# Genome-wide association study of Barley Yellow Dwarf Virus resistance in a Nested Association Mapping population of winter wheat using Skim Exome Capture and Practical Haplotype Graph imputation

by

Enzo G. Unzain Moreno

B.S., Universidad Nacional de Asuncion, 2022

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Interdepartmental Genetics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Approved by:

Major Professor
Allan Fritz

# Copyright

© Enzo Unzain 2025.

# Abstract

Barley Yellow Dwarf Virus (BYDV) is a major constraint to wheat (*Triticum aestivum* L.) production, causing severe yield losses due to stunted growth, chlorosis, and reduced kernel weight. While chemical and cultural control strategies offer limited effectiveness, genetic resistance remains the most sustainable approach. In this study, a genome-wide association study (GWAS) was conducted to identify loci associated with BYDV resistance in a Nested Association Mapping (NAM) population of winter wheat derived from 'Overland'. Phenotypic evaluations were conducted under natural BYDV infection, assessing disease severity, chlorophyll content, plant height, and flowering date. To characterize genetic variation, exome capture sequencing was performed on the parental lines, while skim exome capture sequencing coupled with Practical Haplotype Graph (PHG)-based imputation was used for high-density genotyping of the recombinant inbred lines (RILs). A GWAS approach was employed using Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) and Multi-Locus Mixed Model (MLMM). Significant quantitative trait loci (QTL) were detected on chromosomes 5B, 6A, 7B, and 7D, with BLINK identifying stronger associations for BYDV severity and MLMM capturing additional loci for flowering date. A subset of significant markers was tested for validation through Kompetitive Allele-Specific PCR (KASP) assays. Despite strong marker-trait associations in the discovery population, validation in an independent field trial did not replicate all associations, highlighting the complexity of trait expression and the importance of conducting evaluations across diverse conditions and seasons to ensure robustness of results. These findings provide novel insights into the genetic architecture of BYDV resistance in winter wheat and highlight the value of integrating exome capture sequencing and PHG-based imputation for association mapping in polyploid crops.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgements

I would like to express my deepest gratitude to my committee members for their guidance and support throughout this project. I am especially thankful to Dr. Mary Guttieri for providing the NAM population and for her continuous availability in helping me navigate the complexities of my experimental design and marker analysis. I also sincerely appreciate Dr. Katherine Jordan for her invaluable assistance in constructing the Practical Haplotype Graph (PHG) and for the countless discussions that helped me interpret my genotypic results.

I am incredibly grateful to the KSU Wheat Breeding group for their collaboration, support, and friendship over the past few years. A special thanks to Andy Auld, Shaun Winnie, and Yemane Belaineh for making this journey both productive and entertaining. I would also like to acknowledge Angie Matthews, whose expertise and dedication in the lab were instrumental in processing my DNA samples and running marker analyses. Additionally, Dr. Augusto Tessele, Lucas Batista and Addy Carroll have been tremendously helpful in answering questions, reviewing my work, and encouraging me to think critically about all aspects of my research.

A heartfelt thank you to Dr. Hongliang Wang for his assistance in processing my DNA samples and preparing my sequencing libraries.

I am deeply appreciative of the unwavering support from my family and friends, who have encouraged me throughout this journey. A special thanks to my wife, Cielo, for her patience, understanding, and constant support during my graduate studies.

Finally, I extend my sincere gratitude to Dr. Allan Fritz for the knowledge and mentorship he has provided during my time in his program. I am incredibly thankful for the opportunity to be part of such a successful research group and for his dedication in fostering my growth as a plant breeder.

# Chapter 1 - Literature Review

## Introduction

Wheat (*Triticum aestivum* L.) is a staple cereal crop, but its production is threatened by biotic and abiotic stresses, including Barley Yellow Dwarf Virus (BYDV). Transmitted by aphids, BYDV causes stunted growth, chlorosis, and significant yield losses (Aradottir & Crespo-Herrera, 2021; Bockus et al., 2016). While traditional management strategies such as insecticide use and adjusted planting dates have limited long-term effectiveness, breeding for genetic resistance remains the most sustainable solution (Walls et al., 2019).

Resistance to BYDV is largely polygenic, with only a few resistance loci identified (Ayala, Van Ginkel, et al., 2001; Ordon et al., 2009; Zhang et al., 2009). Genome-wide association studies (GWAS) have proven valuable for mapping quantitative trait loci (QTL) associated with disease resistance (Alseekh et al., 2021; Tibbs Cortes et al., 2021) and Nested Association Mapping (NAM) populations enhance QTL discovery by combining the strengths of linkage and association mapping (Bajgain et al., 2016; Gireesh et al., 2021). However, genotyping accuracy in large populations remains a challenge, especially at low sequencing depths. Practical Haplotype Graph (PHG)-based imputation addresses this by leveraging parental haplotypes to improve genotype calls, maximizing data utility while minimizing costs (Bradbury et al., 2022; Jensen et al., 2020; Jordan et al., 2022).

This chapter provides a comprehensive review of the literature relevant to this study. It begins with an overview of wheat biology and breeding strategies, followed by a discussion of BYDV and its impact on wheat production. Next, the principles of GWAS and NAM populations in disease resistance studies are explored. Finally, genotyping platforms and imputation

techniques, particularly PHG, are discussed in the context of their applications in wheat genomics and breeding.

## Wheat: biology, genetics and importance

### Biology and domestication of wheat

Wheat (*Triticum aestivum* L.) traces its origins to the Tigris and Euphrates River Valley, in what is now modern-day Iraq, where it was domesticated. This domestication was a pivotal event in early agriculture, enabling the transition from hunter-gatherer societies to settled farming communities (Levy & Feldman, 2022).

The evolution of wheat began with *Triticum urartu* (AuAu genome), which diverged from a common ancestor with wild einkorn (*Triticum monococcum boeoticum*, AmAm genome) approximately 10 million years ago. *T. urartu* later hybridized with an extinct species related to *Aegilops speltoides* (SS genome) around 600,000 to 800,000 years ago, forming wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides, AABB* genome) (Jordan et al., 2015; L.-F. Li et al., 2022). Domesticated emmer (*T. dicoccum*) originated from wild emmer populations in southeastern Turkey (Oliveira et al., 2020), with selection favoring traits such as non-shattering spikes and free-threshing grains. Approximately 10,000 years ago, a subsequent hybridization event between domesticated emmer and *Aegilops tauschii* (DD genome) led to the formation of hexaploid wheat (*Triticum aestivum, AABBDD* genome) (Ahmed et al., 2023; Avni et al., 2017).

The addition of the D genome in hexaploid wheat significantly enhanced gluten composition, improving dough elasticity and making it more suitable for leavened bread production, while tetraploid emmer remained an important source for pasta (Avni et al., 2017; Zeibig et al., 2022). This increased functional diversity, combined with the adaptability of

hexaploid wheat, facilitated its spread across diverse environments and its establishment as a globally significant staple crop.

While wheat's domestication and subsequent global expansion significantly improved food security, modern wheat production faces increasing threats from biotic stresses such as diseases and pests. Among these, BYDV poses a major challenge to wheat yield stability, necessitating new breeding strategies to enhance resistance and maintain productivity (Mc Namara et al., 2020; Nancarrow et al., 2018).

**Importance of wheat in global agriculture**

Wheat remains a critical crop in global agriculture, providing around 20% of the calories and protein consumed by humans (Shiferaw et al., 2013). In 2023, global wheat production was approximately 791 million tonnes, with the United States contributing about 49.1 million tonnes (U.S. Department of Agriculture, 2024). As the global population is projected to surpass 10 billion by 2050, it is imperative to significantly increase wheat production to meet future food demands (Van Dijk et al., 2021).

Building on recent insights into wheat productivity, advances in breeding and agronomic practices have significantly contributed to increased attainable yields in key wheat-producing regions. In the U.S. Great Plains, hard red winter wheat yields have shown a steady upward trend, with attainable yields increasing at a higher rate than actual yields. Recent analysis indicates that attainable yield trends surpass actual yield gains by as much as 10–20 kg/ha per year, demonstrating the impact of technological advancements (Hatfield & Beres, 2019). This growing disparity highlights the urgent need for improved breeding and management strategies to bridge the gap between potential and actual yields. Addressing both biotic and abiotic stresses, such as disease pressure, drought, and extreme weather, is essential to enhancing yield stability

and ensuring sustainable wheat production in the face of increasing environmental challenges (Zhao et al., 2022).

BYDV is a significant biotic threat to global wheat production, causing severe stunting, leaf discoloration, and substantial yield losses (J. Singh et al., 2023). A seven-year study of winter wheat cultivars in Kansas by Gaunce and Bockus (2015) found that visual assessments of BYDV incidence accounted for approximately one-third of yield reduction on average. The most susceptible cultivars experienced yield losses ranging from 25% to 86%, although typical losses in susceptible varieties can reach 80%. Regression analysis of grain yield against BYDV severity showed a significant negative relationship, with BYDV severity explaining 19% to 48% of the yield variation per year and 29% across all years. The most susceptible cultivars suffered an average yield loss of 28.5%, while the moderately resistant cultivar ‘Everest’ experienced only 7.8% loss. This suggests that Everest reduced yield loss by approximately 73% relative to the most susceptible lines.

## Barley Yellow Dwarf

Barley yellow dwarf was first observed in barley in California in 1951 (Oswald & Houston, 1953) and was later identified in oat and wheat. The disease is caused by a single virus or a complex of closely related viruses, previously classified into two genera within the family *Luteoviridae*: *Barley yellow dwarf virus* (BYDV, genus Luteovirus) and *Cereal yellow dwarf virus* (CYDV, genus Polerovirus) (F. Liu et al., 2007). However, recent updates in virus taxonomy have reclassified BYDV into distinct species of the family *Tombusviridae*, now recognized as *Luteovirus pavhordei* (BYDV-PAV), *Luteovirus pashordei* (BYDV-PAS), and *Luteovirus mavhordei* (BYDV-MAV) with each species transmitted by a different aphid vector: *Rhopalosiphum padi* transmits BYDV-PAV, *Rhopalosiphum maidis* transmits BYDV-MAV, and

*Sitobion avenae* transmits BYDV-PAS (Miller et al., 2002; Walker et al., 2022). Throughout this document, we will refer to this complex as BYDV unless a more specific distinction is required.

BYDV is transmitted by various aphid species and spreads progressively within infected plants (Aradottir & Crespo-Herrera, 2021; Chain et al., 2005). Infection induces physiological, biochemical, and cytological changes, including disruptions in photosynthate transport, phloem degeneration, and the development of a nutritionally enhanced phloem diet that benefits aphid vectors (Choudhury, Larkin, Meinke, et al., 2019; Rong et al., 2018). Additionally, the virus leads to the formation of distinct cytological inclusions (Y. Liu et al., 2006).

Cereal crops infected with BYDV exhibit symptoms such as leaf discoloration, plant stunting, and a reduction in tillers, kernels per spike, kernel weight, and root growth. These effects contribute to significant yield losses, and in severe cases, infection can result in plant death. (Choudhury, Larkin, Meinke, et al., 2019; Riedell et al., 2003).

Although BYDV can be managed to some extent through agronomic practices, such as adjusting sowing dates or using insecticide seed treatments, the most effective and environmentally sustainable strategy for minimizing yield losses is breeding for resistant cultivars (Bockus et al., 2016; Mc Namara et al., 2020; Walls et al., 2019). The timing of infection plays a crucial role in determining the severity of yield loss. Hesler et al. (2005) found that early-planted winter wheat had a higher incidence of BYDV due to increased aphid infestations in the fall, whereas delaying planting until after September 20 significantly reduced aphid abundance and subsequent BYDV infection.

## Genetic basis of resistance to BYDV

Efforts to develop resistant wheat varieties have focused on identifying resistance loci, introgression of resistance genes from wild relatives, and utilizing molecular breeding tools such

as MAS and genomic prediction. Recent studies have shed light on various genetic mechanisms and breeding strategies that could enhance BYDV resistance in wheat and other species (Ordon et al., 2009; X. Wang et al., 2013).

Resistance refers to the plant's ability to limit or prevent virus infection, replication, or movement within its tissues, thereby reducing viral load and disease severity. Resistance can be mediated by specific genes that hinder virus replication or restrict aphid vectors' ability to transmit the virus. In contrast, tolerance refers to a plant's ability to sustain viral infection with minimal impact on growth, development, or yield (Cooper, 1983). Unlike resistant plants, tolerant plants allow virus replication but exhibit fewer disease symptoms and maintain productivity despite infection.

The only dominant gene reported for BYDV tolerance in wheat is *Bdv1*, originally identified in the cultivar 'Anza' (CI15284, R. P. Singh et al., 1993). Although *Bdv1* confers partial tolerance to BYDV, particularly the MAV strain, its effectiveness varies by serotype and environmental conditions. This limitation has spurred efforts to identify additional resistance sources, especially from more distantly related wheat relatives. No robust BYDV resistance has been detected within the primary or secondary gene pools of wheat; however, resistance genes have been discovered in over ten species of wild wheat relatives, including those from the genera *Thinopyrum*, *Agropyron*, *Elymus*, *Leymus*, *Roegneria*, and *Psathyrostachys* (Brettell et al., 1988; Sharma et al., 1984; Zhang et al., 2009). To date, four resistance genes have been cataloged, each exhibiting different specificities toward various BYDV species and likely operating through distinct mechanisms.

*Thinopyrum intermedium* has been widely utilized as a genetic resource for BYDV resistance in wheat breeding. Disomic addition lines carrying *T. intermedium* chromosomes have

been developed, conferring resistance to BYDV and rust diseases (Larkin et al., 1995). Various studies have characterized the introgression of BYDV resistance from *T. intermedium* into wheat, identifying key resistance loci and their expression in elite wheat backgrounds (Ayala, Van Ginkel, et al., 2001; L. Ayala-Navarrete et al., 2009; Larkin et al., 1995, 1995). *Bdv1* remains the first and most well-characterized gene, offering partial resistance against BYDV-MAV. Notably, *Bdv1* is pleiotropic with the rust resistance *Lr34*/*Yr18*, which confer adult plant resistance to leaf rust (*Puccinia graminis*) and yellow rust (*Puccinia striiformis*) (R. Singh, 1993). However, a recent study in Kansas wheat lines did not find a significant association between *Bdv1* presence and BYDV resistance (Tidakbi et al., 2025). The Italian wheat line 'Mentana' is the original source of the *Bdv1*/*Lr34*/*Yr18* complex, although *Lr34* is also found in CIMMYT's wheat germplasm (Chaves et al., 2020).

Resistance genes *Bdv2* and *Bdv3* were introgressed from *T. intermedium* into wheat lines such as 'Pontin' (L. I. Ayala-Navarrete et al., 2013; Zhang et al., 1999). This translocation lines exhibit strong resistance to the GAV and GPV serotypes of BYDV, which are more prevalent in China. These lines have been extensively used as important resistant parents in Chinese wheat breeding programs. In the United States, *Bdv2* is currently present in only two wheat cultivars, 'Uncharted' from Oklahoma State University and 'Washburn' from the University of Minnesota. *Bdv2* confers strong resistance to several BYDV strains likely by limiting virus replication or movement within the plant. Gene expression analyses in *Bdv2*-resistant lines have revealed the involvement of candidate genes encoding cytochrome P450, HSP70, EGF-CA, carbonic anhydrase, zeaxanthin epoxidase, protein kinases, *NBS-LRR*, and transducin (Ayala, Henry, et al., 2001; Zhang et al., 2004). In contrast, *Bdv3* provides complete resistance to CYDV-RPV and moderate resistance to BYDV-PAV and MAV, indicating potential differences in resistance

mechanisms, with *Bdv2* possibly involving a broader defense response and *Bdv3* exhibiting a more specific resistance spectrum (Aradottir & Crespo-Herrera, 2021). More recently, the resistance gene *Bdv4* was identified in a wheat-*T. intermedium* partial amphiploids located on the chromosome of group 2 (2D-2Ai-2), and it confers resistance to other BYDV strains prevalent in China (Lin et al., 2006).

A GWAS of 335 wheat accessions further uncovered four novel QTLs associated with BYDV resistance, with many of these QTLs derived from Chinese landraces (Choudhury, Larkin, Xu, et al., 2019). Given the partial effects of individual resistance genes, gene pyramiding has emerged as a promising strategy to enhance overall resistance. Research has demonstrated that combining *Bdv2* and *Bdv4* results in an additive effect, significantly increasing resistance levels (Jahier et al., 2009). Moreover, incorporating traits such as resistance to the aphid vectors of BYDV, as observed in *Triticum monococcum*, can help reduce virus transmission and contribute to integrated pest and disease management (Tanguy & Dedryver, 2009). In addition, a novel QTL, *QByd.hwwg-2DL*, was recently discovered on chromosome 2D of wheat cultivar ‘Jagger’, providing stable resistance to BYDV across multiple field experiments. This QTL also overlaps with key yield-related traits, suggesting a potential dual benefit for disease resistance and improved grain production (Bian et al., 2025).

Another GWAS conducted by Tidakbi et al. (2025) identified novel genetic associations with BYDV resistance in winter wheat. The study analyzed 269 winter wheat cultivars and breeding lines in a Kansas disease nursery, using exome and promoter capture sequencing to identify over 640,000 variants for association analysis. Ten significant genomic regions associated with BYDV resistance were discovered, including the previously known *Bdv2*, as well as novel minor-effect loci with resistance contributions ranging from 5% to 22%. Virus

population surveys using nanopore sequencing confirmed that co-infections of BYDV, Soil-borne Wheat Mosaic Virus (SBWMV), and wheat streak mosaic virus (WSMV) were common in Kansas.

Research in other cereal crops also provides valuable insights into BYDV resistance mechanisms. In barley (*Hordeum vulgare*), several resistance genes, including *Ryd2* and *Ryd3*, confer partial resistance by reducing virus replication and symptom severity (Riedel et al., 2011). Additionally, the recently characterized *Ryd4Hb* gene—introgressed from *Hordeum bulbosum*—offers strong resistance (Pidon et al., 2024). In maize (*Zea mays*), resistance to Barley Yellow Dwarf Virus-PAV (BYDV-PAV) has been linked to multiple QTLs across different chromosomes, and promising SNP markers have been identified for marker-assisted selection (Horn et al., 2014). Recent fine mapping efforts have refined a previously identified QTL for BYDV-PAV resistance to a ~0.3 Mbp region on chromosome 10, containing nine genes. Among them, two candidate genes, *Zm00001eb428010* and *Zm00001eb428020*, were strongly associated with resistance (Schmidt et al., 2024). These cross-species insights are invaluable and may inform wheat breeding strategies by revealing novel resistance mechanisms.

Significant progress has been made in understanding the genetic basis of BYDV resistance in wheat through the identification of key resistance loci, the introgression of genes from wild relatives, and the application of advanced molecular breeding techniques such as marker-assisted selection and genomic prediction. However, the polygenic nature of BYDV resistance underscores the need to stack multiple resistance genes for long-term, durable protection against the virus. To effectively identify and characterize these resistance loci, comprehensive genetic analysis is essential, with GWAS serving as a powerful tool for unraveling the genetic architecture of BYDV resistance.

## Genome wide association studies

GWAS is a powerful approach for identifying genetic variants associated with specific traits or diseases. In typical GWAS, both phenotype and genotype data are collected from a large and diverse population, such as a diversity panel. The genotype data primarily consists of genome-wide single nucleotide polymorphisms (SNPs), which are identified using methods like whole-genome resequencing, genotyping-by-sequencing, or array-based genotyping (Tibbs Cortes et al., 2021).

The goal of GWAS is to pinpoint genetic markers that show significant associations with the trait of interest. While some of these markers may be located within genes directly responsible for the trait, this is not always the case. Instead, GWAS leverages the concept of linkage disequilibrium (LD), a nonrandom association of alleles at different loci. Because loci that are physically close on a chromosome are less likely to be separated by recombination, SNPs near a causative mutation may remain in high LD with it, making them statistically associated with the trait (J. Wang et al., 2020).

By analyzing these associations across the genome, GWAS helps identify genomic regions that may harbor causative genes. In genetically diverse populations with deep evolutionary histories, LD blocks tend to be narrow, allowing for high-resolution mapping of functional polymorphisms. This makes GWAS a valuable tool for understanding the genetic architecture of complex traits and informing further functional studies (Alseekh et al., 2021).

Before the development of GWAS, linkage analysis was the primary approach for mapping QTLs. Unlike GWAS, which examines unrelated individuals in diverse populations, linkage mapping focuses on individuals with known relationships, such as $F_2$ populations or recombinant inbred lines (RILs) derived from biparental crosses (Korte & Farlow, 2013). In this

method, genetic markers closely linked to a QTL tend to co-segregate with the trait of interest due to limited recombination events within the studied population. Because of these large linkage blocks, fewer markers are needed compared to GWAS, making linkage analysis particularly effective before next-generation sequencing (NGS) technologies were widely available (Kumar et al., 2017).

The first genome-wide QTL mapping study in tomato (*Solanum lycopersicum*) was conducted in 1988 (Paterson et al., 1988), long before the first GWAS in 2002 (Ozaki et al., 2002). Despite technological advancements, both approaches remain complementary today. Linkage analysis is valuable for studying traits in structured populations with controlled crosses, while GWAS provides high-resolution mapping in genetically diverse populations. Together, these methods improve our ability to dissect complex traits and identify their genetic underpinnings (Korte & Farlow, 2013; Saini, 2022).

All association studies, including GWAS, rely on statistical methods to identify genetic markers associated with phenotypic variation. At its core, GWAS aims to detect SNPs where genotype variation significantly correlates with trait variation. A straightforward approach involves testing each SNP individually using methods such as ANOVA, where the null hypothesis assumes no difference in trait means across genotype groups (Alseekh et al., 2021; Tibbs Cortes et al., 2021). However, testing thousands or millions of SNPs simultaneously increases the likelihood of false positives, necessitating multiple testing corrections such as the False Discovery Rate or the Bonferroni correction (Storey & Tibshirani, 2003). Additionally, population structure and relatedness among individuals can lead to spurious associations if certain SNPs are overrepresented in specific subpopulations. To address this, mixed linear

models (MLM) incorporate both population structure and kinship matrices to control for confounding effects (Yu et al., 2006), improving the reliability of GWAS results.

To further refine GWAS models, multi-locus approaches have been developed, enhancing both power and computational efficiency. The Multi-Locus Mixed Model (MLMM) iteratively incorporates significant SNPs as fixed-effect covariates while adjusting variance components using stepwise regression to optimize model fit (Segura et al., 2012). Another powerful approach, Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK), builds on MLMM but replaces computationally intensive random-effect models with a fixed-effect framework, improving both speed and detection power by optimizing Bayesian Information Criterion selection (Huang et al., 2019). These multi-locus models are particularly useful for detecting complex traits controlled by multiple loci and have been widely adopted for high-resolution GWAS in diverse populations, with recent findings confirming that BLINK outperforms other models in controlling false positives and false negatives in a nested association mapping population of wheat (Chidzanga et al., 2022; Sandhu et al., 2024).

## Nested association mapping populations

Advancements in genetic mapping have led to the development of various generations of mapping populations aimed at improving the resolution and power of QTL detection. First-generation mapping populations primarily consist of biparental populations derived from two contrasting parents, such as $F_2$, RILs, and double haploid populations. These populations have been extensively used for linkage mapping of both qualitative and quantitative traits but are limited by low mapping resolution due to restricted recombination events and reduced allelic diversity, which may introduce ascertainment bias by underrepresenting rare alleles and genetic variation present in broader populations (Darvasi & Soller, 1995; Gireesh et al., 2021). Second-

generation approaches, such as association mapping, utilize diverse natural populations to improve mapping resolution by capturing historical recombination. However, these panels often have lower power in detecting minor-effect QTLs and rare alleles, and population structure can introduce spurious associations (Mackay & Powell, 2007; Xu et al., 2017).

To overcome these limitations, third-generation multiparent mapping populations integrate both linkage and association mapping principles by incorporating multiple diverse parents in structured breeding designs (Samantara et al., 2021; Scott et al., 2020). Applications of GWAS in NAM populations demonstrated its effectiveness in maize for detecting small-effect QTLs shared among families. This approach enables the identification of numerous loci contributing to complex traits, even in the absence of a single large-effect QTL. The ability of NAM-GWAS to capture subtle genetic effects makes it particularly valuable for dissecting polygenic traits, offering a robust framework for understanding genetic architecture across diverse populations (Buckler et al., 2009; Ladejobi et al., 2016).

NAM populations are developed by crossing multiple genetically diverse founder parents with a common founder parent, generating independent RIL populations from each cross. Collectively, these individual RIL populations form the NAM population, with each line representing a mosaic of genetic segments from the donor and common founder parents (Yu et al., 2006). The NAM design integrates historical LD from diverse founder parents with recombination-derived LD generated during population development, thereby improving both mapping resolution and allele diversity (Mackay & Powell, 2007). The reshuffling of parental genomes leads to the segregation of loci, enriching the population with multiple alleles, including rare alleles that are often undetected in traditional biparental or association mapping approaches (Gage et al., 2020; Xavier et al., 2015).

Compared to conventional mapping methods, NAM provides an improved balance between power and resolution, enabling the detection of both common and rare genetic variants (Gireesh et al., 2021). The inclusion of multiple parental lines expands genetic diversity, making NAM particularly effective for dissecting complex traits such as disease resistance and stress adaptation in wheat (Sandhu et al., 2024).

Applications of NAM in wheat breeding have demonstrated its utility in identifying QTLs for disease resistance, a key target for improving crop resilience (Bajgain et al., 2016; Y. Ren et al., 2017). For instance, a multi-reference NAM population in wheat was utilized to identify QTLs associated with stay-green traits, which contribute to drought adaptation (Christopher et al., 2021). Additionally, NAM has been employed to map QTLs related to nitrogen-deficiency tolerance in wheat seedlings, underscoring its potential for enhancing nutrient-use efficiency in breeding programs (D. Ren et al., 2018).

The structured design of NAM has contributed to advancements in modern wheat breeding by improving QTL detection and facilitating the identification of beneficial genetic variation. By enhancing mapping resolution, increasing allele diversity, and expanding the genetic base, NAM provides a framework for studying complex traits. However, the successful implementation of NAM and other genetic mapping approaches depends on the efficiency and accuracy of genotyping methods. As high-throughput genotyping platforms continue to evolve, they play a crucial role in providing the dense marker coverage necessary for precise genetic analysis and trait discovery.

## Genotyping technologies

Genotyping technologies have advanced significantly, providing researchers with various tools to assess genetic variation in wheat, particularly in GWAS within NAM populations. These

technologies include SNP arrays, genotyping-by-sequencing (GBS), whole genome sequencing (WGS), and Kompetitive Allele-Specific PCR (KASP) markers, each with unique advantages and applications in wheat breeding and genetic analysis (Bevan et al., 2017; Elshire et al., 2011).

SNP arrays, such as the Illumina 90K iSelect (S. Wang et al., 2014) and Axiom Wheat Breeder's genotyping arrays (Allen et al., 2017), offer high accuracy, reproducibility, and easy interpretation, making them suitable for genomic selection, population structure analysis, and linkage disequilibrium studies. However, they are limited by ascertainment bias and a fixed set of markers that may not capture novel variants in diverse wheat populations. Additionally, they tend to be relatively expensive (Elbasyoni et al., 2018).

GBS provides a higher marker density but suffers from missing data due to low sequencing depth. WGS offers the most comprehensive approach, but its cost remains a significant limitation for large NAM populations, not only due to sequencing expenses but also the substantial computational resources required for data processing and storage. The highly repetitive nature of the wheat genome further increases computational complexity, making variant calling and genome assembly more challenging and resource-intensive (Burridge et al., 2024; Darrier et al., 2019; Scheben et al., 2017).

An alternative approach, skim exome capture (SEC), reduces sequencing costs while retaining valuable genotypic information. SEC targets low-copy genic regions, allowing for multiplexing of up to 1056 samples per library at 0.32× coverage, providing a balance between cost and resolution (H. Wang et al., 2023).

KASP markers provide a targeted genotyping approach that is also cost-effective, highly specific, and widely used for validating SNPs and conducting MAS (Semagn et al., 2014). KASP assays allow for the discrimination of allelic variants through fluorescence-based detection and

are particularly advantageous for genotyping large numbers of samples with a relatively low cost per assay (B. Kaur et al., 2020). KASP markers are widely used in wheat breeding programs for screening known trait-associated loci, enabling rapid selection of desirable genotypes for disease resistance, yield improvement, and quality traits (Makhoul et al., 2020).

### Imputation techniques

One of the critical challenges in sequencing-based genotyping is missing data, which necessitates the use of imputation techniques to infer unobserved genotypes. Imputation is essential for increasing marker density, enhancing GWAS power, and ensuring the integration of multiple genotyping platforms.

Beagle (Browning & Browning, 2007) is one of the most commonly used imputation tools, efficiently handling large datasets and improving genotype accuracy by leveraging haplotype structure and linkage disequilibrium. Beagle's imputation methods are particularly useful in sequencing-based techniques such as GBS, skim sequencing, and exome capture sequencing, where low sequencing depth lead to significant data gaps (Keeble-Gagnère et al., 2021).

Another powerful imputation technique, the PHG (Jordan et al., 2022), improves upon traditional imputation methods by utilizing a library of haplotypes instead of relying solely on linkage disequilibrium. PHG allows for more accurate genotype predictions, particularly in NAM populations where diverse genetic backgrounds complicate standard imputation approaches. Unlike Beagle, PHG is designed to leverage prior knowledge of haplotypes from reference panels, making it particularly effective in polyploid species like wheat (Bradbury et al., 2022).

Imputation plays a particularly crucial role in sequencing-based techniques such as exome capture sequencing and skim sequencing. Exome capture sequencing focuses on coding regions, providing a high-resolution view of functional variants (P. Kaur & Gaikwad, 2017). Skim sequencing offers a cost-effective method for genome-wide genotyping at reduced sequencing depths (H. Wang et al., 2023). Both methods rely on imputation to reconstruct missing genotypes, ensuring that genetic analyses maintain high accuracy and completeness (P. Kaur & Gaikwad, 2017).

The integration of genotyping technologies, KASP markers, and advanced imputation techniques has significantly improved wheat genomic research. KASP markers provide an efficient and precise method for marker-assisted selection, while imputation methods like Beagle and PHG enhance the utility of sequencing-based approaches by filling in missing data. These innovations collectively support the identification of trait-associated loci, genomic selection, and wheat breeding efforts aimed at improving yield, disease resistance, and climate adaptability (Burridge et al., 2022; Jordan et al., 2022; B. Kaur et al., 2020).

## Research gaps and future directions

Despite significant advancements in wheat genomics and breeding, several challenges remain in the identification and utilization of resistance loci for effective disease management. One major challenge lies in the identification of stable resistance loci for BYDV across diverse genetic backgrounds and environmental conditions. Many resistance loci identified through GWAS and NAM populations exhibit variable effects depending on environmental interactions, genetic background, and pathogen strain diversity (Bian et al., 2025; Choudhury, Larkin, Xu, et al., 2019). Additionally, the polygenic nature of resistance makes it more difficult to stack

multiple resistance genes successfully, highlighting the need for innovative strategies to refine gene selection and deployment in breeding programs.

A promising direction for overcoming these challenges is the integration of PHG with GWAS and NAM. PHG-based imputation enhances genotypic accuracy by leveraging a reference haplotype library, thereby improving the resolution of QTL mapping and association studies (Jordan et al., 2022). The combination of PHG with GWAS allows for more precise identification of trait-associated loci, while its integration with NAM enables better haplotype tracking across multiple founder lines. By refining genotype calls and improving missing data imputation, PHG-based approaches can increase the power of association mapping and provide a more comprehensive understanding of genetic resistance mechanisms. This integration represents a significant step forward in harnessing genomic data to enhance wheat resistance breeding (Bradbury et al., 2022).

Future research should focus on the practical applications of these advancements for the development of germplasm resistant to BYDV. The integration of PHG-imputed datasets into breeding strategies could enhance the accuracy of predicting resistant genotypes (Bian et al., 2025). Ultimately, the integration of advanced genomic tools with conventional breeding approaches will accelerate the development of wheat varieties with long-lasting resistance to BYDV and other biotic stressors, contributing to global food security and sustainable agriculture.

# References


Allen, A. M., Winfield, M. O., Burridge, A. J., Downie, R. C., Benbow, H. R., Barker, G. L. A., Wilkinson, P. A., Coghill, J., Waterfall, C., Davassi, A., Scopes, G., Pirani, A., Webster, T., Brew, F., Bloor, C., Griffiths, S., Bentley, A. R., Alda, M., Jack, P., … Edwards, K. J. (2017). Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat ( *Triticum aestivum* ). *Plant Biotechnology Journal*, *15*(3), 390–401. https://doi.org/10.1111/pbi.12635

Alseekh, S., Kostova, D., Bulut, M., & Fernie, A. R. (2021). Genome-wide association studies: Assessing trait characteristics in model and crop plants. *Cellular and Molecular Life Sciences*, *78*(15), 5743–5754. https://doi.org/10.1007/s00018-021-03868-w

Aradottir, G. I., & Crespo-Herrera, L. (2021). Host plant resistance in wheat to barley yellow dwarf viruses and their aphid vectors: A review. *Current Opinion in Insect Science*, *45*, 59–68. https://doi.org/10.1016/j.cois.2021.01.002

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

Ayala, L., Henry, M., González-de-León, D., Van Ginkel, M., Mujeeb-Kazi, A., Keller, B., & Khairallah, M. (2001). A diagnostic molecular marker allowing the study of Th. Intermedium-derived resistance to BYDV in bread wheat segregating populations: *Theoretical and Applied Genetics*, *102*(6–7), 942–949. https://doi.org/10.1007/s001220000476

Ayala, L., Van Ginkel, M., Khairallah, M., Keller, B., & Henry, M. (2001). Expression of *Thinopyrum intermedium* -Derived *Barley yellow dwarf virus* Resistance in Elite Bread Wheat Backgrounds. *Phytopathology®*, *91*(1), 55–62. https://doi.org/10.1094/PHYTO.2001.91.1.55

Ayala-Navarrete, L. I., Mechanicos, A. A., Gibson, J. M., Singh, D., Bariana, H. S., Fletcher, J., Shorter, S., & Larkin, P. J. (2013). The Pontin series of recombinant alien translocations in bread wheat: Single translocations integrating combinations of *Bdv2*, Lr19 and Sr25

disease-resistance genes from Thinopyrum intermedium and Th. ponticum. *Theoretical and Applied Genetics*, *126*(10), 2467–2475. https://doi.org/10.1007/s00122-013-2147-0

Ayala-Navarrete, L., Tourton, E., Mechanicos, A. A., & Larkin, P. J. (2009). Comparison of *Thinopyrum intermedium* derivatives carrying barley yellow dwarf virus resistance in wheat. *Genome*, *52*(6), 537–546. https://doi.org/10.1139/G09-028

Baenziger, P. S., Beecher, B., Graybosch, R. A., Ibrahim, A. M. H., Baltensperger, D. D., Nelson, L. A., Jin, Y., Wegulo, S. N., Watkins, J. E., Hatchett, J. H., Chen, M.-S., & Bai, G. (2008). Registration of 'NE01643' Wheat. *Journal of Plant Registrations*, *2*(1), 36–42. https://doi.org/10.3198/jpr2007.06.0327crc

Bajgain, P., Rouse, M. N., Tsilo, T. J., Macharia, G. K., Bhavani, S., Jin, Y., & Anderson, J. A. (2016). Nested Association Mapping of Stem Rust Resistance in Wheat Using Genotyping by Sequencing. *PLOS ONE*, *11*(5), e0155760. https://doi.org/10.1371/journal.pone.0155760

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

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

Bevan, M. W., Uauy, C., Wulff, B. B. H., Zhou, J., Krasileva, K., & Clark, M. D. (2017). Genomic innovation for crop improvement. *Nature*, *543*(7645), 346–354. https://doi.org/10.1038/nature22011

Bian, R., Liu, N., Xu, Y., Su, Z., Chai, L., Bernardo, A., St. Amand, P., Rupp, J., Pumphrey, M., Fritz, A., Zhang, G., Jordan, K. W., & Bai, G. (2025). A novel quantitative trait locus for *barley yellow dwarf virus* resistance and kernel traits on chromosome 2D of a wheat cultivar Jagger. *The Plant Genome*, *18*(1), e20548. https://doi.org/10.1002/tpg2.20548

Bockus, W. W., De Wolf, E. D., & Todd, T. C. (2016). Management Strategies for Barley Yellow Dwarf on Winter Wheat in Kansas. *Plant Health Progress*, *17*(2), 122–127. https://doi.org/10.1094/PHP-RS-15-0050

Bradbury, P. J., Casstevens, T., Jensen, S. E., Johnson, L. C., Miller, Z. R., Monier, B., Romay, M. C., Song, B., & Buckler, E. S. (2022). The Practical Haplotype Graph, a platform for

storing and using pangenomes for imputation. *Bioinformatics*, *38*(15), 3698–3702. https://doi.org/10.1093/bioinformatics/btac410

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brettell, R. I. S., Banks, P. M., Cauderon, Y., Chen, X., Cheng, Z. M., Larkin, P. J., & Waterhouse, P. M. (1988). A single wheatgrass chromosome reduces the concentration of barley yellow dwarf virus in wheat. *Annals of Applied Biology*, *113*(3), 599–603. https://doi.org/10.1111/j.1744-7348.1988.tb03337.x

Browning, S. R., & Browning, B. L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics*, *81*(5), 1084–1097. https://doi.org/10.1086/521987

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., Ersoz, E., Flint-Garcia, S., Garcia, A., Glaubitz, J. C., Goodman, M. M., Harjes, C., Guill, K., Kroon, D. E., Larsson, S., Lepak, N. K., Li, H., Mitchell, S. E., Pressoir, G., … McMullen, M. D. (2009). The Genetic Architecture of Maize Flowering Time. *Science*, *325*(5941), 714–718. https://doi.org/10.1126/science.1174276

Burridge, A. J., Winfield, M. O., Wilkinson, P. A., Przewieslik-Allen, A. M., Edwards, K. J., & Barker, G. L. A. (2022). The Use and Limitations of Exome Capture to Detect Novel Variation in the Hexaploid Wheat Genome. *Frontiers in Plant Science*, *13*, 841855. https://doi.org/10.3389/fpls.2022.841855

Burridge, A. J., Winfield, M., Przewieslik-Allen, A., Edwards, K. J., Siddique, I., Barral-Arca, R., Griffiths, S., Cheng, S., Huang, Z., Feng, C., Dreisigacker, S., Bentley, A. R., Brown-Guedira, G., & Barker, G. L. (2024). Development of a next generation SNP genotyping array for wheat. *Plant Biotechnology Journal*, *22*(8), 2235–2247. https://doi.org/10.1111/pbi.14341

Bushnell, B. (2014). *BBMap: A fast, accurate, splice-aware aligner*.

Chain, F., Riault, G., Trottet, M., & Jacquot, E. (2005). Analysis of Accumulation Patterns of Barley yellow dwarf virus-PAV (BYDV-PAV) in Two Resistant Wheat Lines. *European Journal of Plant Pathology*, *113*(4), 343–355. https://doi.org/10.1007/s10658-005-7966-7

Chaves, M. S., Silva, G. B. P. D., Caierão, E., Federizzi, L. C., & Martinelli, J. A. (2020). A century of wheat breeding in Brazil: The origin and inheritance of the *Lr34* locus in wheat varieties released from 1922 to 2016. *Crop Breeding and Applied Biotechnology*, *20*(2), e27952027. https://doi.org/10.1590/1984-70332020v20n2a23

Chidzanga, C., Mullan, D., Roy, S., Baumann, U., & Garcia, M. (2022). Nested association mapping-based GWAS for grain yield and related traits in wheat grown under diverse Australian environments. *Theoretical and Applied Genetics*, *135*(12), 4437–4456. https://doi.org/10.1007/s00122-022-04230-9

Choudhury, S., Larkin, P., Meinke, H., Hasanuzzaman, M. D., Johnson, P., & Zhou, M. (2019). Barley yellow dwarf virus infection affects physiology, morphology, grain yield and flour pasting properties of wheat. *Crop and Pasture Science*, *70*(1), 16. https://doi.org/10.1071/CP18364

Choudhury, S., Larkin, P., Xu, R., Hayden, M., Forrest, K., Meinke, H., Hu, H., Zhou, M., & Fan, Y. (2019). Genome wide association study reveals novel QTL for barley yellow dwarf virus resistance in wheat. *BMC Genomics*, *20*(1), 891. https://doi.org/10.1186/s12864-019-6249-1

Christopher, M., Paccapelo, V., Kelly, A., Macdonald, B., Hickey, L., Richard, C., Verbyla, A., Chenu, K., Borrell, A., Amin, A., & Christopher, J. (2021). QTL identified for stay-green in a multi-reference nested association mapping population of wheat exhibit context dependent expression and parent-specific alleles. *Field Crops Research*, *270*, 108181. https://doi.org/10.1016/j.fcr.2021.108181

Cooper, J. I. (1983). Responses of Plants to Viruses: Proposals for the Use of Terms. *Phytopathology*, *73*(2), 127. https://doi.org/10.1094/Phyto-73-127

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

Darrier, B., Russell, J., Milner, S. G., Hedley, P. E., Shaw, P. D., Macaulay, M., Ramsay, L. D., Halpin, C., Mascher, M., Fleury, D. L., Langridge, P., Stein, N., & Waugh, R. (2019). A Comparison of Mainstream Genotyping Platforms for the Evaluation and Use of Barley Genetic Resources. *Frontiers in Plant Science*, *10*, 544. https://doi.org/10.3389/fpls.2019.00544

Darvasi, A., & Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics*, *141*(3), 1199–1207. https://doi.org/10.1093/genetics/141.3.1199

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

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Emrich, K., Wilde, F., Miedaner, T., & Piepho, H. P. (2008). REML approach for adjusting the Fusarium head blight rating to a phenological date in inoculated selection experiments of wheat. *Theoretical and Applied Genetics*, *117*(1), 65–73. https://doi.org/10.1007/s00122-008-0753-z

Franco, J. A. V., Gage, J. L., Bradbury, P. J., Johnson, L. C., Miller, Z. R., Buckler, E. S., & Romay, M. C. (2020). A Maize Practical Haplotype Graph Leverages Diverse NAM Assemblies. *bioRxiv*, 2020.08.31.268425. https://doi.org/10.1101/2020.08.31.268425

Gage, J. L., Monier, B., Giri, A., & Buckler, E. S. (2020). Ten Years of the Maize Nested Association Mapping Population: Impact, Limitations, and Future Directions. *The Plant Cell*, *32*(7), 2083–2093. https://doi.org/10.1105/tpc.19.00951

Gireesh, C., Sundaram, Raman. M., Anantha, S. M., Pandey, M. K., Madhav, M. S., Rathod, S., Yathish, K. R., Senguttuvel, P., Kalyani, B. M., Ranjith, E., Venkata Subbarao, L., Kumar Mondal, T., Swamy, M., & Rakshit, S. (2021). Nested Association Mapping (NAM) Populations: Present Status and Future Prospects in the Genomics Era. *Critical Reviews in Plant Sciences*, *40*(1), 49–67. https://doi.org/10.1080/07352689.2021.1880019

Hatfield, J. L., & Beres, B. L. (2019). Yield Gaps in Wheat: Path to Enhancing Productivity. *Frontiers in Plant Science*, *10*, 1603. https://doi.org/10.3389/fpls.2019.01603

Hesler, L. S., Riedell, W. E., Langham, M. A. C., & Osborne, S. L. (2005). Insect Infestations, Incidence of Viral Plant Diseases, and Yield of Winter Wheat in Relation to Planting

Date in the Northern Great Plains. *Journal of Economic Entomology*, *98*(6), 2020–2027. https://doi.org/10.1093/jee/98.6.2020

Horn, F., Habekuß, A., & Stich, B. (2014). Genes involved in barley yellow dwarf virus resistance of maize. *Theoretical and Applied Genetics*, *127*(12), 2575–2584. https://doi.org/10.1007/s00122-014-2400-1

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), giy154. https://doi.org/10.1093/gigascience/giy154

Jahier, J., Chain, F., Barloy, D., Tanguy, A. -M., Lemoine, J., Riault, G., Margalé, E., Trottet, M., & Jacquot, E. (2009). Effect of combining two genes for partial resistance to *Barley yellow dwarf virus*- PAV (BYDV-PAV) derived from *Thinopyrum intermedium* in wheat. *Plant Pathology*, *58*(5), 807–814. https://doi.org/10.1111/j.1365-3059.2009.02084.x

Jensen, S. E., Charles, J. R., Muleta, K., Bradbury, P. J., Casstevens, T., Deshpande, S. P., Gore, M. A., Gupta, R., Ilut, D. C., Johnson, L., Lozano, R., Miller, Z., Ramu, P., Rathore, A., Romay, M. C., Upadhyaya, H. D., Varshney, R. K., Morris, G. P., Pressoir, G., … Ramstein, G. P. (2020). A sorghum practical haplotype graph facilitates genome-wide imputation and cost-effective genomic prediction. *The Plant Genome*, *13*(1), e20009. https://doi.org/10.1002/tpg2.20009

Jordan, K. W., Bradbury, P. J., Miller, Z. R., Nyine, M., He, F., Fraser, M., Anderson, J., Mason, E., Katz, A., Pearce, S., Carter, A. H., Prather, S., Pumphrey, M., Chen, J., Cook, J., Liu, S., Rudd, J. C., Wang, Z., Chu, C., … Akhunov, E. D. (2022). Development of the Wheat Practical Haplotype Graph database as a resource for genotyping data storage and genotype imputation. *G3 Genes|Genomes|Genetics*, *12*(2), jkab390. https://doi.org/10.1093/g3journal/jkab390

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*(1), 48. https://doi.org/10.1186/s13059-015-0606-4

Kaur, B., Mavi, G. S., Gill, M. S., & Saini, D. K. (2020). Utilization of KASP technology for wheat improvement. *Cereal Research Communications*, *48*(4), 409–421. https://doi.org/10.1007/s42976-020-00057-6

Kaur, P., & Gaikwad, K. (2017). From Genomes to GENE-omes: Exome Sequencing Concept and Applications in Crop Improvement. *Frontiers in Plant Science*, *8*, 2164. https://doi.org/10.3389/fpls.2017.02164

Keeble-Gagnère, G., Pasam, R., Forrest, K. L., Wong, D., Robinson, H., Godoy, J., Rattey, A., Moody, D., Mullan, D., Walmsley, T., Daetwyler, H. D., Tibbits, 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

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, *9*(1), 29. https://doi.org/10.1186/1746-4811-9-29

Kumar, J., Gupta, D. S., Gupta, S., Dubey, S., Gupta, P., & Kumar, S. (2017). Quantitative trait loci from identification to exploitation for crop improvement. *Plant Cell Reports*, *36*(8), 1187–1213. https://doi.org/10.1007/s00299-017-2127-y

Ladejobi, O., Elderfield, J., Gardner, K. A., Gaynor, R. C., Hickey, J., Hibberd, J. M., Mackay, I. J., & Bentley, A. R. (2016). Maximizing the potential of multi-parental crop populations. *Applied & Translational Genomics*, *11*, 9–17. https://doi.org/10.1016/j.atg.2016.10.002

Larkin, P. J., Banks, P. M., Lagudah, E. S., Appels, R., Xiao, C., Zhiyong, X., Ohm, H. W., & McIntosh, R. A. (1995). Disomic *Thinopyrum intermedium* addition lines in wheat with barley yellow dwarf virus resistance and with rust resistances. *Genome*, *38*(2), 385–394. https://doi.org/10.1139/g95-050

Levy, A. A., & Feldman, M. (2022). Evolution and origin of bread wheat. *The Plant Cell*, *34*(7), 2549–2567. https://doi.org/10.1093/plcell/koac130

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome 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, 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

Lin, Z. S., Huang, D. H., Du, L. P., Ye, X. G., & Xin, Z. Y. (2006). Identification of wheat–*Thinopyrum intermedium* 2Ai-2 ditelosomic addition and substitution lines with resistance to barley yellow dwarf virus. *Plant Breeding*, *125*(2), 114–119. https://doi.org/10.1111/j.1439-0523.2006.01167.x

Liu, F., Wang, X., Liu, Y., Xie, J., Gray, S. M., Zhou, G., & Gao, B. (2007). A Chinese isolate of barley yellow dwarf virus-PAV represents a third distinct species within the PAV serotype. *Archives of Virology*, *152*(7), 1365–1373. https://doi.org/10.1007/s00705-007-0947-8

Liu, Y., Kang, Z., & Buchenauer, H. (2006). Ultrastructural and Immunocytochemical Studies on Effects of Barley Yellow Dwarf Virus – Infection on Fusarium Head Blight, Caused by Fusarium graminearum, in Wheat Plants. *Journal of Phytopathology*, *154*(1), 6–15. https://doi.org/10.1111/j.1439-0434.2005.01048.x

Mackay, I., & Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science*, *12*(2), 57–63. https://doi.org/10.1016/j.tplants.2006.12.001

Makhoul, M., Rambla, C., Voss-Fels, K. P., Hickey, L. T., Snowdon, R. J., & Obermeier, C. (2020). Overcoming polyploidy pitfalls: A user guide for effective SNP conversion into KASP markers in wheat. *Theoretical and Applied Genetics*, *133*(8), 2413–2430. https://doi.org/10.1007/s00122-020-03608-x

Mc Namara, L., Gauthier, K., Walsh, L., Thébaud, G., Gaffney, M., & Jacquot, E. (2020). Management of yellow dwarf disease in Europe in a post-neonicotinoid agriculture. *Pest Management Science*, *76*(7), 2276–2285. https://doi.org/10.1002/ps.5835

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis

Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Miller, W. A., Liu, S., & Beckett, R. (2002). Barley yellow dwarf virus: *Luteoviridae* or *Tombusviridae* ? *Molecular Plant Pathology*, *3*(4), 177–183. https://doi.org/10.1046/j.1364-3703.2002.00112.x

Nancarrow, N., Aftab, M., Freeman, A., Rodoni, B., Hollaway, G., & Trębicki, P. (2018). Prevalence and Incidence of Yellow Dwarf Viruses Across a Climatic Gradient: A Four-Year Field Study in Southeastern Australia. *Plant Disease*, *102*(12), 2465–2472. https://doi.org/10.1094/PDIS-01-18-0116-RE

Ordon, F., Habekuss, A., Kastirr, U., Rabenstein, F., & Kühne, T. (2009). Virus Resistance in Cereals: Sources of Resistance, Genetics and Breeding. *Journal of Phytopathology*, *157*(9), 535–545. https://doi.org/10.1111/j.1439-0434.2009.01540.x

Oswald, J., & Houston, B. (1953). Yellow dwarf disease: A new and damaging virus disease of cereals transmitted by aphids. *California Agriculture*, *7*(10), 3–4.

Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T., Sato, H., Sato, H., Hori, M., & Nakamura, Y. (2002). Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature Genetics*, *32*(4), 650–654.

Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E., & Tanksley, S. D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature*, *335*(6192), 721–726.

Pidon, H., Ruge-Wehling, B., Will, T., Habekuß, A., Wendler, N., Oldach, K., Maasberg-Prelle, A., Korzun, V., & Stein, N. (2024). High-resolution mapping of Ryd4Hb, a major resistance gene to Barley yellow dwarf virus from Hordeum bulbosum. *Theoretical and Applied Genetics*, *137*(3), 60. https://doi.org/10.1007/s00122-024-04542-y

Piepho, H.-P., & Möhring, J. (2007). Computing Heritability and Selection Response From Unbalanced Plant Breeding Trials. *Genetics*, *177*(3), 1881–1888. https://doi.org/10.1534/genetics.107.074229

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Ranabhat, N. B., Fellers, J. P., Bruce, M. A., & Rupp, J. L. S. (2025). High-Throughput Oxford Nanopore Sequencing Unveils Complex Viral Population in Kansas Wheat: Implications for Sustainable Virus Management. *Viruses*, *17*(1), 126. https://doi.org/10.3390/v17010126

Ren, D., Fang, X., Jiang, P., Zhang, G., Hu, J., Wang, X., Meng, Q., Cui, W., Lan, S., Ma, X., Wang, H., & Kong, L. (2018). Genetic Architecture of Nitrogen-Deficiency Tolerance in Wheat Seedlings Based on a Nested Association Mapping (NAM) Population. *Frontiers in Plant Science*, *9*, 845. https://doi.org/10.3389/fpls.2018.00845

Ren, Y., Hou, W., Lan, C., Basnet, B. R., Singh, R. P., Zhu, W., Cheng, X., Cui, D., & Chen, F. (2017). QTL Analysis and Nested Association Mapping for Adult Plant Resistance to Powdery Mildew in Two Bread Wheat Populations. *Frontiers in Plant Science*, *8*, 1212. https://doi.org/10.3389/fpls.2017.01212

Riedel, C., Habekuß, A., Schliephake, E., Niks, R., Broer, I., & Ordon, F. (2011). Pyramiding of Ryd2 and Ryd3 conferring tolerance to a German isolate of Barley yellow dwarf virus-PAV (BYDV-PAV-ASL-1) leads to quantitative resistance against this isolate. *Theoretical and Applied Genetics*, *123*(1), 69–76. https://doi.org/10.1007/s00122-011-1567-y

Riedell, W. E., Kieckhefer, R. W., Haley, S. D., Langham, M. A. C., & Evenson, P. D. (1999). Winter Wheat Responses to Bird Cherry-Oat Aphids and Barley Yellow Dwarf Virus Infection. *Crop Science*, *39*(1), 158–163. https://doi.org/10.2135/cropsci1999.0011183X003900010025x

Riedell, W. E., Kieckhefer, R. W., Langham, M. A. C., & Hesler, L. S. (2003). Root and Shoot Responses to Bird Cherry-Oat Aphids and *Barley yellow dwarf virus* in Spring Wheat. *Crop Science*, *43*(4), 1380–1386. https://doi.org/10.2135/cropsci2003.1380

Rong, W., Wang, X., Wang, X., Massart, S., & Zhang, Z. (2018). Molecular and Ultrastructural Mechanisms Underlying Yellow Dwarf Symptom Formation in Wheat after Infection of Barley Yellow Dwarf Virus. *International Journal of Molecular Sciences*, *19*(4), 1187. https://doi.org/10.3390/ijms19041187

Rosewarne, G., Bonnett, D., Rebetzke, G., Lonergan, P., & Larkin, P. (2015). The Potential of Lr19 and *Bdv2* Translocations to Improve Yield and Disease Resistance in the High

Rainfall Wheat Zones of Australia. *Agronomy*, *5*(1), 55–70. https://doi.org/10.3390/agronomy5010055

Saini, D. K. (2022). Comprehensive evaluation of mapping complex traits in wheat using genome-wide association studies. *Mol Breeding*.

Samantara, K., Reyes, V. P., Agrawal, N., Mohapatra, S. R., & Jena, K. K. (2021). Advances and trends on the utilization of multi-parent advanced generation intercross (MAGIC) for crop improvement. *Euphytica*, *217*(10), 189. https://doi.org/10.1007/s10681-021-02925-6

Sandhu, K. S., Burke, A. B., Merrick, L. F., Pumphrey, M. O., & Carter, A. H. (2024). Comparing performances of different statistical models and multiple threshold methods in a nested association mapping population of wheat. *Frontiers in Plant Science*, *15*, 1460353. https://doi.org/10.3389/fpls.2024.1460353

Saradadevi, R., Palta, J. A., & Siddique, K. H. M. (2017). ABA-Mediated Stomatal Response in Regulating Water Use during the Development of Terminal Drought in Wheat. *Frontiers in Plant Science*, *8*, 1251. https://doi.org/10.3389/fpls.2017.01251

Scheben, A., Batley, J., & Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: Choosing the right tool for the right application. *Plant Biotechnology Journal*, *15*(2), 149–161. https://doi.org/10.1111/pbi.12645

Schmidt, M., Guerreiro, R., Baig, N., Habekuß, A., Will, T., Ruckwied, B., & Stich, B. (2024). Fine mapping a QTL for BYDV-PAV resistance in maize. *Theoretical and Applied Genetics*, *137*(7), 163. https://doi.org/10.1007/s00122-024-04668-z

Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., Clark, M., Dell'Acqua, M., Dixon, L. E., Filippi, C. V., Fradgley, N., Gardner, K. A., Mackay, I. J., O'Sullivan, D., Percival-Alwyn, L., Roorkiwal, M., Singh, R. K., Thudi, M., Varshney, R. K., … Mott, R. (2020). Multi-parent populations in crops: A toolbox integrating genomics and genetic mapping with breeding. *Heredity*, *125*(6), 396–416. https://doi.org/10.1038/s41437-020-0336-6

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

Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology

and its application in crop improvement. *Molecular Breeding*, *33*(1), 1–14. https://doi.org/10.1007/s11032-013-9917-x

Sharma, H. C., Gill, B. S., & Uyemoto, J. K. (1984). High Levels of Resistance in *Agropyron* Species to Barley Yellow Dwarf and Wheat Streak Mosaic Viruses. *Journal of Phytopathology*, *110*(2), 143–147. https://doi.org/10.1111/j.1439-0434.1984.tb03402.x

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., & Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security*, *5*(3), 291–317. https://doi.org/10.1007/s12571-013-0263-y

Silva, P., Evers, B., Kieffaber, A., Wang, X., Brown, R., Gao, L., Fritz, A., Crain, J., & Poland, J. (2022). Applied phenomics and genomics for improving barley yellow dwarf resistance in winter wheat. *G3 Genes|Genomes|Genetics*, *12*(7), jkac064. https://doi.org/10.1093/g3journal/jkac064

Singh, J., Chhabra, B., Raza, A., Yang, S. H., & Sandhu, K. S. (2023). Important wheat diseases in the US and their management in the 21st century. *Frontiers in Plant Science*, *13*, 1010191. https://doi.org/10.3389/fpls.2022.1010191

Singh, R. (1993). Genetic association of gene *Bdv1* for tolerance to barley yellow dwarf virus with genes *Lr34* and *Yr18* for adult plant resistance to rusts in bread wheat. *Plant Disease*, *77*(11), 1103–1106. CABI Databases. https://doi.org/10.1094/PD-77-1103

Singh, R. P., Burnett, P. A., Albarrán, M., & Rajaram, S. (1993). *Bdv1*: A Gene for Tolerance to Barley Yellow Dwarf Virus in Bread Wheats. *Crop Science*, *33*(2), 231–234. https://doi.org/10.2135/cropsci1993.0011183X003300020002x

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, *100*(16), 9440–9445. https://doi.org/10.1073/pnas.1530509100

Tanguy, S., & Dedryver, C.-A. (2009). Reduced BYDV–PAV transmission by the grain aphid in a Triticum monococcum line. *European Journal of Plant Pathology*, *123*(3), 281–289. https://doi.org/10.1007/s10658-008-9365-3

Tibbs Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *The Plant Genome*, *14*(1), e20077. https://doi.org/10.1002/tpg2.20077

Tidakbi, L., Wang, H., Bian, R., Redila, C., Zhu, D., St. Amand, P., Bernardo, A., Bruce, M., Zhang, G., Fritz, A., Guttieri, M., Fellers, J., Bai, G., Noller, J. R., & Jordan, K. W. (2025). Genome-wide association study identifies novel associations with barley yellow dwarf virus and soil-borne wheat mosaic virus resistance in winter wheat association mapping panel. *Crop Science*, *65*(2), e70016. https://doi.org/10.1002/csc2.70016

U.S. Department of Agriculture, F. A. S. (2024). *Production—Wheat*. Production - Wheat. https://www.fas.usda.gov/data/production/commodity/0410000

Van Dijk, M., Morley, T., Rau, M. L., & Saghai, Y. (2021). A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. *Nature Food*, *2*(7), 494–501. https://doi.org/10.1038/s43016-021-00322-9

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Adriaenssens, E. M., Alfenas-Zerbini, P., Dempsey, D. M., Dutilh, B. E., García, M. L., Curtis Hendrickson, R., Junglen, S., Krupovic, M., Kuhn, J. H., Lambert, A. J., Łobocka, M., Oksanen, H. M., Orton, R. J., Robertson, D. L., Rubino, L., … Zerbini, F. M. (2022). Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses (2022). *Archives of Virology*, *167*(11), 2429–2440. https://doi.org/10.1007/s00705-022-05516-5

Walls, J., Rajotte, E., & Rosa, C. (2019). The Past, Present, and Future of Barley Yellow Dwarf Management. *Agriculture*, *9*(1), 23. https://doi.org/10.3390/agriculture9010023

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., Crossa, J., & Gai, J. (2020). Quantitative genetic studies with applications in plant breeding in the omics era. *The Crop Journal*, *8*(5), 683–687. https://doi.org/10.1016/j.cj.2020.09.001

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, R., Plieske, J., International Wheat Genome Sequencing Consortium, 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, X., Liu, Y., Chen, L., Zhao, D., Wang, X., & Zhang, Z. (2013). Wheat resistome in response to barley yellow dwarf virus infection. *Functional & Integrative Genomics*, *13*(2), 155–165. https://doi.org/10.1007/s10142-013-0309-4

Xavier, A., Xu, S., Muir, W. M., & Rainey, K. M. (2015). NAM: Association studies in multiple populations. *Bioinformatics*, *31*(23), 3862–3864. https://doi.org/10.1093/bioinformatics/btv448

Xu, Y., Li, P., Yang, Z., & Xu, C. (2017). Genetic mapping of quantitative trait loci in crops. *The Crop Journal*, *5*(2), 175–184. https://doi.org/10.1016/j.cj.2016.06.003

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*(2), 203–208. https://doi.org/10.1038/ng1702

Zeibig, F., Kilian, B., & Frei, M. (2022). The grain quality of wheat wild relatives in the evolutionary context. *Theoretical and Applied Genetics*, *135*(11), 4029–4048. https://doi.org/10.1007/s00122-021-04013-8

Zhang, Z., Lin, Z., & Xin, Z. (2009). Research progress in BYDV resistance genes derived from wheat and its wild relatives. *Journal of Genetics and Genomics*, *36*(9), 567–573. https://doi.org/10.1016/S1673-8527(08)60148-4

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology*, *7*(1–2), 203–214. https://doi.org/10.1089/10665270050081478

Zhang, Z., Xin, Z., Ma, Y., Chen, X., Xu, Q., & Lin, Z. (1999). Mapping of a BYDV resistance gene fromThinopyrum intermedium in wheat background by molecular markers. *Science in China Series C: Life Sciences*, *42*(6), 663–668. https://doi.org/10.1007/BF02881585

Zhang, Z., Xu, J., Xu, Q., Larkin, P., & Xin, Z. (2004). Development of novel PCR markers linked to the BYDV resistance gene *Bdv2* useful in wheat for marker-assisted selection.

*Theoretical and Applied Genetics*, *109*(2), 433–439. https://doi.org/10.1007/s00122-004-1649-1

Zhao, H., Zhang, L., Kirkham, M. B., Welch, S. M., Nielsen-Gammon, J. W., Bai, G., Luo, J., Andresen, D. A., Rice, C. W., Wan, N., Lollato, R. P., Zheng, D., Gowda, P. H., & Lin, X. (2022). U.S. winter wheat yield loss attributed to compound hot-dry-windy events. *Nature Communications*, *13*(1), 7233. https://doi.org/10.1038/s41467-022-34947-6

# Chapter 2 - Genome wide association for Barley Yellow Dwarf Virus resistance in a nested association mapping population of winter wheat

## Introduction

Wheat (*Triticum aestivum* L.) is a staple cereal crop that plays a crucial role in global food security, providing nearly 20% of human caloric and protein intake (Shiferaw et al., 2013). However, wheat production is frequently challenged by biotic and abiotic stresses, among which Barley Yellow Dwarf Virus (BYDV) represents a significant threat. BYDV is transmitted by aphids and is responsible for substantial yield losses due to stunted growth, leaf discoloration, and reduced kernel weight (Aradottir & Crespo-Herrera, 2021; Choudhury, Larkin, Xu, et al., 2019). Managing BYDV has proven difficult, as traditional control strategies such as delayed planting and insecticide applications provide only partial effectiveness (Hesler et al., 2005; Walls et al., 2019). Thus, genetic resistance remains the most promising and sustainable approach to mitigate its impact on wheat production(Bockus et al., 2016; Mc Namara et al., 2020).

Resistance to BYDV in wheat is predominantly quantitative, with few identified major resistance genes. *Bdv1*, first identified in the cultivar Anza (CI15284), confers partial tolerance to BYDV-MAV, but its effectiveness is strain-dependent (R. Singh, 1993).Other resistance genes, such as *Bdv2* and *Bdv3*, were introgressed from *Thinopyrum intermedium* and provide stronger resistance, particularly in Chinese wheat lines (L. I. Ayala-Navarrete et al., 2013; Zhang et al., 1999). Despite these advances, the limited availability of effective resistance genes has hindered breeding efforts. Moreover, the genetic basis of resistance is further complicated by

environmental variation and aphid population dynamics, which influence virus transmission and severity (Bian et al., 2025; Choudhury, Larkin, Xu, et al., 2019; Tidakbi et al., 2025).

Genome-wide association studies (GWAS) have emerged as a powerful tool for identifying quantitative trait loci (QTL) associated with BYDV resistance (Alseekh et al., 2021; Tibbs Cortes et al., 2021). However, GWAS in diverse populations is often constrained by linkage disequilibrium and genotyping errors, particularly in polyploid crops like wheat. Nested Association Mapping (NAM) populations offer an alternative strategy that combines the advantages of GWAS with linkage mapping to improve QTL resolution (Bajgain et al., 2016; Gireesh et al., 2021). Furthermore, advancements in genotyping technologies, such as skim exome capture (H. Wang et al., 2023), and imputation techniques like the Practical Haplotype Graph (PHG), enhance the power and accuracy of association mapping by reconstructing missing genotype data (Bradbury et al., 2022; Jensen et al., 2020; Jordan et al., 2022).

‘Overland’ winter wheat, a widely grown cultivar in the U.S. Great Plains has exhibited field resistance to BYDV, yet the genetic basis of this resistance remains unclear. To address this gap, a GWAS in a NAM population was conducted to identify SNPs associated with BYDV resistance in Overland-derived lines. The parental population was genotyped using higher coverage skim exome capture sequencing, with lower coverage skim exome capture applied to progeny lines. PHG-based imputation is utilized to refine genotype calls, maximizing statistical power for association mapping. The objective of this study is to identify genetic loci associated with BYDV resistance in Overland winter wheat using a GWAS approach in a NAM population. By leveraging skim exome capture sequencing, and PHG-based imputation, this research seeks to provide genomic resources for breeding wheat cultivars with improved resistance to BYDV, ultimately enhancing disease resistance and yield stability in wheat production.

# Materials and Methods

## Plant materials and field evaluation

The Nested Association Mapping (NAM) population used in this study was developed and provided by the USDA-ARS, Hard Winter Wheat Genetics Research Unit. The population was derived using the variety NE01643 (PI 647959), which was developed cooperatively by the Nebraska Agricultural Experiment Station and the USDA-ARS and released in 2007 under the name Husker Genetics Brand Overland (Baenziger et al., 2008). Overland served as the common parent and was crossed with ten diverse parental lines to generate $F_5$ recombinant inbred lines (RILs) through the single seed descent method. The number of families derived from each cross is detailed in Table 2.1. $F_{5:6}$ seed was multiplied in Yuma, AZ in 2022, and $F_{5:7}$ seed was used to plant the experiment in the fall of 2022. A total of 336 RILs were planted and analyzed under natural field infection conditions at the Kansas State University Ashland Bottoms Research Farm near Manhattan, KS (39°08'38.3"N 96°38'07.1"W). To promote natural infection, the population was planted in mid-September (Hesler et al., 2005). Each line was sown with approximately 30 seeds in a 1.4-meter row plot, with 30 cm spacing between rows. The experiment was analyzed as an augmented randomized complete block design with two replicates of each RIL. Each replicate was divided into five sub-blocks, each containing repeated checks (AgriPro ‘Bob Dole’; ‘McNair 701’; LCS T158; Kansas State University Experimental lines ‘KS05HW14A’; ‘KS89180B-2’ and the recurrent parent Overland). Also, the diverse parental lines were included in each replicate.

Phenotyping was conducted in the spring of 2023 by recording a BYDV visual score rating from 0 to 9, where 0 indicated no visible symptoms and 9 represented severe BYDV infection. These ratings were recorded 8 and 18 days after anthesis (daa). Anthesis was determined when at least 10% of the plot showed extruded anthers. Additionally, chlorophyll content was measured on the same day that the visual score was taken using a SPAD-502 meter (Spectrum Technologies, Plainfield, Illinois, USA) to estimate the amount of chlorophyll present in the leaf. For each plot, SPAD readings were taken from five flag leaves, and the average value was used for analysis.

Check varieties and one progeny line were sampled to confirm the viral population present in the field. This analysis was conducted by the Hard Winter Wheat Genetics Research Unit, USDA-ARS, following the protocol used in Ranabhat et al. (2025). The analysis confirmed the presence of BYDV-PAS in the field, along with detecting Cereal Yellow Dwarf Virus and Wheat Streak Mosaic Virus (John Fellers, personal communication).

## Phenotypic Data Analysis

To obtain unbiased estimates of genotype performance while accounting for experimental design factors, best linear unbiased estimates (BLUEs) were computed from raw field observations using a linear mixed model implemented in the lme4 package in R (Bates et al., 2015). The model used for BLUE estimation was:

$$Y_{ij} = \mu + G_i + R_j + B_{jk} + \varepsilon_{ijk}$$

where $Y_{ij}$ represents the observed BYDV resistance score for the $i^{th}$ genotype in the $j^{th}$ replication, $\mu$ is the overall mean, $G_i$ is the fixed effect of the $i^{th}$ genotype, $R_j$ is the random effect of the $j^{th}$ replication, $B_{jk}$ is the random effect of the $k^{th}$ block nested within the $j^{th}$ replication, and $\varepsilon_{ijk}$ is the residual error term.

To evaluate whether variation in flowering time could confound BYDV severity assessments, an extended linear model was fitted that included flowering dates as a covariate. The extended model was specified as:

$$Y_{ij} = \mu + G_i + R_j + B_{jk} + FD_{ij} + \varepsilon_{ijk}$$

where $FD_{ij}$ is the flowering dates covariate, expressed in Julian days. Including flowering dates in the model allowed for a statistical adjustment of BYDV severity scores based on phenological differences among genotypes. Since genotypes flower at different times, their exposure to BYDV pressure may vary depending on their developmental stage at the time of infection. If not accounted for, this could lead to biased estimates of disease severity (Emrich et al., 2008; Riedell et al., 1999; Silva et al., 2022). The extended model was therefore used to test for the potential confounding effect of flowering time on symptom expression.

**Estimation of heritability**

Variance components were estimated using the lme4 package in R, treating genotype as a random effect. Due to the loss of one sub-block, resulting in 73 lines without replication, $H^2$ estimates were computed following the Cullis method, which effectively addresses unbalanced replication. Unlike traditional heritability estimations that assume equal replication across genotypes, the Cullis method adjusts for prediction error variance (PEV), resulting in a more accurate estimation of the proportion of phenotypic variance explained by genetic effects. The formula used to calculate Cullis-adjusted broad-sense heritability was:

$$H^2 = 1 - \frac{\overline{PEV}}{\sigma_G^2}$$

where $\overline{PEV}$ represents the mean prediction error variance across all genotypes, and $\sigma_G^2$ is the estimated genetic variance. This method ensures that heritability values accurately reflect the precision of genotype predictions rather than being inflated due to data imbalance (Piepho & Möhring, 2007). Importantly, when replication is balanced and equal across genotypes, the Cullis-adjusted heritability yields the same estimates as the traditional method, highlighting its

robustness across both balanced and unbalanced experimental designs. This equivalence was confirmed in a subset of 263 fully replicated lines, where both methods produced nearly identical heritability values across all traits (Table 2.1 Parental lines used in the NAM population and the number of recombinant inbred lines (RILs) derived from each cross with Overland (NE01643)

| Parental Line | Origin | Number of RILs |
|---|---|---|
| PI 157568 | South Korea | 53 |
| PI 157577 | South Korea | 68 |
| PI 182851 | Czech Republic | 27 |
| PI 591997 | China | 50 |
| PI 481725 | Bhutan | 20 |
| PI 94454 | Russia | 12 |
| PI 376469 | Romania | 12 |
| PI 410428 | Ukraine | 26 |
| PI 470438 | Turkey | 17 |
| PI 476784 | Hungary | 51 |

Table 2.2).

## DNA extraction and genotyping

In brief, DNA was extracted from wheat leaf tissue following the method described by H. Wang et al., (2023). Seeds were first grown in trays under greenhouse conditions until the three-leaf stage. Two 3-cm leaf segments were collected from each plant, placed in 96-well tubes, and dehydrated using a freeze-drying system before being ground into a fine powder with zirconia beads. DNA extraction was performed using the sbeadex maxi plant kit (LGC Biosearch Technologies) on an oKtopure automated platform. The extracted DNA was quantified using the Quant-iT PicoGreen dsDNA Assay (Thermo Fisher Scientific), and samples were diluted to a working concentration of 1 ng/µL using automated liquid handling systems.

Tagmentation of genomic DNA was conducted using Illumina Tagment DNA Enzyme (*TDE1*) and buffer, optimized to generate library DNA fragments ranging from 300 to 600 bp. Dual-index adapters (NextSeq-i5 and NextSeq-i7) were then prepared and quantified using a modified OliGreen ssDNA assay. PCR amplification was performed using NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs) with a limited number of cycles to prevent over-amplification.

The resulting libraries were multiplexed into two separate pools: one containing only the 11 parental samples and the other comprising all 336 progeny. These libraries were then purified using QIAquick PCR spin columns (Qiagen) and concentrated before exome capture with the Wheat Exome v1 Kit (Arbor Biosciences), following the manufacturer's protocol.

The parental samples were captured together as a group of 11 to increase the availability of baits for hybridization, thereby enhancing sequencing depth and the diversity of captured DNA per parent. The progeny samples underwent exome capture in a similar manner but as a set of 336, which resulted in lower average sequencing depth and bait diversity per sample.

Following capture, both libraries were purified using QIAquick PCR spin columns (Qiagen), quantified using a Qubit dsDNA HS Assay (Thermo Fisher Scientific), and analyzed for fragment size distribution using an Agilent 2100 Bioanalyzer. The libraries were then normalized and combined in equimolar proportions to create the final sequencing library, which was run on a P2-300 Illumina chip. This equimolar mixing was performed to increase the relative read depth of the parental library compared to the progeny, ensuring higher coverage of parental lines for downstream imputation of the less-covered progeny samples.

## Bioinformatics analysis

### Skim exome capture

Initially, paired-end raw FASTQ files were quality-filtered with bbduk tools v38.79 (Bushnell, 2014) to remove any reads with an average quality score below 20. The filtered reads from the parental samples were then aligned to an Overland PacBio assembly (developed by the USDA Hard Winter Wheat Genotyping Laboratory) using HISAT2 v2.2.1 (Kim et al., 2019), resulting in a SAM file for each accession.

For the parental lines, a custom Perl script was subsequently applied to extract only the uniquely mapping reads from these SAM files. These filtered SAM files were then converted to BAM format for each accession using Samtools v1.9 (H. Li et al., 2009). Following this, the unique BAM files were processed according to GATK best practices: GATK's HaplotypeCaller v4.0.8.1 (McKenna et al., 2010) was run in GVCF mode to generate a g.vcf file for each sample.

Finally, all the g.vcf files were merged in GATK using the tools CombineGVCFs and GenotypeGVCFs to produce a final VCF file, which was further filtered with a custom Perl script to retain only high-quality bi-allelic variant calls present in at least 90% of the sequenced parental samples.

**PHG Construction and imputation**

The Wheat PHG database used in this study was modeled after the framework developed in previous studies (Bradbury et al., 2022; Franco et al., 2020; Jordan et al., 2022). This database comprises 94,175 informative reference ranges which were defined by extending high confidence annotated genes from the Overland assembly (USDA Central Small Grain Genotyping Center) by 2 kb upstream and downstream based on the coordinates from the annotation file.

The database was populated with sequence data from the parental lines. This set of haplotypes was selected to be representative of the haplotypic diversity present in the NAM

population. Finally, we generated consensus haplotypes for these reference ranges using diversity data from the parental accessions. In this step, raw haplotypes were collapsed into consensus haplotypes based on a user-defined maximum divergence (mxDiv) parameter set to 0.0001 for wheat. Additionally, we set minTaxa = 1 to retain haplotypes present in only a single accession, which aids in the imputation of rare haplotypes. This process generated 318,516 consensus haplotypes, which were then combined into a pangenome that served as a reference.

For the progeny lines, quality-filtered reads were uploaded to the PHG database (Bradbury et al., 2022; Franco et al., 2020; Jordan et al., 2022) and aligned to the consensus pangenome reference using the Minimap2 aligner (H. Li, 2018). Default parameters were used, except that minReads = 0, ensuring that the PHG imputed genotypes for all taxa regardless of sequencing coverage within each reference range. The alignment data was then processed to infer the most likely haplotype paths through the graph using hidden Markov models. Missing data were imputed based on transition probabilities, retaining only genotype calls with a transition probability of ≥0.99. Since the taxa were expected to be homozygous, the export was configured to generate homozygous genotype calls. The imputed variants were extracted from the PHG database resulting in 722,414 variants.

## Genome Wide Association Study

### Quality control and filtering of genotypic data

To ensure the accuracy and reliability of genotype data for association analysis, a two-step filtering process was applied. The first step involved family-wide filtering, where the 722,414 SNPs obtained from PHG imputation across 336 RILs were screened to retain only those variants that segregated between the parental lines. Additionally, within each RIL

population, markers with a minor allele frequency (MAF) greater than 0.25 were selected according to the expected segregation patterns in biparental populations.

Following this, the selected variants were aggregated and subjected to LD pruning using PLINK (Purcell et al., 2007). Markers exhibiting an $r^2$ ≥0.25 within a 50 kb sliding window, advancing two SNPs at a time, were removed to reduce redundancy and minimize the effects of highly correlated markers. The threshold of $r^2$ ≥0.25 was chosen to reflect the wheat NAM population's structure, balancing the removal of redundant markers with retaining sufficient coverage for downstream analyses. Although not extremely high, this threshold helps minimize bias from highly correlated markers and refines the dataset to a manageable size. A MAF threshold of ≥0.05 was then applied across the entire population to eliminate rare variants that could lead to spurious associations. After filtering, a final set of 19,647 high-quality SNPs remained and were used in the genome-wide association analysis.

As a control, an additional GWAS was performed using the complete set of markers without applying the LD pruning step. This control analysis retained 393,370 SNPs, filtered only by the family-wide filtering step followed by a MAF ≥ 0.05 threshold. The resulting Manhattan plots from this analysis are presented in Appendix A Figure A. 1 for the BLINK model and Appendix A Figure A. 2 for the MLMM model.

**Principal components and kinship analysis**

To account for population structure and relatedness among genotypes, principal component analysis (PCA) and a kinship matrix were calculated using TASSEL v5.2.93 (Bradbury et al., 2007). The first two principal components (PCA1 and PCA2) explained 6.38% of the genetic variation and were retained for further analysis, as the next principal component accounted for less than 3% of the variance. The kinship matrix was computed using the

VanRaden method, which estimates additive genetic relationships based on marker allele frequencies (VanRaden, 2008). Both PCA components and the kinship matrix were incorporated into GWAS models to control population structure and minimize false positive associations.

**Model selection**

For the GWAS, BLUEs were extracted by treating genotypes as fixed effects. This modeling approach allows for the estimation of genetic effects while minimizing environmental influences. The BLUEs were derived for BYDV severity at both time points, chlorophyll content at both measurements, plant height, and flowering date. Each model accounted for the appropriate experimental design factors, with replication and block effects included as random terms whenever those effects were significant in the fixed effect models.

A GWAS was conducted using the GAPIT3 package in R (J. Wang & Zhang, 2021). The analysis utilized Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) and the Multiple Locus Mixed Model (MLMM) to improve the detection of significant associations. These models were chosen because the distribution of markers across the wheat genome is uneven due to its large and complex genome structure and varying recombination rates. In addition, single-locus models GLM and MLM were also tested to provide a broader comparison of GWAS methodologies. The corresponding Manhattan plots for MLM and GLM are presented in Appendix A Figure A. 3 and Appendix A Figure A. 4, respectively, and a summary table of significant markers identified using MLM is included in Appendix A Table A. 1.

BLINK was selected for its ability to increase statistical power by iteratively selecting markers based on Bayesian information criteria while accounting for linkage disequilibrium, making it more efficient than traditional single-locus models (Huang et al., 2019). MLMM was

used because it considers multiple loci simultaneously, reducing confounding effects from background genetic variation and improving the resolution of QTL mapping (Segura et al., 2012).

Both models incorporated PCA1 and PCA2 as covariates to correct population structure and reduce false positives. To further control for type I errors, the False Discovery Rate (FDR) method was applied, which adjusts significance thresholds based on the proportion of expected false positives among the detected associations. In this study, associations were considered significant if the FDR-adjusted *P*-value was below the cutoff of 0.10. Unlike the more conservative Bonferroni correction, FDR allows for greater statistical power while still effectively minimizing false discoveries, ensuring that robust associations are identified and retained (Benjamini & Hochberg, 1995).

Significant SNPs identified through GWAS were selected for KASP marker development. Kompetitive Allele-Specific Polymorphism (KASP) primers were designed using PolyMarker (http://www.polymarker.info/) based on sequences flanking the SNPs in the Overland PacBio assembly. These primers were used to validate the segregation patterns first in the parental lines; then in a subset of the families in which that marker segregates, and then in an independent population.

**Validation**

The developed KASP markers that showed segregations patterns according to the Exome Capture genotyping data in the parental lines were tested on an independent population consisting of 779 experimental lines from the USDA Hard Winter Wheat Breeding Program. These lines, planted on September 26$^{th}$, 2023 at Hutchinson, KS, represent several stages of the breeding program from $F_3$ to $F_{5:6}$, which emphasizes the integration of desirable genetic traits

into high-yielding winter wheat germplasm for distribution to the breeding community. Visual scoring for BYDV was conducted on May 15th, 2024 (Figure 2.10). Pairwise comparisons of BYDV scores between haplotypes in each environment were conducted using the Wilcoxon test. Additionally, KASP assays targeting known genes conferring resistance to BYDV were tested and compared with the newly developed markers. The PCR program began with an initial denaturation at 94°C for 15 minutes, followed by a 10-cycle touchdown phase of 94°C for 20 seconds and 65°C for 1 minute, decreasing by 0.8°C per cycle. This was followed by 50 cycles of 94°C for 20 seconds and 57°C for 1 minute. An additional 32 cycles were performed with 56°C for 30 seconds, decreasing by 1.0°C per cycle. The program ended with a final extension at 10°C for 5 minutes.

For KASP assays, 3 μL of each genomic DNA sample (~6–30 ng) was loaded into a 384-well PCR plate (Thermo Scientific) and dehydrated at 65°C for 1 hour. A MANTIS liquid handler dispensed 4 μL of reaction mixture containing 1.7 μM of each allele-specific forward primer, 4.2 μM common reverse primer, and 1x PACE Genotyping Master Mix (3CR Bioscience). PCR amplification was performed using a C1000 Touch thermal cycler (Bio-Rad) with the primer sequences and PCR profiles detailed in Appendix A Table A. 2. Fluorescence signals were detected using a FLUOstar Omega Reader (BMG LABTECH), and genotype calls were analyzed using KlusterCaller Version 3.4.1.39 (LGC Biosearch Technologies) and a custom Excel sheet developed by the USDA Hard Winter Wheat Genotyping Laboratory.

# Results

## Phenotypic evaluation

During the establishment of the field trial, an unforeseen issue required one of the sub-blocks to be replanted three weeks later than the rest of the experiment. This delay, effectively

creating a late planting treatment, provided a unique opportunity to assess the influence of planting time on BYDV expression and associated agronomic traits. To evaluate the impact of this treatment, we compared the 73 recombinant inbred lines (RILs) that were present in both the late-planted block and the regularly planted blocks. A fixed-effect linear model was used to analyze differences across treatments for each trait. The analysis revealed that late planting had a statistically significant effect on multiple traits. BYDV severity, measured at both timepoints, was significantly reduced in the late-planted block. Chlorophyll content also showed significant increases in both the first and second measurements. In addition, flowering date was significantly delayed in response to the late planting. In contrast, plant height did not differ significantly between planting times (Figure 2.1).

This delay resulted in significant differences in plant growth stages and disease symptom development, leading to reduced symptom development and decrease variability in that particular sub-block in comparison with the others. Statistical analysis (Figure 2.2) confirmed that this sub-block had significantly different disease severity, chlorophyll content, and flowering date compared to the rest of the experiment. However, differences in height were not significant. Variability in these traits was also reduced. Given those differences, this sub-block was excluded from further analyses to maintain data integrity.

When analyzed as fixed effects (Table 2.3), genotypes showed significant variation for flowering date, plant height, and both BYDV severity measurements, while chlorophyll content displayed weaker genetic effects. Additionally, replication and block effects contributed significantly to the variation in several traits, including flowering date, plant height, BYDV severity, and chlorophyll content. This highlights the importance of the experimental design in

controlling spatial variation, particularly in aphid infestation and virus propagation across the field, which can otherwise obscure the accurate detection of genetic signals.

The remaining sub-blocks exhibited a range of phenotypic responses to natural BYDV infection under field conditions (Table 2.4). BYDV symptoms were first scored 8 days after anthesis for the first measurement and 18 days after anthesis for the second, with disease severity increasing as plants matured. Disease symptoms varied across genotypes, with observed differences in severity, chlorosis intensity, flowering time, and plant height. On average, BYDV severity was moderate, with a mean score of 3.52 (CV = 22.5%) for the first measurement and 3.41 (CV = 37%) for the second, ranging from 1.5 to 6.5 and 1.0 to 8.5, respectively. The recurrent parent, Overland, showed moderate resistance, with a mean BYDV severity of 3.4 in the first measurement and 3.1 in the second. Chlorophyll content varied considerably, with a mean of 36.7 (CV = 18.2%) in the first measurement and 30.9 (CV = 16.8%) in the second, spanning from 17.3 to 54.8 and 11.4 to 48.7, respectively. Flowering date averaged 131 days (CV = 1.17%), ranging from 128 to 134 days, while plant height varied from 64 cm to 112 cm, with a mean of 87.5 cm (CV = 11.4%). Overland exhibited mild chlorosis, with an average chlorophyll content of 36.91 (first measurement) and 28.41 (second measurement), along with moderate plant height (83.11 cm) and a flowering date of 131.11 days.

Among the parental lines, one did not survive the winter (PI 157568), while others exhibited a range of disease responses. The most resistant parental lines, such as PI 157577 and PI 263182, showed lower BYDV severity (3.0–4.0) and relatively high chlorophyll content (31.5–39.0). In contrast, PI 155278 was one of the most susceptible, with a high BYDV severity score of 5.0–8.0, low chlorophyll content (30 – 22.4), and the shortest plant height (56.5 cm).

Among the checks, McNair demonstrated the highest susceptibility to BYDV (severity values of 5.11 and 6.15 for first and second measurements, respectively) and had the lowest chlorophyll content (47.47 and 51.99 SPAD units), highlighting clear symptoms of viral-induced stress. Overland, on the other hand, exhibited the strongest BYDV resistance (severity of 3.23 and 3.28) and high chlorophyll levels, reflecting effective tolerance mechanisms despite its latest flowering date (131.09 days). The early-flowering checks, KS05HW14A (129.43 days) and Bob Dole (129.44 days), showed moderate BYDV tolerance. Notably, Bob Dole, despite early flowering, maintained relatively low severity values (3.43 and 3.60), coupled with good chlorophyll content and taller plant height (97.12 cm), indicating robust adaptability. KS05HW14A, the shortest variety (84.14 cm), had intermediate resistance to BYDV. Lastly, T158 and KS89180B-2 exhibited intermediate performance across most traits (Table 2.5).

Among the RILs, BYDV severity in the second measurement followed a right-skewed distribution, indicating that most lines had moderate resistance while a few exhibited extreme susceptibility. A small proportion of RILs (4.72%) showed extreme resistance with BYDV severity scores below 2, whereas 5.75% of lines displayed high susceptibility with scores above 6. The distribution is in Figure 2.3.

Higher BYDV severity was associated with reduced chlorophyll content ($r$ = -0.37, $P$ = $2.03 \times 10^{-23}$), shorter plant height ($r$ = -0.39, $P$ = $4.78 \times 10^{-26}$), and earlier flowering ($r$ = -0.24, $P$ = $6.37 \times 10^{-10}$), indicating that more severely affected plants exhibited increased leaf chlorosis, were shorter, and reached anthesis earlier (Figure 2.4).

## Heritability estimation

Due to the loss of one sub-block, resulting in 73 lines without replication, $H^2$ estimates were computed using genotype-specific PEV following the Cullis method, which effectively addresses unbalanced replication.

Linear mixed models were implemented using genotypes as random effects. This approach allows for the partitioning of phenotypic variance into genetic and environmental components, accounting for genetic variation while adjusting for environmental effects. In these models, BYDV severity (first and second measurements), chlorophyll content (first and second measurements), plant height, and flowering date were treated as response variables, with genotypes included as a random effect. Additionally, replication and block within replication were modeled as random terms where appropriate (Table 2.3).

Variance component analysis revealed substantial genetic variation for all evaluated traits (Table 2.6). Genotypic variance was largest for plant height and relatively high for flowering date, indicating strong genetic control. For BYDV severity, genotypic variance increased from the first to the second measurement, suggesting greater differentiation among genotypes as disease symptoms progressed. Chlorophyll content, however, showed a slight decrease in both genotypic and residual variance between the two measurements. This reduction in residual variance may reflect a more consistent physiological state among lines during the later stages of growth, potentially due to uniform onset of senescence or more synchronized environmental responses across genotypes.

$H^2$ was highest for plant height (0.74) and flowering date (0.67), indicating strong genetic control. Lower heritability values were obtained for both BYDV severity measurements ($H^2$ = 0.33 and 0.43) and chlorophyll content ($H^2$ = 0.16 and 0.19), suggesting greater influence from

uncontrolled variation such as localized disease pressure, microenvironmental conditions, or other unmeasured sources of error.

The performance of each check variety was evaluated between sub-blocks using mixed linear models, effectively isolating their genotypic values from spatial variation associated with different blocks. Due to their increased replication, checks enabled the estimation of repeatability, which reflects the consistency of a genotype's performance across replicates and serves as an upper bound for broad-sense heritability. As a result, check varieties represent the upper limit of achievable heritability in this experiment and provide an ideal benchmark for assessing the reliability and genetic signal of less-replicated experimental lines. For example, the checks showed notably higher heritabilities for plant height (0.92) and flowering date (0.81), substantially above the broader experimental heritabilities mentioned above. Similarly, BYDV severity for the checks had moderately high heritabilities (0.66–0.70) compared to the lower overall heritabilities observed across the full set of genotypes. Additionally, chlorophyll content heritabilities for checks (0.18–0.37) remained consistently higher than those from the broader experiment, although still relatively low. This suggests that both spatial variability and sampling error contributed to the lower heritability estimates observed in the full set, reinforcing that chlorophyll measurements were particularly sensitive to microenvironmental conditions and other uncontrolled sources of variation (Table 2.5).

## Chlorophyll Content Associated with BYDV Symptom Severity

The impact of BYDV infection was assessed by analyzing the relationship between visual symptom severity and chlorophyll content (SPAD values). Chlorophyll content differed significantly among groups categorized by symptom severity (ANOVA; first measurement: $P$ =

$3.69 \times 10^{-17}$, second measurement: $P = 4.27 \times 10^{-19}$), with a clear trend of reduced chlorophyll as symptoms intensified (Figure 2.5).

A negative linear relationship was observed between BYDV severity and chlorophyll content across both measurement timepoints. As BYDV severity scores increased, SPAD values decreased, indicating progressive chlorosis with higher symptom severity. In the first measurement, the regression slope was $-2.585 \pm 0.293$ ($P = 1.04 \times 10^{-17}$), while in the second measurement, the slope was $-1.652 \pm 0.159$ ($P = 2.03 \times 10^{-23}$), both indicating highly significant negative associations. The corresponding $R^2$ values were 0.11 and 0.14, respectively, with slightly stronger explanatory power observed in the second timepoint. These consistent negative slopes suggest that genotypes experiencing more severe BYDV symptoms tended to have reduced chlorophyll content, reinforcing the physiological impact of disease progression.

## Genotypic Data

### DNA sequencing

This sequencing library produced an average of 11,869,197 paired-end reads per parental line, compared to 851,002 reads per progeny line, resulting in a parental-to-progeny sequencing ratio of 13.95. Quality control metrics showed that approximately 90.9% of parental reads and 90.1% of progeny reads passed filtering. Alignment results indicated that 92.4% of parental and 91.1% of progeny reads aligned to the reference genome, with 85.2% and 80.9%, respectively, aligning uniquely. Assuming that the exome capture bait space in Overland is similar to Chinese Spring wheat, which spans approximately 226 Mb, this sequencing effort resulted in an estimated 7.93× depth of coverage across the targeted regions for the parental lines.

### SNP Data Processing

Genotyping of the NAM population resulted in a total of 722,414 SNPs following PHG imputation. Initially, markers were filtered within each RIL family based on a MAF > 0.25, resulting in 475,036 SNPs after this family-wide filter. Aggregation of filtered SNPs across all families was followed by LD pruning $r^2 > 0.25$ using PLINK, which reduced the dataset to 25,942 markers. Finally, a population-wide MAF > 0.05 filter was applied, resulting in a dataset of 19,647 high-quality SNPs utilized for subsequent genome-wide association analyses. For comparison, applying only the MAF > 0.05 filter without LD pruning yielded 393,369 SNPs; GWAS results using this larger marker set are included in Appendix A.

### Population Structure

Principal Component Analysis (PCA) was performed to assess population structure within the NAM population using a filtered set of 19,647 SNPs. The selection criteria retained two principal components (PCs), which together explained 6.38% of the total genetic variance, with PC1 contributing 4.18% and PC2 contributing 2.20%. Given that subsequent PCs explained less than 2% variance each, only PC1 and PC2 were included in downstream analyses. Despite the relatively low variance explained, the PCA scatter plot (Figure 2.6) revealed clustering of progeny lines by parental origin, confirming genetic structure within the NAM design.

Additionally, a kinship matrix was generated using the VanRaden method to quantify genetic relatedness among lines. The heatmap visualization of the kinship matrix (Figure 2.7) showed some distinct clusters along the diagonal, indicating groups of related progeny lines originating from shared parental lines.

## Genome-wide Association Study

To account for phenological variation that may confound visual assessments of disease severity, BYDV scores were recorded at standardized intervals relative to each line's flowering

date, specifically 8 and 18 days post-anthesis. To further test whether flowering date still influenced disease response, it was included as a fixed effect in the statistical models used to extract BLUEs. The analysis showed no significant effect of flowering date on BYDV severity for either the first ($P = 0.956$) or second ($P = 0.644$) measurement. As the experimental design had already accounted for phenological differences, flowering date was excluded from the final models.

Two GWAS models, BLINK and MLMM, were compared to identify significant marker-trait associations for BYDV symptom severity, chlorophyll content, plant height, and flowering date. Both models incorporated PCA components and kinship matrices to account for population structure and relatedness among lines. Manhattan plots are included in Figure 2.8 and Figure 2.9.

Significance was assessed using FDR adjusted $P$-value < 0.10 and Bonferroni thresholds $P$-value < $5.08 \times 10^{-6}$. BLINK identified more significant markers overall, particularly for plant height, where 12 markers passed the FDR threshold and 10 remained significant under Bonferroni correction. For BYDV severity, BLINK detected three and four significant markers for the first and second measurements, respectively, with a strong association on chromosome 5B. The second measurement captured additional associations on chromosomes 2B, 2D, 7B, and 7D, with the most prominent signal on chromosome 7B. MLMM identified only two significant markers on chromosomes 7B and 7D for the second measurement (Table 2.7).

For flowering date, MLMM identified a substantially higher number of associations than BLINK, detecting 20 markers with FDR correction, primarily on chromosome 2D. However, only five of these remained significant under Bonferroni correction, while BLINK identified just one marker on chromosome 4A.

Chlorophyll content from the first measurement showed three significant associations under the BLINK model and one under MLMM. The strongest signal was detected on chromosome 7B, corresponding to the same SNP that was later identified as significantly associated with BYDV severity in the second visual measurement. While all three associations detected by BLINK passed the FDR threshold, only the 7B SNP remained significant under the more stringent Bonferroni-adjusted threshold.

Overall, BLINK captured more robust associations for most traits, whereas MLMM detected additional loci for flowering date. These results highlight the complex genetic architecture of these traits, particularly for height and flowering date, where multiple loci contribute to phenotypic variation. Significant markers, their genomic positions, allele effects, and statistical significance are summarized in Table 2.7.

## Validation of Significant Markers

### KASP Marker Development

SNP markers on chromosomes 7B and 7D, identified through the MLMM model, exceeded the Bonferroni-adjusted significance threshold of $5.09\times10^{-7}$ ($\alpha = 0.10$) and were therefore prioritized for validation due to their strong association with BYDV severity in the second measurement. An additional marker on chromosome 6A, also detected by MLMM, did not meet the Bonferroni or FDR (<0.10) thresholds but was the third most strongly associated locus for the trait, with a *P*-value of $2.13\times10^{-5}$ and an FDR-adjusted *P*-value of 0.14. Given its close proximity to significance and potential biological relevance, the 6A marker was also selected for validation.

Flanking sequences (100 bp on each side of significant SNPs) were extracted and aligned using BLASTn version 2.16.1 (Zhang et al., 2000) against the IWGSC Chinese Spring RefSeq

v2.1 assembly. The SNP on chromosome 7B aligned perfectly (100% identity) to its corresponding region on chromosome 7B in Chinese spring but also showed high similarity (87–92%) to homeologous regions on chromosomes 7A and 7D, indicating a conserved sequence that may complicate marker specificity. The SNP on chromosome 7D matched its target region with 100% identity but also showed high homology to sequences on chromosomes 7B and 7A, suggesting potential challenges for allele-specific KASP marker development. The region corresponding to the SNP on chromosome 6A aligned most strongly to chromosome 6B (93% identity), followed by 6D and then 6A (92% identity), revealing substantial sequence conservation across homeologs and further complicating marker design.

Significantly associated SNP locations in chromosome 7B and 7D were not suitable for the development of robust KASP markers. Therefore, alternative SNPs within the same LD blocks were identified and tested. However, when these alternative SNPs were aligned to our Overland PacBio assembly, multi-chromosomal alignments were observed, leading to issues in marker specificity.

LD blocks were determined based on the physical positions of flanking SNPs associated with the trait. Using this information, we referenced the unfiltered SNP dataset and identified 87 SNPs within the candidate regions. To develop markers, we used Polymarker, selecting 22 SNPs for further validation based on markers specificity: 10 markers for the 7B locus, 8 for the 6A locus, and 4 for the 7D locus.

These markers were initially tested for consistency with expected genotype calls using their respective parental lines. Of the 22 markers, only 7 successfully amplified and showed genotype call patterns consistent with exome capture sequence data from the parental lines. These 7 markers were then tested in a subset of 55 progeny lines, where genotype calls were

predicted based on skim exome capture and PHG-based imputation. The tested markers included 4 in the 7B locus, 2 in the 6A locus, and 1 in the 7D locus (Table 2.8).

Among these, only 2 markers in the 7B locus exhibited clear separation between fluorescence clusters. The marker 7B:730434717 only misclassified 7.3% of the progeny tested. The marker 7B:730438210 misclassified 5.4%. These markers were further evaluated in the validation dataset, where 7B:730434717 amplified in 684 lines, with 121 lines carrying the SNP allele corresponding to Overland in the NAM population (Figure 2.14). The second marker, 7B:730438210, amplified 684 lines with 308 lines having the same SNP state as Overland (Figure 2.15).

However, single-marker association tests did not show significant differences between any of the six haplotype groups tested for these markers (Figure 2.11 and Figure 2.12). As a control, known KASP markers for *Bdv2* were also tested, and similarly, no significant differences were detected between the 44 lines that were positive for *Bdv2* and the rest of the lines (Figure 2.13).

## Discussion

BYDV remains a significant constraint to wheat production, causing yield losses through stunted growth, chlorosis, and reduced kernel weight (Aradottir & Crespo-Herrera, 2021; Choudhury, Larkin, Meinke, et al., 2019). While chemical control strategies, including insecticide applications and adjusted planting dates, offer partial mitigation, they are neither fully effective nor sustainable (Hesler et al., 2005; Walls et al., 2019). Genetic resistance remains the most viable long-term solution; however, the complex nature of BYDV resistance, which is largely polygenic, has slowed breeding progress (Bockus et al., 2016; Mc Namara et al., 2020).

The most well-characterized resistance loci, *Bdv1*, *Bdv2*, and *Bdv3*, provide varying degrees of protection, yet their effectiveness and agronomic impacts remain uncertain. The *Bdv2* and *Bdv3* resistance genes, introgressed from *Thinopyrum intermedium*, confer strong resistance but have seen limited commercial success in U.S. wheat breeding programs (L. I. Ayala-Navarrete et al., 2013; Zhang et al., 1999). Despite being available for over two decades, *Bdv2* is currently present in only two U.S. wheat cultivars, Uncharted from Oklahoma State University and Washburn from University of Minnesota, neither of which has gained significant commercial adoption. The next successful *Bdv2* wheat in the U.S. will be the first, indicating that this introgression may carry unfavorable agronomic trade-offs limiting its use in breeding programs.

One major concern regarding *Bdv2* is its chromosomal location, as it resides in the same genomic region as *Lr19*, a *Thinopyrum*-derived translocation associated with positive yield effects under irrigation but negative effects under rainfed conditions (Rosewarne et al., 2015). While direct evidence of a yield penalty for *Bdv2* is lacking, its performance in historical field trials suggests a tendency to perform well under optimal conditions but to decline under drought stress.

This study confirmed the importance of later-stage disease assessments when evaluating BYDV resistance. The second measurement, taken 18 days after anthesis, proved to be more reliable than the first measurement at 8 days after anthesis. Although statistical tests indicated a slight decrease in mean BYDV severity between the two measurements, the mean difference of 0.11 was small and not biologically meaningful. Instead of a significant progression in severity, disease expression remained stable over time, suggesting that the virus had reached a plateau within infected plants. However, variability increased in the second measurement (SD = 1.54 vs.

1.06 in the first measurement), indicating a wider range of responses among genotypes. This greater dispersion could be due to differential disease progression, environmental influences, or challenges in distinguishing early BYDV symptoms. A concern with later assessments is the potential for BYDV symptoms to be confounded with natural leaf senescence, which may interfere with the accuracy of visual scores at this stage. While some lines exhibited increased severity, the overall population did not show a consistent trend of worsening symptoms, further reinforcing the reliability of the second measurement for genetic screening.

Chlorophyll content measurements using a SPAD meter offer a quantitative and objective alternative to visual ratings, which can be influenced by scorer subjectivity. Although SPAD meters are not considered high-throughput tools, they are particularly useful in small plot settings, such as head rows in early-generation breeding nurseries. In these environments, UAV-based phenotyping, like the approach used by Silva et al. (2022) can be less effective due to the close spacing of rows, which limits the resolution and accuracy of aerial imagery. Interestingly, the association on chromosome 7B was already detected during the first chlorophyll measurement, indicating that SPAD readings can capture early, subtle physiological differences that are not easily observed with the unaided eye. This suggests that SPAD-based assessments can complement UAV platforms, which are well suited for identifying large-effect loci by improving the ability to detect more refined responses at the individual plant level during the early stages of symptom development.

However, it should be considered that chlorophyll content is not solely influenced by disease severity but is also affected by environmental factors, particularly drought stress. Under water-limited conditions, plants experience stomatal closure to reduce transpiration, leading to decreased $CO_2$ assimilation and an accumulation of reactive oxygen species, which in turn

accelerates chlorophyll degradation (Saradadevi et al., 2017). This physiological response can mimic the chlorosis observed in BYDV-infected plants, making it more difficult to distinguish disease resistance from drought tolerance, particularly in genotypes with variable responses to water stress. Given that drought stress can occur simultaneously with BYDV infection in field environments, differences in SPAD values across genotypes may reflect both stress responses rather than purely disease resistance. Thus, careful consideration of environmental conditions and additional physiological markers may be necessary to ensure the reliability of SPAD measurements as a proxy for BYDV resistance.

This study also compared the performance of MLMM and BLINK models for detecting loci associated with BYDV resistance and other agronomic traits. Although MLMM is designed to capture small-effect QTLs in polygenic traits, BLINK demonstrated superior power in identifying known, biologically meaningful loci with strong effects. For instance, BLINK successfully detected a significant association on chromosome 4B for plant height, aligning with the well-characterized dwarfing gene *Rht-B1b*, as well as a signal on chromosome 6A near the *Rht24* locus—both of which were not detected by MLMM. Similarly, for flowering date, MLMM identified several associations on chromosome 2D that overlap the known position of the *Ppd-D1* gene, highlighting its sensitivity to moderate-effect loci. These results suggest that while MLMM is useful for dissecting complex polygenic traits, BLINK provides a practical advantage in identifying robust, high-confidence QTLs that correspond to major genes, offering immediate value for marker-assisted selection. The ability of BLINK to capture established loci lends credibility to other associations it identifies, including those related to BYDV resistance, and supports its application in breeding programs targeting traits with both major and minor genetic components.

A factor that could be influencing BYDV severity and variability in the validation set is the stark contrast in precipitation between the two testing years. The NAM population at Ashland Bottoms in 2023 received 198 mm of precipitation by the first scoring date, whereas the validation set at Hutchinson in 2024 received only 98 mm over the same period. This two-fold difference in rainfall suggests that the validation set may have been subjected to significant drought stress, which could have interfered with BYDV symptom expression. Drought stress can induce chlorosis, stunt growth, and trigger premature senescence, all of which can mimic BYDV symptoms and impair accurate disease assessment. The presence of environmentally induced chlorosis in drought-stressed plants is likely confounded disease ratings, making it difficult to distinguish true resistance from abiotic stress responses.

While Silva et al. (2022) and Tidakbi et al. (2025) demonstrated a significant resistance effect of *Bdv2* in Kansas environments, the variability observed in the validation set raises questions about the reproducibility of these effects under different environmental conditions. Validated KASP markers for *Bdv2* identified 44 lines containing the introgression in the validation set, but these lines did not show a statistically significant difference in disease response compared to the rest of the set (Figure 2.13). Since visual scores are primarily based on leaf chlorosis patterns, other factors beyond BYDV infection, such as drought stress, could have introduced error in the disease ratings. The extent to which environmental conditions influence disease severity underscores the challenge of relying solely on visual assessments in field evaluations.

In contrast, the NAM population at Ashland Bottoms was grown under more favorable moisture conditions, allowing for a clearer differentiation of genetic resistance. These environmental differences could have also influenced aphid populations, disease pressure, and

overall symptom expression, further complicating comparisons between datasets. The fact that the validation set was evaluated under drought conditions while the NAM population was grown in a wetter year-location suggests that true resistance effects may have been masked in the validation trial due to environmental confounding.

To increase confidence in these results and account for potential genotype-by-environment (GxE) interactions, an additional year of field data should be collected to evaluate the same NAM population under different environmental conditions. A second year of data would help determine the stability of QTL effects across varying climatic conditions, providing a more robust assessment of BYDV resistance loci.

Despite this challenges, this study successfully identified several significant BYDV resistance QTL, with major loci detected on chromosomes 5B, 6A, 7B, and 7D. Unlike *Bdv2*, which has been studied primarily as a single large-effect locus, the resistance identified in Overland-derived lines appears to be polygenic. Given the structure of the NAM population, a joint QTL mapping approach could further improve the identification of smaller-effect QTLs that may not have been well captured in the GWAS analysis. Joint linkage-GWAS approaches have been shown to provide higher confidence in trait-associated regions by integrating both family-based linkage information and population-wide marker associations (Bajgain et al., 2016; Gireesh et al., 2021). Such an approach could be particularly valuable in this study, as it may improve the resolution of QTL regions and enhance the detection of minor alleles contributing to BYDV resistance. By leveraging both linkage and association mapping, future work could provide a clearer understanding of the genetic basis of BYDV tolerance present in Overland.

The identification of native BYDV resistance loci in Overland-derived lines provides a valuable opportunity for developing resistant wheat cultivars without the potential agronomic

drawbacks associated with *Bdv2* introgressions. The ability to select resistance loci within adapted germplasm allows breeders to recombine favorable alleles while avoiding potential yield penalties. Future work should focus on fine-mapping key QTL, validating their stability across diverse environments, integrating a joint QTL mapping approach, and expanding field evaluations to additional years to better capture GxE interactions. These steps will be critical to ensuring that these QTL are broadly applicable and effectively deployed in wheat breeding programs, ultimately improving yield stability and sustainability in wheat production.

## Conclusion

This study identified polygenic resistance to BYDV in Overland-derived wheat lines, with significant QTL on chromosomes 5B, 6A, 7B, and 7D. Unlike *Bdv2*, which has seen limited adoption in U.S. wheat breeding, these native resistance loci offer greater breeding flexibility. The impact of environmental variability, particularly drought stress, highlights the need for multi-year testing to confirm QTL stability. The use of exome capture sequencing and PHG-based imputation provided a cost-effective genotyping approach, but further fine mapping and validation are necessary to integrate these loci into breeding programs.

## References

Alseekh, S., Kostova, D., Bulut, M., & Fernie, A. R. (2021). Genome-wide association studies: Assessing trait characteristics in model and crop plants. Cellular and Molecular Life Sciences, 78(15), 5743–5754. https://doi.org/10.1007/s00018-021-03868-w

Aradottir, G. I., & Crespo-Herrera, L. (2021). Host plant resistance in wheat to barley yellow dwarf viruses and their aphid vectors: A review. Current Opinion in Insect Science, 45, 59–68. https://doi.org/10.1016/j.cois.2021.01.002

Ayala-Navarrete, L. I., Mechanicos, A. A., Gibson, J. M., Singh, D., Bariana, H. S., Fletcher, J., Shorter, S., & Larkin, P. J. (2013). The Pontin series of recombinant alien translocations in bread wheat: Single translocations integrating combinations of *Bdv2*, Lr19 and Sr25

disease-resistance genes from Thinopyrum intermedium and Th. ponticum. Theoretical and Applied Genetics, 126(10), 2467–2475. https://doi.org/10.1007/s00122-013-2147-0

Baenziger, P. S., Beecher, B., Graybosch, R. A., Ibrahim, A. M. H., Baltensperger, D. D., Nelson, L. A., Jin, Y., Wegulo, S. N., Watkins, J. E., Hatchett, J. H., Chen, M.-S., & Bai, G. (2008). Registration of 'NE01643' Wheat. Journal of Plant Registrations, 2(1), 36–42. https://doi.org/10.3198/jpr2007.06.0327crc

Bajgain, P., Rouse, M. N., Tsilo, T. J., Macharia, G. K., Bhavani, S., Jin, Y., & Anderson, J. A. (2016). Nested Association Mapping of Stem Rust Resistance in Wheat Using Genotyping by Sequencing. PLOS ONE, 11(5), e0155760. https://doi.org/10.1371/journal.pone.0155760

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

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

Bian, R., Liu, N., Xu, Y., Su, Z., Chai, L., Bernardo, A., St. Amand, P., Rupp, J., Pumphrey, M., Fritz, A., Zhang, G., Jordan, K. W., & Bai, G. (2025). A novel quantitative trait locus for barley yellow dwarf virus resistance and kernel traits on chromosome 2D of a wheat cultivar Jagger. The Plant Genome, 18(1), e20548. https://doi.org/10.1002/tpg2.20548

Bockus, W. W., De Wolf, E. D., & Todd, T. C. (2016). Management Strategies for Barley Yellow Dwarf on Winter Wheat in Kansas. Plant Health Progress, 17(2), 122–127. https://doi.org/10.1094/PHP-RS-15-0050

Bradbury, P. J., Casstevens, T., Jensen, S. E., Johnson, L. C., Miller, Z. R., Monier, B., Romay, M. C., Song, B., & Buckler, E. S. (2022). The Practical Haplotype Graph, a platform for storing and using pangenomes for imputation. Bioinformatics, 38(15), 3698–3702. https://doi.org/10.1093/bioinformatics/btac410

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Bushnell, B. (2014). BBMap: A fast, accurate, splice-aware aligner.

Choudhury, S., Larkin, P., Meinke, H., Hasanuzzaman, M. D., Johnson, P., & Zhou, M. (2019). Barley yellow dwarf virus infection affects physiology, morphology, grain yield and flour pasting properties of wheat. Crop and Pasture Science, 70(1), 16. https://doi.org/10.1071/CP18364

Choudhury, S., Larkin, P., Xu, R., Hayden, M., Forrest, K., Meinke, H., Hu, H., Zhou, M., & Fan, Y. (2019). Genome wide association study reveals novel QTL for barley yellow dwarf virus resistance in wheat. BMC Genomics, 20(1), 891. https://doi.org/10.1186/s12864-019-6249-1

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

Emrich, K., Wilde, F., Miedaner, T., & Piepho, H. P. (2008). REML approach for adjusting the Fusarium head blight rating to a phenological date in inoculated selection experiments of wheat. Theoretical and Applied Genetics, 117(1), 65–73. https://doi.org/10.1007/s00122-008-0753-z

Franco, J. A. V., Gage, J. L., Bradbury, P. J., Johnson, L. C., Miller, Z. R., Buckler, E. S., & Romay, M. C. (2020). A Maize Practical Haplotype Graph Leverages Diverse NAM Assemblies. bioRxiv, 2020.08.31.268425. https://doi.org/10.1101/2020.08.31.268425

Gireesh, C., Sundaram, Raman. M., Anantha, S. M., Pandey, M. K., Madhav, M. S., Rathod, S., Yathish, K. R., Senguttuvel, P., Kalyani, B. M., Ranjith, E., Venkata Subbarao, L., Kumar Mondal, T., Swamy, M., & Rakshit, S. (2021). Nested Association Mapping (NAM) Populations: Present Status and Future Prospects in the Genomics Era. Critical Reviews in Plant Sciences, 40(1), 49–67. https://doi.org/10.1080/07352689.2021.1880019

Hesler, L. S., Riedell, W. E., Langham, M. A. C., & Osborne, S. L. (2005). Insect Infestations, Incidence of Viral Plant Diseases, and Yield of Winter Wheat in Relation to Planting Date in the Northern Great Plains. Journal of Economic Entomology, 98(6), 2020–2027. https://doi.org/10.1093/jee/98.6.2020

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), giy154. https://doi.org/10.1093/gigascience/giy154

Jensen, S. E., Charles, J. R., Muleta, K., Bradbury, P. J., Casstevens, T., Deshpande, S. P., Gore, M. A., Gupta, R., Ilut, D. C., Johnson, L., Lozano, R., Miller, Z., Ramu, P., Rathore, A., Romay, M. C., Upadhyaya, H. D., Varshney, R. K., Morris, G. P., Pressoir, G., … Ramstein, G. P. (2020). A sorghum practical haplotype graph facilitates genome-wide imputation and cost-effective genomic prediction. The Plant Genome, 13(1), e20009. https://doi.org/10.1002/tpg2.20009

Jordan, K. W., Bradbury, P. J., Miller, Z. R., Nyine, M., He, F., Fraser, M., Anderson, J., Mason, E., Katz, A., Pearce, S., Carter, A. H., Prather, S., Pumphrey, M., Chen, J., Cook, J., Liu, S., Rudd, J. C., Wang, Z., Chu, C., … Akhunov, E. D. (2022). Development of the Wheat Practical Haplotype Graph database as a resource for genotyping data storage and genotype imputation. G3 Genes|Genomes|Genetics, 12(2), jkab390. https://doi.org/10.1093/g3journal/jkab390

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology, 37(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics, 34(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Mc Namara, L., Gauthier, K., Walsh, L., Thébaud, G., Gaffney, M., & Jacquot, E. (2020). Management of yellow dwarf disease in Europe in a post-neonicotinoid agriculture. Pest Management Science, 76(7), 2276–2285. https://doi.org/10.1002/ps.5835

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Piepho, H.-P., & Möhring, J. (2007). Computing Heritability and Selection Response From Unbalanced Plant Breeding Trials. Genetics, 177(3), 1881–1888. https://doi.org/10.1534/genetics.107.074229

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics, 81(3), 559–575. https://doi.org/10.1086/519795

Ranabhat, N. B., Fellers, J. P., Bruce, M. A., & Rupp, J. L. S. (2025). High-Throughput Oxford Nanopore Sequencing Unveils Complex Viral Population in Kansas Wheat: Implications for Sustainable Virus Management. Viruses, 17(1), 126. https://doi.org/10.3390/v17010126

Riedell, W. E., Kieckhefer, R. W., Haley, S. D., Langham, M. A. C., & Evenson, P. D. (1999). Winter Wheat Responses to Bird Cherry-Oat Aphids and Barley Yellow Dwarf Virus Infection. Crop Science, 39(1), 158–163. https://doi.org/10.2135/cropsci1999.0011183X003900010025x

Rosewarne, G., Bonnett, D., Rebetzke, G., Lonergan, P., & Larkin, P. (2015). The Potential of Lr19 and *Bdv2* Translocations to Improve Yield and Disease Resistance in the High Rainfall Wheat Zones of Australia. Agronomy, 5(1), 55–70. https://doi.org/10.3390/agronomy5010055

Saradadevi, R., Palta, J. A., & Siddique, K. H. M. (2017). ABA-Mediated Stomatal Response in Regulating Water Use during the Development of Terminal Drought in Wheat. Frontiers in Plant Science, 8, 1251. https://doi.org/10.3389/fpls.2017.01251

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

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., & Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. Food Security, 5(3), 291–317. https://doi.org/10.1007/s12571-013-0263-y

Silva, P., Evers, B., Kieffaber, A., Wang, X., Brown, R., Gao, L., Fritz, A., Crain, J., & Poland, J. (2022). Applied phenomics and genomics for improving barley yellow dwarf resistance in winter wheat. G3 Genes|Genomes|Genetics, 12(7), jkac064. https://doi.org/10.1093/g3journal/jkac064

Singh, R. (1993). Genetic association of gene *Bdv1* for tolerance to barley yellow dwarf virus with genes *Lr34* and *Yr18* for adult plant resistance to rusts in bread wheat. Plant Disease, 77(11), 1103–1106. CABI Databases. https://doi.org/10.1094/PD-77-1103

Tibbs Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. The Plant Genome, 14(1), e20077. https://doi.org/10.1002/tpg2.20077

Tidakbi, L., Wang, H., Bian, R., Redila, C., Zhu, D., St. Amand, P., Bernardo, A., Bruce, M., Zhang, G., Fritz, A., Guttieri, M., Fellers, J., Bai, G., Noller, J. R., & Jordan, K. W. (2025). Genome-wide association study identifies novel associations with barley yellow dwarf virus and soil-borne wheat mosaic virus resistance in winter wheat association mapping panel. Crop Science, 65(2), e70016. https://doi.org/10.1002/csc2.70016

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science, 91(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Walls, J., Rajotte, E., & Rosa, C. (2019). The Past, Present, and Future of Barley Yellow Dwarf Management. Agriculture, 9(1), 23. https://doi.org/10.3390/agriculture9010023

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., & 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

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A Greedy Algorithm for Aligning DNA Sequences. Journal of Computational Biology, 7(1–2), 203–214. https://doi.org/10.1089/10665270050081478

Zhang, Z., Xin, Z., Ma, Y., Chen, X., Xu, Q., & Lin, Z. (1999). Mapping of a BYDV resistance gene fromThinopyrum intermedium in wheat background by molecular markers. Science in China Series C: Life Sciences, 42(6), 663–668. https://doi.org/10.1007/BF02881585

**Table 2.1 Parental lines used in the NAM population and the number of recombinant inbred lines (RILs) derived from each cross with Overland (NE01643)**

| Parental Line | Origin | Number of RILs |
|---|---|---|
| PI 157568 | South Korea | 53 |
| PI 157577 | South Korea | 68 |
| PI 182851 | Czech Republic | 27 |
| PI 591997 | China | 50 |
| PI 481725 | Bhutan | 20 |
| PI 94454 | Russia | 12 |
| PI 376469 | Romania | 12 |
| PI 410428 | Ukraine | 26 |
| PI 470438 | Turkey | 17 |
| PI 476784 | Hungary | 51 |

**Table 2.2 Comparison of Heritability Estimates Using Entry-Mean and Cullis Methods in Fully Replicated Lines.**

| Trait | Entry-Mean $H^2$ | Cullis-Adjusted $H^2$ |
|---|---|---|
| BYDV Severity (1st Measurement) | 0.37 | 0.37 |
| BYDV Severity (2nd Measurement) | 0.47 | 0.47 |
| Chlorophyll Content (1st Measurement) | 0.18 | 0.19 |
| Chlorophyll Content (2nd Measurement) | 0.22 | 0.22 |
| Flowering Date | 0.71 | 0.71 |
| Plant Height | 0.78 | 0.78 |

This table summarizes broad-sense heritability estimates for six agronomic traits based on 263 fully replicated experimental lines. Heritability was calculated using two approaches: the entry-mean method and the Cullis-adjusted method. Both approaches used mixed models with Genotypes treated as a random effect. The near-identical values obtained from both methods indicate that, under balanced designs with equal replication, both approaches yield equivalent results.

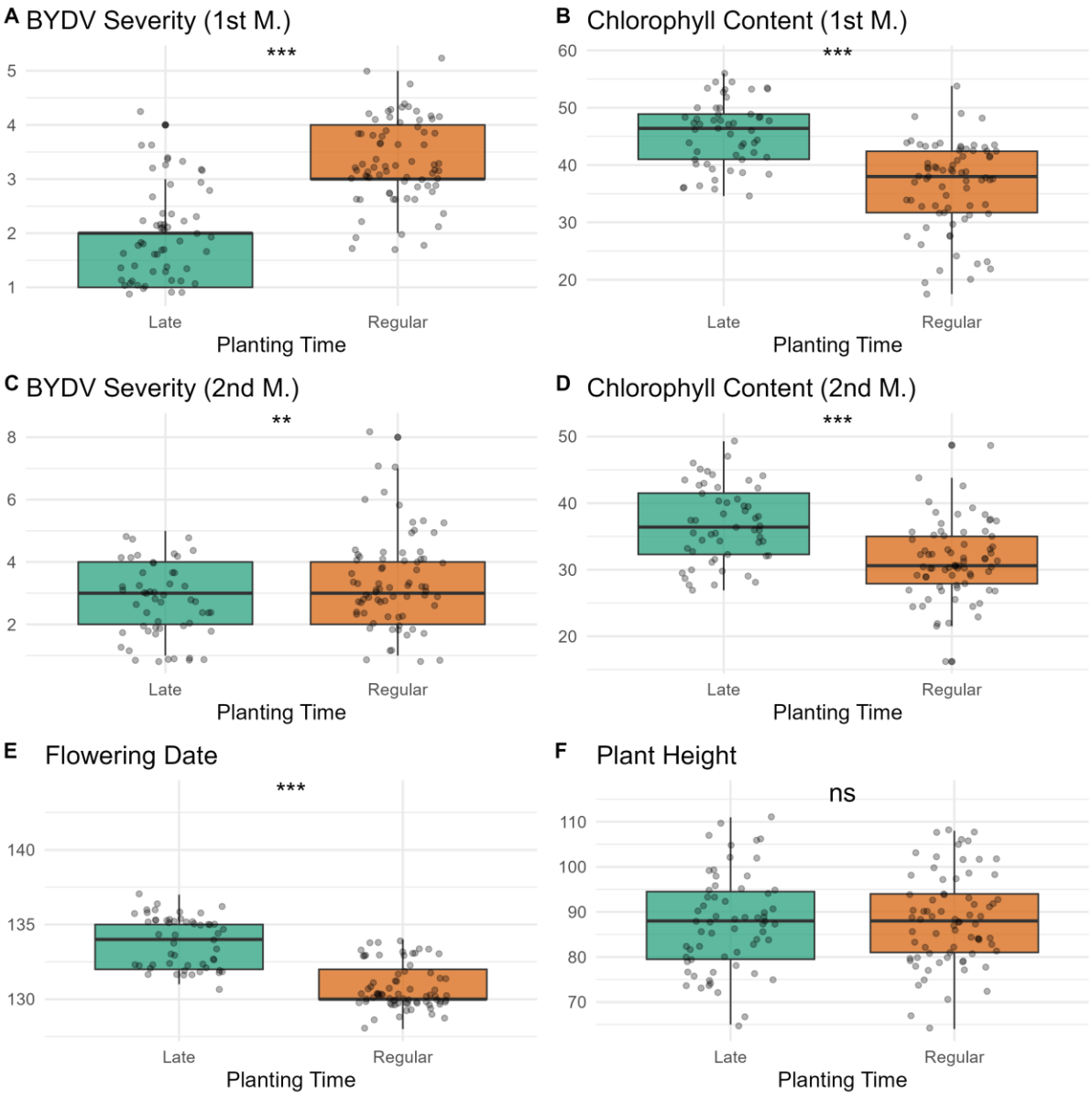



**Figure 2.1 Distribution of phenotypic traits across planting time treatments, comparing regular and late planting (Block 9). Boxplots display BYDV severity (A, C), chlorophyll content (B, D), flowering date (E), and plant height (F) across the two planting dates.**

One block was replanted three weeks later due to an unforeseen issue, is shown as the "Late" treatment, while the remaining blocks represent the "Regular" planting. The median is marked by a horizontal line, with the first and third quartiles forming the edges of the box. Whiskers

extend to the most extreme values within 1.5 times the interquartile range, and outliers are shown as dots. BYDV severity (A, C) represents disease symptom intensity, chlorophyll content (B, D) reflects leaf greenness, flowering date (E) indicates time to anthesis, and plant height (F) measures final plant stature. Asterisks indicate significant differences between treatments: * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$; "ns" indicates non-significant differences. Statistical comparisons were conducted using independent two-sample *T*-tests.

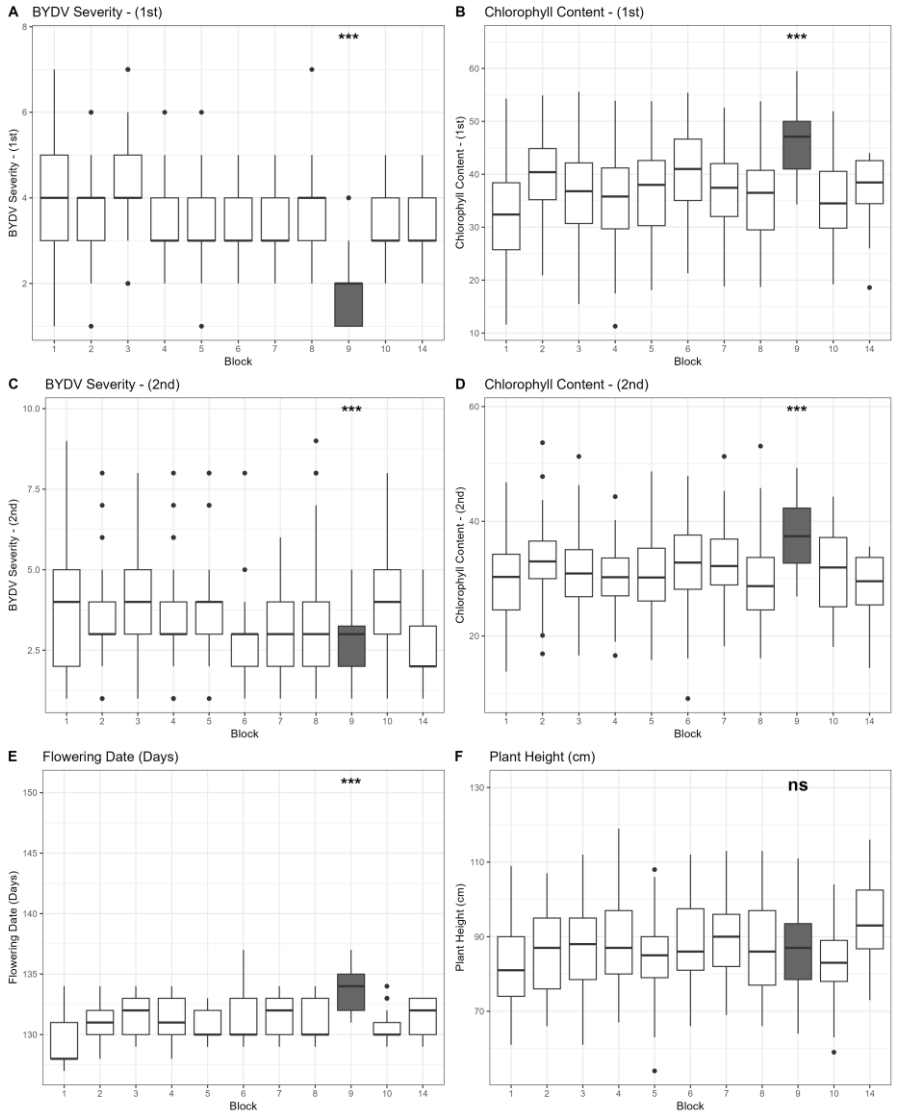



**Figure 2.2 Distribution of phenotypic traits across experimental blocks, highlighting Block 9. Boxplots display BYDV severity (A, B), chlorophyll content (C, D), flowering date (E), and plant height (F) across all blocks.**

Block 9, replanted three weeks later due to an unforeseen issue, is shown in dark gray, while other blocks are white. The median is marked by a horizontal line, with the first and third quartiles forming the edges of the box. Whiskers extend to the most extreme values within 1.5 times the interquartile range, and outliers are indicated as dots. BYDV severity (A, B) represents

disease symptom intensity, chlorophyll content (C, D) reflects leaf greenness, flowering date (E) indicates time to anthesis, and plant height (F) measures final plant stature. Blocks with significant differences have *, **, or *** above the graph for $p < 0.05$, 0.01, and 0.001, respectively, while non-significant differences are labeled "ns". Statistical comparisons were conducted using independent two-sample *T*-tests.

**Table 2.3 Analysis of variance (ANOVA) summary for key traits under BYDV infection.**

| Trait | Source | SS | MS | *F* Value | *P* Value |
|---|---|---|---|---|---|
| Flowering Date (JD) | Genotypes | 1511.141 | 4.367 | 4.53 | 1.20E-37 |
| | Rep | 5.978 | 5.978 | 6.201 | 0.0133 |
| | Rep:Block | 139.956 | 17.494 | 18.146 | 4.00E-22 |
| | Residuals | 290.186 | 0.964 | NA | NA |
| Plant Height | Genotypes | 68833.938 | 198.942 | 4.68 | 4.60E-39 |
| | Rep | 775.906 | 775.906 | 18.254 | 2.60E-05 |
| | Rep:Block | 1401.502 | 175.188 | 4.122 | 1.10E-04 |
| | Residuals | 12794.045 | 42.505 | NA | NA |
| BYDV Severity (1st M.) | Genotypes | 431.069 | 1.246 | 1.583 | 2.30E-05 |
| | Rep | 46.239 | 46.239 | 58.758 | 2.50E-13 |
| | Rep:Block | 34.037 | 4.255 | 5.407 | 2.30E-06 |
| | Residuals | 236.868 | 0.787 | NA | NA |
| BYDV Severity (2nd M.) | Genotypes | 1016.612 | 2.938 | 1.819 | 6.20E-08 |
| | Rep | 41.038 | 41.038 | 25.408 | 8.00E-07 |
| | Rep:Block | 21.598 | 2.7 | 1.672 | 0.105 |
| | Residuals | 486.16 | 1.615 | NA | NA |
| Chlorophyll C. (1st M.) | Genotypes | 26389.246 | 76.269 | 1.254 | 0.0218 |
| | Rep | 225.749 | 225.749 | 3.711 | 0.055 |
| | Rep:Block | 2374.052 | 296.756 | 4.878 | 1.10E-05 |
| | Residuals | 18310.875 | 60.833 | NA | NA |
| Chlorophyll C. (2nd M.) | Genotypes | 16954.556 | 49.002 | 1.173 | 0.0771 |
| | Rep | 40.935 | 40.935 | 0.98 | 0.323 |
| | Rep:Block | 846.853 | 105.857 | 2.534 | 0.0111 |
| | Residuals | 12573.137 | 41.771 | NA | NA |

ANOVA results for BYDV severity (1st and 2nd measurement), chlorophyll content (1st and 2nd measurement), flowering date, and plant height are presented, with genotypes, replication (Rep), and incomplete block structure (Block:Rep) included as sources of variation. The table reports

sum of squares (SS), mean squares (MS), *F*-values, and *P*-values. Genotype effects indicate the level of genetic variation for each trait, while block and replication effects account for experimental design structure.

**Table 2.4 Summary of Phenotypic Responses to BYDV Infection.**

| Trait | Mean | CV (%) | Range | Overland |
|---|---|---|---|---|
| BYDV Severity (1st Measurement) | 3.52 | 22.5 | 1.5 - 6.5 | 3.4 |
| BYDV Severity (2nd Measurement) | 3.41 | 37 | 1.0 - 8.5 | 3.1 |
| Chlorophyll Content (1st Measurement) | 36.7 | 18.2 | 17.3 - 54.8 | 36.91 |
| Chlorophyll Content (2nd Measurement) | 30.9 | 16.8 | 11.4 - 48.7 | 28.41 |
| Flowering Date | 131 | 1.17 | 128 - 134 | 131.11 |
| Plant Height | 87.5 | 11.4 | 64 - 112 | 83.11 |

Summary statistics for key phenotypic traits measured under natural BYDV infection conditions for RILs. BYDV severity was scored at two points: 8 days after anthesis (first measurement) and 18 days after anthesis (second measurement). Chlorophyll content was recorded alongside BYDV severity to assess leaf chlorosis. Flowering date represents the number of days from planting to anthesis, and plant height was measured at physiological maturity. The mean value, coefficient of variation (CV %), and range are provided for each trait. The Overland recurrent parent is included for comparison.

**Table 2.5 Genotypic values and broad-sense heritability ($H^2$) for key check varieties under BYDV infection**

| Check | BYDV (1st M.) | BYDV (2nd M.) | Chlorophyll Content (1st) (SPAD Units) | Chlorophyll Content (2nd) (SPAD Units) | Plant Height (cm) | Flowering date (Julian days) |
|---|---|---|---|---|---|---|
| ***H*²** | **0.70** | **0.66** | **0.37** | **0.18** | **0.81** | **0.92** |
| Overland | 3.2 | 3.5 | 36.3 | 31.7 | 131.4 | 81.1 |
| Bob Dole | 3.33 | 3.89 | 38.63 | 33.17 | 129.33 | 82 |
| KS05HW14A | 4.1 | 4 | 38.8 | 35.3 | 129.6 | 68 |
| KS89180B-2 | 4.7 | 5.4 | 37.6 | 38.3 | 130.4 | 72.7 |
| McNair | 5.1 | 6 | 30.4 | 30 | 130 | 83.6 |
| T158 | 4.2 | 4.1 | 41.5 | 34.4 | 129.5 | 71.3 |

Genotypic values and entry-mean broad-sense heritability ($H^2$) are presented for six check varieties evaluated under Barley Yellow Dwarf Virus (BYDV) infection. Traits include BYDV severity at the first and second measurement (BYDV (1st M.) and BYDV (2nd M.)), chlorophyll content at two time points (SPAD Units), plant height (cm), and flowering date (Julian days). Genotypic values were estimated from linear mixed models that included accession as a fixed effect and block as a random effect, isolating genotype performance within sub-blocks. Broad-sense heritability ($H^2$) was estimated on an entry-mean basis using an average replication of 8.5. Check varieties were more highly replicated than the full population and thus represent the upper bound of achievable heritability. Higher $H^2$ values indicate greater genetic control over the trait, while lower values reflect stronger environmental influence.

**Table 2.6 Genotypic variance, residual variance, and broad-sense heritability ($H^2$) for key traits under BYDV infection.**

| Trait | Genotypic Variance | Residual Variance | $H^2$ |
|---|---|---|---|
| BYDV (1st Measurement) | *0.204* | *0.756* | 0.33 |
| BYDV (2nd Measurement) | 0.675 | 1.552 | 0.43 |
| Chlorophyll Content (1st) | 6.872 | 65.177 | 0.16 |
| Chlorophyll Content (2nd) | 5.321 | 40.957 | 0.19 |
| Plant Height | 74.988 | 43.225 | 0.74 |
| Flowering Date | 1.590 | 1.367 | 0.67 |

Genotypic variance, residual variance, and Cullis-adjusted broad-sense heritability ($H^2$) are presented for BYDV severity (1st and 2nd measurement), chlorophyll content (1st and 2nd measurement), flowering date, and plant height. Genotype was included as a random effect in a mixed model for each trait. Genotypic variance quantifies the genetic contribution to trait variation, while residual variance accounts for environmental and unexplained sources of variation. Cullis-adjusted heritability ($H^2$) incorporates prediction error variance (PEV) to provide an unbiased estimate of the genetic signal under an unbalanced experimental design. Higher $H^2$ values indicate greater genetic control over the trait, while lower values suggest a stronger contribution of residual variation, such as measurement error or uncontrolled within-field variability.

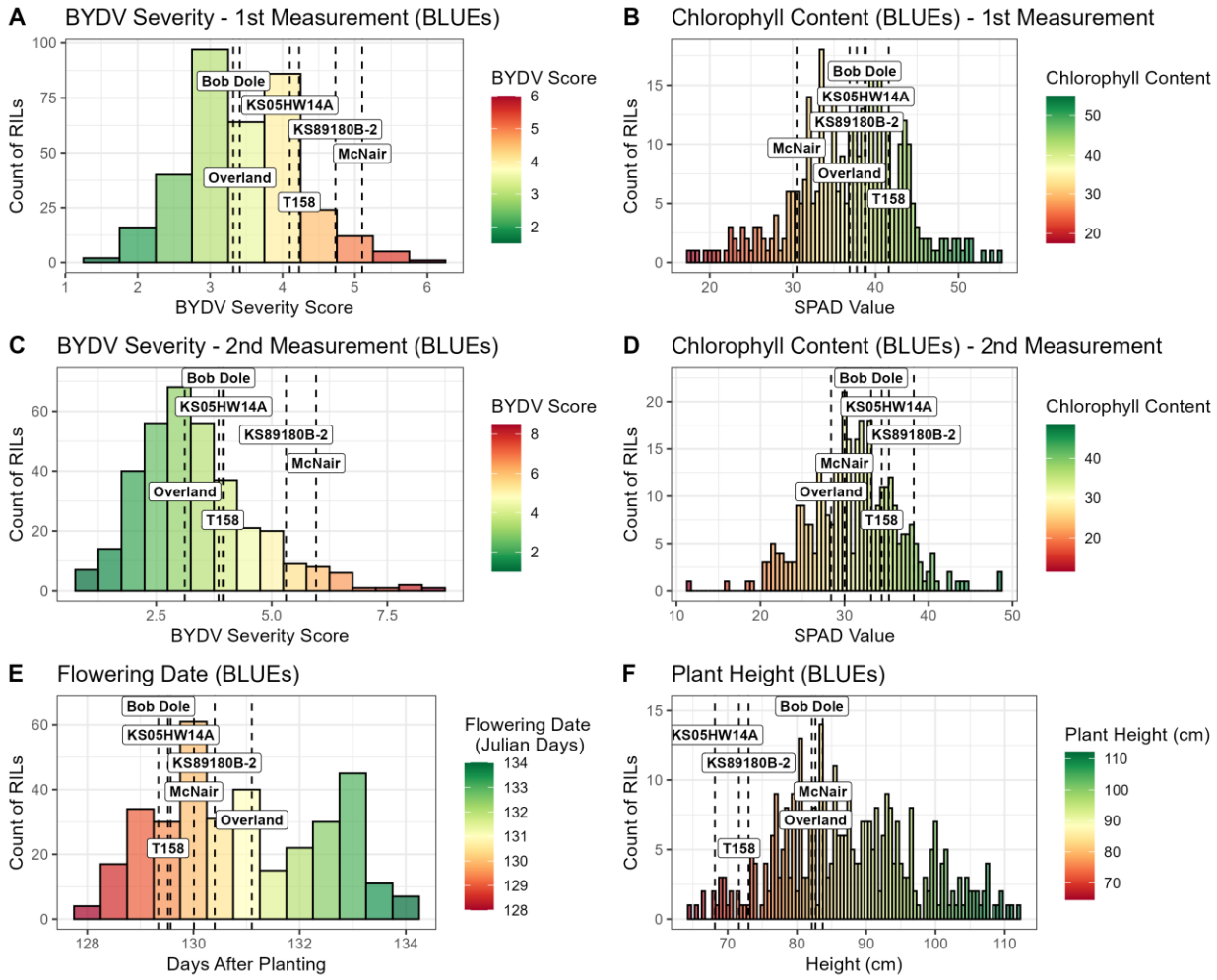



**Figure 2.3 Distribution of BYDV severity, chlorophyll content, flowering date, and plant height in recombinant inbred lines (RILs) based on BLUEs estimates.**

Histograms depict the distribution of BYDV severity (A, C), chlorophyll content (B, D), flowering date (E), and plant height (F) among RILs. The x-axis represents trait values, while the y-axis indicates the number of RILs within each bin. A color gradient visually distinguishes phenotypic variation, with green indicating lower values and red indicating higher values for BYDV severity (A) and chlorophyll content (B), while the inverse pattern is applied for BYDV severity (C), chlorophyll content (D), flowering date (E), and plant height (F), where red

represents lower values and green represents higher values. Dashed vertical lines highlight the positions of key parental lines and checks, and labels indicate genotypes.

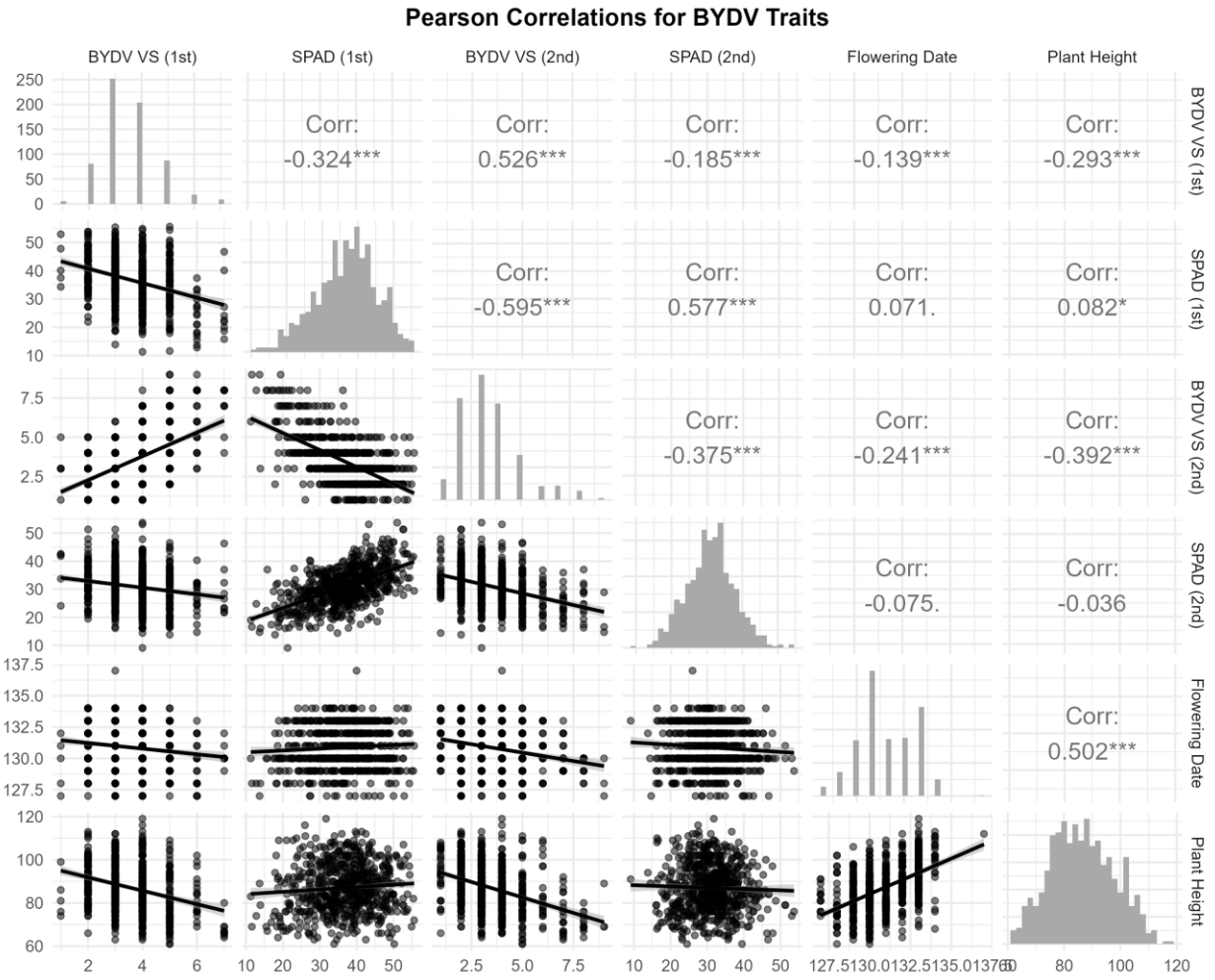



**Figure 2.4 Pearson correlation matrix for BYDV-related traits.**

Pairwise correlations among BYDV severity (1st and 2nd measurement), chlorophyll content (SPAD, 1st and 2nd measurement), flowering date, and plant height are shown. The upper triangle displays Pearson correlation coefficients, with asterisks indicating significance (* $P<$ 0.05, ** $P < 0.01$, *** $P < 0.001$). The lower triangle presents scatter plots with fitted trend lines, illustrating relationships between traits. Diagonal histograms represent the distribution of each trait. Strong correlations highlight associations between disease severity, chlorophyll content, and plant development under BYDV infection conditions.

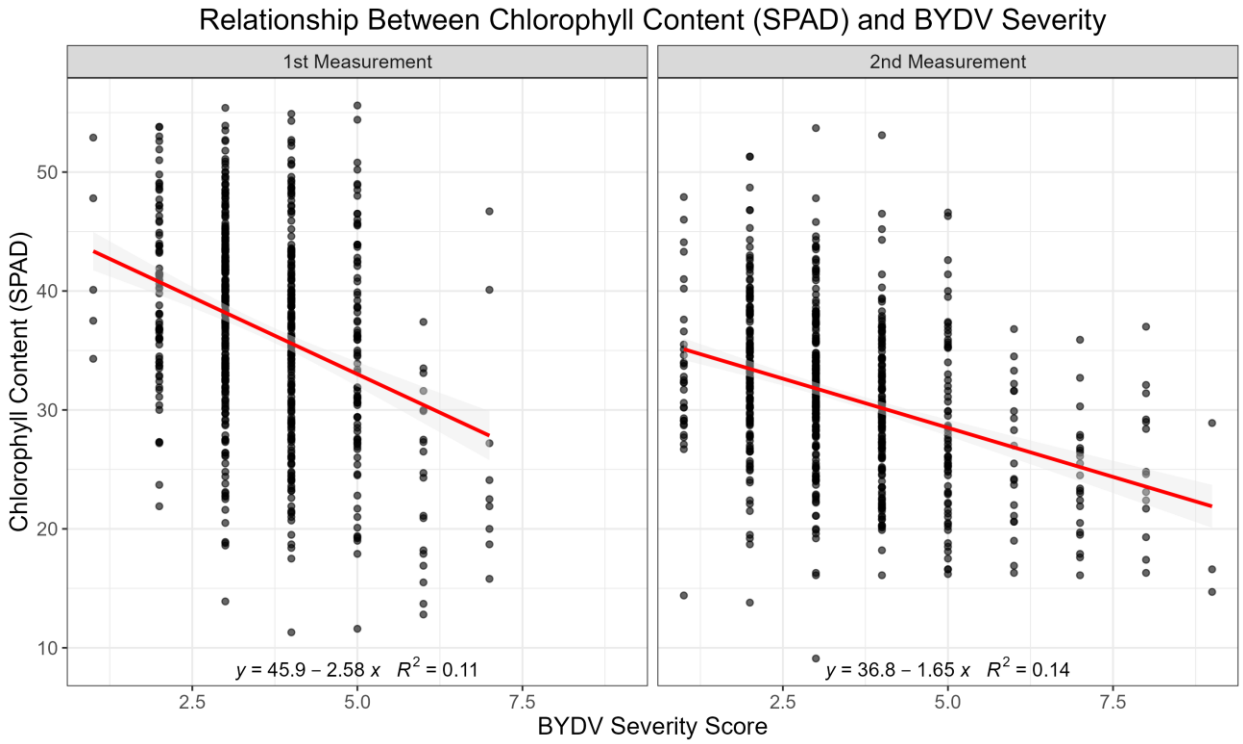



**Figure 2.5 Relationship between chlorophyll content (SPAD) and BYDV severity across two measurements.**

Scatter plots display the relationship between BYDV severity scores and chlorophyll content (SPAD readings) at the first (left) and second (right) measurement. A red linear regression line with a shaded confidence band represents the fitted model, while the equation and $R^2$ value are shown below each panel. The first measurement was taken 8 days after anthesis, and the second was taken 18 days after anthesis, allowing comparison of disease progression and its effect on leaf chlorosis. The observed negative trends indicate that increasing BYDV severity is associated with reduced chlorophyll content, and this relationship becomes slightly stronger over time.

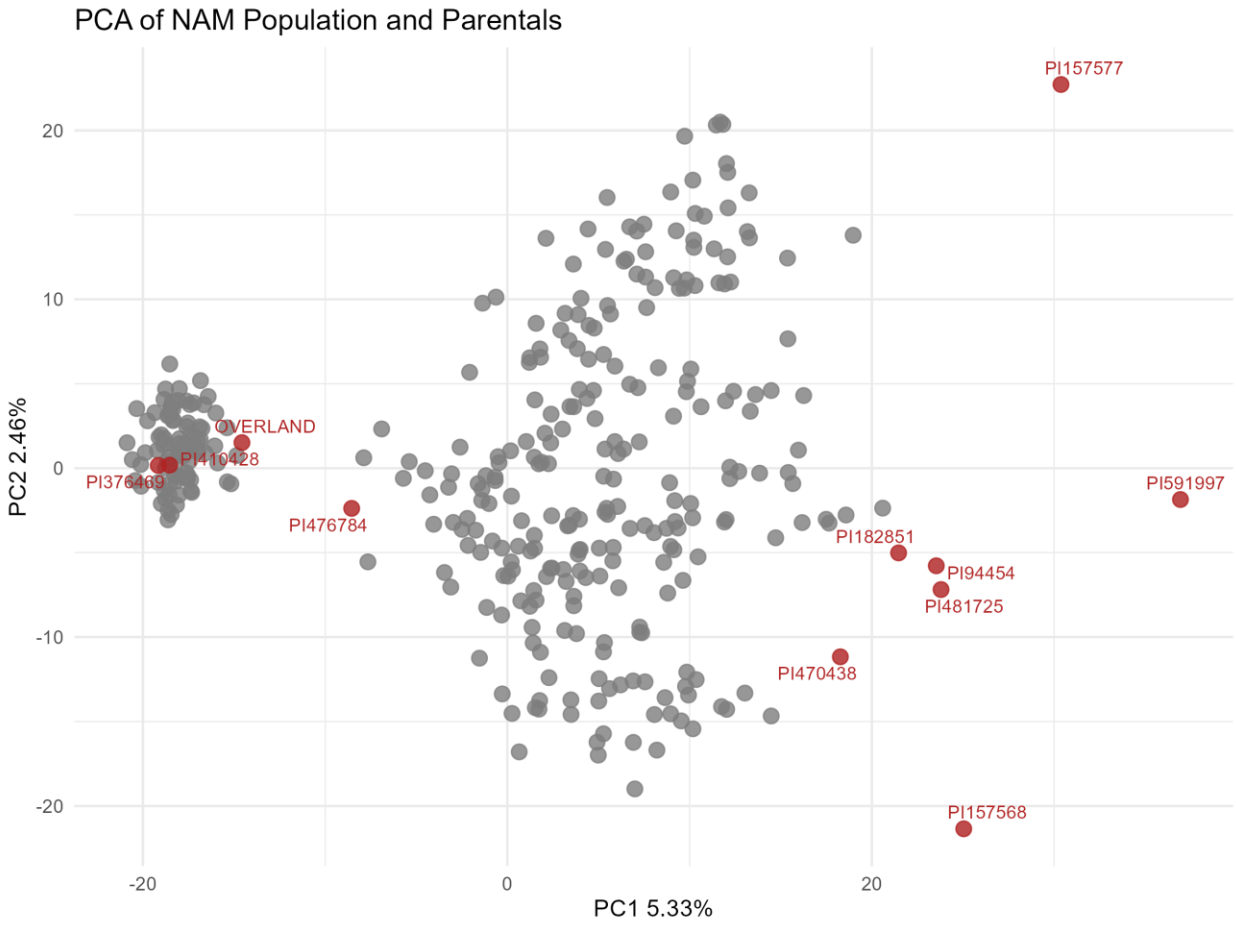



**Figure 2.6 Principal Component Analysis (PCA) of population structure.**

Scatterplot of the first two principal components (PC1 and PC2) derived from genome-wide SNP data, illustrating the genetic structure of the population. Each grey point represents an individual recombinant inbred line (RIL), with clustering patterns reflecting genetic relatedness. Each red point represents a parental line. PC1 and PC2 explain the largest proportions of genetic variance, capturing population stratification and potential substructure. The distribution of genotypes along these axes suggests differentiation among genetic groups, providing insight into the underlying population structure and relatedness among lines.

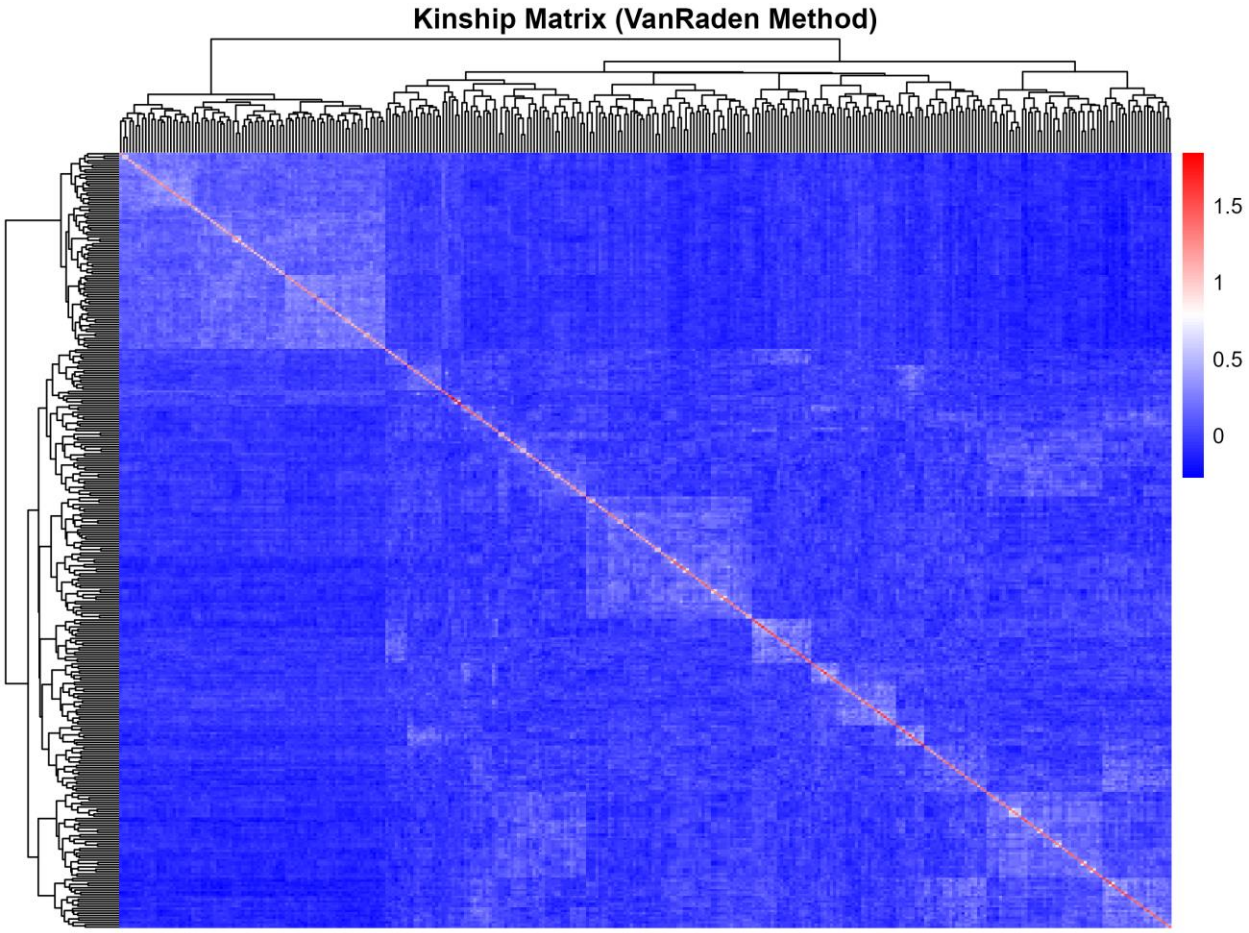



**Figure 2.7 Kinship heatmap of the population based on the VanRaden method.**

Kinship heatmap of the population based on the VanRaden method. Heatmap representation of the kinship matrix computed from genome-wide SNP data using the VanRaden method. The color scale ranges from blue (low genetic similarity) to red (high genetic similarity), with white indicating intermediate values. Rows and columns correspond to individual recombinant inbred lines (RILs), clustered hierarchically to visualize genetic relationships. The presence of distinct clusters suggests varying degrees of relatedness among genotypes, reflecting the underlying population structure.

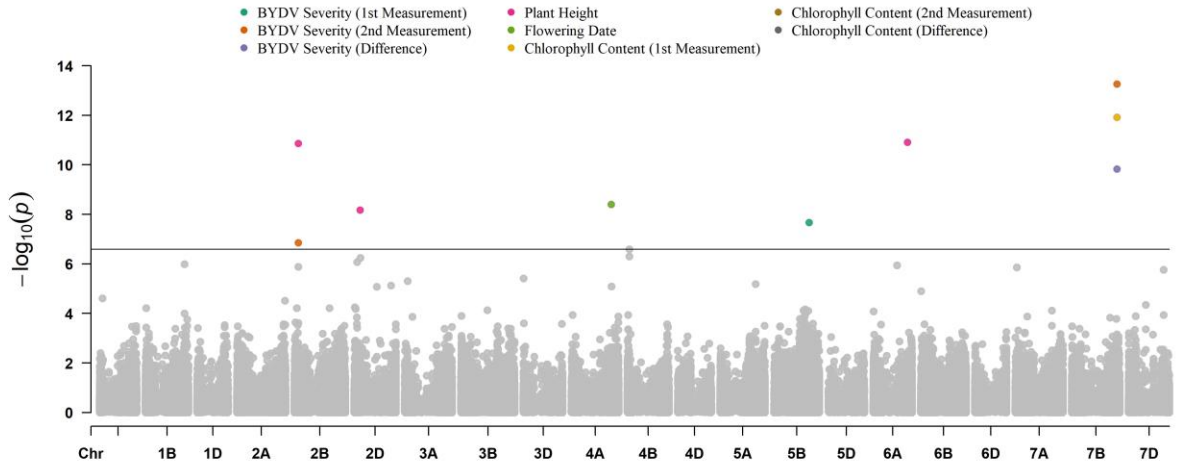



**Figure 2.8 Manhattan plot displaying SNP-trait associations identified through the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) model.**

Each point represents a SNP, plotted by chromosomal position (x-axis) and -$\log_{10}$(*P*-value) (y-axis). The horizontal black line represents the Bonferroni-adjusted threshold at $\alpha < 0.10$, indicating the significance level used to identify associated loci. Significant SNPs surpassing this threshold are highlighted in different colors based on trait associations. BYDV Severity (1st Measurement) and BYDV Severity (2nd Measurement) refer to visual ratings of Barley Yellow Dwarf Virus symptoms taken at 8 and 18 days after flowering, respectively, while the BYDV Severity (Difference) represents the change in symptom severity over time. Plant Height was measured at maturity from the soil surface to the tip of the spike (excluding awns), and Flowering Date corresponds to the number of days from planting until 10% of the plants reached anthesis. Chlorophyll Content (1st and 2nd Measurements) was assessed using a SPAD meter at 8 and 18 days after flowering, respectively, and their difference was used to capture changes in

leaf greenness over time. The plot highlights loci significantly associated with variation in these agronomic and physiological traits under BYDV pressure.

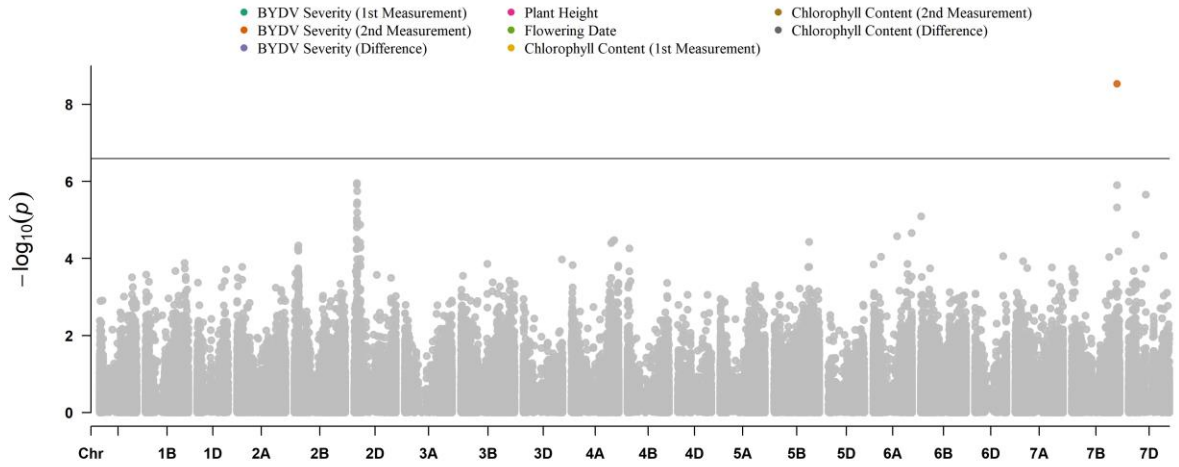



**Figure 2.9 Manhattan plot of genome-wide association study (GWAS) results using the MLMM model.**

Manhattan plot displaying SNP-trait associations identified through the Multi-Locus Mixed Model (MLMM). Each point represents a SNP, plotted by chromosomal position (x-axis) and -$\log_{10}$(*P*-value) (y-axis). The horizontal black line represents the Bonferroni-adjusted threshold at $\alpha < 0.10$, indicating the significance level used to identify associated loci. Significant SNPs surpassing this threshold are highlighted in different colors based on trait associations. BYDV Severity (1st Measurement) and BYDV Severity (2nd Measurement) refer to visual ratings of Barley Yellow Dwarf Virus symptoms taken at 8 and 18 days after flowering, respectively, while the BYDV Severity (Difference) represents the change in symptom severity over time. Plant Height was measured at maturity from the soil surface to the tip of the spike (excluding awns), and Flowering Date corresponds to the number of days from planting until 10% of the plants reached anthesis. Chlorophyll Content (1st and 2nd Measurements) was assessed using a SPAD meter at 8 and 18 days after flowering, respectively, and their difference was used to capture changes in leaf greenness over time. The plot highlights loci significantly associated with variation in these agronomic and physiological traits under BYDV pressure.

**Table 2.7 Summary of significant SNP-trait associations detected using BLINK and MLMM models.**

| Trait | Model | Chr | Physical Position | Parental origin | Effect | *P*-value |
|---|---|---|---|---|---|---|
| BYDV First Measurement (0-9 score) | BLINK | 2D | 595737507 | non-OV | 0.256 | 7.51E-06 |
| | | 5A | 572928016 | non-OV | 0.313 | 6.61E-06 |
| | | 5B | 570073019 | OV | -0.241 | 2.17E-08 |
| BYDV Second Measurement (0-9 score) | BLINK | 2B | 63688651 | non-OV | 0.331 | 1.40E-07 |
| | | 2D | 99417996 | OV | -0.269 | 5.75E-07 |
| | | 7B | 730451609 | non-OV | 0.801 | 5.51E-14 |
| | | 7D | 567312804 | non-OV | 0.459 | 1.76E-06 |
| | MLMM | 7B | 730451609 | non-OV | 0.866 | 2.97E-09 |
| | | 7D | 275211794 | non-OV | 0.474 | 2.24E-06 |
| BYDV Severity (Difference) | BLINK | 2B | 63688651 | non-OV | 0.273 | 1.31E-06 |
| | | 4B | 30760237 | OV | -0.424 | 4.98E-07 |
| | | 7B | 730451609 | non-OV | 0.616 | 1.50E-10 |
| | MLMM | 7B | 730451609 | non-OV | 0.619 | 1.26E-06 |
| Chlorophyll Content First Measurement (SPAD units) | BLINK | 2D | 367387432 | non-OV | -1.85 | 8.50E-06 |
| | | 4A | 646225443 | non-OV | -1.648 | 8.10E-06 |
| | | 7B | 730451609 | non-OV | -4.352 | 1.21E-12 |
| | MLMM | 7B | 730451609 | non-OV | -3.531 | 4.74E-06 |
| Flowering date (days) | BLINK | 4A | 639917966 | OV | 0.367 | 3.95E-09 |
| | MLMM | 2B | 60599283 | non-OV | -0.411 | 5.52E-05 |
| | | 2B | 63688651 | non-OV | -0.456 | 4.65E-05 |
| | | 2D | 41160034 | OV | 0.358 | 3.23E-05 |
| | | 2D | 41572849 | non-OV | -0.365 | 5.69E-05 |
| | | 2D | 42000453 | OV | 0.442 | 1.11E-06 |
| | | 2D | 44499356 | non-OV | -0.388 | 1.13E-05 |
| | | 2D | 44720168 | OV | 0.420 | 3.98E-06 |
| | | 2D | 44807394 | non-OV | -0.401 | 9.47E-06 |
| | | 2D | 45169853 | OV | 0.425 | 1.21E-06 |

| Trait | Model | Chr | Physical Position | Parental Origin | Effect | Min *P*-value |
|---|---|---|---|---|---|---|
| Flowering date (days) | MLMM | 2D | 45497716 | OV | 0.380 | 1.56E-05 |
| | | 2D | 47092124 | non-OV | -0.426 | 3.51E-06 |
| | | 2D | 47435326 | OV | 0.372 | 3.65E-05 |
| | | 2D | 49029785 | OV | 0.371 | 5.31E-05 |
| | | 2D | 49161244 | OV | 0.404 | 1.24E-05 |
| | | 2D | 49762185 | OV | 0.367 | 3.52E-05 |
| | | 2D | 49944668 | non-OV | -0.430 | 1.79E-06 |
| | | 2D | 50187302 | non-OV | -0.398 | 6.39E-06 |
| | | 4A | 639917966 | OV | 0.357 | 4.05E-05 |
| | | 4A | 666029690 | OV | 0.304 | 3.59E-05 |
| | | 4A | 690312469 | OV | 0.333 | 3.37E-05 |
| Plant height (cm) | BLINK | 1B | 631539572 | non-OV | -1.926 | 1.02E-06 |
| | | 2A | 777668426 | non-OV | -1.828 | 3.10E-05 |
| | | 2B | 63688651 | non-OV | -2.884 | 1.39E-11 |
| | | 2D | 50207315 | non-OV | -2.646 | 8.41E-07 |
| | | 2D | 98443066 | non-OV | -2.214 | 6.80E-09 |
| | | 3A | 45347892 | non-OV | 3.392 | 5.03E-06 |
| | | 3D | 101218 | OV | -1.533 | 3.84E-06 |
| | | 4B | 31714889 | non-OV | 3.029 | 2.60E-07 |
| | | 6A | 384938038 | non-OV | 3.612 | 1.15E-06 |
| | | 6A | 552193961 | Non-OV | 3.098 | 1.24E-11 |
| | | 6B | 5239780 | non-OV | 1.802 | 1.29E-05 |
| | | 7A | 22423699 | OV | 2.87 | 1.42E-06 |

Significant marker-trait associations identified for BYDV severity (first and second measurement), chlorophyll content (first measurement), flowering date, and plant height are shown. The chromosome (Chr), physical position, effect size, and *P*-value for each association are reported. SNPs were detected using the BLINK and MLMM models, with significance

determined by the Bonferroni-adjusted threshold at $\alpha < 0.10$ or an FDR-adjusted *P*-value of < 0.10. A column indicating the parental origin of the associated allele is also included: "OV" denotes the allele originated from the recurrent parent Overland, while "non-OV" indicates the allele came from one of the ten diverse PI parental lines. Larger effect sizes indicate stronger associations, while lower *P*-values suggest greater statistical support for the SNP-trait relationship.

**Table 2.8 Genotype call consistency for KASP marker assays and sequencing data in parental and progeny lines.**

| Chromosome | Physical Position | Parental Call Match | Progeny Call Match |
|---|---|---|---|
| 6A | 623291202 | FAILED | |
| | 623291392 | PASSED | FAILED |
| | 623291398 | PASSED | FAILED |
| | 623293247 | FAILED | |
| | 623293615 | FAILED | |
| | 623294481 | FAILED | |
| | 623326268 | FAILED | |
| 7B | 730434717 | PASSED | PASSED |
| | 730438186 | PASSED | FAILED |
| | 730438210 | PASSED | PASSED |
| | 730438273 | FAILED | |
| | 730438358 | FAILED | |
| | 730439269 | FAILED | |
| | 730439309 | FAILED | |
| | 730451688 | PASSED | FAILED |
| | 730451828 | FAILED | |
| | 730451851 | FAILED | |
| 7D | 272288939 | FAILED | |
| | 274609747 | FAILED | |
| | 274989464 | FAILED | |
| | 274989468 | PASSED | FAILED |

List of primers used for KASP genotyping, including chromosome location, physical position, and genotype call concordance with sequencing data. Markers were classified as “PASSED” if the KASP calls matched (>90% concordance) the expected genotypes derived from sequencing data in either the parental or progeny lines, and “FAILED” if they did not.

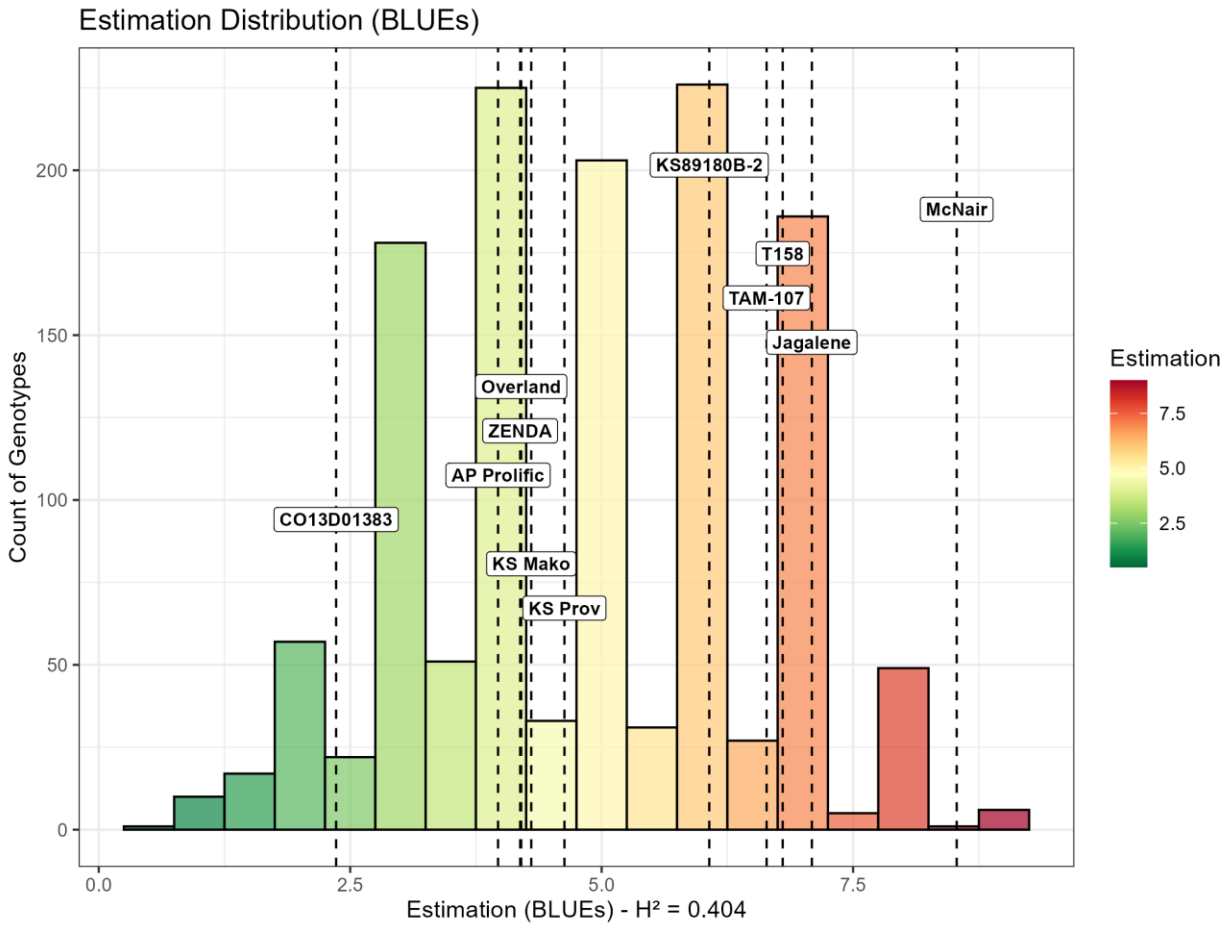



**Figure 2.10 Histogram of the best linear unbiased estimations (BLUEs) for BYDV severity in validation set in Hutchinson, KS (2024).**

The distribution of BLUEs is shown, with dashed vertical lines marking check genotypes. The color gradient represents the variation in estimated values, with green indicating lower severity and red indicating higher severity. The broad-sense heritability ($H^2$ = 0.404) is displayed at the bottom of the plot.

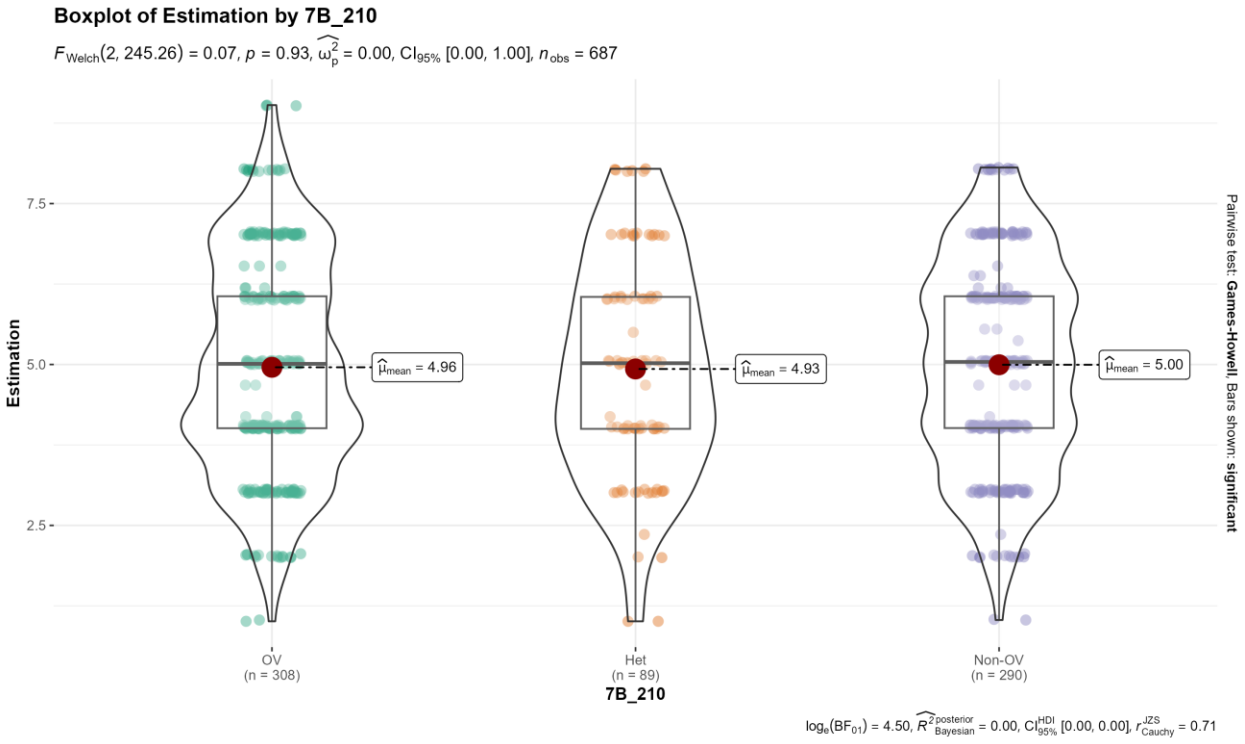



**Figure 2.11 Distribution of BYDV severity based on genotype calls with the developed KASP marker 7B:730438210.**

The violin and boxplots represent the variation in BYDV severity estimation across the three genotypic categories OV (Overland allele positive); Het (Heterozygous) and non-OV (Negative). The red dot denotes the mean estimation per group, with its value indicated. Welch's ANOVA test statistics are displayed above the plot.

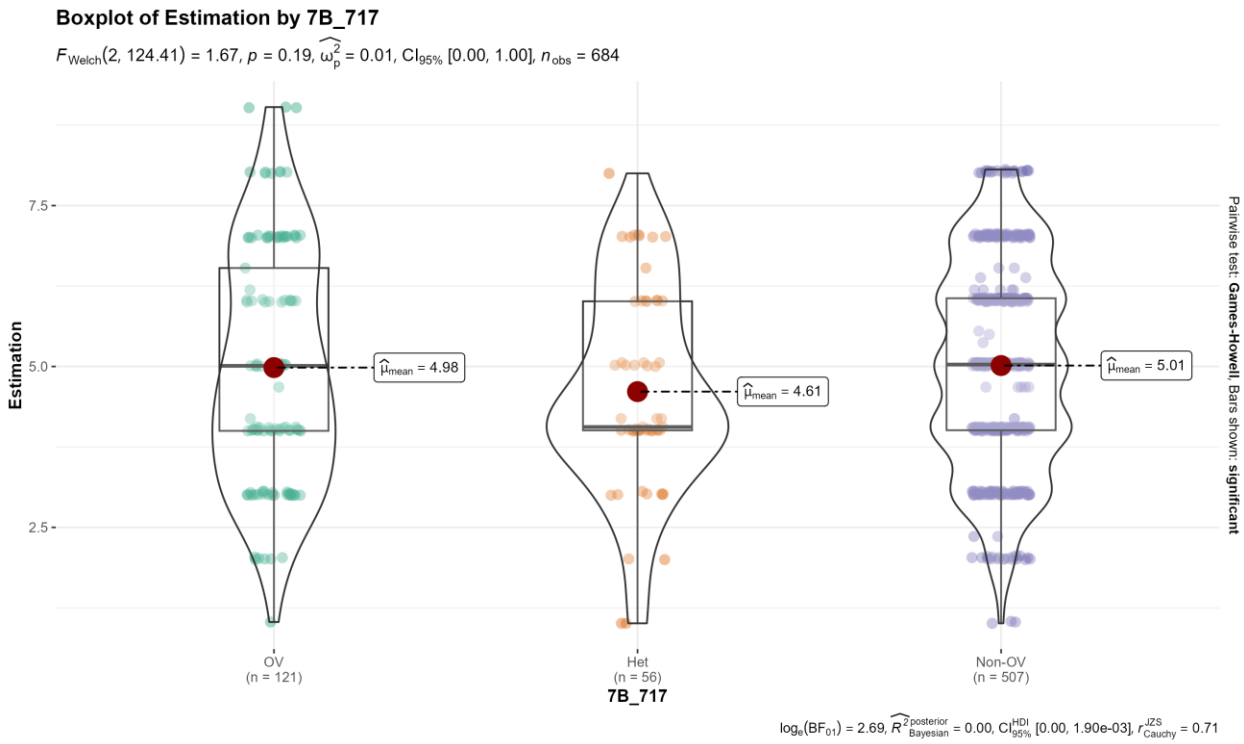



**Figure 2.12 Distribution of BYDV severity based on genotype calls with the developed KASP marker 7B:730434717.**

The violin and boxplots represent the variation in BYDV severity estimation across the three genotypic categories OV (Overland allele positive); Het (Heterozygous) and non-OV (Negative). The red dot denotes the mean estimation per group, with its value indicated. Welch's ANOVA test statistics are displayed above the plot.

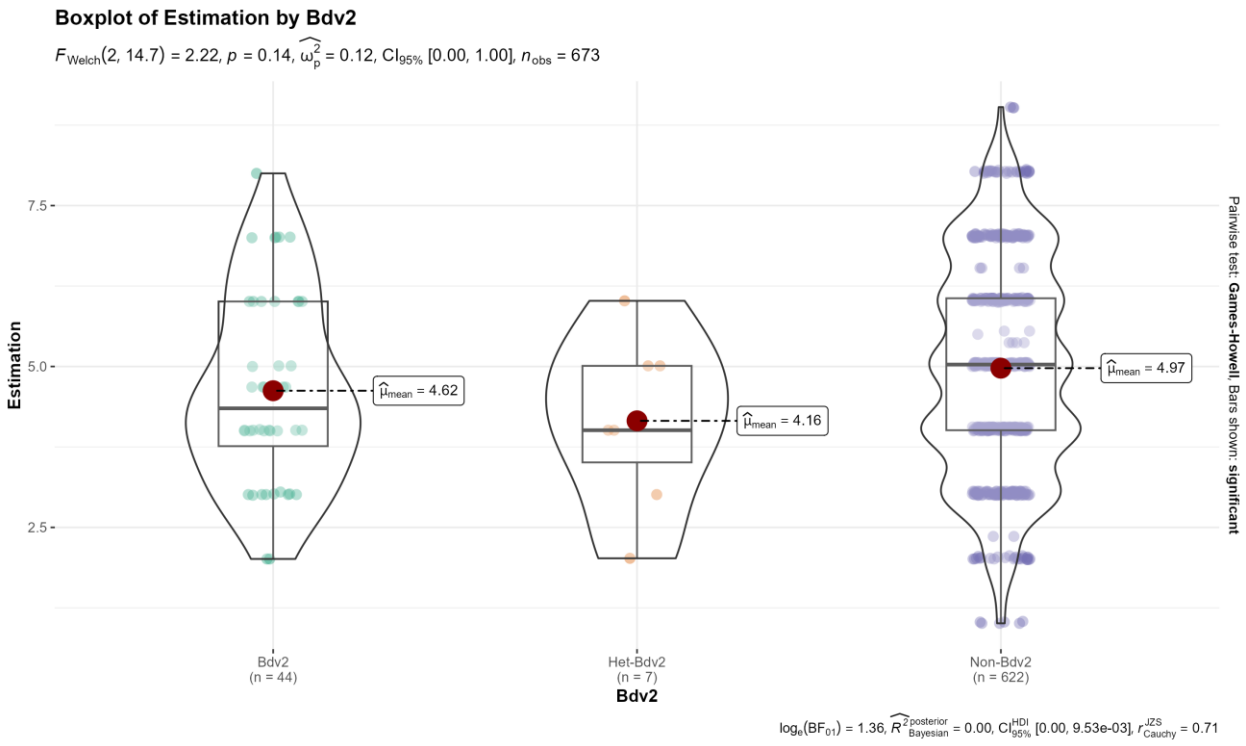



**Figure 2.13 Distribution of BYDV severity based on genotype calls with Bdv2 KASP markers**

Violin and boxplots illustrate the variation in estimated BYDV severity scores among three genotypic groups: *Bdv2* (Positive); Het-*Bdv2* (Heterozygous) and non-*Bdv2* (Negative). The red dot represents the mean estimation, with its value displayed. Statistical comparisons among groups are included, with a Welch's ANOVA test result presented above the plot.

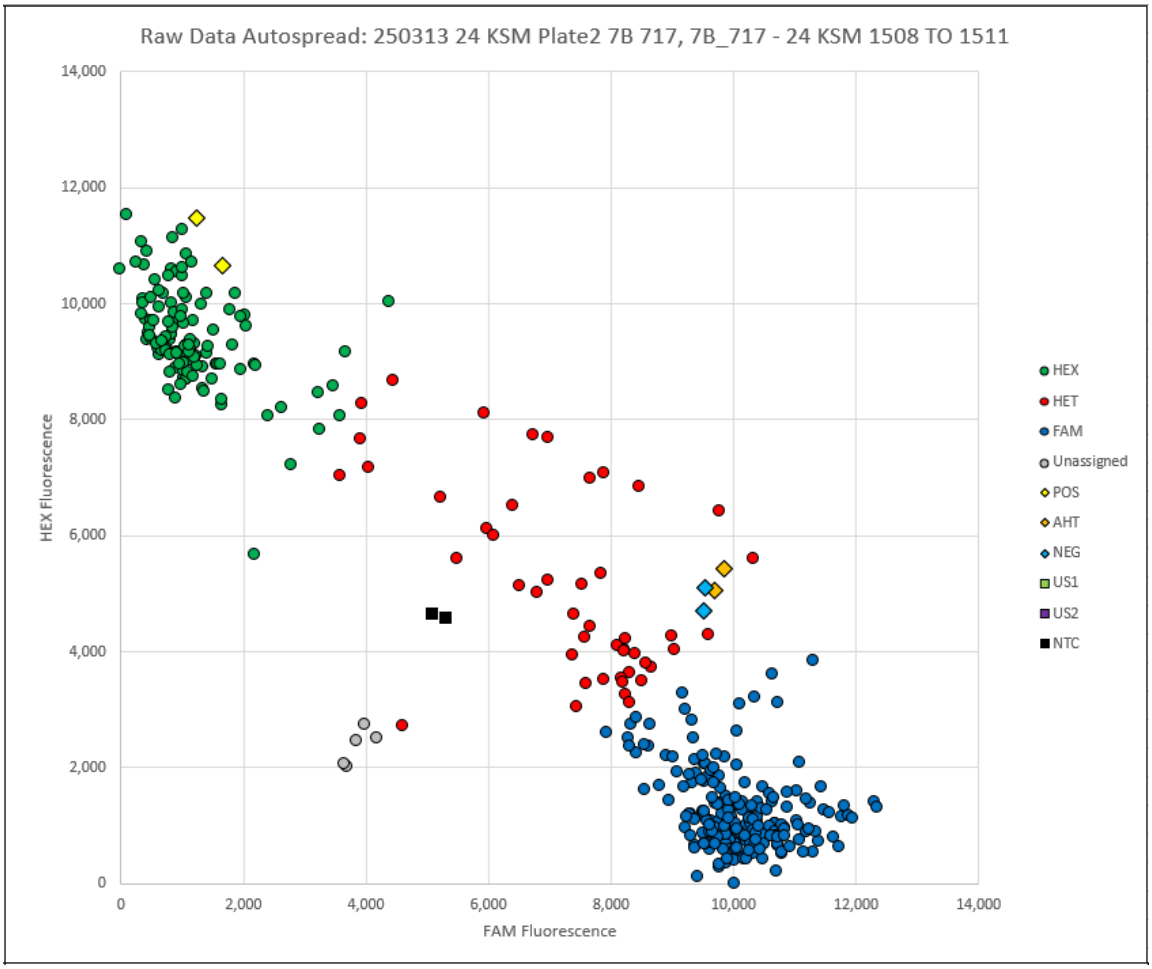



**Figure 2.14 Scatterplot based on the HEX and FAM Fluorescence after KASP assay with 7B:730434717 developed marker tested in the validation set.**

Green points represent an OV (Positive) call; yellow diamonds are Overland DNA as positive controls. Red points are heterozygous calls; Light blue diamonds are negative controls (Alternative parental allele); Mustard diamonds are artificial heterozygous made mixing both positive and negative controls. Black squares are non-template controls. Grey dots are unassign genotypes.

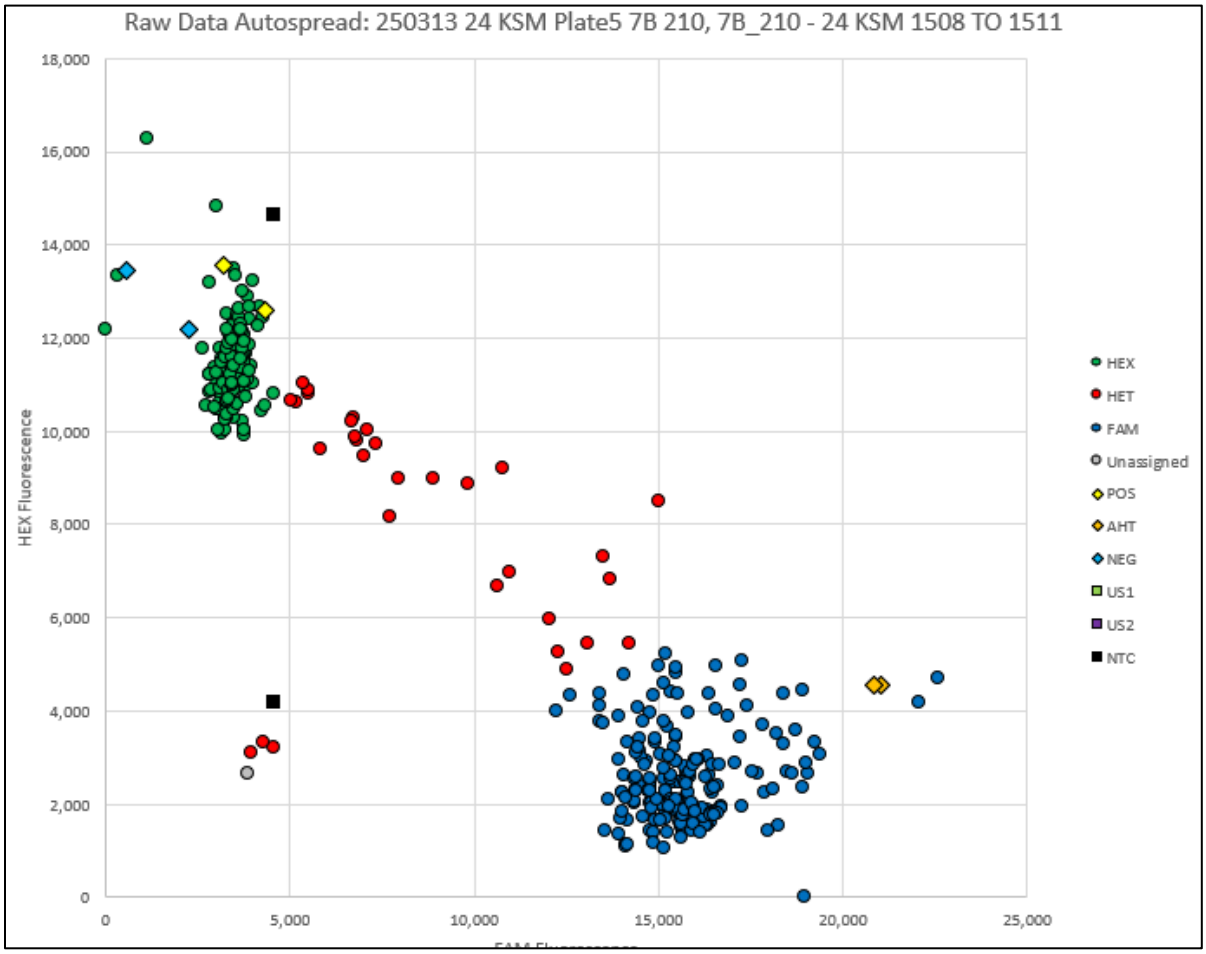
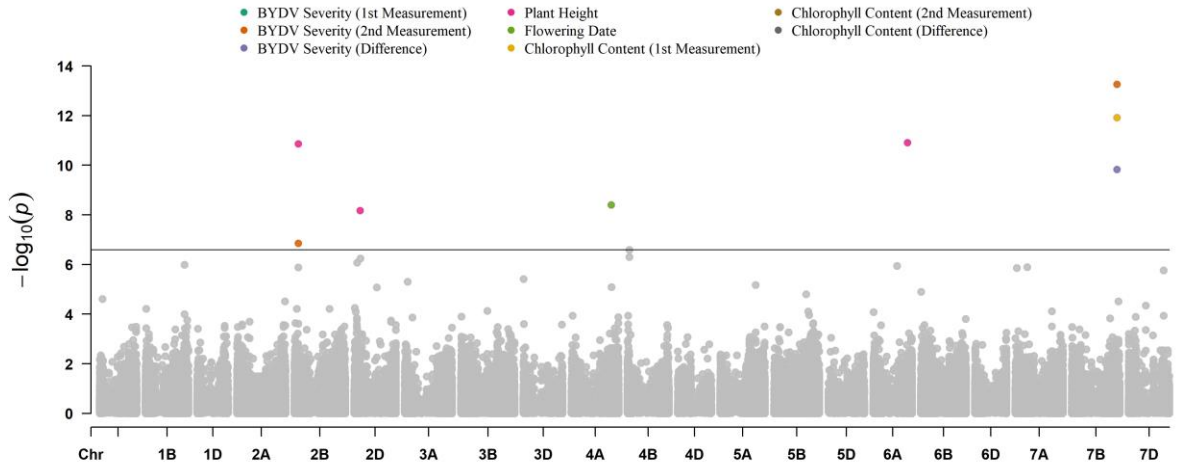



**Figure 2.15 Scatterplot based on the HEX and FAM Fluorescence after KASP assay with 7B:730438210 developed marker tested in the validation set.**

Green points represent an OV (Positive) call; yellow diamonds are Overland DNA as positive controls. Red points are heterozygous calls; light blue diamonds are negative controls (Alternative parental allele); Mustard diamonds are artificial heterozygous made mixing both positive and negative controls. Black squares are non-template controls. Grey dots are unassign genotypes.

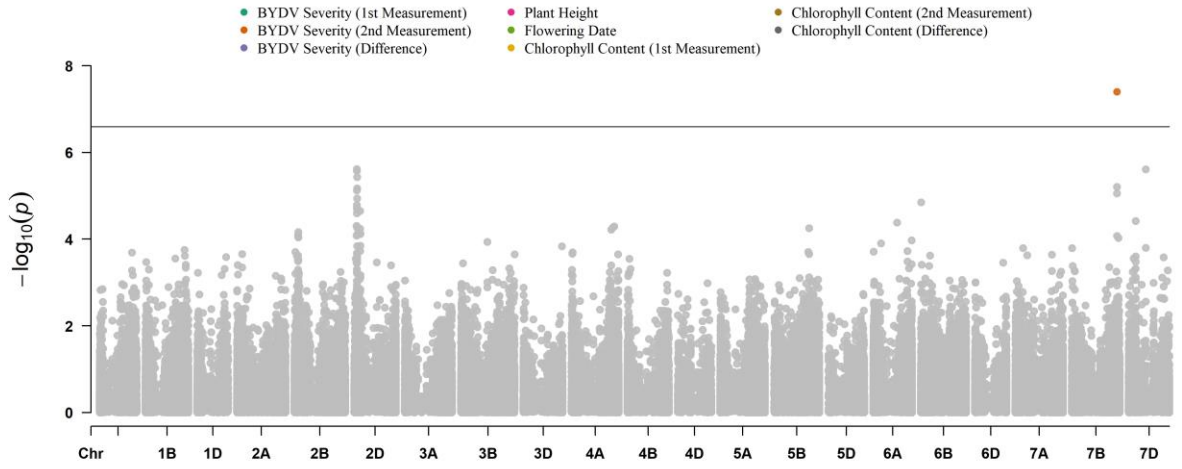
# Appendix A - Supplemental information




**Appendix A Figure A. 1 Manhattan plot displaying SNP-trait associations identified through the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) model.**

The GWAS was performed using a final set of 393,370 SNPs obtained after only applying a minor allele frequency (MAF) filter of 0.05, following the initial family-wide filtering step. Each point represents a SNP, plotted by chromosomal position (x-axis) and -$\log_{10}$(*P*-value) (y-axis). The horizontal black line represents the Bonferroni-adjusted threshold at $\alpha < 0.10$, indicating the significance level used to identify associated loci. Significant SNPs surpassing this threshold are highlighted in different colors based on trait associations. BYDV Severity (1st Measurement) and BYDV Severity (2nd Measurement) refer to visual ratings of Barley Yellow Dwarf Virus symptoms taken at 8 and 18 days after flowering, respectively, while the BYDV Severity (Difference) represents the change in symptom severity over time. Plant Height was measured at maturity from the soil surface to the tip of the spike (excluding awns), and Flowering Date corresponds to the number of days from planting until 10% of the plants reached anthesis.

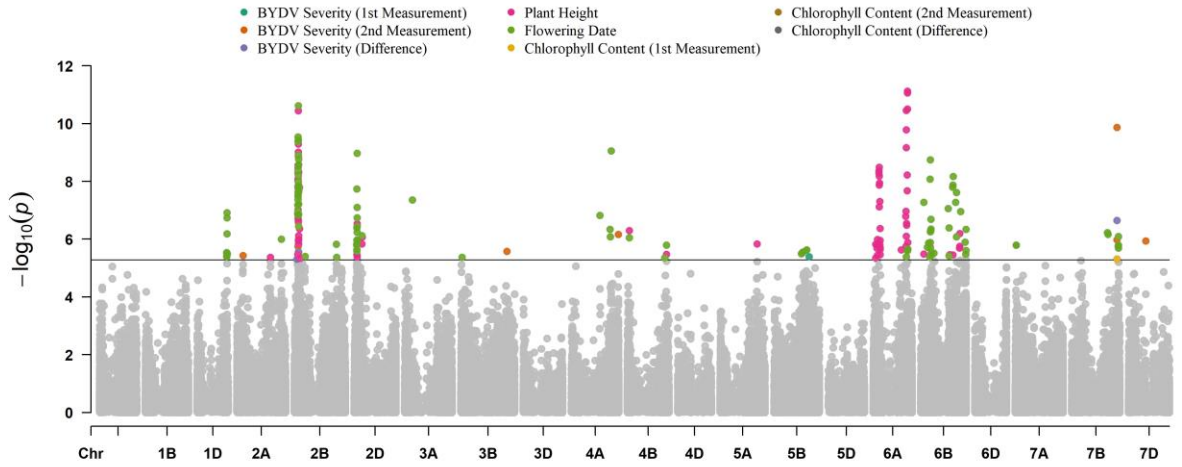
Chlorophyll Content (1st and 2nd Measurements) was assessed using a SPAD meter at 8 and 18 days after flowering, respectively, and their difference was used to capture changes in leaf greenness over time. The plot highlights loci significantly associated with variation in these agronomic and physiological traits under BYDV pressure.




**Appendix A Figure A. 2 Manhattan plot of genome-wide association study (GWAS) results using MLMM.**

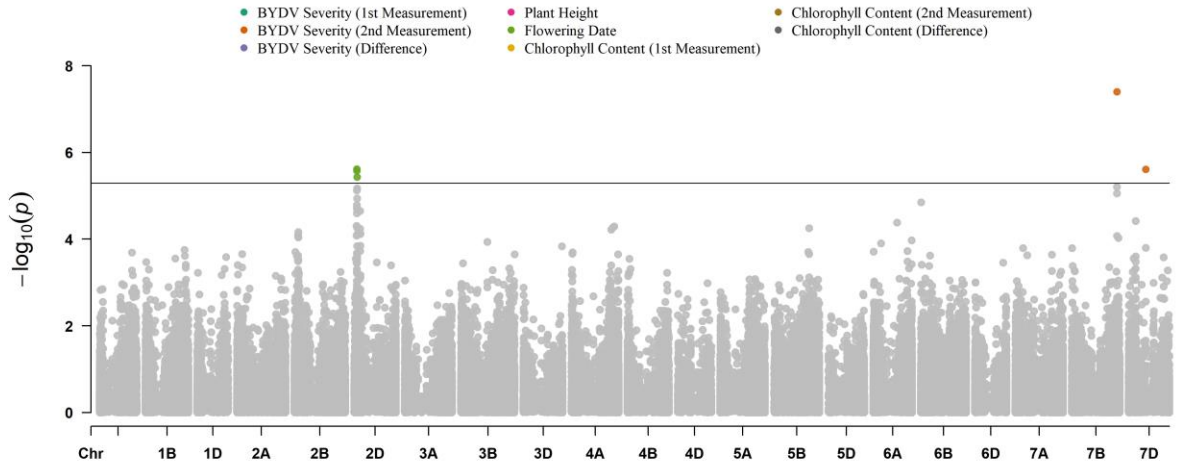
Manhattan plot displaying SNP-trait associations identified through the Multi-Locus Mixed Model (MLMM). The GWAS was performed using a final set of 393,370 SNPs obtained after only applying a minor allele frequency (MAF) filter of 0.05, following the initial family-wide filtering step. Each point represents a SNP, plotted by chromosomal position (x-axis) and -$\log_{10}$(*P*-value) (y-axis). The horizontal black line represents the Bonferroni-adjusted threshold at $\alpha < 0.10$, indicating the significance level used to identify associated loci. Significant SNPs surpassing this threshold are highlighted in different colors based on trait associations. BYDV Severity (1st Measurement) and BYDV Severity (2nd Measurement) refer to visual ratings of Barley Yellow Dwarf Virus symptoms taken at 8 and 18 days after flowering, respectively, while the BYDV Severity (Difference) represents the change in symptom severity over time. Plant Height was measured at maturity from the soil surface to the tip of the spike (excluding awns), and Flowering Date corresponds to the number of days from planting until 10% of the plants reached anthesis. Chlorophyll Content (1st and 2nd Measurements) was assessed using a SPAD

meter at 8 and 18 days after flowering, respectively, and their difference was used to capture changes in leaf greenness over time. The plot highlights loci significantly associated with variation in these agronomic and physiological traits under BYDV pressure.




**Appendix A Figure A. 3 Manhattan plot of genome-wide association study (GWAS) results using GLM.**

Manhattan plot displaying SNP-trait associations identified through Generalized Linear Model (GLM). Each point represents a SNP, plotted by chromosomal position (x-axis) and -$\log_{10}$(*P*-value) (y-axis). The horizontal black line represents the Bonferroni-adjusted threshold at $\alpha <$ 0.10, indicating the significance level used to identify associated loci. Significant SNPs surpassing this threshold are highlighted in different colors based on trait associations. BYDV Severity (1st Measurement) and BYDV Severity (2nd Measurement) refer to visual ratings of Barley Yellow Dwarf Virus symptoms taken at 8 and 18 days after flowering, respectively, while the BYDV Severity (Difference) represents the change in symptom severity over time. Plant Height was measured at maturity from the soil surface to the tip of the spike (excluding awns), and Flowering Date corresponds to the number of days from planting until 10% of the plants reached anthesis. Chlorophyll Content (1st and 2nd Measurements) was assessed using a SPAD meter at 8 and 18 days after flowering, respectively, and their difference was used to capture changes in leaf greenness over time. The plot highlights loci significantly associated with variation in these agronomic and physiological traits under BYDV pressure.




**Appendix A Figure A. 4 Manhattan plot of genome-wide association study (GWAS) results using MLM.**

Manhattan plot displaying SNP-trait associations identified through Mixed Linear Model (MLM). Each point represents a SNP, plotted by chromosomal position (x-axis) and -$\log_{10}$(*P*-value) (y-axis). The horizontal black line represents the Bonferroni-adjusted threshold at $\alpha <$ 0.10, indicating the significance level used to identify associated loci. Significant SNPs surpassing this threshold are highlighted in different colors based on trait associations. BYDV Severity (1st Measurement) and BYDV Severity (2nd Measurement) refer to visual ratings of Barley Yellow Dwarf Virus symptoms taken at 8 and 18 days after flowering, respectively, while the BYDV Severity (Difference) represents the change in symptom severity over time. Plant Height was measured at maturity from the soil surface to the tip of the spike (excluding awns), and Flowering Date corresponds to the number of days from planting until 10% of the plants reached anthesis. Chlorophyll Content (1st and 2nd Measurements) was assessed using a SPAD meter at 8 and 18 days after flowering, respectively, and their difference was used to capture changes in leaf greenness over time. The plot highlights loci significantly associated with variation in these agronomic and physiological traits under BYDV pressure.

**Appendix A Table A. 1. Summary of significant SNP-trait associations detected using MLM.**

| Trait | Chr | Physical Position | Parental origin | Effect | Min *P*-value |
|---|---|---|---|---|---|
| BYDV Second Measurement (0-9 score) | 7B | 730451609 | non-OV | 0.847 | 3.99E-08 |
| | 7D | 567312804 | non-OV | 0.510 | 2.48E-06 |
| Flowering date (days) | 2B | 60599283 | non-OV | -0.411 | 8.05E-05 |
| | 2B | 63688651 | non-OV | -0.456 | 6.88E-05 |
| | 2D | 41160034 | OV | 0.358 | 4.95E-05 |
| | 2D | 41572849 | non-OV | -0.365 | 8.28E-05 |
| | 2D | 42000453 | OV | 0.442 | 2.44E-06 |
| | 2D | 44499356 | non-OV | -0.388 | 1.91E-05 |
| | 2D | 44720168 | OV | 0.420 | 7.54E-06 |
| | 2D | 44807394 | non-OV | -0.401 | 1.63E-05 |
| | 2D | 45169853 | OV | 0.425 | 2.65E-06 |
| | 2D | 45497716 | OV | 0.380 | 2.56E-05 |
| | 2D | 47092124 | non-OV | -0.426 | 6.75E-06 |
| | 2D | 47435326 | OV | 0.372 | 5.52E-05 |
| | 2D | 49029785 | OV | 0.371 | 7.77E-05 |
| | 2D | 49161244 | OV | 0.404 | 2.08E-05 |
| | 2D | 49762185 | OV | 0.367 | 5.35E-05 |
| | 2D | 49944668 | non-OV | -0.430 | 3.72E-06 |
| | 2D | 50187302 | non-OV | -0.398 | 1.15E-05 |
| | 4A | 639917966 | OV | 0.357 | 6.08E-05 |
| | 4A | 666029690 | OV | 0.304 | 5.44E-05 |

Significant marker-trait associations identified for BYDV severity (second measurement, 18 days after flowering) and flowering date are shown. The chromosome (Chr), physical position, effect size, and minimum *P*-value for each association are reported. SNPs were detected using

MLM, with the Bonferroni-adjusted threshold at $\alpha < 0.10$ applied to determine significance. A column indicating the parental origin of the associated allele is also included: "OV" denotes the allele originated from the recurrent parent Overland, while "non-OV" indicates the allele came from one of the ten diverse PI parental lines. Larger effect sizes indicate stronger associations, while lower *P*-values suggest greater statistical support for the SNP-trait relationship.

**Appendix A Table A. 2 Kompetitive Allele Specific PCR (KASP) Marker Details for BYDV-Associated SNPs on Chromosome 7B.**

| Marker | Chr | Physical Position | HEX Marker Primer | FAM Marker Primer | Common Marker Primer | HEX $T_m$ (°C) | FAM $T_m$ (°C) | Common $T_m$ (°C) |
|---|---|---|---|---|---|---|---|---|
| 7B:730434717 | 7B | 730434717 | GAAGGTCGGAGTCAACGGATTtGtatgtggTgcgcttattGttaT | GAAGGTGACCAAGTTCATGCTtGtatgtggTgcgcttattGttaC | cAgggcaatgaaagcatctcc | 59.19 | 59.82 | 58.75 |
| 7B: 730438210 | 7B | 730438210 | GAAGGTCGGAGTCAACGGATTtgaggaagaatatagccacctcG | GAAGGTGACCAAGTTCATGCTtgaggaagaatatagccacctcA | ggcttgctggagttcggatt | 59.12 | 59.91 | 59.59 |

This table provides the sequence and melting temperature information for KASP markers developed from two SNPs significantly associated with BYDV resistance on chromosome 7B. Each marker includes allele-specific primers labeled with HEX and FAM fluorophores, a common reverse primer, and the corresponding melting temperatures ($T_m$) in degrees Celsius. The primer sequences are formatted for compatibility with KASP genotyping platform.