiuk et al., 2018; Marais et al.,

2005; Shahin et al., 2024; Sun et al., 1997). Advances in molecular tools, such as GWAS and MAS, have enabled the efficient utilization of wild emmer-derived genes, such as *Yr15* and *Yr36*. These genes, along with slow rusting loci and co-inherited resistance blocks, form the cornerstone of modern stripe rust resistance breeding strategies.

***Fusarium* head blight**

Fusarium head blight (FHB), primarily caused by *Fusarium graminearum*, is a devastating fungal disease that significantly impacts wheat yield and grain quality. The disease thrives under warm and humid conditions during the flowering and grain-filling stages. Losses are exacerbated by the accumulation of mycotoxins such as deoxynivalenol (DON), which render grain unsuitable for consumption or trade. This dual impact on yield and quality makes FHB one of the most economically significant wheat diseases, particularly in North America, where favorable environmental conditions frequently lead to outbreaks (Bai et al., 2018; Lemes da Silva et al., 2019). While management strategies such as crop rotation and fungicide application provide some level of control, breeding for genetic resistance remains the most effective and sustainable approach to mitigating FHB.

Breeding for FHB resistance is exceptionally challenging due to the complex, quantitative nature of the trait. Unlike many other wheat diseases where major-effect resistance genes have been identified and deployed, FHB resistance is controlled by multiple QTLs, each contributing small to moderate effects. Resistance mechanisms in wheat are categorized into five types: resistance to initial infection (Type I), resistance to fungal spread within the spike (Type II), resistance to mycotoxin accumulation (Type III), resistance to kernel infection (Type IV), and tolerance to yield loss (Type V) (Bai et al., 2018; Bai & Shaner, 1994, 2004; Rudd et al.,

2001). Type II resistance, which limits fungal spread, is the most widely studied and targeted in breeding programs due to its relatively higher heritability and ease of phenotypic evaluation.

The deployment of FHB resistance in breeding programs is complicated by the genetic background's influence on QTL effectiveness and the environment's impact on trait expression. Durable resistance often requires the stacking of multiple QTLs, leveraging their additive and epistatic effects (Kirana et al., 2023). Advances in molecular tools have facilitated the precise identification and deployment of FHB resistance alleles. These technologies are critical for integrating resistance genes from diverse sources, including wild emmer wheat, into modern wheat germplasm (Kumar et al., 2007; Lemes da Silva et al., 2019; Otto et al., 2002).

Wild emmer wheat has been prioritized as a resource for improving FHB resistance; given its adaptation to diverse environments, a reservoir of alleles associated with resistance to both biotic and abiotic stresses has been identified (Garvin et al., 2009; Xie & Nevo, 2008). This allelic diversity is particularly valuable for durum wheat, which has limited natural resistance to FHB (Buerstmayr et al., 2013; Oliver et al., 2007). The close genetic relationship between wild emmer and tetraploid wheat facilitates the transfer of resistance QTLs into durum wheat with minimal genetic incompatibility. Indeed, QTLs originating from wild emmer have been validated in durum wheat backgrounds, with introgressed mapping populations exhibiting measurable improvements in FHB resistance (Buerstmayr et al., 2013; Soresi et al., 2017).

Resistance loci from wild emmer have been mapped to several chromosomes, with significant QTLs identified on chromosomes 1A, 2A, 3A, 5A, and 7A (Garvin et al., 2009; Kumar et al., 2007; Otto et al., 2002). Of these, a QTL on chromosome 7A has been particularly impactful in reducing susceptibility. This locus confers Type II resistance, limiting fungal spread and reducing DON accumulation (Kumar et al., 2007). Another promising QTL on chromosome

3A has been associated with reduced disease severity and kernel infection, underscoring the potential of wild emmer to contribute diverse resistance mechanisms (Otto et al., 2002). These QTLs on chromosomes 7A and 3A with identified allelic diversity in wild emmer have established priority for introgression into durum wheat (Buerstmayr et al., 2013; Kumar et al., 2007; Otto et al., 2002).

The introgression of FHB resistance from wild emmer to durum wheat is not without challenges. Linkage drag – where undesirable agronomic traits are inadvertently co-transferred with resistance alleles – remains a significant obstacle (Hao et al., 2020; Summers & Brown, 2013). Moreover, for FHB specifically, pleiotropic relationships between resistance, maturity, and plant height complicate the interpretation of resistance sources and present difficulties for commercial deployment of these alleles in adapted backgrounds (Buerstmayr et al., 2013; Kirana et al., 2023). The highly polygenic architecture of FHB resistance, coupled with strong environmental sensitivity, further underscores the need for advanced genomic approaches capable of identifying, resolving, and pyramiding multiple QTLs (Kirana et al., 2023).

Fusarium head blight remains one of the most challenging diseases for commercial wheat production, threatening yield and grain quality worldwide. Wild emmer wheat offers a reservoir of allelic diversity for improving FHB resistance, with significant QTLs identified on chromosomes 3A and 7A providing valuable contributions to modern durum breeding efforts (Buerstmayr et al., 2013; Soresi et al., 2017). In hexaploid wheat, however, effective deployment of FHB resistance has been limited, even though resistance does not appear to be suppressed when transferred into a synthetic hexaploid background (Hartel et al., 2004). By leveraging the genetic resources of wild emmer, breeders can develop wheat cultivars with enhanced resistance,

supporting sustainable production and protecting grain quality in the face of ongoing disease pressures.

Wild emmer as a resource for improvement of functional quality

T. dicoccoides (wild emmer) has been extensively characterized as a valuable genetic resource for improving the functional quality of both tetraploid and hexaploid wheats (Peng et al., 2013; Xie & Nevo, 2008). While wild emmer generally exhibits lower yield levels compared to modern cultivated wheats, the strong inverse correlation between yield and protein content in wheat highlights its rich allelic diversity for protein and overall functional quality. Notably, wild emmer wheat exhibits distinct quality profiles compared to *T. aestivum* and *T. durum*, with wild alleles consistently associated with higher grain protein content, both in native germplasm and when introgressed into elite breeding lines (Cabas-Lühmann et al., 2023; Peng et al., 2013).

One of the most significant contributions from wild emmer to cultivated wheat is the *Gpc-B1* allele on chromosome 6B. This allele not only increases grain protein content but also enhances the concentrations of essential micronutrients such as zinc (Zn), iron (Fe), and manganese (Mn) (Distelfeld et al., 2007). The underlying mechanism involves nutrient remobilization from senescing leaves to developing grains during the filling stage. The role of *Gpc-B1* in senescence-driven nutrient remobilization underscores the multifaceted benefits of wild emmer introgressions in enhancing multiple traits simultaneously. Furthermore, *Gpc-B1* is genetically linked to *Yr36*, a high-temperature adult-plant stripe rust resistance gene derived from wild emmer, enabling simultaneous improvements in functional quality, disease resistance, and yield potential, in both tetraploid and hexaploid wheat backgrounds (Uauy et al., 2005).

The adoption of *Gpc-B1* in breeding programs has been limited; however, its consistently elevated expression of grain protein and micronutrient concentrations across diverse

environments demonstrates strong genetic control (Distelfeld et al., 2007). *Gpc-B1*, with its tight linkage to *Yr36* and the associated reduction in yield, harbors a penalty that is unlikely to be tolerated in breeding programs under strong selection for yield productivity (Brevis & Dubcovsky, 2010; Hale et al., 2012). The allele itself is unique to wild emmer germplasm and is absent from modern breeding pools, underscoring both its value and the historical difficulty of overcoming the associated yield disadvantage. Whether the penalty is driven primarily by *Gpc-B1*, by *Yr36*, or by their tight physical linkage remains unclear. Without targeted recombination to break this linkage block, it is difficult to determine whether bidirectional selection at the locus is feasible. Although historical gene flow between tetraploid and hexaploid wheat has shaped the allelic diversity of modern breeding populations, key beneficial alleles such as *Gpc-B1* remain somewhat exclusive to wild emmer (Distelfeld et al., 2007; Peleg et al., 2009a). This highlights the need to continue mining wild emmer diversity – while developing strategies to decouple favorable traits from their associated yield penalties during introgression.

Wild emmer is a vital resource for enhancing macro- and micronutrient concentrations, particularly Zn and Fe, which are essential for addressing widespread global nutrient deficiencies that disproportionately affect women and children. Genetic approaches are critical in this context, as management-based interventions are often limited in regions with nutrient-depleted soils (Cakmak et al., 2004; Distelfeld et al., 2007; Peleg et al., 2009a). For example, *ZIPT*, a gene involved in Zn uptake and transport, is upregulated in wild emmer under Zn-deficient conditions, reflecting its adaptation to poor soils and its potential for improving Zn homeostasis in cultivated wheat (Durmaz et al., 2010).

In addition to *Gpc-B1*, wild emmer harbors numerous QTLs associated with quality traits, such as those on chromosome 5B, which influence both grain protein content and grain yield

(Cakmak et al., 2004; Fatiukha et al., 2020a; Gonzalez-Hernandez et al., 2004; Peng et al., 2013). Transgressive segregation for nutrient accumulation in *T. durum* and *T. aestivum* introgression populations and substitution lines further highlights the feasibility of improving both yield and functional quality simultaneously (Cakmak et al., 2004; Fatiukha et al., 2020a; Peleg et al., 2009a; Xie et al., 2008). However, efforts to decouple beneficial traits, such as Fe and Zn accumulation, from undesirable traits, such as toxic metal uptake (e.g., cadmium and arsenic), are critical for ensuring sustainable biofortification strategies (Cabas-Lühmann et al., 2023).

While enhancing protein content is a primary goal, functional protein quality must also be addressed to ensure superior end-use performance. Wild emmer introgression lines typically exhibit higher protein content, but this does not always translate to improved protein functionality. Deeper characterization of gluten proteins – specifically glutenins and gliadins – is necessary to link protein quantity with quality (Xu et al., 2004).

Among wild emmer accessions, there is substantial allelic variation for protein functionality. For instance, the *Glu-1A* and *Glu-1B* loci each contain multiple unique alleles associated with variations in gluten strength and breadmaking performance (Pourmohammadi et al., 2023; Xu et al., 2004). Additionally, the high-molecular-weight glutenin subunit *Glu-1Ay*, located on chromosome 1A, is silenced in common hexaploid wheat but active in wild emmer. Introgressing an active *Glu-1Ay* allele into elite tetraploid and hexaploid germplasm has been shown to improve flour quality and functionality, demonstrating the potential of wild emmer for enhancing wheat quality (Wang et al., 2018).

The extensive genetic diversity of wild emmer enables the improvement of yield, as well as nutritional and functional quality in modern durum and hexaploid bread wheat. The integration of traits such as higher grain protein content, increased micronutrient density, and

superior gluten functionality offers a pathway to address global challenges in food security and malnutrition (Cabas-Lühmann et al., 2023; Peleg et al., 2009a).

Wild emmer and flowering time variation across vernalization requirements

Flowering time is a crucial trait for regional adaptation and yield potential. Properly timed plant maturation optimizes the balance between biomass development and carbohydrate translocation, ultimately influencing final yield outcomes in wheat (Kato et al., 1998; Peng et al., 2013). Flowering too early may expose plants to early-season cold stress and limit biomass accumulation, while flowering too late risks exposure to late-season heat and drought stress, reducing the time for grain carbohydrate accumulation. Although developmental timing is a quantitative trait, a few major genes related to vernalization and photoperiod account for a significant proportion of variation in modern wheat germplasm (DeWitt et al., 2021).

The primary gene regulating vernalization requirement is *Vrn1*, located on the homoeologous group 5 chromosomes of *T. aestivum*. Vernalization ensures flowering after cold winter temperatures, triggering the transition from vegetative to reproductive growth as temperatures rise (Kato et al., 1997; Strejčková et al., 2023). Another key gene group includes the photoperiod sensitivity genes (*Ppd-1*) on homoeologous group 2 chromosomes. Overexpression of *Ppd-1* loci across all three subgenomes results in photoperiod insensitivity, allowing flowering to occur regardless of day length (DeWitt et al., 2021). Photoperiod-insensitive varieties are valuable in providing stability in flowering time under varying environmental conditions, whereas photoperiod sensitivity is beneficial in adaptive genotypic response.

In addition to these major genes, earliness *per se* (*eps*) genes also contribute to flowering time variation. Unlike *Vrn1* and *Ppd-1*, *eps* genes function independently of environmental

stimuli and have smaller, quantitative effects, modulating flowering time under consistent growing conditions. These QTLs are frequent on chromosomes 2B, 3A, 4D, 5A, 5B, and 7B (Zhou et al., 2016).

Wild emmer populations exhibit substantial variation in developmental timing due to eco-geographical adaptation to diverse climatic conditions (Peng et al., 2013). Early studies on wild emmer classified accessions into four groups based on vernalization requirements: strongly spring, moderately spring, moderately winter, and strongly winter types. While facultative spring/winter types are widely distributed across the Fertile Crescent, strongly spring types are restricted to low-altitude Israeli regions, and strongly winter types are confined to colder, high-altitude areas. Altitude, temperature, and humidity are the strongest correlators with vernalization requirements in wild emmer (Kato et al., 1997).

In addition to vernalization requirement, there is substantial allelic variation at the *Vrn1* locus in wild emmer, influencing the length of vernalization needed for flowering (Strejčková et al., 2023). Beyond vernalization, significant variation exists in flowering time due to QTLs associated with *Ppd-1* and other loci. This variation allows wild emmer to fine-tune developmental timing for regional adaptation (Kato et al., 1998; Peng et al., 2003).

Wild emmer exhibits high phenological plasticity, enabling it to adapt to climatic variability. Nevo et al. (2012) demonstrated this through a 28-year study, showing an average 8.5-day reduction in flowering time in response to rising temperatures. This adaptive shortening highlights wild emmer's value in breeding programs targeting resilience to climate change. Notably, QTLs for both earlier and later flowering have been identified in wild emmer. Even on the same chromosome (e.g., 2B), flowering can be accelerated or delayed depending on

interactions with other loci, providing opportunities for both flexibility and precision in breeding strategies (Zhou et al., 2016).

The allelic diversity present in wild emmer offers extensive opportunities for tetraploid and hexaploid wheat improvement, particularly for modifying developmental timing traits. Although QTL related to crop maturation have not historically been prioritized for introgression from wild emmer, the substantial variability surrounding key developmental loci highlights a reservoir of usable diversity (Kato et al., 1997, 1998; Strejčková et al., 2023; Nevo et al., 2012). Whether the breeding goal is to accelerate or delay flowering, adjust vernalization requirements, or broaden phenological plasticity, wild emmer provides alleles capable of expanding the allelic variation available in modern germplasm pools. Such adaptability is increasingly important under shifting climatic conditions, positioning wild emmer as a critical resource for future wheat breeding efforts.

Wild emmer as a reservoir of abiotic stress tolerance

The increasing unpredictability of climatic patterns has necessitated a shift in breeding programs toward developing crop varieties with enhanced climate resilience while maintaining other desirable agronomic traits. Achieving an optimal balance between yield plasticity – defined as a genotype’s ability to respond favorably to improved growing conditions – and yield potential is critical for ensuring performance in both stressed and non-stressed environments (Peleg et al., 2005). Among abiotic stressors, drought and heat are the most significant threats to global food production, driving associated challenges such as soil aridity, salinity, and nutrient scarcity (Merchuk-Ovnat et al., 2016a; Peleg et al., 2005, 2009). In this context, wild emmer wheat represents an invaluable genetic resource for addressing these challenges (Nevo & Chen, 2010; Peng et al., 2013).

Wild emmer evolved in the Fertile Crescent, an environment dominated by water limitation. The extreme and recurrent drought characteristic of this region has driven the development of both drought escape and drought tolerance mechanisms in wild emmer populations (Lauterberg et al., 2022; Peleg et al., 2005). Drought escape is evident in earlier flowering accessions, which accelerates grain-filling to avoid terminal drought stress. Beyond escape mechanisms, wild emmer exhibits a suite of refined drought tolerance traits, including improved water-use efficiency, osmotic adjustment, and leaf-level adaptations such as rolling and higher chlorophyll content (Peleg et al., 2009b).

Significant variation for traits like carbon isotope ratio (indicative of water-use efficiency), harvest index, and total dry matter, has also been identified (Suneja et al., 2019). Most drought-adaptive alleles are attributed to wild emmer, although some contributions from elite breeding lines have been noted, reflecting the quantitative nature of drought tolerance (Nevo & Chen, 2010; Peleg et al., 2009b). The introgression of quantitative, small-effect loci from wild emmer into cultivated wheat can lead to linkage drag, complicating breeding efforts (Bohra et al., 2022; Hao et al., 2020; Huang et al., 2003; Klindworth et al., 2013; Liu et al., 2006; Merchuk-Ovnat et al., 2016a).

Despite these challenges, several major QTLs from wild emmer have been identified and characterized. QTLs on chromosomes 2B, 7A, and 7B have demonstrated improvements in crop productivity under water-limited conditions while maintaining performance in non-stressed environments (Fatiukha et al., 2021; Peleg et al., 2009b, Peng et al., 2013). However, introgressing these QTLs into elite *T. durum* and *T. aestivum* backgrounds has had mixed success, with some failing to validate attributable to strong linkage drag or unidentified epistatic interactions (Lauterberg et al., 2022; Merchuk-Ovnat et al., 2016a). One notable success is a

QTL on chromosome 5B, associated with improved nitrogen-use efficiency (NUE) under drought conditions. Although this QTL did not directly increase yield, it significantly enhanced grain protein content, demonstrating its utility in improving nutrient-use efficiency and productivity under stress in a hexaploid wheat background (Govta et al., 2024).

Salinity stress, exacerbated by drought, is another major abiotic challenge. Unlike drought tolerance, salinity tolerance involves simpler genetic control mechanisms, primarily the regulation of sodium (Na) accumulation and transport within the plant (Nevo & Chen, 2010). Tetraploid durum wheat, lacking the D genome from *Ae. tauschii*, is particularly vulnerable to salinity stress, as major salinity tolerance genes are often absent in modern A and B subgenomes (Suneja et al., 2019). Wild emmer, however, exhibits significant variability in Na uptake and sequestration, with some accessions demonstrating simultaneous tolerance to both drought and salinity. This underscores the interplay between these stressors and the potential of wild emmer to improve resilience across multiple abiotic challenges (Shavrukov et al., 2010).

A critical measurement for salinity tolerance is the K/Na ratio, with higher Na accumulation negatively affecting plant growth. Wild emmer accessions with superior salinity tolerance maintain a favorable K/Na ratio through mechanisms such as enhanced Na exclusion or sequestration in vacuoles (Nevo & Chen, 2010; Shavrukov et al., 2010). These traits are invaluable for improving tetraploid wheat's performance under saline and arid conditions.

Wild emmer's adaptability to diverse environmental conditions is further evidenced by its plasticity in stress-response. QTLs linked exclusively to plasticity metrics, rather than primary indicator traits, highlight its capacity to adjust to fluctuating environmental stresses (Suneja et al., 2019; Fatiukha et al., 2021). The broad genetic diversity observed across an ecogeographical

gradient offers opportunities for targeted selection of accessions from specific regions for introgression breeding (Peleg et al., 2005).

Wild emmer wheat is a vital reservoir of genetic diversity for abiotic stress tolerance, offering solutions to drought, salinity, and nutrient deficiencies. Its mechanisms of stress adaptation, rich allelic variability, and potential for introgression into modern tetraploid and hexaploid wheat cultivars make it a cornerstone for climate-resilient breeding programs. Although challenges such as linkage drag and complex trait inheritance remain, advances in genomic tools and precision phenotyping are paving the way for more efficient utilization of wild emmer's genetic resources. By harnessing the adaptive traits of this ancestral wheat, breeders can address the pressing need for sustainable and resilient agricultural systems in a changing climate.

Historic prioritization of qualitative traits in introgression breeding

As highlighted in previous sections, wild emmer carries a wide range of beneficial traits that have been well documented with respect to their ecogeographical associations and potential value in breeding programs (Peleg et al., 2005, 2008; Peng et al., 2013; Xie et al., 2008). However, despite the detailed trait mapping and characterization of allelic diversity, introgressions prevalent in modern wheat breeding remain largely confined to loci with large effects, particularly those controlling traits under simpler genetic regulation where major genes can be identified and mapped (Johansson et al., 2024). Moreover, successful examples of introgression from wild emmer are somewhat limited to tetraploid durum wheat, where full chromosomal compatibility enables more straightforward gene transfer (Buerstmayr et al., 2013; Knott et al., 2005; Lhamo et al., 2023; Soresi et al., 2017).

In the case of wild emmer and other species within the wheat gene pool, introgressions have historically focused on disease resistance, with only a few exceptions for large-effect QTL associated with complex traits, such as *Gpc-B1* (grain protein content) and *Glu-1Ay* (glutenin composition), that have been successfully deployed in both tetraploid and hexaploid breeding programs (Distelfeld et al., 2007; Peng et al., 2013; Xu et al., 2004). Notably, transferring wild emmer-derived disease resistance into hexaploid wheat can be complicated by a reduction in resistance efficacy, a phenomenon thought to result from attenuation of expression driven by suppressors across the three hexaploid wheat subgenomes (Chen et al., 2013; Hiebert et al., 2020; Knott et al., 2005; L. Wu et al., 2015).

One of the primary constraints limiting the broader utilization of wild emmer-derived alleles is linkage drag – the introduction of undesirable alleles from donor segments, which remain in genetic linkage with the target locus due to large chromosomal introgressions (Li et al., 2022; Platten & Fritsche-Neto, 2023; Qi et al., 2007; Summers & Brown, 2013). A well-documented example of linkage drag in wheat breeding is the 1BL.1RS translocation from rye (*Secale cereale*) to wheat. While this introgression conferred increased yield potential and enhanced tolerance to abiotic stress and disease through the rye-derived segment, it also resulted in a significant reduction in bread-making quality due to the loss of gluten storage protein genes (Johansson et al., 2024; Katamadze et al., 2023). Because this segment is a translocation that does not freely recombine with wheat chromosomes, it is expected to exhibit especially pronounced linkage drag when homologous recombination is restricted (Bohra et al., 2022; Hao et al., 2020; Klindworth et al., 2013). Another example previously discussed that arises under conditions of free recombination is the reported linkage drag associated with introgressing the *Gpc-B1/Yr36* complex. Here, a yield penalty is most pronounced in tetraploid materials and is

particularly evident in environments conducive to long grain fill periods, as *Gpc-B1* accelerates senescence, thereby shortening the duration available for grain filling and reducing final grain weight (Brevis & Dubcovsky, 2010; Hale et al., 2012).

In the case of wild emmer, introgressions into wheat often carry linkage drag effects related to yield. The introduction of chromosomal segments from wild emmer can disrupt the allelic composition and interactions within elite wheat backgrounds, thereby reducing yield performance due to the incorporation of unadapted alleles (Bohra et al., 2022; Hao et al., 2020; Huang et al., 2003; Klindworth et al., 2013; Liu et al., 2006). As a result, breeding efforts have prioritized the transfer of large-effect QTL while utilizing marker-assisted background selection to accelerate the introgression process while minimizing unwanted genetic baggage (Li et al., 2022). This strategy allows for the efficient transfer of beneficial alleles while limiting negative trade-offs associated with extensive linkage drag.

This challenge is exacerbated when making introgressions for traits controlled by numerous small-effect QTL distributed across the genome. Unlike single major-effect loci, these polygenic traits are difficult to map and selectively introduce without also carrying undesirable alleles from the donor genome. Backcrossing and recombination over multiple generations remain essential to refining introgressed segments, ultimately selecting for the smallest possible chromosomal region that retains a beneficial effect while minimizing disruptions to the genetic background (Nyine et al., 2020; Platten & Fritsche-Neto, 2023).

Despite the challenges, the reintroduction of genetic diversity from less-adapted germplasm is sometimes necessary. The trade-off between short-term genetic gains and the long-term maintenance of genetic variation must be carefully managed. A targeted approach to introgression is particularly useful when existing breeding populations lack genetic diversity for

a given trait or when new selection pressures arise, such as the emergence of novel diseases (Li et al., 2022). In these cases, wild germplasm may provide critical resistance alleles, as has been demonstrated through the successful transfer of multiple disease-resistance loci from wild emmer into tetraploid and hexaploid wheat (Klymiuk et al., 2018; Knott et al., 2005; M. Li et al., 2020; Marais et al., 2005; Peirce et al., 2024; Sun et al., 1997; Tene et al., 2022).

With the increasing emphasis on quantitative breeding approaches – where selection upon genome-wide variation is prioritized rather than focusing solely on individual markers – the role of introgression breeding is evolving. If crop improvement strategies rely on the accumulation of small-effect alleles and complex genetic interactions, the prioritization of major QTL introgression may diminish, especially if those introgressions compromise essential agronomic traits such as yield stability. This raises a fundamental question: How can breeding programs balance the need for allelic diversity to ensure long-term adaptive potential without negatively impacting immediate performance?

Although introgression breeding in wheat has historically been limited to qualitative traits, emerging technologies – such as high-throughput genotyping and phenotyping, improved modeling of allelic effects, and haplotype-based selection – offer new opportunities. Future breeding programs may increasingly target introgressions to capture a broader spectrum of genetic variation, particularly in preparation for climate change challenges. By maintaining allelic diversity across numerous small-effect loci, breeders can enhance trait potential and stability, ensuring that wheat populations remain resilient to future environmental fluctuations. Without these proactive strategies, breeding programs risk facing future scenarios where elite germplasm becomes ill-equipped to handle emerging threats or environmental extremes. Wheat

improvement programs are thus at the pivotal position of meeting both present and future challenges.

Evaluation of approaches to introgression breeding

While linkage drag remains a fundamental challenge in utilizing wild relatives for introgression, the choice of introgression strategy significantly influences the success and effectiveness of the process. Several approaches have been employed in introgression breeding, each with its own advantages and limitations.

The most straightforward approach is direct crossing between the recipient and donor germplasm (Katamadze et al., 2023). Kansas State University, through its Wheat Genetics Resource Center, has successfully expanded wheat's allelic diversity via direct crosses with wild relatives, producing a collection of introgression germplasm (Cox et al., 2017; Gill et al., 2006). This method allows for the rapid transfer of novel genetic material and typically requires subsequent backcrossing to recover the recipient background and minimize the size of introgressed linkage blocks.

However, direct crossing is constrained by chromosomal compatibility, as it relies on meiotic recombination between homologous genomes. As a result, it is primarily limited to species within wheat's secondary gene pool, such as *Ae. tauschii* (D subgenome) and *T. turgidum* (A/B subgenomes) (Boehm & Cai, 2024; Katamadze et al., 2023). When chromosomal complements do not fully align, issues such as low seed set in F₁ progeny arise, requiring multiple backcross generations to improve fertility and stability. Furthermore, this approach does not facilitate simultaneous introgression across all three wheat subgenomes without elaborate, time consuming crossing schemes. The Wheat Genetics Resource Center at Kansas State University has strategically emphasized D subgenome introgressions using *Ae. tauschii* to

address the historically low allelic diversity of this subgenome relative to the A and B subgenomes (Gill et al., 2006; Cox et al., 2017). As such, the structure of D subgenome diversity and the effects of *Ae. tauschii* introgression into hexaploid germplasm are already well defined in prior work (Nyine et al., 2020, 2021, 2025; Singh et al., 2019). Complementary efforts by the USDA-ARS Hard Winter Wheat Genetics Research Unit have focused on directly crossing wild emmer with *T. aestivum* to introduce A and B subgenome alleles for targeted improvement on functional quality traits. This foundational work has provided a clearer understanding of ploidy alterations that occur during direct crosses between wild emmer and hexaploid wheat, as well as recombination patterns in resulting progeny, highlighting regions where large introgressions are retained and areas where the *T. aestivum* background persists due to strong domestication loci signatures (Price et al., 2025a, 2025b).

Another way breeders have achieved widespread introgression across all wheat subgenomes is through an indirect approach – the development of synthetic hexaploid wheat. This process involves recreating *T. aestivum*'s hexaploid genome structure through crosses between its ancestral species, tetraploid *T. turgidum* and diploid *Ae. tauschii* (Warburton et al., 2006). The resulting synthetic wheat lines introduce a significant increase in genetic diversity across all three wheat subgenomes.

This approach has been extensively employed by CIMMYT, primarily using durum wheat as the tetraploid parent, resulting in synthetics that are predominantly of the spring growth type (Cox et al., 2017). Unlike direct crosses, which often yield poorly fertile offspring segregating for chromosome number, synthetic wheats provide fully meiotically stable lines that have full compatibility with any *T. aestivum* parent. Furthermore, CIMMYT's research on synthetic wheat has enhanced understanding of D subgenome variation and helped introduce

valuable diversity into genomic regions of historically low nucleotide diversity (Cox et al., 2017; Warburton et al., 2006).

However, the major drawback of synthetic hexaploids is the reliance on unadapted genetic material, which lacks key domestication traits. Since these lines originate from deeply divergent wild relatives, their agronomic performance can be unpredictable due to disruptions in genetic interactions (Katamadze et al., 2023). While direct crosses allow for relatively quick recovery of recipient background through backcrossing, synthetic wheats require a longer period of selection to re-domesticate and eliminate unfavorable alleles. Therefore, breeders face a trade-off between the speed of deploying introgressions and the degree of genetic diversity introduced (Cox et al., 2017; H. Li et al., 2024).

Gene editing, particularly CRISPR-based approaches, represents a promising but still developing method for gene introgression. However, significant challenges remain, primarily regulatory restrictions and public acceptance surrounding gene editing, which have slowed research and commercialization efforts (Katamadze et al., 2023).

While CRISPR offers unparalleled precision for modifying specific genomic regions, its application in introgression breeding remains constrained by quantitative genetic complexity. Many key agronomic traits are governed by polygenic networks, making it difficult to pinpoint and edit individual loci with large enough effects to be meaningful. This limitation is a key reason why traditional introgression breeding has focused on large-effect QTL rather than more complex, small-effect loci. Until methodologies for selecting and quantifying polygenic variation improve, gene editing will remain complementary rather than a replacement for introgression breeding (Katamadze et al., 2023).

Despite these challenges, CRISPR has already demonstrated its utility in wheat breeding, particularly in targeting gluten proteins associated with celiac disease. Researchers have used multiplex genome editing to knock out specific gliadin components while preserving gluten proteins essential for breadmaking quality (Sánchez-León et al., 2018; Yu et al., 2023). This demonstrates CRISPR's potential for highly targeted modifications with minimal off-target effects. As gene editing techniques evolve, their role in introgression breeding is likely to expand.

When gene introgression involves species outside the wheat primary and secondary gene pools, an alternative approach is the creation of amphiploid hybrids. Amphiploids result from direct crosses between recipient and donor species, followed by chemical chromosome doubling to ensure full genomic representation of each species. This method is particularly useful when meiotic compatibility between species is limited, allowing breeders to bypass traditional crossing barriers (Boehm & Cai, 2024).

One well-known example of amphiploid utilization in wheat breeding is wheat-rye hybrids. These crosses aimed to enhance wheat's abiotic stress tolerance, biomass production, and yield potential by incorporating rye genetic material while maintaining essential wheat traits such as breadmaking quality (Boehm & Cai, 2024; Katamadze et al., 2023;). However, a major limitation of this approach is unexpected reductions in gluten functionality, leading to lower breadmaking quality (Johansson et al., 2024). Despite the benefits of introgressing adaptive traits from rye, linkage drag remains a persistent challenge.

Another application of amphiploid hybrids in wheat breeding involves crossing with *Ae. tauschii*. While direct crosses between hexaploid *T. aestivum* and *Ae. tauschii* are feasible, they often result in poorly fertile F₁ hybrids. Fertility can be restored through backcrossing; however,

this process tends to disproportionately favor the *T. aestivum* D subgenome, leading to a reduced contribution from *Ae. tauschii* as backcrossing progresses (Cox et al., 2017). In contrast, using an amphiploid hybrid approach – where embryos from the cross between wheat and *Ae. tauschii* are rescued and subjected to chemical chromosome doubling – creates a synthetic octoploid that retains the full genomic complement of both parental D subgenomes. Subsequent backcrossing restores hexaploidy while maintaining a higher frequency of *Ae. tauschii* alleles compared to standard direct crosses. While amphiploid hybrids are not strictly necessary for *Ae. tauschii* introgression, they provide a more efficient means of transferring D subgenome diversity, ultimately accelerating genetic gain (H. Li et al., 2024; Nyine et al., 2020).

Conclusion

The genetic improvement of elite wheat breeding germplasm relies on both maintaining high-yielding, well-adapted varieties and ensuring long-term genetic diversity to sustain future gains. Wild emmer represents an invaluable reservoir of genetic variation, offering beneficial alleles for yield improvement, stress resilience, disease resistance, and functional quality. However, challenges such as linkage drag, the complexity of polygenic traits, and disruptions to agronomic stability must be carefully managed to maximize the benefits of introgression breeding.

The evolution of hexaploid wheat highlights the crucial role of *T. dicoccoides* in shaping modern wheat's genetic landscape. While past breeding efforts have primarily focused on major qualitative traits, such as disease resistance, advancements in genomic tools and breeding strategies now provide an opportunity to integrate *T. dicoccoides* alleles for the enhancement of quantitative traits. For example, the use of synthetic octoploid wheat as a bridge for introgression allows for a more efficient transfer of diverse alleles. Although this approach has been prioritized

in the genetically constrained D subgenome, its utility lies in circumventing some of the traditional limitations of direct crossing, enabling a broader and more effective utilization of species like *T. dicoccoides* in wheat breeding.

Modern breeding programs face a critical challenge: balancing immediate agronomic performance with long-term adaptability. The increasing unpredictability of environmental conditions, along with the emergence of new pathogens and changing agricultural demands, underscores the necessity of preserving and leveraging genetic diversity. While historical selection pressures have favored short-term productivity, future wheat improvement must integrate strategies that enhance phenotypic plasticity and ensure stable resilience in dynamic growing conditions.

Ultimately, the successful utilization of *T. dicoccoides* in modern wheat breeding hinges on a strategic, long-term approach – one that balances genetic diversity with agronomic performance. By embracing innovative breeding methodologies and maintaining a commitment to allelic conservation, wheat improvement programs can secure the resilience, adaptability, and productivity necessary to meet global challenges in food production.

Chapter 2 - Genomic signals of introgression using a wheat-*Aegilops tauschii* amphiploid bridge to bring allelic diversity from wild emmer into adapted hexaploid wheat

Abstract

Introgression from wheat wild relatives can broaden the allelic base of breeding germplasm, with donor recovery, genome stability, and distribution of retained haplotypes influenced by founder divergence and population development strategy. This research characterized genome-wide introgression in seven *Triticum dicoccoides* (wild emmer) germplasm development populations produced using a wheat-*Aegilops tauschii* synthetic octoploid amphiploid bridge. Expected donor contributions were 8.33% on the A and B subgenomes and 4.17% on the D subgenome. High-density sequencing-derived variants from population founders were used to perform pedigree-based imputation into introgression lines genotyped using the Illumina Infinium Wheat/Barley 40 K SNP array. Patterns of introgression were assessed using founder allele proportions. Observed donor contributions broadly reflected the population development scheme. Wild emmer donor proportions on the A and B subgenomes ranged from 0.09-0.20, while *Aegilops tauschii* donor proportions on the D subgenome ranged from 0.04-0.07. Introgression was enriched in distal, higher recombination regions and reduced in pericentromeric intervals, consistent with known chromosomal recombination landscapes in wheat. D subgenome introgressions were less frequent, but often longer, likely reflecting low D subgenome progeny marker density, reduced recombination resolution, and greater divergence between *Aegilops tauschii* ssp. *tauschii* and the recurrent parent D subgenomes. Comparison with previous results indicates that amphiploid bridge and direct crossing approaches produce

broadly similar landscapes of introgression from wild emmer. The bridge-mediated strategy differs by introducing allelic diversity in the D subgenome and by theoretically reducing D subgenome aneuploidy expected from direct tetraploid x hexaploid crosses. Overall, this research demonstrates that amphiploid bridge-mediated introgression is a viable strategy for generating hexaploid wheat breeding germplasm with introgressions across all three subgenomes.

Introduction

Bread wheat (*Triticum aestivum*, $2n = 6x = 42$, AABBDD) is an allohexaploid species whose genome is composed of three homoeologous subgenomes derived from distinct ancestral species. This genomic architecture enables the targeted reintroduction of allelic diversity from related taxa in the primary and secondary gene pools – species that share one or more homologous subgenomes with wheat and thus permit homologous recombination – through conventional crossing approaches (Boehm & Cai, 2024; Hao et al., 2020; Nyine et al., 2020). Two such species have been central to wheat introgression breeding: wild emmer (*Triticum dicoccoides*), which carries the A and B subgenomes present in *T. aestivum*, and *Aegilops tauschii*, the diploid progenitor of the wheat D subgenome.

Among the three wheat subgenomes, the D subgenome has been recognized as harboring relatively low allelic diversity compared with the A and B subgenomes, and has therefore been a primary target of introgression breeding (Akhunov et al., 2010; Cox et al., 2017; H. Li et al., 2024; Zhou et al., 2020). However, direct crosses between hexaploid wheat and *Ae. tauschii* are hindered by poor fertility, as only the D subgenome finds homologous pairing partners during meiosis. To overcome this limitation, synthetic hexaploid wheat has been evaluated as an alternative strategy. In this approach, *Ae. tauschii* is crossed with a tetraploid wheat – most commonly durum wheat (*T. durum*) – to recreate a hexaploid genome analogous to the

evolutionary origin of bread wheat (Cox et al., 2017; Dreisigacker et al., 2008; Warburton et al., 2006). While synthetic hexaploids can introduce substantial D subgenome diversity, they also represent a major reset of breeding progress, producing germplasm that is genetically distant from elite hexaploid breeding germplasm.

The disruption of coadapted allele networks that have been refined through decades of selection for agronomic performance has limited the direct use of synthetic hexaploids in commercial variety development (Cox et al., 2017; Dreisigacker et al., 2008; H. Li et al., 2024). The Wheat Genetics Resource Center (WGRC) at Kansas State University assessed an alternative strategy through employment of a wheat-*Aegilops tauschii* amphiploid bridge. In this system, hexaploid wheat is crossed directly with *Ae. tauschii*, followed by embryo rescue and colchicine-induced chromosome doubling to generate synthetic octoploid doubled haploids containing two full complements of the D subgenome (Dale et al., 2017; Nyine et al., 2020). These octoploids could then be backcrossed to the original hexaploid recurrent parent, recovering hexaploidy and an elite background while retaining targeted D subgenome introgressions. Germplasm developed through this amphiploid bridge approach has been extensively characterized, and its utility for introducing D subgenome diversity into elite wheat backgrounds is well established; a detailed discussion of this work is therefore deferred to the existing literature (Nyine et al., 2020, 2021, 2025; Singh et al., 2019).

In contrast to the D subgenome, introgression of allelic diversity on the A and B subgenomes is comparatively more tractable due to the availability of tetraploid wheat species that share full homology with both subgenomes, including wild emmer (*T. dicoccoides*), domesticated emmer (*T. dicoccum*), and durum wheat (*T. durum*). Direct crosses between hexaploid and tetraploid wheats avoid the severe fertility penalties associated with the complete

absence of homoeologs for two subgenomes, as occurs in direct crosses with *Ae. tauschii*.

Nonetheless, much of the existing work on A and B subgenome introgression has focused on durum wheat, where complete chromosomal homology with wild emmer and a comparatively narrow allelic base have facilitated the identification and deployment of beneficial trait diversity (Buerstmayr et al., 2013; Chen et al., 2013; Haudry et al., 2007; Lhamo et al., 2023; Peng et al., 2013; Soresi et al., 2017; Xie & Nevo, 2008).

Efforts to introgress A and B subgenome diversity from wild emmer directly into hexaploid wheat are comparatively fewer. Outside of synthetic hexaploid-based strategies, which primarily emphasize D subgenome variation, notable progress has been made by the USDA-ARS Hard Winter Wheat Genetics Research Unit (HWWGRU). In this work, direct crosses between adapted hexaploid wheat and tetraploid wild emmer, followed by one backcross to the adapted recurrent parent, were used to generate introgression populations in elite backgrounds, enabling characterization of wild emmer introgression within agronomically adapted germplasm (Price et al., 2025a). These efforts have produced a valuable set of derived materials that retain favorable allele combinations already selected for regional performance while incorporating novel diversity from wild emmer. A substantial body of related work from the University of Haifa, Israel, has also characterized wild emmer diversity and developed introgression populations, although those efforts have focused primarily on durum wheat, with relatively few examples involving introgression from wild emmer into hexaploid backgrounds (Bacher et al., 2021; Chatzav et al., 2010; Fatiukha et al., 2020a, 2021; Hu et al., 2015; Kato et al., 1997, 1998; Li et al., 2001; Peleg et al., 2005, 2008a, 2008b, 2009a, 2009b; Peng et al., 2011, 2013; Rahman et al., 2023; Ren et al., 2013; Xie & Nevo, 2008).

The present study builds on and integrates these complementary approaches by combining the amphiploid bridge strategy employed by the WGRC with direct crossing with wild emmer. This project was initiated concurrently with the work described in Price et al. (2025a), which evaluated large-scale, systematic introgression from tetraploid to hexaploid wheat through direct crossing. Conducting these efforts in parallel provided a unique opportunity to directly compare introgression dynamics arising from two distinct crossing schemes and to assess whether patterns of donor genome introgression differ between bridge-mediated and direct approaches.

In the amphiploid bridge design, synthetic octoploid amphiploids were crossed with wild emmer accessions to generate wild introgression lines (WILs). This design is intended to facilitate introgression across all three wheat subgenomes: D subgenome diversity introduced via the amphiploid bridge alongside targeted A and B subgenome diversity derived from wild emmer. Although the expected donor contribution on the D subgenome in the resulting WILs is relatively low, the presence of duplicated D subgenome homologs in the synthetic octoploid is expected to mitigate the D subgenome aneuploidy present in direct hexaploid x tetraploid crosses (Price et al., 2025a). Consequently, this bridge-mediated strategy is expected to produce more fully hexaploid progeny with improved fertility and reduced chromosomal imbalance relative to direct wild emmer crosses.

The primary objective of this study is therefore to characterize the genome-wide architecture of introgression in these WIL populations, with particular emphasis on the successful introduction and retention of wild emmer allelic diversity on the A and B subgenomes within an adapted hexaploid wheat background. This work aims to map introgressed segments across the genome, evaluate differences in introgression patterns among populations, and assess

the extent to which this breeding strategy produces mosaics of introgression spanning all three subgenomes. Together, these analyses provide a comprehensive framework for evaluating the utility of amphiploid bridge-assisted introgression as a means of expanding allelic diversity in elite wheat germplasm.

Methods

Population development

Seven BC₂F_{4:5} *Triticum dicoccoides* germplasm development (TdGD) populations were generated using two hexaploid winter wheat (*T. aestivum*) recurrent parents (KS061406LN~26, KS061278M-4), two diploid *Ae. tauschii* ssp. *tauschii* accessions (TA 2398, TA 10155), and seven wild emmer (*T. dicoccoides*) accessions (PI 414718, PI 470736, PI 487264, PI 428088, PI 466984, PI 471815, PI 471021). The recurrent parents (RPs) were selected based on performance in yield trials in Kansas, whereas accessions of *T. dicoccoides* were selected to maximize geographic range/diversity.

Table 2.1. Source information for *Triticum dicoccoides* accessions used in germplasm development populations

Pedigrees for seven TdGD populations (Pop) are shown, along with geographic origin (source) of the wild emmer donor accession (bolded), growth habit (winter or spring), and elevation at the collection site, expressed as meters above sea level (masl).

Pop	Pedigree	Source	Growth habit	masl
30	KS061406LN~26*2/3/KS061406LN~26/TA 2398// PI 414718	Almagor, Israel	Spring	-74
32	KS061406LN~26*2/3/KS061406LN~26/TA 2398// PI 470736	Elazig, Turkey	Winter	890
34	KS061406LN~26*2/3/KS061406LN~26/TA 2398// PI 487264	Halab, Syria	Winter	350
41	KS061278M-4*2/3/KS061278M-4/TA 2398// PI 428088	Urfa, Turkey	Winter	1200
42	KS061278M-4*2/3/KS061278M-4/TA 2398// PI 466984	Bet Me'ir, Israel	Winter/Spring mix	467
46	KS061278M-4*2/3/KS061278M-4/TA 2398// PI 471815	Israel	Winter	Unk
80	KS061278M-4*2/3/KS061278M-4/TA 10155// PI 471021	Sanhedriyya, Israel	Winter	742

The *Ae. tauschii* accessions chosen in the development of the initial amphiploids were selected as a representation of geographic and allelic diversity (Nyine et al., 2020). Of the

available amphiploids, only two were used in the development of the populations described here. Selected amphiploid plants were visually free of clump dwarfism, a phenotype characterized by severe shoot growth reduction, excessive tillering, and delayed flowering. Clump dwarfism has been reported in interspecific wheat crosses and is hypothesized to arise from disruption of epistatic interactions among subgenomes and chromosomal imbalance associated with aneuploidy in diploid x tetraploid crosses (Matsuda et al., 2017; Okada et al., 2017).

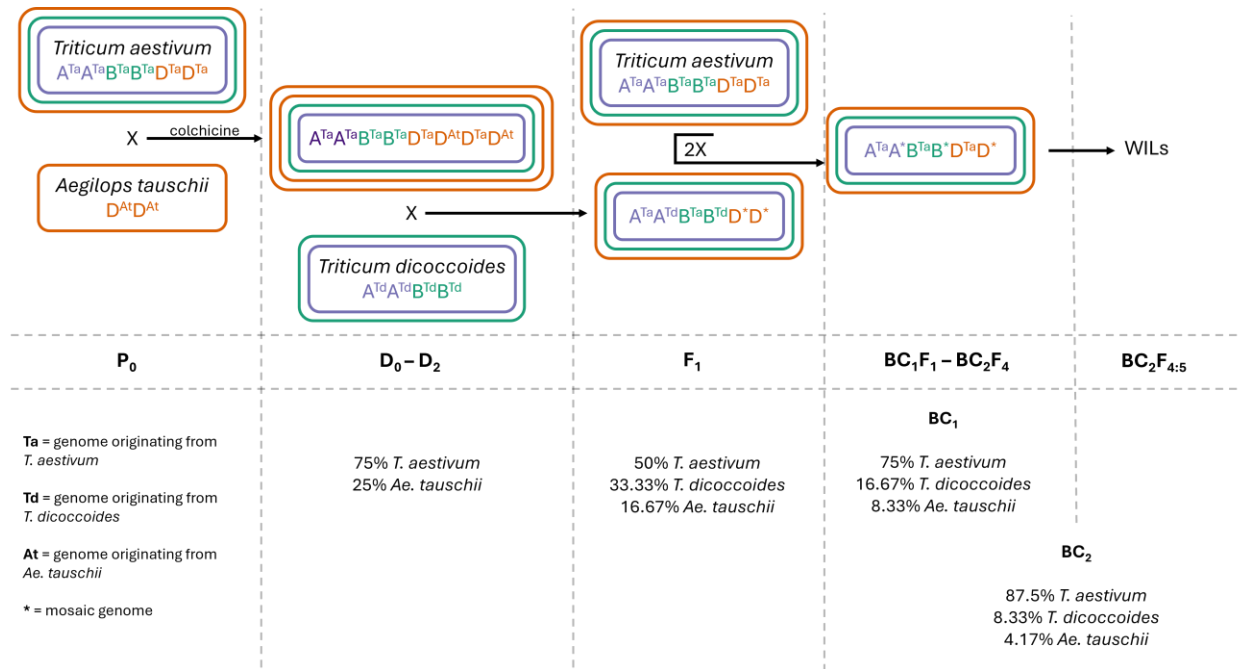


Figure 2.1. Schematic overview of the *Triticum dicoccoides* germplasm development pipeline

Triticum dicoccoides germplasm development (TdGD) populations were developed using an amphiploid bridge species for introgression of wild emmer alleles into hexaploid wheat. P₀: The hexaploid recurrent parent (*T. aestivum*; genome A^{Ta}A^{Ta}B^{Ta}B^{Ta}D^{Ta}D^{Ta}) is crossed with diploid *Ae. tauschii* (D^{At}D^{At}). Colchicine-induced chromosome doubling produces an octoploid amphiploid with genome composition A^{Ta}A^{Ta}B^{Ta}B^{Ta}D^{Ta}D^{Ta}D^{At}D^{At}. One to two generations of self-pollination (D₀-D₂) results in hybrids with an expected allelic contribution of 75% from *T. aestivum* and 25% *Ae. tauschii*. F₁: After selfing generations, the octoploid hybrids are crossed to tetraploid wild emmer (*T. dicoccoides*; genome A^{Td}A^{Td}B^{Td}B^{Td}). Although segregation for ploidy level can occur, the F₁ hexaploids theoretically carry an allelic contribution of 50% from *T. aestivum*, 33.33% from *T. dicoccoides*, and 16.67% from *Ae. tauschii*, with contribution from wild donor alleles on each subgenome. BC₁F₁ - BC₂F₄: The F₁ hybrids are backcrossed twice to the same hexaploid *T. aestivum* recurrent parent, with an expected allelic contribution of 87.5% from *T. aestivum*, 8.33% from *T. dicoccoides*, and 4.17% from *Ae. tauschii* after the second generation of backcrossing. BC₂F_{4:5}: Self-pollination of BC₂ progeny and derivation of individual F_{4:5} wild introgression lines (WILs) fixes recombinant inbred chromosomes, with generated WILs harboring mosaic genomes combining segments from *T. dicoccoides* and *Ae. tauschii* introgressed into the *T. aestivum* background.

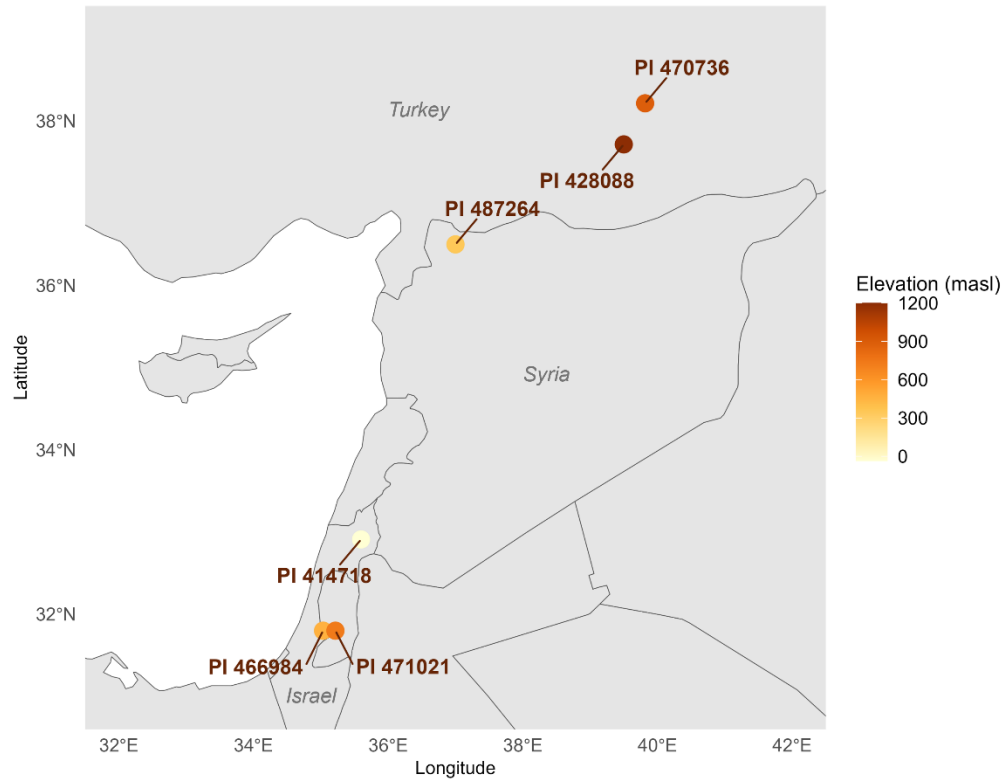


Figure 2.2. Location of origin for accessions used in *Triticum dicoccoides* germplasm development populations

Elevation is represented as meters above sea level (masl). PI 471815 is not represented due to unknown precise location and elevation information.

Development of synthetic octoploids is detailed in Nyine et al. (2020). Briefly, *T. aestivum* and *Ae. tauschii* crosses were made, with resulting hybrid plants regenerated from rescued embryos treated with colchicine (Dale et al., 2017). After one to two generations of self-pollination, amphiploids were used as female plants in crosses with *T. dicoccoides* to produce F₁ seed. To ensure adequate production of BC₁F₁ seed, a minimum of six crosses were made between F₁ plants and the respective hexaploid RP. From these, 10-16 BC₁F₁ plants per population were backcrossed again to the RP to generate BC₂F₁ families. BC₂F₁ plants were selfed to produce BC₂F₂ seed, and equal numbers of BC₂F₂'s derived from each BC₂F₁ family were entered into single seed descent, with a target population size of approximately 500 WILs per population. Attrition during advancement to BC₂F_{4.5} seed reduced realized population sizes

below the target. BC₂F_{4:5} plants were grown under greenhouse conditions to produce pure seed and collect leaf tissue for DNA extraction. Figure 2.1 illustrates the population development scheme.

Table 2.1 provides pedigrees for the seven TdGD populations, along with geographic details for the *T. dicoccoides* accessions. To maximize allelic diversity, the accessions were chosen from a broad geographic range (Figure 2.2). Wild emmer accessions used in populations 32, 34, and 41 were sourced from Turkey and Syria, while populations 30, 42, 46, and 80 were developed using wild emmer accessions from Israel. Both amphiploids used in population development were derived from crosses with *Ae. tauschii* ssp. *tauschii*, a lineage distinct from ssp. *strangulata*, which is recognized as the primary donor of the hexaploid wheat D subgenome (Dvorak et al., 1998; Nyine et al., 2020; Wang et al., 2013).

Array-based genotyping

DNA extraction was performed on greenhouse-grown BC₂F_{4:5} WILs using the BioSprint 96 DNA Plant Kit, following the Qiagen BioSprint DNA Plant Handbook (2005) protocol. Briefly, after grinding the leaf tissue to a fine powder in 96-well plates, lysate was prepared and loaded into plates along with the necessary buffers, wash solutions, and magnetic particles for automated purification. The protocol was initiated to automate DNA binding to magnetic particles, followed by sequential washing with buffer and ethanol to remove contaminants. Finally, purified DNA was eluted in distilled water in a fresh plate, yielding DNA suitable for downstream applications.

DNA quantification was conducted using a low-volume PicoGreen assay, automated on the Biomek and Mantis systems, with fluorescence measurements taken on the Omega plate reader. Each well in a black 384-well plate (shielded from light due to the dye's sensitivity) was

filled with 40 μL of 1x PicoGreen solution, with DNA standards and blanks added manually. DNA samples were diluted 10x using the Biomek system before being combined with PicoGreen in the plate. After a brief incubation, fluorescence was measured on the Omega reader at the appropriate excitation and emission settings. DNA concentrations were calculated using an Excel-based calculator provided by the USDA Central Small Grain Genotyping Lab (Manhattan, KS), and samples were normalized to approximately 50-100 ng/ μL by adding TRIS/Triton X buffer via the Mantis dispenser. The buffer helped prevent DNA degradation during shipment to the genotyping facility.

For shipment, 10 μL of each buffered DNA sample at $\sim$ 50-100 ng/ μL was transferred to new plates and frozen before being shipped to Agriculture Victoria Research (Bundoora, Victoria, Australia) under the care of Dr. Matthew Hayden. The BC₂F_{4:5} WILs were subsequently genotyped using the Illumina Infinium Wheat/Barley 40 K SNP array. This multi-species SNP array includes 25,363 markers for wheat, developed from exome sequencing of a globally diverse wheat germplasm collection. The array is enriched with trait-linked markers, tetraploid-specific markers (from wild and domesticated emmer, durum landraces, and cultivars), and *Ae. tauschii*-specific SNPs to track D subgenome contributions in synthetic derivatives (Keeble-Gagnère et al., 2021).

Quality control and population structuring analysis

Genotyping results were provided as raw SNP calls for all samples, with each genotype coded as 0 (reference), 1 (heterozygous), or 2 (alternate) relative to the IWGSC Chinese Spring assembly v1.0 (IWGSC, 2018). SNP clustering and genotype calling were conducted using GenomeStudio software, with a minimum SNP score threshold of 0.8 applied to ensure clustering accuracy.

After organizing the SNP data by population, quality control filtering was conducted within each of the seven TdGD populations. Markers with unknown chromosome or position were removed, and the resulting matrix was imported into the R package `snpReady` (Granato et al., 2018) for further filtering and an initial population genetic overview. Within each population, monomorphic markers and markers with a call rate below 0.9 (indicating more than 10% missing calls) were removed.

Using the `popgen` function within the `snpReady` package, population genetics metrics were calculated, including major and minor allele frequencies, observed heterozygosity per marker, observed heterozygosity per individual, inbreeding coefficients, effective population sizes, and polymorphic information content. Individual genotypes with heterozygosity levels above 20% were removed on a per-population basis, and markers with heterozygosity levels above 10% were also filtered out. Following these filtering steps within each population, the results were aggregated to retain all markers that passed the filtering criteria in at least one population.

The final filtered matrix, comprising 19,711 markers across 1,912 WILs, was processed in the R package `SNPRelate` (Zheng et al., 2012) to assess population structure. To facilitate efficient analysis, the matrix was first converted to the Genomic Data Structure (GDS) format. Principal components analysis (PCA) was performed to examine genetic structuring across populations, with the first two principal components (PCs) visualized. SNP correlations with the first two PCs were plotted along chromosomes to highlight genome-wide variation patterns and regions contributing to population differentiation. PCA was conducted both across all seven TdGD populations and within subsets based on RP lines (i.e., KS061406LN~26 for populations 30, 32, and 34, and KS061278M-4 for populations 41, 42, 46, and 80).

Sequencing-based genotyping

To enable inference of progeny haplotypes based on pedigree, the two *T. aestivum* RPs, two *Ae. tauschii* accessions, and seven *T. dicoccoides* accessions were deep sequenced. Whole genome shotgun sequencing (WGS) reads for KS061406LN~26 (SRX11411616), KS061278M-4 (SRX11411617), TA 2398 (SRX11411603), and TA 10155 (SRX11411609) were downloaded from NCBI's Sequence Read Archive (SRA) using SRA Toolkit's prefetch v2.10.0 (<https://github.com/ncbi/sra-tools>). These samples were uploaded by Kansas State University in August 2022 as part of a project to increase SNP density in parental lines for imputation into an *Ae. tauschii* introgression population (BioProject ID: PRJNA745927; Nyine et al., 2021). Paired-end whole genome shotgun libraries enriched for 400-600 bp inserts were prepared using the PIPEN Prep system on leaf tissue of two-week-old seedlings, then sequenced as 2x150 bp reads using the Illumina NovaSeq 6000.

SRA Toolkit's prefetch was also used to download raw exome-capture reads for three *T. dicoccoides* accessions – PI 471021 (SRX4978256), PI 428088 (SRX4978292), and PI 487264 (SRX4978300) from data uploaded to NCBI by Kansas State University in November 2018 as part of a wild and domesticated emmer diversity project (BioProject ID: PRJNA503847; Avni et al., 2017). Using shoot tissue of two-week-old seedlings, paired-end libraries were constructed using a whole exome capture protocol – targeting the non-redundant portion of the genome (Jordan et al., 2015) – and sequenced using the Illumina HiSeq2500.

All downloaded SRA files were converted to FASTQ format using SRA Toolkit's fasterq-dump, trimming adaptors and retaining only properly paired reads. Additional FASTQ files for four remaining wild emmer accessions (PI 466984, PI 471815, PI 470736, and PI 414718) were obtained internally via Kansas State University's WGRC. These libraries were

constructed and sequenced using the same Illumina paired-end whole genome sequencing protocol as described by Nyine et al. (2021), with final SRA identifiers pending public release.

The remaining filtered paired-end reads were aligned to the IWGSC RefSeq v1.0 wheat reference genome to maintain coordinate consistency with the array-based markers, which were also anchored to Chinese Spring v1.0 (IWGSC, 2018). Alignment was performed with BWA-MEM v0.7.17 (Li, 2013), supplying sample read-group tags via the -R flag. Resulting SAM files were sorted into BAMs with Picard SortSam, and PCR/optical duplicates were flagged using Picard MarkDuplicates (v2.25.1; <http://broadinstitute.github.io/picard/>). To balance accuracy and retention of true variants, alignments with mapping quality < 50 were removed with SAMtools view v1.16.1 (Li et al., 2009; Danecek et al., 2021). Read-group metadata were standardized via Picard AddOrReplaceReadGroups, and per-sample BAMs were indexed (CSI format) using SAMtools index. SAMtools depth was used to measure per-base sequencing depth, allowing for derivation of a sample-level average depth using all positions with non-zero coverage, compensating for differences in sequencing methodology.

Following alignment and reference-index preparation, variants were jointly called by subgenome and chromosome (21 VCFs total) using BCFtools mpileup (Danecek et al., 2021). D subgenome calls were restricted to the two hexaploid RPs and the two *Ae. tauschii* accessions (n=4), whereas A and B subgenomes included the two RPs and the seven *T. dicoccoides* accessions (n=9). To minimize spurious pileups in repetitive or misaligned regions, a hard cap of 250 reads per site was imposed in mpileup. Since the WGS and exome capture approaches were targeted for relatively low coverage and observed depths rarely approached this limit, sites exceeding 250 reads were interpreted as artifactual alignment rather than true biological variation.

VCF outputs were then filtered with BCFtools view to (1) retain only biallelic SNPs, (2) impose a minimum site-level depth (DP) of 10 – ensuring at least modest support for alternate alleles, as $DP < 10$ typically produced only 1-2 reads for the alternate allele, and (3) enforce a maximum DP of 60, chosen at the inflection where heterozygote allele-balance (AD/DP) began to deviate from ~ 0.5 (Appendix Figure A.1). Although most sites exhibited very high quality scores ($QUAL > 100$), a conservative $QUAL \geq 30$ filter was applied to remove residual low confidence calls. Finally, to guard against sample-specific dropout, variants were required to be non-missing in at least half of the samples per subgenome group ($\geq 2/4$ for D subgenome; $\geq 4/9$ for A/B subgenomes). These sequential filters – biallelic status, depth (10-60), quality ($QUAL \geq 30$), and call rate – yield a high confidence variant set with reliable depth, quality, and representation across samples.

Due to methodological differences in sequencing approaches – specifically differences in the targeted genomic space between WGS and exome capture – additional filtering of variant sites was performed per population. For each population, markers were retained only if they were non-missing, homozygous, and polymorphic in both parents (the recurrent hexaploid and wild emmer donor for A and B subgenome chromosomes, and the recurrent hexaploid and *Ae. tauschii* donor for D subgenome chromosomes). This filtering strategy ensured a complete and informative marker set that prioritizes per population parental divergence, thereby enabling reliable discrimination between recurrent and donor derived alleles in the progeny in downstream analyses.

Despite this filtering, the resulting genotype matrices remained prohibitively large, posing substantial computational constraints during matrix transposition prior to imputation. To further reduce matrix size, marker thinning was performed using a minimum inter-marker

distance criterion. Minimum distances ranging from 50 kb to 250 kb were evaluated to identify a computationally tractable, yet informative, subset of markers.

To independently assess whether KS061278M-4 carries the 1BL.1RS translocation from rye (*Secale cereale*), its WGS reads were additionally aligned using a competitive reference approach following methodology of Parker et al. (2026). A composite reference was constructed by combining the Jagger genome assembly (Walkowiak et al., 2020) with chromosome 1R from the Weining rye assembly (Li et al., 2021). WGS reads for Larry (SRX9970578), a known carrier of the 1BL.1RS translocation, were additionally downloaded from the NCBI SRA and processed in parallel as a positive control. Alignment coverage was summarized in non-overlapping 1 Mb windows across wheat chromosome 1B and rye chromosome 1R. The 1BL.1RS translocation was diagnosed from the combined pattern of reduced coverage across the wheat 1BS arm and increased coverage across the corresponding rye 1RS arm, relative to the coverage profile of the Larry positive control (Parker et al., 2026).

Nucleotide diversity between population founders

Alongside the variant filtering performed for imputation, an additional set of population-specific parental marker matrices was generated to quantify founder divergence within each population. For each population x chromosome x subgenome combination, variants were retained only if both founders were non-missing and homozygous at that site. These files were used to assess pairwise nucleotide diversity between the RP and the corresponding donor parent contributing to each population and subgenome. In contrast to the marker sets curated for imputation, these matrices were not filtered for polymorphism between founders, as restricting analyses to polymorphic sites would trivially yield complete nucleotide diversity across all retained variants. Additionally, marker thinning was not applied, as matrix transposition and

downstream computational constraints were not limiting for the calculation of pairwise founder divergence.

For each population, specific parent pair files (RP vs. A/B subgenome donor; RP vs. D subgenome donor) were generated and variants were assigned to fixed, non-overlapping physical windows of 1 Mb. Within each bin, pairwise nucleotide diversity was estimated as the proportion of callable variants for which the RP and donor allele states differed, calculated as the number of sites with discordant parental alleles divided by the total number of callable sites in the bin. These per-bin values represent local estimates of founder divergence across the genome.

Founder nucleotide diversity was summarized by averaging per-bin estimates within chromosomes, subgenomes, and across the genome for each population. Per-bin divergence estimates were additionally plotted by physical position and faceted by chromosome to visualize spatial patterns of founder divergence across the A, B, and D chromosomes within each population.

Preparation of sequence- and array-based genotypes for imputation

Following across- and within-population filtering of the parental sequence data for imputation, concordance in physical marker positions between the progeny array data and the parental sequence data was limited. To resolve this mismatch and establish a shared marker framework, variants were recalled in the parental sequence data at sites corresponding to markers retained in the progeny array dataset. Variant recalling was performed using the -R option in BCFtools mpileup (Danecek et al., 2021), which restricts variant calling to a user defined set of target regions, with a maximum BAM depth capped at 250 reads per site.

Recalled variants were subsequently used to filter array-based markers across the WIL progeny and population founders within each population. Consistent with earlier quality control

steps, markers were removed if either of the two parents for a given population and subgenome were missing or heterozygous, or if no allelic divergence was observed between the RP and the donor. Retention of this shared and informative marker set between founders and progeny is essential for accurate imputation; markers that are unobserved in founders limit meaningful imputation into sparsely genotyped progeny, as imputation relies on the propagation of founder haplotypes through pedigree structure (Calus et al., 2014; Hickey et al., 2012a, 2012b; Whalen et al., 2020). Moreover, allelic divergence between the RP and donor enables unambiguous discrimination of recurrent- versus donor-derived alleles in the progeny.

Recalled variants were then merged, per chromosome and population, with the existing parental sequence data and the progeny array data to generate a combined genotype matrix. This matrix contained complete genotype information for the parental founders, including recalled variants introduced to establish positional concordance with the progeny array, while progeny genotypes remained sparse and were coded as missing at markers unique to the parental sequence data. During merging, reference (REF; defined relative to Chinese Spring v1.0) and alternate (ALT) allele identities were retained for both datasets, and any markers exhibiting discordance in REF/ALT identity or orientation between parental and progeny sources were removed.

Pedigree-based imputation of sparsely genotyped progeny

Combined marker matrices contained WILs genotyped at all array SNPs, and population founders (RP, wild emmer donor, *Ae. tauschii* donor) genotyped at sequencing-derived variants across the genome as well as at concordant array SNP sites, were used as input for AlphaImpute2. AlphaImpute2 is a pedigree-based imputation algorithm that infers missing genotypes in low-density individuals by propagating shared haplotype information from high-

density relatives using an approximate multi-locus iterative peeling framework (Whalen et al., 2020). This approach was selected for its ability to accommodate complex pedigrees spanning multiple breeding generations, which is essential for introgression populations derived through repeated backcrossing and selfing (Hickey et al., 2012b).

Imputation requires a pedigree file specifying parent-offspring relationships for each WIL, including population founders and intermediate generations across the breeding scheme. A subset of the pedigree structure used is provided in Appendix Table A.1. For each individual at each locus, AlphaImpute2 estimates genotype probabilities as the product of three components: a penetrance term, an anterior term, and a posterior term (Whalen et al., 2020). The penetrance term reflects the likelihood of each genotype given an individual's observed genetic data. The anterior term propagates information from parents to offspring, while the posterior term propagates information from offspring back to parents. In this dataset, where progeny are sparsely genotyped and founders are densely genotyped, genotype inference for WILs is expected to be driven primarily by the anterior term, whereas the posterior term is expected to contribute minimally because WILs have no offspring and founders do not have descendants with higher marker density.

To incorporate linkage information, AlphaImpute2 conditions both anterior and posterior updates on an individual's segregation state along the chromosome, defined as the parental haplotypes inherited at each locus (Whalen et al., 2020). Segregation probabilities are estimated across loci using a hidden Markov model, in which emission probabilities quantify how consistent each segregation state is with the current genotype estimates of the individual and its parents under Mendelian transmission, while transition probabilities penalize changes in segregation state between adjacent loci according to an assumed recombination rate. The

recombination model is parameterized by a specified chromosome map length (in Morgans). Increasing the map length increases the recombination probability assigned to each marker interval, thereby allowing segregation states to change more readily between loci, whereas smaller map lengths enforce stronger continuity of segregation state.

Pedigree structure itself influences the extent to which uncertainty propagates during imputation. In pedigrees where founders and WILs are genotyped but intermediate generations (e.g., BC₁F₁, BC₂F₁) are ungenotyped, uncertainty accumulates across generations. Ungenotyped intermediates are frequently imputed as heterozygous or remain unresolved, depending on genotype calling thresholds; heterozygosity in this context represents uncertainty in haplotype assignment. When segregation states are uncertain and penetrance provides no direct information, inheritance from these intermediates is resolved probabilistically based on their genotype probabilities, which increasingly reflect the dominant haplotype background as allele frequencies across the pedigree become skewed toward the RP. As a result, repeated transmission through uncertain intermediates progressively downweights donor inheritance unless it is repeatedly re-anchored by observed genotype evidence. Over multiple generations, this leads to gradual erosion of donor signal and an apparent collapse toward the RP background.

To limit this effect, WILs were represented as being derived directly from the final backcross generation rather than explicitly modeling all intervening generations of selfing (Appendix Table A.1). This reduces the number of uncertainty resolution steps required during peeling and helps preserve donor haplotypes that are supported by empirical markers. Similarly, restricting founder marker sets to loci informative for distinguishing donor and RP haplotypes reduces reliance on population-level expectations and strengthens segregation state inference.

Final genotype calls were generated after multiple peeling cycles using the default number of iterations and a specified final peeling threshold (FPEEL). This threshold controls the tradeoff between missingness and heterozygosity in the output genotypes. A conservative threshold (e.g., FPEEL = 0.95) yields only high-confidence calls but results in substantial missing data, whereas the default threshold (FPEEL = 0.1) outputs best guess genotypes, producing a complete dataset at the cost of retaining heterozygous calls in regions of unresolved haplotype inheritance (Whalen et al., 2020). Because heterozygosity in this context primarily reflects uncertainty arising from long stretches of missing data between array anchors, heterozygotes were retained in favor of missingness. Therefore, a post imputation heterozygote smoothing step was applied, in which stretches between progeny marker anchors were filled with donor or RP haplotypes when both flanking markers supported the same ancestry. In regions where ancestry could not be determined from flanking markers, heterozygous calls were retained and interpreted as reflecting unresolved segregation states.

Introgression mapping in imputed progeny

Following pedigree-based imputation of WIL genotypes and ancestry-based heterozygote smoothing between flanking anchor markers, donor introgression was quantified in the progeny by computing donor allele proportions in fixed physical windows. Each chromosome was divided into 1 Mb bins, and within each bin the number of alleles matching the population-specific donor (by subgenome) was counted and divided by the total number of alleles observed in that bin. Homozygous genotype calls therefore contributed two alleles, while heterozygous calls contributed one allele to the donor and one to the RP. Under this framework, extended stretches of heterozygosity are interpreted as uncertainty in ancestry assignment rather than definitive evidence of donor presence/absence, reflecting either low progeny anchor marker

density or disagreement between flanking segregation states that prevents confident haplotype resolution.

Although most 1 Mb bins contained sufficient marker support, some bins were sparsely populated or empty, making direct classification of individual bins as donor- or RP-derived inherently noisy. To stabilize ancestry inference across the genome, a moving average (MA) approach adapted from Price et al. (2025a) was used to incorporate information from neighboring bins when calling donor segments. For each focal 1 Mb bin, donor proportions from flanking bins were averaged to compute a smoothed donor proportion. Donor ancestry was then called if this MA value exceeded a specified threshold, using two complementary scales: a short-range MA with a stringent threshold (0.9) and a long-range MA with a more permissive threshold (0.7). For example, donor ancestry could be assigned if the mean donor proportion across the 5 bins upstream and downstream of a focal bin (MA10) exceeded 0.9, or if the mean across the 20 bins upstream and downstream (MA40) exceeded 0.7.

These thresholds were intentionally conservative to prevent donor segments from being called solely due to heterozygous contributions. Because a series of heterozygous bins would yield an expected donor proportion of 0.5, the chosen thresholds ensure that called segments reflect consistent donor support rather than propagated uncertainty. Multiple MA window sizes were evaluated to assess sensitivity of segment detection to the scale of smoothing. Short-range MA windows (threshold = 0.9) were tested at sizes of 2, 6, 10, 20, and 50 bins, while long-range MA windows (threshold = 0.7) were tested at sizes of 10, 20, 40, 60, and 100 bins. The smallest and largest window sizes represent exploratory extremes, with smaller windows expected to be noisier and larger windows increasingly oversmoothed; nonetheless, comparison across parameterizations was informative for assessing robustness of inferred introgression patterns.

Given the overall concordance among MA window size pairs, MA10/MA40 was selected for downstream segment calling as a pragmatic balance between sensitivity and robustness.

After donor bins were identified under a given parameterization, adjacent donor bins were merged to define continuous introgression segments within each WIL. While this MA-based approach introduces uncertainty in the precise locations of recombination breakpoints, it is preferred over bin-level hard calling based on potentially sparse marker information and provides stable, genome-wide estimates of donor ancestry. Given that introgression boundaries are inferred from imputed data rather than direct empirical observation, some ambiguity in segment boundaries is expected, and downstream interpretation of segment locations was therefore conducted with appropriate caution.

Empirical validation of introgression calls was conducted using a KASP assay targeting the domestication locus Q/q (Faris et al., 2003; Simons et al., 2006; Zhang et al., 2020). Because wild emmer carries the wild type Q allele and the RPs carry the domesticated q allele, WILs with a called introgression overlapping the Q/q interval are expected to genotype as Q , whereas WILs without an overlapping introgression are expected to genotype as q . Pop30 and Pop41 were selected for validation, being the two populations with the most WILs that have segments called near or overlapping Q/q .

Results

Array-based genotyping

Array-based SNPs from the Wheat/Barley 40 K SNP chip were filtered within each TdGD population to retain polymorphic markers with call rates $\geq 90\%$, per-marker heterozygosity $\leq 10\%$, and per-individual heterozygosity $\leq 20\%$ (Table 2.2). The retained markers and individuals were then examined through distributions of MAF, observed

heterozygosity at the marker level, and observed heterozygosity at the individual level. Figure 2.3 shows the per-population distributions of MAF, which were largely consistent across populations and strongly skewed toward low frequency alleles. This pattern is expected for BC₂-derived populations, in which recovery of the RP genome results in low expected donor allele frequencies of ~8% on the A and B subgenomes and ~4% on the D subgenome (Figure 2.1). Among populations, Pop30 exhibited the highest average MAF (0.12), while Pop32 showed the lowest (0.07; Table 2.3).

Filtering based on average individual heterozygosity removed very few WILs overall, reflecting the high degree of homozygosity expected in advanced backcross inbred materials (Appendix Figure A.3). Pop80 showed the largest number of individuals exceeding the heterozygosity threshold, and correspondingly also exhibited the greatest number of markers removed due to excessive marker-level heterozygosity (Appendix Figure A.4). Despite these removals, application of polymorphism, heterozygosity, and missingness filters retained a substantial proportion of the original 25 K wheat-specific SNPs from the array in each population. Pop30 retained the largest number of markers (15.7 K), whereas Pop80 retained the fewest (10.6 K).

The coefficient of inbreeding (F) reflects the degree to which individual WILs are more homozygous than expected under a random mating reference model given the observed allele frequencies (Granato et al., 2018), and thus serves as a descriptive index of genome-wide fixation rather than pedigree-based inbreeding. Effective population size (N_e) is calculated from the average inbreeding coefficient and provides a relative measure of overall homozygosity and genetic diversity within each population (Granato et al., 2018), rather than a true demographic estimate. Polymorphic information content (PIC) describes the informativeness of individual

markers based on allele frequencies, with higher values indicating more balanced allele frequencies and, consequently, greater capacity to distinguish allelic variation among WILs (Granato et al., 2018). Together, these metrics provide complementary summaries of genome-wide fixation within individuals and the distribution of allelic variation across markers and populations.

Table 2.2. Number of array-based markers retained in WIL progeny and TdGD population founders

Counts are summarized by chromosome (Chr) and population. Progeny genotypes derived from the wheat-barley SNP array were filtered within each population for call rate $\geq 90\%$, heterozygosity $\leq 10\%$, and removal of monomorphic markers. The physical positions of retained array markers were then used to define targeted variant calling regions in the population founders using BCFtools mpileup. Following merging of parental and progeny datasets, markers were further pruned to retain only those that were non-missing, homozygous, and polymorphic between the RP and donor for each population x subgenome combination. Variant counts shown represent the final shared marker set used for downstream imputation.

Chr	Variant Count						
	Pop30	Pop32	Pop34	Pop41	Pop42	Pop46	Pop80
1A	231	258	228	227	172	170	163
1B	282	165	243	205	192	194	205
1D	149	125	140	90	121	118	70
2A	302	283	361	233	217	292	253
2B	321	290	298	170	195	209	200
2D	196	160	168	135	132	176	89
3A	182	228	296	204	212	212	177
3B	332	284	320	207	215	194	251
3D	136	70	102	122	88	119	66
4A	231	188	209	157	143	138	144
4B	292	191	310	206	265	242	225
4D	63	65	53	35	69	70	35
5A	265	237	257	183	204	227	231
5B	322	296	269	220	248	215	181
5D	133	104	140	107	103	123	53
6A	175	147	153	120	149	140	151
6B	189	196	240	152	205	197	206
6D	104	86	110	78	74	68	36
7A	316	256	310	205	193	213	193
7B	299	260	271	177	217	145	231
7D	135	68	129	109	96	117	74

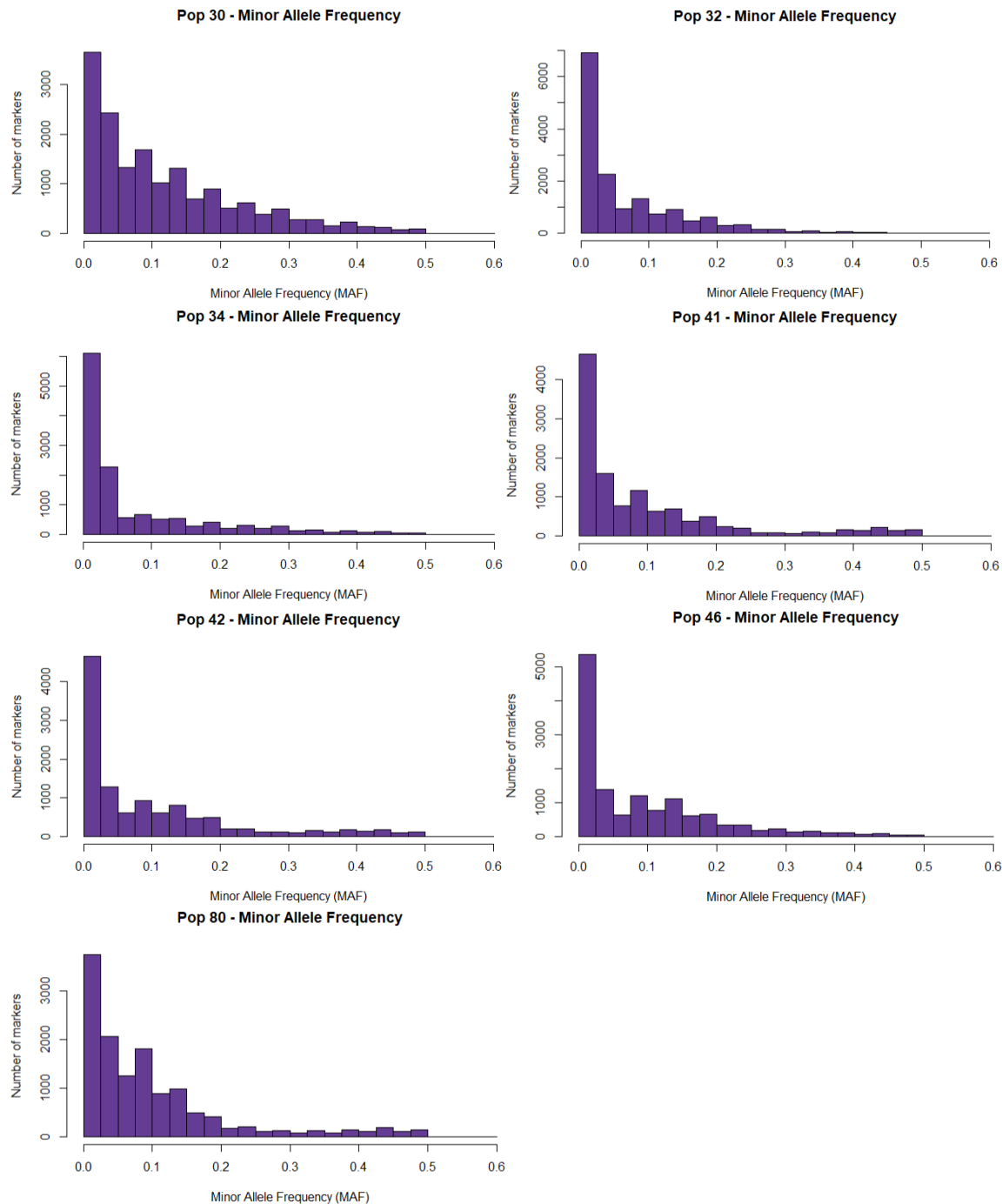


Figure 2.3. Distribution of MAF across markers for each TdGD population

MAF was calculated from observed allele frequencies using the R package *snpReady* (Granato et al., 2018) and summarizes the degree of marker polymorphism within populations.

Population-level differences in these summary statistics were largely consistent with patterns of individual heterozygosity, marker allele frequencies, and population size (Table 2.3).

Pop80, with its highest average individual heterozygosity correspondingly had the lowest

average F (0.62), indicating increased genome-wide residual allelic state segregation within individual WIL genotypes relative to the other populations. As a result of both its lower F and its larger number of retained WILs (334), Pop80 also showed the largest N_e (293). In contrast, Pop34 retained the fewest WILs (189) which consequently resulted in the smallest N_e (127). Pop30 showed the highest F (0.87), consistent with having the lowest average individual heterozygosity; the relatively large number of retained WILs in Pop30 (327) partially offset the effect of increased homozygosity, yielding an intermediate N_e (189) rather than the smallest among populations. Additionally, Pop30 had the highest PIC (0.16) due to having the highest MAF among populations. In contrast, Pop32 had the lowest PIC (0.10), consistent with having the lowest MAF.

Table 2.3. Population genetic summary statistics by TdGD population

Shown are the number of polymorphic markers per population, the number of markers retained after filtering for marker call rate $\geq 90\%$ and marker heterozygosity $\leq 10\%$, and the number of WILs retained after filtering for individual heterozygosity $\leq 20\%$. Population-level averages of observed individual heterozygosity (Het) and minor allele frequency (MAF) are reported, along with the average coefficient of inbreeding (F), effective population size (N_e), and polymorphic information content (PIC). Estimates are derived from allele frequency-based calculations. Summary statistics were computed using the popgen function in the R package snpReady (Granato et al., 2018).

Pop	Marker count - polymorphic	Marker count - filtered	WIL count	Het	MAF	F	N_e	PIC
30	20954	15718	327	0.02	0.12	0.87	189	0.16
32	21516	14585	210	0.03	0.07	0.74	151	0.10
34	16941	12533	189	0.03	0.08	0.79	127	0.11
41	17271	11012	278	0.03	0.09	0.78	186	0.12
42	18143	10633	299	0.03	0.10	0.78	200	0.13
46	17656	12634	272	0.03	0.09	0.78	181	0.13
80	19181	10621	334	0.06	0.10	0.62	293	0.13

Sequencing-based genotyping

Sequencing-based genotyping of TdGD population founders varied across samples in both sequencing methodology and depth. The two hexaploid RPs, the two *Ae. tauschii* D subgenome diploid donors, and four of the seven *T. dicoccoides* tetraploid A and B subgenome

donors were sequenced using WGS sequencing, whereas the remaining three wild emmer tetraploid donors were sequenced using an exome capture approach. Sequencing depth also differed by sample; the hexaploid and diploid parents were sequenced at approximately 8-10x coverage, while tetraploid donors were sequenced at approximately 4-5x (Table 2.4).

Table 2.4. Sequencing information for the founders used in TdGD population development

Each line is broken down by sample/accession identifier, NCBI source information (when available), sequencing methodology (WGS = whole genome shotgun sequencing; EC = exome capture), post-filter read count, in millions (M; after adaptor trimming and removal of unpaired reads), mean read length (bp), mapped read count (M), and average depth across sequenced positions (X).

Sample ID	NCBI SRA ID	Method	Read Count	Read Length	Mapped Read Count	Average Depth
KS061406LN~26	SRX11411616	WGS	901.95 M	144	522.08 M	8.67
KS061278M-4	SRX11411617	WGS	1248.84 M	144	719.64 M	9.16
TA2398	SRX11411603	WGS	392.26 M	144	122.59 M	8.54
TA10155	SRX11411609	WGS	444.55 M	144	145.34 M	10.01
PI487264	SRX4978300	EC	71.66 M	124	40.10 M	4.11
PI428088	SRX4978292	EC	28.44 M	94	15.47 M	2.73
PI471021	SRX4978256	EC	62.05 M	124	34.09 M	3.76
PI466984	-	WGS	485.95 M	144	238.38 M	5.05
PI471815	-	WGS	515.26 M	144	255.92 M	5.15
PI470736	-	WGS	485.48 M	144	263.83 M	5.27
PI414718	-	WGS	377.86 M	144	177.12 M	4.09

Despite these methodological differences, preliminary variant filtering decisions were largely consistent across samples. Reads were aligned to the Chinese Spring v1.0 reference genome, filtered for mapping quality ≥ 50 , and jointly variant called used BCFtools mpileup, with a cap of 250 mapped reads per site. Resulting variants were filtered to retain sites with depth between 10 and 60, variant quality ≥ 30 , and preliminary missingness filtering requiring presence in at least half of the samples within a given subgenome. Following these steps, the majority of variants were retained across all samples (Appendix Table A.2). Among all founders, PI 428088 (the wild emmer donor for Pop41) retained the smallest proportion of variants, consistent with it having the lowest average read depth across genotyped sites (Table 2.4, Appendix Table A.2). Nevertheless, after further filtering within population for homozygosity,

non-missingness, and polymorphism between the RP and donor for each population x subgenome combination, approximately 650 K and 720 K variants remained on the A and B subgenomes, respectively, for Pop41. Founder variant density in Pop41 remained orders of magnitude higher than that available in the WIL progeny. Given that the BC₂F_{4:5} generation has undergone relatively few meiotic recombination events, and that pedigree-based haplotypic imputation relies largely on flanking informative markers and population-level segregation patterns (Hickey et al., 2012a; Whalen et al., 2020), parental marker density was not considered a constraint in this population.

Because imputation accuracy is sensitive to amount of missing data, and redundant marker information can proliferate bias (Hickey et al., 2012a, 2012b; Whalen et al., 2020), parental sequencing-derived variants were further thinned to improve the relative contribution of backbone markers observed in the progeny. Minimum inter-marker distance thinning was performed at thresholds of 50, 100, 150, 200, and 250 kb. Prior to thinning, D subgenome variants were consistently more abundant than A or B subgenome variants, reflecting the greater number of homozygous, non-missing, and polymorphic sites between the hexaploid RPs and *Ae. tauschii* donors compared to the RP x *T. dicoccoides* pairs (Table 2.5). This disparity was most pronounced in populations whose wild emmer donors were sequenced using exome capture, as expected given the reduced genome space targeted by that approach (Jordan et al., 2015, 2018). Progressive thinning reduced these genome-specific disparities. At a minimum distance of 50 kb, fewer D subgenome variants than A or B variants were retained in all populations except Pop41; in Pop41, D subgenome counts did not drop below A and B variant counts until a thinning threshold of 150 kb. Ultimately, a minimum marker distance of 250 kb was selected for downstream analyses. This choice was motivated by the dependence on ancestry-based

imputation; beyond a certain density, additional markers contribute redundancy without improving resolution of recombination breakpoints in the progeny (Whalen et al., 2020).

Table 2.5. Variant counts across successive marker thinning steps, by subgenome and TdGD population

Counts are expressed in millions (M) or thousands (K). Initial counts ($10 \leq DP \leq 60$, $QUAL \geq 30$, $MISS \leq 50\%$) reflect quality filtered variants across all populations prior to filtering by within population pairs of founders. Per population counts following founder filtering include only markers that were homozygous, non-missing, and polymorphic between the recurrent hexaploid parent and the appropriate donor parent (wild emmer for A and B subgenomes; *Ae. tauschii* for the D subgenome). Thinned variant counts are shown for minimum inter-marker distance thresholds ranging from 50 kb to 250 kb, applied to reduce matrix dimensionality prior to imputation.

Subg	Pop	Variant Count		Thinned				
		10≤DP≤60, QUAL≥30, MISS≤50%	Homozygous, non-missing, polymorphic					
				50kb	100kb	150kb	200kb	250kb
A	All	46.94 M	-	-	-	-	-	-
B	All	55.19 M	-	-	-	-	-	-
D	All	28.45 M	-	-	-	-	-	-
A	30	-	19.80 M	92.22 K	47.30 K	31.87 K	24.06 K	19.33 K
B	30	-	22.78 M	93.07 K	48.24 K	32.71 K	24.76 K	19.96 K
D	30	-	23.64 M	72.73 K	37.63 K	25.42 K	19.21 K	15.45 K
A	32	-	13.40 M	91.00 K	46.97 K	31.75 K	24.00 K	19.30 K
B	32	-	15.66 M	92.27 K	48.10 K	32.66 K	24.74 K	19.94 K
D	32	-	23.64 M	72.73 K	37.63 K	25.42 K	19.21 K	15.45 K
A	34	-	2.55 M	84.12 K	44.95 K	30.72 K	23.37 K	18.88 K
B	34	-	2.97 M	84.11 K	45.32 K	31.20 K	23.86 K	19.33 K
D	34	-	23.64 M	72.73 K	37.63 K	25.42 K	19.21 K	15.45 K
A	41	-	0.65 M	59.84 K	35.80 K	25.88 K	20.33 K	16.80 K
B	41	-	0.72 M	61.31 K	36.85 K	26.73 K	21.02 K	17.38 K
D	41	-	23.75 M	72.88 K	37.69 K	25.45 K	19.24 K	15.46 K
A	42	-	16.07 M	93.07 K	47.62 K	32.03 K	24.16 K	19.39 K
B	42	-	19.62 M	94.17 K	48.65 K	32.95 K	24.91 K	20.05 K
D	42	-	23.75 M	72.88 K	37.69 K	25.45 K	19.24 K	15.46 K
A	46	-	11.94 M	92.97 K	47.62 K	32.05 K	24.17 K	19.40 K
B	46	-	14.18 M	94.87 K	48.92 K	33.08 K	25.01 K	20.12 K
D	46	-	23.75 M	72.88 K	37.69 K	25.45 K	19.24 K	15.46 K
A	80	-	2.58 M	84.41 K	45.07 K	30.81 K	23.43 K	18.90 K
B	80	-	3.14 M	84.80 K	45.58 K	31.34 K	23.91 K	19.36 K
D	80	-	24.04 M	72.98 K	37.74 K	25.47 K	19.25 K	15.47 K

Marker thinning disproportionately affected low recombination pericentromeric regions of the final imputation matrices, specifically in WGS data. This effect was partially driven by the post-thinning re-addition of variants called at array-concordant sites; these array markers are biased toward regions of higher recombination and tend to be well captured within the exome, while large gaps in array marker coverage persist near centromeres (Appendix Figure A.2; Keeble-Gagnère et al., 2021). The effects of thinning under different sequencing methodologies are illustrated in Appendix Figure A.5, which compares Pop41 (exome capture) and Pop30 (WGS). At a 50 kb thinning threshold, Pop41 exhibits pronounced regions of reduced variant density on the A and B subgenomes, whereas Pop30 shows relatively uniform coverage, and the D subgenome in Pop41 remains comparatively dense. After further thinning to 100 kb, variant density in Pop30 begins to concentrate away from centromeres, reflecting uneven removal of markers from WGS data. By 250 kb, variant density distributions across chromosomes become broadly comparable between sequencing methodologies, with similar marker counts per 5 Mb bin between approaches.

Overall, the combined filtering and thinning strategy produced imputation matrices with comparable variant density and genome distribution across populations, regardless of sequencing methodology. Importantly, no substantial population-level differences were noticed when observing variant density of imputation matrices and the genomic regions where progeny array backbone markers are concentrated (Appendix Figure A.2).

Allelic differentiation between TdGD populations and founders

Drivers of allelic differentiation within and between TdGD populations, and their contribution to observed population structure, were evaluated using two complementary approaches: principal component analysis (PCA) based on array SNPs in WIL progeny, and

pairwise nucleotide diversity analyses between population founders using sequencing-deriving variants. Together, these approaches were used to characterize genome-wide patterns of differentiation among populations and to assess whether population structure could be attributed to specific genomic regions.

PCA of progeny genotypes

Filtered, population specific SNP array datasets were merged to generate a joint genotype matrix containing approximately 20 K markers across 1,912 WILs. This matrix was used to perform PCA in the R package *snpRelate* (Zheng et al., 2012). PCA was conducted both across all seven TdGD populations and within groups defined by RP: populations derived from KS061416LN~26 (Pop30, Pop32, Pop34) and populations derived from KS061278M-4 (Pop41, Pop42, Pop46, Pop80).

Across all populations, the first principal component (PC1) was associated with 9.8% of the total genotypic variation, while PC2 accounted for 6.5%, together corresponding with 16.3% of the observed variation (Figure 2.4). PCA revealed two distinct clusters corresponding to RP background, with PC1 effectively separating populations derived from KS061406LN~26 from those derived from KS061278M-4. Within RP groupings, differences in the magnitude of population structure were observed. Among the KS061406LN~26 derived populations, PC1 and PC2 explained 12.6% and 5.5% of variation, respectively. Among the KS061278M-4 derived populations, PC1 and PC2 explained 5.7% and 2.8% of variation, respectively (Figure 2.4). In the across population PCA, PC2 further separated a subset of WILs from Pop30 into a cluster distinct from the remaining WILs.

Pop30 is unique among the TdGD populations in that it was derived from a wild emmer accession with a spring growth habit and thus, no vernalization requirement (Table 2.1). Allelic

variation at *Vrn-A1* – a major-effect vernalization locus location on chromosome 1A (DeWitt et al., 2021; Kato et al., 1997; Strejčková et al., 2023) – was used to assess correspondence between vernalization requirement and structuring within Pop30. A trait-linked SNP for *Vrn-A1* present on the Wheat/Barley SNP array segregated in Pop30 and was used to color PCA loadings by genotype class. No clear connection between PCA clustering and *Vrn-A1* allelic state was observed (Appendix Figure A.6).

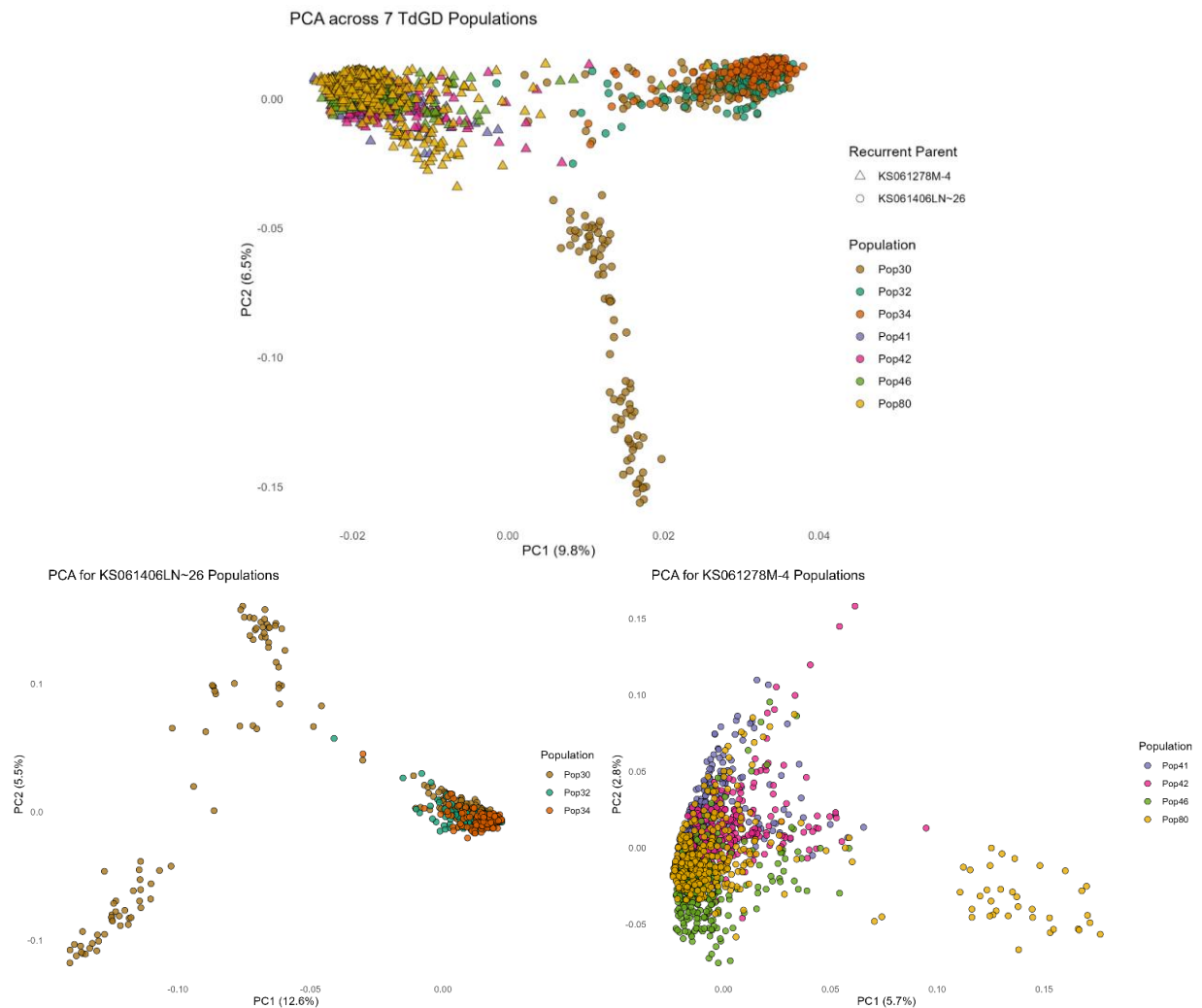


Figure 2.4. Principal component analysis of WIL progeny across TdGD populations

The top panel shows PCA results for all seven TdGD populations analyzed jointly, with points colored by population and shaped by RP background. The bottom panels show PCA results stratified by RP: populations derived from KS061406LN~26 (Pop30, Pop32, Pop34; bottom left) and populations derived from KS061278M-4 (Pop41, Pop42, Pop46, Pop80; bottom right). PCA was performed using filtered array-based SNPs with the R package *snpRelate* (Zheng et al., 2012). The first two principal components are shown in each panel, with the percentage of total variance explained by each component indicated on the axes.

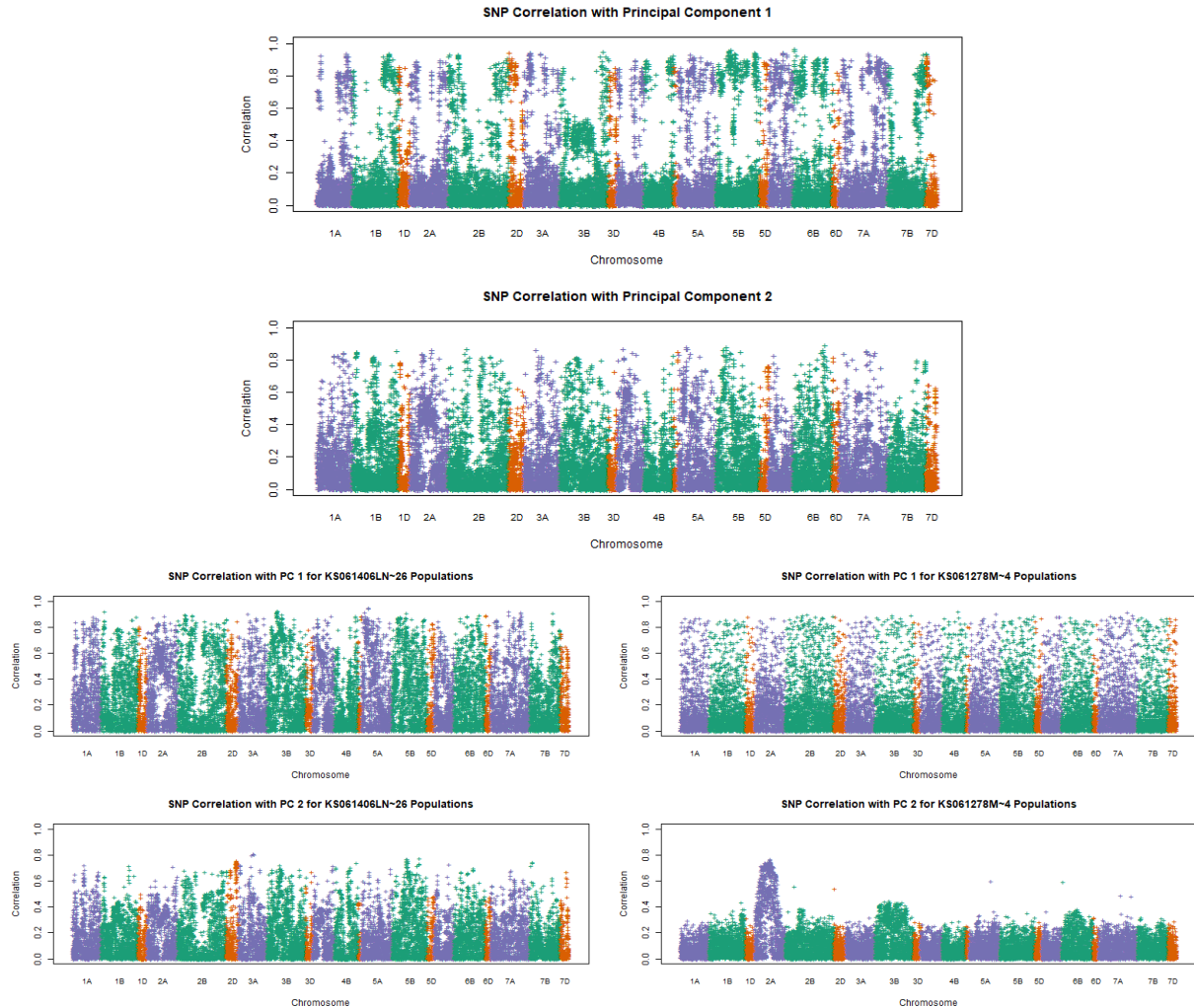


Figure 2.5. SNP-PC correlation plots for the first two PCs derived from population structure analyses of TdGD WILs

Absolute correlations between individual SNPs and PC scores are plotted by genomic position along each chromosome. The top panel shows SNP-PC correlations calculated across all seven TdGD populations analyzed jointly. The bottom panels show SNP-PC correlations stratified by RP background, with populations derived from KS061406LN~26 shown on the left, and populations derived from KS061278M~4 shown on the right. SNPs are ordered by physical position and colored by subgenome. Analyses were conducted using filtered array-based SNP genotypes and the R package *snpRelate* (Zheng et al., 2012).

To further investigate the genomic basis of population structure, SNP correlations with PCs were calculated (Zheng et al., 2012) to identify genomic regions contributing disproportionately to major axes of variation. SNP-PC correlation plots were generated for PCs 1 and 2 both across all populations and within RP groupings (Figure 2.5). To specifically explore the unique structure observed in Pop30, SNP-PC correlation analyses were also performed

within Pop30 alone, with no specific genomic region or chromosome exhibiting an unusually strong contribution to population structure (Appendix Figure A.6).

To assess whether structuring within Pop30 corresponded with levels of donor introgression, WILs were colored by average donor allele proportions calculated from 1 Mb bins derived during introgression mapping. Genome-wide donor proportion showed a gradient of distribution along PC2, whereas D subgenome donor proportion showed a similar distribution along PC1 (Appendix Figure A.6). Despite this correspondence, SNP-PC correlation profiles within Pop30 did not show enrichment of high correlations on D subgenome chromosomes.

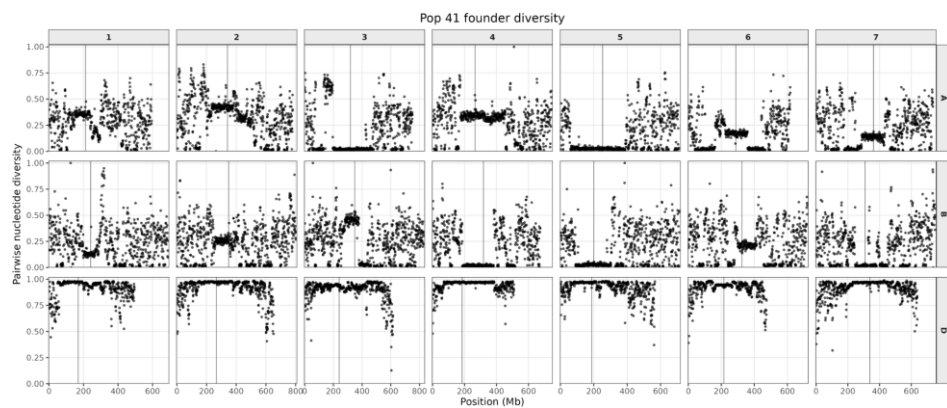
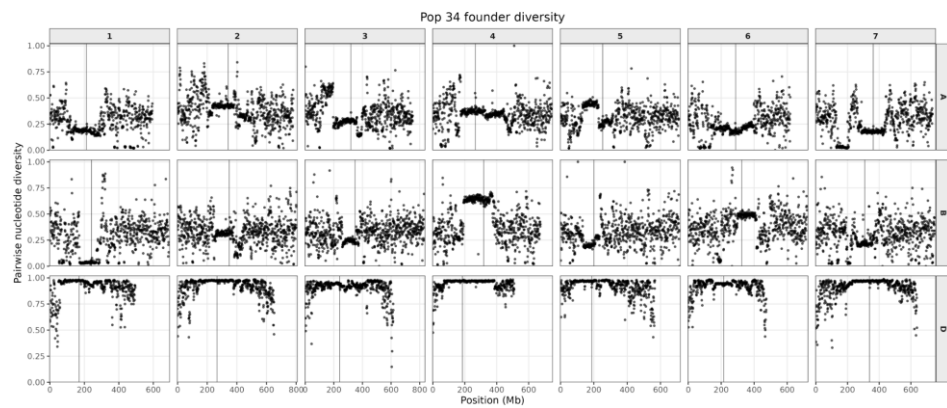
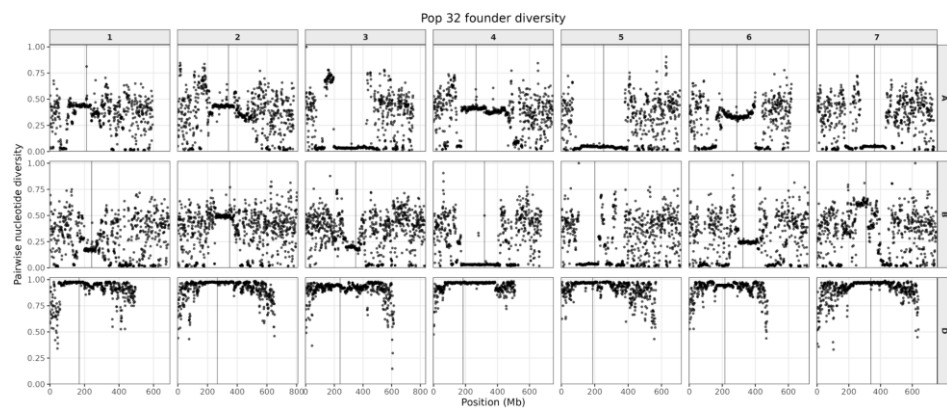
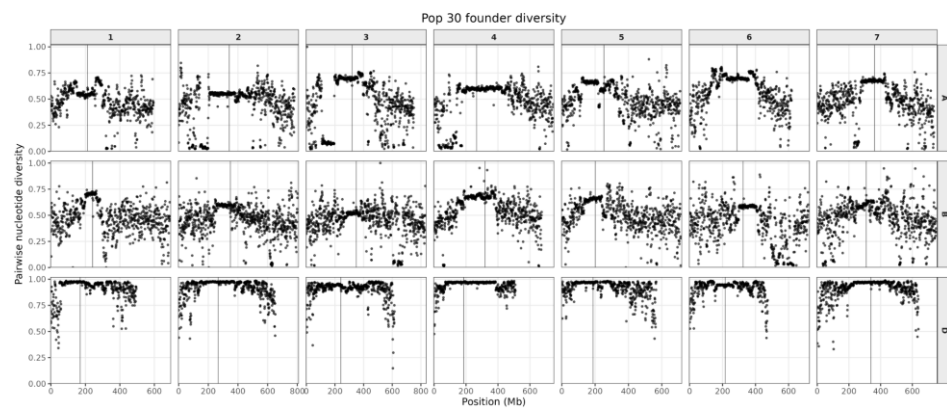
Among KS061278M-4-derived populations, overall structuring was weaker than that observed among KS061406LN~26-derived populations. Although a subset of WILs from Pop80 separated from the main cluster along PC1, SNP-PC correlation analysis for PC1 did not identify any genomic region exhibiting disproportionately high correlations (Figures 2.4-2.5). In contrast, SNP correlations with PC2 revealed elevated correlations along much of chromosome 2A across KS061278M-4-derived populations (Figure 2.5). This observation is biologically plausible given that KS061278M-4 carries the 2NS translocation from *Aegilops ventricosa*, an alien introgression known to suppress recombination in affected regions (Gao et al., 2021; Mustahsan et al., 2023; Price et al., 2025a). Notably, the chromosome 2A signal was not uniform across all KS061278M-4-derived populations. When SNP-PC correlations were examined within individual populations, elevated chromosome 2A correlations with PC2 were apparent only in Pop41 and Pop80 (Appendix Figure A.7). This effect was particularly pronounced in Pop41, whereas Pop80 also exhibited elevated correlations on additional chromosomes, including chromosome 7A.

To evaluate whether separation along PC2 corresponded with quantitative variation in introgression across chromosome 2A, average donor allele proportions on chromosome 2A were calculated for each WIL and used the color the PCA for Pop41 and Pop80 (Appendix Figure A.7). A continuous gradient was observed, with WILs exhibiting high mean donor proportions on chromosome 2A distributed toward one end of PC2 and those with lower donor proportions toward the opposite end. This pattern contrasts with the absence of discrete clustering by 2NS genotype state.

Nucleotide diversity between population founders

In addition to PCA-based assessments of allelic differentiation among WIL progeny using array-derived SNPs, allelic differentiation between population founders was evaluated using sequencing-based variants through calculation of pairwise nucleotide diversity. For each population x subgenome combination, nucleotide diversity was calculated between the RP and its corresponding donor parent (wild emmer for the A and B subgenomes; *Ae tauschii* for the D subgenome). Diversity was quantified as the proportion of differing genotype calls among all callable sites within non-overlapping 1 Mb genomic bins. Table 2.6 summarizes nucleotide diversity values averaged across chromosomes within each subgenome, while Figure 2.6 visualizes nucleotide diversity along physical positions across chromosomes for each population.

Across all populations, nucleotide diversity between founders was substantially higher for the D subgenome than for the A and B subgenomes. For populations developed using *Ae. tauschii* accession TA 2398, average D subgenome nucleotide diversity was 0.91, while the population developed using TA 10155 exhibited a similarly high average of 0.92 (Table 2.6). In contrast, A and B subgenome diversity values ranged from 0.22 (B subgenome in Pop41) to 0.49 (A subgenome in Pop30).



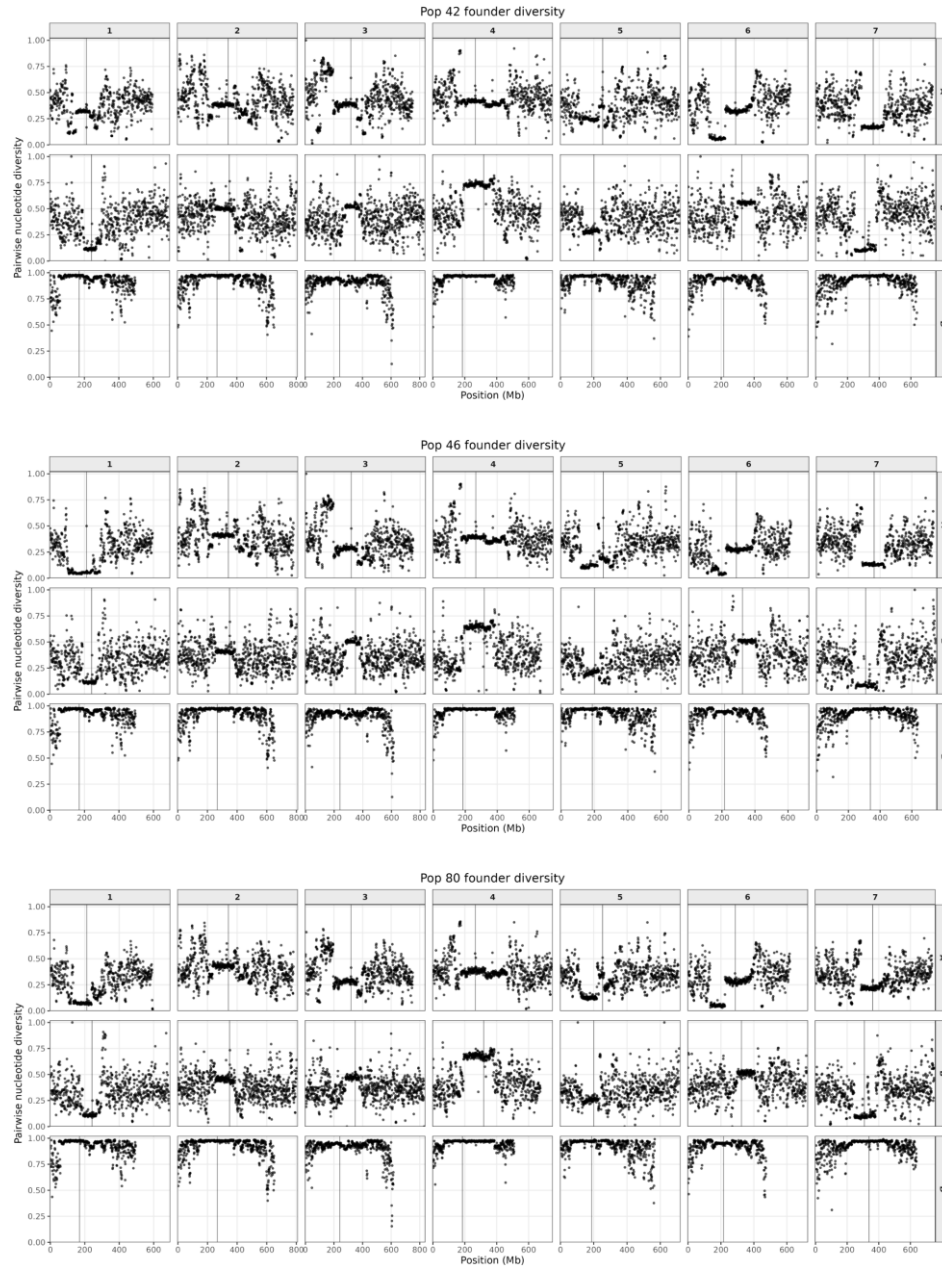


Figure 2.6. Pairwise nucleotide diversity between population founders across TdGD populations, plotted by genomic position and faceted by chromosome

For each population, nucleotide diversity was calculated between the recurrent parent and its corresponding donor parent as the proportion of differing genotype calls among all callable sites within non-overlapping 1 Mb genomic bins. Diversity values are plotted along physical position for each chromosome, with panels arranged by subgenome and chromosome. Vertical solid lines indicate centromere positions based on the Chinese Spring v1.0 reference genome.

Within the A and B subgenomes, modest but meaningful differences in nucleotide diversity were observed among populations. Notably, the wild emmer accession used to develop

Pop30 (PI 414718) exhibited the highest divergence from the RP across populations, with average nucleotide diversity values of 0.49 in the A subgenome and 0.48 in the B subgenome (Table 2.6). This elevated founder divergence is notable given the pronounced within-population structure revealed in Pop30 from PCA (Figure 2.4). In contrast, populations such as Pop41 – where A and B subgenome nucleotide diversity averages were substantially lower (0.23 and 0.22, respectively) – exhibited comparatively weaker population structure.

Patterns of nucleotide diversity also correspond with the geographic origins of wild emmer donors. Populations developed using geographically isolated accessions from Israel (populations 30, 42, 46, and 80) exhibited higher A and B subgenome nucleotide diversity (0.33-0.49) than populations developed using wild emmer accessions from Turkey and Syria (populations 32, 34, and 41; 0.22-0.34). Differences in site count among populations reflect variation in sequencing methodology, as populations 34, 41, and 80 were sequenced using exome capture approaches, yielding lower total site counts relative to populations sequenced using WGS strategies.

Genome-wide visualization of nucleotide diversity further highlights the localized nature of founder divergence (Figure 2.6). Across populations, nucleotide diversity was generally highest near centromeric regions and decreased toward distal chromosomal regions. This pattern was particularly pronounced in the D subgenome, where most chromosomes exhibited uniformly high nucleotide diversity, with reduced divergence observed primarily near chromosome ends. In contrast, A and B subgenome chromosomes displayed more heterogeneous patterns of diversity.

Interestingly, several populations exhibited near complete conservation between founders in centromeric regions of specific chromosomes. In Pop32 and Pop41, the centromeric regions of chromosomes 3A, 5A, 4B, and 5B (and chromosome 7A in Pop32), exhibited markedly low

nucleotide diversity with essentially no divergence between RP and donor (Figure 2.6). In contrast, Pop30 exhibited no conservation in centromeric regions, with any similarities between founders largely restricted to distal, high recombination regions.

Table 2.6. Pairwise nucleotide diversity between population founders across TdGD populations

Pairwise nucleotide diversity was calculated between each recurrent parent (RP) and its corresponding donor parent (wild emmer for the A and B subgenomes, *Ae. tauschii* for the D subgenome) as the proportion of differing variant calls out of all callable variant sites within non-overlapping 1 Mb genomic bins. Reported diversity values represent averages across all chromosomes within each subgenome for a given RP-donor pair. Site count indicates the total number of variant sites used to estimate diversity within each subgenome.

Pop	Subg	RP	Donor	Diversity	Site count
30	A	KS061406LN~26	PI414718	0.49	41.89 K
30	B	KS061406LN~26	PI414718	0.48	48.74 K
30	D	KS061406LN~26	TA2398	0.91	25.70 K
32	A	KS061406LN~26	PI470736	0.30	44.32 K
32	B	KS061406LN~26	PI470736	0.31	51.52 K
32	D	KS061406LN~26	TA2398	0.91	25.70 K
34	A	KS061406LN~26	PI487264	0.32	8.23 K
34	B	KS061406LN~26	PI487264	0.34	9.40 K
34	D	KS061406LN~26	TA2398	0.91	25.70 K
41	A	KS061278M-4	PI428088	0.23	2.85 K
41	B	KS061278M-4	PI428088	0.22	3.40 K
41	D	KS061278M-4	TA2398	0.91	25.83 K
42	A	KS061278M-4	PI466984	0.39	40.88 K
42	B	KS061278M-4	PI466984	0.43	47.02 K
42	D	KS061278M-4	TA2398	0.91	25.83 K
46	A	KS061278M-4	PI471815	0.33	36.14 K
46	B	KS061278M-4	PI471815	0.36	40.64 K
46	D	KS061278M-4	TA2398	0.91	25.83 K
80	A	KS061278M-4	PI471021	0.34	7.68 K
80	B	KS061278M-4	PI471021	0.38	8.73 K
80	D	KS061278M-4	TA10155	0.92	26.05 K

Pedigree-based imputation performance

Imputation of WIL progeny was performed using AlphaImpute2 using a map length parameter of 2 (Morgans) and a final peeling threshold of 0.1 (Whalen et al., 2020). Following imputation, an additional ancestry-based heterozygote smoothing step was applied, with

downstream segment detection thresholds set to avoid calling donor segments based solely on residual heterozygosity.

Using a map length of 2 Morgans effectively controlled genome-wide excess heterozygosity with individual-level heterozygosity being broadly comparable across subgenomes (Appendix Figure A.8). Although some WILs still exhibited elevated subgenome-specific heterozygosity after smoothing, genome-wide average heterozygosity per individual was minimized (Appendix Figure A.8). This effect was most pronounced on the D subgenome, where marker density is lower and donor allele frequencies are expected to be reduced relative to the A and B subgenomes (Figure 2.1). Using the default chromosome length of 1 Morgan, imputation on the D subgenome resulted in persistent uncertainty in progeny genotypes, with neither donor nor RP alleles being consistently imputed across missing regions. Increasing the assumed map length to 2 Morgans mitigated this issue by allowing greater flexibility in segregation state transitions.

Because the set of array markers observed in the progeny is fixed, the density of sequencing-derived markers included in the founders has a direct impact on the amount of information that must be inferred during imputation. Including a large number of founder-only markers, coded as missing in the progeny, increases the proportion of loci for which genotype inference relies entirely on pedigree-derived information rather than empirical genotype evidence. Conversely, restricting the founder marker set to more widely spaced, informative loci reduces the number of uncertain positions that must be imputed and increases reliance on emission evidence at observed progeny markers. In practice, increasing marker density in the founders by using smaller minimum inter-marker distance thinning thresholds (e.g., 50 kb) resulted in reduced donor allele proportions in the imputed progeny, whereas using larger

thinning windows (e.g., 250 kb) increased donor allele retention to more biologically plausible levels. This reflects the fact that reducing the number of loci requiring inference limits the cumulative propagation of uncertainty across the pedigree.

Population-level patterns of residual heterozygosity illustrate the dependence of pedigree-based imputation on anchor marker availability. Pop30 provides a clear example. An extended region on chromosome 3A lacks array-based anchor markers in the progeny (Appendix Figure A.2), requiring long-range imputation without direct evidence for changes in segregation state. Consequently, this region exhibits elevated mean heterozygosity across WILs in Pop30 even after smoothing (Appendix Figure A.9). In this case, flanking markers did not consistently agree in ancestry across individuals, preventing confident haplotype inference and demonstrating the limits of smoothing in regions of marker sparsity. Other regions with elevated population wide heterozygosity were generally smaller and may correspond to localized recombination events between anchor markers where ancestry state cannot be reliably inferred (Appendix Figure A.9).

Table 2.7 presents average marker heterozygosity and MAF by population and subgenome before and after heterozygote smoothing. Average marker heterozygosity predictably decreased following smoothing across all populations and subgenomes. In contrast, changes in MAF were minimal, indicating that smoothing did not systematically bias genotype resolution toward either the RP or donor haplotype. If smoothing preferentially resolved heterozygotes toward the donor, MAF would be expected to increase; conversely, smoothing toward the RP would decrease MAF. Instead, MAF changes were small and inconsistent in direction across population x subgenome, with the largest observed reduction being 0.04 (e.g., from 0.09 to 0.05 in the D subgenome of Pop30; from 0.11 to 0.07 in the A subgenome of Pop34; and from 0.13 to

0.09 in the B subgenome of Pop80; Table 2.7). Distributions of MAF before and after smoothing are shown in Appendix Figure A.10.

Table 2.7. Average marker heterozygosity and minor allele frequency in TdGD WILs following imputation

Heterozygosity (Het) and MAF are summarized after imputation with AlphaImpute2 (Whalen et al., 2020) both before and after heterozygote smoothing based on flanking anchor marker ancestry. Statistics are reported by population and subgenome. Average marker heterozygosity is calculated as the proportion of individuals heterozygous at a given marker, averaged across all markers within each population x subgenome. MAF is defined as the frequency of the less common allele at each marker within a population, averaged across markers.

Pop	Subg	Before het smoothing		After het smoothing	
		Het	MAF	Het	MAF
30	A	0.24	0.15	0.11	0.17
30	B	0.32	0.17	0.12	0.14
30	D	0.16	0.09	0.09	0.05
32	A	0.16	0.09	0.09	0.09
32	B	0.23	0.12	0.11	0.09
32	D	0.09	0.05	0.07	0.05
34	A	0.20	0.11	0.12	0.07
34	B	0.19	0.11	0.11	0.09
34	D	0.15	0.08	0.09	0.06
41	A	0.17	0.10	0.09	0.09
41	B	0.23	0.12	0.11	0.12
41	D	0.12	0.06	0.09	0.06
42	A	0.14	0.08	0.09	0.08
42	B	0.24	0.15	0.12	0.13
42	D	0.12	0.06	0.08	0.04
46	A	0.19	0.11	0.09	0.09
46	B	0.20	0.12	0.11	0.13
46	D	0.07	0.04	0.07	0.04
80	A	0.14	0.08	0.09	0.07
80	B	0.24	0.13	0.11	0.09
80	D	0.09	0.05	0.08	0.04

Overall, these results indicate that pedigree-based imputation with conservative heterozygote smoothing effectively reduced spurious heterozygosity without introducing systematic allele frequency distortions. Residual heterozygosity is largely localized to genomic

regions where marker density limits inference, reinforcing the interpretation that imputation uncertainty is driven by empirical constraints rather than algorithmic bias alone.

Mapping introgression segments in imputed WILs

Following pedigree-based imputation and ancestry-based heterozygote smoothing, donor allele proportions were calculated in non-overlapping 1 Mb bins as the proportion of called alleles matching the population- and subgenome-specific donor. These values were then summarized by WIL and averaged within each population x subgenome to estimate genome-wide donor contribution (Table 2.8).

Population-level donor proportions varied across populations and subgenomes. Pop30 exhibited the highest average donor proportions on the A and B subgenomes (0.20 on both; Table 2.8). This pattern is consistent with the comparatively higher minor allele frequencies in Pop30 (Figure 2.3; Tables 2.3, 2.7). The lowest A and B subgenome donor proportions were observed in Pop80 (0.09 in A; 0.10 in B; Table 2.8). Donor proportions on the D subgenome were consistently lower than on the A and B subgenomes, ranging from 0.04 in Pop46 to 0.07 in Pop34 (Table 2.8).

To identify contiguous introgressed regions, bin-level donor proportions were smoothed using short- and long-range moving averages to reduce noise from bins with limited marker support. Multiple window size combinations were evaluated, and consecutive bins classified as donor were merged into segments. Predictably, increasing MA window size increased called segment lengths (Appendix Table A.3): smaller windows (e.g., MA2) remain sensitive to local stochasticity, whereas very large windows (e.g., MA50/MA100) require sustained donor signal across large physical regions and therefore preferentially detect only the largest introgressions. Across window size pairs, results were broadly concordant in the genomic regions implicated,

although inferred segment boundaries shifted with smoothing intensity. Accordingly, called segments are best interpreted as in support that introgression is present within a region, rather than as precise estimates of recombination breakpoints.

Table 2.8. Summary of donor introgression in TdGD WILs, aggregated by population and subgenome

Donor allele proportion (DonorProp) represents the mean proportion of alleles matching the donor parent across all informative markers and WILs within each population x subgenome combination. Introgression segment lengths are summarized as the mean and median length (Mb) of donor derived segments detected across WILs. Segments were called using moving average smoothing of donor allele proportions with a short-range window of 10 bins and a donor proportion threshold of 0.9, paired with a long-range window of 40 bins and a threshold of 0.7, merging consecutive donor bins to determine segments.

Pop	Subg	DonorProp	SegmentLength	
			Mean	Median
30	A	0.20	43	18
30	B	0.20	35	9
30	D	0.06	40	10
32	A	0.11	41	9
32	B	0.12	45	25
32	D	0.06	80	45
34	A	0.10	18	7
34	B	0.11	61	9
34	D	0.07	78	18
41	A	0.10	51	13
41	B	0.14	56	9
41	D	0.06	70	28
42	A	0.10	58	22
42	B	0.15	60	16
42	D	0.05	39	24
46	A	0.10	53	14
46	B	0.15	53	9
46	D	0.04	34	24
80	A	0.09	52	14
80	B	0.10	34	7
80	D	0.05	68	63

Segment length distributions reflected both biological introgression architecture and technical constraints imposed by anchor marker density (Table 2.8; Figure 2.7). In regions with sparse progeny marker coverage, fewer empirically supported segregation state transitions are

available, increasing the tendency for long continuous haplotype inference. This effect was particularly apparent on the D subgenome, where anchor marker density is lowest (Table 2.2). For example, mean segment length on the D subgenome reached 80 Mb in Pop32 (Table 2.8), consistent with the possibility that multiple introgressions separated by unobserved recombination events can be merged into a single long called segment when intermediate anchor evidence is absent. In contrast, on the A and B subgenomes, where array-based markers are more abundant especially in distal, higher recombination regions, shorter segments were more readily detected. Across populations, mean segment lengths ranged from 18-58 Mb (A subgenome), 34-61 Mb (B subgenome), and 34-80 Mb (D subgenome), while medians ranged from 7-22 Mb (A), 7-25 Mb (B), and 10-63 Mb (D) (Table 2.8), indicating strong right skew and a general bias toward shorter segments (Figure 2.7). Pop30 exhibited a larger number of detected segments than other populations (Figure 2.7), consistent with its higher MAF.

The genomic distribution of introgression was further summarized by the proportion of WILs carrying at least one called segment on each chromosome arm (Table 2.9). Across all population x arm combinations, the highest prevalence occurred on the short arm of chromosome 3B in Pop30, where 98% of WILs carried an introgression somewhere along that arm. In contrast, many D subgenome chromosome arms showed zero WILs with detected introgressions in a given population (Table 2.9), consistent with lower observed D subgenome donor proportions (Table 2.8). Where D subgenome introgressions were detected, they were often long but comparatively infrequent across individuals; the highest D subgenome donor prevalence was observed on the short arm of chromosome 7D in Pop80, where 72% of WILs carried a segment on that arm (Table 2.9). Chromosome 6D exhibited the fewest introgressions across populations.

Chromosome 4B was also notable for low introgression prevalence; the maximum observed for any arm of 4B was 30% (Pop34 long arm; Table 2.9).

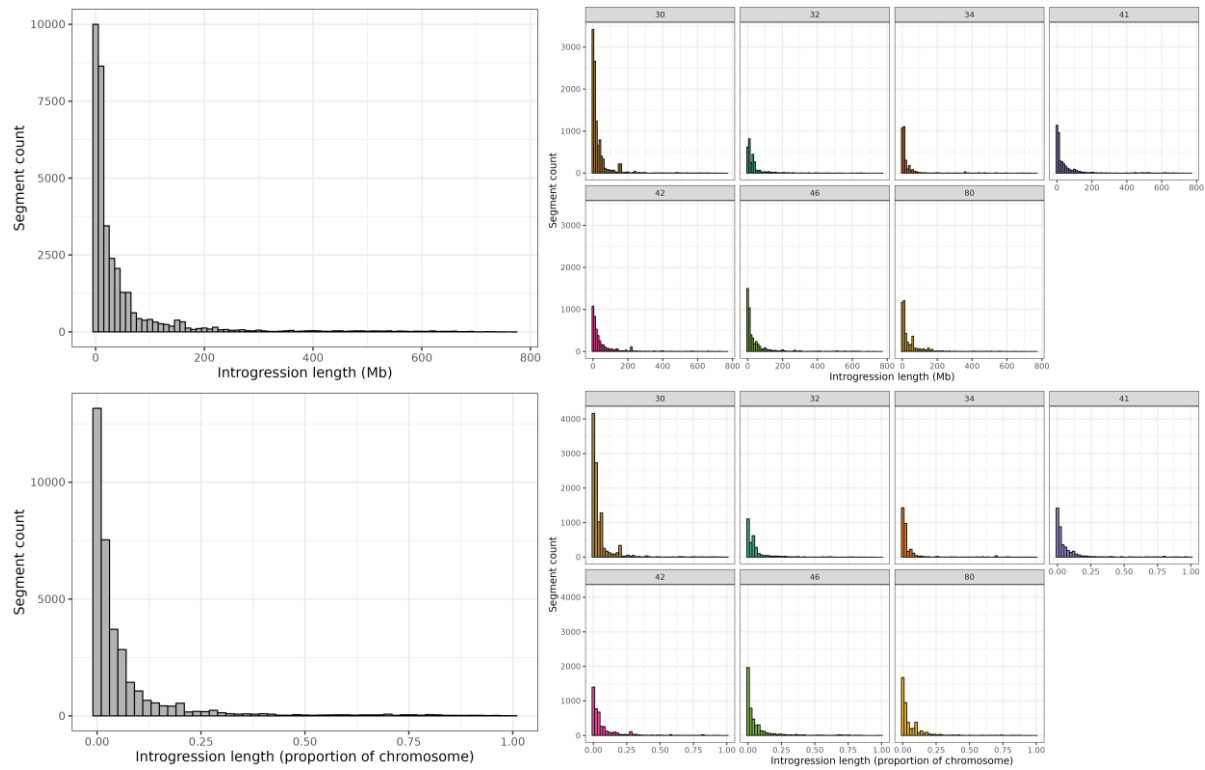
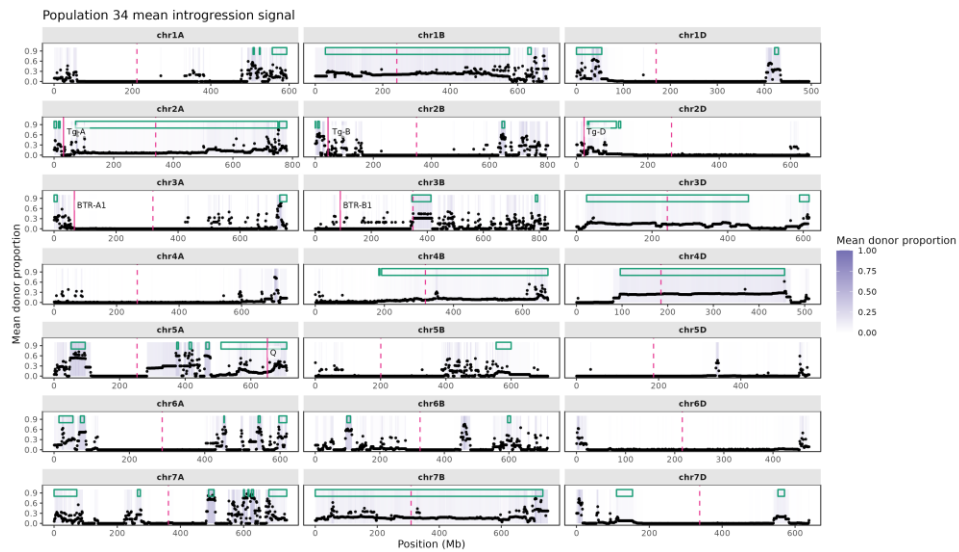
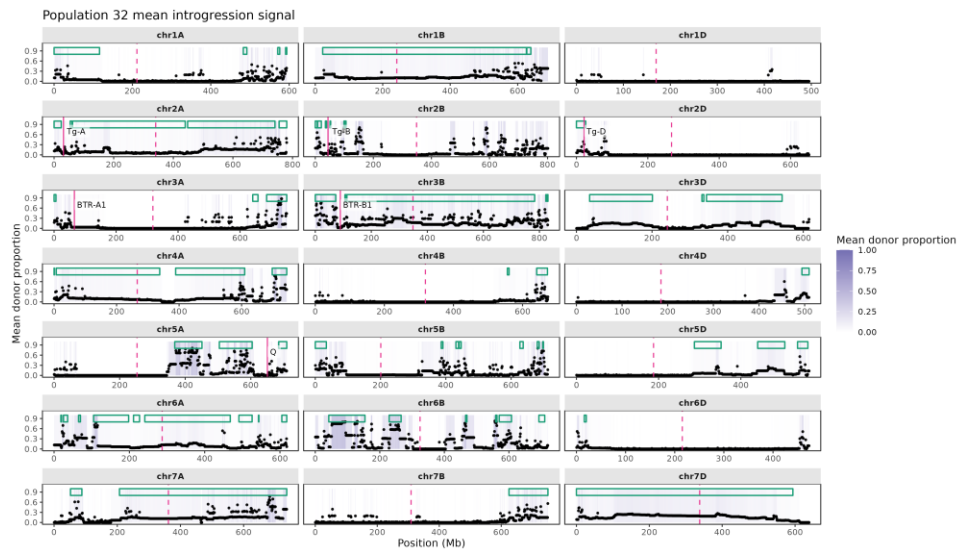
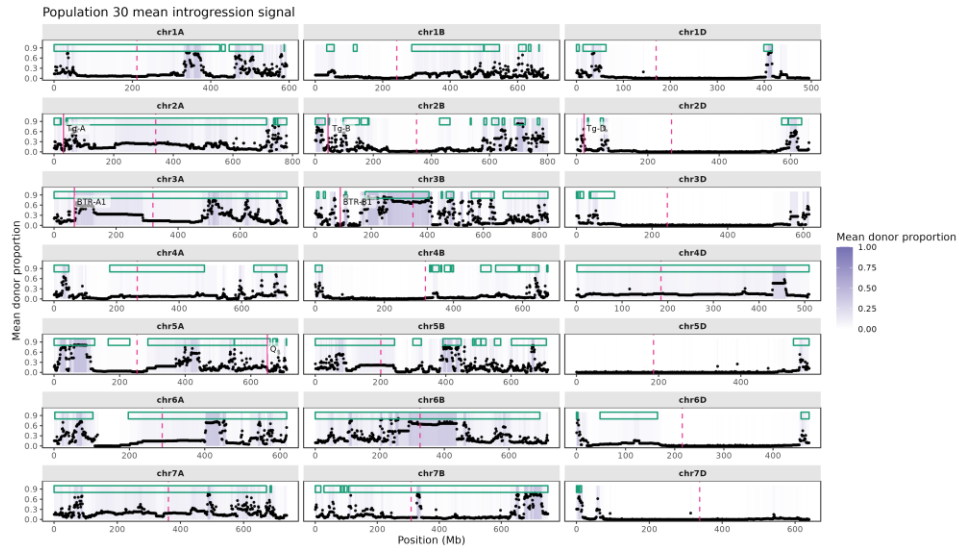
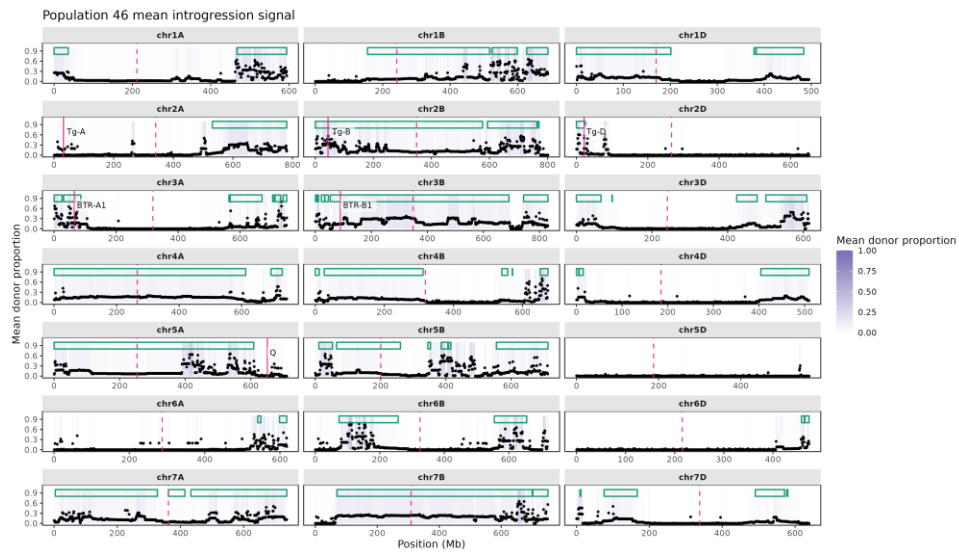
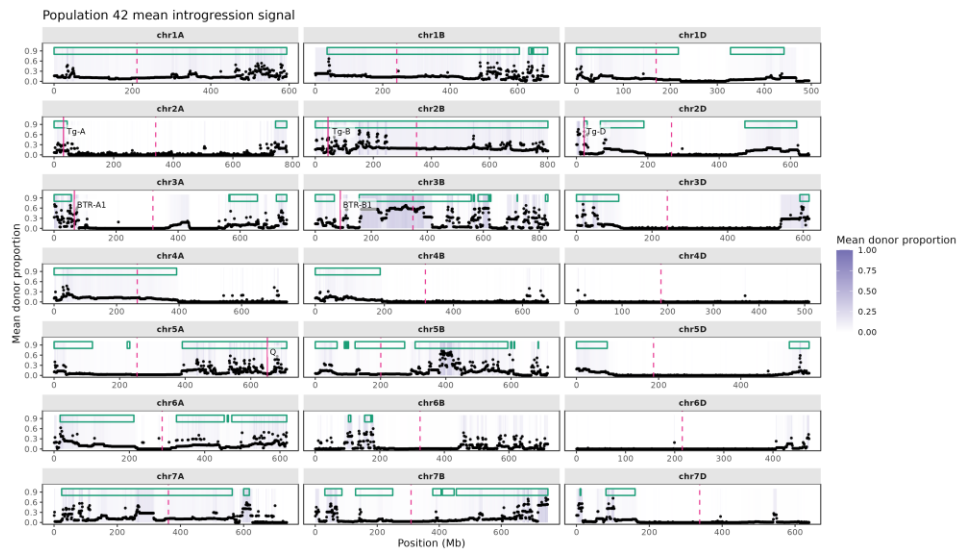
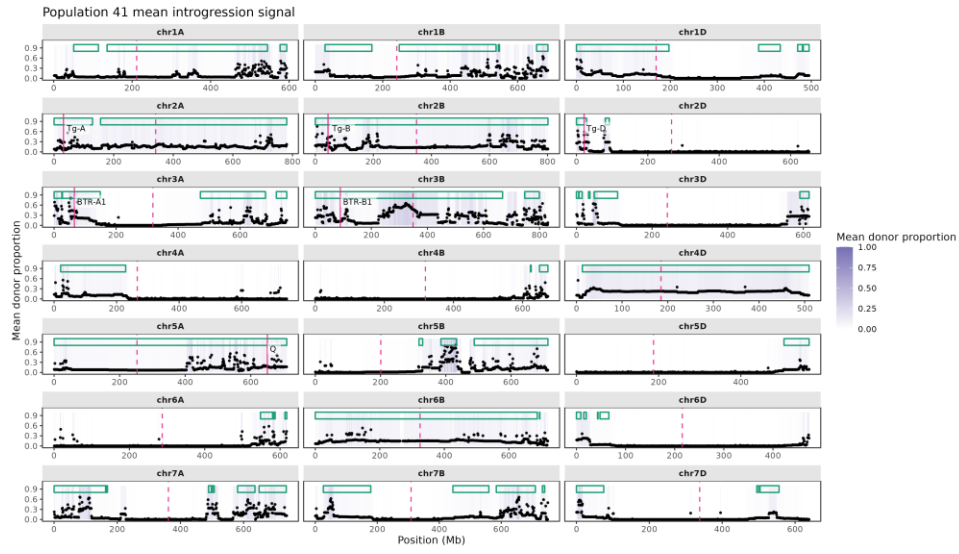


Figure 2.7. Distributions of donor introgression segment lengths in TdGD WILs

Lengths are summarized across all populations (left panels) and split by population (right panels). Segments were identified by applying moving average smoothing to donor allele proportions using a short-range window of 10 bins with a donor proportion threshold of 0.9, paired with a long-range window of 40 bins with a threshold of 0.7, and merging contiguous donor classified bins into continuous segments. Segment lengths are displayed both as physical length (Mb; top panels) and as a proportion of the total chromosome length (bottom panels).

Population averaged donor proportion profiles provide a complementary view of introgression architecture (Figure 2.8). For each chromosome, mean donor proportion was plotted for each 1 Mb bin after averaging across WILs within a population, alongside the genomic positions of called introgression segments. Population mean donor proportions tended to be higher in distal regions away from centromeres, reflecting both the recombination landscape and the concentration of informative progeny anchor markers in higher recombination regions; nonetheless, called segments frequently span centromeric regions.





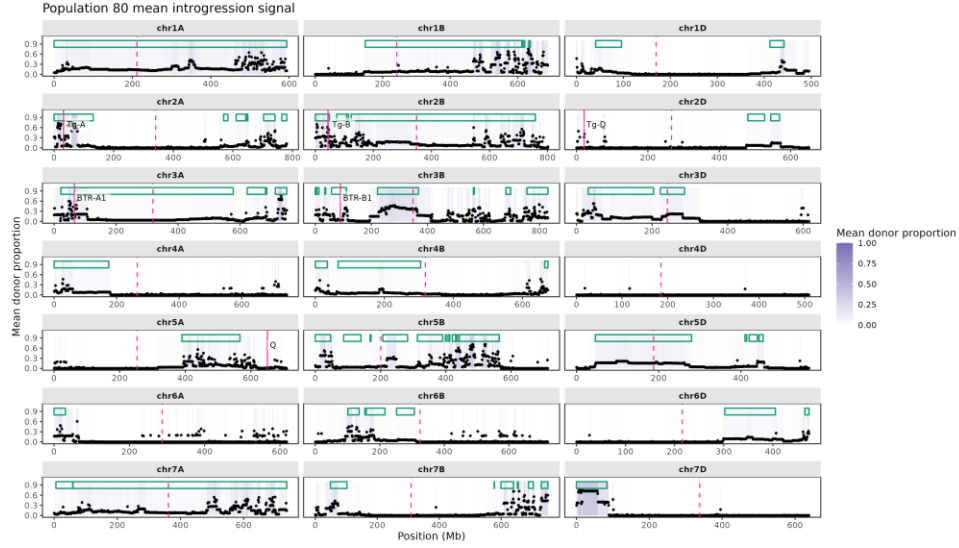


Figure 2.8. Population averaged donor introgression profiles across the genome

Mean donor allele proportions were calculated in 1 Mb bins by averaging across all WILs within each TdGD population and are shown as points along each chromosome facet. Vertical dashed lines indicate centromere positions based on the Chinese Spring v1.0 reference. Major domestication loci – *Q* (free threshing), *Btr* (brittle rachis), and *Tg* (tenacious glume) – are indicated by labeled vertical markers at their physical positions. Horizontal boxes at the top of chromosome panels denote regions where introgression segments were called in WILs using a moving average framework ($MA_{10} \geq 0.9$ and $MA_{40} \geq 0.7$), summarizing population-level consensus segments along each chromosome.

Several localized deviations from the distal enrichment trend were also observed. For example, chromosome 3B exhibited elevated donor signal near the centromere in multiple populations, most prominently in Pop30 (Figure 2.8). While this could reflect true introgression enrichment, it is also a region where uncertainty may be elevated due to sparser anchor coverage, propagating ancestry states across poorly supported intervals, potentially inflating donor proportions or extending inferred segments. However, similar centromere proximal elevation is not uniformly observed across all chromosomes, arguing against a purely global artifact driven by heterozygosity retention.

Population mean donor proportions across D subgenome chromosomes were lower than those on A and B (Figure 2.8; Table 2.8), and chromosome 4B – where few WILs carried called segments – also showed correspondingly low mean donor proportion (Table 2.9; Figure 2.8). In

several populations, called segments collectively span most or all of a chromosome when aggregated across WILs; this reflects population level coverage of introgressions across different individuals rather than chromosome-scale introgression within any single WIL.

Table 2.9. Proportion of WILs with donor introgression segments detected on each chromosome arm, summarized by TdGD population

For each population x chromosome arm combination, proportions were calculated as the number of WILs containing one or more introgression segments overlapping the specified arm divided by the total number of WILs evaluated in that population. Chromosome arms were defined relative to centromere positions of Chinese Spring v1.0. Introgression segments were identified using moving average smoothing of donor allele proportions with a short-range window of 10 bins and a donor proportion threshold of 0.9, paired with a long-range window of 40 bins and a threshold of 0.7.

Chr	Proportion of WILs with introgression													
	Pop30		Pop32		Pop34		Pop41		Pop42		Pop46		Pop80	
	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long
1A	0.19	0.89	0.21	0.28	0.01	0.57	0.06	0.40	0.16	0.37	0.26	0.51	0.22	0.49
1B	0.17	0.42	0.21	0.36	0.32	0.37	0.15	0.37	0.21	0.48	0.07	0.79	0.07	0.69
1D	0.31	0.05	0.00	0.00	0.15	0.34	0.25	0.13	0.21	0.09	0.22	0.20	0.10	0.05
2A	0.45	0.46	0.28	0.37	0.38	0.47	0.24	0.42	0.14	0.26	0.01	0.51	0.64	0.18
2B	0.82	0.91	0.26	0.03	0.53	0.46	0.33	0.57	0.46	0.47	0.36	0.51	0.24	0.46
2D	0.12	0.53	0.14	0.00	0.24	0.00	0.15	0.00	0.20	0.18	0.07	0.00	0.00	0.07
3A	0.76	0.87	0.22	0.64	0.29	0.12	0.38	0.47	0.67	0.28	0.59	0.27	0.08	0.69
3B	0.98	0.87	0.54	0.48	0.08	0.47	0.65	0.60	0.67	0.68	0.36	0.52	0.48	0.48
3D	0.11	0.05	0.16	0.23	0.23	0.36	0.14	0.06	0.18	0.14	0.21	0.40	0.24	0.21
4A	0.39	0.37	0.21	0.21	0.00	0.03	0.12	0.00	0.19	0.11	0.19	0.38	0.12	0.00
4B	0.16	0.23	0.00	0.06	0.10	0.30	0.00	0.16	0.23	0.00	0.15	0.24	0.17	0.11
4D	0.19	0.21	0.00	0.16	0.26	0.28	0.31	0.29	0.00	0.00	0.16	0.20	0.00	0.00
5A	0.86	0.84	0.00	0.78	0.53	0.85	0.12	0.40	0.13	0.24	0.19	0.77	0.00	0.16
5B	0.69	0.91	0.23	0.22	0.00	0.13	0.00	0.54	0.14	0.68	0.14	0.57	0.39	0.66
5D	0.00	0.19	0.00	0.28	0.00	0.01	0.00	0.13	0.17	0.10	0.00	0.00	0.19	0.22
6A	0.88	0.87	0.18	0.35	0.80	0.53	0.00	0.18	0.24	0.40	0.00	0.11	0.05	0.00
6B	0.91	0.83	0.76	0.34	0.55	0.26	0.25	0.22	0.39	0.05	0.29	0.49	0.31	0.00
6D	0.15	0.15	0.04	0.01	0.03	0.03	0.14	0.04	0.00	0.02	0.00	0.07	0.00	0.16
7A	0.48	0.55	0.21	0.37	0.48	0.95	0.20	0.59	0.20	0.44	0.37	0.28	0.26	0.38
7B	0.37	0.90	0.00	0.23	0.33	0.26	0.31	0.28	0.11	0.55	0.31	0.75	0.11	0.44
7D	0.06	0.04	0.37	0.30	0.08	0.19	0.10	0.06	0.52	0.02	0.40	0.09	0.72	0.00

Given the low donor proportions observed on the short arm of chromosome 1B in Pop46 and Pop80 (Table 2.9; Figure 2.8), determining whether the shared RP, KS061278M-4, carried

the 1BL.1RS translocation was particularly important. Comparative alignment coverage analysis using Larry as a known 1BL.1RS carrier (Parker et al., 2026) provided no evidence that KS061278M-4 carries the translocation (Appendix Figure A.11). Therefore, the reduced donor representation and limited segment detection on the short arm of 1B in Pop46 and Pop80 cannot be readily attributed to suppression of recombination due to translocation. Alternative explanations include other structural differences between the respective wild emmer donors and the RP, localized suppression of recombination, or stochastic loss of donor haplotypes arising from the limited number of BC₂F₁ families represented among the derived lines. The available genotypic data do not permit these possibilities to be distinguished conclusively. Nevertheless, the alignment results, together with the detection of donor introgression on the short arm of 1B in Pop41 and Pop42 – which were also developed using KS061278M-4 as the RP – support the conclusion that KS061278M-4 does not carry the 1BL.1RS translocation (Table 2.9; Figure 2.8; Appendix Figure A.11).

Validation of introgression calls was conducted using a KASP assay targeting the domestication locus *Q/q*. In Pop30, 10 WILs had a called segment overlapping *Q/q* and 29 had a segment within 5 Mb of *Q/q*; in Pop41, 28 WILs had segments overlapping *Q/q* and 13 had segments within 5 Mb. Founder controls were included to verify assay behavior; all seven wild emmer samples assayed were called *Q*, and 13 of 14 RP samples were called *q* (one RP control was misclassified). KASP genotypes were broadly consistent with segment calls, with stronger concordance in Pop41 than in Pop30.

In Pop30, 7 of 10 WILs with a segment overlapping *Q/q* were called *Q*, while 3 were called *q*. Among WILs expected to be *q* (i.e., without a segment overlapping *Q/q*), concordance was 86% (219/256), with 37 individuals called *Q* despite expectation of *q*. Within the 5 Mb

window around Q/q , 18 of 29 WILs were called Q and 11 were called q , consistent with uncertainty in inferred boundaries under MA-based calling.

In Pop41, 26 of 28 WILs overlapping Q/q were called Q , and all 13 WILs with a segment within 5 Mb were called Q . Among Pop41 WILs expected to be q , concordance was 93% (151/163), with 12 called Q despite expectation. Overall, classification concordance for Q/q overlap was 70% in Pop30 and 93% in Pop41; within 5 Mb of Q/q , concordance was 62% in Pop30 and 100% in Pop41.

Discussion

Seven BC₂F_{4.5} TdGD populations were evaluated to characterize patterns of donor introgression across the A, B, and D subgenomes of hexaploid wheat. Analytical resolution and biological interpretation were closely intertwined with design decisions made during genotypic data generation and processing. The interpretation of the resulting mosaic genomes requires consideration of both biological and methodological influences. As such, evaluation is made on (1) the implications of differing genotyping platforms and marker densities, (2) the performance of AlphaImpute2 under uneven marker distribution, a heterogeneous recombination landscape, and low donor allele frequencies, (3) founder-level allelic differentiation and patterns of population structure, and (4) the correspondence between an amphiploid bridge-mediated versus direct crossing approach to introgress diversity from wild emmer.

Implications of SNP array-based genotyping of introgression germplasm

The Wheat/Barley 40 K SNP array was selected for its enrichment of markers informative for tetraploid and D subgenome tracking (Keeble-Gagnère et al., 2021). However, SNP array-based genotyping in wheat introduces important constraints that influence pipeline design choices and interpretation of results for introgression mapping.

First, SNP arrays are inherently subject to ascertainment bias: only polymorphisms included during array development can be detected (Crain et al., 2026; Elbasyoni et al., 2018; Keeble-Gagnère et al., 2021; Wang et al., 2014; H. Wang et al., 2023). Because array design was based on alignment to Chinese Spring v1.0 reference genome, variant detection is limited to sites represented in that assembly. Highly divergent haplotypes contributed by wild emmer or *Ae. tauschii* may therefore be underrepresented by the reference (Parker et al., 2026; Walkowiak et al., 2020). In introgression populations derived from wild relatives, this constraint can reduce the number of informative ancestry markers in precisely those regions where divergence is greatest.

Second, probe specificity is complicated by the presence of homoeologous sequences in hexaploid wheat. Although probes are designed to preferentially hybridize to a target locus, high sequence similarity among A, B, and D homoeologs can permit cross-hybridization (Akhunov et al., 2009; Kaur et al., 2012; Keeble-Gagnère et al., 2021; Tiwari et al., 2014; Wang et al., 2014). Spot checking nucleotide BLAST alignments of array probe sequences against multiple reference assemblies (Chinese Spring v1.0, v2.1, and the tetraploid Zavitan genome; Avni et al., 2017; IWGSC, 2018; Zhu et al., 2021) confirmed that while probes aligned most strongly to their annotated target chromosome, substantial secondary alignments to homoeologous chromosomes were common (data not shown). In some cases, probe affinity differed depending on the reference genome queried. This raises the possibility that, in regions of high homoeology, allelic state may occasionally reflect signal from a non-target homoeolog, particularly in wild-derived material where sequence divergence patterns differ from those represented in the reference.

There is no practical mechanism to eliminate this possibility entirely in array-based datasets. In this study, several steps were taken to reduce its impact. Markers exhibiting elevated missingness or heterozygosity in the progeny were removed, under the assumption that such

patterns reflect poor probe performance or ambiguous hybridization. Concordance in REF/ALT identity and orientation between sequencing-derived parental genotypes and array-based progeny calls were enforced during merging; markers showing discordance were excluded entirely. Additionally, only biallelic sites were retained to avoid complications arising from multi-allelic variation that could exacerbate homoeolog misassignment.

Following application of these filters, the final array-derived anchor marker set in the WILs remained sparse and unevenly distributed, particularly in pericentromeric regions of low recombination (Appendix Figure A.2). Consequently, imputation resolution was fundamentally limited by the density and genomic distribution of these empirical anchor markers (Hickey et al., 2012a; Whalen et al., 2020). Introgression detection therefore reflects the resolution of the array backbone rather than the precise underlying haplotype structure. Regions lacking array anchors cannot support exact breakpoint inference, and ancestry interpretation in such intervals is necessarily extrapolated from flanking markers.

These constraints suggest that a sequencing-based approach in the progeny – such as skim exome capture or low-coverage WGS – would provide greater flexibility and reduce dependence on predefined array targets (Crain et al., 2026; H. Wang et al., 2023). Unlike SNP arrays, sequence-based approaches permit variant discovery at sites defined relative to the founders themselves. However, the array-derived genotypes represent an existing and substantial investment, and the present analyses demonstrate that, when coupled with dense founder sequencing and conservative segment calling strategies, meaningful genome-wide introgression patterns can still be inferred.

Overall, while ascertainment bias and homoeolog cross-hybridization are intrinsic limitations of SNP arrays in wheat, the design of the Wheat/Barley 40 K array and the filtering

strategy implemented here aims to minimize their influence sufficiently to support regional introgression inference. Nonetheless, these factors impose a ceiling on haplotype breakpoint resolution and must be considered when interpreting genome-wide donor segment architecture.

Impact of sequencing methodology on founder variant detection

At face value, heterogeneity of variant density and genomic distribution due to sequencing methodology could bias imputation by disproportionately weighting certain founder haplotypes or genomic regions. However, in the context of this study, founder marker density was not a limitation. Even in populations with the lowest retained A/B subgenome founder variant counts, the density of founder genotypes exceeded that of the progeny array anchors by many orders of magnitude. Because pedigree-based imputation is constrained primarily by empirical marker availability in the progeny rather than founder density *per se*, the principal concern was not insufficient founder resolution, but rather overrepresentation of founder only markers that would require inference across excessively large stretches of missing progeny data.

This consideration motivated the implementation of minimum inter-marker distance thinning, which reduced redundancy in founder marker matrices and harmonized variant density across populations and sequencing. After thinning and reintroduction of array concordant recalled variants, the effective imputation matrices were broadly comparable in genome-wide distribution, regardless of whether founders were sequenced via WGS or exome capture. Notably, because the SNP array backbone is enriched for exonic and high recombination regions (Keeble-Gagnère et al., 2021), even WGS-derived founder variants ultimately contributed proportionally more informative markers in distal chromosomal regions than in pericentromeric intervals.

Importantly, the biological structure of BC₂F_{4:5} populations further mitigates concerns about uneven founder sequencing density. These materials derive from a limited number of meiotic events and therefore are expected to exhibit long-range linkage disequilibrium, with introgressed haplotypes expected to span relatively large genomic intervals (Chao et al., 2010; Zhang et al., 2010). Under such conditions, moderate marker density is sufficient to detect major introgression blocks. The objective of imputation in this framework was not to infer precise recombination breakpoints unsupported by empirical data, but rather to interpolate across missing intervals between array anchors and produce complete, genome-wide genotype matrices suitable for ancestry estimation.

Accordingly, while heterogeneity in founder sequencing methodology influenced initial variant discovery, it did not substantively constrain the detection of chromosomal introgression patterns. The primary resolution ceiling remained the density and distribution of progeny anchor markers rather than the absolute depth or platform used for founder sequencing. Nonetheless, detection of smaller introgression segments – particularly in regions with sparse anchor marker coverage – remains limited under this design.

Performance of pedigree-based imputation for introgression detection in BC₂ populations

AlphaImpute2 was originally developed for livestock populations, where recombination landscapes are comparatively uniform and allele frequencies are typically moderate (Calus et al., 2014; Hickey et al., 2012b; Whalen et al., 2020). In contrast, BC₂-derived wheat introgression populations present a markedly different context: uneven recombination across chromosomes, expected long-range linkage disequilibrium, and strongly skewed allele frequencies favoring the

RP (Chao et al., 2010; Zhang et al., 2010). These deviations from underlying assumptions influence how genotype probabilities are resolved during pedigree-based peeling.

Although marker spacing is uneven in physical distance across the wheat genome, this does not necessarily constitute a strong violation of the algorithm's assumptions of even genetic spacing. Regions with low marker density in physical distance are largely confined to regions with suppressed recombination (e.g., pericentromeric regions), whereas regions with higher marker density tend to overlap regions with higher recombination rates (Des Déserts et al., 2021; Gardiner et al., 2019; Jordan et al., 2015, 2018; Keeble-Gagnère et al., 2021; Shaltouki-Rizi et al., 2026). Thus, assuming approximately even spacing in genetic distance is reasonable at the scale relevant for pedigree-based haplotype inference. In practice, high recombining regions in the WIL progeny are anchored by more observed array genotypes, while low recombining regions contain long stretches of missing data that must be imputed based on flanking markers and pedigree information alone. In these regions, segregation probabilities are more strongly influenced by the assumed recombination rate, making the map length parameter particularly relevant.

In BC₂ populations, donor haplotypes are intrinsically rare relative to RP haplotypes. Because AlphaImpute2 iteratively updates genotype probabilities using pedigree information, segregation state inference is necessarily influenced by the predominance of RP haplotypes across related individuals. When anchor marker evidence is sparse or ambiguous, the algorithm preferentially resolves toward the more probable RP background or retains uncertainty as heterozygosity. This dynamic is particularly pronounced on the D subgenome, where both donor allele frequency and empirical anchor marker density are lower than on the A and B subgenomes.

Independent validation at the *Q/q* domestication locus supports the overall suitability of this approach. Concordance between KASP genotypes and bioinformatically detected introgression segments was high, with discrepancies expected due to imprecision in segment boundaries rather than systematic misclassification of ancestry. The observation that concordance remained strong when segments were called within 5 Mb of *Q/q* further underscores that the principal limitation lies in breakpoint resolution rather than systematic errors in ancestry assignment. At the same time, this diagnostic is not without limitations. Population size constrains the power of validation, particularly given the small proportion of WILs expected to carry the wild type *Q* allele. This reflects both the population development scheme and relatively low donor contribution, but may also have been intensified by attrition during population advancement. In particular, advancement through single seed descent may have indirectly selected against plants carrying *Q*, because the bias toward a free threshing spike would make seed retention from wild types less likely.

Genome-wide introgression patterns are also consistent with biological expectations. Donor allele proportions were generally higher in distal chromosome regions characterized by higher recombination and denser anchor marker coverage, while D subgenome donor proportions remained lower across populations. Reduced introgression in centromeric regions has been reported previously (Nyine et al., 2020; Price et al., 2025a). Similarly, the low frequency of introgression detected on chromosome 4B aligns with previous reports (Marcotuli et al., 2025; Price et al., 2025a). Thus, observed introgression profiles are consistent with known recombination landscapes and chromosomal architecture rather than being artifacts of imputation alone.

Although AlphaImpute2 demonstrates sensitivity to uneven marker density and skewed allele frequencies – manifested as longer inferred segments and elevated heterozygosity in low anchor regions – it remains well suited for this application. Its flexibility in accommodating complex pedigrees allows explicit modeling of the breeding scheme and subgenome specific donor relationships. Because high-density genotypes were available only for founders, joint population-pedigree imputation was not feasible; inclusion of densely sequenced WILs could further refine haplotype definition and improve performance in future applications.

The principal limitation of pedigree-based imputation in this context is not founder marker density but the distribution of empirical progeny anchors and the nonuniform recombination landscape of wheat. In genomic regions where nucleotide divergence between founders is elevated – such as in centromeric intervals or across the D subgenome – effective crossover frequency may be reduced (Des Déserts et al., 2021; Gardiner et al., 2019; Jordan et al., 2015, 2018; Nyine et al., 2020; Shaltouki-Rizi et al., 2026). Under such conditions, extended haplotypes can persist across generations, resulting in long-range linkage disequilibrium and larger inherited introgression blocks. This effect, along with fewer empirical anchor markers to support segregation state transitions on the D subgenome, is reflected by the relatively large mean segment sizes observed on the D subgenome. As long as residual heterozygosity is interpreted as uncertainty and conservative thresholds are applied during segment detection, pedigree-based imputation provides a robust framework for identifying genome-wide introgression in BC₂ derived germplasm. Breakpoint precision remains limited, and smaller introgressions may be missed; however, detection of major donor segments and genome-wide ancestry patterns is well supported under the implemented pipeline.

Determinants of population structure in BC₂ introgression germplasm

Across all TdGD populations, the primary axis of differentiation separated WILs according to RP background. This pattern is consistent with the BC₂ design, in which the majority of the genome is expected to derive from the RP. Thus, the dominant component of allelic variance reflects recovery of RP haplotypes.

Pop30 presented a particularly interesting case, exhibiting the greatest founder nucleotide divergence, the highest average MAF, and the most pronounced within-population structuring. A reasonable hypothesis connecting these observations is that founder divergence sets the upper bound on segregating variation entering the BC₂ generation. Populations initiated from more divergent founders inherently contain a broader pool of polymorphic loci. When a large proportion of loci segregate at intermediate frequencies, as reflected in the relatively higher MAF of Pop30, more individuals carry the minor allele at any given polymorphic site. This increases the contribution of those loci to covariance among WILs and enhances the ability of PCA to resolve axes of differentiation (Prokopenko et al., 2016; Turkmen et al., 2019).

However, founder divergence alone may not fully explain the observed MAF and structuring patterns. The number and representation of BC₂F₁ families within each population could also influence allele frequency distributions. If fewer families are disproportionately represented within a population, family-specific haplotypes may occur at more moderate frequencies population-wide, inflating MAF relative to populations in which more families contribute evenly. Conversely, when many families are represented, alleles unique to a single family lineage remain at lower frequency. Although Pop30 ultimately maintained a relatively large population size, suggesting less fertility-related attrition during single seed descent, it remains possible that differential survival occurred between families across generations of

advancement. Such family-level effects could have moderating effects on MAF and promote covariance among related WILs without requiring strong locus-specific drivers.

Importantly, structuring within Pop30 could not be attributed to a specific chromosome, nor was it explained by allelic state at *Vrn-A1*, indicating that differentiation was not driven by a single major-effect locus. Although vernalization requirement is often attributed to major loci, allelic variation persists within wild emmer germplasm and may not be fully captured by a biallelic SNP array, particularly when segregation is observed at only one homoeolog (DeWitt et al., 2021; Kato et al., 1997; Strejčková et al., 2023). A more definitive evaluation would require detailed multi-environment screening of winter tolerance within Pop30. While growth habit differences between donor and RP cannot be excluded as a contributing factor, the absence of pronounced SNP-PC correlation peaks argues against a dominant vernalization locus underlying the pattern. Rather, the data are more consistent with cumulative genome-wide allelic divergence generating mosaic haplotypic differentiation among WILs.

The feature most closely corresponding with PCA axes in Pop30 was donor proportion: genome-wide donor proportion aligned approximately with PC2, while D subgenome donor proportion aligned with PC1. At first glance, this appears inconsistent with SNP-PC correlation plots, which did not show enrichment of high correlations on D subgenome chromosomes. However, SNP-PC correlations reflect locus-specific contributions to PCs, whereas donor proportion represents a summary of ancestry across many loci. One plausible explanation is that D subgenome introgression varies among BC₂F₁ families and covaries with genome-wide haplotypic composition inherited through shared descent. Under this scenario, D subgenome donor proportion could serve as a marker of lineage structure rather than a single genomic driver.

Dispersed SNP-PC correlations would be expected if many large haplotypes across the genome contribute modestly to covariance rather than a single region dominating the signal.

Pop30 was also unique in lacking conserved centromeric regions between founders. In populations where founders share centromeric haplotypes, low recombination intervals contribute little segregating variation. In Pop30, divergence extending into centromeric regions potentially increases the size and persistence of inherited haplotypes. Elevated nucleotide divergence in such regions may reduce effective recombination, allowing longer stretches of co-inherited alleles to persist within family lineages (Des Déserts et al., 2021; Gardiner et al., 2019; Jordan et al., 2015, 2018; Nyine et al., 2020; Shaltouki-Rizi et al., 2026). This would be consistent with both the higher MAF and the diffuse but pronounced PCA structure observed. It may also help explain why D subgenome introgression – characterized by longer segment sizes despite low overall donor proportion – corresponded strongly with PC1. If D subgenome haplotypes were inherited as relatively intact blocks early in population development, variation in their retention among families could disproportionately influence covariance patterns without producing sharp SNP-PC peaks.

These observations raise the possibility that multilocus interactions contribute to the Pop30 pattern. Combinations of alleles retained within families across subgenomes could influence viability, fertility, or developmental performance during generation advancement (Nyine et al., 2020; Tessele et al., 2025b). If certain D subgenome haplotypes were preferentially retained in combination with particular A and B subgenome haplotypes, multilocus effects could generate lineage-level differentiation. Such interactions would manifest as genome-wide covariance rather than isolated locus-specific signals. While this hypothesis cannot be directly

tested with the available data, it remains consistent with the observed alignment between donor proportion and PCA axes alongside diffuse SNP-PC correlations.

This interpretation provides context for chromosome-specific signals observed in other populations. In KS061278M-4-derived populations such as Pop41 and Pop80, within-group structuring was most consistent with variation in donor introgression on chromosome 2A. Although chromosome 2A appears to contribute disproportionately to separation along PC2, it cannot be definitively concluded that introgression level on 2A directly drives structure. Similar to Pop30, chromosome 2A may represent the most visible marker of family-level haplotype differentiation rather than the underlying cause. In both cases, donor introgression could be tagging broader family structure and multilocus covariance established during population development.

It is therefore important to consider the effect of population development mechanics on the observation and interpretation of resulting structuring. Limited numbers of plants were used during crossing and advancement, and WILs derived from the same BC₂F₁ families are expected to share haplotypes at a higher frequency. Without complete traceability of family lineages across generations, the relative contributions of founder divergence, recombination dynamics, and sampling effects cannot be fully disentangled.

The observed gradients in founder divergence align with the geographic distribution of wild emmer. Donor originating from the southern regions exhibited greater divergence from hexaploid germplasm, whereas northern accessions – particularly those from Turkey – showed reduced divergence, consistent with historical gene flow between tetraploids and hexaploids (Cheng et al., 2019; Luo et al., 2007; Zhou et al., 2020). Notably, Pop32 and Pop41, both derived from Turkish accessions, shared conservation of centromeric regions on chromosomes 3A, 5A,

4B, and 5B, suggesting that haplotypes in these regions in WILs will have little segregating variation to use in determining ancestry assignment.

Overall, population structure within TdGD populations reflects the combined effects of RP background recovery, founder nucleotide divergence, quantitative variation in donor haplotype retention, and family-level structuring within populations.

Comparison between amphiploid bridge and direct crossing approaches

The overarching objective of this work was to evaluate the amphiploid bridge-mediated strategy for systematic introgression of allelic diversity from wild emmer into adapted hexaploid wheat backgrounds. The parallel implementation of a direct tetraploid x hexaploid crossing approach in Price et al. (2025a) provides an opportunity for broad comparison between these strategies with respect to introgression frequency and genomic distribution. While differences in genotyping platforms and analytical pipelines limit fine-scale equivalence, general patterns of introgression architecture can be meaningfully compared.

At a genome-wide level, both approaches produced similar introgression landscapes. In both the present study and Price et al. (2025a), introgression segments were enriched in distal, high recombination regions and reduced in pericentromeric intervals. This pattern is consistent with known recombination landscapes in wheat and supports the conclusion that both strategies are effective for transferring novel A and B subgenome haplotypes in regions where crossover frequency is relatively high.

Markedly reduced introgression on chromosome 4B was observed in both studies; however, nuances emerged when considering founder nucleotide divergence. Price et al. reported elevated nucleotide diversity on chromosome 4B when summarizing across founders, whereas in the present study, conservation of 4B centromeric regions was observed specifically in

populations derived from Turkish wild emmer accessions. This discrepancy likely reflects population-level heterogeneity that becomes obscured when founder comparisons are aggregated across geographic origins. It is reasonable to infer that if nucleotide diversity were partitioned by donor origin in Price et al., similar differences might be observed.

The D subgenome introduces a more substantive biological distinction between approaches. The amphiploid bridge strategy inherently introduces D subgenome variation derived from *Ae. tauschii*. In the present study, all amphiploids were derived from ssp. *tauschii*, whereas Nyine et al. (2020) demonstrated that introgression efficiency differs between ssp. *strangulata* and ssp. *tauschii*, with the latter often producing longer introgression segments, particularly in low recombination regions. The longer D subgenome segments observed here are consistent with those findings and may reflect elevated nucleotide divergence between the ssp. *tauschii* donor and the ssp. *strangulata*-derived D subgenome of hexaploid wheat. Greater divergence may reduce effective recombination, resulting in extended haplotypes and longer-range linkage disequilibrium (Jordan et al., 2015, 2018; Nyine et al., 2020).

Notably, Nyine et al. reported substantial variation in D subgenome introgression levels among amphiploids, with retained segments ranging from < 1% to nearly entire chromosome lengths. Because only a subset of amphiploids from that study were used to construct the present populations, variation in D subgenome haplotypic structure among amphiploid parents likely influenced downstream patterns. The D subgenome in the present materials should therefore be interpreted as a byproduct of the bridge strategy rather than as a systematically evaluated introgression target. Comprehensive assessment of D subgenome introgression was previously conducted in Nyine et al. (2020, 2021, 2025), using larger and more diverse amphiploid populations representing both subspecies.

In contrast, the direct crossing approach of Price et al. (2025a) does not introduce novel D subgenome haplotypes but instead conserves the D subgenome of the RP. However, Price et al. reported substantial D subgenome aneuploidy in some introgression lines, an expected outcome of tetraploid x hexaploid crosses where D chromosomes lack homologous pairing partners during meiosis. Aneuploidy screening and selection can mitigate this issue, but it represents an inherent challenge of direct crossing. In the amphiploid bridge approach, the presence of duplicated D subgenomes prior to the F₁ generation theoretically facilitates balanced meiotic segregation and restoration of hexaploidy, reducing the likelihood of persistent D subgenome aneuploidy. Although explicit aneuploidy screening was not feasible here due to limitations in subgenome specificity of the SNP array platform, no evidence suggested widespread D subgenome instability.

From a practical standpoint, the two approaches are broadly equivalent for large scale A and B subgenome introgression. Both generate mosaic introgression patterns constrained by recombination landscapes and structural features. The amphiploid bridge approach offers advantages in maintaining D subgenome stability and enabling simultaneous introgression across all three subgenomes. However, it requires prior development of amphiploids, representing substantial upfront investment. In contrast, direct crossing is faster to implement in the absence of existing amphiploids but may require more intensive selection against aneuploidy.

Thus, choice of approach depends on breeding objectives and available resources. If the goal is systematic A/B subgenome introgression and rapid development of materials, direct crossing is sufficient provided aneuploidy is managed appropriately. If amphiploids are already developed, the bridge strategy offers a stable mechanism for introducing wild emmer diversity while preserving D subgenome integrity and potentially enabling introgression across all three

subgenomes. In either case, both strategies demonstrate that systematic, large-scale introgression of wild emmer alleles into adapted hexaploid backgrounds is feasible and reproducible across independent pipelines.

Conclusion

This work characterized genome-wide introgression architecture in seven BC₂F_{4.5} wheat wild introgression populations developed through integration of a wheat-*Aegilops* amphiploid bridge with direct crossing to wild emmer. Across all populations, introgression was detected on the A, B, and D subgenomes, confirming that this introgression breeding strategy successfully introduces a mosaic of allelic diversity while largely restoring the hexaploid RP background.

Donor allele proportions and segment distributions aligned with expectations derived from population design and founder divergence. Introgression was most prevalent on the A and B subgenomes, with lower but consistently detectable donor contributions on the D subgenome. Genome-wide patterns reflected the recombination landscape of wheat: donor enrichment in distal chromosomal regions, reduced introgression in pericentromeric intervals, and low representation on chromosomes 4A and 4B consistent with known structural and evolutionary features of group 4 chromosomes.

Pedigree-based imputation enabled genome-wide inference of donor haplotypes from sparse progeny genotypes when supported by dense founder sequencing. Conservative parameterization and ancestry-based smoothing limited inflation of heterozygosity and reduced false positive segment calls. Validation at the *Q/q* domestication locus demonstrated high concordance between inferred introgressions and independent marker assays, supporting the reliability of chromosome scale ancestry inference while underscoring the inherent imprecision of breakpoint localization under sparse anchor marker density. Thus, while small introgressions

may be missed, and exact recombination boundaries remain difficult to resolve, major donor haplotypes and population-level ancestry patterns are robustly detected.

Importantly, these results demonstrate that the amphiploid bridge approach facilitates simultaneous introgression across all three wheat subgenomes. The resulting WILs harbor mosaic genomes combining elite hexaploid background with targeted wild diversity, and their introgression profiles are consistent with known evolutionary relationships and chromosomal constraints in wheat. This work establishes a genome-wide framework for interpreting the resulting genomic profiles of bridge-assisted introgression from wild emmer in adapted wheat germplasm. By defining the distribution and frequency of retained donor haplotypes, it provides a critical genomic foundation for evaluating the agronomic and adaptive value of these materials in future work.

Chapter 3 - Quantitative phenotypic evaluation of wild emmer introgression lines developed using an amphiploid bridge for breeding germplasm improvement

Abstract

Wild emmer wheat provides allelic and trait diversity for wheat breeding germplasm improvement, but its use in hexaploid wheat breeding is limited by yield disadvantages, lack of regional adaptation, and uncertainty surrounding the value of introgression for complex quantitative traits. This research evaluated the phenotypic and genotypic impacts of wild emmer introgression using a population developed through a wheat-*Aegilops tauschii* amphiploid bridge. Wild introgression lines (WILs) were evaluated across six Kansas environments from 2021-2023 for agronomic performance, yield and kernel traits, and functional quality. WILs displayed substantially expanded phenotypic variation relative to the recurrent parent and performance checks. WILs exhibited an overall yield disadvantage, reflecting the disruption of elite allele combinations and strong tradeoffs among developmental timing, biomass production, and grain filling dynamics under drought stressed environments. Trait relationships showed that timing and intensity of drought stress strongly shaped performance, with drought during vegetative development limiting tillering and canopy vigor, and terminal drought compressing grain filling with reductions in kernel traits. Nevertheless, transgressive individuals were recovered, with several WILs exceeding the recurrent parent and performance checks for yield, kernel and canopy traits, and functional quality. WILs generally expressed elevated grain protein, although without a corresponding increase in functional quality. Genome-wide association mapping identified biologically credible associations despite low donor allele

frequencies. Significant regions overlapped known or plausible loci for phenology, plant height, canopy traits, kernel size, grain protein, and gluten functionality. Several associations also colocalized with donor introgression regions enriched above genome-wide donor averages, supporting the contribution of retained donor haplotypes to detectable trait variation. Overall, these results indicate that amphiploid bridge-mediated introgression from wild emmer into hexaploid wheat can increase useful quantitative variation and identify candidate genomic regions for future hypothesis-driven research.

Introduction

Triticum dicoccoides (wild emmer) is a valuable genetic resource for broadening the allelic diversity available to select upon in wheat breeding programs. As an AABB tetraploid within the secondary gene pool of hexaploid bread wheat (*Triticum aestivum*, AABBDD), wild emmer enables the reintroduction of ancestral A and B subgenome variation into modern breeding germplasm through homologous recombination (Boehm & Cai, 2024; Hao et al., 2020; Nyine et al., 2020; Qi et al., 2007). This potential is relevant given the pronounced allelic bottlenecks associated with domestication, polyploidization, and continuous selection, which have reduced nucleotide diversity and expanded long-range linkage disequilibrium in modern wheat (Akhunov et al., 2010; Fu & Somers, 2009; Haudry et al., 2007). While historical allele flow between tetraploid and hexaploid wheats has shaped modern breeding germplasm and limited A/B subgenome diversity loss relative to the hexaploid D subgenome (Cheng et al., 2019; He et al., 2019; Zhou et al., 2020), it has not eliminated the need to deliberately mine wild emmer for alleles absent in elite germplasm pools.

The value of wild emmer is tempered by a persistent breeding constraint with yield remaining the strongest axis of phenotypic differentiation between modern wheat and ancestral

germplasm (Chandrasekhar et al., 2017; Deblieck et al., 2022; Hao et al., 2020; Klindworth et al., 2013; Summers & Brown, 2013). Consequently, successful deployment of wild emmer alleles depends not only on identifying useful trait variation, but also on determining whether that variation can be fixed in an adapted background without compromising agronomic performance. Historically, introgression efforts have therefore been biased toward large effect loci that can be tracked and recovered through backcrossing and marker-assisted selection. In contrast, leveraging wild emmer for improvement of quantitative traits is more difficult, because favorable alleles are dispersed genome-wide, introgression segments vary in size and location, and donor alleles may interact unpredictably with the adapted genetic background across environments (Huang et al., 2003; Klindworth et al., 2013; Lauterberg et al., 2022; Li et al., 2022; Merchuk-Ovnat et al., 2016a; Platten & Fritsche-Neto, 2023; Shahin et al., 2024).

Accordingly, most applied use of wild emmer has occurred through introgression into tetraploid backgrounds, particularly durum wheat (*Triticum durum*), where full chromosomal compatibility minimizes ploidy barriers in tetraploid x hexaploid crosses (Boehm & Cai, 2024; Hao et al., 2020; Padmanaban et al., 2017). In that context, wild emmer has been leveraged to expand the limited allelic diversity of modern durum germplasm, most notably for disease resistance and grain quality. The earliest introgression efforts prioritized qualitative resistance traits – such as powdery mildew (*Pm16*; Chen et al., 2005; Reader & Miller, 1991) and rust resistance (*Yr15*, *Yr36*, *Lr53*; Fu et al., 2009; Knott et al., 2005; Marais et al., 2005; Peng et al., 2000a; Sun et al., 1997; Uauy et al., 2005) – because major genes can be readily selected and tracked. Examples of similar deployment into hexaploid *T. aestivum* are comparatively fewer. Interspecific transfer across ploidy levels introduces additional constraints, including reduced fertility, chromosome instability, and background-dependent expression that can attenuate the

effects of otherwise valuable donor alleles (Boehm & Cai, 2024; Chen et al., 2013; Cox et al., 2017; Hiebert et al., 2020; Knott et al., 2005; Peng et al., 2013; L. Wu et al., 2015).

Nonetheless, select cases demonstrate that wild emmer alleles can be introduced to a hexaploid background with notable phenotypic effects. The *Gpc-B1* allele on chromosome 6B, which increases grain protein and micronutrient concentrations through accelerated senescence and nutrient remobilization, has been introgressed in both tetraploid and hexaploid wheat (Distelfeld et al., 2007). However, its application has been constrained by environment-dependent tradeoffs, where accelerated senescence can shorten grain fill and reduce yield potential in environments where long grain fill periods support high kernel weight (Brevis & Dubcovsky, 2010; Hale et al., 2012). Similarly, functional quality improvements have been achieved through the transfer of alleles affecting gluten composition, including activation of the otherwise silenced *Glu-1A γ* high molecular weight glutenin subunit, which can improve flour functionality (Wang et al., 2018; Xu et al., 2004). Although these alleles are unique to wild emmer and have clear effects, net performance across production environments is the most relevant metric when making breeding decisions.

Taken together, prior work establishes two points that motivate the present study. First, wild emmer contains allelic variation that can translate into agronomically beneficial phenotypes when moved into cultivated backgrounds. Second, the primary barrier to broader utilization is not the absence of promising trait variation, but the difficulty of recovering breeding value after introgression, particularly in hexaploid wheat.

At scale, a systematic effort to characterize wild emmer introgression in a hexaploid background across a comprehensive suite of traits remains limited. While such evaluations have been conducted effectively in durum wheat (Buerstmayr et al., 2013; Otto et al., 2002; Peleg et

al., 2009a, 2009b, 2011; Peng et al., 2000b, 2003), comparable large-scale characterization in hexaploid *T. aestivum* is lacking. The goal herein is to conduct a broad phenotypic assessment of wild emmer introgression lines developed in an adapted hexaploid background, to evaluate their utility for generating high performing breeding germplasm, and to assess quantitative changes in trait means and variance while maintaining regional adaptation and yield potential.

The introgression materials evaluated are BC₂F₄ derived lines developed by crossing wild emmer to a wheat-*Aegilops tauschii* amphiploid bridge, providing a vehicle for incorporation of A and B subgenome diversity into an adapted hexaploid background. Screening these wild introgression lines (WILs) across a wide range of traits related to agronomic performance, yield and kernel components, functional quality, and ionomics enables robust characterization across Kansas environments. The objectives of this work are to (1) characterize WILs for trait means, variances, and correlations, (2) associate genomic markers with trait variation using a genome-wide association study (GWAS), (3) assess overlap between marker-trait associations and regions of donor introgression, and (4) evaluate the breeding potential of the developed germplasm.

Methods

Population development

Development of the BC₂F_{4:5} WILs is described in Chapter 2. Briefly, wheat-*Aegilops tauschii* synthetic octoploid amphiploids were first generated and used as a bridge to facilitate hybridization with wild emmer. Resulting F₁s were backcrossed twice to an adapted hexaploid recurrent parent (RP) before advancement to the BC₂F_{4:5} generation via single seed descent.

Using this strategy, seven *T. dicoccoides* germplasm development (TdGD) populations were created, each derived from a distinct wild emmer accession. Given the crossing design,

WIL genomes are expected to be mosaic across all three wheat subgenomes, with an approximate allelic contribution of 8.33% from wild emmer and 4.17% from *Ae. tauschii*. The remaining genome proportion is derived from the RP background.

This design preserves the agronomic framework of the adapted hexaploid parent while enabling systematic evaluation of wild emmer introgression on the A and B subgenomes and its relationship to phenotypic variation. Given the expected low proportion of *Ae. tauschii*-derived alleles, D subgenome introgressions are likely to segregate at low frequency. Alleles at low frequency reduce the power of GWAS because the phenotypic variance attributable to a locus depends jointly on allele frequency and effect size, and variants present in few individuals are more difficult to distinguish from background phenotypic variation (Korte & Farlow, 2013; Spencer et al., 2009; Tibbs Cortes et al., 2020). In addition, donor-derived D subgenome markers with allele frequencies below the minor allele frequency threshold used for marker filtering would be removed prior to association testing. Therefore, the primary analytical focus of this work centers on characterizing trait performance as it relates to A and B subgenome diversity introduced from wild emmer.

Seed increase and population selection

Following three generations of single seed descent and the derivation of BC₂F_{4:5} WILs, seeds from a single spike of each WIL were increased during the 2019-2020 growing season in individual headrows in Yuma County, Arizona. Harvested headrow grain from all WILs across all populations was cleaned and placed into long-term storage, serving as the primary seed source for phenotypic trials in the following season.

Concurrently with the WIL seed increase in Arizona, bulk plots representing each of the seven TdGD populations were grown in Thomas County, Kansas, during the same 2019-2020

season. These plots allowed for assessment of overall field performance at the population level to facilitate selection for further phenotypic testing. Given the large number of derived WILs, rigorously testing all lines was not feasible (Table 3.1). Therefore, a visual assessment of each population was conducted, evaluating adaptation to the target environment, response to abiotic and biotic stresses, plant architecture, and maturity. Based on these criteria, TdGD population 80 (Pop80) was selected for further in-depth evaluation.

Table 3.1. *Triticum dicoccoides* germplasm development (TdGD) populations and corresponding number of wild introgression lines (WILs)

TdGD Population	Population Size (Number of WILs)
30	327
32	210
34	189
41	278
42	299
46	272
80	334

Prior introgression mapping established that TdGD Pop80 carries detectable donor introgression segments across all chromosomes except 4D, with average donor proportions of 0.09, 0.10, and 0.05 on the A, B, and D subgenomes, respectively (Chapter 2). The pedigree of Pop80 is KS061278M-4*2/3/KS061278M-4/TA 10155//PI 471021, where KS061278M-4 serves as the RP and PI 471021 is the wild emmer donor accession. PI 471021 originated from Sanhedriyya, Israel; wild emmer accessions from Israel showed greater nucleotide divergence from the RP than the Turkish and Syrian accessions used to develop other TdGD populations (Chapter 2).

Trialing and phenotypic data collection

Over three growing seasons from 2020 to 2023, TdGD Pop80 was evaluated using 50% partially replicated small-plot yield trials. Although initially consisting of 334 WILs, this set was

reduced to 279 phenotypically screened lines after selection against rachis fragility (shattering prior to mechanical harvest), poor threshability (spikelet kernel retention over 50%), and lack of winter hardiness. Trials consisted of six-row plots with a planted length of 2.44 meters (planted area of 3.83 m²), conducted across three Kansas counties: Ellis (EL), Riley (RL), and Thomas (THD). Data collection was successful in six of the nine planned testing environments, with the remaining three environments experiencing severe environmental stress, compromising reliable data collection and analysis. Trial management varied among environments, with details on crop rotation, tillage, planting and harvest dates, fertilization, and pesticide applications provided in Appendix Table B.1.

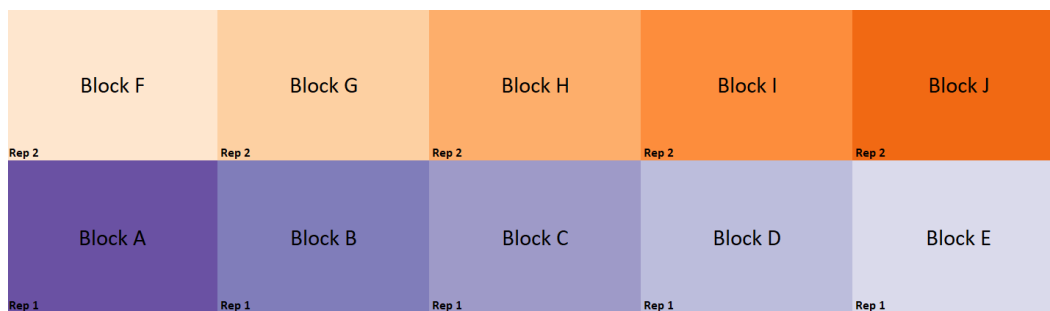


Figure 3.1. Schematic of the experimental design for TdGD Pop80 trials conducted from 2020 to 2023

Each of the ten blocks (A-J) contained complete replications of all performance checks and the RP, while WILs with two replications were evenly distributed across two replicates (Reps 1 and 2).

The partially replicated experimental design was developed using the R package DiGGer (Coombes, 2020). At each trial location, half of the WILs were tested with a single replication, while the other half received two replications. To ensure robust phenotypic comparisons, the recurrent parent (KS061278M-4) and performance checks – “Billings” (PI 656843; Hunger et al., 2014), “Joe” (PI 676270; Zhang et al., 2016), “Showdown” (PI 690085), and “SY Monument” (PI 672581) – were highly replicated in each trial. Trials were organized into ten blocks, with each block containing one replicate of each check and the recurrent parent (RP). WILs receiving

two replications were evenly allocated across two major field replicates (Figure 3.1). The performance checks were selected for their diverse and broad phenotypic responses, serving as benchmarks to evaluate performance of the WILs relative to representative breeding lines. Phenotypic data were systematically collected across the six viable environments for traits related to agronomic performance, yield and kernel components, functional quality, and ionomics.

Phenological measurements – heading date, senescence date, grain fill period

Heading date (HD) was recorded when approximately 50% of spikes within each plot emerged from the flag leaf sheath (Feekes 10.1–10.5). Senescence date (SD) was recorded when approximately 80% of tillers had lost all green tissue in the peduncle, indicating kernels had reached maximum dry weight (Feekes 11.3). In trials with both HD and SD measurements, grain fill period (GFP) was calculated as the difference between SD and HD ($GFP = SD - HD$). All phenological measurements were reported in Julian days.

Plant height

Plant height (PH) was measured at physiological maturity (Feekes 11.3). Height (cm) was averaged from representative tillers, measured from the tiller base to the top of the spike (excluding awns).

Tiller count

Tiller counts (TC) were taken approximately two weeks after heading concluded, near the start of grain filling (trial average of Feekes 10.5.4). TC was recorded as the number of productive tillers (with spikes) in a representative 30.48 centimeters of row from the center of each plot.

Drone-based measurements – NDVI, green reflectance

Spectral reflectance data were collected using unmanned aerial systems (UAS) equipped with RGB and spectral sensors (Appendix Table B.2). Flights occurred at five developmental stages: tillering (Feekes 3), jointing (Feekes 6), booting (Feekes 10), anthesis (Feekes 10.5.1-10.5.3), and grain fill/milk (Feekes 11.1-11.3). Normalized Difference Vegetation Index (NDVI) data, indicating plant health, canopy vigor, and biomass accumulation, were derived from near-infrared and red wavelengths. Green reflectance (GR), a quantification of green pixels (~530-570 nm), was included based on observed mid-season plot color variations during 2021; this metric is considered non-standard in the literature but was taken based on visual observations of plant color differences.

Grain yield

Grain yield (GY) was measured from individual six-row plots using a Zürn small plot combine equipped with an onboard weighing system. Plot grain weights were converted to kilograms per hectare (kg ha^{-1}) based on the planted plot area. Although grain moisture was recorded at harvest, valid moisture readings were not available for all plots, particularly when harvested sample size was too small to generate a reliable measurement. Therefore, to avoid differential treatment of observations during plot-level moisture correction, unadjusted plot weights were retained. A trial average moisture correction could have been applied, however, because this would apply a constant correction factor within each environment, it would not alter within-environment yield rankings. In addition, standardization of trait values within each environment prior to downstream analysis minimized the impact of environment-level differences in mean yield scale. Trial average grain moisture at harvest ranged from 8.7-10.5% (Appendix Table B.1). Grain was cleaned after harvest using an Almaco seed cleaner.

Test weight

Test weight (TW) was determined gravimetrically using a Perten IM9500 grain analyzer. Approximately 450-700 g grain samples per plot were analyzed. TW initially reported in pounds per bushel (lb/bu) was converted to kilograms per hectoliter (kg/hL).

Marvin Kernel Analyzer – kernel width, length, area, TKW

Kernel dimensions and thousand kernel weight (TKW) were measured using the Marvin Kernel Analyzer. A 15-20 g subsample from each harvested plot was weighed and analyzed individually using digital imaging to determine kernel length (KL, mm), kernel width (KW, mm), and projected kernel area (mm²). TKW (g) was automatically calculated based on the subsample weight and digitally counted kernel number. To verify the accuracy of TKW measurements from digital analysis, 5% of the samples from Thomas County in 2021 were randomly selected for manual kernel counting using an Old Mill Model 900 Seed Counter. A strong correlation (0.94) between the manual and digital methods validated the use of the Marvin Analyzer, which provided more comprehensive kernel measurements in significantly less time.

Grain protein

Grain protein (GP, %) was measured using near-infrared (NIR) spectroscopy on the Perten IM9500 grain analyzer, utilizing the same 450-700 g samples analyzed for TW. GP measurements were standardized to a moisture content of 12%. To correct potential biases in NIR predictions, particularly for non-standard introgression material, approximately 5% of samples across all six environments were randomly selected and submitted to the Kansas State University Soil Testing Laboratory for total nitrogen (N) determination via dry combustion analysis. Comparisons between combustion-based (actual) protein measurements and NIR-predicted values indicated that the Perten IM9500 tended to overestimate protein content at

higher values and underestimate at lower values. Consequently, a linear correction was applied to all NIR-based GP measurements using the following formula:

$$\text{Corrected GP} = \text{Observed GP} - (\text{Observed GP} - \text{Crossover Point}) \times (\text{Actual Slope} - 1)$$

The crossover point and slope parameters were derived from regression analysis between the combustion-based and NIR-based protein measurements.

Grain protein deviation

Grain protein deviation (GPD) was calculated to evaluate the relative GP content independent of yield-driven dilution effects. Following methods previously outlined (Haley & Stromberger, 2018; Monaghan et al., 2001; Oury & Godin, 2007), GP content was regressed on GY data within each environment to characterize the typical negative relationship between these traits. For each plot, GPD was defined as the residual, or vertical distance, between the observed GP value and the protein value predicted from the yield-based regression. Positive GPD values indicated higher than expected GP for a given GY level, while negative values indicated lower than expected GP.

Lactic acid SDS-SRC – protein quality score, solvent retention capacity

Protein quality score (PQS) and solvent retention capacity (SRC) were determined using a modified sodium dodecyl sulfate-solvent retention capacity (SDS-SRC) assay adapted from Seabourn et al. (2012). This method represents a small-scale, relatively high-throughput assay for predicting end-use breadmaking quality, correlating closely with lower throughput, conventional quality tests utilized later in breeding programs. Approximately 6 g of grain per harvested plot was milled into whole meal flour using a UDY cyclone mill. Flour samples were equilibrated for a minimum of three days in the laboratory, achieving an estimated standard moisture content of ~9.9%. For the assay, approximately 1 g of flour per sample was sequentially treated with 5 mL of lactic acid, followed by 20 mL of SDS solution. Samples were shaken,

centrifuged, and decanted to measure the weight of the resulting wet pellet. SRC was calculated as the proportion of solvent retained in the pellet, as influenced by swelling of glutenin macropolymers on lactic acid application, providing an estimate of gluten content. PQS was computed as the ratio of SRC to GP content, reflecting gluten strength and functionality. Assays were run in batches of 12 samples, with technical replicates of the wheat variety “Zenda” (PI 683512) included every three batches to ensure reproducibility across assay runs.

Standardization of phenotypic data

To enable direct comparison of trait values across environments, all phenotypic data were standardized within environment using a Z-score transformation. This approach accounted for differences in scale and units among traits, as well as environment-specific differences in measurement approach, trial management, and harvest conditions, including differences in UAS platforms and grain moisture at harvest (Appendix Tables B.1-B.2). Standardizing within environment mitigated the influence of these sources of variation while preserving relative rankings among plot values within each trial.

For each trait, observations were centered by subtracting the environment-specific mean and scaled by dividing by the environment-specific standard deviation, resulting in standardized trait distributions with a mean of 0 and a standard deviation of 1. This transformation allowed traits to be meaningfully compared in downstream analyses, while also improving stability and convergence in mixed modeling, particularly for traits with very large or very small variances. All data transformations were performed using the `dplyr` package and the base R `scale()` function (Wickham et al., 2023).

Within-environment analysis of phenotypic data

Within-environment analyses were conducted for each trait in each environment to obtain best linear unbiased estimates (BLUEs) and corresponding precision weights, thereby characterizing genotype performance after accounting for replication and spatial variation. All models were executed in Mr. Bean (Aparicio et al., 2024), which interfaces with ASReml-R (Butler et al., 2023) to fit:

$$Z_{ijk} = G_{ik} + \varepsilon_{ijk},$$

where Z_{ijk} is the standardized observation of genotype i in replicate j in environment k , G_{ik} is the fixed genotype effect (BLUE), and ε_{ijk} is the residual error. For all environments, except THD 2022, row and column spatial trends were modelled with an AR1 x AR1 residual covariance structure to correct field heterogeneity (Gilmour et al., 1997). Precision weights for each BLUE were defined as the inverse of sampling variance, $w_i = \frac{1}{\widehat{var}(G_i)}$, following the two-stage weighting scheme as described by Möhring and Piepho (2009).

In the 2022 THD trial, the residuals did not conform well to the AR1 x AR1 residual covariance structure, which warranted re-analysis of that trial data using SpATS (Rodríguez-Álvarez et al., 2018). BLUEs and weights were extracted using the same linear model, applying two-dimensional P-spline smoothing instead of using AR1 x AR1 for spatial correction. The outputs of these first-stage analyses – genotype BLUEs and precision weights for each environment – provided the inputs for subsequent across-environment modeling.

Across-environment analysis of phenotypic data

Across-environment analyses were conducted to obtain best linear unbiased predictions (BLUPs) for each genotype using a standard two-stage mixed model framework (Möhring & Piepho, 2009; Smith et al., 2001, 2005). In the first stage, within-environment analyses provided

estimates (BLUES) of genotype performance and their corresponding precision weights, reflecting the reliability of each estimate after accounting for spatial heterogeneity. These weighted BLUES were then used as input data in a second-stage analysis to forecast overall genotypic performance and quantify genotype by environment (GxE) variability across locations and years. The second-stage model was fit for each trait in ASReml-R (Butler et al., 2023), specified by:

$$\hat{G}_{ik} = E_k + g_i + (gE)_{ik} + \varepsilon_{ik},$$

where $\hat{G}_{ik}$ is the weighted within-environment estimate (BLUE) of genotype i in environment k , E_k is the fixed effect mean of environment k , g_i is the random effect of genotype i ($g_i \sim N(0, \sigma_g^2)$), $(gE)_{ik}$ is the random GxE interaction term $N(0, \sigma_{gE}^2)$, and ε_{ik} is the residual error term. A weighted residual structure was specified such that $Var(\varepsilon_{ik}) = \frac{\phi}{w_{ik}}$, where w_{ik} is the first-stage precision weight (the inverse of BLUE sampling variance). The residual scale parameter (ϕ) was held constant, resulting in fixation of residual variance (σ_ε^2 ; Butler et al., 2023). This weighting approach ensured that more precise environment-specific estimates exerted greater influence on the estimation of genotypic and GxE effects, consistent with the two-stage analysis framework described by Möhring and Piepho (2009).

Given the standardization of raw trait values within environment prior to modeling, within-environment variance was effectively constrained to a standard deviation of one. Consequently, absolute magnitudes of variance components were small, but this did not affect the interpretation of variance ratios. Standardization improved comparability among traits of differing scales and mitigated heteroscedasticity across environments. Because the use of standardized BLUES (and their precision weights) reduces heterogeneity among environments, a simple identity covariance structure was specified for both the genotypic and GxE random

effects, assuming equal variance among environments and zero covariance between them. This assumption enabled stable model convergence and facilitated straightforward estimation of variance components.

Although more complex covariance structures (e.g., unstructured or factor-analytic) were explored for some traits, they frequently resulted in convergence failures or matrix singularities due to the low absolute variance of the standardized input data. For traits where convergence was achieved, differences in information criteria were negligible and rank correlations among genotype BLUPs were nearly identical to those from the simpler identity model (data not shown). Therefore, the identity covariance structure was retained for all traits as a parsimonious and computationally robust solution.

Repeatability calculations

Repeatability (R^2) was estimated separately within individual environments and across environments to quantify the proportion of phenotypic variation attributable to genotypic differences under each modeling context. For within-environment analyses, each trait was refit within each environment with genotype treated as a random effect, rather than as a fixed effect used for estimating within-environment BLUEs. Within-environment repeatability therefore represents the extent to which replicated observations within a given environment consistently differentiate genotypes, conditional on the fitted spatial model. Within each environment, repeatability was calculated as:

$$R^2_{within} = \frac{\sigma^2_{g,k}}{\sigma^2_{g,k} + \sigma^2_{\varepsilon,k}},$$

where $\sigma^2_{g,k}$ and $\sigma^2_{\varepsilon,k}$ are the within-environment genotypic and residual variance components, respectively. Higher values indicate that a larger proportion of the remaining phenotypic variation, after accounting for spatial effects, was explained by genetic differences among

entries. Lower values indicate that residual variation was large relative to genotypic variance, suggesting reduced precision for distinguishing genotypes within that environment.

Across-environment genetic repeatability was estimated from the second-stage multi-environment BLUP model. This model partitioned the variation in weighted BLUEs into a genotype variance component (σ_g^2) and a GxE variance component (σ_{gE}^2). Because the BLUP model was fitted using precision weights derived from the first-stage within-environment analysis, uncertainty associated with the estimation of each BLUE was carried forward into the BLUP analysis through the weighting structure. Consequently, the residual variance in the BLUP model was not freely estimable; rather, the residual scale was fixed in the weighted model, while the primary estimable sources of random variation were the genotype and GxE components (Butler et al., 2023). Variation not explained by stable genotype performance is primarily represented through the GxE term, while first-stage estimation uncertainty is accounted for through the precision weights rather than through a freely estimated second-stage residual variance component (Damesa et al., 2017; Endelman, 2023; Möhring and Piepho, 2009).

As such, across-environment repeatability was calculated as:

$$R_{across}^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{gE}^2}$$

where σ_g^2 represents variance attributable to average genotype differences across environments and σ_{gE}^2 represents variation in genotype-specific deviations among environments. Under this formulation, R_{across}^2 describes the proportion of genotypic variance that is expressed consistently across environments, rather than varying by environment. Higher values indicate that genotypic differences were primarily expressed as stable main effects across environments, suggesting that genotype rankings were relatively consistent and that BLUPs provide a strong prediction of genetic value. Lower values indicate that a larger proportion of variation was environment-

specific, suggesting stronger GxE interaction and reduced consistency of genotype performance across environments.

Because a separate residual variance component was not used in the denominator, R^2_{across} should not be interpreted as the proportion of total phenotypic variance explained in the same sense as the R^2_{within} calculation. Instead, it is a variance partitioning diagnostic for the BLUP model that is used to assess the reliability of across-environment genotype summaries and to provide context for interpreting marker-trait associations detected using BLUPs. Traits with higher R^2_{across} are expected to have more stable genetic signals across environments, whereas traits with lower R^2_{across} may produce associations that are more environment dependent.

Assessment of trait relationships

Relationships among traits were assessed using Pearson correlation analysis for independent traits. Pearson correlation coefficients were extracted and visualized using the R package `corrplot` to evaluate the direction and magnitude of associations among trait values (Wei & Simko, 2024). Correlations were calculated separately for BLUEs and BLUPs. BLUE correlations describe relationships among spatially adjusted phenotypic estimates within individual environments. BLUP correlations describe relationships among predicted genotypic values across environments and were therefore interpreted as Pearson correlations of predicted genetic performance.

Derived traits were not interpreted through simple correlations with the component traits from which they were calculated. GFP (derived from HD and SD), GPD (derived from GP and GY), and PQS (derived from GP and SRC) could not be assumed independent in relationship to their respective component traits. GFP, GPD, and PQS were, however, retained in correlation analyses with traits that were independent of their derivation. For GFP, its relationship with HD

and SD was assessed using a regression model, $GFP \sim HD + SD$, describing how variation in modeled GFP was associated with the phenological traits from which it was derived after standardization and BLUE/BLUP modeling. Similarly, the relationship of PQS with its component traits was assessed using a regression model, $PQS \sim GP + SRC$. For GPD, no additional regression with component traits was performed because GPD was itself defined as the residual from the $GP \sim GY$ regression, interpreted as GP variation remaining after accounting for GY variation.

SNP genotyping and GWAS

TdGD Pop80 WILs were genotyped using the Illumina Infinium Wheat/Barley 40 K SNP array, as described in Chapter 2 (Keeble-Gagnère et al., 2021). Markers were filtered to remove monomorphic loci, markers with >10% missing data, and markers with >10% heterozygosity across WILs. After filtering, 10,621 polymorphic SNPs remained, with an average minor allele frequency (MAF) of 0.10. These markers were used as the genotypic input for GWAS conducted in GAPIT version 3 (Wang & Zhang, 2021). Phenotypic inputs consisted of both across-environment BLUPs and within-environment BLUEs.

Marker-trait associations were evaluated using three complementary GWAS models implemented in GAPIT: (1) the mixed linear model (MLM; Yu et al., 2006), (2) the multi-locus mixed model (MLMM; Segura et al., 2012), and (3) Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK; M. Huang et al., 2019). Using multiple statistical frameworks allowed identification of associations that were consistent across modeling assumptions and phenotype parameterizations, thereby increasing confidence in detected signals.

The MLM was used as the standard single marker testing framework. In this model, each SNP is tested individually as a fixed effect while controlling for population structure and

relatedness through inclusion of principal components (PCs) as fixed covariates and a kinship matrix as a random effect (Yu et al., 2006). By modeling genome-wide relatedness through kinship, MLM reduces spurious associations arising from shared ancestry or background genetic similarity among WILs.

Due to the quantitative polygenic architecture of the screened phenotypes, MLMM was also implemented to extend the single marker framework. MLMM performs forward-backward stepwise mixed model regression, iteratively adding significant SNPs as fixed effect cofactors and re-estimating variance components at each step (Segura et al., 2012). By explicitly modeling multiple associated loci as fixed effects, MLMM can increase power to detect additional loci whose effects may be partially masked under single marker testing while retaining control of population structure through the mixed model framework.

BLINK was included as an alternative multi-locus approach. Unlike MLM and MLMM, BLINK replaces the random kinship effect with a set of iteratively selected pseudo quantitative trait nucleotides (QTNs) fitted as fixed covariates (M. Huang et al., 2019). Candidate pseudo QTNs are selected based on association strength, filtered to avoid redundancy using linkage disequilibrium (LD) information, and optimized using Bayesian information criterion (BIC) to control model complexity. The framework reduces computational burden and may improve power when trait architecture includes multiple moderate to large effect loci, while controlling false positives through iterative model updating and penalization.

All GWAS models were run including two PCs. The number of PCs was determined from eigenvalue scree plots (Appendix Figure B.1), where the first two components captured the majority of structured allelic variation and subsequent components each explained <5% of remaining variance. Limiting the model to two PCs balanced control of population structure with

retention of power, avoiding excessive covariate inclusion that could reduce sensitivity to true associations. A MAF threshold of 0.05 was applied during GWAS, resulting in 3,899 SNPs retained for association testing. This reduction from 10,621 implies that most SNPs are retained at very low frequency in the population.

Results

Characterization of trialing environments

The TdGD Pop80 trials were conducted across multiple Kansas locations spanning an east-west precipitation gradient, generally capturing variation in total precipitation amount (Sittel, 2023; Lin et al., 2017; Kansas Mesonet). Although this design was intended to contrast environments differing in water availability across the growing season, persistent drought across Kansas during the trialing years, 2021-2023, constrained this differentiation (Armstrong et al., 2025). Nonetheless, variation in the timing and intensity of drought stress can be assessed rather than total precipitation amount (Figure 3.2).

In 2021 THD, drought stress was particularly acute early in the growing season, extending from post-vernalization through tillering and jointing. Precipitation improved briefly during booting and anthesis but declined again by grain filling. Consequently, 2021 THD (21THD) is best characterized as experiencing moderately severe drought stress, driven by early-season water limitation and late-season dryness.

In 2022, drought was more extreme and sustained. EL received the lowest growing season precipitation of any trial site (Appendix Table B.3), with severe drought persisting through both vegetative and reproductive stages and minimal relief from rainfall. Similarly, 2022 THD (22THD) experienced continuous drought. As such, both 22EL and 22THD represent

severely drought stressed environments, particularly during developmental stages highly sensitive to water limitation.

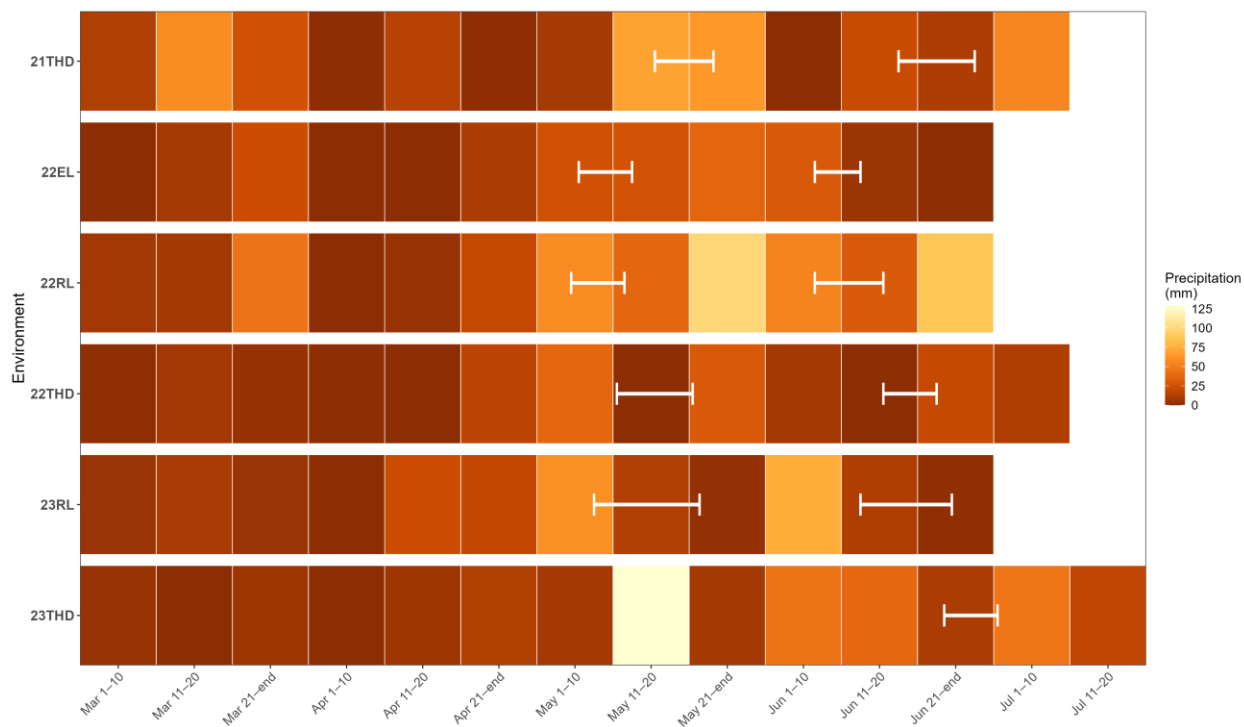


Figure 3.2. Ten-day precipitation totals from March 1 through harvest across TdGD testing environments

Precipitation (expressed in millimeters, mm) was summed within calendar intervals corresponding to days 1-10, 11-20, and 21-end of each month. Intervals are shown for each environment through the respective harvest interval, with blank space indicating periods observed outside the trial window. White horizontal bars indicate the range of plot-level heading and senescence dates, from the first plot to the last plot recorded for each developmental stage, allowing precipitation timing to be compared with key periods of crop development.

At 2022 RL (22RL), drought stress during vegetative growth paralleled that of 22EL and 22THD, but relief from precipitation occurred beginning in May and continued through harvest. However, the specific area where TdGD Pop80 was planted showed notably greater signs of water deficit stress than the surrounding plot area, as evidenced by persistent leaf rolling and accelerated senescence among the more sensitive WILs. This observation was likely due to variability in soil compaction at 22RL. Thus, 22RL is characterized as a moderate drought stress environment.

In 2023, drought persisted throughout both vegetative and reproductive phases at RL (23RL), with no substantial relief from rainfall. The timing of drought stress resembled that of 22EL and 22THD, but its severity more closely matched 21THD, as total precipitation was not as limited as in the 2022 trials (Appendix Table B.3). Therefore, 23RL can be classified as moderately severe in drought stress.

Finally, 2023 THD (23THD) most closely resembled 22RL in drought pattern, with limited precipitation during vegetative development. A rescue irrigation was reportedly applied near the end of April or beginning of May, though the exact amount and timing were not recorded. Subsequent rainfall later in the season supported crop recovery during critical reproductive stages. As a result, 23THD experienced moderately light drought stress overall, with conditions representative of near normal yield potential.

Phenotypic evaluations of TdGD Pop80

Table 3.2 summarizes results from the second stage of the two-stage mixed model analysis for TdGD Pop80 phenotypic data. Because the mixed models were fitted to within-environment standardized data (mean = 0, SD = 1), the fixed-effect estimates for locations are expressed relative to the standard normal distribution. This removes differences among environments in overall mean performance and in the absolute magnitude of phenotypic variance, allowing for evaluation of trait performance in terms of relative deviations from environment-specific means. Under this framework, GxE variance reflects inconsistency in relative genotype performance across environments. This includes crossover interaction, where genotype rankings change among environments, as well as non-crossover changes in the relative spacing among genotypes after standardization. Thus, within-environment standardization shifts the interpretation of GxE toward genotype stability and relative performance across

environments rather than differences in environmental mean or scale. Interpretation therefore focuses on variance component estimates and environment-specific fixed effects rather than on raw trait means (Appendix Table B.4).

Table 3.2. Summary of mixed model results across agronomic performance, yield and kernel component, and functional quality traits evaluated in TdGD Pop80

Fixed environment effects from the across-environment model are reported for each environment, expressed relative to a standard normal distribution. Random effect variance components are partitioned into genotypic and genotype x environment (GxE) terms. Repeatability (R^2_{across}) was calculated as the proportion of genotypic variance relative to total variance.

Trait	Fixed						Random		R ² _{across}
	21THD	22EL	22RL	22THD	23RL	23THD	Genotype	GxE	
Agronomic Performance									
HD	-0.395	-0.062	-0.217	-0.092	-0.071	-	0.468	0.185	0.717
SD	-0.252	-0.473	0.131	-0.051	-0.296	0.290	0.336	0.186	0.644
GFP	0.052	-0.427	0.136	0.049	-0.279	-	0.195	0.299	0.395
PH	0.202	-0.067	0.248	0.034	0.160	0.260	0.459	0.137	0.770
Yield and Kernel Components									
TC	-0.063	-0.481	-0.083	-	-	0.321	0.111	0.186	0.374
NDVI	0.181	-0.527	0.363	-	-	-	0.077	0.102	0.429
GR	0.166	-0.201	0.043	-	-	-	0.139	0.074	0.652
GY	-0.128	-0.452	0.248	-0.044	-0.007	0.249	0.474	0.173	0.733
TKW	-0.029	-0.190	-0.109	0.009	-0.206	-0.106	0.623	0.122	0.836
TW	-0.068	-0.109	0.036	0.076	-0.031	-0.060	0.656	0.207	0.760
KA	-0.020	-0.196	-0.117	-0.018	-0.171	-0.085	0.764	0.079	0.907
KL	-0.015	-0.196	-0.068	-0.036	-0.134	-0.094	0.856	0.042	0.953
KW	-0.076	-0.167	-0.154	0.019	-0.202	-0.085	0.673	0.150	0.818
Functional Quality									
GP	0.036	-0.264	-0.086	0.043	-0.013	0.014	0.396	0.272	0.593
GPD	-0.135	-0.461	0.120	0.017	0.059	0.112	0.290	0.347	0.455
PQS	-0.143	-0.026	0.012	-0.081	0.022	-0.117	0.456	0.243	0.653
SRC	-0.158	-0.126	-0.034	-0.064	0.132	-0.163	0.540	0.146	0.787

Agronomic performance – across environments

Measures of maturity (HD and SD) were relatively stable across environments, with R^2_{across} values of 0.717 for HD and 0.644 for SD (Table 3.2). By contrast, GFP exhibited lower R^2_{across} (0.395), reflecting a stronger environmental influence on grain fill duration. WILs

generally exhibited accelerated maturity overall relative to the RP and checks. For HD and SD BLUPs, the RP and checks clustered toward the upper (later) end of the distributions, and were centered around the RP for GFP (Figure 3.3). All checks had a longer GFP than the RP except for SY Monument (C4), reflecting the consistently later HD of C4 across environments.

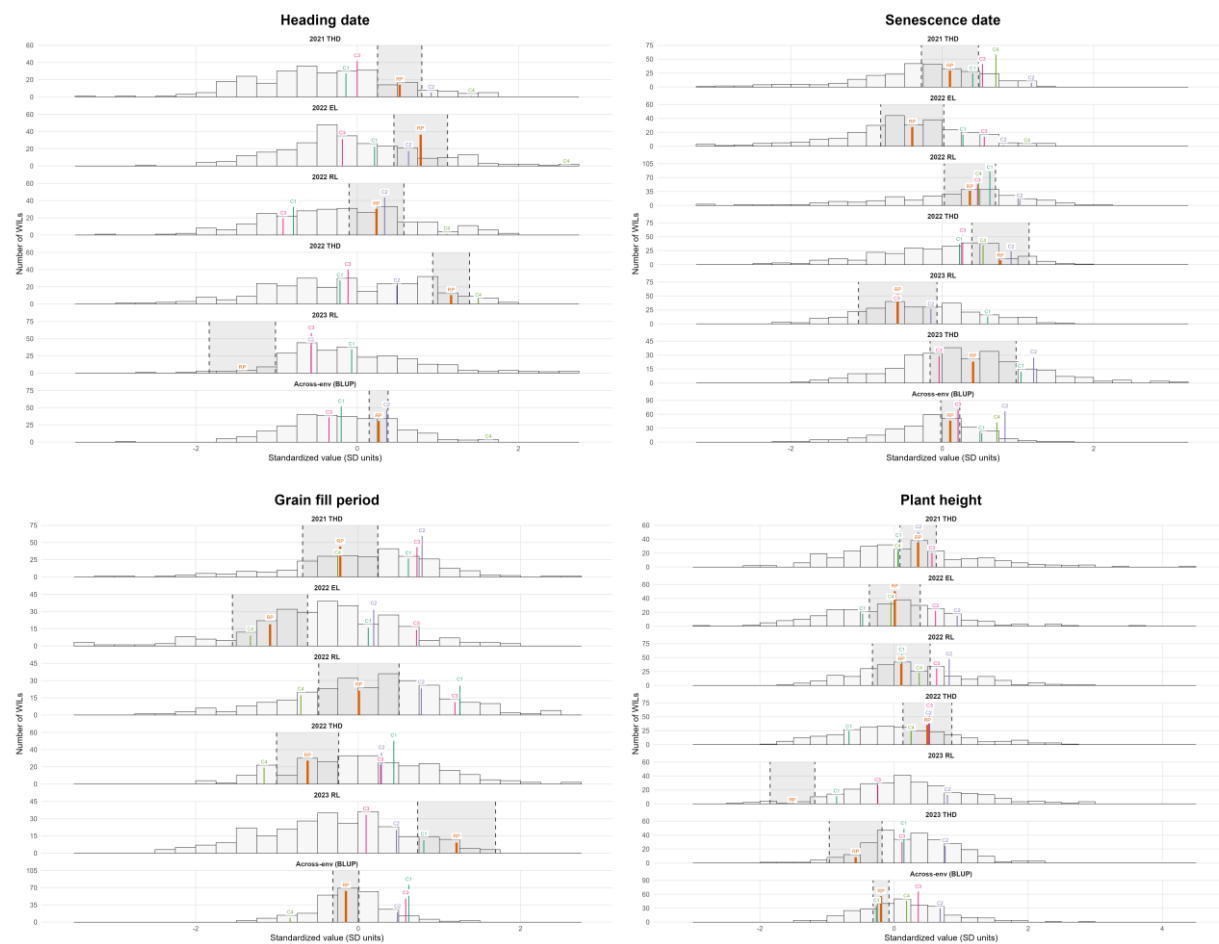


Figure 3.3. Distributions of agronomic performance traits in TdGD Pop80 within and across environments

Environment-specific facets are based on BLUEs, while the across-environment facet represents BLUPs. Colored vertical lines indicate the recurrent parent (RP=KS061278M-4) and performance checks (C1=Billings, C2=Joe, C3=Showdown, C4=SY Monument). Gray shaded areas represent the RP's approximate 95% confidence intervals for BLUEs and prediction intervals for BLUPs, calculated as each estimate ± 2 standard errors. Plots were generated in R using the ggplot2 package (Wickham, 2016).

Multiple linear regression of GFP on its component traits showed that modeled GFP retained the expected dependency on HD and SD through the BLUP modeling process. In the BLUP model, HD and SD jointly explained 93.7% of the variation in GFP (Table 3.3). The

partial regression coefficient for HD was negative (-0.786), while the partial regression coefficient for SD was positive (0.941), consistent with the definition of GFP as the interval between HD and SD. The slightly larger absolute coefficient for SD suggests that GFP was somewhat more strongly aligned with SD than HD in the BLUPs, although both component traits contributed strongly to modeled GFP variation.

For PH, the WILs were generally taller than both the RP and checks – an expected pattern given the long-standing selection for reduced plant stature and improved lodging resistance in modern breeding through *Rht* alleles, in contrast to the taller, more elongated biomass typical of wheat wild relatives (Bacher et al., 2021; Krugman et al., 2010). Lodging was more frequently observed in the WILs than in the RP and checks, particularly following wind or heavy rainfall events; however, much of this lodging was temporal. When lodging occurred, it was typically observed shortly after anthesis during grain filling, before stems were fully desiccated. As plants senesced and lost moisture, most lodged plots recovered, with stems regaining upright stature prior to mechanical harvest. Although terminal lodging scores were slightly higher in the WILs than the RP and checks, differences were not excessive, likely reflecting the relatively dry conditions across trial environments (data not shown).

Agronomic performance – within environments

The direction and magnitude of environmental effects on trait performance highlight the influence of drought timing (Table 3.2). Environmental sensitivity of GFP appears linked to drought conditions during grain filling and senescence, as the most water-limited environment (22EL) showed the largest negative effects on both SD and GFP. In 21THD, environmental effects on both HD and SD were negative and relatively large, but the negative effect on SD was smaller than that on HD, explaining the slight positive increase in GFP. In 22RL, HD also

showed a negative environmental effect, whereas SD was positive, likely reflecting the increase in late season precipitation that supported continued grain filling and extended GFP. In 23RL, both HD and SD exhibited negative environmental effects, though the effect on SD was larger, leading to an overall shortening of GFP.

Table 3.3. Multiple linear regression of GFP on HD and SD

Grain fill period (GFP) was regressed on heading date (HD) and senescence date (SD) using the model $GFP \sim HD + SD$. Environment indicates whether the regression was fitted using across-environment BLUPs or within-environment BLUEs. The partial regression coefficients (Slope) and corresponding standard errors (SE) are presented for HD and SD. VarianceExplained represents the overall model coefficient of determination.

Environment	HD		SD		VarianceExplained
	Slope	SE	Slope	SE	
21THD	-0.777	0.013	1.210	0.016	0.945
22EL	-0.906	0.017	0.980	0.018	0.980
22RL	-0.855	0.010	1.240	0.010	0.956
22THD	-1.071	0.017	0.841	0.016	0.986
23RL	-1.152	0.008	0.996	0.009	0.872
BLUP	-0.786	0.027	0.941	0.031	0.937

Multiple linear regression of GFP on its component traits further indicated that the relative contribution of HD and SD timing to GFP varied among environments (Table 3.3). In 22EL, the HD and SD coefficients were similar in absolute magnitude (-0.906 and 0.980, respectively), suggesting that GFP variation reflected approximately balanced contributions of both component traits under severe, persistent drought. In 21THD and 22RL, the SD coefficients (1.210 and 1.240, respectively) exceeded the absolute magnitude of the HD coefficients (-0.777 and -0.855, respectively), indicating that GFP variation was more strongly aligned with timing of senescence. This pattern is consistent with partial late season recovery or precipitation relief during grain fill, where delayed senescence could contribute more strongly to longer GFP. In contrast, in 22THD and 23RL, the absolute magnitude of the HD coefficients exceeded that of the SD coefficients, indicating that GFP variation was more strongly aligned with HD timing.

Notably, R^2_{within} was low for maturity traits in 23RL relative to the remaining environments, consistent with an overall lower variance explained in 23RL (87.2%; Table 3.3; Appendix Table B.6).

Interestingly, environments where HD contributed more strongly to GFP variation (22THD, 23RL) also had lower R^2_{within} for SD (0.64, 0.48, respectively) than HD (0.90, 0.63). Conversely, when SD contributed more strongly to GFP variation (21THD, 22RL), R^2_{within} for HD (0.84, 0.76) and SD (0.76, 0.79) were more similar. These patterns suggest that GFP variation was driven more strongly by HD when SD was less consistently expressed across genotypes. Persistent drought during grain filling can accelerate senescence, whereas HD has already been established by that point. Thus, lower R^2_{within} for SD in 22THD and 23RL suggests that environmental stress weakened the genotypic signal for senescence timing, causing GFP variation to be driven more strongly by HD (Appendix Table B.6).

Although SD was accelerated in the WILs, GFP was not consistently shortened relative to the RP (Figure 3.3). The direction and magnitude of WIL GFP responses depended on timing and severity of drought stress. In 22EL, the RP had delayed HD relative to the WILs without a corresponding delay in SD, resulting in a shorter GFP than most WILs. This suggests that earlier HD in the WILs may have allowed grain filling to begin sooner under early-season drought, partially buffering GFP duration relative to the RP. In contrast, under moderate drought stress in 23RL, many WILs had later HD than the RP without delayed SD, resulting in shorter GFP. Accelerated senescence of WILs was also evident in 22THD, likely reflecting restricted water availability during late grain fill. However, in 21THD and 22RL, WIL GFP values were more centered around the RP, suggesting that when precipitation relief better aligned with senescence

timing, GFP variation reflected genotypic differences more than stress-induced compression of the GFP.

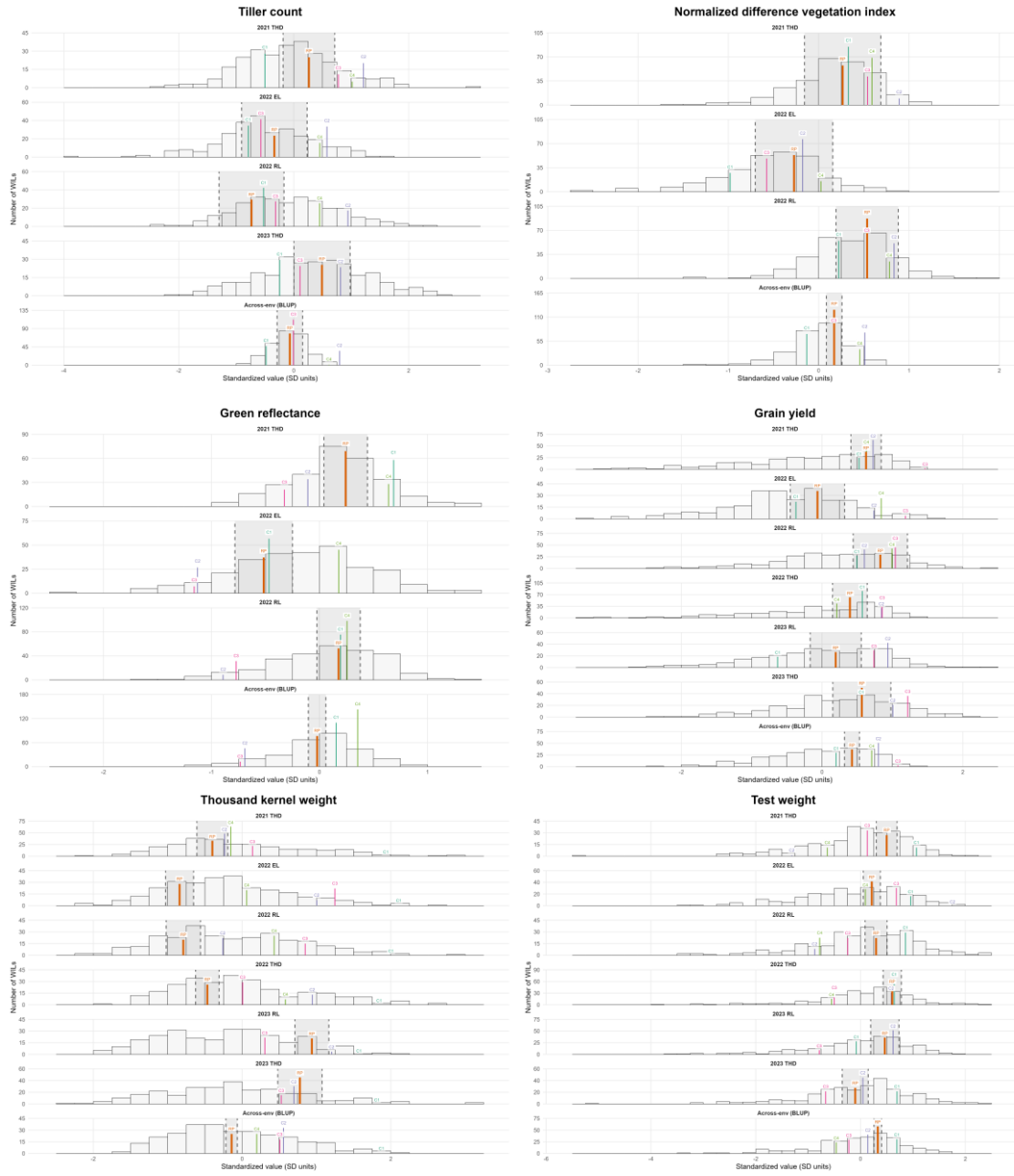
Plant height was reduced in 22EL (Table 3.2), consistent with drought-related reductions in plant height and biomass accumulation (crop stunting) commonly reported in wheat under water stress (Bacher et al., 2021; Hu et al., 2006; Krugman et al., 2010; Lauterberg et al., 2022; Lollato & Knapp, 2020; Peleg et al., 2005; Wang et al., 2017). In 23RL and 23THD, the RP occupied the lower tail of the PH distribution, indicating that WILs maintained greater PH under well-watered or moderately dry conditions (Figure 3.3). Although PH does not directly measure shoot biomass, the ability to maintain height under water stress can indicate reduced drought-induced growth limitation. However, under more extreme drought, WIL height declined more sharply than the RP's, suggesting greater plasticity but also greater sensitivity to severe stress, highlighting a differential response of WIL plant stature to drought intensity.

Yield and kernel components – across environments

Despite its highly quantitative nature and dependence on complex source-sink dynamics and environmental variability (Golan et al., 2015; Merchuk-Ovnat et al., 2016b; Murchie et al., 2023; Reynolds et al., 2017; Slafer et al., 2023), GY exhibited a relatively high R^2_{across} (0.733; Table 3.2), underscoring substantial genotypic differentiation among WILs. Across-environment repeatabilities for source component traits (TC = 0.374; NDVI = 0.429; GR = 0.652) were lower relative to kernel traits (TKW = 0.836; TW = 0.760; KA = 0.907; KL = 0.953; KW = 0.818), indicating higher environmental sensitivity of biomass development and consistent with the greater genetic control of sink components (Table 3.2; Bacher et al., 2021; Golan et al., 2015; Wang et al., 2017). Examining GR and NDVI by growth stage, GR R^2_{across} was the highest mid-season (GR_joint = 0.610; GR_boot = 0.787; GR_anth = 0.733) and lowest at the beginning

(GR_till = 0.226) and end (GR_milk = 0.321) of the growing season (Appendix Table B.5).

Stage-specific NDVI R^2_{across} steadily increased throughout the growing season (NDVI_till = 0.251; NDVI_milk = 0.533). This pattern indicates that canopy vigor, as measured by NDVI, was most environmentally influenced early in the season but became increasingly stable later in development.



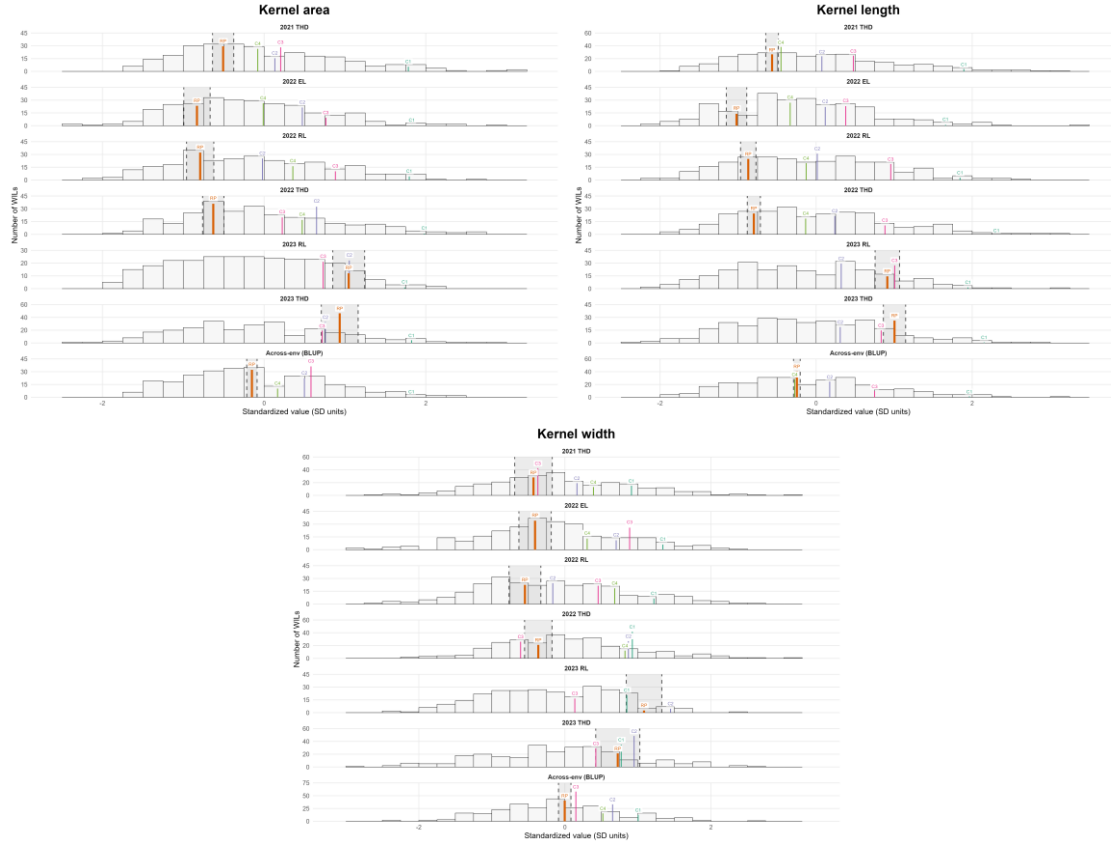


Figure 3.4. Distributions of yield and kernel component traits in TdGD Pop80 within and across environments

Environment-specific facets are based on BLUEs, while the across-environment facet represents BLUPs. Colored vertical lines indicate the recurrent parent (RP=KS061278M-4) and performance checks (C1=Billings, C2=Joe, C3=Showdown, C4=SY Monument). Gray shaded areas represent the RP's approximate 95% confidence intervals for BLUEs and prediction intervals for BLUPs, calculated as each estimate ± 2 standard errors. Plots were generated in R using the ggplot2 package (Wickham, 2016).

Traits related to vegetative biomass (TC, NDVI, GR) were generally positively correlated with GY, whereas kernel traits were negatively correlated, with the exception of TW, which was moderately positively correlated with GY (0.53; Figure 3.6). The highest stage-specific correlation between NDVI and GY was at tillering, with correlations decreasing gradually but remaining positive through subsequent stages. Green reflectance showed stage-dependent directionality across stages, with positive correlations with GY during early vegetative growth (tillering), negative correlations mid-season (jointing-booting), and positive again during reproductive development (anthesis-milk). Notably, GY correlated more strongly with mean

NDVI across stages (0.34) than with GR (0.07), consistent with GR's variable stage-specific relationships.

Looking more closely at comparison of WILs to the RP and checks, WILs had a pronounced yield disadvantage relative to the RP and checks (Figure 3.4). However, several WILs outyielded even the top-performing check, demonstrating the potential for favorable allele combinations from wild introgression. Across environments (BLUPs), no individual WIL exceeded the highest check for TC, though distribution tended to be centered around the RP. For kernel traits, values for WILs spanned and often exceeded the range of values for checks (Figure 3.4). Negative skews relative to the performance checks were observed in WIL distributions for TKW, KA, KW, and KL, although no check exceeded the maximum kernel size obtained by any individual WIL. Notably, the negative distributional deviations in the WILs were not pronounced relative to the RP.

The WILs had a wider range in GR and NDVI performance relative to the RP and checks, with no clear directional skew. Interestingly, C4 consistently scored highest among checks for GR despite having a late HD and shortened GFP (Figures 3.3-3.4). The combination of high GR, late HD, shortened GFP, and strong tillering capacity suggests that C4 may represent a physiologically distinct check, achieving canopy performance through trait combinations not observed in most other lines. This supports the suggestion that GR may be capturing canopy or photosynthetic traits distinct from biomass accumulation. If GR were solely detecting stay-green effects, genotypes with a high GR would be expected to have a longer GFP; instead, C4 combined high GR with shortened grain fill, indicating that GR perhaps reflects canopy vigor or light interception efficiency rather than delayed senescence *per se* (Christopher et al., 2014, 2016; Gurumurthy et al., 2023; Langridge & Reynolds, 2021; Latif et al., 2020; Lauterberg et al.,

2022; Lopes & Reynolds, 2012; Peleg et al., 2005; Shokat et al., 2023). Additional support for this observation comes from correlation structure; GR was negatively correlated with TC (-0.17), whereas NDVI was positively correlated with TC (0.31; Figure 3.6).

Yield and kernel components – within environments

Similar to agronomic performance, variation in yield and kernel component trait responses reflected the timing and severity of drought stress across environments. This pattern was particularly apparent for GY where the strongest negative environmental effect occurred in 22EL, the most water limited site (Table 3.2). Conversely, 22RL and 23THD showed the largest positive effects, reflecting higher yield potential under more favorable moisture conditions. Within-environment repeatability for GY was high across trials, though the lowest occurred in 23THD (0.70; Appendix Table B.6).

Traits corresponding to vegetative biomass development (TC, NDVI, GR) mirrored the environmental patterns observed for PH (Table 3.2). In 22EL, all three traits showed strong negative effects, reflecting the marked reduction in shoot biomass under severe drought stress (Bacher et al., 2021; Hu et al., 2006; Krugman et al., 2010; Lauterberg et al., 2022; Lollato & Knapp, 2020; Peleg et al., 2005; Wang et al., 2017). Conversely, 23THD had the most positive effect on TC, highlighting the capacity of WILs to produce more tillers under adequate water availability. Notably, TC displayed its highest R^2_{within} in 23THD (Appendix Table B.6), suggesting substantial genotypic variation in tillering capacity that becomes most apparent under non-limiting water availability. In both 23THD and 22RL – environments characterized by relatively mild drought – the distribution of TC among WILs extended further above the checks and RP, with the RP in 22RL specifically positioned toward the lower end of the distribution

(Figure 3.4). These patterns demonstrate the considerable potential of WILs for enhanced tillering and vegetative vigor when adequate soil moisture supports biomass development.

When NDVI and GR were evaluated across growth stages (Appendix Table B.5), distinct drought-related trends emerged. In 22EL, NDVI effects became progressively more negative from tillering through grain fill, mirroring the persistent water deficit throughout the season. In contrast, 22RL showed increasingly positive NDVI effects at later stages as precipitation improved. Strong positive effects for GR during booting and anthesis were observed in 21THD. Stage-specific patterns in R^2_{within} for GR varied by environment. In 22RL, GR R^2_{within} was highest at the milk stage (0.84) and lowest during jointing and booting (0.68, 0.73, respectively), whereas in 21THD it peaked from jointing through anthesis (0.80-0.85). In contrast, 22EL exhibited a steady increase in GR R^2_{within} across growth stages. These contrasting trends emphasize that the timing and intensity of drought stress may modulate how consistently genotypes express canopy reflectance traits across development.

Although R^2_{within} was consistently high for kernel traits, timing of drought stress still influenced trait expression. In 23RL, for instance, WILs exhibited delayed HD relative to the RP, which shortened the GFP (Figure 3.3). This corresponded with strong negative effects on TKW and KW (Table 3.2), suggesting that limited time for grain filling constrained final kernel size. Although 23RL was not as limited in terms of water availability, delayed heading in WILs under moderate drought stress reduced sink development. The RP generally fell within the upper tail of the KW distribution in 23RL and 23THD, consistent with longer GFP facilitating greater grain filling. In those same environments, the RP was also high for TKW, KA, and KL, whereas in other environments (particularly 22EL and 22RL) the RP fell within the lower end of these distributions. The high R^2_{across} of TKW, KW, and KL is visually apparent from the consistent

ranking of checks and RP within environments. The variable position of the RP relative to WILs suggests greater phenotypic plasticity among WILs, reflecting heightened responsiveness to environmental variation (Bacher et al., 2021; Peleg et al., 2005; Wang et al., 2017).

An interesting contrast emerged between 22EL and 22RL, where both environments showed large negative effects for TKW and KW, yet their effects on GY diverged – strongly negative in 22EL but positive in 22RL. This suggests that GY differences between the two sites were driven more by source-related canopy traits or yield component compensation than by kernel size alone. Indeed, in 22EL, TC, NDVI, and GR all showed large negative effects, indicating drought-induced reductions in canopy development and source capacity. In contrast, 22RL had a large positive effect for NDVI, a small positive effect for GR, and only a small negative effect for TC, suggesting that greater canopy function may have supported higher GY despite reductions in kernel size (Table 3.2; Beral et al., 2020; Brancourt-Hulmel et al., 2003). This pattern is consistent with GY improvements through greater kernel number per unit area, where smaller kernels may be offset by more kernels or improved maintenance of source activity during grain filling. The modulating effects of source and sink component traits on GY is further supported through correlation, where GY was positively associated with TC, NDVI, GR, and TW, and negatively with TKW, KA, KL, and KW (Appendix Figures B.3-B.8). Across all environments, KA and KL consistently had negative effects – strongest in 22EL – reinforcing the impact of severe drought on grain formation (Table 3.2; Farooq et al., 2011; Pradhan et al., 2012).

Functional quality – across environments

Across-environment repeatability for traits related to functional quality was moderate overall (Table 3.2); GP had an R^2_{across} of 0.593, GPD $R^2_{\text{across}} = 0.455$, PQS $R^2_{\text{across}} = 0.653$, and

SRC $R^2_{\text{across}} = 0.787$. Notably, R^2_{across} for GP and GPD was slightly lower than previously reported (Fatiukha et al., 2020a; Guttieri et al., 2017). This discrepancy implies greater influence of GxE relative to genotypic effects, potentially due to variable drought timing and intensity among environments. Late-season drought occurring during grain filling when both total grain weight and protein proportions are determined, would be expected to have the strongest influence (Singh et al., 2012; Yang et al., 2023).

A pronounced positive tail was observed in the WILs for GP (Figure 3.5), consistent with wild emmer's well-documented contribution to enhanced protein accumulation (Distelfeld et al., 2007; Fatiukha et al., 2020a; Gonzalez-Hernandez et al., 2004; Govta et al., 2024; Peleg et al., 2009a; Wang et al., 2018). Similarly, WIL distributions for GPD showed a positive tail; the RP had higher GP and GPD BLUPs than the checks. These patterns diverged for the functional quality traits (SRC, PQS). In contrast to GP and GPD, relative to the checks or RP, there was no distinct tail for SRC or PQS. Few WILs exceeded the check with the highest SRC (C1), and PQS values of the best WILs matched those of C1 and C4. Thus, while WILs tended to accumulate more protein, that additional protein did not necessarily translate into improved functional quality.

Multiple linear regression of PQS on its component traits indicated that modeled PQS was more strongly aligned with SRC than with GP (Table 3.4). In the BLUP model, the partial regression coefficient for SRC was positive and larger in magnitude (0.866) than the negative coefficient for GP (-0.497), consistent with the calculation of PQS as SRC divided by GP. The stronger alignment of PQS with SRC may partly reflect the higher R^2_{across} of SRC relative to GP, indicating that SRC contributed a more stable genotypic signal to modeled PQS variation. Together, SRC and GP explained 90.7% of the variation in PQS BLUPs, confirming that

modeled PQS retained the expected dependency on its component traits after standardization and BLUP modeling.

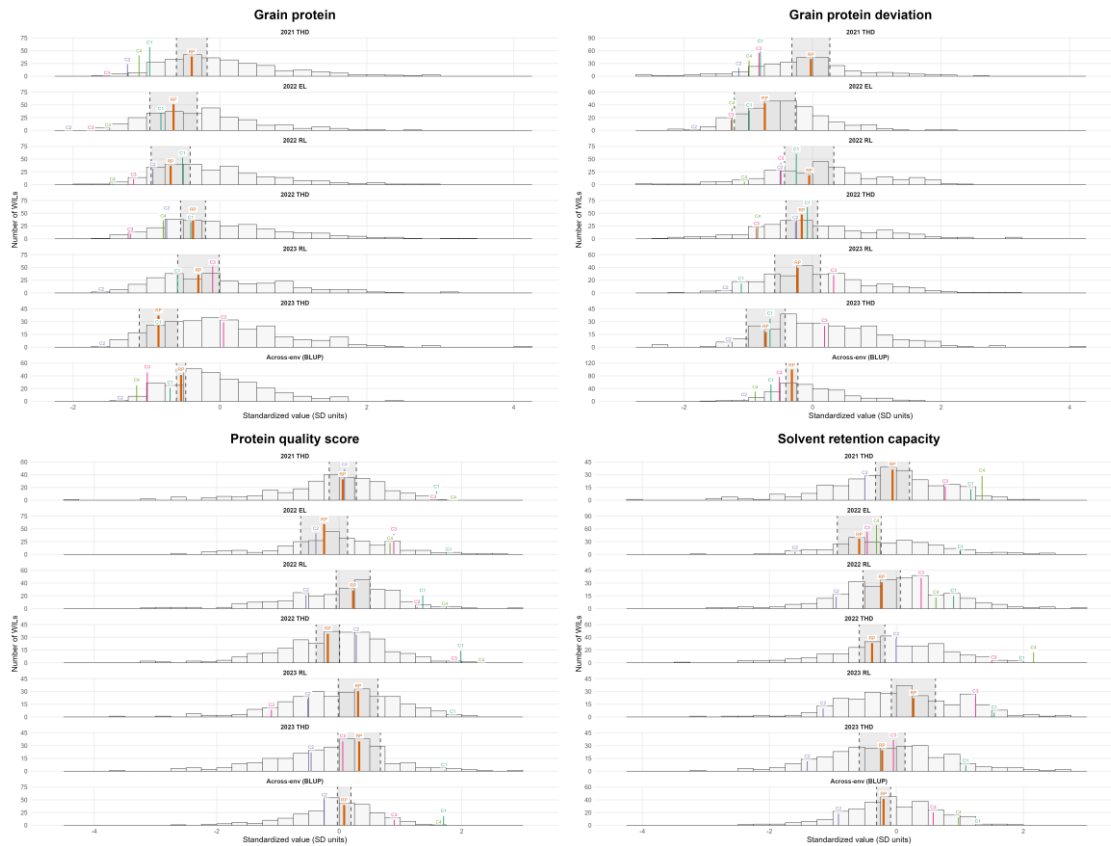


Figure 3.5. Distributions of functional quality traits in TdGD Pop80 within and across environments

Environment-specific facets are based on BLUEs, while the across-environment facet represents BLUPs. Colored vertical lines indicate the recurrent parent (RP=KS061278M-4) and performance checks (C1=Billings, C2=Joe, C3>Showdown, C4=SY Monument). Gray shaded areas represent the RP's approximate 95% confidence intervals for BLUEs and prediction intervals for BLUPs, calculated as each estimate ± 2 standard errors. Plots were generated in R using the ggplot2 package (Wickham, 2016).

Grain protein and GY BLUPs were negatively correlated (-0.69); GP and SRC BLUPs were only weakly correlated (0.20; Figure 3.6). Negative correlations were generally observed between functional quality and agronomic performance (Figure 3.6). Senescence date was negatively correlated with GP (-0.31), GPD (-0.38), PQS (-0.19), and SRC (-0.38). Grain fill period also showed negative correlations with GP (-0.25), GPD (-0.32), and SRC (-0.20), with a very weak negative correlation with PQS (-0.06). Heading date correlations were similarly weak

but negative across all quality traits. Plant height was negatively correlated with PQS (-0.28) and SRC (-0.30), while TC was positively correlated (0.21 and 0.11, respectively). Kernel component traits (TKW, KA, KL, KW) were all negatively correlated with PQS, the strongest being KW (-0.25). Grain protein also had weak negative correlations with TC (-0.25) and NDVI (-0.31). Grain protein and GPD were moderately positively correlated with TKW, KA, KL, and KW, and negatively with TW (Figure 3.6).

Table 3.4. Multiple linear regression of PQS on GP and SRC

Protein quality score (PQS) was regressed on grain protein (GP) and solvent retention capacity (SRC) using the model $PQS \sim GP + SRC$. Environment indicates whether the regression was fitted using across-environment BLUPs or within-environment BLUEs. The partial regression coefficients (Slope) and corresponding standard errors (SE) are presented for GP and SRC. VarianceExplained represents the overall model coefficient of determination.

Environment	GP		SRC		VarianceExplained
	Slope	SE	Slope	SE	
21THD	-0.433	0.019	0.868	0.017	0.977
22EL	-0.770	0.010	1.111	0.009	0.922
22RL	-0.617	0.022	1.069	0.020	0.967
22THD	-0.365	0.013	0.901	0.012	0.993
23RL	-0.354	0.005	0.580	0.005	0.352
23THD	-0.396	0.053	0.521	0.050	0.311
BLUP	-0.497	0.056	0.866	0.052	0.907

Grain protein was consistently negatively correlated with NDVI across all growth stages, strongest during tillering and jointing (-0.29) and weakest during anthesis (-0.18; Figure 3.6). Correlations with GPD were weaker but increased in magnitude through the growing season, with the strongest negative association at NDVI_milk (-0.16). Solvent retention capacity and PQS were also negatively correlated with NDVI_milk (-0.11 and -0.24, respectively), with weaker correlations earlier in development. Correlation patterns for GR by growth stage were similar to GY~GR correlations, with shifting directions of correlation depending on growth stage (Figure 3.6). Correlations with GP shifted from negative during early growth (GR_till = -0.24) to

positive around jointing and booting, and back to negative by anthesis and milk (GR_milk = -0.21).

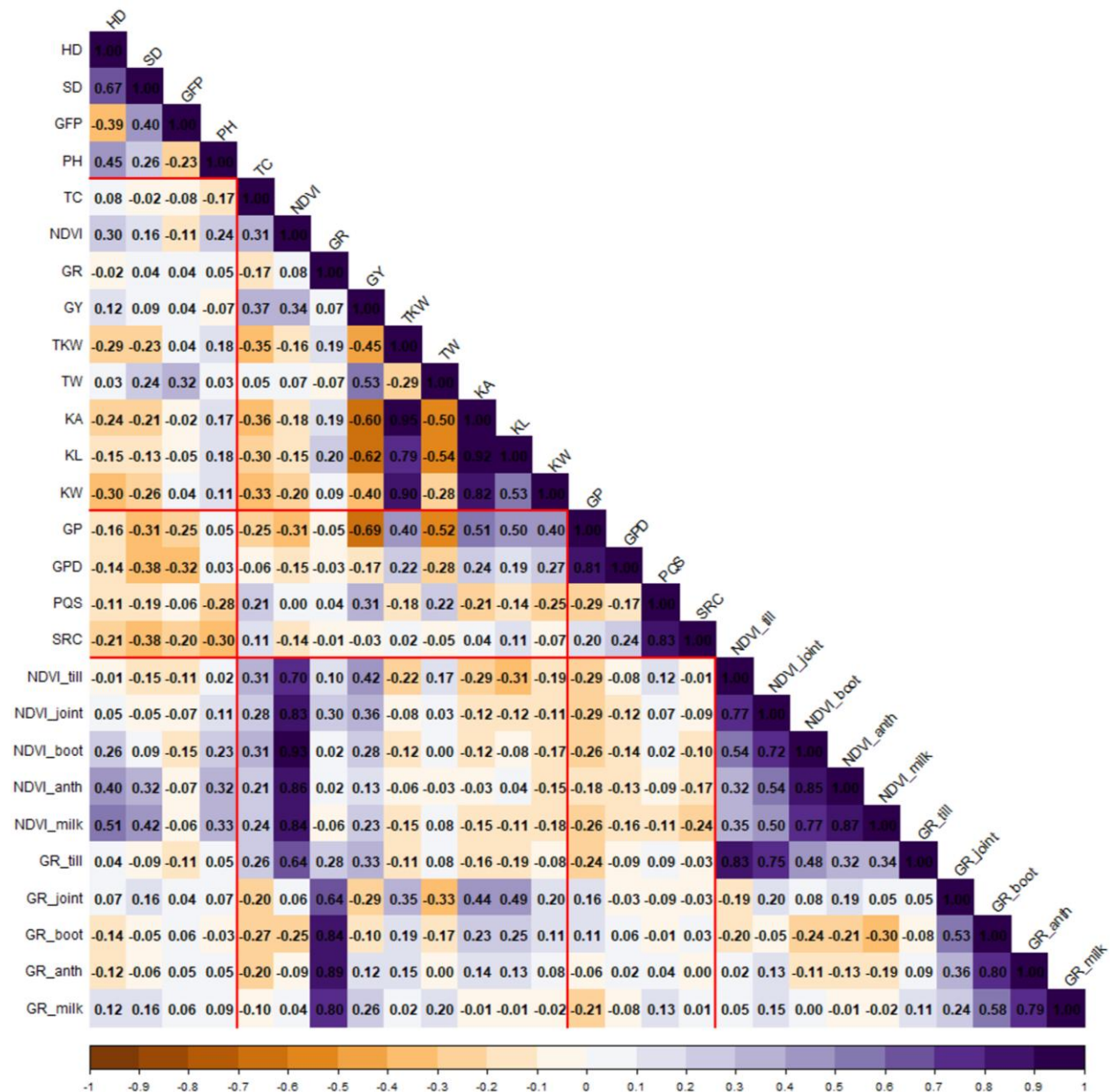


Figure 3.6. Pearson correlations among agronomic performance, yield and kernel component, functional quality, and extended drone data BLUPs in TdGD Pop80 WILs

Values within each cell represent the correlation coefficient, with color indicating both magnitude (darker = stronger) and direction (orange = negative, purple = positive). Red grid lines delimit groupings of traits separated by agronomic performance, yield and kernel components, functional quality, and extended drone data. Correlations were computed and visualized in R using the corrplot package (Wei & Simko, 2024).

Functional quality – within environments

Environmental effect estimates align with the influence of drought on grain quality (Table 3.2). In 21THD and 22EL, environmental effects across quality traits were generally negative, except for GP in 21THD, which was slightly positive. Both environments experienced notable terminal drought, though the 22EL drought was more severe overall, with large negative effects for GY and correspondingly large decreases for GP (-0.264) and GPD (-0.461). In contrast, environments where drought stress was confined to early in the season (22RL, 23THD) showed the largest positive effects for GPD (0.120 and 0.112, respectively) and small effects for GP (-0.086, 0.014). This temporal distinction supports the interpretation that terminal stress limits GP accumulation to a higher degree than drought stress during vegetative development (Singh et al., 2012; Yang et al., 2023).

None of the environments showed a strong positive effect for GP, with the largest being 22THD. The three environments with positive effects for GP were all located in THD (21THD = 0.036; 22THD = 0.043; 23THD = 0.014), supporting the idea that local nitrogen availability at that site enabled higher protein accumulation relative to yield. Interestingly, 22RL, the environment with the highest positive GPD effect (0.120), also had the lowest R^2_{within} for GPD (0.64; Appendix Table B.6). For GP, R^2_{within} ranged from 0.86 in 23THD to 0.94 in 21THD. Consistently high R^2_{within} across quality traits was observed in 22THD, which included the highest R^2_{within} across environments for PQS (0.92) and for SRC (0.91). Because PQS is derived from GP and SRC, it follows that when both component traits have high R^2_{within} , PQS R^2_{within} will likewise be high. Solvent retention capacity showed negative environmental effects across all sites except 23RL, where a relatively large positive effect (0.132) was observed. Protein quality

score also had negative effects across most environments, with small positive effects only in 22RL (0.012) and 23RL (0.022; Table 3.2).

The positive tails for GP and GPD BLUPs were similarly observed in the BLUEs (Figure 3.5). The RP exceeded all checks for GP in 21THD, 22EL, and 22THD. The RP was slightly outperformed by C1 in 22RL, by C1 and C3 in 23THD, and by C3 in 23RL. For GPD, the RP was only outperformed by C1 in 22THD, C3 in 23RL, and both C1 and C3 in 23THD.

Correlations between quality traits in the BLUEs matched the directions of BLUP correlations (Appendix Figures B.3-B.8). Grain protein and SRC were positively correlated in all but 21THD and 22THD, where they were uncorrelated. Grain protein and GY had a consistent negative correlation in each environment. Multiple linear regression of PQS on its component traits in the BLUEs generally matched the pattern observed in the BLUPs; SRC consistently showed a positive coefficient and GP consistently showed a negative coefficient, with the SRC coefficient being larger in absolute magnitude (Table 3.4). The overall model coefficient of determination was greater for each BLUE model than for the BLUP model, except in 23RL and 23THD. In those environments, SRC and GP jointly explained only 35.2% and 31.1% of the variation in PQS, respectively.

The strongest correlations observed between GP or GPD and maturity traits occurred in 21THD ($GP \sim SD = -0.57$, $GP \sim GFP = -0.46$, $GPD \sim SD = -0.48$, $GPD \sim GFP = -0.40$; Appendix Figure B.3). These negative correlations suggest that later senescence and longer GFP were associated with lower GP, likely reflecting dilution of grain nitrogen when grain filling was extended (Bagherikia et al., 2025; De Oliveira Silva et al., 2026). Conversely, earlier senescence or shortened GFP may have increased GP by reducing kernel fill, a pattern commonly observed under drought stress. This interpretation is also consistent with timing of GP deposition

occurring relatively early during grain fill, whereas terminal stress more strongly limits grain size (Guttieri et al., 2001; Martre et al., 2003; Triboi et al., 2003). In contrast, 22EL showed almost no correlation between maturity and protein traits, suggesting that uniformly severe drought conditions may have reduced variation in developmental timing and weakened genotype-specific relationships among these traits (Appendix Figure B.4). Interestingly, functional quality traits were not always affected in the same way as GP or GPD. In 21THD, SRC was uncorrelated with maturity traits, and PQS was only weakly positively correlated ($HD = 0.16$, $SD = 0.23$, $GFP = 0.13$; Appendix Figure B.3). This distinction suggests that drought-related increases in GP do not necessarily correspond to improved functional quality, likely because stress can alter these characteristics differently (Bagherikia et al., 2025; Farooq et al., 2011; Labuschagne et al., 2021).

Correlation trends in the BLUPs with quality and stage-specific NDVI and GR measurements were consistent in the BLUEs. In 21THD, GP had the strongest negative correlation with NDVI_joint (-0.31) and the weakest with NDVI_anth (-0.20 ; Appendix Figure B.3). In 22EL, the strongest negative correlation was with NDVI_till (-0.23 ; Appendix Figure B.4), and in 22RL it was with NDVI_milk (-0.26 ; Appendix Figure B.5). Correlations between GPD and NDVI were weaker overall, ranging from -0.17 in 21THD (NDVI_till; Appendix Figure B.3) to near zero in 22EL (Appendix Figure B.4).

Marker-trait associations in TdGD Pop80

GWAS was conducted to identify genomic regions associated with variation in trait performance within TdGD Pop80. Analyses were performed in GAPIT using MLM, MLMM, and BLINK (M. Huang et al., 2019; Segura et al., 2012; Wang & Zhang, 2021; Yu et al., 2006). For each significant association, results include the physical position of the marker, MAF, and estimated effect of allele substitution (based on standardized phenotypes).

The allelic states of the RP and donor parents were obtained from the founder genome sequence data described in Chapter 2. Associations were further characterized according to the alleles associated with lower and higher mean phenotypic values among observed genotype classes. Counts of WILs carrying the RP versus donor allele were also recorded when parental states could be assigned unambiguously. This framework allowed donor effects to be interpreted in terms of observed genotype class means, indicating whether the donor allele was associated with an increase or decrease in the trait. For A and B subgenome associations, the donor corresponds to wild emmer, whereas for D subgenome associations it corresponds to *Ae. tauschii*.

Finally, donor allele proportions within the 1 Mb genomic bin containing each associated marker, as estimated in Chapter 2, were used to evaluate whether marker-trait associations correspond with regions of donor introgression. This integration of GWAS results with introgression mapping enables assessment of whether phenotypic effects are directly associated with introgressed donor segments.

Agronomic performance

The highest confidence associations for agronomic performance were detected on the short arm of chromosome 2B for HD and SD. For the BLUPs, SD showed strong support across models, with significant markers spanning 45.55-65.23 Mb identified by BLINK, MLM, and MLMM (Table 3.5; Figure 3.7). An association for HD was also detected at 45.55 Mb using BLINK. Similar signals for both HD and SD were detected within individual environments (Appendix Table B.7).

Interpretation of parental allele effects in this region is limited by incomplete coverage of the significant markers in the founder sequencing data. Consequently, it cannot be definitively

determined whether the wild emmer donor contributes earlier or later maturity at most loci. Variation in donor allele proportions across the region suggests potential recombination within the ~20 Mb interval associated with HD and SD. Donor proportions are relatively elevated at 45.55 Mb (0.31) and 65.23 Mb (0.34), but much lower at 59.10 Mb (0.07) and 63.85 Mb (0.08). Among these markers, only the locus at 63.85 Mb contains complete parental allele information; comparison of genotype class means indicates that the donor allele is associated with increased SD (later senescence).

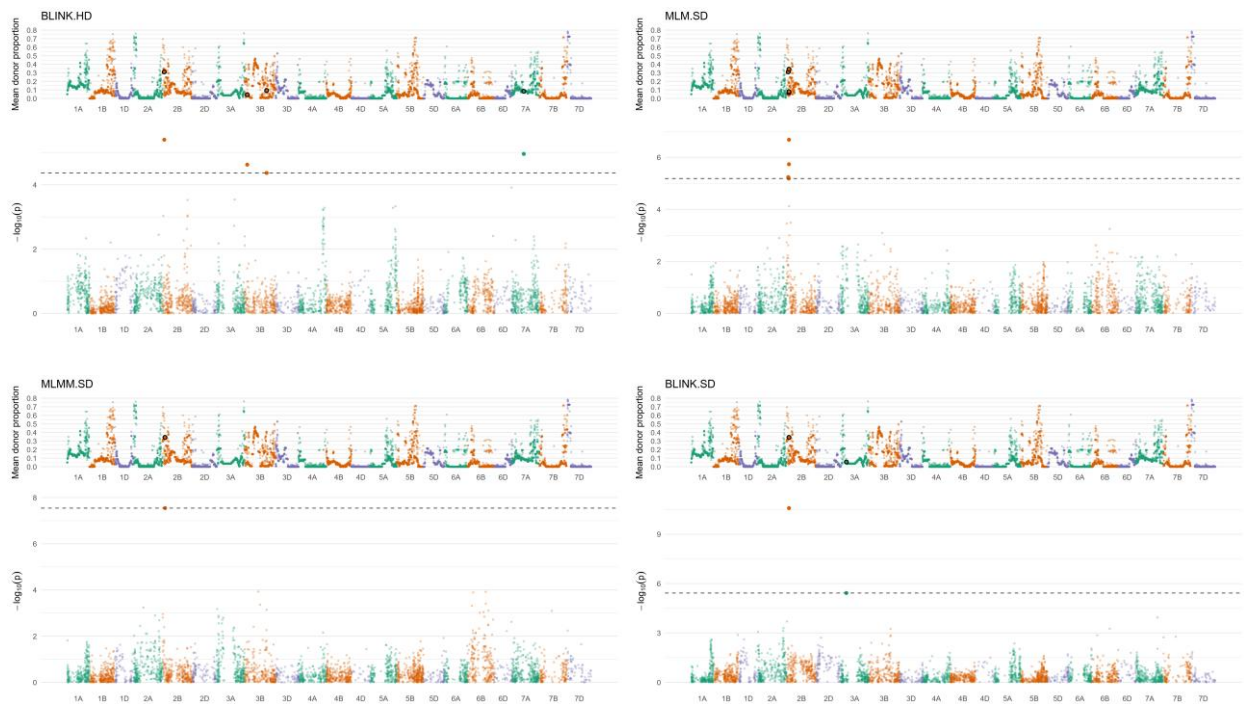


Figure 3.7. GWAS Manhattan plots for HD and SD BLUPs overlaid with mean donor proportions in TdGD Pop80

Significant associations across BLUP models for HD and SD are shown. In each panel, the upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

Two additional BLUP associations for HD were detected on chromosome 3B, located on the short arm (68.40 Mb) and long arm (589.54 Mb). The short arm region on 3B also shows associations with PH BLUPs, as well as HD BLUEs in 21THD and 22THD and PH BLUEs in

23RL (Table 3.5; Appendix Table B.7). The long arm association was additionally detected in PH BLUEs from 22RL. Although several other associations for maturity related traits were identified in the BLUP analysis, these are considered lower confidence due to limited support across models or environments.

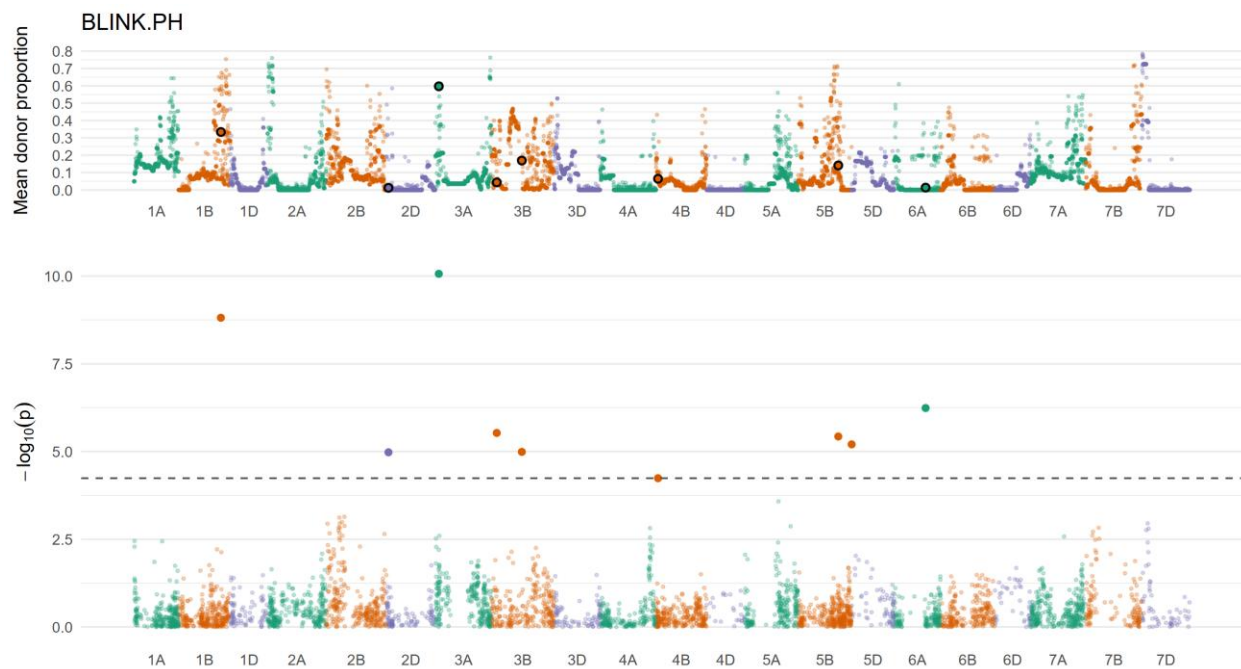


Figure 3.8. GWAS Manhattan plot for PH BLUPs overlaid with mean donor proportions in TdGD Pop80

Significant PH associations detected by BLINK are shown. The upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

For PH, several additional associations were identified by BLINK aside from the locus on chromosome 3B. Notably, the *Rht-B1* region on chromosome 4B was significant in the BLUPs and was also detected in BLUEs from 21THD, 22EL, and 22THD (Figure 3.8; Appendix Table B.7). Genotype class comparisons at these markers indicate that the donor allele is associated with increasing PH. This result is consistent with expectations, as the wild emmer donor lacks the *Rht-B1b* dwarfing allele present in the RP background.

Another notable BLUP signal for PH occurs on chromosome 3A at 47.08 Mb. Within this region, associations with additional maturity related traits were detected using BLUES (Appendix Table B.7), including GFP in 22EL (40.15 Mb), SD in 22RL (47.83 Mb), and HD in 23RL (47.83 Mb).

Table 3.5. GWAS results summary for agronomic performance using BLUPs

For each significant association, the chromosome (Chr), physical position in Mb (Pos), trait, model, MAF, and estimated marker effect are reported. Marker effects are relative to a standard normal distribution. Allelic information includes the allelic states of the RP and donor parents (Don), as well as the alleles associated with the lower (L) and higher (H) mean phenotypic values among genotype classes. Genotype_class columns report the number of WILs carrying the RP allele (WIL_RP) or the donor allele (WIL_Don) at the marker locus. Donor direction (DonDir) indicates whether the donor allele increases or decreases the trait value. Donor allele proportion within the 1 Mb genomic bin containing the marker (DonProp) represents the average proportion of donor alleles across WILs in that genomic region. NA indicates cases where parental alleles could not be unambiguously assigned.

Chr	Pos	Trait	Model	MAF	Effect	Allele		Genotype_class					
						RP	Don	L	H	WIL_RP	WIL_Don	DonDir	DonProp
7A	596.04	GFP	BLINK	0.23	-0.12	NA	NA	C	T	NA	NA	NA	0.27
2B	45.55	HD	BLINK	0.05	-0.43	NA	NA	A	C	NA	NA	NA	0.31
3B	68.40	HD	BLINK	0.25	-0.21	NA	NA	C	T	NA	NA	NA	0.04
3B	589.54	HD	BLINK	0.41	0.35	NA	NA	A	G	NA	NA	NA	0.09
7A	316.01	HD	BLINK	0.48	-0.14	NA	NA	A	C	NA	NA	NA	0.08
1B	563.45	PH	BLINK	0.10	-0.36	NA	NA	T	C	NA	NA	NA	0.33
2D	28.55	PH	BLINK	0.08	0.29	C	A	A	C	207	10	Decrease	0.01
3A	47.08	PH	BLINK	0.06	0.44	G	A	G	A	15	217	Increase	0.60
3B	68.40	PH	BLINK	0.25	-0.22	NA	NA	C	T	NA	NA	NA	0.04
3B	402.47	PH	BLINK	0.44	0.39	NA	NA	A	G	NA	NA	NA	0.17
4B	26.49	PH	BLINK	0.12	-0.21	NA	NA	C	T	NA	NA	NA	0.06
5B	534.32	PH	BLINK	0.07	0.29	T	G	G	T	209	11	Decrease	0.14
5B	712.60	PH	BLINK	0.09	0.31	NA	NA	G	A	NA	NA	NA	0.00
6A	418.54	PH	BLINK	0.07	-0.38	C	T	C	T	211	10	Increase	0.01
2B	65.23	SD	BLINK	0.05	-0.51	NA	NA	A	G	NA	NA	NA	0.34
3A	151.32	SD	BLINK	0.11	0.22	C	T	T	C	199	17	Decrease	0.06
2B	45.55	SD	MLM	0.05	-0.47	NA	NA	A	C	NA	NA	NA	0.31
2B	59.10	SD	MLM	0.06	-0.43	NA	NA	T	C	NA	NA	NA	0.07
2B	63.85	SD	MLM	0.05	-0.49	A	C	A	C	215	7	Increase	0.08
2B	65.23	SD	MLM	0.05	-0.56	NA	NA	A	G	NA	NA	NA	0.34
2B	65.23	SD	MLMM	0.05	-0.56	NA	NA	A	G	NA	NA	NA	0.34

Yield and kernel components

High confidence associations for GR were detected on the distal portion of the short arm of chromosome 2B. In the BLUPs, a marker at 18.26 Mb was significant across all models, with additional markers detected using MLM from 5.53-23.79 Mb (Table 3.6; Figure 3.9). Similar associations were observed in stage-specific models (Appendix Table B.10), first appearing during booting (GR_boot), and continuing through anthesis (GR_anth) and senescence (GR_milk). Environment-specific BLUES also supported this region; across all models, growth stages, and BLUE and BLUP analyses, the peak GR signal consistently occurs between 5.53-33.84 Mb on chromosome 2B (Table 3.6; Appendix Tables B.8, B.10-B.11).

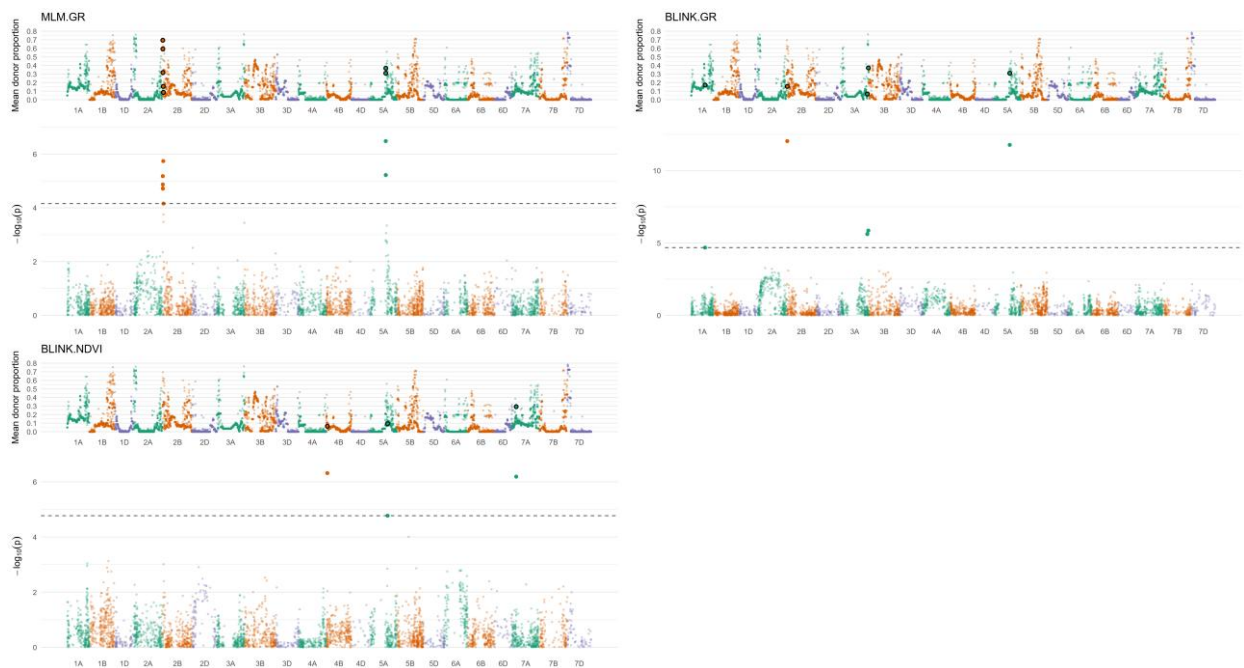


Figure 3.9. GWAS Manhattan plots for GR and NDVI BLUPs overlaid with mean donor proportions in TdGD Pop80

Significant associations for MLM and BLINK models for GR and NDVI are shown. In each panel, the upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

Within this interval, additional associations were detected for SD and PH in 22RL, and for HD in 22THD (Appendix Table B.7). The same region also overlaps associations for NDVI,

including NDVI_anth and NDVI_milk in BLUP analyses (Appendix Table B.10), and NDVI_boot, NDVI_anth, and NDVI_milk in BLUE analyses (Appendix Table B.11). The absence of a strong overall NDVI BLUP signal in this region likely reflects the lower R^2_{across} of NDVI relative to GR when averaged across growth stages (Table 3.2). Consistent with this interpretation, the growth stages at which GR and NDVI associations are detected generally correspond to stages with higher R^2_{across} (Appendix Table B.5).

A second major region affecting GR occurs on the long arm of chromosome 5A. Significant BLUP associations for GR, both averaged across growth stages and evaluated by stage, span 408.91-502.79 Mb (Table 3.6; Appendix Table B.10). Environment-specific BLUEs extend the associated interval further toward the distal end of the chromosome, with GR associations detected up to 671.27 Mb (Appendix Tables B.8, B.11). Across this broader region beginning around 403 Mb, BLUP associations were detected for GR, GR_joint, GR_boot, GR_anth, GR_milk, NDVI, and NDVI_till (Table 3.6; Figure 3.9; Appendix Table B.10). This region is further supported in the BLUEs, with associations detected for GR, NDVI, and multiple stage-specific GR and NDVI traits across 21THD, 22EL, and 22RL (Appendix Tables B.8, B.11).

In addition to canopy growth traits, this chromosome 5A region also contains associations for several kernel size traits. BLUP analyses identified significant associations for KL, KW, KA, and TKW (Table 3.6; Figure 3.10), with additional BLUE support across multiple environments for KA, KL, KW, TKW, and TW, and related agronomic traits (Appendix Tables B.7-B.8).

The direction of donor allele effects for GR differs between the two high confidence loci. In BLUP analyses, chromosome 2B was associated with decreased GR in the presence of the

donor allele, whereas the chromosome 5A association increased GR when the donor allele was present. This pattern was generally consistent across growth stages in BLUPs (Appendix Table B.10). Overall BLUE analyses averaged across growth stages supported the same pattern, although some stage-specific BLUE associations showed deviations; for example, GR_till associations in 22EL and 22RL showed negative donor effects on chromosome 5A (Appendix Table B.11). Interestingly, NDVI associations in these regions showed the opposite pattern, with positive donor effects on chromosome 2B and negative effects on chromosome 5A. Table 3.7 summarizes donor allele directions for drone-derived traits across BLUP and BLUE analyses.

Table 3.6. GWAS results summary for yield and kernel components using BLUPs

For each significant association, the chromosome (Chr), physical position in Mb (Pos), trait, model, MAF, and estimated marker effect are reported. Marker effects are relative to a standard normal distribution. Allelic information includes the allelic states of the RP and donor parents (Don), as well as the alleles associated with the lower (L) and higher (H) mean phenotypic values among genotype classes. Genotype_class columns report the number of WILs carrying the RP allele (WIL_RP) or the donor allele (WIL_Don) at the marker locus. Donor direction (DonDir) indicates whether the donor allele increases or decreases the trait value. Donor allele proportion within the 1 Mb genomic bin containing the marker (DonProp) represents the average proportion of donor alleles across WILs in that genomic region. NA indicates cases where parental alleles could not be unambiguously assigned.

Chr	Pos	Trait	Model	MAF	Effect	Allele		Genotype_class				DonDir	DonProp
						RP	Don	L	H	WIL_RP	WIL_Don		
1A	372.29	GR	BLINK	0.18	0.10	NA	NA	C	T	NA	NA	NA	0.17
2B	18.26	GR	BLINK	0.06	0.26	C	T	T	C	218	14	Decrease	0.16
3A	713.63	GR	BLINK	0.12	0.14	A	G	G	A	199	21	Decrease	0.07
3A	741.24	GR	BLINK	0.36	0.09	NA	NA	A	G	NA	NA	NA	0.37
5A	412.53	GR	BLINK	0.42	-0.13	T	C	T	C	132	97	Increase	0.31
2B	5.53	GR	MLM	0.12	0.20	NA	NA	C	A	NA	NA	NA	0.69
2B	5.67	GR	MLM	0.13	0.17	NA	NA	A	G	NA	NA	NA	0.69
2B	6.23	GR	MLM	0.14	0.16	NA	NA	T	G	NA	NA	NA	0.59
2B	9.73	GR	MLM	0.14	0.17	NA	NA	G	A	NA	NA	NA	0.32
2B	18.26	GR	MLM	0.06	0.25	C	T	T	C	218	14	Decrease	0.16
2B	23.79	GR	MLM	0.08	0.23	A	C	C	A	207	11	Decrease	0.08
5A	412.53	GR	MLM	0.42	-0.15	T	C	T	C	132	97	Increase	0.31
5A	414.01	GR	MLM	0.40	-0.13	A	G	A	G	140	92	Increase	0.37
2B	18.26	GR	MLMM	0.06	0.23	C	T	T	C	218	14	Decrease	0.16
5A	412.53	GR	MLMM	0.42	-0.14	T	C	T	C	132	97	Increase	0.31
1A	34.16	GY	BLINK	0.12	0.23	NA	NA	A	G	NA	NA	NA	0.16
2B	451.89	GY	BLINK	0.11	0.22	NA	NA	G	A	NA	NA	NA	0.06

3A	31.66	GY	BLINK	0.10	0.31	T	C	C	T	195	11	Decrease	0.15
4B	34.72	GY	BLINK	0.06	0.34	C	T	T	C	209	6	Decrease	0.05
6B	171.06	GY	BLINK	0.28	0.15	G	A	A	G	164	63	Decrease	0.31
6B	717.92	GY	BLINK	0.06	0.32	T	G	G	T	209	7	Decrease	0.01
7B	622.25	GY	BLINK	0.16	0.17	NA	NA	A	C	NA	NA	NA	0.04
2A	713.93	KA	BLINK	0.13	-0.42	G	A	G	A	203	29	Increase	0.20
3A	164.61	KA	BLINK	0.11	-0.24	A	G	A	G	191	12	Increase	0.05
3D	433.50	KA	BLINK	0.15	0.22	G	T	T	G	191	27	Decrease	0.02
4B	601.75	KA	BLINK	0.10	-0.64	A	G	A	G	200	14	Increase	0.04
5A	595.55	KA	BLINK	0.06	-0.41	NA	NA	G	A	NA	NA	NA	0.16
5A	609.69	KA	BLINK	0.48	-0.17	NA	NA	T	C	NA	NA	NA	0.19
6B	686.80	KA	BLINK	0.30	0.20	NA	NA	C	T	NA	NA	NA	0.00
7B	141.82	KA	BLINK	0.08	-0.27	A	G	A	G	212	16	Increase	0.09
7D	124.50	KA	BLINK	0.13	-0.22	T	G	T	G	193	22	Increase	0.01
2A	83.01	KL	BLINK	0.08	-0.41	G	A	G	A	206	11	Increase	0.07
4B	28.15	KL	BLINK	0.08	-0.53	A	C	A	C	206	12	Increase	0.06
5A	437.21	KL	BLINK	0.05	-0.56	NA	NA	T	C	NA	NA	NA	0.42
5A	442.71	KL	BLINK	0.44	0.16	G	A	A	G	130	101	Decrease	0.29
6B	481.66	KL	BLINK	0.13	-0.29	NA	NA	G	T	NA	NA	NA	0.00
7D	127.14	KL	BLINK	0.11	-0.33	NA	NA	A	G	NA	NA	NA	0.00
1A	8.21	KW	BLINK	0.11	-0.26	NA	NA	C	T	NA	NA	NA	0.11
2A	713.93	KW	BLINK	0.13	-0.22	G	A	G	A	203	29	Increase	0.20
2B	543.71	KW	BLINK	0.11	-0.31	NA	NA	T	G	NA	NA	NA	0.12
3D	200.38	KW	BLINK	0.14	0.25	A	C	C	A	199	33	Decrease	0.09
4A	695.26	KW	BLINK	0.09	-0.35	NA	NA	A	C	NA	NA	NA	0.00
4B	633.62	KW	BLINK	0.08	-0.41	G	A	G	A	207	10	Increase	0.03
5A	1.20	KW	BLINK	0.11	0.33	A	G	G	A	205	23	Decrease	0.02
5A	609.69	KW	BLINK	0.48	-0.14	NA	NA	T	C	NA	NA	NA	0.19
5B	671.71	KW	BLINK	0.19	-0.19	NA	NA	C	T	NA	NA	NA	0.00
5D	191.34	KW	BLINK	0.14	-0.24	A	C	A	C	199	31	Increase	0.16
6B	690.97	KW	BLINK	0.35	0.21	NA	NA	T	C	NA	NA	NA	0.00
6D	3.03	KW	BLINK	0.22	-0.20	NA	NA	G	A	NA	NA	NA	0.00
7A	527.15	KW	BLINK	0.10	-0.30	T	C	T	C	208	24	Increase	0.09
6B	690.97	KW	MLMM	0.35	0.25	NA	NA	T	C	NA	NA	NA	0.00
4B	28.23	NDVI	BLINK	0.09	0.14	G	T	T	G	208	16	Decrease	0.06
5A	463.07	NDVI	BLINK	0.09	0.12	NA	NA	C	T	NA	NA	NA	0.09
7A	112.98	NDVI	BLINK	0.08	0.18	NA	NA	C	G	NA	NA	NA	0.29
4B	629.95	TC	MLMM	0.06	0.26	NA	NA	A	G	NA	NA	NA	0.02
2A	713.93	TKW	BLINK	0.13	-0.27	G	A	G	A	203	29	Increase	0.20
3D	30.24	TKW	BLINK	0.16	0.21	A	G	G	A	189	33	Decrease	0.24
4B	633.62	TKW	BLINK	0.08	-0.47	G	A	G	A	207	10	Increase	0.03
5A	609.69	TKW	BLINK	0.48	-0.20	NA	NA	T	C	NA	NA	NA	0.19
6B	686.80	TKW	BLINK	0.30	0.19	NA	NA	C	T	NA	NA	NA	0.00

7A	609.55	TKW	BLINK	0.08	-0.26	A	G	G	A	17	214	Decrease	0.35
2A	713.93	TW	BLINK	0.13	0.30	G	A	A	G	203	29	Decrease	0.20
4A	341.07	TW	BLINK	0.11	0.37	T	C	C	T	199	16	Decrease	0.04
4A	695.26	TW	BLINK	0.09	0.55	NA	NA	G	A	NA	NA	NA	0.00
6B	531.53	TW	BLINK	0.06	0.61	A	G	G	A	214	11	Decrease	0.01
2A	713.93	TW	MLM	0.13	0.36	G	A	A	G	203	29	Decrease	0.20
6B	185.67	TW	MLM	0.07	0.65	T	C	C	T	211	10	Decrease	0.08
6B	218.65	TW	MLM	0.06	0.68	A	G	G	A	211	9	Decrease	0.05
6B	333.51	TW	MLM	0.07	0.69	A	G	G	A	209	10	Decrease	0.00
6B	531.53	TW	MLM	0.06	0.70	A	G	G	A	214	11	Decrease	0.01
2A	713.93	TW	MLMM	0.13	0.37	G	A	A	G	203	29	Decrease	0.20
4A	116.47	TW	MLMM	0.07	1.25	NA	NA	C	A	NA	NA	NA	0.08
4A	136.86	TW	MLMM	0.06	-0.90	A	C	C	A	218	14	Decrease	0.07
6B	333.51	TW	MLMM	0.07	0.68	A	G	G	A	209	10	Decrease	0.00

Table 3.7. Summary of donor allele direction for GR and NDVI GWAS associations

For each significant association detected for GR and NDVI, the direction of the donor allele effect is summarized across BLUP analyses and within-environment BLUE analyses. Results are shown for associations averaged across all growth stages (Overall) and for individual developmental stages. Donor direction is indicated as + when the donor allele increases the trait value and - when the donor allele decreases the trait value. Only associations on either the short arm of chromosome 2B or the long arm of chromosome 5A, and for which donor direction could be unambiguously determined, are included.

		BLUP		21THD		22EL		22RL	
		Chr2B	Chr5A	Chr2B	Chr5A	Chr2B	Chr5A	Chr2B	Chr5A
GR	Overall	-	+	-	+		+	-	+
	till						-		-
	joint		+		+				
	boot	-	+		+	-	+	-	
	anth	-	+	-	+/-			-	
	milk	-	+		+		+	-	
NDVI	Overall				-				
	till		-		-		-		
	joint								
	boot					+			
	anth	+				+			
	milk	+				+		+	

Additional associations for NDVI were detected outside these major GR regions. An NDVI BLUP association on chromosome 4B occurs in the same genomic region as the *Rht-B1*

locus and likely reflects variation in NDVI associated with PH rather than canopy vigor independently (Table 3.6; Figure 3.9; Appendix Tables B.7-B.8, B.10-B.11). Another NDVI association on chromosome 7A was detected by BLINK in BLUP analyses and was strongly supported by 21THD BLUES. These associations span 112.98-125.49 Mb on the short arm of chromosome 7A and include signals for overall NDVI as well as stage-specific NDVI traits across multiple growth stages.

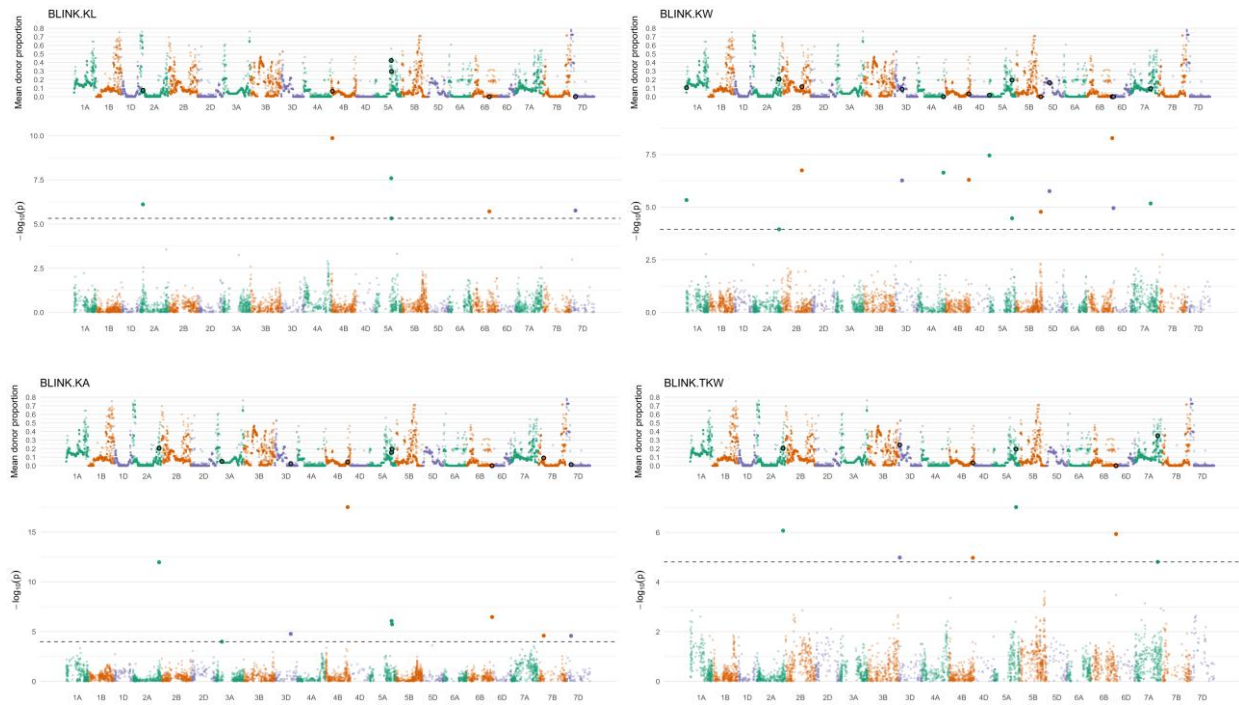


Figure 3.10. GWAS Manhattan plots for kernel size BLUPs (KL, KW, KA, TKW) overlaid with mean donor proportions in TdGD Pop80

Significant associations for BLINK models for kernel size traits are shown. In each panel, the upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

Beyond canopy traits, several genomic regions were associated with kernel size and yield-related traits. A high confidence locus on the long arm of chromosome 2A (713.93 Mb) was associated with KA, KW, TKW, and TW, with the donor allele increasing KA, KW, and TKW while decreasing TW (Table 3.6; Figures 3.10-3.11). This region also showed associations

for KL in some environments (Appendix Table B.8). Additional TW associations were detected on chromosome 4A at 116.47, 136.86, and 341.07 Mb, while a locus near 695.26 Mb on chromosome 4A was associated with both TW and KW BLUPs, although donor direction could not be clearly determined.

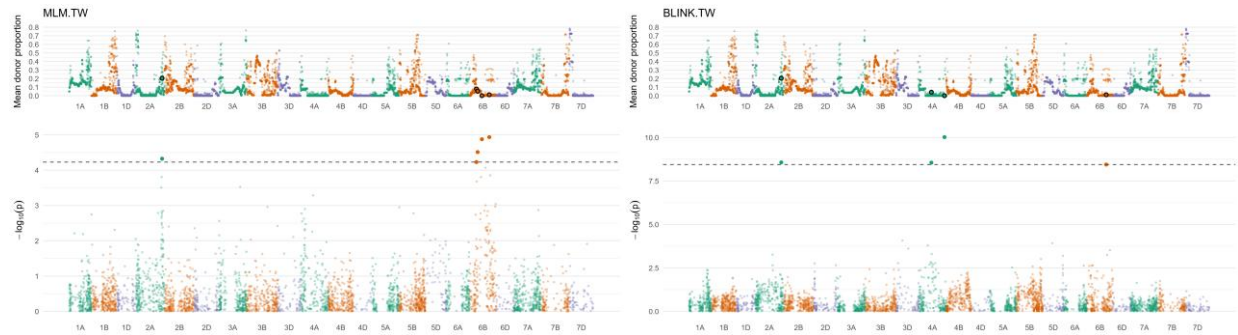


Figure 3.11. GWAS Manhattan plots for TW BLUPs overlaid with mean donor proportion in TdGD Pop80

Significant TW associations for MLM and BLINK models are shown. In each panel, the upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

In addition to NDVI and PH signals near the *Rht-B1* locus, the short arm of chromosome 4B contains additional BLUP associations for yield and kernel traits, including for KL, with donor allele direction being positive (Table 3.6; Figure 3.10). Additionally, an association for GY was detected in the BLUPs at 34.72 Mb (Figure 3.12). The presence of a GY association near *Rht-B1* is consistent with known environment-dependent relationships between PH and GY, particularly under varying drought conditions (Bacher et al., 2021; Ihsan et al., 2016; Liu et al., 2017; Peleg et al., 2005; Wang et al., 2017). This GY signal, however, was poorly supported in the BLUE analyses, only reaching significance in 21THD (Appendix Table B.8). Further along chromosome 4B, an approximately 30 Mb interval between 601.75-633.62 Mb contains BLUP associations for KA, KW, TKW, and TC (Table 3.6; Figures 3.10, 3.13). These associations are

supported by BLUEs across multiple environments and additional traits, including KL and GY (Appendix Table B.8).



Figure 3.12. GWAS Manhattan plot for GY BLUPs overlaid with mean donor proportion in TdGD Pop80

Significant GY associations detected by BLINK are shown. The upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

Significant BLUP associations for yield and kernel components were also distributed along chromosome 6B. The short arm contains associations for GY and TW, while the long arm contains associations for KL, KW, KA, TW, TKW, and GY (Table 3.6; Figures 3.10-3.12). BLUE analyses support these findings and reveal additional associations for stage-specific GR and NDVI traits, as well as SD and GFP (Appendix Tables B.7-B.8, B.11).

The long arm of chromosome 7A contains associations for KW, TKW, and GFP BLUPs, as well as KL, KW, TW, HD, and GFP BLUEs (Tables 3.5-3.6; Figure 3.10; Appendix Tables B.7-B.8). Chromosome 7D also contains an association for KL and KA between 124.50-127.14

Mb, supported by both BLUPs and KL BLUEs across all environments (Table 3.6; Figure 3.10; Appendix Table B.8).



Figure 3.13. GWAS Manhattan plot for TC BLUPs overlaid with mean donor proportion in TdGD Pop80

The significant TC association detected by MLMM is shown. The upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

Although additional associations for yield and kernel traits were detected, the loci described above represent the most consistent signals across traits, models, and environments. In general, donor alleles tended to increase kernel size traits such as KA, KL, KW, and TKW, albeit with some exceptions, whereas donor alleles consistently decrease GY and TW (Table 3.6). This pattern is consistent with the yield disadvantage observed in the WILs (Figure 3.4).

Functional quality

There is overlap between genomic regions associated with functional quality traits and those identified for kernel traits and yield components. This overlap is unsurprising given the influence of kernel size and grain filling dynamics on both the quantity and functional properties

of wheat storage proteins that ultimately determine quality (Figure 3.6; Distelfeld et al., 2007; Singh et al., 2012; Yang et al., 2023). For example, the chromosome 2A locus at 713.93 Mb – associated with KA, KW, TKW, and TW BLUPs – is located near a GP association at 700.54 Mb detected in both BLUPs and BLUEs from 21THD and 22RL (Tables 3.6, 3.8; Figures 3.10-3.11, 3.14; Appendix Table B.9).

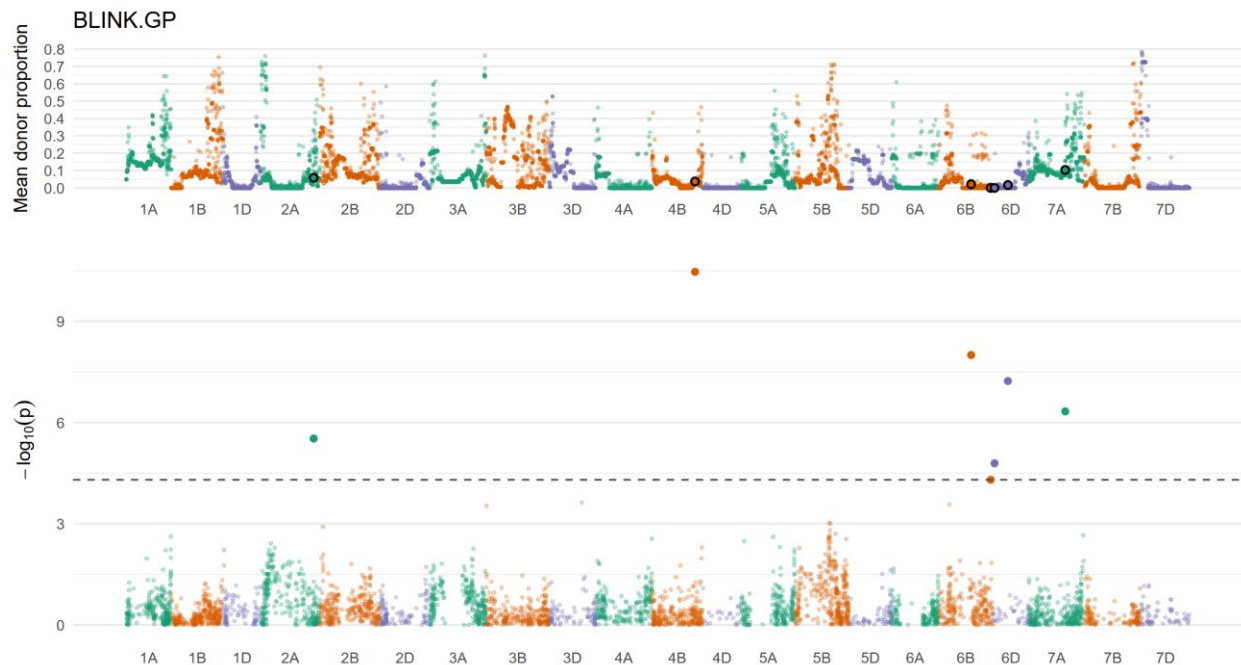


Figure 3.14. GWAS Manhattan plot for GP BLUPs overlaid with mean donor proportion in TdGD Pop80

Significant GP associations detected by BLINK are shown. The upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

Additional colocated associations occur across chromosome 6B, which was noted as harboring signals for multiple maturity, yield, and kernel component traits; GP associations were detected along this chromosome as well (Tables 3.6, 3.8; Figure 3.14; Appendix Tables B.7-B.9). A marker at 431.80 Mb was significantly associated with GP BLUPs and GP BLUEs in 21THD, 22EL, and 22RL. The same marker was also associated with GFP, GR_milk, NDVI_milk, SD, and TW in 21THD BLUEs (Appendix Tables B.7-B.9, B.11). A second GP BLUP association on

the long arm of chromosome 6B occurred at 690.97 Mb, corresponding to the same marker associated with KW BLUPs and KW and TKW BLUEs (Tables 3.6, 3.8; Figures 3.10, 3.14; Appendix Table B.8). Between these loci, an additional signal at 595.29 Mb was detected for SRC BLUPs and SRC BLUEs in 23RL (Table 3.8; Figure 3.15; Appendix Table B.9). Notably, the chromosome 6B regions associated with GP and SRC coincide with segments exhibiting low donor allele proportions (0-0.02), although the donor allele at these loci was associated with increased trait values (Table 3.8).

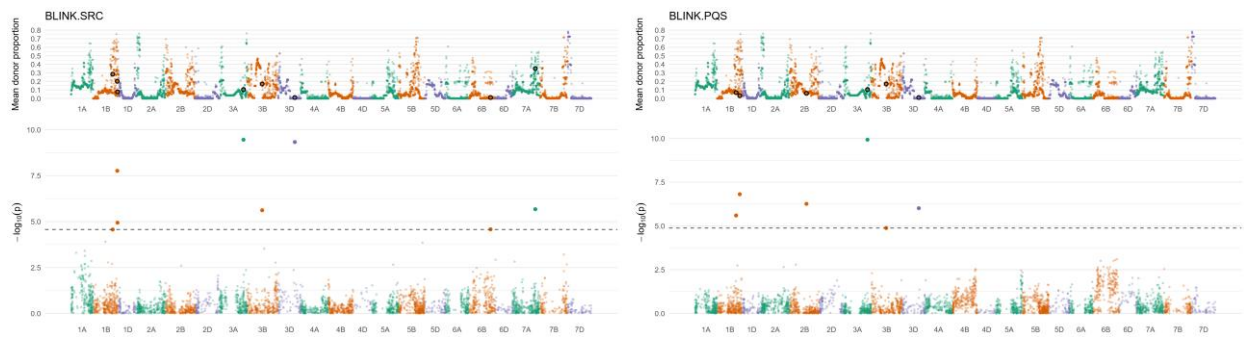


Figure 3.15. GWAS Manhattan plots for SRC and PQS BLUPs overlaid with mean donor proportions in TdGD Pop80

Significant associations for BLINK models for SRC and PQS are shown. In each panel, the upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

The short arm of chromosome 6B contains the *Gpc-B1* allele derived from wild emmer, which regulates senescence-driven nutrient remobilization (Distelfeld et al., 2007; Uauy et al., 2006). Although this locus was not significant in the BLUP analysis, GP BLUEs from 22THD revealed an association at 142.23 Mb (Appendix Table B.9). Additional maturity-related associations were detected nearby toward the distal end of the chromosome arm, including SD and GFP signals in 22EL and SD in 23THD (Appendix Table B.7).

Additional GP associations were detected on chromosome 6D. Two markers significant in the BLUPs were identified at 20.76 and 197.31 Mb, with other GP BLUE associations

occurring at 68.33 and 169.64 Mb (Table 3.8; Figure 3.14; Appendix Table B.9). The association at 20.76 Mb lies near the *Gli-2D* locus, which contains α -gliadin alleles involved in wheat prolamins expression (Pourmohammadi et al., 2023; Rasheed et al., 2014). The association at 197.31 Mb showed a positive donor allele effect for GP, consistent with the BLUE associations detected in this region (Table 3.8; Appendix Table B.9).

Associations were also detected for SRC and PQS, traits reflecting gluten quantity and functional quality (Blanco et al., 1998; Seabourn et al., 2012). On chromosome 1B, BLUP associations for PQS and SRC occurred between 515.05-662.16 Mb, overlapping with a PH BLUP association at 563.45 Mb (Tables 3.5, 3.8; Figures 3.8, 3.15). BLUE analyses further supported this region, with associations spanning 334.24-662.16 Mb for additional traits including GP, GPD, NDVI, and SD (Appendix Tables B.7-B.9, B.10-B.11). This interval encompasses the *Glu-1B* locus encoding high molecular weight glutenin subunits, which exhibits substantial allelic diversity relative to homoeologous loci on chromosomes 1A and 1D (Rasheed et al., 2014; Pourmohammadi et al., 2023; Y. Li et al., 2020). No functional quality associations were detected near the homoeologous locus on chromosome 1A, and the *Glu-1Ay* marker present on the SNP array showed no segregation within Pop80. Patterns of donor allele proportion support recombination across the chromosome 1B interval, with elevated donor proportions at 492.79 Mb (0.37), 532.17-568.53 Mb (0.28-0.33), and 623.67-658.53 Mb (0.20-0.62), while donor proportion elsewhere in the region remained low (Tables 3.5, 3.8; Appendix Tables B.7, B.9). Where donor effects could be unambiguously determined, the donor allele on chromosome 1B was associated with reduced functional quality.

Another notable association occurs on the long arm of chromosome 2B at 451.89 Mb, where BLUP associations were detected for both PQS and GY, with additional BLUE

associations for KA, KW, and TKW in 23RL and TKW in 23THD (Tables 3.6, 3.8; Figures 3.12, 3.15; Appendix Table B.8).

Table 3.8. GWAS results summary for functional quality traits using BLUPs

For each significant association, the chromosome (Chr), physical position in Mb (Pos), trait, model, MAF, and estimated marker effect are reported. Marker effects are relative to a standard normal distribution. Allelic information includes the allelic states of the RP and donor parents (Don), as well as the alleles associated with the lower (L) and higher (H) mean phenotypic values among genotype classes. Genotype_class columns report the number of WILs carrying the RP allele (WIL_RP) or the donor allele (WIL_Don) at the marker locus. Donor direction (DonDir) indicates whether the donor allele increases or decreases the trait value. Donor allele proportion within the 1 Mb genomic bin containing the marker (DonProp) represents the average proportion of donor alleles across WILs in that genomic region. NA indicates cases where parental alleles could not be unambiguously assigned.

Chr	Pos	Trait	Model	MAF	Effect	Allele		Genotype_class				DonDir	DonProp
						RP	Don	L	H	WIL_RP	WIL_Don		
2A	700.54	GP	BLINK	0.10	-0.23	NA	NA	T	C	NA	NA	NA	0.06
4B	570.90	GP	BLINK	0.07	-0.40	C	T	C	T	212	12	Increase	0.04
6B	431.80	GP	BLINK	0.06	-0.46	T	C	T	C	208	3	Increase	0.02
6B	690.97	GP	BLINK	0.35	0.11	NA	NA	T	C	NA	NA	NA	0.00
6D	20.76	GP	BLINK	0.11	-0.18	NA	NA	G	A	NA	NA	NA	0.00
6D	197.31	GP	BLINK	0.11	-0.27	A	C	A	C	197	18	Increase	0.02
7A	481.69	GP	BLINK	0.11	-0.27	T	C	T	C	204	21	Increase	0.10
7A	109.48	GPD	BLINK	0.13	-0.31	G	A	G	A	201	27	Increase	0.19
1B	515.05	PQS	BLINK	0.11	0.46	T	C	C	T	201	19	Decrease	0.07
1B	614.50	PQS	BLINK	0.06	0.34	NA	NA	T	C	NA	NA	NA	0.03
2B	451.89	PQS	BLINK	0.11	0.22	NA	NA	G	A	NA	NA	NA	0.06
3A	647.36	PQS	BLINK	0.08	-0.38	C	T	C	T	213	19	Increase	0.10
3B	402.47	PQS	BLINK	0.44	-0.39	NA	NA	G	A	NA	NA	NA	0.17
3D	454.67	PQS	BLINK	0.10	-0.27	A	G	A	G	201	17	Increase	0.01
3A	647.36	PQS	MLMM	0.08	-0.40	C	T	C	T	213	19	Increase	0.10
3D	454.67	PQS	MLMM	0.10	-0.40	A	G	A	G	201	17	Increase	0.01
1B	532.17	SRC	BLINK	0.19	0.19	NA	NA	G	A	NA	NA	NA	0.28
1B	655.76	SRC	BLINK	0.42	0.18	NA	NA	C	T	NA	NA	NA	0.20
1B	662.16	SRC	BLINK	0.05	0.29	NA	NA	T	C	NA	NA	NA	0.08
3A	647.36	SRC	BLINK	0.08	-0.37	C	T	C	T	213	19	Increase	0.10
3B	402.47	SRC	BLINK	0.44	-0.45	NA	NA	G	A	NA	NA	NA	0.17
3D	454.67	SRC	BLINK	0.10	-0.35	A	G	A	G	201	17	Increase	0.01
6B	595.29	SRC	BLINK	0.06	-0.28	G	A	G	A	216	11	Increase	0.01
7A	609.55	SRC	BLINK	0.08	-0.25	A	G	G	A	17	214	Decrease	0.35

Chromosomal regions on all group 3 chromosomes showed associations for PQS and SRC. BLUP signals occurred at 647.36 Mb on chromosome 3A, 402.47 Mb on chromosome 3B, and 454.67 Mb on chromosome 3D (Table 3.8; Figure 3.15). The chromosome 3A and 3D associations were strongly supported by BLUE analyses from 2021 and 2022 (Appendix Table B.9). The chromosome 3B marker was also associated with PH BLUPs, although the direction of donor allele effect was ambiguous (Table 3.5). In contrast, the donor allele at the chromosome 3A and 3D loci was associated with increased functional quality (Table 3.8).

Discussion

The results presented here represent a large-scale effort to characterize phenotypic performance and variability in WILs developed using an amphiploid bridge to facilitate introgression from wild emmer into an adapted hexaploid background. By integrating a broad suite of traits across multiple environments with analyses of marker-trait associations, this study enables assessment of trait relationships and identification of genomic regions contributing to trait variation.

This work is inherently exploratory in nature, with the primary objective of identifying patterns and generating hypotheses for follow-up work, rather than testing predefined predictions. Consequently, the strength of causal inference is limited, and the interpretation of significant marker-trait associations and phenotypic relationships must be made with appropriate caution. The large number of traits evaluated, combined with genome-wide marker density, increases the likelihood of spurious associations, and observed signals are contingent upon the distribution of allelic variation within the population, the consistency of genotypic performance across environments, and the presence of GxE interactions. Trait expression is further influenced by variation in drought timing and intensity across environments.

Despite these limitations, the breadth of phenotypic characterization across environments provides a valuable framework for evaluating overall performance and stability. In a breeding context, consistent trait expression across environments is critical, as target environments are inherently variable and imperfectly predictable. Traits or indices derived from across environment estimates, such as BLUPs and composite selection indices, are therefore particularly relevant for identifying germplasm with robust performance. Within this framework, the comparison of WILs to the recurrent parent remains essential for identifying favorable phenotypic deviations attributable to wild introgressions within an elite genetic background.

An additional consideration in this population is the balance between maintaining sufficient donor introgression to capture novel allelic variation while avoiding excessive linkage drag that may compromise adaptation. The relatively low frequency of donor alleles across the genome reflects this balance, but also influences the statistical power to detect marker-trait associations, as detectable effects depend on allele frequency within the population.

A key limitation of this study is the variability in drought stress across environments, both in timing and severity, which was not explicitly quantified beyond environmental classification. More detailed characterization of water availability and stress dynamics would improve interpretation of environment-specific responses. However, the range of environmental conditions represented here provides an opportunity to evaluate trait behavior across a gradient of drought stress scenarios, offering insight into stability of trait expression.

While more complex analytical approaches could be applied to further interrogate this dataset, a parsimonious modeling framework was prioritized to facilitate interpretability, particularly given the exploratory objectives of the study. The extensive phenotypic and genomic

data generated nonetheless provide a strong foundation for more targeted, hypothesis-driven analyses in future research.

With this context in mind, the remaining discussion focuses on three primary areas: (1) the overall feasibility of using an amphiploid bridge to introgress useful allelic variation from wild emmer while maintaining adaptive performance, (2) the correspondence between high confidence GWAS signals identified here and previously reported results, and (3) the implications of these findings for future breeding applications and targeted germplasm improvement.

Feasibility of maintaining adaptive performance in bridge-mediated wheat introgression breeding germplasm

Development of BC₂-derived WILs using an amphiploid bridge was broadly effective for incorporating wild emmer allelic diversity while retaining a degree of adaptive performance. Across traits, WILs exhibited increased phenotypic variance relative to the RP, enabling detection of relationships between introgressed alleles and trait expression. Improvements over the RP were most consistently observed for GP and GPD, with many WILs also exceeding the RP for individual yield components and biomass-related traits. However, these gains were accompanied by an overall reduction in GY, indicating a persistent yield disadvantage despite recovery of the RP background by the BC₂ generation.

This yield penalty is consistent with expectations for introgressed germplasm, where recombination disrupts favorable linkage blocks and co-adapted allele complexes present in elite backgrounds (Bacher et al., 2021; Merchunk-Ovnat et al., 2016b; Nyine et al., 2020; Wang et al., 2017). Yield is inherently governed by tightly coordinated interactions among developmental timing, biomass accumulation, and kernel filling, all of which are highly sensitive to

environmental conditions (Golan et al., 2015; Merchuk-Ovnat et al., 2016b; Murchie et al., 2023; Reynolds et al., 2017; Slafer et al., 2023). In this study, drought timing and intensity strongly modulated these relationships, influencing the relative contribution of source (biomass) and sink (kernel) components to final yield. Early season drought limited vegetative growth and tillering, constraining biomass accumulation, whereas late season stress reduced kernel size and density. Under more favorable moisture conditions, WILs frequently expressed increased tillering capacity and biomass production, though often at the expense of kernel size, highlighting the difficulty of simultaneously optimizing source and sink processes.

Kernel components further illustrated the conditional nature of these source-sink relationships. Although WILs generally had lower TKW, KA, KL, and KW than the performance checks, the RP also produced relatively small kernels and occurred near the center of the WIL distributions. Consequently, differences between the WILs and unrelated elite checks partly reflected the kernel phenotype of the recurrent background and do not indicate that wild emmer introgression uniformly reduced kernel size. This interpretation was reinforced by several GWAS associations for which the wild emmer allele increased kernel weight or dimensions.

The relationship between kernel traits and GY was similarly nuanced. Test weight represents grain bulk density and is influenced by kernel filling, density, shape, uniformity, and packing rather than individual kernel size alone (Petingco et al., 2018). The positive association between GY and TW may therefore reflect shared sensitivity to drought, whereby severe stress reduced both harvested grain mass and grain soundness, rather than demonstrating a direct yield advantage from kernel density. Grain number is generally the more environmentally responsive, coarse regulator of wheat yield, whereas kernel weight provides finer adjustment and can compensate for changes in kernel number (Sadras & Slafer, 2011; Slafer et al., 2023). Because

kernel number was not measured in Pop80, it cannot be determined whether wild emmer alleles that increased kernel size were accompanied by fewer kernels or whether their potential effects were constrained by insufficient assimilate supply during grain filling. These alleles should therefore not be considered inherently unfavorable for yield; rather, their breeding value likely depends on combining increased kernel size with sufficient kernel number and source capacity.

A central factor underlying these dynamics was developmental timing. Maturity traits played a dominant role in structuring trait expression across environments, as the timing and severity of drought relative to crop development determined whether yield limitations arose primarily through reduced biomass or impaired grain filling. Many WILs exhibited extended GFP relative to the RP, which supported improved kernel filling in some environments but did not consistently offset reductions in yield. Consequently, the results do not support directional selection for earlier HD, later SD, or longer GFP as individual traits. Rather, developmental timing should be evaluated jointly with yield across environments because the value of a given phenological combination depended on when drought occurred. Wild emmer introgression therefore provides potentially useful variation in developmental responses to water limitation (Fatiukha et al., 2021; Merchuk-Ovnat et al., 2016a; Nevo & Chen, 2010; Peleg et al., 2009b; Suneja et al., 2019), but no single phenological ideotype emerged as consistently favorable. These results emphasize that successful introgression for yield-related traits requires not only favorable alleles for biomass or kernel characteristics, but also appropriate synchronization of developmental timing with environmental conditions.

Taken together, these results illustrate the inherent trade-offs associated with introgressing wild alleles into elite germplasm. While introgression can enhance trait variation and enable improvements in specific traits or trait groups, overall performance is ultimately

determined by the interaction of many small effect loci governing developmental timing, biomass accumulation, resource partitioning, and grain composition. Elite germplasm represents the outcome of long-term selection for favorable combinations of these loci, particularly under regionally relevant stress conditions such as drought. Introducing new diversity disrupts these combinations, leading to increased phenotypic variability and, in many cases, reduced performance.

From a breeding perspective, the value of this germplasm lies in its ability to generate new allele combinations that expand the available genetic space for selection. Rather than expecting immediate improvement in elite performance, introgression populations such as these serve as intermediate material for subsequent recombination and selection. Emphasis on stability across environments – captured through BLUP-based selection – is particularly important given the unpredictability of future production environments. The transgressive phenotypic variance observed in WILs, especially under stress, enhances the effectiveness of selection by providing greater contrast among genotypes.

Overall, these findings support the feasibility of using an amphiploid bridge to introgress wild emmer diversity into hexaploid wheat, even for highly quantitative traits. While yield penalties and trait trade-offs are expected, the identification of high performing WILs and the expansion of trait variability demonstrate that adaptive performance can be maintained and, in some cases, improved through careful selection. The challenge moving forward lies in efficiently identifying and recombining favorable allele combinations to balance productivity, stability, and quality in breeding programs.

Candidate loci for marker-trait associations identified under an exploratory analytical framework

One of the central goals of this work was to determine whether statistically significant marker-trait associations could be detected in a population where donor alleles were expected to occur at relatively low frequency. Among BLUP associations, the mean donor allele proportion in the 1 Mb bin overlapping significant markers was 0.20 for the A subgenome, 0.18 for the B subgenome, and 0.04 for the D subgenome. Comparison to genome-wide mean donor proportions, specifically on the A and B subgenomes (0.09 and 0.10, respectively), suggests that genomic regions contributing statistically detectable trait variation tend to retain donor introgression at frequencies higher than genome-wide donor averages. Thus, although donor alleles were generally uncommon, introgressed segments nonetheless contributed enough segregating variation to enable detection of marker-trait associations.

A useful feature of these results is that several high confidence associations occurred near loci already known to influence the same or related traits. This does not prove that the significant markers tag the causal polymorphisms themselves, but it does provide biological context for the plausibility of the GWAS signals. For example, allelic variation around *Rht-B1* was associated not only with PH, as expected, but also with NDVI, GY, and KL. Such overlap is biologically credible given the broad downstream effects of height on canopy architecture and biomass partitioning, and grain development (Bacher et al., 2021; Ihsan et al., 2016; Liu et al., 2017; Peleg et al., 2005; Wang et al., 2017). Similarly, the high confidence association around *Ppd-B1* for HD and SD is consistent with the established role of this locus in photoperiod sensitivity and developmental timing (DeWitt et al., 2021).

An interesting result from chromosome 2B was the distal region associated with GR BLUPs across the season and from booting through milk, along with NDVI BLUPs from anthesis through milk. Previous studies have reported associations in this region for NDVI, greenness, and biomass-related traits under drought and heat stress, and importantly, these intervals have been distinguished from effects associated with *Ppd-B1*, including in analyses that explicitly controlled for photoperiod-related variation (Concorelli et al., 2018; Pinto et al., 2016). Accordingly, the chromosome 2B GR and NDVI signals identified here are plausibly interpreted as reflecting biomass- or canopy-related effects that are at least partly distinct from phenology *per se*.

A similar interpretation applies to the long arm of chromosome 5A, where a high confidence region for NDVI and GR also colocalized with kernel size traits. Numerous studies have identified chromosome 5AL as a hotspot for yield, biomass, and yield component traits, particularly under drought (Bag et al., 2026; Begum et al., 2020; Gahlaut et al., 2017; Krishnan J. et al., 2024; Puttamadanayaka et al., 2020; Shi et al., 2017; Sukumaran et al., 2015; Vukasovic et al., 2022; Yu et al., 2021). This region overlaps the major vernalization locus *Vrn-A1*, so developmental timing cannot be dismissed as part of the explanation. However, previous work has also suggested that at least some 5AL effects on biomass and yield components can be separated from *Vrn-A1* itself (Shi et al., 2017), and other studies have proposed additional candidate loci in the region (Bag et al., 2026; Begum et al., 2020). In Pop80 specifically, segregation at *Vrn-A1* is essentially negligible, since a trait-linked array marker for the locus fell below the MAF threshold used for GWAS. That observation strengthens the interpretation that the 5AL associations detected here may reflect loci other than *Vrn-A1*, even if developmental timing still contributes indirectly.

Another major hotspot occurred on chromosome 6B, where associations were detected for multiple yield and kernel traits as well as GP and SRC. This region has been repeatedly implicated in variation for yield components and nitrogen uptake or partitioning, and these effects have been reported across normal, drought, heat, and salinity conditions (Garcia et al., 2019; Habash et al., 2007; Katz et al., 2022; Manjunath et al., 2024; Muhu-Din Ahmed et al., 2020; Narjesi et al., 2015; Schmidt et al., 2020; Sukumaran et al., 2018). Prior reporting of SRC associations on 6B (Jernigan et al., 2017) also aligns well with the results here. Although the complexity of GY makes it difficult to resolve exact causal loci across variable environments, the recurring implication of chromosome 6B across related traits suggests that this region contributes to final yield outcomes through integrated effects on grain size, nitrogen dynamics, and grain composition (Garcia et al., 2019; Schmidt et al., 2020).

For functional quality traits, several signals likewise overlapped biologically credible genomic regions. The association of SRC and PQS on chromosome 1B near *Glu-B1*, as well as the GP association on chromosome 6D near *Gli-D2*, is consistent with the expectation that introgression materials can expose substantial variation in storage protein composition and related quality traits (Fatiukha et al., 2020a, Nelson et al., 2006; Schaart et al., 2021; Xu et al., 2004). The 6D GP association is especially notable because it demonstrates that even with reduced D subgenome donor proportions, the population development strategy used here still retained sufficient power to detect D subgenome effects, including donor alleles from *Ae. tauschii* associated with increased GP.

Particularly intriguing are the functional quality associations on the long arms of all group 3 chromosomes. The donor allele increased PQS and SRC at these loci wherever donor direction could be assigned unambiguously, and the 3B association was also pleiotropic for PH.

Because similar signals occurred on 3AL, 3BL, and 3DL, these regions raise the possibility of homoeologous contributions to functional quality. This interpretation should remain tentative, however, as no major effect quality loci analogous to *Glu-1* or *Gli-2* have been established on the long arms of the group 3 chromosomes. Still, chromosome 3B has been implicated repeatedly in end-use quality variation, including reports near the region identified here (Hu et al., 2022; Lou et al., 2021; Wondifraw et al., 2024; Yang et al., 2020), while other studies using different marker systems provide broader but less directly comparable support (Carter et al., 2012; Jernigan et al., 2018; Jiang et al., 2017). Chromosome 3A has also been linked to quality-related variation, though generally less prominently (Jernigan et al., 2018; Law et al., 2005; Lou et al., 2021; Wondifraw et al., 2024; Yang et al., 2020). Evidence for chromosome 3D is sparser, but some reports do exist (Lou et al., 2021; Yang et al., 2020). The overlap between 3B quality associations and PH also has precedent in the literature (Carter et al., 2012), indicating either tight linkage or broader pleiotropic relationships between end-use and agronomic traits in this region.

Another quality-related region of interest was the BLUP association for PQS on the long arm of chromosome 2B, which overlapped a BLUP association for GY. Because PQS is a derived metric rather than a commonly mapped primary phenotype, direct comparison with the literature is difficult. Even so, the overlap with GY and with kernel-related BLUEs suggests that this region may influence functional quality indirectly through effects on grain filling and kernel architecture. This interpretation is consistent with extensive prior evidence that 2BL is a pleiotropic hotspot for yield, kernel traits, and quality-related variation (Blanco et al., 2012; Cui et al., 2014; El-Feki et al., 2013; Garland Campbell et al., 1999; Groos et al., 2003; Jernigan et

al., 2017; Krishnan J. et al., 2024; Marcotuli et al., 2017; Smith et al., 2011; Suprayogi et al., 2009; Zhang et al., 2021).

An additional potentially pleiotropic region of strong interest lies on chromosome 2AL, where BLUP associations were detected for KA, KW, TW, TKW, and GP. The kernel size traits colocalized at the same marker, while GP peaked upstream. This general pattern is well supported by prior literature, including reports for GP and kernel traits (Blanco et al., 2012; Peleg et al., 2009a). This region is therefore credible as a multi-trait locus, although the exact biological basis remains unresolved. *TaGS2-A1*, a gene implicated in nitrogen uptake and accumulation in developing grain and reported to affect both GP and kernel traits under both drought and well-watered conditions, lies downstream of the associations reported here and therefore offers a plausible biological hypothesis (Bazhenov et al., 2022; Li et al., 2011).

Chromosome 4B also emerged as a hotspot, with associations for KA, KW, TKW, and TC. Prior studies provide substantial support for this region as one affecting biomass and kernel traits, including under heat stress and during post-anthesis development (Chen et al., 2021; Deng et al., 2011; Guan et al., 2018; H. Li et al., 2024; Xie et al., 2016). Of particular interest is the report by Chen et al. (2021) of a wild emmer-derived allele increasing effective tiller number in this region, albeit with possible tradeoffs in grain weight and grain number per spike. In the present work, donor direction for TC was ambiguous, but the donor-associated increases in KW, KA, and TKW are consistent with a possible reallocation tradeoff in which fewer tillers are associated with larger grains. This remains speculative, but it is a biologically coherent interpretation.

Test weight associations extended across much of chromosome 4A with additional BLUP associations for KW on the distal long arm. Previous reports support 4A effects on yield

components on both chromosome arms (Ji et al., 2023; Song et al., 2026; Sun et al., 2009; J. Zhao et al., 2023), including a long arm TW association under drought (J. Zhao et al., 2023). The proximity of *TaCwi-4A*, a drought-related gene associated with GY and reproductive performance under stress, further suggests that the distal 4AL region may contain factors relevant to grain size and density, though again positional proximity alone is not sufficient to assign causality (Afzal et al., 2019; Jiang et al., 2015).

Chromosome 7A contained another set of notable associations, with long arm signals for KW, TKW, and GFP and a short arm signal for NDVI. The short arm of 7A has been characterized as a biomass-related hotspot, consistent with the NDVI signal detected here (Hu et al., 2015; Lephuthing et al., 2022; Pretini et al., 2021; Royo et al., 2021), whereas the long arm has been associated more specifically with kernel traits such as TKW and KW (Gao et al., 2015; Hu et al., 2015; Lephuthing et al., 2022). Reports of drought-related effects on photosynthesis, biomass, water use efficiency, and grain filling from wild emmer-derived 7A alleles are especially relevant in the present context, since phenotyping was performed across environments differing in drought timing and severity (Lauterberg et al., 2022; Merchuk-Ovnat et al., 2016a). Thus, chromosome 7A appears to fit well within the broader pattern of drought-responsive loci affecting biomass production, grain filling, and final kernel outcomes (Ehdaie et al., 2016; Gahlaut et al., 2017; Gao et al., 2015; Grela et al., 2025; Kumar et al., 2024; Lephuthing et al., 2022; Liu et al., 2023; Maccaferri et al., 2016; McIntyre et al., 2010; Pretini et al., 2021; Xu et al., 2025).

Compared with the A and B subgenomes, fewer associations were detected on the D subgenome. In addition to the 6D GP associations, chromosome 7D harbored short arm associations for KL and KA. Exact physical overlap with prior reports is limited, but 7D has

been implicated previously in kernel shape variation, including traits related to the KL:KW ratio and kernel perimeter (Li et al., 2015; Xin et al., 2020).

Finally, the short arms of chromosomes 3A and 3B contained high confidence PH associations, with the 3BS region also associated with HD. On 3BS, *Rht5* is an attractive candidate because it is a reduced height locus known to influence plant stature and to show pleiotropic effects on maturity and spike architecture (Chen et al., 2018; Cui et al., 2022; Daoura et al., 2014; Ellis et al., 2005; Rebetzke et al., 2012). Previous reports of PH associations on 3A also support the plausibility of the 3AS signal detected here (Rabbi et al., 2021; Sisi et al., 2022; Zanke et al., 2014).

Overall, the associations detected in this work are broadly consistent with previous primary literature and demonstrate that GWAS was capable of recovering biologically credible signals despite low donor allele frequencies under this bridge-mediated introgression framework. Importantly, multiple strong signals occurred near well-characterized loci such as *Rht*, *Ppd*, *Glu*, and *Gli*, while others mapped to genomic regions repeatedly implicated in drought-responsive biomass production and grain filling, development, and composition. At the same time, these results should be interpreted within the limits of GWAS resolution: associated markers identify candidate intervals, not causative genes, and several of the regions discussed here likely contain multiple linked loci affecting correlated traits. Even with that caution, the repeated overlap between the present results and previously reported QTL-rich regions supports the conclusion that the population development and phenotyping strategy used here successfully generated a tractable system for dissecting introgressed variation. These results therefore provide a strong basis for future fine mapping, haplotype analysis, and targeted drought-specific follow-up in the most compelling candidate regions.

Future directions for hypothesis testing and application in breeding germplasm improvement

These phenotypic and genomic results identify several clear priorities for additional research. One worthy of high priority is more extensive characterization of drought response. In this study, drought timing and severity differed across environments and strongly influenced trait expression, particularly through effects on maturity, grain filling, and kernel development. These patterns were highly informative, but drought itself was not experimentally imposed in a way that allowed specific stress response mechanisms to be identified nor quantified. Future work would benefit from side by side irrigated and dryland experiments, or from managed stress trials designed to impose drought at defined developmental stages. Such work would help determine the extent to which performance of individual WILs reflects broad drought stability or more specific adaptation to stress occurring during vegetative growth, grain filling, or both. Such research under field conditions would not necessarily be straightforward, since natural precipitation can obscure treatment effects, but these results indicate that investment in such research would be worthwhile.

The GR phenotype itself also warrants closer study. Because GR was derived from quantification of green pixels, GR does not map neatly onto more standard canopy or greenness indices, even though it overlaps biologically with NDVI in its apparent capture of canopy condition. The colocalization of GR and NDVI suggests some shared underlying signal, but the exact basis of variation in GR remains unclear. In principle, GR might reflect stay green behavior, canopy density, color differences related to physiology, or some composite of these processes. That GR repeatability peaked midseason, whereas NDVI repeatability peaked later, suggests that the two traits are not interchangeable and may differ in the developmental window

or physiological process they capture. This distinction is further supported by stage-specific correlations with GY and GP. Green reflectance was positively correlated with GY but negatively correlated with GP at tillering, anthesis, and milk/dough stages, whereas the opposite pattern was observed around jointing and booting. This may reflect the changing physiological meaning of canopy greenness across development: early- and late-season GR may primarily capture canopy vigor and photosynthetic source capacity, while midseason GR may be more closely associated with crop nitrogen status during a major period of vegetative nitrogen uptake (De Oliveira Silva et al., 2021; Lollato et al., 2019; Sharma et al., 2023). In this context, positive midseason GR correlations with GP may reflect greater nitrogen accumulation or remobilization potential, whereas positive early- and late-season GR correlations with GY may reflect greater biomass production or grain filling capacity that can dilute GP. Determining whether GR provides unique predictive information for drought adaptation or yield performance, beyond established canopy metrics, would therefore be a useful next step. Of note is that NDVI during tillering had the strongest correlation with GY across growth stages, which was unexpected given that canopy indices often show strong GY associations near anthesis (Duan et al., 2017; Rahman et al., 2021). In this population, however, early NDVI may have captured stand establishment, tillering capacity, or early canopy development that contributed more strongly to final yield under drought stressed environments.

Among the genomic regions identified in this research, chromosome 2B appears especially worthy of further study. The 2BS associations located distally to *Ppd-B1* for GR and NDVI suggest canopy- or biomass-related processes that may be at least partly distinct from phenology. These signals make chromosome 2B a strong target for follow-up physiological

validation to test the hypothesis that the GR and NDVI signals correspond to differences in canopy architecture or photosynthetic maintenance.

Additional genomic regions also merit targeted validation, particularly those that align with previous reports of drought-related variation. The associations on 2AL, 4BL, 5AL, 6BL, 7AS, and 7AL are notable in this regard because they harbor pleiotropic associations across biomass, yield components, and kernel traits, and several overlap regions previously implicated in drought-responsive performance. The 4BL and 7A regions are especially interesting because prior studies have also linked wild emmer-derived alleles in these regions to improved performance under drought (Chen et al., 2021; Lauterberg et al., 2022; Merchuk-Ovnat et al., 2016a). Follow-up studies in managed drought environments could help determine whether these loci contribute primarily to drought tolerance, drought escape, plasticity across stress scenarios, or more general adaptation.

Functional quality traits and their relationships represent another potential area of future work. Several marker-trait associations detected for quality traits in this work overlap previously characterized loci related to glutenin and gliadin composition, supporting the hypothesis that wild emmer introgression can broaden allelic diversity for components of end-use quality. Particularly interesting, however, is the result that there are associations not well explained by known major loci. The PQS and SRC associations detected on the long arms of all group 3 chromosomes are especially noteworthy. Because comparable signals occur on 3AL, 3BL, and 3DL, these regions raise the possibility of homoeologous contributions to functional quality, or alternatively, multilocus regulation of downstream quality traits. Narrowing these intervals and characterizing the underlying haplotypes could be a high value next step.

Despite the focus on specific loci, the broader breeding implication remains that final utility must be judged primarily by performance for GY and secondary breeding targets. Introgression is expected to disrupt co-adapted elite haplotypes, and the overall yield disadvantage of the WILs observed in this population is not surprising. That this population produced a smaller number of WILs with competitive yield performance is an important contrast, indicating that while most recombination events among donor and recurrent alleles will not improve elite performance, some rare combinations do bring competitive value. These results reinforce the importance of maintaining sufficiently large introgression populations so that selection can recover valuable recombinant deviates. The value of introgression is not an expectation of average improvement across all derived lines, but rather in increasing the genetic variance and effective population size and thereby increasing the probability of identifying exceptional lines in the upper tail of the genetic distribution.

Most importantly, this work establishes a tractable system for generating, prioritizing, and assessing hypotheses. The study identified which traits resulted in increased variation via introgression, which combinations of traits were most compatible with high yield performance, and which genomic regions warrant further validation. Key questions to be answered through additional research include whether the group 3 long arm quality signals represent homoeologous control, whether GR captures a distinct physiological process, whether specific loci contribute to yield stability rather than condition-specific plasticity, and whether the detected *Glu* and *Gli* region associations reflect known versus novel allelic diversity. These questions could also be tied to experiments involving managed drought, physiological phenotyping, haplotype analysis, as well as validation of current results through use of other genetic backgrounds.

This study should be viewed not as a final dissection of the mechanisms underlying introgressed variation, but as a large-scale screening effort that demonstrates both the feasibility and the value of bridge-mediated introgression from wild emmer into hexaploid wheat. These results show that this breeding strategy can generate statistically detectable allelic variation for many critical traits, identify candidate chromosomal regions, and produce genetic deviates in the upper tail of genetic distribution that results in lines with strong breeding potential despite the presence of population-wide yield disadvantage. That combination of discovery value and applied relevance makes this population a useful foundation for future hypothesis-driven research as well as germplasm improvement.

Conclusion

This work demonstrates that wild emmer introgression into an adapted hexaploid wheat background, achieved through a wheat-*Aegilops tauschii* amphiploid bridge, can generate substantial and meaningful phenotypic variation for quantitative traits while maintaining the breeding value of current, elite germplasm. Across the range of traits, including agronomic, yield component, and functional quality traits, the Pop80 WILs expressed broad variation relative to the RP, as well as individuals lines that exhibit performance competitive with current elite released varieties, confirming that this breeding strategy successfully expanded the range of competitive phenotypes. At the same time, the results also clearly illustrate that introgression in of itself does not simply provide useful genetic gains. Population-wide yield disadvantages, shifts in developmental timing, tradeoffs between biomass and kernel traits, and strong environmental effects on trait expression all underscore the challenge of recovering useful donor variation while preserving elite adaptation. In this context, this work answers its motivating

hypothesis, but with appropriate nuance: wild emmer can contribute useful diversity in hexaploid wheat, but the value depends on recombination, environmental context, and careful selection.

A major value of this work is that the results clearly expose critical tradeoffs within a multivariate framework. Rather than evaluating introgression through single traits in isolation, these results clearly show that performance is the result of interactions amongst multiple traits such as developmental timing, canopy and biomass traits, grain filling dynamics, and kernel composition. Seasonal timing of drought was a component in delineating the structure of these relationships. These seasonal fluctuations in drought influenced whether limitations to final yield arose more strongly through reduced vegetative growth, shortened grain filling, or otherwise altered resource partitioning into the grain. These results and relationships help explain the relatively poor performance of many WILs despite favorable values for individual traits, while a small subset exhibited strong overall performance across multiple environments.

Results also showed that GWAS in a bridge-mediated introgression population can facilitate the identification of credible marker-trait associations even when donor alleles in the resulting breeding population occur at relatively low frequencies. Multiple high confidence signals overlapped loci or genomic regions previously implicated in developmental timing, plant stature, grain composition, drought-responsive biomass production, and end use quality. Other marker-trait associations suggest that introgression populations may also reveal novel sources of allelic variation that have not been previously characterized in existing elite germplasm. These associations should not be interpreted as proof of causation, but collectively these detected signals indicate that the phenotyping and analytical framework was sufficient to detect structured biological signals beyond simple noise. This project contributed not only useful results from

breeding evaluations, but also a useful map of candidate trait regions as well as creating useful hypotheses for future research.

Ultimately, the broader significance of this research lies in demonstrating that introgression breeding should be viewed as a research problem rather than a guarantee of immediate improvement. Most introgressed lines will not outperform elite material; in elite breeding germplasm, favorable alleles and allelic combinations have been assembled through long-term recombination and selection. Yet the purpose of introgression is not to shift the population mean upward in a single generation, the purpose of introgression is to expand the genetic space available for selection and increase the probability of recovering rare, superior combinations that would not otherwise exist. By that standard, this project was successful in establishing that bridge-mediated introgression from wild emmer into hexaploid wheat preserves sufficient adaptation that results in meaningful field evaluations, generates useful phenotypic and genomic variation, and produces a subset of lines with direct breeding application. Equally importantly, results from this project provides a foundation for additional research on introgression: i.e., targeted validation of identified valuable loci, targeted or managed drought phenotyping, and continued recombination and selection to convert wild emmer diversity into breeding value. Consequently, this population, and this research that helps delineate the genetic structure of the population, represents both an endpoint of one phase of discovery and a beginning for more directed use of wild emmer in hexaploid wheat improvement.

Chapter 4 - Grain ionomic profiling of wild emmer introgression lines developed using an amphiploid bridge

Abstract

Wild emmer wheat is an important source of allelic diversity for wheat germplasm improvement, but its contribution to the grain ionome in hexaploid breeding germplasm remains underexplored. This research evaluated grain elemental variation in introgression lines developed using an amphiploid bridge to introduce allelic and trait diversity from wild emmer. Grain ionomic profiles were characterized in Thomas County, Kansas under varying timing and intensities of natural drought stress. The ionome was interpreted alongside previously measured agronomic, yield, kernel, and quality traits. Wild introgression lines expanded variation for several beneficial macro- and micronutrients, albeit with correlated increases in toxic elements within a codependent ionomic system. Multivariate clustering showed that ionomic traits did not behave as a single trait class; instead, different elemental groups aligned with kernel morphology and protein traits or with maturity traits. Genome-wide association results further supported this integration, with ionomic associations colocalizing with regions affecting phenology, plant architecture, canopy performance, kernel traits, and functional quality. These results demonstrate that wild emmer introgression can generate substantial variation for the grain ionome. Breeding applications will require selection in a multivariate framework for favorable mineral profiles while maintaining yield adaptation, employing precision in identification of favorable recombinant variants.

Introduction

Wheat wild relatives represent important reservoirs of allelic diversity that can be used to expand the phenotypic range of modern breeding germplasm. *Triticum dicoccoides* (wild emmer, AABB) belongs to the primary and secondary gene pools of tetraploid (e.g. *Triticum durum*, AABB) and hexaploid wheats (*Triticum aestivum*, AABBDD), respectively (Boehm & Cai, 2024; Hao et al., 2020; Nyine et al., 2020; Qi et al., 2007). As such, wild emmer has been investigated as a donor of favorable alleles for agronomic performance, stress adaptation, disease resistance, grain quality, and nutritional composition (Bacher et al., 2021; Chatzav et al., 2010; Fatiukha et al., 2020a, 2021; Hu et al., 2015; Kato et al., 1997, 1998; Li et al., 2001; Peleg et al., 2005, 2008a, 2008b, 2009a, 2009b; Peng et al., 2011, 2013; Rahman et al., 2023; Ren et al., 2013; Xie & Nevo, 2008). In the preceding chapters, wild introgression lines (WILs) developed through an amphiploid bridge-mediated approach were evaluated for introgression patterns and broad phenotypic performance. This chapter extends that characterization to the grain ionome, with the goal of determining whether wild emmer introgression provides useful variation for elemental traits in an adapted hexaploid wheat background.

The grain ionome refers to the full complement of mineral nutrient and trace element concentrations measurable in grain tissue (Salt et al., 2008). In wheat, the ionome includes essential macro- and micronutrients, elements that may serve as indicators of plant physiological processes or environmental conditions, along with elements associated with toxicity. Specific elements in the grain ionome have not typically been primary selection targets in wheat breeding programs, although relevant in the context of human nutrition (Garvin et al., 2006; Liu et al., 2021; Peleg et al., 2008b). This is partly because grain elemental composition is governed by complex quantitative inheritance influenced by soil properties, water availability, root

architecture, xylem and phloem transport, remobilization from vegetative tissues, grain filling dynamics, and yield dilution (Carvalho & Vasconcelos, 2013; Fatiukha et al., 2020b; Miner et al., 2022, 2024; Watanabe et al., 2022a; Wysocka et al., 2025). As a result, ionomic traits often exhibit strong GxE interaction and can be difficult to interpret in isolation. However, this complexity also makes the ionome a useful framework for evaluating how wild-derived allelic diversity may influence nutrient acquisition, partitioning, and grain composition.

Wild emmer has been recognized as a source of favorable variation for grain nutrient concentrations. The most prominent example is the *Gpc-B1* locus on chromosome 6B derived from wild emmer, which has been associated with increased grain protein concentration and senescence-driven remobilization of iron, zinc, and manganese (Distelfeld et al., 2007). The successful introgression of this locus into both tetraploid and hexaploid wheat illustrates the potential value of wild emmer for improving grain nutritional quality. However, examples comparable to *Gpc-B1* remain relatively limited. Although wild emmer accessions have been characterized as harboring broad variation for macro- and micronutrient concentrations, fewer studies have evaluated this variation after introgression into adapted breeding germplasm, where wild alleles can be assessed in an agronomically relevant genetic background (Cakmak et al., 2004; Chatzav et al., 2010; Fatiukha et al., 2020b; Liu et al., 2021; Pandey et al., 2016; Peleg et al., 2008b, 2009a; Xie & Nevo, 2008).

Among grain ionomic traits, iron (Fe) and zinc (Zn) are the most prominent targets for wheat biofortification. Both are essential for human health, and deficiencies in Fe and Zn remain major contributors to global micronutrient malnutrition (Cakmak et al., 2004; Carvalho & Vasconcelos, 2013; Guttieri et al., 2015b; Pandey et al., 2016; Peleg et al., 2008b). However, breeding for increased grain yield has often been associated with reduced concentrations of grain

minerals, such as Fe and Zn, due to yield dilution and historical selection on harvest index. Under yield dilution, greater grain biomass can reduce nutrient concentration per unit dry weight when total nutrient uptake or remobilization does not increase proportionally with grain yield (Guttieri et al., 2015b; Miner et al., 2022, 2024; Monasterio & Graham, 2000; Perrier et al., 2016). Therefore, identifying germplasm that combines acceptable yield performance with increased micronutrient concentrations remains a central challenge for biofortification.

Zinc is additionally relevant from a crop physiology perspective, particularly under drought stress. Low soil Zn availability can limit wheat growth and development, and drought can further restrict Zn acquisition by reducing mass flow and diffusion of nutrients to the root surface. Zinc contributes to enzyme function, membrane stability, oxidative stress mitigation, and reproductive development, all of which may influence plant performance under water-limited conditions (Akbarimehr et al., 2022; Bagci et al., 2007; Cakmak et al., 1996; Karim et al., 2012; Peleg et al., 2008b). Consequently, genotypic differences in Zn uptake, partitioning, and remobilization may be relevant not only for nutritional quality but also for adaptation to stressful production environments. Wild emmer has been reported to contain substantial variation for grain Zn and Fe (Cakmak et al., 2004; Durmaz et al., 2011; Peleg et al., 2008b, 2009a), making it a potentially useful donor for improving these traits in wheat breeding germplasm.

A major complication in grain biofortification is that beneficial elements do not vary independently. Phosphorus (P), for example, is an essential macronutrient and a major component of grain reserves, but much of the P in wheat grain is stored as phytic acid. Phytic acid chelates mineral cations, including Zn, Fe, magnesium (Mg), calcium (Ca), and potassium (K), thereby reducing their bioavailability to humans (Fatiukha et al., 2020b; Frittelli et al., 2023; Guttieri et al., 2015a; Peleg et al., 2009a). Because total grain P is often associated with phytate

accumulation, increasing micronutrient concentrations without simultaneously increasing phytate-associated anti-nutritional effects is an important consideration. In this context, the direction and magnitude of correlations among P, Fe, Zn, Mg, Ca, and other cations are biologically informative and evaluating the ionome in a multivariate framework is essential in addition to considering each element independently.

Toxic elements, such as cadmium (Cd), arsenic (As), and aluminum (Al), introduce another layer of complexity. Cadmium is of particular interest in wheat because it can co-accumulate with Zn due to overlapping uptake and transport pathways, creating a potential tradeoff in which selection for increased Zn could inadvertently increase grain Cd concentration (Guttieri et al., 2015b; Perrier et al., 2016; Qin et al., 2021). Arsenic accumulation is similarly undesirable because of its toxicity, with additional impacts on plant health and uptake of desirable nutrients (Zaheer et al., 2024). Aluminum is a marker of soil acidity, where low pH increases soluble Al, which is toxic and inhibits root growth.

Beyond the major biofortification and toxicity minimization targets, several additional elements provide information about plant health, physiology, and environmental response. Potassium, Ca, and Mg are essential macronutrients involved in osmotic regulation, enzyme activation, membrane function, and structural stability. Potassium is particularly important for stomatal regulation and drought response (Akhtar et al., 2022; Patwa et al., 2024; Watanabe et al., 2022a), whereas Mg is central to chlorophyll function and enzymatic processes (Peleg et al., 2009a). Sodium can provide information about ionic balance and stress response, especially where salinity or altered K/Na relationships affect plant performance (Hu et al., 2006; Patwa et al., 2024; Shelden & Roessner, 2013; Watanabe et al., 2022a). Soil acidification, common in intensively managed production systems due to increased nitrogen fertilization (Ballagh et al.,

2023; Lollato et al., 2013, 2019; Schroder et al., 2011), also can reduce the availability of Ca, Mg, sodium (Na), and K (Oh et al., 2014; Souza et al., 2023). Rubidium (Rb) is chemically similar to K and is broadly interpreted as a tracer of phloem activity, whereas strontium (Sr) is chemically similar to Ca and informs xylem activity (Seregin & Kozhevnikova, 2004; Shtangeeva et al., 2011; Vallejo et al., 2005; Zeller & Feller, 2000; Ziegler et al., 2017).

Boron (B), manganese (Mn), copper (Cu), molybdenum (Mo), nickel (Ni), cobalt (Co), and selenium (Se) occur at lower concentrations but may still be informative. Boron is involved in cell wall structure and reproductive development (Akhtar et al., 2022; Brdar-Jokanović, 2020), Mn and Cu participate in redox and enzymatic processes (Karim et al., 2012; Peleg et al., 2009a), Mo is required in nitrogen metabolism (Rana et al., 2025; Watanabe et al., 2022a, 2022b), Ni is associated with urease activity (Gheibi et al., 2009), and Se, while not essential for higher plants, is nutritionally relevant for humans and may vary with soil availability and uptake capacity (Iqbal et al., 2015; Schiavon et al., 2020; Sunic & Spanic, 2024). Together, these elements provide a broader picture of nutrient acquisition and partitioning.

The interpretation of grain elemental concentrations is further complicated by environmental sensitivity. Soil pH, organic matter, carbon content, fertilizer history, water availability, root zone chemistry, and the availability of interacting elements all influence elemental uptake and grain deposition (Carvalho & Vasconcelos, 2013; Fatiukha et al., 2020b; Miner et al., 2022, 2024; Watanabe et al., 2022a; Wysocka et al., 2025). Therefore, ionic phenotypes observed in grain reflect not only genetic capacity for uptake and remobilization, but also the environmental conditions under which plants developed and filled grain. Because of this sensitivity, repeatability for the grain ionome is often relatively low. Previous wheat ionic studies have reported substantial GxE interaction and strong environmental effects on elemental

concentration (Asaro et al., 2016; Guttieri et al., 2015b; Ziegler et al., 2017). Analytical factors can also contribute to uncertainty because ionomics measurements are sensitive to sample handling, contamination, digestion efficiency, and instrument stability (Miner et al., 2024). These challenges make single-environment interpretation difficult and emphasize the need for replicated field evaluation, careful outlier control, and statistical models that separate stable genotypic effects from environment-specific variation. At the same time, the complexity of the ionome makes wild-derived germplasm especially valuable. If the base level of trait variation in elite germplasm is narrow, then identifying favorable combinations of high beneficial nutrients, low toxic elements, and acceptable agronomic performance may be improbable. Introgression populations offer an opportunity to expand the range of segregating variation and potentially identify phenotypic deviates that partially decouple unfavorable relationships among elements (Liu et al., 2021; Peleg et al., 2009a).

The WIL population evaluated in this study provides a useful framework for this purpose. WILs can be evaluated in relation to both the recurrent parent (RP) background and elite performance checks. This allows wild-derived ionomic variation to be assessed within a breeding-relevant context rather than only in unadapted wild accessions. In addition, characterized introgression segments and genome-wide marker data make it possible to connect phenotypic variation in the grain ionome with underlying genomic regions (Chapter 2). When ionomic marker-trait associations overlap regions of elevated introgression, they may indicate genomic intervals where wild-derived alleles contribute to elemental variation. When ionomic associations also overlap agronomic, yield and kernel, or functional quality associations (Chapter 3), they may reveal regions where selection for ionomic traits could interact with broader breeding targets.

Therefore, the objectives of this research were to (1) characterize grain ionomic variation in WILs relative to the RP and performance checks, estimating the repeatability and correlation structure of the ionome across Thomas County, Kansas environments, (2) identify marker-trait associations for elemental concentrations using single- and multi-locus models, and (3) assess whether ionomic associations overlap introgression regions or genomic regions associated with previously characterized phenotypic variation. Collectively, this research evaluates whether amphiploid bridge-mediated introgression from wild emmer can contribute useful ionomic diversity to hexaploid wheat breeding germplasm and whether that variation can be interpreted in the context of broader breeding performance.

Methods

Population development and field trialing

The development of the *Triticum dicoccoides* germplasm development (TdGD) populations, selection of TdGD population 80 (Pop80), and multi-environment field trial design are described in detail in Chapters 2 and 3. Briefly, wheat-*Aegilops tauschii* synthetic octoploid amphiploids were generated and used as a bridge to facilitate hybridization with wild emmer. The resulting F₁ plants were backcrossed twice to an adapted hexaploid recurrent parent (RP), followed by advancement to the BC₂F_{4.5} generation through single seed descent. TdGD Pop80 was selected from the seven derived TdGD populations for comprehensive phenotypic evaluation based on population size, broad agronomic performance, and plant and spike architecture.

TdGD Pop80 was evaluated in partially replicated small-plot field trials across six Kansas environments from 2021-2023, as described in Chapter 3. For the ionomics analysis, grain samples were used from three Thomas County (THD) environments in each of 3 growing

seasons: 21THD, 22THD, and 23THD. Restricting the ionomics analysis to a single county reduced broad geographic environmental heterogeneity while retaining year to year variation in field conditions. Soil testing results from 22THD and 23THD are presented in Appendix Table C.1. The 22THD environment represented the most severe drought stress, with water limitation persisting across vegetative and reproductive development. The 21THD environment experienced moderately severe drought stress, including limited precipitation during vegetative growth and terminal drought stress, although moisture was less limiting near anthesis. The 23THD environment was the least stressed in terms of drought, with water limitation primarily occurring early in vegetative development with more favorable moisture availability during anthesis and grain fill (Chapter 3).

Grain ionomic profiling

Grain samples from harvested plots in 21THD, 22THD, and 23THD were submitted to the Baxter laboratory at the Donald Danforth Plant Science Center, St. Louis, Missouri, for ionomic profiling. Ionomics refers to the high-throughput quantitative measurement of elemental concentrations in biological samples using inductively coupled plasma mass spectrometry (ICP-MS) to characterize mineral nutrient and trace element profiles (Asaro et al., 2016; Ziegler et al., 2017). Grain samples were prepared, digested, and analyzed by ICP-MS following standardized protocols as described by Ziegler et al. (2017).

Elemental concentrations were reported as mg analyte kg⁻¹ tissue for the following elements: boron (B), sodium (Na), magnesium (Mg), aluminum (Al), phosphorus (P), potassium (K), calcium (Ca), iron (Fe), manganese (Mn), cobalt (Co), nickel (Ni), copper (Cu), zinc (Zn), arsenic (As), selenium (Se), rubidium (Rb), strontium (Sr), molybdenum (Mo), and cadmium (Cd). To minimize contamination risk and analytical bias, sample digestion used trace metal

grade nitric acid and ultrapure filtered water. Elemental measurements were corrected using internal standards and blank samples to account for dilution and background signal during digestion and ICP-MS analysis. Pooled digestate controls were included to monitor analytical stability within and across ICP-MS runs.

Because single element outliers within samples are common in ionomics datasets, elemental concentrations were screened prior to mixed model analysis. Extreme single element outliers were identified using a conservative median absolute deviation (MAD) cutoff of 6 (Davies & Gather, 1993). Outlier decisions were evaluated in the context of biological replication so that genotypes with reproducible high elemental concentrations were not excluded solely because they represented potential hyperaccumulating phenotypes. Samples with extreme Al concentrations were additionally flagged because elevated Al may indicate sample metal contamination. Samples showing extreme Al together with widespread elemental outliers across multiple analytes were excluded entirely from downstream analysis.

Phenotypic analysis of elemental concentrations

Analyte data were analyzed using the same two-stage phenotypic modeling framework described in Chapter 3. For each element, plot-level observations were standardized within environment using Z-score transformation prior to modeling, resulting in environment-specific distributions with mean 0 and standard deviation 1. This standardization allowed elemental traits to be compared across environments and reduced scale differences among elements before variance component estimation (Appendix Table C.2).

In the first stage, within-environment analyses were conducted to obtain best linear unbiased estimates (BLUEs) and corresponding precision weights for each genotype in each environment. Spatial correction was performed in Mr. Bean (Aparicio et al., 2024) using

ASReml-R (Butler et al., 2023) for 21THD and 23THD. For these environments, field heterogeneity was modeled using an AR1 x AR1 residual covariance structure (Gilmour et al., 1997). 22THD was analyzed in Mr. Bean using SpATS because residual patterns did not conform well to the AR1 x AR1 structure; two-dimensional P-spline smoothing was therefore used for spatial correction in that environment (Rodríguez-Álvarez et al., 2018). Precision weights were calculated as the inverse of sampling variance of each BLUE and were carried forward into the second stage analysis (Möhring & Piepho, 2009).

In the second stage, across-environment best linear unbiased predictions (BLUPs) were extracted for each genotype in ASReml-R using BLUEs and their corresponding precision weights as input data (Butler et al., 2023; Möhring & Piepho, 2009). The model included environment as a fixed effect and genotype and GxE as random effects. The residual scale parameter was fixed at 1, allowing the precision weights to define the relative contribution of each BLUE to the across-environment model (Butler et al., 2023; Möhring & Piepho, 2009). Under this framework, more precise BLUEs exerted greater influence on the estimation of genotype main effects and GxE variance.

Repeatability was estimated separately within and across environments. Within-environment repeatability (R^2_{within}) was estimated by BLUE models in which genotype was treated as a random effect. These estimates describe the extent to which replicated observations within a given environment consistently differentiated genotypes after accounting for spatial heterogeneity. Across-environment repeatability (R^2_{across}) was estimated by the BLUP model as the ratio of genotypic variance to the sum of genotype and GxE variance. Therefore, R^2_{across} was interpreted as the consistency of genotype performance across years within THD rather than as the proportion of total phenotypic variance explained.

Assessment of trait relationships

Relationships among traits within the ionome were assessed using Pearson correlation analysis. Correlations among BLUEs were calculated within each environment and interpreted as relationships among spatially adjusted phenotypic estimates within a given year. Correlations among BLUPs were calculated across environments and interpreted as relationships among predicted genotypic values for each element. These analyses were used to evaluate whether elements varied independently or as part of a broader ionomic covariance structure.

To place the grain ionome within the broader phenotypic context of TdGD Pop80, hierarchical clustering was conducted across the full set of collected traits. This analysis included BLUPs for agronomic performance, yield and kernel components, and functional quality traits described in Chapter 3, together with BLUPs for the grain elemental traits analyzed herein. When summarizing WIL phenotypic expression across diverse trait classes, interpretation of overall performance becomes challenging when traits differ in direction and magnitude of correlations. Hierarchical clustering provides a way to visualize broad multivariate structure among traits and to identify groups of traits that vary together across genotypes. A dendrogram was constructed using correlation distance, calculated as $1 - r$, where r is the Pearson correlation coefficient among trait BLUPs. Clustering was performed using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA), which, while standard in the literature (Datta & Datta, 2003; Mohammadi & Prasanna, 2003), is intended as an exploratory tool to visualize similarity patterns among traits. Traits that are strongly positively correlated merge at low dendrogram heights, whereas traits with weak, independent, or negative relationships merge only at larger distances.

Genome-wide association analysis and introgression overlap

Genome-wide association studies (GWAS) were conducted using GAPIT version 3 (Wang & Zhang, 2021) with BLUEs and BLUPs as phenotypic inputs. SNP genotyping, marker filtering, and preparation of the genotype matrix are described in Chapters 2 and 3. Briefly, TdGD Pop80 WILs were genotyped using the Illumina Infinium Wheat/Barley 40 K SNP array (Keeble-Gagnère et al., 2021), and markers were filtered prior to association testing to remove low quality and rare variants.

Marker-trait associations were tested using three complementary models implemented in GAPIT: (1) the mixed linear model (MLM; Yu et al., 2006), (2) the multi-locus mixed model (MLMM; Segura et al., 2012), and (3) Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK; M. Huang et al., 2019). The MLM was used as a single-locus mixed model framework controlling for population structure and relatedness. MLMM and BLINK were included as multi-locus approaches to identify associations that may be partially masked in single marker testing and to evaluate consistency of marker-trait signals across model structures. Associations detected across multiple models, across multiple traits, or across both BLUE and BLUP parameterizations were considered higher confidence signals.

Introgression segments identified in Chapter 2 were compared with ionomic GWAS results to determine whether marker-trait associations overlapped genomic regions carrying wild-derived introgressions. Ionomic associations were also compared with associations detected for agronomic performance, yield and kernel components, and functional quality traits described in Chapter 3. These comparisons were used to identify genomic regions where introgression may contribute to variation in grain elemental concentration, either independently or in association with broader phenotypic effects.

Results

Phenotypic evaluation

Table 4.1. Summary of mixed model results across ionic traits evaluated in TdGD Pop80

Fixed environment effects from the across-environment model are reported for each environment, expressed relative to a standard normal distribution. Random effect variance components are partitioned into genotypic and genotype x environment (GxE) terms. Repeatability (R^2_{across}) was calculated as the proportion of genotypic variance relative to total variance.

Trait	Fixed			Random		R^2_{across}
	21THD	22THD	23THD	Genotype	GxE	
B	-0.032	0.002	0.002	0.029	0.231	0.113
Na	-0.049	0.008	0.134	0.037	0.279	0.116
Mg	0.005	0.061	0.020	0.155	0.108	0.588
Al	-0.107	0.045	0.014	0.058	0.237	0.197
P	0.063	0.021	-0.018	0.054	0.283	0.161
K	-0.099	-0.058	0.041	0.117	0.181	0.392
Ca	-0.062	0.026	-0.049	0.066	0.164	0.286
Fe	-0.026	0.031	0.085	0.226	0.148	0.605
Mn	0.109	0.097	0.144	0.204	0.234	0.466
Co	-0.025	-0.015	-0.206	0.099	0.069	0.590
Ni	-0.177	0.005	-0.437	0.160	0.266	0.376
Cu	-0.010	0.068	-0.161	0.191	0.155	0.552
Zn	-0.222	0.027	-0.062	0.046	0.181	0.202
As	0.196	-0.033	-0.201	0.158	0.107	0.595
Se	-0.188	0.099	0.102	0.025	0.157	0.139
Rb	-0.007	-0.010	0.646	0.136	0.191	0.417
Sr	-0.085	0.064	0.232	0.112	0.312	0.264
Mo	-0.040	0.088	0.016	0.254	0.138	0.647
Cd	0.002	-0.013	-0.089	0.134	0.174	0.436

Across environments

In general, R^2_{across} for ionic traits was (0.113-0.647; Table 4.1) and correlations among elements were positive (Figure 4.2), with only a few modest negative correlations (e.g., Rb~Cd = -0.18) existing. The observed positive correlations suggest that variation in kernel concentrations is driven primarily by overall efficiency of nutrient translocation, where hyperaccumulating

genotypes show consistently elevated concentrations across multiple elements with shared transport and metabolic pathways.

Some of the strongest correlations among the BLUPs were between P ($R^2_{\text{across}} = 0.161$), Mg ($R^2_{\text{across}} = 0.588$), Fe ($R^2_{\text{across}} = 0.605$), Cu ($R^2_{\text{across}} = 0.552$), Mn ($R^2_{\text{across}} = 0.466$), Mo ($R^2_{\text{across}} = 0.647$), Cd ($R^2_{\text{across}} = 0.436$), and Zn ($R^2_{\text{across}} = 0.202$). WIL distributions for these elements further highlight shared transport patterns (Figure 4.1). Cadmium exhibited a pronounced positive tail in the BLUPs, indicating that many WILs accumulated higher grain Cd than the RP and checks; Showdown (C3) was the only check that fell within the upper tail. Notably, average grain Cd concentrations in this dataset were low enough to be outside the range of practical relevance (Appendix Table C.2; Guttieri et al., 2015a, 2015b). Similar positive tails were observed for Zn and Fe, with Fe particularly notable; most WILs exceeded the RP and checks for grain Fe. For Zn, C3 was the highest among checks and overlapped with WILs in the upper tail. The positive tail was even stronger for Mn, with nearly all WILs accumulating more Mn than the RP and checks. WILs also had a large positive tail for Mg, with the RP the highest among the checks but still falling below many WILs. Molybdenum stood out as the element with the highest R^2_{across} and had a large positive tail similar to the distribution for Mg. The strongest correlation in these data was between P and Mg (0.70; Figure 4.2); despite the low R^2_{across} for P (Table 4.1), P correlations were relatively strong within the group of metabolic elements. A positive tail was similarly observed for P distribution, where the RP was the highest among checks in the BLUPs. Copper also showed a pronounced positive tail in WIL distributions, consistent with elevated levels of correlated elements in the WILs relative to the checks. Selenium also showed moderate positive correlations with this grouping of correlated elements; a low R^2_{across} (0.139) and less notable transgressive segregation around the RP and checks indicates

low genotypic variation relative to environmental control of Se accumulation (Table 4.1; Figure 4.1).

Additional pairings of elements with moderately strong correlations (Figure 4.2) were K and Rb (0.68) and Ca and Sr (0.62). Potassium and Rb had R^2_{across} of 0.392 and 0.417, respectively (Table 4.1). Calcium and Sr had R^2_{across} of 0.286 and 0.264, respectively. Along with being strongly correlated to Ca, Sr was also moderately correlated with Rb (0.45). Sodium ($R^2_{\text{across}} = 0.116$) was weakly positively correlated with K (0.23) and Rb (0.20); the strongest Na BLUP correlation was with B (0.52). Boron was essentially uncorrelated with K, Rb, Ca, and Sr and had the lowest R^2_{across} (0.113) within the ionome. WILs showed segregation for both Na and B outside the range captured by the checks, with the RP falling lower on the distributions than the remaining checks (Figure 4.1). While the RP was also the lowest check in distributions for K and Rb, pronounced negative tails were observed in the WILs for those traits, with few WILs exceeding the highest checks for Rb, and no WILs exceeding the highest check for K (Billings, C1). For both Sr and Ca, the RP fell in the upper tail of the distribution, with a subset of WILs below the lowest checks (SY Monument, C4 for Sr; C1 for Ca).

In this dataset, Al had low R^2_{across} (0.197; Table 4.1), which likely contributes to its weak overall trait correlations (Figure 4.2). Its strongest correlations were with Na (0.32) and B (0.30). WIL distributions for Al showed a modest positive tail, with the RP consistently below the mean (Figure 4.1). Relative to the ionome, As had high R^2_{across} (0.595) and was correlated most strongly to Rb (0.44) and K (0.42). Similar to K and Rb, As showed a pronounced negative distributional tail among WILs, with the RP being the highest As accumulator among the checks and most WILs accumulating less As than the RP. Similar to As, Co also had high R^2_{across}

(0.590); Co was most strongly correlated to Ni (0.51). Neither Co or Ni ($R^2_{\text{across}} = 0.376$) had clear distributional skews relative to the RP and checks (Figure 4.1).

Within environments

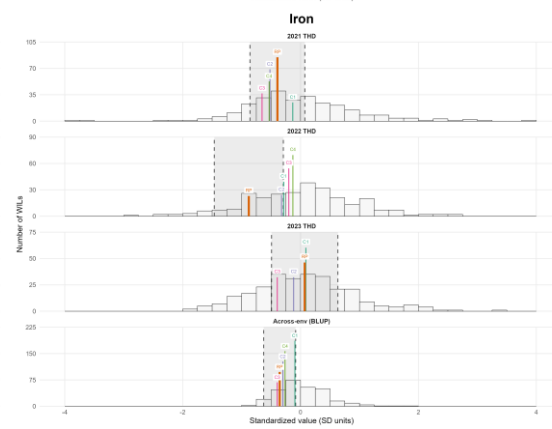
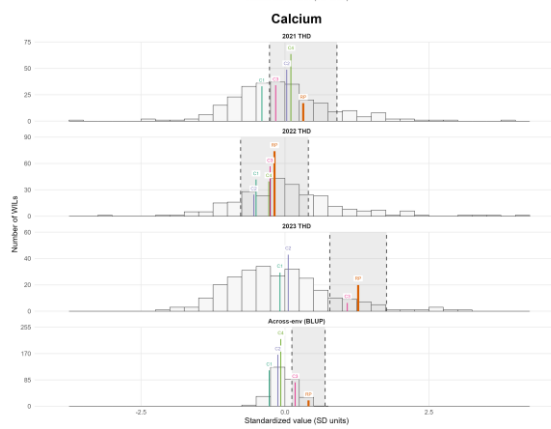
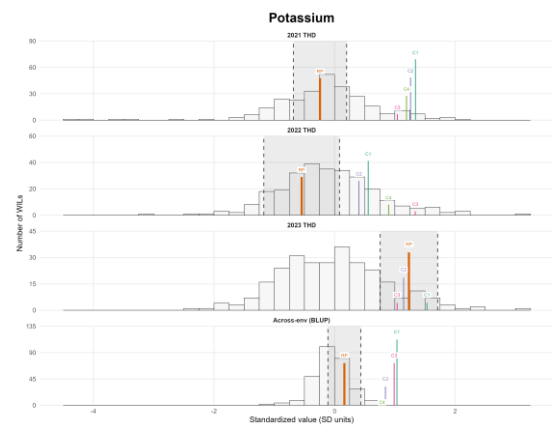
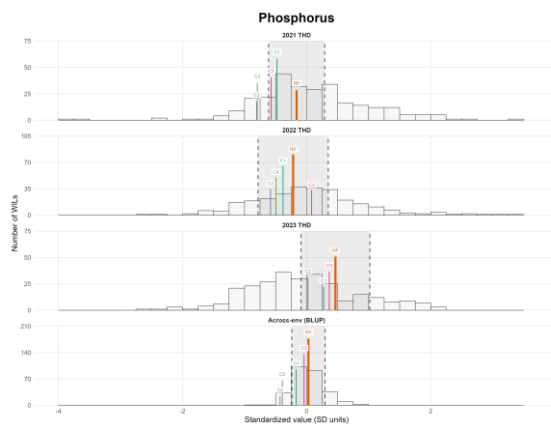
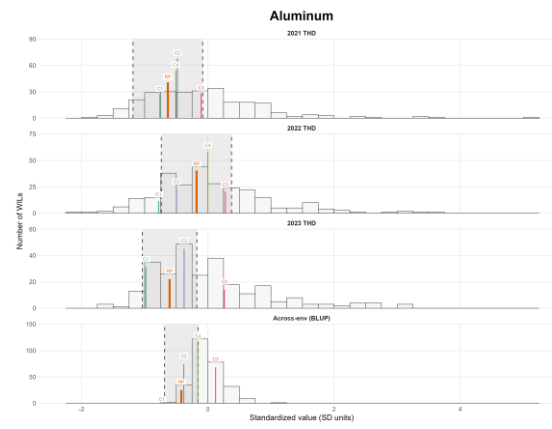
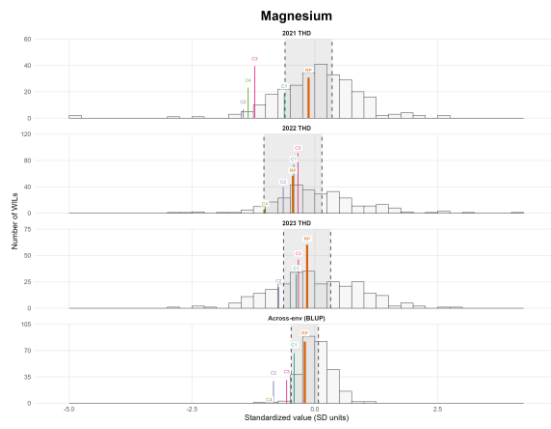
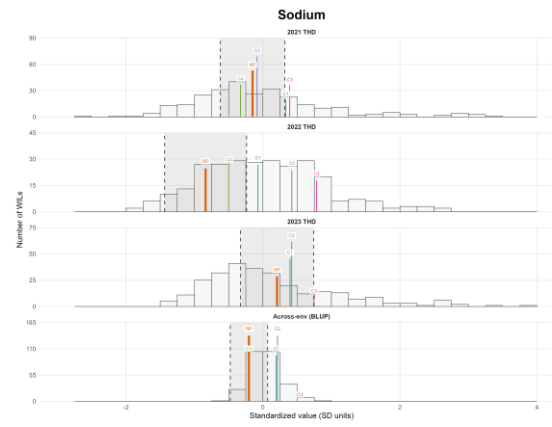
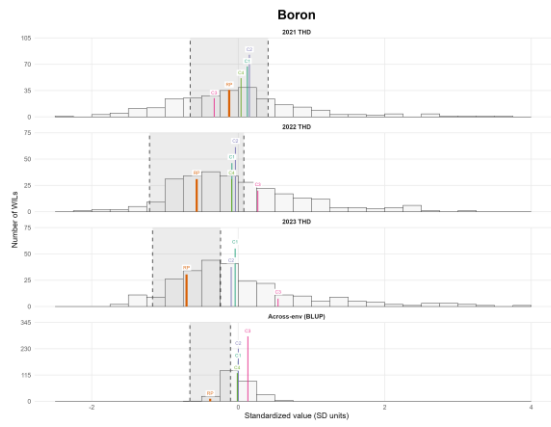
Similar to R^2_{across} , R^2_{within} for ionic traits were also relatively low (Appendix Table C.3). In general, 22THD had the lowest R^2_{within} ; this, combined with generally small environment effects (Table 4.1), support the interpretation that soil water availability in that year primarily constrained nutrient availability rather than revealing consistent genotypic differences in kernel accumulation (Singh et al., 2012; Hu et al., 2006). 23THD generally had the largest environment effects, while 21THD had the highest R^2_{within} . Similar to the BLUPs, BLUE correlations within the ionome were generally positive (Appendix Figures C.1-C.3).

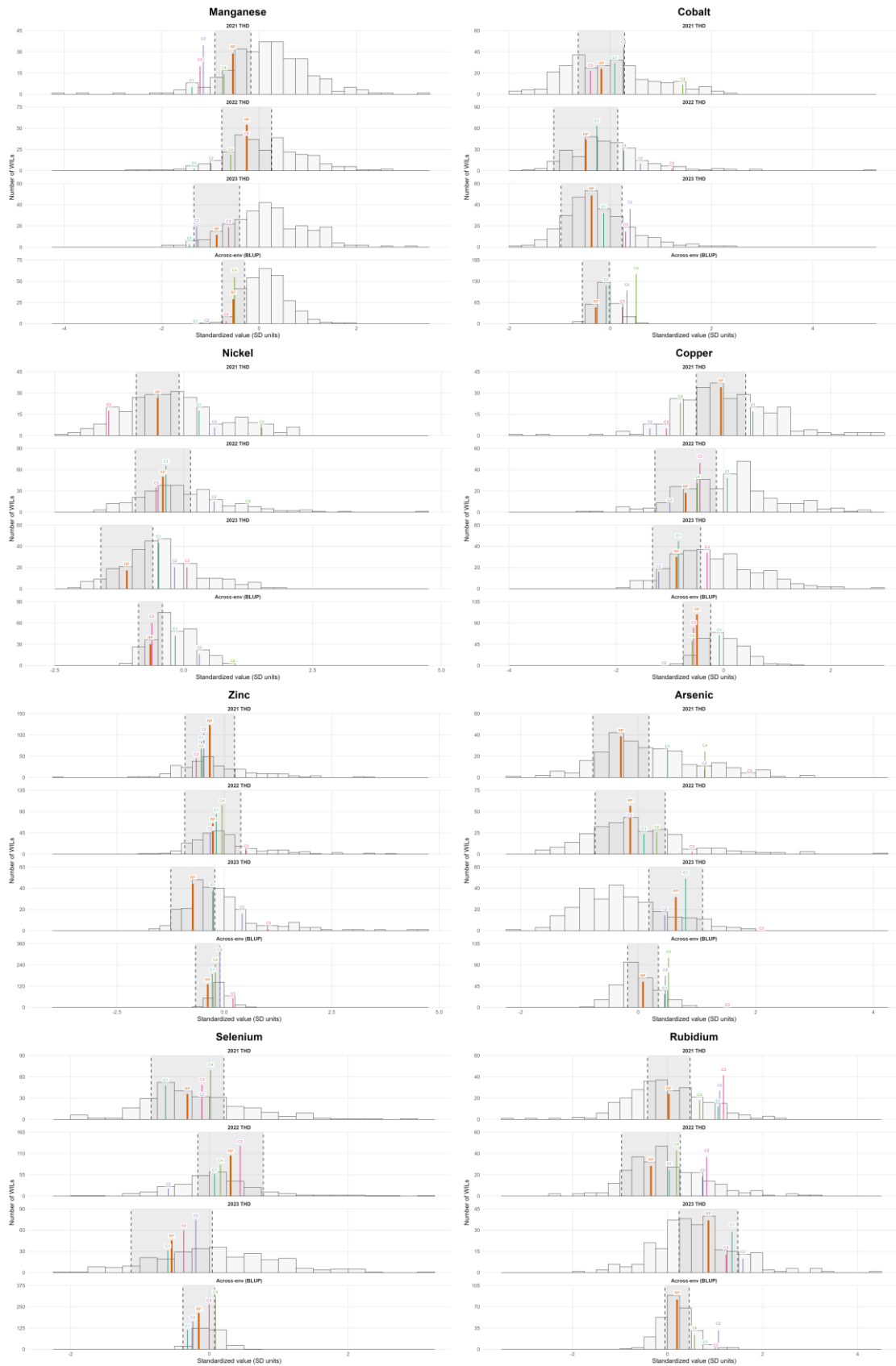
Consistent with performance in the BLUPs, P ($R^2_{\text{within}} 0.24-0.50$), Mg ($R^2_{\text{within}} 0.20-0.46$), Fe ($R^2_{\text{within}} 0.16-0.45$), Mn ($R^2_{\text{within}} 0.41-0.68$), Cu ($R^2_{\text{within}} 0.30-0.51$), Zn ($R^2_{\text{within}} 0.05-0.40$), Cd ($R^2_{\text{within}} 0.32-0.47$), and Mo ($R^2_{\text{within}} 0.28-0.63$) were relatively well correlated within each environment. Traits with low R^2_{across} could be explained by low R^2_{within} and rank changes among the checks in distributions. Zinc, for example, had substantial within-trial heterogeneity in 22THD in addition to the rank ordering of the RP and checks being inconsistent between environments (Figure 4.1; Appendix Table C.3). Low R^2_{across} could likewise be explained by directional differences in effects between environments. The low R^2_{across} for Se, for example, is consistent with its relatively large negative environment effect in 21THD (-0.188) and positive effect in 23THD (0.102; Table 4.1). Although minimally observed in the BLUP distribution, WILs have a positive tail for Se grain concentration, most pronounced in 23THD, with the RP and all the checks in the lower half of the distribution (Figure 4.1). In contrast, elements with

relatively high R^2_{across} and comparably high R^2_{within} were generally consistent in the distributional rankings between environments (e.g., Mo, Mg).

The strongest BLUE correlations were observed for elements involved in vascular activity (Appendix Figures C.1-C.3), K~Rb (0.72 and 0.71 in 21 and 22THD, respectively) and Ca~Sr (0.73 in 23THD). Consistent with the rest of the ionome, R^2_{within} was notably low for K, Rb, Ca, and Sr in 22THD (0.02-0.16), but was reasonably high in 21THD (0.25-0.57) and 23THD (0.38-0.53). The largest positive environment effect across the ionome was observed for Rb in 23THD (0.646); Sr also had a relatively large environment effect in 23THD (0.232). Distributions of K and Rb BLUEs were consistent with the corresponding BLUP distributions, with negative tails observed for the WILs in each environment (Figure 4.1). Consistent with less obvious distributional skews in the WILs for Sr and Ca BLUPs, BLUE distributions for Sr and Ca differed in terms of rank changes among checks and direction of skew, when present. For Sr, the RP was higher than all checks in 23THD but intermediate among checks in 21 and 22THD; most WILs fell below the RP in 23THD, whereas in 21 and 22THD the RP and checks tended to fall in the middle or upper part of the distribution. Calcium showed a modest positive tail in 22THD and a negative tail in 23THD, with the RP above most WILs in 23THD but closer to the distribution center in 21 and 22THD.

The negative distributional skew in the WILs for As BLUPs persisted through the BLUE distributions, with the strongest skew in 23THD (Figure 4.1). Consequently, As had a relatively large negative environment effect in 23THD (-0.201; Table 4.1). Arsenic also had relatively high R^2_{within} in 21 and 23THD (0.44 and 0.45, respectively; Appendix Table C.3).





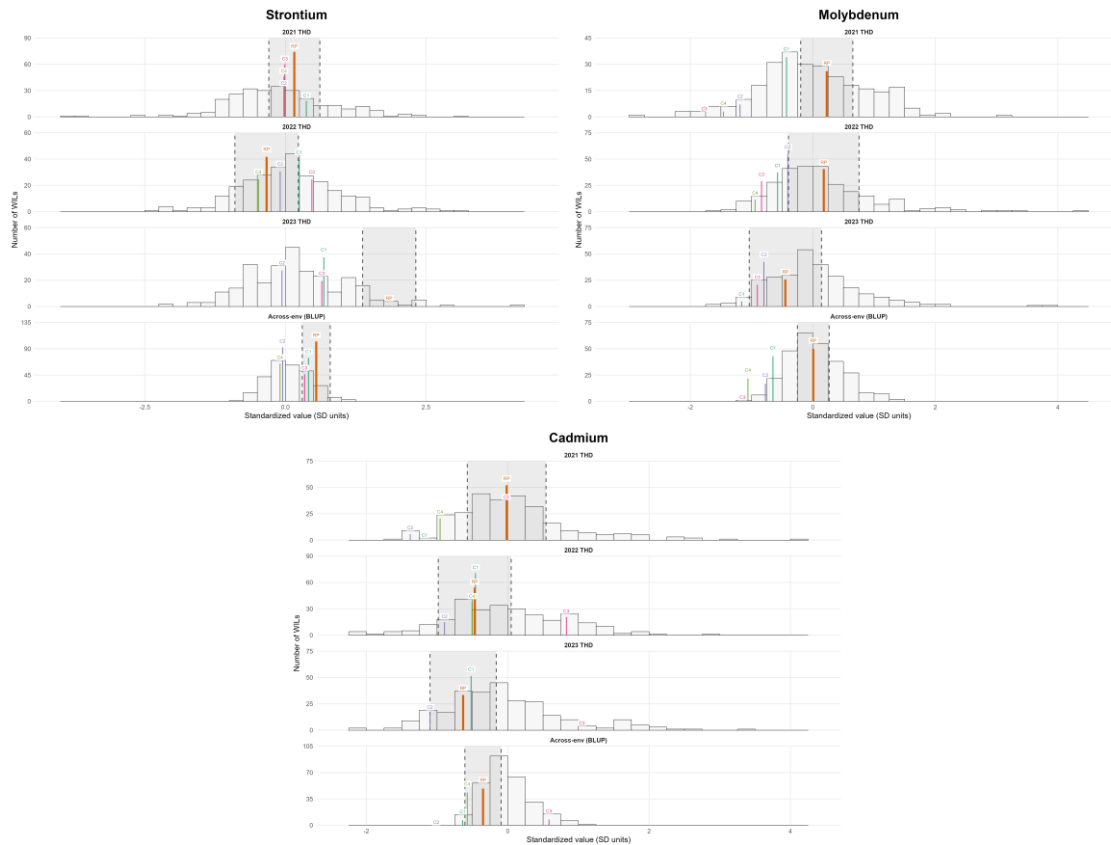


Figure 4.1. Distributions of ionic traits in TdGD Pop80 within and across environments

Environment-specific facets are based on BLUES, while the across-environment facet represents BLUPs. Colored vertical lines indicate the recurrent parent (RP=KS061278M-4) and performance checks (C1=Billings, C2=Joe, C3=Showdown, C4=SY Monument). Gray shaded areas represent the RP's approximate 95% confidence intervals for BLUES and prediction intervals for BLUPs, calculated as each estimate ± 2 standard errors. Plots were generated in R using the ggplot2 package (Wickham, 2016).

Multivariate relationships among traits

Hierarchical clustering was used to place the grain ionome within the broader multivariate phenotypic structure of TdGD Pop80. Cluster assignments were evaluated by cutting the same dendrogram at $k = 2, 3, 4$, and 5 clusters (Appendix Figure C.4).

At the broadest partition, $k = 2$ separated the trait set into two major clusters. One cluster contained grain yield (GY), tiller count (TC), normalized difference vegetation index (NDVI), green reflectance (GR), grain fill period (GFP), test weight (TW), protein quality score (PQS), and solvent retention capacity (SRC). Within this cluster, PQS and SRC merged at the smallest correlation distance, consistent with the derivation of PQS from SRC (Chapter 3). The second

major cluster contained the remaining traits, including grain protein traits, kernel size traits, plant height, maturity traits, and all measured grain elemental concentrations.

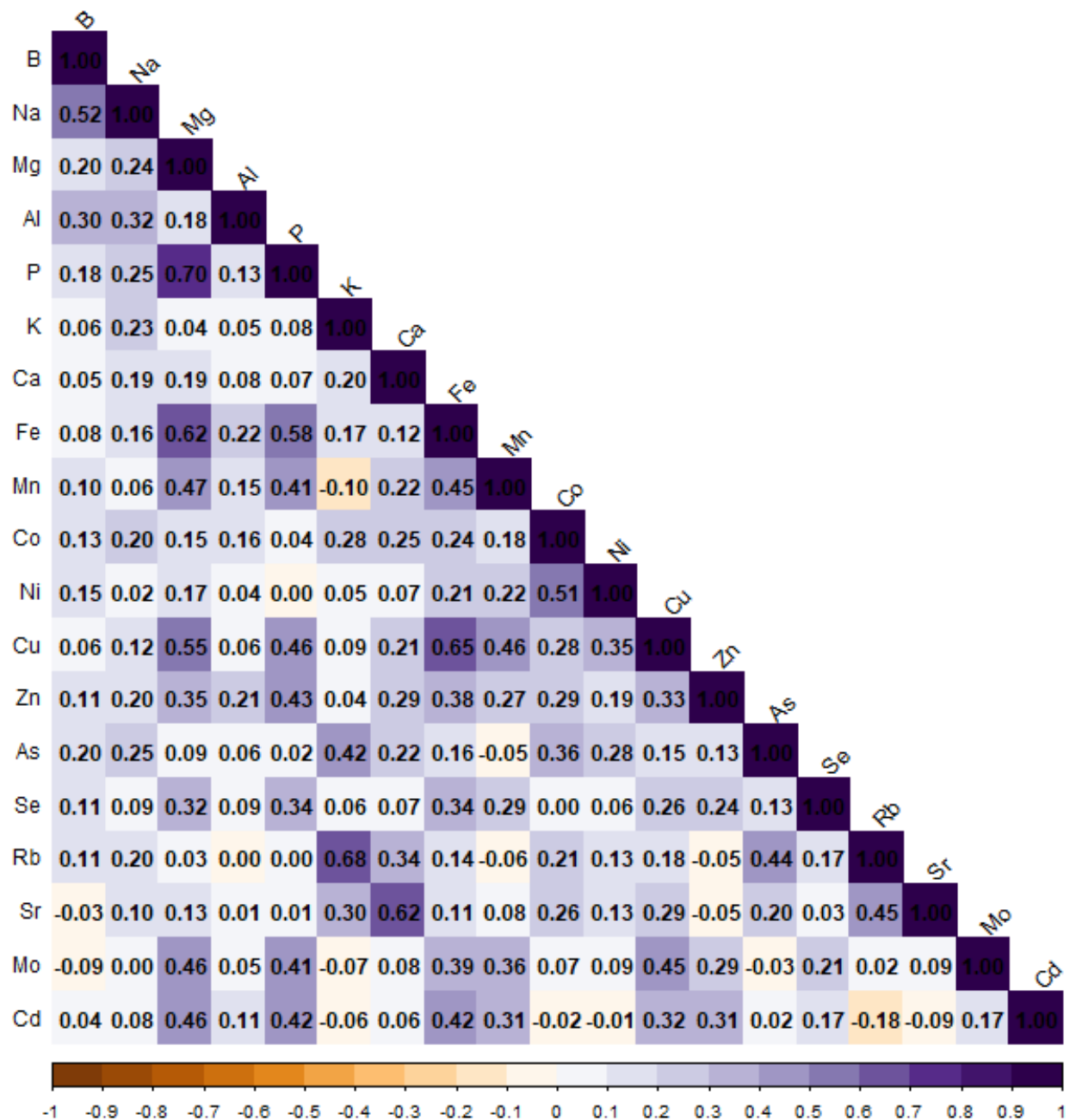


Figure 4.2. Pearson correlations among BLUPs for ionomics traits in TdGD Pop80 WILs

Values within each cell represent the correlation coefficient, with color indicating both magnitude (darker = stronger) and direction (orange = negative, purple = positive). Correlations were computed and visualized in R using the corrplot package (Wei & Simko, 2024).

When the dendrogram was partitioned at $k = 3$, GFP and GR separated from the broader GY-related cluster and formed a distinct module. These two traits were retained within the same major cluster as GY, TC, NDVI, TW, PQS, and SRC at $k = 2$, but they merged with that group only at a larger correlation distance. This indicates that GFP and GR shared some similarity with the broader GY trait group, but were less closely related to the other traits in that cluster than those traits were to one another.

At $k = 4$, the larger cluster containing grain protein, kernel morphology, plant height, maturity, and ionic traits separated into two modules. One group contained grain protein (GP), grain protein deviation (GPD), kernel size traits, and a subset of the ionome, including Se, Cd, Zn, Mo, Mn, Mg, P, Fe, and Cu. Kernel length (KL), kernel width (KW), thousand kernel weight (TKW), and kernel area (KA) merged at relatively small correlation distances, indicating the close relationships among kernel morphology traits. Grain protein and GPD also clustered within this group, as did several micro- and macronutrients that were positively correlated in the phenotypic analysis of the ionome (Figure 4.2).

The second group formed at $k = 4$ contained Al, B, Na, Co, Ni, Ca, Sr, As, K, Rb, plant height (PH), heading date (HD) and senescence date (SD). Within this cluster, K and Rb merged at a small correlation distance, as did Ca and Sr. Heading date and SD also merged with one another before joining PH and the surrounding elemental traits. Thus, this cluster contained both developmental traits and a subset of elements that showed more similar patterns of co-variation with maturity and PH than to the protein, kernel, and highly correlated micronutrient group.

When the dendrogram was partitioned at $k = 5$, Al, B, and Na separated from the PH, maturity, and partial ionome cluster to form a separate module. These three elements had their strongest ionic correlations with each other (Figure 4.2), although each was characterized by

low R^2_{across} . They also showed similar directions of environmental effects in each environment, with negative effects in 21THD and positive effects in 22 and 23THD (Table 4.1).

Overall, the dendrogram showed that ionomic traits did not cluster as a single unified trait class. Instead, different subsets of the grain ionome aligned with distinct portions of the broader Pop80 phenotypic structure (Appendix Figure C.4).

Genome-wide association

Table 4.2. GWAS results summary for ionomics using BLUPs

For each significant association, the chromosome (Chr), physical position in Mb (Pos), trait, model, MAF, and estimated marker effect are reported. Marker effects are relative to a standard normal distribution. Allelic information includes the allelic states of the RP and donor parents (Don), as well as the alleles associated with the lower (L) and higher (H) mean phenotypic values among genotype classes. Genotype_class columns report the number of WILs carrying the RP allele (WIL_RP) or the donor allele (WIL_Don) at the marker locus. Donor direction (DonDir) indicates whether the donor allele increases or decreases the trait value. Donor allele proportion within the 1 Mb genomic bin containing the marker (DonProp) represents the average proportion of donor alleles across WILs in that genomic region. NA indicates cases where parental alleles could not be unambiguously assigned.

Chr	Pos	Trait	Model	MAF	Effect	Allele		Genotype_class				DonDir	DonProp
						RP	Don	L	H	WIL_RP	WIL_Don		
5A	572.40	As	BLINK	0.23	-0.13	C	T	C	T	177	53	Increase	0.21
1A	550.30	Ca	BLINK	0.11	-0.09	NA	NA	T	C	NA	NA	NA	0.20
5A	608.88	Ca	BLINK	0.44	0.08	C	T	C	T	103	129	Increase	0.24
5D	502.03	Ca	BLINK	0.10	0.12	C	T	T	C	200	13	Decrease	0.02
7A	30.13	Ca	BLINK	0.15	0.10	A	G	G	A	191	28	Decrease	0.09
5A	608.88	Ca	MLMM	0.44	0.07	C	T	C	T	103	129	Increase	0.24
4B	28.23	Cd	BLINK	0.09	-0.28	G	T	G	T	208	16	Increase	0.06
3D	29.89	Co	BLINK	0.37	-0.12	T	C	C	T	85	145	Decrease	0.30
3D	29.89	Co	MLM	0.37	-0.12	T	C	C	T	85	145	Decrease	0.30
3D	29.89	Co	MLMM	0.37	-0.13	T	C	C	T	85	145	Decrease	0.30
1A	323.90	Fe	BLINK	0.11	-0.22	A	G	A	G	195	15	Increase	0.14
2A	686.70	Fe	BLINK	0.36	-0.12	NA	NA	G	T	NA	NA	NA	0.02
2D	27.46	Fe	BLINK	0.10	0.22	T	C	C	T	193	9	Decrease	0.02
3D	31.69	Fe	BLINK	0.38	0.14	NA	NA	C	T	NA	NA	NA	0.31
4B	34.72	Fe	BLINK	0.06	-0.38	C	T	C	T	209	6	Increase	0.05
1A	62.48	Mg	BLINK	0.07	-0.19	NA	NA	A	G	NA	NA	NA	0.16
1A	323.90	Mg	BLINK	0.11	-0.12	A	G	A	G	195	15	Increase	0.14
4B	19.74	Mn	BLINK	0.22	-0.20	NA	NA	G	A	NA	NA	NA	0.10
4B	614.28	Mn	BLINK	0.06	-0.35	G	A	G	A	211	7	Increase	0.02
5A	502.58	Mn	BLINK	0.13	0.21	T	C	C	T	184	13	Decrease	0.11

5B	572.39	Mn	BLINK	0.10	-0.29	A	G	A	G	206	20	Increase	0.04
1A	508.84	Mo	BLINK	0.42	-0.14	T	C	T	C	130	94	Increase	0.32
2A	727.62	Mo	BLINK	0.47	-0.14	C	A	A	C	100	115	Decrease	0.37
3A	507.13	Mo	BLINK	0.12	0.22	A	G	G	A	195	17	Decrease	0.09
4A	662.20	Mo	BLINK	0.12	-0.24	T	C	T	C	194	16	Increase	0.02
4B	36.71	Mo	BLINK	0.28	-0.15	NA	NA	G	A	NA	NA	NA	0.10
6B	615.99	Mo	BLINK	0.08	-0.27	NA	NA	A	G	NA	NA	NA	0.23
7D	226.12	Mo	BLINK	0.10	-0.25	T	C	T	C	206	19	Increase	0.01
2B	244.88	Ni	BLINK	0.15	-0.19	G	A	G	A	192	29	Increase	0.18
2D	609.48	Ni	BLINK	0.45	-0.13	NA	NA	G	A	NA	NA	NA	0.01
7A	27.66	Ni	BLINK	0.12	0.17	NA	NA	A	G	NA	NA	NA	0.08
7A	266.64	Ni	BLINK	0.23	-0.15	NA	NA	C	T	NA	NA	NA	0.27
1D	10.54	Rb	BLINK	0.11	0.18	T	C	C	T	204	21	Decrease	0.12
2B	45.55	Rb	BLINK	0.05	-0.38	NA	NA	A	C	NA	NA	NA	0.31
2B	45.55	Rb	MLM	0.05	-0.36	NA	NA	A	C	NA	NA	NA	0.31
2B	63.85	Rb	MLM	0.05	-0.32	A	C	A	C	215	7	Increase	0.08
2B	65.23	Rb	MLM	0.05	-0.34	NA	NA	A	G	NA	NA	NA	0.34
2B	45.55	Rb	MLMM	0.05	-0.35	NA	NA	A	C	NA	NA	NA	0.31
7A	47.04	Se	MLMM	0.38	-0.05	NA	NA	G	A	NA	NA	NA	0.05
2B	63.85	Sr	BLINK	0.05	-0.23	A	C	A	C	215	7	Increase	0.08
7B	602.08	Sr	BLINK	0.05	-0.26	NA	NA	T	G	NA	NA	NA	0.03

Significant associations for ionomic traits frequently colocalized with loci identified for agronomic and kernel traits (Chapter 3). On the short arm of chromosome 2B, a high confidence association for Rb coincides with the identified significant locus for HD and SD, with BLUP associations for Rb spanning 45.55-65.23 Mb (Table 4.2; Figure 4.3). A marker at 63.85 Mb is also associated with Sr, and for all three traits (SD, Rb, Sr), the donor allele is associated with increased trait values (Table 4.2; Chapter 3).

A second region of overlap occurs on chromosome 3D, where a high confidence Co association at 29.89 Mb is detected across MLM, MLMM, and BLINK models (Table 4.2, Figure 4.4). This locus lies in close proximity to associations for TKW (30.24 Mb; Chapter 3) and Fe (31.69 Mb; Figure 4.5). The donor allele in this region is associated with decreased Co

and TKW, while the direction of effect for Fe is ambiguous in BLUPs but is negative in nearby BLUE associations (Appendix Table C.4).

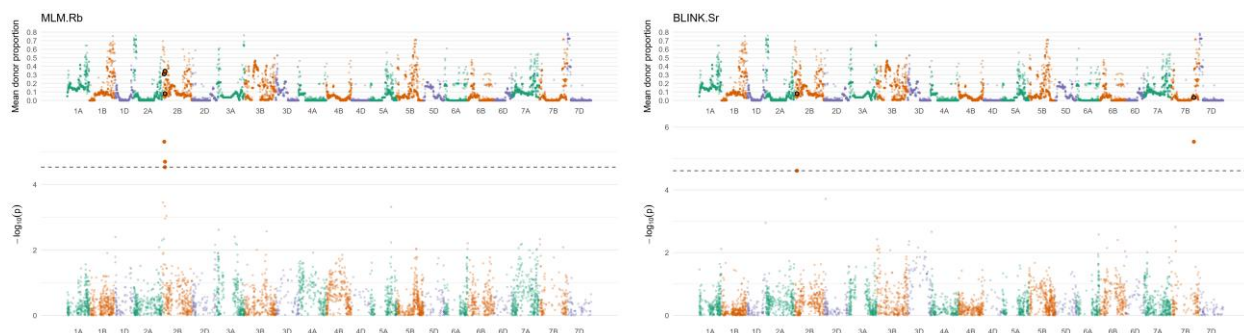


Figure 4.3. GWAS Manhattan plots for vascular transport proxy elements, overlaid with mean donor proportions in TdGD Pop80

Significant associations for Rb and Sr BLUPs are shown. In each panel, the upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

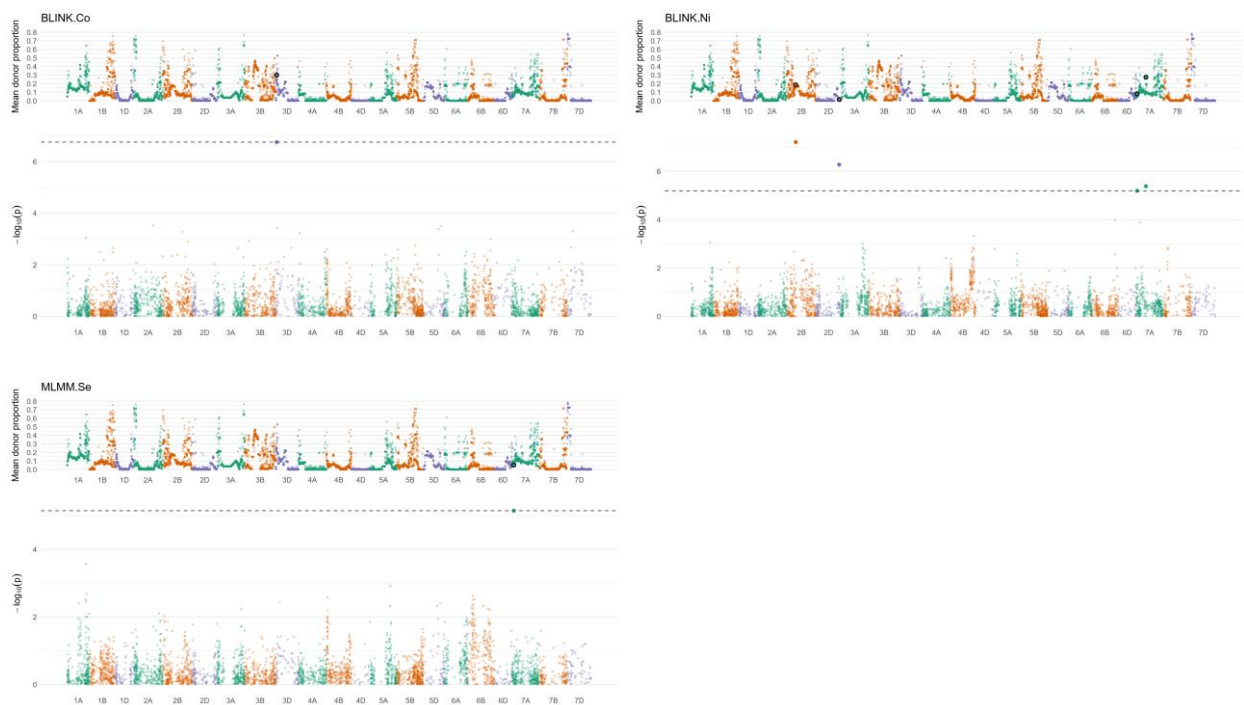


Figure 4.4. GWAS Manhattan plots for trace elements, overlaid with mean donor proportions in TdGD Pop80

Significant associations for Co, Ni, and Se BLUPs are shown. In each panel, the upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

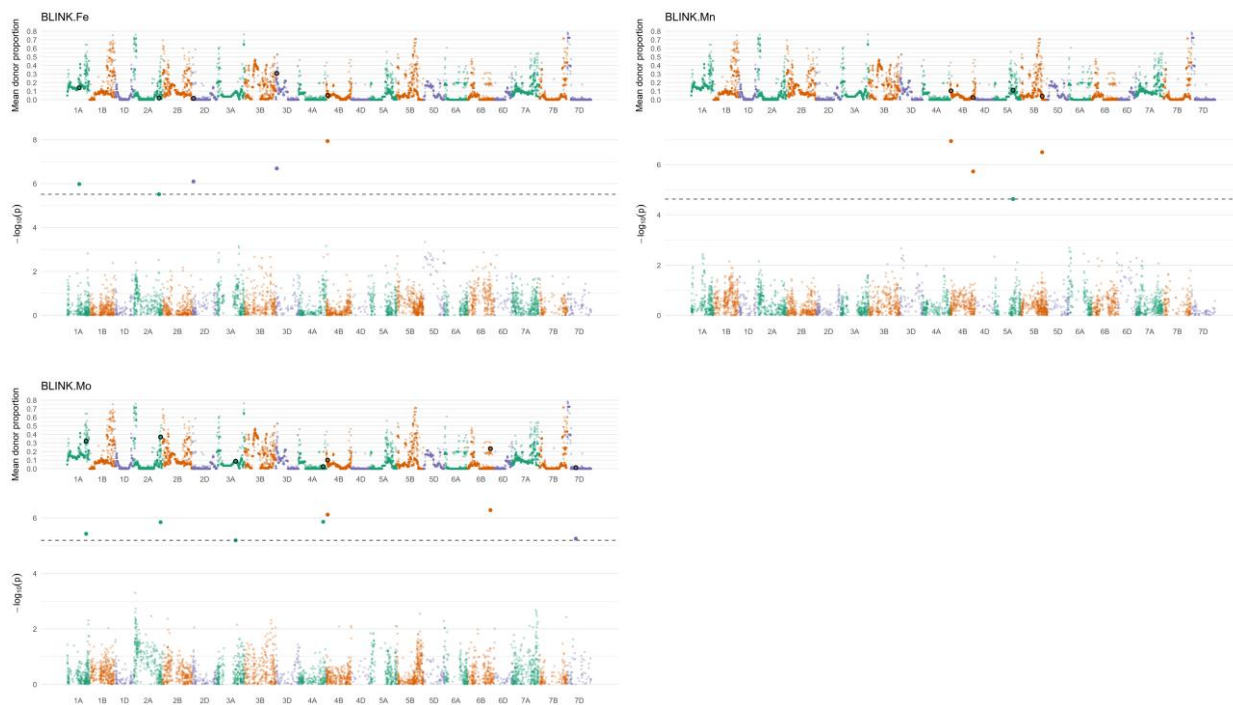


Figure 4.5. GWAS Manhattan plots for metabolic micronutrients, overlaid with mean donor proportions in TdGD Pop80

Significant associations for Fe, Mn, and Mo BLUPs are shown. In each panel, the upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

Pleiotropy is also observed on the short arm of chromosome 4B, encompassing the *Rht-B1* region (19.74-36.71 Mb). This interval contains BLUP associations for multiple agronomic, yield, kernel, and ionic traits, including PH, KL, NDVI, GY, Mn, Cd, Fe, and Mo (Table 4.2; Figures 4.5-4.6; Chapter 3). Environment-specific BLUEs further support this region, with Cu associations detected in 23THD and Se associations in 21THD at 38.79 Mb (Appendix Table C.4). On the long arm of chromosome 4B, an additional Mn BLUP association at 614.28 Mb – where the donor allele increases trait values – coincides with a region significant for KA, KW, TKW, TC, along with nearby Fe BLUE associations at 634.32 Mb (Table 4.2; Figure 4.5; Appendix Table C.4; Chapter 3).

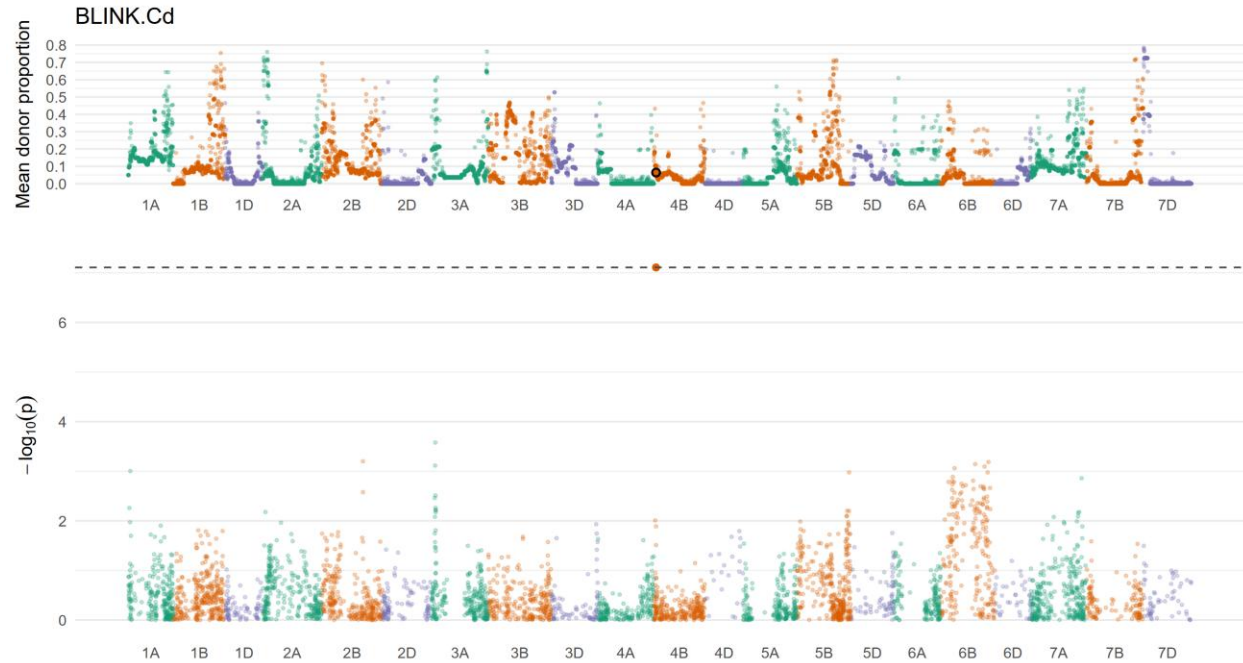


Figure 4.6. GWAS Manhattan plot for Cd BLUPs overlaid with mean donor proportion in TdGD Pop80

The significant Cd association detected by BLINK is shown. The upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

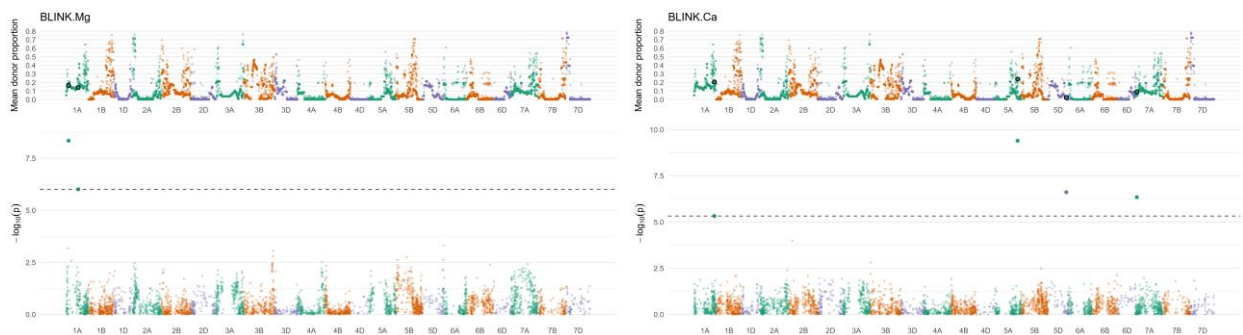


Figure 4.7. GWAS Manhattan plots for structural macronutrients, overlaid with mean donor proportions in TdGD Pop80

Significant associations for Mg and Ca BLUPs are shown. In each panel, the upper track displays the average donor allele proportion across WILs in 1 Mb bins along each chromosome, while the lower track shows the corresponding Manhattan plot of GWAS results. Horizontal dashed lines indicate the significance threshold determined using a false discovery rate (FDR) of 0.05.

Additional overlap between ionomic and other trait groups is evident on chromosome 2A, where Fe (686.70 Mb) and Mo (727.62 Mb) associations occur near a major kernel trait and GP

locus on the long arm (Table 4.2; Figure 4.5; Chapter 3). Molybdenum also exhibits pleiotropic associations on chromosomes 4A (662.20 Mb) and 6B (615.99 Mb), the latter coinciding with a region harboring nearby Fe signals in 21THD BLUEs (Table 4.2; Appendix Table C.4).

On chromosome 1A, a broad interval spanning 323.90-568.47 Mb contains BLUP associations for multiple elements, including Fe, Mg, Mo, and Ca, as well as BLUE associations for Sr and P in 23THD (Table 4.2; Figures 4.5, 4.7; Appendix Table C.4). This region also overlaps loci significant for GR and GR_milk BLUPs, along with stage-specific GR and NDVI BLUE associations (Chapter 3).

Additional ionomic associations are distributed across other chromosomes, including chromosome 7A (Ni at 27.66 Mb, Ca at 30.13 Mb, and Se at 47.04 Mb) and the distal long arm of chromosome 5A, where BLUP associations for Ca (608.88 Mb), Mn (502.58 Mb), and As (572.40 Mb) colocalize with regions controlling GR (overall and separated by growth stage) and kernel size traits (Table 4.2; Figures 4.4, 4.7; Chapter 3).

Discussion

Impact of wild emmer introgression on grain ionomic variation

The objective of this research was to determine whether amphiploid bridge-mediated introgression from wild emmer contributed useful variation for grain elemental traits in an adapted hexaploid wheat background. Overall, the results indicate that introgression altered grain ionomic composition, albeit with directional variation that cannot be interpreted as uniformly beneficial. WILs generally exceeded the RP and performance checks for a broad group of correlated elements, including Fe, Zn, Mn, Mg, Mo, Cu, Cd, and P. WIL distributions suggest that allelic variation increased the range of grain elemental concentrations available in this population, consistent with previous reports that wild emmer harbors substantial diversity for

grain nutritional traits (Cakmak et al., 2004; Chatzav et al., 2010; Distelfeld et al., 2007; Fatiukha et al., 2020b; Liu et al., 2021; Pandey et al., 2016; Peleg et al., 2008b, 2009a; Xie & Nevo, 2008).

The positive correlations among Fe, Zn, Mn, Mg, Mo, Cu, Cd, and P suggest that this portion of the ionome is influenced by shared physiological processes controlling nutrient uptake, remobilization, and grain deposition. From a breeding perspective, this is both useful and limiting. The presence of WILs with elevated Fe, Zn, Mn, Mg, and Mo demonstrate that the population contains transgressive variation for beneficial grain mineral traits; however, the correlated elevation of Cd and P means that simple selection for increased micronutrient concentration could inadvertently increase undesirable components. While Cd is directly toxic, P is being interpreted as an indirect indicator, rather than a direct measurement, of phytic acid. This pattern suggests that future selection would need to consider mineral concentration along with bioavailability, rather than treating elevated total concentrations as inherently favorable.

These results support the value of evaluating the ionome as a multivariate trait system. If the goal is biofortification or nutritional improvement, then the most useful WILs are not necessarily those with the highest absolute concentration of any single beneficial element. Instead, the more valuable materials may be those that partially decouple favorable and unfavorable relationships, such as WILs with increased Fe, Zn, Mn, or Mo without corresponding increases in Cd or excessive P. Identifying these favorable deviates is precisely where introgression populations may be most useful. Even if the average effect of introgression is not uniformly favorable, the expanded segregation created by wild-derived haplotypes increases the likelihood of recovering rare combinations of alleles that would otherwise be difficult to observe with precision in a set of germplasm with a narrower base of trait diversity.

In contrast to the broad group of metabolic and micronutrient elements, K/Rb and Ca/Sr formed distinct correlated pairs and showed a different relationship to broad effects of introgression. Rb and K were strongly correlated, as were Sr and Ca, consistent with their chemical similarities and their interpretation as proxy indicators of phloem and xylem transport, respectively (Seregin & Kozhevnikova, 2004; Shtangeeva et al., 2011; Zeller & Feller, 2000). WIL distributions for K and Rb were generally shifted downward relative to the checks, and Ca and Sr also showed evidence that many WILs accumulated less than the RP or highest checks in some environments.

The relative dearth in existing literature regarding grain Rb and Sr concentrations in wheat make these patterns particularly interesting. Nonetheless, Rb and Sr should not be interpreted as direct measurements of phloem and xylem flux; they are proxy elements whose accumulation may reflect transport, uptake tissue partitioning, and environmental availability. Therefore, the lower Rb/K and Sr/Ca patterns in WILs should be interpreted cautiously as indication of altered transport-associated elemental accumulation, not as direct proof of reduced vascular transport. Still, the separation of these elements from the Fe-Zn-Mn-Mg-Mo-Cu-P-Cd group suggests that different physiological or developmental processes shape different portions of the grain ionome.

Integration of the grain ionome with broader phenotypic performance

The full trait dendrogram provided important context for interpreting the grain ionome relative to the broader Pop80 phenotype set. Ionomeric traits did not cluster as a single group, instead, different subsets of elements aligned with different phenotypic modules. The highly correlated group containing Fe, Zn, Mn, Mg, Mo, Cu, P, Cd, and Se clustered with GP, GPD, and kernel size traits. This supports the interpretation that these elements are tied to grain

composition and kernel development rather than representing an independent nutritional phenotype. The association of this ionomic group with GP and kernel morphology may partly reflect variation in kernel tissue proportions. Because many mineral elements are concentrated in the aleurone and embryo, whereas the starchy endosperm constitutes most kernel volume, differences in the ratio between kernel surface area and volume, along with relative embryo size, could influence whole grain mineral and protein concentrations (De Brier et al., 2015, 2016; Singh et al., 2013). Image-based quantification of kernel shape and embryo area, ideally combined with cross-sectional imaging of aleurone, embryo, and endosperm proportions, could determine whether anatomical differences underlie these relationships.

A second ionomic group, containing K, Rb, Ca, Sr, As, Co, Ni, B, Na, and Al clustered more closely with PH and maturity traits. Within this group, K/Rb and Ca/Sr formed close subclusters, consistent with their interpretation as transport-associated elemental pairs. The clustering of HD and SD with this group suggests that developmental timing may influence final grain concentrations for these elements, either through effects on the timing and duration of nutrient uptake (Distelfeld et al., 2007; Uauy et al., 2006) or exposure to drought during key reproductive stages. This pattern is suggestive that phenology is correlated specifically with this portion of the ionome, especially under the drought stressed environments in which the WILs were evaluated.

The yield-related cluster containing GY, TC, NDVI, TW, PQS, and SRC was largely separate from ionomic traits. This separation is consistent with yield dilution and with the broader challenge of combining high yield and high mineral concentration (Guttieri et al., 2015b; Miner et al., 2022, 2024; Monasterio & Graham, 2000; Perrier et al., 2016). At the same time, the clustering of PQS and SRC with the yield-related cluster rather than with GP and the

metabolic ionome suggests that functional quality in this population was more closely aligned with drought-adapted performance and grain filling than with protein concentration alone (Singh et al., 2012; Yang et al., 2023). Thus, the dendrogram supports the major theme that performance cannot be interpreted on a single trait basis without considering correlated suites of phenotypes and the conditions under which they were collected.

Marker-trait associations for the grain ionome and colocalizations with related traits

The GWAS results further support the conclusion that grain elemental accumulation is integrated with broader developmental and physiological processes. Ionomic associations frequently colocalized with genomic regions associated with maturity, plant height, yield components, or functional quality. Because GWAS identifies marker-trait associations based on linkage disequilibrium, these overlaps cannot distinguish pleiotropy from linkage (Lander & Schork, 1994; Platt et al., 2010; Tibbs Cortes et al., 2020). However, the repeated coincidence of ionomic signals with other trait classes suggests that genomic regions in Pop80 may jointly influence developmental timing, biomass production, grain filling, kernel morphology, nutrient remobilization, and grain ionomic composition.

The chromosome 2B region near *Ppd-B1* is one of the clearest examples. Associations for Rb and Sr colocalized with maturity-related associations described in Chapter 3, and the donor allele at the shared Rb/Sr marker was associated with increased concentration. Since *Ppd-B1* affects photoperiod response and maturity timing, a conservative interpretation is that this region may influence Rb and Sr accumulation indirectly through altered developmental timing. Changes in heading or senescence can affect the timing of nutrient uptake, the duration of grain filling, and the environmental conditions experienced during grain deposition (Bagci et al., 2007;

Cakmak et al., 1996; Hu et al., 2006; Miner et al., 2024; Singh et al., 2012). Under drought, these timing differences may be especially important because water limitation constrains nutrient mobility and vascular transport. Therefore, the Rb/Sr signal near *Ppd-B1* is likely to reflect the interaction of phenology, transport-associated accumulation, and drought response. Notably, a similar Rb/Sr association has been reported previously (Yu et al., 2021). Interestingly, this 2B association for Rb and Sr provide a useful counterpoint to the overall distributional trend. Although many WILs had lower concentrations of these elements, that the donor allele for these associations conferred increased concentrations indicates that population-level effect of introgression and the locus-specific effect of particular donor alleles are not equivalent.

A similar interpretation applies to the 4B region encompassing *Rht-B1*. This interval was associated with PH, NDVI, GY, KL, and multiple ionomic traits, including Mn, Cd, Fe, Mo, and, in some environments, Cu and Se. The broad trait overlap is biologically plausible because PH affects canopy structure, biomass allocation, harvest index, grain filling, and nutrient remobilization (Bacher et al., 2021; Ihsan et al., 2016; Jobson et al., 2018; Liu et al., 2017; Peleg et al., 2005; Velu et al., 2017; Wang et al., 2017). From a breeding standpoint, the 4B region is important because it illustrates how loci already central to adaptation and plant architecture may also influence the ionome, either directly or through correlated effects on growth and grain development.

Chromosome 4B also contained a second notable region on the long arm, where a Mn association coincided with associations for KA, KW, TKW, and TC (Chapter 3). In addition to the hypothesis that this region affects biomass, kernel traits, and effective tiller number under heat stress during post-anthesis development (Chen et al., 2021; Deng et al., 2011; Guan et al.,

2018; H. Li et al., 2024; Xie et al., 2016), the Mn association on 4BL has also been reported previously (Bhatta et al., 2018).

The 5AL region provides another example of ionomic associations occurring within a broader performance hotspot. In this population, Ca associations on 5A colocalized with NDVI, GR, and kernel size traits described in Chapter 3. Although, the region overlaps with the major vernalization locus *Vrn-A1*, segregation at *Vrn-A1* was limited in Pop80 due to low minor allele frequency, suggesting that the observed 5AL associations may not be attributable solely to *Vrn-A1*. This is additionally supported by previous reports of separation of vernalization effects from biomass and yield components and proposition of additional candidate loci in the interval (Bag et al., 2026; Begum et al., 2020; Shi et al., 2017). Previous reports of Ca associations in the same general region, along with hypothesized cation-transporter candidates, further support the idea that 5AL represents a multi-trait hotspot affecting biomass, kernel development, and ion accumulation in concert (Alomari et al., 2017; Alqudah et al., 2025).

Another hotspot occurred on chromosome 6B, with a Mo association overlapping with multiple yield, kernel, and functional quality traits. Chapter 3 suggested this region as broadly influencing grain size and composition through nitrogen uptake or partitioning dynamics (Garcia et al., 2019; Habash et al., 2007; Katz et al., 2022; Manjunath et al., 2024; Muhu-Din Ahmed et al., 2020; Narjesi et al., 2015; Schmidt et al., 2020; Sukumaran et al., 2018). The Mo association supports this inference in the context of previous reports of pleiotropic elemental associations on 6BL (Joukhadar et al., 2025) and the documented role of Mo in nitrogen metabolism (Rana et al., 2025; Watanabe et al., 2022a, 2022b).

An additional genomic region of overlap with kernel traits and GP lies on chromosome 2AL, where BLUP associations were detected for Fe and Mo. The signal for Fe peaked close to

the GP signal, whereas the Mo signal peaked downstream of the remaining colocalized associations. The Mo association in this region has been hypothesized as a candidate for a molybdate transporter or molybdopterin biosynthesis protein (Jin et al., 2022). Additionally, the Fe association has been reported previously (Alomari et al., 2019; Bashir et al., 2025; Peleg et al., 2009a). *TaVITI-2A* is relevant to regional interpretation due to its role in Fe transport, although it maps substantially upstream of the Fe association detected here and is therefore less compelling as a direct positional candidate (Borrill et al., 2014; Connorton et al., 2017). Previous 2A Fe studies have also proposed NAC transcription factors or transmembrane proteins as candidates in nearby intervals (Alomari et al., 2019; Peleg et al., 2009a), and this is especially interesting because NAC-regulated remobilization could plausibly connect Fe, GP, senescence dynamics, and grain filling (Uauy et al., 2006).

Another ionomic hotspot occurred on chromosome 1AL, where Fe, Mg, Mo, and Ca associations colocalized. Prior literature supports chromosome 1A effects for Fe, Mg, and Ca (Cu et al., 2020; Fatiukha et al., 2020b; Ma et al., 2023; Nagi et al., 2025; Srinivasa et al., 2014), and 1A has also been implicated in grain phytic acid variation (Fatiukha et al., 2020b; Wen et al., 2022). Because phytic acid complexes several of these minerals, including Fe, Mg, and Ca, overlap between ionomic associations and previously reported phytic acid loci is biologically reasonable (Fatiukha et al., 2020b; Frittelli et al., 2023; Guttieri et al., 2015b; Peleg et al., 2009a). In contrast, the Mo signal on 1AL appears unsupported by previous reports and may represent a novel signal.

The distal long arm of chromosome 4A may carry another novel Mo signal, given the lack of prior documentation in terms of positional overlap. The colocalization of TW and KW in this region (Chapter 3), along with the repeated colocalization of Mo with kernel traits further

emphasizes the role of Mo in metabolic and physiologic processes contributing to grain development.

Chromosome 7AS contained another set of notable associations, with signals for Ni, Ca, and Se colocalizing with NDVI. Although the NDVI signal is consistent with the characterization of 7AS as a hotspot for biomass-related traits, the ionomic associations are not especially well documented, although Potapova et al. (2023) reported a nearby locus associated with multiple elemental concentrations, including Ca.

Collectively, these GWAS results indicate that the grain ionome in Pop80 can be localized to genomic regions already implicated in phenology, plant stature, canopy performance, kernel morphology, grain protein, or functional quality. This pattern is consistent with the phenotypic clustering results, where ionomic traits separated into different modules rather than clustering as one unified trait class. For breeding, the key implication is that selection for ionomic traits will likely have correlated effects on broader performance; therefore, marker-trait associations for the ionome should be interpreted in the context of multivariate variation rather than as independent nutritional loci.

Limitations and complications for interpretation

Several features of the grain ionomics data complicate interpretation. Elemental values were measured as concentrations, directly quantifying the amount of each analyte per unit grain mass. In that sense, they do not distinguish whether higher concentration reflects greater nutrient uptake, more efficient remobilization, altered partitioning to the grain, or reduced dilution by grain biomass. Therefore, targeted future research is warranted to better dissect the dynamics in individual WILs between elemental acquisition, grain yield, and kernel size. Additionally, collection of samples for ionomic profiling across different plant tissues and developmental

timing, although labor intensive, would posit a more thorough understanding of the aspects of the wheat ionome in Pop80.

Environmental sensitivity further complicates interpretation. Although restricting ionomic profiling to a single research site reduced broad geographic heterogeneity, the three years differed substantially in drought timing and intensity. The low R^2 observed for multiple elements, especially in the most drought stressed environment, highlights the way water availability can restrict nutrient mobility in the soil, reduce mass flow and diffusion to the root surface, and limit uptake or remobilization during reproductive development (Bagci et al., 2007; Cakmak et al., 1996; Hu et al., 2006; Miner et al., 2024; Singh et al., 2012). A useful contrast is the observation that the environment with the highest water availability during key growth stages, supporting greater yield potential, made yield dilution a plausible contributor to the generally negative environmental effects observed in that year. However, this pattern was not universal across the ionome, as Rb and Sr showed positive environmental effects in 23THD. Thus, final grain elemental concentrations likely reflect interacting effects of drought timing, grain biomass, developmental duration, nutrient availability, and element-specific transport or partitioning.

Some interpretations also rely on indirect indicators rather than direct physiological measurements. Total grain P was interpreted in relation to phytic acid given the large proportion of stored grain P as phytate (Cabas-Lühmann et al., 2023; Fatiukha et al., 2020b; Guttieri et al., 2015a; Miner et al., 2022; Peleg et al., 2009a), but phytic acid itself was not measured. Therefore, elevated P should be treated as a potential anti-nutritional concern rather than as direct evidence of increased phytate concentration. Similarly, Rb and Sr were interpreted as proxy elements for vascular transport, however, their concentrations do not directly measure

phloem or xylem flux. Their accumulation may reflect differences in uptake, transport, tissue partitioning, grain loading, or environmental availability. Consequently, the lower K/Rb and Ca/Sr patterns observed in many WILs should be interpreted as altered transport-associated elemental accumulation, not as direct evidence of reduced vascular transport.

The GWAS results should be interpreted with similar caution. Marker-trait associations identify genomic intervals correlated with phenotypic variation, not causal loci. Colocalization between ionic associations and loci for maturity, plant height, canopy traits, kernel morphology, grain protein, or functional quality may reflect pleiotropy, close linkage, shared haplotype structure, or correlated phenotypic responses. These overlaps are nonetheless meaningful in the context of characterized impacts of broader phenotypic structure on the ionome.

Future work should therefore target validation of the most promising phenotypic and genomic signals in ways that disentangle nutrient accumulation from yield dilution, correlated performance effects, and potential increases in toxic elements. Among the genomic regions identified in this research, the chromosome 2B associations appear especially worthy of further study. The region near *Ppd-B1*, including the overlapping maturity and Rb/Sr associations, raises particularly interesting questions about the relationship between developmental timing and vascular transport under drought. Future work could test whether the Rb/Sr association is stable across different drought scenarios, whether it reflects altered uptake versus altered remobilization, or whether it's merely a consequence of phenotypic correlation outside of direct regulation of grain concentrations. Finally, targeted evaluation of recombinants carrying favorable donor introgression at major multi-trait regions would help determine whether

beneficial ionomic effects can be retained without otherwise unacceptable penalties for maturity, yield, or kernel traits.

Breeding implications and conclusions

The results of this chapter indicate that amphiploid bridge-mediated introgression from wild emmer can expand grain ionomic variation in an adapted hexaploid wheat background. The breeding value of the ionomic variation observed in Pop80 lies less in identifying WILs with uniformly superior elemental profiles and more in demonstrating that wild emmer introgression can generate new, selectable combinations of grain elemental traits. The WILs expanded variation for several beneficial mineral elements, but this variation occurred within a correlated ionomic system and relationships with maturity, yield components, and functional quality. For breeding, the relevant target is a balanced profile combining increased beneficial elements with low toxic element accumulation, limited anti-nutritional risk, acceptable yield, and acceptable end-use quality.

The marker-trait associations identified here provide a starting point for that type of selection, but they are best viewed as prioritization tools rather than immediately deployable breeding markers. Additionally, contrasts in average distributional effects of wild introgression and locus-specific effects of individual donor alleles highlight the central breeding value of introgression populations as allowing unfavorable background effects to be broken apart from favorable donor alleles through recombination and selection.

Overall, this research demonstrates that wild emmer introgression altered the grain ionome in ways that are both promising and complex. Although the grain ionome cannot be interpreted as a single trait class or improved through simple directional selection, introgression provides useful selectable variation for identifying recombinant lines with ionomic profiles

outside of what is captured by a current set of elite breeding germplasm. Ultimately, the ionome should be treated as a multivariate, environmentally sensitive component of breeding value in the aim to combine favorable nutritional profiles with acceptable yield, regional adaptation, and grain filling dynamics.

Chapter 5 - Multi-trait index selection of amphiploid bridge-derived wild emmer introgression lines for wheat improvement

Abstract

Wheat wild relatives provide valuable allelic and trait diversity for breeding germplasm improvement, but ploidy barriers, yield disadvantage, and multivariate tradeoffs among quantitative traits constrain their use. This research evaluated whether yield-weighted multi-trait index selection could identify useful germplasm following introgression of wild emmer wheat (*Triticum dicoccoides*) using a wheat-*Aegilops tauschii* amphiploid bridge species. A population of 279 BC₂ introgression lines was evaluated across six rainfed Kansas environments for agronomic, yield and kernel component, functional quality, and ionomic traits. Across-environment performance for grain yield and seven complementary target traits were integrated using a modified grain yield by trait index that maintained primary emphasis on yield. Eighty-two introgression lines exceeded the elite recurrent parent, but only KSTdGD80-175 and KSTdGD80-55 exceeded all four performance checks, illustrating the rarity of favorable recombinant deviates. KSTdGD80-175 ranked first for both grain yield and the overall index and performed stably across environments. KSTdGD80-55 ranked second because favorable stature, tillering, kernel, protein, and functional quality characteristics complemented its competitive yield. Genomic characterization identified distinct wild emmer-derived A and B subgenome introgressions and *Ae. tauschii*-derived D subgenome introgressions in both lines. These results demonstrate the value of large populations and index selection for recovering rare, agronomically competitive introgression germplasm. The two lines provide characterized resources for targeted crossing and validation of wild relative variation across genetic backgrounds and environments.

Introduction

Sustained genetic gain in wheat depends on maintaining sufficient allelic diversity to respond to changing production environments and breeding priorities. Polyploidization, domestication, and continued selection have narrowed portions of the trait and allelic diversity available within modern bread wheat (*Triticum aestivum*, AABBDD) breeding germplasm, prompting interest in the deliberate use of wheat wild relatives for germplasm improvement (Akhunov et al., 2010; Fu & Somers, 2009; Haudry et al., 2007). Wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*, AABB) is particularly valuable because it shares the A and B subgenomes of bread wheat, permitting transfer of donor alleles through homologous recombination (Boehm & Cai, 2024; Hao et al., 2020; Nyine et al., 2020; Qi et al., 2007). Wild emmer has been investigated as a source of diversity for disease resistance, developmental timing, biomass production, kernel characteristics, grain protein, functional quality, mineral composition, and adaptation to environmental stress (Chen et al., 2005; Distelfeld et al., 2007; Dreisigacker et al., 2008; Fu et al., 2009; Fu & Somers, 2009; Golan et al., 2015; Knott et al., 2005; Lauterberg et al., 2022; Marais et al., 2005; Merchuk-Ovnat et al., 2016b; Nevo & Chen, 2010; Peleg et al., 2005; Peng et al., 2000a, 2013; Reader & Miller, 1991; Sun et al., 1997; Uauy et al., 2005; Wang et al., 2018; Xie & Nevo, 2008; Xu et al., 2004). Consequently, its value extends beyond the transfer of individual major effect loci to the broader expansion of quantitative variation available for selection.

The principal constraint to deploying wheat wild relative diversity is not necessarily the absence of valuable alleles, but the difficulty of converting those alleles into breeding value. Wild accessions lack the major domestication alleles conferring suitability for mechanical harvest and are not optimized for yield performance in target environments (Bohra et al., 2022;

Faris et al., 2003; Hao et al., 2020; Huang et al., 2003; Klindworth et al., 2013; Liu et al., 2006; Peleg et al., 2011; Peng et al., 2011; Simons et al., 2006; Tzarfati et al., 2014; Zhao et al., 2019). Introgression can also disrupt favorable linkage blocks and coadapted allele combinations assembled through long-term selection, while introducing linked donor alleles with undesirable effects. These challenges are especially relevant for quantitative traits, for which performance arises from many loci affecting developmental timing, biomass accumulation, resource partitioning, grain filling, and grain composition. The value of a donor allele may therefore depend on the recurrent background, the size and location of the surrounding introgression, and the environments in which it is evaluated (Huang et al., 2003; Klindworth et al., 2013; Lauterberg et al., 2022; Li et al., 2022; Merchuk-Ovnat et al., 2016a; Platten & Fritsche-Neto, 2023; Shahin et al., 2024). Effective utilization of wild relatives requires an intermediate pre-breeding phase in which donor diversity is recombined into adapted backgrounds, unfavorable characteristics are removed, and derived materials are characterized sufficiently to support subsequent selection and crossing.

Several crossing strategies can be used to introduce diversity from the wheat secondary gene pool, but each presents different biological and practical constraints. Direct crosses between tetraploid wild emmer and hexaploid wheat are possible, although the resulting progeny can experience meiotic instability because the D subgenome chromosomes lack homologous pairing partners (Price et al., 2025a). Alternatively, synthetic hexaploid wheat can be created by crossing tetraploid wheat with *Aegilops tauschii* (DD), thereby recreating the genomic composition of bread wheat, but this approach represents a substantial reset from the cumulative adaptation of elite hexaploid germplasm (Cox et al., 2017; Katamadze et al., 2023; Warburton et al., 2006). The strategy evaluated here instead used a wheat-*Aegilops tauschii* synthetic octoploid

amphiploid as a bridge. These amphiploids were originally developed by crossing adapted hexaploid wheat with *Ae. tauschii*, followed by embryo rescue and chromosome doubling to retain complete chromosome complements from both parents (Dale et al., 2017; Nyine et al., 2020). The amphiploids were subsequently crossed with wild emmer and the resulting progeny backcrossed to the elite hexaploid recurrent parent.

The distinctive feature of this approach is that it repurposes an amphiploid developed for D subgenome introgression as a vehicle for introducing wild emmer diversity on the A and B subgenomes. The duplicated D subgenomes of the octoploid theoretically reduce the meiotic imbalance associated with direct tetraploid x hexaploid crossing. At the same time, the approach enables retention of *Ae. tauschii*-derived D subgenome diversity alongside wild emmer-derived A and B subgenome introgressions. Thus, unlike conventional synthetic hexaploid development, the strategy begins with an adapted hexaploid background; and unlike direct wild emmer crossing, it uses the amphiploid to support recovery of fully hexaploid progeny while potentially introducing diversity across all three subgenomes. Direct crossing and amphiploid bridge-mediated approaches can produce broadly similar introgression landscapes (Chapter 2), but the bridge provides a biologically distinct route where suitable amphiploids have already been developed.

Previous characterization of the resulting *T. dicoccoides* germplasm development (TdGD) populations established that this strategy generated mosaic donor haplotypes in an otherwise largely recovered recurrent parent background (Chapter 2). Across seven populations, wild emmer introgressions were retained throughout the A and B subgenomes, with lower but detectable *Ae. tauschii* contributions on the D subgenome. One population was subsequently selected for comprehensive evaluation because of its size and initial agronomic performance.

Wild introgression lines (WILs) were evaluated across environments for agronomic performance, yield and kernel components, functional quality, and ionomic traits (Chapters 3, 4). These analyses showed that the WILs expanded phenotypic variation relative to the recurrent parent and performance checks and generated transgressive individuals for individual traits. However, introgression also produced a population-wide yield disadvantage, environment-dependent changes in maturity and grain filling, and tradeoffs among biomass production, kernel characteristics, grain quality, and elemental composition. These findings demonstrated that bridge-mediated introgression generated meaningful diversity, but also that the value of this diversity could not be determined from single traits or population means alone.

The practical question is therefore whether extensive characterization can be translated into selection of WILs that combine wild-derived diversity with sufficient trait performance for direct use in breeding. Favorable donor effects are expected to occur at relatively low frequencies because recombination must separate useful alleles from linkage drag while restoring coordinated performance across multiple quantitative traits. Large introgression populations increase the opportunity to recover such favorable recombinant deviates, but their identification requires a framework that accounts for correlated traits and prevents improvement in one characteristic from obscuring unacceptable performance in another. This is particularly important when grain yield remains the primary measure of adaptation but must be balanced with developmental timing, performance for individual yield component traits, and functional quality. Grain yield by trait selection indices provide one means of integrating these objectives while maintaining emphasis on yield performance (Yan & Frégeau-Reid, 2018). Complementary evaluation of individual trait profiles and retained donor introgression segments can then determine what additional diversity the selected materials make available for breeding.

The hypothesis for the present research is that backcrossing and selection following amphiploid bridge-mediated introgression from wild emmer would retain sufficient allelic and trait diversity while recovering competitive performance for multi-trait index selection to identify rare WILs with breeding value. Following the integrated multi-environment characterization of the target TdGD population, the objectives of this study were to (1) apply a yield-weighted multi-trait selection index to identify WILs combining competitive grain yield with favorable performance across complementary breeding targets; and (2) characterize the phenotypic profiles and retained wild emmer and *Ae. tauschii* introgression of selected WILs as accessible donor resources for wheat breeding germplasm improvement.

Methods

Germplasm development

The development of the TdGD populations is described in Chapter 2. Briefly, a wheat-*Aegilops tauschii* synthetic octoploid amphiploid was used as a bridge to facilitate introgression of wild emmer diversity into an adapted hexaploid wheat background. For population 80 (Pop80), the amphiploid was generated by crossing the hard red winter wheat line KS061278M-4 with *Ae. tauschii* accession TA 10155 (Panjakent, Tajikistan), followed by artificial chromosome doubling with colchicine. Amphiploid plants visually free of clump dwarfism were crossed as female parents with winter-habit wild emmer accession PI 471021, collected near Sanhedriyya, Israel, at 742 meters above sea level. The resulting F₁ progeny were backcrossed twice to KS061278M-4, which was used as the female parent in both backcross generations.

TdGD Pop80 has the pedigree KS061278M-4*2/3/KS061278M-4/TA 10155//PI 471021. The recurrent parent (RP), KS061278M-4, is an unreleased hard red winter wheat line developed in the Kansas State University wheat breeding program in Manhattan, KS, with the pedigree

Overley*3/Amadina/4/Seri.1B//Kauz/Hevo/3/Amadina/5/Fuller. The RP carries the 2NS translocation from *Aegilops ventricosa* on chromosome 2AS and was chosen because of the line's competitive yield performance and favorable end-use quality profile, providing an elite background in which the effects of the introgressed diversity could be evaluated.

Between 10-15 BC₂F₁ families were generated, and equal numbers of BC₂F₂ plants from each family were advanced through three generations of single seed descent. Although a population size of 500 WILs was targeted, attrition during population advancement resulted in 334 BC₂F_{4:5} WILs. These lines were increased in headrows in Yuma, AZ, during 2020, with concurrent greenhouse increases used for leaf tissue collection. Lines with greater than 50% kernel retention in the spike were removed because of poor threshability. Subsequent selection against rachis fragility, inadequate seed production, and poor winter survival produced a final set of 279 WILs for multi-environment evaluation.

Field evaluation and selection

The field trial design and management are described in detail in Chapter 3. Briefly, the 279 WILs were evaluated in 50% partially replicated small-plot trials across six Kansas environments from 2021-2023: Thomas County in 2021 (21THD); Ellis (22EL), Riley (22RL), and Thomas (22THD) counties in 2022; and Riley (23RL) and Thomas (23THD) counties in 2023. Plots consisted of six rows with a planted length of 2.44 meters (m) and planted area of 3.83 m².

The partially replicated designs were generated using the R package DiGger (Coombes, 2020). Half of the WILs were evaluated in one plot per environment and half in two plots. Trials were organized into 10 blocks, each containing the RP and the performance checks “Billings” (PI 656843; Hunger et al., 2014), “Joe” (PI 676270; Zhang et al., 2016), “Showdown” (PI

690085), and “SY Monument” (PI 672581). WILs assigned two plots were distributed across two major field replicates. The six rainfed environments differed in the timing and severity of water limitation, with relative drought severity characterized as $22EL \approx 22THD > 23RL \approx 21THD > 22RL \gtrsim 23THD$ based on precipitation and visual observation of trial health (Chapter 3).

Phenotypic characterization

Phenotyping procedures are described in detail in Chapters 3 and 4. Traits related to agronomic performance included heading date (HD), senescence date (SD), grain fill period ($GFP = SD - HD$), and plant height (PH). Yield and kernel component traits included tiller count (TC), normalized difference vegetation index (NDVI), green reflectance (GR), grain yield (GY), test weight (TW), thousand kernel weight (TKW), and kernel area, length, and width (KA, KL, and KW). Grain yield was recorded from the onboard weighing system of a Zürn small plot combine. Test weight and grain protein concentration (GP) were measured using a Perten IM9500 (PerkinElmer) instrument. Kernel traits were measured with a Marvin Kernel Analyzer (MARViTECH). Additional functional quality traits included grain protein deviation (GPD), calculated as the residual from a regression of GP on GY (Haley & Stromberger, 2018; Monaghan et al., 2001; Oury & Godin, 2007), and whole grain solvent retention capacity (SRC) and protein quality score (PQS), determined from a lactic acid SDS-SRC assay (Seabourn et al., 2012; Wilson et al., 2020). Ionomics profiling was conducted on all Thomas County grain samples across years using inductively coupled plasma mass spectrometry (ICP-MS) in the Baxter Laboratory at the Donald Danforth Plant Science Center, St. Louis, MO (Asaro et al., 2016; Ziegler et al., 2017).

Statistical analysis

The two-stage mixed model analysis is described in Chapter 3. Plot-level observations were standardized within each environment to a mean of 0 and standard deviation of 1 prior to analysis. This placed traits on a common scale and improved comparability both among traits and among environments in which measurement and trial management differed (Chapter 3). Consequently, positive values indicate performance above the corresponding trial mean and negative values indicate performance below the trial mean.

In the first stage, each trait was analyzed separately within environment to estimate genotype best linear unbiased estimates (BLUEs) and their sampling variances. Genotype was fitted as a fixed effect. Row and column field variation was modeled using an AR1 x AR1 residual covariance structure, except in 22THD, where two-dimensional P-spline smoothing was used because the residual variation did not conform adequately to the AR1 x AR1 structure. Analyses were conducted using Mr. Bean, which interfaces with ASReml-R and SpATS (Aparicio et al., 2024; Butler et al., 2023; Rodríguez-Álvarez et al., 2018). Precision weights were calculated as the inverse of the sampling variance of each BLUE.

In the second stage, BLUEs and their precision weights were used to estimate across-environment genotype performance following the two-stage framework (Möhring & Piepho, 2009; Smith et al., 2001, 2005). Environment was fitted as a fixed effect, whereas genotype and genotype x environment (GxE) interaction were fitted as random effects. The residual scale was fixed so that the precision weights determined the relative contribution of each BLUE. The resulting across-environment best linear unbiased predictions (BLUPs) were used for multi-trait index construction and ranking trait performance within Pop80.

Grain yield by trait selection index

A yield-weighted multi-trait selection index was developed by adapting the genotype grain yield x trait (GYT) framework of Yan and Frégeau-Reid (2018). The original framework combines yield and trait values through multiplication. Direct multiplication was not appropriate here because the standardized values were centered on zero; multiplying two negative values could produce a positive index component even when both GY and the accompanying trait were below their respective population means. Therefore, standardized GY and trait BLUPs were combined additively.

Seven traits were selected to represent complementary components of breeding performance: GFP, PH, TC, TKW, TW, GP, and SRC. Longer GFP, greater TC, TKW, TW, GP, and SRC, and shorter PH were defined as favorable directions. For genotype i and accompanying trait j , the component index was calculated as:

$$C_{ij} = G_i + d_j T_{ij},$$

where G_i is the standardized across-environment GY BLUP, T_{ij} is the standardized BLUP for trait j , and d_j defines the desired direction of selection. The value of d_j was +1 for GFP, TC, TKS, TW, GP, and SRC and was -1 for PH. The overall GYT index was the unweighted mean of the seven component indices:

$$GYT_i = \frac{(G_i + GFP_i) + (G_i - PH_i) + (G_i + TC_i) + (G_i + TKW_i) + (G_i + TW_i) + (G_i + GP_i) + (G_i + SRC_i)}{7}$$

Per this formulation, high GY remained the primary selection criteria, but genotypes were rewarded or penalized according to their broader agronomic, kernel, and functional quality profiles. All WILs, the RP, and performance checks were ranked according to the across-environment index. KSTdGD80-175 and KSTdGD80-55 were selected for detailed characterization because they were the first and second-ranked genotypes, respectively, in the

overall index. Their index rankings were subsequently interpreted alongside traits not included in the index and retained donor introgressions.

Equivalent indices were calculated within individual environments using standardized BLUEs. Because not every component trait was measured in every environment (Chapters 3, 4), these indices contained different subsets of traits and therefore did not represent identical composite measures. Consequently, within-environment GYT indices were used primarily for comparison – to evaluate consistency of WIL rankings across versus within environments – rather than as independent selection metrics.

Genotypic characterization and introgression profiling

Genotyping, imputation, and donor introgression mapping are described in Chapter 2. Briefly, DNA extracted from BC₂F_{4.5} WILs was genotyped using the Illumina Infinium Wheat/Barley 40 K SNP array (Keeble-Gagnère et al., 2021). Population founders were deeply sequenced, with KS061278M-4 and TA 10155 subjected to whole genome sequencing (BioProject ID: PRJNA745927; Nyine et al., 2021) and PI 471021 characterized by exome capture (BioProject ID: PRJNA503847; Avni et al., 2017). Founder sequence information and WIL array genotypes were integrated through pedigree-based imputation to obtain genome-wide marker coverage for introgression analysis (Whalen et al., 2020). Allelic donor proportions were calculated in 1 Mb bins across the genome for each WIL. Donor introgression segments were then identified using a moving average approach, with segments called when the average donor proportion across the five flanking bins on either side of a focal bin was ≥ 0.9 or when the average donor proportion across the 20 flanking bins on either side was ≥ 0.7 . Introgression calls for KSTdGD80-175 and KSTdGD80-55 were used to identify the wild emmer-derived A and B subgenome segments and *Ae. tauschii*-derived D subgenome segments retained by each WIL.

Genome-wide association studies (GWAS) of phenotypic traits were previously conducted within TdGD Pop80, as described in Chapters 3 and 4. At significant GWAS markers, allele states in KSTdGD80-175 and KSTdGD80-55 were classified and evaluated for physical overlap with a called donor introgression segment and for agreement between the estimated direction of the effect of the donor allele and the corresponding trait value of WIL performance. Comparisons outside called introgression segments were retained to account for localized donor alleles that did not meet the moving average criteria for segment identification. These comparisons were used as supporting evidence of potential relationships between retained wild relative diversity and the phenotypic profiles of the selected WILs, rather than as confirmation of causal effects.

Results

Detailed trait-specific performance of TdGD Pop80 was reported in Chapters 3 and 4. Briefly, characterization of the population identified variation across screened traits, while multivariate structuring among these traits indicated that performance reflected correlated suites of characteristics rather than independent responses. Although favorable and transgressive values were observed for individual traits, the WILs exhibited an overall GY disadvantage relative to the RP and adapted checks. Consequently, population-wide variation and single trait performance alone did not clearly identify which WILs offered the greatest utility for wheat breeding.

The present analysis therefore prioritized a yield-weighted multi-trait index to translate the population-level characterization into selection of individual WILs. The index maintained GY as the primary selection criterion while incorporating other breeding targets as complementary sources of information. This approach was used to identify rare WILs that

combined competitive GY with favorable performance across multiple traits despite the broader population-level yield disadvantage.

Grain yield by trait selection index

Using standardized BLUPs for GY and the traits GFP, PH, TC, TKW, TW, GP, and SRC, a yield-weighted selection index was constructed to summarize multi-trait performance and identify competitive introgression germplasm. The GYT index differentiated a small subset of WILs combining high GY with favorable performance across the traits included in the index (Figure 5.1). Of the 279 WILs, 82 exceeded the RP in the across-environment index, whereas 197 ranked below the RP. The RP ranked 90th overall. Two WILs, KSTdGD80-175 and KSTdGD80-55, exceeded all performance checks, with index values of 1.52 and 1.42, respectively (Table 5.1; Figure 5.1). Showdown was the highest-ranking check, with an index value of 1.06, followed by KSTdGD80-130 at 0.95.

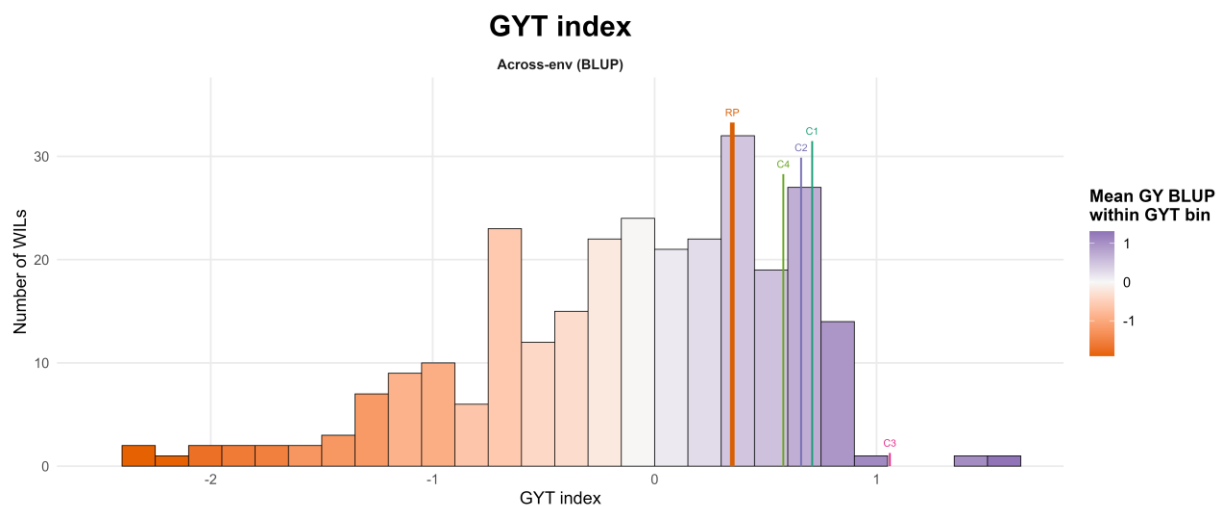


Figure 5.1. Distribution of the BLUP GYT index among TdGD Pop80 WILs

Histogram bins are colored by the mean standardized GY BLUP of WILs within each bin. Vertical lines indicate the recurrent parent (RP=KS061278M-4) and performance checks (C1=Billings, C2=Joe, C3=Showdown, C4=SY Monument). Plots were generated in R using the ggplot2 package (Wickham, 2016).

KSTdGD80-175 ranked first in both the overall index and the BLUP GY analysis. The standardized GY BLUP of 1.31 for KSTdGD80-175 was accompanied by favorable index components for GFP, PH, TW, and SRC. KSTdGD80-55 ranked second in the index despite having lower GY than KSTdGD80-175, Showdown, and several other WILs. Relative to KSTdGD80-175, the lower GY of KSTdGD80-55 was partially compensated by shorter PH, greater TC and TKW, and higher GP. KSTdGD80-130 illustrates the distinction between selection based on GY alone and selection based on the broader trait profile; although GY of KSTdGD80-130 exceeded that of KSTdGD80-55, its taller stature and lower SRC reduced its overall index ranking.

Table 5.1. Subset of grain yield by trait (GYT) index results for RP, performance checks, and top three WILs

Results are ordered by the overall GYT index. Values shown for GY represent standardized BLUPs. Individual GY $\pm$ trait indices were calculated by combining standardized GY with standardized values of each trait, either through addition or subtraction depending on the desired direction of improvement. The overall GYT index was calculated as the mean of the individual GY $\pm$ trait indices for each genotype.

Genotype	GY	GY $\pm$ trait indices							GYT _{index}
		GY + GFP	GY - PH	GY + TC	GY + TKW	GY + TW	GY + GP	GY + SRC	
KSTdGD80-175	1.31	1.88	2.13	1.36	0.34	2.75	0.39	1.79	1.52
KSTdGD80-55	1.05	1.32	2.09	1.25	0.57	2.17	0.81	1.72	1.42
Showdown (C3)	1.08	1.66	0.71	1.07	1.59	0.86	-0.15	1.66	1.06
KSTdGD80-130	1.09	1.29	0.46	1.11	0.80	1.71	0.51	0.80	0.95
Billings (C1)	0.20	0.81	0.47	-0.28	2.08	0.90	-0.48	1.48	0.71
Joe (C2)	0.80	1.28	0.11	1.60	1.36	0.94	-0.52	-0.11	0.66
SY Monument (C4)	0.70	-0.13	0.51	1.32	0.90	0.24	-0.47	1.68	0.58
KS061278M-4 (RP)	0.43	0.27	0.61	0.36	0.29	0.75	-0.09	0.22	0.35

Multi-trait performance of the two highest-ranking WILs was generally consistent across environments. KSTdGD80-175 never ranked below sixth in any within-environment index constructed using BLUEs (data not shown). KSTdGD80-55 ranked first in 22THD, second in 21THD, fourth in 22RL, eighth in 23THD, and fifteenth in 22EL. The lowest ranking was 47th in

23RL, where KSTdGD80-55 had reduced GFP and TKW relative to its performance in the other environments. In comparison, the within-environment index rankings for the RP ranged from 34th in 23RL to 145th in 22THD.

Based on the results from the GYT index, KSTdGD80-175 and KSTdGD80-55 were prioritized as pre-breeding material offering the greatest immediate value for wheat germplasm improvement. The remaining results therefore focus on complementary agronomic, yield, quality, and ionomic characteristics of KSTdGD80-175 and KSTdGD80-175 and the donor-derived introgressions they retain across the A, B, and D subgenomes.

Characterization of KSTdGD80-175 and KSTdGD80-55

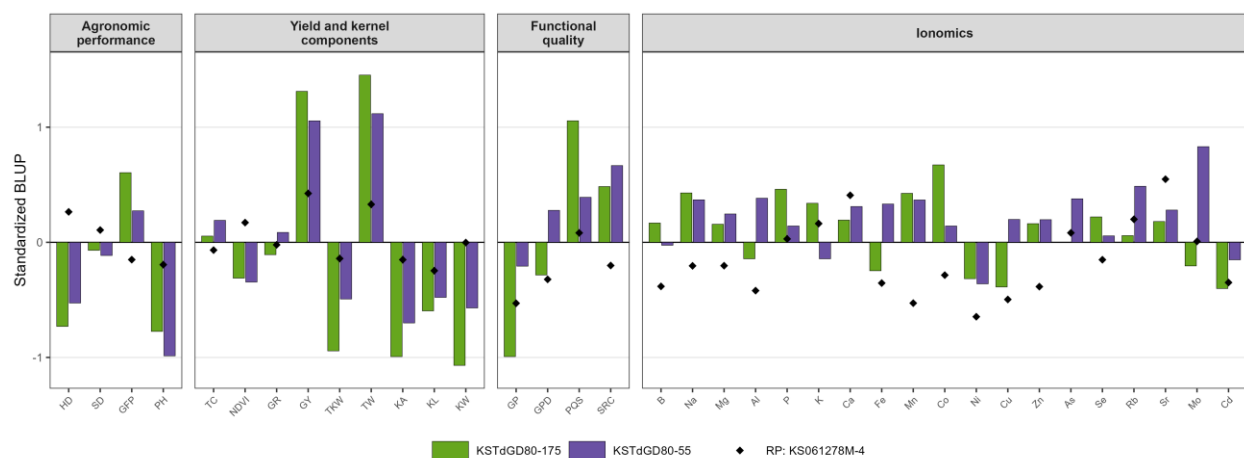


Figure 5.2. Multi-trait profiles of KSTdGD80-175 and KSTdGD80-55

Bars represent standardized across-environment BLUPs for each WIL, and black diamonds indicate the recurrent parent (RP; KS061278M-4). Positive and negative values indicate performance above and below the population mean, respectively. Traits are grouped by agronomic performance, yield and kernel components, functional quality, and grain ionomic composition.

Agronomic performance

KSTdGD80-175 and KSTdGD80-55 are early- to medium-maturing lines characterized by earlier HD and SD than the RP and performance checks (Figure 5.2; Appendix Table D.1). Relative to the population, HD was proportionally earlier than SD in KSTdGD80-175 and KSTdGD80-55, as highlighted by the favorable grain filling periods maintained by both WILs.

KSTdGD80-55 had a longer GFP than the RP and SY Monument (Appendix Table D.1). The GFP of KSTdGD80-55 was especially notable in the two most drought stressed environments, 22EL and 22THD, where the GFP of KSTdGD80-55 exceeded the GFP of all checks and the RP (data not shown). This pattern suggests that KSTdGD80-55 maintains grain fill under drought stress and supports the line's value as a potential source of adaptation to drought stressed environments.

KSTdGD80-175 had a longer GFP than all performance checks and the RP, with the exception of Billings (Appendix Table D.1). Across environments, KSTdGD80-175 generally maintained a longer GFP than KSTdGD80-55, with an especially extended GFP in 22RL and 23RL (data not shown). Relative to KSTdGD80-55, KSTdGD80-175 is therefore distinguished by extended GFP across all environments, whereas KSTdGD80-55 is distinguished by its stability of GFP in more severe drought stress.

Relative magnitudes of lodging resistance were based upon observational field notes, with no formal data being collected for analysis. Based upon these observational notes, KSTdGD80-55 and KSTdGD80-175 both exhibited good lodging resistance. Lodging events occurred primarily mid-season and affected plots showed near complete recovery during grain filling. Both WILs were shorter than the RP and all performance checks (Figure 5.2; Appendix Table D.1). This reduced stature likely contributed to their good standability under storm events.

Yield and kernel components

KSTdGD80-175 and KSTdGD80-55 both exhibited strong GY performance, with KSTdGD80-175 showing the highest across-environment GY BLUP among all evaluated entries (Figures 5.1, 5.2). KSTdGD80-175 consistently matched or exceeded the RP and checks in individual environments, other than Showdown in 21THD, and was distinguished by its

exceptional and stable yield performance across the full range of environments (Appendix Table D.2). KSTdGD80-55 also performed strongly for GY, with especially favorable performance in western Kansas environments at THD. Although KSTdGD80-55 had the weakest relative performance observed in 22EL, KSTdGD80-55 still exceeded the RP and Billings within each environment (Appendix Table D.2). Both WILs consistently and substantially outperformed the RP, demonstrating favorable GY performance across the drought stressed environments evaluated.

Neither WIL exhibited large kernel size; both KSTdGD80-175 and KSTdGD80-55 had lower BLUPs than the RP and all checks for kernel traits (TKW, KA, KL, and KW; Figure 5.2; Appendix Table D.3). In contrast, both WILs had higher TW than the RP and all performance checks, supporting an increase in bulk density arising from smaller kernels.

The two WILs also differed from the RP and checks in other yield component traits. Both KSTdGD80-55 and KSTdGD80-175 had higher TC than the RP, Billings, and Showdown, but lower NDVI than the RP and all checks (Figure 5.2; Appendix Table D.3). KSTdGD80-55 exceeded the RP, Joe, and Showdown for the trait GR, whereas KSTdGD80-175 exceeded Joe and Showdown.

Functional quality

Both KSTdGD80-55 and KSTdGD80-175 exhibited higher functional quality compared to the RP (Figure 5.2; Appendix Table D.4). KSTdGD80-175 had lower GP than the RP, but this reduction was accompanied by improved SRC and a substantially higher PQS. The combination of strong GY gains with reduced GP resulted in a negative GPD, but was still slightly better than that of the RP, indicating that reduced GP in KSTdGD80-175 did not result in poorer functional

quality. In general, the quality profile of KSTdGD80-175 was comparable to the quality profile of Showdown.

KSTdGD80-55 also exhibited better end-use quality compared to the RP, with higher PQS and substantially higher SRC (Figure 5.2; Appendix Table D.4). Of the two WILs, KSTdGD80-55 was distinguished by the highest GP and GPD values, exceeding the values of all performance checks and the RP. Although SRC in KSTdGD80-55 was slightly higher than Showdown, its PQS was lower because of its higher GP. Overall, these results indicate that both lines exhibited better functional quality than the RP, despite KSTdGD80-175 having reduced GP. Notably, KSTdGD80-55 represented improvements across all grain quality traits.

Ionomics

Both KSTdGD80-55 and KSTdGD80-175 generally showed better grain nutrient profiles than the RP, although the two WILs differed in their respective balances between favorable and toxic elemental accumulation (Figure 5.2; Appendix Tables D.5-D.7).

Both WILs had higher Na, Mg, and P concentrations than did the RP, whereas Ca was lower in both WILs relative to the RP (Figure 5.2; Appendix Table D.5). Both WILs had higher Mg and P than did the performance checks. KSTdGD80-175 also had K concentration above that of the RP, whereas KSTdGD80-55 had the lowest K concentration among the RP and checks. Averaged across macronutrient traits, both WILs generally exceeded the RP, although Showdown had a stronger overall macronutrient profile than did KSTdGD80-55.

KSTdGD80-175 was distinguished by elevated concentrations of B, Mn, Co, Zn, and Se, and also had the highest values, compared to the RP and checks, for B, Mn, Co, and Se (Figure 5.2; Appendix Table D.6). KSTdGD80-55 was distinguished by elevated Fe, Cu, and Mo, exceeding the respective values of the RP and all checks. Both WILs had higher average

micronutrient and trace element concentrations than the RP, which showed comparatively low values across this group of elements. This pattern is particularly notable for Fe and Zn, which are important targets in wheat biofortification (Carvalho & Vasconcelos, 2013; Guttieri et al., 2015b; Pandey et al., 2016); both WILs had elevated Zn, and KSTdGD80-55 also had elevated Fe. KSTdGD80-175 had notably low Rb concentration, a proxy element to trace xylem transport (Vallejo et al., 2005; Zeller & Feller, 2000).

The improved nutrient profiles of the two WILs differed, however, in their association with potentially undesirable elements. KSTdGD80-55 had the highest Al concentration, relative to the RP and the performance checks, and also had higher Cd than did the RP (Figure 5.2; Appendix Table D.7). Notably, average grain Cd concentrations in this dataset were low enough to be outside the range of practical relevance (Appendix Table D.7; Guttieri et al., 2015a, 2015b). Among the comparison entries, only Showdown had a higher average concentration of toxic elements than KSTdGD80-55. By contrast, KSTdGD80-175 had reduced As and Cd relative to the RP, although Al for the line was relatively high. KSTdGD80-175 also had lower As compared to all performance checks. Overall, KSTdGD80-175 showed a more favorable balance between elevated beneficial nutrient concentrations and lower accumulation of toxic elements, whereas KSTdGD80-55 combined elevated beneficial nutrient concentrations with less favorable accumulation of toxic elements.

Genomic introgression

KSTdGD80-55 and KSTdGD80-175 both represent novel sources of allelic diversity for wheat breeding programs, with identified introgressions from both wild emmer and *Ae. tauschii*. KSTdGD80-55 carries wild emmer introgressions on chromosomes 1B, 3A, 3B, 4B, and 7A, and *Ae. tauschii* introgressions on chromosomes 3D and 7D (Figure 5.3). The largest introgression in

KSTdGD80-55 occurs on chromosome 3D, spanning much of the short arm and centromeric region, whereas the 7D introgression is also substantial, extending approximately 63 Mb. The smallest introgressions are located near the distal ends of the long arms of chromosomes 1B and 3A. Within subgenomes, average donor proportions in KSTdGD80-55 were 6.2% for the A subgenome, 5.1% for the B subgenome, and 8.9% for the D subgenome.

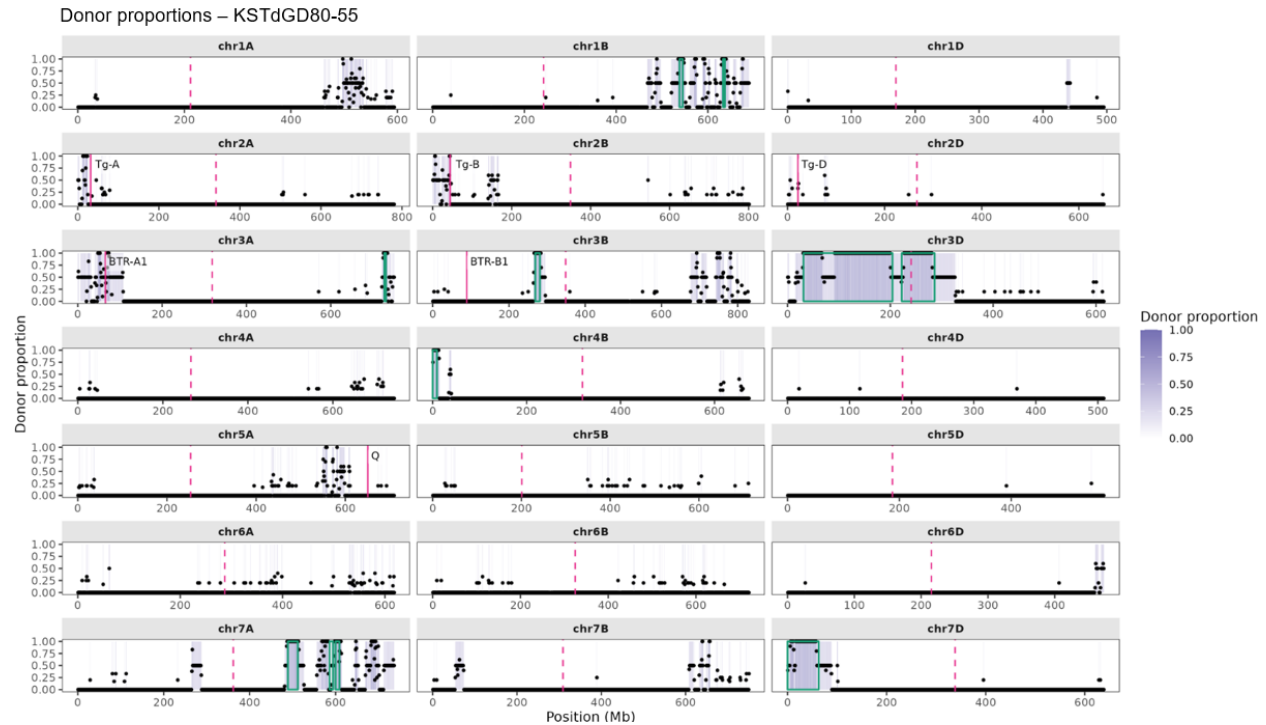


Figure 5.3. Genome-wide introgression profile for KSTdGD80-55

Donor allele proportions were calculated in 1 Mb bins and are shown as points along each chromosome facet. Vertical dashed lines indicate centromere positions based on the Chinese Spring v1.0 reference. Major domestication loci – *Q* (free threshing), *Btr* (brittle rachis), and *Tg* (tenacious glume) – are indicated by labeled vertical markers at their physical positions. Vertical boxes within chromosome panels denote regions where introgression segments were called using a moving average framework ($MA_{10} \geq 0.9$ and $MA_{40} \geq 0.7$).

KSTdGD80-175 carries wild emmer introgressions on chromosomes 1A, 1B, 2A, 3A, 5B, 6B, and 7A, and *Ae. tauschii* introgressions on chromosomes 2D, 3D, and 5D (Figure 5.4).

As in KSTdGD80-55, the largest introgression occurs on chromosome 3D, where a donor segment extends to approximately 64 Mb and spans the centromeric region. The smallest introgressions are located on the long arms of chromosomes 1A and 5B. Average donor

proportions within subgenomes for KSTdGD80-175 were 9.8% for the A subgenome, 6.1% for the B subgenome, and 13.2% for the D subgenome. These values indicate that KSTdGD80-175 carries a greater overall donor contribution than KSTdGD80-55 across all three subgenomes.

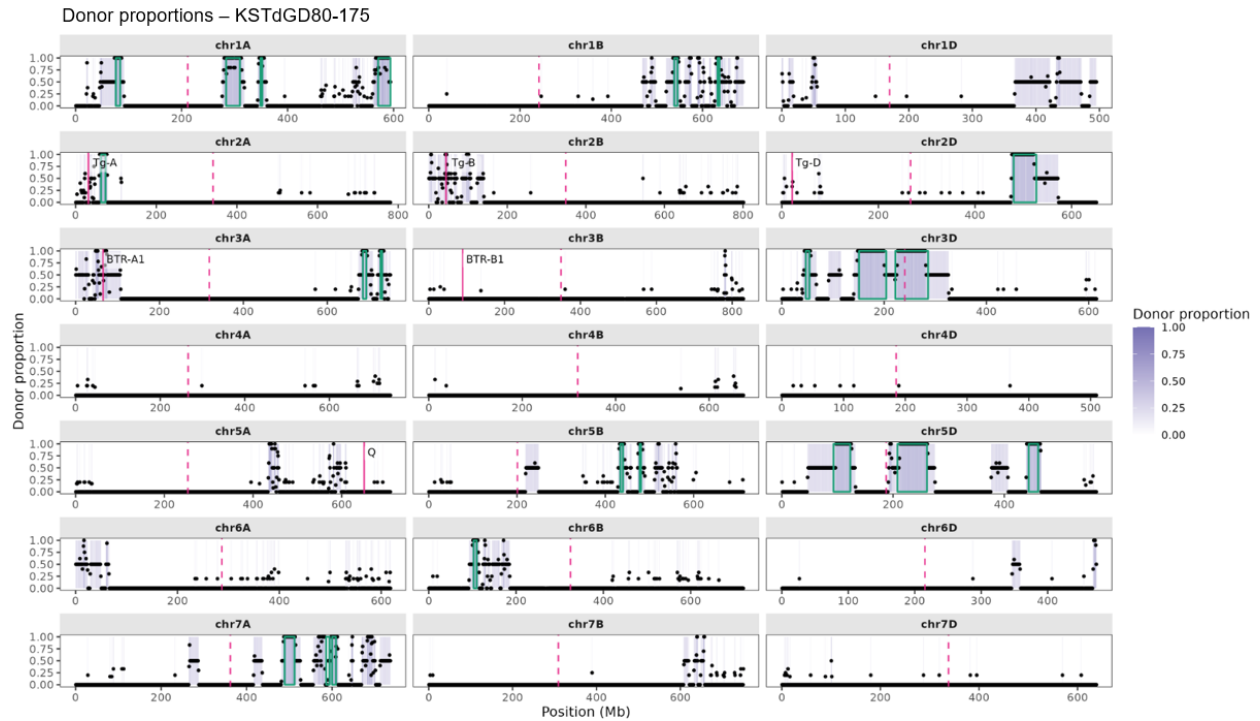


Figure 5.4. Genome-wide introgression profile for KSTdGD80-175

Donor allele proportions were calculated in 1 Mb bins and are shown as points along each chromosome facet. Vertical dashed lines indicate centromere positions based on the Chinese Spring v1.0 reference. Major domestication loci – *Q* (free threshing), *Btr* (brittle rachis), and *Tg* (tenacious glume) – are indicated by labeled vertical markers at their physical positions. Vertical boxes within chromosome panels denote regions where introgression segments were called using a moving average framework ($MA_{10} \geq 0.9$ and $MA_{40} \geq 0.7$).

Allele states in KSTdGD80-175 and KSTdGD80-55 were examined at marker-trait associations previously identified within TdGD Pop80 (Chapters 3, 4). These comparisons identified several instances in which a selected WIL carried a donor allele within or near a mapped introgression segment and the direction of the previously estimated donor effect was consistent with its phenotype. However, concordance varied among traits and loci, emphasizing that the GWAS results provide hypotheses grounded in linkage disequilibrium about the

contributions of retained donor diversity, rather than causal explanations of overall WIL performance.

The clearest concordance occurred within the shared *Ae. tauschii*-derived introgression on chromosome arm 3DS. At 30.24 Mb, both WILs carried the donor allele previously associated with reduced TKW, KW, and KA, consistent with their smaller kernels relative to the RP and performance checks (Figure 5.2; Appendix Table D.3). Nearby elemental associations differentiated the two WILs. At 29.89 Mb, KSTdGD80-55 carried the donor allele associated with lower Co concentrations, whereas KSTdGD80-175 carried the RP allele associated with higher Co concentration, consistent with the comparatively high Co concentration of KSTdGD80-175 (Figure 5.2; Appendix Table D.6).

Table 5.2. High confidence marker-trait association regions containing typed donor alleles in KSTdGD80-175 or KSTdGD80-55

Regions were included when at least one selected WIL carried a typed donor allele at one or more high confidence marker-trait associations within the region. DonDir indicates the direction of the donor allele effect relative to the RP allele. Allele states are designated as donor (Don), recurrent parent (RP), or heterozygous (Het). Segment indicates whether a called introgression segment directly overlapped the association region.

Region	Trait	DonDir	KSTdGD80-175		KSTdGD80-55	
			Allele	Segment	Allele	Segment
Chr1A, 323.90-568.47 Mb	GR	-	RP	Yes	RP	No
	NDVI	+	RP		RP	
	Fe	+	RP		RP	
	P	+	Don		RP	
	Mg	+	RP		RP	
	Mo	+	Don		RP	
	Sr	+	RP		RP	
Chr3A, 40.15-69.67 Mb	GFP	+	RP	No	RP	No
	PH	+	Don		Don	
	GR	+	Don		Don	
	NDVI	+	Don		Don	
Chr3D, 29.89-31.69 Mb	TKW	-	Don	No	Don	Yes
	KA	-	Don		Don	
	KW	-	Don		Don	
	Co	-	RP		Don	
	Fe	-	RP		Don	

Chr3D, 454.67 Mb	PQS	+	Het	No	Don	No
	SRC	+	Het		Don	
Chr4A, 662.20 Mb	HD	-	Het	No	Don	No
	GFP	+	Het		Don	
	Mo	+	Het		Don	
Chr5A, 403.35-671.27 Mb	GR	+	RP	No	RP	No
	NDVI	-	RP		RP	
	TKW	-	Don		Don	
	TW	-	RP		RP	
	KA	+	RP		RP	
	KL	-	Don		Don	
	As	+	Don		Don	
	Ca	+	Don		Don	
	Mn	-	RP		RP	
Chr7A, 527.15-609.55 Mb	TKW	-	Don	Yes	Don	Yes
	KW	+	RP		RP	
	SRC	-	Don		Don	

Additional comparisons also illustrated the quantitative nature of the selected traits. On chromosome arm 7AL, both WILs carried the wild emmer allele at 609.55 Mb, which was associated with reduced TKW and SRC. The reduced TKW of both lines was consistent with this association, whereas their favorable overall SRC performance – particularly that of KSTdGD80-55 – were not (Figure 5.2). On chromosome arm 1AL, KSTdGD80-175 carried donor alleles at 508.84 and 568.01 Mb, which were previously associated with increased Mo and P concentrations, respectively, and occurred near localized donor ancestry in that line. KSTdGD80-55 retained RP alleles at these positions and had no called introgression in the region, although its elevated Mo concentration again indicated contributions from other loci or genomic regions.

The functional quality associations similarly did not provide a single locus explanation for the favorable performance of both WILs. At the chromosome 3D marker at 454.67 Mb associated with PQS and SRC, the genotype call for KSTdGD80-175 was heterozygous, whereas

KSTdGD80-55 carried the donor allele associated with increased functional quality. This marker was outside the called 3D introgression segments in both lines. Neither WIL carried that donor allele at the chromosome 3A functional quality association, despite both exhibited higher SRC and PQS than the RP (Figure 5.2). Thus, the functional quality profiles of the selected WILs likely reflects multilocus variation that was not fully captured by individual GWAS markers or the smoothed introgression calls.

Several additional donor alleles were detected at associations on chromosomes 3A and 4A despite the absence of called introgression segments at those positions. Because introgression segments were defined from moving averages of donor ancestry, isolated donor marker states do not necessarily meet the criteria for a segment call. Table 5.2 summarizes high confidence association regions in which at least one selected WIL carried a typed donor allele at one or more associated markers, together with other associations colocalized within those regions. Regions where both lines carried only RP alleles were excluded because they were less informative for characterizing the donor diversity retained by the selected WILs. Collectively, these comparisons demonstrate that the selected WILs retain donor alleles associated with multiple traits relevant to breeding, while also showing that neither individual marker effects nor broad introgression segments fully explain their integrated trait profiles.

Discussion

Multi-trait selection translated introgression-derived variation into breeding utility

The principal outcome of this research was not simply the generation of additional phenotypic variation, but the recovery of individual WILs in which wheat wild relative diversity was combined with competitive performance across several breeding targets. Introgression populations frequently exhibit greater phenotypic variance while remaining agronomically

inferior on average due to inheritance of unfavorable linked variation alongside useful donor alleles, and because introgression can disrupt allele combinations assembled through long-term selection in adapted germplasm (Huang et al., 2003; Klindworth et al., 2013; Lauterberg et al., 2022; Li et al., 2022; Merchuk-Ovnat et al., 2016a; Platten & Fritsche-Neto, 2023; Shahin et al., 2024). That pattern was evident in TdGD Pop80, where trait-specific characterization identified transgressive variation, but the population-wide yield disadvantage and correlated structure between developmental, yield component, and functional quality traits made the breeding utility of individual WILs difficult to determine from population means or single trait rankings alone (Chapters 3, 4).

The yield-weighted multi-trait index provided a practical means of resolving this ambiguity. Its value arose from maintaining yield as the central measure of adaptation while allowing complementary selection targets to differentiate lines having similar performance for yield. This approach was appropriate for pre-breeding germplasm because favorable secondary characteristics have limited immediate value if accompanied by an unacceptable yield penalty, but yield alone is insufficient to describe the suitability of a line for advancement.

The rankings of the leading WILs demonstrate that the index operated as intended rather than simply reproducing the yield ranking. It preserved yield priority but changed rankings when performance across the accompanying traits provided a defensible reason to do so. KSTdGD80-175 ranked first for both yield and the overall index, confirming that the index retained the highest yielding line when its broader trait profile was also favorable. By comparison, KSTdGD80-55 ranked second in the index despite lower GY than Showdown and several WILs, harboring compensatory advantages in agronomic and quality characteristics.

The index also identified WILs that were competitive relative to elite germplasm rather than merely favorable within an introgression population. Out of 279 WILs, 29% exceeded the RP in the index, demonstrating that the variation generated by introgression was not limited to recovery of the RP genetic background. More importantly, KSTdGD80-175 and KSTdGD80-55 exceeded all four performance checks. This comparison is especially meaningful because the checks represent elite, adapted, winter wheat germplasm against which the practical value of the WILs can be judged. The results therefore support the central hypothesis that competitive multi-trait performance can be recovered following amphiploid bridge-mediated introgression, despite the unfavorable average yield response of the population.

Use of across-environment BLUPs further increased the relevance of the index by emphasizing performance across the target set of rainfed Kansas environments rather than performance within single trials. KSTdGd80-175 remained among the six highest ranking entries in every within-environment index, while KSTdGD80-55 ranked within the top 15 in five of six environments. Although the within-environment indices are not formal estimates of stability, they show that the high across-environment rankings were not driven entirely by exceptional performance in one environment. In particular, the consistency of KSTdGd80-175 supports its use as a source of broadly competitive yield, whereas the performance of KSTdGD80-55 indicates a favorable but somewhat more environmentally responsive combination of target traits.

These results establish the internal validity and practical utility of the index for prioritizing WILs from this population, but they should not be interpreted as independent predictive validation. The same multi-environment dataset contributed to index construction and WIL selection, and standardized values describe performance relative to the germplasm and

environments evaluated. Moreover, this index permits compensation among traits and does not impose minimum thresholds for characteristics that may be non-negotiable in a particular breeding program. The index should therefore be viewed as a prioritization tool rather than a complete product profile. Validation across additional environments and genetic backgrounds, as well as evaluation of ranking sensitivity to alternative trait sets and weights, would further establish the portability of the selection framework.

Population size influenced the opportunity to recover rare favorable recombinants

The low frequency of WILs surpassing all elite checks illustrates the combinatorial challenge of introgression breeding. Recovery of the two best performing WILs required recombination to retain useful donor variation while removing domestication-related deficiencies, reducing yield disadvantage, and restoring coordinated performance across numerous quantitative traits. Each requirement may be achievable independently, but their simultaneous occurrence is probabilistically less frequent.

Consequently, population size is an important design consideration when introgression is intended to improve quantitative trait combinations rather than transfer a single major effect allele. Larger populations increase the number of recombination events sampled and therefore the opportunity to recover favorable combinations of donor and RP alleles. The present results do not define a universal minimum population size, because the required number will depend on donor divergence, backcross generation, introgression size, selection intensity, and the complexity of the target profile. They do, however, demonstrate the risk of judging an introgression strategy from population means or a small sample of derivatives. Had TdGD Pop80 been evaluated at substantially smaller scale, the two most favorable recombinant profiles could feasibly have been missed.

This distinction is particularly important for interpreting the yield disadvantages of introgression breeding populations. The unfavorable population mean accurately describes the general consequence of introgression, but it does not determine the upper limit of performance among recombinant individuals. For pre-breeding, the relevant outcome is whether sufficiently favorable individuals can be recovered and made accessible to breeding programs. The selection of KSTdGD80-175 and KSTdGD80-55 shows that the amphiploid bridge approach produced such individuals, although at a frequency that reinforces the need for population-scale development, empirical evaluation, and multi-trait selection.

Selected introgression lines provide complementary breeding resources

KSTdGD80-175 and KSTdGD80-55 should be viewed as complementary rather than interchangeable resources. KSTdGD80-175 represents a strongly yield-oriented selection. The combination of excellent GY, reduced PH, extended GFP, high TW, and favorable SRC despite lower GP is particularly useful because it demonstrates that donor ancestry can be retained without preventing recovery of competitive regional performance. KSTdGD80-175 may therefore be most immediately useful where breeders seek wild-derived diversity in an adapted background that has already been characterized for improved agronomic and yield component traits under drought stressed environments.

KSTdGD80-55 represents a different balance of traits. Although its GY was lower than that of KSTdGD80-175, its index ranking reflected favorable PH, TC, TW, GP, GPD, and SRC, together with strong performance in several drought stressed western Kansas environments. Its combination of competitive GY and improved functional quality characteristics makes it especially relevant in the context of breeding for improved end-use quality. The favorable SRC of both WILs, despite their differing GP responses, also illustrates why GP alone is an

incomplete measure of end-use potential. The lines therefore offer different starting points for further investigation into the genetic relationship among yield, protein concentration, and protein functionality.

The rainfed environments used in this research differed in the timing and severity of water limitation (Chapter 3). Accordingly, the selected WILs are useful resources for continued investigation of performance under drought-affected production conditions. Their results do not, however, demonstrate drought tolerance as a distinct physiological mechanism because the experiment did not include paired drought stressed and well-watered treatments, and water limitation was not imposed uniformly. The defensible conclusion is that both lines retained favorable performance across the variable rainfed environments evaluated. Controlled or paired field experiments would be necessary to determine whether these responses reflect drought-specific adaptation, general yield potential, drought escape, maintenance of grain filling, or combinations of these processes.

Characterized introgressions increase the value of germplasm releases

The breeding value of KSTdGD80-175 and KSTdGD80-55 extends beyond their phenotypic performance. Both retain wild emmer-derived diversity on the A and B subgenomes and *Ae. tauschii*-derived diversity on the D subgenome, but they carry distinct combinations of introgressions. KSTdGD80-175 retained greater donor proportions across all three subgenomes, whereas KSTdGD80-55 carried a smaller and partially different set of donor segments. Their competitive performance therefore did not arise simply through complete recovery of the RP genome. Instead, the two WILs demonstrate that appreciable donor-derived allelic diversity can remain in agronomically competitive BC₂-derived germplasm.

This genomic characterization makes the selected WILs more immediately useful than uncharacterized wild accessions. Breeders can select between the lines based on both their trait profiles and the donor regions they contain, use them as parents in direct hexaploid crosses, and follow donor segments during subsequent recombination. The lines therefore reduce – although they do not eliminate – the time, population size, and phenotyping required to deploy the original wild emmer and *Ae. tauschii* diversity.

Comparison with previously identified GWAS regions (Chapters 3, 4) further increases the research value of KSTdGD80-175 and KSTdGD80-55, but does not provide a simple causal explanation for their phenotypes. Several donor alleles were directionally consistent while others were discordant with overall trait performance. Such results are expected for quantitative traits governed by multilocus effects and interactions, linkage disequilibrium, and environmental response. They reinforce the interpretation that the value of these lines resides in their integrated allele combinations rather than any single marker or introgression.

For that reason, KSTdGD80-175 and KSTdGD80-55 can serve two related purposes. First, they are practical breeding resources for programs seeking regionally competitive yield performance under drought, variation in developmental timing, improved functional quality, and allelic diversity from wheat wild relatives. Second, they are resources for further research testing whether marker-trait associations and donor allele effects remain consistent after transfer into different genetic backgrounds and environments. Crossing each WIL to multiple adapted parents would permit evaluation of background dependence, while recombination within retained segments could refine their relationships with trait performance.

Conclusion

Amphiploid bridge-mediated introgression generated expanded phenotypic and allelic variation with directed breeding value becoming apparent after selection prioritized multi-trait performance. The modified GYT index provided an effective framework for this purpose by maintaining strong emphasis on yield while distinguishing WILs according to complementary agronomic, kernel component, and functional quality characteristics. The recovery of only two WILs that exceeded all elite performance checks illustrates the rarity of favorable recombinant deviates and the practical importance of developing and evaluating sufficiently large introgression populations.

KSTdGD80-175 and KSTdGD80-55 provide characterized wild emmer and *Ae. tauschii* introgressions in an elite winter wheat background while retaining competitive yield and favorable target trait profiles. Their performance across drought stressed environments supports their use in continued evaluation for yield stability under drought. As germplasm releases, these WILs provide resources for targeted crossing, refinement of donor segments, and validation of wild-derived trait and allelic variation across additional environments and genetic backgrounds. In this way, they move the value of wild relative diversity beyond discovery and toward practical deployment in wheat improvement.

Availability

Breeder seed of KSTdGD80-55 and KSTdGD80-175 will be maintained by the Kansas State University wheat breeding program in Manhattan, KS. Seed is immediately available upon direct request in 2 g quantities for research and breeding purposes. Additionally, seed of both WILs will be available from the USDA-ARS National Laboratory for Genetic Resources Preservation upon publication.

References

- Adhikari, L., Olivera, P. D., Raupp, J., Sela, H., Distelfeld, A., Maccaferri, M., Bozzoli, M., Mazzucotelli, E., Brandolini, A., Tuberosa, R., Özkan, H., Wulff, B. B. H., Steffenson, B. J., & Poland, J. (2026). Dissecting the population structure, diversity and genetic architecture of rust resistance in wild emmer wheat (*Triticum turgidum* subsp. *Dicoccoides*). *BMC Genomics*, 27, 284. <https://doi.org/10.1186/s12864-026-12751-6>
- Afzal, F., Li, H., Gul, A., Subhani, A., Ali, A., Mujeeb-Kazi, A., Ogbonnaya, F., Trethowan, R., Xia, X., He, Z., & Rasheed, A. (2019). Genome-wide analyses reveal footprints of divergent selection and drought adaptive traits in synthetic-derived wheats. *G3: Genes, Genomes, Genetics*, 9(6), 1957-1973. <https://doi.org/10.1534/g3.119.400010>
- Akbarimehr, S., Sayfzadeh, S., Shahsavari, N., Valad Abadi, S. A., & Masouleh, E. H. (2022). Cycocel, iron and zinc effects on yield and physiological characteristics of wheat under drought stress conditions. *Journal of Plant Nutrition*, 45(5), 664-673. <https://doi.org/10.1080/01904167.2021.1936029>
- Akhtar, N., Ilyas, N., Arshad, M., Meraj, T. A., Hefft, D. I., Jan, B. L., & Ahmad, P. (2022). The Impact of Calcium, Potassium, and Boron Application on the Growth and Yield Characteristics of Durum Wheat under Drought Conditions. *Agronomy*, 12(8), 1917. <https://doi.org/10.3390/agronomy12081917>
- Akhunov, E. D., Akhunova, A. R., Anderson, O. D., Anderson, J. A., Blake, N., Clegg, M. T., Coleman-Derr, D., Conley, E. J., Crossman, C. C., Deal, K. R., Dubcovsky, J., Gill, B. S., Hadam, J., Heo, H., Huo, N., Lazo, G. R., Luo, M. C., Ma, Y. Q., Matthews, D. E., ... Dvorak, J. (2010). Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics*, 11, 702. <https://doi.org/10.1186/1471-2164-11-702>
- Akhunov, E., Nicolet, C., & Dvorak, J. (2009). Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theoretical and Applied Genetics*, 119(3), 507-517. <https://doi.org/10.1007/s00122-009-1059-5>
- Ali, S., Leconte, M., Hovmöller, M. S., Enjalbert, J., & de Vallavieille-Pope, C. (2025). Plasticity in initial infection traits in response to high temperature for worldwide representative *Puccinia striiformis* isolates. *Fungal Genetics and Biology*, 181, 104039. <https://doi.org/10.1016/j.fgb.2025.104039>
- Alomari, D. Z., Eggert, K., von Wirén, N., Pillen, K., & Röder, M. S. (2017). Genome-wide association study of calcium accumulation in grains of European wheat cultivars. *Frontiers in Plant Science*, 8(1797). <https://doi.org/10.3389/fpls.2017.01797>
- Alomari, D. Z., Eggert, K., von Wirén, N., Polley, A., Plieske, J., Ganal, M. W., Liu, F., Pillen, K., & Röder, M. S. (2019). Whole-genome association mapping and genomic prediction for iron concentration in wheat grains. *International Journal of Molecular Sciences*, 20(1), 76. <https://doi.org/10.3390/ijms20010076>

- Alqudah, A. M., Elkelish, A., Abu-Elsaoud, A. M., Hassan, S. E.-D., & Thabet, S. G. (2025). Genome-wide association study reveals the genetic basis controlling mineral accumulation in wheat grains under potassium deficiency. *Genetic Resources and Crop Evolution*, 72, 783-796. <https://doi.org/10.1007/s10722-024-01999-0>
- Aparicio, J., Gezan, S. A., Ariza-Suarez, D., Raatz, B., Diaz, S., Heilman-Morales, A., & Lobaton, J. (2024). Mr. Bean: a comprehensive statistical and visualization application for modeling agricultural field trials data. *Frontiers in Plant Science*, 14, 1290078. <https://doi.org/10.3389/fpls.2023.1290078>
- Armstrong, C. L., Baldinelli, S. E., Smith, K. H., Medcalf, S. (2025). Expertise or Experience: Comparing Discussion of Drought between Audiences and Agencies in Nebraska, Kansas, Iowa, and Missouri. *Weather, Climate, and Society*, 17(3), 449-459. <https://doi.org/10.1175/WCAS-D-24-0061.1>
- Asaro, A., Ziegler, G., Ziyomo, C., Hoekenga, O. A., Dilkes, B. P., & Baxter, I. (2016). The interaction of genotype and environment determines variation in the maize kernel ionome. *G3: Genes, Genomes, Genetics*, 6(12), 4175-4183. <https://doi.org/10.1534/g3.116.034827>
- Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., Hale, I., Mascher, M., Spannagl, M., Wiebe, K., Jordan, K. W., Golan, G., Deek, J., Ben-Zvi, B., Ben-Zvi, G., Himmelbach, A., MacLachlan, R. P., Sharpe, A. G., Fritz, A., ... Distelfeld, A. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science*, 357(6346), 93-97. <https://doi.org/10.1126/science.aan0032>
- Bacher, H., Zhu, F., Gao, T., Liu, K., Dhatt, B. K., Awada, T., Zhang, C., Distelfeld, A., Yu, H., Peleg, Z., & Walia, H. (2021). Wild emmer introgression alters root-to-shoot growth dynamics in durum wheat in response to water stress. *Plant Physiology*, 187(3), 1149-1162. <https://doi.org/10.1093/plphys/kiab292>
- Bag, A., Krishna, H., Kumar P.N., V., Meena, S., Devate, N. B., Kumar, S., Patil, R., Reddy, U. G., Singh, A. K., Singh, B., Jain, N., Singh, P. K., & Singh, G. P. (2026). Genetic mapping and candidate gene identification for key physiological traits associated with heat tolerance in wheat (*Triticum aestivum* L.) using a MAGIC population. *PLoS ONE*, 21(1), e0339966. <https://doi.org/10.1371/journal.pone.0339966>
- Bagci, S. A., Ekiz, H., Yilmaz, A., & Cakmak, I. (2007). Effects of Zinc Deficiency and Drought on Grain Yield of Field-grown Wheat Cultivars in Central Anatolia. *Journal of Agronomy and Crop Science*, 193(3), 198-206. <https://doi.org/10.1111/j.1439-037X.2007.00256.x>
- Bagherikia, S., Soughi, H., Khodarahmi, M., & Naghipour, F. (2025). The effect of sowing dates on grain yield and quality in spring wheat (*Triticum aestivum* L.). *Food Science & Nutrition*, 13(5), e70035. <https://doi.org/10.1002/fsn3.70035>
- Bai, G., & Shaner, G. (1994). Scab of wheat: prospects for control. *Plant Disease*, 78(8), 760-766. <https://doi.org/10.1094/PD-78-0760>

- Bai, G., & Shaner, G. (2004). Management and resistance in wheat and barley to *Fusarium* head blight. *Annual Review of Phytopathology*, 42, 135-161. <https://doi.org/10.1146/annurev.phyto.42.040803.140340>
- Bai, G., Su, Z., & Cai, J. (2018). Wheat resistance to *Fusarium* head blight. *Canadian Journal of Plant Pathology*, 40(3), 336-346. <https://doi.org/10.1080/07060661.2018.1476411>
- Ballagh, A., Cox, E. K., Lofton, J., & Arnall, D. B. (2023). Impacts of soil pH and extractable aluminum on winter canola production in the southern Great Plains. *Journal of Plant Nutrition*, 47(2), 257-267. <https://doi.org/10.1080/01904167.2023.2275074>
- Barkley, A., Tack, J., Nalley, L. L., Bergtold, J., Bowden, R., & Fritz, A. (2014). Weather, disease, and wheat breeding effects on Kansas wheat varietal yields, 1985 to 2011. *Agronomy Journal*, 106(1), 227-235. <https://doi.org/10.2134/agronj2013.0388>
- Bashir, L., Budhlakoti, N., Pradhan, A. K., Sharma, D., Jain, A., Rehman, S. S., Kondal, V., Jacob, S. R., Bhardwaj, R., Gaikwad, K., Mishra, D. C., Pandey, A., Kaur, S., Bhati, P. K., Singh, R., Singh, G. P., & Kumar, S. (2025). Identification of quantitative trait nucleotides for grain quality in bread wheat under heat stress. *Scientific Reports*, 15, 6641. <https://doi.org/10.1038/s41598-025-91199-2>
- Bazhenov, M. S., Bespalova, L. A., Kocheshkova, A. A., Chernook, A. G., Puzyrnaya, O. Y., Agaeva, E. V., Nikitina, E. A., Igonin, V. N., Bazhenova, S. S., Vertikova, E. A., Kharchenko, P. N., Karlov, G. I., & Divashuk, M. G. (2022). The association of grain yield and agronomical traits with genes of plant height, photoperiod sensitivity and plastid glutamine synthetase in winter bread wheat (*Triticum aestivum* L.) collection. *International Journal of Molecular Sciences*, 23(19), 11402. <https://doi.org/10.3390/ijms231911402>
- Begum, H., Alam, M. S., Feng, Y., Koua, P., Ashrafuzzaman, M., Shrestha, A., Kamruzzaman, M., Dadshani, S., Ballvora, A., Naz, A. A., & Frei, M. (2020). Genetic dissection of bread wheat diversity and identification of adaptive loci in response to elevated tropospheric ozone. *Plant, Cell, and Environment*, 43(11), 2650-2665. <https://doi.org/10.1111/pce.13864>
- Ben-David, R., Xie, W., Peleg, Z., Saranga, Y., Dinoor, A., & Fahima, T. (2010). Identification and mapping of *PmGl6*, a powdery mildew resistance gene derived from wild emmer wheat. *Theoretical and Applied Genetics*, 121(3), 499-510. <https://doi.org/10.1007/s00122-010-1326-5>
- Beral, A., Rincet, R., Le Gouis, J., Girousse, C., & Allard, V. (2020). Wheat individual grain-size variance originates from crop development and from specific genetic determinism. *PLoS ONE*, 15(3), e0230689. <https://doi.org/10.1371/journal.pone.0230689>
- Bhatta, M., Baenziger, P. S., Waters, B. M., Poudel, R., Belamkar, V., Poland, J., & Morgounov, A. (2018). Genome-wide association study reveals novel genomic regions associated with 10 grain minerals in synthetic hexaploid wheat. *International Journal of Molecular Sciences*, 19(10), 3237. <https://doi.org/10.3390/ijms19103237>

- Blanco, A., Bellomo, M. P., Lotti, C., Maniglio, T., Pasqualone, A., Simeone, R., Troccoli, A., & Di Fonzo, N. (1998). Genetic mapping of sedimentation volume across environments using recombinant inbred lines of durum wheat. *Plant Breeding*, 117(5), 413-417. <https://doi.org/10.1111/j.1439-0523.1998.tb01965.x>
- Blanco, A., Gadaleta, A., Cenci, A., Carluccio, A. V., Abdelbacki, A. M. M., & Simeone, R. (2008). Molecular mapping of the novel powdery mildew resistance gene *Pm36* introgressed from *Triticum turgidum* var. *dicoccoides* in durum wheat. *Theoretical and Applied Genetics*, 117(1), 135-142. <https://doi.org/10.1007/s00122-008-0760-0>
- Blanco, A., Mangini, G., Giancaspro, A., Giove, S., Colasuonno, P., Simeone, R., Signorile, A., De Vita, P., Mastrangelo, A. M., Cattivelli, L., & Gadaleta, A. (2012). Relationships between grain protein content and grain yield components through quantitative trait locus analyses in a recombinant inbred line population derived from two elite durum wheat cultivars. *Molecular Breeding*, 30, 79-92. <https://doi.org/10.1007/s11032-011-9600-z>
- Blanco-Pastor, J. L. (2022). Alternative modes of introgression-mediated selection shaped crop adaptation to novel climates. *Genome Biology and Evolution*, 14(8). <https://doi.org/10.1093/gbe/evac107>
- Boehm, J., Jr., & Cai, X. (2024). Enrichment and diversification of the wheat genome via alien introgression. *Plants*, 13(3), 339. <https://doi.org/10.3390/plants13030339>
- Bohra, A., Kilian, B., Sivasankar, S., Caccamo, M., Mba, C., McCouch, S. R., & Varshney, R. K. (2022). Reap the crop wild relatives for breeding future crops. *Trends in Biotechnology*, 20(4), 412-431. <https://doi.org/10.1016/j.tibtech.2021.08.009>
- Borrill, P., Connorton, J. M., Balk, J., Miller, A. J., Sanders, D., & Uauy, C. (2014). Biofortification of wheat grain with iron and zinc: integrating novel genomic resources and knowledge from model crops. *Frontiers in Plant Science*, 5, 53. <https://doi.org/10.3389/fpls.2014.00053>
- Brancourt-Hulmel, M., Doussinault, G., Lecomte, C., Bérard, P., Le Buanec, B., & Trottet, M. (2003). Genetic improvement of agronomic traits of winter wheat cultivars released in France from 1946 to 1992. *Crop Science*, 43(1), 37-45. <https://doi.org/10.2135/cropsci2003.3700>
- Brdar-Jokanović, M. (2020). Boron Toxicity and Deficiency in Agricultural Plants. *International Journal of Molecular Sciences*, 21(4), 1424. <https://doi.org/10.3390/ijms21041424>
- Brevis, J. C., & Dubcovsky, J. (2010). Effects of the chromosome region including the *Gpc-B1* locus on wheat grain and protein yield. *Crop Science*, 50(1), 93-104. <https://doi.org/10.2135/cropsci2009.02.0057>
- Buerstmayr, M., Alimari, A., Steiner, B., & Buerstmayr, H. (2013). Genetic mapping of QTL for resistance to *Fusarium* head blight spread (type 2 resistance) in a *Triticum dicoccoides* x *Triticum durum* backcross-derived population. *Theoretical and Applied Genetics*, 126(11), 2825-2834. <https://doi.org/10.1007/s00122-013-2174-x>

- Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. G., & Thompson, R. (2023). *ASReml R Reference Manual Version 4.2*. VSN International Ltd.
- Cabas-Lühmann, P., Arriagada, O., Matus, I., Marcotuli, I., Gadaleta, A., & Schwember, A. R. (2023). Comparison of durum with ancient tetraploid wheats from an agronomical, chemical, nutritional, and genetic standpoints: A review. *Euphytica*, 219, 61. <https://doi.org/10.1007/s10681-023-03188-z>
- Cakmak, I., Torun, A., Millet, E., Feldman, M., Fahima, T., Korol, A., Nevo, E., Braun, H. J., & Özkan, H. (2004). *Triticum dicoccoides*: An important genetic resource for increasing zinc and iron concentration in modern cultivated wheat. *Soil Science and Plant Nutrition*, 50(7), 1047-1054. <https://doi.org/10.1080/00380768.2004.10408573>
- Cakmak, I., Yilmaz, A., Kalayci, M., Ekiz, H., Torun, B., Erenoglu, B., & Braun, H. J. (1996). Zinc deficiency as a critical problem in wheat production in Central Anatolia. *Plant and Soil*, 180(2), 165-172. <https://doi.org/10.1007/BF00015299>
- Calus, M. P. L., Bouwman, A. C., Hickey, J. M., Veerkamp, R. F., & Mulder, H. A. (2014). Evaluation of measures of correctness of genotype imputation in the context of genomic prediction: a review of livestock applications. *Animal*, 8(11), 1743-1753. <https://doi.org/10.1017/S1751731114001803>
- Carter, A. H., Garland-Campbell, K., Morris, C. F., & Kidwell, K. K. (2012). Chromosomes 3B and 4D are associated with several milling and baking quality traits in a soft white spring wheat (*Triticum aestivum* L.) population. *Theoretical and Applied Genetics*, 124(6), 1079-1096. <https://doi.org/10.1007/s00122-011-1770-x>
- Carvalho, S. M. P., & Vasconcelos, M. W. (2013). Producing more with less: Strategies and novel technologies for plant-based food biofortification. *Food Research International*, 54(1), 961-971. <https://doi.org/10.1016/j.foodres.2012.12.021>
- Chandrasekhar, K., Nashef, K., & Ben-David, R. (2017). Agronomic and genetic characterization of wild emmer wheat (*Triticum turgidum* subsp. *dicoccoides*) introgression lines in a bread wheat genetic background. *Genetic Resources and Crop Evolution*, 64(8), 1917-1926. <https://doi.org/10.1007/s10722-016-0481-1>
- Chao, S., Dubcovsky, J., Dvorak, J., Luo, M.-C., Baenziger, S. P., Matnyazov, R., Clark, D. R., Talbert, L. E., Anderson, J. A., Dreisigacker, S., Glover, K., Chen, J., Campbell, K., Bruckner, P. L., Rudd, J. C., Haley, S., Carver, B. F., Perry, S., Sorrells, M. E., & Akhunov, E. A. (2010). Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.) *BMC Genomics*, 11, 727. <https://doi.org/10.1186/1471-2164-11-727>
- Chatzav, M., Peleg, Z., Ozturk, L., Yazici, A., Fahima, T., Cakmak, I., & Saranga, Y. (2010). Genetic diversity for grain nutrients in wild emmer wheat: potential for wheat improvement. *Annals of Botany*, 105(7), 1211-1220. <https://doi.org/10.1093/aob/mcq024>

- Chen, L., Li, H., Liu, Y., Yang, J., & Hou, Y. (2021). Genetic identification of a major, novel and stably expressed QTL for effective tiller number from tetraploid wheat. *Biotechnology & Biotechnological Equipment*, 35(1), 1538-1545. <https://doi.org/10.1080/13102818.2021.1996271>
- Chen, L., Yang, Y., Cui, C., Lu, S., Lu, Q., Du, Y., Su, R., Chai, Y., Li, H., Chen, F., Yu, F., & Hu, Y.-G. (2018). Effects of *Vrn-B1* and *Ppd-D1* on developmental and agronomic traits in *Rht5* dwarf plants of bread wheat. *Field Crops Research*, 219, 24-32. <https://doi.org/10.1016/j.fcr.2018.01.022>
- Chen, T., Zhang, L., Zhang, Y., Gao, W., Zhang, P., Guo, L., Yang, D. (2024). Genome-wide identification of the endonuclease family genes implicates potential roles of *TaENDO23* in drought-stressed response and grain development in wheat. *BMC Genomics*, 25, 919. <https://doi.org/10.1186/s12864-024-10840-y>
- Chen, W., Liu, T., & Gao, L. (2013). Suppression of stripe rust and leaf rust resistances in interspecific crosses of wheat. *Euphytica*, 192, 339-346. <https://doi.org/10.1007/s10681-012-0854-2>
- Chen, X. M. (2005). Epidemiology and control of stripe rust (*Puccinia striiformis* f. sp. *tritici*) on wheat. *Canadian Journal of Plant Pathology*, 27(3), 314-337. <https://doi.org/10.1080/07060660509507230>
- Chen, X. M. (2013). High-temperature adult-plant resistance, key for sustainable control of stripe rust. *American Journal of Plant Sciences*, 4(3), 608-627. <http://dx.doi.org/10.4236/ajps.2013.43080>
- Chen, X. M., Luo, Y. H., Xia, X. C., Xia, L. Q., Ren, Z. L., & He, Z. H. (2005). Chromosomal location of powdery mildew resistance gene *Pm16* in wheat using SSR marker analysis. *Plant Breeding*, 124(3), 225-228. <https://doi.org/10.1111/j.1439-0523.2005.01094.x>
- Chen, X. M., Moore, M., Milus, E. A., Long, D. L., Line, R. F., Marshall, D., & Jackson, L. (2002). Wheat stripe rust epidemics and races of *Puccinia striiformis* f. sp. *tritici* in the United States in 2000. *Plant Disease*, 86(1), 39-46. <https://doi.org/10.1094/PDIS.2002.86.1.39>
- Cheng, H., Liu, J., Wen, J., Nie, X., Xu, L., Chen, N., Li, Z., Wang, Q., Zheng, Z., Li, M., Cui, L., Liu, Z., Bian, J., Wang, Z., Xu, S., Appels, R., Han, D., Song, W., & Jiang, Y. (2019). Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. *Genome Biology*, 20(1), 136. <https://doi.org/10.1186/s13059-019-1744-x>
- Christopher, J. T., Christopher, M. J., Borrell, A. K., Fletcher, S., & Chenu, K. (2016). Stay-green traits to improve wheat adaptation in well-watered and water-limited environments. *Journal of Experimental Botany*, 67(17), 5159-5172. <https://doi.org/10.1093/jxb/erw276>
- Christopher, J. T., Veyradier, M., Borrell, A. K., Harvey, G., Fletcher, S., & Karine, C. (2014). Phenotyping novel stay-green traits to capture genetic variation in senescence dynamics. *Functional Plant Biology*, 41(11), 1035-1048. <https://doi.org/10.1071/FP14052>

- Coletta, R. D., Qiu, Y., Ou, S., Hufford, M. B., & Hirsch, C. N. (2021). How the pan-genome is changing crop genomics and improvement. *Genome Biology*, 22, 3. <https://doi.org/10.1186/s13059-020-02224-8>
- Condorelli, G. E., Maccaferri, M., Newcomb, M., Andrade-Sanchez, P., White, J. W., French, A. N., Sciara, G., Ward, R., & Tuberosa, R. (2018). Comparative aerial and ground based high throughput phenotyping for the genetic dissection of NDVI as a proxy for drought adaptive traits in durum wheat. *Frontiers in Plant Science*, 9. <https://doi.org/10.3389/fpls.2018.00893>
- Connorton, J. M., Jones, E. R., Rodríguez-Ramiro, I., Fairweather-Tait, S., Uauy, C., & Balk, J. (2017). Wheat vacuolar iron transporter *TaVIT2* transports Fe and Mn and is effective for biofortification. *Plant Physiology*, 174(4), 2434-2444. <https://doi.org/10.1104/pp.17.00672>
- Coombes, N. E. (2020). DiGger design search tool in R. Retrieved from <https://nswdpiobiom.org/austatgen/software/>
- Cox, T. S., Wu, J., Wang, S., Cai, J., Zhong, Q., & Fu, B. (2017). Comparing two approaches for introgression of germplasm from *Aegilops tauschii* into common wheat. *The Crop Journal*, 5(4), 355-362. <https://doi.org/10.1016/j.cj.2017.05.006>
- Crain, J. L., Crossa, J., DeHaan, L., Dreisigacker, S., Poland, J., Singh, R. P., & Vitale, P. (2026). Skim-sequencing for genomic selection in wheat: a comparison of marker platforms. *The Plant Genome*, 19(1), e70197. <https://doi.org/10.1002/tpg2.70197>
- Cruppe, G., DeWolf, E., Jaenisch, B. R., Andersen Onofre, K., Valent, B., Fritz, A. K., & Lollato, R. P. (2021). Experimental and producer-reported data quantify the value of foliar fungicide to winter wheat and its dependency on genotype and environment in the U.S. central Great Plains. *Field Crops Research*, 273, 108300. <https://doi.org/10.1016/j.fcr.2021.108300>
- Cu, S. T., Guild, G., Nicolson, A., Velu, G., Singh, R., & Stangoulis, J. (2020). Genetic dissection of zinc, iron, copper, manganese and phosphorous in wheat (*Triticum aestivum* L.) grain and rachis at two developmental stages. *Plant Science*, 291, 110338. <https://doi.org/10.1016/j.plantsci.2019.110338>
- Cui, C., Lu, Q., Zhao, Z., Lu, S., Duan, S., Yang, Y., Qiao, Y., Chen, L., & Hu, Y.-G. (2022). The fine mapping of dwarf gene *Rht5* in bread wheat and its effects on plant height and main agronomic traits. *Planta*, 255, 114. <https://doi.org/10.1007/s00425-022-03888-1>
- Cui, F., Zhao, C., Ding, A., Li, J., Wang, L., Li, X., Bao, Y., Li, J., & Wang, H. (2014). Construction of an integrative linkage map and QTL mapping of grain yield-related traits using three related wheat RIL populations. *Theoretical and Applied Genetics*, 127, 659-675. <https://doi.org/10.1007/s00122-013-2249-8>
- Dale, Z., Jie, H., Luyu, H., Cancan, Z., Yun, Z., Yarui, S., & Suoping, L. (2017). An advanced backcross population through synthetic octaploid wheat as a “bridge”: Development and

- QTL detection for seed dormancy. *Frontiers in Plant Science*, 8, 2123. <https://doi.org/10.3389/fpls.2017.02123>
- Damesa, T. M., Möhring, J., Worku, M., & Piepho, H.-P. (2017). One step at a time: Stage-wise analysis of a series of experiments. *Agronomy Journal*, 109(3), 848-857. <https://doi.org/10.2134/agronj2016.07.0395>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), 1-4. <https://doi.org/10.1093/gigascience/giab008>
- Daoura, B. G., Chen, L., Du, Y., & Hu, Y.-G. (2014). Genetic effects of dwarfing gene *Rht-5* on agronomic traits in common wheat (*Triticum aestivum* L.) and QTL analysis on its linked traits. *Field Crops Research*, 156, 22-29. <https://doi.org/10.1016/j.fcr.2013.10.007>
- Datta, S., & Datta, S. (2003). Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics*, 19(4), 459-466. <https://doi.org/10.1093/bioinformatics/btg025>
- Davies, L., & Gather, U. (1993). The Identification of Multiple Outliers. *Journal of the American Statistical Association*, 88(423), 782-792. <https://doi.org/10.2307/2290763>
- Deblieck, M., Szilagy, G., Andrii, F., Saranga, Y., Lauterberg, M., Neumann, K., Krugman, T., Perovic, D., Pillen, K., & Ordon, F. (2022). Dissection of a grain yield QTL from wild emmer wheat reveals sub-intervals associated with culm length and kernel number. *Frontiers in Genetics*, 13, 955295. <https://doi.org/10.3389/fgene.2022.955295>
- De Brier, N., Gomand, S. V., Donner, E., Paterson, D., Delcour, J. A., Lombi, E., & Smolders, E. (2015). Distribution of minerals in wheat grains (*Triticum aestivum* L.) and in roller milling fractions affected by pearling. *Journal of Agricultural and Food Chemistry*, 63(4), 1276-1285. <https://doi.org/10.1021/jf5055485>
- De Brier, N., Gomand, S. V., Donner, E., Paterson, D., Smolders, E., Delcour, J. A., & Lombi, E. (2016). Element distribution and iron speciation in mature wheat grains (*Triticum aestivum* L.) using synchrotron X-ray fluorescence microscopy mapping and X-ray absorption near-edge structure (XANES) imaging. *Plant, Cell and Environment*, 39, 1835-1847. <https://doi.org/10.1111/pce.12749>
- Deng, S., Wu, X., Wu, Y., Zhou, R., Wang, H., Jia, J., & Liu, S. (2010). Characterization and precise mapping of a QTL increasing spike number with pleiotropic effects in wheat. *Theoretical and Applied Genetics*, 122, 281-289. <https://doi.org/10.1007/s00122-010-1443-1>
- De Oliveira Silva, A., Jaenisch, B. R., Ciampitti, I. A., & Lollato, R. P. (2021). Wheat nitrogen, phosphorus, potassium, and sulfur uptake dynamics under different management practices. *Agronomy Journal*, 113(3), 2752-2769. <https://doi.org/10.1002/agj2.20637>

- De Oliveira Silva, A., Slafer, G. A., Alderman, P. D., & Carver, B. F. (2026). Grain protein-yield relationship in wheat: sources of variation. *Field Crops Research*, 342, 110453. <https://doi.org/10.1016/j.fcr.2026.110453>
- Des Déserts, A. D., Bouchet, S., Sourdille, P., & Servin, B. (2021). Evolution of recombination landscapes in diverging populations of bread wheat. *Genome Biology and Evolution*, 13(8). <https://doi.org/10.1093/gbe/evab152>
- DeWitt, N., Guedira, M., Lauer, E., Murphy, J. P., Marshall, D., Mergoum, M., Johnson, J., Holland, J. B., & Brown-Guedira, G. (2021). Characterizing the oligogenic architecture of plant growth phenotypes informs genomic selection approaches in a common wheat population. *BMC Genomics*, 22(1), 402. <https://doi.org/10.1186/s12864-021-07574-6>
- De Wolf, E. D., Andersen Onofre, K. F., & Lollato, R. P. (2023). Early season environmental indicators of wheat stripe rust epidemics in Kansas and the central Great Plains region of the United States. *Plant Disease*, 107(7), 2119-2125. <https://doi.org/10.1094/PDIS-08-22-1873-RE>
- Distelfeld, A., Cakmak, I., Peleg, Z., Ozturk, L., Yazici, A. M., Budak, H., Saranga, Y., & Fahima, T. (2007). Multiple QTL-effects of wheat Gpc-B1 locus on grain protein and micronutrient concentrations. *Physiologia Plantarum*, 129(3), 635-643. <https://doi.org/10.1111/j.1399-3054.2006.00841.x>
- Dreisigacker, S., Kishii, M., Lage, J., & Warburton, M. L. (2008). Use of synthetic hexaploid wheat to increase diversity for CIMMYT bread wheat improvement. *Australian Journal of Agricultural Research*, 59(5), 413-420. <https://doi.org/10.1071/AR07225>
- Duan, T., Chapman, S. C., Guo, Y., & Zheng, B. (2017). Dynamic monitoring of NDVI in wheat agronomy and breeding trials using an unmanned aerial vehicle. *Field Crops Research*, 210, 71-80. <https://doi.org/10.1016/j.fcr.2017.05.025>
- Dubcovsky, J., & Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*, 316(5833), 1862-1866. <https://doi.org/10.1126/science.1143986>
- Durmaz, E., Coruh, C., Dinler, G., Grusak, M. A., Peleg, Z., Saranga, Y., Fahima, T., Yazici, A., Ozturk, L., Cakmak, I., & Budak, H. (2011). Expression and cellular localization of ZIP1 transporter under zinc deficiency in wild emmer wheat. *Plant Molecular Biology Reporter*, 29(4), 582-596. <https://doi.org/10.1007/s11105-010-0264-3>
- Dvorak, J., Akhunov, E. D., Akhunov, A. R., Deal, K. R., & Luo, M. C. (2006). Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Molecular Biology and Evolution*, 23(7), 1386-1396. <https://doi.org/10.1093/molbev/msl004>
- Dvorak, J., Deal, K. R., Luo, M. C., You, F. M., von Borstel, K., & Dehghani, H. (2012). The origin of spelt and free-threshing hexaploid wheat. *Journal of Heredity*, 103(3), 426-441. <https://doi.org/10.1093/jhered/esr152>

- Dvorak, J., di Terlizzi, P., Zhang, H. B., & Resta, P. (1993). The evolution of polyploid wheats: Identification of the A genome donor species. *Genome*, 36(1), 21-31. <https://doi.org/10.1139/g93-004>
- Dvorak, J., Luo, M. C., Yang, Z. L., & Zhang, H. B. (1998). The structure of the *Aegilops tauschii* gene pool and the evolution of hexaploid wheat. *Theoretical and Applied Genetics*, 97(4), 657-670. <https://doi.org/10.1007/s001220050942>
- Ehdaie, B., Mohammadi, S. A., & Nouraein, M. (2016). QTLs for root traits at mid-tillering and for root and shoot traits at maturity in a RIL population of spring bread wheat grown under well-watered conditions. *Euphytica*, 211, 17-38. <https://doi.org/10.1007/s10681-016-1670-x>
- El-Feki, W. M., Byrne, P., Reid, S. D., Lapitan, N. L. V., & Haley, S. D. (2013). Quantitative trait locus mapping for end-use quality traits in hard winter wheat under contrasting soil moisture levels. *Crop Science*, 53, 1953-1967. <https://doi.org/10.2135/cropsci2012.12.0674>
- Elbasyoni, I. S., Lorenz, A. J., Guttieri, M., Frels, K., Baenziger, P. S., Poland, J., & Akhunov, E. (2018). A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant Science*, 270, 123-130. <https://doi.org/10.1016/j.plantsci.2018.02.019>
- Ellis, M. H., Rebetzke, G. J., Azanza, F., Richards, R. A., & Spielmeier, W. (2005). Molecular mapping of gibberellin-responsive dwarfing genes in bread wheat. *Theoretical and Applied Genetics*, 111(3), 423-430. <https://doi.org/10.1007/s00122-005-2008-6>
- Endelman, J. B. (2023). Fully efficient, two-stage analysis of multi-environment trials with directional dominance and multi-trait genomic selection. *Theoretical and Applied Genetics*, 136, 65. <https://doi.org/10.1007/s00122-023-04298-x>
- Faris, J. D., Fellers, J. P., Brooks, S. A., & Gill, B. S. (2003). A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene. *Genetics*, 164(1), 311-321. <https://doi.org/10.1093/genetics/164.1.311>
- Farooq, M., Bramley, H., Palta, J. A., & Siddique, K. H. M. (2011). Heat stress in wheat during reproductive and grain-filling phases. *Critical Reviews in Plant Sciences*, 30(6), 491-507. <https://doi.org/10.1080/07352689.2011.615687>
- Fatiukha, A., Deblieck, M., Klymiuk, V., Merchuk-Ovnat, L., Peleg, Z., Ordon, F., Fahima, T., Korol, A., Saranga, Y., & Krugman, T. (2021). Genomic architecture of phenotypic plasticity in response to water stress in tetraploid wheat. *International Journal of Molecular Sciences*, 22(4), 1723. <https://doi.org/10.3390/ijms22041723>
- Fatiukha, A., Filler, N., Lupo, I., Lidzbarsky, G., Klymiuk, V., Korol, A. B., Pozniak, C., Fahima, T., & Krugman, T. (2020a). Grain protein content and thousand kernel weight QTLs identified in a durum $\times$ wild emmer wheat mapping population tested in five

- environments. *Theoretical and Applied Genetics*, 133(1), 119-131.
<https://doi.org/10.1007/s00122-019-03444-8>
- Fatiukha, A., Klymiuk, V., Peleg, Z., Saranga, Y., Cakmak, I., Krugman, T., Korol, A. B., & Fahima, T. (2020b). Variation in phosphorous and sulfur content shapes the genetic architecture and phenotypic associations within the wheat grain ionome. *The Plant Journal*, 101(3), 555-572. <https://doi.org/10.1111/tpj.14554>
- Frittelli, A., Botticella, E., Palombieri, S., Masci, S., Celletti, S., Fontanella, M. C., Astolfi, S., De Vita, P., Volpato, M., & Sestili, F. (2023). The suppression of TdMRP3 genes reduces the phytic acid and increases the nutrient accumulation in durum wheat grain. *Frontiers in Plant Science*, 14, 1079559. <https://doi.org/10.3389/fpls.2023.1079559>
- Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., Sela, H., Fahima, T., & Dubcovsky, J. (2009). A Kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *Science*, 323(5919), 1354-1357.
<https://doi.org/10.1126/science.1166289>
- Fu, Y. B., & Somers, D. J. (2009). Genome-wide reduction of genetic diversity in wheat breeding. *Crop Science*, 49(1), 161-168. <https://doi.org/10.2135/cropsci2008.03.0125>
- Gahlaut, V., Jaiswal, V., Tyagi, B. S., Singh, G., Sareen, S., Balyan, H. S., & Gupta, P. S. (2017). QTL mapping for nine drought-responsive agronomic traits in bread wheat under irrigated and rain-fed environments. *PLoS ONE*, 12(8), e0182857.
<https://doi.org/10.1371/journal.pone.0182857>
- Gao, F., Wen, W., Liu, J., Rasheed, A., Yin, G., Xia, X., Wu, X., & He, Z. (2015). Genome-wide linkage mapping of QTL for yield components, plant height and yield-related physiological traits in the Chinese wheat cross Zhou 8425B/Chinese Spring. *Frontiers in Plant Science*, 6, 1099. <https://doi.org/10.3389/fpls.2015.01099>
- Gao, L., Koo, D.-H., Juliana, P., Rife, T., Singh, D., Lemes da Silva, C., Lux, T., Dorn, K. M., Clinesmith, M., Silva, P., Wang, X., Spannagl, M., Monat, C., Friebe, B., Steuernagel, B., Muehlbauer, G. J., Walkowiak, S., Pozniak, C., Singh, R., ... Poland, J. (2021). The *Aegilops ventricosa* 2N^VS segment in bread wheat: cytology, genomics and breeding. *Theoretical and Applied Genetics*, 134, 529-542. <https://doi.org/10.1007/s00122-020-03712-y>
- Garcia, M., Eckermann, P., Haefele, S., Satija, S., Sznajder, B., Timmins, A., Baumann, U., Wolters, P., Mather, D. E., & Fleury, D. (2019). Genome-wide association mapping of grain yield in a diverse collection of spring wheat (*Triticum aestivum* L.) evaluated in southern Australia. *PLoS ONE*, 14(2), e0211730.
<https://doi.org/10.1371/journal.pone.0211730>
- Gardiner, L.-J., Wingen, L. U., Bailey, P., Joynson, R., Brabbs, T., Wright, J., Higgins, J. D., Hall, N., Griffiths, S., Clavijo, B. J., & Hall, A. (2019). Analysis of the recombination landscape of hexaploid bread wheat reveals genes controlling recombination and gene

- conversion frequency. *Genome Biology*, 20, 69. <https://doi.org/10.1186/s13059-019-1675-6>
- Gardner, H., Onofre, K. F. A., & De Wolf, E. D. (2023). Characterizing the response of *Puccinia striiformis* f. sp. *tritici* to periods of heat stress that are common in Kansas and the Great Plains regions of North America. *Phytopathology*, 113(8), 1457-1464. <https://doi.org/10.1094/PHYTO-12-22-0475-R>
- Garland Campbell, K., Bergman, C. J., Gualberto, D. G., Anderson, J. A., Giroux, M. J., Hareland, G., Fulcher, R. G., Sorrells, M. E., & Finney, P. L. (1999). Quantitative trait loci associated with kernel traits in a soft x hard wheat cross. *Crop Science*, 39, 1184-1195. <https://doi.org/10.2135/cropsci1999.0011183X003900040039x>
- Garvin, D. F., Stack, R. W., & Hansen, J. M. (2009). Quantitative trait locus mapping of increased *Fusarium* head blight susceptibility associated with a wild emmer wheat chromosome. *Phytopathology*, 99(4), 447-452. <https://doi.org/10.1094/PHYTO-99-4-0447>
- Garvin, D. F., Welch, R. M., & Finley, J. W. (2006). Historical shifts in the seed mineral micronutrient concentration of US hard red winter wheat germplasm. *Journal of the Science of Food and Agriculture*, 86(13), 2213-2220. <https://doi.org/10.1002/jsfa.2601>
- Genc, Y., Verbyla, A. P., Torun, A. A., Cakmak, I., Willsmore, K., Wallwork, H., & McDonald, G. K. (2009). Quantitative trait loci analysis of zinc efficiency and grain zinc concentration in wheat using whole genome average interval mapping. *Plant and Soil* 314, 49-66. <https://doi.org/10.1007/s11104-008-9704-3>
- Gheibi, M. N., Malakouti, M. J., Kholdebarin, B., Ghanati, F., Teimouri, S., & Sayadi, R. (2009). Significance of nickel supply for growth and chlorophyll content of wheat supplied with urea or ammonium nitrate. *Journal of Plant Nutrition*, 32(9), 1440-1450. <https://doi.org/10.1080/01904160903092655>
- Gill, B. S., Friebe, B., Raupp, W. J., Wilson, D. L., Cox, T. S., Sears, R. G., Brown-Guedira, G. L., & Fritz, A. K. (2006). Wheat Genetics Resource Center: The first 25 years. In *Advances in Agronomy* (Vol. 89, pp. 73-136). [https://doi.org/10.1016/S0065-2113\(05\)89002-9](https://doi.org/10.1016/S0065-2113(05)89002-9)
- Gilmour, A. R., Cullis, B. R., & Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(3), 269-293. <https://doi.org/10.2307/1400446>
- Golan, G., Oksenberg, A., & Peleg, Z. (2015). Genetic evidence for differential selection of grain and embryo weight during wheat evolution under domestication. *Journal of Experimental Botany*, 66(19), 5703-5711. <https://doi.org/10.1093/jxb/erv249>
- Gonzalez-Hernandez, J. L., Elias, E. M., & Kianian, S. F. (2004). Mapping genes for grain protein concentration and grain yield on chromosome 5B of *Triticum turgidum* var. *dicoccoides*. *Euphytica*, 139(3), 217-225. <https://doi.org/10.1007/s10681-004-3157-4>

- Govta, N., Fatiukha, A., Govta, L., Pozniak, C., Distelfeld, A., Fahima, T., Beckles, D. M., & Krugman, T. (2024). Nitrogen deficiency tolerance conferred by introgression of a QTL derived from wild emmer into bread wheat. *Theoretical and Applied Genetics*, 137(8), 187. <https://doi.org/10.1007/s00122-024-04692-z>
- Granato, I. S., Galli, G., Couto, E. G., Sousa, M. B., Mendonça, L. F., & Fritsche-Neto, R. (2018). snpReady: A tool to assist breeders in genomic analysis. *Molecular Breeding*, 38, 1-6. <https://doi.org/10.1007/s11032-018-0844-8>
- Grela, M., Quarrie, S., Cyganek, K., Bocianowski, J., Karbarz, M., Tyrka, M., Habash, D., Dziurka, M., Kowalczyk, E., Szarski, W., & Czyczyło-Mysza, I. M. (2025). Impact of soil drought on yield and leaf sugar content in wheat: genotypic and phenotypic relationships compared using a doubled haploid population. *International Journal of Molecular Sciences*, 26(16), 7833. <https://doi.org/10.3390/ijms26167833>
- Groos, C., Robert, N., Bervas, E., & Charmet, G. (2003). Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. *Theoretical and Applied Genetics*, 106, 1032-1040. <https://doi.org/10.1007/s00122-002-1111-1>
- Guan, P., Lu, L., Jia, L., Kabir, M. R., Zhang, J., Lan, T., Zhao, Y., Xin, M., Hu, Z., Yao, Y., Ni, Z., Sun, Q., & Peng, H. (2018). Global QTL analysis identifies genomic regions on chromosomes 4A and 4B harboring stable loci for yield-related traits across different environments in wheat (*Triticum aestivum* L.). *Frontiers in Plant Science*, 9, 529. <https://doi.org/10.3389/fpls.2018.00529>
- Gurumurthy, S., Arora, A., Krishna, H., Chinnusamy, V., & Hazra, K. K. (2023). Genotypic capacity of post-anthesis stem reserve mobilization in wheat for yield sustainability under drought and heat stress in the subtropical region. *Frontiers in Genetics*, 14. <https://doi.org/10.3389/fgene.2023.1180941>
- Guttieri, M. J., Baenziger, P. S., Frels, K., Carver, B., Arnall, B., Wang, S., Akhunov, E., & Waters, B. M. (2015a). Prospects for Selecting Wheat with Increased Zinc and Decreased Cadmium Concentration in Grain. *Crop Science*, 55(4), 1712-1728. <https://doi.org/10.2135/cropsci2014.08.0559>
- Guttieri, M. J., Baenziger, S., Frels, K., Carver, B., Arnall, B., & Waters, B. M. (2015b). Variation for Grain Mineral Concentration in a Diversity Panel of Current and Historical Great Plains Hard Winter Wheat Germplasm. *Crop Science*, 55(3), 1035-1052. <https://doi.org/10.2135/cropsci2014.07.0506>
- Guttieri, M. J., Frels, K., Regassa, T., Waters, B. M., & Baenziger, P. S. (2017). Variation for nitrogen use efficiency traits in current and historical great plains hard winter wheat. *Euphytica*, 213, 87. <https://doi.org/10.1007/s10681-017-1869-5>
- Guttieri, M. J., Stark, J. C., O'Brien, K., & Souza, E. (2001). Relative sensitivity of spring wheat grain yield and quality parameters to moisture deficit. *Crop Science*, 41(2), 283-607. <https://doi.org/10.2135/cropsci2001.412327x>

- Habash, D. Z., Bernard, S., Schondelmaier, J., Weyen, J., & Quarrie, S. A. (2007). The genetics of nitrogen use in hexaploid wheat: N utilization, development and yield. *Theoretical and Applied Genetics*, 114, 403-419. <https://doi.org/10.1007/s00122-006-0429-5>
- Hale, I., Zhang, X., Fu, D., & Dubcovsky, J. (2012). Registration of wheat lines carrying the partial stripe rust resistance gene *Yr36* without the *Gpc-B1* allele for high grain protein content. *Journal of Plant Registrations*, 7(1), 108-112. <https://doi.org/10.3198/jpr2012.03.0150crg>
- Haley, S., & Stromberger, J. (2018). A More Meaningful Way to Think About Protein – Grain Protein Deviation. Soil and Crop Sciences Department, Colorado State University.
- Hao, M., Zhang, L., Ning, S., Huang, L., Yuan, Z., Wu, B., Yan, Z., Dai, S., Jiang, B., Zheng, Y., & Liu, D. (2020). The resurgence of introgression breeding, as exemplified in wheat improvement. *Frontiers in Plant Science*, 11, 252. <https://doi.org/10.3389/fpls.2020.00252>
- Hartel, K. D., Berzonsky, W. A., Kianian, S. F., & Ali, S. (2004). Expression of a *Triticum turgidum* var. *dicoccoides* source of *Fusarium* head blight resistance transferred to synthetic hexaploid wheat. *Plant Breeding*, 123(6), 516-519. <https://doi.org/10.1111/j.1439-0523.2004.01025.x>
- Haudry, A., Cenci, A., Ravel, C., Bataillon, T., Brunel, D., Poncet, C., Hochu, I., Poirier, S., Santoni, S., Glémin, S., & David, J. (2007). Grinding up wheat: A massive loss of nucleotide diversity since domestication. *Molecular Biology and Evolution*, 24(7), 1506-1517. <https://doi.org/10.1093/molbev/msm077>
- He, F., Pasam, R. K., Shi, F., Kant, S., Keeble-Gagnère, G., Kay, P., Forrest, K., Fritz, A. K., Hucl, P., Wiebe, K., Wong, D., & Edwards, J. (2019). Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nature Genetics*, 51(5), 896-904. <https://doi.org/10.1038/s41588-019-0382-2>
- Herrera-Foessel, S. A., Singh, R. P., Lillemo, M., Huerta-Espino, J., Bhavani, S., Singh, S., Lan, C., Calvo-Salazar, V., & Lagudah, E. S. (2014). *Lr67/Yr46* confers adult plant resistance to stem rust and powdery mildew in wheat. *Theoretical and Applied Genetics*, 127(4), 781-789. <https://doi.org/10.1007/s00122-013-2256-9>
- Hickey, J. M., Crossa, J., Babu, R., & de los Campos, G. (2012a). Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. *Crop Science*, 52(2), 654-663. <https://doi.org/10.2135/cropsci2011.07.0358>
- Hickey, J. M., Kinghorn, B. P., Tier, B., van der Werf, J. H. J., & Cleveland, M. A. (2012b). A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. *Genetics Selection Evolution*, 44(1), 9. <https://doi.org/10.1186/1297-9686-44-9>
- Hiebert, C. W., Moscou, M. J., Hewitt, T., Steuernagel, B., Hernández-Pinzón, I., Green, P., Pujol, V., Zhang, P., Rouse, M. N., Jin, Y., McIntosh, R. A., Upadhyaya, N., Zhang, J.,

- Bhavani, S., Vrána, J., Karafiátová, M., Huang, L., Fetch, T., Doležel, J., ... Spielmeier, W. (2020). Stem rust resistance in wheat is suppressed by a subunit of the mediator complex. *Nature Communications*, 11, 1123. <https://doi.org/10.1038/s41467-020-14937-2>
- Hu, J., Xiao, G., Jiang, P., Zhao, Y., Zhang, G., Ma, X., Yao, J., Xue, L., Su, P., & Bao, Y. (2022). QTL detection for bread wheat processing quality in a nested association mapping population of semi-wild and domesticated wheat varieties. *BMC Plant Biology*, 22, 129. <https://doi.org/10.1186/s12870-022-03523-x>
- Hu, X., Ren, J., Ren, X., Huang, S., Sabiel, S. A. I., Luo, M., Nevo, E., Fu, C., Peng, J., & Sun, D. (2015). Association of agronomic traits with SNP markers in durum wheat (*Triticum turgidum* L. *durum* (Desf.)). *PLoS ONE*, 10(6), e0130854. <https://doi.org/10.1371/journal.pone.0130854>
- Hu, Y., Burucs, Z., & Schmidhalter, U. (2006). Short-Term Effect of Drought and Salinity on Growth and Mineral Elements in Wheat Seedlings. *Journal of Plant Nutrition*, 29(12), 2227-2243. <https://doi.org/10.1080/01904160600975111>
- Huang, L., Feng, L., He, Y., Tang, Z., He, J., Sela, H., Krugman, T., Fahima, T., Liu, D., & Wu, B. (2019). Variation in stripe rust resistance and morphological traits in wild emmer wheat populations. *Agronomy*, 9(2), 44. <https://doi.org/10.3390/agronomy9020044>
- Huang, M., Liu, X., Zhou, Y., Summers, R. M., & Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience*, 8(2). <https://doi.org/10.1093/gigascience/giy154>
- Huang, X. Q., Cöster, H., Ganai, M. W., & Röder, M. S. (2003). Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, 106, 1379-1389. <https://doi.org/10.1007/s00122-002-1179-7>
- Huerta-Espino, J., Singh, R. P., Germán, S., McCallum, B. D., Park, R. F., Chen, W. Q., Bhardwaj, S. C., & Goyeau, H. (2011). Global status of wheat leaf rust caused by *Puccinia triticina*. *Euphytica*, 179(1), 143-160. <https://doi.org/10.1007/s10681-011-0361-x>
- Hunger, R. M., Edwards, J. T., Bowden, R. L., Yan, L., Rayas-Duarte, P., Bai, G., Horn, G. W., Kolmer, J. A., Giles, K. L., Chen, M.-S., Jin, Y., Osburn, R. D., Bayles, M. B., Seabourn, B. W., Klatt, A. R., & Carver, B. F. (2014). 'Billings' wheat combines early maturity, disease resistance, and desirable grain quality for the Southern Great Plains, USA. *Journal of Plant Registrations*, 8(1), 22-31. <https://doi.org/10.3198/jpr2012.11.0053crc>
- Ihsan, M. Z., El-Nakhlawy, F. S., Ismail, S. M., Fahad, S., & daur, I. (2016). Wheat phenological development and growth studies as affected by drought and late season high temperature stress under arid environment. *Frontiers in Plant Science*, 7(795). <https://doi.org/10.3389/fpls.2016.00795>

- International Wheat Genome Sequencing Consortium. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, 345(6194), 1251788. <https://doi.org/10.1126/science.1251788>
- International Wheat Genome Sequencing Consortium. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403). <https://doi.org/10.1126/science.aar7191>
- Iqbal, M., Hussain, I., Liaqat, H., Ashraf, M. A., & Rasheed, R. (2015). Exogenously applied selenium reduces oxidative stress and induces heat tolerance in spring wheat. *Plant Physiology and Biochemistry*, 94, 95-103. <https://doi.org/10.1016/j.plaphy.2015.05.012>
- Jernigan, K. L., Godoy, J. V., Huang, M., Zhou, Y., Morris, C. F., Garland-Campbell, K. A., Zhang, Z., & Carter, A. H. (2018). Genetic dissection of end-use quality traits in adapted soft white winter wheat. *Frontiers in Plant Science*, 9, 271. <https://doi.org/10.3389/fpls.2018.00271>
- Jernigan, K. L., Morris, C. F., Zemetra, R., Chen, J., Garland-Campbell, K., & Carter, A. H. (2017). Genetic analysis of soft white wheat end-use quality traits in a club by common wheat cross. *Journal of Cereal Science*, 76, 148-156. <https://doi.org/10.1016/j.jcs.2017.06.005>
- Ji, G., Xu, Z., Fan, X., Zhou, Q., Chen, L., Yu, Q., Liao, S., Jiang, C., Feng, B., & Wang, T. (2023). Identification and validation of major QTL for grain size and weight in bread wheat (*Triticum aestivum* L.). *The Crop Journal*, 11(2), 564-572. <https://doi.org/10.1016/j.cj.2022.06.014>
- Jiang, P., Zhang, P.-P., Zhang, X., & Ma, H.-X. (2017). Genetic diversity and association analysis for solvent retention capacity in the accessions derived from soft wheat Ningmai 9. *International Journal of Genomics*, 2017(1), 2413150. <https://doi.org/10.1155/2017/2413150>
- Jiang, Y., Chen, Q., Wang, Y., Guo, Z., Xu, B., Zhu, J., Zhang, Y., Gong, X., Luo, C., Wu, W., Liu, C., Kong, L., Deng, M., Jiang, Q., Lan, X., Wang, J., Chen, G., Zheng, Y., Wei, Y., & Qi, P. (2019). Re-acquisition of the brittle rachis trait via a transposon insertion in domestication gene *Q* during wheat de-domestication. *The New Phytologist*, 224(2), 961-973. <https://doi.org/10.1111/nph.15977>
- Jiang, Y., Jiang, Q., Hao, C., Hou, J., Wang, L., Zhang, H., Zhang, S., Chen, X., & Zhang, X. (2015). A yield-associated gene *TaCWI*, in wheat: its function, selection and evolution in global breeding revealed by haplotype analysis. *Theoretical and Applied Genetics*, 128, 131-143. <https://doi.org/10.1007/s00122-014-2417-5>
- Jin, X., Zou, Z., Wu, Z., Liu, C., Yan, S., Peng, Y., Lei, Z., & Zhou, Z. (2022). Genome-wide association study reveals genomic regions associated with molybdenum accumulation in wheat grains. *Frontiers in Plant Science*, 13, 854966. <https://doi.org/10.3389/fpls.2022.854966>

- Jobson, E. M., Martin, J. M., Schneider, T. M., & Giroux, M. J. (2018). The impact of the *Rht-B1b*, *Rht-D1b*, and *Rht-8* wheat semi-dwarfing genes on flour milling, baking, and micronutrients. *Cereal Chemistry*, 95(6), 770-778. <https://doi.org/10.1002/cche.10091>
- 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. <https://doi.org/10.3389/fnut.2024.1393357>
- Johnston, P. A., Munro, C., Butler, R. C., Browne, J., Gibbs, A., & Shorter, S. (2017). The future of *Lr34* in modern, high-input wheat breeding programs. *Crop Science*, 57(2), 671-680. <https://doi.org/10.2135/cropsci2016.03.0158>
- Jordan, K. W., Wang, S., He, F., Chao, S., Lun, Y., Paux, E., Sourdille, P., Sherman, J., Akhunova, A., Blake, N. K., Pumphrey, M. O., Glover, K., Dubcovsky, J., Talbert, L., & Akhunov, E. D. (2018). The genetic architecture of genome-wide recombination variation in allopolyploid wheat revealed by nested association mapping. *The Plant Journal*, 95, 1039-1054. <https://doi.org/10.1111/tpj.14009>
- Jordan, K. W., Wang, S., Lun, Y., Gardiner, L.-J., MacLachlan, R., Hucl, P., Wiebe, K., Wong, D., Forrest, K. L., IWGS Consortium, Sharpe, A. G., Sidebottom, C. H., Hall, N., Toomajian, C., Close, T., Dubcovsky, J., Akhunova, A., Talbert, L., Bansal, U. K., ... Akhunov, E. (2015). A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biology*, 16, 48. <https://doi.org/10.1186/s13059-015-0606-4>
- Joukhadar, R., Trethowan, R. M., Thistlethwaite, R., Hayden, M. J., Stangoulis, J., Cu, S., Tibbits, J., & Daetwyler, H. D. (2025). Stable pleiotropic loci controlling the accumulation of multiple nutritional elements in wheat. *Theoretical and Applied Genetics*, 138(5), 95. <https://doi.org/10.1007/s00122-025-04877-0>
- Kansas Mesonet. (2025). Kansas Mesonet historical data. Accessed February 2025. <http://mesonet.k-state.edu/weather/historical>
- Kansas Mesonet. Climate basics. Kansas State University. <https://mesonet.k-state.edu/climate/basics/>
- Karim, M. R., Zhang, Y.-Z., Zhao, R.-R., Chen, X.-P., Zhang, F.-S., & Zou, C.-Q. (2012). Alleviation of drought stress in winter wheat by late foliar application of zinc, boron, and manganese. *Journal of Plant Nutrition and Soil Science*, 175, 142-151. <https://doi.org/10.1002/jpln.201100141>
- Kastori, R., Maksimović, I., Denčić, S., Kádár, I., Putnik-Delić, M., & Momčilović, V. (2017). Strontium accumulation in whole grain of *Aegilops* and *Triticum* species. *Journal of Plant Nutrition and Soil Science*, 180(2), 212-219. <https://doi.org/10.1002/jpln.201600435>

- Katamadze, A., Vergara-Díaz, O., Uberegui, E., Yoldi-Achalandabaso, A., Araus, J. L., & Vicente, R. (2023). Evolution of wheat architecture, physiology, and metabolism during domestication and further cultivation: Lessons for crop improvement. *The Crop Journal*, 11(6), 1080-1096. <https://doi.org/10.1016/j.cj.2023.06.006>
- Kato, K., Mori, Y., Beiles, A., & Nevo, E. (1997). Geographical variation in heading traits in wild emmer wheat, *Triticum dicoccoides*. I. Variation in vernalization response and ecological differentiation. *Theoretical and Applied Genetics*, 95(4–5), 546-552. <https://doi.org/10.1007/s001220050595>
- Kato, K., Tanizoe, C., Beiles, A., & Nevo, E. (1998). Geographical variation in heading traits in wild emmer wheat, *Triticum dicoccoides*. II. Variation in heading date and adaptation to diverse eco-geographical conditions. *Hereditas*, 128(1), 33-39. <https://doi.org/10.1111/j.1601-5223.1998.00033.x>
- Katz, A., Byrne, P., Reid, S., Bratschun, S., Haley, S., & Pearce, S. (2022). Identification and validation of a QTL for spikelet number on chromosome arm 6BL of common wheat (*Triticum aestivum* L.). *Molecular Breeding*, 42, 17. <https://doi.org/10.1007/s11032-022-01288-7>
- Kaur, S., Francki, M. G., & Forster, J. W. (2012). Identification, characterization and interpretation of single-nucleotide sequence variation in allopolyploid crop species. *Plant Biotechnology Journal*, 10(2), 125-138. <https://doi.org/10.1111/j.1467-7652.2011.00644.x>
- Keeble-Gagnère, G., Pasam, R., Forrest, K. L., Wong, D., Robinson, H., Godoy, J., ... & Hayden, M. J. (2021). Novel design of imputation-enabled SNP arrays for breeding and research applications supporting multi-species hybridization. *Frontiers in Plant Science*, 12, 756877. <https://doi.org/10.3389/fpls.2021.756877>
- Kirana, R. P., Michel, S., Moreno-Amores, J., Prat, N., Lemmens, M., Buerstmayr, M., Buerstmayr, H., & Steiner, B. (2023). Pyramiding *Fusarium* head blight resistance QTL from *T. aestivum*, *T. dicoccum* and *T. dicoccoides* in durum wheat. *Theoretical and Applied Genetics*, 136(9), 201. <https://doi.org/10.1007/s00122-023-04426-7>
- Klindworth, D. L., Hareland, G. A., Elias, E. M., & Xu, S. S. (2013). Attempted compensation for linkage drag affecting agronomic characteristics of durum wheat 1AS/1DL translocation lines. *Crop Science*, 53(2), 422-429. <https://doi.org/10.2135/cropsci2012.05.0310>
- Klymiuk, V., Yaniv, E., Huang, L., Raats, D., Fatiukha, A., Chen, S., Feng, L., Frenkel, Z., Krugman, T., Lidzbarsky, G., Chang, W., Jääskeläinen, M. J., Schudoma, C., Paulin, L., Laine, P., Bariana, H., Sela, H., Saleem, K., Sørensen, C. K., ... Fahima, T. (2018). Cloning of the wheat *Yr15* resistance gene sheds light on the plant tandem kinase-pseudokinase family. *Nature Communications*, 9(1), 3735. <https://doi.org/10.1038/s41467-018-06138-9>

- Knott, D. R., Bai, D., & Zale, J. (2005). The transfer of leaf and stem rust resistance from wild emmer wheats to durum and common wheat. *Canadian Journal of Plant Science*, 85(1), 49-57. <https://doi.org/10.4141/P03-212>
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9, 29. <https://doi.org/10.1186/1746-4811-9-29>
- Krattinger, S. G., Lagudah, E. S., Spielmeier, W., Singh, R. P., Huerta-Espino, J., McFadden, H., Bossolini, E., Selter, L. L., & Keller, B. (2009). A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science*, 323(5919), 1360-1363. <https://doi.org/10.1126/science.1166453>
- Krishnan J., N., Kaur, S., Kumar, U., Singh, R., Dhillon, G. S., Bhati, P. K., & Chhuneja, P. (2024). Mapping heat tolerance QTLs in *Triticum durum-Aegilops speltoides* backcross introgression lines to enhance thermotolerance in wheat. *Frontiers in Plant Science*, 15, 1485914. <https://doi.org/10.3389/fpls.2024.1485914>
- Krugman, T., Chagué, V., Peleg, Z., Balzergue, S., Just, J., Korol, A. B., Nevo, E., Saranga, Y., Chalhou, B., & Fahima, T. (2010). Multilevel regulation and signalling processes associated with adaptation to terminal drought in wild emmer wheat. *Functional & Integrative Genomics*, 10(2), 187-209. <https://doi.org/10.1007/s10142-010-0166-3>
- Kumar, S., Kumar, S., Sharma, H., Singh, V. P., Rawale, K. S., Kahlon, K. S., Gupta, V., Bhatt, S. K., Vairamani, R., Gill, K. S., & Balyan, H. S. (2024). Physical map of QTL for eleven agronomic traits across fifteen environments, identification of related candidate genes, and development of KASP markers with emphasis on terminal heat stress tolerance in common wheat. *Theoretical and Applied Genetics*, 137, 235. <https://doi.org/10.1007/s00122-024-04748-0>
- Kumar, S., Stack, R. W., Friesen, T. L., & Faris, J. D. (2007). Identification of a novel *Fusarium* head blight resistance quantitative trait locus on chromosome 7A in tetraploid wheat. *Phytopathology*, 97(5), 592-597. <https://doi.org/10.1094/PHYTO-97-5-0592>
- Labuschagne, M., Guzmán, C., Phakela, K., Wentzel, B., & van Biljon, A. (2021). Solvent retention capacity and gluten protein composition of durum wheat flour as influenced by drought and heat stress. *Plants*, 10(5), 1000. <https://doi.org/10.3390/plants10051000>
- Lander, E. S., & Schork, N. J. (1994). Genetic dissection of complex traits. *Science*, 265(5181), 2037-2048. <https://doi.org/10.1126/science.8091226>
- Langridge, P., & Reynolds, M. (2021). Breeding for drought and heat tolerance in wheat. *Theoretical and Applied Genetics*, 134, 1753-1770. <https://doi.org/10.1007/s00122-021-03795-1>
- Latif, S., Wang, L., Khan, J., Ali, Z., Sehgal, S. K., Babar, M. A., Wang, J., & Quraishi, U. M. (2020). Deciphering the Role of Stay-Green Trait to Mitigate Terminal Heat Stress in Bread Wheat. *Agronomy*, 10(7), 1001. <https://doi.org/10.3390/agronomy10071001>

- Lauterberg, M., Saranga, Y., Deblieck, M., Klukas, C., Krugman, T., Perovic, D., Ordon, F., Graner, A., & Neumann, K. (2022). Precision phenotyping across the life cycle to validate and decipher drought-adaptive QTLs of wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) introduced into elite wheat varieties. *Frontiers in Plant Science*, 13, 965287. <https://doi.org/10.3389/fpls.2022.965287>
- Law, C. N., Bhandari, D. G., Salmon, S. E., Greenwell, P. W., Foot, I. M., Cauvain, S. P., Sayers, E. J., & Worland, A. J. (2005). Novel genes on chromosome 3A influencing breadmaking quality in wheat, including a new gene for loaf volume, *Lvl 1*. *Journal of Cereal Science*, 41(3), 317-326. <https://doi.org/10.1016/j.jcs.2004.11.003>
- Lemes da Silva, C., Fritz, A., Clinesmith, M., Poland, J., Dowell, F., & Peiris, K. (2019). QTL mapping of *Fusarium* head blight resistance and deoxynivalenol accumulation in the Kansas wheat variety 'Everest'. *Molecular Breeding*, 39(35). <https://doi.org/10.1007/s11032-019-0937-z>
- Lephuthing, M. C., Khumalo, T. P., Tolmay, V. L., Dube, E., & Tsilo, T. J. (2022). Genetic mapping of quantitative trait loci associated with plant height and yield component traits in a wheat (*Triticum aestivum* L.) doubled haploid population derived from Tugela-DN x Elands. *Agronomy*, 12(10), 2283. <https://doi.org/10.3390/agronomy12102283>
- Lhamo, D., Sun, Q., Zhang, Q., Li, X., Fiedler, J. D., Xia, G., ... & Xu, S. S. (2023). Genome-wide association analyses of leaf rust resistance in cultivated emmer wheat. *Theoretical and Applied Genetics*, 136(1), 20. <https://doi.org/10.1007/s00122-023-04281-6>
- Li, G., Wang, L., Yang, J., He, H., Jin, H., Li, X., Ren, T., Ren, Z., Li, F., Han, X., Zhao, X., Dong, L., Li, Y., Song, Z., Yan, Z., Zheng, N., Shi, C., Wang, Z., Yang, S., ... Wang, D. (2021). A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nature Genetics*, 53, 574-584. <https://doi.org/10.1038/s41588-021-00808-z>
- Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genomes Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, H., Zhang, P., Luo, M., Hoque, M., Chakraborty, S., Brooks, B., Li, J., Singh, S., Forest, K., Binney, A., Zhang, L., Mather, D., & Ayliffe, M. (2023). Introgression of the bread wheat D genome encoded *Lr34/Yr18/Sr57/Pm38/Ltn1* adult plant resistance gene into *Triticum turgidum* (durum wheat). *Theoretical and Applied Genetics*, 136(11), 226. <https://doi.org/10.1007/s00122-023-04466-z>
- Li, H., Zhu, L., Fan, R., Li, Z., Liu, Y., Shaheen, A., Nie, F., Li, C., Liu, X., Li, Y., Liu, W., Yang, Y., Guo, T., Zhu, Y., Bu, M., Li, C., Liang, H., Bai, S., Ma, F., Guo, G., Zhang, Z., Huang, J., Zhou, Y., & Song, C.-P. (2024). A platform for whole-genome speed

- introgression from *Aegilops tauschii* to wheat for breeding future crops. *Nature Protocols*, 19, 281-312. <https://doi.org/10.1038/s41596-023-00922-8>
- Li, L. F., Zhang, Z. B., Wang, Z. H., Li, N., Sha, Y., Wang, X. F., Ding, N., Li, Y., Zhao, J., Wu, Y., Gong, L., Mafessoni, F., Levy, A. A., & Liu, B. (2022). Genome sequences of five Sitopsis species of *Aegilops* and the origin of polyploid wheat B subgenome. *Molecular Plant*, 15(3), 488-503. <https://doi.org/10.1016/j.molp.2021.12.019>
- Li, M., Dong, L., Li, B., Wang, Z., Xie, J., Qiu, D., Li, Y., Shi, W., Yang, L., Wu, Q., Chen, Y., Lu, P., Guo, G., Zhang, H., Zhang, P., Zhu, K., Li, Y., Zhang, Y., Wang, R., ... Liu, Z. (2020). A CNL protein in wild emmer wheat confers powdery mildew resistance. *New Phytologist*, 228(3), 1027-1037. <https://doi.org/10.1111/nph.16761>
- Li, M., Wang, Z., Liang, Z., Shen, W., Sun, F., Xi, Y., & Liu, S. (2015). Quantitative trait loci analysis for kernel-related characteristics in common wheat (*Triticum aestivum* L.). *Crop Science*, 55, 1485-1493. <https://doi.org/10.2135/cropsci2014.09.0616>
- Li, Y., Fu, J., Shen, Q., & Yang, D. (2020). High-molecular-weight glutenin subunits: Genetics, structures, and relation to end use qualities. *International Journal of Molecular Sciences*, 22(1), 184. <https://doi.org/10.3390/ijms22010184>
- Li, Y., Hu, J., Qu, Y., Qiu, D., Lin, H., Du, J., Hou, L., Ma, L., Wu, Q., Zhou, Y., Zhang, H., Yang, L., Liu, H., Liu, Z., Zhou, Y., & Li, H. (2024). Alleles on locus chromosome 4B from different parents confer tiller number and the yield-associated traits in wheat. *BMC Plant Biology*, 24, 454. <https://doi.org/10.1186/s12870-024-05079-4>
- Li, X.-P., Zhao, X.-Q., He, X., Zhao, G.-Y., Li, B., Liu, D.-C., Zhang, A.-M., Zhang, X.-Y., Tong, Y.-P., & Li, Z.-S. (2011). Haplotype analysis of the genes encoding glutamine synthetase plastic isoforms and their association with nitrogen-use and yield-related traits in bread wheat. *New Phytologist*, 189(2), 449-458. <https://doi.org/10.1111/j.1469-8137.2010.03490.x>
- Li, Y. C., Krugman, T., Fahima, T., Beiles, A., Korol, A. B., & Nevo, E. (2001). Spatiotemporal allozyme divergence caused by aridity stress in a natural population of wild wheat, *Triticum dicoccoides*, at the Ammiad microsite, Israel. *Theoretical and Applied Genetics*, 102(6-7), 853-864. <https://doi.org/10.1007/s001220000474>
- Li, Y., Shi, F., Lin, Z., Robinson, H., Moody, D., Rattey, A., Godoy, J., Mullan, D., Keeble-Gagnere, G., Hayden, M. J., Tibbits, J. F. G., & Daetwyler, H. D. (2022). Benefit of introgression depends on level of genetic trait variation in cereal breeding programmes. *Frontiers in Plant Science*, 13, 786452. <https://doi.org/10.3389/fpls.2022.786452>
- Lin, X., Harrington, J. Jr., Ciampitti, I., Gowda, P., Brown, D., & Kisekka, I. (2017). Kansas Trends and Changes in Temperature, Precipitation, Drought, and Frost-Free Days from the 1890s to 2015. *Journal of Contemporary Water Research & Education*, 162(1), 18-30. <https://doi.org/10.1111/j.1936-704X.2017.03257.x>

- Liu, B., Li, L., Fu, C., Zhang, Y., Bai, B., Du, J., Zeng, J., Bian, Y., Liu, S., Song, J., Luo, X., Xie, L., Sun, M., Xu, X., Xia, X., & Cao, S. (2023). Genetic dissection of grain morphology and yield components in a wheat line with defective grain filling. *Theoretical and Applied Genetics*, 136, 165. <https://doi.org/10.1007/s00122-023-04410-1>
- Liu, J., Huang, L., Li, T., Liu, Y., Yan, Z., Tang, G., Zheng, Y., Liu, D., & Wu, B. (2021). Genome-wide association study for grain micronutrient concentrations in wheat advanced lines derived from wild emmer. *Frontiers in Plant Science*, 12, 651283. <https://doi.org/10.3389/fpls.2021.651283>
- Liu, S., Zhou, R., Dong, Y., Li, P., & Jia, J. (2006). Development, utilization of introgression lines using a synthetic wheat as a donor. *Theoretical and Applied Genetics*, 112, 1360-1373. <https://doi.org/10.1007/s00122-006-0238-x>
- Liu, Y., Zhang, J., Hu, Y.-G., & Chen, J. (2017). Dwarfing genes *Rht4* and *Rht-B1b* affect plant height and key agronomic traits in common wheat under two water regimes. *Field Crops Research*, 204, 242-248. <https://doi.org/10.1016/j.fcr.2017.01.020>
- Liu, Z., Sun, Q., Ni, Z., Nevo, E., & Yang, T. (2002). Molecular characterization of a novel powdery mildew resistance gene *Pm30* in wheat originating from wild emmer. *Euphytica*, 123(1), 21-29. <https://doi.org/10.1023/A:1014471113511>
- Lollato, R. P., Edwards, J. T., & Zhang, H. (2013). Effect of Alternative Soil Acidity Amelioration Strategies on Soil pH Distribution and Wheat Agronomic Response. *Soil Science Society of America Journal*, 77(5), 1831-1841. <https://doi.org/10.2136/sssaj2013.04.0129>
- Lollato, R. P., Figueiredo, B. M., Dhillon, J. S., Arnall, D. B., & Raun, W. R. (2019). Wheat grain yield and grain-nitrogen relationships as affected by N, P, and K fertilization: a synthesis of long-term experiments. *Field Crops Research*, 236, 42-57. <https://doi.org/10.1016/j.fcr.2019.03.005>
- Lollato, R., & Knapp, M. (2020). Drought stress taking its toll on Kansas wheat production in 2020. Kansas State University Agronomy eUpdate. <https://eupdate.agronomy.ksu.edu/article/drought-stress-taking-its-toll-on-kansas-wheat-production-in-2020-386-1>
- Lollato, R. P., Ochsner, T. E., Arnall, D. B., Griffin, T. W., & Edwards, J. T. (2019). From Field Experiments to Regional Forecasts: Upscaling Wheat Grain and Forage Yield Response to Acidic Soils. *Agronomy Journal*, 111(1), 287-302. <https://doi.org/10.2134/agronj2018.03.0206>
- Lopes, M. S., & Reynolds, M. P. (2012). Stay-green in spring wheat can be determined by spectral reflectance measurements (normalized difference vegetation index) independently from phenology. *Journal of Experimental Botany*, 63(10), 3789-3798. <https://doi.org/10.1093/jxb/ers071>

- Lou, H., Zhang, R., Liu, Y., Guo, D., Zhai, S., Chen, A., Zhang, Y., Xie, C., You, M., Peng, H., Liang, R., Ni, Z., Sun, Q., & Li, B. (2021). Genome-wide association study of six quality-related traits in common wheat (*Triticum aestivum* L.) under two sowing conditions. *Theoretical and Applied Genetics*, 134, 399-418. <https://doi.org/10.1007/s00122-020-03704-y>
- Luo, M. C., Yang, Z. L., You, F. M., Kawahara, T., Waines, J. G., & Dvorak, J. (2007). The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. *Theoretical and Applied Genetics*, 114(6), 947-959. <https://doi.org/10.1007/s00122-006-0474-0>
- Ma, J., Ren, J., Yuan, X., Yuan, M., Zhang, D., Li, C., Zeng, Q., Wu, J., Han, D., & Jiang, L. (2023). Genome-wide association study reveals the genetic variation and candidate gene for grain calcium content in bread wheat. *Plant Cell Reports*, 42, 1379-1390. <https://doi.org/10.1007/s00299-023-03036-3>
- Maccaferri, M., El-Feki, W., Nazemi, G., Salvi, S., Canè, M. A., Colalongo, M. C., Stefanelli, S., & Tuberosa, R. (2016). Prioritizing quantitative trait loci for root system architecture in tetraploid wheat. *Journal of Experimental Botany*, 67(4), 1161-1178. <https://doi.org/10.1093/jxb/erw039>
- Manjunath, K. K., Krishna, H., Devate, N. B., Sunilkumar, V. P., Patil, S. P., Chauhan, D., Singh, S., Kumar, S., Jain, N., Singh, G. P., & Singh, P. K. (2024). QTL mapping: insights into genomic regions governing component traits of yield under combined heat and drought stress in wheat. *Frontiers in Genetics*, 14, 1282240. <https://doi.org/10.3389/fgene.2023.1282240>
- Marais, G. F., Pretorius, Z. A., Wellings, C. R., McCallum, B., & Marais, A. S. (2005). Leaf rust and stripe rust resistance genes transferred to common wheat from *Triticum dicoccoides*. *Euphytica*, 143(1-2), 115-123. <https://doi.org/10.1007/s10681-005-2911-6>
- Marcotuli, I., Gadaleta, A., Mangini, G., Signorile, A. M., Zacheo, S. A., Blanco, A., Simeone, R., & Colasuonno, P. (2017). Development of a high-density SNP-based linkage map and detection of QTL for β -glucans, protein content, grain yield per spike and heading time in durum wheat. *International Journal of Molecular Sciences*, 18(6), 1329. <https://doi.org/10.3390/ijms18061329>
- Marcotuli, I., Soriano, J. M., Colasuonno, P., Muciaccia, S., & Gadaleta, A. (2025). Identification of new QTLs for quality traits and yield using tetraploid wheat interspecific backcross inbred lines. *BMC Genomics*, 26(1), 1111. <https://doi.org/10.1186/s12864-025-12323-0>
- Martre, P., Porter, J. R., Jamieson, P. D., & Triboï, E. (2003). Modeling grain nitrogen accumulation and protein composition to understand the sink/source regulations of nitrogen remobilization for wheat. *Plant Physiology*, 133(4), 1959-1967. <https://doi.org/10.1104/pp.103.030585>

- Matsuda, R., Iehisa, J. C. M., Sakaguchi, K., Ohno, R., Yoshida, K., & Takumi, S. (2017). Global gene expression profiling related to temperature-sensitive growth abnormalities in interspecific crosses between tetraploid wheat and *Aegilops tauschii*. *PLoS ONE*, 12(5), e0176497. <https://doi.org/10.1371/journal.pone.0176497>
- Matsuoka, Y., & Nasuda, S. (2004). Durum wheat as a candidate for the unknown female progenitor of bread wheat: An empirical study with a highly fertile F1 hybrid with *Aegilops tauschii* Coss. *Theoretical and Applied Genetics*, 109(8), 1710-1717. <https://doi.org/10.1007/s00122-004-1806-6>
- McFadden, E. S., & Sears, E. R. (1946). The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *Journal of Heredity*, 37, 81-89. <https://doi.org/10.1093/oxfordjournals.jhered.a105590>
- McIntyre, C. L., Mathews, K. L., Rattey, A., Chapman, S. C., Drenth, J., Ghaderi, M., Reynolds, M., & Shorter, R. (2010). Molecular detection of genomic regions associated with grain yield and yield-related components in an elite bread wheat cross evaluated under irrigated and rainfed conditions. *Theoretical and Applied Genetics*, 120(3), 527-541. <https://doi.org/10.1007/s00122-009-1173-4>
- Merchuk-Ovnat, L., Barak, V., Fahima, T., Ordon, F., Lidzbarsky, G. A., Krugman, T., & Saranga, Y. (2016a). Ancestral QTL alleles from wild emmer wheat improve drought resistance and productivity in modern wheat cultivars. *Frontiers in Plant Science*, 7, 452. <https://doi.org/10.3389/fpls.2016.00452>
- Merchuk-Ovnat, L., Fahima, T., Krugman, T., & Saranga, Y. (2016b). Ancestral QTL alleles from wild emmer wheat improve grain yield, biomass, and photosynthesis across environments in modern wheat. *Plant Science*, 251, 23-34. <https://doi.org/10.1016/j.plantsci.2016.05.003>
- Milus, E. A., Kristensen, K., & Hovmøller, M. S. (2009). Evidence for increased aggressiveness in a recent widespread strain of *Puccinia striiformis* f. sp. *tritici* causing stripe rust of wheat. *Phytopathology*, 99(1), 89-94. <https://doi.org/10.1094/PHYTO-99-1-0089>
- Milus, E. A., & Seyran, E. (2006). Aggressiveness of *Puccinia striiformis* f. sp. *tritici* isolates in the south-central United States. *Plant Disease*, 90(7), 847-852. <https://doi.org/10.1094/PD-90-0847>
- Miñana-Posada, S., Lorrain, C., McDonald, B. A., & Feurtey, A. (2025). Thermal adaptation in worldwide collections of a major fungal pathogen. *Molecular Plant-Microbe Interactions*, 38(2), 252-264. <https://doi.org/10.1094/MPMI-09-24-0112-FI>
- Miner, G. L., Delgado, J. A., Ippolito, J. A., Johnson, J. J., Kluth, D. L., & Stewart, C. E. (2022). Wheat grain micronutrients and relationships with yield and protein in the U.S. Central Great Plains. *Field Crops Research*, 279, 108453. <https://doi.org/10.1016/j.fcr.2022.108453>

- Miner, G. L., Stewart, C. E., Delgado, J. A., Ippolito, J. A., Mason, R. E., Haley, S. D., Guttieri, M. J., Ainsworth, E. A., McGrath, J. M., & Beebout, S. J. (2024). Global change impacts on mineral nutritional quality of cereal grains: Coordinated datasets and analyses to advance a systems-based understanding. *Field Crops Research*, 310, 109338. <https://doi.org/10.1016/j.fcr.2024.109338>
- Mohammadi, S. A., & Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants – salient statistical tools and considerations. *Crop Science*, 43, 1235-1248. <https://doi.org/10.2135/cropsci2003.1235>
- Möhring, J., & Piepho, H.-P. (2009). Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Science*, 49(2), 823-829. <https://doi.org/10.2135/cropsci2009.02.0083>
- Monaghan, J. M., Snape, J. W., Chojceki, A. J. S., & Kettlewell, P. S. (2001). The use of grain protein deviation for identifying wheat cultivars with high grain protein concentration and yield. *Euphytica*, 122, 309-317. <https://doi.org/10.1023/A:1012961703208>
- Monasterio, I., & Graham, R. D. (2000). Breeding for trace minerals in wheat. *Food and Nutrition Bulletin*, 21(4), 392-396. <https://doi.org/10.1177/156482650002100409>
- Mosleth, E. F., Lillehammer, M., Pellny, T. K., Wood, A. J., Riche, A. B., Hussain, A., Griffiths, S., Hawkesford, M. J., & Shewry, P. R. (2020). Genetic variation and heritability of grain protein deviation in European wheat genotypes. *Field Crops Research*, 255, 107896. <https://doi.org/10.1016/j.fcr.2020.107896>
- Muhu-Din Ahmed, H. G., Sajjad, M., Zeng, Y., Iqbal, M., Habibullah Khan, S., Ullah, A., & Nadeem Akhtar, M. (2020). Genome-wide association mapping through 90K SNP array for quality and yield attributes in bread wheat against water-deficit conditions. *Agriculture*, 10(9), 392. <https://doi.org/10.3390/agriculture10090392>
- Munaro, L. B., Hefley, T. J., DeWolf, E., Haley, S., Fritz, A. K., Zhang, G., ... & Lollato, R. P. (2020). Exploring long-term variety performance trials to improve environment-specific genotype × management recommendations: A case-study for winter wheat. *Field Crops Research*, 255, 107848. <https://doi.org/10.1016/j.fcr.2020.107848>
- Murchie, E. H., Reynolds, M. P., Slafer, G. A., Foulkes, M. J., Acevedo-Siaca, L., McAusland, L., Sharwood, R., Griffiths, S., Flavell, R. B., Gwyn, J., Sawkins, M., & Carmo-Silva, E. (2023). A ‘wiring diagram’ for source strength traits impacting wheat yield potential. *Journal of Experimental Botany*, 74(1), 72-90. <https://doi.org/10.1093/jxb/erac415>
- Mustahsan, W., Guttieri, M. J., Bowden, R. L., Garland-Campbell, K., Jordan, K., Bai, G., & Zhang, G. (2023). Mapping the quantitative field resistance to stripe rust in a hard winter wheat population “Overley” x “Overland”. *Crop Science*, 63(4), 2050-2066. <https://doi.org/10.1002/csc2.20977>
- Nalam, V. J., Vales, M. I., Watson, C. J. W., Kianian, S. F., & Riera-Lizarazu, O. (2006). Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (*Triticum*

- turgidum* L.). *Theoretical and Applied Genetics*, 112(2), 373-381.
<https://doi.org/10.1007/s00122-005-0140-y>
- Narjesi, V., Mardi, M., Hervan, E. M., Azadi, A., Naghavi, M. R., Ebrahimi, M., & Zali, A. A. (2015). Analysis of quantitative trait loci (QTL) for grain yield and agronomic traits in wheat (*Triticum aestivum* L.) under normal and salt stress conditions. *Plant Molecular Biology Reporter*, 33, 2030-2040. <https://doi.org/10.1007/s11105-015-0876-8>
- Negi, C., Kumar, K., Dhariwal, R., Vyas, P., & Vasistha, N. K. (2025). Genome-wide association studies for grain micronutrient concentration in spring wheat (*Triticum aestivum* L.). *Plants*, 14(22), 3472. <https://doi.org/10.3390/plants14223472>
- Nelson, J. C., Andreescu, C., Breseghello, F., Finney, P. L., Gualberto, D. G., Bergman, C. J., Peña, R. J., Perretant, M. R., Leroy, P., Qualset, C. O., & Sorrells, M. E. (2006). Quantitative trait locus analysis of wheat quality traits. *Euphytica*, 149, 145-159. <https://doi.org/10.1007/s10681-005-9062-7>
- Nevo, E., & Chen, G. (2010). Drought and salt tolerances in wild relatives for wheat and barley improvement. *Plant, Cell & Environment*, 33(4), 670-685. <https://doi.org/10.1111/j.1365-3040.2009.02107.x>
- Nevo, E., Fu, Y.-B., Pavlicek, T., Khalifa, S., Tavasi, M., & Beiles, A. (2012). Evolution of wild cereals during 28 years of global warming in Israel. *Proceedings of the National Academy of Sciences*, 109(9), 3412-3415. <https://doi.org/10.1073/pnas.1121411109>
- Nyine, M., Adhikari, E., Clinesmith, M., Aiken, R., Betzen, B., Wang, W., Davidson, D., Yu, Z., Guo, Y., He, F., Akhunova, A., Jordan, K. W., Fritz, A. K., & 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. <https://doi.org/10.3389/fpls.2021.716955>
- Nyine, M., Adhikari, E., Clinesmith, M., Jordan, K. W., Fritz, A. K., & Akhunov, E. (2020). Genomic patterns of introgression in interspecific populations created by crossing wheat with its wild relative. *G3: Genes, Genomes, Genetics*, 10(10), 3651-3663. <https://doi.org/10.1534/g3.120.401479>
- Nyine, M., Davidson, D., Adhikari, E., Clinesmith, M., Wang, H., Akhunova, A., Fritz, A., & Akhunov, E. (2025). Genomic signals of ecogeographic adaptation in a wild relative are associated with improved wheat performance under drought stress. *Genome Biology*, 26, 35. <https://doi.org/10.1186/s13059-025-03500-1>
- Oh, M. W., Roy, S. K., Kamal, A. H. M., Cho, K., Cho, S.-W., Park, C.-S., Choi, J.-S., Komatsu, S., & Woo, S.-H. (2014). Proteome analysis of roots of wheat seedlings under aluminum stress. *Molecular Biology Reports*, 41(2), 671-681. <https://doi.org/10.1007/s11033-013-2905-8>

- Okada, M., Yoshida, K., & Takumi, S. (2017). Hybrid incompatibilities in interspecific crosses between tetraploid wheat and its wild diploid relative *Aegilops umbellulata*. *Plant Molecular Biology*, 95, 625-645. <https://doi.org/10.1007/s11103-017-0677-6>
- Oliver, R. E., Stack, R. W., Miller, J. D., & Cai, X. (2007). Reaction of wild emmer wheat accessions to *Fusarium* head blight. *Crop Science*, 47(2), 893-899. <https://doi.org/10.2135/cropsci2006.08.0531>
- Otto, C. D., Kianian, S. F., Elias, E. M., Stack, R. W., & Joppa, L. R. (2002). Genetic dissection of a major *Fusarium* head blight QTL in tetraploid wheat. *Plant Molecular Biology*, 48(5-6), 625-632. <https://doi.org/10.1023/a:1014821929830>
- Oury, F.-X., & Godin, C. (2007). Yield and grain protein concentration in bread wheat: how to use the negative relationship between the two characters to identify favorable genotypes? *Euphytica*, 157, 45-57. <https://doi.org/10.1007/s10681-007-9395-5>
- Padmanaban, S., Zhang, P., Hare, R. A., Sutherland, M. W., & Martin, A. (2017). Pentaploid wheat hybrids: applications, characterization, and challenges. *Frontiers in Plant Science*, 8. <https://doi.org/10.3389/fpls.2017.00358>
- Pandey, A., Khan, M. K., Hakki, E. E., Thomas, G., Hamurcu, M., Gezgin, S., Gizlenci, O., Akkaya, M. S. (2016). Assessment of genetic variability for grain nutrients from diverse regions: potential for wheat improvement. *SpringerPlus*, 5(1), 1912. <https://doi.org/10.1186/s40064-016-3586-2>
- Parker, K., Wang, Z., Raul, Y., Gao, L., Poland, J., Rudd, J., Ibrahim, A., & Liu, S. (2026). Diverse reference genomes detect variants in the US winter wheat. *The Plant Genome*, 19, e70160. <https://doi.org/10.1002/tpg2.70160>
- Patwa, N., Pandey, V., Gupta, O. P., Yadav, A., Meena, M. R., Ram, S., & Singh, G. (2024). Unravelling wheat genotypic responses: insights into salinity stress tolerance in relation to oxidative stress, antioxidant mechanisms, oxmolyte accumulation and grain quality parameters. *BMC Plant Biology*, 24, 875. <https://doi.org/10.1186/s12870-024-05508-4>
- Peleg, Z., Cakmak, I., Ozturk, L., Yazici, A., Jun, Y., Budak, H., Korol, A. B., Fahima, T., & Saranga, Y. (2009a). Quantitative trait loci conferring grain mineral nutrient concentrations in durum wheat × wild emmer wheat RIL population. *Theoretical and Applied Genetics*, 119(2), 353-369. <https://doi.org/10.1007/s00122-009-1044-z>
- Peleg, Z., Fahima, T., Abbo, S., Krugman, T., Yakir, D., Nevo, E., & Saranga, Y. (2005). Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. *Plant, Cell & Environment*, 28(2), 176-191. <https://doi.org/10.1111/j.1365-3040.2005.01259.x>
- Peleg, Z., Fahima, T., Korol, A. B., Abbo, S., & Saranga, Y. (2011). Genetic analysis of wheat domestication and evolution under domestication. *Journal of Experimental Botany*, 62(14), 5051-5061. <https://doi.org/10.1093/jxb/err206>

- Peleg, Z., Fahima, T., Krugman, T., Abbo, S., Yakir, D., Korol, A. B., & Saranga, Y. (2009b). Genomic dissection of drought resistance in durum wheat $\times$ wild emmer wheat recombinant inbred line population. *Plant, Cell & Environment*, 32(7), 758-779. <https://doi.org/10.1111/j.1365-3040.2009.01956.x>
- Peleg, Z., Saranga, Y., Krugman, T., Abbo, S., Nevo, E., & Fahima, T. (2008a). Allelic diversity associated with aridity gradient in wild emmer wheat populations. *Plant, Cell & Environment*, 31(1), 39-49. <https://doi.org/10.1111/j.1365-3040.2007.01731.x>
- Peleg, Z., Saranga, Y., Yazici, A., Fahima, T., Ozturk, L., Cakmak, I. (2008b). Grain zinc, iron and protein concentrations and zinc-efficiency in wild emmer wheat under contrasting irrigation regimes. *Plant and Soil*, 306, 57-67. <https://doi.org/10.1007/s11104-007-9417-z>
- Peng, J. H., Fahima, T., Röder, M. S., Huang, Q. Y., Dahan, A., Li, Y. C., Grama, A., & Nevo, E. (2000a). High-density molecular map of chromosome region harboring stripe-rust resistance genes *YrH52* and *Yr15* derived from wild emmer wheat, *Triticum dicoccoides*. *Genetica*, 109(3), 199-210. <https://doi.org/10.1023/a:1017573726512>
- Peng, J. H., Fahima, T., Röder, M. S., Li, Y. C., Grama, A., & Nevo, E. (2000b). Microsatellite high-density mapping of the stripe rust resistance gene *YrH52* region on chromosome 1B and evaluation of its marker-assisted selection in the F2 generation in wild emmer wheat. *New Phytologist*, 146(1), 141-154. <https://doi.org/10.1046/j.1469-8137.2000.00617.x>
- Peng, J. H., Sun, D. F., Peng, Y. L., & Nevo, E. (2013). Gene discovery in *Triticum dicoccoides*, the direct progenitor of cultivated wheats. *Cereal Research Communications*, 41(1), 1-22. <https://doi.org/10.1556/CRC.2012.0030>
- Peng, J., Ronin, Y., Fahima, T., Röder, M. S., Li, Y., Nevo, E., & Korol, A. (2003). Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. *Proceedings of the National Academy of Sciences*, 100(5), 2489-2494. <https://doi.org/10.1073/pnas.252763199>
- Peng, J., Sun, D., & Nevo, E. (2011). Wild emmer wheat, *Triticum dicoccoides*, occupies a pivotal position in wheat domestication process. *Australian Journal of Crop Science*, 5(9), 1127-1137. <https://search.informit.org/doi/10.3316/informit.045025500764821>
- Perrier, F., Yan, B., Candaudap, F., Pokrovsky, O. S., Gourdain, E., Meleard, B., Bussière, S., Coriou, C., Robert, T., Nguyen, C., & Cornu, J. Y. (2016). Variability in grain cadmium concentration among durum wheat cultivars: impact of aboveground biomass partitioning. *Plant and Soil*, 404, 307-320. <https://doi.org/10.1007/s11104-016-2847-8>
- Petersen, G., Seberg, O., Yde, M., & Berthelsen, K. (2006). Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution*, 39(1), 70-82. <https://doi.org/10.1016/j.ympev.2006.01.023>
- Petingco, M. C., Casada, M. E., Maghirang, R. G., Thompson, S. A., McNeill, S. G., Montross, M. D., & Turner, A. P. (2018). Influence of kernel shape and size on the packing ratio

- and compressibility of hard red winter wheat. *Transactions of the American Society of Agricultural and Biological Engineers*, 61(4), 1437-1448. <https://doi.org/10.13031/trans.12648>
- Peirce, E. S., Evers, B., Winn, Z. J., Raupp, W. J., Guttieri, M., Fritz, A. K., Poland, J., Akhunov, E., Haley, S., Mason, E., & Nachappa, P. (2024). Identifying novel sources of resistance to wheat stem sawfly in five wild wheat species. *Pest Management Science*, 80, 2976-2990. <https://doi.org/10.1002/ps.8008>
- Pinto, R. S., Lopes, M. S., Collins, N. C., Reynolds, M. P. (2016). Modelling and genetic dissection of staygreen under heat stress. *Theoretical and Applied Genetics*, 129, 2055-2074. <https://doi.org/10.1007/s00122-016-2757-4>
- Platt, A., Vilhjálmsson, B. J., & Nordborg, M. (2010). Conditions under which genome-wide association studies will be positively misleading. *Genetics*, 186(3), 1045-1052. <https://doi.org/10.1534/genetics.110.121665>
- Platten, J. D., & Fritsche-Neto, R. (2023). Optimizing quantitative trait loci introgression in elite rice germplasms: Comparing methods and population sizes to develop new recipients via stochastic simulations. *Plant Breeding*, 142(4), 439-448. <https://doi.org/10.1111/pbr.13118>
- Pont, C., Leroy, T., Seidel, M., Tondelli, A., Duchemin, W., Armisen, D., Lang, D., Bustos-Korts, D., Goué, N., Balfourier, F., Kilian, B., Özkan, H., Waite, D., Dyer, S., Letellier, T., Alaux, M., Waugh, R., Keller, B., & Mayer, K. F. X. (2019). Tracing the ancestry of modern bread wheats. *Nature Genetics*, 51(5), 905-911. <https://doi.org/10.1038/s41588-019-0393-z>
- Potapova, N. A., Timoshchuk, A. N., Tiys, E. S., Vinichenko, N. A., Leonova, I. N., Salina, E. A., & Tsepilov, Y. A. (2023). Multivariate Genome-wide association study of concentrations of seven elements in seeds reveals four new loci in Russian wheat lines. *Plants*, 12(17), 3019. <https://doi.org/10.3390/plants12173019>
- Pourmohammadi, K., Abedi, E., & Hashemi, S. M. B. (2023). Gliadin and glutenin genomes and their effects on the technological aspect of wheat-based products. *Current Research in Food Science*, 7, 100622. <https://doi.org/10.1016/j.crfs.2023.100622>
- Pradhan, G. P., Prasad, P. V. V., Fritz, A. K., Kirkham, M. B., & Gill, B. S. (2012). Effects of drought and high temperature stress on synthetic hexaploid wheat. *Functional Plant Biology*, 39(3), 190-198. <https://doi.org/10.1071/FP11245>
- Pretini, N., Vanzetti, L. S., Terrile, I. I., Donaire, G., & González, F. G. (2021). Mapping QTL for spike fertility and related traits in two doubled haploid wheat (*Triticum aestivum* L.) populations. *BMC Plant Biology*, 21, 353. <https://doi.org/10.1186/s12870-021-03061-y>
- Price, J. H., Guttieri, M. J., Nyine, M., & Akhunov, E. (2025a). Development and characterization of a wild emmer wheat backcross introgression population for hard

- winter wheat improvement. *The Plant Genome*, 18(3), e70104.
<https://doi.org/10.1002/tpg2.70104>
- Price, J. H., Guttieri, M. J., & Stutz, S. (2025b). Random forest regression to predict Farinograph traits from GlutoPeak output in wheat wild relative backcross lines. *Cereal Chemistry*, 102, 466-476. <https://doi.org/10.1002/cche.10860>
- Prokopenko, D., Hecker, J., Silverman, E. K., Pagano, M., Nöthen, M. M., Dina, C., Lange, C., & Fier, H. L. (2016). Utilizing the Jaccard index to reveal population stratification in sequencing data: a simulation study and an application to the 1000 Genomes Project. *Bioinformatics*, 32(9), 1366-1372. <https://doi.org/10.1093/bioinformatics/btv752>
- Puttamadanayaka, S., Harikrishna, Balaramaiah, M., Biradar, S., Parmeshwarappa, S. V., Sinha, N., Prasad, S. V. S., Mishra, P. C., Jain, N., Singh, P. K., Singh, G. P., & Prabhu, K. V. (2020). Mapping genomic regions of moisture deficit stress tolerance using backcross inbred lines in wheat (*Triticum aestivum* L.). *Scientific Reports*, 10, 21646.
<https://doi.org/10.1038/s41598-020-78671-x>
- Qi, L., Friebe, B., Zhang, P., & Gill, B. S. (2007). Homoeologous recombination, chromosome engineering and crop improvement. *Chromosome research*, 15, 3-19.
<https://doi.org/10.1007/s10577-006-1108-8>
- Qiagen. (2005). BioSprint DNA plant handbook. Qiagen.
- Qie, Y., Liu, Y., Wang, M., Li, X., See, D. R., An, D., & Chen, X. (2019). Development, validation, and re-selection of wheat lines with pyramided genes *Yr64* and *Yr15* linked on the short arm of chromosome 1B for resistance to stripe rust. *Plant Disease*, 103(1), 51-58. <https://doi.org/10.1094/PDIS-03-18-0470-RE>
- Qin, X., Xia, Y., Hu, C., Yu, M., Shabala, S., Wu, S., Tan, Q., Xu, S., & Sun, X. (2021). Ionomics analysis provides new insights into the co-enrichment of cadmium and zinc in wheat grains. *Ecotoxicology and Environmental Safety*, 223, 112623.
<https://doi.org/10.1016/j.ecoenv.2021.112623>
- Rabbi, S. M. H. A., Kumar, A., Naraghi, S. M., Simsek, S., Sapkota, S., Solanki, S., Alamri, M. S., Elias, E. M., Kianian, S., Missaoui, A., & Mergoum, M. (2021). Genome-wide association mapping for yield and related traits under drought stressed and non-stressed environments in wheat. *Frontiers in Genetics*, 12, 649988.
<https://doi.org/10.3389/fgene.2021.649988>
- Raghunandan, K., Tanwar, J., Patil, S. N., Chandra, A. K., Tyagi, S., Agarwal, P., Mallick, N., Murukan, N., Kumari, J., Sahu, T. K., Jacob, S. R., Kumar, A., Yadav, S., Nyamgoud, S., Vinod, Singh, A. K., & Jha, S. K. (2022). Identification of novel broad-spectrum leaf rust resistance sources from Khapli wheat landraces. *Plants*, 11(15), 1965.
<https://doi.org/10.3390/plants11151965>
- Rahman, M. M., Crain, J., Haghighattalab, A., Singh, R. P., & Poland, J. (2021). Improving wheat yield prediction using secondary traits and high-density phenotyping under heat-

- stressed environments. *Frontiers in Plant Science*, 12, 633651. <https://doi.org/10.3389/fpls.2021.633651>
- Rahman, S., Islam, S., Nevo, E., Saieed, M. A. U., Liu, Q., Varshney, R. K., & Ma, W. (2023). Characterizing agronomic and shoot morphological diversity across 263 wild emmer wheat accessions. *Agriculture*, 13, 759. <https://doi.org/10.3390/agriculture13040759>
- Rahman, S., Islam, S., Yu, Z., She, M., Nevo, E., & Ma, W. (2020). Current progress in understanding and recovering the wheat genes lost in evolution and domestication. *International Journal of Molecular Sciences*, 21(16), 5836. <https://doi.org/10.3390/ijms21165836>
- Rana, M., Sankhyan, N. K., Thakur, P., Babal, B., Anjali, Sharma, S., Kumari, S., & Kumar, P. (2025). Molybdenum in soil-plant system: bioavailability, dynamics and implications for sustainable crop production. *Discover Soil*, 2, 89. <https://doi.org/10.1007/s44378-025-00118-4>
- Rasheed, A., Xia, X., Yan, Y., Appels, R., Mahmood, T., & He, Z. (2014). Wheat seed storage proteins: Advances in molecular genetics, diversity and breeding applications. *Journal of Cereal Science*, 60(1), 11-24. <https://doi.org/10.1016/j.jcs.2014.01.020>
- Reader, S. M., & Miller, T. E. (1991). The introduction into wheat of a major gene for resistance to powdery mildew from *Triticum dicoccoides*. *Euphytica*, 53(1), 57-60. <https://doi.org/10.1007/BF00032033>
- Rebetzke, G. J., Ellis, M. H., Bonnett, D. G., Mickelson, B., Condon, A. G., & Richards, R. A. (2012). Height reduction and agronomic performance for selected gibberellin-responsive dwarfing genes in bread wheat (*Triticum aestivum* L.). *Field Crops Research*, 126, 87-96. <https://doi.org/10.1016/j.fcr.2011.09.022>
- Ren, J., Chen, L., Sun, D., You, F. M., Wang, J., Peng, Y., Nevo, E., Beiles, A., Sun, D., Luo, M. C., & Peng, J. (2013). SNP-revealed genetic diversity in wild emmer wheat correlates with ecological factors. *BMC Evolutionary Biology*, 13, 169. <https://doi.org/10.1186/1471-2148-13-169>
- Reynolds, M. P., Pask, A. J. D., Hoppitt, W. J. E., Sonder, K., Sukumaran, S., Molero, G., Saint Pierre, C., Payne, T., Singh, R. P., Braun, H. J., Gonzalez, F. G., Terrile, I. I., Barma, N. C. D., Hakim, A., He, Z., Fan, Z., Novoselovic, D., Maghraby, M., Gad, K. I. M., ... Joshi, A. K. (2017). Strategic crossing of biomass and harvest index – source and sink – achieves genetic gains in wheat. *Euphytica*, 213(11), 257. <https://doi.org/10.1007/s10681-017-2040-z>
- Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A., & Eilers, P. H. C. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics*, 23, 52-71. <https://doi.org/10.1016/j.spasta.2017.10.003>

- Royo, C., Ammar, K., Villegas, D., & Soriano, J. M. (2021). Agronomic, physiological and genetic changes associated with evolution, migration and modern breeding in durum wheat. *Frontiers in Plant Science*, 12, 674470. <https://doi.org/10.3389/fpls.2021.674470>
- Rudd, J. C., Horsley, R. D., McKendry, A. L., & Elias, E. M. (2001). Host plant resistance genes for *Fusarium* head blight: sources, mechanisms, and utility in conventional breeding systems. *Crop Science*, 41(3), 620-627. <https://doi.org/10.2135/cropsci2001.413620x>
- Sadras, V. O., & Slafer, G. A. (2011). Environmental modulation of yield components in cereals: Heritabilities reveal a hierarchy of phenotypic plasticities. *Field Crops Research*, 127, 215-224. <https://doi.org/10.1016/j.fcr.2011.11.014>
- Salt, D. E., Baxter, I., & Lahner, B. (2008). Ionomics and the study of the plant ionome. *Annual Review of Plant Biology*, 59, 709-733. <https://doi.org/10.1146/annurev.arplant.59.032607.092942>
- Sánchez-León, S., Gil-Humanes, J., Ozuna, C. V., Giménez, M. J., Sousa, C., Voytas, D. F., & Barro, F. (2018). Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnology Journal*, 16(4), 902-910. <https://doi.org/10.1111/pbi.12837>
- Schaart, J. G., Salentijn, E. M. J., Goryunova, S. V., Chidzanga, C., Esselink, D. G., Gosman, N., Bentley, A. R., Gilissen, L. J. W. J., & Smulders, M. J. M. (2021). Exploring the alpha-gliadin locus: the 33-mer peptide with six overlapping coeliac disease epitopes in *Triticum aestivum* is derived from a subgroup of *Aegilops tauschii*. *The Plant Journal*, 106(1), 86-94. <https://doi.org/10.1111/tpj.15147>
- Schiavon, M., Nardi, S., dalla Vecchia, F., & Ertani, A. (2020). Selenium biofortification in the 21st century: status and challenges for healthy human nutrition. *Plant and Soil*, 453, 245-270. <https://doi.org/10.1007/s11104-020-04635-9>
- Schmidt, J., Tricker, P. J., Eckermann, P., Kalambettu, P., Garcia, M., & Fleury, D. (2020). Novel alleles for combined drought and heat stress tolerance in wheat. *Frontiers in Plant Science*, 10, 1800. <https://doi.org/10.3389/fpls.2019.01800>
- Schroder, J. L., Zhang, H., Girma, K., Raun, W. R., Penn, C.J., & Payton, M. E. (2011). Soil Acidification from Long-Term Use of Nitrogen Fertilizers on Winter Wheat. *Soil Science Society of America Journal*, 75(3), 957-964. <https://doi.org/10.2136/sssaj2010.0187>
- Seabourn, B. W., Xiao, Z. S., Tilley, M., Herald, T. J., & Park, S. H. (2012). A Rapid, Small-Scale Sedimentation Method to Predict Breadmaking Quality of Hard Winter Wheat. *Crop Science*, 52(3), 1306-1315. <https://doi.org/10.2135/cropsci2011.04.0210>
- Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., & Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature genetics*, 44(7), 825-830. <https://doi.org/10.1038/ng.2314>
- Sela, H., Ezrati, S., Ben-Yehuda, P., Manisterski, J., Akhunov, E., Dvorak, J., Breiman, A., & Korol, A. (2014). Linkage disequilibrium and association analysis of stripe rust resistance

- in wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) population in Israel. *Theoretical and Applied Genetics*, 127(11), 2453-2463. <https://doi.org/10.1007/s00122-014-2389-5>
- Sela, H., Ezrati, S., Manisterski, J., Korol, A. B., & Fahima, T. (2011). Rapid linkage disequilibrium decay in the *Lr10* gene in wild emmer wheat (*Triticum dicoccoides*) populations. *Theoretical and Applied Genetics*, 122(4), 175-187. <https://doi.org/10.1007/s00122-010-1434-2>
- Seregin, I. V., & Kozhevnikova, A. D. (2004). Strontium Transport, Distribution, and Toxic Effects on Maize Seedling Growth. *Russian Journal of Plant Physiology*, 51(2), 215-221. <https://doi.org/10.1023/B:RUPP.0000019217.89936.e7>
- Shahin, A. A., Omara, R. I., Omar, H. A., El-Din, H. S., Sehsah, M. D., Essa, T., Zayton, M. A., & Omar, H. S. (2024). Evaluation of effectiveness of resistance genes in wheat genotypes using marker-assisted selection for stripe rust resistance breeding. *BMC Plant Biology*, 24(1), 785. <https://doi.org/10.1186/s12870-024-05470-1>
- Shaltouki-Rizi, M., Dewitt, N., & Mohammadi, M. (2026). Mapping recombination cold spots in wheat via meiotic recombination in a large biparental population. *G3: Genes, Genomes, Genetics*, jkag097. <https://doi.org/10.1093/g3journal/jkag097>
- Shan, Y., & Osborne, C. P. (2024). Diversification of quantitative morphological traits in wheat. *Annals of Botany*, 133(3), 413-426. <https://doi.org/10.1093/aob/mcad202>
- Sharma, S., Kumar, T., Foulkes, M. J., Orford, S., Singh, A. M., Wingen, L. U., Karnam, V., Nair, L. S., Mandal, P. K., Griffiths, S., Hawkesford, M. J., Shewry, P. R., Bentley, A. R., & Pandey, R. (2023). Nitrogen uptake and remobilization from pre- and post-anthesis stages contribute towards grain yield and grain protein concentration in wheat grown in limited nitrogen conditions. *CABI Agriculture and Bioscience*, 4(1), 12. <https://doi.org/10.1186/s43170-023-00153-7>
- Shavrukov, Y., Langridge, P., Tester, M., & Nevo, E. (2010). Wide genetic diversity of salinity tolerance, sodium exclusion and growth in wild emmer wheat, *Triticum dicoccoides*. *Breeding Science*, 60(4), 426-435. <https://doi.org/10.1270/jsbbs.60.426>
- Shelden, M. C., & Roessner, U. (2013). Advances in functional genomics for investigating salinity stress tolerance mechanisms in cereals. *Frontiers in Plant Science*, 4, 123. <https://doi.org/10.3389/fpls.2013.00123>
- Shi, S., Azam, F. I., Li, H., Chang, X., Li, B., & Jing, R. (2017). Mapping QTL for stay-green and agronomic traits in wheat under diverse water regimes. *Euphytica*, 213, 246. <https://doi.org/10.1007/s10681-017-2002-5>
- Shokat, S., Großkinsky, D. K., Singh, S., & Liu, F. (2023). The role of genetic diversity and pre-breeding traits to improve drought and heat tolerance of bread wheat at the reproductive stage. *Food and Energy Security*, 12(6), e478. <https://doi.org/10.1002/fes3.478>

- Shtangeeva, I., Steinnes, E., & Lierhagen, S. (2011). Macronutrients and trace elements in rye and wheat: Similarities and differences in uptake and relationships between elements. *Environmental and Experimental Botany*, 70(2-3), 259-265. <https://doi.org/10.1016/j.envexpbot.2010.09.013>
- Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z., Tai, Y. S., Gill, B. S., & Faris, J. D. (2006). Molecular characterization of the major wheat domestication gene *Q*. *Genetics*, 172(1), 547-555. <https://doi.org/10.1534/genetics.105.044727>
- Singh, N., Wu, S., Tiwari, V., Sehgal, S., Raupp, J., Wilson, D., Abbasov, M., Gill, B., & Poland, J. (2019). Genomic analysis confirms population structure and identifies inter-lineage hybrids in *Aegilops tauschii*. *Frontiers in Plant Science*, 10, 9. <https://doi.org/10.3389/fpls.2019.00009>
- Singh, R. P., Huerta-Espino, J., & William, H. M. (2005). Genetics and breeding for durable resistance to leaf and stripe rusts in wheat. *Turkish Journal of Agriculture and Forestry*, 29(2), 121-127. Available at: <https://journals.tubitak.gov.tr/agriculture/vol29/iss2/4>
- Singh, S., Gupta, A. K., & Kaur, N. (2012). Influence of Drought and Sowing Time on Protein Composition, Antinutrients, and Mineral Contents of Wheat. *The Scientific World Journal*, 2012, 485751. <https://doi.org/10.1100/2012/485751>
- Singh, S. P., Vogel-Mikuš, K., Arčon, I., Vavpetič, P., Jeromel, L., Pelicon, P., Kumar, J., & Tuli, R. (2013). Pattern of iron distribution in maternal and filial tissues in wheat grains with contrasting levels of iron. *Journal of Experimental Botany*, 64(11), 3249-3260. <https://doi.org/10.1093/jxb/ert160>
- Sisi, N. A., Stein, N., Himmelbach, A., & Mohammadi, S. A. (2022). High-density linkage mapping of agronomic trait QTLs in wheat under water deficit condition using genotyping by sequencing (GBS). *Plants*, 11(19), 2533. <https://doi.org/10.3390/plants11192533>
- Sittel, M. (2023). Average Precipitation Distribution: A Kansas Climate Primer. Kansas State University Agronomy eUpdate. <https://eupdate.agronomy.ksu.edu/article/average-precipitation-distribution-a-kansas-climate-primer-594-3>
- Sivasamy, M., Aparna, M., Kumar, J., Jayaprakash, P., Vikas, V. K., John, P., Nisha, R., Saryaprakash, Sivan, K., & Punniakotti. (2014). Phenotypic and molecular confirmation of durable adult plant leaf rust resistance (APR) genes *Lr34+*, *Lr46+* and *Lr67+* linked to leaf tip necrosis (LTN) in select registered Indian wheat (*T. aestivum*) genetic stocks. *Cereal Research Communications*, 42(2), 262-273. <https://doi.org/10.1556/CRC.2013.0054>
- Slafer, G. A., Foulkes, M. J., Reynolds, M. P., Murchie, E. H., Carmo-Silva, E., Flavell, R., Gwyn, J., Sawkins, M., & Griffiths, S. (2023). A 'wiring diagram' for sink strength traits impacting wheat yield potential. *Journal of Experimental Botany*, 74(1), 40-71. <https://doi.org/10.1093/jxb/erac410>

- Smith, A., Cullis, B., & Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*, 57(4), 1138-1147. <https://doi.org/10.1111/j.0006-341X.2001.01138.x>
- 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. <https://doi.org/10.1017/S0021859605005587>
- Smith, N., Guttieri, M., Souza, E., Shoots, J., Sorrells, M., & Sneller, C. (2011). Identification and validation of QTL for grain quality traits in a cross of soft wheat cultivars Pioneer Brand 25R26 and Foster. *Crop Science*, 51, 1424-1436. <https://doi.org/10.2135/cropsci2010.04.0193>
- Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online. Accessed July 2026.
- Song, Q., Cui, W., Song, J., Zhou, B., Chen, L., Xu, K., & Jin, Y. (2026). Uncover the genetic basis of processing quality related traits in common wheat (*Triticum aestivum* L.) using genome-wide association mapping. *Frontiers in Plant Science*, 17, 1755182. <https://doi.org/10.3389/fpls.2026.1755182>
- Soresi, D., Zappacosta, D., Garayalde, A., Irigoyen, I., Basualdo, J., & Carrera, A. (2017). A valuable QTL for *Fusarium* head blight resistance from *Triticum turgidum* L. ssp. *dicoccoides* has a stable expression in durum wheat cultivars. *Cereal Research Communications*, 45(2), 234-247. <https://doi.org/10.1556/0806.45.2017.007>
- Souza, J. L. B., Antonangelo, J. A., Zhang, H., Reed, V., Finch, B., & Arnall, B. (2023). Impact of long-term fertilization in no-till on the stratification of soil acidity and related parameters. *Soil & Tillage Research*, 228, 105624. <https://doi.org/10.1016/j.still.2022.105624>
- Spencer, C. C. A., Su, Z., Donnelly, P., & Marchini, J. (2009). Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. *PLoS Genetics*, 5(5), e1000477. <https://doi.org/10.1371/journal.pgen.1000477>
- Srinivasa, J., Arun, B., Mishra, V. K., Singh, G. P., Velu, G., Babu, R., Vasistha, N. K., & Joshi, A. K. (2014). Zinc and iron concentration QTL mapped in a *Triticum spelta* x *T. aestivum* cross. *Theoretical and Applied Genetics*, 127, 1643-1651. <https://doi.org/10.1007/s00122-014-2327-6>
- Strejčková, B., Mazzucotelli, E., Čegan, R., Milec, Z., Brus, J., Çakır, E., Mastrangelo, A. M., Özkan, H., & Šafář, J. (2023). Wild emmer wheat, the progenitor of modern bread wheat, exhibits great diversity in the VERNALIZATION1 gene. *Frontiers in Plant Science*, 13, 1106164. <https://doi.org/10.3389/fpls.2022.1106164>
- Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P., & Reynolds, M. P. (2015). Genome-wide association study for grain yield and related traits in an elite spring wheat

- population grown in temperate irrigated environments. *Theoretical and Applied Genetics*, 128, 353-363. <https://doi.org/10.1007/s00122-014-2435-3>
- Sukumaran, S., Lopes, M., Dreisigacker, S., & Reynolds, M. (2017). Genetic analysis of multi-environmental spring wheat trials identifies genomic regions for locus-specific trade-offs for grain weight and grain number. *Theoretical and Applied Genetics*, 131, 985-998. <https://doi.org/10.1007/s00122-017-3037-7>
- Summers, R. W., & Brown, J. K. M. (2013). Constraints on breeding for disease resistance in commercially competitive wheat cultivars. *Plant Pathology*, 62, 115-121. <https://doi.org/10.1111/ppa.12165>
- Sun, G. L., Fahima, T., Korol, A. B., Turpeinen, T., Grama, A., Ronin, Y. I., & Nevo, E. (1997). Identification of molecular markers linked to the *Yr15* stripe rust resistance gene of wheat originated in wild emmer wheat, *Triticum dicoccoides*. *Theoretical and Applied Genetics*, 95(5-6), 622-628. <https://doi.org/10.1007/s001220050604>
- Sun, X.-Y., Wu, K., Zhao, Y., Kong, F.-M., Han, G.-Z., Jiang, H.-M., Huang, X.-J., Li, R.-J., Wang, H.-G., & Li, S.-S. (2009). QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. *Euphytica*, 165, 615-624. <https://doi.org/10.1007/s10681-008-9794-2>
- Suneja, Y., Gupta, A. K., & Bains, N. S. (2019). Stress adaptive plasticity: *Aegilops tauschii* and *Triticum dicoccoides* as potential donors of drought-associated morpho-physiological traits in wheat. *Frontiers in Plant Science*, 10, 211. <https://doi.org/10.3389/fpls.2019.00211>
- Sunic, K., & Spanic, V. (2024). Genetic Biofortification of Winter Wheat with Selenium (Se). *Plants*, 13(13), 1816. <https://doi.org/10.3390/plants13131816>
- Suprayogi, Y., Pozniak, C. J., Clarke, F. R., Clarke, J. M., Knox, R. E., & Singh, A. K. (2009). Identification and validation of quantitative trait loci for grain protein concentration in adapted Canadian durum wheat populations. *Theoretical and Applied Genetics*, 119, 437-448. <https://doi.org/10.1007/s00122-009-1050-1>
- Tene, M., Adhikari, E., Cobo, N., Jordan, K. W., Matny, O., del Blanco, I. A., Roter, J., Ezrati, S., Govta, L., Manisterski, J., et al. (2022). GWAS for stripe rust resistance in wild emmer wheat (*Triticum dicoccoides*) population: Obstacles and solutions. *Crops*, 2(1), 42-61. <https://doi.org/10.3390/crops2010005>
- Tessele, A., González-Diéguez, D., Crossa, J., Johnson, B. E., Morris, G. P., & Fritz, A. K. (2025a). Improving genomic selection in hexaploid wheat with sub-genome additive and epistatic models. *G3: Genes, Genomes, Genetics*, 15(4), jkaf031. <https://doi.org/10.1093/g3journal/jkaf031>
- Tessele, A., Morris, G., Akhunov, E., Johnson, B., Clinesmith, M., Carroll, A. H., & Fritz, A. (2025b). Interchromosomal linkage disequilibrium analysis reveals strong indications of

- sign epistasis in wheat breeding families. *G3: Genes, Genomes, Genetics*, 16(1), jkaf265. <https://doi.org/10.1093/g3journal/jkaf265>
- Tibbs Cortes, L., Zhang, Z., & Yu, J. (2020). Status and prospects of genome-wide association studies in plants. *The Plant Genome*, 14(1), e20077. <https://doi.org/10.1002/tpg2.20077>
- Tiwari, V. K., Wang, S., Sehgal, S., Vrána, J., Friebe, B., Kubaláková, M., Chhuneja, P., Doležel, J., Akhunov, E., Kalia, B., Sabir, J., & Gill, B. S. (2014). SNP discovery for mapping alien introgressions in wheat. *BMC Genomics*, 15, 273. <https://doi.org/10.1186/1471-2164-15-273>
- Triboï, E., Martre, P., & Triboï-Blondel, A.-M. (2003). Environmentally-induced changes in protein composition in developing grains of wheat are related to changes in total protein content. *Journal of Experimental Botany*, 54(388), 1731-1742. <https://doi.org/10.1093/jxb/erg183>
- Turkmen, A. S., Yuan, Y., & Billor, N. (2019). Evaluation of methods for adjusting population stratification in genome-wide association studies: Standard versus categorical principal component analysis. *Annals of Human Genetics*, 83(6), 454-464. <https://doi.org/10.1111/ahg.12339>
- Tzarfati, R., Barak, V., Krugman, T., Fahima, T., Abbo, S., Saranga, Y., Korol, A. B. (2014). Novel quantitative trait loci underlying major domestication traits in tetraploid wheat. *Molecular Breeding*, 34, 1613-1628. <https://doi.org/10.1007/s11032-014-0182-4>
- Uauy, C., Brevis, J. C., Chen, X., Khan, I., Jackson, L., Chicaiza, O., Distelfeld, A., Fahima, T., & Dubcovsky, J. (2005). High-temperature adult-plant (HTAP) stripe rust resistance gene *Yr36* from *Triticum turgidum* ssp. *dicoccoides* is closely linked to the grain protein content locus *Gpc-B1*. *Theoretical and Applied Genetics*, 112(1), 97-105. <https://doi.org/10.1007/s00122-005-0109-x>
- Uauy, C., Distelfeld, A., Fahima, T., Blechl, A., & Dubcovsky, J. (2006). A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat. *Science*, 314(5803), 1298-1301. <https://doi.org/10.1126/science.1133649>
- Vallejo, A. J., Peralta, M. L., & Santa-Maria, G. E. (2005). Expression of potassium-transporter coding genes, and kinetics of rubidium uptake, along a longitudinal root axis. *Plant, Cell and Environment*, 28(7), 850-862. <https://doi.org/10.1111/j.1365-3040.2005.01334.x>
- Velu, G., Singh, R. P., Huerta, J., & Guzmán, C. (2017). Genetic impact of *Rht* dwarfing genes on grain micronutrients concentration in wheat. *Field Crops Research*, 214, 373-377. <https://doi.org/10.1016/j.fcr.2017.09.030>
- Vukasovic, S., Alahmad, S., Christopher, J., Snowdon, R. J., Stahl, A., & Hickey, L. T. (2022). Dissecting the genetics of early vigour to design drought-adapted wheat. *Frontiers in Plant Science*, 12, 754439. <https://doi.org/10.3389/fpls.2021.754439>

- Vuorinen, A. L., Kalendar, R., Fahima, T., Korpelainen, H., Nevo, E., & Schulman, A. H. (2018). Retrotransposon-based genetic diversity assessment in wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*). *Agronomy*, 8(7), 107. <https://doi.org/10.3390/agronomy8070107>
- Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., Ramirez-Gonzalez, R. H., Kolodziej, M. C., Delorean, E., Thambugala, D., Klymiuk, V., Byrns, B., Gundlach, H., Bandi, V., Siri, J. N., Nilsen, K., Aquino, C., Himmelbach, A., Copetti, D., ... Pozniak, C. J. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature*, 588, 277-283. <https://doi.org/10.1038/s41586-020-2961-x>
- Wang, H., Bernardo, A., St. Amand, P., Bai, G., Bowden, R. L., Guttieri, M. J., & Jordan, K. W. (2023). Skim exome capture genotyping in wheat. *The Plant Genome*, 16(4), e20381. <https://doi.org/10.1002/tpg2.20381>
- Wang, J., Luo, M.-C., Chen, Z., You, F. M., Wei, Y., Zheng, Y., & Dvorak, J. (2013). *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytologist*, 198(3), 925-937. <https://doi.org/10.1111/nph.12164>
- Wang, J., Shi, X., Zhou, Z., Qin, M., Wang, Y., Li, W., Yang, P., Wu, Z., & Lei, Z. (2022). Genetic dissection of grain iron concentration in hexaploid wheat (*Triticum aestivum* L.) using a genome-wide association analysis method. *PeerJ*, 10, e13625. <https://doi.org/10.7717/peerj.13625>
- Wang, J.-Y., Xiong, Y.-C., Li, F.-M., Siddique, K. H. M., & Turner, N. C. (2017). Effects of Drought Stress on Morphophysiological Traits, Biochemical Characteristics, Yield, and Yield Components in Different Ploidy Wheat: A Meta-Analysis. *Advances in Agronomy*, 143, 139-173. <https://doi.org/10.1016/bs.agron.2017.01.002>
- Wang, J., Zhang, Z. (2021). GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics & Bioinformatics*, 19(4), 629-640. <https://doi.org/10.1016/j.gpb.2021.08.005>
- Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., Maccaferri, M., Salvi, S., Milner, S. G., Cattivelli, L., Mastrangelo, A. M., Whan, A., Stephen, S., Barker, G., Wieseke, F., Plieske, J., IWGSC, Lillemo, M., Mather, D., ... Akhunov, E. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal*, 12(6), 787-796. <https://doi.org/10.1111/pbi.12183>
- Wang, Z., Huang, L., Wu, B., Hu, J., Jiang, Z., Qi, P., Zheng, Y., & Liu, D. (2018). Characterization of an integrated active *Glu-1Ay* allele in common wheat from wild emmer and its potential role in flour improvement. *International Journal of Molecular Sciences*, 19(4), 923. <https://doi.org/10.3390/ijms19040923>

- Wang, Z., Miao, L., Chen, Y., Peng, H., Ni, Z., Sun, Q., & Guo, W. (2023). Deciphering the evolution and complexity of wheat germplasm from a genomic perspective. *Journal of Genetics and Genomics*, 50(8), 846-860. <https://doi.org/10.1016/j.jgg.2023.08.002>
- Warburton, M. L., Crossa, J., Franco, J., Kazi, M., Trethowan, R., Rajaram, S., Pfeiffer, W., Zhang, P., Dreisigacker, S., & van Ginkel, M. (2006). Bringing wild relatives back into the family: Recovering genetic diversity in CIMMYT improved wheat germplasm. *Euphytica*, 149(3), 289-301. <https://doi.org/10.1007/s10681-005-9077-0>
- Watanabe, T., Okada, R., & Urayama, M. (2022a). Differences in ionic responses to nutrient deficiencies among plant species under field conditions. *Journal of Plant Nutrition*, 45(10), 1493-1503. <https://doi.org/10.1080/01904167.2021.2020837>
- Watanabe, T., Okada, R., Tokunaga, S., Maruyama, H., Urayama, M., & Shinano, T. (2022b). Nitrogen deficiency-induced molybdenum accumulation in wheat. *Journal of Plant Nutrition*, 45(9), 1413-1424. <https://doi.org/10.1080/01904167.2021.2020838>
- Wei, T., & Simko, V. (2024). R package ‘corrplot’: Visualization of a Correlation Matrix. (Version 0.95), <https://github.com/taiyun/corrplot>
- Wen, Z., Juliana, P., Dhugga, H. S., Pacheco, M., Martínez, U. I., Aguilar, A., Ibba, M. I., Govindan, V., Singh, R. P., & Dhugga, K. S. (2022). Genome-wide association study of phytic acid in wheat grain unravels markers for improving biofortification. *Frontiers in Plant Science*, 13, 830147. <https://doi.org/10.3389/fpls.2022.830147>
- Whalen, A., Gorjanc, G., & Hickey, J. (2020). AlphaImpute2: Fast and accurate pedigree and population based imputation for hundreds of thousands of individuals in livestock populations. *bioRxiv*. <https://doi.org/10.1101/2020.09.16.299677>
- Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation (R package version 1.1.4).
- Wilson, T. L., Guttieri, M. J., Nelson, N. O., Fritz, A., & Tilley, M. (2020). Nitrogen and sulfur effects on hard winter wheat quality and asparagine concentration. *Journal of Cereal Science*, 93, 102969. <https://doi.org/10.1016/j.jcs.2020.102969>
- Wondifraw, M. A., Winn, Z. J., Haley, S. D., Stromberger, J. A., Hudson-Arns, E. E., & Mason, R. E. (2024). Elucidation of the genetic architecture of water absorption capacity in hard winter wheat through genome wide association study. *The Plant Genome*, 17, e20500. <https://doi.org/10.1002/tpg2.20500>
- Wu, L., Xia, X., Rosewarne, G. M., Zhu, H., Li, S., Zhang, Z., & He, Z. (2015). Stripe rust resistance gene *Yr18* and its suppressor gene in Chinese wheat landraces. *Plant Breeding*, 134(6), 634-640. <https://doi.org/10.1111/pbr.12311>

- Wu, Q.-H., Chen, Y.-X., Zhou, S.-H., Fu, L., Chen, J.-J., Xiao, Y., Zhang, D., Ouyang, S.-H., Zhao, X.-J., Cui, Y., Zhang, D.-Y., Liang, Y., Wang, Z.-Z., Xie, J.-Z., Qin, J.-X., Wang, G.-X., Li, D.-L., Huang, Y.-L., Yu, M.-H., ... Liu, Z.-Y. (2015). High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population Yanda1817 x Beinong6. *PLoS ONE*, 10(2), e0118144. <https://doi.org/10.1371/journal.pone.0118144>
- Wysocka, K., Cacak-Pietrzak, G., & Sosulski, T. (2025). Mineral Concentration in Spring Wheat Grain Under Organic, Integrated, and Conventional Farming Systems and Their Alterations During Processing. *Plants*, 14(7), 1003. <https://doi.org/10.3390/plants14071003>
- Xie, C., Sun, Q., Ni, Z., Yang, T., Nevo, E., & Fahima, T. (2003). Chromosomal location of a *Triticum dicoccoides*-derived powdery mildew resistance gene in common wheat by using microsatellite markers. *Theoretical and Applied Genetics*, 106(2), 341-345. <https://doi.org/10.1007/s00122-002-1022-1>
- Xie, Q., Mayes, S., & Sparkes, D. L. (2016). Preanthesis biomass accumulation of plant and plant organs defines yield components in wheat. *European Journal of Agronomy*, 81, 15-26. <https://doi.org/10.1016/j.eja.2016.08.007>
- Xie, W., & Nevo, E. (2008). Wild emmer: Genetic resources, gene mapping, and potential for wheat improvement. *Euphytica*, 164(3), 603-614. <https://doi.org/10.1007/s10681-008-9703-8>
- Xin, F., Zhu, T., Wei, S., Han, Y., Zhao, Y., Zhang, D., Ma, L., & Ding, Q. (2020). QTL mapping of kernel traits and validation of a major QTL for kernel length-width ratio using SNP and bulked segregant analysis in wheat. *Scientific Reports*, 10, 25. <https://doi.org/10.1038/s41598-019-56979-7>
- Xu, H., Dong, J., Meng, X., Jiang, H., Ding, G., Wang, Z., Wang, F., Wang, Y., Wu, M., Ma, L., Miao, G., Qin, R., Zhao, C., Sun, H., Cui, F., & Wu, Y. (2025). *TaHAL3-7A* increases grain yield by enhancing photosynthetic pigment content in wheat (*Triticum aestivum* L.). *The Plant Journal*, 122, e70274. <https://doi.org/10.1111/tpj.70274>
- Xu, S. S., Khan, K., Klindworth, D. L., Faris, J. D., & Nygard, G. (2004). Chromosomal location of genes for novel glutenin subunits and gliadins in wild emmer wheat (*Triticum turgidum* L. var. *dicoccoides*). *Theoretical and Applied Genetics*, 108(7), 1221-1228. <https://doi.org/10.1007/s00122-003-1555-y>
- Yan, W., & Frégeau-Reid, J. (2018). Genotype by Yield*Trait (GYT) biplot: a novel approach for genotype selection based on multiple traits. *Scientific Reports*, 8, 8242. <https://doi.org/10.1038/s41598-018-26688-8>
- Yang, J., Yang, R., Liang, X., Marshall, J. M., & Neibling, W. (2023). Impact of drought stress on spring wheat grain yield and quality. *Agroecosystems, Geosciences & Environment*, 6(1), e20351. <https://doi.org/10.1002/agg2.20351>

- Yang, Y., Chai, Y., Zhang, X., Lu, S., Zhao, Z., Wei, D., Chen, L., & Hu, Y.-G. (2020). Multi-locus GWAS of quality traits in bread wheat: mining more candidate genes and possible regulatory network. *Frontiers in Plant Science*, 11, 1091. <https://doi.org/10.3389/fpls.2020.01091>
- Yaniv, E., Raats, D., Ronin, Y., Korol, A. B., Grama, A., Bariana, H., Dubcovsky, J., Schulman, A. H., & Fahima, T. (2015). Evaluation of marker-assisted selection for the stripe rust resistance gene *Yr15*, introgressed from wild emmer wheat. *Molecular Breeding*, 35(1), 43. <https://doi.org/10.1007/s11032-015-0238-0>
- Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., & Liu, X. (2021). rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteomics, & Bioinformatics*, 19(4), 619-628. <https://doi.org/10.1016/j.gpb.2020.10.007>
- 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. <https://doi.org/10.1038/ng1702>
- Yu, S., Assanga, S. O., Awika, J. M., Ibrahim, A. M. H., Rudd, J. C., Xue, Q., Guttieri, M. J., Zhang, G., Baker, J. A., Jessup, K. E., & Liu, S. (2021). Genetic mapping of quantitative trait loci for end-use quality and grain minerals in hard red winter wheat. *Agronomy*, 11(12), 2519. <https://doi.org/10.3390/agronomy11122519>
- Yu, Z., Yunusbaev, U., Fritz, A., Tilley, M., Akhunova, A., Trick, H., & Akhunov, E. (2024). CRISPR-based editing of the ω - and γ -gliadin gene clusters reduces wheat immunoreactivity without affecting grain protein quality. *Plant Biotechnology Journal*, 22(4), 892-903. <https://doi.org/10.1111/pbi.14231>
- Zaheer, M. S., Aijaz, N., Hameed, A., Buttar, N. A., Rehman, S., Riaz, M. W., Ahmad, A., Manzoor, M. A., & Asaduzzaman, M. (2024). Cultivating resilience in wheat: mitigating arsenic toxicity with seaweed extract and *Azospirillum brasilense*. *Frontiers in Microbiology*, 15, 1441719. <https://doi.org/10.3389/fmicb.2024.1441719>
- Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., Argillier, O., Stiewe, G., Hinze, M., Neumann, K., Ganai, M. W., & Röder, M. S. (2014). Whole genome association mapping of plant height in winter wheat (*Triticum aestivum* L.). *PLoS ONE*, 9(11), e113287. <https://doi.org/10.1371/journal.pone.0113287>
- Zeller, S., & Feller, U. (2000) Long-Distance Transport of Alkali Metals in Maturing Wheat. *Biologia Plantarum*, 43(4), 523-528. <https://doi.org/10.1023/A:1002806522138>
- Zhang, B., Zhao, J., Huang, J., Wang, X., Guo, Z., Jia, Q., Cao, S., Sun, Z., Luo, H., Kang, Z., & Jin, S. (2024). Aggressiveness of *Puccinia striiformis* f. sp. *tritici* isolates at high temperatures: a study case in core overwintering area of Gansu as inoculum source. *Plants*, 13(24), 3518. <https://doi.org/10.3390/plants13243518>

- Zhang, D., Bai, G., Zhu, C., Yu, J., & Carver, B. F. (2010). Genetic diversity, population structure, and linkage disequilibrium in U.S. elite winter wheat. *The Plant Genome*, 3(2), 117-127. <https://doi.org/10.3835/plantgenome2010.03.0004>
- Zhang, G., Martin, T. J., Fritz, A. K., Miller, R., Chen, M.-S., Bowden, R. L., & Bai, G. (2016). Registration of ‘Joe’ hard white winter wheat. *Journal of Plant Registrations*, 10(3), 283-286. <https://doi.org/10.3198/jpr2016.02.0007crc>
- Zhang, J., She, M., Yang, R., Jiang, Y., Qin, Y., Zhai, S., Balotf, S., Zhao, Y., Anwar, M., Alhabbar, Z., Juhász, A., Chen, J., Liu, H., Liu, Q., Zheng, T., Yang, F., Rong, J., Chen, K., Lu, M., ... Ma, W. (2021). Yield-related QTL clusters and the potential candidate genes in two wheat DH populations. *International Journal of Molecular Science*, 22(21), 11934. <https://doi.org/10.3390/ijms222111934>
- Zhang, Z., Li, A., Song, G., Geng, S., Gill, B. S., Faris, J. D., & Mao, L. (2020). Comprehensive analysis of *Q* gene near-isogenic lines reveals key molecular pathways for wheat domestication and improvement. *The Plant Journal*, 102, 299-310. <https://doi.org/10.1111/tpj.14624>
- Zhao, J., Li, X., Qiao, L., Zheng, X., Wu, B., Guo, M., Feng, M., Qi, Z., Yang, W., & Zheng, J. (2023). Identification of structural variations related to drought tolerance in wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, 136, 37. <https://doi.org/10.1007/s00122-023-04283-4>
- Zhao, L., Zheng, Y., Wang, Y., Wang, S., Wang, T., Wang, C., Chen, Y., Zhang, K., Zhang, N., Dong, Z., & Chen, F. (2023). A *HST1*-like gene controls tiller angle through regulating endogenous auxin in common wheat. *Plant Biotechnology Journal*, 21(1), 122-135. <https://doi.org/10.1111/pbi.13930>
- Zhao, Y., Xie, P., Guan, P., Wang, Y., Li, Y., Yu, K., Xin, M., Hu, Z., Yao, Y., Ni, Z., Sun, Q., Xie, C., & Peng, H. (2019). *Btr1-A* induces grain shattering and affects spike morphology and yield-related traits in wheat. *Plant & Cell Physiology*, 60(6), 1342-1353. <https://doi.org/10.1093/pcp/pcz050>
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), 3326-3328. <https://doi.org/10.1093/bioinformatics/bts606>
- Zhou, W., Wu, S., Ding, M., Li, J., Shi, Z., Wei, W., Guo, J., Zhang, H., Jiang, Y., & Rong, J. (2016). Mapping of *Ppd-B1*, a major candidate gene for late heading on wild emmer chromosome arm 2BS and assessment of its interactions with early heading QTLs on 3AL. *PLoS ONE*, 11(2), e0147377. <https://doi.org/10.1371/journal.pone.0147377>
- Zhou, Y., Bai, S., Li, H., Sun, G., Zhang, D., Ma, F., Zhao, X., Nie, F., Li, J., Chen, L., Lv, L., Zhu, L., Fan, R., Ge, Y., Shaheen, A., Guo, G., Zhang, Z., Ma, J., Liang, H., ... Song, C. (2021). Introgressing the *Aegilops tauschii* genome into wheat as a basis for cereal improvement. *Nature Plants*, 7(6), 774-786. <https://doi.org/10.1038/s41477-021-00934-w>

- Zhou, Y., Zhao, X., Li, Y., Xu, J., Bi, A., Kang, L., Xu, D., Chen, H., Wang, Y., Wang, Y., Liu, S., Jiao, C., Lu, H., Wang, J., Yin, C., Jiao, Y., & Lu, F. (2020). *Triticum* population sequencing provides insights into wheat adaptation. *Nature Genetics*, 52(12), 1412-1422. <https://doi.org/10.1038/s41588-020-00722-w>
- Zhu, T., Wang, L., Rimbart, H., Rodriguez, J. C., Deal, K. R., De Oliveira, R., Choulet, F., Keeble-Gagnere, G., Tibbits, J., Rogers, J., Eversole, K., Appels, R., Gu, Y. Q., Mascher, M., Dvorak, J., & Luo, M.-C. (2021). Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *The Plant Journal*, 107, 303-314. <https://doi.org/10.1111/tpj.15289>
- Ziegler, G., Kear, P. J., Wu, D., Ziyomo, C., Lipka, A. E., Gore, M., Hoekenga, O., & Baxter, I. (2017). Elemental accumulation in kernels of the maize Nested Association Mapping panel reveals signals of gene by environment interactions. *bioRxiv*, 096446. <https://doi.org/10.1101/164962>

Appendix A - Supplementary tables and figures for Chapter 2

Appendix Table A.1. Subset of a pedigree input file for TdGD population 80 used for imputation with AlphaImpute2

Hierarchical structure for two WILs is shown within each breeding generation, using parentage assignments to define family structure and inform genotype imputation through pedigree-based inference.

Name	ID	Female	Male
KS061278M-4	RP1	0	0
PI471021	WE80	0	0
TA10155	TA1	0	0
DH.80	DH.80	RP1	TA1
F1.80	F1.80	DH.80	WE80
BC1F1.80	BC1F1.80	RP1	F1.80
BC2F1.80	BC2F1.80	RP1	BC1F1.80
KSTdGD80.130	WIL80.130	BC2F1.80	BC2F1.80
KSTdGD80.55	WIL80.55	BC2F1.80	BC2F1.80

Appendix Table A.2. Variant counts for TdGD population founders across successive filtering steps

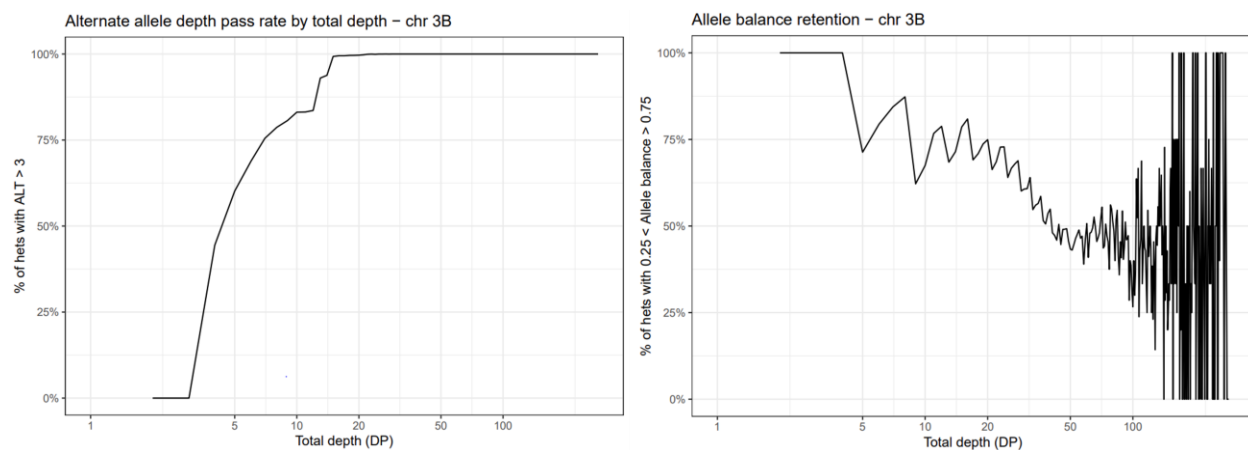
Counts are expressed in millions (M) per sample at successive filtering steps: raw count of biallelic SNPs, count within depth window ($10 \leq DP \leq 60$), and count of variants with $QUAL \geq 30$ and missingness (MISS) $\leq 50\%$.

Sample ID	Variant Count		
	Biallelic SNPs	$10 \leq DP \leq 60$	$QUAL \geq 30$, $MISS \leq 50\%$
KS061406LN~26	178.52 M	165.34 M	128.65 M
KS061278M-4	180.23 M	166.56 M	129.11 M
TA2398	31.82 M	29.64 M	27.01 M
TA10155	32.28 M	30.04 M	27.27 M
PI487264	27.63 M	24.11 M	18.32 M
PI428088	10.76 M	8.44 M	6.50 M
PI471021	25.33 M	21.93 M	17.03 M
PI466984	136.65 M	126.56 M	98.32 M
PI471815	141.77 M	131.00 M	100.16 M
PI470736	138.77 M	128.63 M	98.84 M
PI414718	129.25 M	119.98 M	93.93 M

Appendix Table A.3. Mean and median lengths of donor introgression segments detected in WILs within each TdGD population under moving average (MA) based segment calling parameterizations

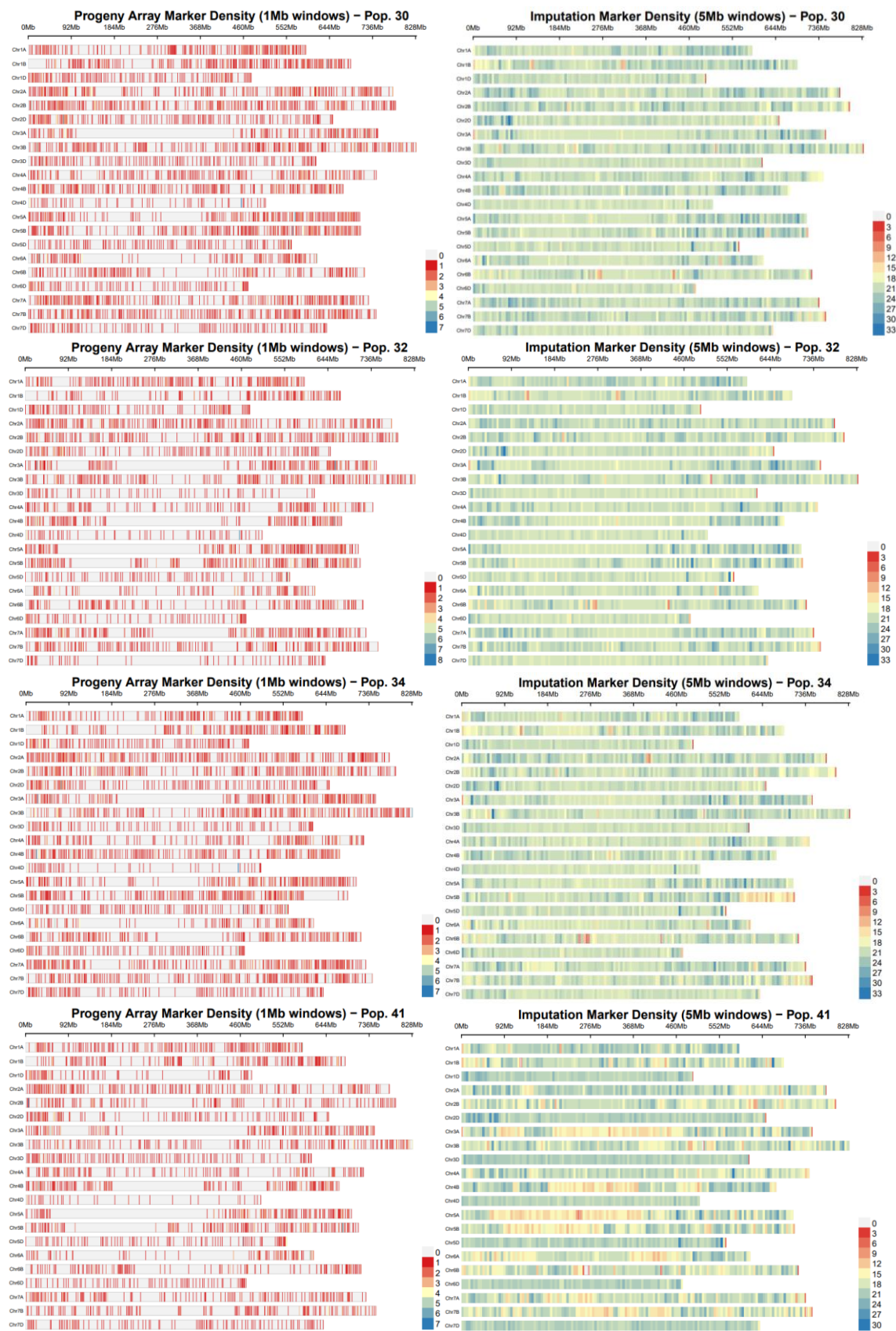
Segments were identified using paired short- and long-range MA smoothing windows applied to donor allele proportions along the genome. Short-range MA windows were evaluated at sizes of 2, 6, 10, 20, and 50 bins using a donor proportion threshold of 0.9, while long-range MA windows were evaluated at sizes of 10, 20, 40, 60, and 100 bins using a threshold of 0.7. Each column reports the mean and median segment length (Mb) across all WILs within a population for a given short-/long-range MA window combination.

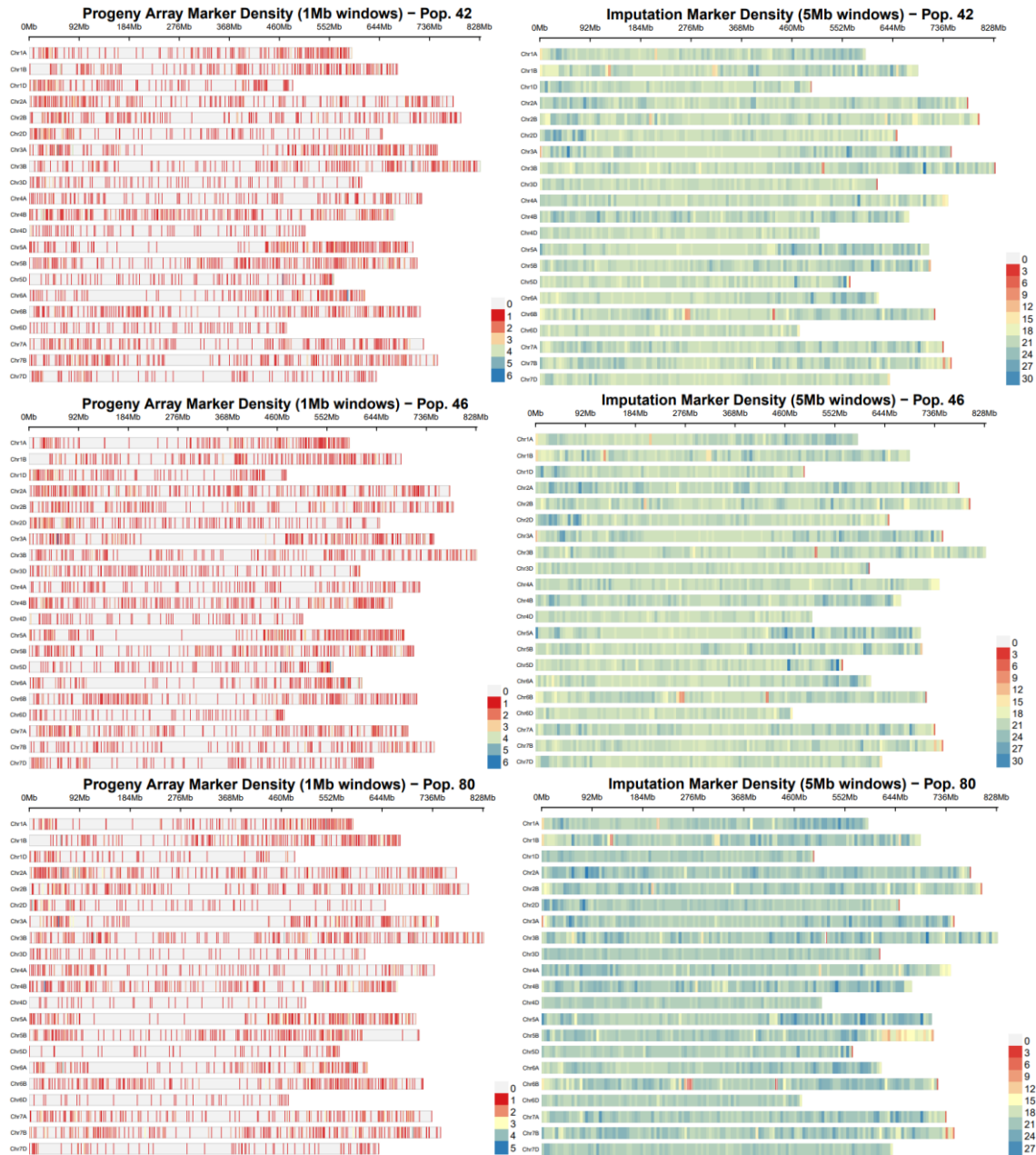
Pop	MA2/MA10		MA6/MA20		MA10/MA40		MA20/MA60		MA50/MA100	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
30	21	8	30	10	39	11	60	17	113	53
32	25	7	37	15	48	21	67	24	105	49
34	22	9	29	11	38	9	71	16	119	32
41	30	10	45	15	56	12	88	38	137	67
42	27	8	41	15	56	20	85	35	117	52
46	30	10	40	13	51	11	87	42	119	56
80	26	11	36	13	44	11	73	41	106	57



Appendix Figure A.1. Representative depth thresholding analyses on chromosome 3B, used to inform VCF DP filters

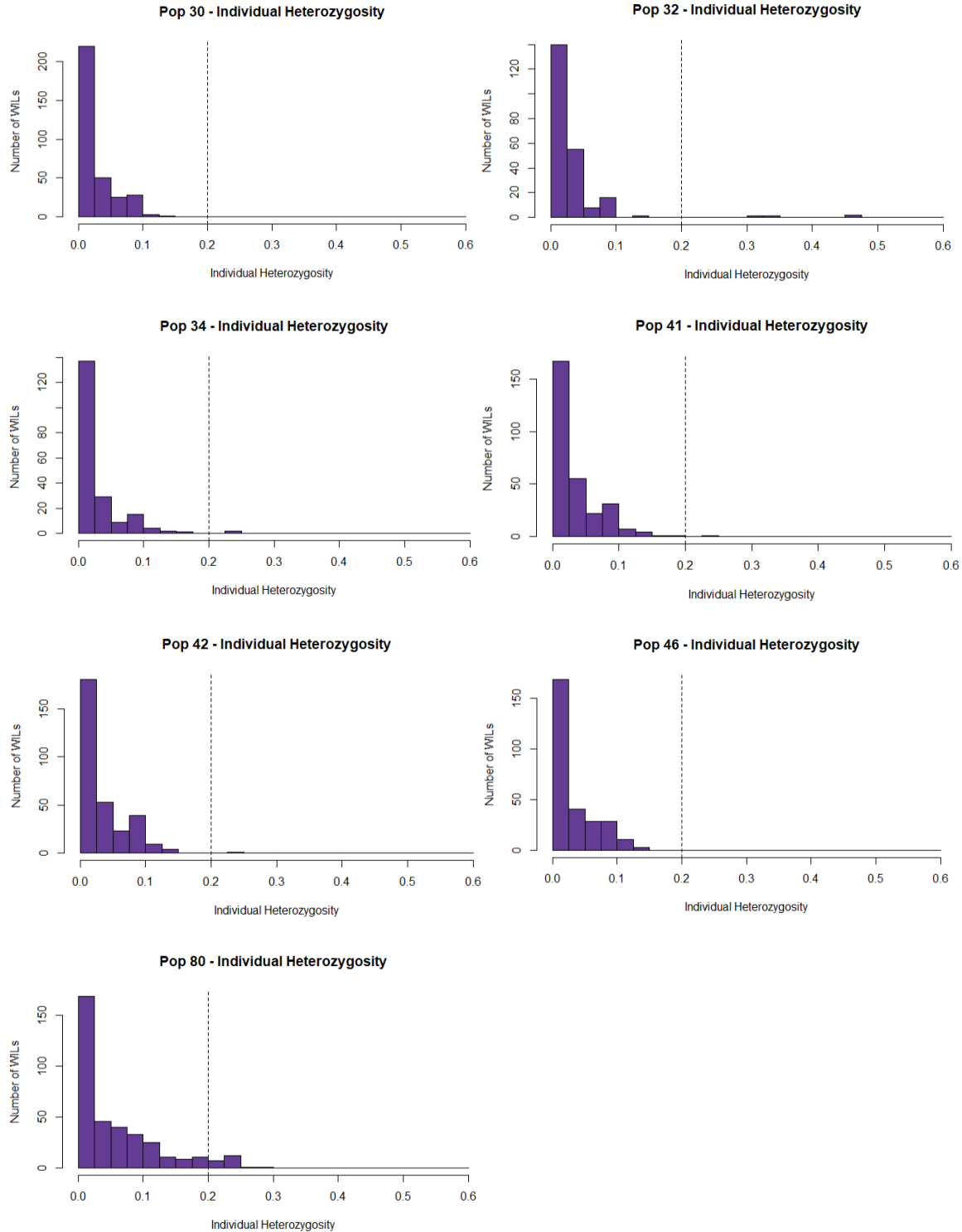
Although only 3B is shown here, identical patterns of alternate allele depth support and allele balance decay were observed across all 21 chromosomes. Left: Percentage of variant sites with ≥ 3 reads supporting the alternate allele, plotted against total site depth (DP; $\log_{10}$ scale). Right: Percentage of variant sites with balanced allele ratios (alternate allele count/total depth; between 0.25-0.75), plotted against total site depth (DP; $\log_{10}$ scale).





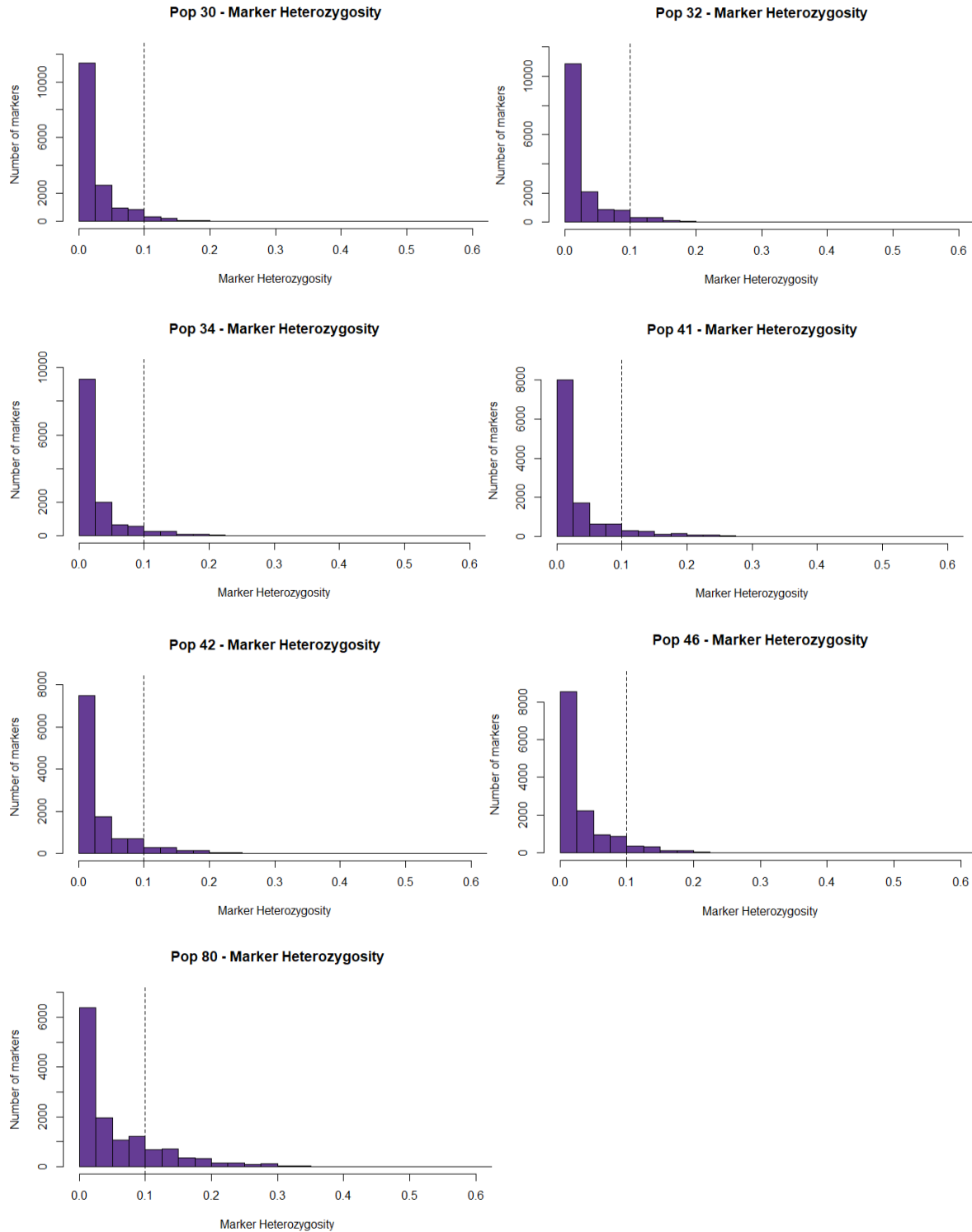
Appendix Figure A.2. Marker density of genotypic matrices used for pedigree-based imputation in TdGD populations

Progeny array marker density (left panels) shows the distribution of SNP array-derived markers genotyped in WILs, summarized in 1 Mb physical windows along each chromosome. These markers provide the primary empirical anchors for segregation state inference in the progeny. Imputation marker density (right panel) shows the distribution of markers observed in the population founders, including sequencing-derived variants and array SNPs concordant between founders and progeny, summarized in 5 Mb windows. These markers define the set of loci available for propagating founder haplotypes through the pedigree during imputation. Marker densities are plotted along physical chromosome lengths using the R package CMplot (Yin et al., 2021). Differences in marker density between panels highlight regions where haplotype inference in the progeny relies primarily on pedigree-based propagation rather than direct genotype observation. All markers are homozygous, non-missing, and polymorphic with respect to population founders, with REF/ALT concordance between array- and sequence-derived markers. Sequencing-derived variants are thinned to a minimum distance of 250 kb.



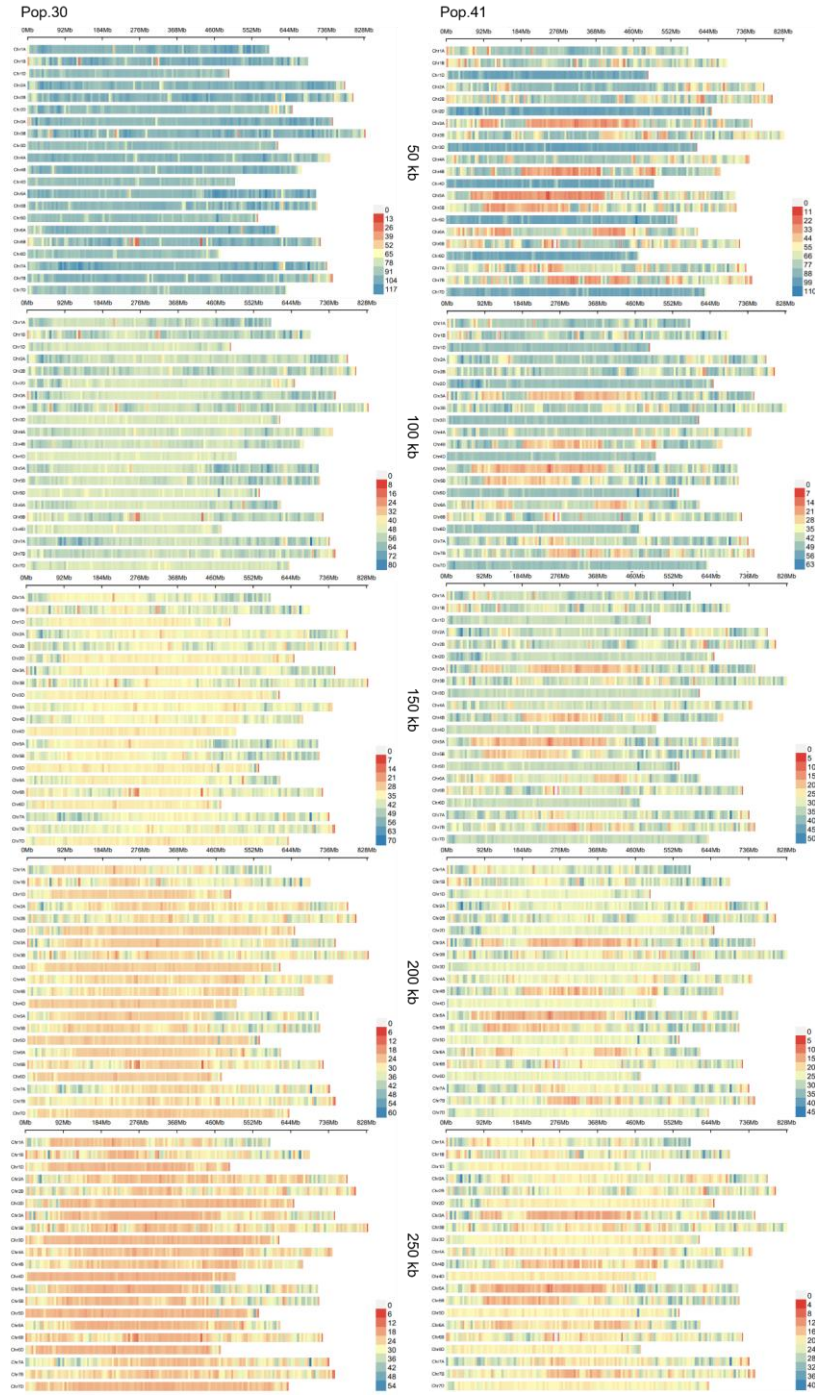
Appendix Figure A.3. Distribution of observed heterozygosity per WIL across markers within each TdGD population

Individual heterozygosity was calculated from observed heterozygous genotype calls using the R package *snpReady* (Granato et al., 2018). The dashed vertical line indicates the heterozygosity threshold above which individuals were classified as excessively heterozygous and removed from downstream analysis.



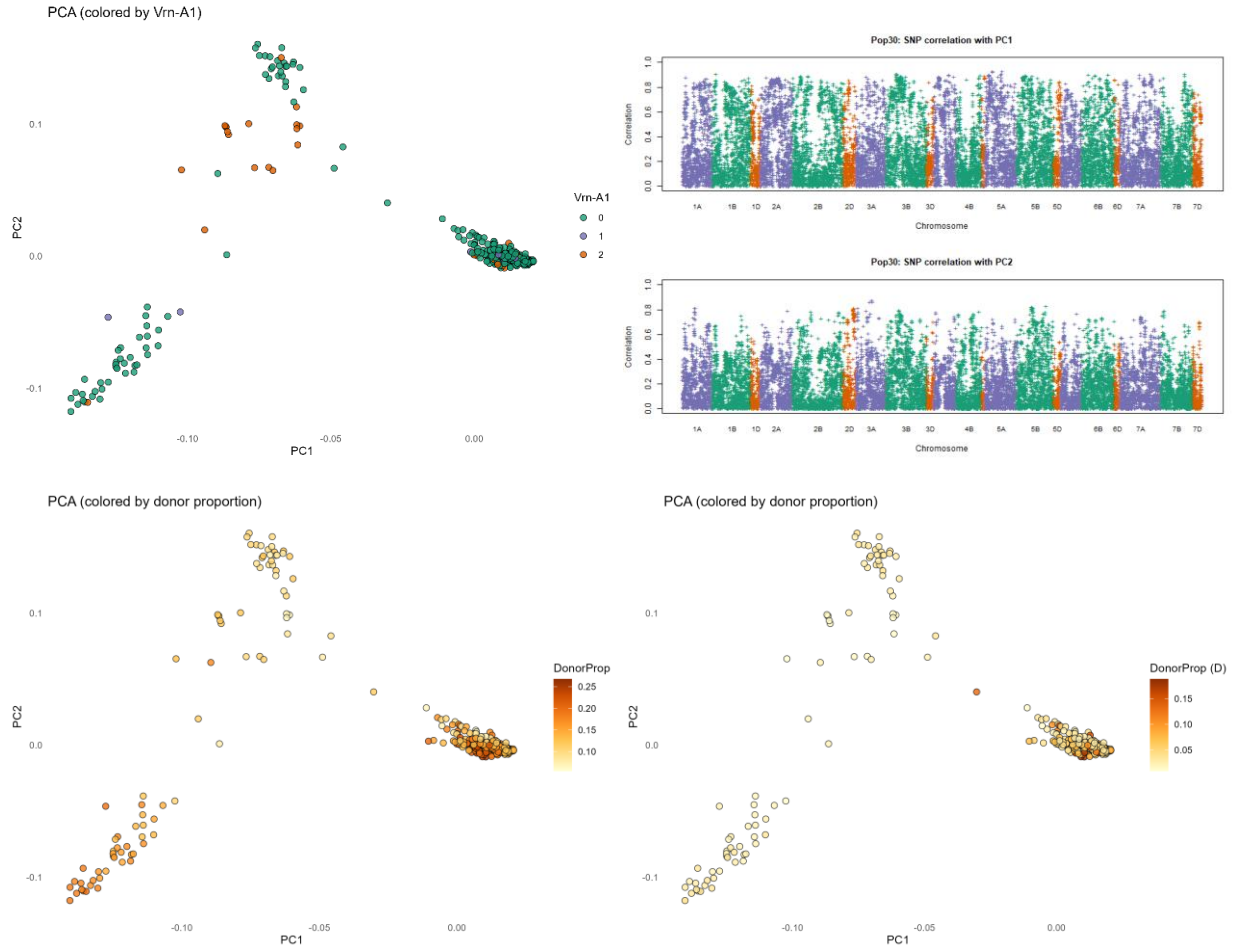
Appendix Figure A.4. Distribution of observed heterozygosity per marker across WILs within each TdGD population

Marker heterozygosity was calculated from observed heterozygous genotype calls using the R package *snpReady* (Granato et al., 2018). The dashed vertical line indicates the heterozygosity threshold above which markers were classified as excessively heterozygous and removed from downstream analysis.



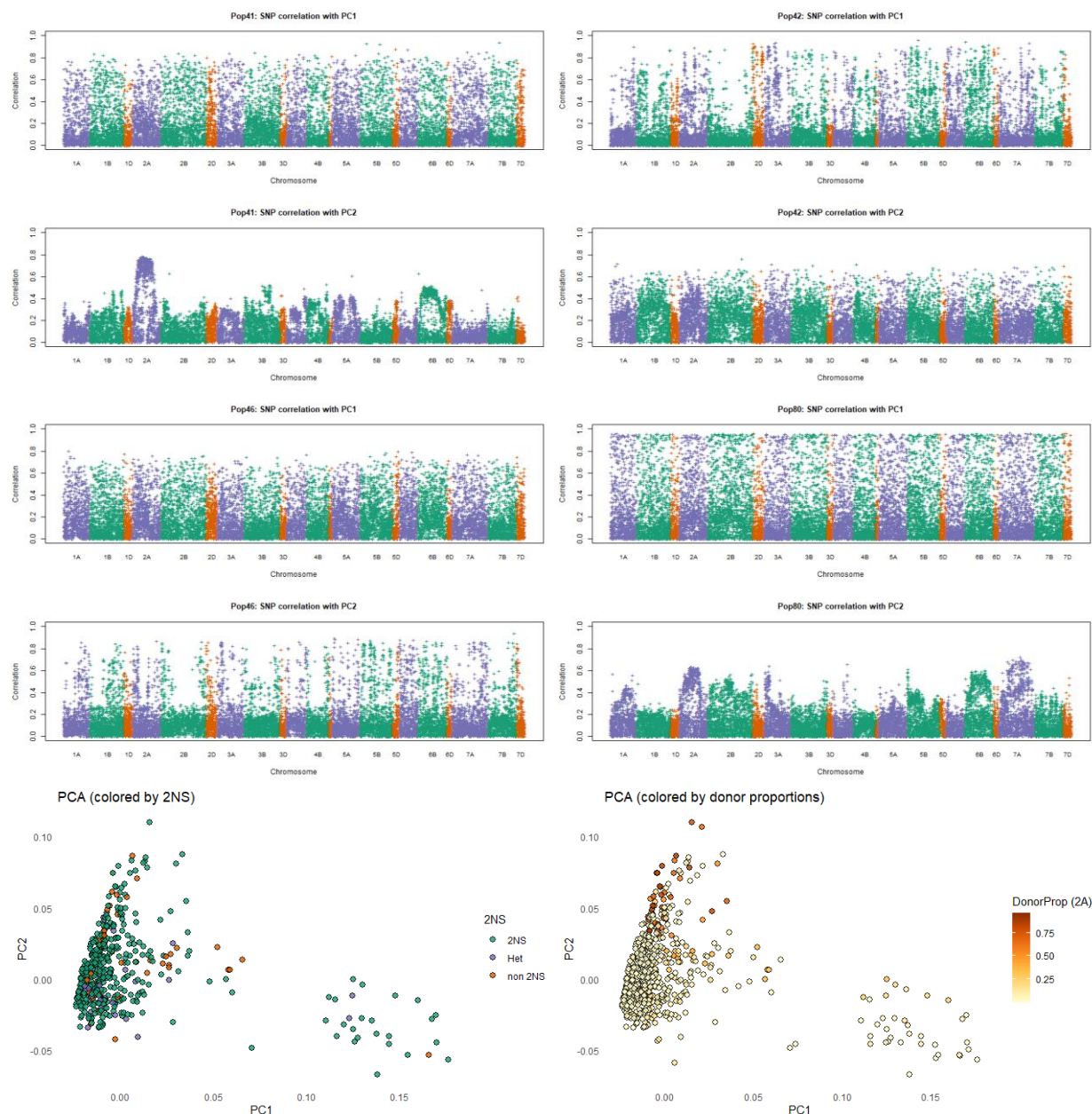
Appendix Figure A.5. Marker density of imputation genotype matrices for TdGD Pop31 and Pop41

Pop30 is shown on the left panels and represents a whole genome shotgun sequencing approach. Pop41 is shown on the right panels and represents an exome capture sequencing approach. Sequencing-derived variants were filtered and thinning using a minimum inter-marker distance of 50, 100, 150, 200, or 250 kb (top to bottom). At each thinning threshold, additional variant sites recalled based on physical overlap with progeny array SNPs are included. Marker density is summarized in 5 Mb physical windows along each chromosome and plotted using the R package CMplot (Yin et al., 2021).



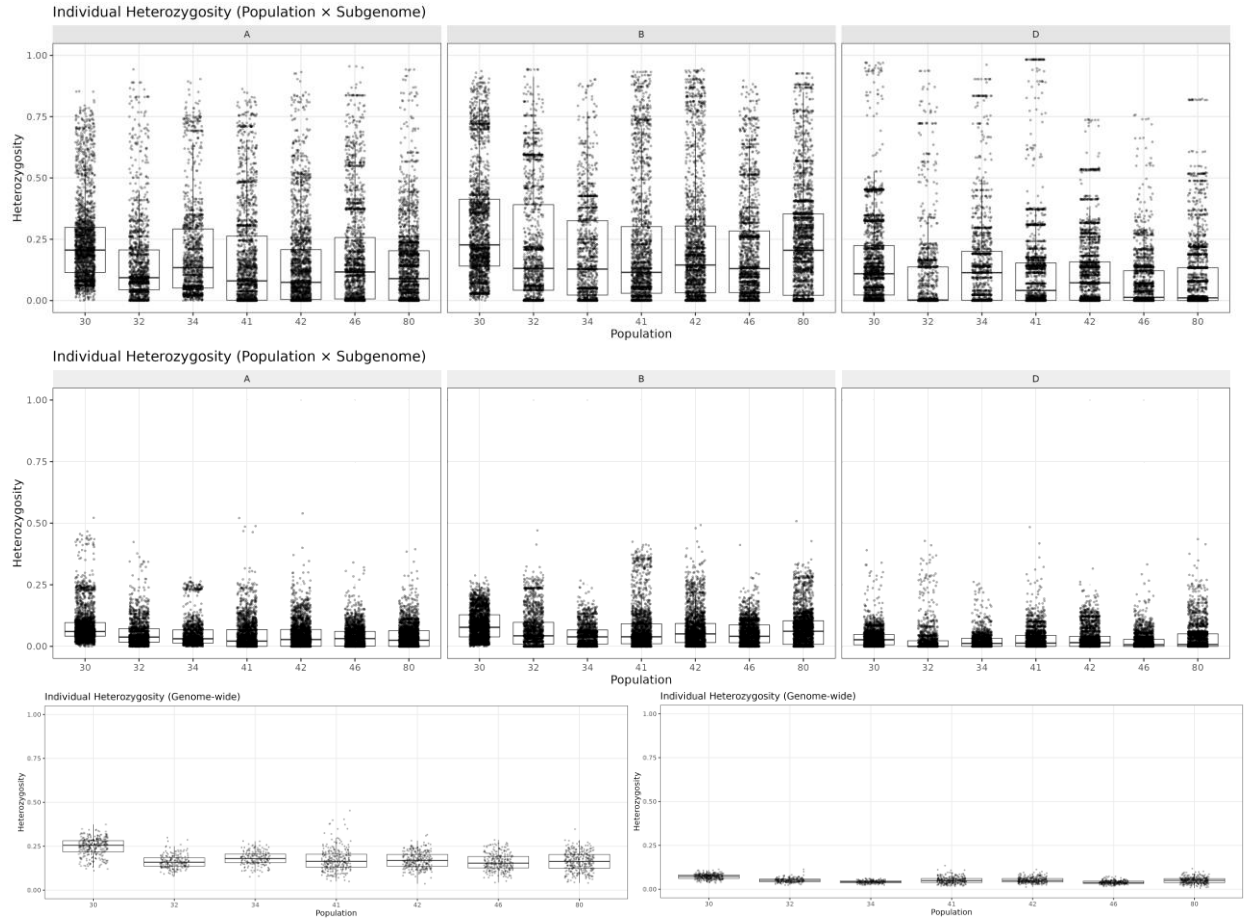
Appendix Figure A.6. PCA results for TdGD Pop30

Analyses were conducted using filtered array-based SNPs and the R package *snpRelate* (Zheng et al., 2012). The top left panel shows the first two PCs, with points colored by allelic state at the major vernalization locus *Vrn-A1*. Genotype classes are coded as 0 for the homozygous allele matching the RP (KS061406LN~26), 2 for the alternate homozygous allele, and 1 for heterozygotes. The top right panel shows absolute correlations between individual SNPs and the first two PCs, plotted by physical genomic position along each chromosome. The lower panels display PC1 and PC2 with WILs colored by average donor allele proportion calculated from 1 Mb introgression mapping bins, shown genome-wide (bottom left) and restricted to the D subgenome (bottom right).



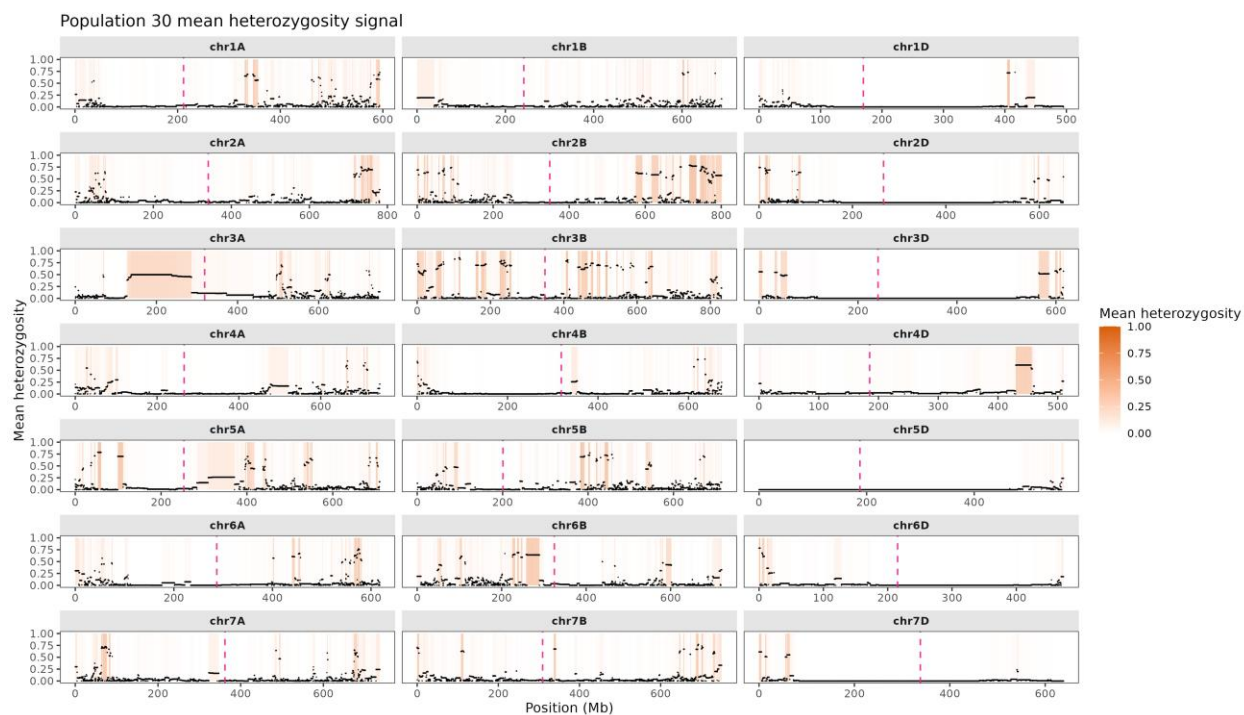
Appendix Figure A.7. PCA and SNP-PC correlation results for TdGD populations derived from KS061278M-4

PCA was conducted using filtered array-based SNPs in WIL progeny with the R package *snpRelate* (Zheng et al., 2012). For each population (Pop41, Pop42, Pop46, and Pop80), SNP correlation plots display the absolute correlations between individual SNPs and the first two PCs, plotted by genomic position across chromosomes to identify genomic regions contributing to within population structure. Elevated correlations with PC2 are observed on chromosome 2A in Pop41 and Pop80. PCA plots, subset to Pop41 and Pop80, are shown colored by results from a competitive allele-specific PCR (KASP) assay targeting the 2NS translocation (lower left) and by average donor proportion across chromosome 2A calculated from 1 Mb introgression mapping bins (lower right).



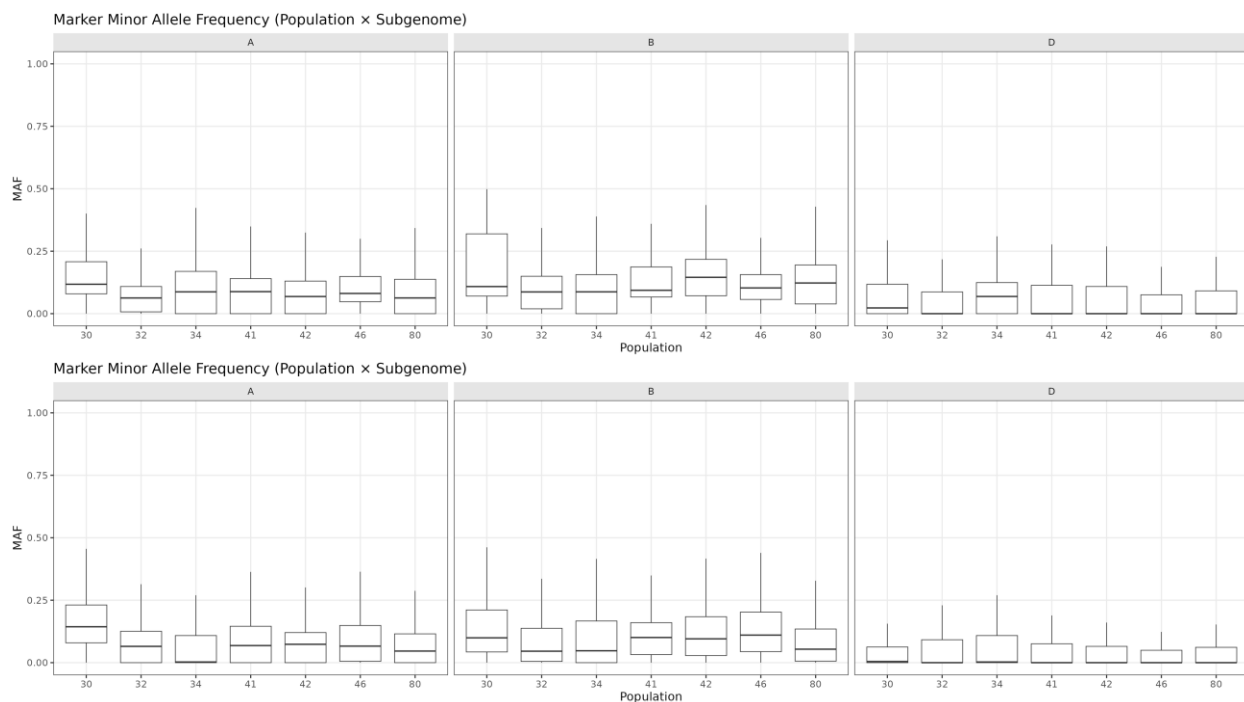
Appendix Figure A.8. Per-individual heterozygosity in TdGD WILs following imputation

Individual heterozygosity is defined as the proportion of heterozygous sites among all callable sites. Heterozygosity is summarized by population x subgenome and genome-wide. The top panel shows heterozygosity prior to heterozygote smoothing, in which heterozygous calls are retained regardless of flanking marker ancestry. The middle panel shows heterozygosity after conservative heterozygote smoothing, applied only when flanking anchor markers agree in ancestry state, illustrating the reduction in residual heterozygosity following this step. The bottom panels present genome-wide heterozygosity aggregated across all chromosomes, shown before smoothing (left) and after smoothing (right).



Appendix Figure A.9. Mean marker heterozygosity in TdGD Pop30 following imputation and heterozygote smoothing

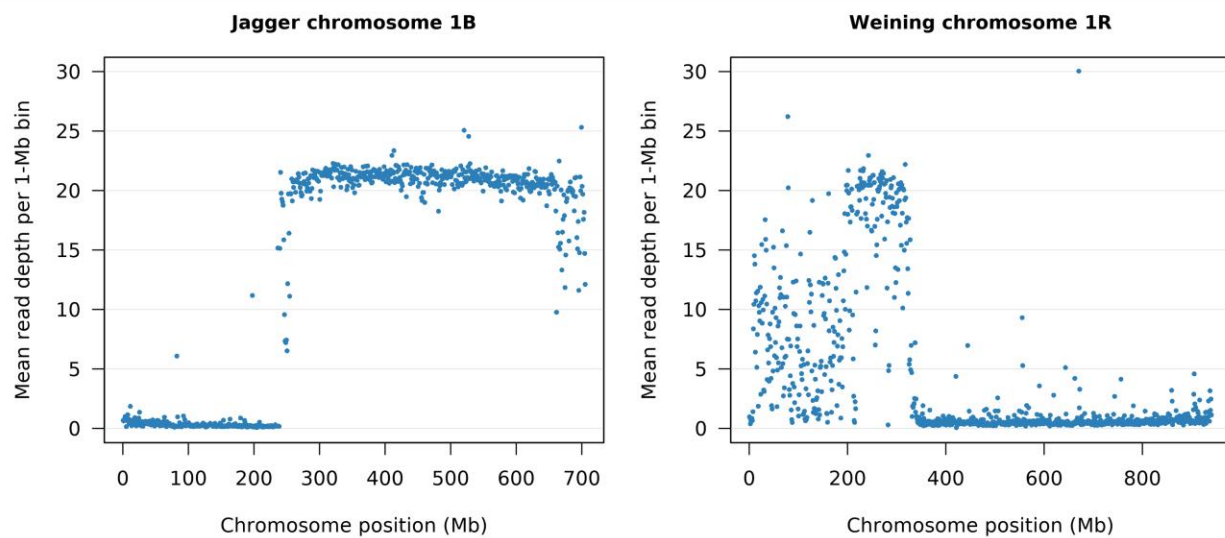
Mean heterozygosity at each marker is calculated as the proportion of individuals heterozygous at that site among all WILs in population 30. Values are plotted by physical position along each chromosome and faceted by chromosome to highlight genomic regions exhibiting elevated residual heterozygosity across the population after heterozygote smoothing.



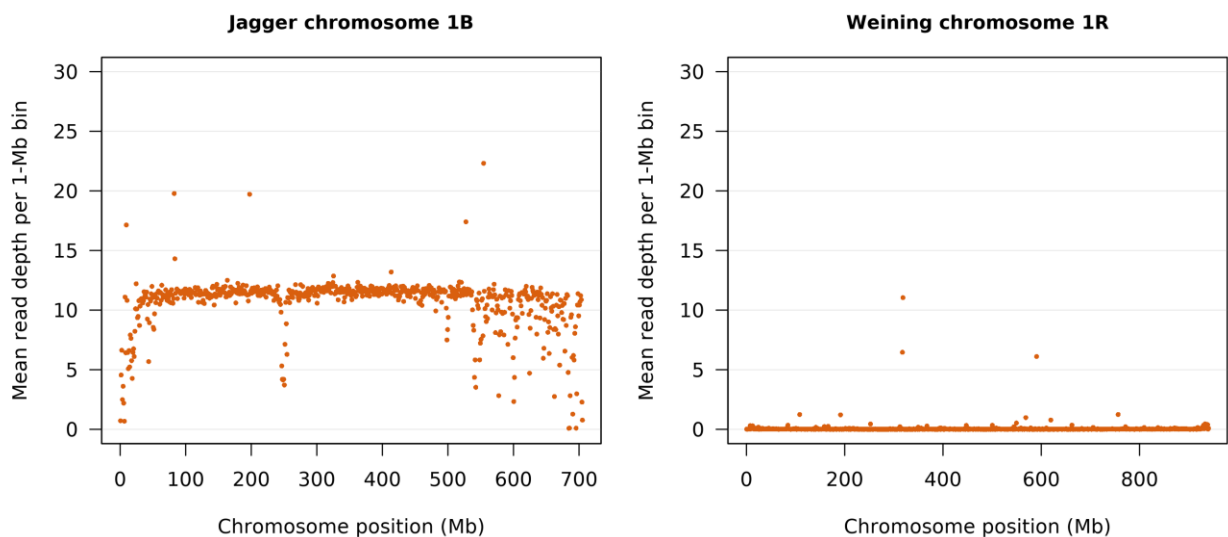
Appendix Figure A.10. Marker MAF in TdGD WILs following imputation, summarized by population and subgenome

For each marker, MAF is defined as the frequency of the less common allele within a population. The top panel shows MAF distributions prior to ancestry-based heterozygote smoothing, and the bottom panel shows MAF distributions after heterozygote smoothing within each population.

Larry – coverage profiles



KS061278M-4 – coverage profiles



Appendix Figure A.11. Comparative alignment coverage profiles to assess 1BL.1RS presence in KS061278M-4

Whole genome sequencing reads from Larry (top), a known 1BL.1RS-positive control, and KS061278M-4 (bottom) were aligned to a composite reference containing Jagger and Weining rye chromosome 1R. Points represent mean read depth within non-overlapping 1 Mb windows.

Appendix B - Supplementary tables and figures for Chapter 3

Appendix Table B.1. Trial management details for TdGD testing environments

Trial information is shown for Thomas (THD, 2021-2023), Ellis (EL, 2022), and Riley (RL, 2022-2023) Counties, Kansas. Information is provided for crop rotation, tillage practices, soil type, planting date, pre-planting seed treatment, grams of seed planted per plot (g plot⁻¹; 3.83 m² planted plot area), fertilizer inputs, herbicide and fungicide applications, harvest date, and trial mean grain moisture at harvest. Gaucho XT Flowable (imidacloprid, matalaxyl, and tebuconazole; Bayer Crop Science) was applied with Impress IR, a red seed treatment additive that improves treatment coverage and retention (KALO). Powerflex (pyroxsulam; Corteva Agriscience), Zidua (pyroxa-sulfone; BASF), and MCPA (2-methyl-4-chlorophenoxyacetic acid) were used for weed control. Nexicor (fluxapyroxad, pyraclostrobin, and propiconazole; BASF) was used for foliar disease control. DAP (diammonium phosphate) fertilizer was applied at planting (in-furrow). Seed treatment rates are expressed as milliliters of product per kilogram of seed (mL kg⁻¹). Fertilizer and pesticide application rates are expressed as dry weights, in kilograms per hectare (kg ha⁻¹). Fungicide applications were timed according to crop developmental stage and recorded as “flag leaf fully exposed” when calendar date information was unavailable. Soil information for each trial location was obtained from the USDA Natural Resources Conservation Service Web Soil Survey (Soil Survey Staff).

Trial		21THD	22EL	22RL	22THD	23RL	23THD
Rotation		Wheat-Fallow	Sorghum-Fallow-Wheat	Soybean-Soybean-Wheat	Wheat-Fallow	Soybean-Soybean-Wheat	Wheat-Fallow
Tillage		Conventional till	Conventional till	No till	Conventional till	No till	Conventional till
Soil type		Richfield silty clay loam	63% Harney silt loam 37% Crete silty clay loam	Reading silt loam	87% Keith silt loam 13% Richfield silty clay loam	Reading silt loam	Richfield silty clay loam
Planting	Date	9/22/2020	10/5/2021	10/19/2021	9/28/2021	10/17/2022	9/28/2022
	Seed treatment	2.40 mL kg ⁻¹ Gaucho XT 1.20 mL kg ⁻¹ Impress IR	2.40 mL kg ⁻¹ Gaucho XT 1.20 mL kg ⁻¹ Impress IR	2.40 mL kg ⁻¹ Gaucho XT 1.20 mL kg ⁻¹ Impress IR	2.40 mL kg ⁻¹ Gaucho XT 1.20 mL kg ⁻¹ Impress IR	2.40 mL kg ⁻¹ Gaucho XT 1.20 mL kg ⁻¹ Impress IR	2.40 mL kg ⁻¹ Gaucho XT 1.20 mL kg ⁻¹ Impress IR
	g plot ⁻¹	45	35	35	35	30	30
	Pre-planting fertilizer	-	67.25 kg ha ⁻¹ N	56.04 kg ha ⁻¹ N	-	56.04 kg ha ⁻¹ N	-
	In-furrow fertilizer	30.30 kg ha ⁻¹ DAP	30.30 kg ha ⁻¹ DAP	30.30 kg ha ⁻¹ DAP	30.30 kg ha ⁻¹ DAP	30.30 kg ha ⁻¹ DAP	30.30 kg ha ⁻¹ DAP
Fall herbicide	Timing	-	-	11/16/2021	-	-	-
	Info	-	-	0.14 kg ha ⁻¹ PowerFlex	-	-	-
Spring herbicide	Timing	3/31/2021	4/11/2022	4/12/2022	4/4/2022	4/4/2023	4/11/2023
	Info	0.21 kg ha ⁻¹ Zidua 0.84 kg ha ⁻¹ MCPA	0.21 kg ha ⁻¹ Zidua 0.84 kg ha ⁻¹ MCPA	0.21 kg ha ⁻¹ Zidua 0.84 kg ha ⁻¹ MCPA	0.21 kg ha ⁻¹ Zidua 0.84 kg ha ⁻¹ MCPA	0.21 kg ha ⁻¹ Zidua 0.84 kg ha ⁻¹ MCPA	0.21 kg ha ⁻¹ Zidua 0.84 kg ha ⁻¹ MCPA
Top dress fertilizer	Timing	3/31/2021	-	3/29/2022	4/5/2022	4/4/2023	4/11/2023
	Info	56.04 kg ha ⁻¹ N	-	119.93 kg ha ⁻¹ N	60.53 kg ha ⁻¹ N	99.76 kg ha ⁻¹ N	19.50 kg ha ⁻¹ N

Fungicide	Timing	5/12/2021	Flag leaf fully exposed	Flag leaf fully exposed	Flag leaf fully exposed	Flag leaf fully exposed	Flag leaf fully exposed
	Info	0.70 kg ha ⁻¹ Nexicor	0.70 kg ha ⁻¹ Nexicor	0.70 kg ha ⁻¹ Nexicor	0.70 kg ha ⁻¹ Nexicor	0.70 kg ha ⁻¹ Nexicor	0.70 kg ha ⁻¹ Nexicor
Harvest	Date	7/9/2021	6/24/2022	6/30/2022	7/6/2022	6/29/2023	7/17/2023
	Average moisture	9.4%	8.7%	10.1%	10.5%	9.5%	9.0%

Appendix Table B.2. UAS platform, sensor, altitude, and image resolution specifications used for drone-based data collection in 2021 and 2022

Platform, flight altitude, RGB and spectral imagers and ground sampling distance (GSD) are reported separately for each year due to differences in equipment between years. GSD represents the ground area captured by each image pixel, with smaller values indicating finer spatial resolution. RGB imagery was used to derive green reflectance (GR), whereas red and near-infrared (NIR) imagery were used to calculate normalized difference vegetation index (NDVI).

Year	2021	2022
Platform	DJI Phantom 4 Pro	DJI Matrice 100
Altitude	20 m	30 m
RGB		
Imager	Integrated Phantom 4 Pro RGB camera	DJI Zenmuse X3 RGB camera
GSD	0.5 cm pixel ⁻¹	1 cm pixel ⁻¹
Spectral		
Imager	Sentera High Precision Single sensor (red/NIR)	Micasense RedEdge-M Multispectral sensor
GSD	1.65 cm pixel ⁻¹	3.5 cm pixel ⁻¹

Appendix Table B.3. Cumulative precipitation for TdGD field trials

Cumulative precipitation (TotPrecip, mm) was calculated for each trial from January 1 (Jan1) through the trial-specific harvest date and from March 1 (Mar1) through harvest. Precipitation data were obtained from the Kansas Mesonet (2025).

Trial	TotPrecip	
	Jan1	Mar1
21THD	352.79	340.09
22EL	168.61	163.54
22RL	453.61	449.82
22THD	145.02	137.4
23RL	289.53	223.5
23THD	361.67	332.73

Appendix Table B.4. Trait means and standard deviations across TdGD testing environments

Means and standard deviations (StdDev) were calculated from raw, unstandardized observations within each environment and are provided as a reference for comparing traits among environments before within-environment standardization and subsequent two-stage mixed model analysis. Traits are grouped into agronomic performance, yield and kernel components, and functional quality categories. Trait abbreviations (Abb) and measurement units are provided. Due to differences in sensors and measurement platforms between environments, only standardized values are reported for NDVI and GR.

Trait	Abb	Units	21THD		22EL		22RL		22THD		23RL		23THD	
			Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev
Agronomic performance														
Heading date	HD	Julian days	140.77	1.34	129.43	1.18	128.90	1.22	137.09	2.05	134.85	2.38	-	-
Senescence date	SD	Julian days	175.52	2.03	161.43	1.26	164.58	1.69	170.40	1.61	171.42	2.17	178.50	1.35
Grain fill period	GFP	Days	34.74	1.66	32.00	1.28	35.67	1.33	33.31	1.90	36.56	1.90	-	-
Plant height	PH	cm	96.06	7.09	69.72	4.70	68.26	5.17	81.59	5.45	75.66	6.54	84.53	9.34
Yield and kernel components														
Tiller count	TC	count	56.93	13.75	66.56	11.99	64.77	10.27	-	-	-	-	75.10	15.08
Normalized difference vegetation index	NDVI	-	0.06	0.76	0.08	0.78	0.06	0.72	-	-	-	-	-	-
Green	GR	-	-0.01	0.61	0.01	0.69	0.03	0.57	-	-	-	-	-	-
Grain yield	GY	kg/ha	5223.24	1181.26	3633.75	635.53	3485.63	964.22	5265.47	1034.66	3847.67	608.21	5016.57	919.20
Thousand kernel weight	TKW	g	29.05	2.65	27.64	2.23	27.79	2.47	25.62	2.55	29.53	2.74	31.58	2.58
Test Weight	TW	kg/hL	72.94	1.86	75.14	1.55	72.34	1.52	73.79	1.57	74.92	1.78	72.92	1.25
Kernel area	KA	mm2	13.29	1.00	12.67	0.79	12.93	0.92	12.31	0.91	13.14	0.95	13.70	0.88
Kernel length	KL	mm	5.94	0.34	5.71	0.27	5.85	0.29	5.90	0.30	5.86	0.28	5.96	0.28
Kernel width	KW	mm	3.02	0.09	2.98	0.08	2.99	0.10	2.84	0.10	3.04	0.11	3.10	0.10
Functional quality														
Grain protein	GP	%	15.52	0.79	15.60	0.93	17.08	1.03	16.52	0.56	16.56	0.61	15.93	0.62
Grain protein deviation	GPD	-	0.00	0.61	0.00	0.85	0.00	0.67	0.00	0.44	0.00	0.46	0.00	0.54
Protein quality score	PQS	-	17.22	1.87	15.35	1.12	15.72	1.40	15.35	1.36	17.38	1.18	17.47	1.12
Solvent retention capacity	SRC	%	266.54	26.73	239.42	20.65	268.21	26.97	253.30	20.39	287.61	20.64	278.20	18.58

Appendix Table B.5. Summary of mixed model results for extended drone data, split by growth stage, for TdGD Pop80

Fixed environment effects from the across-environment model are reported for each environment, expressed relative to a standard normal distribution. Random effect variance components are partitioned into genotypic and genotype x environment (GxE) terms. Repeatability (R^2_{across}) was calculated as the proportion of genotypic variance relative to total variance.

Trait	Fixed			Random		R^2_{across}
	21THD	22EL	22RL	Genotype	GxE	
NDVI_till	0.408	-0.371	0.165	0.072	0.215	0.251
NDVI_joint	0.352	-0.365	0.336	0.084	0.178	0.320
NDVI_boot	0.165	-0.511	0.442	0.073	0.151	0.327
NDVI_anth	0.241	-0.700	0.467	0.116	0.168	0.408
NDVI_milk	0.039	-0.615	0.466	0.147	0.128	0.533
GR_till	-0.039	-0.143	0.131	0.099	0.337	0.226
GR_joint	-0.170	-0.231	-0.025	0.285	0.182	0.610
GR_boot	0.458	0.066	0.072	0.308	0.083	0.787
GR_anth	0.599	-0.327	-0.126	0.386	0.141	0.733
GR_milk	0.056	-0.267	-0.186	0.179	0.379	0.321

Appendix Table B.6. Within-environment estimates of repeatability for agronomic performance, yield and kernel component, functional quality, and extended drone data traits in TdGD Pop80

Repeatability (R^2_{within}) was estimated using SpATS models fitted separately for each trait and environment, with genotype included as a random effect during spatial (row x column) correction. R^2_{within} was calculated as the ratio of genotypic variance to total variance (genotypic + residual). These estimates reflect the reproducibility of genotype performance among replicate plots within each environment, where higher spatial heterogeneity is expected to reduce R^2_{within} .

Trait	R^2_{within}					
	21THD	22EL	22RL	22THD	23RL	23THD
HD	0.84	0.76	0.76	0.90	0.63	-
SD	0.76	0.68	0.79	0.64	0.48	0.39
GFP	0.62	0.59	0.57	0.69	0.46	-
PH	0.89	0.68	0.67	0.74	0.77	0.73
TC	0.37	0.26	0.31	-	-	0.46
NDVI	0.66	0.55	0.64	-	-	-
GR	0.89	0.73	0.84	-	-	-
GY	0.90	0.74	0.72	0.87	0.77	0.70
TKW	0.93	0.93	0.90	0.94	0.90	0.81
TW	0.95	0.96	0.92	0.93	0.86	0.91
KA	0.97	0.95	0.95	0.96	0.92	0.88
KL	0.99	0.97	0.98	0.99	0.95	0.95
KW	0.91	0.92	0.92	0.93	0.87	0.81
GP	0.94	0.87	0.90	0.93	0.90	0.86
GPD	0.85	0.70	0.64	0.88	0.81	0.78
PQS	0.88	0.70	0.84	0.92	0.75	0.77
SRC	0.84	0.75	0.83	0.91	0.76	0.74
NDVI_till	0.32	0.57	0.71	-	-	-
NDVI_joint	0.56	0.53	0.67	-	-	-
NDVI_boot	0.63	0.47	0.47	-	-	-
NDVI_anth	0.70	0.58	0.59	-	-	-
NDVI_milk	0.63	0.67	0.64	-	-	-
GR_till	0.57	0.46	0.79	-	-	-
GR_joint	0.83	0.69	0.68	-	-	-
GR_boot	0.80	0.67	0.73	-	-	-
GR_anth	0.85	0.71	0.81	-	-	-
GR_milk	0.76	0.73	0.84	-	-	-

Appendix Table B.7. GWAS results summary for agronomic performance using BLUES

For each significant association, the chromosome (Chr), physical position in Mb (Pos), trait, model, MAF, and estimated marker effect are reported. Marker effects are relative to a standard normal distribution. Allelic information includes the allelic states of the RP and donor parents (Don), as well as the alleles associated with the lower (L) and higher (H) mean phenotypic values among genotype classes. Genotype_class columns report the number of WILs carrying the RP allele (WIL_RP) or the donor allele (WIL_Don) at the marker locus. Donor direction (DonDir) indicates whether the donor allele increases or decreases the trait value. Donor allele proportion within the 1 Mb genomic bin containing the marker (DonProp) represents the average proportion of donor alleles across WILs in that genomic region. NA indicates cases where parental alleles could not be unambiguously assigned.

Chr	Pos	Trait	Model	MAF	Effect	Allele		Genotype_class				DonDir	DonProp
						RP	Don	L	H	WIL_RP	WIL_Don		
21THD													
2A	10.60	GFP	BLINK	0.12	-0.36	NA	NA	C	T	NA	NA	NA	0.11
4B	28.15	GFP	BLINK	0.05	0.44	NA	NA	C	T	NA	NA	NA	0.06
4B	28.18	GFP	MLM	0.08	0.63	NA	NA	T	C	NA	NA	NA	0.06
4B	28.41	GFP	MLM	0.08	0.63	T	C	C	T	210	17	Decrease	0.06
4B	32.17	GFP	MLM	0.08	0.65	NA	NA	T	C	NA	NA	NA	0.05
6B	163.23	GFP	MLM	0.08	0.63	NA	NA	C	T	NA	NA	NA	0.05
6B	185.67	GFP	MLM	0.07	0.74	T	C	C	T	210	10	Decrease	0.08
6B	217.17	GFP	MLM	0.07	0.69	NA	NA	C	T	NA	NA	NA	0.06
6B	218.65	GFP	MLM	0.06	0.78	A	G	G	A	210	9	Decrease	0.05
6B	309.77	GFP	MLM	0.08	0.72	A	G	G	A	206	10	Decrease	0.05
6B	333.51	GFP	MLM	0.07	0.76	A	G	G	A	208	10	Decrease	0.00
6B	431.80	GFP	MLM	0.09	0.60	A	G	G	A	198	8	Decrease	0.02
6B	431.80	GFP	MLM	0.06	0.99	T	C	C	T	207	3	Decrease	0.02
6B	431.81	GFP	MLM	0.06	0.70	T	G	G	T	211	10	Decrease	0.02
6B	450.59	GFP	MLM	0.06	0.77	A	G	G	A	214	10	Decrease	0.01
6B	490.21	GFP	MLM	0.12	0.52	NA	NA	A	T	NA	NA	NA	0.00
6B	491.62	GFP	MLM	0.11	0.63	NA	NA	G	A	NA	NA	NA	0.00
6B	618.67	GFP	MLM	0.10	0.59	NA	NA	G	A	NA	NA	NA	0.31
6B	431.80	GFP	MLMM	0.06	0.99	T	C	C	T	207	3	Decrease	0.02
3A	481.52	HD	BLINK	0.15	0.28	NA	NA	T	C	NA	NA	NA	0.08
3B	68.40	HD	BLINK	0.25	-0.29	NA	NA	C	T	NA	NA	NA	0.04
7A	584.99	HD	BLINK	0.35	0.18	NA	NA	G	A	NA	NA	NA	0.10
3A	41.37	PH	BLINK	0.05	0.69	NA	NA	A	G	NA	NA	NA	0.20
4B	28.41	PH	BLINK	0.08	-0.81	T	C	T	C	210	17	Increase	0.06
6A	578.89	PH	BLINK	0.06	-0.48	A	G	A	G	209	8	Increase	0.01
7D	76.48	PH	BLINK	0.08	-0.49	NA	NA	A	G	NA	NA	NA	0.40
4B	25.17	PH	MLM	0.05	-0.67	NA	NA	A	C	NA	NA	NA	0.08
4B	26.49	PH	MLM	0.11	-0.66	NA	NA	C	T	NA	NA	NA	0.06
4B	28.15	PH	MLM	0.05	-0.68	NA	NA	T	C	NA	NA	NA	0.06
4B	28.15	PH	MLM	0.08	-0.71	A	C	A	C	205	12	Increase	0.06

4B	28.15	PH	MLM	0.07	-0.85	A	G	A	G	211	14	Increase	0.06
4B	28.18	PH	MLM	0.08	-0.91	NA	NA	C	T	NA	NA	NA	0.06
4B	28.23	PH	MLM	0.09	-0.80	G	T	G	T	207	16	Increase	0.06
4B	28.41	PH	MLM	0.08	-0.89	T	C	T	C	210	17	Increase	0.06
4B	28.41	PH	MLM	0.07	-0.76	T	C	T	C	209	12	Increase	0.06
4B	30.36	PH	MLM	0.08	-0.80	A	G	A	G	210	14	Increase	0.05
4B	30.86	PH	MLM	0.09	-0.70	NA	NA	T	C	NA	NA	NA	0.05
4B	31.18	PH	MLM	0.07	-0.86	C	T	C	T	211	14	Increase	0.05
4B	32.17	PH	MLM	0.08	-0.74	NA	NA	C	T	NA	NA	NA	0.05
4B	33.07	PH	MLM	0.07	-0.71	A	C	A	C	209	12	Increase	0.04
4B	38.79	PH	MLM	0.06	-0.69	C	T	C	T	216	15	Increase	0.11
4B	480.92	PH	MLM	0.07	-0.88	T	C	T	C	207	10	Increase	0.02
4B	28.41	PH	MLMM	0.08	-0.90	T	C	T	C	210	17	Increase	0.06
2A	18.02	SD	BLINK	0.07	-0.31	T	C	T	C	212	12	Increase	0.30
2B	65.23	SD	BLINK	0.05	-0.61	NA	NA	A	G	NA	NA	NA	0.34
4B	28.41	SD	BLINK	0.08	0.33	T	C	C	T	210	17	Decrease	0.06
6B	431.80	SD	BLINK	0.06	0.74	T	C	C	T	207	3	Decrease	0.02
7A	436.70	SD	BLINK	0.08	0.34	A	C	C	A	212	17	Decrease	0.08
7D	6.38	SD	BLINK	0.13	0.24	NA	NA	G	A	NA	NA	NA	0.72
2B	65.23	SD	MLMM	0.05	-0.69	NA	NA	A	G	NA	NA	NA	0.34
6B	431.80	SD	MLMM	0.06	0.91	T	C	C	T	207	3	Decrease	0.02
22EL													
3A	40.15	GFP	BLINK	0.05	-0.43	A	G	A	G	218	11	Increase	0.16
5A	689.61	GFP	BLINK	0.06	-0.62	NA	NA	G	T	NA	NA	NA	0.00
6B	135.12	GFP	BLINK	0.09	0.59	NA	NA	A	C	NA	NA	NA	0.31
7B	3.50	GFP	BLINK	0.06	-0.44	C	T	C	T	215	10	Increase	0.23
2B	66.20	HD	BLINK	0.07	-0.68	T	C	T	C	205	8	Increase	0.18
3A	481.52	HD	BLINK	0.15	0.39	NA	NA	T	C	NA	NA	NA	0.08
1B	426.44	PH	BLINK	0.07	-0.33	A	G	A	G	214	14	Increase	0.07
1B	623.67	PH	BLINK	0.38	0.25	C	G	G	C	82	136	Decrease	0.34
1B	632.42	PH	BLINK	0.29	-0.25	NA	NA	A	G	NA	NA	NA	0.62
2D	28.55	PH	BLINK	0.08	0.42	C	A	A	C	206	10	Decrease	0.01
3B	829.20	PH	BLINK	0.16	0.33	T	G	G	T	188	29	Decrease	0.10
4B	28.15	PH	BLINK	0.08	-0.42	A	C	A	C	205	12	Increase	0.06
6A	552.94	PH	BLINK	0.07	-0.64	NA	NA	T	C	NA	NA	NA	0.17
6B	717.92	PH	BLINK	0.06	0.53	T	G	G	T	208	7	Decrease	0.01
2B	63.85	SD	BLINK	0.05	-0.63	A	C	A	C	214	7	Increase	0.08
3A	31.66	SD	BLINK	0.10	-0.42	T	C	T	C	195	11	Increase	0.15
3A	438.61	SD	BLINK	0.11	0.54	A	G	G	A	199	18	Decrease	0.07
5B	644.87	SD	BLINK	0.11	0.32	NA	NA	G	A	NA	NA	NA	0.00
6B	120.65	SD	BLINK	0.09	0.49	T	C	C	T	201	12	Decrease	0.04
2B	63.85	SD	MLM	0.05	-0.66	A	C	A	C	214	7	Increase	0.08

2B	63.85	SD	MLMM	0.05	-0.68	A	C	A	C	214	7	Increase	0.08
22RL													
7A	553.99	GFP	BLINK	0.12	-0.42	NA	NA	C	T	NA	NA	NA	0.07
3A	151.32	HD	BLINK	0.11	0.55	C	T	T	C	199	17	Decrease	0.06
1B	440.27	PH	BLINK	0.10	-0.56	NA	NA	G	A	NA	NA	NA	0.08
2B	18.26	PH	BLINK	0.06	-0.48	C	T	C	T	217	14	Increase	0.16
3B	581.49	PH	BLINK	0.40	0.56	T	C	C	T	129	81	Decrease	0.39
3B	581.70	PH	BLINK	0.39	0.38	NA	NA	T	C	NA	NA	NA	0.39
3B	691.71	PH	BLINK	0.42	0.28	A	G	G	A	129	90	Decrease	0.28
2B	6.23	SD	BLINK	0.14	-0.35	NA	NA	G	T	NA	NA	NA	0.59
2B	63.85	SD	BLINK	0.05	-0.48	A	C	A	C	214	7	Increase	0.08
3A	47.83	SD	BLINK	0.12	0.44	NA	NA	G	A	NA	NA	NA	0.60
3A	47.83	SD	MLM	0.12	0.57	NA	NA	G	A	NA	NA	NA	0.60
2B	6.23	SD	MLMM	0.14	-0.41	NA	NA	G	T	NA	NA	NA	0.59
3A	47.83	SD	MLMM	0.12	0.58	NA	NA	G	A	NA	NA	NA	0.60
22THD													
4A	662.20	GFP	BLINK	0.12	-0.42	T	C	T	C	193	16	Increase	0.02
2B	23.79	HD	BLINK	0.06	-0.38	C	G	C	G	217	14	Increase	0.08
2D	67.58	HD	BLINK	0.16	0.29	T	C	C	T	187	28	Decrease	0.03
3A	438.62	HD	BLINK	0.11	0.47	NA	NA	C	T	NA	NA	NA	0.07
3B	0.99	HD	BLINK	0.15	-0.25	NA	NA	C	T	NA	NA	NA	0.19
3B	63.47	HD	BLINK	0.34	-0.25	NA	NA	C	T	NA	NA	NA	0.22
4A	664.45	HD	BLINK	0.11	0.43	NA	NA	C	T	NA	NA	NA	0.00
5A	672.74	HD	BLINK	0.07	0.37	T	C	C	T	209	11	Decrease	0.04
6A	502.68	HD	BLINK	0.07	-0.49	C	T	C	T	209	9	Increase	0.01
7A	316.01	HD	BLINK	0.48	-0.22	NA	NA	A	C	NA	NA	NA	0.08
4A	662.20	HD	MLM	0.12	0.59	T	C	C	T	193	16	Decrease	0.02
4A	662.20	HD	MLMM	0.12	0.58	T	C	C	T	193	16	Decrease	0.02
3B	566.84	PH	BLINK	0.42	0.34	G	A	A	G	134	97	Decrease	0.36
4B	28.18	PH	BLINK	0.08	-0.40	NA	NA	C	T	NA	NA	NA	0.06
5B	448.12	PH	BLINK	0.09	0.33	NA	NA	A	G	NA	NA	NA	0.28
2B	65.23	SD	BLINK	0.05	-0.59	NA	NA	A	G	NA	NA	NA	0.34
4B	34.72	SD	BLINK	0.06	0.41	C	T	T	C	208	6	Decrease	0.05
23RL													
2B	63.85	HD	BLINK	0.05	-0.80	A	C	A	C	214	7	Increase	0.08
3A	481.52	HD	BLINK	0.15	0.31	NA	NA	T	C	NA	NA	NA	0.08
5B	498.29	HD	BLINK	0.13	-0.27	T	G	T	G	197	28	Increase	0.12
7B	332.80	HD	BLINK	0.06	-0.40	A	G	A	G	216	15	Increase	0.02
2B	45.55	HD	MLM	0.05	-0.88	NA	NA	A	C	NA	NA	NA	0.31
2B	59.10	HD	MLM	0.06	-0.77	NA	NA	T	C	NA	NA	NA	0.07
2B	63.85	HD	MLM	0.05	-0.89	A	C	A	C	214	7	Increase	0.08
2B	65.23	HD	MLM	0.05	-0.91	NA	NA	A	G	NA	NA	NA	0.34

2B	66.20	HD	MLM	0.07	-0.62	T	C	T	C	205	8	Increase	0.18
2B	63.85	HD	MLMM	0.05	-0.90	A	C	A	C	214	7	Increase	0.08
3A	47.83	HD	MLMM	0.12	0.38	NA	NA	G	A	NA	NA	NA	0.60
7B	332.80	HD	MLMM	0.06	-0.49	A	G	A	G	216	15	Increase	0.02
1B	563.45	PH	BLINK	0.10	-0.47	NA	NA	T	C	NA	NA	NA	0.33
3B	68.40	PH	BLINK	0.25	-0.31	NA	NA	C	T	NA	NA	NA	0.04
6A	467.04	PH	BLINK	0.07	-0.68	NA	NA	T	C	NA	NA	NA	0.00
2B	65.23	SD	BLINK	0.05	-0.56	NA	NA	A	G	NA	NA	NA	0.34
5B	498.29	SD	BLINK	0.13	-0.34	T	G	T	G	196	28	Increase	0.12
5D	502.03	SD	BLINK	0.09	0.41	C	T	T	C	199	12	Decrease	0.02
6D	460.74	SD	BLINK	0.07	-0.69	A	G	A	G	201	2	Increase	0.10
23THD													
2B	184.39	PH	BLINK	0.15	-0.21	A	C	A	C	196	34	Increase	0.19
5A	700.35	PH	BLINK	0.07	0.31	NA	NA	G	T	NA	NA	NA	0.00
5B	439.27	PH	BLINK	0.09	0.29	NA	NA	T	C	NA	NA	NA	0.53
6A	418.54	PH	BLINK	0.07	-0.36	C	T	C	T	210	10	Increase	0.01
1B	655.76	SD	BLINK	0.42	-0.19	NA	NA	T	C	NA	NA	NA	0.20
2B	65.23	SD	BLINK	0.05	-0.78	NA	NA	A	G	NA	NA	NA	0.34
6B	115.95	SD	BLINK	0.06	0.41	G	A	A	G	213	9	Decrease	0.15
7B	6.15	SD	BLINK	0.05	-0.55	NA	NA	T	C	NA	NA	NA	0.16
2B	45.55	SD	MLM	0.05	-0.69	NA	NA	A	C	NA	NA	NA	0.31
2B	59.10	SD	MLM	0.06	-0.59	NA	NA	T	C	NA	NA	NA	0.07
2B	63.85	SD	MLM	0.05	-0.67	A	C	A	C	214	7	Increase	0.08
2B	65.23	SD	MLM	0.05	-0.80	NA	NA	A	G	NA	NA	NA	0.34
2B	65.23	SD	MLMM	0.05	-0.80	NA	NA	A	G	NA	NA	NA	0.34

Appendix Table B.8. GWAS results summary for yield and kernel components using BLUEs

For each significant association, the chromosome (Chr), physical position in Mb (Pos), trait, model, MAF, and estimated marker effect are reported. Marker effects are relative to a standard normal distribution. Allelic information includes the allelic states of the RP and donor parents (Don), as well as the alleles associated with the lower (L) and higher (H) mean phenotypic values among genotype classes. Genotype_class columns report the number of WILs carrying the RP allele (WIL_RP) or the donor allele (WIL_Don) at the marker locus. Donor direction (DonDir) indicates whether the donor allele increases or decreases the trait value. Donor allele proportion within the 1 Mb genomic bin containing the marker (DonProp) represents the average proportion of donor alleles across WILs in that genomic region. NA indicates cases where parental alleles could not be unambiguously assigned.

Chr	Pos	Trait	Model	MAF	Effect	Allele		Genotype_class				DonDir	DonProp
						RP	Don	L	H	WIL_RP	WIL_Don		
21THD													
2B	23.79	GR	BLINK	0.08	0.24	A	C	C	A	206	11	Decrease	0.08
3B	457.98	GR	BLINK	0.48	0.20	NA	NA	T	C	NA	NA	NA	0.02
5A	412.53	GR	BLINK	0.43	-0.15	T	C	T	C	131	97	Increase	0.31
5A	412.53	GR	MLM	0.43	-0.15	T	C	T	C	131	97	Increase	0.31
5A	414.01	GR	MLM	0.40	-0.14	A	G	A	G	139	92	Increase	0.37
5A	412.53	GR	MLMM	0.43	-0.15	T	C	T	C	131	97	Increase	0.31
3A	552.32	GY	BLINK	0.14	0.46	T	C	C	T	188	20	Decrease	0.08
4B	34.72	GY	BLINK	0.06	0.52	C	T	T	C	208	6	Decrease	0.05
6B	717.92	GY	BLINK	0.06	0.47	T	G	G	T	208	7	Decrease	0.01
1A	28.62	KA	BLINK	0.24	-0.20	C	T	C	T	175	53	Increase	0.18
2A	83.01	KA	BLINK	0.08	-0.24	G	A	G	A	205	11	Increase	0.07
2A	103.38	KA	BLINK	0.09	0.60	T	C	T	C	205	17	Increase	0.07
2A	117.39	KA	BLINK	0.09	0.58	C	A	C	A	208	18	Increase	0.07
2A	133.34	KA	BLINK	0.06	0.74	C	T	C	T	210	9	Increase	0.05
2A	162.42	KA	BLINK	0.06	0.84	A	C	A	C	210	9	Increase	0.01
2A	686.86	KA	BLINK	0.11	-0.32	NA	NA	G	A	NA	NA	NA	0.02
3A	438.62	KA	BLINK	0.11	-0.23	NA	NA	T	C	NA	NA	NA	0.07
3D	30.24	KA	BLINK	0.16	0.20	A	G	G	A	188	33	Decrease	0.24
4B	628.30	KA	BLINK	0.06	-0.51	A	G	A	G	207	5	Increase	0.02
5A	510.03	KA	BLINK	0.07	-0.36	A	G	A	G	209	11	Increase	0.09
5A	609.69	KA	BLINK	0.48	-0.22	NA	NA	T	C	NA	NA	NA	0.19
6B	686.80	KA	BLINK	0.30	0.17	NA	NA	C	T	NA	NA	NA	0.00
6B	688.31	KA	BLINK	0.05	-0.34	NA	NA	A	G	NA	NA	NA	0.00
7A	108.96	KA	BLINK	0.21	-0.20	C	T	C	T	180	46	Increase	0.32
1A	28.62	KL	BLINK	0.24	-0.26	C	T	C	T	175	53	Increase	0.18
2A	83.01	KL	BLINK	0.08	-0.40	G	A	G	A	205	11	Increase	0.07
3A	438.62	KL	BLINK	0.11	-0.35	NA	NA	T	C	NA	NA	NA	0.07
4A	644.49	KL	BLINK	0.09	-0.30	T	G	T	G	204	16	Increase	0.02
4B	28.23	KL	BLINK	0.09	-0.52	G	T	G	T	207	16	Increase	0.06
5A	437.21	KL	BLINK	0.05	-0.69	NA	NA	T	C	NA	NA	NA	0.42

5A	446.83	KL	BLINK	0.49	0.20	NA	NA	T	C	NA	NA	NA	0.14
5A	595.55	KL	BLINK	0.05	-0.49	NA	NA	G	A	NA	NA	NA	0.16
5A	608.88	KL	BLINK	0.44	-0.17	C	T	T	C	102	129	Decrease	0.24
7A	646.99	KL	BLINK	0.07	-0.33	NA	NA	A	G	NA	NA	NA	0.46
7D	127.14	KL	BLINK	0.11	-0.36	NA	NA	A	G	NA	NA	NA	0.00
4B	617.28	KW	BLINK	0.05	-0.72	G	T	G	T	217	11	Increase	0.22
6A	597.35	KW	BLINK	0.10	0.40	NA	NA	C	A	NA	NA	NA	0.01
1A	272.61	NDVI	BLINK	0.11	0.13	NA	NA	G	A	NA	NA	NA	0.14
1B	662.16	NDVI	BLINK	0.05	0.18	NA	NA	T	C	NA	NA	NA	0.08
1D	41.57	NDVI	BLINK	0.14	-0.09	T	C	T	C	198	32	Increase	0.14
2A	581.58	NDVI	BLINK	0.16	0.14	G	A	A	G	185	30	Decrease	0.06
2B	129.66	NDVI	BLINK	0.06	-0.17	NA	NA	A	G	NA	NA	NA	0.24
4A	19.66	NDVI	BLINK	0.10	-0.15	A	G	A	G	207	23	Increase	0.08
5A	469.27	NDVI	BLINK	0.05	0.23	A	G	G	A	219	12	Decrease	0.10
7A	112.98	NDVI	BLINK	0.08	0.38	NA	NA	C	G	NA	NA	NA	0.29
3A	44.60	NDVI	MLM	0.07	0.37	A	G	A	G	4	202	Increase	0.23
3A	44.61	NDVI	MLM	0.07	-0.34	T	C	T	C	203	5	Increase	0.23
3A	45.78	NDVI	MLM	0.06	0.44	NA	NA	C	T	NA	NA	NA	0.03
3A	47.08	NDVI	MLM	0.06	0.32	G	A	G	A	15	216	Increase	0.60
3A	50.89	NDVI	MLM	0.05	0.33	A	G	A	G	7	214	Increase	0.59
3A	52.35	NDVI	MLM	0.08	0.32	NA	NA	A	C	NA	NA	NA	0.37
3A	56.14	NDVI	MLM	0.05	0.34	NA	NA	T	C	NA	NA	NA	0.17
7A	112.98	NDVI	MLM	0.08	0.47	NA	NA	C	G	NA	NA	NA	0.29
7A	117.78	NDVI	MLM	0.05	0.40	NA	NA	C	T	NA	NA	NA	0.15
7A	125.49	NDVI	MLM	0.09	0.31	NA	NA	C	T	NA	NA	NA	0.10
7A	112.98	NDVI	MLMM	0.08	0.48	NA	NA	C	G	NA	NA	NA	0.29
4B	14.11	TC	BLINK	0.12	0.36	NA	NA	C	T	NA	NA	NA	0.11
5A	502.79	TC	BLINK	0.09	0.40	T	C	C	T	202	13	Decrease	0.11
4B	14.11	TC	MLMM	0.12	0.39	NA	NA	C	T	NA	NA	NA	0.11
3D	433.50	TKW	BLINK	0.15	0.25	G	T	T	G	190	27	Decrease	0.02
4B	622.78	TKW	BLINK	0.06	-0.61	NA	NA	C	T	NA	NA	NA	0.10
5A	510.03	TKW	BLINK	0.07	-0.36	A	G	A	G	209	11	Increase	0.09
5A	609.69	TKW	BLINK	0.48	-0.21	NA	NA	T	C	NA	NA	NA	0.19
6B	686.80	TKW	BLINK	0.30	0.20	NA	NA	C	T	NA	NA	NA	0.00
7A	303.97	TKW	BLINK	0.13	-0.31	NA	NA	A	G	NA	NA	NA	0.09
2A	713.93	TW	BLINK	0.13	0.32	G	A	A	G	202	29	Decrease	0.20
3B	112.27	TW	BLINK	0.41	0.21	NA	NA	C	T	NA	NA	NA	0.34
4A	341.07	TW	BLINK	0.10	0.82	T	C	C	T	199	16	Decrease	0.04
4A	365.62	TW	BLINK	0.08	-0.87	T	C	C	T	208	14	Decrease	0.01
6B	431.80	TW	BLINK	0.06	0.91	T	C	C	T	207	3	Decrease	0.02
7A	654.77	TW	BLINK	0.10	0.32	A	G	G	A	209	22	Decrease	0.06
2A	691.23	TW	MLM	0.08	0.50	NA	NA	C	T	NA	NA	NA	0.27
2A	700.54	TW	MLM	0.10	0.55	NA	NA	C	T	NA	NA	NA	0.06

2A	700.54	TW	MLM	0.09	0.55	NA	NA	G	A	NA	NA	NA	0.06
2A	713.93	TW	MLM	0.13	0.42	G	A	A	G	202	29	Decrease	0.20
4A	25.80	TW	MLM	0.10	0.43	NA	NA	C	T	NA	NA	NA	0.16
6B	185.67	TW	MLM	0.07	0.78	T	C	C	T	210	10	Decrease	0.08
6B	431.80	TW	MLM	0.06	0.95	T	C	C	T	207	3	Decrease	0.02
6B	531.53	TW	MLM	0.06	0.77	A	G	G	A	213	11	Decrease	0.01
7A	654.77	TW	MLM	0.10	0.55	A	G	G	A	209	22	Decrease	0.06
2A	700.54	TW	MLMM	0.10	0.52	NA	NA	C	T	NA	NA	NA	0.06
4A	25.80	TW	MLMM	0.10	0.41	NA	NA	C	T	NA	NA	NA	0.16
6B	431.80	TW	MLMM	0.06	0.88	T	C	C	T	207	3	Decrease	0.02
22EL													
2A	205.92	GR	BLINK	0.07	-0.31	NA	NA	A	G	NA	NA	NA	0.00
5A	412.53	GR	BLINK	0.43	-0.20	T	C	T	C	131	97	Increase	0.31
5B	680.99	GY	BLINK	0.10	0.49	NA	NA	C	T	NA	NA	NA	0.00
6B	472.95	GY	BLINK	0.09	0.38	G	A	A	G	204	15	Decrease	0.02
2A	713.93	KA	BLINK	0.13	-0.34	G	A	G	A	202	29	Increase	0.20
4B	628.30	KA	BLINK	0.06	-0.42	A	G	A	G	207	5	Increase	0.02
5A	442.35	KA	BLINK	0.42	-0.19	NA	NA	A	C	NA	NA	NA	0.29
6B	686.80	KA	BLINK	0.30	0.18	NA	NA	C	T	NA	NA	NA	0.00
3A	438.62	KL	BLINK	0.11	-0.35	NA	NA	T	C	NA	NA	NA	0.07
4A	644.49	KL	BLINK	0.09	-0.36	T	G	T	G	204	16	Increase	0.02
4B	28.15	KL	BLINK	0.08	-0.34	A	C	A	C	205	12	Increase	0.06
5A	446.83	KL	BLINK	0.49	0.21	NA	NA	T	C	NA	NA	NA	0.14
5A	547.80	KL	BLINK	0.08	-0.31	A	G	A	G	212	19	Increase	0.07
6B	686.80	KL	BLINK	0.30	0.17	NA	NA	C	T	NA	NA	NA	0.00
6B	688.31	KL	BLINK	0.05	-0.47	NA	NA	A	G	NA	NA	NA	0.00
7D	124.50	KL	BLINK	0.13	-0.32	T	G	T	G	192	22	Increase	0.01
2A	713.93	KW	BLINK	0.13	-0.24	G	A	G	A	202	29	Increase	0.20
3D	30.24	KW	BLINK	0.16	0.34	A	G	G	A	188	33	Decrease	0.24
4A	673.35	KW	BLINK	0.08	-0.37	A	G	A	G	211	19	Increase	0.03
5A	1.20	KW	BLINK	0.11	0.40	A	G	G	A	204	23	Decrease	0.02
5B	650.01	KW	BLINK	0.07	0.40	NA	NA	T	C	NA	NA	NA	0.00
5D	132.41	KW	BLINK	0.13	-0.43	T	C	T	C	198	29	Increase	0.16
6B	686.80	KW	BLINK	0.30	0.21	NA	NA	C	T	NA	NA	NA	0.00
7A	527.15	KW	BLINK	0.10	-0.41	T	C	T	C	207	24	Increase	0.09
2A	476.22	NDVI	BLINK	0.06	0.34	NA	NA	T	G	NA	NA	NA	0.00
5B	8.34	NDVI	BLINK	0.06	0.34	NA	NA	T	C	NA	NA	NA	0.04
2A	715.39	TKW	BLINK	0.23	-0.28	NA	NA	C	A	NA	NA	NA	0.22
2B	614.91	TKW	BLINK	0.07	0.42	NA	NA	A	G	NA	NA	NA	0.06
4B	39.30	TKW	BLINK	0.05	-0.56	A	C	A	C	217	11	Increase	0.06
5A	609.69	TKW	BLINK	0.48	-0.17	NA	NA	T	C	NA	NA	NA	0.19
6B	686.80	TKW	BLINK	0.30	0.19	NA	NA	C	T	NA	NA	NA	0.00

3D	422.47	TW	BLINK	0.10	-0.25	G	A	G	A	207	24	Increase	0.02
4A	341.07	TW	BLINK	0.10	0.39	T	C	C	T	199	16	Decrease	0.04
5B	34.76	TW	BLINK	0.09	0.29	NA	NA	A	G	NA	NA	NA	0.04
5D	406.59	TW	BLINK	0.12	0.39	A	G	G	A	196	19	Decrease	0.07
6B	531.53	TW	BLINK	0.06	1.10	A	G	G	A	213	11	Decrease	0.01
6B	567.61	TW	BLINK	0.05	-0.90	A	G	G	A	218	11	Decrease	0.01
7A	273.02	TW	BLINK	0.23	-0.20	NA	NA	C	T	NA	NA	NA	0.19
22RL													
2B	6.23	GR	BLINK	0.14	0.16	NA	NA	T	G	NA	NA	NA	0.59
2B	18.26	GR	BLINK	0.06	0.31	C	T	T	C	217	14	Decrease	0.16
3A	713.63	GR	BLINK	0.11	0.23	A	G	G	A	199	21	Decrease	0.07
5A	412.53	GR	BLINK	0.43	-0.11	T	C	T	C	131	97	Increase	0.31
2B	5.53	GR	MLM	0.12	0.33	NA	NA	C	A	NA	NA	NA	0.69
2B	5.67	GR	MLM	0.13	0.31	NA	NA	A	G	NA	NA	NA	0.69
2B	6.23	GR	MLM	0.14	0.29	NA	NA	T	G	NA	NA	NA	0.59
2B	9.73	GR	MLM	0.15	0.26	NA	NA	G	A	NA	NA	NA	0.32
2B	18.26	GR	MLM	0.06	0.41	C	T	T	C	217	14	Decrease	0.16
2B	23.79	GR	MLM	0.08	0.37	A	C	C	A	206	11	Decrease	0.08
2B	23.79	GR	MLM	0.06	0.35	C	G	G	C	217	14	Decrease	0.08
2B	23.81	GR	MLM	0.06	0.36	T	C	C	T	216	11	Decrease	0.08
5A	412.53	GR	MLM	0.43	-0.15	T	C	T	C	131	97	Increase	0.31
2B	5.67	GR	MLMM	0.13	0.31	NA	NA	A	G	NA	NA	NA	0.69
1D	38.87	GY	BLINK	0.19	0.27	T	C	C	T	185	43	Decrease	0.15
2B	451.89	GY	BLINK	0.11	0.46	NA	NA	G	A	NA	NA	NA	0.06
3A	53.03	GY	BLINK	0.15	0.39	NA	NA	G	A	NA	NA	NA	0.54
2A	713.93	KA	BLINK	0.13	-0.50	G	A	G	A	202	29	Increase	0.20
3A	12.84	KA	BLINK	0.06	0.85	NA	NA	T	C	NA	NA	NA	0.20
3A	31.66	KA	BLINK	0.10	-0.43	T	C	T	C	195	11	Increase	0.15
4A	644.49	KA	BLINK	0.09	-0.34	T	G	T	G	204	16	Increase	0.02
4B	601.75	KA	BLINK	0.10	-0.60	A	G	A	G	199	14	Increase	0.04
5B	508.80	KA	BLINK	0.11	0.26	NA	NA	T	C	NA	NA	NA	0.11
6B	686.80	KA	BLINK	0.30	0.22	NA	NA	C	T	NA	NA	NA	0.00
2A	713.93	KL	BLINK	0.13	-0.43	G	A	G	A	202	29	Increase	0.20
3A	438.62	KL	BLINK	0.11	-0.42	NA	NA	T	C	NA	NA	NA	0.07
5A	446.83	KL	BLINK	0.49	0.20	NA	NA	T	C	NA	NA	NA	0.14
5A	547.80	KL	BLINK	0.08	-0.27	A	G	A	G	212	19	Increase	0.07
6B	686.80	KL	BLINK	0.30	0.22	NA	NA	C	T	NA	NA	NA	0.00
6B	688.31	KL	BLINK	0.05	-0.40	NA	NA	A	G	NA	NA	NA	0.00
7D	124.50	KL	BLINK	0.13	-0.29	T	G	T	G	192	22	Increase	0.01
4B	626.63	KW	BLINK	0.06	-0.60	NA	NA	G	T	NA	NA	NA	0.02
5A	3.65	KW	BLINK	0.11	0.36	A	G	A	G	14	196	Increase	0.15
5A	608.88	KW	BLINK	0.44	-0.19	C	T	T	C	102	129	Decrease	0.24
5D	0.35	KW	BLINK	0.19	-0.37	T	G	T	G	178	34	Increase	0.03

6B	686.80	KW	BLINK	0.30	0.28	NA	NA	C	T	NA	NA	NA	0.00
6B	690.97	KW	MLMM	0.35	0.31	NA	NA	T	C	NA	NA	NA	0.00
4B	28.23	NDVI	BLINK	0.09	0.20	G	T	T	G	207	16	Decrease	0.06
5A	470.19	NDVI	BLINK	0.09	0.18	NA	NA	T	C	NA	NA	NA	0.28
5B	249.07	NDVI	BLINK	0.05	0.30	A	G	G	A	209	3	Decrease	0.04
2A	713.93	TKW	BLINK	0.13	-0.39	G	A	G	A	202	29	Increase	0.20
3A	56.14	TKW	BLINK	0.05	-0.40	NA	NA	C	T	NA	NA	NA	0.17
4B	633.62	TKW	BLINK	0.07	-0.62	G	A	G	A	207	10	Increase	0.03
5A	609.69	TKW	BLINK	0.48	-0.21	NA	NA	T	C	NA	NA	NA	0.19
6B	686.80	TKW	BLINK	0.30	0.26	NA	NA	C	T	NA	NA	NA	0.00
6B	684.27	TKW	MLM	0.37	0.26	NA	NA	C	T	NA	NA	NA	0.00
6B	686.80	TKW	MLM	0.30	0.29	NA	NA	C	T	NA	NA	NA	0.00
6B	690.97	TKW	MLM	0.35	0.26	NA	NA	T	C	NA	NA	NA	0.00
6B	686.80	TKW	MLMM	0.30	0.30	NA	NA	C	T	NA	NA	NA	0.00
2A	713.93	TW	BLINK	0.13	0.38	G	A	A	G	202	29	Decrease	0.20
3D	422.47	TW	BLINK	0.10	-0.51	G	A	G	A	207	24	Increase	0.02
5B	640.56	TW	BLINK	0.18	0.26	NA	NA	T	C	NA	NA	NA	0.00
5D	263.15	TW	BLINK	0.16	0.28	A	G	G	A	176	19	Decrease	0.12
6B	147.58	TW	BLINK	0.05	-0.74	NA	NA	C	T	NA	NA	NA	0.14
6B	165.07	TW	BLINK	0.05	-0.95	T	G	G	T	217	10	Decrease	0.05
6B	417.85	TW	BLINK	0.06	-1.04	NA	NA	T	C	NA	NA	NA	0.00
6B	531.53	TW	BLINK	0.06	0.74	A	G	G	A	213	11	Decrease	0.01
6B	196.68	TW	MLM	0.17	0.50	T	C	C	T	191	39	Decrease	0.08
6B	333.51	TW	MLM	0.07	0.85	A	G	G	A	208	10	Decrease	0.00
6B	531.53	TW	MLM	0.06	0.87	A	G	G	A	213	11	Decrease	0.01
6B	196.68	TW	MLMM	0.17	0.49	T	C	C	T	191	39	Decrease	0.08
6B	218.65	TW	MLMM	0.06	0.91	A	G	G	A	210	9	Decrease	0.05
6B	333.51	TW	MLMM	0.07	0.93	A	G	G	A	208	10	Decrease	0.00
6B	531.53	TW	MLMM	0.06	0.91	A	G	G	A	213	11	Decrease	0.01
6B	554.79	TW	MLMM	0.11	0.59	T	C	C	T	202	20	Decrease	0.02
22THD													
2B	63.85	GY	BLINK	0.05	0.46	A	C	C	A	214	7	Decrease	0.08
3A	80.09	GY	BLINK	0.09	0.37	T	C	C	T	204	15	Decrease	0.11
2A	713.93	KA	BLINK	0.13	-0.37	G	A	G	A	202	29	Increase	0.20
4B	28.23	KA	BLINK	0.09	-0.72	G	T	G	T	207	16	Increase	0.06
5A	609.69	KA	BLINK	0.48	-0.16	NA	NA	T	C	NA	NA	NA	0.19
6B	686.80	KA	BLINK	0.30	0.17	NA	NA	C	T	NA	NA	NA	0.00
3A	438.62	KL	BLINK	0.11	-0.32	NA	NA	T	C	NA	NA	NA	0.07
3B	3.91	KL	BLINK	0.14	-0.22	NA	NA	C	T	NA	NA	NA	0.34
4B	28.15	KL	BLINK	0.08	-0.49	A	C	A	C	205	12	Increase	0.06
5A	442.71	KL	BLINK	0.44	0.15	G	A	A	G	129	101	Decrease	0.29
5A	547.80	KL	BLINK	0.08	-0.41	A	G	A	G	212	19	Increase	0.07

6B	688.31	KL	BLINK	0.05	-0.53	NA	NA	A	G	NA	NA	NA	0.00
7D	124.50	KL	BLINK	0.13	-0.33	T	G	T	G	192	22	Increase	0.01
1A	10.08	KW	BLINK	0.44	-0.16	NA	NA	T	C	NA	NA	NA	0.09
2A	700.54	KW	BLINK	0.10	-0.39	NA	NA	T	C	NA	NA	NA	0.06
2B	249.21	KW	BLINK	0.09	-0.29	NA	NA	T	G	NA	NA	NA	0.18
3D	30.24	KW	BLINK	0.16	0.24	A	G	G	A	188	33	Decrease	0.24
4B	28.23	KW	BLINK	0.09	-0.38	G	T	G	T	207	16	Increase	0.06
4B	601.75	KW	BLINK	0.10	-0.39	A	G	A	G	199	14	Increase	0.04
5A	609.69	KW	BLINK	0.48	-0.16	NA	NA	T	C	NA	NA	NA	0.19
5B	550.39	KW	BLINK	0.08	0.32	NA	NA	T	C	NA	NA	NA	0.17
6B	690.97	KW	BLINK	0.35	0.19	NA	NA	T	C	NA	NA	NA	0.00
6D	3.03	KW	BLINK	0.22	-0.21	NA	NA	G	A	NA	NA	NA	0.00
2D	223.90	TKW	BLINK	0.07	-0.37	C	T	C	T	202	5	Increase	0.01
4B	28.23	TKW	BLINK	0.09	-0.67	G	T	G	T	207	16	Increase	0.06
5A	558.97	TKW	BLINK	0.26	0.15	A	G	G	A	170	58	Decrease	0.21
5A	595.55	TKW	BLINK	0.05	-0.36	NA	NA	G	A	NA	NA	NA	0.16
5A	609.69	TKW	BLINK	0.48	-0.20	NA	NA	T	C	NA	NA	NA	0.19
6B	690.97	TKW	BLINK	0.35	0.21	NA	NA	T	C	NA	NA	NA	0.00
1A	517.47	TW	BLINK	0.10	-0.25	G	A	G	A	201	16	Increase	0.16
2A	713.93	TW	BLINK	0.13	0.23	G	A	A	G	202	29	Decrease	0.20
2B	110.48	TW	BLINK	0.15	0.33	NA	NA	G	T	NA	NA	NA	0.06
2D	32.79	TW	BLINK	0.10	-0.31	NA	NA	T	C	NA	NA	NA	0.01
3B	419.56	TW	BLINK	0.35	0.27	T	C	C	T	150	81	Decrease	0.01
4A	695.26	TW	BLINK	0.09	0.43	NA	NA	G	A	NA	NA	NA	0.00
5B	677.15	TW	BLINK	0.10	0.25	NA	NA	C	T	NA	NA	NA	0.00
6B	531.53	TW	BLINK	0.06	0.50	A	G	G	A	213	11	Decrease	0.01
6B	704.04	TW	BLINK	0.06	0.31	NA	NA	A	G	NA	NA	NA	0.00
5B	10.43	TW	MLMM	0.08	-0.77	G	A	A	G	204	9	Decrease	0.04
5B	11.31	TW	MLMM	0.07	0.87	A	G	G	A	215	16	Decrease	0.05
6B	704.04	TW	MLMM	0.06	0.56	NA	NA	A	G	NA	NA	NA	0.00
23RL													
3A	25.49	GY	BLINK	0.44	-0.17	G	A	G	A	124	95	Increase	0.42
3A	31.66	GY	BLINK	0.10	0.44	T	C	C	T	195	11	Decrease	0.15
5B	454.89	GY	BLINK	0.10	0.38	T	C	C	T	195	12	Decrease	0.06
2B	451.89	KA	BLINK	0.12	-0.51	NA	NA	A	G	NA	NA	NA	0.06
3B	691.27	KA	BLINK	0.47	0.24	NA	NA	G	A	NA	NA	NA	0.28
3D	433.50	KA	BLINK	0.15	0.17	G	T	T	G	189	27	Decrease	0.02
4B	28.41	KA	BLINK	0.08	-0.39	T	C	T	C	209	17	Increase	0.06
4B	601.75	KA	BLINK	0.10	-0.35	A	G	A	G	198	14	Increase	0.04
6B	686.80	KA	BLINK	0.30	0.21	NA	NA	C	T	NA	NA	NA	0.00
2A	715.39	KL	BLINK	0.23	-0.30	NA	NA	C	A	NA	NA	NA	0.22
4B	601.75	KL	BLINK	0.10	-0.57	A	G	A	G	198	14	Increase	0.04

5A	442.35	KL	BLINK	0.42	-0.19	NA	NA	A	C	NA	NA	NA	0.29
5A	547.80	KL	BLINK	0.08	-0.33	A	G	A	G	211	19	Increase	0.07
5A	595.55	KL	BLINK	0.05	-0.41	NA	NA	G	A	NA	NA	NA	0.16
6B	554.79	KL	BLINK	0.11	-0.26	T	C	T	C	201	20	Increase	0.02
7D	127.14	KL	BLINK	0.11	-0.27	NA	NA	A	G	NA	NA	NA	0.00
2B	451.89	KW	BLINK	0.12	-0.44	NA	NA	A	G	NA	NA	NA	0.06
3B	691.27	KW	BLINK	0.47	0.25	NA	NA	G	A	NA	NA	NA	0.28
3D	433.50	KW	BLINK	0.15	0.23	G	T	T	G	189	27	Decrease	0.02
6B	690.97	KW	BLINK	0.35	0.20	NA	NA	T	C	NA	NA	NA	0.00
2B	451.89	TKW	BLINK	0.12	-0.40	NA	NA	A	G	NA	NA	NA	0.06
3D	433.50	TKW	BLINK	0.15	0.19	G	T	T	G	189	27	Decrease	0.02
4B	28.15	TKW	BLINK	0.08	-0.39	A	C	A	C	204	12	Increase	0.06
4B	601.75	TKW	BLINK	0.10	-0.38	A	G	A	G	198	14	Increase	0.04
2A	713.93	TW	BLINK	0.13	0.33	G	A	A	G	202	29	Decrease	0.20
2B	110.48	TW	BLINK	0.15	0.36	NA	NA	G	T	NA	NA	NA	0.06
4A	312.67	TW	BLINK	0.11	0.33	A	G	G	A	194	14	Decrease	0.03
5A	547.80	TW	BLINK	0.08	0.43	A	G	G	A	212	19	Decrease	0.07
6B	618.67	TW	BLINK	0.10	0.34	NA	NA	G	A	NA	NA	NA	0.31
7B	13.87	TW	BLINK	0.07	0.35	NA	NA	G	A	NA	NA	NA	0.02
23THD													
4B	607.11	GY	BLINK	0.08	0.36	G	A	A	G	210	17	Decrease	0.04
7B	622.25	GY	BLINK	0.16	0.33	NA	NA	A	C	NA	NA	NA	0.04
2A	713.93	KA	BLINK	0.13	-0.32	G	A	G	A	201	29	Increase	0.20
4B	633.62	KA	BLINK	0.07	-0.65	G	A	G	A	206	10	Increase	0.03
5A	609.69	KA	BLINK	0.48	-0.18	NA	NA	T	C	NA	NA	NA	0.19
2A	715.39	KL	BLINK	0.23	-0.28	NA	NA	C	A	NA	NA	NA	0.22
4B	28.15	KL	BLINK	0.05	-0.63	NA	NA	T	C	NA	NA	NA	0.06
5A	595.55	KL	BLINK	0.05	-0.44	NA	NA	G	A	NA	NA	NA	0.16
7D	124.50	KL	BLINK	0.13	-0.35	T	G	T	G	191	22	Increase	0.01
2B	175.54	KW	BLINK	0.18	-0.33	NA	NA	G	A	NA	NA	NA	0.14
2B	543.71	KW	BLINK	0.11	-0.48	NA	NA	T	G	NA	NA	NA	0.12
3D	56.71	KW	BLINK	0.18	0.36	C	T	T	C	188	39	Decrease	0.20
4B	633.62	KW	BLINK	0.07	-0.61	G	A	G	A	206	10	Increase	0.03
5A	595.55	KW	BLINK	0.05	-0.50	NA	NA	G	A	NA	NA	NA	0.16
5D	132.41	KW	BLINK	0.13	-0.35	T	C	T	C	197	29	Increase	0.16
7A	540.20	KW	BLINK	0.12	-0.46	T	C	T	C	192	17	Increase	0.07
1A	30.18	TKW	BLINK	0.17	-0.26	NA	NA	T	C	NA	NA	NA	0.16
2A	715.39	TKW	BLINK	0.23	-0.28	NA	NA	C	A	NA	NA	NA	0.22
2B	451.89	TKW	BLINK	0.12	-0.44	NA	NA	A	G	NA	NA	NA	0.06
4B	28.15	TKW	BLINK	0.05	-0.70	NA	NA	T	C	NA	NA	NA	0.06
5A	608.88	TKW	BLINK	0.44	-0.18	C	T	T	C	101	129	Decrease	0.24
6B	686.80	TKW	BLINK	0.30	0.22	NA	NA	C	T	NA	NA	NA	0.00

2A	691.23	TW	BLINK	0.08	0.54	NA	NA	C	T	NA	NA	NA	0.27
3A	13.00	TW	BLINK	0.07	0.63	NA	NA	G	A	NA	NA	NA	0.20
3B	112.27	TW	BLINK	0.41	0.33	NA	NA	C	T	NA	NA	NA	0.34
4A	16.46	TW	BLINK	0.12	0.34	C	T	T	C	190	15	Decrease	0.07
6A	550.08	TW	BLINK	0.09	-0.40	NA	NA	A	G	NA	NA	NA	0.00

Appendix Table B.9. GWAS results summary for functional quality using BLUEs

For each significant association, the chromosome (Chr), physical position in Mb (Pos), trait, model, MAF, and estimated marker effect are reported. Marker effects are relative to a standard normal distribution. Allelic information includes the allelic states of the RP and donor parents (Don), as well as the alleles associated with the lower (L) and higher (H) mean phenotypic values among genotype classes. Genotype_class columns report the number of WILs carrying the RP allele (WIL_RP) or the donor allele (WIL_Don) at the marker locus. Donor direction (DonDir) indicates whether the donor allele increases or decreases the trait value. Donor allele proportion within the 1 Mb genomic bin containing the marker (DonProp) represents the average proportion of donor alleles across WILs in that genomic region. NA indicates cases where parental alleles could not be unambiguously assigned.

Chr	Pos	Trait	Model	MAF	Effect	Allele		Genotype_class				DonDir	DonProp
						RP	Don	L	H	WIL_RP	WIL_Don		
21THD													
2A	700.54	GP	BLINK	0.10	-0.36	NA	NA	T	C	NA	NA	NA	0.06
3D	174.08	GP	BLINK	0.15	0.15	C	A	A	C	196	35	Decrease	0.13
4B	28.18	GP	BLINK	0.08	-0.36	NA	NA	C	T	NA	NA	NA	0.06
4B	601.75	GP	BLINK	0.10	-0.38	A	G	A	G	199	14	Increase	0.04
5B	462.73	GP	BLINK	0.07	-0.29	C	T	C	T	214	17	Increase	0.18
6B	431.80	GP	BLINK	0.06	-0.78	T	C	T	C	207	3	Increase	0.02
6D	197.31	GP	BLINK	0.11	-0.41	A	C	A	C	196	18	Increase	0.02
7A	481.69	GP	BLINK	0.11	-0.30	T	C	T	C	203	21	Increase	0.10
7A	719.57	GP	BLINK	0.09	-0.31	NA	NA	G	A	NA	NA	NA	0.18
6B	431.80	GP	MLMM	0.06	-0.78	T	C	T	C	207	3	Increase	0.02
6D	169.64	GP	MLMM	0.10	-0.49	NA	NA	G	A	NA	NA	NA	0.00
2B	25.28	GPD	BLINK	0.05	0.38	NA	NA	C	T	NA	NA	NA	0.35
4B	32.17	GPD	BLINK	0.08	-0.57	NA	NA	C	T	NA	NA	NA	0.05
4B	664.51	GPD	BLINK	0.06	0.39	C	A	A	C	209	8	Decrease	0.13
6B	688.31	GPD	BLINK	0.05	-0.46	NA	NA	A	G	NA	NA	NA	0.00
7A	109.48	GPD	BLINK	0.13	-0.31	G	A	G	A	200	27	Increase	0.19
5D	548.37	GPD	MLMM	0.08	0.45	NA	NA	T	C	NA	NA	NA	0.00
1B	492.79	PQS	BLINK	0.29	0.32	NA	NA	C	T	NA	NA	NA	0.37
1B	658.53	PQS	BLINK	0.05	0.57	NA	NA	T	C	NA	NA	NA	0.37
3A	647.36	PQS	BLINK	0.08	-0.47	C	T	C	T	212	19	Increase	0.10
3D	454.67	PQS	BLINK	0.10	-0.52	A	G	A	G	200	17	Increase	0.01
4B	634.33	PQS	BLINK	0.09	0.42	C	T	T	C	210	19	Decrease	0.06
6A	61.89	PQS	BLINK	0.06	-0.38	NA	NA	T	C	NA	NA	NA	0.61
7A	275.92	PQS	BLINK	0.33	0.23	NA	NA	T	G	NA	NA	NA	0.19
1B	492.79	PQS	MLMM	0.29	0.33	NA	NA	C	T	NA	NA	NA	0.37
3D	454.67	PQS	MLMM	0.10	-0.57	A	G	A	G	200	17	Increase	0.01
1B	380.89	SRC	BLINK	0.14	0.56	A	C	C	A	193	28	Decrease	0.10
1B	406.04	SRC	BLINK	0.14	-0.41	A	G	G	A	187	19	Decrease	0.08
1B	492.79	SRC	BLINK	0.29	0.25	NA	NA	C	T	NA	NA	NA	0.37
1B	514.16	SRC	BLINK	0.10	-2.08	T	C	C	T	201	18	Decrease	0.06

1B	515.05	SRC	BLINK	0.11	2.14	T	C	C	T	200	19	Decrease	0.07
1B	658.53	SRC	BLINK	0.05	0.67	NA	NA	T	C	NA	NA	NA	0.37
3A	647.36	SRC	BLINK	0.08	-0.53	C	T	C	T	212	19	Increase	0.10
3D	454.67	SRC	BLINK	0.10	-0.49	A	G	A	G	200	17	Increase	0.01
4B	2.04	SRC	BLINK	0.16	-0.26	T	C	T	C	191	33	Increase	0.17
6B	217.17	SRC	BLINK	0.07	-0.47	NA	NA	T	C	NA	NA	NA	0.06
22EL													
1B	512.14	GP	BLINK	0.37	0.16	NA	NA	C	T	NA	NA	NA	0.06
6B	431.80	GP	BLINK	0.06	-0.69	T	C	T	C	207	3	Increase	0.02
6B	684.27	GP	BLINK	0.37	0.16	NA	NA	C	T	NA	NA	NA	0.00
7B	622.25	GP	BLINK	0.16	-0.29	NA	NA	C	A	NA	NA	NA	0.04
1B	655.76	PQS	BLINK	0.42	0.19	NA	NA	C	T	NA	NA	NA	0.20
3A	647.36	PQS	BLINK	0.08	-0.43	C	T	C	T	212	19	Increase	0.10
1B	655.76	SRC	BLINK	0.42	0.20	NA	NA	C	T	NA	NA	NA	0.20
6B	217.17	SRC	BLINK	0.07	-0.46	NA	NA	T	C	NA	NA	NA	0.06
22RL													
2A	700.54	GP	BLINK	0.10	-0.37	NA	NA	T	C	NA	NA	NA	0.06
3A	438.62	GP	BLINK	0.11	-0.37	NA	NA	T	C	NA	NA	NA	0.07
4B	192.79	GP	BLINK	0.23	-0.19	T	C	T	C	174	49	Increase	0.18
5B	439.27	GP	BLINK	0.09	-0.37	NA	NA	G	T	NA	NA	NA	0.53
6A	617.18	GP	BLINK	0.21	0.18	T	C	C	T	183	48	Decrease	0.32
6B	431.80	GP	BLINK	0.06	-0.52	T	C	T	C	207	3	Increase	0.02
6D	68.33	GP	BLINK	0.13	-0.25	A	G	A	G	191	20	Increase	0.02
7A	659.55	GP	BLINK	0.06	-0.44	NA	NA	T	C	NA	NA	NA	0.04
1B	492.79	PQS	BLINK	0.29	0.29	NA	NA	C	T	NA	NA	NA	0.37
1B	568.53	PQS	BLINK	0.06	0.40	NA	NA	A	G	NA	NA	NA	0.33
1B	655.76	PQS	BLINK	0.42	0.18	NA	NA	C	T	NA	NA	NA	0.20
3A	46.83	PQS	BLINK	0.10	-0.39	T	G	T	G	202	16	Increase	0.08
3A	647.36	PQS	BLINK	0.08	-0.41	C	T	C	T	212	19	Increase	0.10
3D	454.67	PQS	BLINK	0.10	-0.35	A	G	A	G	200	17	Increase	0.01
4B	171.09	PQS	BLINK	0.06	0.46	C	T	T	C	217	12	Decrease	0.07
1B	655.76	SRC	BLINK	0.42	0.16	NA	NA	C	T	NA	NA	NA	0.20
3A	484.11	SRC	BLINK	0.08	-0.33	T	G	T	G	213	18	Increase	0.08
7A	470.89	SRC	BLINK	0.11	-0.32	C	T	C	T	204	25	Increase	0.09
22THD													
6B	142.23	GP	BLINK	0.07	-0.35	NA	NA	G	A	NA	NA	NA	0.03
6D	68.33	GP	BLINK	0.13	-0.33	A	G	A	G	191	20	Increase	0.02
1B	424.55	GPD	BLINK	0.08	0.52	C	A	A	C	210	18	Decrease	0.08
5B	573.22	GPD	BLINK	0.11	-0.43	T	C	T	C	190	12	Increase	0.05
3A	642.94	PQS	BLINK	0.13	-0.45	G	A	G	A	186	17	Increase	0.08
3D	454.67	PQS	BLINK	0.10	-0.49	A	G	A	G	200	17	Increase	0.01
3D	454.67	PQS	MLM	0.10	-0.57	A	G	A	G	200	17	Increase	0.01

3A	642.94	PQS	MLMM	0.13	-0.50	G	A	G	A	186	17	Increase	0.08
3D	454.67	PQS	MLMM	0.10	-0.57	A	G	A	G	200	17	Increase	0.01
1A	140.37	SRC	BLINK	0.12	0.25	A	G	G	A	203	26	Decrease	0.13
1B	515.05	SRC	BLINK	0.11	0.48	T	C	C	T	200	19	Decrease	0.07
1B	655.76	SRC	BLINK	0.42	0.15	NA	NA	C	T	NA	NA	NA	0.20
3A	642.94	SRC	BLINK	0.13	-0.53	G	A	G	A	186	17	Increase	0.08
3D	454.67	SRC	BLINK	0.10	-0.48	A	G	A	G	200	17	Increase	0.01
4B	7.04	SRC	BLINK	0.48	-0.15	NA	NA	A	G	NA	NA	NA	0.18
7A	562.22	SRC	BLINK	0.35	0.17	G	A	A	G	144	73	Decrease	0.24
3D	454.67	SRC	MLM	0.10	-0.60	A	G	A	G	200	17	Increase	0.01
3D	454.67	SRC	MLMM	0.10	-0.59	A	G	A	G	200	17	Increase	0.01
23RL													
1B	655.76	SRC	BLINK	0.42	0.20	NA	NA	C	T	NA	NA	NA	0.20
6B	595.29	SRC	BLINK	0.06	-0.45	G	A	G	A	215	11	Increase	0.01
23THD													
7A	138.96	GPD	MLMM	0.10	-0.51	A	C	A	C	208	23	Increase	0.13
7A	172.20	GPD	MLMM	0.09	-0.54	C	T	C	T	210	21	Increase	0.12
7A	187.44	GPD	MLMM	0.10	-0.56	C	A	C	A	206	21	Increase	0.11
1B	334.24	SRC	BLINK	0.16	0.29	G	A	A	G	194	36	Decrease	0.16
1B	566.76	SRC	BLINK	0.08	0.47	NA	NA	G	A	NA	NA	NA	0.33
2D	613.44	SRC	BLINK	0.29	0.19	NA	NA	G	A	NA	NA	NA	0.01
3A	141.01	SRC	BLINK	0.09	-0.53	G	A	G	A	200	13	Increase	0.09
3B	809.60	SRC	BLINK	0.08	-0.42	T	C	T	C	204	13	Increase	0.10
6B	182.59	SRC	BLINK	0.09	-0.56	A	G	A	G	198	10	Increase	0.13
7B	648.92	SRC	BLINK	0.09	0.30	NA	NA	T	C	NA	NA	NA	0.12

Appendix Table B.10. GWAS results summary for extended drone data using BLUPs

For each significant association, the chromosome (Chr), physical position in Mb (Pos), trait, model, MAF, and estimated marker effect are reported. Marker effects are relative to a standard normal distribution. Allelic information includes the allelic states of the RP and donor parents (Don), as well as the alleles associated with the lower (L) and higher (H) mean phenotypic values among genotype classes. Genotype_class columns report the number of WILs carrying the RP allele (WIL_RP) or the donor allele (WIL_Don) at the marker locus. Donor direction (DonDir) indicates whether the donor allele increases or decreases the trait value. Donor allele proportion within the 1 Mb genomic bin containing the marker (DonProp) represents the average proportion of donor alleles across WILs in that genomic region. NA indicates cases where parental alleles could not be unambiguously assigned.

Chr	Pos	Trait	Model	MAF	Effect	Allele		Genotype_class				DonDir	DonProp
						RP	Don	L	H	WIL_RP	WIL_Don		
2B	9.73	GR_anth	BLINK	0.14	0.30	NA	NA	G	A	NA	NA	NA	0.32
2B	23.79	GR_anth	BLINK	0.08	0.42	A	C	C	A	207	11	Decrease	0.08
5A	414.01	GR_anth	BLINK	0.40	-0.16	A	G	A	G	140	92	Increase	0.37
6D	8.66	GR_anth	BLINK	0.08	-0.23	G	A	G	A	210	16	Increase	0.01
2B	5.53	GR_anth	MLM	0.12	0.36	NA	NA	C	A	NA	NA	NA	0.69
2B	5.67	GR_anth	MLM	0.13	0.30	NA	NA	A	G	NA	NA	NA	0.69
2B	6.23	GR_anth	MLM	0.14	0.27	NA	NA	T	G	NA	NA	NA	0.59
2B	9.73	GR_anth	MLM	0.14	0.31	NA	NA	G	A	NA	NA	NA	0.32
2B	18.26	GR_anth	MLM	0.06	0.45	C	T	T	C	218	14	Decrease	0.16
2B	23.79	GR_anth	MLM	0.08	0.44	A	C	C	A	207	11	Decrease	0.08
2B	23.79	GR_anth	MLM	0.06	0.39	C	G	G	C	218	14	Decrease	0.08
2B	23.81	GR_anth	MLM	0.06	0.37	T	C	C	T	217	11	Decrease	0.08
5A	412.53	GR_anth	MLM	0.42	-0.20	T	C	T	C	132	97	Increase	0.31
5A	414.01	GR_anth	MLM	0.40	-0.19	A	G	A	G	140	92	Increase	0.37
2B	18.26	GR_anth	MLMM	0.06	0.44	C	T	T	C	218	14	Decrease	0.16
2B	18.26	GR_boot	BLINK	0.06	0.40	C	T	T	C	218	14	Decrease	0.16
3A	713.63	GR_boot	BLINK	0.12	0.23	A	G	G	A	199	21	Decrease	0.07
5A	502.79	GR_boot	BLINK	0.09	-0.24	T	C	T	C	203	13	Increase	0.11
2B	5.67	GR_boot	MLM	0.13	0.24	NA	NA	A	G	NA	NA	NA	0.69
2B	6.23	GR_boot	MLM	0.14	0.23	NA	NA	T	G	NA	NA	NA	0.59
2B	18.26	GR_boot	MLM	0.06	0.37	C	T	T	C	218	14	Decrease	0.16
2B	23.79	GR_boot	MLM	0.08	0.33	A	C	C	A	207	11	Decrease	0.08
2B	23.79	GR_boot	MLM	0.06	0.31	C	G	G	C	218	14	Decrease	0.08
2B	23.81	GR_boot	MLM	0.06	0.32	T	C	C	T	217	11	Decrease	0.08
2B	18.26	GR_boot	MLMM	0.06	0.37	C	T	T	C	218	14	Decrease	0.16
2A	718.72	GR_joint	BLINK	0.09	-0.26	NA	NA	C	T	NA	NA	NA	0.43
3B	3.91	GR_joint	BLINK	0.15	-0.20	NA	NA	C	T	NA	NA	NA	0.34
3B	399.97	GR_joint	BLINK	0.49	0.29	NA	NA	C	T	NA	NA	NA	0.17
4B	667.68	GR_joint	BLINK	0.06	-0.23	A	G	A	G	214	10	Increase	0.11
5A	412.53	GR_joint	BLINK	0.42	-0.15	T	C	T	C	132	97	Increase	0.31
5A	408.91	GR_joint	MLM	0.44	0.17	NA	NA	A	G	NA	NA	NA	0.21

5A	408.91	GR_joint	MLMM	0.44	0.17	NA	NA	A	G	NA	NA	NA	0.21
1A	495.70	GR_milk	BLINK	0.11	0.28	G	C	C	G	194	12	Decrease	0.14
2B	18.26	GR_milk	BLINK	0.06	0.32	C	T	T	C	218	14	Decrease	0.16
3B	474.76	GR_milk	BLINK	0.48	0.19	C	T	T	C	121	111	Decrease	0.32
5A	412.53	GR_milk	BLINK	0.42	-0.17	T	C	T	C	132	97	Increase	0.31
6A	554.82	GR_milk	BLINK	0.08	-0.27	C	T	C	T	207	11	Increase	0.18
2B	5.53	GR_milk	MLM	0.12	0.30	NA	NA	C	A	NA	NA	NA	0.69
2B	9.73	GR_milk	MLM	0.14	0.26	NA	NA	G	A	NA	NA	NA	0.32
2B	18.26	GR_milk	MLM	0.06	0.36	C	T	T	C	218	14	Decrease	0.16
5A	412.53	GR_milk	MLM	0.42	-0.20	T	C	T	C	132	97	Increase	0.31
2B	5.53	GR_milk	MLMM	0.12	0.31	NA	NA	C	A	NA	NA	NA	0.69
1B	383.00	NDVI_anth	BLINK	0.14	-0.14	T	C	T	C	186	21	Increase	0.10
2A	29.59	NDVI_anth	BLINK	0.06	-0.31	G	A	G	A	211	5	Increase	0.06
2D	83.17	NDVI_anth	BLINK	0.07	0.17	NA	NA	A	C	NA	NA	NA	0.00
5B	13.19	NDVI_anth	BLINK	0.16	0.15	NA	NA	A	G	NA	NA	NA	0.06
7A	112.98	NDVI_anth	BLINK	0.08	0.25	NA	NA	C	G	NA	NA	NA	0.29
2B	6.23	NDVI_anth	MLM	0.14	-0.17	NA	NA	G	T	NA	NA	NA	0.59
2B	18.26	NDVI_anth	MLM	0.06	-0.24	C	T	C	T	218	14	Increase	0.16
2B	23.81	NDVI_anth	MLM	0.06	-0.26	T	C	T	C	217	11	Increase	0.08
2B	26.10	NDVI_anth	MLM	0.06	-0.24	NA	NA	T	C	NA	NA	NA	0.35
3A	44.60	NDVI_anth	MLM	0.07	0.33	A	G	A	G	4	203	Increase	0.23
3A	45.78	NDVI_anth	MLM	0.06	0.37	NA	NA	C	T	NA	NA	NA	0.03
3A	52.35	NDVI_anth	MLM	0.08	0.30	NA	NA	A	C	NA	NA	NA	0.37
3A	61.32	NDVI_anth	MLM	0.08	0.36	C	T	C	T	9	206	Increase	0.43
3A	69.67	NDVI_anth	MLM	0.07	0.35	T	C	T	C	5	204	Increase	0.38
3A	661.48	NDVI_boot	BLINK	0.08	0.15	NA	NA	G	T	NA	NA	NA	0.10
4B	34.72	NDVI_boot	BLINK	0.06	0.18	C	T	T	C	209	6	Decrease	0.05
1B	59.60	NDVI_joint	BLINK	0.10	-0.11	T	C	T	C	209	22	Increase	0.02
1B	511.21	NDVI_joint	BLINK	0.20	-0.06	NA	NA	T	C	NA	NA	NA	0.06
2A	241.97	NDVI_joint	BLINK	0.05	-0.14	NA	NA	T	C	NA	NA	NA	0.00
3A	661.48	NDVI_joint	BLINK	0.08	0.11	NA	NA	G	T	NA	NA	NA	0.10
4B	28.23	NDVI_joint	BLINK	0.09	0.13	G	T	T	G	208	16	Decrease	0.06
4B	28.23	NDVI_milk	BLINK	0.09	0.13	G	T	T	G	208	16	Decrease	0.06
6B	276.77	NDVI_milk	BLINK	0.09	0.16	NA	NA	A	G	NA	NA	NA	0.06
2B	18.26	NDVI_milk	MLM	0.06	-0.21	C	T	C	T	218	14	Increase	0.16
2B	33.84	NDVI_milk	MLM	0.13	-0.16	T	C	T	C	203	29	Increase	0.30
3A	61.32	NDVI_milk	MLM	0.08	0.29	C	T	C	T	9	206	Increase	0.43
2B	18.26	NDVI_milk	MLMM	0.06	-0.19	C	T	C	T	218	14	Increase	0.16
3A	61.32	NDVI_milk	MLMM	0.08	0.35	C	T	C	T	9	206	Increase	0.43
7B	730.23	NDVI_milk	MLMM	0.11	-0.24	G	T	T	G	26	205	Decrease	0.52
5A	511.40	NDVI_till	BLINK	0.05	0.24	A	G	G	A	215	8	Decrease	0.08
5A	511.40	NDVI_till	MLMM	0.05	0.23	A	G	G	A	215	8	Decrease	0.08

Appendix Table B.11. GWAS results summary for extended drone data using BLUES

For each significant association, the chromosome (Chr), physical position in Mb (Pos), trait, model, MAF, and estimated marker effect are reported. Marker effects are relative to a standard normal distribution. Allelic information includes the allelic states of the RP and donor parents (Don), as well as the alleles associated with the lower (L) and higher (H) mean phenotypic values among genotype classes. Genotype_class columns report the number of WILs carrying the RP allele (WIL_RP) or the donor allele (WIL_Don) at the marker locus. Donor direction (DonDir) indicates whether the donor allele increases or decreases the trait value. Donor allele proportion within the 1 Mb genomic bin containing the marker (DonProp) represents the average proportion of donor alleles across WILs in that genomic region. NA indicates cases where parental alleles could not be unambiguously assigned.

Chr	Pos	Trait	Model	MAF	Effect	Allele		Genotype_class				DonDir	DonProp
						RP	Don	L	H	WIL_RP	WIL_Don		
21THD													
2B	23.79	GR_anth	BLINK	0.08	0.44	A	C	C	A	206	11	Decrease	0.08
3A	713.63	GR_anth	BLINK	0.09	0.24	T	C	C	T	206	15	Decrease	0.07
4A	140.13	GR_anth	BLINK	0.08	0.31	C	T	T	C	209	14	Decrease	0.08
5A	412.53	GR_anth	BLINK	0.43	-0.13	T	C	T	C	131	97	Increase	0.31
5A	470.03	GR_anth	BLINK	0.05	-0.33	T	G	T	G	215	8	Increase	0.28
5A	658.87	GR_anth	BLINK	0.08	0.36	A	G	G	A	203	10	Decrease	0.02
2B	5.53	GR_anth	MLM	0.12	0.36	NA	NA	C	A	NA	NA	NA	0.69
2B	5.67	GR_anth	MLM	0.13	0.34	NA	NA	A	G	NA	NA	NA	0.69
2B	6.23	GR_anth	MLM	0.14	0.30	NA	NA	T	G	NA	NA	NA	0.59
2B	9.73	GR_anth	MLM	0.15	0.32	NA	NA	G	A	NA	NA	NA	0.32
2B	18.26	GR_anth	MLM	0.06	0.44	C	T	T	C	217	14	Decrease	0.16
2B	23.79	GR_anth	MLM	0.08	0.44	A	C	C	A	206	11	Decrease	0.08
2B	23.79	GR_anth	MLM	0.06	0.42	C	G	G	C	217	14	Decrease	0.08
2B	5.67	GR_anth	MLMM	0.13	0.34	NA	NA	A	G	NA	NA	NA	0.69
1A	42.67	GR_boot	BLINK	0.14	0.22	T	C	C	T	194	29	Decrease	0.15
2A	52.75	GR_boot	BLINK	0.08	-0.26	T	C	T	C	205	11	Increase	0.07
2B	6.23	GR_boot	BLINK	0.14	0.26	NA	NA	T	G	NA	NA	NA	0.59
2B	241.25	GR_boot	BLINK	0.12	0.21	A	G	G	A	202	28	Decrease	0.17
3B	761.01	GR_boot	BLINK	0.06	-0.22	NA	NA	C	T	NA	NA	NA	0.31
5A	502.79	GR_boot	BLINK	0.09	-0.29	T	C	T	C	202	13	Increase	0.11
5B	680.99	GR_boot	BLINK	0.10	-0.24	NA	NA	T	C	NA	NA	NA	0.00
2A	718.72	GR_joint	BLINK	0.09	-0.50	NA	NA	C	T	NA	NA	NA	0.43
3B	3.91	GR_joint	BLINK	0.14	-0.20	NA	NA	C	T	NA	NA	NA	0.34
3B	561.10	GR_joint	BLINK	0.33	-0.24	A	G	A	G	154	74	Increase	0.32
3B	685.48	GR_joint	BLINK	0.43	-0.21	NA	NA	C	T	NA	NA	NA	0.21
3D	56.71	GR_joint	BLINK	0.18	0.37	C	T	T	C	189	39	Decrease	0.20
5A	412.53	GR_joint	BLINK	0.43	-0.30	T	C	T	C	131	97	Increase	0.31
5A	595.55	GR_joint	BLINK	0.05	-0.36	NA	NA	G	A	NA	NA	NA	0.16
6B	146.65	GR_joint	BLINK	0.07	-0.44	NA	NA	T	C	NA	NA	NA	0.14
3A	61.32	GR_milk	BLINK	0.08	0.82	C	T	C	T	9	205	Increase	0.43

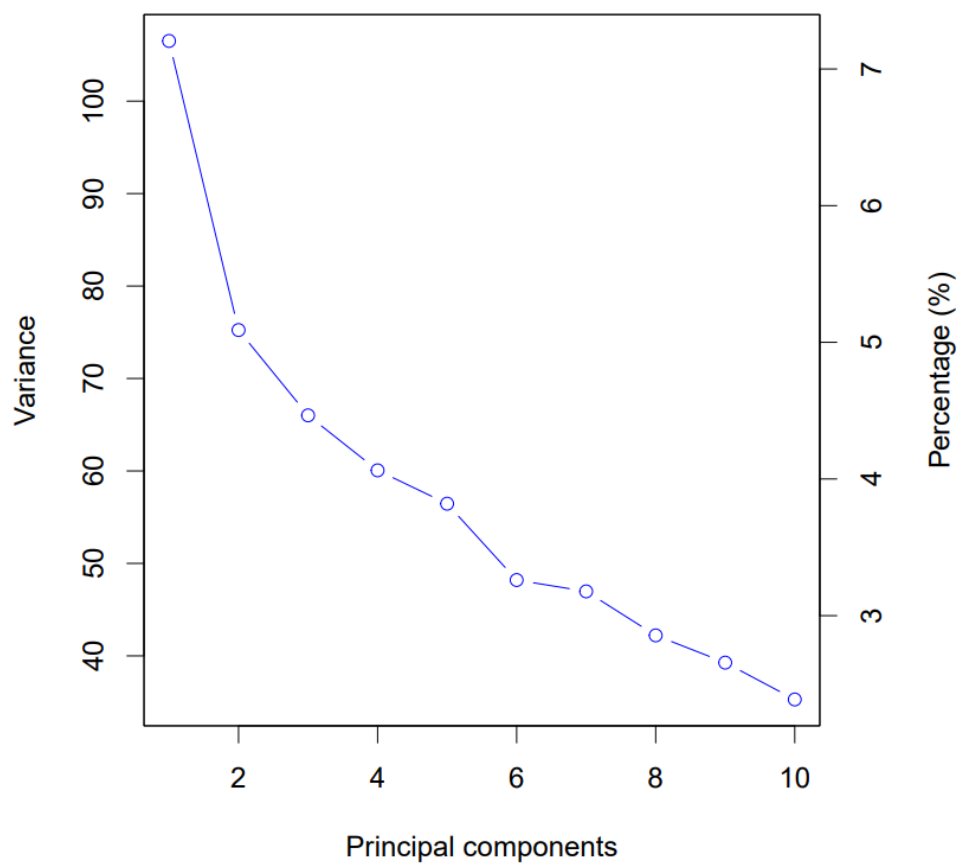
3D	613.70	GR_milk	BLINK	0.44	-0.20	NA	NA	C	T	NA	NA	NA	0.00
5A	657.04	GR_milk	BLINK	0.06	-0.66	T	G	T	G	205	4	Increase	0.02
5A	671.27	GR_milk	BLINK	0.09	0.31	NA	NA	G	A	NA	NA	NA	0.02
6A	496.38	GR_milk	BLINK	0.05	-0.33	C	T	C	T	217	10	Increase	0.14
6B	431.80	GR_milk	BLINK	0.06	0.65	T	C	C	T	207	3	Decrease	0.02
3A	61.32	GR_milk	MLM	0.08	0.68	C	T	C	T	9	205	Increase	0.43
1A	495.70	NDVI_anth	BLINK	0.11	-0.22	G	C	G	C	193	12	Increase	0.14
3A	45.78	NDVI_anth	BLINK	0.06	0.43	NA	NA	C	T	NA	NA	NA	0.03
4A	400.75	NDVI_anth	BLINK	0.11	-0.22	NA	NA	C	T	NA	NA	NA	0.00
7A	112.98	NDVI_anth	BLINK	0.08	0.39	NA	NA	C	G	NA	NA	NA	0.29
2B	6.23	NDVI_anth	MLM	0.14	-0.25	NA	NA	G	T	NA	NA	NA	0.59
2B	9.73	NDVI_anth	MLM	0.15	-0.27	NA	NA	A	G	NA	NA	NA	0.32
3A	42.38	NDVI_anth	MLM	0.07	0.48	NA	NA	T	C	NA	NA	NA	0.20
3A	44.60	NDVI_anth	MLM	0.07	0.57	A	G	A	G	4	202	Increase	0.23
3A	44.61	NDVI_anth	MLM	0.07	-0.49	T	C	T	C	203	5	Increase	0.23
3A	45.78	NDVI_anth	MLM	0.06	0.66	NA	NA	C	T	NA	NA	NA	0.03
3A	47.08	NDVI_anth	MLM	0.06	0.49	G	A	G	A	15	216	Increase	0.60
3A	50.89	NDVI_anth	MLM	0.05	0.49	A	G	A	G	7	214	Increase	0.59
3A	52.35	NDVI_anth	MLM	0.08	0.50	NA	NA	A	C	NA	NA	NA	0.37
3A	56.14	NDVI_anth	MLM	0.05	0.42	NA	NA	T	C	NA	NA	NA	0.17
3A	61.32	NDVI_anth	MLM	0.08	0.46	C	T	C	T	9	205	Increase	0.43
3A	69.67	NDVI_anth	MLM	0.07	0.55	T	C	T	C	5	204	Increase	0.38
4A	4.24	NDVI_anth	MLM	0.06	0.55	G	A	G	A	4	207	Increase	0.28
7A	94.78	NDVI_anth	MLM	0.06	0.46	NA	NA	C	T	NA	NA	NA	0.16
7A	112.98	NDVI_anth	MLM	0.08	0.54	NA	NA	C	G	NA	NA	NA	0.29
7A	125.49	NDVI_anth	MLM	0.09	0.37	NA	NA	C	T	NA	NA	NA	0.10
3A	45.78	NDVI_anth	MLMM	0.06	0.65	NA	NA	C	T	NA	NA	NA	0.03
1D	41.57	NDVI_boot	BLINK	0.14	-0.34	T	C	T	C	198	32	Increase	0.14
1D	198.02	NDVI_boot	BLINK	0.13	0.36	C	A	C	A	193	21	Increase	0.01
2D	60.61	NDVI_boot	BLINK	0.15	-0.22	NA	NA	C	T	NA	NA	NA	0.00
3A	47.08	NDVI_boot	BLINK	0.06	0.25	G	A	G	A	15	216	Increase	0.60
7A	112.98	NDVI_boot	BLINK	0.08	0.36	NA	NA	C	G	NA	NA	NA	0.29
3A	45.78	NDVI_boot	MLM	0.06	0.52	NA	NA	C	T	NA	NA	NA	0.03
3A	47.08	NDVI_boot	MLM	0.06	0.40	G	A	G	A	15	216	Increase	0.60
3A	50.89	NDVI_boot	MLM	0.05	0.45	A	G	A	G	7	214	Increase	0.59
3A	52.35	NDVI_boot	MLM	0.08	0.41	NA	NA	A	C	NA	NA	NA	0.37
3A	56.14	NDVI_boot	MLM	0.05	0.38	NA	NA	T	C	NA	NA	NA	0.17
7A	112.98	NDVI_boot	MLM	0.08	0.48	NA	NA	C	G	NA	NA	NA	0.29
7A	125.49	NDVI_boot	MLM	0.09	0.32	NA	NA	C	T	NA	NA	NA	0.10
7A	112.98	NDVI_boot	MLMM	0.08	0.48	NA	NA	C	G	NA	NA	NA	0.29
1A	518.08	NDVI_milk	BLINK	0.10	-0.25	C	T	C	T	203	19	Increase	0.23
5A	403.35	NDVI_milk	BLINK	0.06	0.45	NA	NA	C	T	NA	NA	NA	0.08
6B	431.80	NDVI_milk	BLINK	0.06	0.60	T	C	C	T	207	3	Decrease	0.02

7A	112.98	NDVI_milk	BLINK	0.08	0.55	NA	NA	C	G	NA	NA	NA	0.29
7B	338.12	NDVI_milk	BLINK	0.07	-0.29	NA	NA	T	C	NA	NA	NA	0.00
3A	42.38	NDVI_milk	MLM	0.07	0.52	NA	NA	T	C	NA	NA	NA	0.20
3A	44.60	NDVI_milk	MLM	0.07	0.57	A	G	A	G	4	202	Increase	0.23
3A	45.78	NDVI_milk	MLM	0.06	0.58	NA	NA	C	T	NA	NA	NA	0.03
3A	47.08	NDVI_milk	MLM	0.06	0.52	G	A	G	A	15	216	Increase	0.60
3A	56.14	NDVI_milk	MLM	0.05	0.49	NA	NA	T	C	NA	NA	NA	0.17
3A	61.32	NDVI_milk	MLM	0.08	0.68	C	T	C	T	9	205	Increase	0.43
6B	199.72	NDVI_milk	MLM	0.10	0.50	A	G	G	A	199	15	Decrease	0.05
6B	206.83	NDVI_milk	MLM	0.08	0.54	T	C	C	T	208	14	Decrease	0.04
6B	431.80	NDVI_milk	MLM	0.06	0.71	T	C	C	T	207	3	Decrease	0.02
6B	554.79	NDVI_milk	MLM	0.11	0.45	T	C	C	T	202	20	Decrease	0.02
6B	645.35	NDVI_milk	MLM	0.06	0.49	A	G	G	A	216	15	Decrease	0.02
7A	112.98	NDVI_milk	MLM	0.08	0.51	NA	NA	C	G	NA	NA	NA	0.29
3A	61.32	NDVI_milk	MLMM	0.08	0.60	C	T	C	T	9	205	Increase	0.43
6B	554.79	NDVI_milk	MLMM	0.11	0.42	T	C	C	T	202	20	Decrease	0.02
5A	699.47	NDVI_till	BLINK	0.09	0.27	A	G	G	A	210	19	Decrease	0.02
7A	112.98	NDVI_till	BLINK	0.08	0.33	NA	NA	C	G	NA	NA	NA	0.29
5A	699.47	NDVI_till	MLMM	0.09	0.27	A	G	G	A	210	19	Decrease	0.02
7A	112.98	NDVI_till	MLMM	0.08	0.33	NA	NA	C	G	NA	NA	NA	0.29
22EL													
2A	543.76	GR_anth	BLINK	0.08	-0.48	A	G	A	G	204	9	Increase	0.01
2B	442.52	GR_anth	BLINK	0.07	0.50	T	G	G	T	212	12	Decrease	0.07
4A	60.04	GR_anth	BLINK	0.12	0.42	G	A	A	G	193	19	Decrease	0.09
7B	63.17	GR_anth	BLINK	0.40	-0.18	C	T	T	C	90	138	Decrease	0.43
1A	483.03	GR_boot	BLINK	0.31	0.25	NA	NA	C	T	NA	NA	NA	0.30
1A	556.45	GR_boot	BLINK	0.11	-0.33	NA	NA	T	C	NA	NA	NA	0.14
2A	385.90	GR_boot	BLINK	0.10	-0.29	T	C	T	C	209	22	Increase	0.02
2B	23.79	GR_boot	BLINK	0.08	0.76	A	C	C	A	206	11	Decrease	0.08
3A	713.63	GR_boot	BLINK	0.11	0.31	A	G	G	A	199	21	Decrease	0.07
5A	414.01	GR_boot	BLINK	0.40	-0.27	A	G	A	G	139	92	Increase	0.37
2B	18.26	GR_boot	MLM	0.06	0.64	C	T	T	C	217	14	Decrease	0.16
2B	23.79	GR_boot	MLM	0.08	0.73	A	C	C	A	206	11	Decrease	0.08
2B	23.79	GR_boot	MLM	0.06	0.61	C	G	G	C	217	14	Decrease	0.08
2B	23.81	GR_boot	MLM	0.06	0.67	T	C	C	T	216	11	Decrease	0.08
2B	23.79	GR_boot	MLMM	0.08	0.72	A	C	C	A	206	11	Decrease	0.08
5A	414.01	GR_boot	MLMM	0.40	-0.28	A	G	A	G	139	92	Increase	0.37
1A	39.65	GR_joint	BLINK	0.14	0.59	C	T	T	C	190	22	Decrease	0.13
1A	140.37	GR_joint	BLINK	0.12	-0.38	A	G	A	G	203	26	Increase	0.13
2A	718.72	GR_joint	BLINK	0.09	-0.44	NA	NA	C	T	NA	NA	NA	0.43
5A	419.83	GR_joint	BLINK	0.33	-0.31	NA	NA	C	A	NA	NA	NA	0.21
5A	595.55	GR_joint	BLINK	0.05	-0.46	NA	NA	G	A	NA	NA	NA	0.16
5B	541.16	GR_joint	BLINK	0.09	-0.45	A	C	A	C	201	13	Increase	0.24

5A	408.91	GR_joint	MLMM	0.45	0.27	NA	NA	A	G	NA	NA	NA	0.21
2B	545.27	GR_milk	BLINK	0.10	0.39	NA	NA	G	A	NA	NA	NA	0.60
5A	414.01	GR_milk	BLINK	0.40	-0.18	A	G	A	G	139	92	Increase	0.37
5B	47.57	GR_milk	BLINK	0.06	0.37	A	G	A	G	15	216	Increase	0.38
1B	511.21	GR_till	BLINK	0.20	-0.26	NA	NA	T	C	NA	NA	NA	0.06
5A	512.46	GR_till	BLINK	0.07	0.53	T	C	C	T	207	10	Decrease	0.08
2B	23.81	NDVI_anth	BLINK	0.06	-0.49	T	C	T	C	216	11	Increase	0.08
3A	52.35	NDVI_anth	BLINK	0.08	0.48	NA	NA	A	C	NA	NA	NA	0.37
4B	34.72	NDVI_anth	BLINK	0.06	0.42	C	T	T	C	208	6	Decrease	0.05
5B	8.34	NDVI_anth	BLINK	0.06	0.41	NA	NA	T	C	NA	NA	NA	0.04
6A	415.53	NDVI_anth	BLINK	0.06	-0.32	A	G	A	G	216	15	Increase	0.01
7B	730.23	NDVI_anth	BLINK	0.11	-0.23	G	T	T	G	26	204	Decrease	0.52
2B	23.81	NDVI_anth	MLMM	0.06	-0.52	T	C	T	C	216	11	Increase	0.08
3A	69.67	NDVI_anth	MLMM	0.07	0.57	T	C	T	C	5	204	Increase	0.38
5B	8.34	NDVI_anth	MLMM	0.06	0.48	NA	NA	T	C	NA	NA	NA	0.04
7B	730.23	NDVI_anth	MLMM	0.11	-0.40	G	T	T	G	26	204	Decrease	0.52
1A	391.25	NDVI_boot	BLINK	0.11	0.22	A	G	G	A	198	16	Decrease	0.17
5B	8.34	NDVI_boot	BLINK	0.06	0.34	NA	NA	T	C	NA	NA	NA	0.04
2B	23.81	NDVI_boot	MLMM	0.06	-0.37	T	C	T	C	216	11	Increase	0.08
2B	26.10	NDVI_boot	MLMM	0.06	-0.33	NA	NA	T	C	NA	NA	NA	0.35
5B	8.34	NDVI_boot	MLMM	0.06	0.41	NA	NA	T	C	NA	NA	NA	0.04
2A	476.22	NDVI_joint	BLINK	0.06	0.36	NA	NA	T	G	NA	NA	NA	0.00
2B	23.81	NDVI_milk	BLINK	0.06	-0.47	T	C	T	C	216	11	Increase	0.08
4B	34.72	NDVI_milk	BLINK	0.06	0.53	C	T	T	C	208	6	Decrease	0.05
6A	415.53	NDVI_milk	BLINK	0.06	-0.42	A	G	A	G	216	15	Increase	0.01
6B	490.65	NDVI_milk	BLINK	0.12	0.30	NA	NA	A	G	NA	NA	NA	0.00
2B	18.26	NDVI_milk	MLM	0.06	-0.45	C	T	C	T	217	14	Increase	0.16
2B	23.79	NDVI_milk	MLM	0.06	-0.45	C	G	C	G	217	14	Increase	0.08
2B	23.81	NDVI_milk	MLM	0.06	-0.51	T	C	T	C	216	11	Increase	0.08
2B	23.81	NDVI_milk	MLMM	0.06	-0.51	T	C	T	C	216	11	Increase	0.08
2A	476.22	NDVI_till	BLINK	0.06	0.31	NA	NA	T	G	NA	NA	NA	0.00
3D	31.55	NDVI_till	BLINK	0.49	-0.13	C	A	C	A	118	112	Increase	0.31
4B	285.41	NDVI_till	BLINK	0.10	0.20	A	G	G	A	200	13	Decrease	0.05
5A	512.46	NDVI_till	BLINK	0.07	0.35	T	C	C	T	207	10	Decrease	0.08
22RL													
2B	18.26	GR_anth	BLINK	0.06	0.63	C	T	T	C	217	14	Decrease	0.16
2B	5.53	GR_anth	MLM	0.12	0.59	NA	NA	C	A	NA	NA	NA	0.69
2B	5.67	GR_anth	MLM	0.13	0.47	NA	NA	A	G	NA	NA	NA	0.69
2B	6.23	GR_anth	MLM	0.14	0.46	NA	NA	T	G	NA	NA	NA	0.59
2B	9.73	GR_anth	MLM	0.15	0.45	NA	NA	G	A	NA	NA	NA	0.32
2B	18.26	GR_anth	MLM	0.06	0.73	C	T	T	C	217	14	Decrease	0.16
2B	23.79	GR_anth	MLM	0.08	0.67	A	C	C	A	206	11	Decrease	0.08

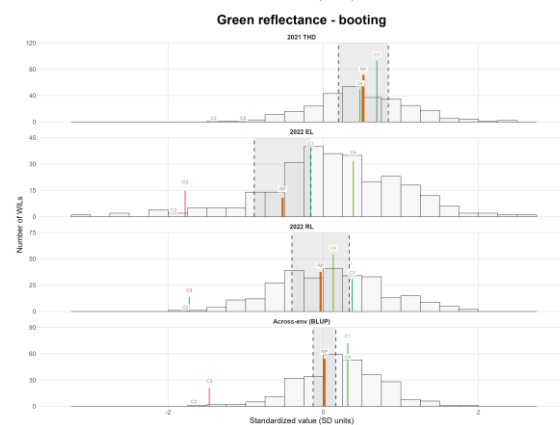
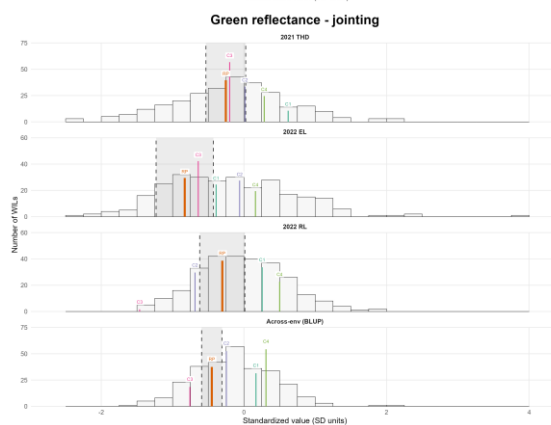
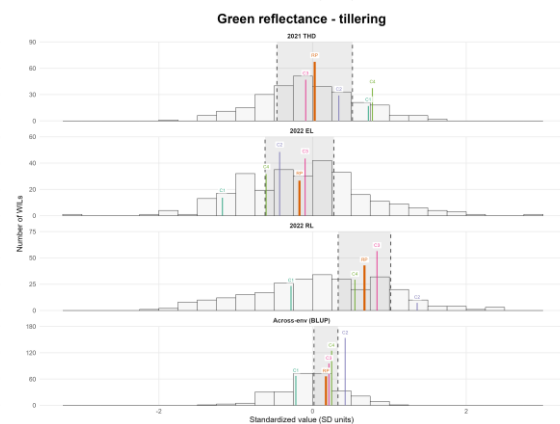
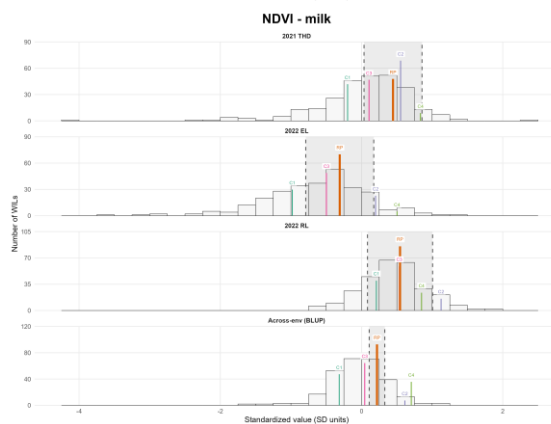
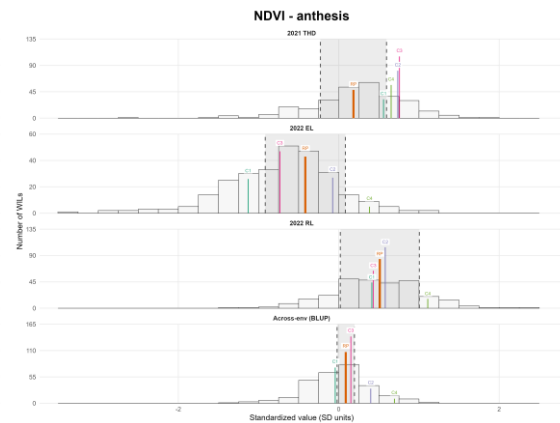
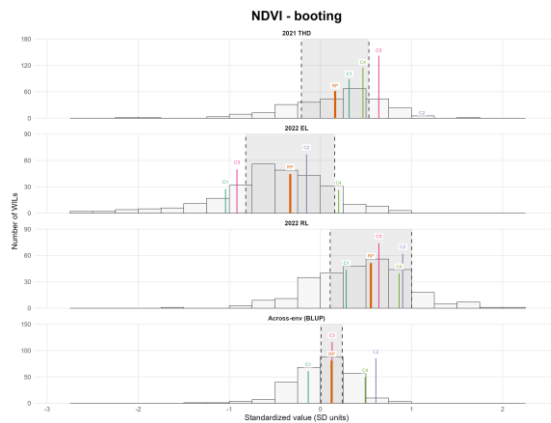
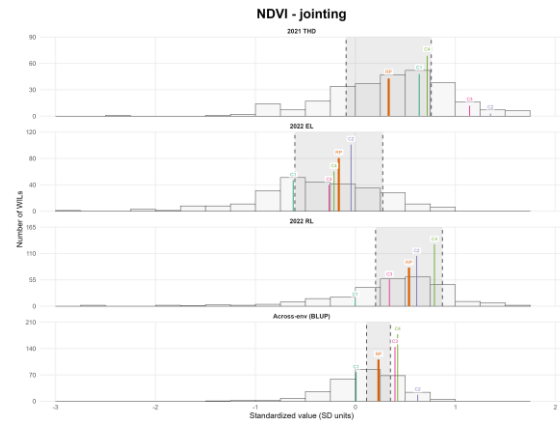
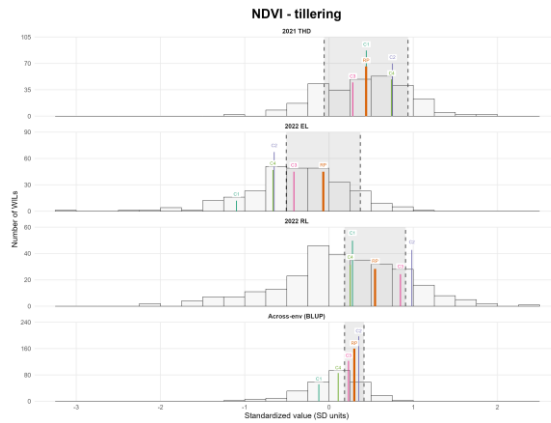
2B	23.79	GR_anth	MLM	0.06	0.62	C	G	G	C	217	14	Decrease	0.08
2B	23.81	GR_anth	MLM	0.06	0.60	T	C	C	T	216	11	Decrease	0.08
2B	18.26	GR_anth	MLMM	0.06	0.73	C	T	T	C	217	14	Decrease	0.16
2B	18.26	GR_boot	BLINK	0.06	0.45	C	T	T	C	217	14	Decrease	0.16
3A	557.71	GR_boot	BLINK	0.12	-0.26	NA	NA	C	T	NA	NA	NA	0.08
5B	47.11	GR_boot	BLINK	0.09	-0.30	NA	NA	T	C	NA	NA	NA	0.38
5B	670.83	GR_boot	BLINK	0.42	-0.19	NA	NA	T	C	NA	NA	NA	0.00
2B	18.26	GR_boot	MLM	0.06	0.45	C	T	T	C	217	14	Decrease	0.16
2B	18.26	GR_boot	MLMM	0.06	0.45	C	T	T	C	217	14	Decrease	0.16
3B	3.91	GR_joint	BLINK	0.14	-0.21	NA	NA	C	T	NA	NA	NA	0.34
5A	408.91	GR_joint	BLINK	0.45	0.17	NA	NA	A	G	NA	NA	NA	0.21
5B	529.61	GR_joint	BLINK	0.07	-0.28	T	C	T	C	208	9	Increase	0.22
6A	452.28	GR_joint	BLINK	0.07	-0.25	NA	NA	C	A	NA	NA	NA	0.00
5A	408.91	GR_joint	MLMM	0.45	0.18	NA	NA	A	G	NA	NA	NA	0.21
2B	9.73	GR_milk	BLINK	0.15	0.44	NA	NA	G	A	NA	NA	NA	0.32
2B	18.26	GR_milk	BLINK	0.06	0.43	C	T	T	C	217	14	Decrease	0.16
2B	33.84	GR_milk	BLINK	0.13	0.30	T	C	C	T	202	29	Decrease	0.30
2B	5.53	GR_milk	MLM	0.12	0.60	NA	NA	C	A	NA	NA	NA	0.69
2B	5.67	GR_milk	MLM	0.13	0.51	NA	NA	A	G	NA	NA	NA	0.69
2B	6.23	GR_milk	MLM	0.14	0.50	NA	NA	T	G	NA	NA	NA	0.59
2B	9.73	GR_milk	MLM	0.15	0.56	NA	NA	G	A	NA	NA	NA	0.32
2B	18.26	GR_milk	MLM	0.06	0.79	C	T	T	C	217	14	Decrease	0.16
2B	23.79	GR_milk	MLM	0.08	0.68	A	C	C	A	206	11	Decrease	0.08
2B	23.79	GR_milk	MLM	0.06	0.64	C	G	G	C	217	14	Decrease	0.08
2B	23.81	GR_milk	MLM	0.06	0.62	T	C	C	T	216	11	Decrease	0.08
2B	26.10	GR_milk	MLM	0.06	0.53	NA	NA	C	T	NA	NA	NA	0.35
2B	33.84	GR_milk	MLM	0.13	0.49	T	C	C	T	202	29	Decrease	0.30
2B	9.73	GR_milk	MLMM	0.15	0.42	NA	NA	G	A	NA	NA	NA	0.32
2B	18.26	GR_milk	MLMM	0.06	0.63	C	T	T	C	217	14	Decrease	0.16
5A	561.17	GR_till	BLINK	0.07	0.46	A	C	C	A	206	8	Decrease	0.06
5B	455.17	GR_till	BLINK	0.10	0.55	A	G	G	A	195	11	Decrease	0.21
5B	588.88	GR_till	BLINK	0.09	0.46	A	C	C	A	203	14	Decrease	0.01
7A	162.90	GR_till	BLINK	0.06	-0.52	NA	NA	A	G	NA	NA	NA	0.12
4B	28.23	NDVI_anth	BLINK	0.09	0.24	G	T	T	G	207	16	Decrease	0.06
7B	608.72	NDVI_anth	BLINK	0.05	-0.47	G	A	A	G	3	210	Decrease	0.35
4B	34.72	NDVI_boot	BLINK	0.06	0.32	C	T	T	C	208	6	Decrease	0.05
5A	470.19	NDVI_boot	BLINK	0.09	0.25	NA	NA	T	C	NA	NA	NA	0.28
6D	65.64	NDVI_boot	BLINK	0.12	0.24	G	A	A	G	190	14	Decrease	0.02
7A	679.88	NDVI_boot	BLINK	0.07	-0.27	NA	NA	A	G	NA	NA	NA	0.18
2A	65.40	NDVI_joint	BLINK	0.08	-0.23	NA	NA	A	G	NA	NA	NA	0.72
4B	28.23	NDVI_joint	BLINK	0.09	0.28	G	T	T	G	207	16	Decrease	0.06
5B	588.88	NDVI_joint	BLINK	0.09	0.32	A	C	C	A	203	14	Decrease	0.01

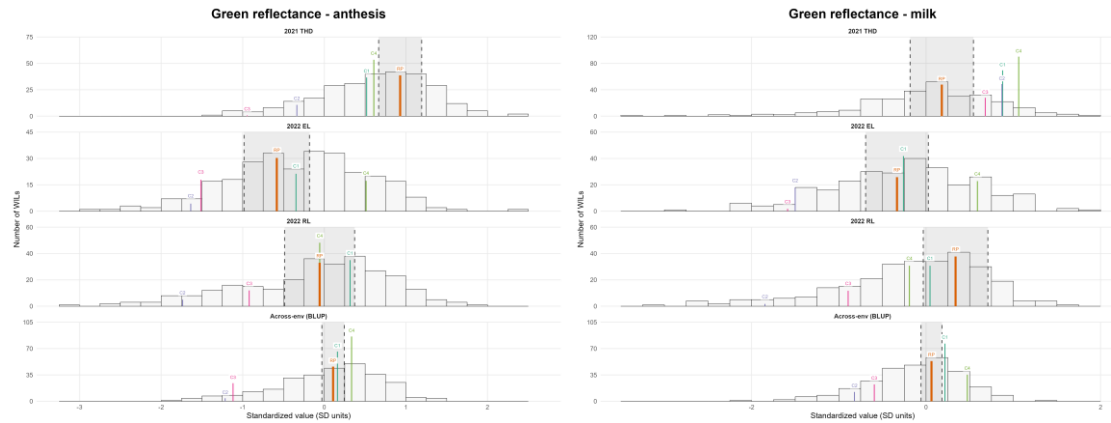
6D	10.45	NDVI_joint	BLINK	0.23	-0.19	NA	NA	G	A	NA	NA	NA	0.00
6D	458.14	NDVI_joint	BLINK	0.07	0.26	NA	NA	G	A	NA	NA	NA	0.06
1A	495.70	NDVI_milk	BLINK	0.11	-0.17	G	C	G	C	193	12	Increase	0.14
2B	33.84	NDVI_milk	BLINK	0.13	-0.23	T	C	T	C	202	29	Increase	0.30
2D	67.58	NDVI_milk	BLINK	0.16	0.17	T	C	C	T	187	28	Decrease	0.03
5B	7.94	NDVI_milk	BLINK	0.07	0.25	A	C	C	A	212	15	Decrease	0.06
2B	33.84	NDVI_milk	MLM	0.13	-0.22	T	C	T	C	202	29	Increase	0.30
2B	33.84	NDVI_milk	MLMM	0.13	-0.22	T	C	T	C	202	29	Increase	0.30
1B	59.60	NDVI_till	BLINK	0.10	-0.36	T	C	T	C	208	22	Increase	0.02
2A	364.40	NDVI_till	BLINK	0.06	-0.32	NA	NA	G	A	NA	NA	NA	0.00
5B	588.88	NDVI_till	BLINK	0.09	0.41	A	C	C	A	203	14	Decrease	0.01
5B	588.88	NDVI_till	MLM	0.09	0.47	A	C	C	A	203	14	Decrease	0.01
5B	588.88	NDVI_till	MLMM	0.09	0.48	A	C	C	A	203	14	Decrease	0.01



Appendix Figure B.1. Scree plot of the first ten principal components from genome-wide SNP marker data

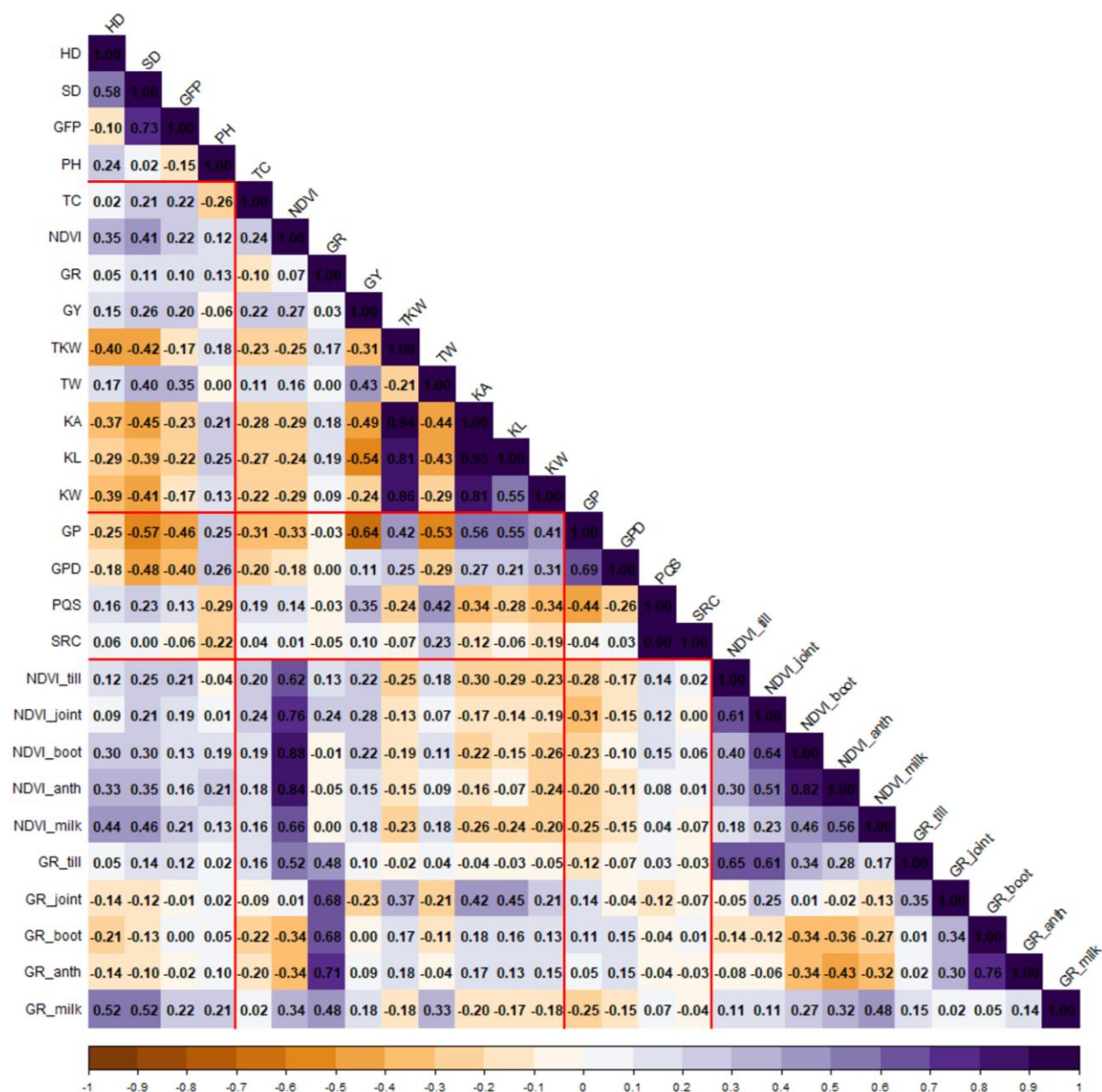
The left axis indicates the eigenvalue (variance), and the right axis shows the corresponding percentage of total variance explained by each component. Plot was generated using GAPIT version 3 (Wang & Zhang, 2021).





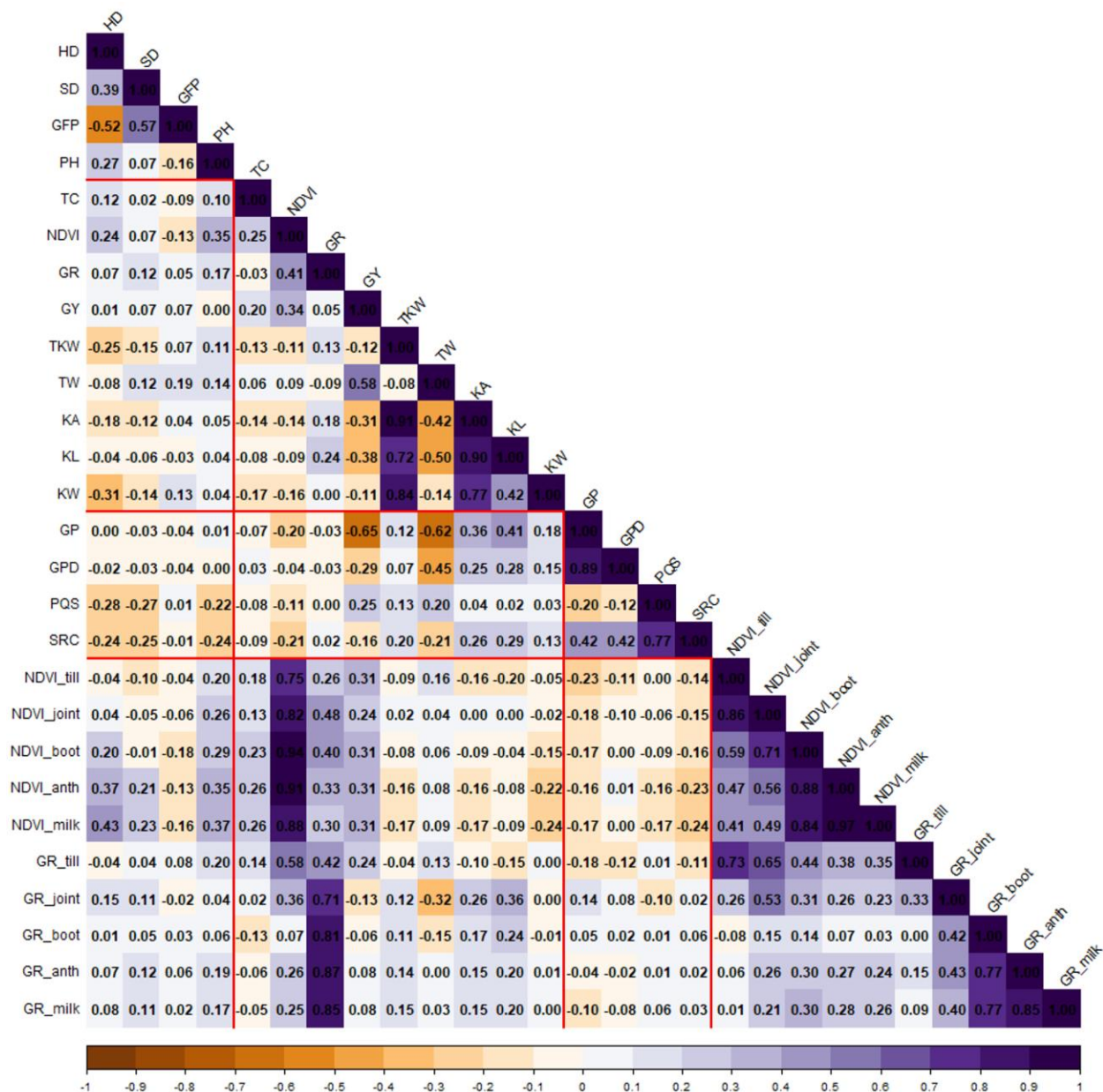
Appendix Figure B.2. Distributions of normalized difference vegetation index (NDVI) and green reflectance performance by growth stage

Measurements were collected at tillering (Feekes 3), jointing (Feekes 6), booting (Feekes 10), anthesis (Feekes 10.5.1-10.5.3), and milk/dough (Feekes 11.1-11.3). Facets by environment show within environment BLUEs, while the across environment facet represents BLUPs. Colored vertical lines indicate the recurrent parent (RP=KS061278M-4) and performance checks (C1=Billings, C2=Joe, C3=Showdown, C4=SY Monument). Gray shaded areas represent the RP's approximate 95% confidence intervals for BLUEs and prediction intervals for BLUPs, calculated as each estimate ± 2 standard errors. Plots were generated in R using the ggplot2 package (Wickham, 2016).



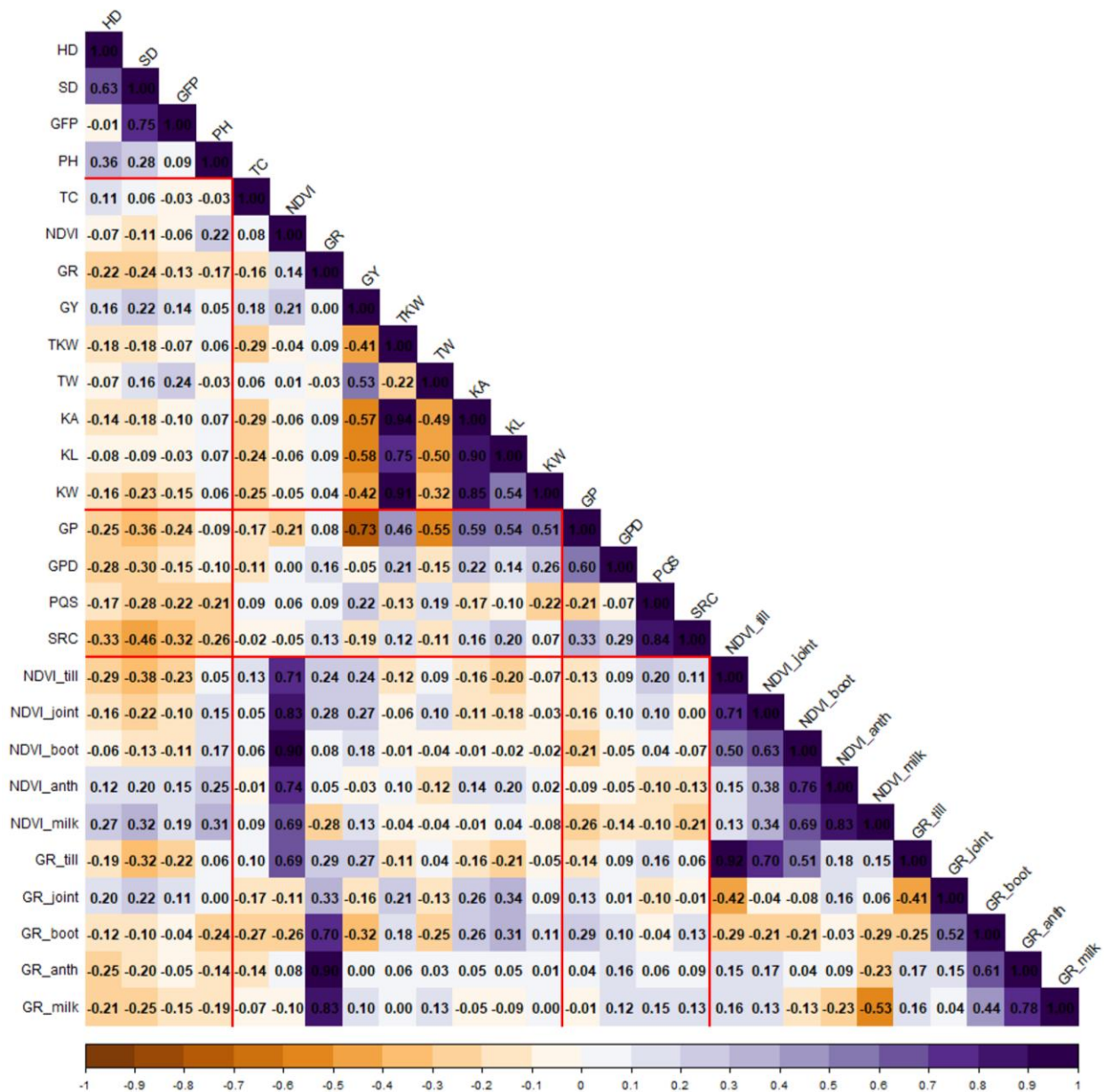
Appendix Figure B.3. Pearson correlations for agronomic performance, yield and kernel component, functional quality, and extended drone data BLUEs in 21THD for TdGD Pop80 WILs

Values within each cell represent the correlation coefficient, with color indicating both magnitude (darker = stronger) and direction (orange = negative, purple = positive). Red grid lines delimit groups of traits separated by agronomic performance, yield and kernel components, functional quality, and extended drone data. Correlations were computed and visualized in R using the corrplot package (Wei & Simko, 2024).



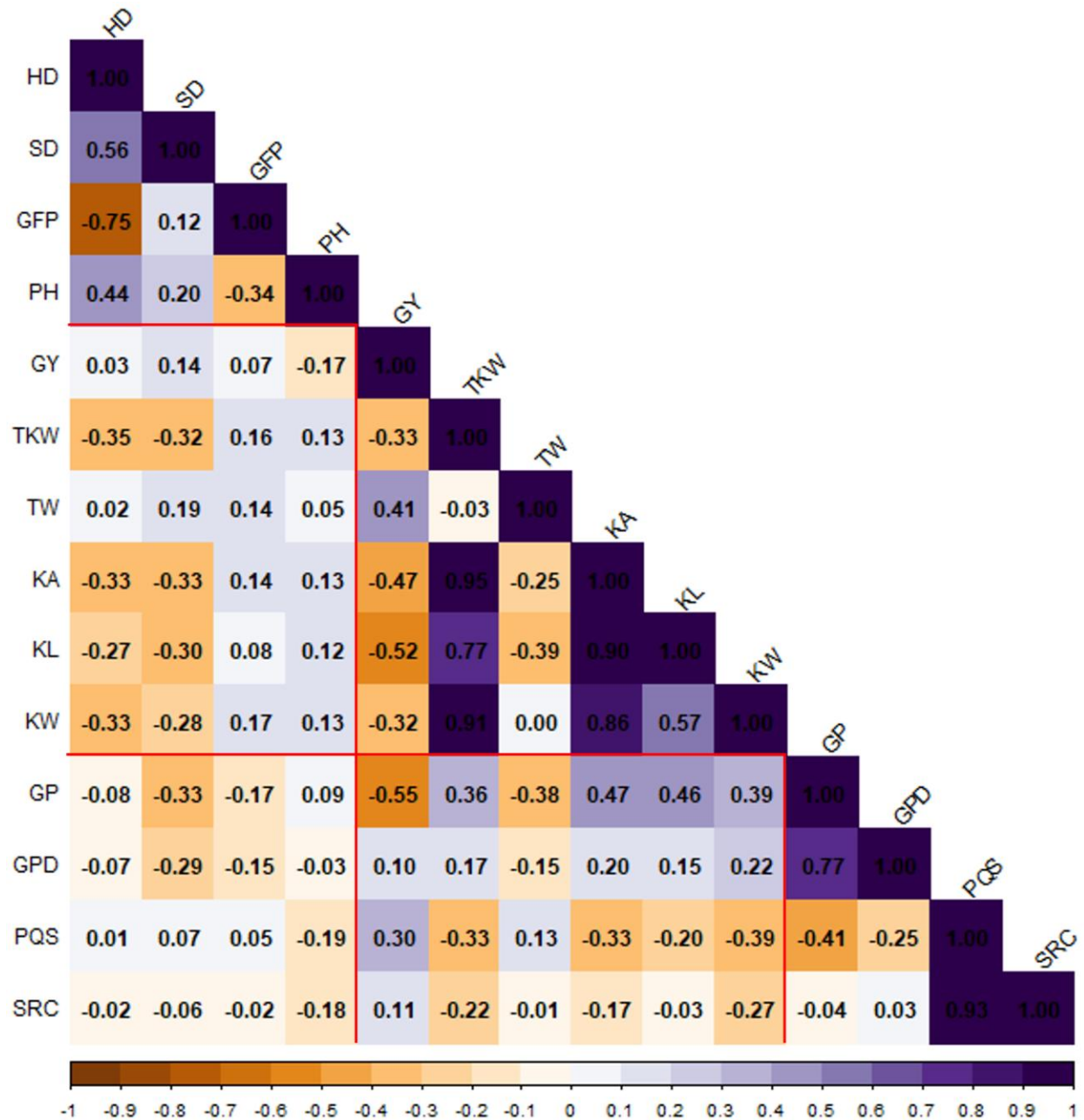
Appendix Figure B.4. Pearson correlations for agronomic performance, yield and kernel component, functional quality, and extended drone data BLUEs in 22EL for TdGD Pop80 WILs

Values within each cell represent the correlation coefficient, with color indicating both magnitude (darker = stronger) and direction (orange = negative, purple = positive). Red grid lines delimit groups of traits separated by agronomic performance, yield and kernel components, functional quality, and extended drone data. Correlations were computed and visualized in R using the corplot package (Wei & Simko, 2024).



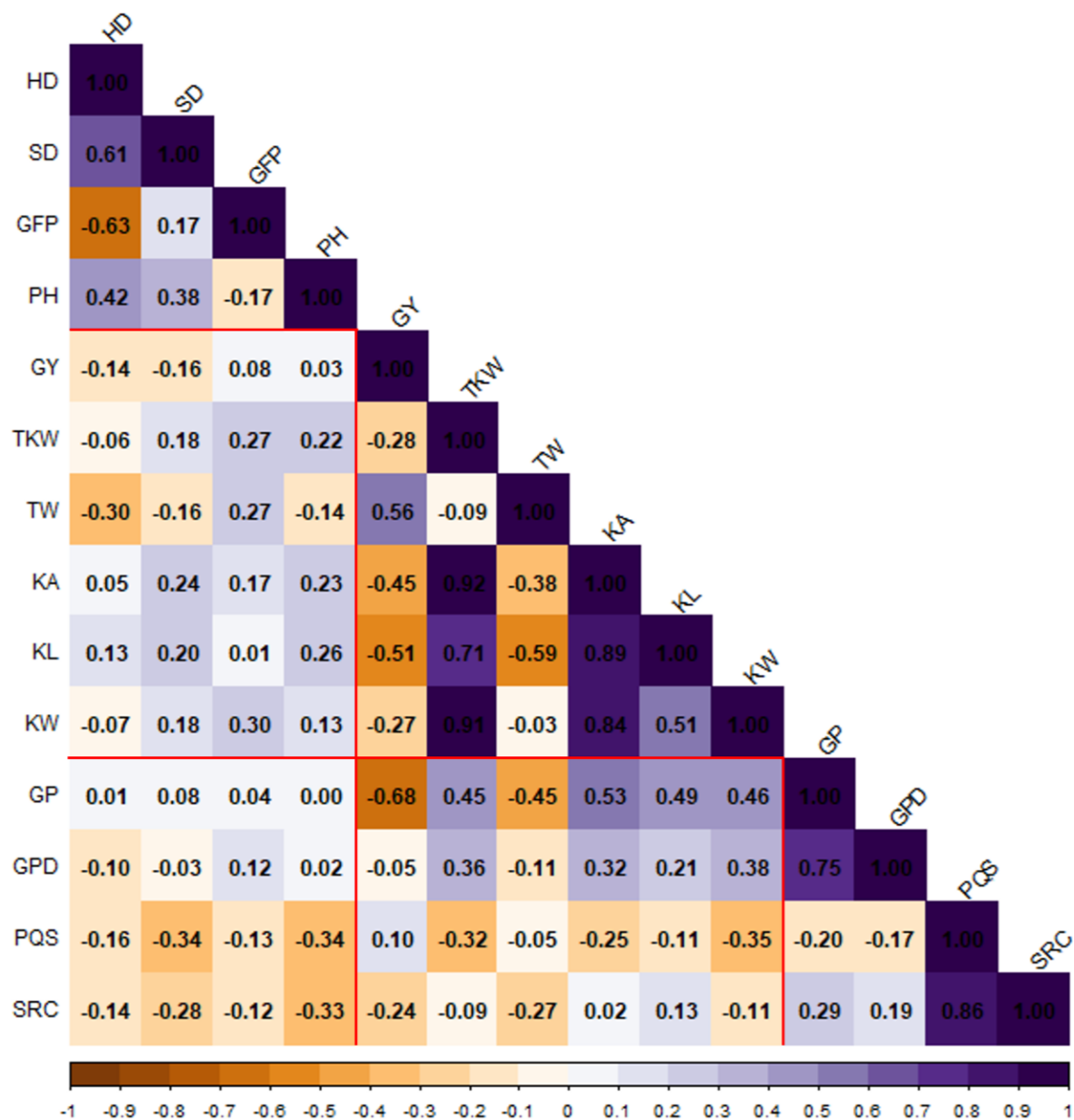
Appendix Figure B.5. Pearson correlations for agronomic performance, yield and kernel component, functional quality, and extended drone data BLUEs in 22RL for TdGD Pop80 WILs

Values within each cell represent the correlation coefficient, with color indicating both magnitude (darker = stronger) and direction (orange = negative, purple = positive). Red grid lines delimit groups of traits separated by agronomic performance, yield and kernel components, functional quality, and extended drone data. Correlations were computed and visualized in R using the corplot package (Wei & Simko, 2024).



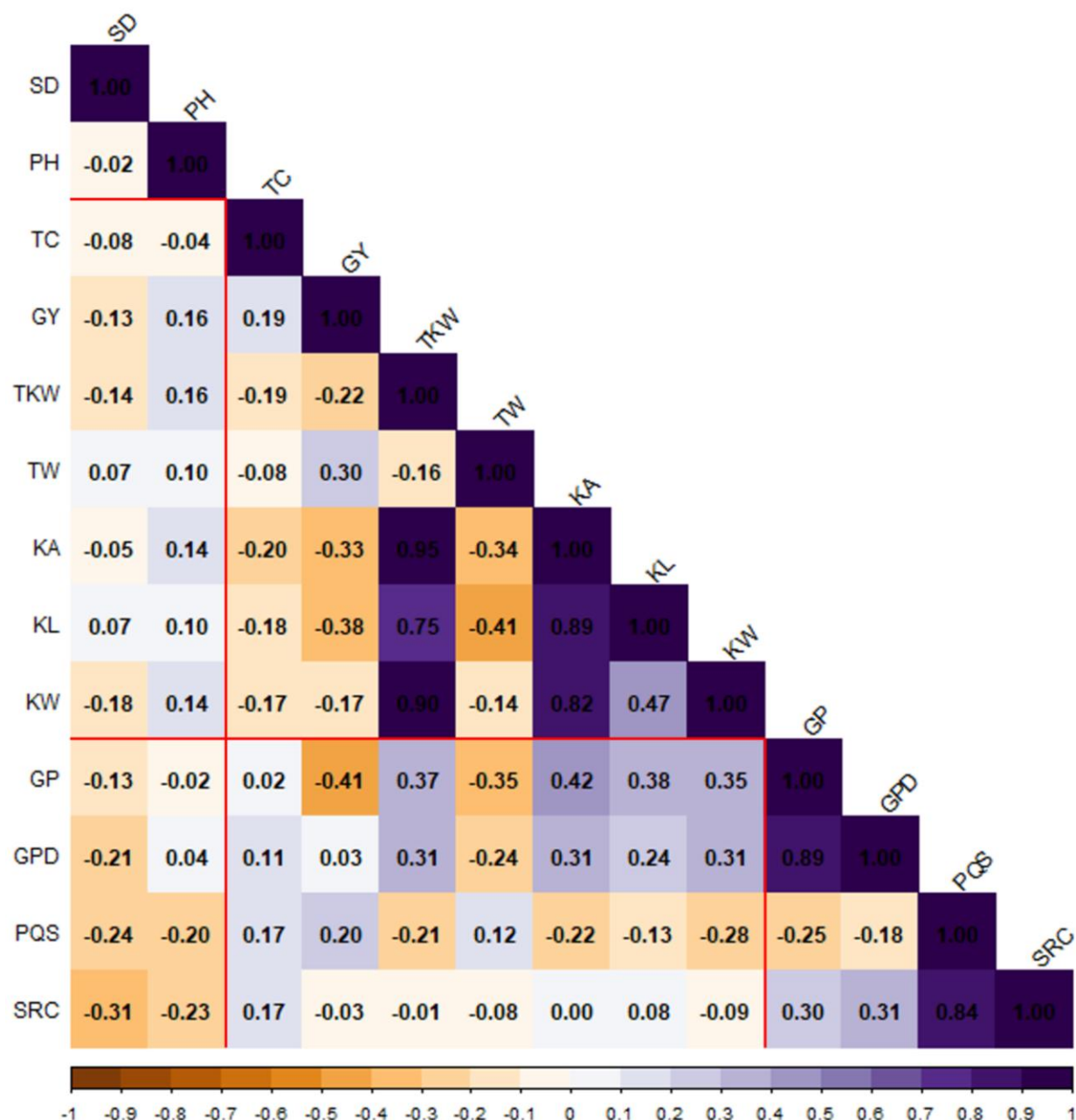
Appendix Figure B.6. Pearson correlations for agronomic performance, yield and kernel component, and functional quality BLUEs in 22THD for TdGD Pop80 WILs

Values within each cell represent the correlation coefficient, with color indicating both magnitude (darker = stronger) and direction (orange = negative, purple = positive). Red grid lines delimit groups of traits separated by agronomic performance, yield and kernel components, and functional quality. Correlations were computed and visualized in R using the corrplot package (Wei & Simko, 2024).



Appendix Figure B.7. Pearson correlations for agronomic performance, yield and kernel component, and functional quality BLUEs in 23RL for TdGD Pop80 WILs

Values within each cell represent the correlation coefficient, with color indicating both magnitude (darker = stronger) and direction (orange = negative, purple = positive). Red grid lines delimit groups of traits separated by agronomic performance, yield and kernel components, and functional quality. Correlations were computed and visualized in R using the corrplot package (Wei & Simko, 2024).



Appendix Figure B.8. Pearson correlations for agronomic performance, yield and kernel component, and functional quality BLUEs in 23THD for TdGD Pop80 WILs

Values within each cell represent the correlation coefficient, with color indicating both magnitude (darker = stronger) and direction (orange = negative, purple = positive). Red grid lines delimit groups of traits separated by agronomic performance, yield and kernel components, and functional quality. Correlations were computed and visualized in R using the corrplot package (Wei & Simko, 2024).

Appendix C - Supplementary tables and figures for Chapter 4

Appendix Table C.1. Soil testing data for Thomas County, Kansas in 2022 and 2023

Soil testing was conducted by the Kansas State University Soil Testing Laboratory using soil samples collected to a depth of 30.48 centimeters. Soil pH was measured using a 1:1 slurry of prepared soil and deionized water. Organic matter (OM) was estimated by loss on ignition. Cation exchange capacity (CEC), reported in milliequivalents per 100 grams of soil (meq 100 g⁻¹), was estimated using a displacement method with saturating ammonium acetate. Phosphorus (P) was obtained using Melich III extraction to estimate plant availability of soil P. Plant availability of soil sulfur (S) was measured using a monocalcium phosphate extraction of sulfate. Exchangeable aluminum (Al) was extracted using potassium chloride. Cadmium (Cd) was obtained using nitric acid digestion. Inorganic nitrogen (N) forms, nitrate-N (NO₃⁻) and ammonium-N (NH₄⁺), was extracted using potassium chloride. Available N, expressed in pounds per acre (lb acre⁻¹), was calculated from NO₃⁻ and sampling depth conversion. Plant availability of soil exchangeable cations potassium (K⁺), magnesium (Mg²⁺), calcium (Ca²⁺), and sodium (Na⁺) were extracted using ammonium acetate. DTPA extraction was used to measure Plant availability of soil micronutrients iron (Fe³⁺), zinc (Zn²⁺), copper (Cu²⁺), and manganese (Mn²⁺).

Test	Units	Environment	
		22THD	23THD
pH	-	7.8	7.7
OM	%	-	2.2
CEC	meq 100 g ⁻¹	26.0	-
P	ppm	16.0	32.9
S	ppm	6.4	6.5
Al	ppm	-	2.1
Cd	ppm	1.0	1.0
Nitrogen			
NO ₃ ⁻	ppm	23.2	35.9
NH ₄ ⁺	ppm	-	2.2
Available N	lb acre ⁻¹	92.8	143.6
Cations			
K ⁺	ppm	700.0	765.0
Mg ²⁺	ppm	493.0	429.2
Ca ²⁺	ppm	3973.0	4909.2
Na ⁺	ppm	44.0	12.1
Micronutrients			
Fe ³⁺	ppm	-	10.4
Zn ²⁺	ppm	-	0.6
Cu ²⁺	ppm	-	0.8
Mn ²⁺	ppm	-	6.3

Appendix Table C.2. Ionomic trait means and standard deviations in Thomas County across years

Means and standard deviations (StdDev) were calculated from raw, unstandardized observations within each environment and are provided as a reference for comparing traits among environments before within-environment standardization and subsequent two-stage mixed model analysis. Trait abbreviations (Abb) are provided; measurement units are mg kg⁻¹.

Trait	Abb	2021 THD		2022 THD		2023 THD	
		Mean	StdDev	Mean	StdDev	Mean	StdDev
Boron	B	2.562	1.030	2.858	1.015	2.193	1.122
Sodium	Na	3.140	0.815	3.723	0.826	3.081	0.928
Magnesium	Mg	10218.145	1351.008	9805.044	1179.031	10152.738	1131.038
Aluminum	Al	2.807	1.212	4.163	1.423	3.598	1.501
Phosphorus	P	4563.160	790.197	4665.651	767.976	5001.316	655.481
Potassium	K	5865.703	886.158	6851.096	983.776	6430.839	799.786
Calcium	Ca	817.097	159.318	871.658	155.284	889.823	163.906
Iron	Fe	49.080	8.440	55.668	7.382	50.879	6.910
Manganese	Mn	109.913	19.715	104.955	18.351	98.730	14.930
Cobalt	Co	0.010	0.003	0.007	0.002	0.006	0.003
Nickel	Ni	1.149	0.354	0.751	0.239	0.541	0.191
Copper	Cu	9.549	1.536	8.706	1.245	7.725	1.283
Zinc	Zn	34.067	7.409	24.212	7.481	27.908	7.531
Arsenic	As	0.022	0.008	0.019	0.006	0.015	0.005
Selenium	Se	0.710	0.275	0.819	0.301	1.003	0.235
Rubidium	Rb	1.867	0.368	1.786	0.304	1.758	0.280
Strontium	Sr	5.182	0.919	5.545	0.952	5.016	0.843
Molybdenum	Mo	2.936	0.719	1.400	0.420	1.544	0.387
Cadmium	Cd	0.006	0.002	0.013	0.005	0.011	0.004

Appendix Table C.3. Within-environment estimates of repeatability for ionic traits in TdGD Pop80

Repeatability (R^2_{within}) was estimated using SpATS models fitted separately for each trait and environment, with genotype included as a random effect during spatial (row x column) correction. R^2_{within} was calculated as the ratio of genotypic variance to total variance (genotypic + residual). These estimates reflect the reproducibility of genotype performance among replicate plots within each environment, where higher spatial heterogeneity is expected to reduce R^2_{within} .

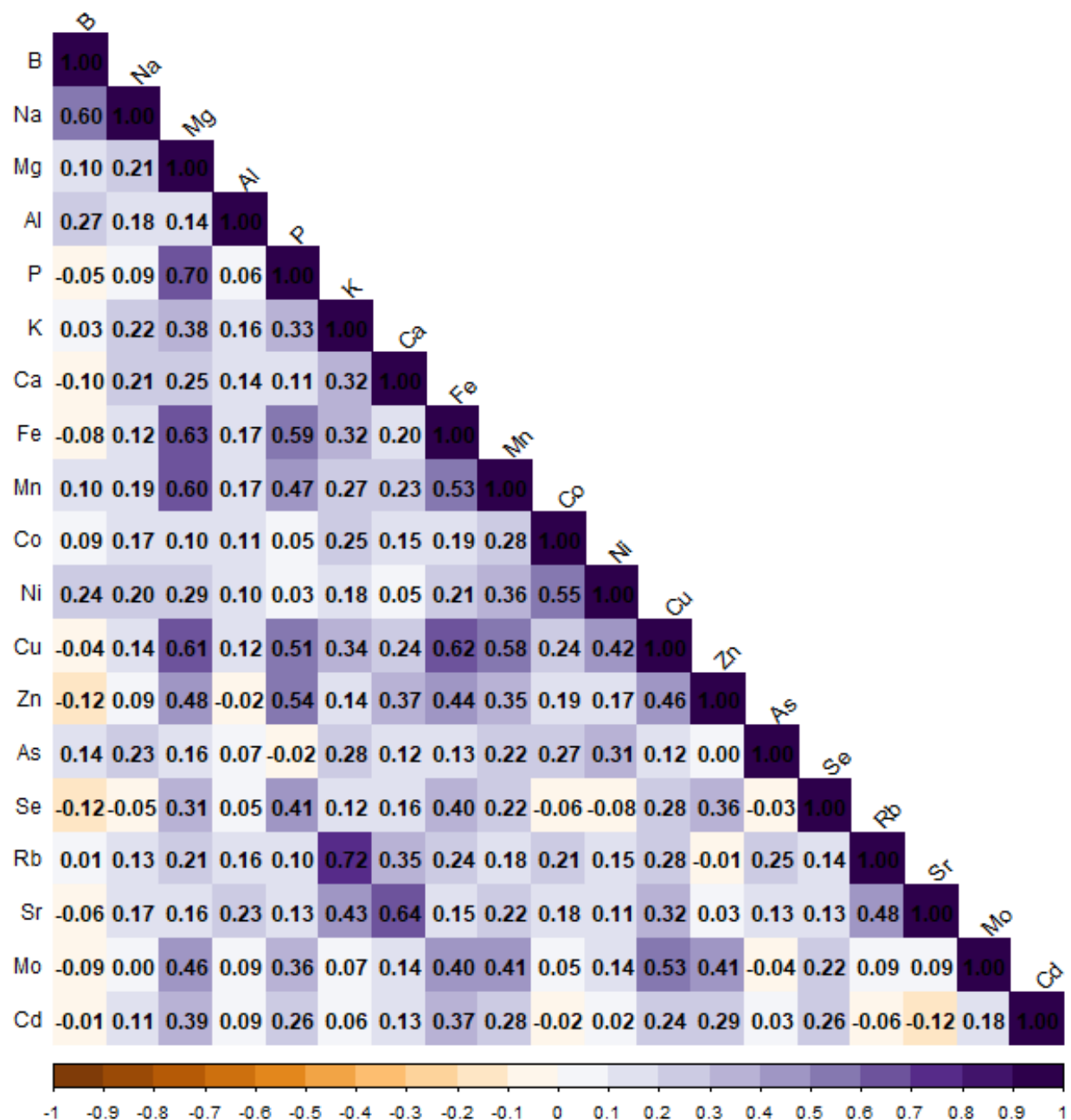
Trait	R^2_{within}		
	21THD	22THD	23THD
B	0.26	0.00	0.40
Na	0.40	0.15	0.32
Mg	0.46	0.20	0.36
Al	0.21	0.22	0.45
P	0.50	0.24	0.32
K	0.50	0.16	0.47
Ca	0.25	0.09	0.38
Fe	0.45	0.16	0.23
Mn	0.68	0.41	0.55
Co	0.44	0.12	0.19
Ni	0.67	0.35	0.43
Cu	0.49	0.30	0.51
Zn	0.29	0.05	0.40
As	0.44	0.19	0.45
Se	0.35	0.03	0.31
Rb	0.57	0.10	0.38
Sr	0.54	0.02	0.53
Mo	0.63	0.28	0.30
Cd	0.32	0.32	0.47

Appendix Table C.4. GWAS results summary for ionomics using BLUEs

For each significant association, the chromosome (Chr), physical position in Mb (Pos), trait, model, MAF, and estimated marker effect are reported. Marker effects are relative to a standard normal distribution. Allelic information includes the allelic states of the RP and donor parents (Don), as well as the alleles associated with the lower (L) and higher (H) mean phenotypic values among genotype classes. Genotype_class columns report the number of WILs carrying the RP allele (WIL_RP) or the donor allele (WIL_Don) at the marker locus. Donor direction (DonDir) indicates whether the donor allele increases or decreases the trait value. Donor allele proportion within the 1 Mb genomic bin containing the marker (DonProp) represents the average proportion of donor alleles across WILs in that genomic region. NA indicates cases where parental alleles could not be unambiguously assigned.

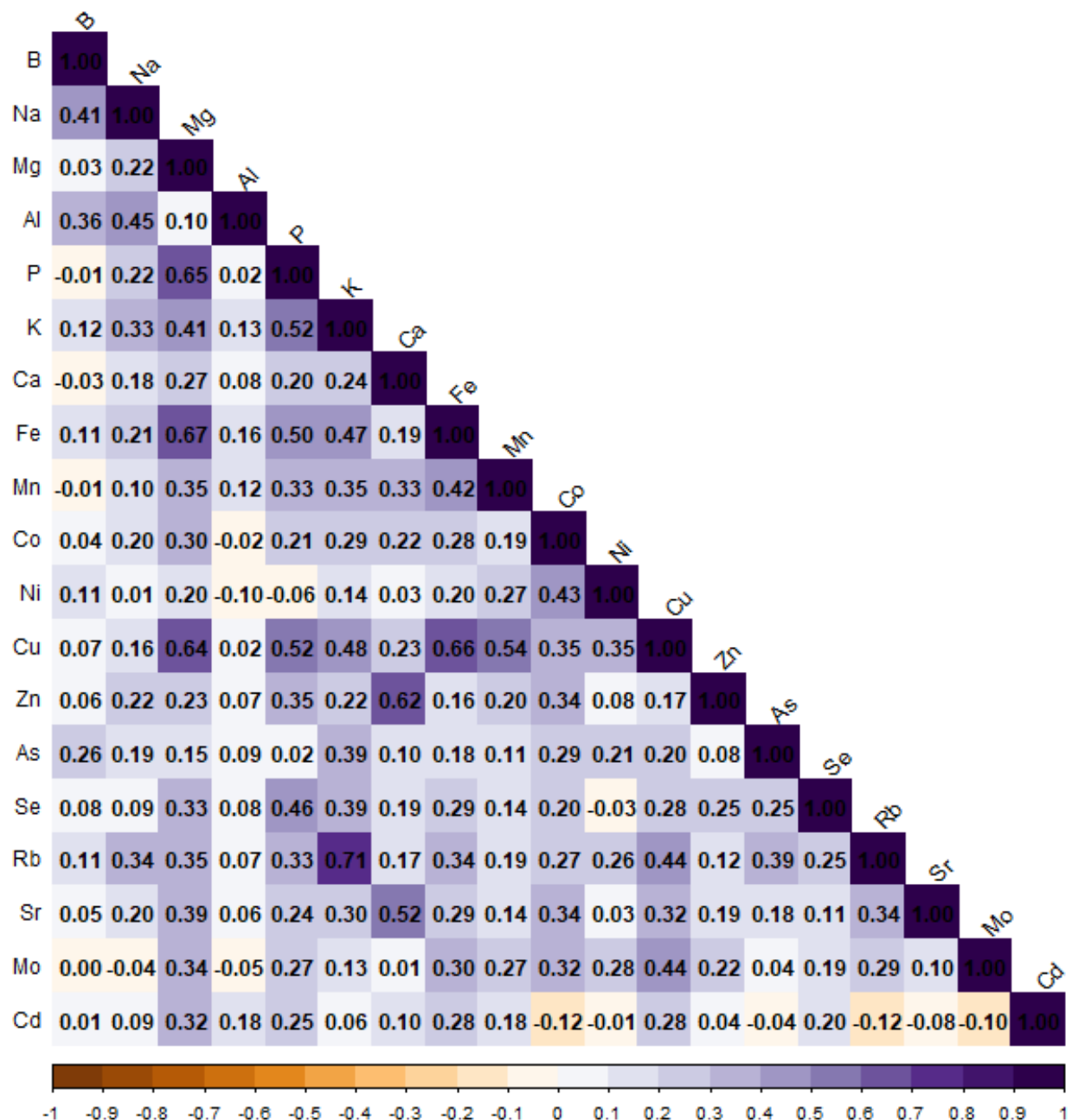
Chr	Pos	Trait	Model	MAF	Effect	Allele		Genotype_class				DonDir	DonProp
						RP	Don	L	H	WIL_RP	WIL_Don		
21THD													
3B	693.64	Cu	BLINK	0.37	0.35	G	A	A	G	130	73	Decrease	0.31
3D	31.55	Fe	BLINK	0.48	0.24	C	A	A	C	114	107	Decrease	0.31
3D	487.81	Fe	BLINK	0.14	-0.35	G	A	G	A	181	19	Increase	0.01
4B	634.32	Fe	BLINK	0.08	-0.70	G	A	G	A	199	14	Increase	0.06
6B	431.80	Fe	BLINK	0.06	-0.76	T	C	T	C	198	3	Increase	0.02
6B	616.32	Fe	BLINK	0.07	0.58	G	T	G	T	196	7	Increase	0.19
6B	618.67	Fe	BLINK	0.07	-0.55	NA	NA	T	C	NA	NA	NA	0.31
4B	634.32	Fe	MLMM	0.08	-0.75	G	A	G	A	199	14	Increase	0.06
4B	38.79	Se	BLINK	0.07	-0.47	C	T	C	T	207	15	Increase	0.11
4B	38.79	Se	MLMM	0.07	-0.47	C	T	C	T	207	15	Increase	0.11
22THD													
6A	578.89	Mo	BLINK	0.07	-0.49	A	G	A	G	204	8	Increase	0.01
23THD													
1A	568.47	Ca	BLINK	0.24	-0.30	NA	NA	G	A	NA	NA	NA	0.22
3B	251.18	Ca	BLINK	0.43	0.39	NA	NA	T	C	NA	NA	NA	0.41
1A	568.47	Ca	MLMM	0.24	-0.31	NA	NA	G	A	NA	NA	NA	0.22
3B	251.18	Ca	MLMM	0.43	0.51	NA	NA	T	C	NA	NA	NA	0.41
2D	27.46	Cu	BLINK	0.10	0.36	T	C	C	T	190	9	Decrease	0.02
4B	30.86	Cu	BLINK	0.09	-0.42	NA	NA	T	C	NA	NA	NA	0.05
5A	442.35	Cu	BLINK	0.41	0.20	NA	NA	C	A	NA	NA	NA	0.29
2D	27.46	Fe	BLINK	0.10	0.36	T	C	C	T	190	9	Decrease	0.02
3D	31.69	Fe	BLINK	0.37	0.20	NA	NA	C	T	NA	NA	NA	0.31
4B	34.72	Fe	BLINK	0.06	-0.77	C	T	C	T	205	6	Increase	0.05
5B	548.44	Mo	BLINK	0.07	-0.33	A	C	A	C	209	12	Increase	0.07
6A	67.06	Mo	BLINK	0.06	-0.48	NA	NA	G	A	NA	NA	NA	0.01
6B	626.21	Mo	BLINK	0.10	-0.33	NA	NA	T	C	NA	NA	NA	0.01
7D	400.87	Mo	BLINK	0.09	-0.36	T	C	T	C	208	20	Increase	0.01
1A	568.01	P	BLINK	0.23	-0.29	C	T	C	T	174	52	Increase	0.22
5A	673.74	P	BLINK	0.08	-0.40	NA	NA	T	C	NA	NA	NA	0.25
5B	392.50	P	BLINK	0.10	-0.39	NA	NA	C	T	NA	NA	NA	0.11

7A	96.54	P	BLINK	0.43	0.24	NA	NA	G	A	NA	NA	NA	0.16
1D	10.54	Rb	BLINK	0.11	0.33	T	C	C	T	201	21	Decrease	0.12
2B	45.55	Rb	BLINK	0.05	-0.76	NA	NA	A	C	NA	NA	NA	0.31
2B	45.55	Rb	MLM	0.05	-0.70	NA	NA	A	C	NA	NA	NA	0.31
2B	45.55	Rb	MLMM	0.05	-0.70	NA	NA	A	C	NA	NA	NA	0.31
1A	518.08	Sr	BLINK	0.10	-0.35	C	T	C	T	201	18	Increase	0.23
4A	29.08	Sr	BLINK	0.14	-0.27	NA	NA	C	T	NA	NA	NA	0.18
6B	690.97	Sr	BLINK	0.34	-0.21	NA	NA	C	T	NA	NA	NA	0.00



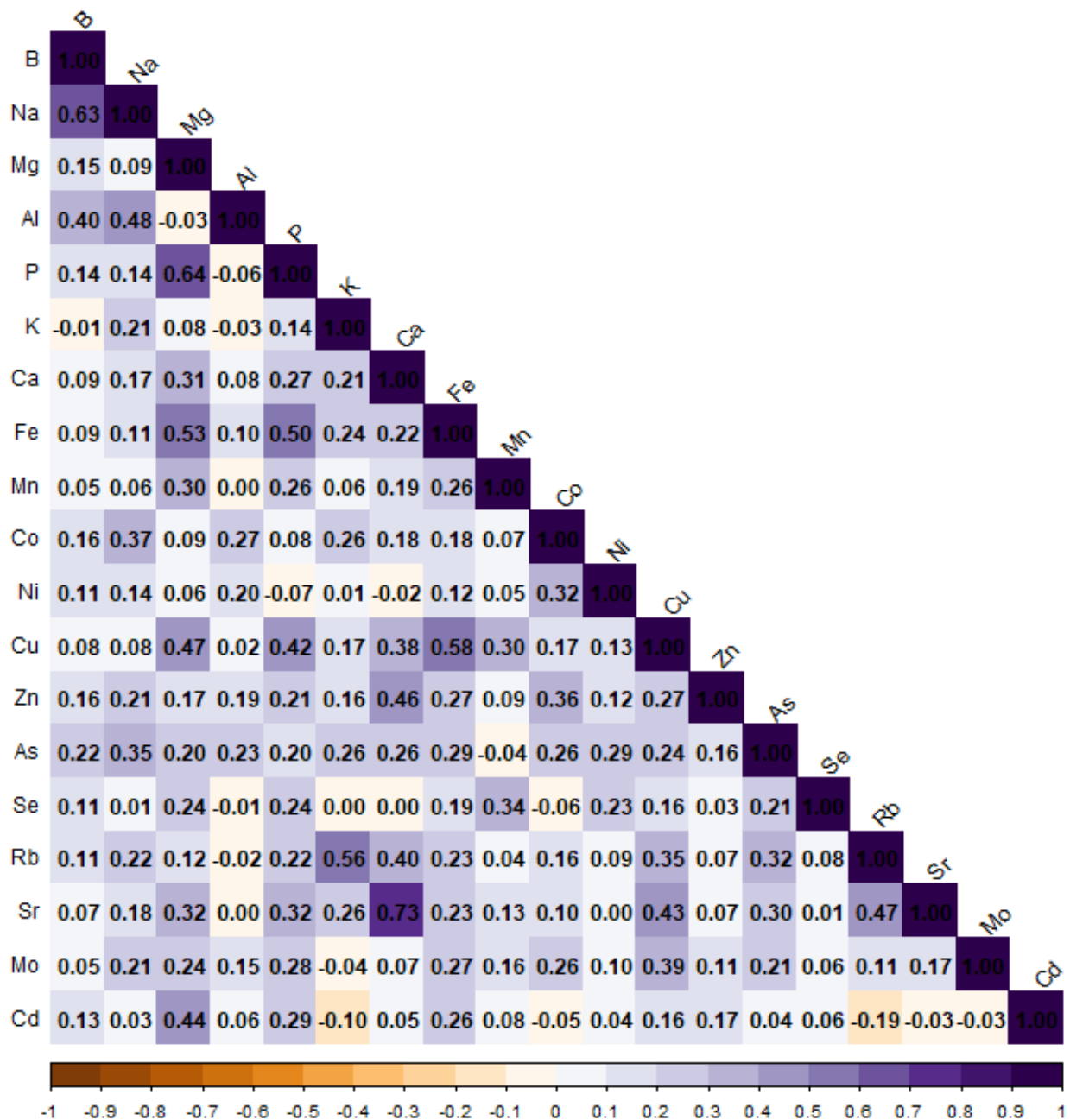
Appendix Figure C.1. Pearson correlations among BLUEs for ionomics traits measured in 21THD for TdGD Pop80 WILs

Values within each cell represent the correlation coefficient, with color indicating both magnitude (darker = stronger) and direction (orange = negative, purple = positive). Correlations were computed and visualized in R using the corplot package (Wei & Simko, 2024).



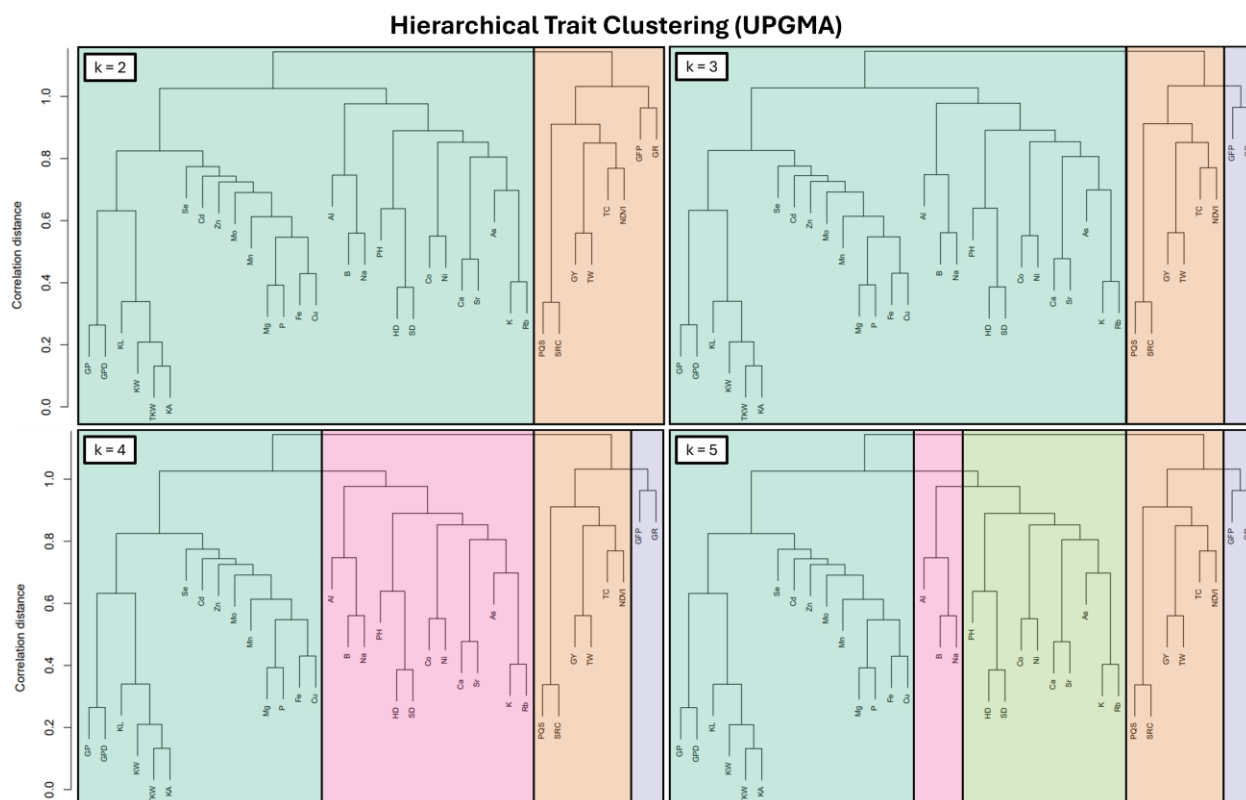
Appendix Figure C.2. Pearson correlations among BLUEs for ionomics traits measured in 22THD for TdGD Pop80 WILs

Values within each cell represent the correlation coefficient, with color indicating both magnitude (darker = stronger) and direction (orange = negative, purple = positive). Correlations were computed and visualized in R using the corrplot package (Wei & Simko, 2024).



Appendix Figure C.3. Pearson correlations among BLUEs for ionomics traits measured in 23THD for TdGD Pop80 WILs

Values within each cell represent the correlation coefficient, with color indicating both magnitude (darker = stronger) and direction (orange = negative, purple = positive). Correlations were computed and visualized in R using the corplot package (Wei & Simko, 2024).



Appendix Figure C.4. Hierarchical clustering of TdGD Pop80 trait BLUPs

Dendrograms show hierarchical clustering of across-environment BLUPs for the full set of traits evaluated in TdGD Pop80 using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and correlation distance (calculated as $1 - r$, where r is the Pearson correlation coefficient). The y-axis represents the correlation distance at which clusters merge, where smaller distances indicate strong positive correlations among traits, and where larger distances represent weak or negative correlations. The same dendrogram structure is shown under four cluster partition scenarios, with shaded boxes indicating trait assignments when the tree is cut at $k = 2, 3, 4$, or 5 clusters. Traits included in the clustering analysis were (1) agronomic performance: heading date (HD), senescence date (SD), grain fill period (GFP), plant height (PH); (2) yield and kernel components: tiller count (TC), normalized difference vegetation index (NDVI), green reflectance (GR), grain yield (GY), thousand kernel weight (TKW), test weight (TW), kernel area (KA), kernel length (KL), kernel width (KW); (3) functional quality: grain protein (GP), grain protein deviation (GPD), protein quality score (PQS), solvent retention capacity (SRC); and (4) grain ionic panel: B, Na, Mg, Al, P, K, Ca, Fe, Mn, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo, and Cd.

Appendix D - Supplementary tables for Chapter 5

Appendix Table D.1. Agronomic performance BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks

BLUPs were obtained from a two-stage mixed model analysis of standardized data for heading date (HD), senescence date (SD), grain fill period (GFP), and plant height (PH), and therefore represent performance relative to the population trait means. Low and high trial means indicate the range of environment-specific means in their original measurement units. Repeatability represents the proportion of genotypic variance relative to total variance.

Genotype		HD	SD	GFP	PH
Billings		-0.200	0.523	0.624	-0.264
Joe		0.365	0.830	0.485	0.687
KS061278M-4		0.264	0.107	-0.151	-0.194
KSTdGD80-175		-0.731	-0.069	0.605	-0.774
KSTdGD80-55		-0.527	-0.114	0.274	-0.986
Showdown		-0.351	0.210	0.586	0.361
SY Monument		1.627	0.723	-0.838	0.187
Repeatability		0.717	0.644	0.395	0.770
Trial Mean	Units	Julian days	Julian days	days	cm
	Low	128.904	161.427	32.000	68.260
	High	140.773	178.496	36.565	96.059

Appendix Table D.2. Grain yield BLUEs and BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks

BLUEs and BLUPs were obtained from a two-stage mixed model analysis of standardized data for grain yield (GY), and therefore represent performance relative to the trial means. BLUEs are shown by environment. Trial means are included in their original measurement units. Repeatability represents the proportion of genotypic variance relative to total variance.

Genotype	21THD	22EL	22RL	22THD	23RL	23THD	BLUP
Billings	0.526	-0.371	0.494	0.571	-0.632	0.559	0.200
Joe	0.726	0.737	0.602	0.844	0.935	1.005	0.801
KS061278M-4	0.625	-0.066	0.828	0.396	0.194	0.565	0.425
KSTdGD80-175	1.176	1.247	1.257	1.123	1.306	1.961	1.312
KSTdGD80-55	1.439	0.155	1.148	1.239	0.872	1.572	1.055
Showdown	1.464	1.186	1.041	0.853	0.738	1.217	1.079
SY Monument	0.630	0.841	0.993	0.209	-	-	0.705
Repeatability	0.900	0.740	0.720	0.870	0.770	0.700	0.733
Trial Mean (kg ha⁻¹)	5223.242	3633.751	3485.628	5265.467	3847.667	5016.573	

Appendix Table D.3. Yield and kernel component BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks

BLUPs were obtained from a two-stage mixed model analysis of standardized data for tiller count (TC), normalized difference vegetation index (NDVI), green reflectance (GR), thousand kernel weight (TKW), test weight (TW), kernel area, length, and width (KA, KL, KW), and therefore represent performance relative to the population trait means. Low and high trial means indicate the range of environment-specific means in their original measurement units. Due to differences in sensors and measurement platforms between environments, only standardized means are reported for NDVI and GR. Repeatability represents the proportion of genotypic variance relative to total variance.

Genotype		TC	NDVI	GR	TKW	TW	KA	KL	KW
Billings		-0.480	-0.133	0.154	1.874	0.694	1.822	1.964	1.004
Joe		0.797	0.510	-0.691	0.557	0.139	0.498	0.174	0.653
KS061278M-4		-0.068	0.171	-0.022	-0.141	0.330	-0.151	-0.247	-0.002
KSTdGD80-175		0.054	-0.311	-0.107	-0.943	1.453	-0.993	-0.596	-1.069
KSTdGD80-55		0.191	-0.345	0.086	-0.493	1.117	-0.701	-0.478	-0.570
Showdown		-0.006	0.167	-0.734	0.507	-0.224	0.580	0.750	0.150
SY Monument		0.612	0.455	0.354	0.197	-0.463	0.166	-0.274	0.520
Repeatability		0.374	0.429	0.652	0.836	0.760	0.907	0.953	0.818
Trial Mean	Units	count	-	-	g	kg hL ⁻¹	mm ²	mm	mm
	Low	56.928	0.059	-0.015	25.622	72.335	12.312	5.708	2.839
	High	75.101	0.085	0.033	31.576	75.136	13.698	5.961	3.104

Appendix Table D.4. Functional quality BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks

BLUPs were obtained from a two-stage mixed model analysis of standardized data for grain protein (GP), grain protein deviation (GPD), protein quality score (PQS) and solvent retention capacity (SRC), and therefore represent performance relative to the population trait means. Low and high trial means indicate the range of environment-specific means in their original measurement units. Repeatability represents the proportion of genotypic variance relative to total variance.

Genotype		GP	GPD	PQS	SRC
Billings		-0.680	-0.648	1.706	1.281
Joe		-1.358	-1.065	-0.244	-0.912
KS061278M-4		-0.530	-0.322	0.082	-0.202
KSTdGD80-175		-0.993	-0.284	1.056	0.484
KSTdGD80-55		-0.207	0.278	0.391	0.667
Showdown		-0.990	-0.517	0.899	0.581
SY Monument		-1.135	-0.892	1.626	0.979
Repeatability		0.593	0.455	0.653	0.787
Trial Mean	Units	%	-	-	%
	Low	15.516	0.000	15.348	239.417
	High	17.079	0.000	17.474	287.608

Appendix Table D.5. Macronutrient BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks

BLUPs were obtained from a two-stage mixed model analysis of standardized data for sodium (Na), magnesium (Mg), phosphorus (P), potassium (K), and calcium (Ca), and therefore represent performance relative to the population trait means. Low and high trial means indicate the range of environment-specific means in their original measurement units. Repeatability represents the proportion of genotypic variance relative to total variance.

Genotype		Na	Mg	P	K	Ca
Billings		0.195	-0.421	-0.167	1.035	-0.272
Joe		0.213	-0.841	-0.392	0.844	-0.124
KS061278M-4		-0.204	-0.203	0.029	0.163	0.409
KSTdGD80-175		0.429	0.157	0.461	0.339	0.193
KSTdGD80-55		0.369	0.246	0.142	-0.143	0.311
Showdown		0.541	-0.572	-0.040	0.988	0.176
SY Monument		-0.206	-0.928	-0.428	0.787	-0.073
Repeatability		0.116	0.588	0.161	0.392	0.286
Trial Mean (mg kg ⁻¹)	Low	3.081	9805.044	4563.160	5865.703	817.097
	High	3.723	10218.145	5001.316	6851.096	889.823

Appendix Table D.6. Micronutrient and trace element BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks

BLUPs were obtained from a two-stage mixed model analysis of standardized data for boron (B), iron (Fe), manganese (Mn), cobalt (Co), nickel (Ni), copper (Cu), zinc (Zn), selenium (Se), rubidium (Rb), strontium (Sr), and molybdenum (Mo), and therefore represent performance relative to the population trait means. Low and high trial means indicate the range of environment-specific means in their original measurement units. Repeatability represents the proportion of genotypic variance relative to total variance.

Genotype		B	Fe	Mn	Co	Ni	Cu	Zn	Se	Rb	Sr	Mo
Billings		-0.001	-0.096	-1.303	-0.081	-0.164	-0.081	-0.283	-0.314	0.795	0.410	-0.650
Joe		0.001	-0.300	-1.072	0.328	0.305	-1.107	-0.104	-0.237	1.070	-0.049	-0.775
KS061278M-4		-0.383	-0.354	-0.528	-0.285	-0.646	-0.496	-0.385	-0.151	0.200	0.548	0.008
KSTdGD80-175		0.168	-0.247	0.426	0.673	-0.316	-0.388	0.162	0.221	0.058	0.181	-0.205
KSTdGD80-55		-0.026	0.333	0.368	0.142	-0.361	0.199	0.197	0.056	0.487	0.280	0.831
Showdown		0.131	-0.395	-0.676	0.240	-0.620	-0.561	0.202	-0.002	1.013	0.342	-1.146
SY Monument		-0.014	-0.263	-0.505	0.513	0.949	-0.586	-0.208	0.089	0.569	-0.095	-1.059
Repeatability		0.113	0.605	0.466	0.590	0.376	0.552	0.202	0.139	0.417	0.264	0.647
Trial Mean (mg kg ⁻¹)	Low	2.193	49.080	98.730	0.006	0.541	7.725	24.212	0.710	1.758	5.016	1.400
	High	2.858	55.668	109.913	0.010	1.149	9.549	34.067	1.003	1.867	5.545	2.936

Appendix Table D.7. Toxic element BLUPs for KSTdGD80-55, KSTdGD80-175, the recurrent parent, and performance checks

BLUPs were obtained from a two-stage mixed model analysis of standardized data for aluminum (Al), arsenic (As), and cadmium (Cd), and therefore represent performance relative to the population trait means. Low and high trial means indicate the range of environment-specific means in their original measurement units. Repeatability represents the proportion of genotypic variance relative to total variance.

Genotype	Al	As	Cd
Billings	-0.729	0.450	-0.639
Joe	-0.380	0.462	-1.011
KS061278M-4	-0.420	0.082	-0.349
KSTdGD80-175	-0.143	0.000	-0.401
KSTdGD80-55	0.383	0.378	-0.152
Showdown	0.124	1.517	0.584
SY Monument	-0.158	0.515	-0.573
Repeatability	0.197	0.595	0.436
Trial Mean	Low	2.807	0.015
(mg kg ⁻¹)	High	4.163	0.022
		0.006	0.013