ABSTRACT

Title of Dissertation: **GENOMICS ENABLED GENE DISCOVERY IN DIPLOID AND POLYPLOID WHEAT**

Inderjit Singh Yadav, Doctor of Philosophy,2024

Dissertation directed by: Dr. Vijay Tiwari,
Assistant Professor,
Genomics of Small Grain Cereal Crops,
Department of Plant Science and Landscape Architecture

Hexaploid bread wheat (*Triticum aestivum*) is one of the most important staple food crops for humans. Sustainable genetic improvement in wheat is critical for ensuring global food security and requires the introduction of new genes and alleles into elite wheat cultivars. The progenitor species and wild wheat relatives are a reservoir of genetic diversity for wheat improvement. This doctoral thesis demonstrates the application of genomic resources and bioinformatics pipelines to characterize the wild germplasm and to streamline the gene discovery pipeline using five diverse species involving wheat progenitor, wild, and related species. Genomics-assisted characterization of the genetic diversity present in gene banks is a major step towards the systematic utilization of unexploited germplasm to ensure the sustainable development of new varieties. Toward this end, we used genomics datasets to curate wild and related accessions of tetraploid wheat from two distinct species *Triticum turgidum* and *Triticum timopheevii.* Using Genotyping by sequencing

(GBS) data and a unique similarity matrix and powercore analysis, a set of 102 accessions were identified as the core set accessions that represent 20 and 35 percent of the total accessions of the WGRC tetraploid wheat collection of *T. turgidum* and *T. timopheevii,* respectively. Further, three distinct centers of rich genetic diversity were identified for wild and domesticated emmer and *T. timopheevii* in the Fertile Crescent. GWAS analysis of the genotypic and phenotypic dataset identified a novel QTL for leaf rust resistance on chromosome 2B in *T. timopheevii*.

Triticale is a man-made cereal derived from a cross between tetraploid and hexaploid wheat with diploid rye. There are large numbers of triticale germplasm available in different gene banks; however, in many cases, the ploidy information is not accurate and affects the quality of work with large triticale germplasm. In this work, using the low-cost GBS datasets, a pipeline was developed to detect contamination in the UMD triticale collection and facilitated the accurate classification of ploidy, ensuring the purity of the triticale germplasm. This approach identified contamination of 11 wheat accessions and enabled the correct classification of 236 hexaploid and 12 octoploid triticale, these results were further confirmed through GISH experiments.

Wild and related germplasms are considered as the goldmine of genetic diversity for wheat improvement. The modern wheat cultivars have gone through several rounds of heavy selections for yield related traits and have lost the genetic diversity against several abiotic and biotic stresses. On the other hand, wild relatives of wheat have been growing naturally without any substantial artificial selection pressure and it allowed them to preserve their genetic diversity. This study investigates the genetic diversity of a selected set of genes to visualize the differences in wild wheat relatives and polyploid wheat cultivars. To study these differences, group 5 chromosome of *Aegilops geniculata* and *Aegilops umbellulata*, belonging to the tertiary gene pool, were assembled. Comparative analysis revealed a higher rate of pseudogenization in bread wheat

compared to these two wild relatives, primarily due to the difference in exon/intron length between the genes, rendering these genes non-functional.

Diploid einkorn wheat (*Triticum monococcum*), with inherent disease resistance, offers a valuable resource for wheat improvement. To facilitate its proper utilization, two of the reference genomes-one wild (*T. monococcum* ssp. *aegilopoides*) and one domesticated (*T. monococcum* ssp. *monococcum*) were assembled in the study. Kmer-GWAS identified seven novel QTLs associated with powdery mildew resistance, three for leaf rust resistance, and two for stem rust resistance. These QTLs harbor diverse gene classes encoding for resistance gene analogs, cysteine-rich receptor kinases, transcription factors, and G-type lectins. Overall, the knowledge and resources developed in this research would contribute to the characterization of vast germplasm and the development of climate-resilient wheat.

GENOMICS ENABLED GENE DISCOVERY IN DIPLOID AND POLYPLOID WHEAT

by

Inderjit Singh Yadav

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Advisory Committee:
Dr. Vijay Tiwari, Chair
Dr. James Culver
Dr. Nidhi Rawat
Dr. Yiping Qi
Dr. Abani Pradhan

Dedication

I dedicate this dissertation to my parents. Their support and belief in me have fueled my passion to address the problem of global hunger.

Acknowledgments

First and foremost, I would like to express my gratitude to Dr. Vijay Tiwari for giving me this opportunity to work on a multidisciplinary and impactful research project. Thank you for your guidance and continuous support throughout this journey.

I am also grateful to my esteemed committee members, Dr. James Culver, Dr. Nidhi Rawat, Dr. Yiping Qi, and Dr. Abani Pradhan. Their guidance and support have been invaluable, enriching my understanding of various methods and concepts.

Many thanks to present and past members of Tiwari Lab and Rawat Lab, especially Anmol, Bhavit, Lovepreet, Adam, Gautam, Sydney, Taylor, and Janelle, for constant companionship and collaboration and for making the lab environment a welcoming and encouraging workplace. I am thankful to the undergraduate interns in the lab for their help in harvesting, threshing, and phenotyping.

I am very grateful to our collaborators Dr. Javier Martin, Dr. James Kolmer, Dr. Matthew N Rouse, Dr. Dal-Hoe Koo, Dr. Thomas Lux, Dr. Manuel Spannagl, Dr. Michael Abrouk, Dr. Jesse Poland, Dr. Simon G. Krattinger and all others who have contributed their expertise, and support to this endeavor. Your valuable contributions have enriched the scope, depth, and success of my research.

I extend my heartfelt gratitude to Dr. Parveen Chhuneja and Punjab Agricultural University for granting me study leave, which enabled me to pursue this research endeavor. I am thankful to the Indian Council of Agricultural Research (ICAR), New Delhi, for awarding me the Netaji Subhas - ICAR International scholarship, which provided invaluable support during this academic journey.

I wish to express my gratitude to Dr. Erwin and all staff members for sustaining a welcoming and respectful professional environment.

Some individuals have profoundly impacted my life. I extend my heartfelt thanks to Anshul, Anoop, Pawan, Naveen, Prashant, Amandeep, Gurupkar, and my other colleagues at PAU.

Finally, I would like to extend my deepest gratitude to my family. My parents and brother have been my unwavering anchor through both the highs and lows of this journey. Their love, support, and encouragement have been invaluable. I am also grateful to every person who has played a part in my academic and personal journey.

Inderjit

Contributions to this Dissertation

Chapter 3.

Dr. Parveen Chhuneja provided flow-sorted chromosome sequence of $5U^u$ and RNAseq of RILs.

Chapter 4.

Dr. Xue-Feng Ma provided the triticale genotyping data.

Dr. Dal-Hoe Koo for performing GISH analysis.

Chapter 5.

Dr. Thomas Lux and Dr. Manuel Spannagl performed the repeat annotation and gene prediction.

Chapter 6.

Dr. Beat Keller and Dr. Javier Sánchez Martín provided the powdery mildew phenotype data.

Dr. James Kolmer provided the leaf rust phenotype data

Dr. Matthew N Rouse provided the stem rust phenotype data

Dr Gautam Saripalli performed the interval mapping of the RILs population.

Table of Contents

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Figures

Chapter 2

Chapter 3

Chapter 6

Chapter 1: Introduction

1.1 Wheat

Wheat (*Triticum aestivum* L), a member of the *Poaceae* family, serves as the cornerstone of global food security. It provides 40% dosage of the daily calorie intake for humans worldwide (Venske et al., 2019). With a projected global population of 9.8 billion by 2050 (https://www.un.org), the pressure to ensure adequate wheat production is immense. The projected output of 784.9 million metric tons for 2023–2024 is expected to be insufficient to meet the estimated 796.4 million metric tons of global demand (https://www.statista.com/). This gap highlights the urgency of finding solutions to bridge the supply and demand for this essential crop.

1.2 Origin of wheat

Domestication of wheat occurred in the Fertile Crescent around 10,000 ago and paved the pace for stable human settlements and civilizations from the hunter-gatherers. The modern hexaploid wheat has evolved through a complex evolutionary journey characterized by two distinguished hybridization events involving three different diploid progenitor species. The first hybridization involved the association of *Triticum urartu* (AA,2n=2x=14) and an unknown extinct species from the *Sitopsis* section that contributed to the B genome. *Aegilops speltoides* (SS, 2n=2x=14) is considered the closest relative to the donor of B-genome in wheat. This hybridization resulted in allotetraploid wheat, known as wild emmer wheat (*Triticum dicoccum*; AABB, 2n = 4x = 28). A subsequent hybridization took place between *Triticum dicoccum* and *Aegilops tauschii* (DD, 2n=2x=14), a wild diploid known as goat grass,

resulting in the development of the allohexaploid (*Triticum aestivum*, AABBDD, 2n = 6x = 42). The polyploidization led to the development of a very large wheat genome (~17Gb) and was characterized by a high percentage of repetitive elements accounting for nearly 85% of the wheat genome.(Appels et al., 2018; Wicker et al., 2018)

1.3 Genetic erosion in modern cultivars

Polyploidization events involving a limited number of ancestral species led to the genetic bottleneck and gene loss (Dubcovsky & Dvorak, 2007; Nazco et al., 2012). The D genome of hexaploid wheat has significantly less genetic diversity than its progenitor, likely due to the bottleneck effect associated with polyploidization. Further, the present-day wheat has undergone a significant genetic erosion driven by modern agricultural practices climate change, and socioeconomic pressure (Arzani & Ashraf, 2017; Fayaz et al., 2019; Khoury et al., 2022). With the discovery of dwarfing genes and the development of new high-yielding semi-dwarf cultivars led to a significant decline in the use of local landraces in the global wheat breeding program (Soriano et al., 2016), which further accelerated the loss of genetic diversity in the modern cultivars (S. Rahman et al., 2020). Wheat faces continuous threats from changing climatic conditions and ever evolving biotic and abiotic stresses (Enghiad et al., 2017). Rust diseases of wheat poses a major challenge as they yield losses up to 30 percent (Juliana et al., 2018). Powdery mildew is estimated to cause a worldwide yield loss of nearly 5% annually (Savary et al., 2019; R. P. Singh et al., 2016). Wheat production needs to increase by 60 to 70% to feed 10 billion people by 2050. Modern agriculture faces a dual challenge: mitigating yield losses from biotic and abiotic stresses while simultaneously increasing production to meet the demands of a burgeoning population.

This pressure is particularly acute in developing countries, where wheat production needs to double just to keep pace with population growth (Ray et al., 2013)

1.4 Wheat gene pool

Wild and related wheat germplasm is considered the gold mine for genetic diversity and can help recharge the already depleted pool of modern wheat cultivars. Gene introgression of favorable alleles and gene/gene complexes from wild relatives offers an excellent and efficient opportunity to enhance the genetic base of the cultivated gene pool for various desirable traits. The wheat gene pool is divided into three distinct categories based on the genetic makeup of its constituent species: primary, secondary, and tertiary. The primary gene pool comprises closely related species with shared genomes (diploid and tetraploid progenitors). The secondary gene pool comprises species sharing one of the genomes with wheat and consists of polyploid *Triticum* and *Aegilops* species. Species carrying non-homologous genomes constitute the tertiary gene pool pool (Chaudhary et al., 2014). The progenitor species and wild germplasm co-existed for millions of years in the Fertile Crescent region and accumulated variations through mutations and natural introgressions between the members of the different gene pools (Vardi & Zohary, 1967).

1.5 Gene bank collections

The genetic diversity of the crops has been valued, and multiple gene banks have been established worldwide to preserve these resource (McCouch et al., 2020). For instance, the gene bank at the University of Maryland preserves an extensive collection of wheat gene pools. In recent years, gene bank curators have explored novel genetic variations in the *Poaceae* family, including *T. monococcum* (Adhikari et al.,

2022), *T. aestivum* (Balfourier et al., 2007; Hao et al., 2006, 2008; Pascual et al., 2020), *Ae. tauschii* (N. Singh et al., 2019) and *T. turgidum* (Ruiz et al., 2013), for trait improvement. Progenitor species and alien germplasm have contributed valuable agronomic traits to wheat, including enhanced grain quality (Xie & Nevo, 2008) and resistance to biotic stresses (Xie & Nevo, 2008).

1.6 Genomics resources in wheat

As a critical global food source, wheat has been the focus of extensive research efforts to understand its genetic makeup. The International Wheat Genome Sequencing Consortium (IWGSC, 2014), a collaboration of 73 research institutions from 20 countries, successfully assembled the first reference genome of wheat using a BAC-by-BAC approach (Marcussen et al., 2014) following the draft wheat genome (The International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018). In recent years, several wheat relatives, *Ae. tauschii* (Jia et al., 2013), wild emmer (Avni et al., 2017; Dvorak et al., 2018; Maccaferri et al., 2019; Zhu et al., 2019), *T. urartu* (Ling et al., 2018), *T. monococcum* (Ahmed et al., 2023) have been sequenced and are available to the scientific community. The availability of these genomic sequences, along with the development of advanced tools in functional genomics like transcriptomics, proteomics, and metabolomics, provides researchers with a powerful toolkit to unravel the complex genetic architecture of wheat traits. This comprehensive understanding of the wheat genome and its relatives paves the way for the development of improved wheat varieties with enhanced agronomic performance, disease resistance, and stress tolerance, ultimately contributing to global food security.

1.7 Genome-wide association studies

Cost-effective next-generation short-read sequencing technologies offer opportunities for the characterization, classification, and comparative analysis of germplasm for efficient utilization in breeding programs. Crop improvement is a continuous and iterative process that involves the incorporation of new allelic variants found in germplasm collection or the incorporation of *de novo* induced variations (Martínez-Fortún et al., 2022).

Since the first successful study of myocardial infarction in 2002 (Ozaki et al., 2002), genome-wide association study (GWAS) has become a ubiquitous technique in understanding the complex traits in plants. It has been widely used for investigating agriculturally important traits in crops like rice, soybean, sorghum, barley, cotton, wheat, and many others crops (Ersoz et al., 2007; H. J. Liu & Yan, 2019; Sukumaran & Yu, 2014). In recent years, GWAS have been successfully applied to investigate the resistance against devastating diseases such as stem rust and leaf rust (Aoun et al., 2021; Genievskaya et al., 2022; Saccomanno et al., 2018), stripe rust (W. Liu et al., 2017), FHB (Sari et al., 2020), and powdery mildew resistance (Simeone et al., 2020) in tetraploid wheat. While SNP-based Genome-Wide Association Studies (GWAS) have been a powerful tool for identifying genetic variants associated with traits, they have limitations. One major limitation is the requirement for a high-quality reference genome. Which is problematic for species with no reference genome. Additionally, SNP-based GWAS primarily focus on single nucleotide polymorphisms and may miss other important variations, such as insertions/deletions (indels) and copy number variations (CNVs). Furthermore, sequencing errors can also introduce noise in the

analysis. Alignment-free methods offer a promising alternative, adaptable to a variety of research scenarios. Their independence from a reference genome makes them particularly valuable for studying non-model organisms or those with limited genomic resources, providing reassurance about their applicability. This method has been incredibly useful in detecting rare disease variants. Some examples of tools include HAWK (A. Rahman et al., 2018) and kmdiff (Lemane et al., 2022). The method uses a string of sequences of defined length known as k-mer rather than the SNP as a marker for association. Several tools are available in plants for k-mer based association mapping i.e., kmersGWAS (Voichek & Weigel, 2020), and kGWAS (Gaurav et al., 2022). Gaurav et al. (2022) successfully utilized the kGWAS method in *Ae. tauschii* to identify genes related to disease resistance and quality traits (Gaurav et al., 2022). The method is efficient and highly applicable to polyploid crops such as wheat.

The constant emergence of novel pathotype/races and increased pest pressure, coupled with declining yield components due to climate change, necessitate the urgent development of climate-resilient crops to ensure global food security. Hexaploid wheat can significantly benefit from gene discovery efforts from its wild relatives. These distant relatives harbor a more comprehensive range of genetic diversity, potentially containing genes for disease resistance and stress tolerance not yet found in hexaploid wheat. By leveraging advancements in genomic technologies, researchers can identify and transfer these beneficial genes into hexaploid wheat varieties, ultimately developing more robust and productive crops.

Dissertation overview

The dissertation focuses on developing computational methodology and pipelines for the efficient management of biological germplasm and their utilization in novel gene discovery for biotic stresses. The project will investigate the natural sources of variation in the progenitor species and wild wheat germplasm for trait improvement using bioinformatics methods and tools. The pipelines developed in the study will help in the management and categorization of germplasm. The study aims to establish the diploid einkorn wheat as a model for trait discovery in the *Poaceae* family. The germplasm studied includes five different species from the gene pool i.e., wild and domesticated emmer wheat (*T. turgidum*), *T. timopheevii*, *Ae. geniculata*, *Ae. umbellulata*, manmade cereal triticale, wild and domesticated *T. monococcum.*

Characterization of the germplasm is critical for proper management and their utilization in the breeding program. The redundancy found in gene bank collections creates challenges for management and long-term storage. Chapter 2 characterizes 908 accessions of *T. turgidum* and *T. timopheevii* from the University of Maryland gene bank collection into a set of manageable accessions using the cost-effective GBS technology. The genetic diversity of the two species was explored to detect the center of rich genetic diversity or center of origin and their potential route of migration into European continent. The content of chapter 2 was published in the "Journal of Advance Research" (Yadav et al., 2023).

Several studies have introduced genes from the tertiary gene pool to the elite cultivars, adding novel sources of resistance or substituting inefficient/inactive ortholog in polyploid wheat. Comparative genome analysis is required for the proper

characterization of these genes. Leaf rust resistance (*Lr57)* gene has been introduced in the bread wheat from *Ae. geniculata* on chromosome 5D. Chapter 3 deals with the comparative genomic analysis of group 5 chromosomes of *Ae. geniculata* ($5M^g$), *Ae. umbellulata* ($5U^u$) and *T. aestivum* cv. Chinese Spring (5D) to study the introgressed locus *Lr57*. The study focused on the difference in the genetic makeup (gene structure) of cultivated wheat and wild relatives using orthologous genes. The content of chapter 3 was published in the "Frontiers in Plant Science" (Yadav et al., 2023).

The effectiveness of vast germplasm in breeding programs is compromised by the presence of admixture, contamination of related species and misclassification/nomenclature errors of the original collections. Present-day triticale represents second-generation hybrids developed from the cross of two primary triticale, with chromosomal substitutions/translocations. Chapter 4 addresses these challenges through the development of computational pipeline using GBS data. The pipeline efficiently classifies and determines the ploidy level in triticale germplasm (270 accessions). The use of GBS offers an effective method over labor-intensive techniques like *in-situ* hybridization. The methodology developed can detect the ploidy, large and small introgressions, and chromosomal substitutions and can also be extended to other germplasm collections with contamination.

High-quality reference genome assemblies are essential for precise gene identification and characterization and facilitates the discovery of genes associated with beneficial traits for crop improvement. Chapter 5 focuses on the development of reference genome for two diverse einkorn wheat accessions: one wild (*T. monococcum* L. subsp. *aegilopoides*; TA391) and one domesticated (*T. monococcum* L. subsp.

monococcum; TA10868) and their comparative analysis with the wheat genome. Comparison of the assembled genome was done with other progenitor species to study the evolutionary divergence.

Einkorn wheat (*T. monococcum*), with its simple genome architecture compared hexaploid wheat, can serve as a model for discovering and characterizing novel sources of disease resistance. *T. monococcum* exhibits natural resistance to several devastating diseases, including powdery mildew and rust diseases (leaf and stem rust). Chapter 6 deals with the utilization of the TmGWAS (*T. monococcum* GWAS panel) panel of 220 diverse accessions for the discovery of novel sources of disease resistance using the kmer GWAS. The genome assembly developed in Chapter 5 was utilized for the gene discovery. This work demonstrates the application of diverse GWAS panel complemented with reference genomes to discover new genes and alleles for wheat improvement.

Together, these objectives will generate genomic resources for the scientific community, potentially speeding up gene discovery for trait improvement in wheat. The study will deliver novel knowledge about new genes and alleles for disease resistance and enhance our understanding of available genetic diversity in wheat germplasm.

Chapter 2: Exploring Genetic Diversity of Wild and Related Tetraploid Wheat Species *Triticum turgidum* and *Triticum timopheevii*

The work presented in this chapter has been published as:

Yadav, I. S., Singh, N., Wu, S., Raupp, J., Wilson, D. L., Rawat, N., Gill, B. S., Poland, J., & Tiwari, V. K. (2023). Exploring genetic diversity of wild and related tetraploid wheat species *Triticum turgidum* and *Triticum timopheevii*. *Journal of advanced research*, *48*, 47–60. https://doi.org/10.1016/j.jare.2022.08.020

Headings and figure numbers were modified for the consistency within this dissertation. Supplementary information is included in the Appendix. Hyperlinks to the article are included in the main text for supplementary figures and tables.

2.1 Abstract

The domestication bottleneck has reduced genetic diversity in wheat, necessitating the use of wild relatives in breeding programs. Wild tetraploid wheat (*T. turgidum & Triticum timopheevii)* are widely used in the breeding programs but with shared morphological characters, it is difficult to distinguish these, resulting in misclassification/mislabeling or duplication of accessions in the Gene bank. The study aims to explore Genotyping by sequencing (GBS) to characterize 908 wild and domesticated tetraploid wheat accessions to generate a core set of accessions to be used in the breeding program. TASSEL-GBS pipeline was used for SNP discovery. The tool fastStructure was used to determine the population structure. PowerCore was used to generate a core set of accessions. Nucleotide diversity matrices of Nie's and *F*-statistics ($F_{ST)}$ index were used to determine the center of genetic diversity.

Analysis revealed a high proportion of duplicated accessions in both *T. timopheevii* (65%) and *T. turgidum* (47%). Genome-wide nucleotide diversity and F_{ST} scan uncovered a lower intra and higher inter-species differentiation. Distinct F_{ST} regions were identified in genomic regions belonging to domestication genes: non-brittle rachis (*Btr1*) and vernalization (*VRN-1*). Our results suggest that Israel, Jordan, Syria, and Lebanon are the hub of wild emmer genetic diversity; Turkey, and Georgia for *T. durum*; and Iraq, Azerbaijan, and Armenia for the *T. timopheevii*. Identified core set accessions preserved more than 93% of the available genetic diversity. Genome wide association study (GWAS) indicated the potential chromosomal segment for resistance to leaf rust in *T. timopheevii.* The present study explored the potential of GBS technology in data reduction while maintaining the significant genetic diversity

of the species. Wild germplasm showed more differentiation than domesticated accessions, indicating the availability of sufficient diversity for crop improvement. With reduced complexity, the core set preserves the genetic diversity of the genebank collections and will aid in a more robust characterization of wild germplasm.

2.2 Introduction

Wheat is one of the most important crop plants for the human population. Two major wheat species grown worldwide are *T. aestivum*, a hexaploid species usually called "common" or "bread" wheat, and *T. turgidum* ssp. *durum*, a tetraploid species adapted to hot and dry conditions surrounding the Mediterranean Sea and similar climates in other regions. Tetraploid wheat (*T. turgidum* L.) is an important species within the genus *Triticum*, and its numerous subspecies including wild forms harbor many desirable agronomic traits for durum and bread wheat improvement. Hexaploid bread wheat has evolved through two distinct past hybridization events, which included three diploid progenitor species. The wheat genome comprises three sub-genomes namely A, B, and D (2n=6x=42). A genome of the bread wheat was contributed by *T. urartu* (AA), while *Ae. speltoides* (SS) is considered the closest match for B- genome donor. The D-genome was contributed by *Aegilops tauschii* (DD). *T. urartu* contributed A genomes to the tetraploid wheat species *T. turgidum* (BBAA, 2n=4x=28) and *T. timopheevii* (GGAA, 2n=4x=28) (Dvorak et al., 1988, 1993), and *Ae. speltoides* (SS) as maternal parent contributed towards the B and G genomes (Gornicki et al., 2014). These two tetraploid wheat species (*T. turgidum* and *T. timopheevii*) are reproductively isolated (crosses between these are sterile and show reduced chromosome pairing) and belong to distinct evolutionary lineages (Zohary et al., 2012).

Evidence for the first cultivation of wild emmer dates to ~10,300 to 9,500 years ago in the southern Levant region, while the cultivation of domesticated emmer began ~9500-9000 years ago and was grown as a mixture along with the wild emmer for a longer period (Feldman & Kislev, 2007). Over the years the durum wheat i.e., *T. turgidum ssp. durum* had remained an important economic crop majorly used for making pasta, leading to the whole-genome sequencing of a modern-day cultivar (Maccaferri et al., 2019).

Wild *T. turgidum,* represented by *T. dicoccoides* is the only true wild polyploid wheat of the lineage, found in dry and saline regions of the Middle East (Nevo et al., 1992). The genetic diversity present in *T. dicoccoides* is a rich source for agronomic traits related to grain quality and abiotic/biotic stress (W. Xie & Nevo, 2008). Domesticated lineages of *T. dicoccoides* consist of subspecies *T. turgidum* spp. *dicoccum, T. turgidum* spp. *dicoccon, T. turgidum* spp. *durum, T. turgidum* spp. *turgidium, T. turgidum* spp. *polonicum, T. turgidum* spp. *carthlicum, T. turgidum* spp. *turanicum and T. turgidum* spp. *palaeocolchicum* (Dvořák, 2013). *T. timopheevii* belongs to the secondary gene pool of wheat and its domesticated form is grown as a minor crop in Georgia. *T. timopheevii* is an important resource for the improvement of durum and bread wheat (Mori et al., 2009). *T. timopheevii* consists of two sub-species i.e., wild *T. timopheevii* spp. *armeniacum* and cultivated *T. timopheevii* spp. *timopheevii* (Dvořák, 2013). The two lineages show a spread in the Transcaucasian region at the border of Europe and Asia (Matsuoka, 2011).

Maestra and coworkers (1999) studied chromosomal structure in wild and cultivated populations of *T. timopheevii*, durum, and hexaploid wheat. They reported

similar homologous pairing between A-genome and lesser association between chromosomes of the B-G genome. Structural similarities were observed for the A genome chromosomes (1, 2, 5, and 7) and G-genome chromosomes (2,3,5, and 6) with their respective A and B genome counterparts in durum and hexaploid wheat (Maestra & Naranjo, 1999). The higher similarity between the A-genome of the two species was attributed to their common ancestor *T. urartu* (Dvorak et al., 1993). The B-genome of *T. turgidum* showed higher divergence from the S-genome of *Ae. speltoides* compared to the G-genome of *T. timopheevii* (Dvořák, 2013). *T. turgidum* and *T. timopheevii* share similar morphological characters and are generally mislabeled or misclassified. Molecular markers are routinely applied for whole-genome characterization. Through barcoded multiplexing, low-cost high-throughput techniques like GBS can generate a significant number of SNP markers for large populations. The method involves a set of restriction enzymes to digest the genome into fragments, which are then sequenced using Illumina single or paired-end sequencing (Poland et al., 2012). SNP calls can be made from the sequencing data using bioinformatics pipelines i.e., UNEAK, TASSEL-GBS, and Fast-GBS (Glaubitz et al., 2014; F. Lu et al., 2013; Torkamaneh et al., 2017). Czajkowska and coworkers have reclassified 10 *T. turgidum* accessions to *T. timopheevii* using the gene-based PCR and GBS markers (Czajkowska et al., 2019).

Change of seed dispersal mode (reduced shattering) is considered as significant mutation leading to the domestication of wheat and a key trait for domestication syndrome (Abbo et al., 2014; Peleg et al., 2011). The non-brittle rachis was considered a major domestication marker for cereals, where the spike shattering is an observable trait in wild relatives compared to the intact non-shattered rachis of domesticated

cereals (Harlan & Zohary, 1966; Tanno & Willcox, 2006). In barley, two tightly linked genes *Brittle rachis 1* (*Btr1*) and *Brittle rachis 2* (*Btr2*) were associated with the brittle rachis trait, where genetic mutation leading to non-functional genes results in non-brittle spikes (Civáň & Brown, 2017; Pourkheirandish et al., 2015). In wild emmer, *Btr1* and *Btr2* gene homologs on chromosomes 3A and 3B were identified, where mutation in the *Btr1* gene was found to be associated with the non-shattering trait (Avni et al., 2017). Single point mutation of *Btr1* in domesticated einkorn was associated with the non-shattering trait (Pourkheirandish et al., 2018). The adaptability of wheat to diverse environments was attributed to the genetic changes in vernalization-related genes *VRN-1* mapping to the long arm of chromosome 5 (Law et al., 1976; Muterko & Salina, 2018). *VRN-1* gene encodes MADS-box transcription factor controlling the flowering time and regulates plant's transition from vegetative to reproductive phase (Shcherban et al., 2015; Shitsukawa et al., 2007).

There is an urgent need for the preservation of the wild germplasm from the danger of genetic erosion. Wild germplasm holds immense potential for the enhancement of genetic traits like mineral content, grain protein, and resistance to biotic/abiotic stress (Arzani, 2011) For efficient utilization of the germplasm in the breeding program, it is necessary to characterize and classify the available genetic diversity to a small set of accessions representing the maximal diversity. With the advent of parallel sequencing technologies and methods like genotyping-by-sequencing, cost-effective strategies can be developed for the curation of diverse germplasm present in gene banks (S. R. McCouch et al., 2012; Poland & Rife, 2012). The present study was undertaken to utilize the low-cost GBS technology to create a

reference core set for *T. turgidum* and *T. timopheevii* accessions present in Wheat genetics resource center (WGRC) gene bank for the effective utilization of wild and related tetraploid Triticum species in the breeding programs.

2.3 Material and Methods

Plant material

The present study included 908 accessions of tetraploid wheat from two different species, *T. turgidum* and *T. timopheevii* present in the Wheat genetics resource center (WGRC) at Kansas State University (K-State), Manhattan, KS, USA. Table S2.1 presents passport data and collection site; Figure 2.1B shows the geographic location and distribution of the accessions.

Plant tissue collection and genotyping-by-sequencing

For the DNA extraction, plant tissue collection was done from a single plant from each accession after growing them in the greenhouse for 2-3 weeks. A small piece of young leaf from a single plant of each accession was cut and collected in 96-well tissue collection box. Plant samples were then lyophilized, followed by genomic DNA extraction using Qiagen BioSprint a 96 DNA Plant Kit (QIAGEN, Hilden, Germany). Extracted DNA was quantified with Quant-iT^{TM} PicoGreen$^{®}$ dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, United States). After DNA quantification and normalization, the DNA samples were transferred to 96-well DNA collection plates with one random well per plate left blank for quality control and library integrity. Genotyping of the DNA samples was performed using genotyping-by-sequencing (GBS) as described (Poland et al., 2012). GBS libraries were prepared in 96 plexing using two restriction enzymes - a rare cutter *PstI* (5′-CTGCAG-3′), and a frequent

cutter *MspI* (5′-CCGG-3′) with a standard reverse adapter ligated. GBS libraries were sequenced using the Illumina HiSeq2000 (Illumina, San Diego, CA, United States) platform at the University of Missouri (UMC; Columbia, Missouri), generating 100bp long reads.

SNP genotyping and data filtering

TASSEL-GBS V5.0 pipeline was used for SNP discovery (Glaubitz et al., 2014). The pipeline was modified for SNP discovery without reference genome. More than 199k SNPs were discovered and used for further analysis. Population-level SNP filtering was performed and SNPs with minor allele frequency (MAF) less than 0.01 and missing data of more than 20% were removed. Further, SNPs with heterozygosity greater than 5% were removed because all the used tetraploid wheat accessions were highly inbred. Fisher's exact test at alpha 0.001 with Bonferroni correction was performed to determine if the putative SNPs were from allelic tags as described (Poland et al., 2012). Individual samples with more than 80% missing SNP calls and more than 5% heterozygosity were also removed. Retained markers and samples were used for further analyses. After all the filtering steps, 65,535 tags were eventually used for the diversity study.

Genetic diversity analysis and population structure

GBS tags were blast searched against the *T. turgidum* spp. *durum* genome considering a maximum of three mismatches and no gap for locating the genomic position. As a measure of average heterozygosity over multiple SNPs in a given population, Nei's diversity index (Nei, 1973) was computed for the whole population using VCFtools (Danecek et al., 2011). Pairwise F_{ST} between subpopulations and

subspecies were computed and plotted using custom R scripts. An R-package *rrBLUP* was used for missing data imputation using the 'mean' method for PCA. The Principal Component Analysis was performed in R. Eigenvalues and eigenvectors were computed with the 'e' function using the 'A' matrix output of the rrBLUP package (Endelman, 2011). Package *ape* was used to draw neighbor-joining tree for all accessions (Paradis et al., 2004). Species Patterson's D (ABBA-BABA) statistic was calculated using Dsuite (Malinsky et al., 2021). Population structure was analyzed at different values of sub-population runs ranging from K=2 to K=15 using fastStructure for *T. turgidum* and *T. timopheevii* at both inter and intra-species levels (Raj et al., 2014). The best population substructure was selected using the chooseK.py script of fastStructure. FastStructure output was graphically visualized using R package pophelper (Francis, 2017). PAUP4.0a163 was used to conduct coalescent SVDquartets analysis ("PAUP (Phylogenetic Analysis Using Parsimony)," 2008). Genome-wide association study (GWAS) for leaf rust in *T. timopheevii* using GAPIT (J. Wang & Zhang, 2021). FarmCPU, GLM, MLM, MLMM and SUPER models present in GAPIT were used. The Significant SNP associations were identified at FDR < 0.05.

Diversity of domestication genes

Triticum aestivum VRN-A1 and *VRN-1B* (AY747600 and AY747606) gene sequence was used in blast search to detect the corresponding homologs in the *T. turgidum* spp. *durum* genome, *T. dicoccoides* and *T. aestivum*. Similarly, *H. vulgare* homologs of domestication gene *non-brittle rachis 1* (*btr1*) and *non-brittle rachis 2* (*btr2*) were used for locating the homologous segment for brittle rachis in the tetraploid

genome. For *photoperiod response locus* (Ppd-1) gene *T. aestivum* (XP_037475580.1) and for *Q-locus* gene *T. turgidum* (AY702955) sequences were used in the blast search.

Genetically diverse representative core set selection

The selected set of SNPs was used to specify a representative core-set from the *T. turgidum* and *T. timopheevii* collections. The core sets of tetraploid wheat accessions were identified in two steps using the integration of genotypic and phenotypic datasets. In the first step, identified SNPs were first used with the software package PowerCore to identify the core set using default settings (K. W. Kim et al., 2007) (N. Singh, Wu, Tiwari, et al., 2019), selecting the lines to retain the most diverse alleles by implementing advanced M (maximization) strategy. One of the limitations of this software was its inability to use more than 15000 SNPs, so only SNPs with less than 10-15% missing data were used in the study. In the second step the number of selected accessions was further reduced by phenotypically guided selection using the available phenotypic data for growth patterns, leaf rust composite, stem rust (race TTKSK) (Rouse et al., 2011) and hessian fly biotype D resistance. The entire genetic diversity captured by the MiniCore was assessed by the percent segregating SNPs in the selected accessions relative to the whole collection.

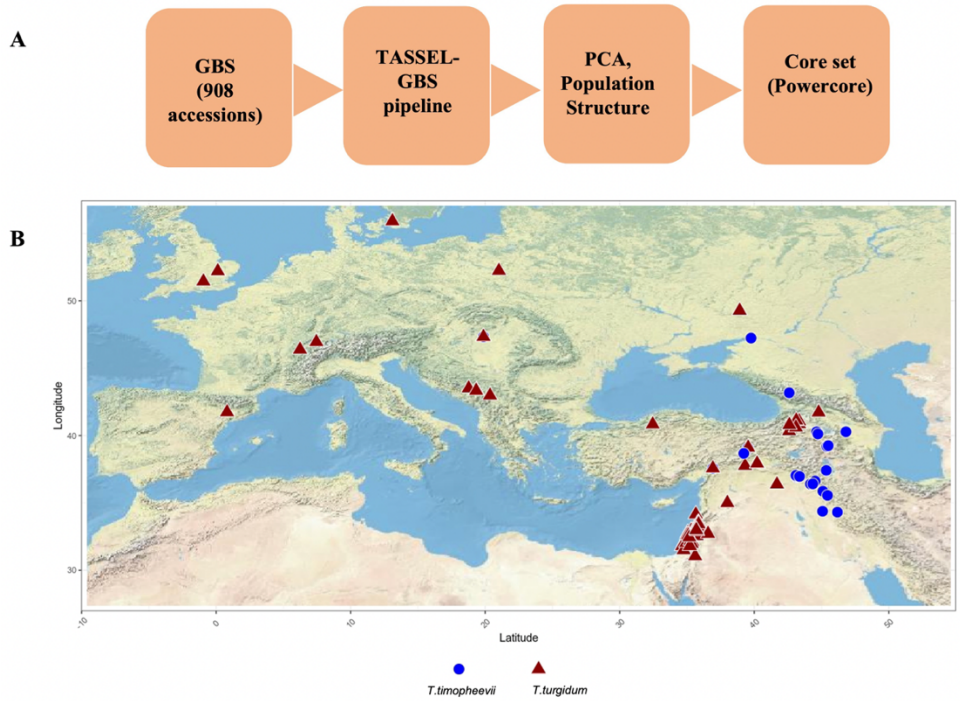


Figure 2.1 A) Flowchart of the method used. B) Geographical distribution of tetraploid wheats collection of WGRC. Red triangle represent *T. turgidum* and blue circle represent the distribution of *T. timopheevii*. Accessions from the same region are clustered together.

2.4 Results

The present study investigated the genetic diversity confined in a diverse collection of tetraploid wheat accessions available at the Wheat genetics resource center (WGRC) at Kansas State University, Manhattan, USA (Table 2.1). Table S2.1 shows species passport information, where 17 duplicated accessions were removed based on the passport and phenotypic data. A total of 908 accessions from *T. turgidum* (primary gene pool) and *T. timopheevii* (secondary gene pool) were used in the study. Figure 2.1 shows the flowchart of the methodology used and geographical locations of germplasm collection. *T. turgidum* was represented by eight subspecies: including the

wild form *T. turgidum* spp. *dicoccoides* and eight other domesticated subspecies (Figure 2.2A).

Table 2.1 Tetraploid germplasm collection of *T. turgidum* and *T. timopheevii* at WGRC gene bank.

Species	Genome	Chromosome (2n)	Number of Accessions
Triticum turgidum	BBA^UA^U	28	124
Triticum turgidum spp. *dicoccoides*	BBA^UA^U	28	384
Triticum turgidum spp. *turgidum*	BBA^UA^U	28	2
Triticum turgidum spp. *carthlicum*	BBA^UA^U	28	81
Triticum turgidum spp. *dicoccon*	BBA^UA^U	28	10
Triticum turgidum spp. *turanicum*	BBA^UA^U	28	1
Triticum turgidum spp. *polonicum*	BBA^UA^U	28	1
Triticum turgidum spp. *durum*	BBA^UA^U	28	1
Triticum turgidum spp. *paleocolchicum*	BBA^UA^U	28	2
Triticum timopheevii	GGA^MA^M	28	11
Triticum timopheevii spp. *timopheevii*	GGA^MA^M	28	7
Triticum timopheevii spp. *armeniacum*	GGA^MA^M	28	284
Total			**908**

The domesticated *T. turgidum* subspecies included in the study were *T. turgidum* ssp. *durum*, *T. turgidum* ssp. *paleocolchicum*, *T. turgidum* ssp. *turanicum, T. turgidum* ssp. *polonicum, T. turgidum* ssp. *dicoccum, T. turgidum* ssp. *dicoccon, T. turgidum* ssp. *carthilicum,* and *T. turgidum* ssp. *turgidum.* Wild *T. timopheevii* was represented *by T. timopheevii* ssp. *armeniacum,* whereas *T. timopheevii* ssp. *timopheevii* was the domesticated timopheevii subspecies. A set of 10 wheat lines were included in the study to demonstrate the relationship between tetraploid and hexaploid wheat. Figures 2.2B and 2.2C displays the country-wise distribution of the accessions used in the study.

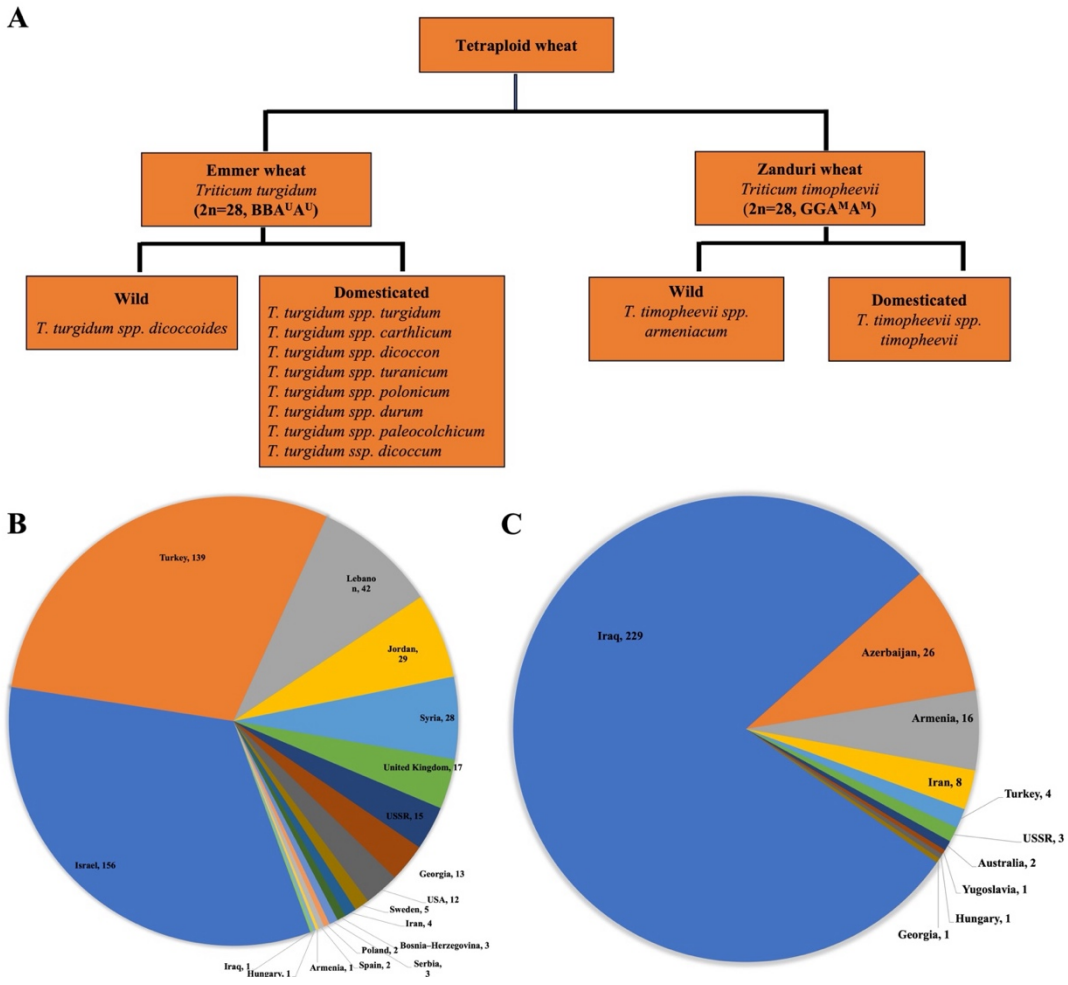


Figure 2.2 A) Tetraploid wheat accessions used in the study. Country-wise distribution of B) *T. turgidum* and C) *T. timopheevii* accessions

Principal component and cluster analysis

After filtering the SNPs with a minor allele frequency (MAF) of less than 0.01 and missing data >20, a set of 65,535 SNPs was used for the PCA. The first two components represented 27 % and 7% of the total variability, respectively (Figure 2.3A). Accessions were grouped into two major clusters: the first cluster comprised of *T. turgidum* and *T. aestivum* accessions, while the second cluster consisted of *T.*

timopheevii with an intermix of a few *T. turgidum* accessions. PC1 separated the complexes primarily based on the species type (cluster1 and cluster2), while the PC2 exhibited differentiation of the *T. turgidum* subspecies into five sub-clusters along the Y-axis. Further, PCA of tetraploid collection of *T. turgidum* and *T. timopheevii* separated the accessions based on their origin, where PC1 and PC2 contributed 28 % and 7% of sample variance, respectively (Figure 2.3B). Loose clustering was observed for the accessions not belonging to Middle Eastern origin. *T. timopheevii* ssp. *armeniacum* and domesticated *T. timopheevii* ssp. *timopheevii* grouped into one cluster attributed to their Middle Eastern origin (Armenia, Azerbaijan, Iran, and Turkey).

For *T. turgidum*, a considerable variance was observed among the accessions belonging to *T. turgidum* ssp. *dicoccoides* (dark green) from the Middle East (Israel, Lebanon, Turkey, Syria, Jordan, and Iraq), Europe (England and Sweden), and the USA. *T. turgidum* ssp. *carthilicum* clustered out with the most significant variance on PC2. While the accessions of *T. turgidum* ssp. *durum, T. turgidum* ssp. *paleocolchicum, T. turgidum* ssp. *turanicum, T. turgidum* ssp. *polonicum, T. turgidum* ssp. *dicoccum, T. turgidum* ssp. *turgidum* displayed loose clustering. The genotype-based phylogenetic tree formed two distinct clusters; the largest cluster consisted of *T. turgidum* separated from *T. timopheevii* accessions (Figure S2.1), wheat accessions formed a subgroup within the largest *T. turgidum* cluster. Five accessions were reclassified based on the clustering pattern in the PCA plot and their proximity to different clusters in the population substructure. These include three *T. timopheevii* accessions (TA9, TA875, and TA105) identified as belonging to *T. turgidum* and two *T. turgidum* (TA58 and TA987) accessions were reclassified to *T. timopheevii*. Chromosome-wise distribution

of SNPs belonging to *T. turgidum* and *T. timopheevii* are shown in Figure 2.3C and 2.3D, respectively.

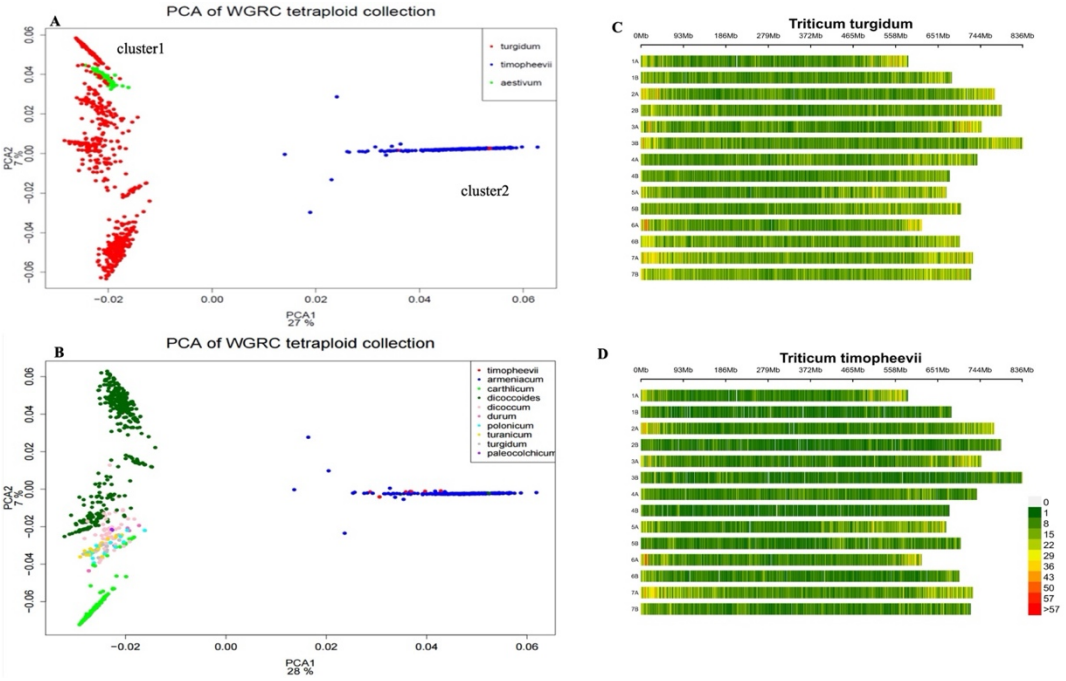


Figure 2.3 Principal Component Analysis of A) tetraploid collection of three distinct *Triticum* species: *T. aestivum* (green dots), *T. turgidum* (red dots) and *T. timopheevii* (blue dots). B) Tetraploid subspecies are a collection of wild and cultivated subspecies of *T. turgidum* and *T. timopheevii.* Distribution of the SNPs C) SNPs belonging to *T. turgidum* accessions and D) SNPs belonging to *T. timopheevii* accessions mapped to *T. turgidum* genome.

Population structure of tetraploid wheat accessions

The fastStructure based population assignment of *T. turgidum* and *T. timopheevii* accessions was performed at different sub-population levels ranging from K=2 to K=15. Using the choose.py script associated with fastStructure, the optimal sub-populational level was observed at K=14, indicating the best available population sub-structure for the accessions. (Table 2.2, Figure 2.4). At optimal K-value (14), 11 and 3 species-specific subclusters were observed for *T. turgidum* and *T. timopheevii,*

respectively. The highest level of admixture was found in four *T. turgidum* accessions from England and Turkey (TA1002, TA1003, TA1009, and TA1148). *T. timopheevii* accessions were grouped into three distinct clusters (cluster5, cluster8, and cluster14, highlighted with a thick line in Figure 2.4B), of which *T. timopheevii* spp. *armeniacum* group consisted of cluster8 and cluster14, while cluster5 majorly consisted of *T. timopheevii* spp. *timopheevii* accessions. Cluster8 constituted the largest group with the majority of accessions hailing from Iraq and a few from Azerbaijan and Turkey. Accessions from Azerbaijan were grouped into cluster14, whereas cluster5 was found to be composed of a mixture of accessions from the Middle East, European countries (Yugoslavia, Hungary, and Russian Federation), and Australia (Table 2.2, Figure 2.4). *T. turgidum* was grouped into three species-specific classes, *T. turgidum* spp. *dicoccoides* formed clusters 1,2,3,4,6,7,9,12, and 13, *T. turgidum* spp. *carthlicum* (cluster10), while cluster11 was found to be composed of a mixture with other domesticated sub-species (Table 2.2). Cluster2 consisted of *T. turgidum* spp. *dicoccoides* accessions from the Middle East (Israel, Syria, Iraq, Lebanon, and Jordan), while the majority of *T. turgidum* spp. *carthlicum* accessions in cluster10 were from Turkey, Georgia, and the Russian Federation, with fewer accessions from other European countries (Armenia, Hungary, Poland, Spain, Sweden, and the United Kingdom).

A slightly different clustering pattern was observed during species-specific population clustering (Table 2.3, Figure 2.4). *T. timopheevii* accessions clustered into two sub-populations at the optimal K-value of 2 compared to the previous 3 clusters. The majority of *T. timopheevii* spp. *armeniacum* accessions (288) formed the largest

group, while the *T. timopheevii* spp. *timopheevii* (13) grouped into the second cluster. Cluster1 consisted of timopheevii accession from Azerbaijan, Iran, Armenia, Israel, Turkey, Iraq, and Russian Federation, while cluster2 consisted of accessions from Turkey, Georgia, Hungary, Iraq, Azerbaijan, Russian Federation, and Australia.

For *T. turgidum,* cluster3 and cluster4 composed the largest group with 89 accessions each. Cluster1 was found to be the most diverse and consisted of accessions belonging to *T. turgidum* spp. *dicoccoides, T. turgidum* spp. *dicoccum*, *T. turgidum* spp. *carthlicum*, *T. turgidum* spp. *polonicum*, *T. turgidum* spp. 'blue' *dicoccum*, *T. turgidum* spp. 'blue' *durum*, *T. turgidum* spp. 'white' *dicoccum*, *T. turgidum* spp. *turanicum*, *T. turgidum* spp. 'short' *turanicum*, *T. turgidum* spp. *turgidum*, *T. turgidum* spp. *durum*, *T. turgidum* spp. *turgidum.* Cluster6 consisted of *T. aestivum* accession grouped with *T. turgidum* spp. *carthlicum*.

Genome-wide nucleotide diversity, FST scan, and introgression signal

Nucleotide diversity (Nei's Π) demonstrates the average pairwise distance between all possible pairs of individuals, calculated based on 76 % (152,769) of 199,349 tags mapping to *T. turgidum* spp. *durum* genome with criteria of 3 mismatches and zero gap per tag. A higher genetic differentiation of 0.27 was observed between the two tetraploid species *T. turgidum* and *T. timopheevii* (Table 2.4**).** Nei's index values for the individual genome (A and B genome) were found in concordance with the global diversity of 0.273 (A genome) and 0.279 (B genome). Within-species comparison resulted in lower nucleotide diversity of 0.18 and 0.04 for *T. turgidum* and *T. timopheevii,* respectively. Species-specific diversity was found to be consistent with the regional distribution of the two species. The pairwise fixation index (F_{ST}) revealed

a high level of divergence between the two species. *T. turgidum* and *T. timopheevii* populations demonstrated a high F_{ST} value (0.72). Between the two species, the B-genome showed lower F_{ST} (0.70) than the A-genome (0.74), owing to fewer tags mapping to the B-genome from *T. timopheevii*.

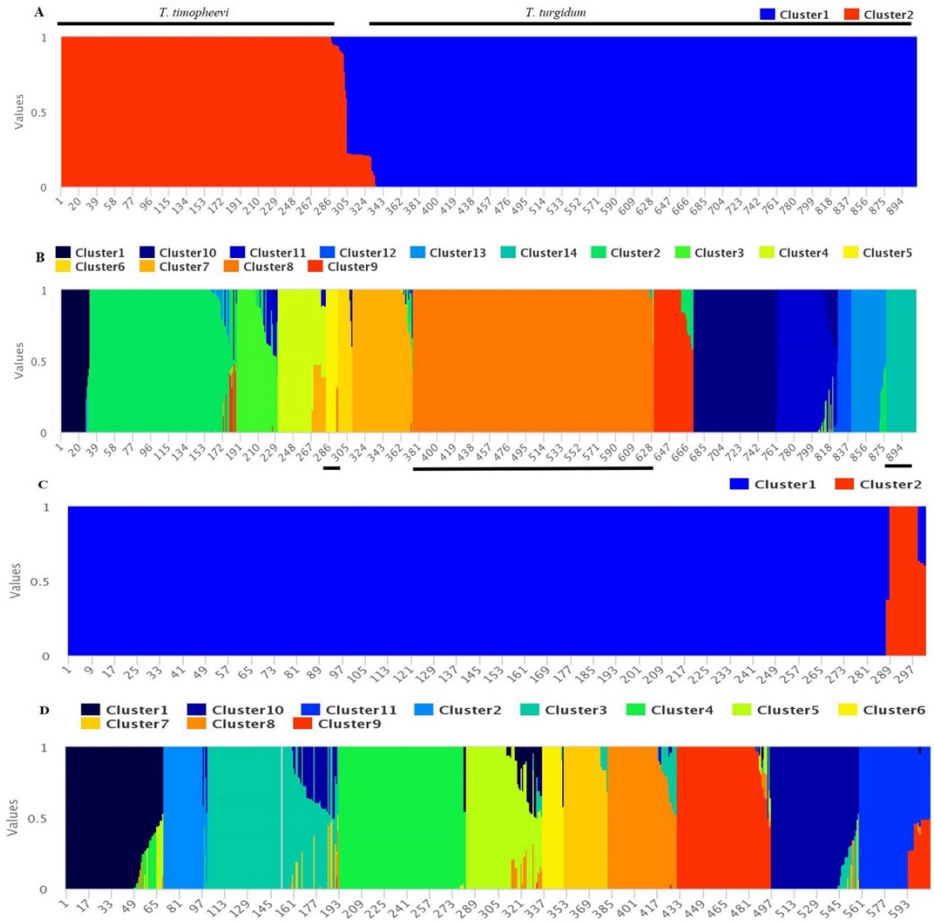


Figure 2.4 fastStucture based population sub-clustering of *T. turgidum* and *T. timopheevii* accessions. Using complete accessions, A) sub-clustering at K=2, B) optimal population substructure at K=14. *T. timopheevii* clusters are highlighted with the thick line. Species-specific population sub-clustering C) *T. timopheevii* at optimal K=2 and D) *T. turgidum* at optimal K=11.

Table 2.2 Population sub-clustering of *Triticum turgidum and Triticum timopheevii* accessions at optimal sub-population (K=14).

Population group	Sub-species	Number of accessions	Country of Collection	Nei's genetic diversity index
Triticum turgidum				
Cluster1	*dicoccoides*	30	Israel, Syria, Iraq	0.059
Cluster2	*dicoccoides*	156	Israel, Syria, Iraq, Lebanon, Jordan	0.126
Cluster3	*dicoccoides (minor: dicoccon, paleocolchicum)*	42	England, Sweden, Bosnia–Herzegovina, Serbia, Georgia, Turkey	0.106
Cluster4	*dicoccoides*	50	Turkey	0.028
Cluster6	*dicoccon*	15	Serbia	0.043
Cluster7	*dicoccoides*	64	Turkey, Iraq (1 accession)	0.085
Cluster9	*dicoccoides*	42	Lebanon	0.082
Cluster10	*carthlicum*	89	Major (Iraq, Turkey, Armenia), Georgia, Russian Federation, Hungary, Iran, Poland, Spain, Sweden, United Kingdom(UK), United State of America(USA)	0.024
Cluster11	*turgidum,dicoccon,turanicum,polonicum, durum, carthlicum*	65	England, USA, Spain	0.084
Cluster12	*dicoccoides, dicoccon*	14	Not available	0.033
Cluster13	*dicoccoides*	37	Israel	0.063
Triticum timopheevii				
Cluster5	*timopheevii*	13	Turkey, Iraq, Georgia, Hungary, Russian Federation, Australia	0.016
Cluster8	*armeniacum*	256	Iraq, Iran, Azerbaijan, Turkey	0.028
Cluster14	*armeniacum*	32	Israel, Azerbaijan, Armenia, Russian Federation	0.010

Despite regional proximity, two of the *T. timopheevii* sub-populations, *T. timopheevii* spp. *armeniacum* and *T. timopheevii* spp. *timopheevii* displayed a higher F_{ST} (0.60), indicating more sub-species differentiation due to selection and domestication. (Figure S2.2). A comparative lower F_{ST} index of 0.44 was observed between *T. turgidum* spp. *dicoccoides* and *T. turgidum* spp. *carthlicum.*

Table 2.3 Species-specific population sub-clustering of *Triticum timopheevii* (K=2) *and Triticum turgidum* (K=11).

Population group	Sub-species	Number of accessions	Country of collection	Nei's genetic diversity index
Triticum timopheevii				
Cluster1	*armeniacum*	288	Azerbaijan, Iran, Armenia, Israel, Turkey, Iraq, Russian Federation	0.028764
Cluster2	*timopheevii*	13	Azerbaijan, Turkey, Georgia, Hungary, Iraq, Russian Federation, Australia	0.0157207
Triticum turgidum				
Cluster1	*carthlicum, dicoccoides, dicoccon, durum, polonicum, turanicum, turgidum*	68	USA, Spain, England	0.0855099
Cluster2	*dicoccoides*	30	Iraq, Syria, Jordan, Israel, USA	0.0592321
Cluster3	*dicoccoides*	89	Israel, Jordan	0.118767
Cluster4	*carthlicum*	89	USA, Sweden, Russian Federation, Georgia, Spain, United Kingdom, Poland, Turkey, Armenia, Hungary, Iran, Poland	0.0239709
Cluster5	*dicoccoides, dicoccon, paleocolchicum*	50	England, Sweden, Turkey, Bosnia–Herzegovina, Serbia, Sweden, Georgia	0.111123
Cluster6	*dicoccon*	15	Serbia	0.0438263
Cluster7	*dicoccoides*	31	Israel	0.0513653
Cluster8	*dicoccoides*	49	Iraq, Syria, Lebanon	0.0925869
Cluster9	*dicoccoides*	64	Turkey	0.0839453
Cluster10	*dicoccoides*	61	Armenia, Israel, Syria, Jordan, USA, Lebanon, Israel	0.0952499
Cluster11	*dicoccoides*	50	Turkey	0.0294869

A statistically significant introgression signal with positive D-statistics and Z-score of more than three was observed between *T. turgidum* spp. *dicoccoides* and its domesticated accessions with the four-taxon f_d statistic. No significant introgression was observed between the *T. timopheevii* accessions. Relative introgression was detected between the wild and domesticated *T. turgidum* and *T. timopheevii* (Table S2.2). Analysis of the available data did not reveal any evidence of introgression from the tetraploid accession into the wheat genome.

Table 2.4 Genome and species specific Nei's diversity indices and pairwise F_{ST} coefficients of tetraploid species and core collection.

Species	Nei's genetic diversity index	F_{ST}
T. turgidum and T. timopheevii	0.270	0.72
A-genome	0.273	0.74
B-genome	0.279	0.70
T. timopheevii	0.040	0.60
Wild	0.028	
Domesticated	0.016	
T. turgidum	0.180	0.44 (*dicoccoides & carthilicum*)
Wild	0.180	
Domesticated (all accessions)	0.130	
T. turgidum spp. *carthilicum*	0.030	
Core collection		
T. timopheevii	0.048	
T. turgidum (wild)	0.169	
T. turgidum (domesticated)	0.120	

Identification of duplicated accessions and mini core set of T. turgidum and T. timopheevii accessions

Monoallelic SNPs were filtered using stringent criteria of keeping sites with MAF >0.01 and <20% of missing data. Keeping the diversity, accessions sharing 99% or more of similar bases in pairwise comparison, were treated as duplicated. Among *T. timopheevii*, 105 (*T. timopheevii* spp. *armeniacum*:94*, T. timopheevii* spp. *timopheevii*:11) unique accessions were identified. Among the *T. turgidum,* 319 (*T. turgidum* spp. *dicoccoides*: 181 and domesticated emmer: 138) unique accessions were identified (Table S2.3). Based on the Powercore analysis and phenotypic data, 37 core accessions were identified for *T. timopheevii*, 27 accessions for *T. turgidum* spp. *dicoccoides* and 38 accessions for domesticated *T. turgidum* from the WGRC collection (Figure S2.3, S2.4, Table S2.4). Core set preserved ~95%, 93%, and 98% of allelic

diversity present in wild emmer, domesticated emmer, and *T. timopheevii,* respectively. Lower Nei's diversity values for the core set indicated retention of a rich level of nucleotide diversity for *T. timopheevii* (Π:0.04)*,* wild emmer (Π:0.169) and domesticated emmer (Π:0.12). These were comparable to the species-specific Nei's values. (Table 2.4).

Center of diversity

Nei's nucleotide diversity for different admixture clusters was found to be lower except for clusters2 and cluster3, which showed a relatively higher value for *T. turgidum* (Table 2.2). Cluster2 consisted of dicoccoides accessions from Israel, Syria, Iraq, Lebanon, and Jordan and may be considered as the epicenter of *T. turgidum* diversity. While cluster 3 also consisted of dicoccoides accessions from the European region. Cluster 10 contained the maximum diversity of domesticated accessions. For, *T. timopheevii* spp. *armeniacum* cluster 8 displayed most of the diversity in comparison to the other two clusters, predominantly accessions from Iraq and Azerbaijan (Table 2.2). Comparison of F_{ST} values indicates species differentiation. The pairwise F_{ST} was calculated for each sub-population cluster/group. Significantly, high F_{ST} values ranging from 0.242 to 0.965 were observed between the different sub-population clusters (Figure S2.9). F_{ST} based clustering grouped cluster 2 and cluster 3 into independent groups of similar diversity profiles. Group I consisted of clusters 1, 2, 9, and 13 from the Mediterranean region (Israel, Syria, Iraq, Lebanon, Jordan), while Group II consisted of clusters 3, 6, 11, and 12 from the European region. Group III consisted of *T. turgidum* ssp. *dicoccoides* (clusters 4,7) and *T. turgidum* ssp. *carthlicum* (cluster10) with many accessions from Turkey and Georgia. Cluster10 was found to be the most

diverse group, with a collection of accessions from Central Europe, England, and the USA.

A species tree was developed using the SVDquartets coalescent model implemented in PAUP. The species tree demonstrated distinct evolution of *T. timopheevii* from *T. turgidum* and a clear distinction was observed between wild and domesticated emmer wheat. *T. aestivum* was clustered with *T. turgidum* ssp. *carthlicum* owing to its recent evolution (Figure S2.10A). A consensus tree based on population structure groups from Table 2.2, categorized the species clusters according to their geographic location. (Figure S2.10B). Clusters 2, 9, and 13 from Group I of F_{ST} clustering were grouped into a single cluster; clusters4 and 7 from Group III of F_{ST} formed a separate group; and clusters6, 10, and 11 with domesticated accessions were congregated with hexaploid wheat. Accessions from cluster10 shared 71–81% of alleles with accessions from cluster4 and 7, with most of the accessions belonging to Turkey. All the *T. turgidum* accessions from Turkey belonging to clusters 3, 4, 7, and 10 were clustered into one group to calculate the nucleotide diversity, which was found to be significantly higher (0.104) than individual groups.

Similarly, a higher genetic distance (F_{ST}: 0.64) was observed between the wild and domesticated accessions in this region. The *T. turgidum* ssp. *carthlicum* accessions from England, the United States, Sweden, and Poland were classified as belonging to cluster 10, except for two accessions from the United States (TA2843 and TA2884) and one accession from Spain (TA2868), which showed admixture between clusters 10 and 11. Cluster10 consisted of accessions majorly from Turkey and with an admixture coefficient of 0.99 indicates Turkey as the center of diversity. *T. turgidum* ssp.

dicoccoides accessions from England exhibited admixture with samples from clusters 2, 3, and 11, while accessions from the United States were assigned to cluster 2 and 3, accessions from Sweden showed admixture between clusters 3 and 2 (Figure S2.11). This distinct pattern of dicoccoides accessions may be due to the regional isolation and was also observed in the cluster-specific tree (Figure S2.10B). Cluster2, with the maximum diversity of wild emmer, may be considered as the center of species diversity for *T. turgidum* spp. *dicoccoides* and comprised of the region of Israel, Jordan (West Bank), Syria (Golan Heights) and Lebanon (Al Biqa). Domesticated emmer wheat was found in highest diversity in the regions of Turkey (Kars) and Georgia (Tiflis). With Turkey's highest genetic diversity of accession, domesticated accessions have found their way into central Europe through Turkey. For *T. timopheevii* spp. *armeniacum,* the region of higher genetic diversity comprised Iraq (Arbil, Dahuk, and As Sulaymaniyah), Azerbaijan (Nakhichevanskaya), and Armenia. A comparison of the nucleotide diversity of accessions from these three areas revealed that accessions from Iraq (0.29) had the highest diversity, followed by Azerbaijan (0.19) and Armenia (0.07). F_{ST} comparison revealed lesser divergence (F_{ST}:0.30) between the populations of Iraq and Azerbaijan compared to the Iraq-Armenian population (F_{ST}: 0.41). Based on the higher genetic diversity and resistant accessions the region comprising Arbil, Dahuk, and As Sulaymaniyah in Iraq may be the potential site of origin for the wild-type *T. timopheevii.*

Signatures of domestication and adaptation-related genes in tetraploid wheat

VRN-A1 and *VRN-1B* gene sequences of *T. aestivum* (NCBI accessions AY747600 and AY747606) were used for the detection of corresponding candidate

regions in the *T. turgidum* genome (Golovnina et al., 2010). Blast search identified the gene positions mapping to chromosome 5A (549156384-549152141) and 5B (570844281-570831393). Differential but similar F_{ST} distribution was observed in the candidate region for *VRN-1* on chromosomes 5A and 5B in both *T. turgidum* and *T. timopheevii* accessions. For the domestication gene *brittle rachis* (*Btr*), two physically linked genes *non-brittle rachis 1* (*btr1*) and *non-brittle rachis 2* (*btr2*) of *H. vulgare* (NCBI accession: KR813335.1) were used in the blast search. (Pourkheirandish et al., 2015). Two chromosomal segments corresponding to 3A and 3B were identified (Table S2.5**)**. An elevated F_{ST} was observed for both *T. turgidum* and *T. timopheevii* (Figure S2.6, S2.7). *VRN* and *Btr* genes of *T. dicoccoides* and *T. turgidum* showed higher similarity to *T. aestivum* genes. Figure S2.8 displays F_{ST} distribution of *Ppd-1* and *Q-locus*. For *Ppd*-1A, single blast hit was observed on chromosome 2A, while no similarity was detected on chromosome 2B of *T. turgidum*. *Q-locus* showed hits on chromosomes 5A and 5B.

Distribution of disease resistant T. timopheevii accessions

Out of 105 unique *T. timopheevii* accessions, phenotypic data was made available for 95 accessions related to leaf rust, Septoria, tan spot, and powdery mildew (data not shown). Figure S2.5 shows the distribution of phenotypic data for *T. timopheevii.* Most accessions showing resistance to all four diseases were found belong to Iraq (region of Arbil, Dahuk, and As Sulaymaniyah). Detailed examination revealed accessions showing resistance to Septoria and tan spot belonging to Armenia, Azerbaijan, Iran, and Iraq. Accessions from Armenia and Azerbaijan showed moderate resistance to Powdery mildew and susceptibility to leaf rust, while accessions from Iran

were susceptible to both Powdery mildew and leaf rust. Single accessions from Australia and Hungary were resistant to leaf rust only or the collectors selected these for being resistant to leaf rust. Genome-wide association study (GWAS) for leaf rust identifies four significant marker-trait associations (MTAs) using MLMM and three significant MTAs using FarmCPU on chromosome 2B. One MTA (SNP_ 29001) on distal region of chromosome 2BS was found to be common with both the methods (Table S2.6; Figure S2.12). Therefore, this MTA was considered more important and hence was further examined for putative candidate genes (CGs). For identification of putative CGs, the MTA region was extended to 100 kb in both directions and CGs were identified in this 200 kb region. In this region, four genes were identified which encoded for two NBS-LRR-like resistance proteins (TRITD2Bv1G010030 and TRITD2Bv1G010060), and two Exocyst complex component 2C putative proteins (TRITD2Bv1G010050 and TRITD2Bv1G010090). NBS-LRR-like resistance proteins play a critical role in providing resistance against leaf and stripe rust of wheat. Interestingly, two of the reported leaf rust (*Lr*) resistance genes, *Lr16* (McCartney et al., 2005) and *LrZH2* (C. Wang et al., 2016) were also reported earlier in distal region of the chromosome 2BS. Therefore, the above two NBS-LRR genes may be the potential candidates for either of the two *Lr* genes; however, this certainly needs further validation. The results need to be further examined with the inclusion of more phenotypic data in the future.

Accession	*Septoria*	Tan Spot	Powdery Mildew	Leaf Rust
TA105	R	R	5	9
TA1008	R	R	9	8
TA6	R	R	-	8
TA23	-	-	5	8
TA25	R	-	4	2
TA48	-	-	4	8
TA101	R	R	5	9
TA145	R	VR	4	2
TA153	-	-	2	2
TA170	-	-	0	3
TA861	R	R	4	5
TA877	R	R	4	8
TA921	-	-	4	7
TA946	R	R	4	2
TA1486	R	R	1	2
TA1488	R	R	5	8
TA1489	-	-	3	2
TA1499	R	R	7	8
TA1510	-	-	4	9
TA1514	-	-	4	8
TA1520	R	R	3	1
TA1521	-	-	4	1
TA1526	R	R	4	5
TA1536	R	R	5	3
TA1571	R	R	4	9
TA892	-	-	1	2
TA905L1	-	-	5	2
TA44	R	R	4	-
TA49	R	R	4	8
TA1564	R	R	5	5
TA1565	-	-	3	9
TA1569	R	R	4	9
TA2892	R	R	6	8

Figure 2.5 Heatmap of selected *T. timopheevii* accessions in response to Septoria, tan spot, powdery mildew, and leaf rust. VR: very resistant or immune; R: resistant; Powdery mildew (0: immune,1: resistant, 2-4: Moderately resistant, 5-9:Susceptible) Leaf rust (0;immune, 1-2:resistant, 3-4: Moderately resistant, 5-9:Susceptible)

2.5 Discussion

The WGRC tetraploid wheat collection included germplasm from all over the world, especially from the Fertile Crescent region of the Middle East (Figure 2.1). Archaeological evidence suggests that wild emmer and free-threshing tetraploid domesticated emmer were first cultivated in the southern Levant region as early as 10000 years ago (Feldman & Kislev, 2007). Presently, wild emmer is found growing in the western region of the fertile crescent (Israel, Jordan, and Syria), south-eastern

region of Turkey, mountain ranges of east Iraq and west of Iran (Özkan et al., 2011). Jordan Valley, with the highest genetic diversity, was considered to be the epicenter of wild emmer. (Harlan & Zohary, 1966). *T. timopheevii* spp. *armeniacum (T. araraticum)* displayed a spread in Iran, Iraq, Georgia, Armenia, Azerbaijan, and Turkey (Badaeva et al., 1990), whereas *T. timopheevii* spp. *timopheevii* found to be endemic to Georgia (Mitrofanova et al., 2016). Both the tetraploid species are self-pollinated. Crosses between *T. turgidum* and *T. timopheevii* result in a non-fertile progeny due to failed pairing of chromosomes (Wagenaar, 1966). These two tetraploid species share highly similar morphological characteristics (Czajkowska et al., 2019) and to some extent chromosomal similarities (Maestra & Naranjo, 1999) that can result in misclassification, mislabeling, and duplication of accessions from different countries. In the absence of passport information, sometimes accession identification becomes difficult. The WGRC gene bank provides germplasm collections to the breeders for its efficient utilization in the breeding program for wheat improvement. Considering the accession duplication, efficient strategies were needed to narrow these to a core set of accessions representing most diversity for their easy utilization in breeding programs. The availability of a core set will reduce the effort needed for the phenotyping of the large number of accessions. An efficient but time-consuming strategy involves using the combination of phenotypic and passport information to limit the number of accessions. Clustering-based methods like PCA can efficiently be used with next-generation sequencing data for their characterization using the genome-wide tags generated through GBS (N. Singh, Wu, Raupp, et al., 2019). The present study used

908 tetraploid wheat accessions from the WGRC gene bank to define a core collection of genotypes covering most of the available genetic diversity.

Tetraploid wheat races and their distribution

Wild emmer germplasm has been explored in many studies using whole-genome sequencing (Avni et al., 2017; Dvorak, Wang, Zhu, Jorgensen, Deal, et al., 2018; Maccaferri et al., 2019; Zhu et al., 2019) and exome sequencing (He et al., 2019). Walkowiak and coworkers explored the secondary gene pool species *i.e., T. timopheevii* for identifying introgression segment on chromosome 2B of wheat variety LongReach Lancer (Walkowiak et al., 2020). Cytological methods were widely used to detect *T. timopheevii* introgression signals in the wheat (Badaeva et al., 2016; Padmanaban et al., 2018). Wild emmer accessions were previously classified into wild race I (*T. turgidum ssp. dicoccoides* (Körn. ex Asch. & Graebner) Schweinf.) race II *(T. turgidum ssp. dicoccum* Schrank ex Schübler), race IV (*T. turgidum ssp. durum* Desf.), race V (*T. turgidum ssp. polonicum* L.) and race VI (*T. turgidum* L.) (Percival, 1921). Races were reclassified as *T. turgidum* and *T. timopheevii* (Zhuk) Zhuk subspecies (Van, 1994). Hulled wild tetraploids include *T. timopheevii* spp. *armeniacum* and *T. turgidum* spp. *dicoccoides*, while *T. timopheevii* spp. *timopheevii* and *T. turgidum* spp. *paleocolchicum* constitutes members of domesticated hulled subspecies. Domesticated free-threshing emmer species consists of *T. turgidum* spp. *carthlicum*, *T. turgidum* spp. *durum*, *T. turgidum* spp. *polonicum*, *T. turgidum* spp. *turanicum*, *T. turgidum* spp. *turgidum* and *T. turgidum* spp. *turgidum var uniaristatum.*

Diversity analysis and population structure

Our results showed higher but almost similar nucleotide diversity between the A-genome (0.273) and B-genome (0.279) of *T. turgidum* and *T. timopheevii*. Although literature reports a higher dissimilarity in the B-genome compared to A-genome between the two species. A potential reason for inconsistency may be the lesser (40%) mapping of G-genome tags on B-genome used in blast search. A reduced or low genetic diversity was observed in the species-specific accessions (Table 2.4). A higher Nei's diversity index was observed for wild accessions (*T. turgidum* spp. *dicoccoide*: 0.180*, T. timopheevii* spp. *armeniacum*: 0.028) compared to domesticated accessions (*T. turgidum*: 0.13, *T. timopheevii*: 0.016). Sub-population cluster3 and cluster4 among the *T. turgidum* spp. *dicoccoide* showed the highest and lowest Nei's index diversity, respectively (Table 2.3). Diversity analysis based on fastStructure and F_{ST} estimates grouped two tetraploid species into distinct clusters supporting the species-specific differentiation between the *T. turgidum* and *T. timopheevii*. Rodriguez and coworkers reported higher homologous pairing of B genome of *Ae. speltoides* and G genome of *T. timopheevii* compared to B-genome of *T. aestivum*, suggesting the most recent emergence *T. timopheevii* (Rodríguez et al., 2000). In the absence of migration, population divergence increases significantly, resulting in higher differentiation (F_{ST} values). Different threshold values were reported for significant F_{ST} estimates among the populations, such as >0.25 (Hartl & Clark, 1997) or >0.15 (Frankham et al., 2002). Intra-species differentiation was observed between wild and domesticated *T. timopheevii* accessions belonging to *T. timopheevii* spp. *armeniacum* and *T. timopheevii* spp. *timopheevii* and also between wild and domesticated *T. turgidum*

accessions i.e*., T. turgidum* spp. *dicoccoides* and *T. turgidum* spp. *carthlicum* resulting from domestication (Table 2.4). *T. timopheevii* spp. *armeniacum,* grows widely from Armenia to Azerbaijan region of the former Russian Federation, while domesticated *timopheevii* remained endemic to a small region in the west of Georgia with cultivation limited to a few villages (Badaeva et al., 2022). The higher genetic differentiation observed within the *T. timopheevii* accessions may be due to their region-specific cultivation or may be attributed to the lesser diversity captured by fewer accessions of domesticated *T. timopheevii* under study.

Out of the 11 clusters of emmer wheat, *T. turgidum* spp. *dicoccoides* majorly or solely contributed to the seven clusters (Table 2.3). Lower nucleotide diversity was observed within the domesticated accession clusters (cluter4 and cluster6) compared to the wild *dicoccoides*. Clustering distinction observed for *T. turgidum* ssp. *carthlicum* and *T. turgidum* ssp. *dicoccon* was an indication of speciation, compared to *T. turgidum* ssp. *durum*, *T. turgidum* ssp. *polonicum, T. turgidum* ssp. *turanicum* and *T. turgidum* ssp. *turgidum* (Figure S2.10B). Previous studies on durum using molecular markers i.e., SSR, DarT and SNPs suggested a lesser genetic differentiation within the domesticated turgidum accessions (Laidò et al., 2013; Oliveira et al., 2014, 2020). Separate sub-population clustering and low nucleotide diversity of *T. turgidum* ssp. *carthlicum* supports its genetic differentiation from other tetraploid wheat sub-species. Riefolo and coworkers (2011) reported separate phylogenetic positions for *T. turgidum* ssp. *durum* and *T. turgidum* ssp. *carthlicum* (Riefolo et al., 2011). *T. turgidum* ssp. *carthlicum* (also known as Persian wheat) was supposed to be originated from the cross of domesticated emmer wheat and hexaploid wheat (Kuckuck, 1979; Matsuoka, 2011).

Genetic markers also reported separate phylogenetic clades for *T. turgidum* ssp. *carthlicum* and *T. turgidum ssp. dicoccon* (Hyun et al., 2020; Laidò et al., 2013) and provide support for the separate *T. turgidum* ssp. *dicoccon* cluster in population structure analysis and species tree. A recent study reported three distinct lineages for *T. araraticum* based on the genomic and cytogenetic diversity of 862 tetraploid accessions (Badaeva et al., 2022). Two lineages for *T. araraticum* were designated as ARA-0 and ARA-1 and one domesticated *T. timopheevii.* Lineage ARA-0 had wider geographic distribution compared to ARA-1. In our results, population clustering using the whole set of *T. timopheevii* and *T. turgidum* accessions resulted in 3 sub-population clusters (cluster5, cluster8 and cluster14) attributed to the regional specificity of cluster8 to Iraq and cluster 14 to Azerbaijan and Armenia., but the pattern diminishes during comparison of *T. timopheevii* only accessions, where lower Nei's diversity (0.028) was observed for *T. timopheevii* spp. *armeniacum* and 0.015 for *T. timopheevii* spp. *timopheevii.*

Phylogenetic analysis of wheat chloroplast genomes also showed species-specific distinctions of Poaceae members, specifically the tetraploid wheat (Gornicki et al., 2014). In another GBS based study, two major classes of *T. timopheevii* spp. *armeniacum* and *T. timopheevii* spp. *timopheevii* were reported (Hyun et al., 2020). These results were consistent with our finding of two major lineages of *T. timopheevii.* More timopheevii accessions may be required to support three clade classification. The potential reason for the differential clustering may be the closeness of ARA-1 class to the *T. turgidum* as discussed by Badaeva *et al.,* (Badaeva et al., 2022). Based on nucleotide diversity and differentiation, the region comprising Israel, Jordan (West

Bank), Syria (Golan Heights) and Lebanon (Al Biqa) were rich in diversity for *T. turgidum* spp. *dicoccoides*. *Carthilicum* displayed rich diversity in the region of Turkey (Kars) and Georgia (Tiflis). Regions of Iraq (Arbil, Dahuk, and As Sulaymaniyah), Azerbaijan (Nakhichevanskaya), and Armenia can be considered as diversity rich areas for *T. timopheevii* spp. *armeniacum* with the maximum genetic diversity in Iraq. Figure 2.6 summarize major identified center of genetic diversity, potential center of domestication and route of cultivation for tetraploid wheat.

FST signal surrounding the domestication genes

A variable and low F_{ST} signal was observed in the potential candidate region for the genes: the non-brittle rachis (*Btr1*), vernalization (*VRN-1*), developmental signaling gene *Ppd-1A* and *Q locus.* However, no clear haplotype was identified in the genic region in the absence of tags mapping on the gene, but a variable and similar F_{ST} signal between the wild and domesticated tetraploids on A and B genome provides evidence for the emergence of domestication and environmental adaptation traits in domesticated accessions. Variability in *VRN-1* (MADS-box transcription factor) controls the plant adaptability to different environments (Muterko & Salina, 2018). *VRN-1* alleles in *T. dicoccodes* and *T. timopheevii* displayed differential origin with insertion in the main gene in *T. timopheevii* and a region of variability in the promotor region of *T. dicoccoides* (Shcherban & Salina, 2017).

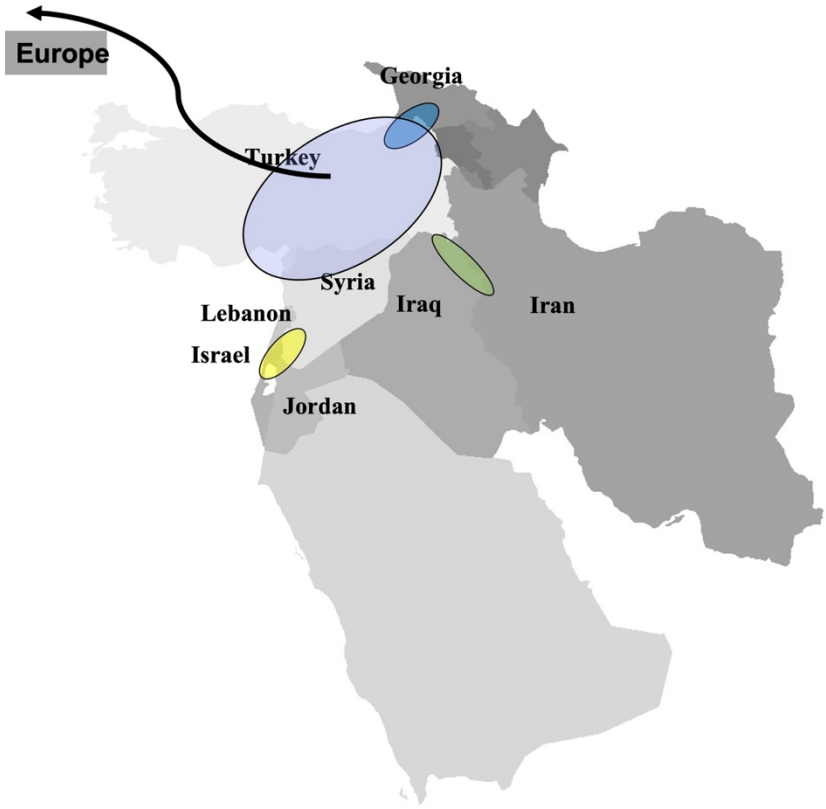


Figure 2.6 Centre of rich genetic diversity for tetraploid wheat. Highlighted regions are not on scale, represents major cities. Blue circle: *T. turgidum* spp*. carthlicum* (Kars, Turkey; Tiflis, Georgia); Yellow circle : *T. turgidum spp. dicoccoides* (Israel, Jordan, Syria , Lebanon)*;* Green circle : *T. timopheevii* spp. *armeniacum (*Iraq (Arbil, Dahuk, As Sulaymaniyah)). Purple circle in Turkey may be considered a region of domestication and the cultivation of domesticated emmer. Arrow represents the potential route of migration into the Europe.

In another study of gene-based haplotypes using our core accessions, domesticated *T. turgidum* was found to share haplotypes for the *Btr1-A* and *Btr1-B.* In *T. timopheevii,* novel mutations were identified in the *Brt1-A* allele, displaying partial brittleness of spikes. At the same time, no amplification was observed for the allele on the G-genome, suggesting the differential domestication event for *T. timopheevii* (Nave et al., 2021). The region surrounding the genes related to domestication traits showed

modest differentiation between wild and domesticated accessions, demonstrating the preservation of diversity in the region (Figure S2.8).

Core collection

The study identified a core set of accessions from wild and domesticated accession of *T. timopheevii* and *T. turgidum* accessions. The selected core set of accessions represents more than 93% of the available allelic diversity, with nucleotide diversity comparable to individual population used in the study (Table 2.4). From the original number of accessions, a nearly 10x sample reduction was achieved in the core set. Wild emmer consisted of a majority of the accessions from Israel (11) followed by Turkey (6), Jordan (5), and one each from Sweden, Syria, England, and Lebanon, representing the sub-population clusters 1,2,3,4,7, and 13. Domesticated emmer consisted of accessions from Turkey (7), Portugal (3) and one each from Jordan, Sweden, Syria, Afghanistan, Armenia, Former USSR, Greece, Hungary, Kyrgyzstan, Omen, Spain, United States, and Yemen, representing sub-population clusters 2,3,6,10,11 and 12. *T. timopheevii* spp. *araminicam* core set consisted of accessions from Iraq (21), Armenia (5), Azerbaijan (4), Turkey(2), Russian Federation (2) and Iran (1) from cluster8 and 14, while *T. timopheevii* spp. *timopheevii* consisted of two accessions from Russian federation. Our *T. timopheevii* core set capture a rich source of diversity against fungal diseases like *Septoria*, leaf rust, tan spot, and powdery mildew (Figure 2.5). For fifty percent accessions, phenotypic data is available. The present core set will serve as a rich source of genetic variations to be utilized in the breeding program and should be coupled with careful phenotyping. In recent years, gene bank curation has been done to explore the novel genetic variations in *T.*

monococcum (Adhikari et al., 2022), *T. aestivum* (Balfourier et al., 2007; C. Y. Hao et al., 2006, 2008; Pascual, Fernández, et al., 2020), *Ae. tauschii* (N. Singh, Wu, Tiwari, et al., 2019), *T. turgidum* L. (Ruiz et al., 2013), rye (Sidhu et al., 2019), sorghum (Cuevas et al., 2017) , melon (X. Wang et al., 2021) , pumpkin (Nguyen et al., 2020), capsicum (Tripodi et al., 2021) for trait improvement.

GBS is a relatively inexpensive method for genotyping large numbers of samples and provides more SNPs than SNP arrays. Although the domesticated varieties are high yielding, they lack resistance to biotic and abiotic stress. Wild germplasm with rich and unexplored genetic diversity holds the potential for traits improvement related to resistance, mineral content, and protein content (Arzani, 2011). Several resistance genes for powdery mildew resistance *Pm*6 (Jorgensen & Jensen, 1973), *Pm*27 (Järve et al., 2000), *Pm*37 (Srnić et al., 2005), *MlAG*12 (Maxwell et al., 2009) , leaf rust *Lr*50 (Brown-Guedira et al., 2003), *LrTt*2 (Leonova et al., 2010), stem rust *Sr*40 (Dyck, 1992) and fusarium head blight resistance (Malihipour et al., 2016, 2017) has been introgressed into the bread wheat from *T. timopheevii.* In recent years, *Pm*36 (Blanco et al., 2008), *Pm*41 (G. Li et al., 2009) and leaf and stripe rust resistance (Ullah et al., 2016) genes were contributed by the wild emmer. The present tetraploid core set will serve as a rich resource of genetic variations. Researchers can effectively phenotype the small set of accessions for several traits under field or greenhouse conditions. SNP identified from the collection will facilitate future wheat breeding, association mapping, and introgression studies. Present core set accessions were already utilized to demonstrate independent domestication of *T. timopheevii* using *Brittle rachis* 1 gene

(Nave et al., 2021). These genetically distinct tetraploid accessions, coupled with quality phenotyping, can potentially develop elite wheat varieties.

2.6 Conclusion

The present study explored the potential of GBS technology in data reduction while maintaining the significant genetic diversity of the species. NGS technology can reclassify species with similar morphological characters that are difficult to distinguish phenotypically. GBS based diversity analysis leads to identify of diversity rich pockets of tetraploid wheat in Fertile Crescent. Population structure supported the genetic differentiation of both wild and domesticated tetraploids. Wild germplasm showed more differentiation than domesticated accessions, indicating the availability of sufficient diversity for crop improvement. With reduced complexity, a core-set preserves the genetic diversity of the gene bank collections and will aids in more robust characterization of wild germplasm. The study preserves ~93 % of the allelic diversity of *T. timopheevii* and *T. turgidum* accessions. GBS proved to be an efficient technique of gene bank curation and can be effectively used to explore the rich diversity for species with huge genomes or higher ploidy not suitable for resequencing. With 10x reduction, and rich germplasm diversity for disease resistance, the tetraploid core-set will cut expenses associated with storage and maintenance and act as a resource to speed global breeding initiatives.

2.7 Supplementary Material

The supplementary figures and tables are hyperlinked to the published manuscript https://www.frontiersin.org/articles/10.3389/fpls.2023.1144000/full#supplementary-material

Chapter 3: Comparative Genomic Analysis of $5M^g$ Chromosome of *Aegilops geniculata* and $5U^u$ Chromosome of *Aegilops umbellulata* Reveal Genic Diversity in the Tertiary Gene Pool

The work presented in this chapter has been published as follows:

Yadav, I. S., Rawat, N., Chhuneja, P., Kaur, S., Uauy, C., Lazo, G., Gu, Y. Q., Doležel, J., & Tiwari, V. K. (2023). Comparative genomic analysis of $5M^g$ chromosome of *Aegilops geniculata* and $5U^u$ chromosome of *Aegilops umbellulata* reveal genic diversity in the tertiary gene pool. *Frontiers in plant science*, *14*, 1144000. https://doi.org/10.3389/fpls.2023.1144000

Headings and figure numbers were modified for consistency within this dissertation. Supplementary information is included in the Appendix. Hyperlinks to the article are included in the main text for supplementary figures and table.

3.1 Abstract

Wheat is one of the most important cereal crops for the global food security. Due to its narrow genetic base, modern bread wheat cultivars face challenges from increasing abiotic and biotic stresses. Since genetic improvement is the most sustainable approach, finding novel genes and alleles is critical for enhancing the genetic diversity of wheat. The tertiary gene pool of wheat is considered a gold mine for genetic diversity as novel genes, and alleles can be identified and transferred to wheat cultivars. *Ae. geniculata* and *Ae. umbellulata* are the key members of the tertiary gene pool for wheat and harbor important genes against abiotic and biotic stresses. Homoeologous-group five chromosomes ($5U^u$ and $5M^g$) have been extensively studied from *Ae. geniculata* and *Ae. umbellulata* as they harbor several important genes including *Lr57, Lr76, Yr40, Yr70, Sr53* and chromosomal pairing loci. In the present study, using chromosome DNA sequencing and RNAseq datasets, comparative analysis was done to study homoeologous gene evolution in $5M^g$, $5U^{u,}$ and group 5 wheat chromosomes. Current findings highlight the diversity of transcription factors and resistance genes, resulting from the differential expansion of the gene families. Both the chromosomes were found to be enriched with the "response to stimulus" category of genes providing resistance against biotic and abiotic stress. Phylogenetic study positioned the M genome closer to the D genome, with higher proximity to the A genome than the B genome. Over 4000 genes were impacted by SNPs on 5D, with 4-5% of those genes displaying non-disruptive variations that affect gene function.

3.2 Introduction

The United Nation's current projections for global human population growth are alarming, with 8.6 billion people by 2030 and 9.8 billion people by 2050 (https://www.un.org). Being a staple source of calories for around 40% of the human population, common wheat (hexaploid bread wheat, 2n=6x=42; *Triticum aestivum* L.) is a critical crop for global food security (http://www.fao.org/worldfoodsituation/csdb/en/). Wheat faces continuous threats from changing climatic conditions and evolving biotic and abiotic stresses (Enghiad et al., 2017). To meet the ever-increasing demands, there is a need to increase wheat production and breed for high-yielding and more resistant cultivars (Appels et al., 2018). To meet these challenges, genetic improvement of crop plants is the most sustainable agricultural approach, requiring continuous identification, characterization, and deployment of agronomically important genes and their useful allelic variants.

The hexaploid bread wheat has a complex genome architecture consisting of three sub-genomes (A, B, and D) arising from different hybridization events. The first hybridization resulted in tetraploid wheat *Triticum turgidum* (2n=4x=28, AABB), where the A genome was contributed by the diploid progenitor *Triticum urartu* (2*n* = 2x = 14, AA). *Aegilops speltoides* (2n=14, SS) is thought to be a potential donor for the B genome. A subsequent hybridization between *T. turgidum* and *Aegilops tauschii* (DD) leads to the emergence of hexaploid wheat *Triticum aestivum* (2n = 6x = 42, AABBDD) (The International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018). The wheat gene pool is divided into three major categories based on the genomic constitution of the species i.e., primary, secondary, and tertiary. The primary gene pool

consists of closely related species with shared genomes consisting of diploid and tetraploid progenitors. The secondary gene pool is made up of species sharing at least one of the genomes with wheat and consists of polyploid *Triticum* and *Aegilops* species. Species carrying non-homologous genomes are part of the tertiary gene pool (Chaudhary et al., 2014). *Aegilops* species are compatible with wheat and can be effectively used in trait enhancement using chromosome engineering approaches (Fernandez-Calvin & Orellana, 1992; X.-Y. Zhang et al., 1996).

Aegilops geniculata (2n=4x=28, $M^gM^gU^gU^g$) and *Aegilops umbellulata* (UU) are members of the tertiary gene pool and show wide distribution in the Mediterranean Basin (Arrigo et al., 2010) and Western Asiatic region (Meimberg et al., 2009), respectively. Leaf rust resistance (*Lr57),* stripe rust resistance (*Yr40*) (Kuraparthy et al., 2007), stem rust resistance (*Sr53)* (W. Liu et al., 2011), and powdery mildew resistance (*Pm29*) (Zeller et al., 2002) genes have been introgressed into bread wheat over the years from *Ae. geniculata.* Similarly, *the* rust resistance gene *Lr*9 (Sears, 1956)*, Lr76, and Yr70* genes have been introgressed from *Ae. umbellulata* (Bansal et al., 2017) and utilized commercially (A. Sharma et al., 2021) . The U- and M-genomes of *Aegilops* species are also a rich source of genes for nutritional quality (Bálint et al., 2001). Introgression of 2U, 4U, 5U, 7U, 2M, 5M, and 7M chromosomes of *Ae. geniculata* increased the seed protein content of bread wheat (Rakszegi et al., 2017). Chromosomes 1U and 1M, increase the polymeric glutenin and demonstrate a positive effect on wheat dough strength (Garg et al., 2016; Rakszegi et al., 2017).

The high-throughput and cost-effective next-generation sequencing technology holds the potential to generate a vast number of molecular markers for the detection of

alien introgression. In polyploid crops such as wheat, flow cytometric sorting of individual chromosomes reduces the genomic complexity of the whole genome (Doležel et al., 2012). Flow-sorting has been widely used to study the relationship between different chromosomes and wild-species introgressions, i.e., *Ae. geniculata*, *Ae. umbellulata*, *Ae. comosa*, *Ae. biuncialis, Ae. markgrafii*, *Ae. triuncialis*, *Ae.cylindrica*, *T. urartu*, *Ae. speltoides*, *Ae. tauschii* (Molnár et al., 2011, 2014; Tiwari et al., 2014; Bansal et al., 2020). With a speedy pace of delivery of genomic resources and availability of germplasm from distant tertiary gene pool members, it becomes increasingly important to analyze gene level variations between cultivated wheat germplasm and its distant relatives. Understanding these gene-based variations will allow us to employ new tools, such as gene editing to perform the next generation of crop improvement.

The present study uses the DNA sequences from the flow-sorted $5M^g$ chromosome of *Ae. geniculata* and $5U^u$ chromosome of *Ae. umbellulata* available in the public database to study the gene evolution of homologous chromosomes in comparison to the wheat genome (Chinese Spring assembly and the available wheat cultivar genomes) (Bansal et al., 2020; Tiwari et al., 2014, 2015). These two chromosomes were selected for being the source of the rust resistance gene (*Lr57*, *Lr76*, *Yr40*, *Lr70,* and *Sr53*). The study aims to identify the pattern of gene evolution for resistance genes and transcription factors in $5M^g$ and $5U^u$ chromosomes with respect to the homoeologous genes in modern wheat cultivars.

3.3 Material and Methods

Assembly

The flow-shorted $5M^g$ chromosome sequence for the accession TA7670 was downloaded from the SRA database (SRX1167449). The Flow-sorted chromosome sequence of $5U^u$ was provided by Dr. Parveen Chhuneja (personal communication). SRA toolkit was used to partition the SRA file of $5M^g$ into individual fastq files (The SRA Toolkit Development Team). Fastqc was employed to check the quality of the raw reads (Andrews, 2010). Trimmomatic was used to remove adaptor sequences and filter low-quality reads (Bolger et al., 2014). The *de novo* sequence assembly was performed using SOAPdenovo2 at four different k-mers of k41, k51, k61, and k71 (R. Luo et al., 2012). BBmap was used to generate assembly statistics (Bushnell, 2014). The quality of the assembly was assessed using BUSCO-conserved genes from poales_odb10 (n:4896) (Simão et al., 2015). Contigs shorter than 500bp were removed from the assembly. A hierarchical repeat-finding approach was used, where the initial repeat finding was performed with Poaceae-specific repeats from Repbase (Bao et al., 2015). Genome-specific *de-novo* repeat families were identified using RepeatModeller (Flynn et al., 2020). Poaceae-specific repeats and *de-novo* repeats were masked using RepeatMasker (Tarailo-Graovac & Chen, 2009). *Ab initio* gene prediction was done using AUGUSTUS (Hoff and Stanke 2019). Genes with a length more than or equal to 150 bp or 50 amino acids were selected for further study. Full-length genes were selected based on the presence of both transcription start and termination sites in the prediction results for the study.

Diamond BLAST was used to search for the homologs of predicted genes from the NCBI non-redundant database (Buchfink et al. 2021). Blast2GO was used for the assessment of gene ontology (GO) terms and function assignment (Conesa and Götz 2008). A Hidden Markov Model (HMM) based tool, pfamscan, was used for protein domain identification using the PfamA database (Mistry et al. 2007). OrthoFinder was used to detect the orthologs (Emms and Kelly 2019). UpsetR was used to visualize orthogroups (Conway et al., 2017). Mercator4 was used to assign MapMan functional annotations and pathways to *Ae. geniculata, Ae. umbellulata*, chr5A, chr5B, and chr5D (Schwacke et al., 2019) . Transcription factors were annotated using PlantTFDB (Jin et al., 2017). The alignment program BWA was used to align $5M^{g}$ and $5U^{u}$ reads to the Chinese Spring genome (Li and Durbin 2010). The sam files were converted to bam format using samtools (Li et al., 2009). BCFtools was used for SNP calling and filtration (Danecek et al., 2021). SNPeff was used to determine the effect of mutations on the gene structure (Cingolani et al., 2012). Variant files were merged with the exome capture data from 890 wheat exomes (He et al. 2019). Plink and ggplot2 were used to generate a PCA plot from the merged vcf file for chr5D (Wickham 2016). A syntenic map of 10Mb introgressed region was prepared using syntenyPlotteR (Farré et al., 2019). SSR markers were identified using MISA (Thiel et al. 2003). Primer3 was used to design primers for the SSR marker (Koressaar and Remm 2007). Polymarker was used to design Kompetitive allele-specific PCR (KASP) markers for $5M^{g}$ and $5U^{u}$ SNPs mapped on chromosome 5D of CS (Ramirez-Gonzalez et al., 2015).

Comparison of gene structure between 5M^{g}, 5U^{u,} and wheat

Two approaches were used to compare the gene structure between 5M^{g}, 5U^{u,} and 5D. In the first approach, single copy-orthologs were compared for the differences in gene structure based on the number of exons, number of introns, and gene length. In the second approach, full-length genes of 5M^{m} and 5U^{u} were mapped to the chromosome 5D of CS using exonerate program (Slater & Birney, 2005). The mapped fragments were filtered based on exonerate similarity score >1000. Genes mapping to a genomic region longer than 15kb were removed from the comparison.

Phylogenetic analysis

For the phylogenetic study, a species tree was generated using the single copy-genes based on the conserved gene alignments of the species i.e., *Oryza sativa* spp. japonica, *Brachypodium distachyon*, *Hordeum vulgare*, *Secale cereale*, *T. urartu*, *T. turgidum*, *T. turgidum* ssp. *dicoccoides*, *Ae. tauschii,* D genome of hexaploid wheat species (*T. aestivum* cv. Chinese Spring, *T. aestivum* cv. Arinalrfor, *T. aestivum* cv. Jagger, *T. aestivum* cv. Julius, *T. aestivum* cv. Lancer, *T. aestivum* cv. Landmark, *T. aestivum* cv. Mace, *T. aestivum* cv. Mattis, *T. aestivum* cv. Norin61, *T. aestivum* cv. Stanley, *T. spelta).* Alignment free distance-based method implemented in JolyTree (Criscuolo, 2019) was used to generate phylogenetic tree from the chromosome level alignment of 5M^{g} and 5U^{u} with the group 5D chromosome of different wheat cultivars from the 10+ wheat genome project (Walkowiak et al., 2020).

Evolution of gene families

Orthologous assignment of the Orthofinder results was converted into a presence-absence matrix using the orthocount2bin program (Natsidis et al. 2021). Gene families were classified *via* all-vs-all BLASTp search of 5M^{g}, 5U^{u}, 5D (Chinese spring), and 5D chromosomes of *Ae. tauschii* (Camacho et al. 2009). BLASTp results were clustered into gene families using MCL (Enright et al. 2002), and family size counts were determined. The species tree generated by Orthofinder was converted into an ultrametric tree. Computational Analysis of gene Family Evolution (de Bie et al. 2006) was used to identify gene families that have undergone significant expansion or contraction in the genome. Gene family evolution was studied for the defined classes of resistance genes and transcription factors.

Differential expression analysis of NIL carrying 5M^{g} introgression.

To compare the structure, expression, and sequence level variations between wheat and its homoeologous chromosomes (with respect to 5M^{g}), RNAseq expression data of small translocations of chromosome 5M^{g} on wheat chromosomes 5D in leaf rust and stripe susceptible wheat cultivar WL711 from previous studies on leaf rust (Yadav et al. 2016) and a second RNAseq dataset was provided by Dr. Parveen Chhuneja (personal communication) was used. The susceptible wheat cultivar WL711 and two introgression lines, pau16062 (syn. T756, WL711 + *Lr57/Yr40*) and pau16055 (syn. T598, WL711 + *Lr57/Yr40*), in the WL711 background, were inoculated with leaf rust. RNA samples were collected at 6-time intervals after inoculation: 0 HPI (control), 12, 24, 48, 72, and 96 HPI. Hisat2 was used to map filtered reads to the Chinese Spring (CS) genome (Kim et al. 2019). FeatureCounts was used for populating expression

counts (Liao et al. 2014). Differential expression analysis was performed at respective HPI between the introgression lines (T756 and T598) and the susceptible cultivar WL711 using edgeR (McCarthy et al. 2012). WEGO was used to plot the gene ontology (Ye et al., 2018). Gene ontology enrichment was performed using the goseq package (Young et al., 2010).

3.4 Results

Sequence assembly

Chromosome assembly in the present study was performed using the 40 Gb (53X coverage) high-quality filtered data of *Ae. geniculata* ($5M^g$ chromosome) downloaded from NCBI and 14.61 Gb (18X coverage) data of *Ae. umbellulata* ($5U^u$ chromosome). The fragment size for $5M^g$ reads was 500 bp, while the fragment size for $5U^u$ reads was 600 bp. The highest N50 was obtained at k71 and k61 for $5M^g$ and $5U^u$, with an N50 of 1.078 kb and 559 bp, respectively. For chromosomes $5M^g$ and $5U^u$, a total of 675 Mb and 499 Mb of the sequence was assembled, respectively. The contigs with a length of more than 500 bp were selected, resulting in 397 Mb and 243 Mb of sequence for $5M^g$ and $5U^u$, respectively (Table 3.1). BUSCO-based assembly evaluation reported 736 and 383 complete BUSCO genes in $5M^g$ and $5U^u$, respectively. (Figure S3.1). A total of 79 and 63 percent repetitive elements were detected using the hierarchical repeat identification approach; these included 77 and 60 percent repeats that were specific to *Poaceae* and 1.75 and 2.67 percent *de novo* repeats, in two assembled chromosomes $5M^g$ and $5U^u$, respectively (Table S3.1). Scaffolding using Ragoo (Alonge et al. 2019) resulted in 284 Mb of scaffolded contigs for $5M^g$, whereas for $5U^u$ only 63 Mb of the contigs were scaffolded (Figure 3.1A).

Table 3.1 Sequence assembly statistics of *Ae. geniculata* (5M^{g}) and *Ae. umbellulata* (5U^{u})

	***Ae. geniculata* (5M)**				***Ae. umbellulata* (5U)**			
Assembly	**Kmer 71**		**Gapfilling (>500bp)**		**Kmer 61**		**Gapfilling (>500bp)**	
Total no. of scaffold	1,854,032		202,327		1,608,425		219,000	
Total no. of contigs	2,619,128		222,657		1,831,974		315,403	
Genome in scaffold sequence	675.070 Mb		397.503 Mb		499.800 Mb		243.154 Mb	
Genome in contig sequence	595.063 Mb (11.852% gap)		395.173 Mb (0.586% gap)		413.541 Mb (17.259% gap)		215.904 Mb (11.207% gap)	
Main genome scaff old N50/L50:	98246/1.078 Kb		27129/3.404 Kb		204532/559		57394/1.134 Kb	
Max scaffold lengt h:	55.822 Kb		55.666 Kb		31.308 Kb		30.444 Kb	
Max contig length:	32.207 Kb		55.666 Kb		16.486 Kb		30.443 Kb	
Contig length	No of scaffold	No of bases	No of scaffold	No of bases	No of scaffold	No of bases	No of scaffold	No of bases
All	1,854,032	675,070,048	20,232,700	397,502,599	1,608,425	499,799,616	21,900,000	243,153,551
1 Kb	105,913	345,500,735	102,022	328,272,111	87,787	165,760,371	76,064	141,598,981
25 Kb	299	9,072,840	279	8,438,200	7	191,104	5	137,978
50 Kb	5	271,480	5	267,609	-	-	-	-
Repetitive elements								
			Repetitive base	**percentage**			**Repetitive base**	**percentage**
Poaceae specific repeats			308,149,023	77.52 %			148,107,104	60.91 %
Novel *de novo* repeat			6,947,009	1.75%			6,488,548	2.67%
Total masked based			315,096,032	79.20 %			154,595,652	63.58 %

Gene prediction and annotation

Gene prediction tool AUGUSTUS was used to perform *de novo* gene prediction in scaffolds larger than 1 Kb. In total, 9,702 and 7,258 genes were predicted in 5M^{g} and 5U^{u}, respectively (Table 3.2). Homologs were detected for 82% of genes in the two assemblies. Gene ontology (GO) terms were identified for 5,504 (57%) and 4,199 (58%) genes; KEGG enzymes were annotated for 2,389 (24%) and 1,952 (27%) genes,

and protein domains were detected for 4,836 (50%) and 3,431 (48%) genes in 5M^{g} and 5U^{u}, respectively. Due to significant variation in total number of genes across the assemblies, the study narrowed its focus to a subset of 5,100 and 2,592 full-length genes in 5M^{g} and 5U^{u}, respectively.

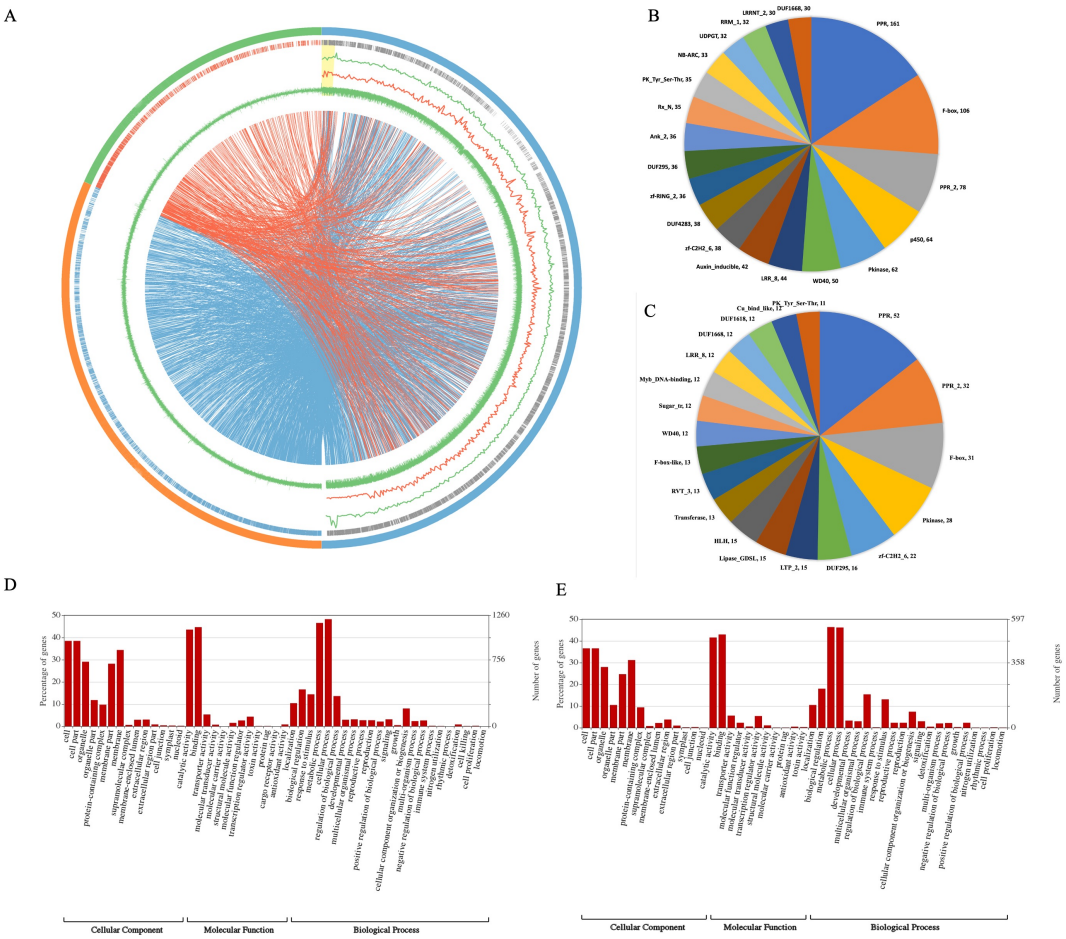


Figure 3.1 A) Chromosome similarity plot of *Ae. geniculata* and *Ae. umbellulata* chromosomes in comparison to chromosome 5D of wheat. Outer circle: 5D (blue),5M^{g} (orange), 5U^{u} (green). Second circle shows the gene density. Line graphs 5M^{g} (green), and 5U^{u} (orange) show SNP density on 5D. GC content (inner green circle). Connecting lines display orthologous genes. B) Top 20 protein family domains of *Ae. geniculata* (5M^{g}) and C) *Ae. umbellulata* (5U^{u}). Classification of genes in gene ontology (GO) classes D) *Ae. geniculata* (5M^{g}) and E) *Ae. umbellulata* (5U^{u}).

Table 3.2 Genes predicted in *Ae. geniculata (*5M^{g}) and *Ae. umbellulata* (5U^{u}) using augustus.

	***Ae. geniculata* (5M^{g})**		***Ae. umbellulata* (5U^{u})**	
Total no of predicted genes	9,702		7,258	
Full length genes with transcription start site and termination site	5,100		2,592	
Functional annotation				
	No. of gene	Full length gene	No. of gene	Full length gene
Blast homologs	8,123	4,294	5,882	1,399
Gene Ontology	5,504	2,918	4,199	1,399
Enzyme	2,389	2,521	1,952	1,195
Pfam	4,836	4,302	3,431	1,730

A total of 4,294 (84%) and 1,399 (54%) full-length genes were found to have BLAST homologs in 5M^{g} and 5U^{u}, respectively. These genes also included the full-length BUSCO genes. Protein family domains were identified for 4,302 (84%) and 1,730 (66%) genes in two chromosomes (Figure 3.1B and 3.1C). The major protein classes consisted of PPR, F-box, and protein kinase. Gene ontology classes were assigned to 2,918 (57%) and 1,399 (54%) of the genes in 5M^{g} and 5U^{u}, respectively (Figure 3.1D and 3.1E). In both chromosomes, the major portion of the genes were found to be involved in cellular activity, metabolic processes, binding, and catalytic activity. Additionally, it was discovered that both chromosomes were enriched for the "response to stimulus" category of genes. Nearly 2,521 (49%) and 1,195 (46%) genes in 5M^{g} and 5U^{u}, respectively, were assigned to KEGG enzymes (Table S3.2).

Transcription factor and Resistance gene

In the 5M^{g} and 5U^{u} chromosomes, 176 and 109 transcription factors (TF) belonging to 33 and 28 TF classes were identified. There were approximately 38%

more TF classes on chromosome 5M^{g} compared to chromosome 5U^{u} (Figure 3.2A, Table S3.6). The majority of TFs belong to C2H2, bHLH, B3, MYB-related, bZIP, M-type MADS, WRKY, and ERF classes and constitute 67% and 60% of the TF classes in 5M^{g} and 5U^{u}, respectively. Protein family annotations based on NB-ARC and LRR domains revealed 78 (out of total 136) and 21 (out of total 80) full-length R-genes in 5M^{g} and 5U^{u}, respectively.

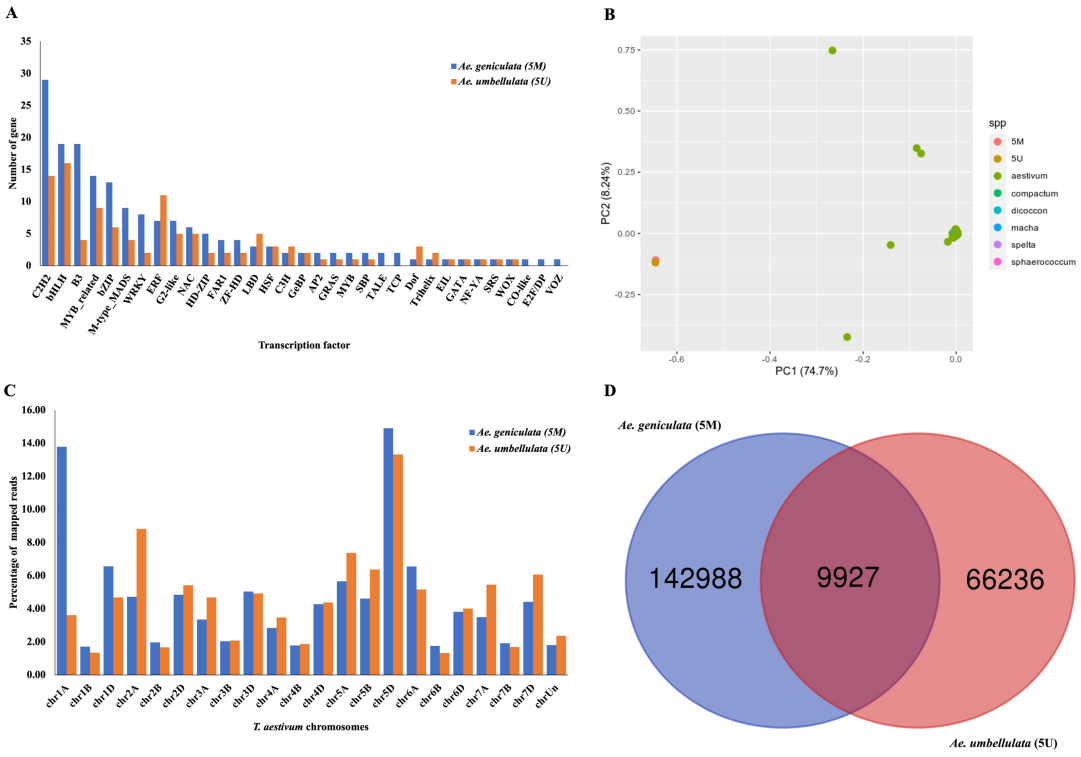


Figure 3.2 A) Distribution of transcription factor classes identified from PlantTFDB. B) Principal component analysis (PCA) of *Ae. geniculata* (5M^{g}) and *Ae. umbellulata* (5U^{u}) with tetraploid and hexaploid wheat accession. C) Percentage alignment statistics of 5M^{g} and 5U^{u} reads on the complete genome of Chinese spring displaying partial read mapping on chromosomes other than 5D. D) Shared high-quality homozygous SNPs of *Ae. geniculata* (5M^{g}) and *Ae. umbellulata* (5U^{u}).

Relationship of 5M^{g} and 5U^{u} chromosomes with homoeologous bread wheat chromosomes

Biologically related organisms share orthologous genes with similar functions. Chromosome 5M^{g} shared 1,400, 2,844, 3,223, and 3,266 homoeologous genes with the chromosomes 5U^{u}, 5A, 5B, and 5D, while the chromosome 5U^{u} shared 1618, 1740, 1614, and 1604 orthologs with the chromosomes 5M^{g}, 5A, 5B, and 5D. The number of shared orthogroups between the five chromosomes (5M^{m}, 5U^{u}, 5A, 5B, and 5D) are shown in figure 3.3A. There were 878 shared orthogroups between 5M^{g}, 5U^{u}, 5A, 5B, and 5D chromosomes, where 77 and 21 orthogroups were specific to 5M^{g} and 5U^{u} chromosomes, respectively. A total of 140 orthogroups were common between the 5M^{g} and 5U^{u} genomes. While *Ae. umbellulata* shared 234 orthogroups with group 5 chromosomes of wheat, *Ae. geniculata* shared the highest number of orthogroups (1,090) with A, B, and D genomes. There were 514 single-copy orthologs between the five chromosomes. Singleton representation constituted 893 and 573 genes in *Ae. geniculata* and *Ae. umbellulata.* The orthologous relationship of 5M^{g} and 5U^{u} chromosomes with *Ae. tauschii* and ten other wheat cultivars are shown in Figure 3.3B (M.C. Luo et al., 2017; Walkowiak et al., 2020).

Comparison of 5M^{g} and 5U^{u} chromosome sequences with wheat chromosome 5D

To see the divergence of the tertiary gene pool from the available germplasm resource, 5M^{g} and 5U^{u} were compared with the exome capture data from a previous study (He et al., 2019). In the comparison with group 5D chromosomes from 890 diverse hexaploid and tetraploid wheat accessions, *Ae. geniculata* and *Ae. umbellulata* clustered into a separate group, as depicted in Figure 3.2B**.** Based on the results of read

mapping onto the whole wheat genome, the highest percentage of $5M^{g}$ reads were located on chr5D and chr1A, whereas the highest percentage of $5U^{u}$ reads mapped to chr5D and chr2A. Significantly lower read mapping was observed on the B genome (Figure 3.2C). According to phylogenetic studies and whole-genome alignment of reads, $5M^{g}$ and $5U^{u}$ chromosomes showed the most resemblance with group D, lesser with group A, and the least similarity to the B genome. Figure S3.2 displays the coverage of uniquely mapped reads of *Ae. geniculata* and *Ae. umbellulata* on chromosome 5D of Chinese Spring. Both the chromosomes showed higher coverage of mapped reads in a 1Mb window. A dip in the coverage was observed at 150 Mb on chromosome 5D for the two aligned chromosomes. The distal region on the long arm of chromosome 5D displayed differential coverage for *Ae. geniculata* and *Ae. umbellulata*. The elevated region in the coverage plot indicates ancient introgression of the $5M^{g}$ fragment on chromosome 5D of the Chinese Spring, supported by the lower coverage for the corresponding segment from $5U^{u}$ of *Ae. umbellulata.*

From the read mapping, 7,776,990 (2.57 %) of the concordantly paired reads of $5M^{g}$ and 3,625,329 (3.47 %) reads of $5U^{u}$ mapped on the coding region of Chinese Spring (170Mb,1.17% of genome). Chr5D showed the highest coverage of reads mapped on the coding region, followed by chr5A and chr5B. Chr7D and Unassigned chromosome sequence exhibited good coverage of the $5U^{u}$ read, while Chr1A showed higher coverage of the $5M^{g}$ read (Figure S3.3). These results indicate genetic similarity and hybridization. A higher percentage of read mapping was observed in the intergenic region compared to the genic region on chromosome 5D. Read mapping to the coding region accounts for 2.45% at $5M^{g}$ and 3.08% at $5U^{u}$.

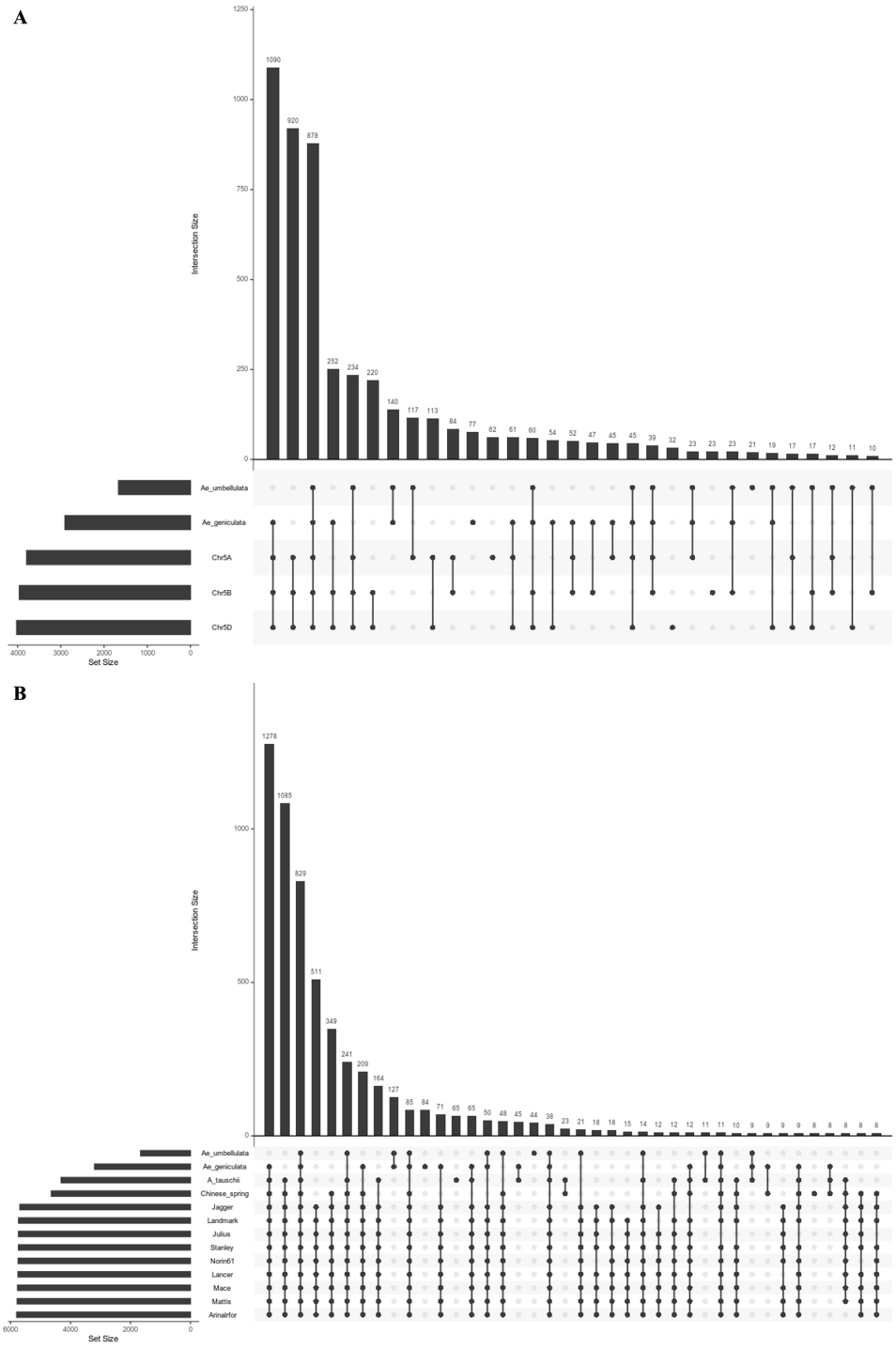


Figure 3.3 Upset plot of orthologous groups A) Orthogroup between *Ae. geniculata* (5M^{g}), *Ae. umbellulata* (5U^{u}), chr5A and chr5B and chr5D of Chinese spring. B) Orthogroup between *Ae. geniculata* (5M^{g}), *Ae. umbellulata* (5U^{u}), *Aegilops tauschii, T. aestivum* cv. Chinese Spring, *T. aestivum* cv. Arinalrfor, *T. aestivum* cv. Jagger, *T.*

aestivum cv. Julius*, T. aestivum* cv. Lancer*, T. aestivum* cv. Landmark*, T. aestivum* cv. Mace*, T. aestivum* cv. Mattis*, T. aestivum* cv. Norin61*, T. aestivum* cv. Stanley.

There were 249,041 and 114,485 high-quality SNPs identified for $5M^{g}$ and $5U^{u}$, respectively, with indels accounting for a larger portion (24%) of the detected variations. In $5M^{g}$, 25 % of SNPs were heterozygous compared to only 15 % in $5U^{u}$ (Table 3.3). Figure 3.2D displays the shared and unique SNPs between the two chromosomes mapped to chromosome 5D of Chinese Spring. Out of the total 5,545 high-confidence genes present on chromosome 5D, 5,090 and 4,482 genes were affected by the SNPs from the $5M^{g}$ and $5U^{u}$ reads, respectively. The majority of constituent variants were missense variants, accounting for 35% of all variant sites. An estimated 4 to 5 % of the proteins were affected by the "moderate effect", which leads to the non-disruptive variant that could alter protein efficacy. These non-disruptive variants may play an important role and may result in lowering the gene efficiency or may make a gene less effective to pathogens in wheat. A moderate impact on gene function was observed for 3,152 and 2,329 genes of 5D, impacted by $5M^{g}$ and $5U^{u}$ reads, respectively. There were 49 and 46 resistance genes impacted by $5M^{g}$ and $5U^{u}$ SNPs, respectively.

Differences in the gene structure between wheat and tertiary gene pool

Gene structure statistics of $5M^{g}$, $5U^{u}$, 5A, 5B, and 5D chromosomes are shown in table 3.4. The longest genes in $5M^{g}$ and $5U^{u}$ were 22 kb long, whereas those in Chinese Spring vary from 47 kb to 63 kb. The longest intron in $5M^{g}$ was 7.3 kb, in $5U^{u}$ it was 5.2 kb, and in Chinese Spring intron length ranged from 34 to 44 kb. Lesser

number of exons and introns were observed per gene in 5M^{g} and 5U^{u} compared to that of Chinese Spring.

Table 3.3 Single-nucleotide polymorphism (SNP) based on the *Ae. geniculata* (5M^{g}) and *Ae. umbellulata* (5U^{u}) chromosome-specific reads mapping to chromosome 5D of Chinese spring.

	***Ae. geniculata* (5M^{g})**		***Ae. umbellulata* (5U^{u})**	
	Number of identified SNP	**Number of homozygous alternate allele**	**Number of identified SNP**	**Number of homozygous alternate allele**
Total number of variant sites	249,041	152,915	114,485	76,163
SNPs	200,654	152,915	91,715	76,163
Indels	48,387	0	22,770	0
SNP Transitions/Transversions	1.74 (229667/131822)	1.75 (194622/111208)	1.72 (108316/63054)	1.72 (96338/55988)

Fewer intron per mRNA and smaller mean intron length, suggest a shorter gene length in alien germplasm compared to that of Chinese spring. Further investigation of the homoeologous gene structure between the chromosomes revealed 712 orthologs with an identical number of exons from 1,466 single-copy orthologs between 5M^{g} and 5D. Nearly, 61 out of 712 genes showed identical protein lengths in both 5M^{g} and 5D. Thirteen genes showed a difference of more than 200 bp and up to 900 bp, whereas the length difference for 443 proteins was less than 200 bp. Between 5U^{u} and 5D, OrthoFinder detected 615 single-copy orthologs, 286 of which showed the same number of exons, where 93 were of equal length. For 40 of these genes, the length difference was less than 200bp, while the difference for the remaining 53 genes ranged from 200 to 900bp. For the single-copy orthologs, gene length varied significantly between the two genomes, ranging from 2 bp to 7 Kb.

The similarity percentage between the orthologs varied from 25 to 100%. More than 90 percent similarity was observed for 69% and 57% of genes in $5M^g$ and $5U^u$ with the CS genes, respectively. In comparison to the homoeologous genes of CS for $5M^g$, 753 genes showed shorter transcript length and 692 showed longer transcript length, whereas in $5U^u$, 398 genes showed shorter, and 252 genes showed longer transcript length. Among the single exon genes, 127 genes were shorter, and 161 genes were longer in $5M^g$ whereas, in $5U^u$ 132 genes showed shorter and 520 genes showed larger genes length compared to CS homologs. For multi-exon genes, more genes were found with comparatively smaller genes length ($5M^g$: 625, $5U^u$:255) compared to genes with larger lengths ($5M^g$: 520 higher, $5U^u$: 167 higher) in both the studied chromosomes with respect to CS genes (Table S3.3).

Out of 5,100 genes, a total of 4,031 genes were mapped on 5D using exonerate. Similarity-based filtering selected 2,355 genes, out of which finally 1,946 genes were used for analysis after removing erroneous alignments of larger than 15kb. For 1,673 genes, the number of exons in $5M^g$ matched exactly with the predicted number of exons from exonerate mapping. About 1,017 genes showed an equal number of exons between the exonerate mapping result and CS gene annotations. Out of these, 524 genes showed longer gene lengths in $5M^g$ and 493 demonstrated higher gene length in CS (Table S3.4). Concordance of the results was observed for 1,426 genes from similarity search and exonerate mapping results out of 1,446 single-copy genes compared. Although our results were based on gene length comparison of single copy-orthologs in a single chromosome study, these results indicate differences in gene structure between cultivated wheat and distinct wild relatives (Figure 3.4).

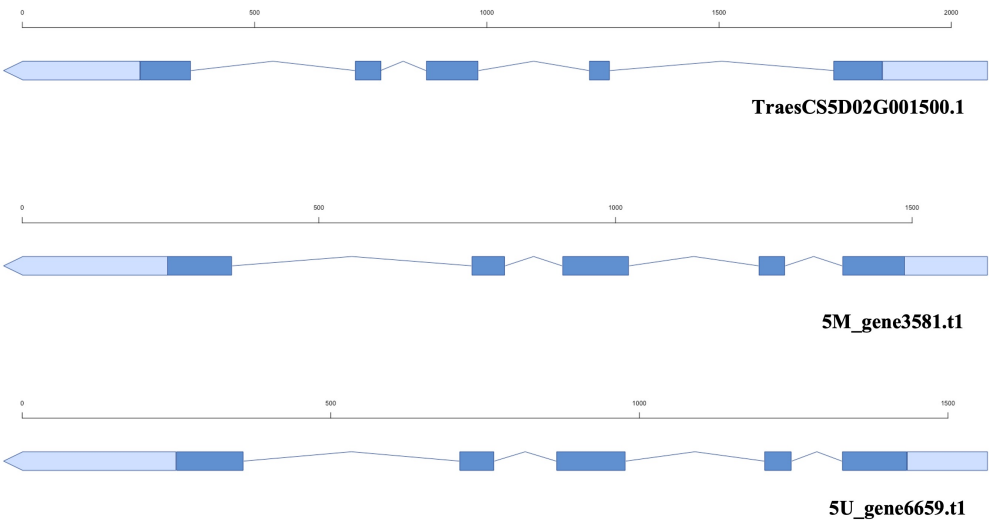
Table 3.4 Gene structure comparison of *Ae. geniculata* ($5M^g$) and *Ae. umbellulata* ($5U^u$) chromosomes with group 5 chromosomes of Chinese spring.

	***Ae. geniculata* ($5M^g$)**	***Ae. umbellulata* ($5U^u$)**	**CS (chr5A)**	**CS (chr5B)**	**CS (chr5D)**
Number of full length gene	5100	2592	5429	5576	5563
Total number of exons	16885	6716	25836	26364	26015
Total number of introns	11129	3905	20387	20762	20437
Number of single exon gene	1836	1279	1316	1361	1392
Mean exons per mRNA	3.3	2.6	4.8	4.7	4.7
Mean introns per mRNA	2.2	1.5	3.8	3.7	3.7
Total gene length (bp)	12473015	4579550	18449098	19962808	18985249
Total exon length (bp)	6911779	2977095	8372686	8984820	8953505
Total intron length (bp)	4695503	1378321	9970751	10884969	9908071
Mean gene length (bp)	2445	1766	3398	3580	3412
Mean exon length (bp)	409	443	324	340	344
Mean intron length (bp)	421	352	489	524	484
Length of longest gene (bp)	22527	22543	47936	63855	58200
Length of longest exon (bp)	4322	3853	10983	6448	7292
Length of longest intron (bp)	7380	5261	38985	34008	44713
Length of shortest gene (bp)	351	325	108	105	105
Length of shortest exon (bp)	5	7	1	1	1
Length of shortest intron (bp)	60	60	2	2	2

Phylogenetic Analysis

Species tree based on the conserved gene alignments of the species i.e., *Oryza sativa* spp. japonica, *Brachypodium distachyon*, *Hordeum vulgare*, *Secale cereale*, *T. urartu*, *T. turgidum*, *T. turgidum* ssp. *dicoccoides*, *Ae. tauschii,* D genome of hexaploid wheat species (*T. aestivum cv.* Chinese Spring, *T. aestivum* cv. Arinalrfor, *T. aestivum* cv. Jagger, *T. aestivum* cv. Julius, *T. aestivum* cv. Lancer, *T. aestivum* cv. Landmark, *T. aestivum* cv. Mace, *T. aestivum* cv. Mattis, *T. aestivum* cv. Norin61, *T. aestivum* cv.

Stanley, *T. spelta)* demonstrated close clustering of *Ae. geniculata* and *Ae. umbellulata* with the D genome group (Figure 3.5A).



Gene id	Gene length	No . of exons	Total exon length	No. of introns	Total intron length	No. of CDS	Total CDS length
TraesCS5D02G001500.1	2079	5	907	4	1172	5	426
5M_gene3581.t1	1628	5	811	4	817	5	426
5U_gene6659.t1	1565	5	807	4	758	5	426

Figure 3.4 Comparison of gene structure in Chinese spring, *Ae. geniculata* and *Ae. umbellulata*.

Genes are the most conserved elements during evolution compared to the intergenic regions and serve as the main driving force for species evolution. Intergenic regions consist of repetitive elements and evolve at different rates compared to their genic counterparts. In the absence of the whole genome sequence, chromosome alignment of $5M^{g}$ and $5U^{u}$ with the group 5D chromosome of different wheat cultivars from the 10+ wheat genome project was done (Walkowiak et al., 2020). The Kmer-based phylogenetic tree reveals the closeness of the M and U chromosomes to the D group of chromosomes (Figure 3.5B). The k-mer-based tree also provides more resolution separating the M and U chromosomes at branches. There was a positive

correlation between the phylogenetic tree and whole-genome alignments of the $5M^g$ and $5U^u$ reads to the wheat genome, where a higher percentage of read mapping was observed for the A genome compared to that of the B genome. These two observations support the closeness between M/U and A genomes.

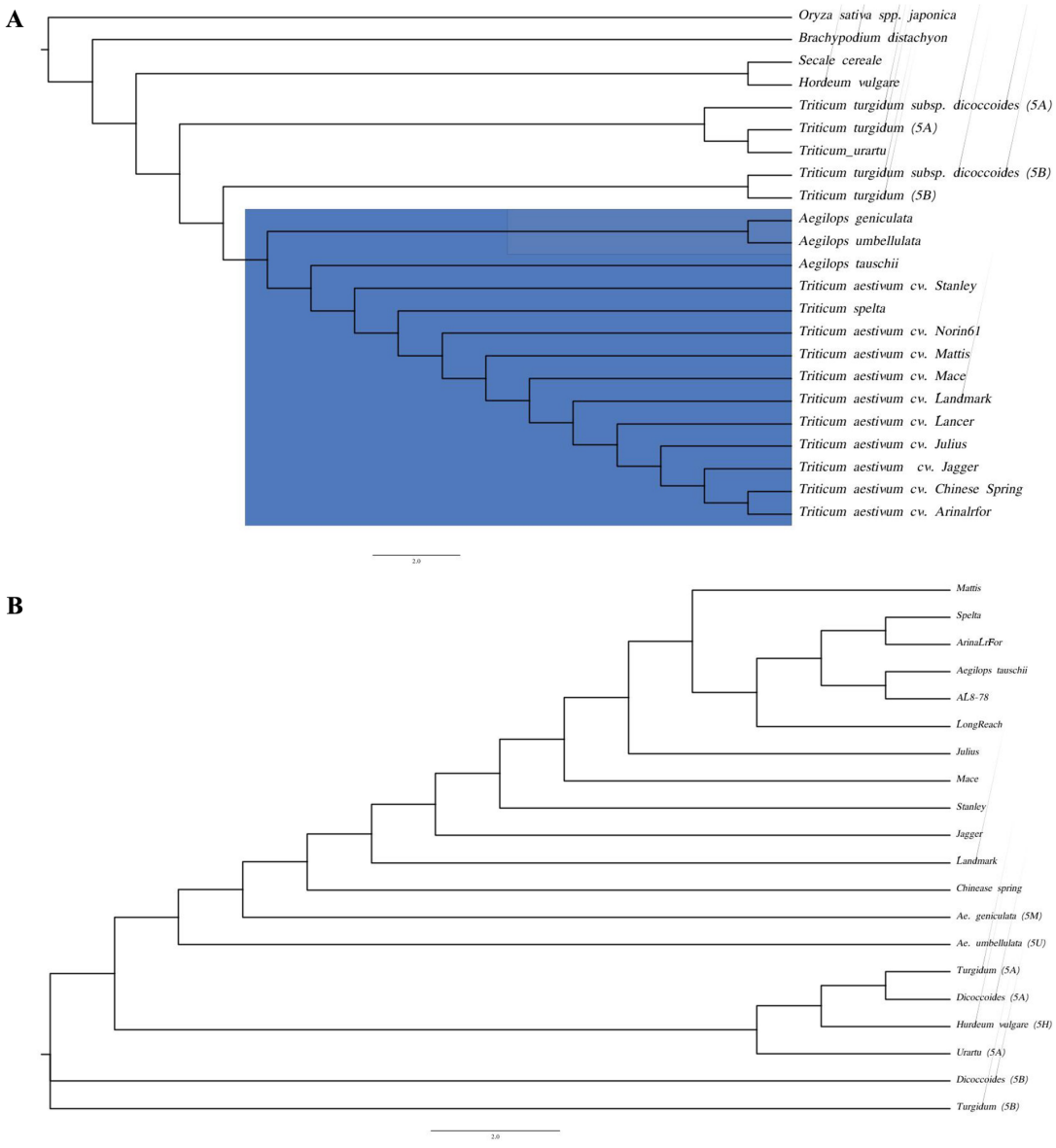


Figure 3.5 Phylogenetic trees based on homologous sequence A) *Ae. geniculata* ($5M^g$) and *Ae. umbellulata* ($5U^u$) species tree based on monocot genomes. B) Kmer based species tree of wheat relatives.

Evolution of gene families

The evolution of gene families shaped the present genomic landscape. The evolution of resistance genes and transcription factors was examined in the $5M^g$ and $5U^u$ chromosomes. Gene gain and loss events can be inferred based on the presence/absence of orthologs in the clades. It is believed that these events reflect origins or the loss of key characteristics within those clades (Natsidis et al., 2021). Figure 3.6A shows the heatmap of the presence/absence of R genes. In the presence of A and B genomes of tetraploid species, R genes formed two distinct clusters, where the *Ae. umbellulata* profile clustered with the A group of chromosomes and *Ae. geniculata* clustered with the D group of chromosomes (Figure 3.6B**)**. In the absence of a tetraploid genome, *Ae. geniculata* and *Ae. umbellulata* together formed a group distinct from wheat and *Ae. tauschii*. In the case of transcription factors, *Ae. geniculata* and *Ae. umbellulata* formed a separate group from the rest of the studied genomes. The TF profile of $5M^g$ and $5U^u$ chromosomes was more similar to the tetraploid genome. A comparison of only the D-group of chromosomes also showed a distinct cluster for these genomes, where *Ae. tauschii* branched out separately (Figure 3.6C, D**)**.

This study compared gene evolution patterns among four species i.e., *Ae. geniculata*, *Ae. umbellulata*, *T. aestivum cv*. Chinese Spring and *Ae. tauschii.* Expansion of the TF gene families was observed on all the chromosomes. Figure S3.4 displays expanded and rapidly evolving TF families. Nealy 16 and 36 TF gene families showed expansion in *Ae. umbellulata* and *Ae. geniculata,* respectively. For the R genes, 12 and 8 gene families showed expansion, whereas 12 and 4 gene families were found

to be rapidly evolving in *Ae. geniculata* and *Ae. umbellulata*, respectively. The contraction was observed for two of the gene families.

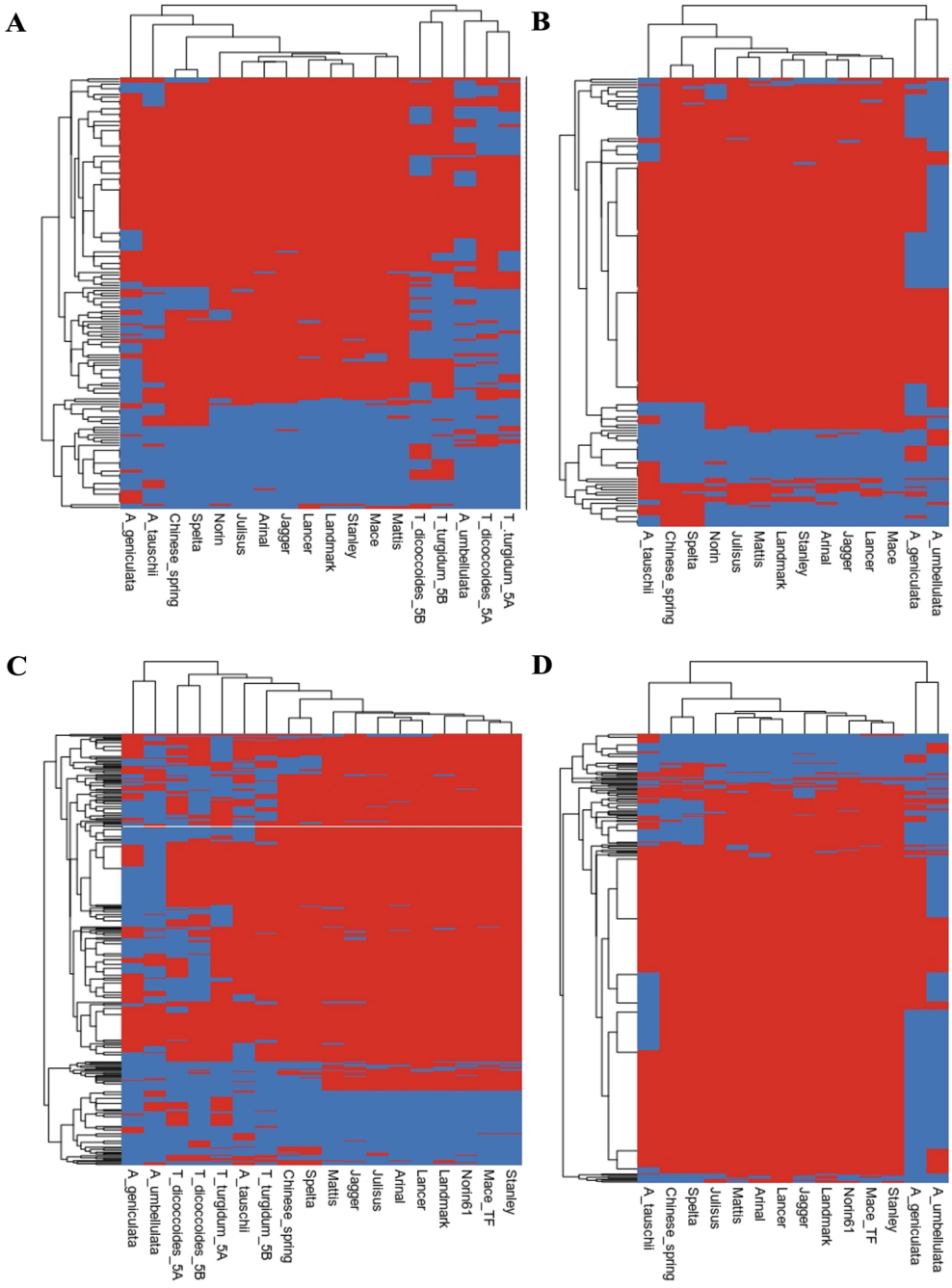


Figure 3.6 Heatmap inferring the presence and absence of gene families of A, B) resistance genes and C, D) transcription factors.

Genes involved in ETI and PTI response.

Mercator4 was used to assign MapMan functional annotations and pathways to *Ae. geniculata, Ae. umbellulata*, chr5A, chr5B, and chr5D. Genes involved in response to biotic stress for effector-triggered immunity (ETI) and pattern-triggered immunity (PTI) are shown in Table 3.5. These classifications are important as these will help us in categorizing the gene in response to pathogen attacks based on their similarity to known genes in Arabidopsis and rice. Chromosome 5D was found to be more enriched with pathogenicity-related genes compared to other chromosomes, owing its origin to *Ae. tauschii.* Comparatively fewer genes were reported for $5M^{g}$ and $5U^{u}$ chromosomes, while these chromosomes are well known to be a rich source of resistance genes. Chromosomes 5B and 5D displayed more genes with the "putative disease resistance protein" annotation compared to $5M^{g}$ and $5U^{u}$. A possible explanation may be the presence of new sources of resistance in these two genomes, while most of the genes in CS may have lost their functional significance with sequence divergence. This requires further exploration through whole genome studies. Figure 3.7 displays the putative genes involved in pathogen response to ETI and PTI response along with putative resistance genes.

Table 3.5 Mapman-based assignment of functional gene categories of effector-triggered immunity (ETI) and pattern-triggered immunity (PTI) response.

	Category	Category	5M[g]	5U[u]	5A	5B	5D
Pattern-triggered immunity (PTI)	Nematode elicitor response	receptor protein kinase (NILR)	0	0	0	0	2
	Fungal elicitor response	protein kinase (PBL27/RLCK185)	0	0	0	0	3
	Fungal elicitor response	chitin receptor protein kinase (CEBiP)	1	2	2	1	0
	Bacterial elicitor response	protein kinase (PCRK)	0	0	0	0	1
		FLS2-BAK1 flagellin receptor complex.flagellin receptor protein kinase component (FLS2)	0	0	0	0	1
		FLS2-BAK1 flagellin receptor complex.dynamically associated protein kinase (BIK1)	1	0	1	1	0
		FLS2-BAK1 flagellin receptor complex.co-receptor kinase component (BAK1)	0	0	0	0	1
	network.bacterial elicitor response.protein kinase (BSK1)	network.bacterial elicitor response.protein kinase (BSK1)'	0	0	1	0	0
Effector-triggered immunity (ETI)	TNL-mediated effector-triggered immunity	regulatory protein (ADR)	1	0	1	1	0
		posttranscriptional regulatory protein (FPA)	1	0	1	1	0
	RIN4-RPM1 immune signalling	regulatory factor (RIN4) guarded by RPM1/RPS2 activities	1	0	0	1	1
		CC-NLR-type effector receptor (RPM1)	0	0	0	0	1
	regulatory protein (RAR1)	regulatory protein (RAR1)	0	0	0	0	1
	immunity suppressor (SOBER/TIPSY)	immunity suppressor (SOBER/TIPSY)	0	0	0	0	3
	network regulatory protein (EIJ1)	network regulatory protein (EIJ1)	0	0	1	1	1
	EDS1-PAD4/SAG101 signalling heterodimers	component EDS1	1	1	1	1	0
Systematic acquired resistance	regulatory protein (CBP60/SARD)	regulatory protein (CBP60/SARD)	5	1	7	9	1
	pipecolic acid metabolism	pipecolate oxidase (SOX)	4	2	3	4	0
	NPR1-interactive transcription factor (TGA)	NPR1-interactive transcription factor (TGA)	0	0	0	0	1
	pathogen polygalacturonase inhibitor (PGIP)	pathogen polygalacturonase inhibitor (PGIP)	2	1	1	1	3
	defensin activities.defensin (PDF2)	defensin (PDF2)	1	2	4	1	3
Unassigned	not assigned.annotated	not assigned.annotated	1	5	2	0	0

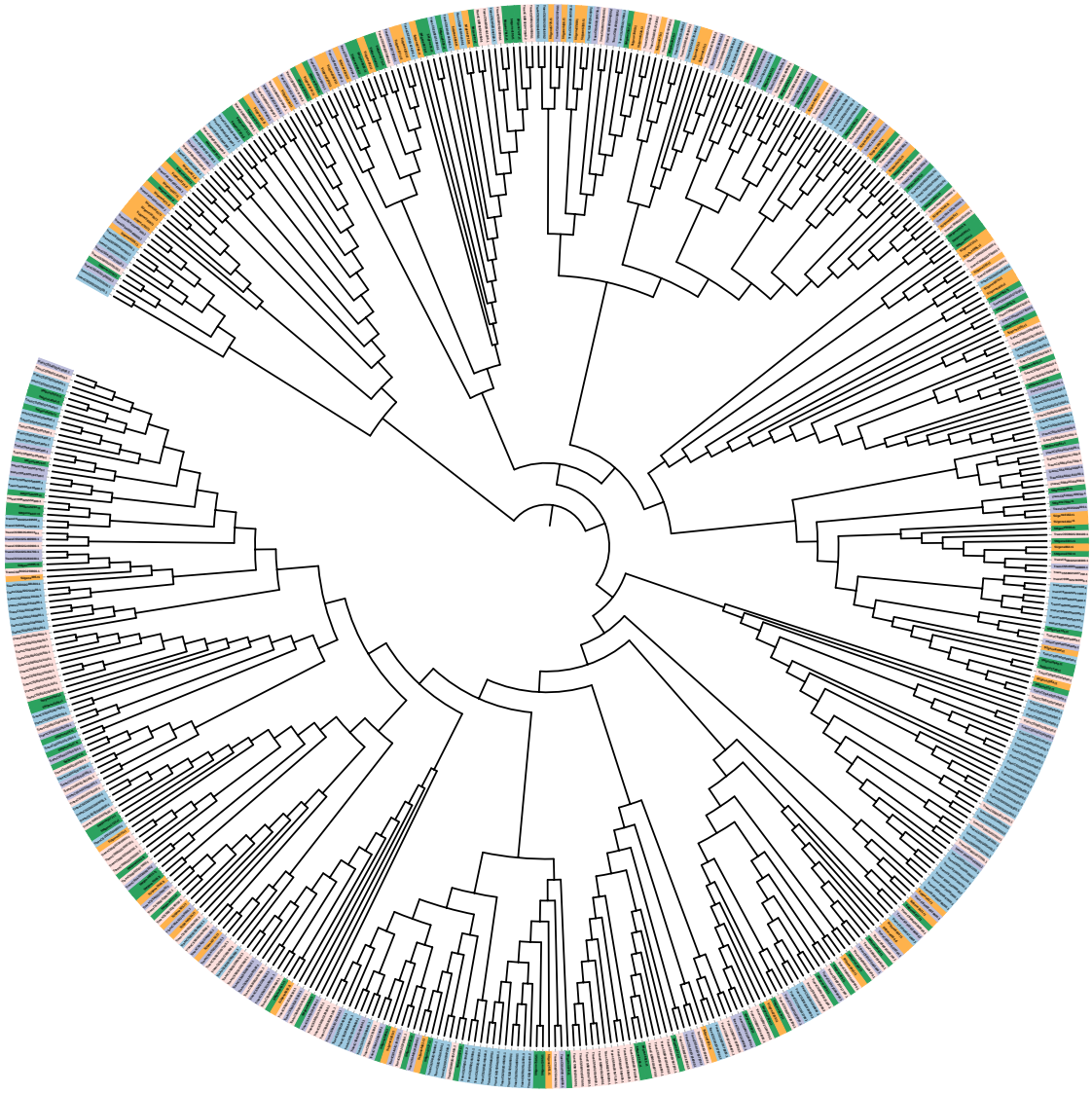


Figure 3.7 Mapman-based phylogenetic tree of genes involved in ETI and PTI response and resistance in *Ae. geniculata* (5M^{g}: green) and *Ae. umbellulata* (5U^{u}: orange), 5A(purple),5B (pink), and 5D (blue) chromosomes of Chinese spring.

Comparative analysis of 5M^{g} and 5U^{u} introgressions providing resistance against leaf rust using transcriptome data under leaf rust stress

To compare the gene expression between the homoeologous genes from 5M^{g}, 5U^{u,} and wheat, highly specific translocation germplasm available to us was used

(Tiwari et al., 2014, 2015; Bansal et al., 2020; Yadav et al., 2016; Steadham et al., 2021). Differential expression analysis was performed using these resources to determine potential candidate genes translocated from $5M^g$ of *Ae. geniculata* to the terminal distal region (10Mb) on the 5D chromosome of wheat using two translocation lines T756 & T598 (pau16052 & pau16055) challenged with leaf rust inoculum. Differential expression analysis resulted in 3,196 differentially expressed genes (DEGs) between T756 and WL711, and 5,099 DEGs between T598 and WL711. A total of 6,512 genes were differentially expressed in two ILs, and 1,784 genes were common in both studies. T756 showed 2,058 up-regulated and 1,138 down-regulated genes, while T598 displayed 2,833 up-regulated and 2,265 down-regulated genes (Table 3.6**,** Table S3.5). DEGs at different HPIs intervals between T756 and WL711, and T598 and WL711 are shown in Figure 3.8. Notably, 21 genes in T756 and 283 genes in T598 were shared among all HPI intervals. In both the ILs, catalytic activity and transferase activity were found to have a higher number of up-regulated genes, while DNA binding and biological regulation exhibited down-regulation. Response to the biotic stimulus was up regulated in T756, while no corresponding GO term was detected in T598. About of 283 genes were shared at different HPI in T598 compared to only 21 shared genes in T756.

The gene ontology (GO) plot using WEGO indicated a higher number of genes involved in binding and catalytic activity under molecular functions, whereas biological processes showed higher number of genes under metabolic and cellular processes followed by a response to stimulus process in both introgression lines (Figure S3.5). Gene ontology enrichment using goseq demonstrated a higher percentage of

genes being involved in ATP binding, ADP binding, and protein kinase activity under the molecular function category playing a significant role in the signaling process. (Figure S3.6). Genes belonging to chitinase activity were also found enriched. These genes may play a role in degrading the membrane-bound carbohydrate structures on the fungal cell wall. Monooxygenase and oxidoreductase activity was also found to be enriched under defense response. In biological processes, protein phosphorylation and defense response categories were prevalent in both the introgression lines. In T756, the integral component of membrane was found enriched under cellular component, while the same was not detected in T598.

Nearly 417 genes were found differentially expressed on chromosome 5D collectively in both ILs, with 78 up-regulated and 53 down-regulated genes in T756; 95 up-regulated and 267 down-regulated genes in T598. DEGs in the 10 Mb distal region of chromosome 5DS, with reported introgression for Lr57/Yr40 (Kuraparthy et al., 2007; Steadham et al., 2021; Yadav et al., 2016) were filtered. Out of 173 high-confidence genes in the region, 54 genes were found to be differentially expressed. Most genes were downregulated in the candidate region with prominent transferase and dephosphorylation activity shown in Figure S3.5. *Ae. geniculata* is a member of the tertiary gene pool and there are chances that homologous genes may not be present on the CS designated chromosomes.

Table 3.6 Differentially expressed genes in introgression lines T756 and T598 in comparison to WL711.

HPI interval	Total DE gene	DEG (FDR<0.001)	Upregulated	Down regulated
T756 (WL711+*Lr57*)				
0HPI	32052	184	101	83
12HPI	36214	313	187	126
24HPI	34150	256	136	120
48HPI	34406	262	150	112
72HPI	34685	196	110	86
96HPI	40457	2820	1847	973
T598 (WL711+*Yr40*)				
0HPI	33610	570	344	226
12HPI	36681	1054	691	363
24HPI	35529	1188	670	518
48HPI	35597	1349	866	483
72HPI	36242	1564	943	621
96HPI	40799	3302	1713	1589

Analysis of differentially expressed genes (DEGs) on the unanchored chromosomes of Chinese Spring (CS) revealed 191 DEGs, including 146 genes in T598 and 90 genes in T756 Among these DEGs, 26 genes were identified that contained protein kinase, NB-ARC, or LRR domains, potentially linking them to disease resistance. There were four different genes in each ILs, which were up-regulated in at least four HPI intervals. These genes appeared to be potential candidates for the *Lr57* because they were significantly upregulated, with expression changes ranging from 2 to 9-fold. (Table S3.5). However, BLAST searches against the 5M^{g} assembly of *Ae. geniculata* did not yield significant homologs, making it challenging to definitively establish their origin from this wild relative. These genes may be the source of resistance with higher up-regulation, but it is difficult to establish their origin from *Ae.*

geniculata. The syntenic map of the 10Mb introgressed region created using the contigs of the mapped genes is shown in Figure S3.8.

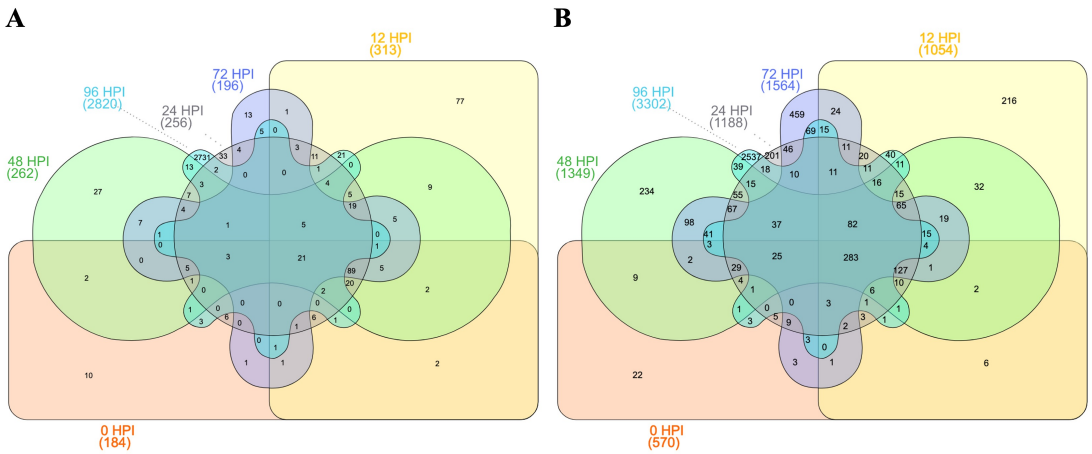


Figure 3.8 Differentially expressed genes under different time intervals A) between T756 and WL711 and B) between T598 and WL711.

Development of SSR and SNP markers

The MISA program detected 47,358 and 28,612 SSRs in $5M^g$ and $5U^u$ chromosomes, respectively. The majority of the SSRs were either monomeric or dimeric types (Figure S3.7). Compound SSR constituted 10 % and 8% of the total SSRs in $5M^g$ and $5U^u$, respectively. Primers were designed for 31,750 and 13,199 SSRs for the two chromosomes, respectively. To identify genome-specific primers, the $5M^g$ primer pairs were mapped to $5U^u$ and 5D, similarly, $5U^u$ primer pairs were mapped to $5M^g$ and 5D. With the stringent criteria of two mismatches and zero gaps, 4,051 and 2,389 primers of $5M^g$ mapped to 5D and $5U^u$, respectively. In $5U^u$, 1,672 and 1,675 uniquely mapped primer pairs were detected for 5D and $5M^g$, respectively. Chromosome 5D and $5M^g$ shared 1,044 primer pairs from $5M^g$, while 5D and $5M^g$

shared 884 primer pairs for $5U^u$ **(**Table S3.7**).** Kompetitive allele specific PCR (KASP) markers were developed from the homozygous SNPs between 5M and 5D and 5U and 5D (Table S3.8).

3.5 Discussion

Characterization of genetic diversity underlying a phenotype is critical in improving wheat against several biotic and abiotic stresses. Polyploidization enhanced the wheat adaptability to diverse environments outside its center of origin. At the same time, domestication and selective breeding for high yield have reduced the genetic base of cultivated wheat varieties (Feuillet et al., 2008). Wild and distant relatives of wheat contain rich genetic diversity for its improvement and many of these wild relatives are well-documented to provide resistance to biotic and abiotic stresses. *Ae. geniculata* and *Ae. umbellulata* are two such species from the tertiary gene pool of wheat and harbor several genes and alleles for wheat improvement. There is consensus about the conservation of large-scale synteny of the homoeologous genes and chromosomes, or chromosomal regions based on available genomics datasets and comparative genome analysis of wheat's wild relatives (Walkowiak et al., 2020). Whole genome sequencing of *Ae. comosa* and *Ae. umbellulata* demonstrated the collinearity of the M and U genome with the D genome (Said et al., 2021). Our initial results have indicated that the $5M^g$ chromosome is homoeologous to wheat chromosome 5D. Previous analysis of translocations lines TA5602 (5% translocation of $5M^g$ short arm on 5D of wheat) and pau3732 (5% translocation of $5U^u$ short arm on 5D of wheat) showed no major rearrangements in the introgressed 9.5 Mb region (Bansal et al., 2017; Steadham et al., 2021). These lines and available datasets from shotgun sequencing of chromosomes

5U^{u}, 5M^{g}, and RNAseq results from translocation lines and availability of reference genomes of wheat lines, provide us an excellent opportunity to look at the evolutionary patterns in wild and domesticated homoeologous accessions, especially for the disease resistance genes in targeted physical regions.

The present study reported 397 Mb and 243 Mb assembled sequences of flow-sorted chromosomes 5M^{g} and 5U^{u} of *Ae. geniculata* and *Ae. umbellulata* retrieved from public sources to study the pattern of gene evolution in group 5 chromosomes. An N50 of 3.4 kb and 1.13 kb was observed for both 5M^{g} and 5U^{u} chromosomes. Assembled sequence demonstrated a repeat content of 79% and 63%, with 736 and 383 conserved BUSCO genes for 5M^{g} and 5U^{u}, respectively. The N50 achieved was higher than that of the previous study for 5M^{g} (1.1kb) (Tiwari et al. 2015), but was lower for 5U^{u} (2.9kb) (Bansal et al. 2020). Due to the fragmented assembly and partial genes, analysis was restricted to the full-length genes carrying both transcription start and termination sites, which still give us access to more genes than the previous studies. The full-length genes reported showed sequence homologs for 84% and 54% of sequences in 5M^{g} and 5U^{u} in the NCBI non-redundant database. Functional annotations (protein domain, enzyme classes, and GO annotations) were reported for more than 50% of these genes. The majority of TFs identified belong to C2H2, bHLH, B3, MYB-related, bZIP, M-type MADS, WRKY, and ERF classes and constitute 67% and 60% of the TF classes in 5M^{g} and 5U^{u}, respectively. C2H2 zinc finger transcription factors play essential roles in plant growth, development, and biotic and abiotic stress resistance (Han et al., 2020). Similarly, bHLH plays a significant role in drought, salt, and chilling stress (Sun et al. 2018). MYB transcription factors are known for their role in plant development,

secondary metabolism, hormone signal transduction, disease resistance, and abiotic stress tolerance (Katiyar et al. 2012). The basic leucine zipper (bZIP) gene family is one of the most prominent transcription factor families in plants, and members of this family play essential roles in multiple biological processes such as light signaling, seed maturation, flower development as well as abiotic and biotic stress responses (Wang et al. 2018). MADS-box transcription factors plays role in floral organ identity determination (Theißen et al. 2016). ERF transcription factors play many diverse functions in cellular processes, such as hormonal signal transduction, response to biotic and abiotic stresses, regulation of metabolism, and in developmental processes in many plant species (Nakano et al. 2006). The richness of these TFs classes makes 5M^{g} and 5U^{u} chromosomes an attractive target for introgressions, also evident from the previous introgression studies. In the present study, both the chromosomes displayed higher number of resistance genes.

Despite the restricted growth environment for *Ae. umbellulata*, it showed higher genetic diversity than *Ae. tauschii* (Okada et al. 2018). According to Stoilova et. al., (2006) chromosomal substitution of 6U has been shown to provide resistance to powdery mildew at both the seedling and adult plant stages (Stoilova and Spetsov 2006). Synthetic lines generated through crossing between durum and *Ae. umbellulata* showed increased plant height and the number of spikes and smaller in spike length (Okada et al. 2020). Grain hardness is regulated by the Hardness (Ha) locus on 5D, which contains the genes Pina-D1 and Pinb-D1 that code for puroindoline proteins. For hard-textured common wheat, variations in *Ae. umbellulata* is considered a useful resource for enlarging the diversity of grain hardness (Okada et al. 2018). Both the ILs

used in the present study have higher grain hardness than the recipient parent (data not shown). Wheat-*Ae. umbellulata* introgression carrying $1U^u$ demonstrated accumulation of glutenin macropolymer (GMP) compared to the recurrent parent Chinese Spring (Du et al. 2019), implying its potential application for improving the end-use property of the wheat flour.

During evolution, gene order or collinearity tends to be conserved across species (Huynen and Bork 1998). Syntenic block represents the regions with shared homologous genes between the two genomes derived from the common ancestor (Tang et al. 2011). These blocks are often used to investigate gene evolution that might have shaped the speciation. Natural or artificial plant hybridizations typically result in the addition of new gene combinations that enhance the genic density of the available germplasm (Arrigo et al. 2011). Ramírez-González et al., (2018) demonstrated the biased expression in nearly 30% of homoeologous genes in wheat (Ramírez-González et al., 2018). DEGs from RNA expression analysis of susceptible and resistant introgression lines provided an array of genes involved in resistance belonging to classes of DNA binding, catalytic activity, and protein kinase activity. Chitinase-related genes were also found to be enriched in DEGs. DEGs in the reported introgressed region were found to be downregulated. Four genes belonging to resistance gene class were found to be highly expressed at most of the HPI intervals. BLAST searches against the $5M^g$ assembly of *Ae. geniculata* did not identify significant homologs for these candidate genes. This lack of detection could potentially be attributed to the fragmented nature of the 5Mg assembly itself.

Evolutionary studies revealed the closeness between *Ae. geniculata* and *Ae. tauschii* (Middleton et al. 2014; Li et al. 2015). Chloroplast genome-based studies have shown the divergence of *Ae. tauschii* and *Ae. geniculata* nearly around 1.62 mya (Middleton et al. 2014). Li *et.al.,* (2015) derived distinct clusters for M, S, and D genomes using the chloroplast sequence of the progenitors (Li et al. 2015). A higher number (60%) of homoeologous genes were shared between $5M^{g}$, $5U^{u}$, and group5 chromosomes (A, B, D genome) of wheat. There were nearly 514 single-copy orthologs attributed to a lower gene count of *Ae. umbellulata*. A comparison of the $5M^{g}$ and $5U^{u}$ with the *Ae. tauschii* and genomes of 10 other wheat cultivars is shown in Figure 3.2B**.** Comparison of $5M^{g}$ and $5U^{u}$ with the publicly available exome data of hexaploid and tetraploid wheat, positioned the two assembled chromosomes into a separate cluster, demonstrating the distinctness of the U and M genome from the D genome. The phylogenetic analysis suggests an intermediate position of the M and U genomes between the A and D genomes. This was further supported by the read-mapping study, which showed a considerably higher number of read mapping on the A genome compared to the B genome. Although our observations were based on the single chromosome data, a purity of ~95% of flow-sorted chromosomes, reduces the possibility of contamination in both chromosomes, which come from different studies. But still, further exploration is needed in the presence of higher similarity to A genome.

Read mapping demonstrated some sort of similarity between $5M^{g}$ and chr1A, and $5U^{u}$ and chr2A. These sites can be considered the potential hotspot region for introgression or hybridization although no reported introgression was found in the support. A similar observation was depicted through the chromosome-level phylogeny,

which grouped the M genome with the D genome and ordered these closer to the A genome. The likelihood of contamination from other chromosomes is reduced by the 95% purity of the flow-sorted chromosomes. Chromosome flow-sorting made a significant contribution in lowering the complexity of the wheat genome (Marcussen et al., 2014). Previously natural hybrids of *Ae. geniculata* and wheat have been reported to exist under field conditions in central Spain (Loureiro et al., 2006). However, given the increased similarity to the A genome, further genome-scale exploration is needed in the future. Mena et. al., (1993) demonstrated 5A/5M^{v} introgression of *Ae. ventricosa* in the wheat-Aegilops system. (Mena et al., 1993). Researchers have reported substitution lines where 7A and 7B were replaced by 7M providing resistance to Fusarium head blight (FHB), powdery mildew, and stripe rust (Yang et al., 2022). The significant number of mutations observed with the Chinese Spring provide distinction to the genomes. Indels accounted for one-fourth of the total variants, responsible for differences in the gene models between alien chromosomes and Chinese Spring. Nearly 90% and 80% of the genes on chr5D were impacted by SNP variants from 5M^{g} and 5U^{u}, of which 35% were missense-type variants. A total of 4-5 percent of variants showed a moderate effect on the structure and function of the genes.

During the evolution, expansion in the gene sequence resulting from the incorporation of large insertional elements in the intronic region has been observed. Sela and colleagues demonstrated 60% of TE insertions within introns compared to the transcribed region within the mammalian genome (Sela et al., 2010). Significant difference in gene length was observed for the 5M^{g} and 5U^{u} genes compared to chr5D. Tertiary gene pool species displayed a shorter gene structure compared to their

hexaploid counterpart, evidenced by the fewer intron per mRNA and smaller mean intron length. Differences were also observed in single copy orthologs, where genes with the same number of exons and equal protein length showed differences in total gene length. Mapping of $5M^g$ genes on 5D chromosome resulted in similar gene structure, with a different exon-intron model (Figure 3.4).

Genome-wide AFLP markers have proved reliable to detect hybridization and introgression events. Natural hybridization in *wheat-Aegilops* species has been reported when present or grown in proximity (Arrigo et al., 2011). Whole genome/chromosome sequencing can be very helpful in detecting introgressions (Walkowiak et al., 2020).

Analysis revealed evidence of substantial introgression on the telomeric region of wheat chromosome 5D. This introgression appears to originate from the $5M^g$ chromosome of *Ae. geniculata*, spanning a region between 546 Mb and 566 Mb. Notably, the read coverage density within this region in a window of 1 Mb approaches nearly equal to 1, suggesting a high degree of sequence similarity between the two chromosomes. The region consisted of 394 high confidence genes. However, no literature-based evidence was found for the introgression. The developed SSR markers can be used for detecting the present and other introgressions on chromosome 5D (Table S3.3).

3.6 Conclusion

With their extensive genetic diversity, *Ae. geniculata* and *Ae. umbellulata* represent a valuable resource for wheat improvement. This study identified a wide range of orthologs, both shared and species-specific, which can be used to enhance disease resistance and quality traits. Both chromosomes harbor an abundance of

transcription factors that might play a regulatory role in the pathways for tolerance to drought, salt, and chilling stresses. The study highlighted the differences in the gene structure between the members of tertiary gene pool and cultivated wheat. These may be the results of incorporation of large TE insertions in wheat genome. Gene families in the tertiary gene pool are still under evolutionary pressure and are expanding. Phylogenetic and read-mapping studies demonstrated the higher resemblance of M and U genomes to the D genome, followed by A genome. The remarkably diverse set of TFs with a role in biotic and abiotic stress and resistance genes make these two chromosomes interesting targets for introgression studies.

3.7 Supplementary Material

The supplementary figures and tables are hyperlinked to the published manuscript https://www.frontiersin.org/articles/10.3389/fpls.2023.1144000/full#supplementary-material

Chapter 4: Curation and Characterization of the UMD Triticale Collection

4.1 Abstract

Triticale is an important multi-use crop due to its diverse adaptability and resistance against abiotic and biotic stresses. Genetic diversity is essential for the improvement of crop plants for targeted phenotype. Due to the incorporation of different wheat and rye parental lines and the outcrossing nature of the rye genome, triticale germplasm exhibits a limited range of genetic diversity. Gene bank collection presents a valuable resource for the plant breeding community as a reservoir of hidden genetic and phenotypic diversity. However, storage, maintenance, phenotypic characterization, and utilization of germplasm require resources and put forward a monetary constraint to the curators. Hexaploid triticale faces a severe problem of seed contamination from octoploid triticale and wheat. Genotypic characterization of germplasm allows us to gauge the extent of genetic diversity and can be used systematically to deploy available genetic variations in crop improvement programs. In the present study, we utilized genotypic data from 260 triticale accessions to develop an efficient pipeline using a pseudo-reference of triticale. This method effectively identifies ploidy levels and seed contaminations, and it can detect large chromosomal insertions, deletions, and small introgressions in triticale. A core set of 89 accessions representing 97 percent of the total allelic diversity was developed. This core set offers significant advantages for triticale breeding, serving as valuable source material for developing high-yielding varieties and facilitating easier characterization.

4.2 Introduction

Triticale is a generic name given to the hybrid developed through artificial hybridization of wheat (*Triticum*) and rye (*Secale cereale*), collectively known as *x Triticosecale* Wittmack (Stace, 1987). Wilson et al. (1873) reported the first triticale hybrid with *Triticum* as the female parent and *Secale* as the pollen donor (Wilson, 1873). After the discovery of colchicine in 1937 development of amphidiploid triticale has increased. Present-day triticale germplasm is a collection of second-generation hybrids developed from the cross of two primary triticale (Gupta and Priyadarshan, 1982). The development of triticale combines the yield potential of wheat with the stress tolerance abilities of rye, resulting in a new climate-resilient crop (Mergoum & Gómez Macpherson, 2004). Global triticale acreage covers more than 10 million acres of land with an average production of ~17 million tons. In the United States, triticale is grown on roughly one million acres, primarily for forage and pastures (Gupta & Priyadarshan, 1982; Kumssa et al., 2022) . Triticale is a crop with prolific growth and adaptation to various environmental conditions (especially in nutrient-deficient environments with various biotic and abiotic stresses) as it combines the hardiness of rye and high grain yield and nutritional qualities of wheat (Mergoum & Gómez Macpherson, 2004).

Initially, efforts were focused on developing octoploid hybrids between *T. aestivum* (2n=42=AABBDD) and *S. cereale* (2n=14=RR) (Wilson, 1873). However, hexaploid triticale, developed using *Triticum durum* (AABB) as the recipient parent, later became more prevalent due to its higher genomic stability (Mergoum et al., 2009). Triticale can be classified as primary, secondary, and substituted type based on their

genetic makeup (Furman et al., 1997). Primary triticale are the direct amphidiploid of wheat and rye, the secondary triticale was generated by intercrossing of primary triticale themselves or primaries with wheat in various combinations. Secondary triticale is the most common form and provides an excellent resource for understanding genetics and genomics of agronomically important genes and traits from the rye genome. Triticale is a versatile crop with applications as both forage and a grain source (Wójcik-Gront & Studnicki, 2021). Triticale consists of three different groups: spring, winter, and intermediate (facultative) types. Spring triticale yields more forage during the early stage, while winter types are known for their higher forage yield (Mergoum et al., 2009; Mergoum & Gómez Macpherson, 2004; Salmons et al., 2002). In 2018, Europe alone accounted for 89% of the global triticale production. According to the market reports in the next five-year triticale, production would increase with a CAGR of 2.8% annually (https://www.mordorintelligence.com/industry-reports/triticale-market).

Germplasm collections are always considered the primary genetic resource for sustainable crop improvement (Wang et al., 2018). Our Triticale germplasm collection at the University of Maryland presents a rich resource of genetic variations for triticale breeding and improvement programs. Storage and utilization of gene bank pose both space and monetary constraints involving the planting of germplasm every year for its maintenance and continuation. These collections are regularly evaluated for their performance of phenotypic traits like height, grain number, spike length, and tolerance to biotic and abiotic stresses. Seed contamination can occur in triticale due to various factors during seed production, processing, or storage. Contamination affects genetic

purity, performance, and utilization of the crop. Hexaploid triticale may contain a mixture of octoploid triticale and wheat, which can significantly alter the germplasm collection, leading to dilution of the genetic makeup, reduced yield, and forage quality of hexaploid triticale. Reducing redundancy within these collections is necessary to streamline evaluation and optimize resource utilization. This can be done by identifying a core set of closely related samples; the rest can serve as a "reserve collection" for future needs (Frankel et al., 1984).

The present study assesses a varied collection of triticale through genotypic evaluations utilizing short-read sequencing. We aim to gauge the genetic diversity within the current germplasm collection present at the University of Maryland. The study focuses on developing a method for the characterization of triticale germplasm, identification of misclassified accessions (hexaploid vs. octoploid), and generation of a core set for the breeding program.

4.3 Material and Methods

Plant material and sequencing

GBS sequencing data for 260 triticale accessions, involving 75 spring and 185 winter triticale was provided by Dr Xue-Feng Ma, Nobel Research Institute (Ayalew et al., 2022). The accessions were evaluated for plant height, plant biomass, root biomass, spike weight, seed weight, and leaf nitrogen content in the year 2020 at Beltsville Research Fields, University of Maryland, College Park, MD (data not shown)

Classification of ploidy and detection of seed contamination in triticale

The R-genome of *Secale cereale* and the A, B, and D genomes of *Triticum aestivum* were concatenated together to create a pseudo-reference for triticale. Fastqc

was employed to check the quality of the raw reads (Andrews, 2010). Trimmomatic was used to remove adaptor sequences and filter low-quality reads (Bolger et al., 2014). BWA was used to map high-quality reads to the pseudo-reference (Li & Durbin, 2010). BCFtools was used for SNP calling and filtration (Danecek et al. 2021). bedtools was used to calculate the genome-wide read coverage in a window of 1 Mb span along the genome (Quinlan & Hall, 2010). Ploidy was determined by considering a 10% cutoff of read coverage for the respective genome (A, B, D, and R). Chromosomes with more than 2% read coverage were considered potential translocations. Small segmental introgressions were detected using an exponential rolling average for each chromosome. The score method in the R package "outlier" was used to detect the deviated segments using the MAD module (Komsta, 2022).

Cluster analysis

Plink was used to prune data PCA analysis (Purcell et al., 2007). The unweighted pair group method with an arithmetic mean (UPGMA) based phylogenetic tree was generated using vcfR (Knaus & Grünwald, 2017). Admixture was used to determine the population structure (Alexander & Lange, 2011) and the results were visualized using pophelper (Francis, 2017). Nucleotide diversity (Nei's) and population differentiation index (Fixation Index, F_{ST}) were calculated using vcftools (Danecek et al., 2011). Positively selected regions in triticale were determined at 99^{th} percentile of likelihood values using SweeD (Pavlidis et al., 2013). Gene Ontology terms of positively selected genes were displayed using WEGO (Ye et al., 2018). A core set of accessions was developed using GenoCore with default parameters (Jeong et al., 2017).

4.4 Results

Diversity analysis of triticale

GBS data of the single seed descended plants of 260 accessions (from National Small Grains Collection, Idaho, PRJNA715663) was used in the study. These accessions were selected from a collection of 1,400 triticale accessions based on the phenotypic data (plant architecture, biomass, grain yield and winter hardiness) (Ayalew et al., 2022). The collection includes 74 accessions of spring triticale and 186 accessions belonging to winter triticale. In the absence of reference genome, a pseudo-reference was generated for triticale by combining the genomes of *T. aestivum* and *Secale cereale*. High-quality filtered reads were mapped to the pseudo-reference. The SNPs vcf file was filtered considering a minor allele frequency of < 0.01 and 10% maximum missing data. After the stringent filtering, a total of 38,573 high-quality SNPs were selected for the analysis. One sample (SRR14023521), with more than 30% of missing data, was excluded from the analysis. Figure 4.1a shows the SNP density plot of composite triticale SNPs on the A, B, D, and R chromosomes. For the D-genome, lower SNP density was observed compared to the A, B, and R genomes. SNP percentage for each chromosome ranged from 3 to 6 percent. Principal component analysis (PCA) of the genotypic data revealed that the first two components explained 11.2% and 8.7% of the total variance, respectively (Figure 4.1b). An intermixed pattern was observed between the winter and spring accessions, with two distinct groups evident along the x-axis. Being an inter-species cross of wheat and rye from a diverse set of parents, this variability seems to be acceptable. The unweighted Pair Group Method with Arithmetic Mean (UPGMA) based phylogenetic tree also displayed a

quite complex phylogeny relating to the population structure of the group described later (Figure 4.1c). Despite their variation in winter/spring growth habit, no significant genetic differences were observed between triticale accessions.

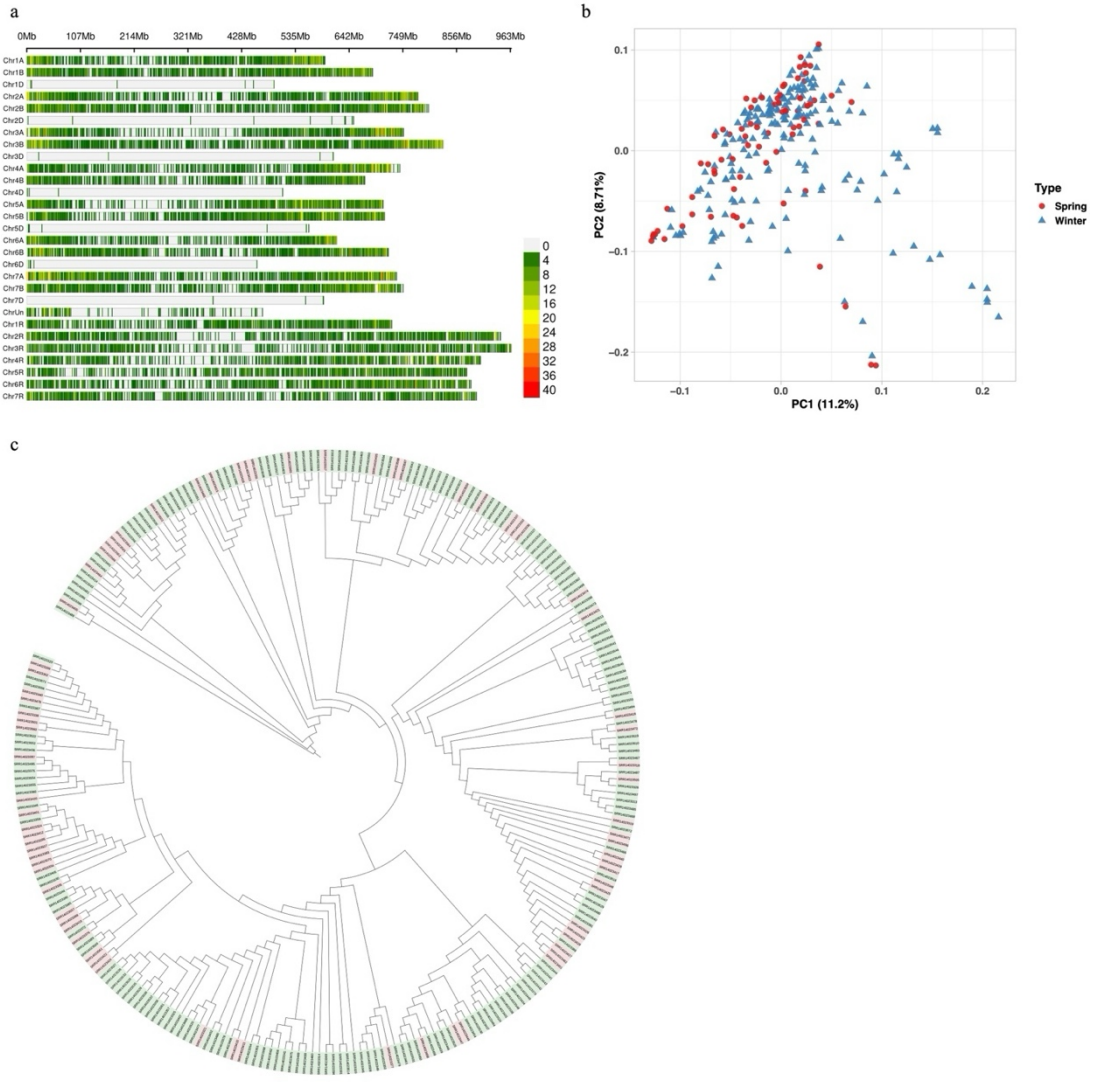


Figure 4.1 a) Chromosome-wise SNP density of triticale. b) Principal component analysis of the 260 selected winter and spring triticale accessions. C) UPGMA-based

phylogenetic tree of triticale accessions. Spring triticale are shown in pink, and winter triticale are shown in green.

Genetic purity assessment and seed mixture detection in triticale accessions

We performed the read coverage analysis to assess genetic purity and identify potential seed contamination within the triticale germplasm. This analysis aimed to detect any seed mixing or foreign genetic material among the accessions. Average read coverage was calculated in 1Mb windows from the mapped bam files to evaluate uniformity**.** The analysis revealed 236 accessions as hexaploid and 12 as octoploid, based on a 10% read mapping threshold for the respective A, B, D, and R genomes. Additionally, low or absence of coverage for the rye genome was considered as wheat contamination, leading to the identification of 11 wheat accessions within the collection. Figure 4.2 displays the read coverage for representative accessions of octoploid and hexaploid triticale, and the wheat accessions within the triticale collection.

Based on the new ploidy level classification, the phylogenetic tree was re-evaluated. Wheat accessions formed a distinct cluster, while most octoploid triticale clustered together (Figure 4.3). However, a few octoploid accessions showed closer affinity to wheat. To focus on true triticale diversity, these wheat lines and the octoploid accessions were excluded from further analyses.

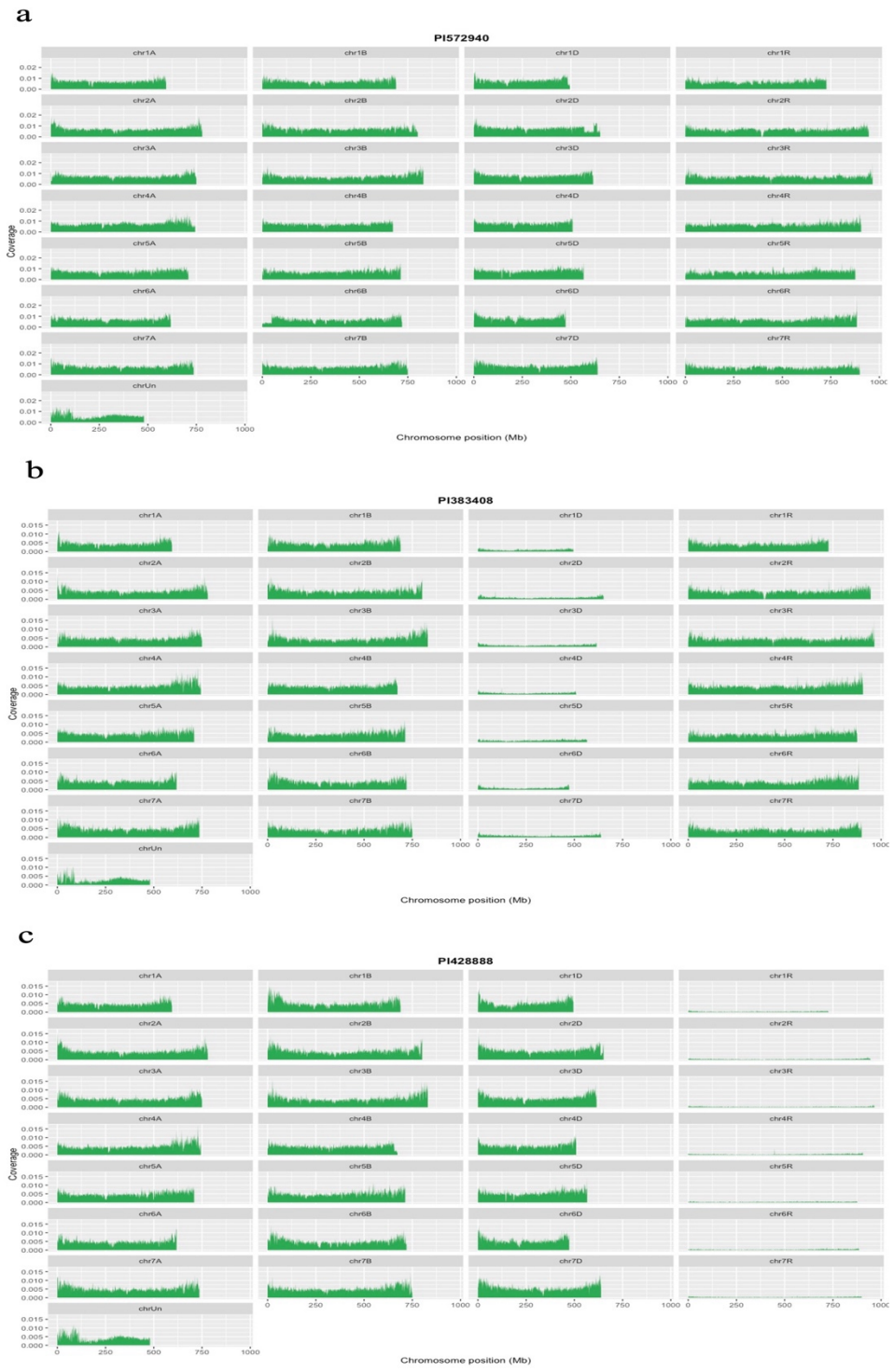


Figure 4.2 Classification of triticale based on the read coverage analysis: a) Octoploid (PI572940), b) hexaploid (PI383408), and c) wheat (PI42888). These are representative figures of each category.

Potential chromosomal aberrations in triticale

Analysis of chromosome-wise read coverage from GBS data revealed potential chromosomal aberrations in triticale, including whole chromosome additions and deletions. Table 4.1 summarizes the observed translocations and deletions in triticale. Notably, chromosome 2D appeared frequently added in hexaploid triticale, while an octoploid accession displayed a deletion of chromosome 6D. These findings suggest variations in chromosome number within the triticale germplasm. Figure 4.4 demonstrates the read coverage of the addition/deletion of the chromosome in the identified accessions.

Table 4.1 Identified chromosomal translocation and deletion in triticale germplasm.

S.No	Cultivar/Accession	Translocation	Deletion	Type
1	PI587326	6D	-	Hexaploid
2	PI542528	2D	-	Hexaploid
3	PI414968	2D	-	Hexaploid
4	CIxt96	-	6D	Octoploid
5	PI386142	2D	-	Hexaploid

Potential short introgressions

To detect the small introgression, regions with elevated and decreased read coverage were studied with a focus on group D chromosomes. A total of 14 small translocations were detected with a higher percentage of 3AL:3DL translocations. Table 4.2 shows the identified translocations in the dataset. Notably, accession NT17410 displayed a putative 1BS:1RS translocation (Figure 4.5). The reason for these small translocations is not well understood and needs further exploration.

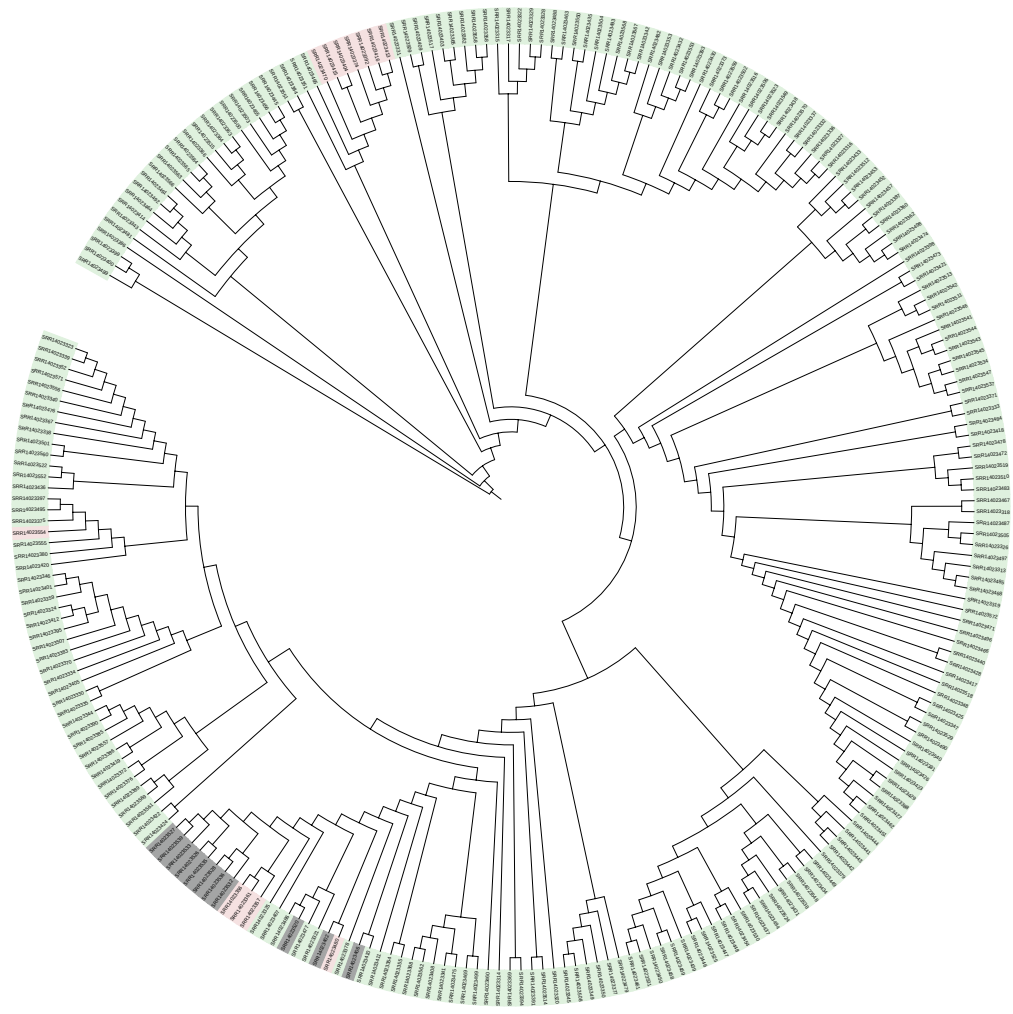


Figure 4.3 Phylogenetic tree depicting the reclassified triticale accessions. Hexaploid triticale are represented in green, octoploid triticale in pink, and wheat accessions in grey.

Figure 4.4 Putative chromosomal additions and deletions in triticale accessions. a) Addition of chromosome 6D in hexaploid accession PI587326. b-d) Addition of chromosome 2D in hexaploid accessions PI542528, PI414968, and PI386142, respectively, e) Deletion of chromosome 6D in octoploid accession CIxt96.

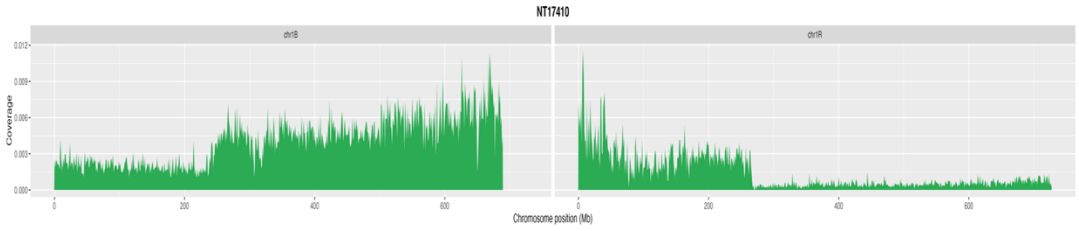


Figure 4.5 Potential 1BS:1RS introgression in hexaploid accession NT17410.

Table 4.2 Potential small introgressions detected in hexaploid triticale accessions.

S. No	Cultivar	Expected Ploidy	A	B	D	R	Event
1	NT17410	Hexaploid	-	1B	-	1R	1BS:1RS
2	PI587326	Hexaploid	2A	-	2D	-	2AS:2DS
3	PI564727	Hexaploid	2A	-	2D	-	2AL:2DL
4	PI587495	Hexaploid	3A	-	3D	-	3AL:3DL
5	PI587450	Hexaploid	3A	-	3D	-	3AL:3DL
6	PI587448	Hexaploid	3A	-	3D	-	3AL:3DL
7	PI611782	Hexaploid	3A	-	3D	-	3AL:3DL
8	NT16406	Hexaploid	3A	-	3D	-	3AL:3DL
9	NT15440	Hexaploid	3A	-	3D	-	3AL:3DL
10	NF97233	Hexaploid	3A	-	3D	-	3AL:3DL
11	NF9	Hexaploid	3A	-	3D	-	3AL:3DL
12	PI414972	Hexaploid	5A	-	5D	-	5AS:5DS
13	PI587453	Hexaploid	7A	-	7D	-	7AL:7DL
14	PI429270	Hexaploid	7A	-	7D	-	7AL:7DL

Population structure of triticale germplasm

ADMIXTURE was used to infer sub-populations and individual ancestry at different levels, ranging from K=2 to K=20. The optimal sub-population structure was observed at K=11, based on the lowest cross-validation error. At the sub-population level of K=2, two clusters were identified. Cluster 1 comprised 14 homozygous accessions, and Cluster 2 consisted of 34 homozygous accessions. The remaining 211 accessions displayed heterozygosity. Further inspection revealed that all 14 homozygous accessions in Cluster 1 were hexaploid triticale. Cluster 2 included 23 homozygous accessions of hexaploid triticale, 3 accessions belonging to octoploid triticale, and 8 wheat accessions. At the optimal sub-population level of K=11, a significant increase in heterozygosity was observed among the accessions (Figure 4.6).

Out of the total accessions, 47 displayed homozygosity: 1 wheat accession, 4 octoploid triticale accessions, and 42 hexaploid triticale accessions. Notably, 17% of hexaploid triticale accessions were homozygous at K=11. Cluster 7 was found to be the most heterozygous group.

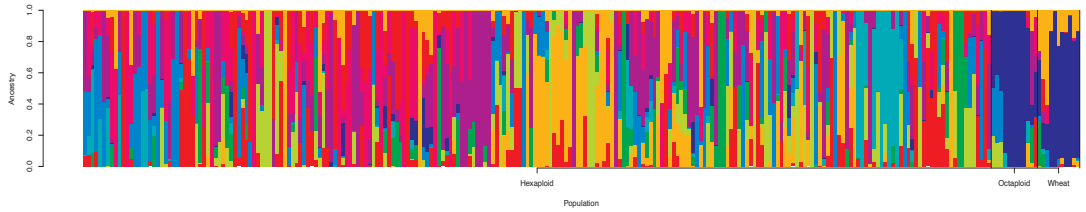


Figure 4.6**.** Admixture-based population structure assignment at the optimal K value (K=11). Population are highlighted as hexaploid, octoploid, and wheat.

Genome-wide nucleotide diversity and F_{ST} scan of triticale accessions

We observed a low average nucleotide diversity (Nei's) for the whole triticale population (4.26406e-06), indicating minimal differentiation between hexaploid and octoploid populations. Hexaploid triticale exhibited a similar pattern with low nucleotide diversity (4.26977e-06). Lower Nei's diversity values were also evident at the chromosome level (Figure 4.7). Additionally, triticale showed a low Tajima's D value (0.0400355), suggesting that the population is likely evolving neutrally, with no strong selective pressures acting on it.

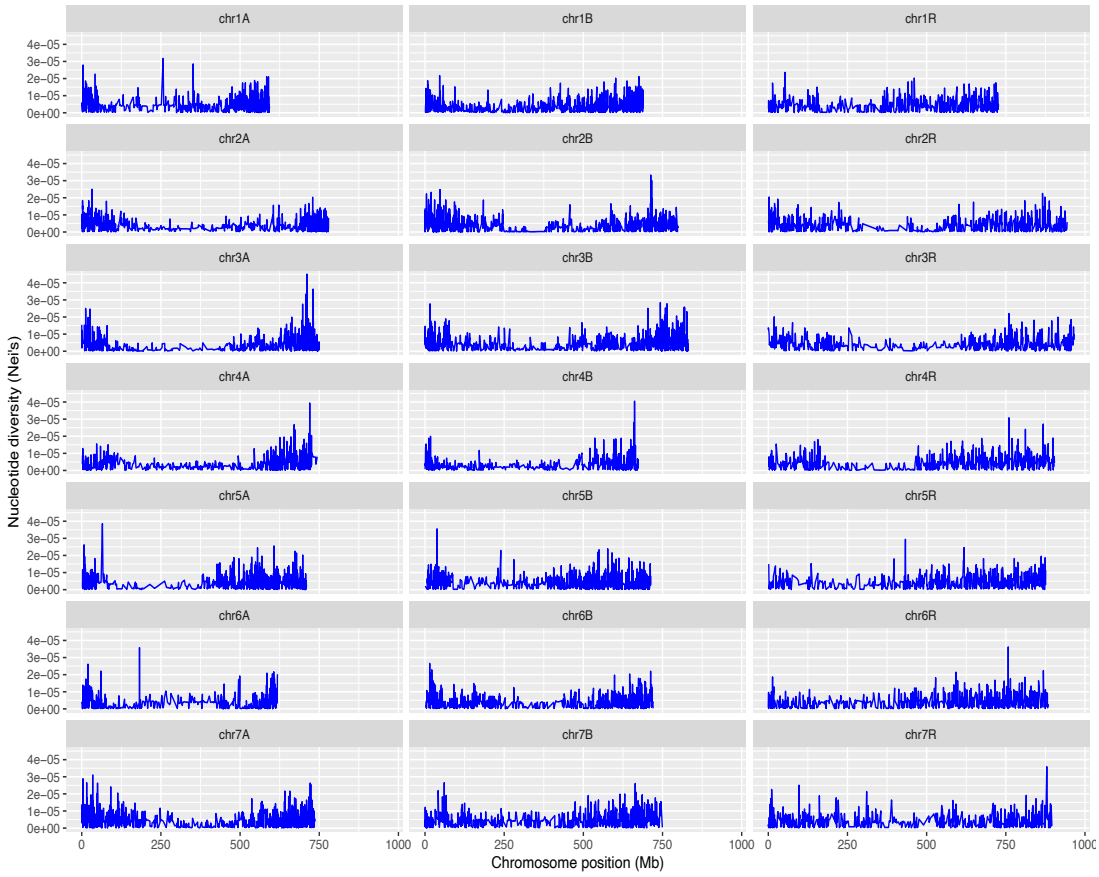


Figure 4.7 The distribution of Nei's nucleotide diversity in a window of 100kb for hexaploid triticale.

Fixation index (F_{ST}) analysis based on Weir and Cockerham's method revealed low genetic differentiation (F_{ST} = 0.0078) between winter and spring accessions within the hexaploid triticale collection (Figure 4.8). This finding suggests minimal genetic variation associated with growth habit.

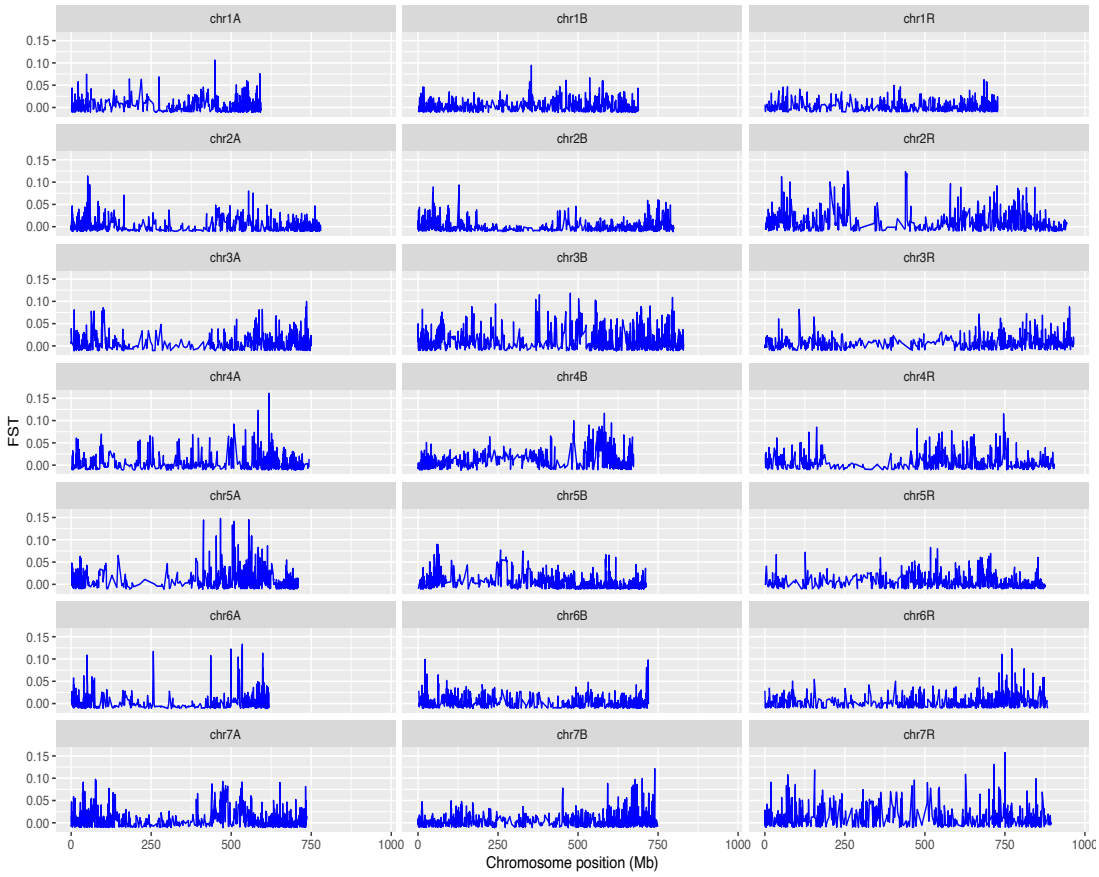


Figure 4.8 Distribution of F_{ST} (Fixation index) values of population differentiation in a window of 100kb in hexaploid triticale.

Identification of core set of triticale accessions

Using GenoCore's default parameters, 89 core set accessions were identified from 236 hexaploid triticale at 99% shared alleles (Figure 4.9). These accessions effectively captured 97% of the allelic diversity present in the original dataset. The present core set demonstrated lesser nucleotide diversity, suggesting a remnant of higher similarity among accessions. Despite a modest reduction in accession number, the core set effectively captured a significant portion of the triticale population's diversity.

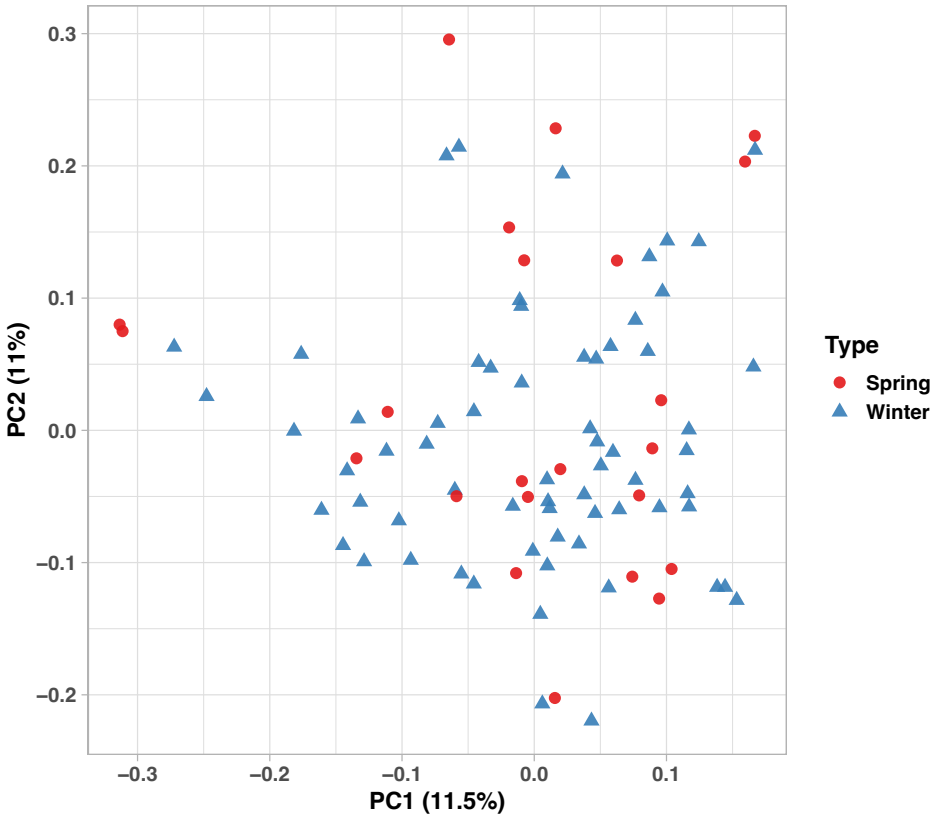


Figure 4.9 PCA plot of 89 core set triticale accessions. Spring triticale are represented as circles, and winter triticale are represented by triangles.

Genomic In Situ Hybridization

GISH (Genomic In Situ Hybridization) is a powerful technique for visualizing the chromosomes of different genomes within a hybrid organism. GISH utilizes fluorescently labeled probes specific to the genomes of wheat (blue) and rye (red) genomes. Hybridization of these probes with the chromosomes of triticale accessions allows for visualization of their origin. In the case of pure wheat accession (AABBDD), all chromosomes should exhibit signals from the wheat, while hexaploid (AABBRR) and octoploid triticale (AABBDDRR) shows a combination of signals from wheat and rye (R) genomes with respective ploidy. GISH results validated the ploidy of three wheat accessions (NT09423, NT17403 and NT17410), four octoploid triticale

(PI611706, PI429295, PI572937 and CIxt29) accessions and four hexaploid (PI564727, PI587453, NT15440 and PI587326) accessions shown in figure 4.10.

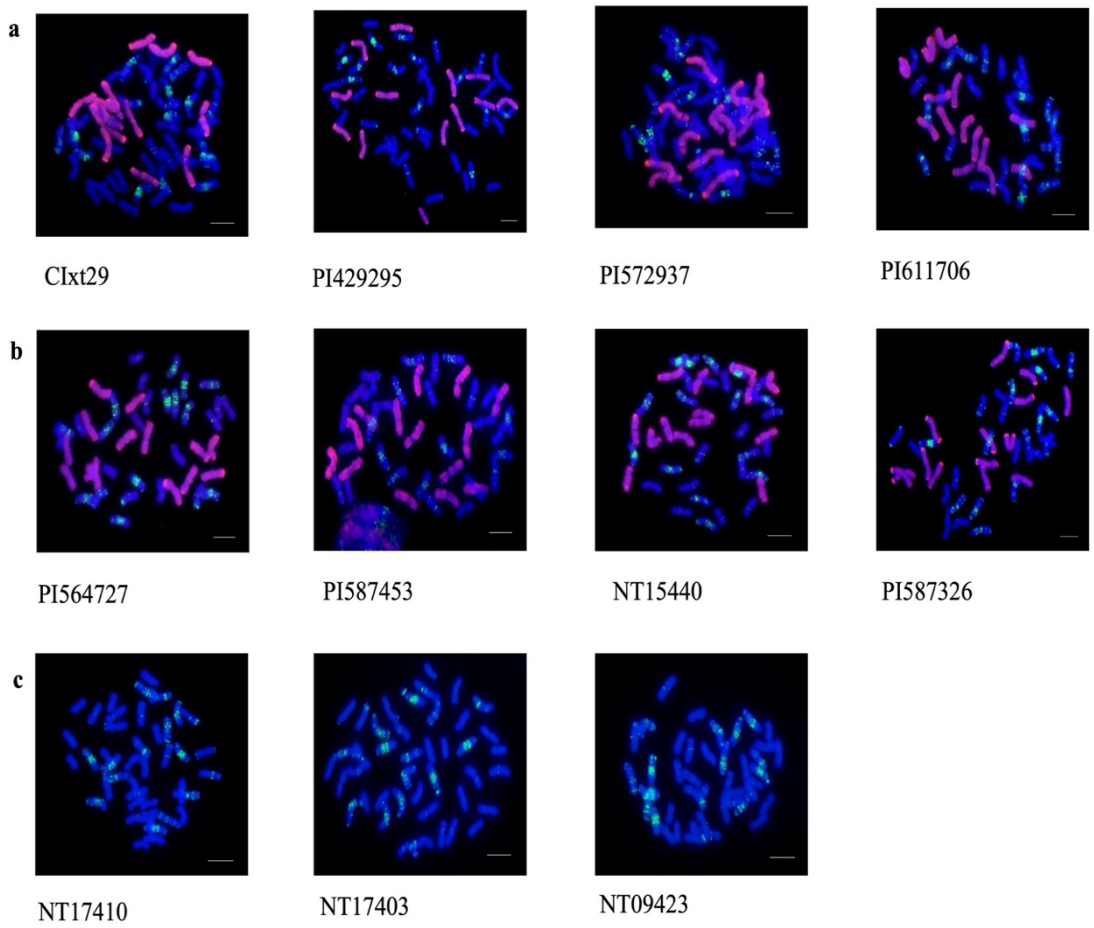


Figure 4.10 GISH images: a) Octoploid triticale, b) hexaploid triticale, and c) wheat. Wheat chromosomes are shown in blue, and rye chromosomes are shown in red.

4.5 Discussion

Triticale is a minor cereal primarily cultivated in Europe for animal feed and fodder (Wafer, 2017) and is gaining popularity in the United States as a cover crop. Triticale harbors yield potential of wheat, tolerance, and resistance characteristics of rye, with a wide adaptive range for different environments (Furman, 2004). Triticale germplasm represents a diverse collection with many variations in phenological

characters inherited from its diverse parent used in the initial hybridization. Triticale germplasm faces the problem of seed contamination from octoploid triticale and wheat, which may result in improper classification of triticale germplasm. In the current project, we developed a pipeline using GBS to classify the ploidy levels and determine wheat contamination.

The triticale collection at the University of Maryland demonstrates rich variability for traits like grain yield, weight, and nitrogen usage (Ayalew et al., 2022). Clustering based on the principal components, although explaining the diversity, could not clearly distinguish between the spring and winter triticale. No distinction was observed for the octoploid accessions and wheat. The potential reason might be the lower number of SNPs in the D-genome, as evidenced by the chromosome-wise SNP density.

In the absence of desirable SNP density for the D genome, we used the read mapping approach to determine potential sources of contamination. With stringent criteria of 10% read mapping in a 100Kb window, 236 hexaploid, 12 octoploid, and 11 wheat accessions were categorized. Hexaploid triticale shows more robust phenotypic traits compared to octoploid triticale and is preferable for breeding due to its genetic stability. The categorization will enhance the use of hexaploid accession in the breeding program. Chromosome-wise read coverage revealed potential whole chromosomal translocations and deletions, which can only be detected utilizing cytomolecular methods. Additionally, smaller translocations were also identified within the triticale collection. Using GISH, we demonstrated the correctness of the predictions done using GBS. The simpler approach based on the read mapping can effectively identify the

contamination. The approach developed can be easily extended for the characterization of other similar hybrids.

Admixture-based population structure analysis showed 11 sub-populations within the triticale germplasm and demonstrated a higher intermix of different accessions (Figure 4.6). As the cultivated triticale represents second-generation hybrids, which may have resulted in higher allele sharing in the germplasm collection. A total of 47 accessions showed homozygosity, which consisted of 42 hexaploid triticale accessions. These homozygous and pure accessions would be very useful in the breeding program. A low nucleotide diversity (Nei's diversity) and Fixation index (F_{ST}) indicate lesser differentiation among the triticale accessions, with less than a century of active ancestry.

Plant breeders and pathologists explore the available genetic diversity for improving the existing elite cultivars/varieties through breeding methods for associated biotic and abiotic stress tolerance (Pandey et al., 2017). A thorough characterization of the genebank collections for genomic, agronomic, biochemical, and other related traits is the key to its efficient utilization. Enriching the genebank accession with the phenotypic data enhances its utilization by researchers (Anglin et al., 2018). We developed a core set of the triticale accession (89 accessions) to perform robust phenotyping for trait discovery. The core set of accessions representing most of the available germplasm diversity is of utmost importance in plant breeding as it saves a lot of money and resources needed to maintain germplasm. The core set displays a significant performance advantage in efficiently utilizing the resources (Odong et al., 2013).

4.6 Conclusion

This study highlights the importance of germplasm characterization for efficient utilization in breeding programs. The initial germplasm collection contained redundant entries, seed contamination, and misclassified accessions, necessitating the development of a pipeline for categorization. We successfully employed a GBS-based pipeline to achieve this goal. The developed pipeline effectively characterizes hexaploid triticale and identifies potential wheat contamination within the germplasm. The resulting core set of 89 accessions, chosen from the initial collection of 236 triticale, effectively captures a significant portion of the original collection's allelic diversity. This streamlined resource offers a manageable and informative foundation for future phenotypic characterization efforts, ultimately accelerating breeding program progress.

4.7 Acknowledgement

Authors are thankful to Dr. Xuefeng Ma, National Small Grains Collection (NSGC), Nobel Research Institute, Oklahoma, for providing genotype data of triticale accession. Authors are thankful to Dr**.** Dal-Hoe Koo, WGRC, Kansas State University for performing GISH analysis.

Chapter 5: Whole Genome Assembly and Comparative Analysis of Two Einkorn Genomes

5.1 Abstract

Einkorn wheat (*Triticum monococcum*) is one of the oldest crops, cultivated by humans since the early days of civilization. Einkorn is closely related to the A-genome progenitor (*Triticum urartu*) of the hexaploid wheat. It holds a unique status as a sole diploid wheat species with naturally occurring both wild and domesticated forms in the Fertile Crescent. The present study investigates the genomic feature of the einkorn genome by assembling one wild and one domesticated accession using PacBio HiFi reads. A total of ~ 5Gb of the genome was assembled, with an N50 of 47Mb and 35Mb for the wild and domesticated accessions, respectively. Reference-based scaffolding assigns the contigs to pseudomolecules, leaving only a small portion as unassigned contigs. The BUSCO-based evaluation confirmed the completeness of 97% of conserved genes. Both genomes showed a high percentage of repetitive elements, accounting for ~83% of the assembled genomes. Einkorn genomes demonstrate a strong synteny with the *T. urartu* and A-genome of wheat. Transcription factors and R genes form the major gene classes. A relatively high number of species-specific orthogroups were observed in each accession. Gene family evolution detected a higher number of rapidly evolving gene families in the einkorn genome. Collectively, these findings suggest that the assembled einkorn genomes harbor novel genomic elements, including transcription factors and R genes, potentially associated with adaptation and stress tolerance. This comprehensive genomic characterization of einkorn wheat

provides valuable resources for understanding its evolutionary history, adaptation mechanisms, and potential for wheat improvement programs.

5.2 Introduction

Being a staple source of calories for around 40% of the human population, bread wheat (*Triticum aestivum*, 2n=6x=42) is a critical crop for global food security. The genome of bread wheat is highly complex, consisting of three sub-genomes (A, B, and D) as it arose from two different natural hybridization events. The first hybridization resulted in the formation of tetraploid wheat *T. turgidum* spp. *dicoccoides* (2n=4x=28, AABB), where the A genome was contributed by the diploid progenitor *T. urartu* (2n = 2x = 14, AA). *Ae. speltoides* (2n=14, SS) is considered to be the closest source for the B genome. Another subsequent hybridization between the tetraploid *T. turgidum* and diploid goat grass *Ae. tauschii* (DD) leads to the emergence of present-day hexaploid wheat *T. aestivum* (2n = 6x = 42, AABBDD) (The International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018). The progenitor species and wild germplasm co-existed for millions of years in the Fertile Crescent region and accumulated variations through mutations as well as through natural introgressions between the members of the different gene pools of wheat (Vardi & Zohary, 1967).

Einkorn wheat (*T. monococcum* ssp. *monococcum*: 2n=2x=14, A^mA^m) is the only A-genome-containing diploid wheat that occurs both in wild and domesticated forms in nature. *T. monococcum* holds the distinction of being one of the first crops domesticated by humans, cultivated over 12,000 years ago. It is closely related to *T. urartu*, the A-genome progenitor of hexaploid wheat. *T. monococcum* genome is under-represented, with its contribution limited to *T. zhukovskyi* (A^mA^mAAGG), whereas *T.*

urartu shows dominance with its contributions towards species like *T. turgidum* (AABB), *T. timopheevii* (AAGG) and *T. aestivum* (AABBDD) (Dvorak et al., 1993).

With its long history of cultivation across diverse geographical and environmental regions, einkorn wheat (*T. monococcum*) stands as an important reservoir of genes for enhancing stress resistance in modern wheat varieties. Research suggests einkorn possesses genetic variations conferring resistance to a range of biotic stresses, such as fungal diseases (Chen et al., 2018; Elkot et al., 2015; Chhuneja et al., 2008; Jing et al., 2007) and abiotic stresses, including drought and salinity (Sharma et al., 2020; Zaharieva & Monneveux, 2014). This potential for stress tolerance makes einkorn a valuable resource for breeders seeking to improve the resilience of cultivated wheat in the face of a changing environment.

The genetic similarity and conservation between einkorn and the A-genome of *T. aestivum* have been well-documented in various studies (Devi et al., 2019; Fox et al., 2014; Ling et al., 2013; Marcussen et al., 2014; The International Wheat Genome Sequencing Consortium (IWGSC) et al., 2018). The presence of highly variable association panels (Fricano et al., 2014; Heun et al., 1997), genetic populations, and a TILLING population (Rawat et al., 2012) from a robust foundation for delving into gene function and unraveling the genetic underpinnings of crucial traits in wheat. These resources provide researchers with the means to explore the genetic basis of various traits and accelerate the development of improved wheat varieties through targeted breeding efforts.

In the present study, two einkorn genomes, one wild *T. monococcum* spp. *aegilopoides* (TA391) and one domesticated *T. monococcum* ssp. *monococcum*

(TA10898) was selected from two distant clades of the phylogenetic tree generated using the GBS (Adhikari et al., 2022). These two accessions were selected based on the drastic phenotypic differences i.e., resistance to powdery mildew (TA391: resistant and TA10868: susceptible), variability in spikelet number per spike (TA391: high spikelet count and TA10868: low spikelet count), brittle rachis (TA391) and non-brittle (TA10868), blue aleurone (TA391: blue and TA10868: brown) and coleoptile color (TA391: brown and TA10868: normal). Another reason for the selection was the availability of more than 13 mapped QTL and genes from a biparental RIL population generated using these accessions as the parental lines. Gene bank at UMD possesses 135 introgression lines with A-genome introgression from these two parental lines. Additionally, a large TILLING population was generated by mutagenizing accession TA10868. The reference level assemblies of *T. monococcum* accessions TA391 and TA10868 will provide a foundation to integrate all the resources available to us to perform gene discovery, validation, and deployment of genes in the elite cultivars.

T. monococcum provides a unique opportunity with its unexploited genetic diversity and can serve as a model species for novel gene discovery for quality traits, and biotic and abiotic stresses for wheat genetic improvement (Jing et al., 2007). The study selected two very contrasting einkorn wheat accessions for genome assembly: a wild *T. monococcum* ssp. *aegilopoides* (TA391) and a domesticated *T. monococcum* ssp. *monococcum* (TA10868). These high-quality genome assemblies will serve as valuable resources for investigating genome dynamics and their applications in cereal breeding programs.

5.3 Material and Methods

Plant materials

A collection of 881 diverse sets of accession of wild and domesticated *T. monococcum* present at the gene bank at University of Maryland was used in the study. From the panel, two most diverse accessions one wild (*T. monococcum* ssp. *aegilopoides*; TA391) and one domesticated (*T. monococcum* ssp. *monococcum*; TA10868) were selected for the development of reference level assemblies. These two accessions were selected based on phenotypic traits i.e., resistance to powdery mildew (TA391: resistant and TA10868: susceptible), variability in spikelet number per spike (TA391: high spikelet count: and TA10868: low spikelet count), brittle rachis (TA391) and non-brittle (TA10868), blue aleurone (TA391: blue and TA10868: brown) and coleoptile color (TA391: brown and TA10868: normal).

Genome sequencing

High-molecular-mass (HMM) genomic DNA was isolated from young leaves obtained from 3-week-old seedlings after dark treatment for 48 h. DNA was extracted according to the HMM DNA extraction protocol for long-read sequencing (Mayjonade et al., 2016). DNA quantification was performed using the Qubit dsDNA HS Assay (Q32851, Thermo Fisher Scientific), purity was confirmed using the Nanodrop spectrophotometer by checking the 260/280 and 260/230 ratios, and the DNA size was validated by using the FemtoPulse system (Agilent). HiFi libraries were then prepared according to the manual 'Procedure & Checklist - Preparing HiFi SMRTbell Libraries using the SMRTbell Express Template Prep Kit 2.0' (PN 101-853-100, Pacific Biosciences) with 10 µg DNA sheared by using the Megaruptor 2 system (Diagenode)

to obtain a 15–20 kb average size. Four size-selected libraries were sequences on PacBio Sequel II for each of the two accessions. Raw data of 99.26 Gb and 93.77 Gb PacBio HiFi reads was obtained for TA391 and TA10868, respectively, corresponding to a coverage of nearly 15-fold (Table S5.1).

Genome assembly and validation

PacBio HiFi long reads from both accessions were assembled into a primary contig using Hifiasm v.15.1 (Cheng et al., 2021). Subsequently, the assembly was scaffolded using the previously assembled genome sequence of *T. monococcum* ssp. *aegilopoides* (accession TA299) (Ahmed et al., 2023) as a reference using RagTag (Alonge et al., 2022). The quality of the two einkorn assemblies was assessed using two approaches. First, PacBio HiFi long reads (mapped with minimap2 v.2.21) and Illumina short reads generated for this study (mapped with bwa mem v.0.7.17) were aligned back to the final assemblies (H. Li, 2018; H. Li & Durbin, 2010). In the second step, BUSCO v.5.0.0 with the plant dataset (poales_odb10) was employed to evaluate the completeness of the assemblies by assessing the presence of conserved single-copy orthologous genes (Simão et al., 2015). Merqury (v.1.3) was utilized to assess the consensus quality (QV) and completeness of the assembly by comparing k-mers (Rhie et al., 2020).

Repetitive element annotation and gene prediction

For both TA391 and TA10868, repetitive elements were annotated using the EDTA pipeline (Ou et al., 2019). Gene prediction was performed using the method described previously by (Mascher et al., 2021), incorporating the seedling stage RNAseq and Isoseq from TA391 and TA10868. Additionally, RNAseq and Isoseq data

from six tissues (aerial, flag leaf, glume, grain, root, and spike) belonging to *T. monococcum* (accessions TA299 and TA10622) were also used in for gene prediction (Ahmed et al., 2023).

Functional annotation of the genes

The diamond BLAST was used to search for the homologs of predicted genes from the NCBI non-redundant database (Buchfink et al., 2021). Blast2GO was used for the assessment of gene ontology (GO) terms and function assignment (Conesa & Stefan, 2008). A Hidden Markov Model (HMM) based tool, pfamscan, was used for protein domain identification using the PfamA database (Mistry et al., 2007). Mercator4 was used to assign MapMan functional annotations and pathways (Schwacke et al., 2019). Transcription factors were annotated using the transcription factor prediction module of PlantTFDB (Jin et al., 2017).

Ortholog determination and evolution of gene families

OrthoFinder (v. 2.5.5) was used to assign orthologous groups (Emms & Kelly, 2019). First, orthogroups were assigned to TA391 and TA10868. Secondly, orthogroup**s** were assigned for the four einkorn genomes (TA391, TA10868, TA299 and TA10622). Finally, the progenitor species of wheat i.e., *Ae. tauschii*, *Ae. speltoides*, *T. urartu*, and A-genome of each *Triticum dicoccoides*, *T. turgidum**,*** and *T. aestivum* cv. Chinese spring were used for the ortholog assignment. UpsetR was used to visualize orthogroups (Conway et al., 2017). For gene family determination, all-vs-all protein blast search was performed using the longest gene isoform (Camacho et al., 2009). MCL clustering was applied to the blast results to group genes into families, and the size of each family was determined (Enright et al., 2002). The OrthoFinder species tree

was converted to ultrametric tree. Computational Analysis of Gene Family Evolution (CAFÉ) was used to assess the expansion and contraction of gene families (De Bie et al., 2006).

NLR repertoire

NLR (nucleotide-binding and leucine-rich repeat) genes were identified within the einkorn genomes using NLR-Annotator (Steuernagel et al., 2020). The study assessed the sequence similarity between NLR (nucleotide-binding leucine-rich repeat) genes within each einkorn wheat genome (intra-genomic redundancy) and between the two einkorn genomes (inter-genomic redundancy) at three levels of sequence identity 95, 98, and 100 percent.

Synteny and genome alignment

MCScanX (Y. Wang et al., 2012) was used to identify the syntenic blocks and duplicated genes between the four einkorn genomes (TA391, TA10868, TA299, and TA10622). Whole genome alignments were performed using CHROMEISTER (Pérez-Wohlfeil et al., 2019).

5.4 Results

Chromosome assembly

The present study sequenced and assembled the genomes of two einkorn wheat accessions: a wild *T. monococcum* ssp. *aegilopoides* (TA391) and a domesticated *T. monococcum* ssp. *monococcum* (TA10868) using Pacbio HiFi sequencing. The assembled contigs consisted of 5.18 Gb and 5.15 Gb sequences with an N50 of 47Mb and 35Mb for the TA391 and TA10868, respectively (Table 5.1, Table S5.2). Reference-based scaffolding of the primary assembly was performed with one of our

previously published genomes of *T. monococcum* ssp. *aegilopoides* (TA299) (Ahmed et al., 2023) using Ragtag (Alonge et al., 2022) generating 7 pseudomolecules and only 65 Mb and 35 Mb of sequence remained unassigned in TA391 and TA10868, respectively. BUSCO analysis of conserved genes identified 97% complete BUSCO genes in both assemblies (Figure 5.2)(Simão et al., 2015). Additionally, merqury (Rhie et al., 2020) was used for reference-free assembly evaluation, revealing 97% solid k-mer completeness between the raw reads and the assembled sequences. Both methods confirm the high level of completeness achieved in our assemblies.

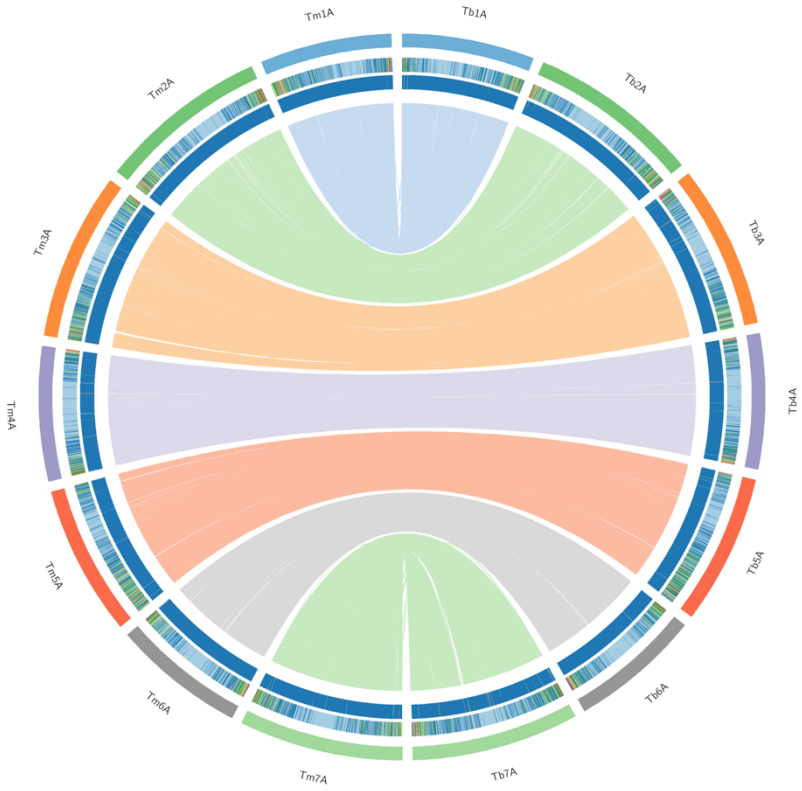


Figure 5.1 Circos plot showing annotation feature of TA391(shown as Tb) and TA10868. (shown as Tm). Tracks from the outside side to the inside represent the chromosome length, gene density, and repeat density. The connecting links show the syntenic genes between two genomes.

Table 5.1 Genome assembly statistics of TA391 and TA10868.

Genome Assembly	**TA391**	**TA10868**
Number of primary contigs	1236	887
Total assembled sequence	5182.825 Mb	5155.417 Mb
N50	47.66 Mb	35.079 Mb
L50	31	44
Longest contig	231.55 Mb	135.206 Mb
Number of contigs > 50 Mb	29 (2511 Mb)	25 (1799Mb)
GC %	0.4629	0.4625
Repetitive elements (%)	83.82 %	83.98 %
Scaffolded into chromosome	5,117 Mb	5,119 Mb
Non scaffolded	65 Mb	35 Mb
Number of high confidence gene	34,717	33,876
Number of low confidence gene	36,795	35,547
Number of Transcription factors	1,989	2,012

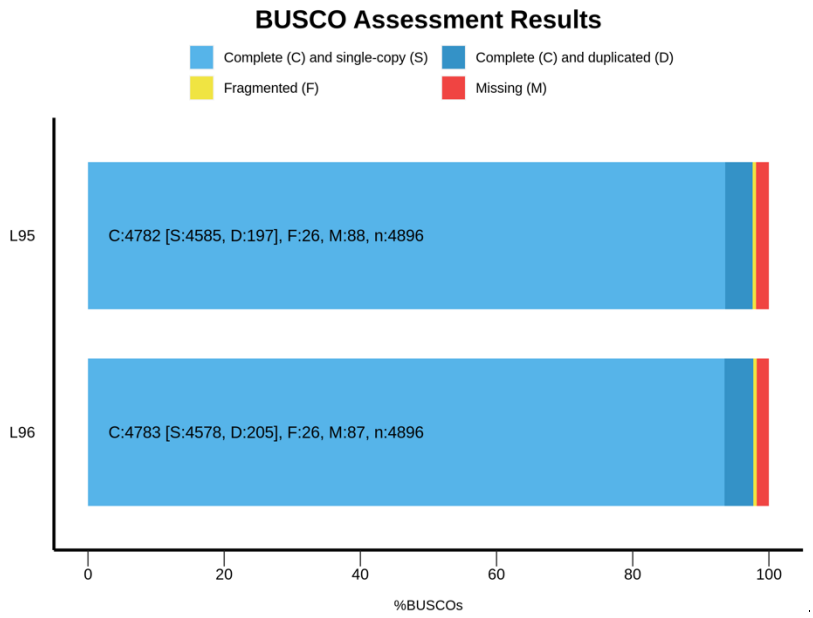


Figure 5.2 Conserved BUSCO-gene-based evaluation of TA391 and TA10868.

Whole genome alignments revealed a higher degree of synteny between the two assembled genomes (TA391 and TA10868) and the A-genome of bread wheat (*T. aestivum* cv. Chinese Spring) (Figure 5.3). This observation is consistent with the previously reported rearrangement in chromosome 4A between wheat and einkorn genomes (Ahmed et al., 2023). Furthermore, the paracentric inversions Inv(4AS;4AL)1 and Inv(4AL;4AL)1 identified in our analysis have been documented earlier in the wheat progenitor *T. urartu* (Dvorak, Wang, Zhu, Jorgensen, Luo, et al., 2018). Notably, both einkorn genomes displayed higher synteny with *T. urartu*, except in the centromeric regions, where some differences were observed compared to bread wheat (Figure 5.3). These findings suggest a complete assembly of the centromeric regions in our assembled genomes.

A total of 34,717 high-confidence and 36,795 low-confidence genes were predicted in TA391, and 33,876 high-confidence and 35,547 low-confidence genes were predicted in TA10868. The average transcript length was 3.3 kb, with a mean of ~ 4.6 exons per mRNA. Notably, 14,002 and 12,973 single-exon genes were identified in TA391 and TA10868, respectively **(**Table 5.1 and Table S5.3**).** Furthermore, a significant number of high-confidence genes exhibited alternate splicing. Accession TA391 harbored 2,616 genes with alternative splicing, while TA10868 possessed 2,585 genes undergoing the splicing process.

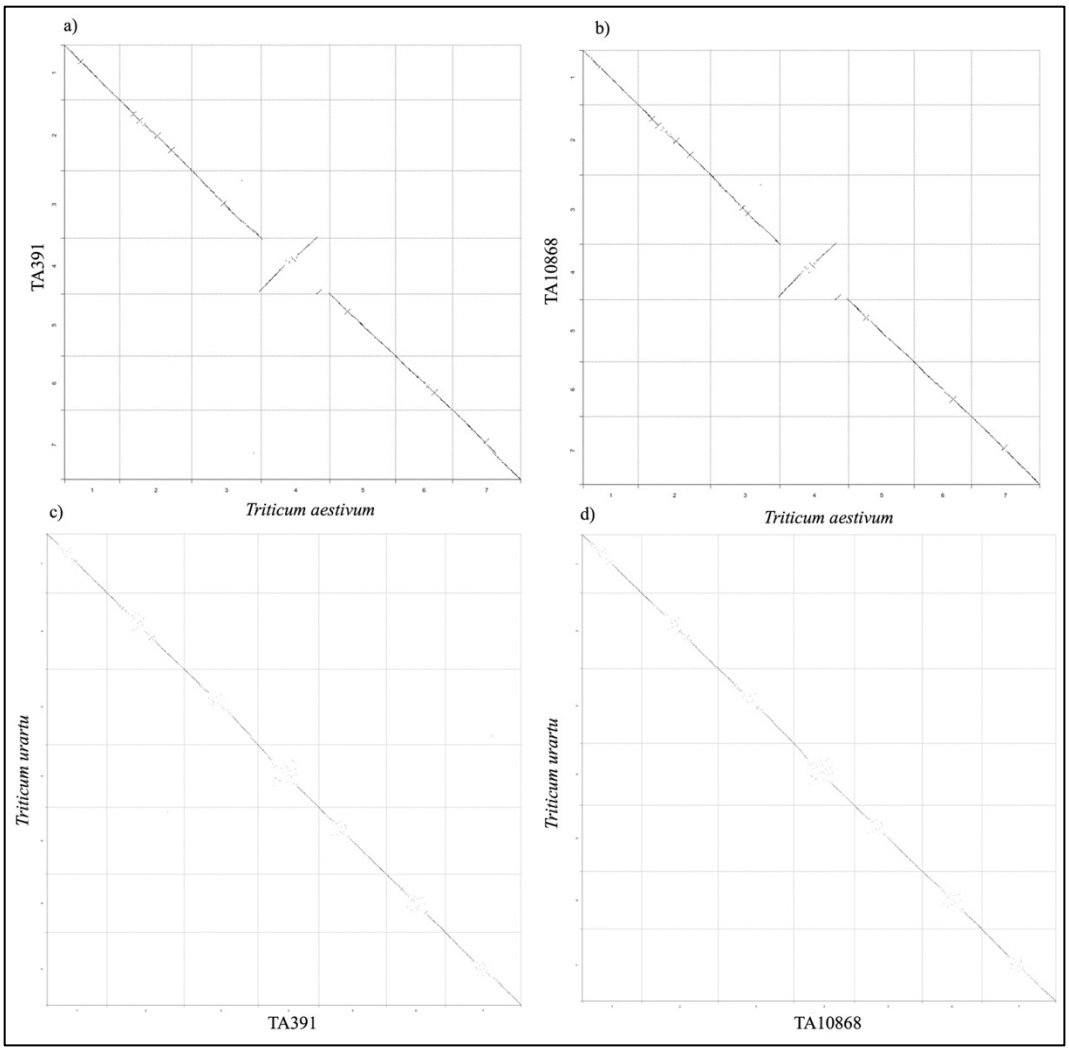


Figure 5.3 Whole genome alignment of a) TA391 with *T. aestivum* cv. Chinese Spring, b) TA10868 with *T. aestivum* cv. Chinese Spring, c) TA391 with *T. urartu*, and d) TA10868 with *T. urartu.*

Repetitive elements

Both einkorn wheat accessions displayed a high proportion of repetitive elements in their genomes, with 83.82% and 83.98% for TA391 and TA10868, respectively (Figure 5.4). Retrotransposons were the most abundant class of repetitive elements, accounting for approximately 68% (TA391) and 67% (TA10868) of the total repeats. DNA transposons constituted 25% and 26% of the total repeats in TA391 and

TA10868, respectively. Among LTR retrotransposons, Gypsy superfamilies were the most abundant, comprising 28.29 % and 27.39 % of the total repeats in TA391 and TA10868, respectively. Copia superfamily showed a lower abundance, constituting 14.92 % and 15.59 % of the repetitive elements in the two genomes, respectively.

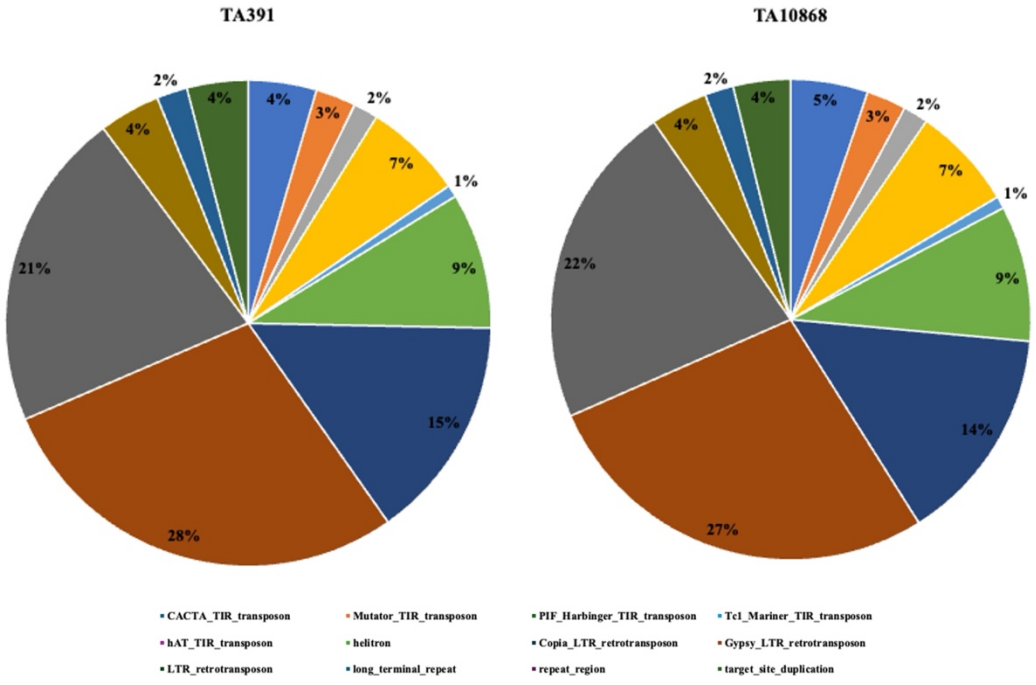


Figure 5.4 Distribution of repetitive elements in TA391 and TA10868.

-Transcription factors

There were 1,989 transcription factors and 572 transcriptional regulators in TA391, while TA10868 harbored 2,012 transcription factors and 568 transcriptional regulators. Notably, both genomes exhibited a consistent distribution of transcription factor classes (Table S5.4). Specifically, 19 genes belonging to the reproductive meristem (REM) sub-family of the B3 class of transcription factors were detected based on interpro domain (IPR039218). The count signifies a 50% reduction compared to previously reported REM family members in *T. urartu* and *T. aestivum* (Ling et al.,

2018). Functionally, the REM subfamily is associated with vernalization and flower development processes (Levy et al., 2002; Swaminathan et al., 2008). These findings suggest a potential link between the reduced number of REM family members in einkorn wheat and its vernalization response. This is particularly interesting considering the absence of vernalization requirements observed in *T. monococcum* compared t*o T. urartu* and *T. aestivum.*

Gene family analysis

The present einkorn wheat genomes (TA391 and TA10868) exhibit a substantial overlap of 32,628 shared orthogroup with previous einkorn assemblies (TA299 and TA10622) and demonstrated 1,589 species-specific orthogroup and 9,024 single copy ortholog containing orthogroup, indicating both shared and unique genetic elements within the einkorn genomes. Notably, 13,282 orthogroups were shared among all four einkorn genomes, with 8,590 ortho-groups observed between TA391 and TA10868, and 5,436 orthogroups between TA299 and TA10622 (Figure 5.5). Comparison of homologous genes between the einkorn assemblies (TA391 and TA10868) and other wheat species (*Ae. tauschii*, *Ae. speltoides*, *T. urartu*, and A-genomes of *T. dicoccoides*, *T. turgidum*, and *T. aestivum* cv. Chinese Spring) identified a total of 86,693 orthogroup with substantial number of species-specific orthogroup (19,747) and 8,658 shared orthogroup (Figure 5.6).

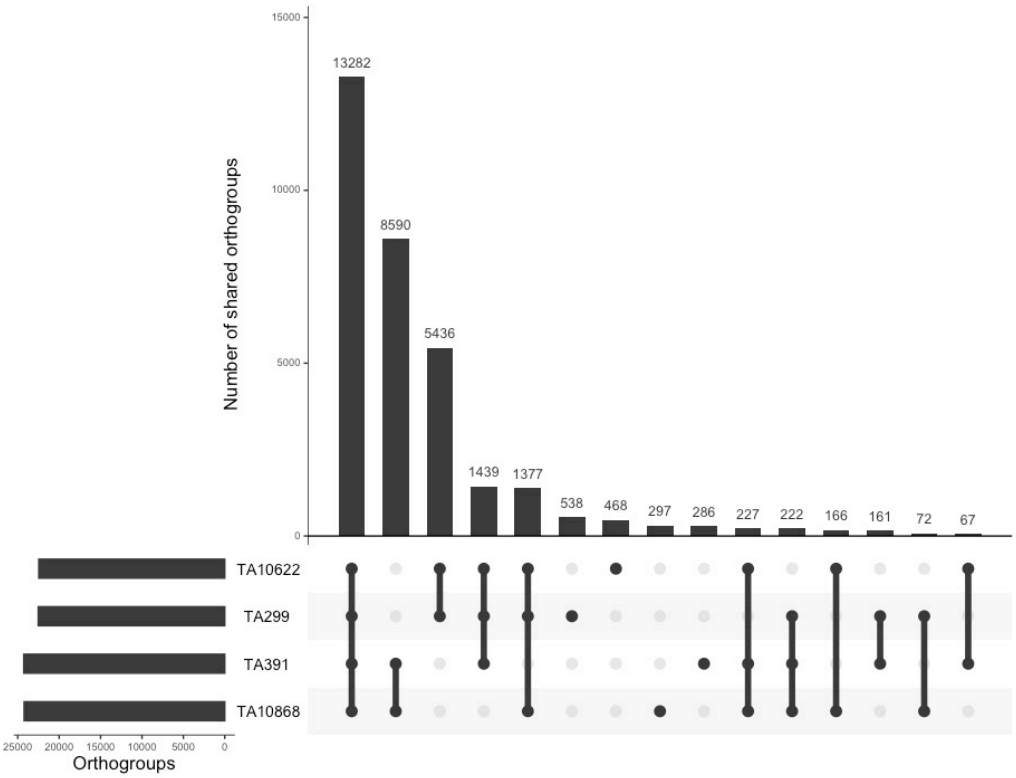


Figure 5.5 Upset plot of orthologous groups between TA391, TA10868, TA299 and TA10622 genomes.

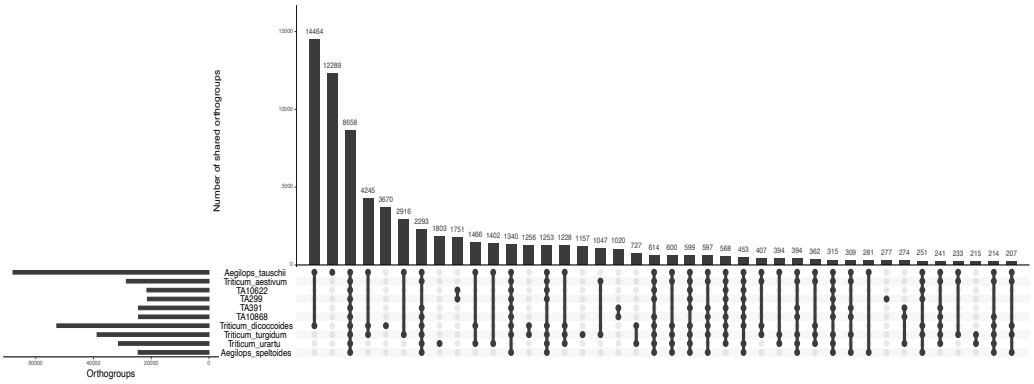


Figure 5.6 Upset plot of orthologous groups of einkorn genome and progenitor wheat species (*Ae. tauschii*, *Ae. speltoides*, *T. urartu*), and the A-genomes of *T. dicoccoides*, *T. turgidum*, and *T. aestivum* cv. Chinese Spring.

A total of 18,503 gene families were initially identified through mcl-based gene clustering (Enright et al., 2002). To simplify the analysis, gene families containing more than 100 members were excluded, resulting in 17,965 families. A uniform birth-death parameter (λ) of 0.4049 was detected for the species tree constructed using OrthoFinder. Significant changes were observed in 1,020 gene families, characterized by a family-wide p-value of ≤ 0.05. Specifically, 645 gene families were found rapidly evolving with family-wide p ≤ 0.01 and Viterbi p-value ≤ 0.01. Interestingly, internodes displayed a higher frequency of gene family loss compared to gene family gain. Notably, among the einkorn genomes investigated, TA391 and TA10868 demonstrated a notable pattern of expansion and contraction for gene families (Figure 5.7).

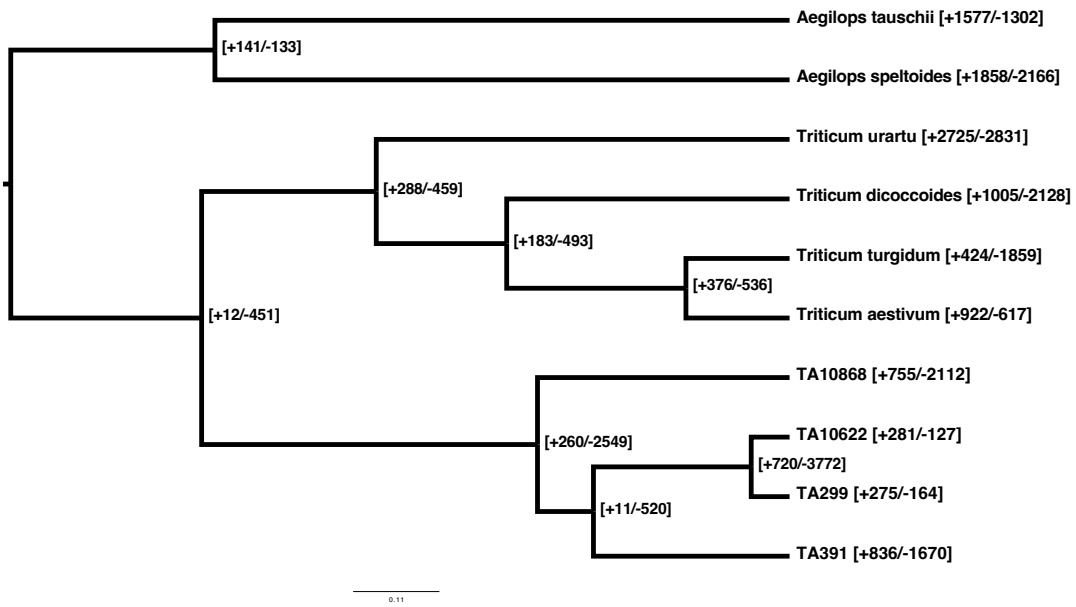


Figure 5.7 Evolution of gene families in einkorn and progenitor species of wheat. Numbers in the square bract show the gene family gain (positive) and gene family loss (negative).

Divergence time estimation for einkorn genome

The divergence time estimates of the species were retrieved from TIMETREE5 (http://www.timetree.org/) at diverging nodes *Ae. speltoides* and *T. monococcum* (3.54 MYA). The substitution rate per unit time was calculated to be approximately 0.004447 substitutions per million years based on 603 single-copy orthologs and 215609 sites. The estimation resulted in the inferred divergence time of approximately 1.532 million years for the separation of species *T. urartu* and *T. monococcum* (Figure 5.8).

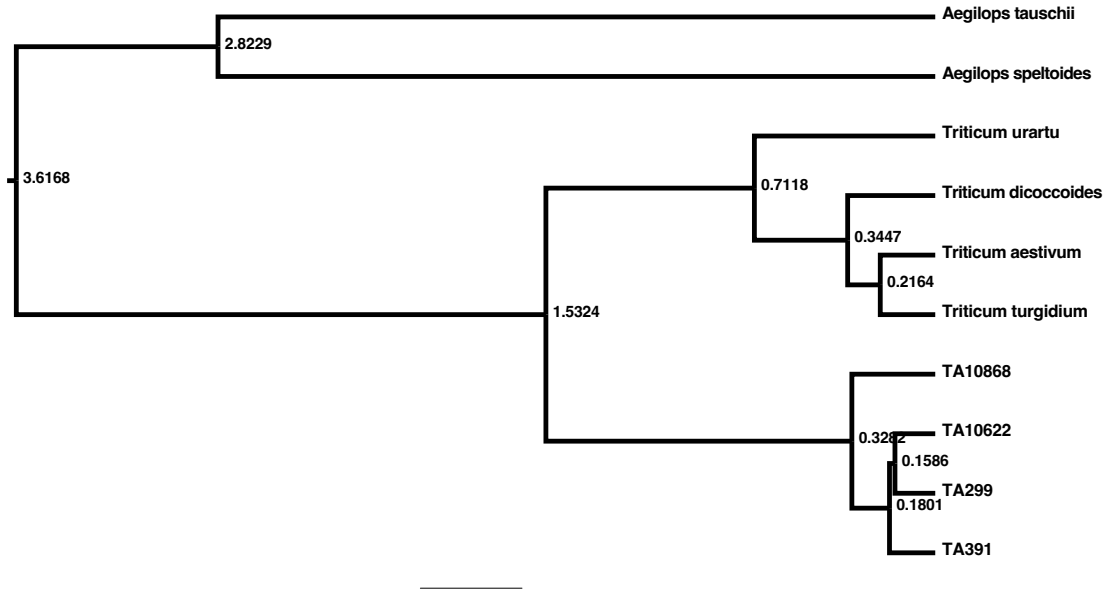


Figure 5.8 Divergence time estimate for *T. monococcum* and *T. urartu*. Numbers at the nodes represents the evolution time MYR.

NLR repertoire of einkorn genome

Eliminating identical or duplicated predictions of R genes at 100% sequence identity, about 948, 958, 908, and 915 unique R-genes were identified in TA391, TA10868, TA299, and TA10622, respectively. Interestingly, only 10 identical R-genes were detected between TA391 and TA10868, whereas comparisons with other einkorn genomes displayed a higher degree of similarity within subspecies a*egilopoides* and

subspecies *monococcum* (Table 5.2). Clustering of the NLR repository retained 1,955 genes at 98 % sequence similarity and 3,346 genes at 100% sequence similarity (Figure 5.9). Orthologous clustering revealed 414 single-copy orthologs, while 423 orthogroup consisted of more than two accessions. Additionally, 13, 15, 12, and 12 R genes were identified as singletons in TA391, TA10868, TA299, and TA10622, respectively. Notably, a relatively low number of species-specific orthogroup (TA391: 22, TA10868: 24, TA299: 16, and TA10622: 12) were detected. The phylogenetic tree of NLR genes in TA391, TA10868, TA299, and TA10622 illustrates NLR diversity of the einkorn genomes (Figure 5.10).

Table 5.2 Number of identical R-genes between TA391, TA10868, TA299 and TA10622 genomes.

Genome	**Other genomes**	**Number of identical R-gene**
TA391	TA10868	10
	TA299	159
	TA10622	10
TA10868	TA299	14
	TA10622	189

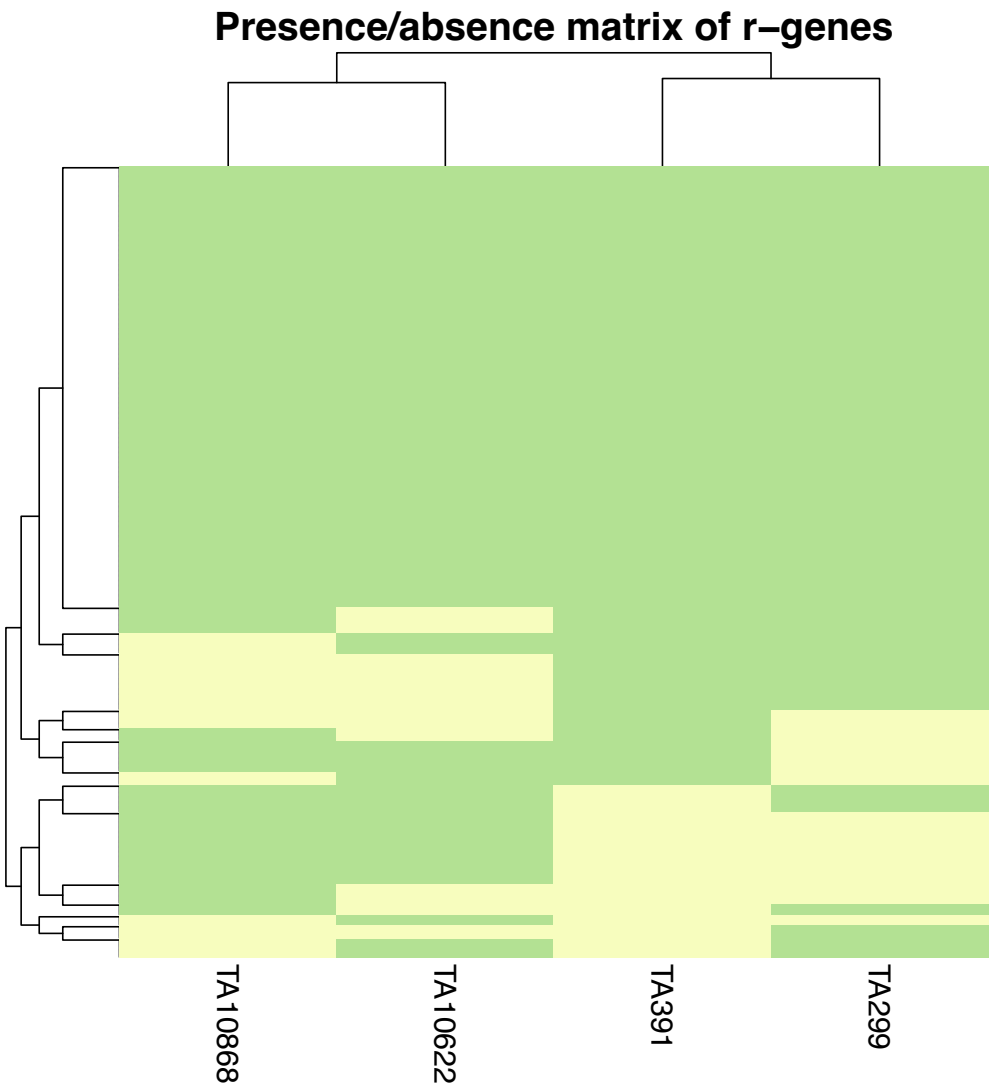


Figure 5.9 Heatmap of shared and species-specific R-genes clusters between four einkorn genomes. Green color shows the presence of the R-gene cluster, and yellow shows absence.

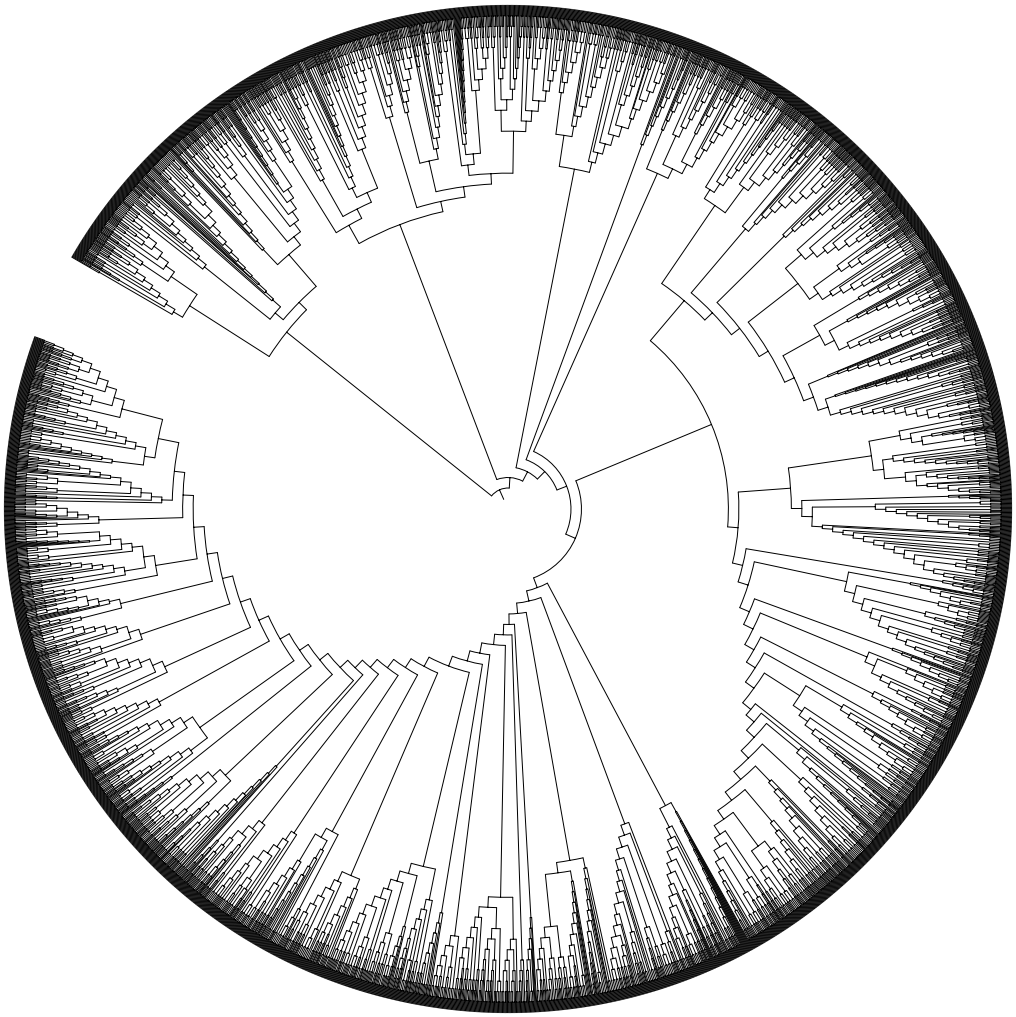


Figure 5.10 Phylogenetic tree of NLR genes in four einkorn genomes illustrates NLR diversity.

Synteny analysis

Gene duplication events constitute a significant source of genetic variation and can facilitate the emergence of new functions. Both TA391 and TA10868 displayed a similar distribution of duplication types, with dispersed duplications being the most abundant category, followed by tandem duplications. Notably, whole genome duplications (WGD) and segmental duplications also showed consistency between the genomes (Table 5.3). The percentage of WGD genes ranged from 4% to 8% per

chromosome in both genomes, while tandemly duplicated genes varied from 12 to 18 percent (Table 5.4). The results suggest a potential role for these duplications in shaping the genome architecture. Additionally, both assemblies exhibited a similar pattern of whole genome duplications (Figure 5.11). Furthermore, tandemly duplicated genes were enriched in Gene Ontology (GO) categories associated with protein binding, ATP binding, protein phosphorylation, protein kinase activity, DNA binding, regulation of DNA-templated transcription, ADP binding, proteolysis, and carbohydrate metabolic processes.

Table 5.3 Whole genome duplication of genes on the chromosomes in two einkorn genomes.

Duplication Type	**TA391**	**TA10868**
Singleton	2569	2532
Dispersed	24735	23855
Proximal	2681	2694
Tandem	5011	5059
WGD or segmental	1893	1879

Table 5.4 Chromosome-wise distribution of duplicated genes in TA391 and TA10868.

	Whole genome duplication		**Tandem gene duplication**	
Chromosome	**TA391**	**TA10868**	**TA391**	**TA10868**
chr1A	331	341	566	540
chr2A	233	214	984	959
chr3A	321	317	771	807
chr4A	180	176	419	442
chr5A	267	267	800	853
chr6A	325	318	540	549
chr7A	236	246	877	875
total	1893	1879	4957	5025

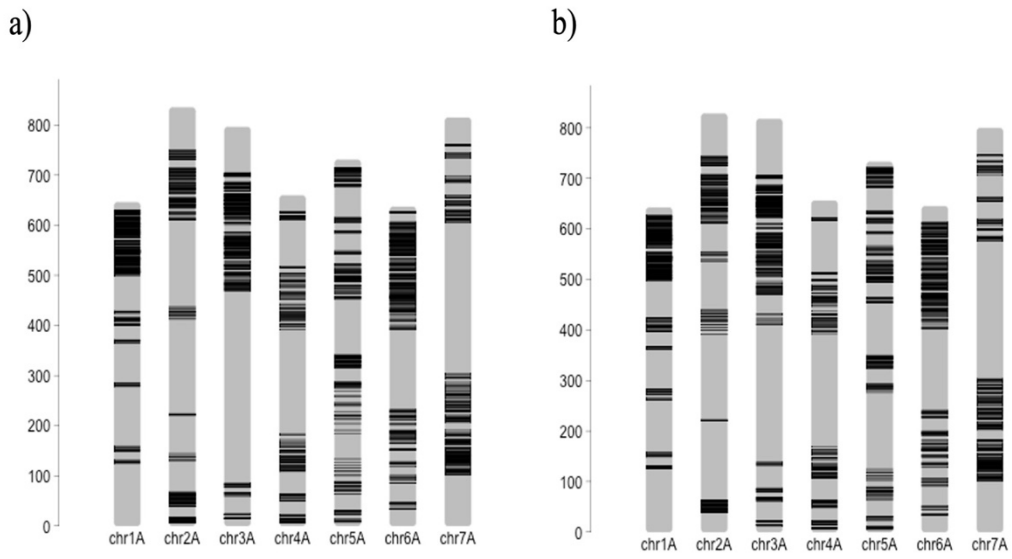


Figure 5.11 Distribution of genes involved in whole genome duplications a) TA391 and b) TA10868.

Large tandem repeats and Large segmental duplications

Self-alignment of TA391 and TA10868 genomes did not reveal any translocation or inversion events. TA391 exhibited 12 tandem repeats exceeding 500Kb in length, with four repeats surpassing 1Mb, while TA10868 displayed 16 tandem repeats exceeding 500Kb and four repeats exceeding 1Mb, with a higher prevalence on chr7A in TA391 and chr3A in TA10868 (Table 5.5). Using SDDetector, large segmental duplications were identified. Accession TA10868 showed 47 duplications of larger than 100 Kb ranging up to 770 Kb, while TA391 exhibited 40 regions ranging from 100 Kb to 400 Kb. Notably, both studies revealed a higher frequency of duplicated regions in TA10868 compared to TA391.

Table 5.5 Chromosome-wise distribution of tandem repeats of more than 500Kb identified using tandem repeat finder.

Chromosome	Number of tandem repeats	
	TA391	TA10868
chr1A	1	1
chr3A	2	4
chr4A	1	3
chr5A	2	2
chr6A	1	3
chr7A	6	3

Analysis of chromosomal alignments using CHROMIESTER revealed a total of 115 structural rearrangements between the TA10868 and TA391 genomes. These rearrangements can further be classified into 64 transpositions, 40 inversions, and 11 inverted transpositions (Table 5.6). Notably, a higher frequency of inversion was observed on chromosome 4, while chromosome 3 exhibited greater number of inverted transpositions. Chromosome 7, on the other hand, showed a higher prevalence of transposition compared to other chromosomes (Figure 5.12).

Table 5.6 Chromosomal rearrangements between TA10868 and TA391.

Chromosome	Inversion	Inverted transposition	Transposition
chr1A	4	2	-
chr2A	7	3	-
chr3A	4	5	26
chr4A	10	-	-
chr5A	3	-	-
chr6A	9	1	-
chr7A	3	-	38

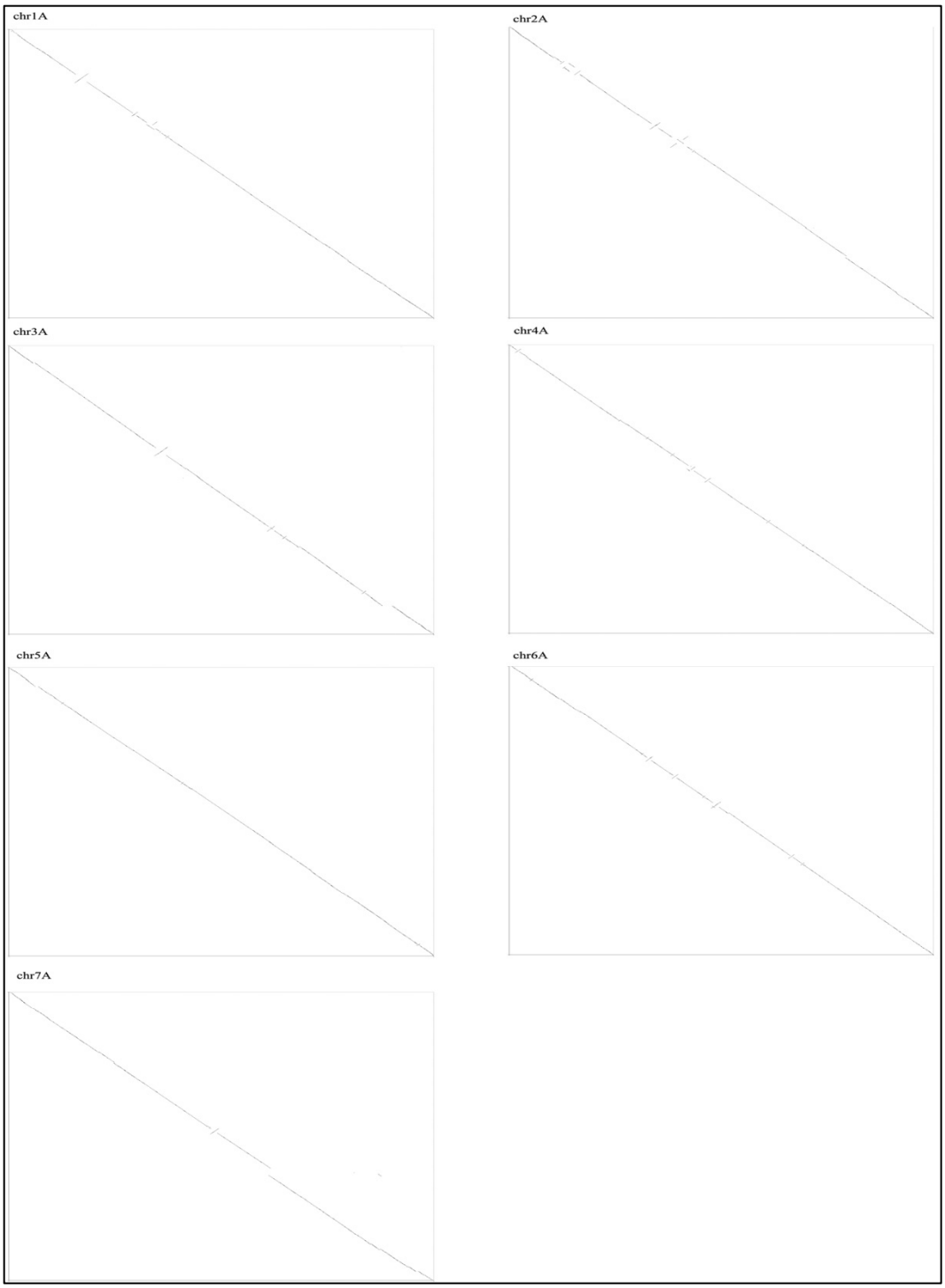


Figure 5.12 Chromosomal rearrangements between TA10868 and TA391 identified using CHROMIESTER.

5.5 Discussion

Einkorn wheat, the first domesticated wheat species, can offer a glimpse into the genetic makeup of the ancestors of modern hexaploid wheat. Studying its genome provides insights into wheat domestication and evolution. Being ancient and with its wild growing habitat einkorn possesses unique genes and variations not found in modern wheat. These variations might offer potential for breeding programs to improve disease resistance, stress tolerance, or grain quality in cultivated wheat. Comparing einkorn genomes to modern wheat varieties helps us to understand how the wheat genome has changed over time through selection and breeding practices. Einkorn wheat's diploid genome offers a simple model system for wheat genetics research compared to the complex hexaploid genome of bread wheat. The present study generated high-quality reference genome assemblies for two einkorn wheat accessions: a wild *T. monococcum* ssp. *aegilopoides* (TA391) and a domesticated *T. monococcum* ssp. *monococcum* (TA10868). These assemblies achieved a high level of completeness, with a BUSCO score exceeding 97 percent. Additionally, both the assemblies exhibited high levels of repetitive elements (83%) and a high gene content (TA391: 34,717 high confidence genes and TA10868: 33,876 high confidence genes).

Analysis of whole genome alignments revealed a strong synteny between the einkorn genomes (TA391 and TA10868) and the A-genome of bread wheat (*T. aestivum* cv. Chinese Spring) except for the rearrangements identified in chromosome 4A. This aligns with previous reports of rearrangements in chromosome 4A between einkorn and bread wheat (Ahmed et al., 2023). We detected the paracentric inversions (Inv(4AS;4AL)1 and Inv(4AL;4AL)1) on chromosome 4A, documented earlier in the

wheat progenitor *T. urartu* (Dvorak, Wang, Zhu, Jorgensen, Luo, et al., 2018). Transcription factors and transcriptional regulators constitute the highly abundant class, comprising 7.37% and 7.61% of the total high-confidence genes in TA391 and TA10868, respectively. Both the assemblies harbored nearly 1,589 species-specific orthogroups, potentially representing genes associated with adaptations. Our analysis revealed higher prevalence of gene family loss compared to gene gain during the divergence of einkorn wheat from other wheat progenitor species, which includes 645 rapidly evolving gene families, exhibiting strong evidence of positive selection. Our analysis revealed a greater level of similarity within R gene class belonging to the same einkorn subspecies (*T. aegilopoides* and *T. monococcum*). This suggests a potentially conserved repertoire of resistance genes within these sub-species. Analysis of gene duplications revealed a higher percentage of tandemly duplicated genes, which varies from 12-18 percent per chromosome. These findings suggest that tandem duplications might have played a substantial role in shaping the genetic landscape of the einkorn genome.

Overall, this study provides a valuable foundation for further research on einkorn wheat. The high-quality genome assemblies, coupled with insights into gene content, regulatory elements, and gene duplication events, pave the way for a deeper understanding of einkorn biology and its potential contribution to wheat improvement.

5.6 Conclusion

The study delivers high-quality reference assemblies of two einkorn genomes (TA391 and TA10868), providing valuable insights into their unique genetic makeup. The presence of species-specific genes, R gene clusters, transcription factors, and

rapidly evolving gene families underscores the distinct genetic potential of einkorn wheat. This resource offers an intriguing opportunity to explore and identify agriculturally significant traits in diploid wheat. The identified loci/markers hold promise for application in breeding programs to enhance biotic and abiotic stress tolerance in tetraploid and hexaploid wheat varieties.

5.7 Supplementary Material

The supplementary information is provided in the Proquest.

Chapter 6: Trait Genomics in Einkorn Wheat for Accelerating Gene Discovery

6.1 Abstract

To feed the ever-growing human population, wheat breeders have focused on developing high-yielding varieties, which reduce genetic diversity and increase its vulnerability to various biotic and abiotic stresses. Einkorn wheat (*T. monococcum*), with its diploid genome, high disease resistance, and similarity to the A-genome of bread wheat, represents an excellent model for gene discovery. The present study employed kmer-based genome-wide association studies (GWAS) on a comprehensive association panel encompassing wild and domesticated *T. monococcum* accessions. The study aimed to identify the novel sources of resistance against the three most devastating wheat diseases: powdery mildew, leaf rust, and stem rust. Seven novel Quantitative Trait Loci (QTLs) associated with disease resistance were identified: six for highly virulent powdery mildew races, three for leaf rust resistance, and two for stem rust resistance. These genes present novel sources of resistance and includes RGA, cysteine-rich receptor kinases, transcription factors and G-type lectins. The integration of these novel QTLs into breeding programs holds immense potential for developing new disease-resistant wheat varieties.

6.2 Introduction

Breeding for the high-yielding traits leads to reduced genetic diversity and increased susceptibility in cultivated wheat, which faces continuous threats from changing climates and evolving biotic and abiotic stresses (Enghiad et al., 2017). Wild

and related germplasm of wheat is considered a gold mine for genetic diversity, and can help in replenishing the depleted genetic diversity of bread wheat. Gene introgression of favorable alleles and gene/gene complexes from wild relatives offers an excellent and efficient opportunity to enhance the genetic resistance of the cultivated gene pool for various desirable traits (Leigh et al., 2022). The wider wheat gene pool consists of three groups based on the genomic constitution of the species, and comprised of primary, secondary, and tertiary member species. The primary gene pool includes closely related species with shared genomes (diploid and tetraploid progenitors). The secondary gene pool comprises species sharing one of the genomes with wheat and consists of polyploid *Triticum* and *Aegilops* species. The tertiary gene pool encompasses species with non-homologous genomes, representing the most distant genetic relatives (Chaudhary et al., 2014). The progenitor species and wild germplasm co-existed for millions of years in the Fertile Crescent and accumulated variations through mutations as well as through natural introgressions between the members of the different gene pools (Vardi & Zohary, 1967). To uncover economically important genes, Zohary et. al. (1969) recommended the breeders and geneticists to use diploid and tetraploid wheat progenitors in the breeding program (Zoharyi et al., 1969).

Einkorn wheat (*T. monococcum*) is the only A-genome-containing diploid wheat that occurs both in wild and domesticated forms in nature. It is closely related to *Triticum urartu*, the A-genome progenitor of hexaploid wheat. *T. monococcum* genome is under-represented in the gene pool, with its contribution limited to *T. zhukovskyi* (A^mA^mAAGG), whereas *T. urartu* shows dominance with its contributions towards multiple species like *T. turgidum* (AABB), *T. timopheevii* (AAGG) and *T. aestivum*

(AABBDD) (Dvorak et al., 1993). Einkorn wheat holds considerable breeding potential of a sustainable crop (Cox, 1997). Bread wheat is constantly challenged by the biotic stresses from diseases i.e., powdery mildew and rust. Powdery mildew (PM), caused by *Blumeria graminis* f. sp. *tritici*, is one of the most destructive foliar diseases worldwide (Wu et al., 2021). Several genes (*Pm1a, Pm1b,* and *Pm25*) have been introgressed into the hexaploid wheat from the diploid progenitor for powdery mildew resistance. Till date, sixteen genes belonging to diverse classes of NBS-LRR, ABC transporter, hexose transporter and tandem kinase classes have been cloned i.e., *Pm3* (Yahiaoui et al., 2004), *Pm38* (Krattinger et al., 2009), *Pm8* (Hurni et al., 2013), *Pm46* (Moore et al., 2015) , *Pm2* (Sánchez-Martín et al., 2016) , *Pm21*(Cao et al., 2011), *Pm17* (S. P. Singh et al., 2018) , *Pm60* (Zou et al., 2018), *Pm5* (J. Xie et al., 2020), *Pm24* (Lu et al., 2020) and *Pm41* (M. Li et al., 2020). Senic et. al. (2005) reported two novel dominant resistance genes from *T. monococcum* and *T. timopheevii* (Srnić et al., 2005). Yao et al. (2007) reported two new alleles for one of the cloned gene *Pm1* in *T. monococcum* (Yao et al., 2007). In another study, a recessive powdery mildew resistance gene *pm2026* was identified and introgressed on chromosome 5AL in wheat from einkorn wheat (H. Xu et al., 2008). Chhuneja et al. (2012) reported two novel powdery mildew genes (*PmTb7A.1* and *PmTb7A.2*), which were found linked to *Pm1* from *T. boeoticum* acc. pau5088 (Chhuneja et al., 2012). Wang et al. (2023) recently identified a novel resistance gene, *PmNJ3946*, located on chromosome 3A of einkorn wheat accession TA2032 (P. Wang et al., 2023).

Leaf rust caused by *Puccina triticina* can yield losses up to 60% in wheat (Kolmer, 2013). Nearly 82 *Lr* genes have been designated in the literature (X. Xu et

al., 2022). The B-genome relatives have been shown to contribute higher number of rust resistance genes to wheat compared to the A and D genome relatives (Kumar et al., 2022). Till date 10 *Lr* genes have been cloned, which includes *Lr10* (Feuillet et al., 2003), *Lr21* (Huang et al., 2003), *Lr1* (Cloutier et al., 2007), *Lr34* (Krattinger et al., 2009), *Lr67* (Moore et al., 2015), *Lr22a* (Thind et al., 2017), *Rph1a* (Dracatos et al., 2019), *Lr9* (Y. Wang et al., 2023), *Lr14a* (Kolodziej et al., 2021), *Lr13*(Hewitt et al., 2021), *Lr42* (Lin et al., 2022).

Stem rust caused by *Puccinia graminis* f. sp. *tritici* (*Pgt*) also poses challenges to wheat production. In 1999, a highly virulent race named TTKSK (Ug99 isolate) emerged, rendering most existing wheat cultivars susceptible in Africa, Asia, and the Middle East, posing significant threat to global food security (Rouse & Jin, 2011). Another aggressive race, TKTTF, caused severe stem rust epidemics in Ethiopia between 2012 and 2014 (P. Olivera et al., 2015). Another closely related race, RRTTF, has overcome the resistance conferred by genes *Sr36*, *SrTmp*, and *Sr1RSAmigo* (P. D. Olivera et al., 2019). Similarly, race TTRTF triggered an epidemic in southern Italy by overcoming resistance to genes *Sr13b*, *Sr21*, *Sr35*, *Sr45*, and *Sr50* (Bhattacharya, 2017). Despite the identification of nearly 60 designated resistance genes (*Sr* genes), the ongoing emergence of new virulent races highlights the need for continuous development of resistant cultivars (G. Yu et al., 2022). Einkorn wheat has contributed five *Sr* genes- *Sr*21 (Chen et al., 2015), *Sr*22a/*Sr*22b(Luo et al., 2022) and *Sr*35 (Saintenac et al., 2013), *Sr*60 (Chen et al., 2018), *SrTm4* (H. Li et al., 2023). Notably, all except *SrTm4* have been successfully cloned from *T. monococcum*.

The identification of numerous Quantitative Trait Loci (QTL) for resistance against various biotic and abiotic stresses underscores its potential as a valuable resource for wheat improvement (K. Singh, Chhuneja, et al., 2007; K. Singh, Ghai, et al., 2007; K. Yu et al., 2017). Breeders can leverage techniques like introgression breeding and distant hybridization to incorporate resistance traits from *T. monococcum* into cultivated wheat varieties, leading to the development of more resilient crops (Saripalli et al., 2023; Wulft & Moscou, 2014).

Genome-wide association studies (GWAS) have emerged as a powerful tool for identifying quantitative trait loci (QTL) associated with disease resistance in wheat. Recent successes include the discovery of QTL for leaf rust (Aoun et al., 2021; Genievskaya et al., 2022; Saccomanno et al., 2018), stripe rust (Liu et al., 2017), FHB (Sari et al., 2020), and powdery mildew resistance (Simeone et al., 2020). However, SNP-based GWAS has limitations i.e., need of a reference genome which makes its use difficult for species without a reference genome or incomplete genomes. Additionally, SNP-based GWAS may overlook structural variations like insertions/deletions (indels) and copy number variations (CNVs), as well as potential biases due to sequencing errors. These limitations may be potentially overcome by using alignment-free methods, which have been proven useful in detecting rare disease variants. The method uses a string of sequence of defined length known as k-mer rather than the SNP as a marker. Gaurav et. al. (2022) has successfully used the kGWAS in *Ae. tauschii* for the identification of disease resistance and quality traits-related genes (Gaurav et al., 2022). The method is highly efficient and applicable to polyploid crops like wheat.

The present study utilizes K-mer based genome-wide association studies (GWAS) on a diverse *T. monococcum* association panel (TmGWAS) to unlock the hidden potential of einkorn wheat for resistance against major diseases like powdery mildew, leaf rust, and stem rust.

6.3 Material and Methods

Whole-genome sequencing of the einkorn diversity panel

Genomic DNA was extracted from one or two young leaves of 220 einkorn accessions using CTAB extraction (Abrouk et al., 2020). DNA quantification was performed using the Qubit dsDNA HS Assay (Q32851, Thermo Fisher Scientific), the purity was assessed using the Nanodrop spectrophotometer by checking the 260/280 and 260/230 ratios, and the integrity was confirmed by analyzing 1 µl per sample on a 1% TAE agarose gel. Library preparation and sequencing (150 bp paired-end libraries) were performed by Novogene using the Illumina NovaSeq 6000 system. Whole-genome sequencing data for the 220 einkorn accessions were first trimmed using trimmomatic99 (v.0.38; parameters: SLIDINGWINDOW:5:20 MINLEN:5)(Bolger et al., 2014).

Phenotyping of T. monococcum panel for Powdery mildew (B. graminis f. sp. tritici)

Phenotyping for the six most virulent powdery mildew isolates (Bgt96224, Bgt98230, Bgt98411, BgtIRN_GOR-5, Bgt_KAZ-1b and Bgt_THUN12) was done in Switzerland using the protocol described previously (Sánchez-Martín et al., 2016), provided by the collaborators. The total percentage of leaf area covered with postulates was accessed in replicates of four for each *T. monococcum* accession. Disease levels

were assessed 7–9 d after inoculation, and were categorized into five classes: Resistance (R, covering 0–10% of leaf area), Intermediate Resistance (IR, covering 10–25% of leaf area), Intermediate (I, covering 25–50% of leaf area), Intermediate Susceptible (IS, covering 50–75% of leaf area), and Susceptible (S, covering more than 75% of leaf area).

Phenotyping of T. monococcum panel for Leaf rust (Puccinia triticina)

The association panel was inoculated with a mixture of common leaf rust races found in the US and are virulent to all the known *Lr* genes in the US wheat cultivars. Two tests were conducted for the infection type (IT) on dates 5-5-2022 and 5-25-2022. The plants were inoculated at the second leaf stage about 10 days after planting and IT for the first leaf was recorded. The phenotypic data provided by the collaborators was used in analysis.

Phenotyping of T. monococcum panel for stem rust (Puccinia graminis)

The *T. monococcum* association panel was phenotyped for the three most virulent stem rust races in US i.e., TTKSK, TKKTF+ and TTRTF. The disease response was recorded in two replicates after 14 dpi (4-18-2022 and 4-27-2022). The severity of the disease was initially measured on the stackman scale (Stakman et al., 1962) and was converted into 0-9 scale for easy interpretation (Riaz et al., 2016), where the average was considered for complex scores. Disease levels were categorized into five classes based on IT scores: resistant (< =1), moderately resistant (>1 and <3), intermediate (>3 and <6), moderately susceptible (>6 and <7) and suspectable (>7). The GWAS panel indicated higher susceptibility for the race TTRTF, which was excluded from the analysis.

K-mer-based association mapping

The K-mer presence/absence matrix was created following the protocol described by (Gaurav et al., 2022) with some modifications in the original method. Jellyfish (jellyfish-2.3.0) was used to count the k-mer of size 51 from the filtered fastq reads. *k*-mers with a count of less than two in an accession were discarded immediately. Changes in the original protocol involves developing Jellyfish database using "–disk" parameter to make databases of similar sizes. Later, the utility script count_in_file (https://github.com/gmarcais/Jellyfish/tree/master/examples/count_in_file) was used to make a composite matrix of the kmers. A custom python script was used to convert the matrix into presence/absence matrix by replacing the occurrence of k-mer with 1 and absence with 0. Matrix was further filtered for k-mers occurring in less than two accessions or all but one accession. For the rest of the analysis protocol described by (Gaurav et al., 2022) was used. Association mapping plots were generated by mapping the associated k-mer to one of the reference assemblies: TA391 (powdery mildew), TA10868 (leaf rust, stem rust and stripe rust).

Analysis of the GWAS results

The identified k-mer block of 10Kb was extended to 163kb on both sides based on the LD (163 Kb) of the einkorn genome. Genes were searched from the genome annotation file. Orthofinder (Emms & Kelly, 2019) was used to identify the orthologs of the genes among the four of the studied reference genomes (TA391, TA10868, TA299 and TA10622). In the presence of multiple orthologs of a gene, the reciprocal best blast hit method was used to select the best ortholog as described in the method (http://archive.sysbio.harvard.edu/CSB/resources/computational/scriptome/UNIX/Pro

tocols/Sequences.htm). For the identification of potential candidates involved in disease resistance, genes were searched for their potential role in plant defense using functional annotation and GO terms. Genes with 100% identical orthologs in the four reference genomes were removed from the analysis, and the rest of the genes were analyzed for variations in gene structure between domesticated and wild einkorn wheat. The haplotype of genes was studied using CandiHap (X. Li et al., 2023).

6.4 Results

Gene bank collections of germplasm are a good resource for the identification of genes conferring resistance to diseases. The study utilizes a GWAS panel comprised of 220 *T. monococcum* accessions (wild and domesticated, Table S6.1) to identify the novel sources of resistance for powdery mildew (*Blumeria graminis* f. sp. *tritici*) , leaf rust (*Puccinia triticina*) , stem rust (*Puccinia graminis* f. sp. *tritici)* using kmer-GWAS method.

Powdery mildews (PW)

For the PW, 6 most virulent races of *Bgt* were used in the study, which include 1) BgtIRN_GOR-5, 2) Bgt96224, 3) Bgt98230, 4) Bgt98411, and 5) Bgt_KAZ-1b and 6) Bgt_THUN12. Figure 6.1 shows the phenotypic distribution of the powdery mildew races. Most of the domesticated einkorn accessions were found to be resistant to all the studied races compared to wild type einkorn accessions. The severity scores were recorded as the percentage of leaf area covered by the fungal pustules. Data was recorded in replicates of four with scores ranging from 0 (highly resistant) to 100 (highly susceptible). Which was further categorized into a five-step scale; a) resistant

(0-10%), b) Intermediate Resistant (10-25%), c) Intermediate (25-50%) d) Intermediate Susceptible (50-75%), and e) Susceptible > 75% for the analysis, shown in figure 6.2.

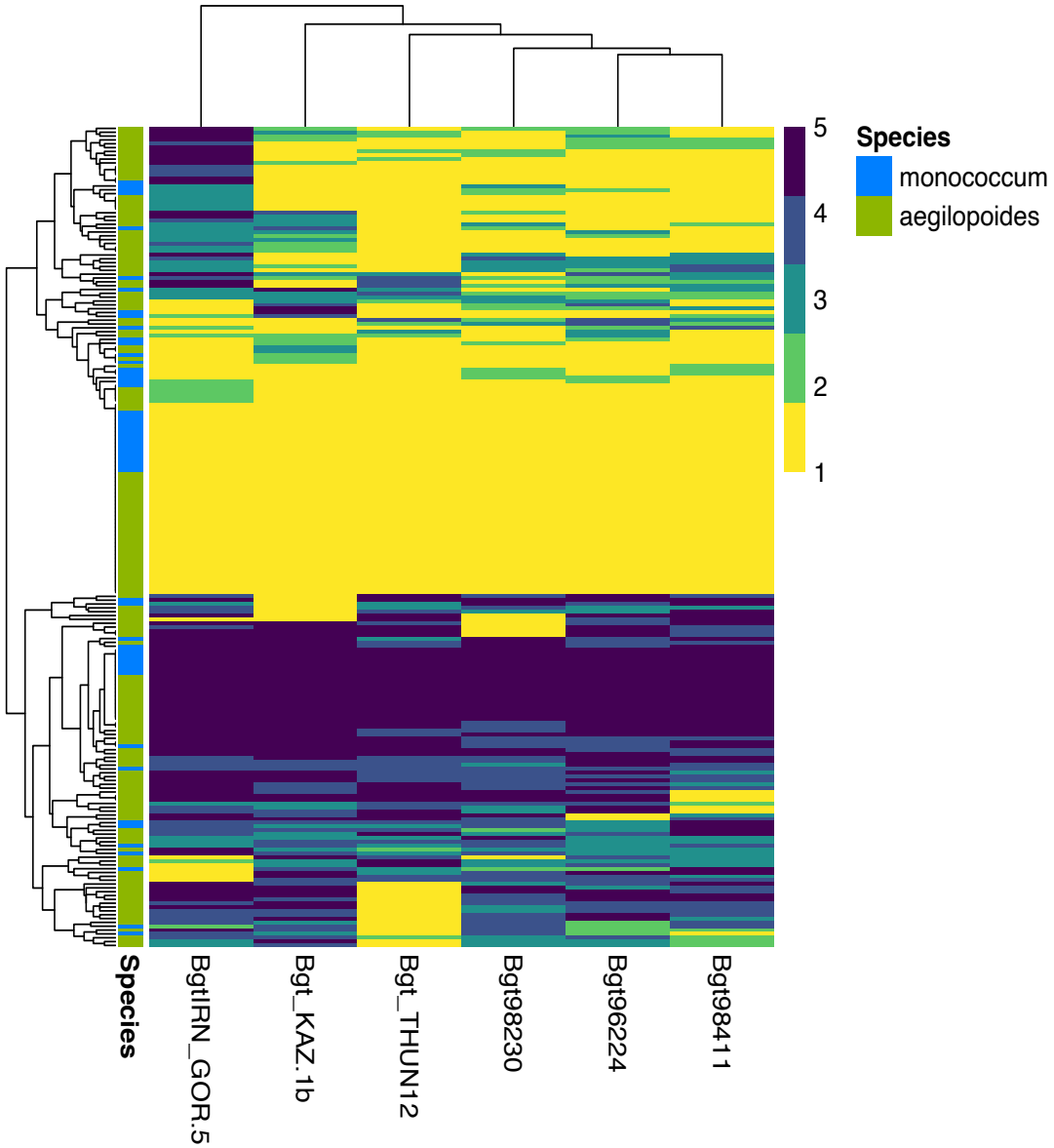


Figure 6.1 Heat map showing the clustering of the *T. monococcum* accessions based on their phenotype scores. The five-step scale; a) Resistant (0-10%, 5), b) Intermediate Resistant (10-25% ,4), c) Intermediate (25-50%, 3) d) Intermediate Susceptible (50-75%, 2), and e) Susceptible (> 75%, 1).

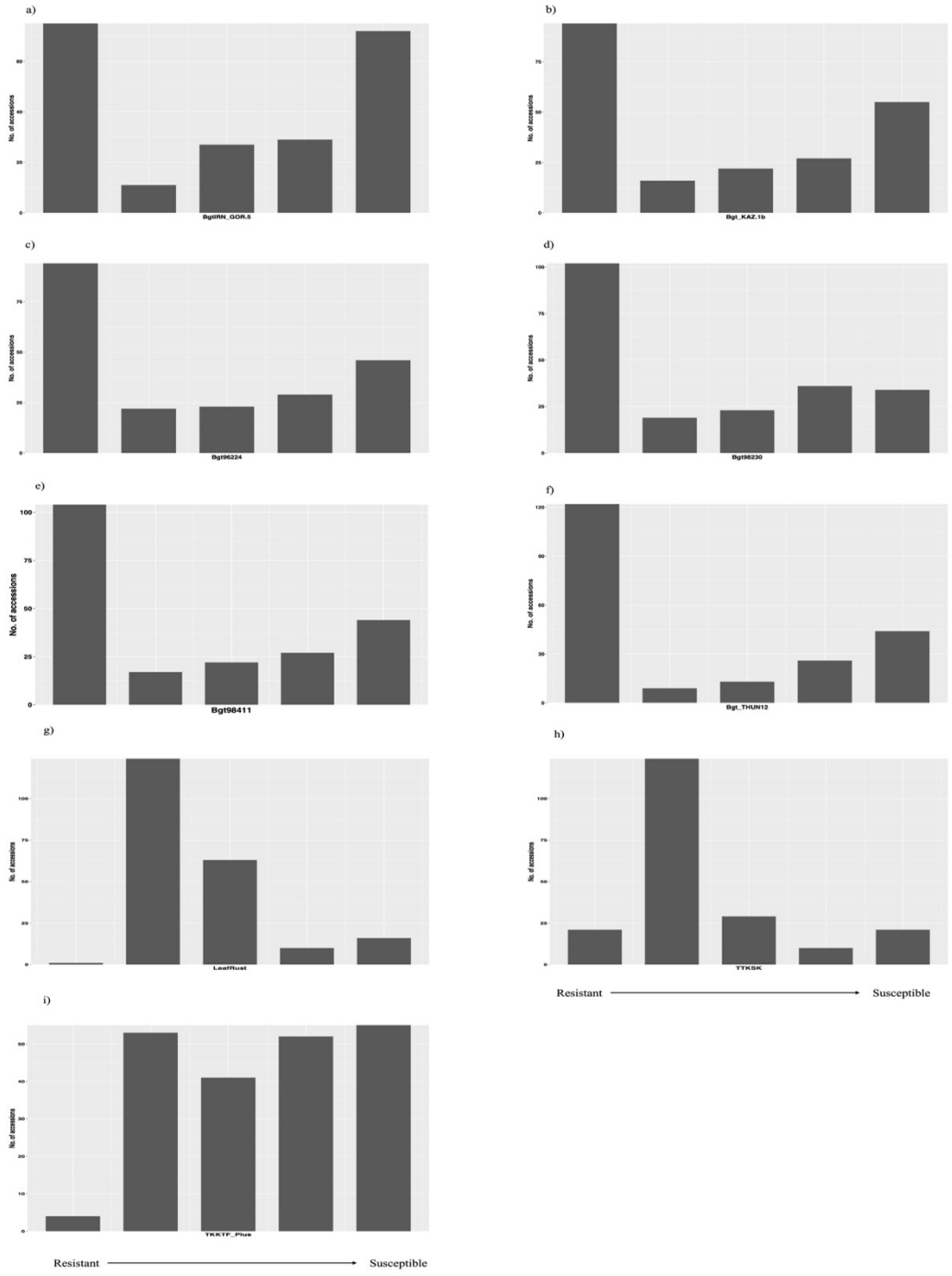


Figure 6.2 Frequency distribution of a) powdery mildew race BgtIRN_GOR-5, b) Bgt_KAZ-1b, c) Bgt96224, d) Bgt98230, e) Bgt98411, f) Bgt_THUN12, g) Leaf rust, e) Stem rust race TTKSK and f) TKKTF+. Data categorized into a five-step scale; resistant, intermediate resistant, intermediate, intermediate susceptible, and susceptible.

Powdery mildew race BgtIRN_GOR-5

For the race BgtIRN_GOR-5, one major QTL was detected on the long arm of chromosome 7A (Figure 6.3). The mapped region spans 747kb (769-770Mb) and harbors 12 high-confidence genes (Table 6.1). The region overlaps with an already cloned stem rust gene *Sr22* from *T. monococcum,* providing resistance to Ug99 (Luo et al., 2022). Notably, this region also coincides with a mapped QTL *PmTb7A.1* from *Triticum boeoticum*, reported to provide broad-spectrum resistance against multiple powdery mildew races (Chhuneja et al., 2015). The genes in the region were compared with the corresponding orthologs of one susceptible accession (TA10868) and three resistant accessions i.e., TA391, TA299, and TA10622. Comparative analysis of gene structures between susceptible and resistant accessions revealed two promising candidate genes within the QTL region. One of the candidate genes belongs to the RGA5 gene category, while the other was found to be the cysteine-rich receptor-like protein kinase. Table 6.1 displays the gene structure information.

The potential candidate gene RGA5 showed two alternate splice forms in resistant accessions TA391 and TA299, while the susceptible einkorn accessions showed the presence of only one isoform of the gene. Since the previously identified *Sr22* gene also resides within this QTL region (Luo et al., 2022), we performed the blast search of *Sr22* on TA391 assembly, which resulted in only a partial match on the genome, confirming the novelty of the current RGA gene.

The protein kinase candidate showed the presence of a premature stop codon in the susceptible accession TA10868, which resulted in the formation of a smaller mRNA of 948bp compared to 1758bp mRNA in resistant accessions. The gene

consisted of two domains: Pkinase domain and PK_Tyr_Ser-Thr (Protein Kinase Tyrosine-Serine-Threonine Domain). The susceptible (TA10868) accession showed the loss of PK_Tyr_Ser-Thr domain (Table 6.1). This domain belongs to a conserved protein family and plays a significant role in protein phosphorylation, an essential process in numerous cellular activities. Therefore, the loss of this critical domain would result in the formation of defective or inactive proteins in the susceptible cultivars/accessions. These results make protein kinase as the most potential candidate for the locus *PmTb7A.1* providing broad-spectrum resistance against multiple *Bgt* races. Interestingly, haplotype analysis revealed a single cluster for both wild einkorn accessions, while the two domesticated accessions formed distinct clusters.

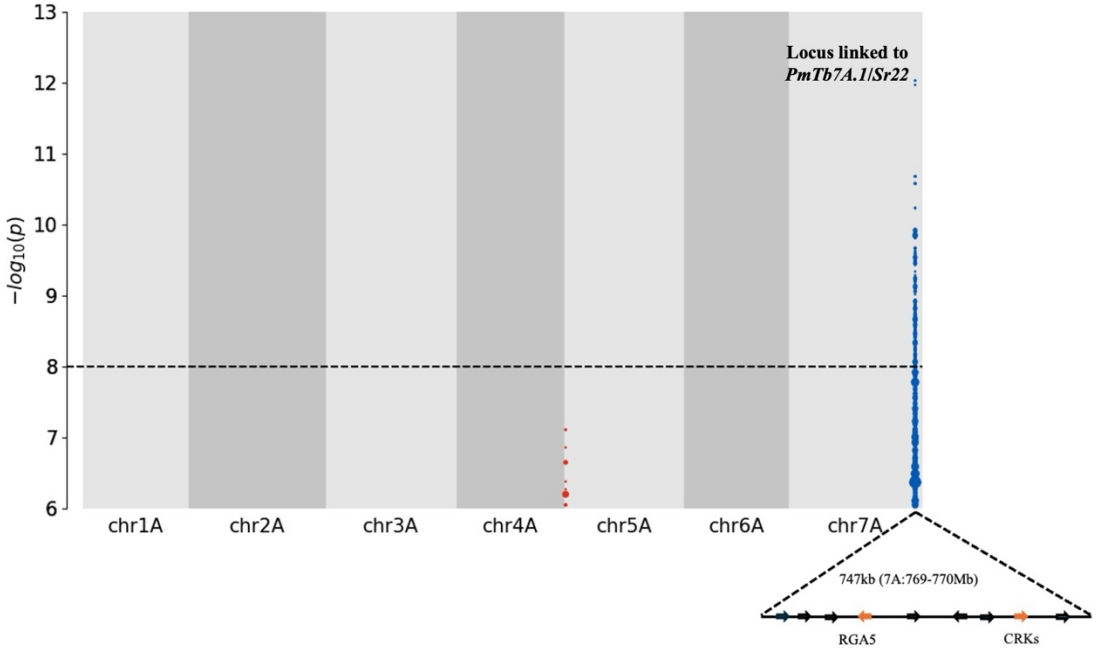


Figure 6.3 *k*-mer–based association mapping of powdery mildew race BgtIRN_GOR-5 on TA391 reference assembly. Genes in the region are shown as pointing arrows depicting the gene orientation. Potential candidates (RGA5 and CRKs) are shown in orange color.

Table 6.1 List of identified potential candidate genes of TA391 and their homologs in TA10868, TA299, and TA10622 genomes for the race BgtIRN_GOR-5. Description of Protein domains as follows: a) Rx_N (plant resistance proteins), b) NB-ARC (nucleotide-binding protein containing apoptotic protease-activating factor 1), c) Pkinase (Protein kinase domain), d) PK_Tyr_Ser-Thr (Protein tyrosine and serine/threonine kinase).

Bgt race	Gene	Accession	Gene_ID	Gene Length (bp)	Coding sequence (bp)	Protein domain
BgtIRN_GOR-5	Disease resistance protein RGA5	TA391	TRIMO_AEGILO_L95.r1.7AG00430040.1	9769	2808	Rx_N, NB-ARC
			TRIMO_AEGILO_L95.r1.7AG00430040.2		2823	Rx_N, NB-ARC
		TA10868	TRIMO_MONO_L96.r1.7AG00579480.1	5771	2823	Rx_N, NB-ARC
		TA299	Tm.TA299.r1.7AG0168310.1	9700	808	Rx_N,NB-ARC
			Tm.TA299.r1.7AG0168310.2		2823	Rx_N,NB-ARC
		TA10622	Tm.TA10622.r1.7AG0162130.1	20806	2817	Rx_N,NB-ARC
	Putative cysteine-rich receptor-like protein kinase	TA391	TRIMO_AEGILO_L95.r1.7AG00430180.2	11169	1758	Pkinase,PK_Tyr_Ser-Thr
		TA10868	TRIMO_MONO_L96.r1.7AG00579510.1	11096	948	Pkinase
		TA299	Tm.TA299.r1.7AG0168430.1	11171	1758	Pkinase,PK_Tyr_Ser-Thr
		TA10622	Tm.TA10622.r1.7AG0162240_1.1	34959	1758	Pkinase,PK_Tyr_Ser-Thr

Powdery mildew race Bgt_KAZ-1b

For the race Bgt_KAZ-1b, two QTLs were identified on chromosome 7A (Figure 6.4). The QTL on the short arm was found to be more prominent with the higher number of mapped k-mers and was also unique compared to the QTL region on the long arm, which was shared with the race Bgt_IRN_GOR-5. A literature search did not reveal any mapped QTLs on the telomeric end of 7AS in wheat, suggesting this discovery represents a novel resistance locus. The genomic region spans 337 kb (12.5Kb-12.8Kb) with a total of six high-confidence genes. These genes consisted of membrane protein, cytochrome p450-related protein, and a single RGA5 type of gene

(Table 6.2). Single copy orthologs were observed for the RGA5 in the four reference genomes. There were no differences in the protein domains as they were found intact (Table 6.2). The gene sequence in accession TA10622 was removed from the analysis as it was found to be fragmented into two distinct genes, with a separation of 3Kb between two fragments. A premature stop codon was observed in the c-terminal region of the RGA in susceptible accession (TA10868) compared to the other resistant accessions, making it a suitable candidate. Haplotype analysis identified distinct haplotypes for RGA5 in the four reference genomes.

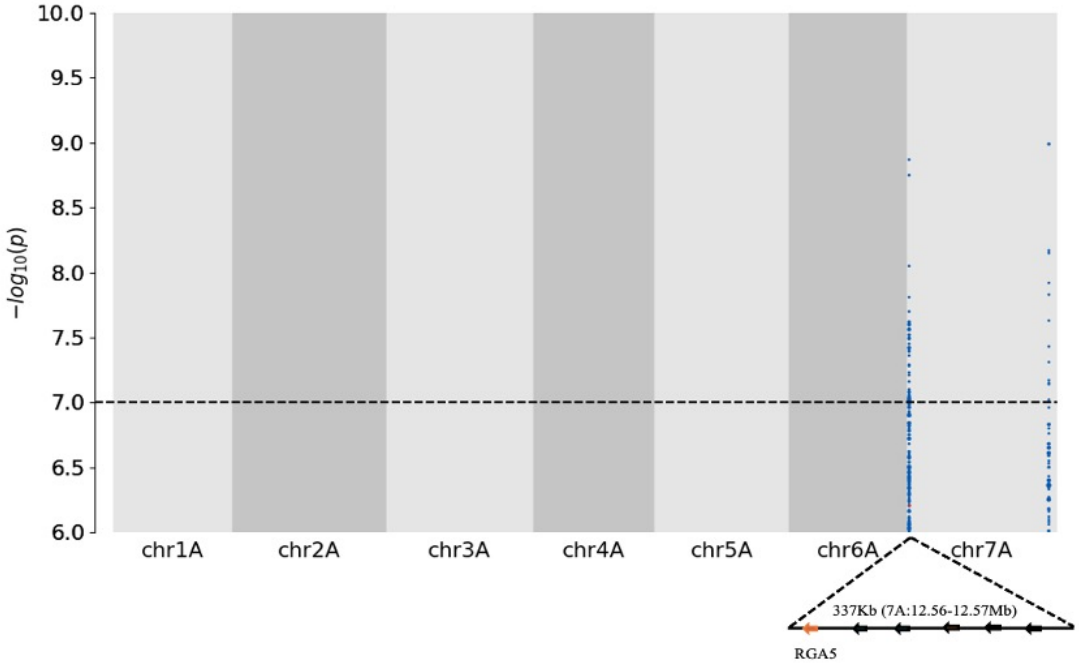


Figure 6.4 *k*-mer–based association mapping of powdery mildew race Bgt_KAZ-1b on TA391 reference assembly. Genes in the region are shown as pointing arrows depicting the gene orientation. Potential candidate (RGA5) is shown in orange color.

Table 6.2 List of identified potential candidate genes of TA391 and their homologs in TA10868, TA299, and TA10622 genomes for the race Bgt_KAZ-1b. Description of Protein domains is as follows: a) Rx_N (plant resistance proteins), b) NB-ARC (nucleotide-binding protein containing apoptotic protease-activating factor 1).

Bgt race	Gene	Accession	Gene_ID	Gene Length (bp)	Coding sequence (bp)	Protein domain
Bgt_KAZ-1b	NB-ARC domain containing protein	TA391	TRIMO_AEGILO_L95.r1.7AG00324600.1	12140	3267	Rx_N, NB-ARC
		TA10868	TRIMO_MONO_L96.r1.7AG0048 4400.1	4623	3102	Rx_N, NB-ARC
		TA299	Tm.TA299.r1.7AG0005710.1	12638	3267	Rx_N,NB-ARC
		TA10622	Tm.TA10622.r1.7AG0006200.1	10309	1533	
			Tm.TA10622.r1.7AG0006180.1	1740	1641	

Powdery mildew race Bgt96224

For the race Bgt96224, multiple QTLs were detected on chromosomes 2A, 3A, 4A, 6A, and 7A (Figure 6.5). QTL present on chromosome 7A was shared with the previous two races. An investigation of QTL on 3AS identified a single gene encoding for transcription factor BHLH3 (basic helix-loop-helix). Single copy orthologs were observed in the four studied reference genomes for BHLH3. Susceptible accession (TA10868) showed smaller protein compared to resistant accessions, making it a suitable candidate for the study. However, no domain loss was observed in the protein domain analysis (Table 6.3). The region also consisted of two cloned genes *Pm44* (Alam et al., 2011) and *Pm13* (H. Li et al., 2024). The gene *Pm13* was cloned recently and encodes for a mixed lineage kinase domain-like (MLKL) protein. Blast search did not show any significant hit for these genes in the reference genome TA391, confirming the novelty of identified QTL. Investigation of genes within other minor QTLs revealed

the presence of a Serine/threonine-protein phosphatase 5 on 2AL, exhibiting a single non-synonymous change in susceptible accession TA10868. Serine/threonine-protein phosphatases are known to play a significant role in oxidative stress signaling (Máthé et al., 2019). Further, a gene cluster consisting of wall-associated receptor kinase 5 (1 gene), wall-associated receptor kinase 2 (1 gene), disease resistance protein Piks-2 (2 genes), and protein LAZY 1 (1 gene) were observed on the long arm of chromosome 6A. These two QTL need further exploration; however, they were not considered in the current study based on the lower p-value cutoff.

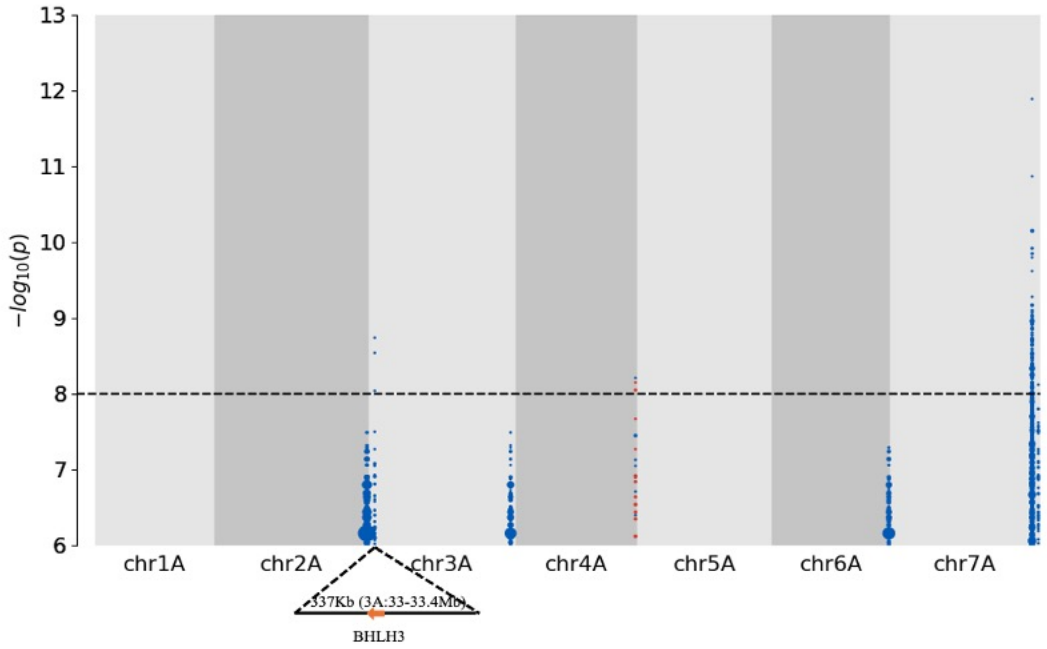


Figure 6.5 *k*-mer–based association mapping of powdery mildew race Bgt96224 on TA391 reference assembly. Genes in the region are shown as pointing arrows depicting the gene orientation. Potential candidate (BHLH3) is shown in orange color.

Table 6.3 List of identified potential candidate gene of TA391 and their homologs in TA10868, TA299, and TA10622 genomes for the race Bgt96224. Description of Protein domain is as follows: a) BHLH (basic Helix Loop Helix (bHLH) domain superfamily).

Bgt race	Gene	Accession	Gene_ID	Gene Length (bp)	Coding sequence (bp)	Protein domain
Bgt96224	Transcription factor BHLH3	TA391	TRIMO_AEGILO_L95.r1.3AG00014280.1	1203	999	bHLH
		TA10868	TRIMO_MONO_L96.r1.3AG00598000.1	1217	777	bHLH
		TA299	Tm.TA299.r1.3AG0015900.1	1203	999	bHLH
		TA10622	Tm.TA10622.r1.3AG0014450.1	1203	999	bHLH

Powdery mildew race Bgt98230

For the race Bgt98230, a major and novel QTL was detected on chromosome 6A (Figure 6.6). Few genes have been reported on the long arm of the group 6 chromosome in wheat, which includes two of the genes *Pm54* (Hao et al., 2015) and *Pm69* (W. Xie et al., 2012) on 6B. The QTL on chromosome 6AL spanned 337kb (585.6Mb-585.9Mb) and showed 8 high confidence genes, most of which were unannotated, except one gene encoding the "Nuclear transcription factor Y subunit A-5 (Arabidopsis thaliana)" as a potential candidate in the region (Table 6.4). One non-synonymous SNP was observed between the domesticated and wild accessions. TA10622 was found to be moderately resistant for this race. Haplotype analysis revealed that two resistant accessions TA299 and TA391 clustered together, while accessions TA10622 and TA10868 belonged to distinct haplotypes. The candidate gene was primarily considered based on the differences between TA391 and TA10868. Further exploration of the promotor region is needed for the candidate gene. There were no differences in the protein domains (Table 6.4).

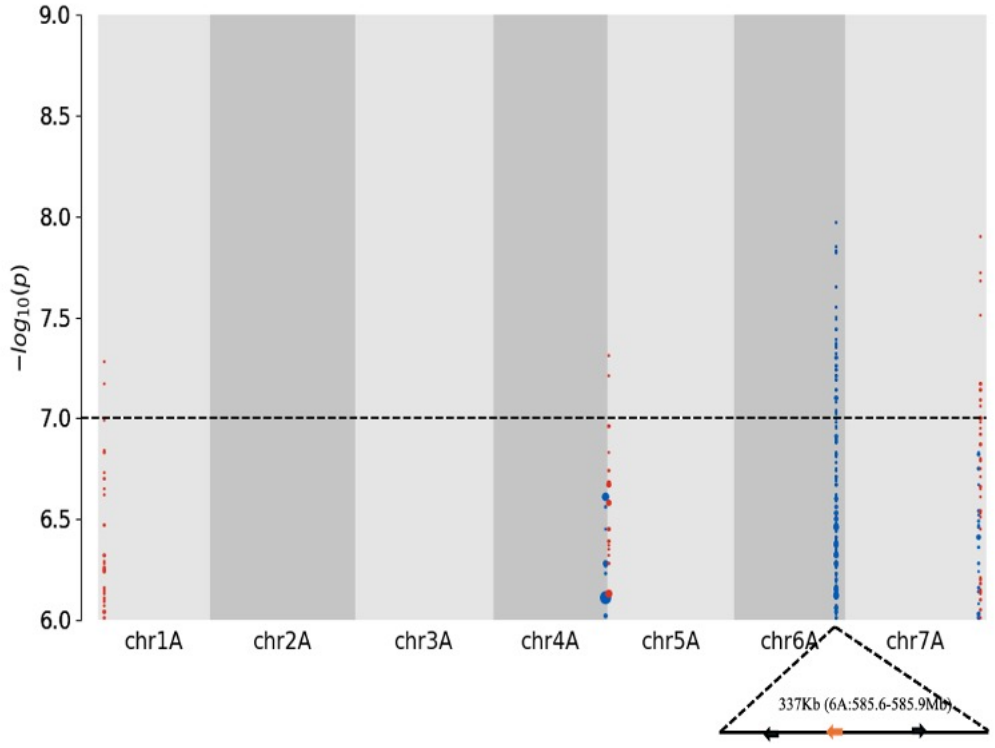


Figure 6.6 *k*-mer–based association mapping of powdery mildew race Bgt98230 on TA391 reference assembly. Genes in the region are shown as pointing arrows depicting the gene orientation. Potential candidate (Nuclear transcription factor Y subunit A-5) is shown in orange color.

Table 6.4 List of identified potential candidate gene of TA391 and their homologs in TA10868, TA299, and TA10622 genomes for the race Bgt98230. Description of Protein domain is as follows a) CBFB_NFYA (CCAAT-binding transcription factor (CBFB/NFYA) subunit B).

Bgt race	**Gene**	**Accession**	**Gene_ID**	**Gene Length (bp)**	**Coding sequence (bp)**	**Protein domain**
Bgt98230	Similar to NFYA5: Nuclear transcription factor Y subunit A-5	TA391	Tm.L95_00020952.1	3837	765	CBFB_NFYA
		TA10868	Tm.L96_00021164.1	3829	765	CBFB_NFYA
		TA299	Tm.TA299.r1.6AG0123920.1	3817	765	CBFB_NFYA
		TA10622	Tm.TA10622.r1.6AG0118950.1	3829	765	CBFB_NFYA

Powdery mildew race Bgt98411

For the race Bgt98411, two QTLs were detected, one on chromosome 5A and the other on chromosome 6A (Figure 6.7). In the absence of other mapped QTLs in the region, these can be considered as novel QTLs. Both the QTLs span a broader region of 337Kb. The QTL on 5AL spans from 571.7 Mb to 572.1Mb and encompasses five genes, including a cluster of three "G-type lectin S-receptor-like serine/threonine-protein kinase" (Table 6.5). One of the G-type lectins showed no difference in gene sequence in TA391 and TA10868. In contrast, the other two displayed numerous differences, rendering them as potential race-specific candidates for the powdery mildew resistance. Further analysis revealed, the presence of a shorter mRNA for the candidate gene 1 (TRIMO_MONO_L96.r1.5AG00363210.1) in susceptible (TA10868) accession compared to other resistant accessions, attributed to the presence of premature stop codon. The second gene only exhibited a single nonsynonymous change (G->S) between resistant and susceptible accessions. Haplotype analysis further delineated different haplotype groups for each of the four accessions. Notably, the presence of premature stop codon in candidate 1 elevates its potential for gene validation studies. QTL on chromosome 6A consists of six high-confidence genes. No significant differences were observed among the orthologs in the resistant and susceptible accessions.

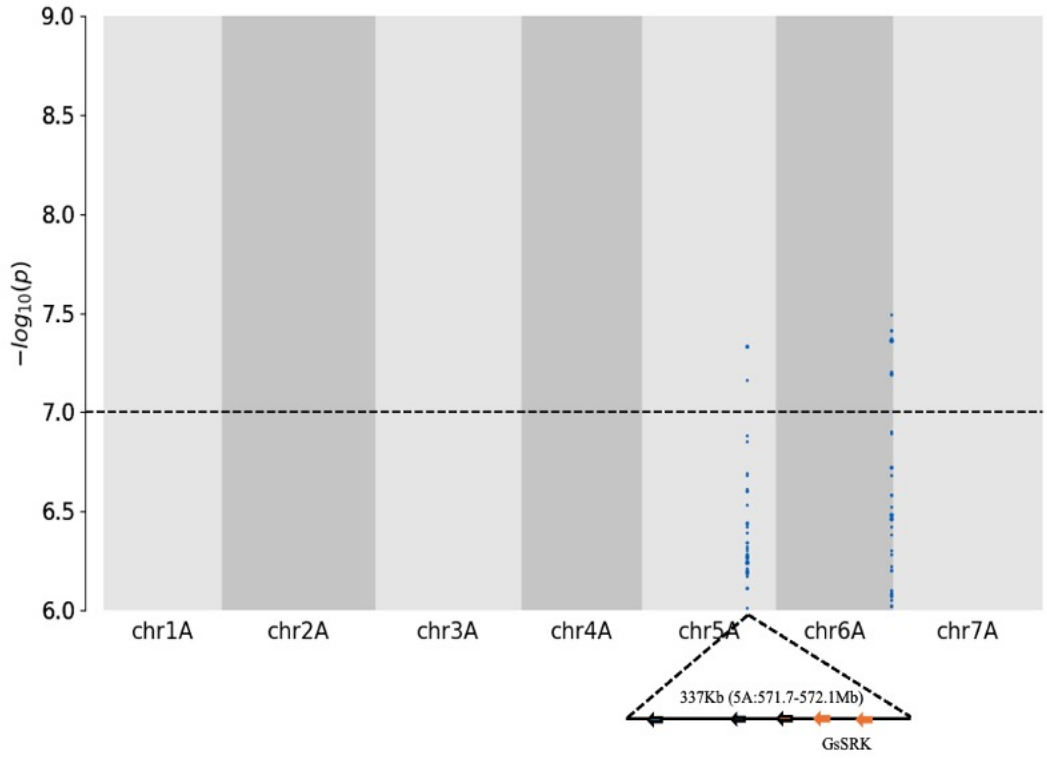


Figure 6.7 *k*-mer–based association mapping of powdery mildew race Bgt98411 on TA391 reference assembly. Genes in the region are shown as pointing arrows depicting the gene orientation. Two potential candidates (GsSRK) are shown in orange color.

Leaf rust

The association panel was evaluated for its resistance against leaf rust. Figure 6.2 depicts the distribution of infection scores for *Puccinia triticina*. To comprehensively identify leaf rust resistance genes, we employed a three-pronged approach. This included: 1) K-mer-based GWAS using TA391 reference, 2) K-mer-based GWAS using TA299 reference, and 3) analysis of a recombinant inbred line (RIL) population derived from resistant (TA391) and susceptible (TA10868) einkorn accessions. This multifaceted approach broadened the scope for uncovering potential resistance genes. The RIL population consisted of 812 lines derived from the cross of

T. monococcum ssp. *aegilopoides* (TA391) and *T. monococcum* ssp. *monococcum* (TA10868*)* accessions originally described by Singh et al (K. Singh, Ghai, et al., 2007).

Table 6.5 List of identified potential candidate gene of TA391 and their homologs in TA10868, TA299, and TA10622 genomes for the race Bgt98411. Description of protein domain is as follows: a) B_lectin (D-mannose binding lectin), b) S_locus_glycop (S-locus glycoprotein domain), c) PK_Tyr_Ser-Thr (Protein tyrosine and serine/threonine kinase).

Bgt race	**Gene**	**Accession**	**Gene_ID**	**Gene Length (bp)**	**Coding sequence (bp)**	**Protein domain**
Bgt98411	Putative G-type lectin S-receptor-like serine/threonine-protein kinase	TA391	TRIMO_AEGILO_L95.r1.5AG0060 1590.1	4078	2544	B_lectin,S_locus_glycop,PAN_2,PK_Tyr_Ser-Thr
			TRIMO_AEGILO_L95.r1.5AG0060 1590.2	4078	1581	B_lectin,S_locus_glycop,PAN_2
		TA10868	TRIMO_MONO_L96.r1.5AG00363 210.1	7129	2394	B_lectin,S_locus_glycop,PAN_2,PK_Tyr_Ser-Thr
		TA299	Tm.TA299.r1.5AG 0143770.1	3832	2544	B_lectin,S_locus_glycop,PAN_2,PK_Tyr_Ser-Thr
		TA10622	Tm.TA10622.r1.5 AG0138330.1	3832	2544	B_lectin,S_locus_glycop,PAN_2,PK_Tyr_Ser-Thr
	G-type lectin S-receptor-like serine/threonine-protein kinase	TA391	TRIMO_AEGILO_L95.r1.5AG0060 1600.1	3765	2547	B_lectin,S_locus_glycop,PAN_2,PK_Tyr_Ser-Thr
		TA10868	TRIMO_MONO_L96.r1.5AG00363 240.1	3756	2547	B_lectin,S_locus_glycop,PAN_2,PK_Tyr_Ser-Thr
		TA299	Tm.TA299.r1.5AG 0143820.1	3756	2547	B_lectin,S_locus_glycop,PAN_2,PK_Tyr_Ser-Thr
		TA10622	Tm.TA10622.r1.5 AG0138380.1	3756	2547	B_lectin,S_locus_glycop,PAN_2,PK_Tyr_Ser-Thr

Using the Kmer-GWAS with reference TA391, a significant QTL was identified on 5AL, spanning 7.23Mb (560-568Mb) (Figure 6.8). Within this region, 38 high-confidence genes were identified, predominantly comprised of unannotated or hypothetical proteins. Among the annotated genes were two leucine-rich repeat

receptor-like protein kinases, one ankyrin repeat-containing protein, and three cysteine-rich receptor-like protein kinases (Table 6.6). Notably, the cysteine-rich receptor-like protein kinase did not exhibit any differences between resistant and susceptible accessions. No orthologs were observed for the ankyrin repeat-containing protein in TA299 and TA10622, rendering these unsuitable for further investigation. However, three non-synonymous mutations were detected in TA10868 and TA391. Subsequently, only two leucine-rich repeat receptor-like protein kinase genes remained, with only one gene showing orthologs in all four reference genomes. Among these, the potential candidate gene (TRIMO_MONO_L96.r1.5AG00361130.1) demonstrated three isoforms in TA10868, two in TA391, and only one in TA299 and TA10622 (Table 6.6). As all four accessions displayed resistance, and TA391 demonstrated moderate resistance, the gene was considered a potential candidate. It is anticipated that the gene will exhibit proportional expression of isoforms across the four accessions.

GWAS analysis using GAPIT on the RIL population revealed the identification of a major marker-trait association (MTA) on chromosome 2A, consistently detected across various methods, including blink, MLM, GLM, and MLMM (Figure 6.8). The MTA exhibited peaks around 12Mb, with a span of approximately 4.3 kb among different methods. Subsequently, the genomic region was extended to 163 kb on both sides of the peak based on the LD and leading to the discovery of two high-confidence genes within the interval: one encoding an NB-ARC domain-containing protein and the other an exocyst subunit exo family protein (Table 6.6). Notably, the NB-ARC gene showed two non-synonymous changes in accession TA391 and not in other accessions.

For the exocyst subunit exo family protein, no orthologs were observed in TA299 and TA10622, but the differences in the gene length among TA391 and TA10868 render it an intriguing candidate for leaf rust resistance. The function of this gene in endomembrane trafficking suggests a potential link to disease resistance, warranting further investigation.

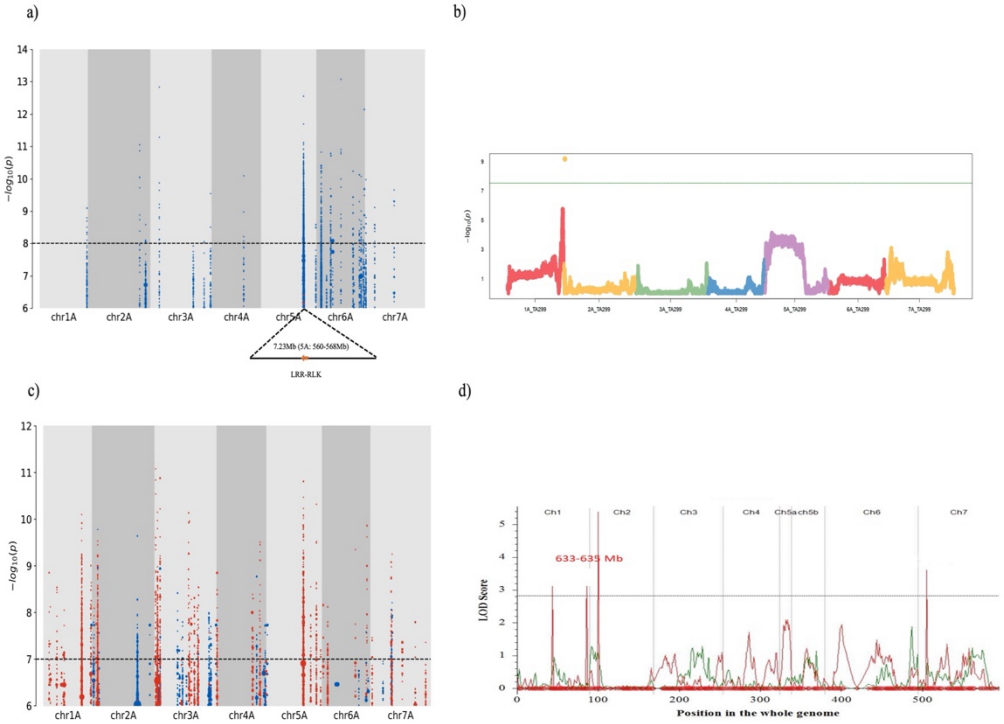


Figure 6.8 a) *k*-mer–based association mapping for leaf rust using TA10868 reference assembly b) SNP-GWAS based on TA391 and TA10868 RIL population using model BLINK, c) *k*-mer–based association mapping for leaf rust using TA299 reference assembly, and d) Interval mapping of leaf rust using TA299 as reference.

K-mer GWAS and Interval mapping performed using the RIL population with TA299 as a reference (Figure 6.8) led to the identification of a shared genomic region on the long arm of chromosome 1A (633Mb). The region consists of mapped gene (*Lr2K38*) associated with leaf rust resistance in the known QLr.ags-1AL. The region is

delimited by the flanking markers IWB20487 and IWA4022, as outlined in Sapkota et al. (2020), and is situated approximately at 626.6Mb on TA299 (Sapkota et al., 2020). The region identified through K-mer GWAS spanned across three segments: 625.1-625.4Mb (containing 9 genes), 629.6-630Mb (with 2 genes), and 633.4-633.7Mb (encompassing 4 genes). Interval mapping results for leaf rust resistance on TA299 aligned with the interval of 633-635Mb, suggesting the third region as the most promising candidate region for locus QLr.ags-1AL. This region notably comprises only one potential candidate gene, RGA5 (Tm.TA299.r1.1AG0155200), potentially linked with *Lr2K38*. Orthologs of the RGA5 were present in TA299, TA10622, and TA10868, but absent in the TA391.

Table 6.6 List of identified potential candidate genes for leaf rust considering resistant accession TA10868 as reference and their homologs in TA391, TA299, and TA10622 genomes.

Gene	**Accessions**	**Gene_ID**	**Gene Length (bp)**	**Coding sequence (bp)**
Leucine-rich repeat receptor-like protein kinase	TA10868	TRIMO_MONO_L96.r1.5AG00361130.1	5042	3336
		TRIMO_MONO_L96.r1.5AG00361130.2		3336
		TRIMO_MONO_L96.r1.5AG00361130.3		3531
	TA391	TRIMO_AEGILO_L95.r1.5AG00599820.1	5106	3423
		TRIMO_AEGILO_L95.r1.5AG00599820.2		3423
	TA10622	Tm.TA10622.r1.5AG0133780.1	4910	3336
	TA299	Tm.TA299.r1.5AG0138990.1	3440	3336
NBARC domain-containing	TA391	TRIMO_AEGILO_L95.r1.2AG00115320.1	4383	2538
		TRIMO_AEGILO_L95.r1.2AG00115320.2		3360
	TA10868	TRIMO_MONO_L96.r1.2AG00091920.1	4496	2538
		TRIMO_MONO_L96.r1.2AG00091920.2		3360
	TA299	Tm.TA299.r1.2AG0007270.1	4407	3360
	TA10622	Tm.TA10622.r1.2AG0006960.1	4446	3360
exocyst subunit exo family	TA391	TRIMO_AEGILO_L95.r1.2AG00115350.1	3563	1596
	TA10868	TRIMO_MONO_L96.r1.2AG00091950.1	3669	1602

RGA5	TA10868	Tm.TA10622.r1.1AG0152360.1	5806	2691
	TA299	Tm.TA299.r1.1AG0155200.1	5713	2691
	TA10622	TRIMO_MONO_L96.r1.1AG00083480.1	5763	2691

Stem rust

K-mer GWAS analyses were conducted for two of the stem races i.e., TTKSK (Ug99) and TKKTF+. Figure 6.2 shows the distribution of the phenotype score for these races. Although multiple quantitative trait loci (QTLs) were identified, the two most significant associations were observed on the long arms of chromosomes 2A and 7A (Figure 6.9). Notably, the region on 2AL coincided with the cloned gene *Sr21*, mapped to the gene (TRIMO_MONO_L96.r1.2AG00185870.1) at 750.72 Mb. Interestingly, the QTL region lies outside the mapped *Sr21* gene, mapped at 750.72Mb on reference TA10868. For the race TTKSK, the region on 2A spans around 1.16Mb (750.3-751.5Mb), while for the race TKKTF+ it spans 1.10Mb (751Mb-752.1Mb), with an overlapping region of 557 Kb. As *Sr21* mapped outside of the region on TKKTF+, the 751Mb region was examined for potentially resistance genes. The region encompasses 17 high-confidence genes in TKKTF+ and 15 high-confidence genes in TTKSK, including a cluster of resistance gene analogs (RGAs), transcription factors (F-box and B3), and BED-type domain-containing proteins. This enrichment of potential defense-related genes suggests a hotspot region for stem rust resistance. Most of these genes were exclusive to TA10868 and TA391, with no orthologs found in the other two reference genomes, complicating gene identification.

For TKKTF+, three distinct regions were identified on chromosome 7A, spanning from 234.3-243.2 Mb, 537.6-538.9 Mb, and 551.6-552.5 Mb (Figure 6.9). These regions encompass numerous genes associated with various functional classes,

including probable L-type lectin-domain containing receptor kinase, pentatricopeptide repeats, receptor-like protein kinase HERK 1, auxin response factor, and multiple transcription factors. In contrast, for TTKSK, significant regions were identified at positions 676.5-676.9 Mb and 792.7-793.1Mb and consisted of multiple genes belonging to diverse classes such as f-box, pentatricopeptide repeats, and transcription factors (B3, bHLH25). This complexity in gene composition complicates the identification process and needs further exploration. The *Sr21* was also conformed through interval mapping based on RIL population.

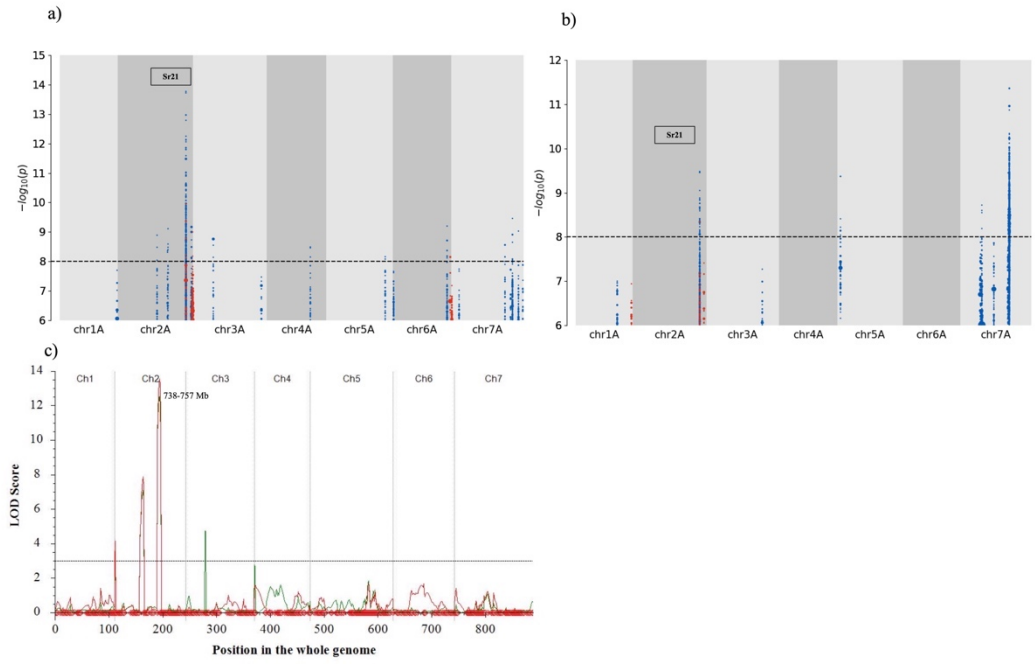


Figure 6.9 a) k-mer–based association mapping for stem rust race a) TTKSK and b) TKKTF+ and c) interval mapping based on RIL population of TA391 and TA10868 RIL population.

6.5 Discussion

Genome-wide association studies (GWAS) play a pivotal role in modern plant breeding by facilitating the discovery of novel genes and alleles from a wide panel of

germplasm. The natural variation serves as the foundation for GWAS to detect associations between genetic markers and the target trait. Utilizing the novel genetic variations from rich germplasm, GWAS can identify associations of genetic markers for disease resistance. GWAS can pinpoint genomics regions harboring important candidate genes underlying the studied trait and can provide insight into the genetic basis of the complex trait. The present study leverages an association panel of 220 diverse *T. monococcum* accessions to investigate resistance against the most devastating diseases of wheat i.e., powdery mildew, leaf rust, and stem rust.

K-mer-based GWAS revealed quantitative trait loci associated with resistance to powdery mildew (PW) on almost all chromosomes (1A-7A) for five races. Notably, a QTL on chromosome 7AL exhibited consistent association across four PW races. Furthermore, 7AL QTL appeared to be the most influential resistance locus for the races BgtIRN_GOR-5 and Bgt96224 races. Additionally, race-specific QTLs were identified for Bgt98230 (6AL) and Bgt98411 (5AL). These QTL regions collectively harbor a total of 98 high-confidence genes potentially involved in disease resistance i.e., Bgt_KAZ-1b (17 genes), BgtIRN_GOR-5 (22 genes), Bgt98411 (10 genes), Bgt98230 (32 genes), and Bgt96224 (53 genes).

Our investigation identified three distinct categories of genes associated with resistance to powdery mildew (PW) for the studied races. Six genes located on the long arm of chromosome 7A (7AL) exhibited consistent presence across four accessions, suggesting its potential for broad-spectrum resistance against multiple PW races. An additional set of six genes, also located on 7AL, was shared by three accessions. Conversely, the distribution of genes on chromosomes 2A, 3A, 4A, and 6A displayed

race specificity, potentially reflecting co-evolutionary dynamics between the host and pathogen. Notably, genes identified on chromosome 5A were shared between two races. This characterization of resistance gene distribution provides valuable insights for selecting candidate genes that might confer broad-spectrum resistance against PW. Traditionally, resistance (R) genes often mediate race-specific resistance, while transcription factors and kinases are more likely to be associated with broad-spectrum resistance (Jones Jonathan D.G. & Dangl Jeffery L., 2006; Macho & Zipfel, 2014; S. P. Pandey & Somssich, 2009).

The investigation identified a set of promising candidate genes associated with resistance to specific powdery mildew (PW) races. For race BgtIRN_GOR-5, the candidate includes disease-resistance protein RGA5 and a putative cysteine-rich receptor-like protein kinase. We reported a novel RGA5 gene associated with race Bgt_KAZ-1b and a transcription factor BHLH3 for the race Bgt96224. For race Bgt98230, a Nuclear transcription factor Y subunit A-5 coding gene, potentially involved in regulating stress responses, and for race Bgt98411, gene encoding G-type lectin S-receptor-like serine/threonine-protein kinase involved in pathogen recognition and signaling were identified.

Analysis of the 16 cloned genes associated with powdery mildew resistance reveals the predominant presence of nucleotide-binding site and leucine-rich repeat (NLR) immune receptor encoding genes, constituting 12 out of the 16 identified genes. Notable examples include *Pm3*, *Pm8*, *Pm2*, *Pm21*, *Pm60*, *Pm17*, *Pm4*1, *Pm5e*, *Pm1*a, *MlIW172*/*MlWE18*, *Pm12*, and *Pm69*, recognized for their race-specific resistance (P. Gupta et al., 2022). The remaining four cloned genes belongs to diverse functional

classes: *Pm38* encodes an ATP-binding cassette (ABC) transporter, *Pm46* encodes a hexose transporter, *Pm24* encodes a tandem kinase protein (WTK3 and WTK4), and *Pm4* encodes a chimeric protein featuring a serine/threonine kinase domain, multiple C2 domains, and transmembrane regions. In addition to NLR immune receptors, alternative gene classes also contributed to powdery mildew resistance. Notably, *MLO* and *EDR1* exemplify this diversity, offering various levels of recessively inherited powdery mildew (Sánchez-Martín & Keller, 2021; Y. Wang et al., 2014; Y. Zhang et al., 2017). *MLO* encodes a membrane-localized protein, whereas *EDR1* encodes a cytoplasmic Raf-like mitogen-activated protein kinase (MAPKKK).

Several cystine kinases have been reported for fungal diseases i.e., *Pm13* (MLKL) (H. Li et al., 2024), *HvCRK1* a negative (K. Yang et al., 2013) regulator (*Hordeum vulgare* Cysteine-rich Receptor-like Kinase 1) (Rayapuram et al., 2012), *TaCRK1* (*Triticum aestivum* Cysteine-rich Receptor-like Kinase 1) identified in wheat, provides resistance against *Rhizoctonia cerealis* (K. Yang et al., 2013). A novel cysteine-rich receptor-like kinase gene, *TaCRK2*, contributes to leaf rust resistance in wheat (Gu et al., 2020). Upon perception of signals, Cystine-Rich Receptor Kinases (CRKs) initiate a complex signaling cascade within the plant cell. This cascade often involves phosphorylation events, leading to the activation of various defense mechanisms (Chern et al., 2016). Studies have shown that CRKs are involved in diverse defense responses, including the production of reactive oxygen species, activation of defense genes, and even programmed cell death in response to pathogens (Kamel et al., 2023). G-type lectin S-receptor-like serine/threonine-protein kinases (*GsSTKs*) bind specific carbohydrates of the invading pathogen and initiate a signal cascade that

triggers phosphorylation of other proteins within plant cells (Afzal et al., 2008; Y. Zhang et al., 2022). *GsSTKs* activate the production of reactive oxygen species, expression of defense genes, and strengthening of cell walls to restrict pathogen access (Wan et al., 2021)**.** *GsSRK* (Serine/Threonine Receptor-like Kinase) acts as a positive regulator for salt stress tolerance in wild soyabean (X. L. Sun et al., 2013). *NbERK1* (*Nicotiana benthamiana* ERK1) is a G-type lectin receptor-like kinase, positively regulates chitin signaling and contributes to resistance against the fungal pathogen *Sclerotinia sclerotiorum* (Pi et al., 2023)**.** BHLH transcription factors (Basic Helix-Loop-Helix) is a diverse family known to regulate various plant developmental processes (Guo et al., 2021). These factors can act as positive regulators, activating genes involved in defense responses like the production of secondary metabolites or the hypersensitive response upon pathogen attack (Falak et al., 2021; Song et al., 2013). Ma et. al., (2023) showed enhanced resistance for powdery mildew by overexpressing *MdbHLH093* (*Malus domestica* bHLH093) in *Arabidopsis thaliana* (Ma et al., 2023). Nuclear transcription factor Y complexes are involved in gametogenesis, embryogenesis, seed development, flowering time regulation, primary root elongation, abscisic acid signaling, and drought resistance (Zhao et al., 2017; Zhang et al., 2023).

Investigation of resistance gene for leaf rust, utilizing three distinct approaches, identified three candidate genes: two encoding resistance genes (RGAs) and one encoding an exocyst subunit protein belonging to the exo family. Exocyst proteins are crucial for vesicle trafficking in plants. They facilitate the fusion of secretory vesicles with the plasma membrane, allowing the delivery of defense compounds like enzymes, antimicrobial peptides, and signaling molecules to the cell wall and external

environment. In rice, effector ZiF from *Maganaporthe oryzae* has been shown to bind to the hydrophobic pocket of Exo70 protein and modulate the immune response (De La Concepcion et al., 2022). Du Y. et. al. (2018) reported exocyst subunits in *Solanaceae* to be involved in callose deposition, suggesting that they play a role in basal plant defense (Y. Du et al., 2018). Stegmann et. al. (2014) showed *Exo70B1* subunit of the exocyst complex plays important role in defense response against *Pst* (Stegmann et al., 2014). In a recent study, a genetic module consisted of *Pur1* and *Exo70FX12* was shown to provide resistance against *Puccinia striiformis* f. sp. *tritici* in barley. *Pur1* was orthologous to rice *Xa21* gene and consisted of leucine-rich repeat receptor kinase, whereas *Exo70FX12* belongs to the Poales-specific *Exo70FX* clade (Holden et al., 2022). Our findings for leaf rust resistance align with this concept with the presence of a module consisting of NB-ARC (nucleotide-binding site and leucine-rich repeat) domain and an exocyst subunit exo family protein in the resistant lines TA391 and TA10868. Interestingly, only the NB-ARC gene cluster was observed in TA299 and TA10622 and not exocyst subunit. This contrasting presence of the exocyst subunit gene in resistant lines suggests a potential role for this gene in mediating resistance within the RIL population.

These novel set of genes if implemented in the breeding program would serve as a basis for the novel disease resistance for powdery mildew and leaf rust. Functional validation can be achieved through Virus-Induced Gene Silencing (VIGS) and expression analysis in resistant and susceptible einkorn accessions. Subsequently, cDNA clones of validated candidates can be transformed into susceptible wheat

cultivars (e.g., Fielder) using Agrobacterium-mediated transformation or other suitable methods.

6.6 Conclusion

The present study employed the K-mer GWAS to identify the novel sources of resistance against powdery mildew, leaf rust, and stem rust from a diverse collection of wild (alpha, beta, and gamma races) and domesticated *T. monococcum* accessions. The TmGWAS panel exhibited remarkable genetic diversity for disease resistance with systematic phenotyping. Our analysis pinpointed QTLs and most potential candidates within the LD region of 163 Kb in *T. monococcum* associated with powdery mildew and leaf rust resistance. K-mer GWAS proved to be effective in the detection of novel sources of resistance, resulting in seven QTLs for powdery mildew, three for leaf rust and two for stem rust. These novel sets of genes hold promise for developing wheat cultivars with robust resistance to powdery mildew and leaf rust upon functional validation and transformation of susceptible wheat cultivars.

6.7 Supplementary Material

The supplementary information is provided in the Proquest.

Chapter 7 Conclusions and Future Directions

Wheat is an important crop for the global human population. However, due to domestication bottlenecks and selective breeding, wheat has a narrow genetic base (Venske et al., 2019). Elite wheat cultivars display reduced genetic diversity, rendering them vulnerable to increasing abiotic and biotic stresses. Wheat progenitors, wild, and related species possess a wide genetic diversity, essential for replenishing the allelic loss in bread wheat (Peleg et al., 2011). This dissertation represents a systematic research work on the curation and characterization of useful germplasm from three different species, including wild and related tetraploid wheat species *T. turgidum* and *T. timopheevii* and man-made cereal triticale. Further, the work explores chromosome-specific assembles of wild wheat species *Ae. geniculata*, and *Ae. umbellulata* to compare and characterize gene-based variations to validate the utility of wild and related germplasm for crop improvement. The study also identified novel genes conferring resistance to powdery mildew, leaf rust, and stem rust from the *T. monococcum* GWAS panel.

Chapter 2 describes the genome-assisted evaluation of the genetic diversity in the wild and related tetraploid wheat species *T. turgidum* and *T. timopheevii*. Work was conducted on 908 accessions of *T. turgidum* and *T. timopheevii* from the WGRC gene bank, Kansas State University. GBS-based genotyping was performed on these 908 accessions, and analysis revealed significant levels of duplication within germplasm, accounting for 65 and 47 percent duplication in *T. timopheevii* and *T. turgidum*, respectively. The principal component analysis identified two distinct clusters, demonstrating the species-level differentiation and isolation of two lineages.

Population structure analysis revealed the best available population sub-structure at K=14, where 11 and 3 species-specific subclusters were observed for *T. turgidum* and *T. timopheevii,* respectively. These clusters correspond to the regional isolation and sub-species differentiation in tetraploid wheat. Additionally, both species demonstrated a higher genetic differentiation of 0.27 based on nucleotide diversity (Nei's Π). Based on the clustering pattern, hotspots of genetic diversity were identified for wild emmer in Israel, Jordan, Syria, and Lebanon; for *T. durum* in Turkey and Georgia; and for *T. timopheevii* in Iraq, Azerbaijan, and Armenia. Core set identification accomplished a 10x reduction in size while capturing a high proportion (>93%) of the available genetic diversity. A novel QTL identified at the telomeric end of 2BS consists of four genes encoding two "NBS-LRR-like resistance proteins" and two "Exocyst complex component 2C" proteins. With the reduced complexity, the core set can be used to characterize tetraploid germplasm while maintaining its genetic variability. Furthermore, the potential candidate genes identified in the MTA can be validated using TILLING (Targeting Induced Local Lesions in Genomes), Virus-Induced Gene Silencing (VIGS), and gene complementation approaches.

Chapter 3 describes the differences in the gene structures observed between two members of the tertiary gene pool (*Ae. geniculata* and *Ae. umbellulata*) and domesticated wheat, using the publicly available data from NCBI. *Ae. umbellulata* is the donor for leaf rust resistance (*Lr57)* gene introduced into bread wheat. Both *Ae. geniculata* and *Ae. umbellulata* shared a higher percentage of genes with cultivated wheat and possess species-specific genes, including transcription factors and pathogenicity-related genes. These offer novel avenues for trait enhancement in wheat.

The expansion of the gene families in the tertiary gene pool indicates the evolutionary pressure on wild germplasm. Phylogenetic analysis revealed a closer relationship between the M and U genomes and the D genome of wheat, followed by the A genome. The members of the tertiary gene pool exhibited fewer introns per mRNA and a small average intron length, indicating a smaller overall gene structure compared to the Chinese spring. The potential explanation for this may be the occurrence of insertion elements in the wheat genes compared to other members of the tertiary gene pool. There are limitations of reads collapsing in the repeat region when assembling highly repetitive genomes like *Ae. geniculata* using short-read sequencing. Further validation can be performed by sequencing and assembling the whole genome using PacBio long-read sequencing.

Chapter 4 addressed the problem of determining contamination admixture and translocations in triticale germplasm. For the study, a pseudo-reference was generated by combining *T. aestivum* and *S. cereale* genomes. The pipeline has successfully identified and classified the triticale germplasm, revealing 236 hexaploid and 12 octoploid accessions. It also detected contamination by eleven wheat accessions within the triticale germplasm collection. Chromosomal read alignment plots revealed four whole chromosomal translocations, one deletion, and 15 small translocations in the triticale germplasm. The variability in the germplasm was depicted by a higher population sub-structure of K=11. Additionally, lower genetic differentiation was observed in the B-genome compared to the A- and R-genome. Winter and spring triticale showed lower genetic diversity based on the Fixation index (F_{ST}), indicating less differentiation. To help in the phenotyping, a core set of 89 triticale accessions was

created that retains 97% of the allelic diversity. In the study, GBS data proved valuable for characterizing germplasm, identifying hexaploid and octoploid triticale, detecting misclassified accessions, and wheat contamination. The small core set of the triticale accession allows for efficient phenotyping of agronomic traits, both in the greenhouse and in the field.

Chapter 5 focuses on whole genome assembly and comparative genomics of the wild (TA391) and domesticated (TA10868) accessions of *T. monococcum*. The high-quality assemblies achieved completeness, demonstrating ~97% of complete BUSCO genes. Both genomes harbor a higher percentage (83%) of repetitive elements. Einkorn genomes exhibit a higher degree of synteny with the A-genome of *T. urartu* and *T. aestivum.* The assembled genome harbors a unique set of R genes and transcription factors. The reduced presence of REM family members indicated the lack of vernalization treatment of einkorn wheat. The evolution of gene families suggested a dominant trend of gene loss during divergence, with 645 rapidly evolving gene families. The loss/gain of gene families may have contributed towards the better adaptability of the species. By uncovering genes and their evolutionary patterns, this research provides valuable insights into the einkorn genome. This knowledge can be used to improve the resilience of other wheat varieties. Further, the assembled genomes can be used in genome-wide association studies coupled with the RIL population mapping for rapid gene discovery. The assembled genomes can further be used for comparative genome analysis with other wild relatives and progenitors. Further Pangenome comparisons can be performed to study the genome dynamics.

Chapter 6 explored the potential of einkorn wheat (*T. monococcum*) to address the problem of replenishing the depleted genetic resistance in wheat. A highly diverse association panel was screened and explored for diseases like powdery mildew, leaf rust, and stem rust using Kmer-GWAS. The study successfully identified a total of nine novel QTLs with few potential candidate genes, i.e., powdery mildew (5), leaf rust (3), and stem rust (2). These QTLs harbor some novel genes encoding for RGA, cysteine-rich receptor kinases, transcription factors, and G-type lectins. Study identifies a potential candidate for a broad-spectrum resistance to powdery mildew, co-located in the mapping region of one of the known QTL *PmTb7A.1.* Race-specific QTLs were identified for the powdery mildew races Bgt98230 (6AL) and Bgt98411 (5AL). Three QTLs were identified for leaf rust, two encoding resistance genes (RGAs) and one encoding an exocyst subunit protein belonging to the exo family. The identified genes provide a good resource for the trait enhancement in wheat. These can be validated using TILLING or RIL populations coupled with VIGS and genome editing methods.

The well-categorized and promising accessions can further be used to develop TILLING resources for gene validation. Our group has progressed in sequencing the *Ae. geniculata* genome, which will form a promising resource for the comparative study of the gene structures between the alien and domesticated germplasm. Further, Kmer-GWAS can be extended to study other agronomic traits like heading date, spike morphology, trichome length, trichome density and diseases like fusarium head blight **(**FHB) and wheat blast. The availability of new sequenced genomes opens a window for pangenome comparisons. These comparisons can reveal how gene functions have been gained or lost over time, providing valuable insights into wheat genome evolution.

Overall, this doctoral research shows the potential of wild and related germplasm for wheat breeding and provides a way forward to narrow down smaller and practically applicable number of accessions to channelize the genetic diversity in wheat improvement programs.

Using the diploid genetic, genomic and germplasm resources from einkorn wheat, this study established a nice pipeline for gene discovery. Through an integrated approach, study identified nine QTLs associated with disease resistance for most important diseases including powdery mildew, leaf rust and stem rust. This study opens opportunities to perform functional validation of agronomically important genes for these traits. In conclusion, the knowledge and resources generated from this work contribute to the understanding of the available genetic diversity of wild wheat germplasm and their effective utilization in gene discovery and overall wheat improvement. This knowledge can be utilized to develop more resilient and robust wheat varieties to address future climate change and food security challenges.

Appendices

List of Supplementary tables

Chapter 2

Chapter 3

Table S3.4. Mapped location of *Ae. geniculata* genes on chromosome 5D of wheat and comparison of gene structure.

Table S3.5. List of differentially expressed genes in T756 and T598.

Table S3.6. List of transcription factor classes identified from PlantTFDB in *Ae. geniculata* ($5M^g$) and *Ae. umbellulata* ($5U^u$).

Table S3.7. Primer sequence of uniquely mapped primer of $5M^g$ on 5D and $5U^u$, and $5U^u$ primers mapped on 5D and $5M^g$.

Table S3.8. Uniquely mapped KASP markers mapping of $5M^g$ and $5U^u$ SNPs on 5D chromosomes of Chinese Spring (CS).

Chapter 5

Table S5.1. Raw read data statistics for TA391 and TA10868.

Table S5.2. Genome assembly statistics of TA391 and TA10868.

Table S5.3. Gene feature statistics of high and low confidence genes for TA391 and TA10868.

Table S5.4. List of transcription factors identified in TA391 and TA10868.

Chapter 6

Table S6.1. Passport information of einkorn diversity panel. Accession name, country of collection, latitude (Lat.), longitude (Long.), and elevation are included. Precise GPS coordinates were obtained from Genesys (https://www.genesys-pgr.org/). If GPS coordinates were not provided, they were inferred from the respective name of the collection site (GPS type 'inferred'). If no collection site was provided, the GPS coordinates were placed in the middle of the respective country (GPS type 'country').

List of Supplementary figures

Chapter 2

Figure S2.10. SVDquartets coalescent-based tree **A**) Species tree based on the available species-specific annotation *T. turgidum* and *T. timopheevii* **B**) sub-population tree of fast structure population clusters.

Figure S2.11. Heatmap displaying the clustering and admixture pattern of *Triticum turgidum* accession from England, the United States, and Sweden.

Figure S2.12. Single nucleotide polymorphism (SNP) significantly associated with leaf rust resistance in *T. timopheevii* identified by genome-wide association study (GWAS) with FarmCPU model. (A) Manhattan plot, significantly linked SNP is marked inside the box on chromosome 2B, (B) Quantile-quantile plot.

Chapter 3

Figure S3.1. Conserved BUSCO-gene-based evaluation of *Ae. geniculata* ($5M^{g}$) and *Ae. umbellulata* ($5M^{u}$).

Figure S3.2. Whole chromosome display of density of mapped reads of *Ae. geniculata* and *Ae. umbellulata* on chromosome 5D of Chinese spring.

Figure S3.3. Coverage of mapped reads of *Ae. geniculata* ($5M^{g}$) and *Ae. umbellulata* ($5M^{u}$) on the genic region of Chinese spring.

Figure S3.4. Evolution of gene families of resistance gene and transcription identified using CAFÉ. A) expansion of resistance genes, B) contraction of resistance genes, C) rapidly evolving resistance genes, D) expansion of TFs, E) contraction of TFs, and F) rapidly evolving TFs genes.

Figure S3.5. Gene ontology classification of differentially expressed genes in A) introgression line T756 and B) introgression line T598.

Figure S3.6. Gene ontology enrichment total differentially expressed genes TA756 and WL711and B) T598 and WL711

Figure S3.7. Distribution of simple sequence repeats in *Ae. geniculata* ($5M^g$) and *Ae. umbellulata* ($5U^u$).

Figure S3.8. Synteny of *Ae. geniculata* ($5M^g$) and *Ae. umbellulata* ($5U^u$) chromosomes with chromosome 5D of Chinese spring in 10Mb introgression from introgression lines. Connecting lines represents the contig location on 5D.

Note: The supplementary tables and figures from the published data can be accessed through the hyperlink.

Bibliography

Abbo, S., Pinhasi van-Oss, R., Gopher, A., Saranga, Y., Ofner, I., & Peleg, Z. (2014). Plant domestication versus crop evolution: A conceptual framework for cereals and grain legumes. *Trends in Plant Science*, *19*(6), 351–360. https://doi.org/10.1016/j.tplants.2013.12.002

Abrouk, M., Ahmed, H. I., Cubry, P., Šimoníková, D., Cauet, S., Pailles, Y., Bettgenhaeuser, J., Gapa, L., Scarcelli, N., Couderc, M., Zekraoui, L., Kathiresan, N., Čížková, J., Hřibová, E., Doležel, J., Arribat, S., Bergès, H., Wieringa, J. J., Gueye, M., … Krattinger, S. G. (2020). Fonio millet genome unlocks African orphan crop diversity for agriculture in a changing climate. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-18329-4

Adhikari, L., Raupp, J., Wu, S., Wilson, D., Evers, B., Koo, D. H., Singh, N., Friebe, B., & Poland, J. (2022). Genetic characterization and curation of diploid A-genome wheat species. *Plant Physiology*, *188*(4), 2101–2114. https://doi.org/10.1093/plphys/kiac006

Ahmed, H. I., Heuberger, M., Schoen, A., Koo, D. H., Quiroz-Chavez, J., Adhikari, L., Raupp, J., Cauet, S., Rodde, N., Cravero, C., Callot, C., Lazo, G. R., Kathiresan, N., Sharma, P. K., Moot, I., Yadav, I. S., Singh, L., Saripalli, G., Rawat, N., … Krattinger, S. G. (2023). Einkorn genomics sheds light on history of the oldest domesticated wheat. *Nature*, *620*(7975), 830–838. https://doi.org/10.1038/s41586-023-06389-7

Akhunov, E., Nicolet, C., and Dvorak, J. (2009). Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theoretical and Applied Genetics* 119, 507–517. https://doi.org/10.1007/s00122-009-1059-5

Alam, Md. A., Xue, F., Wang, C., & Ji, W. (2011). Powdery Mildew Resistance Genes in Wheat: Identification and Genetic Analysis. *Journal of Molecular Biology Research*, *1*(1). https://doi.org/10.5539/jmbr.v1n1p20

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*. https://doi.org/10.1186/1471-2105-12-246

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., & Schatz, M. C. (2019). RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome biology*, *20*(1), 224. https://doi.org/10.1186/s13059-019-1829-6

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C., & Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology*, *23*(1). https://doi.org/10.1186/s13059-022-02823-7

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. *Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc*.

Anglin, N. L., Amri, A., Kehel, Z., & Ellis, D. (2018). A Case of Need : Linking Traits to Genebank Accessions. *16*(5), 337–349. https://doi.org/10.1089/bio.2018.0033

Aoun, M., Rouse, M. N., Kolmer, J. A., Kumar, A., & Elias, E. M. (2021). Genome-Wide Association Studies Reveal All-Stage Rust Resistance Loci in Elite Durum

Wheat Genotypes. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.640739

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C. J., Choulet, F., Distelfeld, A., Poland, J., Ronen, G., Barad, O., Baruch, K., Keeble-Gagnère, G., Mascher, M., Ben-Zvi, G., Josselin, A. A., Himmelbach, A., Balfourier, F., Wang, L. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, *361*(6403). https://doi.org/10.1126/science.aar7191

Arrigo, N., Felber, F., Parisod, C., Buerki, S., Alvarez, N., David, J., & Guadagnuolo, R. (2010). Origin and expansion of the allotetraploid Aegilops geniculata, a wild relative of wheat. *The New phytologist*, *187*(4), 1170–1180. https://doi.org/10.1111/j.1469-8137.2010.03328.x

Arrigo, N., Guadagnuolo, R., Lappe, S., Pasche, S., Parisod, C., and Felber, F. (2011). Gene flow between wheat and wild relatives: Empirical evidence from Aegilops geniculata, Ae. neglecta and Ae. triuncialis. *Evol Appl* 4, 685–695. https://doi.org/10.1111/j.1752-4571.2011.00191.x.

Arzani, A. (2011). Emmer (Triticum turgidum spp. dicoccum) Flour and Breads. In R. R. W. and V. B. P. Victor R. Preedy (Ed.), *Flour and Breads and their Fortification in Health and Disease Prevention* (pp. 69–78). https://doi.org/10.1016/C2009-0-30556-5

Arzani, A., & Ashraf, M. (2017). Cultivated Ancient Wheats (Triticum spp.): A Potential Source of Health-Beneficial Food Products. In *Comprehensive Reviews in Food Science and Food Safety* (Vol. 16, Issue 3, pp. 477–488). Blackwell Publishing Inc. https://doi.org/10.1111/1541-4337.12262

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., Komatsuda, T., … Distelfeld, A. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science*, *97*(July), 93–97. https://doi.org/10.1126/science.aan0032

Ayalew, H., Anderson, J. D., Krom, N., Tang, Y., Butler, T. J., Tiwari, V., Rawat, N., & Ma, X. F. (2022). Genotyping-by-sequencing and genomic selection applications in hexaploid triticale. *G3: Genes, Genomes, Genetics*, *12*(2). https://doi.org/10.1093/g3journal/jkab413

Badaeva, E. D., Boguslavsky, R. L., Badaev, N. S., & Zelenin, A. V. (1990). Intraspecific chromosomal polymorphism of Triticum araraticum (Poaceae) detected by C-banding technique. *Plant Systematics and Evolution*, *169*(1–2), 13–24. https://doi.org/10.1007/BF00935980

Badaeva, E. D., Ruban, A. S., Zoshchuk, S. A., & Kilian, B. (2016). Molecular cytogenetic characterization of Triticum timopheevii chromosomes provides new insight on genome evolution of T . zhukovskyi. *Plant Syst Evol*, *302*, 943–956. https://doi.org/10.1007/s00606-016-1309-3

Badaeva, E. D., Konovalov, F. A., Knüpffer, H., Fricano, A., Ruban, A. S., Kehel, Z., Zoshchuk, S. A., Surzhikov, S. A., Neumann, K., Graner, A., Hammer, K., Filatenko, A., Bogaard, A., Jones, G., Özkan, H., & Kilian, B. (2022). Genetic diversity, distribution and domestication history of the neglected GGAtAt genepool

of wheat. *Theoretical and Applied Genetics*, *135*(3), 755–776. https://doi.org/10.1007/s00122-021-03912-0

Balfourier, F., Roussel, V., Strelchenko, P., Exbrayat-Vinson, F., Sourdille, P., Boutet, G., Koenig, J., Ravel, C., Mitrofanova, O., Beckert, M., & Charmet, G. (2007). A worldwide bread wheat core collection arrayed in a 384-well plate. *Theoretical and Applied Genetics*, *114*(7), 1265–1275. https://doi.org/10.1007/s00122-007-0517-1

Bálint, A. F., Kovács, G., Erdei, L., and Sutka, J. (2001). Comparison of the Cu, Zn, Fe, Ca and Mg contents of the grains of wild, ancient and cultivated wheat species. *Cereal Res Commun* 29, 375–382. https://doi.org/10.1007/BF03543684

Bansal, M., Kaur, S., Dhaliwal, H. S., Bains, N. S., Bariana, H. S., Chhuneja, P.,& Bansal, U.K. (2017). Mapping of Aegilops umbellulata-derived leaf rust and stripe rust resistance loci in wheat. *Plant Pathol* 66, 38–44. https://doi.org/10.1111/ppa.12549

Bansal, M., Adamski, N. M., Toor, P. I., Kaur, S., Molnár, I., Holušová, K., Vrána, J., Doležel, J., Valárik, M., Uauy, C., & Chhuneja, P. (2020). Aegilops umbellulata introgression carrying leaf rust and stripe rust resistance genes Lr76 and Yr70 located to 9.47-Mb region on 5DS telomeric end through a combination of chromosome sorting and sequencing. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, *133*(3), 903–915. https://doi.org/10.1007/s00122-019-03514-x

Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6, 4–9. https://doi.org/10.1186/s13100-015-0041-9.

Bhattacharya, S. (2017). Deadly new wheat disease threatens Europe's crops. *Nature*, *542*(7640), 145–146. https://doi.org/10.1038/nature.2017.21424

Blanco, A., Gadaleta, A., Cenci, A., Carluccio, A. v., Abdelbacki, A. M. M., & Simeone, R. (2008). Molecular mapping of the novel powdery mildew resistance gene Pm36 introgressed from Triticum turgidum var. dicoccoides in durum wheat. *Theoretical and Applied Genetics*, *117*(1), 135–142. https://doi.org/10.1007/s00122-008-0760-0

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Genome analysis Trimmomatic : a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brown-Guedira, G. L., Singh, S., & Fritz, A. K. (2003). *Performance and Mapping of Leaf Rust Resistance Transferred to Wheat from Triticum timopheevii subsp. armeniacum*. *Phytopathology*, 93 (2003), 784-789, https://doi.org/10.1094/PHYTO.2003.93.7.784

Buchfink, B., Reuter, K., & Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner. in, ed. B. C. No. LBNL-7065E. Ernest Orlando Lawrence Berkeley National Laboratory.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST plus: architecture and applications. *BMC*

Bioinformatics, *10*(421), 1. https://doi.org/Artn 421\nDoi 10.1186/1471-2105-10-421

Cao, A., Xing, L., Wang, X., Yang, X., Wang, W., Sun, Y., Qian, C., Ni, J., Chen, Y., Liu, D., Wang, X., & Chen, P. (2011). Serine/threonine kinase gene Stpk-V, a key member of powdery mildew resistance gene Pm21, confers powdery mildew resistance in wheat. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(19), 7727–7732. https://doi.org/10.1073/pnas.1016981108

Chaudhary, H. K., Kaila, V., Rather, S. A., Badiyal, A., Hussain, W., Jamwal, N. S., et al. (2014). "Wheat," in *Alien Gene Transfer in Crop Plants*. 1–26. https://doi.org/10.1007/978-1-4614-9572-7_1

Chen, S., Guo, Y., Briggs, J., Dubach, F., Chao, S., Zhang, W., Rouse, M. N., & Dubcovsky, J. (2018). Mapping and characterization of wheat stem rust resistance genes SrTm5 and Sr60 from Triticum monococcum. *Theoretical and Applied Genetics*, *131*(3), 625–635. https://doi.org/10.1007/s00122-017-3024-z

Chen, S., Rouse, M. N., Zhang, W., Jin, Y., Akhunov, E., Wei, Y., & Dubcovsky, J. (2015). Fine mapping and characterization of Sr21, a temperature-sensitive diploid wheat resistance gene effective against the Puccinia graminis f. sp. tritici Ug99 race group. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, *128*(4), 645–656. https://doi.org/10.1007/s00122-015-2460-x

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), 170–175. https://doi.org/10.1038/s41592-020-01056-5

Chhuneja, P., Kaur, S., Garg, T., Ghai, M., Kaur, S., Prashar, M., Bains, N. S., Goel, R. K., Keller, B., Dhaliwal, H. S., & Singh, K. (2008). Mapping of adult plant stripe rust resistance genes in diploid A genome wheat species and their transfer to bread wheat. *Theoretical and Applied Genetics*, *116*(3), 313–324. https://doi.org/10.1007/s00122-007-0668-0

Chhuneja, P., Kumar, K., Stirnweis, D., Hurni, S., Keller, B., Dhaliwal, H. S., & Singh, K. (2012). Identification and mapping of two powdery mildew resistance genes in Triticum boeoticum L. *Theoretical and Applied Genetics*, *124*(6), 1051–1058. https://doi.org/10.1007/s00122-011-1768-4

Chhuneja, P., Yadav, B., Stirnweis, D., Hurni, S., Kaur, S., Elkot, A. F., Keller, B., Wicker, T., Sehgal, S., Gill, B. S., & Singh, K. (2015). Fine mapping of powdery mildew resistance genes PmTb7A.1 and PmTb7A.2 in Triticum boeoticum (Boiss.) using the shotgun sequence assembly of chromosome 7AL. *Theoretical and Applied Genetics*, *128*(10), 2099–2111. https://doi.org/10.1007/s00122-015-2570-5

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80–92. https://doi.org/10.4161/fly.19695

Civáň, P., & Brown, T. A. (2017). A novel mutation conferring the nonbrittle phenotype of cultivated barley. *New Phytologist*, *214*(1), 468–472. https://doi.org/10.1111/nph.14377

Cloutier, S., McCallum, B. D., Loutre, C., Banks, T. W., Wicker, T., Feuillet, C., Keller, B., & Jordan, M. C. (2007). Leaf rust resistance gene Lr1, isolated from bread wheat (Triticum aestivum L.) is a member of the large psr567 gene family. *Plant Molecular Biology*, *65*(1–2), 93–106. https://doi.org/10.1007/s11103-007-9201-8

Conesa, A., & Stefan, G. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, *2008*, 1687. https://doi.org/10.1155/2008/619832

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Cox, T. S. (1997). Deepening the Wheat Gene Pool. *Journal of Crop Production*, *1*(1), 1–25. https://doi.org/10.1300/J144v01n01_01

Criscuolo, A. (2019). A fast alignment-free bioinformatics procedure to infer accurate distance-based phylogenetic trees from genome assemblies. *Res Ideas Outcomes* 5. https://doi.org/10.3897/rio.5.e36178

Cuevas, H. E., Rosa-Valentin, G., Hayes, C. M., Rooney, W. L., & Hoffmann, L. (2017). Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: Implications for germplasm conservation, evaluation, and utilization in crop improvement. *BMC Genomics*, *18*, 108. https://doi.org/10.1186/s12864-016-3475-7

Czajkowska, B. I., Oliveira, H. R., & Brown, T. A. (2019). A discriminatory test for the wheat B and G genomes reveals misclassified accessions of Triticum timopheevii and Triticum turgidum. *PLoS ONE*, *14*(4), 1–10. https://doi.org/10.1371/journal.pone.0215175

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: A computational tool for the study of gene family evolution. *Bioinformatics*, *22*(10), 1269–1271. https://doi.org/10.1093/bioinformatics/btl097

Devi, U., Grewal, S., Yang, C., Hubbart-edwards, S., Scholefield, D., Ashling, S., Burridge, A., King, I. P., & King, J. (2019). Development and characterisation of interspecific hybrid lines with genome-wide introgressions from Triticum timopheevii in a hexaploid wheat background. *BMC plant biology*, *19*(1), 183. https://doi.org/10.1186/s12870-019-1785-z

Doležel, J., Vrána, J., Šafář, J., Bartoš, J., Kubaláková, M., and Šimková, H. (2012). Chromosomes in the flow to simplify genome analysis. *Funct Integr Genomics* 12, 397–416. https://doi.org/10.1007/s10142-012-0293-0

Dracatos, P. M., Bartov, J., Elmansour, H., Singh, D., Karafiátová, M., Zhang, P., Steuernagel, B., Svačina, R., Cobbin, J. C. A., Clark, B., Hoxha, S., Khatkar, M.

S., Doležel, J., Wulff, B. B., & Park, R. F. (2019). The coiled-coil NLR RPH1, confers leaf rust resistance in barley cultivar sudan. *Plant Physiology*, *179*(4), 1362–1372. https://doi.org/10.1104/pp.18.01052

Du, X., Jia, Z., Yu, Y., Wang, S., Che, B., Ni, F., et al. (2019). A wheat-Aegilops umbellulata addition line improves wheat agronomic traits and processing quality. *Breed Sci* 69, 503–507. https://doi.org/10.1270/jsbbs.18200

Dubcovsky, J., & Dvorak, J. (2007). Genome Plasticity a Key Factor in the Success of Polyploid Wheat Under Domestication. *Science*, *316*(5833), 1862–1866. https://doi.org/10.1126/science.1143986

Dvorak, J., McGuire, P. E., & Cassidy, B. (1988). Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. *Genome*, *30*(5), 680–689. https://doi.org/10.1139/g88-115

Dvořák, J. (2013). Triticum Species (wheat). In *Encyclopedia of Genetics*, 2060 (2013)

Dvorak, J., Wang, L., Zhu, T., Jorgensen, C. M., Deal, K. R., Dai, X., & Dawson, M. W. (2018). Structural variation and rates of genome evolution in the grass family seen through comparison of sequences of genomes greatly differing in size. *The Plant Journal*, *95*, 487–503. https://doi.org/10.1111/tpj.13964

Dvorak, J., Wang, L., Zhu, T., Jorgensen, C. M., Luo, M. C., Deal, K. R., Gu, Y. Q., Gill, B. S., Distelfeld, A., Devos, K. M., Qi, P., & McGuire, P. E. (2018). Reassessment of the evolution of wheat chromosomes 4A, 5A, and 7B. *Theoretical and Applied Genetics*, *131*(11), 2451–2462. https://doi.org/10.1007/s00122-018-3165-8

Dyck, P. L. (1992). Transfer of a gene for stem rust resistance from Triticum araraticum to hexaploid wheat. *Genomce*, *35*, 788–792. https://doi.org/10.1139/g92-120

Elkot, A. F. A., Chhuneja, P., Kaur, S., Saluja, M., Keller, B., & Singh, K. (2015). Marker assisted transfer of two powdery mildew resistance genes PmTb7A.1 and PmTb7A.2 from Triticum boeoticum (Boiss.) to Triticum aestivum (L.). *PLoS ONE*, *10*(6). https://doi.org/10.1371/journal.pone.0128297

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 1–14. https://doi.org/10.1186/s13059-019-1832-y

Endelman, J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome*, *4*(3), 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Enghiad, A., Ufer, D., Countryman, A. M., & Thilmany, D. D. (2017). An Overview of Global Wheat Market Fundamentals in an Era of Climate Concerns. *International Journal of Agronomy*, *2017*, 1–15. https://doi.org/10.1155/2017/3931897

Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, *30*(7), 1575–1584. https://doi.org/10.1093/nar/30.7.1575

Ersoz, E. S., Yu, J., & Buckler, E. S. (2007). Applications of Linkage Disequilibrium and Association Mapping in Crop Plants. In *Genomics-Assisted Crop Improvement* (pp. 97–119). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6295-7_5

Farré, M., Kim, J., Proskuryakova, A. A., Zhang, Y., Kulemzina, A. I., Li, Q., et al. (2019). Evolution of gene regulation in ruminants differs between evolutionary breakpoint regions and homologous synteny blocks. *Genome Res* 29, 576–589. https://doi.org/10.1101/gr.239863.118

Fayaz, F., Aghaee Sarbarzeh, M., Talebi, R., & Azadi, A. (2019). Genetic Diversity and Molecular Characterization of Iranian Durum Wheat Landraces (Triticum turgidum durum (Desf.) Husn.) Using DArT Markers. *Biochemical Genetics*, *57*(1), 98–116. https://doi.org/10.1007/s10528-018-9877-2

Feldman, M., & Kislev, M. E. (2007). Domestication of emmer wheat and evolution of free-threshing tetraploid wheat. *Israel Journal of Plant Sciences*, *55*(3–4), 207–221. https://doi.org/10.1560/IJPS.55.3-4.207

Fernandez-Calvin, B., & Orellana, J. (1992). Relationship between pairing frequencies and genome affinity estimations in Aegilops ovata × Triticum aestivum hybrid plants. *Heredity*, *68*(2), 165–172. https://doi.org/10.1038/hdy.1992.25

Feuillet, C., Travella, S., Stein, N., Albar, L., Nublat, A., & Keller, B. (2003). Map-based isolation of the leaf rust disease resistance gene Lr10 from the hexaploid wheat (Triticum aestivum L.) genome. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(25), 15253–15258. https://doi.org/10.1073/pnas.2435133100

Feuillet, C., Langridge, P., & Waugh, R. (2008). Cereal breeding takes a walk on the wild side. *Trends in genetics*, *24*(1), 24–32. https://doi.org/10.1016/j.tig.2007.11.001

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

Fox, S. E., Geniza, M., Hanumappa, M., Naithani, S., Sullivan, C., Preece, J., Tiwari, V. K., Elser, J., Leonard, J. M., Sage, A., Gresham, C., Kerhornou, A., Bolser, D., McCarthy, F., Kersey, P., Lazo, G. R., & Jaiswal, P. (2014). De novo transcriptome assembly and analyses of gene expression during photomorphogenesis in diploid wheat Triticum monococcum. *PLoS ONE*, *9*(5). https://doi.org/10.1371/journal.pone.0096855

Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

Frankel, O. H., Arber, W., Llimensee, K., Peacock, W. J., & Starlinger, P. (1984). Genetic manipulation: impact on man and society. *Genetic Perspective of Germplasm Conservation*.

Frankham, R., Ballou, J. D., Briscoe, D. A., & McInnes, K. H. (2002). *Introduction to Conservation Genetics*. Cambridge University Press. *Genetical Research*. 2004;83(3):221-222. https://doi.org/10.1017/S0016672304216913.

Fricano, A., Brandolini, A., Rossini, L., Sourdille, P., Wunder, J., Effgen, S., Hidalgo, A., Erba, D., Piffanelli, P., & Salamini, F. (2014). Crossability of Triticum urartu and Triticum monococcum wheats, homoeologous recombination, and description of a panel of interspecific introgression lines. *G3: Genes, Genomes, Genetics*, *4*(10), 1931–1941. https://doi.org/10.1534/g3.114.013623

Furman, B. J., Qualset, C. O., Skovmand, B., Heaton, J. H., Corke, H., & Wesenberg, D. M. (1997). Characterization and analysis of North American Triticale Genetic Resources. *Crop Science*, *37*(6), 1951–1959. https://doi.org/10.2135/cropsci1997.0011183X003700060046x

Furman, B. J. (2004). Triticale. In *Encyclopedia of Grain Science,* (pp. 298–303). Academic Press.

Garg, M., Tsujimoto, H., Gupta, R. K., Kumar, A., Kaur, N., Kumar, R., et al. (2016). Chromosome specific substitution lines of aegilops geniculata alter parameters of bread making quality of wheat. *PLoS One* 11. https://doi.org/10.1371/journal.pone.0162350

Gaurav, K., Arora, S., Silva, P., Sánchez-Martín, J., Horsnell, R., Gao, L., Brar, G. S., Widrig, V., John Raupp, W., Singh, N., Wu, S., Kale, S. M., Chinoy, C., Nicholson, P., Quiroz-Chávez, J., Simmonds, J., Hayta, S., Smedley, M. A., Harwood, W., & Wulff, B. B. H. (2022). Population genomic analysis of Aegilops tauschii identifies targets for bread wheat improvement. *Nature Biotechnology*, *40*(3), 422–431. https://doi.org/10.1038/s41587-021-01058-4

Genievskaya, Y., Pecchioni, N., Laidò, G., Anuarbek, S., Rsaliyev, A., Chudinov, V., Zatybekov, A., Turuspekov, Y., & Abugalieva, S. (2022). Genome-Wide Association Study of Leaf Rust and Stem Rust Seedling and Adult Resistances in Tetraploid Wheat Accessions Harvested in Kazakhstan. *Plants*, *11*(15). https://doi.org/10.3390/plants11151904

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0090346

Golovnina, K. A., Kondratenko, E. Y., Blinov, A. G., & Goncharov, N. P. (2010). Molecular characterization of vernalization loci VRN1 in wild and cultivated wheats. *BMC Plant Biology*, *10*, 168. https://doi.org/10.1186/1471-2229-10-168

Gornicki, P., Zhu, H., Wang, J., Challa, G. S., Zhang, Z., Gill, B. S., & Li, W. (2014). The chloroplast view of the evolution of polyploid wheat. *The New phytologist*, *204*(3), 704–714. https://doi.org/10.1111/nph.12931

Gupta, P. K., & Priyadarshan, P. M. (1982). Triticale: Present status and future prospects. *Advances in Genetics*, *21*(C), 255–345. https://doi.org/10.1016/S0065-2660(08)60300-4

Han, G., Lu, C., Guo, J., Qiao, Z., Sui, N., Qiu, N., et al. (2020). C2H2 Zinc Finger Proteins: Master Regulators of Abiotic Stress Responses in Plants. *Front Plant Sci* 11. https://doi.org/10.3389/fpls.2020.00115

Hao, C. Y., Zhang, X. Y., Wang, L. F., Dong, Y. S., Shang, X. W., & Jia, J. Z. (2006). Genetic diversity and core collection evaluations in common wheat germplasm from the Northwestern Spring Wheat Region in China. *Molecular Breeding*, *17*(1), 69–77. https://doi.org/10.1007/s11032-005-2453-6

Hao, C. Y., Dong, Y. C., Wang, L. F., You, G. X., Zhang, H. N., Ge, H. M., Jia, J. Z., & Zhang, X. Y. (2008). Genetic diversity and construction of core collection in Chinese wheat genetic resources. *Chinese Science Bulletin*, *53*(10), 1518–1526. https://doi.org/10.1007/s11434-008-0212-x

Hao, Y., Parks, R., Cowger, C., Chen, Z., Wang, Y., Bland, D., Murphy, J. P., Guedira, M., Brown-Guedira, G., & Johnson, J. (2015). Molecular characterization of a new

powdery mildew resistance gene Pm54 in soft red winter wheat. *Theoretical and Applied Genetics*, *128*(3), 465–476. https://doi.org/10.1007/s00122-014-2445-1

Harlan, J. R., & Zohary, D. (1966). Distribution of wild wheats and barley. *Science (New York, N.Y.)*, *153*(3740), 1074–1080. https://doi.org/10.1126/science.153.3740.1074

Hartl, D. L., & Clark, A. G. (1997). *Principles of Population Genetics* (3rd ed.). Sinauer Associates.https://doi.org/10.1002/(SICI)1521-1878(199812)20:12<1055::AID-BIES14>3.0.CO;2-X

He, F., Pasam, R., Shi, F., Kant, S., Keeble-gagnere, G., Kay, P., Forrest, K., Fritz, A., Hucl, P., Wiebe, K., Knox, R., Cuthbert, R., Pozniak, C., Akhunova, A., Morrell, P. L., Davies, J. P., Webb, S. R., Spangenberg, G., Hayes, B., Akhunov, E. (2019). Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nature Genetics*, *51*(May), 896–904. https://doi.org/10.1038/s41588-019-0382-2

Heun, M., Schäfer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B., & Salamini, F. (1997). Site of Einkorn Wheat Domestication Identified by DNA Fingerprinting. *Science*, *278*(5341), 1312–1314. https://doi.org/10.1126/science.278.5341.1312

Hewitt, T., Zhang, J., Huang, L., Upadhyaya, N., Li, J., Park, R., Hoxha, S., McIntosh, R., Lagudah, E., & Zhang, P. (2021). Wheat leaf rust resistance gene Lr13 is a specific Ne2 allele for hybrid necrosis. *Molecular Plant*, *14*(7), 1025–1028. https://doi.org/https://doi.org/10.1016/j.molp.2021.05.010

Hoff, K. J., and Stanke, M. (2019). Predicting Genes in Single Genomes with AUGUSTUS. *Curr Protoc Bioinformatics* 65. https://doi.org/10.1002/cpbi.57

Huang, L., Brooks, S. A., Li, W., Fellers, J. P., Trick, H. N., & Gill, B. S. (2003). Map-Based Cloning of Leaf Rust Resistance Gene Lr21 From the Large and Polyploid Genome of Bread Wheat. *Genetics*, *164*, 655–664. https://academic.oup.com/genetics/article/164/2/655/6050382

Hurni, S., Brunner, S., Buchmann, G., Herren, G., Jordan, T., Krukowski, P., Wicker, T., Yahiaoui, N., Mago, R., & Keller, B. (2013). Rye Pm8 and wheat Pm3 are orthologous genes and show evolutionary conservation of resistance function against powdery mildew. *Plant Journal*, *76*(6), 957–969. https://doi.org/10.1111/tpj.12345

Hussien, T., Bowden, R. L., Gill, B. S., & Cox, T. S. (1998). Chromosomal locations in common wheat of three new leaf rust resistance genes from Triticum monococcum. In *Euphytica* (Vol. 101).

Huynen, M. A., and Bork, P. (1998). Measuring genome evolution. *Proc. Natl. Acad. Sci. USA* 95, 5849–5856. https://doi.org/10.1073/pnas.95.11.584

Hyun, D. Y., Sebastin, R., Lee, K. J., Lee, G. A., Shin, M. J., Kim, S. H., Lee, J. R., & Cho, G. T. (2020). Genotyping-by-Sequencing Derived Single Nucleotide Polymorphisms Provide the First Well-Resolved Phylogeny for the Genus Triticum (Poaceae). *Frontiers in Plant Science*, *11*(June), 1–15. https://doi.org/10.3389/fpls.2020.00688

Järve, K., Peusha, H. O., Tsymbalova, J., Tamm, S., Devos, K. M., & Enno, T. M. (2000). Chromosomal location of a Triticum timopheevii--derived powdery

mildew resistance gene transferred to common wheat. *Genome*, *43*(2), 377–381. https://doi.org/10.1139/g99-141

Jeong, S., Kim, J. Y., Jeong, S. C., Kang, S. T., Moon, J. K., & Kim, N. (2017). GenoCore: A simple and fast algorithm for core subset selection from large genotype datasets. *PLoS ONE*, *12*(7). https://doi.org/10.1371/journal.pone.0181420

Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W., Appels, R., Pfeifer, M., Tao, Y., Zhang, X., Jing, R., Zhang, C., Ma, Y., Gao, L., Gao, C., Spannagl, M., Mayer, K. F. X., Li, D., Pan, S., … Wang, J. (2013). Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*, *496*(7443), 91–95. https://doi.org/10.1038/nature12028

Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., & Gao, G. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, *45*(D1), D1040–D1045. https://doi.org/10.1093/nar/gkw982

Jing, H. C., Kornyukhin, D., Kanyuka, K., Orford, S., Zlatska, A., Mitrofanova, O. P., Koebner, R., & Hammond-Kosack, K. (2007). Identification of variation in adaptively important traits and genome-wide analysis of trait-marker associations in Triticum monococcum. *Journal of Experimental Botany*, *58*(13), 3749–3764. https://doi.org/10.1093/jxb/erm225

Jorgensen, J. H., & Jensen, C. J. (1973). Gene Pm6 for resistance to powdery mildew in wheat. *Euphytica*, *22*, 423. https://doi.org/10.1007/BF00022656

Juliana, P., Singh, R. P., Singh, P. K., Poland, J. A., Bergstrom, G. C., Huerta-Espino, J., Bhavani, S., Crossa, J., & Sorrells, M. E. (2018). Genome-wide association mapping for resistance to leaf rust, stripe rust and tan spot in wheat reveals potential candidate genes. *Theoretical and Applied Genetics*, *131*(7), 1405–1422. https://doi.org/10.1007/s00122-018-3086-6

Katiyar, A., Smita, S., Lenka, S. K., Rajwanshi, R., Chinnusamy, V., and Bansal, K. C. (2012). Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genomics* 13. https://doi.org/10.1186/1471-2164-13-544

Khoury, C. K., Brush, S., Costich, D. E., Curry, H. A., de Haan, S., Engels, J. M. M., Guarino, L., Hoban, S., Mercer, K. L., Miller, A. J., Nabhan, G. P., Perales, H. R., Richards, C., Riggins, C., & Thormann, I. (2022). Crop genetic erosion: understanding and responding to loss of crop diversity. In *New Phytologist* (Vol. 233, Issue 1, pp. 84–118). John Wiley and Sons Inc. https://doi.org/10.1111/nph.17733

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

Kim, K. W., Chung, H. K., Cho, G. T., Ma, K. H., Chandrabalan, D., Gwag, J. G., Kim, T. S., Cho, E. G., & Park, Y. J. (2007). PowerCore: A program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics*, *23*(16), 2155–2162. https://doi.org/10.1093/bioinformatics/btm313

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53. https://doi.org/10.1111/1755-0998.12549

Kolmer, J. (2013). Leaf rust of wheat: Pathogen biology, variation and host resistance. In *Forests* (Vol. 4, Issue 1, pp. 70–84). https://doi.org/10.3390/f4010070

Kolodziej, M. C., Singla, J., Sánchez-Martín, J., Zbinden, H., Šimková, H., Karafiátová, M., Doležel, J., Gronnier, J., Poretti, M., Glauser, G., Zhu, W., Köster, P., Zipfel, C., Wicker, T., Krattinger, S. G., & Keller, B. (2021). A membrane-bound ankyrin repeat protein confers race-specific leaf rust disease resistance in wheat. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20777-x

Komsta, L. (2022). *outliers: A collection of some tests commonly used for identifying outlier.* https://CRAN.R-project.org/package=outliers. r package version 0.14

Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291. https://doi.org/10.1093/bioinformatics/btm091

Krattinger, S. G., Lagudah, E. S., Spielmeyer, W., Singh, R. P., Huerta-Espino, J., McFadden, H., Bossolini, E., Selter, L. L., & Keller, B. (2009). A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science*, *323*(5919), 1360–1363. https://doi.org/10.1126/science.1166453

Kuckuck, H. (1979). On the origin of triticum carthlicum neyski (triticum persicum vav.). *Wheat Inf. Serv.*, *50*, 1–5.

Kumar, K., Jan, I., Saripalli, G., Sharma, P. K., Mir, R. R., Balyan, H. S., & Gupta, P. K. (2022). An Update on Resistance Genes and Their Use in the Development of Leaf Rust Resistant Cultivars in Wheat. In *Frontiers in Genetics* (Vol. 13). Frontiers Media S.A. https://doi.org/10.3389/fgene.2022.816057

Kumssa, T. T., Anderson, J. D., Johnson, J. P., Norton, S., Saha, M. C., Trammell, M. A., Rogers, J. K., Butler, T. J., & Ma, X. F. (2022). Trends of seasonal forage yield changes of triticale in the southern Great Plains of the United States. *Grassland Research*, *1*(3), 166–173. https://doi.org/10.1002/glr2.12027

Kuraparthy, V., Chhuneja, P., Dhaliwal, H. S., Kaur, S., Bowden, R. L., and Gill, B. S. (2007). Characterization and mapping of cryptic alien introgression from Aegilops geniculata with new leaf rust and stripe rust resistance genes Lr57 and Yr40 in wheat. *Theoretical and Applied Genetics* 114, 1379–1389. https://doi.org/10.1007/s00122-007-0524-2

Laidò, G., Mangini, G., Taranto, F., Gadaleta, A., Blanco, A., Cattivelli, L., Marone, D., Mastrangelo, A. M., Papa, R., & De Vita, P. (2013). Genetic Diversity and Population Structure of Tetraploid Wheats (Triticum turgidum L.) Estimated by SSR, DArT and Pedigree Data. *PLoS ONE*, *8*(6), e67280. https://doi.org/10.1371/journal.pone.0067280

Law, C. N., Worland, A. J., & Giorgi, B. (1976). The genetic control of ear-emergence time by chromosomes 5A and 5D of wheat. *Heredity*, *36*(1), 49–58. https://doi.org/10.1038/hdy.1976.5

Leigh, F. J., Wright, T. I. C., Horsnell, R. A., Dyer, S., & Bentley, A. R. (2022). Progenitor species hold untapped diversity for potential climate-responsive traits

for use in wheat breeding and crop improvement. In *Heredity* (Vol. 128, Issue 5, pp. 291–303). Springer Nature. https://doi.org/10.1038/s41437-022-00527-z

Lemane, T., Chikhi, R., & Peterlongo, P. (2022). k mdiff, large-scale and user-friendly differential k-mer analyses. *Bioinformatics (Oxford, England)*, *38*(24), 5443–5445. https://doi.org/10.1093/bioinformatics/btac689

Leonova, I. N., Budashkina, E. B., Flath, K., Weidner, A., Börner, A., & Röder, M. S. (2010). Microsatellite mapping of a leaf rust resistance gene transferred to common wheat from Triticum timopheevii. *Cereal Research Communications*, *38*(2), 211–219. https://doi.org/10.1556/CRC.38.2010.2.7

Levy, Y. Y., Mesnage, S., Mylne, J. S., Gendall, A. R., & Dean, C. (2002). Multiple Roles of *Arabidopsis VRN1* in Vernalization and Flowering Time Control. *Science*, *297*(5579), 243–246. https://doi.org/10.1126/science.1072147

Li, G., Fang, T., Zhang, H., Xie, C., Li, H., Yang, T., Nevo, E., Fahima, T., Sun, Q., & Liu, Z. (2009). Molecular identification of a new powdery mildew resistance gene Pm41 on chromosome 3BL derived from wild emmer (Triticum turgidum var. dicoccoides). *Theoretical and Applied Genetics*, *119*(3), 531–539. https://doi.org/10.1007/s00122-009-1061-y

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

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, H., Luo, J., Zhang, W., Hua, L., Li, K., Wang, J., Xu, B., Yang, C., Wang, G., Rouse, M. N., Dubcovsky, J., & Chen, S. (2023). High-resolution mapping of SrTm4, a recessive resistance gene to wheat stem rust. *Theoretical and Applied Genetics*, *136*(5). https://doi.org/10.1007/s00122-023-04369-z

Li, H., Men, W., Ma, C., Liu, Q., Dong, Z., Tian, X., Wang, C., Liu, C., Gill, H. S., Ma, P., Zhang, Z., Liu, B., Zhao, Y., Sehgal, S. K., & Liu, W. (2024). Wheat powdery mildew resistance gene Pm13 encodes a mixed lineage kinase domain-like protein. *Nature Communications*, *15*(1), 2449. https://doi.org/10.1038/s41467-024-46814-7

Li, L.F., Liu, B., Olsen, K. M., and Wendel, J. F. (2015). A re-evaluation of the homoploid hybrid origin of Aegilops tauschii, the donor of the wheat D-subgenome. *New Phytologist* 208, 4–8. https://doi.org/10.1111/nph.13294

Li, M., Dong, L., Li, B., Wang, Z., Xie, J., Qiu, D., Li, Y., Shi, W., Yang, L., Wu, Q., Chen, Y., Lu, P., Guo, G., Zhang, H., Zhang, P., Zhu, K., Li, Y., Zhang, Y., Wang, R., … Liu, Z. (2020). A CNL protein in wild emmer wheat confers powdery mildew resistance. *New Phytologist*, *228*(3), 1027–1037. https://doi.org/10.1111/nph.16761

Li, X., Shi, Z., Gao, J., Wang, X., & Guo, K. (2023). CandiHap: a haplotype analysis toolkit for natural variation study. *Molecular Breeding*, *43*(3). https://doi.org/10.1007/s11032-023-01366-4

Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

Lin, G., Chen, H., Tian, B., Sehgal, S. K., Singh, L., Xie, J., Rawat, N., Juliana, P., Singh, N., Shrestha, S., Wilson, D. L., Shult, H., Lee, H., Schoen, A. W., Tiwari, V. K., Singh, R. P., Guttieri, M. J., Trick, H. N., Poland, J., … Liu, S. (2022). Cloning of the broadly effective wheat leaf rust resistance gene Lr42 transferred from Aegilops tauschii. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-30784-9

Ling, H. Q., Wang, J., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., Fan, H., Li, D., Dong, L., Tao, Y., Gao, C., Wu, H., Li, Y., Cui, Y., Guo, X., Zheng, S., Wang, B., Yu, K., … Zhang, A. (2013). Draft genome of the wheat A-genome progenitor Triticum urartu. *Nature*, *496*(7443), 87–90. https://doi.org/10.1038/nature11997

Ling, H. Q., Ma, B., Shi, X., Liu, H., Dong, L., Sun, H., Cao, Y., Gao, Q., Zheng, S., Li, Y., Yu, Y., Du, H., Qi, M., Li, Y., Lu, H., Yu, H., Cui, Y., Wang, N., Chen, C., … Liang, C. (2018). Genome sequence of the progenitor of wheat A subgenome Triticum urartu. *Nature*, *557*(7705), 424–428. https://doi.org/10.1038/s41586-018-0108-0

Liu, H. J., & Yan, J. (2019). Crop genome-wide association study: a harvest of biological relevance. *The Plant journal : for cell and molecular biology*, *97*(1), 8–18. https://doi.org/10.1111/tpj.14139

Liu, W., Maccaferri, M., Bulli, P., Rynearson, S., Tuberosa, R., Chen, X., & Pumphrey, M. (2017). Genome-wide association mapping for seedling and field resistance to Puccinia striiformis f. sp. tritici in elite durum wheat. *Theoretical and Applied Genetics*, *130*(4), 649–667. https://doi.org/10.1007/s00122-016-2841-9

Liu, W., Rouse, M., Friebe, B., Jin, Y., Gill, B., and Pumphrey, M. O. (2011). Discovery and molecular mapping of a new gene conferring resistance to stem rust, Sr53, derived from Aegilops geniculata and characterization of spontaneous translocation stocks with reduced alien chromatin. *Chromosome Research* 19, 669–682. https://doi.org/10.1007/s10577-011-9226-3

Loureiro, I., Escorial, M. C., García-Baudín, J. M., and Chueca, M. C. (2006). Evidence of natural hybridization between Aegilops geniculata and wheat under field conditions in Central Spain. *Environ Biosafety Res* 5, 105–109. https://doi.org/10.1051/ebr:2006020

Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., Buckler, E. S., & Costich, D. E. (2013). Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PLoS Genetics*, *9*(1). https://doi.org/10.1371/journal.pgen.1003215

Lu, P., Guo, L., Wang, Z., Li, B., Li, J., Li, Y., Qiu, D., Shi, W., Yang, L., Wang, N., Guo, G., Xie, J., Wu, Q., Chen, Y., Li, M., Zhang, H., Dong, L., Zhang, P., Zhu, K., … Liu, Z. (2020). A rare gain of function mutation in a wheat tandem kinase confers resistance to powdery mildew. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-14294-0

Luo, J., Rouse, M. N., Hua, L., Li, H., Li, B., Li, T., Zhang, W., Gao, C., Wang, Y., Dubcovsky, J., & Chen, S. (2022). Identification and characterization of Sr22b, a new allele of the wheat stem rust resistance gene Sr22 effective against the Ug99

race group. *Plant Biotechnology Journal*, *20*(3), 554–563. https://doi.org/10.1111/pbi.13737

Luo, M. C., Gu, Y. Q., Puiu, D., Wang, H., Twardziok, S. O., Deal, K. R., Huo, N., Zhu, T., Wang, L., Wang, Y., McGuire, P. E., Liu, S., Long, H., Ramasamy, R. K., Rodriguez, J. C., Van, S. L., Yuan, L., Wang, Z., Xia, Z., Xiao, L., … Dvořák, J. (2017). Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. *Nature*, *551*(7681), 498–502. https://doi.org/10.1038/nature24486

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., … Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, *1*(1), 18. https://doi.org/10.1186/2047-217X-1-18

Maccaferri, M., Harris, N. S., Twardziok, S. O., Pasam, R. K., Gundlach, H., Spannagl, M., Ormanbekova, D., Lux, T., Prade, V. M., Milner, S. G., Himmelbach, A., Mascher, M., Bagnaresi, P., Faccioli, P., Cozzi, P., Lauria, M., Lazzari, B., Stella, A., Manconi, A., Cattivelli, L. (2019). Durum wheat genome highlights past domestication signatures and future improvement targets. *Nature Genetics*, *51*(5), 885–895. https://doi.org/10.1038/s41588-019-0381-3

Maestra, B., & Naranjo, T. (1999). Structural chromosome differentiation between Triticum timopheevii and T. turgidum and T. aestivum. *Theoretical and Applied Genetics*, *98*(5), 744–750. https://doi.org/10.1007/s001220051130

Malihipour, A., Gilbert, J., Fedak, G., Brûlé-Babel, A., & Cao, W. (2016). Characterization of agronomic traits in a population of wheat derived from Triticum timopheevii and their association with Fusarium head blight. *European Journal of Plant Pathology*, *144*(1), 31–43. https://doi.org/10.1007/s10658-015-0744-2

Malihipour, A., Gilbert, J., Fedak, G., Brûlé-Babel, A., & Cao, W. (2017). Mapping the a genome for qtl conditioning resistance to Fusarium head blight in a wheat population with Triticum Timopheevii background. *Plant Disease*, *101*(1), 11–19. https://doi.org/10.1094/PDIS-02-16-0144-RE

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, *21*(2), 584–595. https://doi.org/10.1111/1755-0998.13265

Marcussen, T., Sandve, S. R., Heier, L., Pfeifer, M., Kugler, K. G., Zhan, B., Spannagl, M., Pfeifer, M., Jakobsen, K. S., Wulff, B. B. H., Steuernagel, B., Mayer, K. F. X., & Olsen, O.A. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science 345*(6194), 1250092. https://doi.org/10.1126/science.1251788

Martínez-Fortún, J., Phillips, D. W., & Jones, H. D. (2022). Natural and artificial sources of genetic variation used in crop breeding: A baseline comparator for genome editing. *Frontiers in Genome Editing*, *4*. https://doi.org/10.3389/fgeed.2022.937853

Mascher, M., Wicker, T., Jenkins, J., Plott, C., Lux, T., Koh, C. S., Ens, J., Gundlach, H., Boston, L. B., Tulpová, Z., Holden, S., Hernández-Pinzón, I., Scholz, U., Mayer, K. F. X., Spannagl, M., Pozniak, C. J., Sharpe, A. G., Simková, H., Moscou, M. J., … Stein, N. (2021). Long-read sequence assembly: A technical

evaluation in barley. *Plant Cell*, *33*(6), 1888–1906. https://doi.org/10.1093/plcell/koab077

Máthé, C., Garda, T., Freytag, C., & Hamvas, M. M. (2019). The role of serine-threonine protein phosphatase pp2a in plant oxidative stress signaling—facts and hypotheses. In *International Journal of Molecular Sciences* (Vol. 20, Issue 12). MDPI AG. https://doi.org/10.3390/ijms20123028

Matsuoka, Y. (2011). Evolution of polyploid triticum wheats under cultivation: The role of domestication, natural hybridization and allopolyploid speciation in their diversification. *Plant and Cell Physiology*, *52*(5), 750–764. https://doi.org/10.1093/pcp/pcr018

Maxwell, J. J., Lyerly, J. H., Cowger, C., Marshall, D., Brown-Guedira, G., & Murphy, J. P. (2009). MlAG12: A Triticum timopheevii-derived powdery mildew resistance gene in common wheat on chromosome 7AL. *Theoretical and Applied Genetics*, *119*(8), 1489–1495. https://doi.org/10.1007/s00122-009-1150-y

Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., Langlade, N., & Muños, S. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *BioTechniques*, *61*(4), 203–205. https://doi.org/10.2144/000114460

McCartney, C. A., Somers, D. J., McCallum, B. D., Thomas, J., Humphreys, D. G., Menzies, J. G., & Brown, P. D. (2005). Microsatellite tagging of the leaf rust resistance gene Lr16 on wheat chromosome 2BSc. *Molecular Breeding*, *15*(4), 329–337. https://doi.org/10.1007/s11032-004-5948-7

McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40, 4288–4297. https://doi.org/10.1093/nar/gks042

McCouch, S. R., McNally, K. L., Wang, W., & Hamilton, R. S. (2012). Genomics of gene banks: A case study in rice. *American Journal of Botany*, *99*(2), 407–423. https://doi.org/10.3732/ajb.1100385

McCouch, S., Navabi, Z. K., Abberton, M., Anglin, N. L., Barbieri, R. L., Baum, M., Bett, K., Booker, H., Brown, G. L., Bryan, G. J., Cattivelli, L., Charest, D., Eversole, K., Freitas, M., Ghamkhar, K., Grattapaglia, D., Henry, R., Valadares Inglis, M. C., Islam, T., Rieseberg, L. H. (2020). Mobilizing Crop Biodiversity. *Molecular Plant*, *13*(10), 1341–1344. https://doi.org/10.1016/j.molp.2020.08.011

Mergoum, M., Singh, P. K., Pena, R. J., Lozano-del Rı́o, A. J., Cooper, K. V., Salmon, D. F., & Macpherson, H. G. (2009). Triticale: A "New" Crop with Old Challenges. In *Cereals* (Vol. 3, pp. 267–287). Springer US. https://doi.org/10.1007/978-0-387-72297-9

Mergoum, Mohamed., & Gómez Macpherson, Helena. (2004). *Triticale improvement and production*. Food and Agriculture Organization of the United Nations.

Middleton, C. P., Senerchia, N., Stein, N., Akhunov, E. D., Keller, B., Wicker, T., et al. (2014). Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the triticeae tribe. *PLoS One* 9. https://doi.org/10.1371/journal.pone.0085761

Mistry, J., Bateman, A., & Finn, R. D. (2007). Predicting active site residue annotations in the Pfam database. *BMC Bioinformatics*, *8*, 298. https://doi.org/10.1186/1471-2105-8-298

Mitrofanova, O., Badaeva, E., & Salina, E. A. (2016). T. timopheevii and T. zhukovskii, the bread wheat cousins and their contribution to T. aestivum improvement. In V. J. Bonjean Alain P., Angus William J. & Maarten (Eds.), *The World Wheat Book: A History of Wheat Breeding* (Volume 3, pp. 1167–228). Tec & Doc Lavoisier.

Molnár, I., Kubaláková, M., Šimková, H., Cseh, A., Molnár-Láng, M., and Doležel, J. (2011). Chromosome isolation by flow sorting in Aegilops umbellulata and Ae. comosa and their allotetraploid hybrids Ae. biuncialis and Ae. geniculata. *PLoS One* 6. https://doi.org/10.1371/journal.pone.0027708

Molnár, I., Kubaláková, M., Šimková, H., Farkas, A., Cseh, A., Megyeri, M., et al. (2014). Flow cytometric chromosome sorting from diploid progenitors of bread wheat, T. urartu, Ae. speltoides and Ae. tauschii. *Theoretical and Applied Genetics* 127, 1091–1104. https://doi.org/10.1007/s00122-014-2282-2

Moore, J. W., Herrera-Foessel, S., Lan, C., Schnippenkoetter, W., Ayliffe, M., Huerta-Espino, J., Lillemo, M., Viccars, L., Milne, R., Periyannan, S., Kong, X., Spielmeyer, W., Talbot, M., Bariana, H., Patrick, J. W., Dodds, P., Singh, R., & Lagudah, E. (2015). A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. *Nature Genetics*, *47*(12), 1494–1498. https://doi.org/10.1038/ng.3439

Mori, N., Kondo, Y., Ishii, T., Kawahara, T., Valkoun, J., & Nakamura, C. (2009). Genetic diversity and origin of timopheevi wheat inferred by chloroplast DNA fingerprinting. *Breeding Science*, *59*, 571–578

Muterko, A., & Salina, E. (2018). Origin and distribution of the VRN-A1 exon 4 and exon 7 haplotypes in domesticated wheat species. *Agronomy*, *8*(8), 1–14. https://doi.org/10.3390/agronomy8080156

Nakano, T., Suzuki, K., Fujimura, T., and Shinshi, H. (2006). Genome-wide analysis of the ERF gene family in arabidopsis and rice. *Plant Physiol* 140, 411–432. https://doi.org/10.1104/pp.105.073783

Natsidis, P., Kapli, P., Schiffer, P. H., and Telford, M. J. (2021). Systematic errors in orthology inference and their effects on evolutionary analyses. *iScience* 24. https://doi.org/10.1016/j.isci.2021.102110

Nave, M., Taş, M., Raupp, J., Tiwari, V. K., Ozkan, H., Poland, J., Hale, I., Komatsuda, T., & Distelfeld, A. (2021). The independent domestication of timopheev's wheat: Insights from haplotype analysis of the brittle rachis 1 (btr1-a) gene. *Genes*, *12*(3), 1–10. https://doi.org/10.3390/genes12030338

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(12 (I)), 3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Nevo, E., Gorham, J., & Beiles, A. (1992). Variation for 11Na uptake in wild emmer wheat, Triticum dicoccoides in israel: Salt tolerance resources for wheat improvement. *Journal of Experimental Botany*, *43*(4), 511–518. https://doi.org/10.1093/jxb/43.4.511

Nguyen, N. N., Kim, M., Jung, J. K., Shim, E. J., Chung, S. M., Park, Y., Lee, G. P., & Sim, S. C. (2020). Genome-wide SNP discovery and core marker sets for assessment of genetic variations in cultivated pumpkin (Cucurbita spp.). *Horticulture Research*, *7*(1), 108. https://doi.org/10.1038/s41438-020-00342-9

Odong, T. L., Jansen, J., Eeuwijk, F. A. Van, & Hintum, T. J. L. Van. (2013). *Quality of core collections for effective utilisation of genetic resources review , discussion and interpretation*. 289–305. https://doi.org/10.1007/s00122-012-1971-y

Okada, M., Ikeda, T. M., Yoshida, K., & Takumi, S. (2018). Effect of the U genome on grain hardness in nascent synthetic hexaploids derived from interspecific hybrids between durum wheat and Aegilops umbellulata. *J Cereal Sci* 83, 153–161. https://doi.org/10.1016/j.jcs.2018.08.011

Okada, M., Yoshida, K., Nishijima, R., Michikawa, A., Motoi, Y., Sato, K., & Takumi, S. (2018). RNA-seq analysis reveals considerable genetic diversity and provides genetic markers saturating all chromosomes in the diploid wild wheat relative Aegilops umbellulata. *BMC plant biology*, *18*(1), 271. https://doi.org/10.1186/s12870-018-1498-8

Oliveira, H. R., Hagenblad, J., Leino, M. W., Leigh, F. J., Lister, D. L., Penã-Chocarro, L., & Jones, M. K. (2014). Wheat in the Mediterranean revisited - tetraploid wheat landraces assessed with elite bread wheat Single Nucleotide Polymorphism markers. *BMC Genetics*, *15*(54), 1–13. https://doi.org/10.1186/1471-2156-15-54

Oliveira, H. R., Jacocks, L., Czajkowska, B. I., Kennedy, S. L., & Brown, T. A. (2020). Multiregional origins of the domesticated tetraploid wheats. *PLoS ONE*, *15*(1), 1–20. https://doi.org/10.1371/journal.pone.0227148

Olivera, P. D., Sikharulidze, Z., Dumbadze, R., Szabo, L. J., Newcomb, M., Natsarishvili, K., Rouse, M. N., Luster, D. G., & Jin, Y. (2019). Presence of a Sexual Population of Puccinia graminis f. Sp. Tritici in Georgia Provides a Hotspot for Genotypic and Phenotypic Diversity. *Phytopathology*, *109*(12), 2152–2160. https://doi.org/10.1094/PHYTO-06-19-0186-R

Olivera, P., Newcomb, M., Szabo, L. J., Rouse, M., Johnson, J., Gale, S., Luster, D. G., Hodson, D., Cox, J. A., Burgin, L., Hort, M., Gilligan, C. A., Patpour, M., Justesen, A. F., Hovmøller, M. S., Woldeab, G., Hailu, E., Hundie, B., Tadesse, K., & Jin, Y. (2015). Phenotypic and genotypic characterization of race TKTTF of Puccinia graminis f. sp. tritici that caused a wheat stem rust epidemic in southern Ethiopia in 2013-14. *Phytopathology*, *105*(7), 917–928. https://doi.org/10.1094/PHYTO-11-14-0302-FI

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-019-1905-y

Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T., Sato, H., Sato, H., Hori, M., Nakamura, Y., & Tanaka, T. (2002). Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature Genetics*, *32*(4), 650–654. https://doi.org/10.1038/ng1047

Özkan, H., Willcox, G., Graner, A., Salamini, F., & Kilian, B. (2011). Geographic distribution and domestication of wild emmer wheat (Triticum dicoccoides). *Genetic Resources and Crop Evolution*, *58*(1), 11–53. https://doi.org/10.1007/s10722-010-9581-5

Padmanaban, S., Zhang, P., Sutherland, M. W., & Martin, A. (2018). Association between presence of Triticum timopheevii introgression and D-genome retention

in hexaploid / tetraploid wheat crosses. *Molecular Breeding*, *38*(84), 1–8. https://doi.org/10.1007/s11032-018-0838-6

Pandey, P., Irulappan, V., Bagavathiannan, M. V, & Sanchez, L. E. (2017). *Impact of Combined Abiotic and Biotic Stresses on Plant Growth and Avenues for Crop Improvement by Exploiting Physio-morphological Traits*. *8*(April), 1–15. https://doi.org/10.3389/fpls.2017.00537

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Pascual, L., Fernández, M., Aparicio, N., López-Fernández, M., Fité, R., Giraldo, P., & Ruiz, M. (2020). Development of a multipurpose core collection of bread wheat based on high-throughput genotyping data. *Agronomy*, *10*(4). https://doi.org/10.3390/agronomy10040534

PAUP (phylogenetic analysis using parsimony). (2008). In *Encyclopedia of Genetics, Genomics, Proteomics and Informatics* (p. 1455). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6754-9_12413

Pavlidis, P., Živković, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, *30*(9), 2224–2234. https://doi.org/10.1093/molbev/mst112

Peleg, Z., Fahima, T., Korol, A. B., Abbo, S., & Saranga, Y. (2011). Genetic analysis of wheat domestication and evolution under domestication. *Journal of Experimental Botany*, *62*(14), 5051–5061. https://doi.org/10.1093/jxb/err206

Percival, J. (1921). *The wheat plant : a monograph*. London: Duckworth. https://doi.org/10.1038/109366a0

Pérez-Wohlfeil, E., Diaz-del-Pino, S., & Trelles, O. (2019). Ultra-fast genome comparison for large-scale genomic experiments. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-46773-w

Poland, J. A., & Rife, T. W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome*, *5*(3), 92–102. https://doi.org/10.3835/plantgenome2012.05.0005

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by- Sequencing Approach. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0032253

Pourkheirandish, M., Hensel, G., Kilian, B., Senthil, N., Chen, G., Sameri, M., Azhaguvel, P., Sakuma, S., Dhanagond, S., Sharma, R., Mascher, M., Himmelbach, A., Gottwald, S., Nair, S. K., Tagiri, A., Yukuhiro, F., Nagamura, Y., Kanamori, H., Matsumoto, T., … Komatsuda, T. (2015). Evolution of the Grain Dispersal System in Barley. *Cell*, *162*(3), 527–539. https://doi.org/10.1016/j.cell.2015.07.002

Pourkheirandish, M., Dai, F., Sakuma, S., Kanamori, H., Distelfeld, A., Willcox, G., Kawahara, T., Matsumoto, T., Kilian, B., & Komatsuda, T. (2018). On the origin of the non-brittle rachis trait of domesticated einkorn wheat. *Frontiers in Plant Science*, *8*(January), 1–10. https://doi.org/10.3389/fpls.2017.02031

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. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rahman, A., Hallgrímsdó ttir, I., Eisen, M., & Pachter, L. (2018). *Association mapping from sequencing reads using k-mers*. eLife 7:e32920 https://doi.org/10.7554/eLife.32920.001

Rahman, S., Islam, S., Yu, Z., She, M., Nevo, E., & Ma, W. (2020). Current progress in understanding and recovering the wheat genes lost in evolution and domestication. *International Journal of Molecular Sciences*, *21*(16), 1–19. https://doi.org/10.3390/ijms21165836

Raj, A., Stephens, M., & Pritchard, J. K. (2014). FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Rakszegi, M., Molnár, I., Lovegrove, A., Darkó, É., Farkas, A., Láng, L., et al. (2017). Addition of aegilops U and M chromosomes affects protein and dietary fiber content of wholemeal wheat flour. *Front Plant Sci* 8. https://doi.org/10.3389/fpls.2017.01529.

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The transcriptional landscape of polyploid wheat. *Science (1979)* 361. https://doi.org/10.1126/science.aar6089

Ramirez-Gonzalez, R. H., Uauy, C., and Caccamo, M. (2015). PolyMarker: A fast polyploid primer design pipeline. *Bioinformatics* 31, 2038–2039. https://doi.org/10.1093/bioinformatics/btv069

Rawat, N., Sehgal, S. K., Joshi, A., Rothe, N., Wilson, D. L., Mcgraw, N., Vadlani, P. V, Li, W., & Gill, B. S. (2012). *A diploid wheat TILLING resource for wheat functional genomics*. http://www.biomedcentral.com/1471-2229/12/205

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0066428

Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02134-9

Riaz, A., Periyannan, S., Aitken, E., & Hickey, L. (2016). A rapid phenotyping method for adult plant resistance to leaf rust in wheat. *Plant Methods*, *12*(1). https://doi.org/10.1186/s13007-016-0117-7

Riefolo, C., Ficco, D., Cattivelli, L., & Vita, P. (2011). Genetic diversity of gluten proteins in T. turgidum L. *Cereal Research Communications*, *39*(3), 405–414. https://doi.org/10.1556/CRC.39.2011.3.10

Rodríguez, S., Maestra, B., Perera, E., Díez, M., & Naranjo, T. (2000). Pairing affinities of the B- and G-genome chromosomes of polyploid wheats with those of Aegilops speltoides. *Genome*, *43*(5), 814–819. https://doi.org/10.1139/g00-055

Rouse, M. N., & Jin, Y. (2011). Stem rust resistance in A-genome diploid relatives of wheat. *Plant Disease*, *95*(8), 941–944. https://doi.org/10.1094/PDIS-04-10-0260

Rouse, M. N., Olson, E. L., Gill, B. S., Pumphrey, M. O., & Jin, Y. (2011). Stem rust resistance in *Aegilops Tauschii* germplasm. *Crop Science*, 51(5), 2074–2078. https://doi.org/10.2135/cropsci2010.12.0719

Ruiz, M., Giraldo, P., Royo, C., & Carrillo, J. M. (2013). Creation and validation of the spanish durum wheat core collection. *Crop Science*, *53*(6), 2530–2537. https://doi.org/10.2135/cropsci2013.04.0238

Saccomanno, A., Matny, O., Marone, D., Laidò, G., Petruzzino, G., Mazzucotelli, E., Desiderio, F., Blanco, A., Gadaleta, A., Pecchioni, N., De Vita, P., Steffenson, B., & Mastrangelo, A. M. (2018). Genetic mapping of loci for resistance to stem rust in a tetraploid wheat collection. *International Journal of Molecular Sciences*, *19*(12). https://doi.org/10.3390/ijms19123907

Said, M., Holušová, K., Farkas, A., Ivanizs, L., Gaál, E., Cápal, P., et al. (2021). Development of DNA Markers From Physically Mapped Loci in Aegilops comosa and Aegilops umbellulata Using Single-Gene FISH and Chromosome Sequences. *Front Plant Sci* 12. https://doi.org/10.3389/fpls.2021.689031

Saintenac, C., Zhang, W., Salcedo, A., Rouse, M. N., Trick, H. N., Akhunov, E., & Dubcovsky, J. (2013). Identification of Wheat Gene *Sr35* That Confers Resistance to Ug99 Stem Rust Race Group. *Science*, *341*(6147), 783–786. https://doi.org/10.1126/science.1239022

Salmons, D., Temelh, F., & Spence, S. (2002). *Chemical composition of Western Canadian triticale varieties*. Proceedings of the 5thInternational Triticale Symposium,Volume II June30-July5,2002,Radzikow,Poland.

Sánchez-Martín, J., Steuernagel, B., Ghosh, S., Herren, G., Hurni, S., Adamski, N., Vrána, J., Kubaláková, M., Krattinger, S. G., Wicker, T., Doležel, J., Keller, B., & Wulff, B. B. H. (2016). Rapid gene isolation in barley and wheat by mutant chromosome sequencing. *Genome Biology*, *17*(1) 221. https://doi.org/10.1186/s13059-016-1082-1

Sapkota, S., Mergoum, M., Kumar, A., Fiedler, J. D., Johnson, J., Bland, D., Lopez, B., Sutton, S., Ghimire, B., Buck, J., Chen, Z., & Harrison, S. (2020). A novel adult plant leaf rust resistance gene Lr2K38 mapped on wheat chromosome 1AL. *Plant Genome*, *13*(3). https://doi.org/10.1002/tpg2.20061

Sari, E., Knox, R. E., Ruan, Y., Henriquez, M. A., Kumar, S., Burt, A. J., Cuthbert, R. D., Konkin, D. J., Walkowiak, S., Campbell, H. L., Singh, A. K., Ross, J., Lokuruge, P., Hsueh, E., Boyle, K., Sidebottom, C., Condie, J., Yates, S., Pozniak, C. J., & Fobert, P. R. (2020). Historic recombination in a durum wheat breeding panel enables high-resolution mapping of Fusarium head blight resistance quantitative trait loci. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-64399-1

Saripalli, G., Adhikari, L., Amos, C., Kibriya, A., Ahmed, H. I., Heuberger, M., Raupp, J., Athiyannan, N., Wicker, T., Abrouk, M., Wallace, S., Hosseinirad, S., Chhuneja, P., Livesay, J., Rawat, N., Krattinger, S. G., Poland, J., & Tiwari, V. (2023). Integration of genetic and genomics resources in einkorn wheat enables precision mapping of important traits. *Communications Biology*, *6*(1). https://doi.org/10.1038/s42003-023-05189-z

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution*, *3*(3), 430–439. https://doi.org/10.1038/s41559-018-0793-y

Schwacke, R., Ponce-Soto, G. Y., Krause, K., Bolger, A. M., Arsova, B., Hallab, A., Gruden, K., Stitt, M., Bolger, M. E., & Usadel, B. (2019). MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Molecular Plant*, *12*(6), 879–892. https://doi.org/10.1016/j.molp.2019.01.003

Sears, E. R. (1956). *The transfer of leaf-rust resistance from Aegilops umbellulata to wheat.* Available at: https://www.cabdirect.org/cabdirect/abstract/19581600184

Sela, N., Kim, E., and Ast, G. (2010). The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome Biol* 11. https://doi.org/10.1186/gb-2010-11-6-r59

Sharma, A., Srivastava, P., Mavi, G. S., Kaur, S., Kaur, J., Bala, R., Singh, T. P., Sohu, V. S., Chhuneja, P., Bains, N. S., & Singh, G. P. (2021). Resurrection of Wheat Cultivar PBW343 Using Marker-Assisted Gene Pyramiding for Rust Resistance. *Frontiers in plant science*, *12*, 570408. https://doi.org/10.3389/fpls.2021.570408

Sharma, P., Bawa, P., Yadav, B., Kaur, P., Jindal, S., Yadav, I., Kaur, S., Singh, K., & Chhuneja, P. (2020). Physical mapping of an adult plant stripe rust resistance gene from Triticum monococcum. *Journal of Plant Biochemistry and Biotechnology*, *29*(1), 47–55. https://doi.org/10.1007/s13562-019-00511-5

Shcherban, A. B., Strygina, K. V., & Salina, E. A. (2015). VRN-1 gene- associated prerequisites of spring growth habit in wild tetraploid wheat T. dicoccoides and the diploid A genome species. *BMC Plant Biology*, *15*(1), 1–13. https://doi.org/10.1186/s12870-015-0473-x

Shcherban, A. B., & Salina, E. A. (2017). Evolution of VRN-1 homoeologous loci in allopolyploids of Triticum and their diploid precursors. *BMC Plant Biology*, *17*(Suppl 1), 188. https://doi.org/10.1186/s12870-017-1129-9

Shitsukawa, N., Ikari, C., Shimada, S., Kitagawa, S., Sakamoto, K., Saito, H., Ryuto, H., Fukunishi, N., Abe, T., Takumi, S., Nasuda, S., & Murai, K. (2007). The einkorn wheat (Triticum monococcum) mutant, maintained vegetative phase, is caused by a deletion in the VRN1 gene. *Genes and Genetic Systems*, *82*(2), 167–170. https://doi.org/10.1266/ggs.82.167

Sidhu, J. S., Ramakrishnan, S. M., Ali, S., Bernardo, A., Bai, G., Abdullah, S., Ayana, G., & Sehgal, S. K. (2019). Assessing the genetic diversity and characterizing genomic regions conferring Tan Spot resistance in cultivated rye. *PLoS ONE*, *14*(3). https://doi.org/10.1371/journal.pone.0214519

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simeone, R., Piarulli, L., Nigro, D., Signorile, M. A., Blanco, E., Mangini, G., & Blanco, A. (2020). Mapping powdery mildew (Blumeria graminis f. sp. tritici) resistance inwild and cultivated tetraploid wheats. *International Journal of Molecular Sciences*, *21*(21), 1–19. https://doi.org/10.3390/ijms21217910

Singh, K., Chhuneja, P., Ghai, M., Kaur, S., Goel, R. K., Bains, N. S., Keller, B., & Dhaliwal, H. S. (2007). *Wheat Production in Stressed Environments*. http://www.usda.gov/nass

Singh, K., Ghai, M., Garg, M., Chhuneja, P., Kaur, P., Schnurbusch, T., Keller, B., & Dhaliwal, H. S. (2007). An integrated molecular linkage map of diploid wheat based on a Triticum boeoticum x T. monococcum RIL population. *Theoretical and Applied Genetics*, *115*(3), 301–312. https://doi.org/10.1007/s00122-007-0543-z

Singh, N., Wu, S., Raupp, W. J., Sehgal, S., Arora, S., Tiwari, V., Vikram, P., Singh, S., Chhuneja, P., Gill, B. S., & Poland, J. (2019). Efficient curation of genebanks using next generation sequencing reveals substantial duplication of germplasm accessions. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-018-37269-0

Singh, N., Wu, S., Tiwari, V., Sehgal, S., Raupp, J., Wilson, D., Abbasov, M., Gill, B., & Poland, J. (2019). Genomic analysis confirms population structure and identifies inter-lineage hybrids in Aegilops tauschii. *Frontiers in Plant Science*, *10, 9*. https://doi.org/10.3389/fpls.2019.00009

Singh, R. P., Singh, P. K., Rutkoski, J., Hodson, D. P., He, X., Jørgensen, L. N., Hovmøller, M. S., & Huerta-Espino, J. (2016). Disease Impact on Wheat Yield Potential and Prospects of Genetic Control. *Annual Review of Phytopathology*, *54*, 303–322. https://doi.org/10.1146/annurev-phyto-080615-095835

Singh, S. P., Hurni, S., Ruinelli, M., Brunner, S., Sanchez-Martin, J., Krukowski, P., Peditto, D., Buchmann, G., Zbinden, H., & Keller, B. (2018). Evolutionary divergence of the rye Pm17 and Pm8 resistance genes reveals ancient diversity. *Plant Molecular Biology*, *98*(3), 249–260. https://doi.org/10.1007/s11103-018-0780-3

Slater, G. S., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics*, *6*, 31. https://doi.org/10.1186/1471-2105-6-31

Soriano, J. M., Villegas, D., Aranzana, M. J., García Del Moral, L. F., & Royo, C. (2016). Genetic Structure of Modern Durum Wheat Cultivars and Mediterranean Landraces Matches with Their Agronomic Performance. *PloS one*, *11*(8), e0160983. https://doi.org/10.1371/journal.pone.0160983

Srnić, G., Murphy, J. P., Lyerly, J. H., Leath, S., & Marshall, D. S. (2005). Inheritance and chromosomal assignment of powdery mildew resistance genes in two winter wheat germplasm lines. *Crop Science*, *45*(4), 1578–1586. https://doi.org/10.2135/cropsci2004.0530

Stace, C. A. (1987). Triticale: A Case of Nomenclatural Mistreatment. *Taxon 36*, 445-452. https://doi.org/10.2307/1221447

Stakman, E. C., Stewart, D. M., & Loegering, W. Q. (1962). Identification of physiologic races of Puccinia graminis var. tritici. Available at: https://www.ars.usda.gov/ARSUserFiles/50620500/Cerealrusts/Pgt/Stakman_code_Pgt.pdf

Steadham, J., Schulden, T., Kalia, B., Koo, D. H., Gill, B. S., Bowden, R., et al. (2021). An approach for high-resolution genetic mapping of distant wild relatives of bread wheat: example of fine mapping of Lr57 and Yr40 genes. *Theoretical and Applied Genetics* 134, 2671–2686. https://doi.org/10.1007/s00122-021-03851-w

Steuernagel, B., Witek, K., Krattinger, S. G., Ramirez-Gonzalez, R. H., Schoonbeek, H. J., Yu, G., Baggs, E., Witek, A. I., Yadav, I., Krasileva, K. V., Jones, J. D. G., Uauy, C., Keller, B., Ridout, C. J., & Wulff, B. B. H. (2020). The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiology*, *183*(2), 468–482. https://doi.org/10.1104/pp.19.01273

Stoilova, T., and Spetsov, P. (2006). Chromosome 6U from Aegilops geniculata Roth Carrying Powdery Mildew Resistance in Bread Wheat. Breeding Science 56, 351-357. https://doi.org/10.1270/jsbbs.56.351

Sukumaran, S., & Yu, J. (2014). Association Mapping of Genetic Resources: Achievements and Future Perspectives. In *Genomics of Plant Genetic Resources* (pp. 207–235). Springer Netherlands. https://doi.org/10.1007/978-94-007-7572-5_9

Sun, X., Wang, Y., and Sui, N. (2018). Transcriptional regulation of bHLH during plant response to stress. *Biochem Biophys Res Commun* 503, 397–401. https://doi.org/10.1016/j.bbrc.2018.07.123

Swaminathan, K., Peterson, K., & Jack, T. (2008). The plant B3 superfamily. In *Trends in Plant Science* (Vol. 13, Issue 12, pp. 647–655). https://doi.org/10.1016/j.tplants.2008.09.006

Tang, H., Lyons, E., Pedersen, B., Schnable, J. C., Paterson, A. H., and Freeling, M. (2011). Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC Bioinformatics* 12. https://doi.org/10.1186/1471-2105-12-102

Tanno, K. I., & Willcox, G. (2006). How fast was wild wheat domesticated? *Science*, *311*(5769), 1886. https://doi.org/10.1126/science.1124635

Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics*, *Chapter 4*, 4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25

International Wheat Genome Sequencing Consortium (IWGSC) (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, *361*(6403), eaar7191. https://doi.org/10.1126/science.aar7191

The SRA Toolkit Development Team (2013). *https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software*.

Theißen, G., Melzer, R., and Ruümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: Linking plant development and evolution. *Development (Cambridge)* 143, 3259–3271. https://doi.org/10.1242/dev.134080

Thiel, T., Michalek, W., Varshney, R. K., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theoretical and Applied Genetics* 106, 411–422. https://doi.org/10.1007/s00122-002-1031-0

Thind, A. K., Wicker, T., Šimková, H., Fossati, D., Moullet, O., Brabant, C., Vrána, J., Doležel, J., & Krattinger, S. G. (2017). Rapid cloning of genes in hexaploid wheat using cultivar-specific long-range chromosome assembly. *Nature Biotechnology*, *35*(8), 793–796. https://doi.org/10.1038/nbt.3877

Tiwari, V. K., Rawat, N., Chhuneja, P., Neelam, K., Aggarwal, R., Randhawa, G. S., Dhaliwal, H. S., Keller, B., & Singh, K. (2009). Mapping of quantitative trait loci

for grain iron and zinc concentration in diploid a genome wheat. *Journal of Heredity*, *100*(6), 771–776. https://doi.org/10.1093/jhered/esp030

Tiwari, V. K., Wang, S., Danilova, T., Koo, D. H., Vrána, J., Kubaláková, M., Hribova, E., Rawat, N., Kalia, B., Singh, N., Friebe, B., Doležel, J., Akhunov, E., Poland, J., Sabir, J. S., & Gill, B. S. (2015). Exploring the tertiary gene pool of bread wheat: sequence assembly and analysis of chromosome 5M(g) of Aegilops geniculata. *The Plant journal : for cell and molecular biology*, *84*(4), 733–746. https://doi.org/10.1111/tpj.13036

Tiwari, V. K., Wang, S., Sehgal, S., Vrána, J., Friebe, B., Kubaláková, M., Chhuneja, P., Doležel, J., Akhunov, E., Kalia, B., Sabir, J., & Gill, B. S. (2014). SNP Discovery for mapping alien introgressions in wheat. *BMC genomics*, *15*, 273. https://doi.org/10.1186/1471-2164-15-273

Torkamaneh, D., Laroche, J., Bastien, M., Abed, A., & Belzile, F. (2017). Fast-GBS: A new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data. *BMC Bioinformatics*, *18*(1). https://doi.org/10.1186/s12859-016-1431-9

Tripodi, P., Rabanus-Wallace, M. T., Barchi, L., Kale, S., Esposito, S., Acquadro, A., Schafleitner, R., van Zonneveld, M., Prohens, J., Diez, M. J., Börner, A., Salinier, J., Caromel, B., Bovy, A., Boyaci, F., Pasev, G., Brandt, R., Himmelbach, A., Portis, E., Finkers, R., … Stein, N. (2021). Global range expansion history of pepper (*Capsicum* spp.) revealed by over 10,000 genebank accessions. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(34), e2104315118. https://doi.org/10.1073/pnas.2104315118

Ullah, N., Asif, M., Badshah, H., Bashir, T., & Mumtaz, A. S. (2016). Introgression lines obtained from the cross between triticum aestivum and triticum turgidum (Durum wheat) as a source of leaf and stripe (yellow) rust resistance geneswe. *Turkish Journal of Biology*, *40*(3), 547–553. https://doi.org/10.3906/biy-1501-99

van Poecke, R. M., Maccaferri, M., Tang, J., Truong, H. T., Janssen, A., van Orsouw, N. J., Salvi, S., Sanguineti, M. C., Tuberosa, R., & van der Vossen, E. A. (2013). Sequence-based SNP genotyping in durum wheat. *Plant biotechnology journal*, *11*(7), 809–817. https://doi.org/10.1111/pbi.12072

Van SM. (1994) Wild wheats: a monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae). Wageningen Agriculture University Papers.

Vardi, A., & Zohary, D. (1967). Introgression in wheat via triploid hybrids. *Heredity*, *22*, 541–560. https://doi.org/https://doi.org/10.1038/hdy.1967.69

Venske, E., Dos Santos, R. S., Busanello, C., Gustafson, P., & Costa de Oliveira, A. (2019). Bread wheat: a role model for plant domestication and breeding. In *Hereditas* (Vol. 156, p. 16). NLM (Medline). https://doi.org/10.1186/s41065-019-0093-9

Voichek, Y., & Weigel, D. (2020). Identifying genetic variants underlying phenotypic variation in plants without complete genomes. *Nature Genetics*, *52*(5), 534–540. https://doi.org/10.1038/s41588-020-0612-7

Wagenaar, E. B. (1966). Studies on the genome constitution of triticum timopheevi zhuk. Ii. The t. Timopheevi complex and its origin. *Evolution*, *20*(2), 150–164. https://doi.org/10.1111/j.1558-5646.1966.tb03351.x

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Delorean, E., Thambugala, D., Klymiuk, V., Byrns, B., Clavijo, B., Koo, D., Ens, J., Wiebe, K., & Diaye, A. N. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature*, *588*(April), 277–283. https://doi.org/10.1038/s41586-020-2961-x

Wang, C., Yin, G., Xia, X., He, Z., Zhang, P., Yao, Z., Qin, J., Li, Z., & Liu, D. (2016). Molecular mapping of a new temperature-sensitive gene LrZH22 for leaf rust resistance in Chinese wheat cultivar Zhoumai 22. *Molecular Breeding*, *36*(2), 1–10. https://doi.org/10.1007/s11032-016-0437-3

Wang, J., & Zhang, Z. (2021). GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics, Proteomics and Bioinformatics*, *19*(4), 629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Wang, P., Huang, J., Li, N., Zhang, J., Gu, C., Yuan, Y., Wen, Z., Jia, H., Kong, Z., & Ma, Z. (2023). Identification and fine mapping of PmNJ3946 for powdery mildew resistance in einkorn wheat. *Crop Journal*, *11*(6), 1846–1851. https://doi.org/10.1016/j.cj.2023.05.010

Wang, X., Bao, K., Reddy, U. K., Bai, Y., Hammar, S. A., Jiao, C., Wehner, T. C., Ramírez-Madera, A. O., Weng, Y., Grumet, R., & Fei, Z. (2018). The USDA cucumber (Cucumis sativus L.) collection: genetic diversity, population structure, genome-wide association studies, and core collection development. *Horticulture Research*, *5*(1). https://doi.org/10.1038/s41438-018-0080-8

Wang, X., Ando, K., Wu, S., Reddy, U. K., Tamang, P., Bao, K., Hammar, S. A., Grumet, R., McCreight, J. D., & Fei, Z. (2021). Genetic characterization of melon accessions in the U.S. National Plant Germplasm System and construction of a melon core collection. *Molecular Horticulture*, *11*(1). https://doi.org/10.1186/s43897-021-00014-9

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, *40*(7), e49. https://doi.org/10.1093/nar/gkr1293

Wang, Y., Zhang, Y., Zhou, R., Dossa, K., Yu, J., Li, D., Liu, A., Mmadi, M. A., Zhang, X., & You, J. (2018). Identification and characterization of the bZIP transcription factor family and its expression in response to abiotic stresses in sesame. *PloS one*, *13*(7), e0200850. https://doi.org/10.1371/journal.pone.0200850

Wang, Y., Abrouk, M., Gourdoupis, S., Koo, D. H., Karafiátová, M., Molnár, I., Holušová, K., Doležel, J., Athiyannan, N., Cavalet-Giorsa, E., Jaremko, Ł., Poland, J., & Krattinger, S. G. (2023). An unusual tandem kinase fusion protein confers leaf rust resistance in wheat. *Nature Genetics*, *55*(6), 914–920. https://doi.org/10.1038/s41588-023-01401-2

Wicker, T., Gundlach, H., Spannagl, M., Uauy, C., Borrill, P., Ramírez-González, R. H., De Oliveira, R., Mayer, K. F. X., Paux, E., & Choulet, F. (2018). Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1479-0

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York Available at: https://ggplot2.tidyverse.org.

Wilson, S. (1873). II . Wheat and Rye Hybrids. *Transactions of the Botanical Society of Edinburgh*, *12*(1–4), 286–288. https://doi.org/10.1080/03746607309469536

Winfield, M. O., Allen, A. M., Burridge, A. J., Barker, G. L. A., Benbow, H. R., Wilkinson, P. A., et al. (2016). High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol J* 14, 1195–1206. https://doi.org/10.1111/pbi.12485.

Wójcik-Gront, E., & Studnicki, M. (2021). Long-term yield variability of triticale (×triticosecale wittmack) tested using a cart model. *Agriculture (Switzerland)*, *11*(2), 1–12. https://doi.org/10.3390/agriculture11020092

Wu, X., Bian, Q., Gao, Y., Ni, X., Sun, Y., Xuan, Y., Cao, Y., & Li, T. (2021). Evaluation of resistance to powdery mildew and identification of resistance genes in wheat cultivars. *PeerJ*, *9*. https://doi.org/10.7717/peerj.10425

Wulft, B. B. H., & Moscou, M. J. (2014). Strategies for transferring resistance into wheat: From wide crosses to GM cassettes. In *Frontiers in Plant Science* (Vol. 5, Issue DEC). Frontiers Media S.A. https://doi.org/10.3389/fpls.2014.00692

Xie, J., Guo, G., Wang, Y., Hu, T., Wang, L., Li, J., Qiu, D., Li, Y., Wu, Q., Lu, P., Chen, Y., Dong, L., Li, M., Zhang, H., Zhang, P., Zhu, K., Li, B., Deal, K. R., Huo, N., Zhang, Y., … Liu, Z. (2020). A rare single nucleotide variant in Pm5e confers powdery mildew resistance in common wheat. *The New phytologist*, *228*(3), 1011–1026. https://doi.org/10.1111/nph.16762

Xie, W., & Nevo, E. (2008). Wild emmer: Genetic resources, gene mapping and potential for wheat improvement. *Euphytica*, *164*(3), 603–614. https://doi.org/10.1007/s10681-008-9703-8

Xie, W., Ben-David, R., Zeng, B., Distelfeld, A., Röder, M. S., Dinoor, A., & Fahima, T. (2012). Identification and characterization of a novel powdery mildew resistance gene PmG3M derived from wild emmer wheat, Triticum dicoccoides. *Theoretical and Applied Genetics*, *124*(5), 911–922. https://doi.org/10.1007/s00122-011-1756-8

Xu, H., Yao, G., Xiong, L., Yang, L., Jiang, Y., Fu, B., Zhao, W., Zhang, Z., Zhang, C., & Ma, Z. (2008). Identification and mapping of pm2026: A recessive powdery mildew resistance gene in an einkorn (Triticum monococcum L.) accession. *Theoretical and Applied Genetics*, *117*(4), 471–477. https://doi.org/10.1007/s00122-008-0791-6

Xu, X., Kolmer, J., Li, G., Tan, C., Carver, B. F., Bian, R., Bernardo, A., & Bai, G. (2022). Identification and characterization of the novel leaf rust resistance gene Lr81 in wheat. *Theoretical and Applied Genetics*, *135*(8), 2725–2734. https://doi.org/10.1007/s00122-022-04145-5

Yadav, I. S., Sharma, A., Kaur, S., Nahar, N., Bhardwaj, S. C., Sharma, T. R., & Chhuneja, P. (2016). Comparative Temporal Transcriptome Profiling of Wheat near Isogenic Line Carrying *Lr*57 under Compatible and Incompatible Interactions. *Frontiers in plant science*, *7*, 1943. https://doi.org/10.3389/fpls.2016.01943

Yahiaoui, N., Srichumpa, P., Dudler, R., & Keller, B. (2004). Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene Pm3b from hexaploid wheat. *Plant Journal*, *37*(4), 528–538. https://doi.org/10.1046/j.1365-313X.2003.01977.x

Yang, X., Xu, M., Wang, Y., Cheng, X., Huang, C., Zhang, H., Li, T., Wang, C., Chen, C., Wang, Y., & Ji, W. (2022). Development and Molecular Cytogenetic Identification of Two Wheat-*Aegilops geniculata* Roth $7M^{g}$ Chromosome

Substitution Lines with Resistance to Fusarium Head Blight, Powdery Mildew and Stripe Rust. *International journal of molecular sciences*, *23*(13), 7056. https://doi.org/10.3390/ijms23137056

Yao, G., Zhang, J., Yang, L., Xu, H., Jiang, Y., Xiong, L., Zhang, C., Zhang, Z., Ma, Z., & Sorrells, M. E. (2007). Genetic mapping of two powdery mildew resistance genes in einkorn (Triticum monococcum L.) accessions. *Theoretical and Applied Genetics*, *114*(2), 351–358. https://doi.org/10.1007/s00122-006-0438-4

Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., Xu, H., Huang, X., Li, S., Zhou, A., Zhang, X., Bolund, L., Chen, Q., Wang, J., Yang, H., Fang, L., & Shi, C. (2018). WEGO 2.0: A web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Research*, *46*(W1), W71–W75. https://doi.org/10.1093/nar/gky400

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Available at: http://genomebiology.com/2010/11/2/R14

Yu, G., Matny, O., Champouret, N., Steuernagel, B., Moscou, M. J., Hernández-Pinzón, I., Green, P., Hayta, S., Smedley, M., Harwood, W., Kangara, N., Yue, Y., Gardener, C., Banfield, M. J., Olivera, P. D., Welchin, C., Simmons, J., Millet, E., Minz-Dub, A., … Wulff, B. B. H. (2022). Aegilops sharonensis genome-assisted identification of stem rust resistance gene Sr62. *Nature Communications*, *13*(1), 1607. https://doi.org/10.1038/s41467-022-29132-8

Yu, K., Liu, D., Wu, W., Yang, W., Sun, J., Li, X., Zhan, K., Cui, D., Ling, H., Liu, C., & Zhang, A. (2017). Development of an integrated linkage map of einkorn wheat and its application for QTL mapping and genome sequence anchoring. *Theoretical and Applied Genetics*, *130*(1), 53–70. https://doi.org/10.1007/s00122-016-2791-2

Zaharieva, M., & Monneveux, P. (2014). Cultivated einkorn wheat (Triticum monococcum L. subsp. monococcum): the long life of a founder crop of agriculture. *Genetic Resources and Crop Evolution*, *61*(3), 677–706. https://doi.org/10.1007/s10722-014-0084-7

Zeller, F. J., Kong, L., Hartl, L., Mohler, V., and Hsam, S. L. K. (2002). Chromosomal location of genes for resistance to powdery mildew in common wheat (Triticum aestivum L. em Thell.) 7. Gene Pm29 in line Pova. *Euphytica* 123, 187–194. https://doi.org/10.1023/A:1014944619304

Zhang, X.-Y., Wang, R.-C., and Dong, Y.-S. (1996). RAPD polymorphisms in Aegilops geniculata Roth (Ae. ovata auct. non L.). *Genet Resour Crop Evol* 43, 429. https://doi.org/10.1007/BF00123733

Zhu, T., Wang, L., Rodriguez, J. C., Deal, K. R., Avni, R., Distelfeld, A., Mcguire, P. E., Dvorak, J., & Luo, M. (2019). *Improved Genome Sequence of Wild Emmer Wheat Zavitan with the Aid of Optical Maps*. *9*(March), 619–624. https://doi.org/10.1534/g3.118.200902

Zohary, D., Harlan, J.R. & Vardi, A. The wild diploid progenitors of wheat and their breeding value. *Euphytica* **18**, 58–65 (1969). https://doi.org/10.1007/BF00021982

Zohary, D., Hopf, M., & Weiss, E. (2012). Cereals. In *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin* (p. 51). https://doi.org/10.1093/acprof:osobl/9780199549061.003.0003

Zou, S., Wang, H., Li, Y., Kong, Z., & Tang, D. (2018). The NB-LRR gene Pm60 confers powdery mildew resistance in wheat. *New Phytologist*, *218*(1), 298–309. https://doi.org/10.1111/nph.14964