uOttawa

A Study of *Aegilops* Genomes: Towards Understanding Their Evolution and Potential for Wheat Improvement

Hamna Shazadee

Thesis submitted to the University of Ottawa in

partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Biology

Department of Biology
Faculty of Science
University of Ottawa

Abstract

Crop improvement has historically relied on the continuous enhancement of yield, adaptability, and resilience, yet polyploidization, domestication and modern breeding have narrowed the genetic diversity of cultivated wheat. Wild relatives, particularly species within the genus *Aegilops*, represent a critical reservoir of untapped genetic variation with the potential to address emerging agricultural challenges. Despite their evolutionary and agronomic importance, genomic resources for many *Aegilops* species have remained limited, restricting comprehensive investigations of their genome evolution and utilization in wheat improvement.

In this study, high-quality genome assemblies and gene annotations were generated for 18 diploid, tetraploid, and hexaploid *Aegilops* species, which, together with previously published genomes, complete a reference set for all 25 species within the genus. Assembly sizes ranged from 5.24 Gb in diploids to 12.65 Gb in polyploids, with high contiguity enabling detailed analyses of complex and repeat-rich genomic regions. Gene annotation provided a comprehensive overview of protein-coding gene content across all assembled *Aegilops* genomes. The annotated gene sets offer a consistent resource for comparative genomic analyses and functional exploration across the genus.

Phylogenetic reconstruction based on near single-copy orthologs across *Triticum* and *Aegilops* genomes provided a well-resolved evolutionary framework and confirmed relationships between polyploid genomes and their diploid progenitors. Divergence time estimates further contextualized species evolution within the genus. Comparative analyses of structural variation revealed extensive genome restructuring following polyploidization, including insertions, deletions, duplications, and rearrangements, while maintaining overall syntenic conservation.

These patterns highlight the dynamic nature of polyploid genomes and their divergence from ancestral states.

Genome-wide characterization of transposable elements revealed that *Aegilops* genomes are dominated by long terminal repeat (LTR) retrotransposons, with evidence of both conserved and subgenome-specific patterns of expansion and contraction. Temporal analyses of LTR insertions indicated variation in insertion activity over time, reflecting differences in transposable element dynamics among species and subgenomes.

Together, these results provide a comprehensive and integrative view of genome evolution in *Aegilops*, linking genome structure, transposable element dynamics, and evolutionary history across ploidy levels. The genomic resources and insights generated in this study not only advance our understanding of polyploid genome evolution but also provide a valuable foundation for harnessing *Aegilops* diversity to improve the resilience, adaptability, and productivity of cultivated wheat.

Acknowledgements

This journey has been one of the most challenging yet deeply meaningful periods of my life, and it would not have been possible without the support, kindness, and belief of so many people.

First and foremost, I would like to express my deepest gratitude to my supervisor and thesis advisor, Dr. Sylvie Cloutier, Principal Research Scientist at Agriculture and Agri-Food Canada (AAFC) and Adjunct Professor at the University of Ottawa, for her exceptional guidance, unwavering support, and mentorship throughout my PhD. Thank you for giving me this dream opportunity and for shaping my research mindset and academic skills, particularly in scientific writing and critical thinking. Your dedication and passion for science continue to inspire me. Even after retirement, your continued presence in the lab and commitment to research are truly admirable. You will always remain a role model in my academic and research career. I am especially grateful for encouraging me to attend and present at the third International Wheat Congress, an experience that greatly strengthened my confidence and remains one of the most memorable moments of my journey. I will always be deeply grateful to you.

I am also sincerely grateful to my co-supervisor, Dr. Frank M. You, Senior Research Scientist in Bioinformatics at Agriculture and Agri-Food Canada and Adjunct Professor at the University of Manitoba, for his invaluable guidance, insightful discussions, and continuous support. His expertise in bioinformatics and genomics, along with his thoughtful feedback, greatly strengthened this research. I truly appreciate his patience, generosity with his time, and willingness to guide me whenever I needed direction. I would also like to extend a special thanks to Dr. Frank M. You and Dr. Sylvie Cloutier for allowing me to complete my research remotely and for organizing lab meetings at times that were suitable for my circumstances.

I would like to extend my sincere thanks to my thesis advisory committee members, Dr. Catherine Cullingham (Associate Professor of Biology, Carleton University) and Dr. Julian R. Starr (Professor of Biology, University of Ottawa), for their valuable feedback, thoughtful discussions, and continuous encouragement. Their critical insights significantly contributed to my scientific growth. I also gratefully acknowledge Dr. Bahram Samanfar, Research Scientist at Agriculture and Agri-Food Canada and Adjunct Professor at Carleton University, for serving as an external examiner for my comprehensive examination. I am also greatly thankful to Dr. Wen Chen,

Research Scientist at Agriculture and Agri-Food Canada and Adjunct Professor at the University of Ottawa, for kindly agreeing to serve as an internal examiner for my thesis. My sincere thanks to Dr. Julie King (Professor of Cereal Genetics, University of Nottingham, UK) for kindly agreeing to serve as the external examiner of my thesis.

I am very grateful to the senior research technicians in Dr. Sylvie Cloutier's lab, Tara Edwards and Madeleine Lévesque-Lemay, for their invaluable assistance with experiments and for their support in reviewing my research proposal and presentations. I would also like to thank Dr. Bijendra Khadka (Postdoctoral Fellow) for his guidance on the gene annotation pipeline and for his support, as well as Bosen Jia for his assistance with running analyses when needed.

I am also thankful to members of Dr. Frank You's lab, especially Dr. Chunfang Zheng (Biologist in Computational Biology), for her technical assistance, management of sequence data, and support with data analysis whenever needed, and Dr. Pingchuan Li (Postdoctoral Fellow) for always being willing to assist with running analyses whenever required.

I also extend my sincere gratitude to the current and former members of Dr. Sylvie Cloutier's lab at AAFC, Tara Edwards, Madeleine Lévesque-Lemay, Dawn Chi, Dr. Bijendra Khadka, Dr. Akshaya Vasudevan, Sampurna Bartaula, and Bosen Jia for their support and friendship throughout this journey.

I would also like to sincerely thank Dr. Sara Martin, Research Scientist at Agriculture and Agri-Food Canada, and Tracey James, Research Technician at Agriculture and Agri-Food Canada, for their invaluable assistance with conducting the flow cytometry experiments.

I would like to sincerely thank Dr. Curtis Pozniak (Professor and Director, Crop Development Centre, University of Saskatchewan) and Jennifer Ens (Research Assistant, University of Saskatchewan) for generating ONT and Hi-C data, assisting with draft genome assemblies, and contributing to manuscript review.

I gratefully acknowledge the financial support from the 4D Wheat Project: Diversity, Domestication, Discovery and Delivery, a partnership between Genome Canada, Agriculture and Agri-Food Canada, and stakeholders in the wheat industry, with administrative support from

Genome Prairie, as well as funding through the Canadian Agricultural Partnership. I am also sincerely grateful to Agriculture and Agri-Food Canada (AAFC) for the opportunity to work as a Research Affiliate Program student. During my time here, I gained valuable experience that has contributed significantly to both my research career and personal growth. I am equally thankful to the University of Ottawa for financial support through the International Doctoral Scholarship and the PhD Admission Scholarship.

My heartfelt thanks to Dr. Sridhar Ravichandran and his wife, Dr. Garima, for their kindness and generosity in welcoming me into their home when I first arrived in Canada during the COVID-19 pandemic. Your care, warmth, and support during that time meant so much to me.

I am sincerely thankful to my cousin sisters, Sana Nazir and Saba Azam, for their constant encouragement and support throughout this journey.

I also extend my gratitude to my dear friends in Pakistan, Faiza Idress, my childhood best friend, Amara Arif, and Tehreem Iqbal, for their support and encouragement despite the distance.

I express my deepest gratitude to my family. To my dearest parents, Muhammad Latif and Sajida Parveen, thank you for your unconditional love, sacrifices, and unwavering belief in me. My beloved father, I wish you were here to see me reach this milestone. I wish you were still waiting for me and asking how many days are left until I come home. Until we meet again! To my sister, Amrin Latif, words are not enough to express my gratitude. Your endless support, sacrifices, and strength have carried me through this journey. You took care of everything at home so I could pursue my dreams; this would not have been possible without you. To my brothers, Fahad Bajwa, Hamad Bajwa, Jawad Bajwa, and Hamid Bajwa, thank you for your love, trust, and encouragement, and for always supporting my ambitions.

A very special and heartfelt thanks to my husband, Nadeem Khan. Your unwavering support, patience, and encouragement made this journey possible. Managing a PhD while raising our child was not easy, but you stood by me through it all. Thank you for waking up early, leaving for work at 6 a.m. every day, and taking care of our baby immediately upon returning home so I could focus on my work. In moments when it felt impossible to continue, you reminded me that I could do it, and because of you, I did.

And finally, to my little one, my greatest source of strength and joy. You taught me resilience in ways I never imagined and showed me that even the impossible can be achieved.

And to myself, thank you for not giving up.

Dedicated to

My Beloved Father

I Miss You More Than Words

Could Ever Express

Table of contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

Bp	Base pair
BUSCO	Benchmarking Universal Single-Copy Orthologs
CDS	Coding DNA sequence
CGIAR	Consultative Group on International Agricultural Research
CNV	Copy number variation
DNA	Deoxyribonucleic acid
EDTA	Extensive *de novo* transposable element annotator
env	Envelope protein
FAO	Food and Agriculture Organization
gag	Group-specific antigen
Gb	Gigabase
Gbp	Gigabase pairs
GMAP	Genomic mapping and alignment program
GTR	General time-reversible (model)
Gamma	Gamma distribution (rate variation among sites)
HC	High confidence
Hi-C	High-throughput chromosome conformation capture
HapHiC	Haplotype-aware Hi-C scaffolding tool
HMW gDNA	High-molecular-weight genomic DNA
HyPo	Hybrid polisher
INT	Integrase
I	Invariant sites

Iso-seq	Isoform sequencing (PacBio technology)
IWGSC	International Wheat Genome Sequencing Consortium
kb	Kilobase
LC	Low confidence
LINE	Long interspersed nuclear element
LTR	Long terminal repeat
MAFFT	Multiple Alignment using Fast Fourier Transform
Mb	Megabase
Mbp	Megabase pairs
MCMC	Markov Chain Monte Carlo
MITE	Miniature inverted-repeat transposable element
MT	Million tonnes
MUMmer	Maximal unique matches
MYA	Million years ago
NCBI	National Center for Biotechnology Information
ONT	Oxford Nanopore Technologies
Omni-C	Chromosome conformation capture technology (Dovetail Genomics)
PacBio	Pacific Biosciences sequencing technology
PAL2NAL	Protein alignment to nucleic acid alignment
PAML	Phylogenetic analysis by maximum likelihood
PAV	Presence/absence variation
PFGE	Pulsed-field gel electrophoresis
PhyKIT	Phylogenomics toolkit

pol	Polyprotein
PR	Protease
PTREP	Protein database of TREP-derived sequences
RH	Ribonuclease H
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RT	Reverse transcriptase
SINE	Short interspersed nuclear element
SNPs	Single nucleotide polymorphisms
Subsp.	Subspecies
SV	Structural variation
SyRI	Synteny and rearrangement identifier
TE	Transposable element
TIR	Terminal inverted repeat
TSD	Target site duplication
TREP	Triticeae repeat sequence database
YaHS	Yet another Hi-C scaffolding tool

Chapter 1

General introduction

1.1 Literature review

1.1.1 Wheat

Wheat, a staple crop cultivated on more than 219 million hectares worldwide, provides approximately one-fifth of the global population's total caloric and protein intakes (CGIAR, 2018; FAO, 2024). According to the Food and Agriculture Organization (FAO), global wheat production reached an average of approximately 798 million tonnes (MT) in 2024 (FAO, 2024). Despite its global importance, recent wheat production has struggled to keep pace with increasing demand, leading to price instability and food insecurity in many regions.

By 2050, the global population is projected to reach nearly 9 billion, with wheat demand expected to increase by approximately 60% (Godfray et al., 2010; Tadesse et al., 2016). To meet this demand, annual wheat yield gains must rise from the current rate of less than 1% to at least 1.6% per year (Tadesse et al., 2016). Achieving such gains is increasingly challenging due to constraints imposed by water scarcity, climate change, and the continual emergence of new pathogens and pests, all of which threaten yield stability and productivity (Asseng et al., 2015; Chaves et al., 2013; Rasheed et al., 2018). Addressing these challenges requires substantial improvements in wheat productivity, resilience to abiotic and biotic stresses, and the adoption of improved agronomic practices. Strengthening the genetic and genomic foundations of wheat improvement is therefore essential to ensure sustainable wheat production and long-term global food security.

1.1.2 Wheat evolution

Bread wheat (*Triticum aestivum* L.) is an allohexaploid species comprising the homoeologous subgenomes A, B, and D (2n = 6x = 42). Its evolutionary history within the *Triticum–Aegilops*

complex involved two successive allopolyploidization events (Figure 1.1) (Venske et al., 2019). The first allopolyploidization event occurred approximately 500,000 years ago and involved hybridization between the diploid species *Triticum urartu*, which contributed the A subgenome, and a species closely related to *Aegilops speltoides*, which contributed the B subgenome. This event gave rise to tetraploid wild emmer wheat (*Triticum turgidum* subsp. *dicoccoides*) (Li et al., 2022; Marcussen et al., 2014). Domestication of wild emmer gave rise to domesticated emmer wheat (*T. turgidum* subsp. *dicoccum*), from which subsequent selection led to the development of durum wheat (*T. turgidum* subsp. *durum*) (Charmet, 2011; Kimber, 2001). The second allopolyploidization event involved hybridization between an AB allotetraploid wheat lineage and *Aegilops tauschii* (Tausch's goatgrass), the donor of the D subgenome, resulting in the formation of hexaploid bread wheat (Kihara, 1944; McFadden and Sears, 1946; Salamini et al., 2002).

The diploid progenitors of the A and D subgenomes are widely accepted as being *T. urartu* and *Ae. tauschii*, respectively. In contrast, the precise origin of the B subgenome progenitor remains unresolved, as no extant diploid species exhibits complete chromosomal affinity with the B genome of tetraploid or hexaploid wheats. Whole-genome sequencing analyses suggest that *Ae. speltoides* diverged from the B subgenome donor approximately 4.5 million years ago, supporting the hypothesis that the B subgenome originated from an extinct or yet-unidentified diploid species closely related to *Ae. speltoides* (Li et al., 2022).

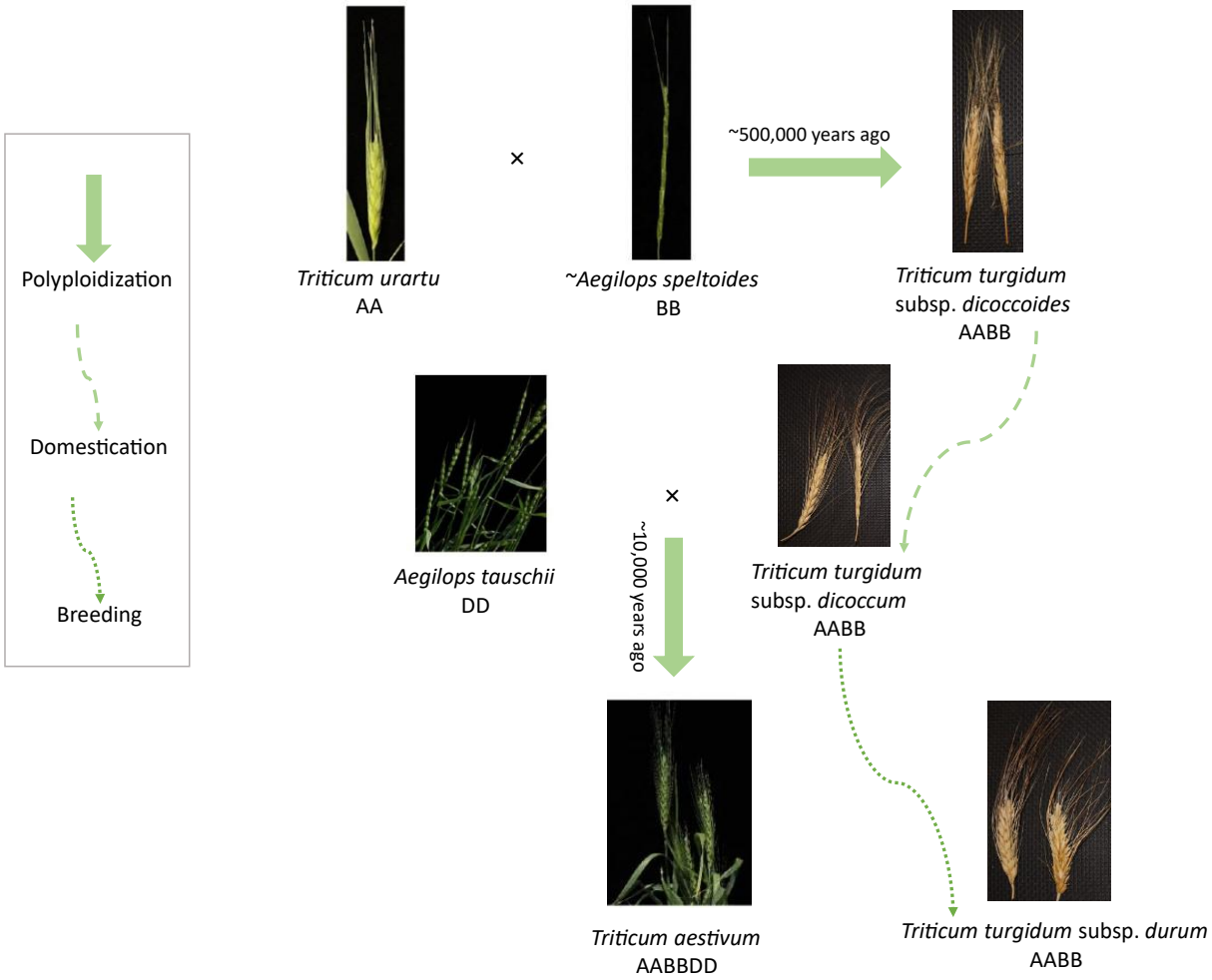


Figure 1.1 Evolution of hexaploid wheat.

Photographs courtesy of Dr. Sylvie Cloutier and Dr. Curtis Pozniak.

1.1.3 Genetic diversity bottlenecks in wheat

The wheat genome has undergone a series of genetic diversity bottlenecks during its evolutionary history as a result of speciation, domestication, natural selection and, more recently, intensive artificial selection (breeding). These bottlenecks have reduced the genetic variation present in modern bread wheat and are reflected in the limited diversity observed within contemporary wheat germplasm.

1.1.3.1 Speciation

Speciation is the evolutionary process through which new and distinct species arise. It occurs when populations become genetically distinct. Such divergence may result from geographic isolation or from genomic events such as polyploidization. Three major modes of speciation have been described: allopatric, sympatric, and parapatric (Mayr, 1999). Allopatric speciation occurs when populations are physically separated and gene exchange is prevented, whereas sympatric speciation takes place without spatial isolation. Parapatric speciation involves divergence in adjacent populations that maintain limited gene flow during the speciation process (Abbott et al., 2008).

Polyploidization has played a particularly important role in plant evolution and speciation (Soltis and Soltis, 2009). Polyploid species can arise either through autopolyploidization, involving chromosome duplication within a single species, or through allopolyploidization, which results from hybridization between divergent species followed by genome doubling. Allopolyploidization represents a rapid and profound form of speciation, generating new species with combined genomes from distinct progenitors (Levy and Feldman, 2022). Wheat is a prime example of allopolyploid evolution (Feldman et al., 2012).

During the formation of polyploid wheat, only a limited number of individuals participated in the original hybridization events, leading to a pronounced genetic bottleneck known as the founder effect (Dubcovsky and Dvorak, 2007). Also, few *Aegilops tauschii* individuals contributed to the formation of hexaploid bread wheat, leading to reduced representation of *Ae. tauschii* genetic diversity in the wheat D subgenome (Akhunov et al., 2010; Dvorak et al., 1998). Similarly, the A and B subgenomes of bread wheat capture only approximately 31% of the genetic diversity found in the wild emmer wheat gene pool (Haudry et al., 2007). Together, these speciation-related

bottlenecks have reduced genetic diversity in modern wheat, underscoring the need to explore wild relatives for the potential to broaden genetic diversity in wheat.

1.1.3.2 Domestication

Domestication is the process by which plants are selected by humans for traits that are advantageous for agriculture. Most major crop species were domesticated between 4,000 and 10,000 years ago from wild progenitor populations (Doebley et al., 2006). Domestication typically results in a reduction of genetic diversity, as cultivated plants retain only a subset of the variation present in their wild ancestors (Tanksley and McCouch, 1997).

Wheat was among the earliest domesticated crops, with domestication occurring approximately 10,000 years ago (Charmet, 2011; Levy and Feldman, 2022). Key domestication traits in wheat include the non-brittle rachis, which prevents seed shattering, and non-hulled glumes, which confer free-threshing ability. During early wheat cultivation, farmers repeatedly selected and propagated plants expressing these favorable traits, leading to their fixation in domesticated populations (Harlan et al., 1973; Levy and Feldman, 2022). However, this selective process also imposed a genetic bottleneck. Domestication involved a limited number of founding genotypes, resulting in reduced genetic variability in domesticated tetraploid and hexaploid wheats compared with their wild diploid and tetraploid progenitors (Charmet, 2011). This domestication-driven bottleneck, together with earlier speciation events, has contributed to the narrow genetic base of modern wheat.

1.1.3.3 Artificial selection/breeding

The development and application of modern wheat breeding over the past century have led to the release of high-yielding cultivars that underpin contemporary agricultural production. However, it has been argued that intensive breeding practices have also further narrowed the genetic base

available for crop improvement (Reif et al., 2005). Continuous selection for a limited set of agronomically favorable traits has reduced genetic diversity in wheat beyond the bottlenecks imposed by speciation and domestication, thereby constraining the pool of variation available for future breeding efforts (Tanksley and McCouch, 1997).

In bread wheat, breeding-associated bottlenecks are estimated to have reduced effective population size by approximately 6% (Cavanagh et al., 2013). Moreover, many breeding programs have relied heavily on a small number of elite genotypes, resulting in the loss or underutilization of genetic variation that may be critical for adaptation to emerging challenges such as climate change and its associated impacts on biotic and abiotic stresses (Lopes et al., 2015). Together, these effects highlight how modern breeding practices, while successful in increasing yield, have further constrained wheat genetic diversity and reinforce the need to broaden the germplasm base using wild and underexploited genetic resources.

1.1.4 Genetic diversity in wild relatives of wheat

Wild relatives of wheat represent a critical reservoir of genetic diversity that can be harnessed to improve crop resilience, productivity, and adaptation. Among these wild relatives, species of the genus *Aegilops* (goatgrasses) are the closest relatives of cultivated wheat and constitute a prime source of diversity. The genus *Aegilops* comprises 25 species, including 11 diploid and 14 polyploid taxa. These species are classified into five sections based on genome composition: *Aegilops* (U genome and its combinations), *Cylindropyron* (C and DC genomes), *Vertebrata* (D genome and its combinations), *Comopyrum* (M or N genomes), and *Sitopsis* (S genome) (Van Slageren, 1994) . In addition, *Aegilops mutica* (T genome) is a diploid species that does not fall within these five sections and is placed in the separate *Amblyopyrum* subgenus (Figure 1.2) (Van Slageren, 1994). The arrows in Figure 1.2 point to the evolutionary relationships between species

within the genus. Despite sharing genomes, these 25 species are highly diverse as can be seen by the extent of the morphological variation at spike levels (Figure 1.3).

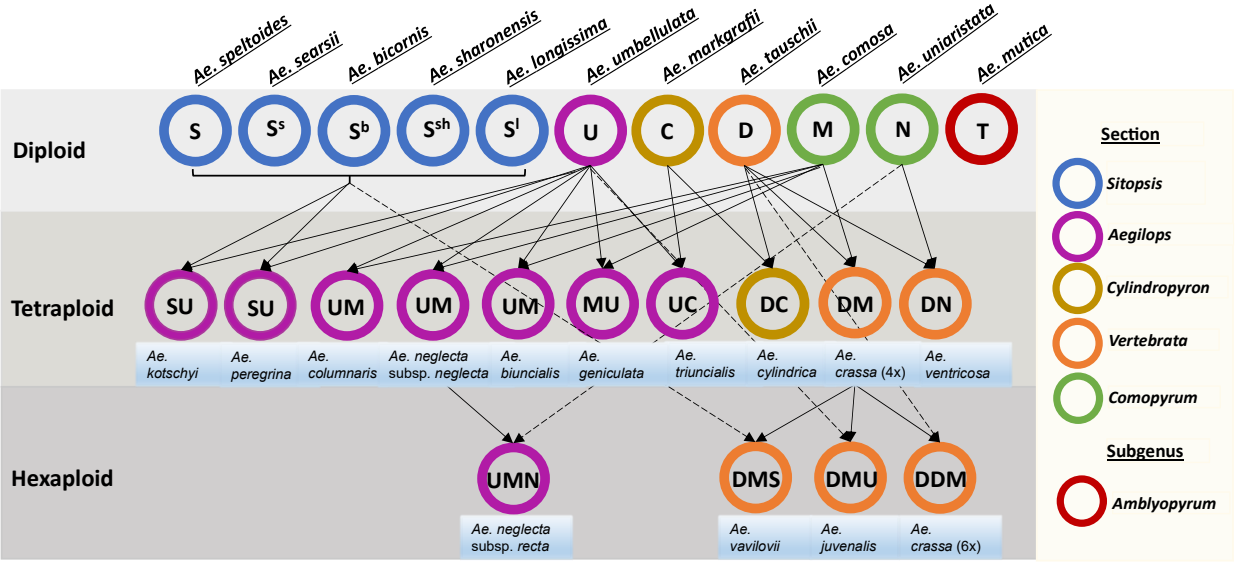


Figure 1.2 *Aegilops* species classification, their nuclear genomes and relationships across ploidy level.

Solid lines represent progenitor relationships involved in the formation of tetraploid genomes and the tetraploid ancestors that contributed to hexaploid genome formation. Dashed lines represent diploid progenitor species that contributed genomes during the polyploidization events leading to hexaploid genomes. Figure adapted from Przewieslik-Allen et al. (2019), with modifications.

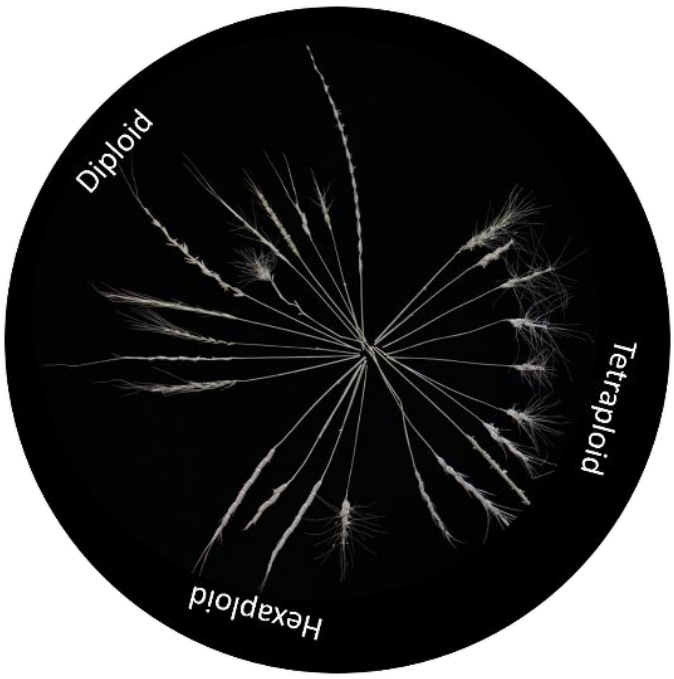


Figure 1.3 Morphological diversity of *Aegilops* species.

Spike (head) morphology illustrating differences in spike length, spikelet compactness, and awn characteristics. These features reflect the broad phenotypic diversity present within the genus.

Numerous *Aegilops* species have been exploited as sources of resistance to major biotic stresses in wheat. For example, *Ae. umbellulata*, *Ae. speltoides*, *Ae. kotschyi*, *Ae. neglecta*, *Ae. geniculata* and *Ae. peregrina* have contributed resistance to fungal diseases including leaf, stem, and stripe rusts and powdery mildew, highlighting the importance of *Aegilops* as a reservoir of disease resistance genes (Bansal et al., 2017; Liu et al., 2011; Marais et al., 2009; Marais et al., 2008; Marais et al., 2005; Petersen et al., 2015). In addition to biotic stress resistance, *Aegilops* species also harbor substantial variation for tolerance to abiotic stresses. Several species, including *Ae. speltoides*, *Ae. tauschii*, and *Ae. geniculata* have been associated with enhanced tolerance to drought (Baalbaki et al., 2006; Zaharieva et al., 2001), while tolerance to salinity has been reported in species such as *Ae. tauschii* and *Ae. neglecta* (Ahmadi et al., 2020). Variation for heat tolerance affecting grain yield has been observed among *Aegilops* species, with *Ae. geniculata* and *Ae.*

speltoides showing greater tolerance to elevated temperatures (Pradhan et al., 2012), whereas cold tolerance has been documented in species such as *Ae. tauschii* (Masoomi-Aladizgeh et al., 2015). Together, these examples illustrate the diversity of adaptive traits present in *Aegilops* species, underscoring their value for improving wheat resilience under diverse environmental conditions.

1.1.5 The concept of gene pools

A gene pool refers to the total genetic diversity present within a species or group of related species and plays a key role in their ability to adapt to environmental changes and disease pressures (Singh et al., 2021). In wheat, the concept of gene pools is commonly used to describe the evolutionary and breeding relationships among species within the Triticeae tribe which comprises approximately 20-27 genera, including *Triticum*, *Aegilops*, *Hordeum* (includes barley), and *Secale* (includes rye), which together represent important sources of genetic variation for wheat improvement (Kellogg, 2015; Soreng et al., 2015).

Based on genome composition and crossability with cultivated wheat, relatives are classified into primary, secondary, and tertiary gene pools. The primary gene pool includes species containing the A, B, or D genomes and combinations thereof, encompassing cultivated hexaploid and tetraploid wheats as well as their diploid genome donors (Chaudhary et al., 2013). The secondary gene pool consists of species that carry at least one A, B, or D genome together with additional closely related genomes. This group includes species of the Sitopsis section: *Ae. bicornis* (S^b), *Ae. sharonensis* (S^{sh}), *Ae. longissima* (S^l), *Ae. searsii* (S^s), and *Ae. speltoides* (S) (Chaudhary et al., 2013; Harlan, 1992). The tertiary gene pool comprises more distantly related Triticeae species that lack the A, B, or D genomes (Chaudhary et al., 2013). *Aegilops* species are represented across all three wheat gene pools (Figure 1.4) and exhibit extensive genomic diversity (Figures 1.2 and 1.3). This diversity makes *Aegilops* a particularly valuable genetic reservoir for

wheat improvement, and considerable efforts have been made to assess and exploit their breeding potential to enhance wheat productivity and resilience (Kilian et al., 2011).

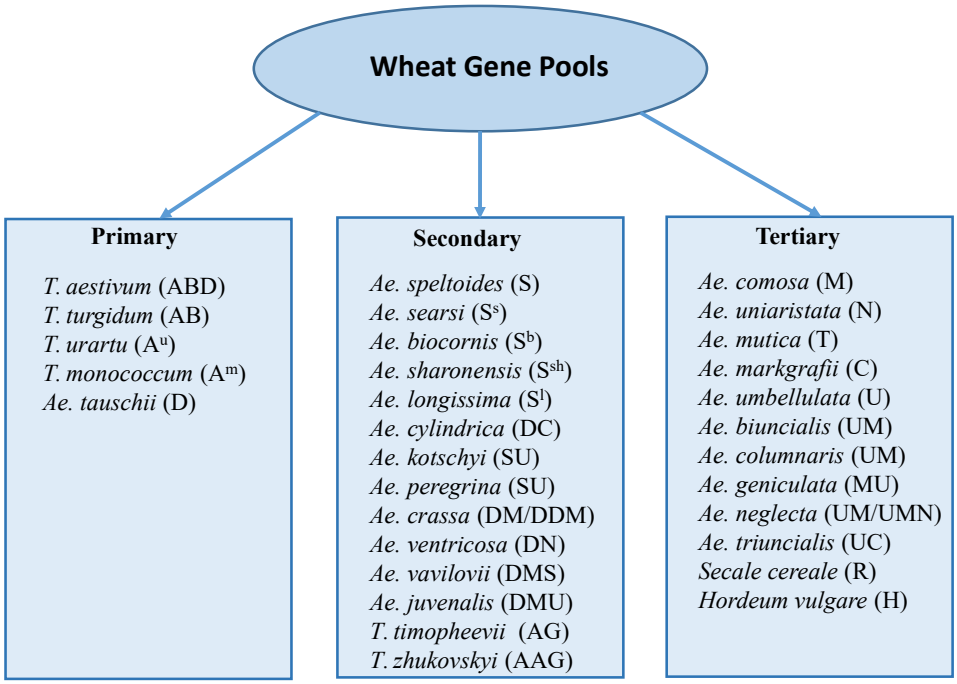


Figure 1.4 The gene pools of wheat.

1.1.6 Challenges in utilizing wild genetic diversity for wheat improvement

The successful use of genetic diversity from wheat wild relatives requires not only the identification of beneficial genes but also their effective transfer into cultivated wheat. Conventional wheat breeding primarily relies on wheat x wheat crosses; however, interspecific hybridization with species from the primary, secondary, and tertiary gene pools offers opportunities to introduce novel or previously lost genetic variation. Despite its potential, gene transfer from wild relatives such as *Aegilops* remains challenging due to multiple biological barriers, including asynchronous flowering, poor pollen viability, sexual incompatibility, hybrid sterility, the frequent need for embryo rescue, and the persistence of linkage drag (Dempewolf et al., 2012).

Recent advances in genomics and molecular technologies are helping to overcome these limitations by enabling more precise characterization of genetic diversity across large germplasm collections and facilitating the identification of candidate genes and genomic regions associated with desirable traits. These developments have the potential to accelerate the discovery and deployment of beneficial alleles from *Aegilops* and other wild relatives, thereby enhancing the efficiency of wheat improvement programs (Langridge and Waugh, 2019; Milner et al., 2019; Müller et al., 2018).

1.1.7 Structural Variation (SV): types and evolutionary significance

Genetic variation in genomes occurs at multiple scales, ranging from single nucleotide substitutions to large chromosomal rearrangements and whole genome duplications (Vihinen, 2018). While single nucleotide polymorphisms (SNPs) have been extensively characterized due to their ease of detection using short-read sequencing technologies, they do not capture the full spectrum of genomic variation underlying phenotypic diversity (Butler, 2012). In contrast, structural variations (SVs) represent a major but historically underexplored source of genetic diversity (Springer et al., 2009).

SVs are changes that affect long DNA sequences (> 50 base pairs), including deletions, insertions, duplications, inversions and translocations (Gorkovskiy and Verstrepen, 2021). Insertions, duplications, and deletions can generate copy number variations (CNVs) and presence-absence variations (PAVs), which alter gene content. SVs may affect both coding and non-coding regions of the genome, and can lead to phenotypic changes by modifying gene structure, gene copy number, and in some cases by generating novel genes (Gamazon and Stranger, 2015; Huang et al., 2018; Lavington and Kern, 2017; Radke and Lee, 2015; Stewart and Rogers, 2019). In plants, SVs play important roles in genome evolution, adaptation, and agronomically relevant traits, including

responses to biotic and abiotic stresses (Lye and Purugganan, 2019). In wheat, CNVs affecting key developmental genes such as *Ppd-B1*, *Vrn-A1*, and *Rht-D1* influence photoperiod response, vernalization requirement, and plant height, respectively, highlighting the functional importance of structural variation in agronomically relevant traits (Díaz et al., 2012; Li et al., 2012; Würschum et al., 2015). Similarly, deletions of large gene clusters at the *FR-2* locus, including *CBF-B12*, *CBF-B14*, and *CBF-B15*, have been shown to reduce cold tolerance in durum and bread wheats, providing further evidence for the role of SVs in abiotic stress adaptation (Díaz et al., 2012; Knox et al., 2010; Pearce et al., 2013).

In the Triticeae, large, repeat-rich genomes, together with a history of polyploidization, contribute to the high prevalence and evolutionary importance of structural variations. Studying SVs in *Aegilops* species can therefore improve our understanding of genome divergence and speciation and reveal genetic variation with potential value for wheat improvement.

1.1.8 Transposable elements (TEs) as drivers of plant genome evolution

Transposable elements (TEs) are ubiquitous components of eukaryotic genomes and represent major drivers of genome evolution (Feschotte et al., 2002). TEs are DNA sequences capable of moving or duplicating within the genome and are broadly classified into two main classes based on their transposition mechanisms. Class I retrotransposons transpose via a 'copy-and-paste' mechanism involving an RNA intermediate, whereas Class II DNA transposons move through a 'cut-and-paste' mechanism (Figure 1.5) (Wicker et al., 2007).

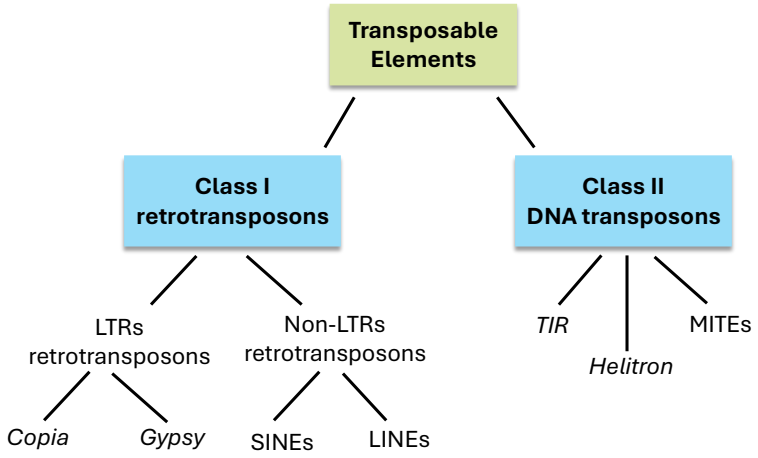


Figure 1.5 Classification of transposable elements in plant genomes.

Abbreviations: SINEs, short interspersed nuclear elements; LINEs, long interspersed nuclear elements; TIR, terminal inverted repeat; MITEs, miniature inverted-repeat transposable elements.

In plant genomes, TEs constitute a substantial proportion of the total genome content, ranging from approximately 14% to 84%, with Class I retrotransposons accounting for more than 80% of TEs in many species, especially in monocots (Ragupathy et al., 2013). Among Class I elements, long terminal repeat (LTR) retrotransposons are abundant and widespread in plant genomes and account for up to 70% of all retrotransposons in previously characterized *Aegilops* and *Triticum* species (Table 1.1) (You et al., 2015).

Table 1.1 Transposable element composition in *Aegilops* and *Triticum* genomes

Species	Genome	Genome size (Gb)	TE (%)	Class I LTR (%)	References
Ae. tauschii	D	>4.2	86.88	64.39	Wang et al. (2021)
Ae. bicornis	S^b	5.73	85.99	66.29	Li et al. (2022)
Ae. longissima	S^l	6.22	88.08	70.37	Li et al. (2022)
Ae. searsii	S^s	5.55	86.86	65.40	Li et al. (2022)
Ae. sharonensis	S^{sh}	6.07	88.11	70.30	Li et al. (2022)
Ae. speltoides	S	4.60	86.13	69.96	Li et al. (2022)
Ae. ventricosa	DN	8.67	83.52	67.21	Liu et al. (2025)
T. urartu	A	4.94	81.42	68.61	Ling et al. (2018)
T. turgidum	AB	10.45	82.20	70.00	Maccaferri et al. (2019); Avni et al. (2017)
T. aestivum	ABD	14.57	85.00	65.86	IWGSC (2021)

LTR retrotransposons are characterized by the presence of two long terminal repeat sequences flanking an internal coding region (Figure 1.6) (Makałowski et al., 2019). A key feature of LTR retrotransposons is that the sequence of the two LTRs are identical at the time of insertion but gradually accumulate mutations over time. This property has been widely exploited to estimate the insertion time of individual LTR retrotransposons using a molecular clock approach based on nucleotide divergence between LTR pairs (Dangel et al., 1995; SanMiguel et al., 1998; SanMiguel et al., 1996). Assuming a known substitution rate, the divergence between LTR pairs provides an estimate of the time of insertion of retrotransposons and allows reconstruction of past transpositional activity (Kijima and Innan, 2010; Ma et al., 2004).

Analysis of LTR retrotransposons and their insertion dynamics has provided valuable insights into genome evolution, particularly in large and polyploid plant genomes. In *Aegilops*, examining the abundance, classification, and time of insertion of LTR retrotransposons provide an effective framework for understanding genome expansion, lineage-specific repeat activity, and the evolutionary processes shaping these diploid and polyploid genomes.

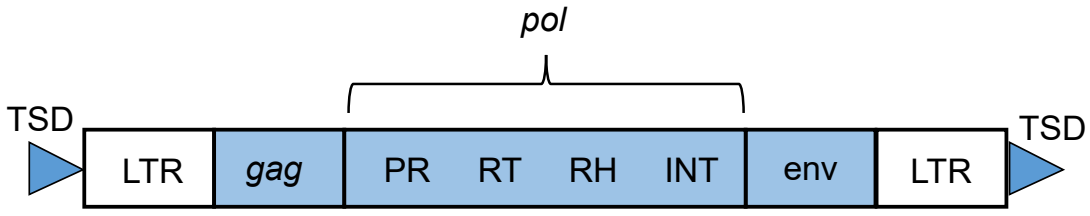


Figure 1.6 Structure of long terminal repeat (LTR) retrotransposon.

Abbreviations: *gag*, group-specific antigen; *pol*, polyprotein; PR, protease; RT, reverse transcriptase; RH, ribonuclease H; INT, integrase; env, envelope; TSD, target site duplication.

1.1.9 Challenges in *Triticum* and *Aegilops* genomics

Advances in genomics have accelerated genetic research and crop improvement across many plant species; however, progress in wheat has historically lagged behind other major crops. This disparity is primarily due to the size and complexity of the wheat genome (IWGSC, 2018). Bread wheat has an estimated genome size of ~16–17 gigabases (Gb) and is an allohexaploid composed of three subgenomes, each contributing approximately 5.5 Gb. In addition, more than 80% of the genome consists of repetitive sequences, predominantly transposable elements (Marcussen et al., 2014). These features have long posed major challenges for genome sequencing, assembly, and downstream analyses.

Comparable challenges extend to wheat's wild relatives, including species of the genus *Aegilops*. Although most sequenced *Aegilops* genomes are diploid—with the exception of the tetraploid *Aegilops ventricosa*—they remain large (approximately 4.5–12 Gb) and are similarly enriched in repetitive DNA (Avni et al., 2022; Li et al., 2022; Liu et al., 2025; Luo et al., 2017; Wang et al., 2021). These features have historically limited the resolution and accuracy of genome-wide analyses, particularly for detecting structural variations (SVs), which play critical roles in genome evolution, adaptation, and phenotypic diversity (Huddleston and Eichler, 2016; Wellenreuther et al., 2019; Wolf and Ellegren, 2017).

Structural variations are inherently difficult to identify using short-read sequencing technologies because short reads do not span large insertions, deletions, inversions, and other complex rearrangements, especially within repeat-rich regions, and their assemblies often fail in resolving them (Chaisson et al., 2015; Sedlazeck et al., 2018). As a result, reliable detection of SVs requires highly contiguous and accurate genome assemblies that preserve long-range genomic structure (Peona et al., 2021; Peona et al., 2018; Weckselblatt and Rudd, 2015). Until recently, the

limited availability of such high-quality reference assemblies has constrained comprehensive investigations of SVs and transposable elements in both *Triticum* and *Aegilops* genomes.

Recent advances in long-read sequencing technologies, particularly Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), combined with chromosome-scale scaffolding approaches such as Hi-C and related methods, can overcome these limitations. These technologies enable the generation of highly contiguous, pseudomolecule-scale genome assemblies, even for large and repeat-rich plant genomes. Consequently, they provide unprecedented opportunities to characterize structural variations, transposable element dynamics, and genome organizations on a genome-wide scale, thereby deepening our understanding of genome evolution and diversification in wheat and its wild relatives.

1.2 Purpose of the study

Several diploid *Aegilops* genome assemblies have recently been produced, providing valuable insights into genome structure and diversity (Abrouk et al., 2023; Avni et al., 2022; Grewal et al., 2025; Li et al., 2024; Li et al., 2022; Luo et al., 2017; Wang et al., 2021). However, comparable genomic resources for polyploid *Aegilops* species is unavailable, except for *Aegilops ventricosa* (Liu et al., 2025). Genome assemblies for all other tetraploid and hexaploid *Aegilops* species remain unavailable, which limits our understanding of genome evolution following polyploidization and constrains the effective utilization of wild genetic diversity for wheat improvement. This study aims to address this gap by first generating high-quality genome assemblies for 18 *Aegilops* species, including four diploid species and all fourteen tetraploid and hexaploid species within the genus, thereby completing the sequencing of all 25 species of the *Aegilops* genus.

Using these genomic resources, genome evolution in *Aegilops* is then examined through comparative analyses of structural variations, transposable element dynamics, and polyploid divergence. Collectively, this work provides a comprehensive genomic framework for understanding *Aegilops* genome diversity and supports the informed utilization of genetic variation from wild relatives in wheat improvement.

1.3 Objectives

The specific objectives of this study include:

1. Generate high-quality genome assemblies for diploid and polyploid *Aegilops* species.
2. Perform genome annotation across all *Aegilops* assemblies to enable comparative analyses.
3. Infer phylogenetic relationships and estimate divergence times among *Aegilops* species.
4. Identify and characterize structural variations among *Aegilops* species, with emphasis on comparisons between polyploid genomes and their diploid progenitors.
5. Analyze the composition, abundance, and evolutionary dynamics of transposable elements across *Aegilops* genomes and assess patterns of conservation across ploidy levels and among species.

1.4 Hypotheses

1. Structural variation contributes to genome diversification and evolutionary divergence in *Aegilops* species.

2. *Aegilops* subgenomes follow different evolutionary trajectories after polyploidization, leading to variation in genome size, gene content, and genome structure.

Chapter 2

Evolutionary dynamics of *Aegilops* revealed through comparative genome assembly of all 25 species

Hamna Shazadee[1,2], Tara Edwards[2], Madeleine Lévesque-Lemay[2], Chunfang Zheng[2], Jennifer Ens[3], Curtis J. Pozniak[3], Frank M. You[2,*], Sylvie Cloutier[1,2,*]

[1]Department of Biology, University of Ottawa, Ottawa, ON, K1N 6N5, Canada

[2]Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, K1A 0C6, Canada

[3]Crop Development Centre, University of Saskatchewan, Saskatoon, SK, S7N 5A8, Canada

*Correspondence: Sylvie Cloutier and Frank M. You

This chapter has been accepted with revisions for publication in *Nature Communications* on 4 August 2026.

Author contributions

H.S. performed data analyses and interpretation, generated figures, and drafted the manuscript. T.E., M.L.L., and J.E. performed extraction, library construction, and data management. C.Z. performed data management. F.M.Y., C.J.P., and S.C. performed project management, study design, data generation, and manuscript editing. All co-authors read and approved the final manuscript.

2.1 Abstract

Aegilops species are the closest wild relatives of wheat and an important reservoir of genetic diversity for its improvement. Despite their potential, many *Aegilops* genomes remain poorly characterized. Here we present high-quality assemblies of 18 diploid, tetraploid, and hexaploid *Aegilops* genomes, which, along with the previously published genomes, complete the production of reference assemblies for all 25 genomes in this genus. Assembly sizes ranged from 5.24 Gb in diploids to 12.65 Gb in hexaploids, with scaffold N50 values up to 749.2 Mb. Gene annotation identified 53,035-156,779 protein-coding genes, of which 21,865-60,490 were classified as high-confidence. Orthogroup-based pangenome analysis across the 25 *Aegilops* genomes identified 80,521 orthogroups, including 15,809 core, 61,735 dispensable, and 2,977 species-specific orthogroups, highlighting substantial gene content variation among genomes. Phylogenetic analysis of 63 *Triticum* and *Aegilops* genomes/subgenomes based on near single-copy orthologs defines the phylogenetic relationships within the *Triticum/Aegilops* complex and supported the proposed diploid progenitor relationships of polyploid lineages. *Ae. mutica* (T) and *Ae. speltoides* (S) belong to the B lineage while the remaining Sitopsis grouped within the D lineage. Structural variation analyses using diploid progenitors as references revealed extensive large-scale rearrangements following polyploidization, emphasizing the dynamics of their evolution. Transposable element (TE) annotation further highlighted subgenome-specific TE expansions and contractions, providing insights into the mechanisms shaping genome structure after polyploidization. Collectively, these genomic resources provide a comprehensive framework for exploring *Aegilops* diversity, understanding polyploid evolution, and accelerating wheat improvement.

Key words: *Aegilops*, genome assembly, polyploid, structural variation, transposable elements

2.2 Introduction

Bread wheat (*Triticum aestivum* L.) is a major staple crop, providing approximately 20% of global calories and proteins of the Human population (Gustafsson et al., 2009). As the human population continues to grow, wheat production must increase by more than 50% over current levels by 2050 to meet future food demands (Tadesse et al., 2019). However, increasing productivity while maintaining food safety poses a significant challenge to scientists as agriculture simultaneously faces climate change, evolving pathogens, and environmental stressors. Wheat's genetic diversity has been reduced by successive bottlenecks during its evolution, especially through polyploidization, domestication, and intensive modern breeding. Overcoming these limitations may be key to addressing major challenges in wheat production (Cavanagh et al., 2013; Charmet, 2011; Dubcovsky and Dvorak, 2007)

Wild relatives, particularly *Aegilops* species, offer a rich source of genetic variation (Dempewolf et al., 2017; Schneider et al., 2008). The genus includes 25 species across eleven genome types (D, S, S^b, S^l, S^{sh}, S^s, U, C, N, M and T), with polyploids arising via interspecific hybridization and chromosome doubling (Kilian et al., 2011; Van Slageren, 1994). Following polyploidization, genomes undergo structural changes, gene expansion or loss, and subgenome divergence, producing genomic structures distinct from diploid progenitors despite sharing genome labels (Kilian et al., 2011; Van Slageren, 1994). While phylogenetic relationships have been studied using morphological and molecular data (Dvorak et al., 1998; Dvorak et al., 1993; Petersen et al., 2006; Van Slageren, 1994), genome-scale analyses of structural variation and polyploid evolution remain limited. *Aegilops* species have played a pivotal role in wheat evolution, with *Ae. tauschii* donating the D subgenome (Kihara, 1944; McFadden and Sears, 1946; Salamini et al., 2002), and a species related to *Ae. speltoides* contributing the B subgenome (Li et al., 2022;

Marcussen et al., 2014). They also provide valuable alleles for disease resistance and stress tolerance (Colmer et al., 2006; Helguera et al., 2003; Nevo and Chen, 2010; Olivera et al., 2018; Weidner et al., 2012), though their use in breeding is limited by crossing barriers and linkage drag.

Transposable elements (TEs), especially long-terminal repeat (LTR) retrotransposons (*Copia* and *Gypsy*), dominate large grass genomes and contribute to genome size variation, structural remodeling, and gene regulation (Bennetzen and Wang, 2014; Feschotte et al., 2002; Wicker et al., 2018; Wicker et al., 2007). TE dynamics often shift following polyploidization due to genomic shock, relaxed selection, or epigenetic reprogramming (Bennetzen and Wang, 2014; de Tomás and Vicient, 2024; Vicient and Casacuberta, 2017). Despite their importance, the role of TEs in subgenome divergence across *Aegilops* has not been studied because genome assemblies were not available for all species.

The genome assembly of several diploid (7 chromosomes) species have been sequenced, including *Ae. tauschii* (D), *Ae. speltoides* (S), *Ae. longissima* (S^l), *Ae. bicornis* (S^b), *Ae. searsii* (S^s), *Ae. sharonensis* (S^{sh}), *Ae. umbellulata* (U), *Ae. comosa* (M), and *Ae. mutica* (T) (Abrouk et al., 2023; Avni et al., 2022; Grewal et al., 2025; Li et al., 2024; Li et al., 2022; Luo et al., 2017; Wang et al., 2021). Our lab has recently completed the assemblies of the remaining diploid species, including *Ae. uniaristata* (N) and *Ae. markgrafii* (C) (Khadka et al. 2026, unpublished). However, the only genome assembly of polyploid (14 and 21 chromosomes) species published to date is that of *Ae. ventricosa* (DN) (Liu et al., 2025), leaving a knowledge gap in polyploid *Aegilops* structural genomics.

To address this gap, we generated high-quality assemblies for 18 *Aegilops* species (4 diploid, 10 tetraploid and 4 hexaploid) using Oxford Nanopore, PacBio, Illumina, Hi-C, and Omni-C technologies. Consistent machine learning–based gene annotation, phylogenomic analyses,

structural variation, and LTR characterization enabled a comprehensive view of subgenome evolution, TE dynamics, and divergence from diploid progenitors. Collectively, this genomic resource provides a foundation for understanding *Aegilops* genome evolution and exploiting wild diversity for wheat improvement.

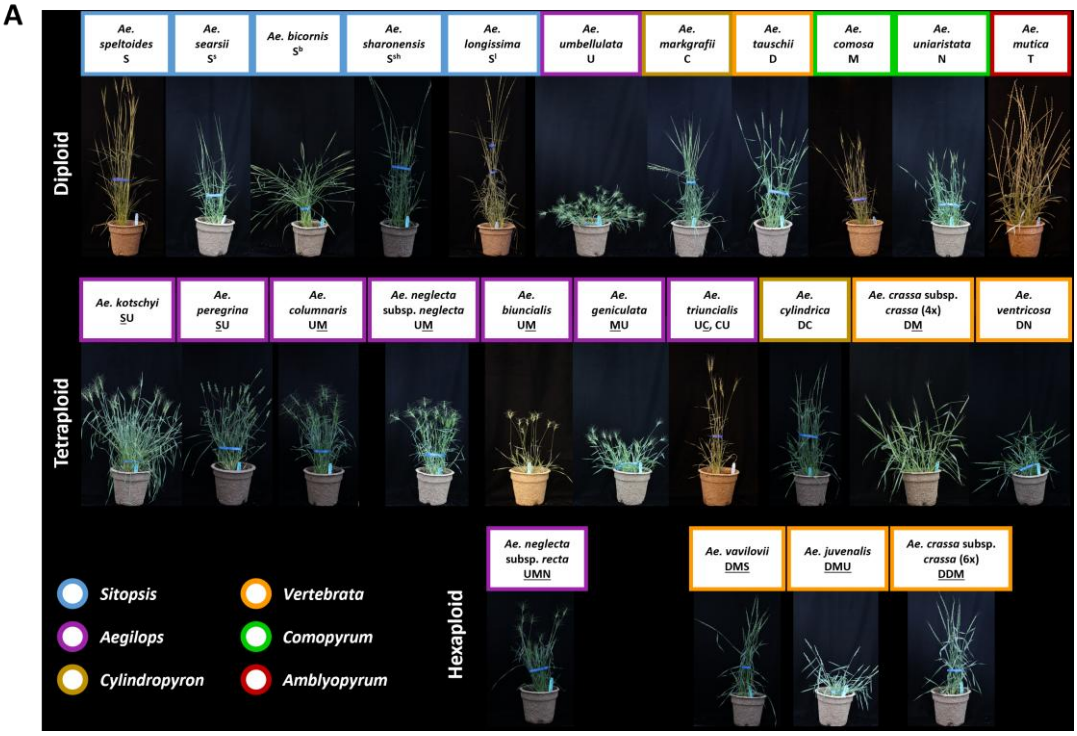
2.3 Results

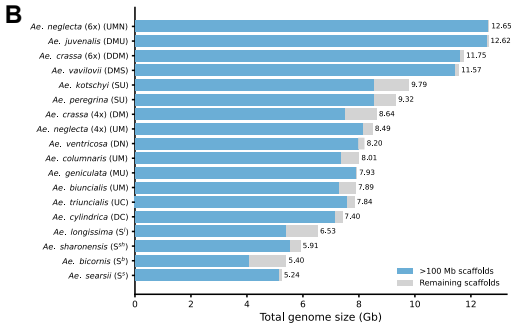
2.3.1 Genome assembly

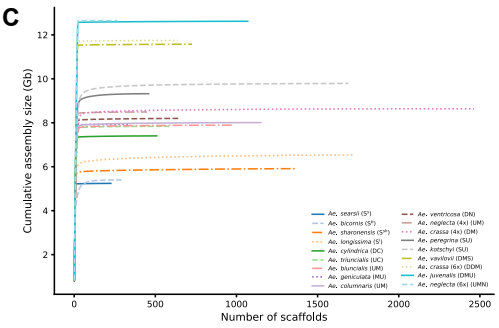
Aegilops species exhibit substantial morphological diversity at the whole-plant level, reflecting their wide evolutionary range across diploid, tetraploid, and hexaploid lineages (Figure 2.1A). To investigate the genomic basis underlying this diversity, we generated high-quality genome assemblies for 18 *Aegilops* species, representing four of the 11 diploid, all ten tetraploid, and all four hexaploid species, using a single representative accession for each species (Appendix 1). Assemblies were produced using Oxford Nanopore Technologies (ONT, Oxford, UK) long reads, polished with PacBio (Pacific Biosciences, Menlo Park, CA, USA) HiFi long reads and Illumina (San Diego, CA, USA) short reads, and scaffolding with Hi-C and Omni-C (Dovetail Genomics, Scotts Valley, CA, USA) short-read data. Sequencing coverage for each platform and genome assembly is summarized in Appendix 2. Diploid genome sizes ranged from 5.24 Gb (*Ae. searsii*) to 6.53 Gb (*Ae. longissima*), with N50 values of 560.0–749.2 Mb and N90 values of 23.2–397.7 Mb (Appendix 3). Scaffold numbers ranged from 226 to 1,720, with 8–10 scaffolds exceeding 100 Mb (Figures 2.1B, 2.1C and Appendix 3). Tetraploid genomes ranged from 7.40 Gb (*Ae. cylindrica*) to 9.79 Gb (*Ae. kotschyi*), with N50 values of 367.2–644.9 Mb and N90 values of 62.5–339.7 Mb (Figure 2.1B and Appendix 3). Scaffold counts ranged from 349 to 2,463, with 14–24 scaffolds >100 Mb (Figures 2.1B-2.1C and Appendix 3). Hexaploid assemblies ranged from 11.57 Gb (*Ae. vavilovii*) to 12.65 Gb (*Ae. neglecta* subsp. *recta*), with N50 values of 623.8–705.7 Mb

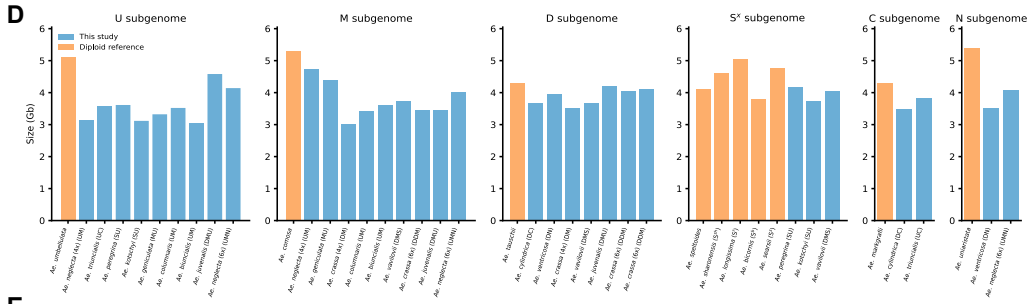
and N90 values of 316.1–390.8 Mb (Figure 2.1B and Appendix 3). Scaffold numbers ranged from 266 to 1,070, with 21–23 scaffolds >100 Mb (Figures 2.1B-2.1C and S Appendix 3). Scaffolds >100 Mb accounted for 75.62–98.27% of diploid assemblies, 86.98–99.79% of tetraploid assemblies, and 98.95–99.89% of hexaploid assemblies, whereas scaffolds <100 Mb contributed the remaining 1.73–24.38%, 0.21–13.02%, and 0.11–1.05% of the total assembly size, respectively, illustrating the high contiguity of all genome assemblies. Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis using the embryophyta_odb10 dataset (Manni et al., 2021) showed high completeness (98.4–99.8%), supporting the overall quality of the assemblies (Appendices 3 and 4). In addition, LTR Assembly Index (LAI) scores ranged from 11.5 to 24.5 across assemblies, indicating reference-quality representation of repetitive regions further supporting the overall quality of the genome assemblies (Appendix 3).

Comparative analyses of genome size between polyploid subgenomes and their corresponding diploid reference genomes revealed a consistent reduction in subgenome size following polyploidization, indicating genome contraction relative to their putative progenitors (Figure 2.1D). A similar trend was observed when comparing mean genome sizes across ploidy levels, where diploid genomes were consistently larger, whereas tetraploid and hexaploid subgenomes exhibited reduced average genome sizes across all genome groups (U, M, D, S^x, C and N) (Figure 2.1E). Together, these results indicate that the generated assemblies achieved high contiguity and completeness across all ploidy levels while also revealing consistent patterns of genome downsizing associated with polyploid evolution in *Aegilops*.


A
Ae. speltoides S
Ae. searsii S^s
Ae. bicornis S^b
Ae. sharonensis S^sh
Ae. longissima S^l
Ae. umbellulata U
Ae. markgrafii C
Ae. tauschii D
Ae. comosa M
Ae. uniaristata N
Ae. mutica T
Diploid
Ae. kotschyi SU
Ae. peregrina SU
Ae. columnaris UM
Ae. neglecta subsp. neglecta UM
Ae. biuncialis UM
Ae. geniculata MU
Ae. triuncialis UC, CU
Ae. cylindrica DC
Ae. crassa subsp. crassa (4x) DM
Ae. ventricosa DN
Tetraploid
Ae. neglecta subsp. recta UMN
Ae. vavilovii DMS
Ae. juvenalis DMU
Ae. crassa subsp. crassa (6x) DDM
Hexaploid
Sitopsis
Aegilops
Cylindropyron
Vertebrata
Comopyrum
Amblyopyrum


B
Ae. neglecta (6x) (UMN) 12.65
Ae. juvenalis (DMU) 12.62
Ae. crassa (6x) (DDM) 11.75
Ae. vavilovii (DMS) 11.57
Ae. kotschyi (SU) 9.79
Ae. peregrina (SU) 9.32
Ae. crassa (4x) (DM) 8.64
Ae. neglecta (4x) (UM) 8.49
Ae. ventricosa (DN) 8.20
Ae. columnaris (UM) 8.01
Ae. geniculata (MU) 7.93
Ae. biuncialis (UM) 7.89
Ae. triuncialis (UC) 7.84
Ae. cylindrica (DC) 7.40
Ae. longissima (S^l) 6.53
Ae. sharonensis (S^sh) 5.91
Ae. bicornis (S^b) 5.40
Ae. searsii (S^s) 5.24
>100 Mb scaffolds
Remaining scaffolds
Total genome size (Gb)


C
Cumulative assembly size (Gb)
Number of scaffolds


D
U subgenome
M subgenome
D subgenome
S^x subgenome
C subgenome
N subgenome
This study
Diploid reference
Size (Gb)

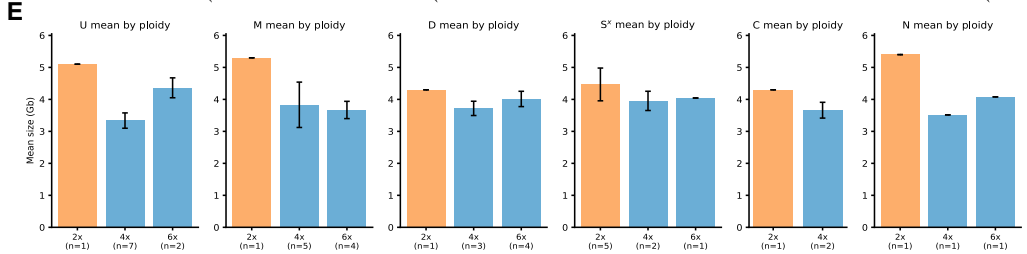

E
U mean by ploidy
M mean by ploidy
D mean by ploidy
S^x mean by ploidy
C mean by ploidy
N mean by ploidy
Mean size (Gb)
2x (n=1)
4x (n=7)
6x (n=2)
2x (n=1)
4x (n=5)
6x (n=4)
2x (n=1)
4x (n=3)
6x (n=4)
2x (n=5)
4x (n=2)
6x (n=1)
2x (n=1)
4x (n=2)
2x (n=1)
4x (n=1)
6x (n=1)

Figure 2.1 Morphological diversity and genome assembly statistics across *Aegilops* species.

(A) Whole-plant morphology showing variation in plant architecture across diploid, tetraploid, and hexaploid *Aegilops* species, highlighting extensive phenotypic diversity across the genus.

(B) Total genome size estimates for each species, with contributions from scaffolds >100 Mb shown alongside remaining scaffolds.

(C) Assembly contiguity illustrated by the relationship between cumulative assembly size and number of scaffolds across genomes.

(D) Comparative genome size variation across subgenomes (U, M, D, S^x, C and N) in polyploid *Aegilops* species relative to their corresponding diploid progenitors, highlighting subgenome-specific expansion and contraction patterns.

(E) Mean genome size variation across ploidy levels (diploid, tetraploid, and hexaploid) for each subgenome group, showing trends in genome size evolution following polyploidization.

2.3.2 Gene annotation and orthogroup-based pangenome analysis

The total numbers of protein-coding genes annotated increased with ploidy, ranging from 53,035–59,817 in diploids, 88,422–109,856 in tetraploids and 143,429–156,779 in hexaploids. *Ae. crassa* subsp. *crassa* (4x) was an outlier with 149,108 genes. HC gene counts ranged from 21,865–27,783 (diploids), 33,590–60,490 (tetraploids), and 44,149–55,945 (hexaploids). BUSCO completeness scores (84.5%–98.9%) indicate that most core genes were captured (Figure 2.2A and Appendix 5).

To further resolve gene content evolution at the subgenome level, we compared total protein-coding gene numbers in each polyploid subgenome against their corresponding diploid progenitors (Figure 2.2B). This revealed highly heterogeneous patterns of gene retention and loss across subgenomes following polyploidization. The U subgenome showed both expansion and contraction relative to its diploid reference, indicating lineage-dependent gene content evolution rather than a uniform trend. In contrast, the M subgenome generally exhibited gene loss across most polyploid species, with notable exceptions of gene expansion in *Ae. crassa* subsp. *crassa*

(4x) and *Ae. neglecta* subsp. *neglecta*. The D subgenome displayed the highest variability among all subgenomes. Strong gene expansion was observed exclusively in *Ae. crassa* subsp. *crassa* (4x), whereas all other comparisons showed reduced gene numbers relative to diploid *Ae. tauschii*. This pattern highlights highly dynamic and lineage-specific evolutionary trajectories within the D subgenome. In comparison, the S subgenome remained largely unchanged, with gene contents closely matching its diploid counterpart. Similarly, the C subgenome showed high conservation, with *Ae. cylindrica* maintaining nearly identical gene numbers to diploid *Ae. markgrafii*, while *Ae. triuncialis* exhibited a moderate expansion. The N subgenome showed intermediate behaviour, with gene loss in *Ae. ventricosa* and gene expansion in *Ae. neglecta* subsp. *recta*, indicating moderate but lineage-specific variation without a consistent directional trend (Figure 2.2B). Overall, these results indicate that gene content evolution in *Aegilops* polyploids is strongly subgenome-dependent, with distinct evolutionary trajectories ranging from high conservation (S and C subgenomes), moderate plasticity (N and U subgenomes), to pronounced structural and gene content remodeling (D and M subgenomes).

Orthogroup-based pangenome analysis further resolved gene content variation across the 25 *Aegilops* genomes. A total of 80,521 orthogroups were identified, including 15,809 (19.63%) core orthogroups shared across all genomes, 61,735 (76.67%) dispensable orthogroups present in 2–24 genomes, and 2,977 (3.70%) species-specific orthogroups, unique to individual genomes where the latter ranged from 16 in *Ae. sharonensis* to 548 in *Ae. mutica* (Figure 2.2C and Appendix 6). These orthogroups corresponded to 18,587–55,519 core genes, 24,913–95,073 dispensable genes, and 36–2,283 species-specific genes per genome (Figure 2.2D and Appendix 6) with both core and dispensable gene fractions generally increasing with ploidy level.

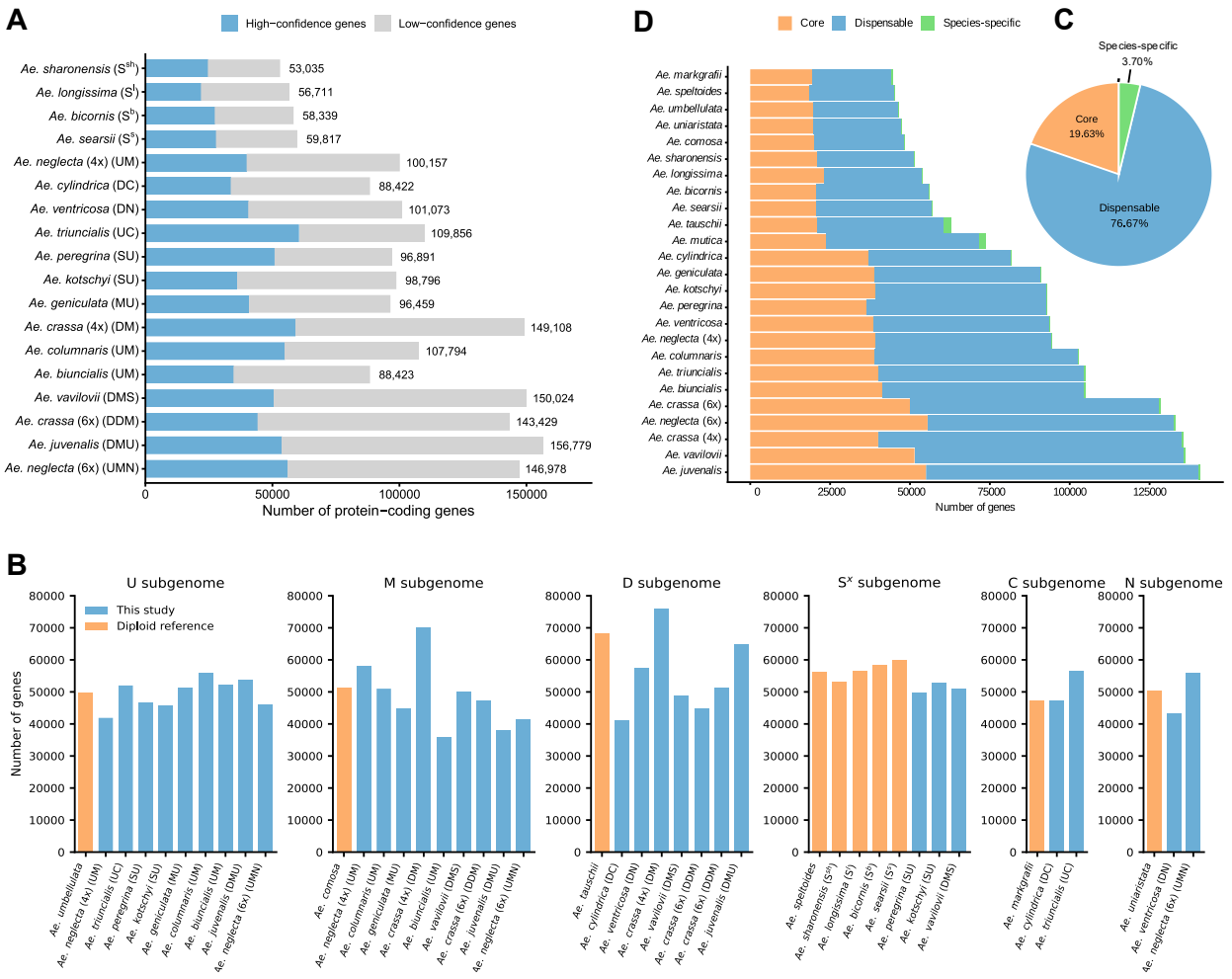


Figure 2.2 Gene annotation statistics and pan-genome analysis of *Aegilops* species based on protein-coding genes.

(A) Bar plot showing the total number of protein-coding genes in each species, including high-confidence and low-confidence gene models, highlighting variation in gene content across the genus.

(B) Comparison of total protein-coding gene numbers across subgenome groups (U, M, D, S^x, C and N) relative to their corresponding diploid progenitor species, illustrating changes in gene content associated with polyploidization.

(C) Pie chart showing the proportions of core, dispensable, and species-specific orthogroups across the *Aegilops* pan-genome, summarizing gene conservation and diversification across species.

(D) Distribution of core, dispensable, and species-specific genes across individual genomes, highlighting variation in gene sharing and lineage-specific gene content. The following species names were shortened: *Ae. neglecta* subsp. *neglecta* is shown as *Ae. neglecta* (4x); *Ae. neglecta*

subsp. *recta* as *Ae. neglecta* (6x); *Ae. crassa* subsp. *crassa* (4x) as *Ae. crassa* (4x); *Ae. crassa* subsp. *crassa* (6x) as *Ae. crassa* (6x).

2.3.3 Phylogenetic relationships and divergence time

Phylogenetic relationships were inferred from near single-copy orthologs (773) across 36 *Aegilops* genomes and subgenomes generated in this study, together with 27 previously published *Aegilops* and *Triticum* genomes/subgenomes. *Brachypodium distachyon*, *Hordeum vulgare* and *Thinopyrum elongatum* were included as outgroups to root the tree. The resulting phylogeny resolved A, B, D, S^x (Sitopsis) N, M, C, and U clades (Figure 2.3). Within the A clade, the A genomes of *T. urartu* and *T. monococcum* clustered with the A subgenomes of polyploid *Triticum* species. The B genome clade included the B subgenomes of *T. aestivum* and *T. turgidum*, the S genome of the *Ae. speltoides*, the T genome of *Ae. mutica* and the G genome of *T. timopheevii*. Most polyploid *Aegilops* subgenomes clustered with their expected diploid progenitors. For example, the D subgenomes grouped with the D genome of *Ae. tauschii* and D subgenome of *Triticum aestivum*, reflecting their shared ancestry. The N subgenomes of *Ae. ventricosa* (DN) and *Ae. neglecta* subsp. *recta* (UMN) grouped with *Ae. uniaristata* (N), while the C subgenomes of *Ae. triuncialis* (UC) and *Ae. cylindrica* (DC) clustered with *Ae. markgrafii* (C). Similarly, all U subgenomes formed a well-supported group with *Ae. umbellulata* (U). The "S" subgenomes of *Ae. kotschyi* and *Ae. peregrina* clustered within the Sitopsis group, close to *Ae. longissima* (S^l) and *Ae. sharonensis* (S^{sh}), hinting that one or the other are the most likely progenitors of these species rather than *Ae. speltoides* (Figure 2.3). These concordant groupings confirm the involvement of the corresponding diploid species as progenitors of the respective subgenomes in polyploid *Aegilops* species.

Conversely, the M subgenomes were more phylogenetically dispersed. Only the M subgenomes of *Ae. geniculata* (MU) and *Ae. biuncialis* (UM) clustered with *Ae. comosa* (M). The

M subgenomes of *Ae. neglecta* subsp. *neglecta* (UM), *Ae. neglecta* subsp. *recta* (UMN) and *Ae. columnaris* (UM) clustered within the U-genome clade, while that of *Ae. crassa* subsp. *crassa* (4x, DM), *Ae. crassa* subsp. *crassa* (6x, DDM), *Ae. juvenalis* (DMU), and *Ae. vavilovii* (DMS) nested within the D clade (Figure 2.3). These interspersed patterns suggest that M subgenomes have undergone extensive interchromosomal rearrangements and genome restructuring following polyploidization, resulting in reduced phylogenetic affinity to their putative diploid progenitors and closer association with co-resident subgenomes within the polyploid nucleus. Similarly, the S subgenome of *Ae. vavilovii* clustered with the D clade rather than with *Ae. speltoides* or any other Sitopsis species. Its close relationship to *Ae. crassa* subsp. *crassa* (6x, DDM) supports the possibility that this *Ae. vavilovii* accession may be misclassified.

Using the same phylogenomic dataset, divergence times among *Triticum–Aegilops* lineages revealed a clear temporal structure. The earliest separations between the A and B genomes were estimated at ~6.99 MYA, followed by the divergence of the B and D lineages at ~6.65 MYA. Within the D clade, divergence ranged from 3.03 to 5.44 MYA when considering the D genomes of *Ae. tauschii*, *Triticum aestivum*, and polyploid *Aegilops* species. The most recent split occurred between *Ae. tauschii* and *Triticum aestivum*, whereas the deepest divergence reflects separation among more distantly related D subgenomes. The Sitopsis lineage diverged from the D lineage at ~6.30 MYA, with internal divergence spanning 1.73–5.22 MYA (Figure 2.3).

The N, M, C, and U genomes share a common ancestor at ~5.87 MYA. N genomes exhibit relatively narrow divergence (3.91–4.43 MYA), with the N subgenome of *Ae. neglecta* subsp. *recta* diverging at 3.91 MYA and that of *Ae. ventricosa* slightly earlier at 4.43 MYA. Diploid *Ae. comosa* clustered with the M subgenomes of *Ae. biuncialis* and *Ae. geniculata*, which diverged from each other at 3.62 MYA and from *Ae. comosa* at 4.44 MYA. The M subgenomes that grouped

within the D or U clades showed divergence times of 3.89–4.92 MYA (D clade) and 3.51–4.56 MYA (U clade), highlighting their deeper and more variable evolutionary placement. In the C clade, the C subgenomes of *Ae. cylindrica* and *Ae. triuncialis* diverged from the diploid *Ae. markgrafii* at approximately 3.77 and 4.56 MYA, respectively. In contrast, U genomes exhibited a broader divergence range (2.86–5.09 MYA), although most U subgenomes diverged more recently (<4 MYA) (Figure 2.3). Overall, phylogenetic relationships largely reflect expected diploid–polyploid associations, with exceptions—particularly in the M lineage—indicating a complex, reticulate evolutionary history. Divergence estimates further support earlier origins for M-genome–containing species relative to other lineages.

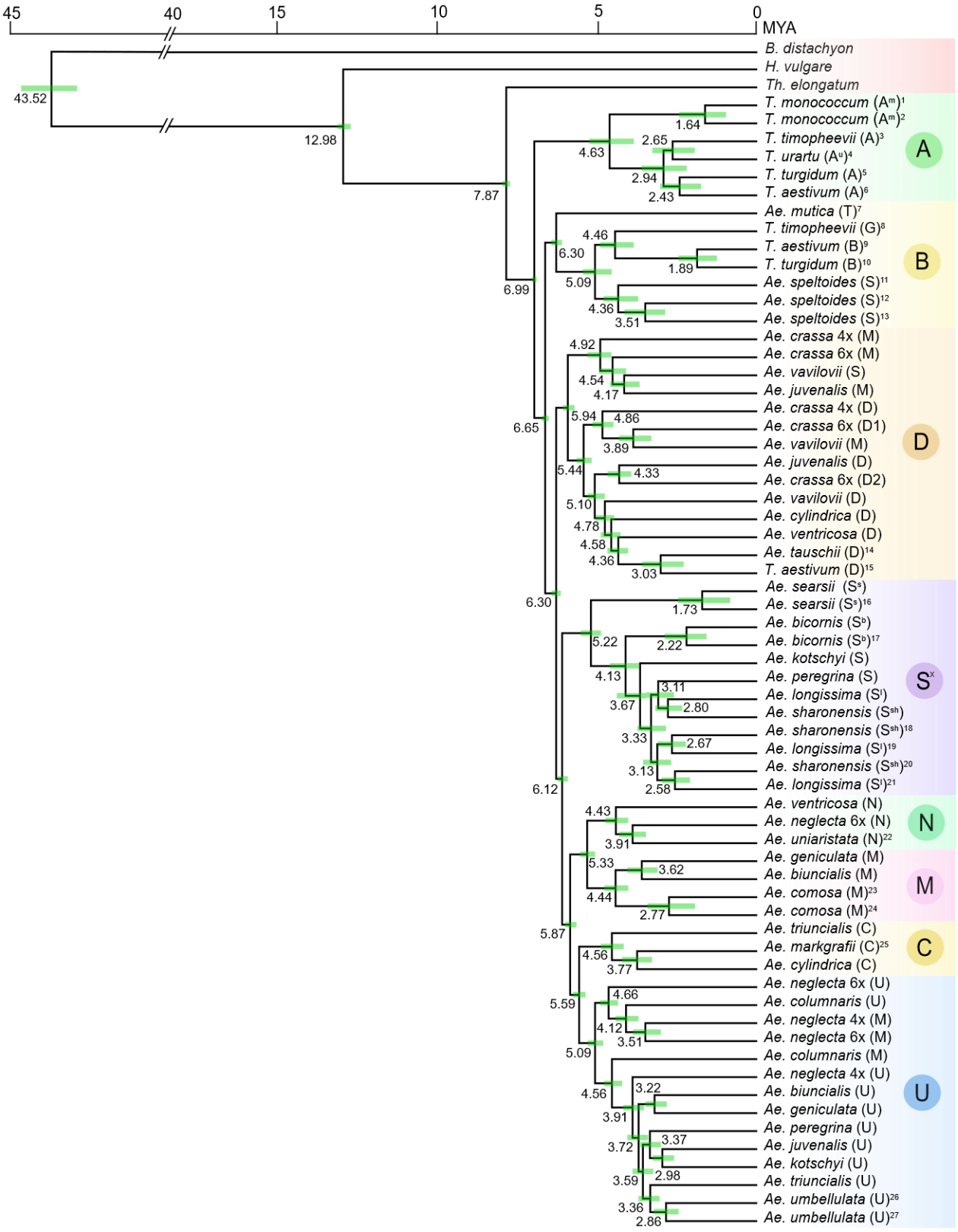


Figure 2.3 Phylogenetic relationships and divergence-time estimates in the *Triticum*/*Aegilops* species complex.

Phylogeny and divergence times inferred from 773 near single-copy orthologs. Horizontal bars represent the 95% highest posterior density (HPD) intervals for each node. Superscript numbers correspond to previously published genome references (e.g., 1, 2 — Ahmed et al. (2023); 3, 8 — Grewal et al. (2024); 4 —Ling et al. (2018); 5, 10 — Zhu et al. (2019); 6, 9, 15 — IWGSC (2021); 7 — Grewal et al. (2025); 11 — Yang et al. (2023); 12, 16, 17, 20, 21 — Li et al. (2022); 13, 18, 19 — Avni et al. (2022); 14 — Wang et al. (2021); 22, 24, 25, 27 — Khadka et al. (2026, unpublished); 23 — Li et al. (2024); 26 — Abrouk et al. (2023). Genomes without superscript numbers represent assemblies generated in this study. For simplicity, the following species names were shortened: *Ae. crassa* subsp. *crassa* (4x) is shown as *Ae. crassa* 4x; *Ae. crassa* subsp. *crassa* (6x) as *Ae. crassa* 6x; *Ae. neglecta* subsp. *neglecta* as *Ae. neglecta* 4x; and *Ae. neglecta* subsp. *recta* as *Ae. neglecta* 6x. A, B, D, S^x, N, M, C, and U indicate the corresponding clades.

2.3.4 Genome structural variation (SV) landscape

To investigate structural variation, tetraploid and hexaploid *Aegilops* subgenomes were compared with their respective diploid progenitors from which synteny, inversions, translocations, duplications, insertions, and deletions were tallied. Many subgenome–progenitor pairs retained extensive synteny, reflecting overall genome conservation. Among the 34 comparisons, fourteen were highly syntenic with over 2,000 shared syntenic blocks, ten showed moderate synteny with 1,000–2,000 blocks, and the remaining comparisons had fewer than 1,000 blocks, indicating varying levels of structural divergence (Figure 2.4A and Appendix 7).

Inversions were less frequent than other SV types but often spanned large regions ranging from 507 bp to 351 Mb, and accounted for the largest proportion of genome span among all SV classes (2.6–30.9%) (Figure 2.4B and Appendices 7-9), with the largest detected on chromosome 2M in the *Ae. geniculata–Ae. comosa* comparison (Figure 2.4C). Translocations were widespread, ranging from 439 bp to 67 Mb, and accounted for 0.0–13.2% of the genome span, with most comparisons showing over 2,000 events (Appendices 7-9). The largest intrachromosomal translocation occurred on chromosome 5C in *Ae. cylindrica–Ae. markgrafii* (Figure 2.4D) while the largest interchromosomal event involved chromosomes 5S of *Ae. peregrina* and $7S^l$ of *Ae.*

longissima (Figure 2.4E). Duplications were the most abundant SV type, with ten comparisons exceeding 10,000 events and up to 36,853 in the U genome comparison between *Ae. peregrina* and *Ae. umbellulata* (Figure 2.4A and Appendix 7). Despite this high abundance, duplications accounted for only a small proportion of genome span (0.0–8.2%) (Appendix 7). Duplication sizes ranged from 397 bp to 100 Mb (Appendix 8), with the largest events observed between chromosomes 3U–5U (100 Mb) and 3U–4U (80 Mb) in *Ae. neglecta* subsp. *neglecta* and *Ae. umbellulata* (Figure 2.4F). Insertions and deletions were less frequent, generally below 2,000 events per comparison, and both spanned only a very small portion of the genome, with sizes ranging from 51 bp to 11 Mb and 51 bp to 2 Mb, respectively (Appendices 7 and 8).

Lineage-specific patterns were also evident. Several M subgenomes showed low synteny with *Ae. comosa*, resulting in few detectable SVs, consistent with their phylogenetic placement alongside co-resident U or D subgenomes. Subsequent SV analyses of these subgenomes using the reference genome corresponding to their coexisting counterparts revealed higher synteny and more SVs (Appendices 10A-10G and 11), indicating extensive post-polyploidization restructuring. Similarly, the putative S subgenome of *Ae. vavilovii* showed limited synteny with *Ae. speltoides* (Appendix 9Y) and other Sitopsis genomes, but higher synteny with *Ae. tauschii* (Appendices 10H-10L and 11). In agreement with its phylogenetic placement, this SV pattern supports the interpretation that this subgenome corresponds to a D genome related to *Ae. crassa* subsp. *crassa* (6x, DDM), indicating potential misclassification of this accession. In contrast, the S subgenomes of *Ae. peregrina* and *Ae. kotschyi* exhibited greater synteny with *Ae. longissima* and *Ae. sharonensis* (Appendices 9I-9J and 9L-9M) than with *Ae. speltoides* (Appendices 10M-10N and 11), consistent with their phylogenetic relationships. Together, SV patterns closely mirror

phylogenetic relationships and highlight extensive interchromosomal restructuring following polyploidization.

To assert that the extent of the variation observed in polyploid species compared to their progenitor was the result of post-polyploidization restructuring, we quantified SV in four diploid species where assemblies for multiple accessions were available, namely *Ae. sharonensis, Ae. longissima*, *Ae. bicornis* and *Ae. searsii*. These diploid pangenome comparisons revealed substantial structural variation within species, indicating that SVs were already present among diploid lineages prior to polyploid formation (Appendices 12 and 13). The higher synteny and greater number of detected SVs in diploid comparisons likely reflect increased sequence similarity, which enables more extensive alignment and facilitates SV identification. Therefore, SVs identified in polyploid–diploid comparisons may represent a combination of ancestral standing variation retained from diploid progenitors and structural changes that occurred after polyploidization.

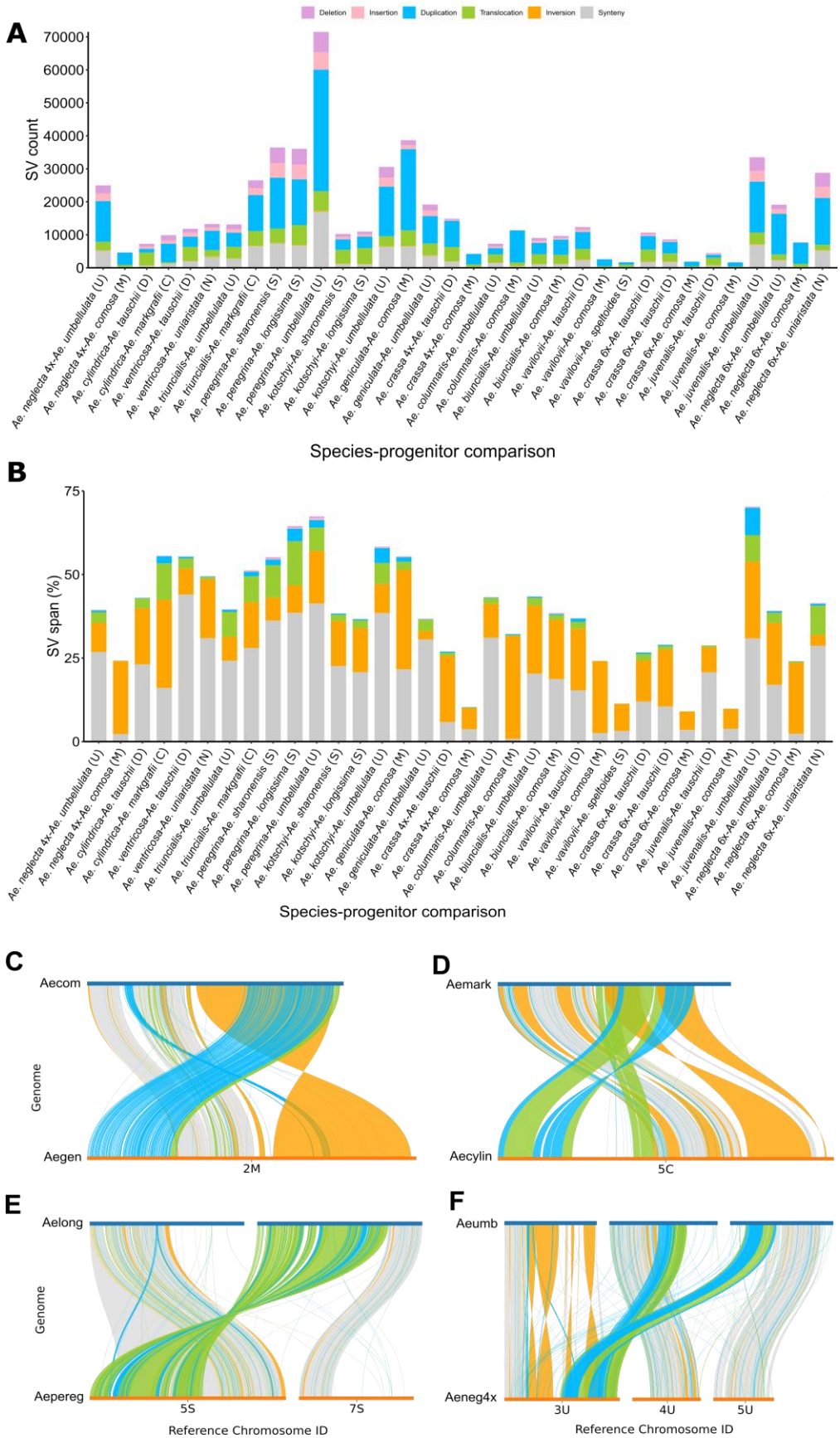

A
Deletion
Insertion
Duplication
Translocation
Inversion
Synteny
SV count
70000
60000
50000
40000
30000
20000
10000
0
Ae. neglecta 4x-Ae. umbellulata (U)
Ae. neglecta 4x-Ae. comosa (M)
Ae. cylindrica-Ae. tauschii (D)
Ae. cylindrica-Ae. markgrafii (C)
Ae. ventricosa-Ae. tauschii (D)
Ae. ventricosa-Ae. uniaristata (N)
Ae. triuncialis-Ae. umbellulata (U)
Ae. triuncialis-Ae. markgrafii (C)
Ae. peregrina-Ae. sharonensis (S)
Ae. peregrina-Ae. longissima (S)
Ae. kotschyi-Ae. umbellulata (U)
Ae. kotschyi-Ae. sharonensis (S)
Ae. kotschyi-Ae. longissima (S)
Ae. geniculata-Ae. umbellulata (U)
Ae. geniculata-Ae. comosa (M)
Ae. crassa 4x-Ae. umbellulata (U)
Ae. crassa 4x-Ae. tauschii (D)
Ae. crassa 4x-Ae. comosa (M)
Ae. columnaris-Ae. umbellulata (U)
Ae. columnaris-Ae. comosa (M)
Ae. biuncialis-Ae. umbellulata (U)
Ae. biuncialis-Ae. comosa (M)
Ae. vavilovii-Ae. tauschii (D)
Ae. vavilovii-Ae. comosa (M)
Ae. vavilovii-Ae. speltoides (S)
Ae. crassa 6x-Ae. tauschii (D)
Ae. crassa 6x-Ae. tauschii (D)
Ae. crassa 6x-Ae. comosa (M)
Ae. juvenalis-Ae. tauschii (D)
Ae. juvenalis-Ae. comosa (M)
Ae. juvenalis-Ae. umbellulata (U)
Ae. neglecta 6x-Ae. umbellulata (U)
Ae. neglecta 6x-Ae. comosa (M)
Ae. neglecta 6x-Ae. uniaristata (N)
Species-progenitor comparison
B
SV span (%)
75
50
25
0
Species-progenitor comparison
C
Aecom
Genome
Aegen
2M
D
Aemark
Aecylin
5C
E
Aelong
Genome
Aepereg
5S
7S
Reference Chromosome ID
F
Aeumb
Aeneg4x
3U
4U
5U
Reference Chromosome ID

Figure 2.4 Synteny patterns and structural variation in *Aegilops* subgenomes relative to their diploid progenitors.

(A) Stacked bar plots showing the distribution of syntenic regions and structural variant (SV) types—including inversions, translocations, duplications, insertions, and deletions—across subgenome–progenitor comparisons. Each bar represents the relative abundance of synteny and SV classes for a given comparison. The following species names were shortened: *Ae. neglecta* subsp. *neglecta* is shown as *Ae. neglecta* 4x; *Ae. crassa* subsp. *crassa* (4x) as *Ae. crassa* 4x; *Ae. crassa* subsp. *crassa* (6x) as *Ae. crassa* 6x; and *Ae. neglecta* subsp. *recta* as *Ae. neglecta* 6x.

(B) Stacked bar plots showing the percentage of each genome spanned by syntenic regions and SV classes across subgenome–progenitor comparisons. Each bar represents the proportion of genomic sequence covered by each category for a given comparison. Species abbreviations are the same as in (A).

(C–F) Representative large-scale structural variants identified in polyploid *Aegilops* subgenomes:

(C) the largest inversion on chromosome 2M in the *Ae. geniculata–Ae. comosa* comparison; **(D)** the largest intrachromosomal translocation on chromosome 5C in *Ae. cylindrica–Ae. markgrafii*;

(E) the largest interchromosomal translocation involving chromosome 5S of *Ae. peregrina* and chromosome $7S^{l}$ of *Ae. longissima*;

(F) the largest interchromosomal duplications between 3U–5U and 3U–4U in *Ae. neglecta* subsp. *neglecta–Ae. umbellulata*. To improve visualization, only SVs larger than 10 kb were included in plots. Species abbreviations used in the plots are: Aecom (*Ae. comosa*), Aegen (*Ae. geniculata*), Aemark (*Ae. markgrafii*), Aecylin (*Ae. cylindrica*), Aelong (*Ae. longissima*), Aepereg (*Ae. peregrina*), Aeumb (*Ae. umbellulata*) and Aeneg4x (*Ae. neglecta* subsp. *neglecta*).

2.3.5 Genome-wide transposable element (TE) composition

Genome-wide transposable element (TE) content was analyzed across all 18 *Aegilops* assemblies. TEs account for 85.44–88.10% of the genomes. This repetitive fraction is predominantly composed of Class I LTR retrotransposons (*Gypsy* and *Copia*) and Class II DNA transposons of the CACTA superfamily, whereas other TE superfamilies contribute marginally. *Gypsy* elements are the most abundant (38.0–46.84%), followed by *Copia* (13.97–19.04%) and CACTA elements (8.54–16.01%) (Figure 2.5A and Appendices 14-15).

Despite similar overall TE content, variations in composition were observed among species. In particular, *Ae. crassa* subsp. *crassa* (4x), *Ae. vavilovii*, and *Ae. crassa* subsp. *crassa* (6x) had fewer *Gypsy* (38.00%, 38.90%, 39.71%) and *Copia* (13.97%, 14.90%, 14.72%) , but more CACTA elements (16.01%, 14.32%, 15.29%) relative to other genomes. In addition, *Ae. crassa* subsp. *crassa* (4x) has the highest proportion of unclassified LTR elements (9.25%) (Appendix 14). These results indicate lineage-specific differences in TE composition across *Aegilops* species.

2.3.6 Comparative analysis of LTR clade dynamics across subgenomes

To assess LTR retrotransposon composition across *Aegilops* genomes, LTRs were classified into distinct clades using a domain-based HMM approach. Across all genomes, we identified four *Copia* (*Angela*, *SIRE*, *Ale*, *TAR*) and five *Gypsy* (*Retand*, *Athila*, *Tekay*, *CRM*, *Reina*) clades. For each clade, LTR counts were normalized per gigabase (Gb) to enable comparison across genomes and subgenomes (Appendices 16-21), with polyploid subgenomes compared to their diploid references (Figures 2.5B-2.5G).

Angela was the most abundant *Copia* element, followed by *SIRE*, whereas *Ale* and *TAR* were less represented. Among *Gypsy* elements, *Retand* and *Athila* were predominant, with *Tekay* and *CRM* showing intermediate levels, and *Reina* remaining low. Overall, LTR composition was broadly conserved, though total abundance and specific clade proportions showed lineage- and subgenome-specific variations. M and D subgenomes showed the greatest variability, whereas U and S subgenomes closely resembled their diploid progenitors, with minor variation in C and N subgenomes (Figures 2.5B-2.5G and Appendices 16-21).

Log_2 fold change analysis relative to diploid genomes revealed lineage-specific dynamics (Figure 2.5H). *Angela* consistently contracted in several M subgenomes including *Ae. crassa* subsp. *crassa* (4x and 6x) and *Ae. juvenalis*, while *SIRE* and *Ale* varied moderately, with *Ale*

expanding in some D subgenomes. *TAR* displayed localized expansion in certain M subgenomes while remaining stable across other subgenomes. Among *Gypsy* lineages, *Retand* and *Athila* expanded in multiple M subgenomes, and *Tekay* contracted within the D lineage. *CRM* displayed contrasting patterns, with depletion in some U subgenomes such as *Ae. geniculata*, *Ae. biuncialis*, but expanded in several M and D subgenomes. In contrast, *Reina* consistently expanded across all D subgenomes (Figure 2.5H). Independent of species identity, fold-change patterns clustered subgenomes with similar LTR dynamics. While overall composition is conserved, individual LTR clades underwent subgenome-specific expansion and contraction.

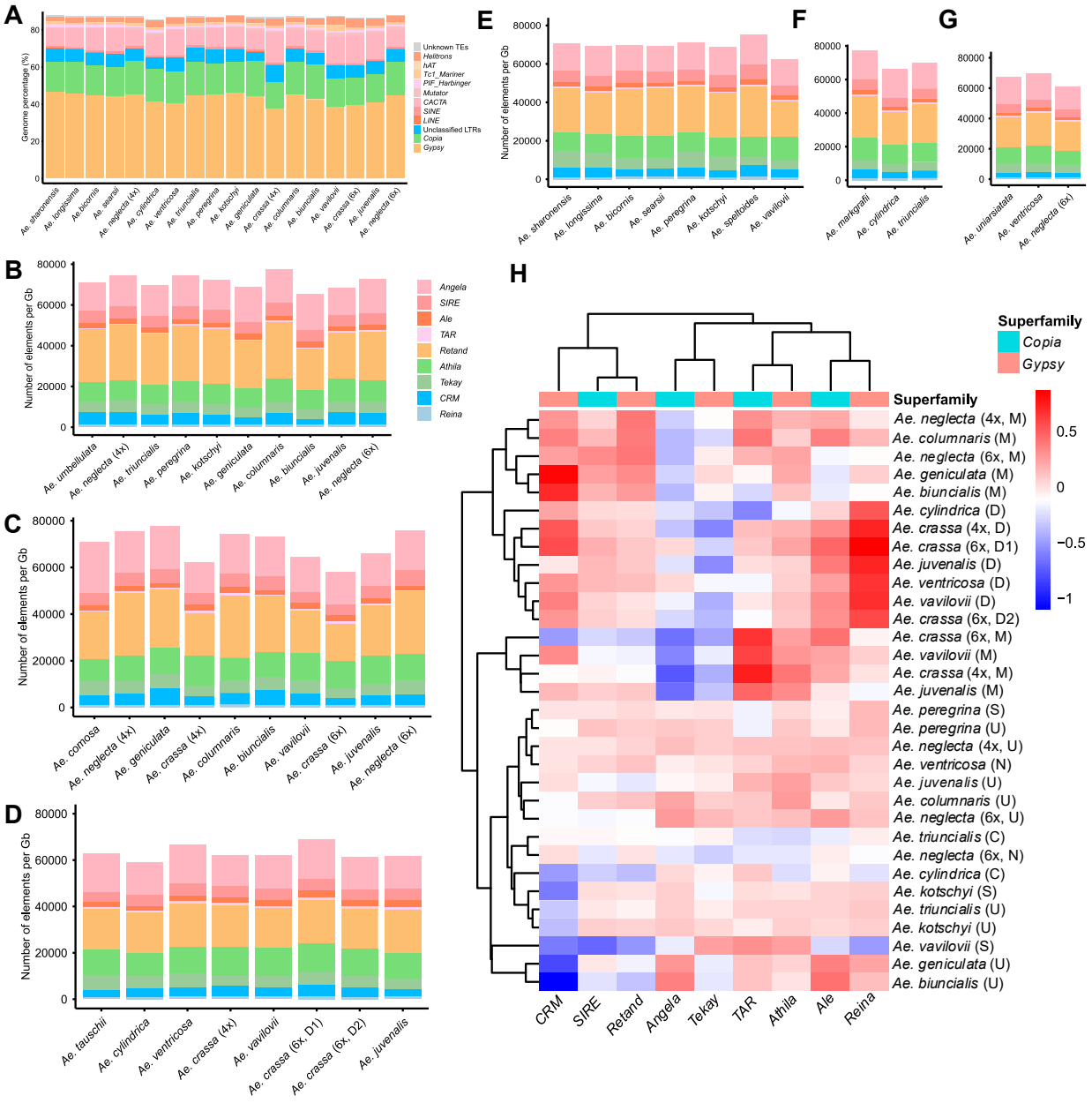


Figure 2.5 Transposable element composition and LTR clade dynamics across *Aegilops* genomes and subgenomes.

(A) Stacked bar plot showing the proportional contribution of each transposable element class to overall *Aegilops* genome composition.

(B–G) Normalized LTR clade abundance (per Gb) across U, M, D, S, C, and N subgenomes, respectively.

(H) Heatmap showing $\log_2$ fold changes in LTR clade abundance relative to corresponding diploid reference genomes. The following species names were shortened: *Ae. neglecta* subsp. *neglecta* is shown as *Ae. neglecta* (4x); *Ae. crassa* subsp. *crassa* (4x) as *Ae. crassa* (4x); *Ae. crassa* subsp. *crassa* (6x) as *Ae. crassa* (6x); and *Ae. neglecta* subsp. *recta* as *Ae. neglecta* (6x).

2.3.7 Subgenome-specific LTR dynamics within polyploid *Aegilops* species

To investigate LTR dynamics within polyploid genomes, we compared LTR clade abundance between subgenomes using $\log_2$ fold-change. For tetraploids, subgenome pairs were compared (e.g., U vs M in *Ae. biuncialis*), while in hexaploids all pairwise comparisons were performed (e.g., D vs M, D vs S, M vs S in *Ae. vavilovii*). Overall, most clade-level differences were modest ($\log_2$ fold-change < 0.5), indicating similar LTR composition within species (Figure 2.6 and Appendices 22-24).

Despite this overall similarity, consistent subgenome-specific patterns were observed. In *Ae. geniculata* and *Ae. biuncialis*, the *Ale* and *TAR Copia* clades were enriched in the U subgenome, while the *CRM Gypsy* clade was consistently enriched in the M subgenome (Figures 2.6G and 2.6J), suggesting conserved dynamics. In *Ae. peregrina* and *Ae. kotschyi*, the *Tekay Gypsy* clade was enriched in the S subgenomes, with additional *CRM* expansion in the U subgenome of *Ae. kotschyi* (Figures 2.6E-2.6F). In tetraploids *Ae. neglecta* subsp. *neglecta*, *Ae. cylindrica*, *Ae. ventricosa*, *Ae. triuncialis*, *Ae. crassa* subsp. *crassa* (4x), and *Ae. columnaris*, minimal differences were observed for most clades, supporting overall homogenization (Figures 2.6A–2.6D and 2.6H–2.6I). However, localized expansions were still evident, including *TAR* in the C subgenome of *Ae. cylindrica* (Figure 2.6B), *Ale* in the M subgenome of *Ae. columnaris* (Figure 2.26I), and *CRM* enrichment in multiple U and D subgenomes (Figures 2.6A, 2.6C and 2.6H).

In hexaploids, subgenome-specific differences were limited but more pronounced in some species. In *Ae. juvenalis*, *CRM* showed clear expansion in the U subgenome relative to D and M (Figures 2.6Q–2.6S), while in *Ae. crassa* subsp. *crassa* (6x), its expansion was in D subgenomes relative to M (Figures 2.6N-2.6P). In *Ae. neglecta* subsp. *recta*, *CRM* was enriched in the U and M subgenomes compared with the N subgenome, whereas *Ale* expanded in the M subgenome relative to the U subgenome (Figures 2.6T–2.6V). In contrast, *Ae. vavilovii* showed minimal variation, indicating a highly homogenized LTR landscape (Figures 2.6K–2.6M). Overall, LTR composition is largely homogenized among subgenomes within species, likely reflecting shared nuclear environments after polyploidization. However, repeated expansion of specific lineages—particularly *Gypsy CRM*—across multiple species suggests ongoing or preferential activity in certain genomic contexts, highlighting its role in recent LTR dynamics in polyploid *Aegilops*.

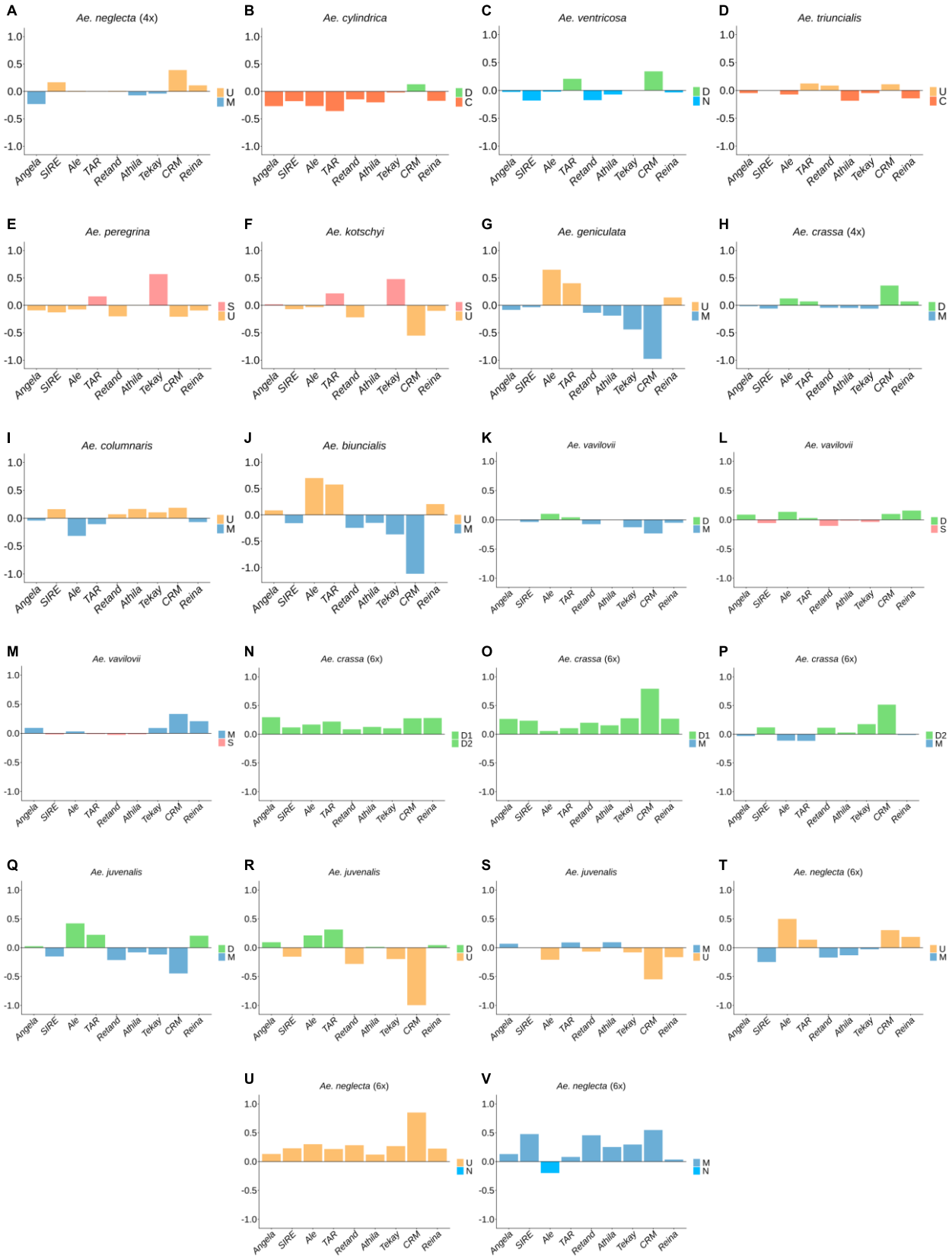

A
Ae. neglecta (4x)
B
Ae. cylindrica
C
Ae. ventricosa
D
Ae. triuncialis
E
Ae. peregrina
F
Ae. kotschyi
G
Ae. geniculata
H
Ae. crassa (4x)
I
Ae. columnaris
J
Ae. biuncialis
K
Ae. vavilovii
L
Ae. vavilovii
M
Ae. vavilovii
N
Ae. crassa (6x)
O
Ae. crassa (6x)
P
Ae. crassa (6x)
Q
Ae. juvenalis
R
Ae. juvenalis
S
Ae. juvenalis
T
Ae. neglecta (6x)
U
Ae. neglecta (6x)
V
Ae. neglecta (6x)
1.0
0.5
0.0
-0.5
-1.0
Angela
SIRE
Ale
TAR
Retand
Athila
Tekay
CRM
Reina
U
M
D
C
D
N
U
C
S
U
U
M
D
M
U
M
D
M
D
S
M
S
D1
D2
D1
M
D2
M
D
M
D
U
M
U
U
N
M
N

Figure 2.6 Subgenome-specific LTR dynamics within polyploid *Aegilops* species.

(A–V) Log_2 fold changes in LTR lineage abundance between subgenomes within tetraploid and hexaploid species, highlighting subgenome-specific expansions and contractions of individual LTR clades. For each panel, species names are at the top and subgenomes compared are specified in the legend. The following species names were shortened: *Ae. neglecta* subsp. *neglecta* is shown as *Ae. neglecta* (4x); *Ae. crassa* subsp. *crassa* (4x) as *Ae. crassa* (4x); *Ae. crassa* subsp. *crassa* (6x) as *Ae. crassa* (6x); and *Ae. neglecta* subsp. *recta* as *Ae. neglecta* (6x).

2.3.8 Insertion time dynamics of LTR retrotransposons across subgenomes

To investigate temporal dynamics of LTR retrotransposons, insertion times were estimated for Full-length elements in each subgenome and compared with their diploid references. Across all genomes, a consistent temporal separation between LTR superfamilies was observed. *Copia* elements showed predominantly recent insertions, peaking at ~0–1 MYA, whereas *Gypsy* elements exhibited older insertion profiles, generally peaking at ~0.5–1.5 MYA (Figure 2.7 and Appendices 25-30).

Subgenome-specific variation in full-length LTR abundance was also evident. Several subgenomes, including the U and M subgenomes of *Ae. biuncialis*, the M subgenome of *Ae. crassa* subsp. *crassa* (6x), and the D and C subgenomes of *Ae. cylindrica*, exhibited lower numbers of full-length LTRs, indicating variation in intact LTR representation across subgenomes (Figures 2.7A–2.7C and 2.7E). Differences were also apparent at the ploidy level. Tetraploid subgenomes showed LTR insertion time distributions broadly similar to their diploid progenitors, whereas hexaploid genomes exhibited lower overall frequencies of full-length LTR elements and less pronounced peaks, indicating differences in the abundance and preservation of intact LTR retrotransposons across ploidy levels.

To place these patterns in an evolutionary context, insertion time distributions were compared with estimated divergence times between polyploid subgenomes and their

corresponding diploid progenitors. Divergence time analysis indicated that subgenomes diverged more than 2 MYA (Figure 2.3), whereas many peak LTR insertions occurred within the last 0–1.5 MY (Figure 2.7 and Appendices 25-30). This temporal separation indicates that most LTR activity postdates subgenome divergence and therefore represents lineage- and subgenome-specific transposition. Consistent with this, recent *Copia* activity (~0–1 MYA) suggests ongoing amplification after polyploidization. Although *Gypsy* elements show slightly older insertion profiles, their activity also largely falls within the post-divergence period, suggesting independent amplification across lineages. Overall, these results indicate that LTR dynamics in *Aegilops* are driven primarily by both lineage-specific insertion histories and the differential retention or degradation of older elements.

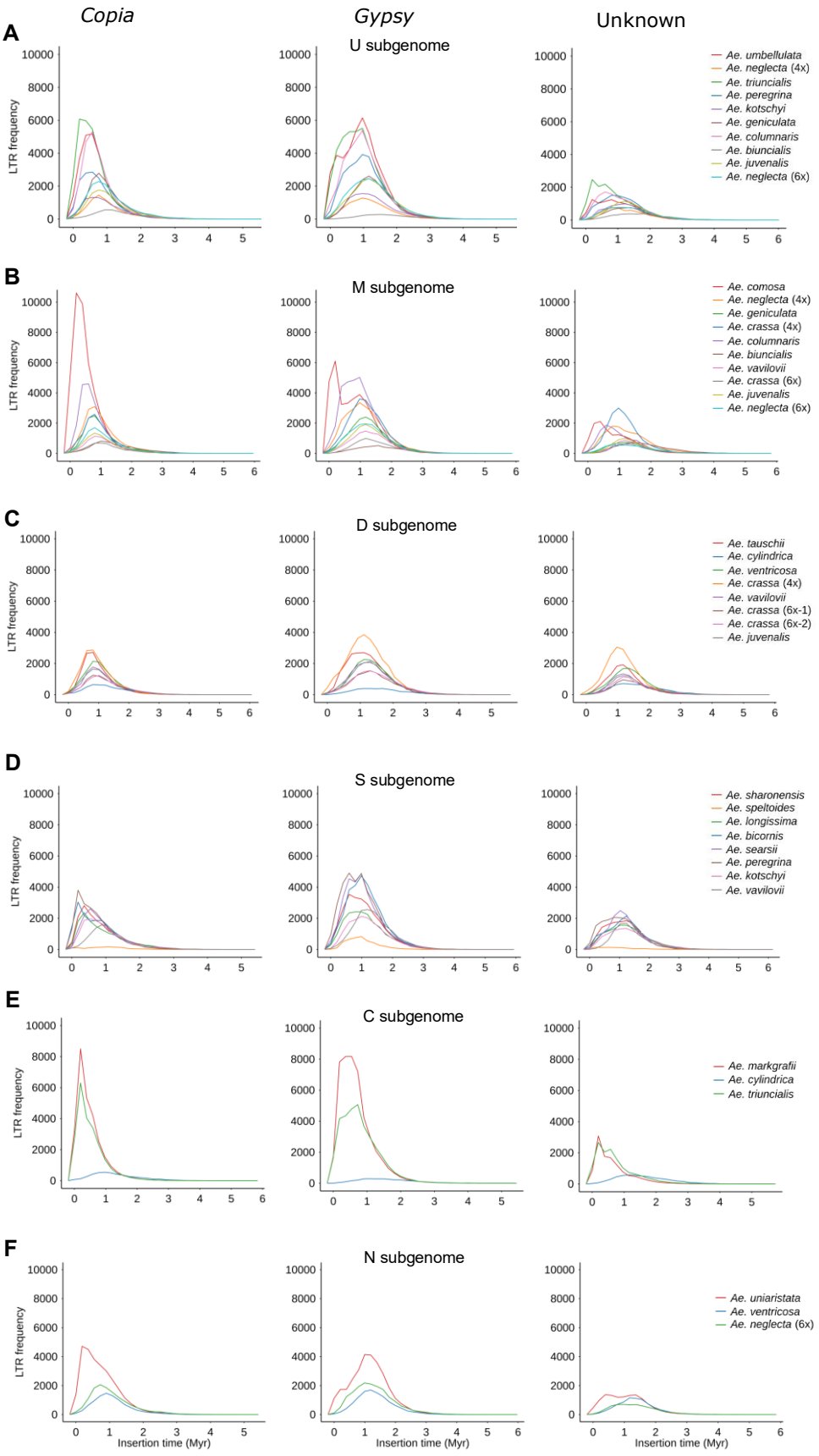

Copia
Gypsy
Unknown
A
U subgenome
B
M subgenome
C
D subgenome
D
S subgenome
E
C subgenome
F
N subgenome
LTR frequency
Insertion time (Myr)
Ae. umbellulata
Ae. neglecta (4x)
Ae. triuncialis
Ae. peregrina
Ae. kotschyi
Ae. geniculata
Ae. columnaris
Ae. biuncialis
Ae. juvenalis
Ae. neglecta (6x)
Ae. comosa
Ae. neglecta (4x)
Ae. geniculata
Ae. crassa (4x)
Ae. columnaris
Ae. biuncialis
Ae. vavilovii
Ae. crassa (6x)
Ae. juvenalis
Ae. neglecta (6x)
Ae. tauschii
Ae. cylindrica
Ae. ventricosa
Ae. crassa (4x)
Ae. vavilovii
Ae. crassa (6x-1)
Ae. crassa (6x-2)
Ae. juvenalis
Ae. sharonensis
Ae. speltoides
Ae. longissima
Ae. bicornis
Ae. searsii
Ae. peregrina
Ae. kotschyi
Ae. vavilovii
Ae. markgrafii
Ae. cylindrica
Ae. triuncialis
Ae. uniaristata
Ae. ventricosa
Ae. neglecta (6x)

Figure 2.7 Insertion time dynamics of LTR retrotransposons across *Aegilops* subgenomes.

(A–F) Insertion time distributions of LTR retrotransposons in **(A)** U, **(B)** M, **(C)** D, **(D)** S, **(E)** C, and **(F)** N subgenomes, including comparisons with the corresponding diploid reference genomes. The following species names were shortened: *Ae. neglecta* subsp. *neglecta* is shown as *Ae. neglecta* (4x); *Ae. crassa* subsp. *crassa* (4x) as *Ae. crassa* (4x); *Ae. crassa* subsp. *crassa* (6x) as *Ae. crassa* (6x); and *Ae. neglecta* subsp. *recta* as *Ae. neglecta* (6x). *Ae. crassa* (6x-1) and *Ae. crassa* (6x-2) denote the D1 and D2 subgenomes of *Ae. crassa* subsp. *crassa* (6x), respectively.

2.4 Discussion

We generated high quality genome assemblies of four diploid, ten tetraploid, and four hexaploid species of *Aegilops*, which, together with the previously published ones, complete the production of assemblies for all 25 species of the genus. These assemblies substantially expand genomic resources for the genus, particularly for polyploid taxa, for which only *Ae. ventricosa* (Liu et al., 2025) was previously available. They provide a foundation for investigating genome evolution, polyploidization, structural variation, and traits relevant to wheat improvement.

Overall, assembly sizes are consistent with previous estimates while showing improved contiguity and completeness. Diploid assemblies closely match earlier reports (Avni et al., 2022; Li et al., 2022), with higher N50 and N90 values indicating improved continuity. Tetraploid genome sizes are concordant with cytogenetic estimates (Eilam et al., 2008; Eilam et al., 2007) and our flow cytometry data (Appendix 31), with S^x-containing species (e.g., *Ae. kotschyi*, *Ae. peregrina*) having larger genomes, whereas C-genome species (e.g., *Ae. cylindrica*, *Ae. triuncialis*) exhibit smaller genome sizes (Eilam et al., 2008; Eilam et al., 2007). The *Ae. ventricosa* assembly is comparable to that of Liu et al. (2025). The four hexaploid assemblies represent the first scaffold-level genomic resources for these species, with consistently high contiguity despite the challenges of assembling large, repetitive polyploid genomes. Comparative analyses further revealed a consistent pattern of genome downsizing across nearly all polyploid subgenomes relative to their diploid progenitors, indicating that genome contraction represents a pervasive feature of polyploid

genome evolution in *Aegilops*. Similar patterns have been reported in other Triticeae polyploids and are generally interpreted as consequences of post-polyploid diploidization involving sequence elimination and genome restructuring (Eilam et al., 2008; Leitch and Bennett, 2004).

Building on these assemblies, protein-coding genes were annotated across all ploidy levels. High-confidence (HC) gene counts were generally lower than in previous studies. In diploids, HC gene numbers ranged from 21,865–27,783, compared with 31,183–40,222 reported in prior studies (Avni et al., 2022; Li et al., 2022; Yu et al., 2022). Similarly, tetraploid genomes had fewer HC genes than *Ae. ventricosa* (69,000) (Liu et al., 2025) and tetraploid wheat (65,012–66,559) (Avni et al., 2017; Maccaferri et al., 2019), although *Ae. triuncialis* (60,490) showed comparable values. In hexaploids, HC gene counts were also lower than the 106,913 genes reported for the Chinese Spring wheat genome (IWGSC, 2021). These differences could stem from methodological variation. Previous annotations incorporated RNA-Seq, Iso-Seq, and homology-based evidence, whereas our approach used a machine-learning model without transcriptomic support, but applied a stricter 95% similarity threshold (vs. 80–90% in earlier studies). Despite lower HC gene counts, BUSCO completeness (84.5%–98.9%) indicates that vast majority of the conserved genes were captured, supporting their use in comparative and evolutionary analyses.

Annotation of protein-coding genes in *Ae. crassa* subsp. *crassa* (4x) revealed higher gene content than in other tetraploid assemblies. Because elevated gene numbers can raise concerns about ploidy misclassification, nuclear DNA content was assessed by flow cytometry, confirming a genome size of 10.39 Gb, consistent with other tetraploids (8.83–12.01 Gb) (Appendix 31). The increased gene number suggests lineage-specific gene retention and reduced post-polyploidization gene loss, a pattern commonly observed in polyploid plant genomes following whole-genome duplication (Cheng et al., 2018). Its deep phylogenetic divergence further suggests that extended

evolutionary time contributed to the accumulation and retention of lineage-specific genes relative to other tetraploids.

Subgenome-level comparisons revealed that gene content evolution following polyploidization was highly heterogeneous. The M and D subgenomes displayed the greatest variability, including substantial gene loss in several lineages, consistent with the strongest genome downsizing observed in the M subgenomes. In comparison, the U and N subgenomes exhibited intermediate patterns characterized by moderate lineage-specific expansion and contraction, indicating a more limited but still detectable degree of post-polyploid remodeling. The S and C subgenomes remained comparatively conserved, retaining gene numbers similar to their diploid progenitors, consistent with genome size comparisons showing minimal deviation from diploid genome sizes in these lineages. Overall, these results indicate that genome downsizing following polyploidization is associated with widespread and uneven gene fractionation across subgenomes, reflecting coordinated but lineage-specific reductions in gene content.

The phylogenomic framework provides a comprehensive assessment of evolutionary relationships across all 25 recognized *Aegilops* species together with key *Triticum* genomes. The inferred relationships are largely consistent with established cytogenetic and taxonomic frameworks (Gill and Friebe, 2002; Kilian et al., 2011; Lilienfeld, 1951; Van Slageren, 1994) while also revealing lineage-specific complexities of genome evolution in the *Aegilops–Triticum* complex. Within the A clade, *T. urartu* and *T. monococcum* cluster with A subgenomes of polyploid wheats, consistent with their established roles as A-genome donors (Kilian et al., 2011; Li et al., 2022; Marcussen et al., 2014). The B-genome clade includes the B subgenomes of *T. aestivum* and *T. turgidum*, together with the S genome of *Ae. speltoides* and the G genome of *T. timopheevii*. The T genome of *Ae. mutica* also groups within this lineage, substantiating previous

studies that suggested it as an early-diverging member of the B lineage. (Avni et al., 2022; Bernhardt et al., 2020; Glémin et al., 2019; Li et al., 2022)

The D genomes form a well-supported clade, with all D subgenomes clustering with *Ae. tauschii* and the D genome of hexaploid wheat, consistent with previous evidence for their origin (Chen et al., 2020; Lilienfeld and Kihara, 1934; Liu et al., 2025). Notably, divergence patterns suggest a more ancient origin of D subgenomes in polyploid *Aegilops* than previously proposed (Liu et al., 2025), with *Ae. crassa* subsp. *crassa* (4x) showing the earliest divergence, supporting its status as one of the oldest polyploid species (Dubcovsky and Dvorak, 1995; Kroupin et al., 2022). The "S" subgenomes of *Ae. peregrina* and *Ae. kotschyi* cluster with *Ae. longissima* (S^l) and *Ae. sharonensis* (S^{sh}), rather than *Ae. speltoides* (S), consistent with earlier studies (Badaeva et al., 2004; Friebe, 1996; Jewell, 1979; Jewell and Driscoll, 1983; Zhang et al., 1992; Zhao et al., 2016). Structural variation patterns further support *Ae. sharonensis* as the most likely progenitor. Similarly, the N, C, and U subgenomes group with their respective diploid donors (*Ae. uniaristata*, *Ae. markgrafii*, and *Ae. umbellulata*). For the N lineage, divergence estimates indicate a more ancient origin for *Ae. ventricosa* compared to previous reports (Liu et al., 2025) possibly reflecting an earlier split and subsequent independent evolution. Overall, these patterns support largely conserved relationships between polyploid subgenomes and their diploid progenitors.

The M genome displays the most complex and phylogenetically dispersed pattern. Only the M subgenomes of *Ae. geniculata* and *Ae. biuncialis* cluster with the diploid donor *Ae. comosa*, whereas others group within U- or D-genome clades. This pattern differs from the expected donor–subgenome relationships and suggests a more complex evolutionary history than previously assumed. Although alternative explanations, including subgenome misclassification, paralogous sequence inclusion, or assembly-related artifacts, could contribute to unexpected phylogenetic

placements, several factors support a biological interpretation. All polyploid genomes were assigned to subgenomes using a consistent comparative approach across species, reducing the likelihood of systematic subgenome misassignment. Furthermore, the high contiguity and completeness of the assemblies, supported by high N50 values, BUSCO completeness, and LTR assembly quality scores ranging from 11.5 to 24.5, indicate that assembly errors are unlikely to fully explain the observed phylogenetic patterns. The analyses were also based on near single-copy orthologs, reducing the potential influence of paralogous sequences. In addition, SV and synteny analyses provided further support for a biological basis underlying these phylogenetic patterns. M subgenomes that were phylogenetically displaced from their diploid progenitor showed reduced synteny and fewer detectable structural variants in direct comparisons with *Ae. comosa*, consistent with their divergence from the ancestral lineage. Most M subgenomes show deep divergence (>4 MYA), placing them among the oldest polyploid-derived subgenomes in *Aegilops*. Notably, the M subgenome of *Ae. crassa* subsp. *crassa* (4x; DM) shows the earliest divergence, consistent with earlier cytogenetic studies identifying this species as one of the most ancient polyploids (Dubcovsky and Dvorak, 1995; Kroupin et al., 2022). This phylogenetic displacement, together with deep divergence, suggests post-polyploidization extensive genomic remodeling, potentially associated with structural variation and homoeologous exchanges during prolonged coexistence with D or U genomes (Feldman and Levy, 2012; Zhang et al., 2023). An alternative hypothesis is that some polyploid subgenomes currently assigned to the M genome may not have originated from extant *Ae. comosa*, but instead from a closely related, unsampled, or extinct diploid lineage, or that their genomic ancestry has been obscured by subsequent evolutionary processes. Consequently, the identity of the diploid donor for these subgenomes may require re-evaluation using broader taxonomic sampling to provide additional genomic evidence.

Similarly, the S subgenome of *Ae. vavilovii* clustered with the D clade rather than with *Ae. speltoides* or other Sitopsis species. Previous cytogenetic and molecular studies suggest that *Ae. vavilovii* originated from a tetraploid *Ae. crassa*-like (DM) ancestor following hybridization with an S-genome donor, indicating a close evolutionary relationship between *Ae. vavilovii* (DMS) and *Ae. crassa* subsp. *crassa* (6x) (DDM) through their shared tetraploid DM ancestry (Baum et al., 2012; Edet et al., 2018). The placement of the putative S subgenome within the D lineage is unexpected and is consistent with the possibility that the assigned S subgenome may instead represent a D subgenome, suggesting potential accession misidentification or taxonomic uncertainty. Alternatively, given the close relationship of several Sitopsis species to the D lineage, the S subgenome of *Ae. vavilovii* may have originated from another Sitopsis lineage, such as *Ae. longissima*, *Ae. bicornis*, *Ae. searsii*, or *Ae. sharonensis*. This possibility is consistent with contributions of *Ae. longissima* and/or *Ae. sharonensis* to the S subgenomes of *Ae. kotschyi* and *Ae. peregrina*. The observed placement may also reflect the complex evolutionary history of DMS-genome polyploids, including homoeologous exchanges, incomplete lineage sorting, or historical introgression that influenced the phylogenetic signal. Additional genomic sampling of related species and multiple accessions will be required to determine whether this pattern reflects evolutionary divergence or accession misidentification.

Comparative analyses of the assemblies revealed that extensive structural variation is a pervasive feature of polyploid *Aegilops* genomes and a major driver of post-polyploidization evolution. While many subgenome–progenitor comparisons retain large syntenic regions, widespread inversions, translocations, duplications, insertions, and deletions indicate genome reorganization following polyploid formation, consistent with observations in wheat and other *Aegilops* polyploids (Badaeva et al., 2004; Badaeva et al., 2021; Friebe, 1996). Large inversions,

though less frequent, produce major structural changes, while abundant translocations reflect homoeologous pairing and exchanges documented cytogenetically (Badaeva et al., 2004; Friebe, 1996; Lilienfeld, 1951), indicating that interchromosomal rearrangements have been a persistent feature of their evolutionary history. Duplications, the most prevalent SV type, often involve large segments and likely contribute to gene copy number variation, dosage effects, and functional diversification, all recognized sources of evolutionary novelty in polyploid genomes (Flagel and Wendel, 2009).

Patterns of structural variation closely mirror phylogenetic relationships. Subgenomes phylogenetically displaced from their diploid progenitors—particularly M subgenomes—show reduced synteny and few detectable SVs in direct comparisons, whereas comparisons with references of co-resident subgenomes reveal higher synteny and more SVs, suggesting that prolonged coexistence of divergent subgenomes facilitates interchromosomal rearrangements and homoeologous exchanges, widening divergence from diploid progenitors.

Considering the limited availability of comprehensive pan-genome resources for *Aegilops* species, it is not possible to unambiguously distinguish structural variants that are unique to polyploidization from those already present within natural diploid populations. However, additional comparisons among multiple diploid *Aegilops* genomes from the same species indicate that substantial structural variation is already present within diploid gene pools. This suggests that at least a portion of the SV observed in polyploid–progenitor comparisons may reflect pre-existing variation within ancestral diploid populations, in addition to structural changes accumulated following polyploidization.

Genome-wide TE annotation across 18 *Aegilops* species revealed that TEs constitute a major fraction of these genomes (85–88%), dominated by LTR retrotransposons (*Gypsy* and

Copia) and *CACTA* DNA transposons, consistent with other Triticeae species (Vicient and Casacuberta, 2017; Wicker et al., 2018). While overall TE composition is broadly conserved, lineage-specific variations—such as reduced *Gypsy* and *Copia* content and elevated *CACTA* abundance in several polyploids—highlight differential retention and amplification, reflecting the dynamic behavior of TEs following polyploidization (Fedoroff, 2000; Parisod et al., 2010).

Comparative analysis of LTR retrotransposon clades across *Aegilops* genomes revealed TE landscape dominated by a limited number of highly abundant lineages. Across all genomes, *Angela* represented the most abundant *Copia* element, whereas within the *Gypsy* superfamily, *Retand* and *Athila* were consistently predominant, indicating that these clades constitute the major components of LTR retrotransposon diversity in *Aegilops*. Comparative LTR analyses across subgenomes relative to their diploid progenitors reveal that post-polyploidization TE dynamics are highly lineage- and subgenome-specific, rather than following a uniform trajectory. Instead, distinct LTR clades exhibit contrasting patterns depending on genomic context. For example, the *Copia* element *Angela* shows consistent contraction across most M subgenomes, whereas *TAR* displays localized expansions in specific M subgenomes, indicating heterogeneous *Copia* dynamics within the same subgenomic background. In contrast, the *Gypsy* lineage *Reina* shows a consistent expansion across all D subgenomes, suggesting a shared D-genome-specific propensity for this TE family. Similarly, *CRM* exhibits highly contrasting behavior across genomes, with contraction in several U subgenomes (e.g., *Ae. biuncialis* and *Ae. geniculata*), but expansion in selected M and D subgenomes. These lineage-specific LTR dynamics were broadly consistent with patterns of genome size evolution across subgenomes. Subgenomes exhibiting stronger genome downsizing, particularly the M subgenomes, generally also showed greater shifts in LTR composition and abundance, suggesting that TE turnover contributed, at least in part, to genome contraction

following polyploidization. In contrast, the S subgenomes remained comparatively stable in LTR composition and abundance, closely resembling their diploid progenitors, consistent with their overall conservation in genome size and gene content. Together, these findings indicate that different subgenomes followed distinct evolutionary trajectories after polyploidization, with TE dynamics contributing differentially to genome remodeling and subgenome divergence in *Aegilops*.

Subgenome-level analyses within polyploids further indicate that while most LTR clades are homogenized, certain families display localized expansions or contractions. Notably, the *Gypsy CRM* consistently exhibits subgenome-specific expansion, suggesting preferential activity or retention. Temporal analyses show *Copia* elements driving recent genome evolution, whereas *Gypsy* elements largely reflect older amplification events, likely influenced by transposition activity, epigenetic regulation, and element removal via recombination or deletion (Bennetzen and Wang, 2014; Fedoroff, 2000; Lisch, 2013; Pulido and Casacuberta, 2023). Overall, these findings underscore the central role of TEs in shaping *Aegilops* genome structure and their contribution to subgenome differentiation and polyploid genome evolution.

In conclusion, we provide a comprehensive genomic framework for 18 *Aegilops* species, integrating high-quality assemblies, gene annotation, phylogenetic, and structural variation analyses. Polyploid subgenomes show substantial divergence from their diploid progenitors, associated with extensive interchromosomal remodeling. LTR analyses reveal lineage- and subgenome-specific expansions and contractions, highlighting the dynamic role of TEs upon polyploidization. The multi-faceted analysis provides insights into polyploid *Aegilops* genome evolution and can be capitalized upon to inform wheat improvement through novel alleles, subgenome variation, and structural diversity.

2.5 Methods

2.5.1 Plant materials

One accession each of 18 *Aegilops* species (Appendix 1) were grown in root trainers in a growth cabinet with a 16 h light and 8 h dark photoperiod regime, and temperature settings of 24°C during the day and 18°C at night. Seedlings were vernalized at the 4-5 leaf stage for ten weeks in a 4°C growth chamber with a 12 h photoperiod. After vernalization, seedlings were transplanted and returned to the original growth chamber. At heading, plants were bagged to prevent cross-pollination and seeds were harvested at maturity. Single-seed descents were grown for 3-5 generations to increase homozygosity prior to genome sequencing initiation.

2.5.2 Isolation of high-molecular-weight genomic DNA

Ten to 58 seedlings from each of the 18 single-seed descent accessions (Appendix 1) were grown in root trainers as described above. At the 2-4 leaf stage, plants were maintained in darkness for 2-3 days, after which tissue collection was performed by weighing the tissue prior to flash-freezing and storage in liquid nitrogen. Seedlings were allowed to regrow and the sampling process was repeated until approximately 15 g was collected per sample. Pooling tissue from multiple seedlings ensured sufficient high-molecular-weight DNA yield for long-read sequencing. Heterogeneity due to residual heterozygosity within pools was minimized through single-seed descent of each accession for three or more generations prior to tissue sampling.

High-molecular-weight genomic DNA (HMW gDNA) was extracted using a modified nuclei isolation protocol (Zhang et al., 2012) where the nuclei isolation buffer (NIB) was supplemented immediately before use with spermine trihydrochloride and spermidine tetrahydrochloride to final concentrations of 1 mM each. Briefly, the tissue was ground to a fine powder in liquid nitrogen using a mortar and pestle and homogenized in the supplemented NIB.

The homogenate was filtered and divided into four 50 mL Falcon tubes. Nuclei were pelleted by centrifugation at 2,786 *g* for 20 minutes at 4°C. Pellets were gently resuspended in 200 µL ice-cold NIB using a paintbrush and combined into a single 50 mL Falcon tube. The nuclei were washed four additional times with 40 mL supplemented NIB. After the final wash, nuclei were resuspended in 20 mL of 1× homogenization buffer (HB) containing 1% (w/v) polyvinylpyrrolidone (PVP40), followed by centrifugation. The final pellet was gently resuspended in 3 mL of 1× HB + PVP using minimal pipetting and a wide-bore tip.

The nuclei suspension was divided into six 15-mL Falcon tubes (~500 µL per tube), and nuclei were pelleted by centrifugation at 2,450 *g* for 10 minutes at 4°C. Each pellet was resuspended in 3.6 mL lysis buffer (100 mM Tris-HCl pH 9.0, 25 mM Na_2EDTA, 138 mM NaCl, 1% PVP40, 0.3 mg/mL proteinase K, 1% sodium lauryl sarcosine) and incubated at 50°C with gentle agitation for 75–90 minutes, with manual inversion every 10 minutes. Following lysis, 1.4 mL 5M NaCl was added (final concentration 1.5 M), and samples were gently mixed and centrifuged at 2,755 *g* for 15 minutes at 4°C to pellet the debris. The gDNA-containing supernatants were transferred to new tubes, and the DNA was precipitated by adding 2.5 volumes cold absolute ethanol (~11 mL). Tubes were gently rocked for 1-2 min until DNA strands were visible, followed by mixing on a HulaMixer (ThermoFisher Scientific, Waltham, MA, USA) for an additional 10 minutes. The DNA was pelleted by centrifugation at 4,260 *g* for 15 minutes at 4°C. The DNA pellets were washed with 10 mL cold 70% ethanol, mixed again on the HulaMixer for ten minutes, centrifuged, and air-dried horizontally in a laminar flow hood for 30 minutes. Any residual ethanol was carefully removed with a long-stem swab. A total of 200 µL “low-TE” buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) was added to each tube, mixed by gently flicking the tubes and stored at 4°C for 2–4 days to allow complete resuspension of the HMW gDNA before

quantification using the Quant-iT™ PicoGreen® dsDNA assay kit (ThermoFisher Scientific, Burlington, ON, Canada).

DNA quality was assessed by pulsed-field gel electrophoresis (PFGE). DNA samples were loaded on a 1% UltraPure™ agarose gel (ThermoFisher Scientific) in 0.5× TBE buffer and resolved on a CHEF DR-II electrophoresis system (Bio-Rad, Mississauga, Ontario, Canada) at 6 V/cm with a pulse time from 5 to 15 seconds over 16 hours at 12.5°C. Lambda PFG Ladder and Midrange PFG Marker (New England Biolabs, Ipswich, MA, USA) were used as molecular size standards.

2.5.3 Sequencing

Multiple library types were constructed depending on the species: (1) Oxford Nanopore Technologies (ONT) long-read, (2) PacBio SMRTbell, (3) Illumina short-read, (4) Hi-C and (5) Omni-C. Libraries were sequenced on the corresponding platforms to achieve the following coverages: 21–63x for ONT, 3–6x for PacBio HiFi, 10–23x for Illumina short reads, 10–40x for Hi-C, and 9–23x for Omni-C (Appendix 2).

2.5.3.1 Oxford Nanopore Technologies (ONT)

ONT 1D ligation libraries were prepared using the LSK109 kit, following the manufacturer's protocol with minor modifications to maximize retention of ultra-long DNA fragments. The HMW gDNA was size-selected using a BluePippin system (Sage Science, MA, USA) with the high-pass protocol to enrich for fragments larger than 30 kb. Following size selection, magnetic bead purification was used to clean and concentrate the DNA. The concentrations were determined using a Qubit 2.0 Fluorometer (ThermoFisher Scientific, Burlington, ON, Canada), and integrity was assessed using a TapeStation 2200 system (Agilent Technologies, Mississauga, ON, Canada). The samples were stored at 4°C until library preparation. For each library, 1.2–1.5 µg of end-

repaired, size-selected HMW gDNA was used. Sequencing was performed on a PromethION (Oxford Nanopore Technologies, Oxford, UK) using R9 flow cells with high accuracy base calling.

2.5.3.2 PacBio HiFi

For PacBio sequencing, HMW gDNA was sent to the Centre d'expertise et de services Génome Québec (Montréal, QC, Canada). PacBio gDNA libraries were prepared following the Pacific Biosciences SMRTbell® prep kit 3.0 protocol for whole genome and metagenome library construction. The HMW gDNA was sheared using the Diagenode Megaruptor 3 instrument (Diagenode Inc., Denville, NJ, USA). DNA damage repair, end repair, and SMRTbell adapter ligation steps were performed according to the manufacturer's instructions with SMRTbell Template Prep Kit 3.0 reagents (Pacific Biosciences). Libraries were size-selected using a diluted AMPure PB bead protocol. Sequencing primers (version 3.2) were annealed to the libraries, and the Sequel II 2.2 polymerase was bound to the templates. Libraries were cleaned using SMRTbell cleanup beads following the SMRTlink v11 calculator protocol. Sequencing was performed on a PacBio Sequel II instrument at a loading concentration of 80 pM using the adaptive loading protocol, Sequel II Sequencing Kit 2.0, SMRT Cell 8M, with a 2-hour pre-extension time and a 30-hour movie collection time.

2.5.3.3 Illumina short reads

HMW gDNA at 50ng/µl was sent to the Centre d'expertise et de services Génome Québec (Montréal, QC, Canada). Libraries were prepared from 700 ng gDNA using the TruSeq DNA PCR-Free Library Preparation Kit (Lucigen, Wisconson, USA), following the manufacturer's recommended protocol. Library quantification was performed using the KAPA Library Quantification Kit—Complete (Universal) (Kapa Biosystems). The average fragment size of each

library was determined using the Agilent 5300 Fragment Analyzer. Libraries were normalized, pooled, and denatured in 0.02 N NaOH, then neutralized with HT1 buffer. The pooled libraries were loaded at 400 pM onto an Illumina NovaSeq 6000 S4 flow cell using the Xp workflow according to the manufacturer's instructions. Sequencing was conducted in paired-end mode (2 × 150 bp cycles). A 1% PhiX control library served as internal quality control. Base calling and demultiplexing were performed using BCL Convert version 4.2.4, generating high-quality FASTQ files for downstream analysis.

2.5.3.4 Hi-C

Hi-C libraries were prepared following the protocol described by Padmarasu et al. (2019). Fresh young leaf tissue was crosslinked with formaldehyde, and chromatin was digested using the restriction enzyme *Dpn*II. The restricted DNA was blunt-end repaired and proximity ligation was performed, incorporating biotinylated nucleotides to enrich for ligation junctions. The resulting DNA was purified and used for library construction following the Illumina DNA Prep protocol (https://support-docs.illumina.com/LP/IlluminaDNAPrep/Content/LP/FrontPages/IlluminaDNAPrep.htm). Libraries were sent to the Centre d'expertise et de services Génome Québec (Montréal, QC, Canada) for sequencing.

At the sequencing center, libraries were indexed using TruSeq LT single indices (Illumina), PCR amplified and quantified using the KAPA Library Quantification Kit—Complete (Universal) (Kapa Biosystems), and average fragment size was confirmed on an Agilent 5300 Fragment Analyzer. Libraries were normalized, pooled (3–4 libraries per pool), denatured in 0.02 N NaOH, and neutralized using a pre-load buffer. The pooled libraries were loaded at 200 pM onto an

Illumina NovaSeq 6000 S4 flow cells. Sequence and processing were as described for Illumina short reads above.

2.5.3.5 Omni-C

Young leaf tissue was harvested from seedlings at the 2–4 leaf stage, following a dark treatment of approximately 72 hours. A total of 300 mg of fresh leaf tissue was collected and immediately frozen in liquid nitrogen. Omni-C libraries were prepared using the Dovetail® Omni-C® Kit (Cantata Bio, Scotts Valley, CA, USA), following the manufacturer's user guide with plant-specific modifications. To obtain a good fragment size distribution, the Nuclease Enzyme Mix (Nucl C) was further diluted 2-4 folds relative to the standard Nucl C working concentration. Final Omni-C libraries were quantified using the Qubit™ dsDNA High Sensitivity Assay Kit (ThermoFisher Scientific), and fragment size distributions were assessed using a TapeStation 4200 system with High Sensitivity D5000 ScreenTape (Agilent Technologies). Libraries were sequenced at the Centre d'expertise et de services Génome Québec (Montréal, QC, Canada) using the same Illumina protocol described for Hi-C libraries above, including indexing, PCR amplification, quantification, and fragment-size assessment. Sequencing was performed on an Illumina NovaSeq 6000 S4 flow cell in 2 × 150 bp paired-end mode across two lanes: Lane 1 included libraries for *Ae. geniculata*, *Ae. juvenalis*, and *Ae. vavilovii*; Lane 2 included libraries for *Ae. crassa* subsp. *crassa* (6x), *Ae. vavilovii*, and *Ae. neglecta* subsp. *recta*. A 1% PhiX library was used as sequencing control, and raw data processing was performed as described for Hi-C libraries.

2.5.4 Genome assembly

ONT long reads were assembled using the SMARTdenovo v1.0.0 (Liu et al., 2021) for all accessions except for *Ae. crassa* ssp. *crassa* (4x), *Ae. columnaris*, and *Ae. triuncialis*, which were assembled with hifiasm v0.19.8 (Cheng et al., 2021) due to suboptimal assemblies with

SMARTdenovo. Initial draft assemblies were first polished with PacBio long reads using Inspector v1.0.1 (Chen et al., 2021), followed by a second round of polishing with Illumina short reads using HyPo v1.0.3 (Kundu et al., 2019). Hi-C data were used to scaffold the polished assemblies with YaHS v1.2.2 (Zhou et al., 2022). Upon Hi-C scaffolding, assemblies of *Ae. geniculata*, *Ae. vavilovii*, *Ae. crassa* subsp. *crassa* (6x), *Ae. juvenalis*, and *Ae. neglecta* subsp. *recta* were more fragmented than the other 13; hence, Omni-C data were generated for these species, and scaffolding was performed with the HapHiC v1.0.6 (Zeng et al., 2024). Following scaffolding and contamination check for chloroplast and mitochondria sequences, all assemblies were further screened for potential contamination as part of the NCBI genome submission process, and any identified contaminant sequences were removed prior to final submission. Assembly completeness and quality were assessed using BUSCO v5.2.2 (Manni et al., 2021), with the embryophyta_odb10 database comprising 1,614 conserved single-copy orthologs. LTR assembly quality was additionally assessed using the LTR Assembly Index (LAI) (Ou et al., 2018). LAI scores were calculated using the corresponding genome assembly, EDTA-identified intact LTR retrotransposon list, and EDTA transposable element annotation file as inputs.

2.5.5 Gene annotation

Ab initio gene prediction of all *Aegilops* genome assemblies was performed using Helixer v0.3.5, a deep learning-based tool for gene identification, using the land_plant lineage model with a peak threshold of 0.9 (Stiehler et al., 2020). Predicted gene models were subsequently classified into high-confidence (HC) and low-confidence (LC) categories based on sequence homology to curated protein databases. All predicted gene models were compared against UniPoa (Poaceae database of annotated proteins from the UniProt database), the UniMag database (validated Magnoliophyta proteins from SwissProt) (Boeckmann et al., 2003), and PTREP (a database of hypothetical

proteins derived from the non-redundant set of transposable elements in TREP [Triticeae Repeat Sequence]) (Wicker et al., 2002). Sequence homology to each database was determined using BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with e-value < $10e^{-5}$. Classification into HC and LC gene categories was based on the best hit obtained using the following criteria: (i) alignments with at least 95% overlap between query and subject sequences for UniPoa and UniMag, and (ii) alignments with at least 95% query coverage for PTREP. Gene models were classified as HC if they had a significant hit in UniPoa and/or UniMag and were not present in PTREP. Gene models that did not meet these criteria were designated as LC.

2.5.6 Orthogroup inference and pangenome analysis

Protein sequences derived from the predicted gene models of all 25 *Aegilops* species were used for orthogroup inference using OrthoFinder v3.1.0 (Emms and Kelly, 2019). Orthogroups were defined as sets of genes descended from a single gene in the last common ancestor. Based on their distribution across species, orthogroups were classified into three categories: (i) core orthogroups, present in all 25 species; (ii) dispensable orthogroups, present in two or more but not all species; and (iii) species-specific orthogroups, present in only one species. Corresponding genes within each category were assigned based on their inclusion in these orthogroups.

2.5.7 Phylogenetic tree construction and divergence time estimation

Phylogenetic relationships were inferred from single-copy orthologs identified using OrthoFinder v3.1.0 (Emms and Kelly, 2019). For polyploid species, subgenomes were separated and analyzed as independent entries. The dataset comprised 36 *Aegilops* genomes and subgenomes generated in this study, along with 27 previously published *Aegilops* and *Triticum* genomes/subgenomes. The outgroup species *Hordeum vulgare* (Beier et al., 2017), *Thinopyrum elongatum* (Wang et al., 2020), and *Brachypodium distachyon* (Vogel et al., 2010), were included to root the phylogenetic

tree. These species were chosen based on their evolutionary relationships with the Triticeae lineage. *Aegilops* species belong to the tribe Triticeae within the grass family Poaceae, and *Hordeum vulgare* (barley) and *Thinopyrum elongatum* (tall wheatgrass) are closely related Triticeae species that diverged from the *Triticum–Aegilops* lineage after their common ancestor. Therefore, they provide an appropriate phylogenetic reference for rooting relationships within Triticeae while remaining sufficiently divergent to avoid being placed within the ingroup. *Brachypodium distachyon*, belonging to the more distantly related tribe Brachypodieae, was included as an additional outgroup to provide an external reference for rooting the phylogeny and polarizing evolutionary relationships. Reference genome assemblies for the previously published *Triticum/Aegilops* species were *T. urartu* (Ling et al., 2018), *T. monococcum* TA10622 and TA299 (Ahmed et al., 2023), A and B subgenomes of *T. turgidum* (Zhu et al., 2019), A and G subgenomes of *T. timopheevii* (Grewal et al., 2024), A, B and D subgenomes of *T. aestivum* (IWGSC, 2021), *Ae. tauschii* (Wang et al., 2021), *Ae. speltoides* (Yang et al., 2023), *Ae. sharonensis*, *Ae. longissima, Ae. speltoides* (Avni et al., 2022), *Ae. speltoides*, *Ae. bicornis*, *Ae. longissima*, *Ae. searsii*, *Ae. sharonensis* (Li et al., 2022), *Ae. mutica* (Grewal et al., 2025), *Ae. comosa* (Li et al., 2024), *Ae. umbellulata* (Abrouk et al., 2023), *Ae. comosa*, *Ae. umbellulata*, *Ae. markgrafii* and *Ae. uniaristata* (Khadka et al. 2026, unpublished). Because of the large size and diversity of this 66 genome dataset, no strict set of universal single-copy orthologs was initially identified. To address this, orthologous groups were defined as follows: single-copy genes must be present in at least 80% of the genomes, while allowing the remaining 20% of genomes to contain multiple copies. In the latter cases, only the longest isoform was retained for analysis. This approach yielded a final set of 773 near single-copy orthologs. Protein sequences from each orthologous group were aligned using MAFFT v7.525 (Katoh and Standley, 2013), and the resulting alignments were back-

translated to coding sequences (CDSs) with PAL2NAL (Suyama et al., 2006). Finally, conserved regions of each CDS alignment were extracted using Gblocks v0.91b (Castresana, 2000).

For phylogenetic inference, the filtered CDS alignments were concatenated into a single supermatrix using PhyKIT (https://jlsteenwyk.com/PhyKIT/usage/index.html#create-concatenation-matrix). Phylogenetic trees were constructed using the maximum-likelihood method implemented in RAxML-NG v1.2.2 (Kozlov et al., 2019) with the GTR+I+GAMMA model and 1000 bootstrap replicates.

Divergence times were estimated using a relaxed molecular clock implemented in MCMCTree from the PAML v4.10.9 package (Dos Reis and Yang, 2019), using the correlated rate model (clock = 3) and the HKY85 substitution model. Because evolutionary rates differ across codon positions, the concatenated supermatrix was partitioned by codon position, treating the first, second, and third positions as independent partitions. The MCMC analysis was run for 10,000,000 iterations following a burn-in of 3,000,000 iterations. Convergence was assessed by verifying that the effective sample size (ESS) for all parameters exceeded 200. Date calibrations were applied to major nodes based on previously published estimates: (i) the divergence between *Brachypodium distachyon* and the *Triticum*/*Aegilops* lineage (mean 44.4 ± 3.53 MYA) (Marcussen et al., 2014), (ii) the divergence between *Hordeum vulgare* (barley) and *Triticum*/*Aegilops* (~13 MYA) (Gaut, 2002), and (iii) the divergence between the A- and B-genome lineages of wheat and *Aegilops* (~7 MYA) (Marcussen et al., 2014). For each dataset, two independent MCMC runs were performed to confirm consistency and convergence of the posterior estimates. Tree topologies were visualized using FigTree v1.4.4 (https://github.com/rambaut/figtree).

2.5.8 Identification of structural variation

Previously generated diploid *Aegilops* genome assemblies for *Ae. comosa*, *Ae. umbellulata*, *Ae. markgrafii*, *Ae. uniaristata* (Khadka et al. 2026, unpublished), *Ae. speltoides* (Avni et al., 2022) and *Ae*. *tauschii* (Wang et al., 2021) were used as reference genomes for structural variation (SV) analysis. For each polyploid genome, the corresponding diploid progenitor genomes served as references. Subgenomes within the polyploid assemblies were assigned based on gene collinearity with their respective diploid progenitor genomes. Coding sequences from the diploid progenitor assemblies were mapped to the polyploid assemblies using GMAP (Wu and Watanabe, 2005) with minimum identity and coverage thresholds of 70%. Gene collinearity information was processed with a custom Perl script to assign scaffolds to their corresponding subgenomes. The seven largest scaffolds from each subgenome were selected for SV analyses. Scaffolds were renamed based on synteny with the assigned reference genome and standardized to match the chromosome naming of the diploid progenitor species. Pairwise whole-genome alignments between subgenomes and their respective diploid progenitors were performed using MUMmer v3.23 with the parameters --maxmatch -l 500 (Marçais et al., 2018). The alignment output was processed with the command line show-coords -THrd to generate coordinate files, which were then used as input for SV detection using the SyRI v1.7.0 (Goel et al., 2019) pipeline with default settings. All SVs > 50 bp were tallied. Visualization of SV plots was performed using the plotsr tool v1.1.1 (Goel and Schneeberger, 2022).

2.5.9 Transposable element identification and LTR analysis

Transposable elements (TEs) were annotated using the Extensive *de novo* TE Annotator (EDTA) v2.2.2 pipeline**,** which performs genome-wide *de novo* identification and classification of repetitive sequences (Ou et al., 2019). LTR retrotransposons identified by EDTA were further

classified into lineages using TEsorter v1.4.7 (Zhang et al., 2022) with the rexdb-plant database (Neumann et al., 2019) and the 70-30-80 rule, assigning elements to specific superfamilies and lineages based on conserved protein domains detected through HMM-based searches. Insertion times of intact LTR retrotransposons were estimated from the divergence between their 5′ and 3′ LTR sequences, as implemented in EDTA, assuming a substitution rate of 1.3×10^{-8} substitutions per site per year, following estimates from rice (Ou et al., 2019). To ensure methodological consistency across all genomes included in the comparative analyses, transposable elements were re-annotated in all previously published genomes using the same EDTA pipeline and parameters applied to the newly assembled genomes in this study. Likewise, LTR lineage classification and insertion age estimation were performed using the same TEsorter and EDTA workflows for all genomes, minimizing methodological differences and enabling direct comparisons of TE composition and evolutionary dynamics across species.

2.5.10 Flow cytometry

Leaf tissue samples were collected from each of the *Aegilops* species at the 3-5 leaf stage and each of the size standard plants, namely pea (*Pisum sativum*) for diploid, rye (*Secale cereale*) and corn (*Zea mays*) for tetraploid and faba bean (*Vicia faba*) for hexaploid species. To ensure optimal preservation, the samples were stored in damp paper towels on ice until they were processed on the same day. Leaf pieces (~5 mm) of the sample and standard were chopped together with a razor blade in 750 µl cold lysis buffer LB01 (15 mM Tris, 2 mM Na_2EDTA, 0.5 mM spermine tetrahydrochloride, 80 mM KCl, 20 mM NaCl, 0.1% (v/v) Triton X-100, 15 mM β-mercaptoethanol) in a petri dish. Samples were filtered into glass test tubes through a 30 µm nylon mesh filter to which 50 µl of RNase (1 mg/mL) was added. The samples were then stained for 30 minutes in the dark with 250 µl propidium iodide (100 µg/mL in Galbraith buffer: 45 mM MgCl2,

20 mM MOPS, 30 mM sodium citrate, 0.1% (v/v) Triton X-100). Three replicates of each *Aegilops* species were run at low speed on three different days on a Gallios flow cytometer (Beckman Coulter, Mississauga, ON, Canada) using the procedure described in Martin et al. (2017). The DNA content of samples was estimated using a fluorescence 585/42 nm detector. All characteristic of the fluorescence area peaks from the flow cytometer List Mode Data (LMD) files, including the mean, coefficient of variation and nuclei counts were processed on R Studio using the flowPloidy package (Smith et al., 2018).

2.5.11 Data availability

All 18 genome assemblies generated in this study have been deposited at the National Center for Biotechnology Information under BioProject PRJNA1449007. Accession numbers are as follows: *Ae. sharonensis* (JCAHDG000000000), *Ae. longissima* (JCAHDH000000000), *Ae. bicornis* (JCAHDI000000000), *Ae. searsii* (JBXSOC000000000), *Ae. neglecta* subsp. *neglecta* (JBXSOD000000000), *Ae. cylindrica* (JBXSOE000000000), *Ae. ventricosa* (JBXSQR000000000), *Ae. triuncialis* (JBXSQS000000000), *Ae. peregrina* (JBXSQT000000000), *Ae. kotschyi* (JBXSQU000000000), *Ae. geniculata* (JBXSQV000000000), *Ae. crassa* subsp. *crassa* (4x) (JBXSQW000000000), *Ae. columnaris* (JBXSQX000000000), *Ae. biuncialis* (JBXSQY000000000), *Ae. vavilovii* (JBXUDD000000000), *Ae. crassa* subsp. *crassa* (6x) (JBXUOA000000000), *Ae. juvenalis* (JBXUOB000000000), and *Ae. neglecta* subsp. *recta* (JBXUOC000000000). ONT and PacBio HiFi long read and Illumina short read sequencing data are being deposited in the NCBI Sequence Read Archive (SRA) under the above-mentioned associated BioProject accessions along with the Illumina short read data generated from the Hi-C and Omni-C libraries. Gene annotations are also being deposited within the same BioProject records.

2.5.12 Code availability

All computational analyses were performed using established bioinformatics tools, with software versions, parameters, and analytical procedures documented in the Methods section to facilitate reproducibility. Detailed sample information and accession records are provided in Appendix 1.

Chapter 3

General discussion

The transition from early agriculture to modern crop production has been driven by continuous efforts to enhance yield, adaptability, and resilience in staple crops (Fuller et al., 2023; Pingali, 2012). Among these, bread wheat (*Triticum aestivum*) is one of the most widely cultivated crops globally and plays a central role in ensuring food security for a growing population (Shewry and Hey, 2015). However, the processes of polyploidization, domestication, and intensive breeding that have underpinned its success have also led to a reduction in genetic diversity limiting the capacity of modern wheat to respond to rapidly changing environmental conditions, including climate change, emerging pathogens, and abiotic stresses (Cavanagh et al., 2013; Charmet, 2011; Dubcovsky and Dvorak, 2007).

To address these limitations, increasing attention has been directed toward crop wild relatives as reservoirs of untapped genetic variation. Species within the genus *Aegilops*, which are closely related to wheat and have contributed directly to its evolutionary history, represent a valuable genetic resource (Schneider et al., 2008). These species harbor extensive genomic diversity that remains largely underexplored, despite their potential to provide beneficial alleles for traits such as stress tolerance, disease resistance, and environmental adaptability (Ahmadi et al., 2020; Bansal et al., 2017; Masoomi-Aladizgeh et al., 2015; Pradhan et al., 2012). Understanding the genomic architecture and evolutionary relationships among *Aegilops* species can be leveraged not only for elucidating patterns of genome evolution but also for facilitating their effective utilization in wheat improvement programs.

The evolutionary history of wheat is characterized by multiple hybridization and polyploidization events involving distinct diploid progenitors, resulting in the formation of tetraploid and hexaploid lineages (Marcussen et al., 2014; Salamini et al., 2002). While polyploidization enables the combination of diverse germplasm, it also triggers extensive genomic reorganization, including structural variation, chromosomal rearrangements and the activation and proliferation of transposable elements (Soltis et al., 2015; Vicient and Casacuberta, 2017). These processes collectively shape genome structure and function, yet their relative contributions and evolutionary dynamics remain incompletely understood, particularly across diverse wild relatives such as *Aegilops*.

Recent advances in long-read sequencing technologies, combined with chromosome conformation capture approaches such as Hi-C and Omni-C, have enabled the generation of highly contiguous, chromosome-scale genome assemblies for complex plant genomes (Cetin and Sefer, 2025; Guiglielmoni, 2025; Pucker et al., 2022). These developments provide an unprecedented opportunity to investigate genome organization, evolutionary relationships, and structural variation at high resolution, particularly in large and repeat-rich genomes where repeats constitute a substantial proportion of genomes.

In this context, the present study established a comprehensive genomic framework for multiple diploid and polyploid *Aegilops* species to investigate patterns of genome evolution across ploidy levels. By integrating high-quality genome assemblies, gene annotation, phylogenetic analyses, divergence time estimation, and analyses of structural variation and transposable elements, this thesis delivers a multidimensional understanding of genome diversification in wild wheat relatives. Importantly, the inclusion of both polyploid genomes and their corresponding

diploid progenitors creates a comparative framework to disentangle ancestral genomic features from changes that have occurred following polyploidization.

Specifically, this study addressed the following objectives:

- Generating and refining high-quality genome assemblies and gene annotations for a diverse set of *Aegilops* species using long-read sequencing and chromatin interaction data, thereby expanding genomic resources and enabling high-resolution comparative analyses.
- Reconstructing phylogenetic relationships and estimating divergence times among *Aegilops* species to resolve their evolutionary history and provide a temporal framework for genome diversification.
- Investigating patterns of structural variation and assessing syntenic relationships between polyploid subgenomes and their diploid progenitors to evaluate genome conservation and divergence following polyploidization.
- Characterizing the composition, abundance, and evolutionary dynamics of transposable elements, with a particular emphasis on LTR retrotransposons, and exploring their contribution to genome expansion and variation across diploid and polyploid genomes.

Together, these analyses provide an integrated perspective on genome evolution across *Aegilops* species, linking genome structure, evolutionary history, and sequence dynamics. The genomic resources and insights described in this thesis enhance our understanding of polyploid genome evolution and lay a foundation for the effective utilization of *Aegilops* genetic diversity in wheat improvement.

3.1 Genome resources for comparative and evolutionary genomics in *Aegilops*

A major outcome of this study is the generation of high-quality, near chromosome-scale genome assemblies and corresponding annotations for multiple *Aegilops* species. Despite their importance as key contributors to wheat evolution and reservoirs of genetic diversity, genome assemblies of polyploid *Aegilops* species have not been produced, with the exception of *Ae. ventricosa* which has relatively well-developed genomic resources (Liu et al., 2025). Thus, the assemblies generated in this study expand the genomic resources available for wild wheat relatives and complete the production of genome assemblies for all 25 *Aegilops* species. The improved contiguity and structural resolution of these assemblies enabled more accurate characterization of repetitive and structurally complex genomic regions, which are often poorly resolved, particularly in large grass genomes rich in transposable elements.

In addition, gene annotations provide a robust basis for comparative genomic analyses, enabling assessment of gene content and conservation across species. Collectively, these resources enable downstream analyses of genome evolution and provide a foundation for investigating phylogenetic relationships, structural variation, and transposable element dynamics in *Aegilops*.

3.2 Phylogenetic insights into *Aegilops* genome evolution and polyploid origins

Phylogenetic analyses of *Aegilops* species provided a clearer understanding of their evolutionary relationships and genomic affinities across diploid and polyploid lineages. The use of genome-wide data revealed well-resolved relationships among species, improving upon previous studies that were often limited by sparse molecular markers or partial genomic information (Huynh et al., 2019; Petersen et al., 2006; Yamane and Kawahara, 2005). Overall, the inferred phylogeny mostly aligned with established genome classifications, supporting the current understanding of lineage relationships within the genus (Kilian et al., 2011; Van Slageren, 1994). There were, however,

some unexpected clustering patterns, indicating that several subgenomes, including multiple M subgenomes, do not fully correspond to their inferred diploid ancestry. These divergences may reflect post-polyploidization evolutionary processes such as differential transposable element expansion among subgenomes, homoeologous exchanges leading to sequence reshuffling between related chromosomes, and independent subgenome-specific evolutionary trajectories following polyploid formation. Such processes can progressively obscure the signal of the extant diploid progenitors and result in subgenomes that cluster outside their expected ancestral lineages.

Divergence time estimates provided a temporal perspective on species diversification within the phylogeny. Together, the phylogenetic and divergence time analyses constitute a refined evolutionary framework for *Aegilops*, which is essential for contextualizing patterns of structural variation and transposable element dynamics observed across diploid and polyploid genomes.

3.3 Structural variation and genome remodeling during polyploid evolution

Structural variation (SV) represents a major source of genomic diversity in plants, contributing to genome evolution through changes in gene order, copy number and chromosome structure, and playing a key role in adaptation and speciation (Gorkovskiy and Verstrepen, 2021; Lye and Purugganan, 2019; Saxena et al., 2014). Analysis of structural variation (SV) across *Aegilops* genomes revealed substantial genome divergence alongside overall conservation of synteny between polyploid subgenomes and their diploid progenitors. While many comparisons retained broad structural similarities, the presence of numerous insertions, deletions, duplications, and rearrangements highlights extensive genome remodeling following polyploidization. Importantly, patterns of structural variation were closely aligned with phylogenetic relationships. Subgenomes that clustered with their diploid progenitors showed higher degrees of synteny and more detectable

structural variants, whereas subgenomes that were phylogenetically more distant or grouped with alternative lineages were less syntenic and had fewer detectable SVs when compared to their progenitors. As such, the extent of detectable structural variation depends on the evolutionary relatedness between genomes and reflects lineage-specific patterns of divergence. Overall, structural variation in *Aegilops* is shaped by both ancestral divergence and post-polyploidization genome restructuring. This close correspondence between SV patterns and phylogenetic relationships underscores the importance of evolutionary context in interpreting genome structural variation.

3.4 Role of transposable elements in shaping *Aegilops* genome evolution

Transposable elements (TEs) are major drivers of genome evolution in plants, contributing to genome size expansion, structural variation, and regulatory innovation (Bennetzen and Wang, 2014; Wicker et al., 2018). In grass genomes, including wheat and its wild relatives, TEs constitute the largest genomic fraction and play a central role in shaping genome architecture over evolutionary time (Wicker et al., 2018). In this study, genome-wide TE analyses revealed that *Aegilops* genomes are highly TE-rich, with *Gypsy* and *Copia* LTR retrotransposons being the predominant components, consistent with other species of the *Triticum*/*Aegilops* complex (Avni et al., 2022; Li et al., 2022; Liu et al., 2025; Wicker et al., 2018).

Comparisons between polyploid subgenomes and diploid progenitors revealed largely conserved LTR landscapes, but subgenome-specific differences were observed in certain TE families, indicating that TE dynamics are partly inherited, but are also shaped by post-polyploid evolutionary processes. Subgenome-level analyses within species further indicated that while most LTR clades are homogenized, certain families display localized expansions or contractions, with

the *Gypsy CRM* lineage consistently showing subgenome-specific expansion, suggesting preferential activity or retention.

LTR insertion time analyses determined TE evolutionary events, with *Copia* elements associated with more recent activity and *Gypsy* elements reflecting older amplification events. This pattern reflects differences in the accumulation and removal of TE families over time, as influenced by their activity and genome silencing processes (Senerchia et al., 2013). Overall, these findings highlight TEs as key contributors to genome diversification in *Aegilops*, influencing both genome expansion and subgenome differentiation following polyploidization.

3.5 An evolutionary model of subgenome evolution following polyploidization

One of the major outcomes of this study is the establishment of an integrated evolutionary framework describing the evolutionary paths of *Aegilops* subgenomes following polyploidization. By combining genome size comparisons, gene content analyses, phylogenetic relationships, structural variation, and transposable element dynamics across all 25 *Aegilops* species, it became evident that polyploid genome evolution did not follow a single uniform trajectory. Instead, subgenomes exhibited distinct evolutionary patterns shaped by different degrees of genome contraction, gene fractionation, structural rearrangement, and TE turnover.

A pervasive pattern observed across nearly all polyploid species was genome downsizing relative to their diploid progenitors. Although polyploidization initially combines complete chromosome sets from multiple progenitors, long-term evolution of co-resident subgenomes was generally associated with their size reduction. This pattern is consistent with diploidization processes previously reported in wheat and other polyploid plants, where sequence elimination and genome restructuring progressively reduce redundant genomic content following whole-

genome duplication (Eilam et al., 2008; Leitch and Bennett, 2004). Importantly, the extent of downsizing differed substantially among subgenomes, indicating lineage-specific evolutionary responses after polyploidization.

Subgenome-level analyses revealed that reductions in genome size were frequently accompanied by reductions in gene content, indicating coordinated contraction of both genic and non-genic genomic components. The strongest effects were observed in the M subgenomes, which consistently showed extensive genome downsizing, substantial gene loss, high structural divergence, and pronounced TE variation relative to the diploid progenitor *Ae. comosa*. These patterns suggest that M subgenomes experienced extensive post-polyploidization remodeling and fractionation. Phylogenetic analyses further supported this interpretation, as several M subgenomes no longer clustered closely with their presumed diploid donor and instead grouped with co-resident subgenomes, indicating prolonged genomic interactions and evolutionary divergence. Together, these observations suggest that the M lineage followed one of the most dynamic evolutionary trajectories within *Aegilops* polyploids.

The D subgenomes also displayed substantial evolutionary variability, although their patterns differed from those of the M lineage. Most D subgenomes exhibited reduced gene content relative to diploid *Ae. tauschii*, but several LTR lineages showed strong expansion. In particular, the *Gypsy Reina* lineage expanded consistently across all D subgenomes, suggesting shared D-genome-specific LTR dynamics following polyploidization. Despite maintaining overall phylogenetic affinity with the *Ae. tauschii* progenitor species, the D subgenomes displayed considerable structural and LTR diversification, indicating that substantial genomic remodeling occurred even in lineages retaining clear donor relationships.

In contrast, the S and C subgenomes followed comparatively conservative evolutionary trajectories. Both genome groups retained genome sizes and gene contents broadly similar to their diploid progenitors while also showing relatively stable LTR landscapes. LTR composition in S subgenomes remained highly similar to diploid relatives, with comparatively limited evidence of large-scale LTR expansion or contraction. Similarly, the C subgenome of *Ae. cylindrica* remained highly conserved relative to diploid *Ae. markgrafii*. These patterns suggest that some subgenomes experienced relatively limited post-polyploid restructuring and retained greater ancestral genomic stability. This conservation may reflect differences in the evolutionary dynamics among subgenomes following polyploidization, where certain subgenomes may have experienced stronger selective constraints maintaining gene content and genome structure. In addition, reduced transposition activity or more efficient regulation of repetitive elements may have limited genome expansion and structural changes in these subgenomes.

The U and N subgenomes exhibited intermediate evolutionary patterns compared to the above-described cases. Both displayed substantial variation in genome size, together with moderate lineage-specific variation in gene content and LTR abundance, including lineage-specific expansions and contractions, rather than consistent directional trends. For example, *CRM* elements contracted in U subgenomes of *Ae. biuncialis* and *Ae. geniculata*, whereas other LTR lineages remained comparatively stable. These observations indicate that TE evolution following polyploidization is context dependent and differs among LTR lineages even within the same subgenomic background.

TE analyses further revealed that different LTR lineages contributed unequally to genome evolution across subgenomes. Across all genomes, *Angela* represented the dominant *Copia*

lineage, whereas *Retand* and *Athila* were the predominant *Gypsy* elements, indicating that these lineages constitute the major components of the *Aegilops* repeat landscape. However, individual lineages followed distinct evolutionary trajectories after polyploidization. For example, *Angela* consistently contracted in most M subgenomes, whereas TAR showed localized expansion in several M lineages. *CRM* exhibited highly contrasting patterns among genomes, with contractions in certain U subgenomes but expansions in selected M and D subgenomes. These lineage-specific responses indicate that post-polyploid LTR evolution is shaped not only by subgenome identity but also by intrinsic differences among TE families in activity, retention, and removal.

Temporal analyses of LTR insertions further suggest that TE superfamilies contributed differently across evolutionary timescales. *Copia* elements were generally associated with more recent activity, whereas *Gypsy* elements reflected older amplification events. Together with observed contractions in major TE families, these findings suggest that both TE amplification and TE removal contributed to genome remodeling after polyploidization. The association between strong TE turnover and extensive genome downsizing, particularly in M subgenomes, indicates that TE dynamics likely contributed to genome contraction alongside gene fractionation and structural rearrangements.

Overall, the results support a model in which *Aegilops* subgenomes followed distinct evolutionary trajectories after polyploidization rather than converging toward a common genomic state. Some subgenomes, particularly S and C, remained comparatively conserved, whereas others, especially M and portions of the D lineage, underwent extensive restructuring characterized by genome downsizing, gene loss, structural variation, and dynamic TE turnover. These findings indicate that polyploid genome evolution in *Aegilops* is lineage dependent and shaped by complex

interactions among gene retention, structural variation, TE activity, and long-term subgenome coexistence. Although this framework provides a comprehensive model for interpreting lineage-dependent patterns of subgenome evolution in *Aegilops*, it represents a hypothesis generated from comparative genomic observations rather than a formally tested evolutionary model. Future studies incorporating multiple accessions per species, broader taxonomic sampling, population-level variation, and functional analyses will be required to further evaluate the mechanisms underlying the differential evolutionary trajectories of *Aegilops* subgenomes.

3.6 Limitations of the study

Although this study provides a comprehensive comparative framework for understanding genome evolution across *Aegilops* species, several limitations should be considered when interpreting the evolutionary patterns identified. First, the comparative analyses were based primarily on a single accession per species, which limits the ability to distinguish evolutionary changes associated specifically with polyploidization from pre-existing variation among ancestral populations, donor lineage divergence, or accession-specific differences. Therefore, observed differences in genome size, gene content, structural variation, and transposable element composition should be interpreted as comparative patterns rather than definitive evidence of changes that occurred exclusively after polyploid formation. Future studies incorporating multiple accessions per species and population-level sampling will be essential to separate lineage-specific evolutionary changes from standing genetic variation.

A second limitation concerns the inference of gene loss, genome contraction, and structural remodeling following polyploidization. Comparisons with a single extant diploid relative provide an approximation of ancestral genomic states but cannot fully capture the diversity of ancestral

genomes that contributed to polyploid formation. Differences between diploid and polyploid genomes may therefore reflect divergence between sampled lineages rather than direct consequences of polyploidization. Expanding analyses to include additional representatives of ancestral lineages, and branch-aware comparative methods will help better resolve the timing and direction of genomic changes.

Differences in gene content and annotation completeness among genomes may also be influenced by technical factors, including assembly quality, annotation methodology, and the availability of transcriptomic evidence. While several previously published *Aegilops* genomes were annotated using transcriptomic and homology-based evidence, gene annotation in this study relied primarily on *ab initio* gene prediction. Consequently, some differences observed in comparative gene content analyses may reflect methodological variation in addition to biological differences. Future genome annotation efforts incorporating accession-matched transcriptome data, homology-based evidence, and functional validation will further improve annotation accuracy and facilitate more robust comparisons among *Aegilops* genomes.

The interpretation of phylogenetic relationships also requires consideration of methodological limitations. Concatenated phylogenetic analyses provide a dominant evolutionary signal but may not fully resolve complex evolutionary histories involving incomplete lineage sorting, hybridization, or reticulate evolution. Future analyses incorporating gene-tree/species-tree approaches and explicit tests of gene-tree discordance will provide additional resolution of these evolutionary relationships.

Interpretation of LTR insertion times requires caution because these estimates rely on a substitution rate constant, which may vary among species, genomic regions, and evolutionary

lineages. In addition, processes such as gene conversion between the two LTR copies can reduce sequence divergence and lead to underestimation of insertion times. Therefore, estimated LTR insertion times should be interpreted as approximate indicators of the relative timing of TE accumulation rather than precise dates of individual TE bursts or transpositional activity. Additional analyses incorporating multiple accessions, population-level variation, and complementary approaches for assessing TE dynamics will be required to better understand the mechanisms underlying repeat evolution.

Together, these limitations highlight that the evolutionary framework proposed in this study represents a testable model based on comparative genomic patterns. Future investigations integrating broader sampling, population genomics, transcriptomic evidence, functional analyses, and complementary evolutionary approaches will be necessary to distinguish correlation from causation and to further elucidate the mechanisms shaping subgenome evolution following polyploidization.

3.7 Future perspectives

The genomic resources and comparative analyses generated in this study provide a strong foundation for advancing both fundamental and applied research in *Aegilops* and wheat evolution. However, several avenues remain to be explored to fully leverage these findings and further deepen our understanding of polyploid genome dynamics.

One important direction is the functional characterization of genomic variation identified across *Aegilops* species. While this study highlights extensive structural variation, lineage-specific gene content differences, and transposable element dynamics, linking these genomic features to gene function and phenotypic traits remains a key challenge. Integrating transcriptomic,

epigenomic, proteomic, and phenotypic datasets will help identify structural variants and TE insertions that influence gene expression, regulatory networks, and traits associated with stress tolerance, adaptation, and agronomic performance. Such approaches will enable the prioritization of candidate genes and genomic regions for functional validation and provide a framework for translating evolutionary genomic discoveries into potential breeding applications.

From an applied perspective, translating and leveraging genomic insights into practical breeding strategies remains a major goal. The importance of *Aegilops* species in wheat improvement is already well established through the successful transfer of numerous resistance genes conferring resistance against major pests and diseases. These include key genes for powdery mildew resistance such as *Pm29* from *Ae. geniculata* (Zeller et al., 2002) and *Pm57* from *Ae. searsii* (Liu et al., 2017), and stem rust resistance genes such as *Sr47* (*Ae. speltoides*) (Klindworth et al., 2012) and *Sr53* (*Ae. geniculata*) (Liu et al., 2011). In addition, *Aegilops* species have contributed important leaf rust resistance genes, including *Lr9* (*Ae. umbellulata*) (Wang et al., 2023) and *Lr54* (*Ae. kotschyi*) (Marais et al., 2005), as well as stripe rust resistance gene *Yr42* (*Ae. neglecta*) (Marais et al., 2009). They have also provided resistance to other important stresses, such as cyst nematode resistance (*CreX*, *Ae. peregrina*) (Barloy et al., 2007) and Hessian fly resistance (*H30*, *Ae. triuncialis*) (Martín-Sánchez et al., 2003). Traditionally, beneficial genes from wild relatives were transferred into wheat through interspecific hybridization followed by embryo rescue, chromosome doubling, and successive backcrossing, often combined with cytogenetic or molecular marker-based selection to retain the desired genomic segments. While effective, these approaches are frequently limited by sexual incompatibility barriers and the co-introduction of large linked genomic regions, resulting in linkage drag that can negatively affect agronomic performance (Dempewolf et al., 2012). Genome editing technologies such as CRISPR/Cas have

the potential to bypass several of these limitations by enabling targeted modification of specific genes identified in wild relatives, rather than transferring large chromosomal segments (Pan et al., 2021; Wang and Doudna, 2023). This allows for more precise utilization of beneficial alleles and reduces the burden of linked undesirable traits. However, genome editing is not yet a fully comprehensive solution, as most current applications are limited to gene knockouts or targeted sequence modifications. Emerging approaches such as base editing and prime editing are beginning to expand this scope by enabling more precise nucleotide-level changes (Anzalone et al., 2019; Komor et al., 2016), offering increasing potential for the direct functional deployment of alleles discovered in wild species, providing the presence of a suitable ortholog target in wheat. Building on this, the genomic resources generated in this study—particularly the complete set of high-quality assemblies, gene annotations, structural variation catalogs, and TE landscapes across *Aegilops* species—provide a foundation for systematically identifying genomic regions underlying agronomically important traits. Integration of these resources with phenotypic datasets and wheat genomic resources will enable the discovery of candidate genes, functional alleles, and trait-associated structural variants that can be developed into molecular markers for marker-assisted selection (MAS), genomic prediction, or targeted genome editing approaches. For example, lineage-specific genes or alleles associated with disease resistance, stress adaptation, or environmental resilience can be prioritized for functional validation and potential deployment in wheat improvement programs. Similarly, structural variants affecting gene dosage, regulation, or genomic architecture represent potential targets for further investigation as sources of beneficial variation. Together, these approaches will facilitate more precise utilization of *Aegilops* genetic diversity while reducing reliance on large chromosomal introgressions that often introduce undesirable linked regions.

Another promising avenue is the exploration of regulatory evolution following polyploidization. Polyploid genomes are known to undergo extensive regulatory reprogramming, including changes in gene expression dominance, subgenome bias, and epigenetic modifications. Future studies focusing on chromatin accessibility, DNA methylation, and histone modifications across subgenomes have the potential to elucidate the mechanisms underlying subgenome differentiation and stabilization. In particular, understanding transposable elements' contributions to regulatory changes and genome plasticity is an obvious next step and will likely be the topic of further investigations.

The expanded set of high-quality genome assemblies also enables more detailed analyses of population-level variation within *Aegilops* species. Incorporating multiple accessions per species will allow the construction of pangenomes, providing insights into core and dispensable genomic components and revealing the extent of intraspecific diversity. Such approaches will be critical for identifying rare or lineage-specific alleles that may have been overlooked in single-reference genome studies.

Finally, continued improvements in sequencing technologies and scaffolding approaches are predicted to improve the completeness and accuracy of genome assemblies, including the resolution of highly repetitive and structurally complex regions such as centromeres and telomeres. Gap-free, telomere-to-telomere assemblies across multiple *Aegilops* species will further enhance the resolution with which we can study genome organization and evolution.

Overall, the integration of high-quality genomic resources with functional and applied studies will be essential for unlocking the full potential of *Aegilops* diversity. Such efforts will not only advance our understanding of polyploid genome evolution but also contribute to the

development of more resilient and adaptable wheat varieties in the face of global agricultural challenges.

3.8 Conclusion

This study provides a comprehensive genomic and evolutionary analysis of multiple diploid and polyploid *Aegilops* species, significantly expanding the available genomic resources for wild relatives of wheat. By generating high-quality, near chromosome-scale genome assemblies and corresponding gene annotation, this work establishes a robust framework for investigating genome structure, evolution, and diversity across the entire *Aegilops* genus. These resources addressed a major gap in the genomic characterization of *Aegilops*, particularly for polyploid species, and enabled high-resolution comparative analyses that were previously not feasible.

Phylogenetic relationships and divergence time estimation offered a refined view of evolutionary relationships within *Aegilops*, largely supporting established genome classifications while also revealing deviations in subgenome affinities. These findings highlight the complexity of polyploid genome evolution, demonstrating that subgenomes do not always retain a simple or direct correspondence with their diploid progenitors. Instead, they reflect dynamic evolutionary trajectories shaped by both ancestral divergence and lineage-specific processes following polyploidization.

Comparative analyses of structural variation revealed that, despite broad conservation of synteny between polyploid subgenomes and their diploid progenitors, *Aegilops* genomes have undergone extensive structural remodeling. The distribution of insertions, deletions, duplications, and rearrangements underscores the significant role of structural variation in genome diversification. Moreover, the strong correspondence between patterns of structural variation and

phylogenetic relationships emphasizes the importance of evolutionary context in interpreting genome-level differences.

Transposable element analyses further revealed that *Aegilops* genomes are highly dynamic and dominated by LTR retrotransposons, particularly *Gypsy* and *Copia* elements. While overall TE landscapes are largely conserved between diploid and polyploid genomes, subgenome-specific differences and evidence of differential expansion highlight the role of TEs in shaping genome structure following polyploidization. Temporal patterns of LTR activity indicate distinct waves of TE amplification, contributing to genome expansion and diversification over evolutionary time.

In summary, this study advances both the genomic resources and the conceptual understanding of polyploid genome evolution in the wild wheat relative species of the *Aegilops* genus. It highlights the dynamic nature of *Aegilops* genomes and underscores their potential as a rich reservoir of genetic variation for future wheat improvement.

References

Abbott, R.J., Ritchie, M.G., and Hollingsworth, P.M. (2008). Introduction. Speciation in plants and animals: pattern and process. Philos. Trans. R. Soc. B, Biol. Sci. **363**:2965-2969.

Abrouk, M., Wang, Y., Cavalet-Giorsa, E., Troukhan, M., Kravchuk, M., and Krattinger, S.G. (2023). Chromosome-scale assembly of the wild wheat relative *Aegilops umbellulata*. Sci. Data **10**:739.

Ahmadi, J., Pour-Aboughadareh, A., Fabriki Ourang, S., Khalili, P., and Poczai, P. (2020). Unraveling salinity stress responses in ancestral and neglected wheat species at early growth stage: A baseline for utilization in future wheat improvement programs. Physiol. Mol. Biol. Plants **26**:537-549.

Ahmed, H.I., Heuberger, M., Schoen, A., Koo, D.-H., Quiroz-Chavez, J., Adhikari, L., Raupp, J., Cauet, S., Rodde, N., Cravero, C., et al. (2023). Einkorn genomics sheds light on history of the oldest domesticated wheat. Nature **620**:830-838.

Akhunov, E.D., Akhunova, A.R., Anderson, O.D., Anderson, J.A., Blake, N., Clegg, M.T., Coleman-Derr, D., Conley, E.J., Crossman, C.C., and Deal, K.R. (2010). Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genom. **11**:702.

Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature **576**:149-157.

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., and White, J.W. (2015). Rising temperatures reduce global wheat production. Nat. Clim. Chang. **5**:143-147.

Avni, R., Lux, T., Minz-Dub, A., Millet, E., Sela, H., Distelfeld, A., Deek, J., Yu, G., Steuernagel, B., Pozniak, C., et al. (2022). Genome sequences of three *Aegilops* species of the section Sitopsis reveal phylogenetic relationships and provide resources for wheat improvement. Plant J. **110**:179-192.

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S.O., Gundlach, H., Hale, I., Mascher, M., Spannagl, M., Wiebe, K., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science (New York, N.Y.) **357**:93-97.

Baalbaki, R., Hajj-Hassan, N., and Zurayk, R. (2006). *Aegilops* species from semiarid areas of Lebanon: variation in quantitative attributes under water stress. Crop Sci. **46**:799-806.

Badaeva, E., Amosova, A., Samatadze, T., Zoshchuk, S., Shostak, N., Chikida, N., Zelenin, A., Raupp, W., Friebe, B., and Gill, B. (2004). Genome differentiation in *Aegilops*. 4. Evolution of the U-genome cluster. Plant System. Evol. **246**:45-76.

Badaeva, E.D., Chikida, N.N., Belousova, M.K., Ruban, A.S., Surzhikov, S.A., and Zoshchuk, S.A. (2021). A new insight on the evolution of polyploid *Aegilops* species from the complex Crassa: Molecular-cytogenetic analysis. Plant System. Evol. **307**:3.

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

Barloy, D., Lemoine, J., Abelard, P., Tanguy, A.M., Rivoal, R., and Jahier, J. (2007). Marker-assisted pyramiding of two cereal cyst nematode resistance genes from *Aegilops variabilis* in wheat. Mol. Breed. **20**:31-40.

Baum, B.R., Edwards, T., Mamuti, M., and Johnson, D.A. (2012). Phylogenetic relationships among the polyploid and diploid *Aegilops* species inferred from the nuclear 5S rDNA sequences (Poaceae: Triticeae). Genome **55**:177-193.

Beier, S., Himmelbach, A., Colmsee, C., Zhang, X.-Q., Barrero, R.A., Zhang, Q., Li, L., Bayer, M., Bolser, D., Taudien, S., et al. (2017). Construction of a map-based reference genome sequence for barley, *Hordeum vulgare* L. Sci. Data **4**:170044.

Bennetzen, J.L., and Wang, H. (2014). The contributions of transposable elements to the structure, function, and evolution of plant genomes. Ann. Rev. Plant Biol. **65**:505-530.

Bernhardt, N., Brassac, J., Dong, X., Willing, E.M., Poskar, C.H., Kilian, B., and Blattner, F.R. (2020). Genome-wide sequence information reveals recurrent hybridization among diploid wheat wild relatives. Plant J. **102**:493-506.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. **31**:365-370.

Butler, J.M. (2012). Chapter 12 - Single Nucleotide Polymorphisms and Applications. In Advanced Topics in Forensic DNA Typing: Methodology, J.M. Butler, ed. (Academic Press: San Diego), pp. 347-369.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17**:540-552.

Cavanagh, C.R., Chao, S., Wang, S., Huang, B.E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G.L., Akhunova, A., et al. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. USA. **110**:8057-8062.

Cetin, S., and Sefer, E. (2025). The significance of chromosome conformation capture in 3D genome architecture comprehension. Comput. Biol. Chem. **119**:108534.

CGIAR (2018). The CGIAR research program on wheat 2017 annual report https://hdl.handle.net/10568/97477.

Chaisson, M.J., Wilson, R.K., and Eichler, E.E. (2015). Genetic variation and the *de novo* assembly of human genomes. Nat. Rev. Genet. **16**:627-640.

Charmet, G. (2011). Wheat domestication: lessons for the future. C. R. Biol. **334**:212-220.

Chaudhary, H.K., Kaila, V., Rather, S.A., Badiyal, A., Hussain, W., Jamwal, N.S., and Mahato, A. (2013). Wheat. In Alien Gene Transfer in Crop Plants, Volume 2: Achievements and Impacts, (Springer: pp. 1-26.

Chaves, M.S., Martinelli, J.A., Wesp-Guterres, C., Graichen, F.A.S., Brammer, S.P., Scagliusi, S.M., da Silva, P.R., Wiethölter, P., Torres, G.A.M., and Lau, E.Y. (2013). The importance for food security of maintaining rust resistance in wheat. Food Sec. **5**:157-176.

Chen, N., Chen, W.-J., Yan, H., Wang, Y., Kang, H.-Y., Zhang, H.-Q., Zhou, Y.-H., Sun, G.-L., Sha, L.-N., and Fan, X. (2020). Evolutionary patterns of plastome uncover diploid-polyploid maternal relationships in Triticeae. Mol. Phylogenet. Evol. **149**:106838.

Chen, Y., Zhang, Y., Wang, A.Y., Gao, M., and Chong, Z. (2021). Accurate long-read *de novo* assembly evaluation with Inspector. Genome Biol. **22**:312.

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. Nat. Plants **4**:258-268.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature Methods **18**:170-175.

Colmer, T.D., Flowers, T.J., and Munns, R. (2006). Use of wild relatives to improve salt tolerance in wheat. J. Exp. Bot. **57**:1059-1078.

Dangel, A.W., Baker, B.J., Mendoza, A.R., and Yu, C.Y. (1995). Complement component C4 gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-K (C4) are a molecular clock of evolution. Immunogenetics **42**:41-52.

de Tomás, C., and Vicient, C.M. (2024). The genomic shock hypothesis: genetic and epigenetic alterations of transposable elements after interspecific hybridization in plants. Epigenomes **8**:2.

Dempewolf, H., Hodgins, K.A., Rummell, S.E., Ellstrand, N.C., and Rieseberg, L.H. (2012). Reproductive isolation during domestication. Plant Cell **24**:2710-2717.

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. Crop Sci. **57**:1070-1082.

Díaz, A., Zikhali, M., Turner, A.S., Isaac, P., and Laurie, D.A. (2012). Copy number variation affecting the Photoperiod-B1 and Vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*). PloS one **7**:e33234.

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. Cell **127**:1309-1321.

Dos Reis, M., and Yang, Z. (2019). Bayesian molecular clock dating using genome-scale datasets. Methods Mol. Biol. **1910**:309-330.

Dubcovsky, J., and Dvorak, J. (1995). Genome identification of the *Triticum crassum* complex (Poaceae) with the restriction patterns of repeated nucleotide sequences. Am. J. Bot. **82**:131-140.

Dubcovsky, J., and Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. Science (New York, N.Y.) **316**:1862-1866.

Dvorak, J., Terlizzi, P., Zhang, H.B., and Resta, P. (1993). The evolution of polyploid wheats: identification of the A genome donor species. Genome **36**:21-31.

Dvorak, J., Luo, M.C., Yang, Z.L., and Zhang, H.B. (1998). The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor. Appl. Genet. **97**:657-670.

Edet, O.U., Gorafi, Y.S.A., Nasuda, S., and Tsujimoto, H. (2018). DArTseq-based analysis of genomic relationships among species of tribe Triticeae. Sci. Rep. **8**:16397.

Eilam, T., Anikster, Y., Millet, E., Manisterski, J., and Feldman, M. (2008). Nuclear DNA amount and genome downsizing in natural and synthetic allopolyploids of the genera *Aegilops* and *Triticum*. Genome **51**:616-627.

Eilam, T., Anikster, Y., Millet, E., Manisterski, J., Sagi-Assif, O., and Feldman, M. (2007). Genome size and genome evolution in diploid Triticeae species. Genome **50**:1029-1037.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**:238.

FAO (2024). http://www.fao.org/faostat/en/#data/QCL/visualize (accessed on 8 January 2026).

Fedoroff, N. (2000). Transposons and genome evolution in plants. Proc. Natl. Acad. Sci. USA **97**:7002-7007.

Feldman, M., and Levy, A.A. (2012). Genome evolution due to allopolyploidization in wheat. Genetics **192**:763-774.

Feldman, M., Levy, A.A., Fahima, T., and Korol, A. (2012). Genomic asymmetry in allopolyploid plants: wheat as a model. J. Exp. Bot. **63**:5045-5059.

Feschotte, C., Jiang, N., and Wessler, S.R. (2002). Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet. **3**:329-341.

Flagel, L.E., and Wendel, J.F. (2009). Gene duplication and evolutionary novelty in plants. New Phytol. **183**:557-564.

Friebe, B. (1996). Chromosome banding and genome analysis in diploid and cultivated polyploid wheats. In Methods of genome analysis in plants, P.P. Jauhar, ed. (CRC Press: Boca Raton, FL), pp. 39-60.

Fuller, D.Q., Denham, T., and Allaby, R. (2023). Plant domestication and agricultural ecologies. Curr. Biol. **33**:R636-R649.

Gamazon, E.R., and Stranger, B.E. (2015). The impact of human copy number variation on gene expression. Brief. funct. genomics **14**:352-357.

Gaut, B.S. (2002). Evolutionary dynamics of grass genomes. New Phytologist **154**:15-28.

Gill, B., and Friebe, B. (2002). Cytogenetics, phylogeny and evolution of cultivated wheats. Bread wheat: improvement and production. Food and Agriculture Organization of the United Nations, Rome:71-88.

Glémin, S., Scornavacca, C., Dainat, J., Burgarella, C., Viader, V., Ardisson, M., Sarah, G., Santoni, S., David, J., and Ranwez, V. (2019). Pervasive hybridizations in the history of wheat relatives. Science Adv. **5**:eaav9188.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., and Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. Science (New York, N.Y.) **327**:812-818.

Goel, M., and Schneeberger, K. (2022). plotsr: visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics (Oxford, England) **38**:2922-2926.

Goel, M., Sun, H., Jiao, W.-B., and Schneeberger, K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. **20**:277.

Gorkovskiy, A., and Verstrepen, K.J. (2021). The role of structural variation in adaptation and evolution of yeast and other fungi. Genes (Basel) **12**:699.

Grewal, S., Yang, C.-y., Scholefield, D., Ashling, S., Ghosh, S., Swarbreck, D., Collins, J., Yao, E., Sen, T.Z., Wilson, M., et al. (2024). Chromosome-scale genome assembly of bread wheat's wild relative *Triticum timopheevii*. Sci. Data **11**:420.

Grewal, S., Yang, C.-y., Krasheninnikova, K., Collins, J., Wood, J.M.D., Ashling, S., Scholefield, D., Kaithakottil, G.G., Swarbreck, D., Yao, E., et al. (2025). Chromosome-level haplotype-resolved genome assembly of bread wheat's wild relative *Aegilops mutica*. Sci. Data **12**:438.

Guiglielmoni, N. (2025). *De novo* genome assembly using long reads and chromosome conformation capture. Methods Mol. Biol. **2935**:1-27.

Gustafsson, Ö., Kruså, M., Zencak, Z., Sheesley, R.J., Granat, L., Engström, E., Praveen, P.S., Rao, P.S.P., Leck, C., and Rodhe, H. (2009). Brown Clouds over South Asia: Biomass or Fossil Fuel Combustion? Science (New York, N.Y.) **323**:495-498.

Harlan, J.R. (1992). Crops and Man. American Society of Agronomy. Crop Science Society of America, Madison, Wisconsin **16**:63-262.

Harlan, J.R., de Wet, J.M., and Price, E.G. (1973). Comparative evolution of cereals. Evolution:311-325.

Haudry, A., Cenci, A., Ravel, C., Bataillon, T., Brunel, D., Poncet, C., Hochu, I., Poirier, S., Santoni, S., and Glémin, S. (2007). Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mole. biol. evol. **24**:1506-1517.

Helguera, M., Khan, I.A., Kolmer, J., Lijavetzky, D., Zhong-qi, L., and Dubcovsky, J. (2003). PCR assays for the *Lr37-Yr17-Sr38* cluster of rust resistance genes and their use to develop isogenic hard red spring wheat lines. Crop Sci. **43**:1839-1847.

Huang, Y.-C., Dang, V.D., Chang, N.-C., and Wang, J. (2018). Multiple large inversions and breakpoint rewiring of gene expression in the evolution of the fire ant social supergene. Proc. Biol. Sci **285**

Huddleston, J., and Eichler, E.E. (2016). An incomplete understanding of human genetic variation. Genetics **202**:1251-1254.

Huynh, S., Marcussen, T., Felber, F., and Parisod, C. (2019). Hybridization preceded radiation in diploid wheats. Mol. Phylogenet. Evol. **139**:106554.

IWGSC (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science (New York, N.Y.) **361**:eaar7191.

IWGSC (2021). Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. Plant J. **107**:303-314.

Jewell, D. (1979). Chromosome banding in *Triticum aestivum* cv. Chinese Spring and *Aegilops variabilis*. Chromosoma **71**:129-134.

Jewell, D., and Driscoll, C. (1983). The addition of *Aegilops variabilis* chromosomes to *Triticum aestivum* and their identification. Can. J. Genet. Cytol. **25**:76-84.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. **30**:772-780.

Kellogg, E.A. (2015). Flowering plants, monocots, poaceae (Springer).

Kihara, H. (1944). Discovery of the DD-Analyser, one of the ancestors of *Triticum vulgare*. Agric. Hortic. **19**:13-14.

Kijima, T., and Innan, H. (2010). On the estimation of the insertion time of LTR retrotransposable elements. Mole. biol. evol. **27**:896-904.

Kilian, B., Mammen, K., Millet, E., Sharma, R., Graner, A., Salamini, F., Hammer, K., and Ozkan, H. (2011). *Aegilops* L. In Wild crop relatives: genomic and breeding resources, cereals, C. Kole, ed. (Springer: Berlin, Heidelberg), pp. 1-76.

Kimber, F.M. (2001). Wild wheat. An introduction. Special Report 353, College of Agriculture, University of Missouri, Colombia:129-131.

Klindworth, D.L., Niu, Z., Chao, S., Friesen, T.L., Jin, Y., Faris, J.D., Cai, X., and Xu, S.S. (2012). Introgression and characterization of a goatgrass gene for a high level of resistance to Ug99 stem rust in tetraploid wheat. G3 Genes|Genomes|Genetics **2**:665-673.

Knox, A.K., Dhillon, T., Cheng, H., Tondelli, A., Pecchioni, N., and Stockinger, E.J. (2010). CBF gene copy number variation at Frost Resistance-2 is associated with levels of freezing tolerance in temperate-climate cereals. Theor. Appl. Genet. **121**:21-35.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature **533**:420-424.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics (Oxford, England) **35**:4453-4455.

Kroupin, P.Y., Badaeva, E.D., Sokolova, V.M., Chikida, N.N., Belousova, M.K., Surzhikov, S.A., Nikitina, E.A., Kocheshkova, A.A., Ulyanov, D.S., and Ermolaev, A.S. (2022). *Aegilops crassa* Boiss. repeatome characterized using low-coverage NGS as a source of new FISH markers: Application in phylogenetic studies of the *Triticeae*. Front. Plant Sci. **13**:980764.

Kundu, R., Casey, J., and Sung, W.-K. (2019). HyPo: super fast & accurate polisher for long read genome assemblies. bioRxiv:2019.2012.2019.882506.

Langridge, P., and Waugh, R. (2019). Harnessing the potential of germplasm collections. Nat. Genet. **51**:200-201.

Lavington, E., and Kern, A.D. (2017). The effect of common inversion polymorphisms In (2L) t and In (3R) Mo on patterns of transcriptional variation in *Drosophila melanogaster*. G3 **7**:3659-3668.

Leitch, I.J., and Bennett, M.D. (2004). Genome downsizing in polyploid plants. Biol. J. Linn. Soc. **82**:651-663.

Levy, A.A., and Feldman, M. (2022). Evolution and origin of bread wheat. Plant Cell **34**:2549-2567.

Li, H., Rehman, S.u., Song, R., Qiao, L., Hao, X., Zhang, J., Li, K., Hou, L., Hu, W., Wang, L., et al. (2024). Chromosome-scale assembly and annotation of the wild wheat relative *Aegilops comosa*. Sci. Data **11**:1454.

Li, L.F., Zhang, Z.B., Wang, Z.H., Li, N., Sha, Y., Wang, X.F., Ding, N., Li, Y., Zhao, J., Wu, Y., et al. (2022). Genome sequences of five Sitopsis species of *Aegilops* and the origin of polyploid wheat B subgenome. Mol. Plant **15**:488-503.

Li, Y., Xiao, J., Wu, J., Duan, J., Liu, Y., Ye, X., Zhang, X., Guo, X., Gu, Y., and Zhang, L. (2012). A tandem segmental duplication (TSD) in green revolution gene Rht-D1b region underlies plant height variation. New Phytol. **196**:282-291.

Lilienfeld, F., and Kihara, H. (1934). Genomanalyse bei *Triticum* und *Aegilops*. Cytologia **6**:87-122.

Lilienfeld, F.A. (1951). H. Kihara: Genome-analysis in *Triticum* and *Aegilops*. X Concluding review. Cytologia **16**:101-123.

Ling, H.Q., Ma, B., Shi, X., Liu, H., Dong, L., Sun, H., Cao, Y., Gao, Q., Zheng, S., Li, Y., et al. (2018). Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. Nature **557**:424-428.

Lisch, D. (2013). How important are transposons for plant evolution? Nat. Rev. Genet. **14**:49-61.

Liu, H., Wu, S., Li, A., and Ruan, J. (2021). SMARTdenovo: a *de novo* assembler using long noisy reads. GigaByte **2021**:gigabyte15.

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 Res **19**:669-682.

Liu, W., Koo, D.-H., Xia, Q., Li, C., Bai, F., Song, Y., Friebe, B., and Gill, B.S. (2017). Homoeologous recombination-based transfer and molecular cytogenetic mapping of powdery mildew-resistant gene *Pm57* from *Aegilops searsii* into wheat. Theor. Appl. Genet. **130**:841-848.

Liu, Z., Yang, F., Wan, H., Deng, C., Hu, W., Fan, X., Wang, J., Yang, M., Feng, J., Wang, Q., et al. (2025). Genome architecture of the allotetraploid wild grass *Aegilops ventricosa* reveals its evolutionary history and contributions to wheat improvement. Plant Commun. **6**:101131.

Lopes, M.S., El-Basyoni, I., Baenziger, P.S., Singh, S., Royo, C., Ozbek, K., Aktas, H., Ozer, E., Ozdemir, F., Manickavelu, A., et al. (2015). Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. J. Exp. Bot. **66**:3477-3486.

Luo, M.-C., Gu, Y.Q., Puiu, D., Wang, H., Twardziok, S.O., Deal, K.R., Huo, N., Zhu, T., Wang, L., Wang, Y., et al. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. Nature **551**:498-502.

Lye, Z.N., and Purugganan, M.D. (2019). Copy number variation in domestication. Trends Plant Sci. **24**:352-365.

Ma, J., Devos, K.M., and Bennetzen, J.L. (2004). Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res. **14**:860-869.

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., et al. (2019). Durum wheat genome highlights past domestication signatures and future improvement targets. Nat. Genet. **51**:885-895.

Makałowski, W., Gotea, V., Pande, A., and Makałowska, I. (2019). Transposable elements: classification, identification, and their use as a tool for comparative genomics. In Evolutionary genomics: statistical and computational methods, (Springer: pp. 177-207.

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. **38**:4647-4654.

Marais, F., Marais, A., McCallum, B., and Pretorius, Z. (2009). Transfer of leaf rust and stripe rust resistance genes *Lr62* and *Yr42* from *Aegilops neglecta* Req. ex Bertol. to common wheat. Crop Sci. **49**:871-879.

Marais, G.F., McCallum, B., and Marais, A.S. (2008). Wheat leaf rust resistance gene *Lr59* derived from *Aegilops peregrina*. Plant Breed. **127**:340-345.

Marais, G.F., McCallum, B., Snyman, J.E., Pretorius, Z.A., and Marais, A.S. (2005). Leaf rust and stripe rust resistance genes *Lr54* and *Yr37* transferred to wheat from *Aegilops kotschyi*. Plant Breed. **124**:538-541.

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. PLoS Comput. Biol. **14**:e1005944.

Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K.S., Wulff, B.B., Steuernagel, B., Mayer, K.F., and Olsen, O.A. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. Science (New York, N.Y.) **345**:1250092.

Martín-Sánchez, J.A., Gómez-Colmenarejo, M., Del Moral, J., Sin, E., Montes, M.J., González-Belinchón, C., López-Braña, I., and Delibes, A. (2003). A new Hessian fly resistance gene (*H30*) transferred from the wild grass *Aegilops triuncialis* to hexaploid wheat. Theor. Appl. Genet. **106**:1248-1255.

Martin, S.L., Smith, T.W., James, T., Shalabi, F., Kron, P., and Sauder, C.A. (2017). An update to the Canadian range, abundance, and ploidy of Camelina spp. (Brassicaceae) east of the Rocky Mountains. Botany **95**:405-417.

Masoomi-Aladizgeh, F., Aalami, A., Esfahani, M., Aghaei, M.J., and Mozaffari, K. (2015). Identification of *CBF14* and *NAC2* genes in *Aegilops tauschii* associated with resistance to freezing stress. Appl. Biochem. Biotechnol. **176**:1059-1070.

Mayr, E. (1999). Systematics and the origin of species, from the viewpoint of a zoologist (Harvard University Press).

McFadden, E., and Sears, E. (1946). The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J. Hered. **37**:81-107.

Milner, S.G., Jost, M., Taketa, S., Mazón, E.R., Himmelbach, A., Oppermann, M., Weise, S., Knüpffer, H., Basterrechea, M., König, P., et al. (2019). Genebank genomics highlights the diversity of a global barley collection. Nat. Genet. **51**:319-326.

Müller, T., Schierscher-Viret, B., Fossati, D., Brabant, C., Schori, A., Keller, B., and Krattinger, S.G. (2018). Unlocking the diversity of genebanks: whole-genome marker analysis of Swiss bread wheat and spelt. Theor. Appl. Genet. **131**:407-416.

Neumann, P., Novák, P., Hoštáková, N., and Macas, J. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mob. DNA **10**:1.

Nevo, E., and Chen, G. (2010). Drought and salt tolerances in wild relatives for wheat and barley improvement. Plant Cell Environ. **33**:670-685.

Olivera, P.D., Rouse, M.N., and Jin, Y. (2018). Identification of new sources of resistance to wheat stem rust in *Aegilops* spp. in the tertiary genepool of wheat. Front. Plant Sci. **9**:1719.

Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res. **46**:e126-e126.

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., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. **20**:275.

Padmarasu, S., Himmelbach, A., Mascher, M., and Stein, N. (2019). In situ Hi-C for plants: an improved method to detect long-range chromatin interactions. Methods Mol. Biol. **1933**:441-472.

Pan, C., Wu, X., Markel, K., Malzahn, A.A., Kundagrami, N., Sretenovic, S., Zhang, Y., Cheng, Y., Shih, P.M., and Qi, Y. (2021). CRISPR–Act3.0 for highly efficient multiplexed gene activation in plants. Nat. Plants **7**:942-953.

Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., Ainouche, M., Chalhoub, B., and Grandbastien, M.A. (2010). Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol. **186**:37-45.

Pearce, S., Zhu, J., Boldizsár, Á., Vágújfalvi, A., Burke, A., Garland-Campbell, K., Galiba, G., and Dubcovsky, J. (2013). Large deletions in the CBF gene cluster at the Fr-B2 locus are associated with reduced frost tolerance in wheat. Theor. Appl. Genet. **126**:2683-2697.

Peona, V., Weissensteiner, M.H., and Suh, A. (2018). How complete are "complete" genome assemblies?—An avian perspective. Mole. Ecol. Resour. **18**:1188-1195.

Peona, V., Blom, M.P., Xu, L., Burri, R., Sullivan, S., Bunikis, I., Liachko, I., Haryoko, T., Jønsson, K.A., and Zhou, Q. (2021). Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. Mole. Ecol. Resour. **21**:263-286.

Petersen, G., Seberg, O., Yde, M., and Berthelsen, K. (2006). Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). Mol. Phylogenet. Evol. **39**:70-82.

Petersen, S., Lyerly, J.H., Worthington, M.L., Parks, W.R., Cowger, C., Marshall, D.S., Brown-Guedira, G., and Murphy, J.P. (2015). Mapping of powdery mildew resistance gene *Pm53* introgressed from *Aegilops speltoides* into soft red winter wheat. Theor. Appl. Genet. **128**:303-312.

Pingali, P.L. (2012). Green revolution: Impacts, limits, and the path ahead. Proc. Natl. Acad. Sci. U. S. A. **109**:12302-12308.

Pradhan, G.P., Prasad, P.V., Fritz, A.K., Kirkham, M.B., and Gill, B.S. (2012). Response of *Aegilops* species to drought stress during reproductive stages of development. Funct. Plant Biol. **39**:51-59.

Przewieslik-Allen, A.M., Burridge, A.J., Wilkinson, P.A., Winfield, M.O., Shaw, D.S., McAusland, L., King, J., King, I.P., Edwards, K.J., and Barker, G.L. (2019). Developing a high-throughput SNP-based marker system to facilitate the introgression of traits from *Aegilops* species into bread wheat (*Triticum aestivum*). Front. Plant Sci. **9**:1993.

Pucker, B., Irisarri, I., de Vries, J., and Xu, B. (2022). Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. Quant. Plant Biol. **3**:e5.

Pulido, M., and Casacuberta, J.M. (2023). Transposable element evolution in plant genome ecosystems. Curr. Opin. Plant Biol. **75**:102418.

Radke, D.W., and Lee, C. (2015). Adaptive potential of genomic structural variation in human and mammalian evolution. Brief. Funct. Genomics **14**:358-368.

Ragupathy, R., You, F.M., and Cloutier, S. (2013). Arguments for standardizing transposable element annotation in plant genomes. Trends Plant Sci. **18**:367-376.

Rasheed, A., Mujeeb-Kazi, A., Ogbonnaya, F.C., He, Z., and Rajaram, S. (2018). Wheat genetic resources in the post-genomics era: promise and challenges. Ann. Bot. **121**:603-616.

Reif, J.C., Zhang, P., Dreisigacker, S., Warburton, M.L., van Ginkel, M., Hoisington, D., Bohn, M., and Melchinger, A.E. (2005). Wheat genetic diversity trends during domestication and breeding. Theor. Appl. Genet. **110**:859-864.

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., and Martin, W. (2002). Genetics and geography of wild cereal domestication in the near east. Nat. Rev. Genet. **3**:429-441.

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L. (1998). The paleontology of intergene retrotransposons of maize. Nat. genet. **20**:43-45.

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., and Avramova, Z. (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science (New York, N.Y.) **274**:765-768.

Saxena, R.K., Edwards, D., and Varshney, R.K. (2014). Structural variations in plant genomes. Brief. Funct. Genomics **13**:296-307.

Schneider, A., Molnár, I., and Molnár-Láng, M. (2008). Utilisation of *Aegilops* (goatgrass) species to widen the genetic diversity of cultivated wheat. Euphytica **163**:1-19.

Sedlazeck, F.J., Lee, H., Darby, C.A., and Schatz, M.C. (2018). Piercing the dark matter: bioinformatics of long-range sequencing and mapping. Nat. Rev. Genet. **19**:329-346.

Senerchia, N., Wicker, T., Felber, F., and Parisod, C. (2013). Evolutionary dynamics of retrotransposons assessed by high-throughput sequencing in wild relatives of wheat. Genome Biol. Evol. **5**:1010-1020.

Shewry, P.R., and Hey, S.J. (2015). The contribution of wheat to human diet and health. Food Energy Secur. **4**:178-202.

Singh, D.P., Singh, A.K., and Singh, A. (2021). Plant breeding and cultivar development (Academic Press).

Smith, T.W., Kron, P., and Martin, S.L. (2018). flowPloidy: An R package for genome size and ploidy assessment of flow cytometry data. Appl. Plant Sci. **6**:e01164.

Soltis, P.S., and Soltis, D.E. (2009). The role of hybridization in plant speciation. Ann. rev. Plant Biol. **60**:561-588.

Soltis, P.S., Marchant, D.B., Van de Peer, Y., and Soltis, D.E. (2015). Polyploidy and genome evolution in plants. Curr. Opin. Genet. Dev. **35**:119-125.

Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Zuloaga, F.O., Judziewicz, E.J., Filgueiras, T.S., Davis, J.I., and Morrone, O. (2015). A worldwide phylogenetic classification of the Poaceae (Gramineae). J. Syst. Evol. **53**:117-137.

Springer, N.M., Ying, K., Fu, Y., Ji, T., Yeh, C.T., Jia, Y., Wu, W., Richmond, T., Kitzman, J., Rosenbaum, H., et al. (2009). Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet. **5**:e1000734.

Stewart, N.B., and Rogers, R.L. (2019). Chromosomal rearrangements as a source of new gene formation in *Drosophila yakuba*. PLoS Genet. **15**:e1008314.

Stiehler, F., Steinborn, M., Scholz, S., Dey, D., Weber, A.P.M., and Denton, A.K. (2020). Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. Bioinformatics (Oxford, England) **36**:5291-5298.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. **34**:W609-612.

Tadesse, W., Solh, M., Braun, H., Oweis, T., and Baum, M. (2016). Approaches and strategies for sustainable wheat production. Amman, Jordan: International Center for Agricultural Research in the Dry Areas (ICARDA).

Tadesse, W., Sanchez-Garcia, M., Assefa, S.G., Amri, A., Bishaw, Z., Ogbonnaya, F., and Baum, M. (2019). Genetic gains in wheat breeding and its role in feeding the world. Crop Breed. Genet. Genom. **1**:e190005.

Tanksley, S.D., and McCouch, S.R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. Science (New York, N.Y.) **277**:1063-1066.

Van Slageren, M. (1994). Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig (*Poaceae*) (Agricultural University Wageningen).

Venske, E., dos Santos, R.S., Busanello, C., Gustafson, P., and Costa de Oliveira, A. (2019). Bread wheat: a role model for plant domestication and breeding. Hereditas **156**:16.

Vicient, C.M., and Casacuberta, J.M. (2017). Impact of transposable elements on polyploid plant genomes. Ann. Bot. **120**:195-207.

Vihinen, M. (2018). Systematics for types and effects of DNA variations. BMC genomics **19**:974.

Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D., Bevan, M.W., Barry, K., Lucas, S., Harmon-Smith, M., Lail, K., et al. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature **463**:763-768.

Wang, H., Sun, S., Ge, W., Zhao, L., Hou, B., Wang, K., Lyu, Z., Chen, L., Xu, S., Guo, J., et al. (2020). Horizontal gene transfer of *Fhb7* from fungus underlies *Fusarium* head blight resistance in wheat. Science (New York, N.Y.) **368**:eaba5435.

Wang, J.Y., and Doudna, J.A. (2023). CRISPR technology: A decade of genome editing is only the beginning. Science (New York, N.Y.) **379**:eadd8643.

Wang, L., Zhu, T., Rodriguez, J.C., Deal, K.R., Dubcovsky, J., McGuire, P.E., Lux, T., Spannagl, M., Mayer, K.F.X., Baldrich, P., et al. (2021). *Aegilops tauschii* genome assembly Aet v5.0 features greater sequence contiguity and improved annotation. G3 (Bethesda) **11**:jkab325.

Wang, Y., Abrouk, M., Gourdoupis, S., Koo, D.-H., Karafiátová, M., Molnár, I., Holušová, K., Doležel, J., Athiyannan, N., Cavalet-Giorsa, E., et al. (2023). An unusual tandem kinase fusion protein confers leaf rust resistance in wheat. Nat. genet. **55**:914-920.

Weckselblatt, B., and Rudd, M.K. (2015). Human structural variation: mechanisms of chromosome rearrangements. Trends Genet. **31**:587-599.

Weidner, A., Röder, M.S., and Börner, A. (2012). Mapping wheat powdery mildew resistance derived from *Aegilops markgrafii*. Plant Genet. Resour. **10**:137-140.

Wellenreuther, M., Mérot, C., Berdan, E., and Bernatchez, L. (2019). Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. Mol. Ecol. **28**

Wicker, T., Matthews, D.E., and Keller, B. (2002). TREP: a database for Triticeae repetitive elements. Trends Plant Sci. **7**:561-562.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., and Panaud, O. (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. **8**:973-982.

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., et al. (2018). Impact of transposable elements on genome structure and evolution in bread wheat. Genome Biol. **19**:103.

Wolf, J.B., and Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. Nat. Rev. Genet. **18**:87-100.

Wu, T.D., and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics (Oxford, England) **21**:1859-1875.

Würschum, T., Boeven, P.H., Langer, S.M., Longin, C.F.H., and Leiser, W.L. (2015). Multiply to conquer: copy number variations at *Ppd-B1* and *Vrn-A1* facilitate global adaptation in wheat. BMC Genet. **16**:96.

Yamane, K., and Kawahara, T. (2005). Intra- and interspecific phylogenetic relationships among diploid *Triticum-Aegilops* species (Poaceae) based on base-pair substitutions, indels, and microsatellites in chloroplast noncoding sequences. Am. J. Bot. **92**:1887-1898.

Yang, Y., Cui, L., Lu, Z., Li, G., Yang, Z., Zhao, G., Kong, C., Li, D., Chen, Y., Xie, Z., et al. (2023). Genome sequencing of *Sitopsis* species provides insights into their contribution to the B subgenome of bread wheat. Plant Commun. **4**:100567.

You, F.M., Cloutier, S., Shan, Y., and Ragupathy, R. (2015). LTR annotator: automated identification and annotation of LTR retrotransposons in plant genomes. Int. J. Biosci. Biochem. Bioinforma. **5**:165.

Yu, G., Matny, O., Champouret, N., Steuernagel, B., Moscou, M.J., Hernández-Pinzón, I., Green, P., Hayta, S., Smedley, M., Harwood, W., et al. (2022). *Aegilops sharonensis* genome-assisted identification of stem rust resistance gene *Sr62*. Nat. Commun. **13**:1607.

Zaharieva, M., Gaulin, E., Havaux, M., Acevedo, E., and Monneveux, P. (2001). Drought and heat responses in the wild wheat relative *Aegilops geniculata* Roth: potential interest for wheat improvement. Crop Sci. **41**:1321-1329.

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.

Zeng, X., Yi, Z., Zhang, X., Du, Y., Li, Y., Zhou, Z., Chen, S., Zhao, H., Yang, S., Wang, Y., et al. (2024). Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. Nat. Plants **10**:1184-1200.

Zhang, H.-B., Dvořák, J., and Waines, J.G. (1992). Diploid ancestry and evolution of *Triticum kotschyi* and *T. peregrinum* examined using variation in repeated nucleotide sequences. Genome **35**:182-191.

Zhang, M., Zhang, Y., Scheuring, C.F., Wu, C.-C., Dong, J.J., and Zhang, H.-B. (2012). Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. Nat. Protoc. **7**:467-478.

Zhang, R.-G., Li, G.-Y., Wang, X.-L., Dainat, J., Wang, Z.-X., Ou, S., and Ma, Y. (2022). TEsorter: An accurate and fast method to classify LTR-retrotransposons in plant genomes. Hort. Res. **9**:uhac017.

Zhang, Z., Lv, R., Wang, B., Xun, H., Liu, B., and Xu, C. (2023). Effects of allopolyploidization and homoeologous chromosomal segment exchange on homoeolog expression in a synthetic allotetraploid wheat under variable environmental conditions. Plants **12**:3111.

Zhao, L., Ning, S., Yu, J., Hao, M., Zhang, L., Yuan, Z., Zheng, Y., and Liu, D. (2016). Cytological identification of an *Aegilops variabilis* chromosome carrying stripe rust resistance in wheat. Breed. Sci. **66**:522-529.

Zhou, C., McCarthy, S.A., and Durbin, R. (2022). YaHS: yet another Hi-C scaffolding tool. Bioinformatics (Oxford, England) **39**:btac808.

Zhu, T., Wang, L., Rodriguez, J.C., Deal, K.R., Avni, R., Distelfeld, A., McGuire, P.E., Dvorak, J., and Luo, M.C. (2019). Improved genome sequence of wild emmer wheat Zavitan with the aid of optical maps. G3 (Bethesda) **9**:619-624.

Appendices

Supplementary materials for Chapter 2

Appendix 1 Description of the 18 *Aegilops* accessions from which genome assemblies were obtained, including the number of seedlings used to create the DNA pools.

Species	Genome	Accession	Source[1]/Alias	PGRC No.[5]	NCVP DAO No.[6]	Entry[7]	No. seedlings
Ae. sharonensis	S^{sh}	AE133	IPK	TMP 29646	01-01669407	1043-1-1-4	35
Ae. longissima	S^{l}	AE334	IPK	TMP 29645	01-01669406	1042-1-1-3	50
Ae. bicornis	S^{b}	AE105	IPK	TMP 29634	01-01669405	1044-1-1-4	53
Ae. searsii	S^{s}	AE1071	IPK	TMP 29652	01-01669408	1049-1-1-1	39
Ae. neglecta subsp. *neglecta*	UM	AE729	IPK	TMP 29649	01-01669397	977-1-1-3	10
Ae. cylindrica	DC	RL5716	EKC135/PGR6493;CN34409[2]	TMP 29641	01-01669404	801-1-1-1-1	48
Ae. ventricosa	DN	AE1568	IPK	TMP 29657	01-01669412	944-1-1-2	11
Ae. triuncialis	UC	RL5011	EKC165[3]	TMP 29654	01-01669399	812-1-1-1-4	37
Ae. peregrina	SU	RL5017	EKC163[3]	TMP 29651	01-01669398	810-1-1	57
Ae. kotschyi	SU	AE737	IPK	TMP 29644	01-01669396	1032-1-1-3	25
Ae. geniculata	MU	RL5859	EKC160[4]	TMP 29642	01-01669395	808-1-3-5	50
Ae. crassa subsp. *crassa* (4x)	DM	AE290	IPK	TMP 29638	01-01669411	829-1-4	23
Ae. columnaris	UM	AE113	IPK	TMP 29636	01-01669394	823-1-1	31
Ae. biuncialis	UM	RL5847	EKC186[4]	TMP 29635	01-01669393	815-1-1-33	45
Ae. vavilovii	DMS	AE299	IPK	TMP 29640	01-01669414	962-1-2	58
Ae. crassa subsp. *crassa* (6x)	DDM	AE8	IPK	TMP 29639	01-01669413	819-1-3-1	20
Ae. juvenalis	DMU	AE182	IPK	TMP 29643	01-01669415	828-1-1	51
Ae. neglecta subsp. *recta*	UMN	AE129	IPK	TMP 29650	01-01669400	982-1-1-2	25

[1] IPK = Leibniz Institute of Plant Genetics and Crop Plant Research, and EKC = Dr. Eric Kerber Collection (Agriculture and Agri-Food Canada) [2] CN = accession number from Plant Gene Resource Canada (PGRC); [3] originally obtained in 1983 from Dr. Ernie Sears; [4] unknown original source; [5] Plant Gene Resource Canada accession numbers for seed request; [6] National Collection of Vascular Plants-Department of Agriculture, Ottawa (Canada) specimen numbers; [7] Dashes (-) indicate the number of generations of selfing prior to pooling. For example, 1043-1-1-4 means that entry 1043 was selfed three times and the pool of 35 plants corresponded to the fourth generation of selfing.

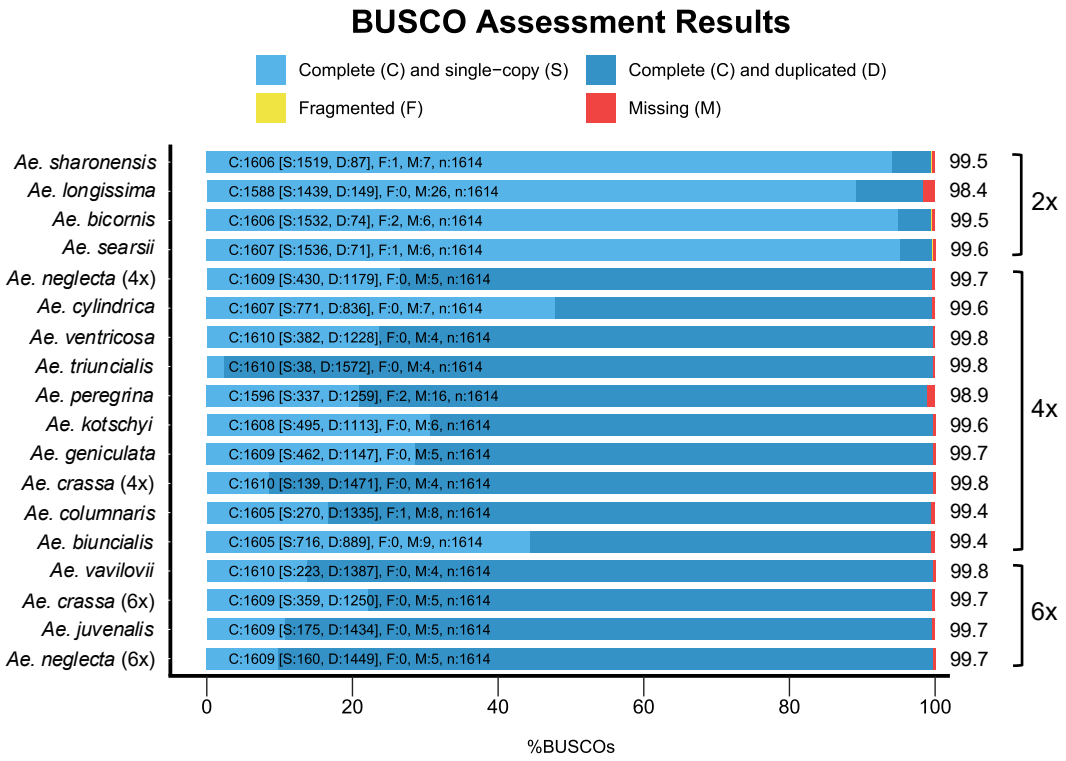
Appendix 2 Sequence coverage (X) across technologies for 18 *Aegilops* species.

Dashes (-) indicate species for which no Omni-C data were generated.

Species	Genome	ONT Coverage (X)	PacBio Coverage (X)	Illumina Coverage (X)	Hi-C Coverage (X)	Omni-C Coverage (X)
Ae. sharonensis	S^{sh}	23.9	3.7	11.9	32.1	-
Ae. longissima	S^{l}	37.1	3.1	10.7	40.7	-
Ae. bicornis	S^{b}	25.4	4.9	12.5	38.0	-
Ae. searsii	S^{s}	29.1	4.2	14.2	32.3	-
Ae. neglecta subsp. *neglecta*	UM	63.8	5.5	13.2	21.8	-
Ae. cylindrica	DC	38.5	3.9	18.8	27.5	-
Ae. ventricosa	DN	38.2	6.6	16.3	18.4	-
Ae. triuncialis	UC	36.8	4.6	23.5	23.9	-
Ae. peregrina	SU	31.9	4.4	20.1	18.3	-
Ae. kotschyi	SU	34.9	4.9	13.7	16.8	-
Ae. geniculata	MU	61.0	5.8	21.4	25.6	14.2
Ae. crassa subsp. *crassa* (4x)	DM	37.9	4.5	20.3	23.9	-
Ae. columnaris	UM	43.3	4.9	19.8	24.8	-
Ae. biuncialis	UM	34.3	3.8	23.1	36.5	-
Ae. vavilovii	DMS	29.9	6.1	19.4	13.1	23.4
Ae. crassa subsp. *crassa* (6x)	DDM	30.9	4.7	16.0	13.1	13.0
Ae. juvenalis	DMU	35.4	5.6	16.1	15.1	9.4
Ae. neglecta subsp. *recta*	UMN	37.4	5.5	15.7	10.3	9.9

Appendix 3 Statistics of the 18 *Aegilops* genome assemblies.

Species	Genome	Assembly size (Gbp)	No. of scaffolds	Largest scaffold (Mbp)	No. of scaffolds > 100 Mbp	N50 (Mbp)	N90 (Mbp)	BUSCO%	LTR Assembly Index (LAI)
Ae. sharonensis	S^{sh}	5.91	1373	820.4	10	726.6	364.8	99.5	17.9
Ae. longissima	S^{l}	6.53	1720	810.2	8	680.6	41.0	98.4	16.2
Ae. bicornis	S^{b}	5.40	292	851.2	9	560.0	23.2	99.5	20.2
Ae. searsii	S^{s}	5.24	226	835.1	8	749.2	397.7	99.6	20.2
Ae. neglecta subsp. *neglecta*	UM	8.49	449	859.2	16	644.9	272.1	99.7	14.1
Ae. cylindrica	DC	7.40	509	788.9	14	573.3	339.7	99.6	11.9
Ae. ventricosa	DN	8.19	642	842.6	17	640.8	193.5	99.8	14.9
Ae. triuncialis	UC	7.84	585	892.6	15	563.9	210.2	99.8	24.5
Ae. peregrina	SU	9.32	459	875.5	19	508.7	111.7	98.9	19.6
Ae. kotschyi	SU	9.79	1689	836.8	24	367.2	62.5	99.6	16.4
Ae. geniculata	MU	7.92	349	840.9	15	616.1	306.2	99.7	15.2
Ae. crassa subsp. *crassa* (4x)	DM	8.63	2463	846.9	20	391.8	85.3	99.8	17.5
Ae. columnaris	UM	8.00	1150	842.0	17	513.3	132.8	99.4	20.2
Ae. biuncialis	UM	7.89	975	829.1	19	588.8	125.8	99.4	11.5
Ae. vavilovii	DMS	11.57	726	859.6	21	633.9	344.3	99.8	12.1
Ae. crassa subsp. *crassa* (6x)	DDM	11.74	638	912.8	21	633.9	332.5	99.7	12.6
Ae. juvenalis	DMU	12.62	1070	888.0	22	623.8	390.8	99.7	13.1
Ae. neglecta subsp. *recta*	UMN	12.65	266	840.9	23	705.7	316.1	99.7	11.8



Appendix 4 BUSCO assessment of genome assembly completeness for the analyzed *Aegilops* genomes using the *embryophyta_odb10* database, which contains 1,614 core plant genes (n).

Each bar represents the proportion of BUSCOs classified as complete (C)[single-copy (S) or duplicated (D)], fragmented (F), or missing (M). The numbers inside each segment of the bars indicate the count of BUSCO genes in each category, while the numbers on the right of each bar show the corresponding BUSCO completeness score as a percentage. Genome ploidy, indicated on the right as 2x, 4x, and 6x, corresponds to diploid, tetraploid, and hexaploid assemblies, respectively. The following species names were shortened: *Ae. neglecta* subsp. *neglecta* is shown as *Ae. neglecta* (4x); *Ae. crassa* subsp. *crassa* (4x) as *Ae. crassa* (4x); *Ae. crassa* subsp. *crassa* (6x) as *Ae. crassa* (6x); and *Ae. neglecta* subsp. *recta* as *Ae. neglecta* (6x).

Appendix 5 Gene annotation summary for 18 *Aegilops* genome assemblies.

Species	Genome	Predicted protein-coding genes	High confidence genes	Low confidence genes	BUSCO (%)
Ae. sharonensis	S^{sh}	53,035	24,588	28,447	91.8
Ae. longissima	S^{l}	56,711	21,865	34,846	84.5
Ae. bicornis	S^{b}	58,339	27,291	31,048	95.1
Ae. searsii	S^{s}	59,817	27,783	32,034	95.0
Ae. neglecta subsp. *neglecta*	UM	100,157	39,848	60,309	90.2
Ae. cylindrica	DC	88,422	33,590	54,832	85.7
Ae. ventricosa	DN	101,073	40,473	60,600	92.5
Ae. triuncialis	UC	109,856	60,490	49,366	98.9
Ae. peregrina	SU	96,891	50,857	46,034	97.8
Ae. kotschyi	SU	98,796	36,127	62,669	90.6
Ae. geniculata	MU	96,459	40,797	55,662	92.6
Ae. crassa subsp. *crassa* (4x)	DM	149,108	59,100	90,008	98.4
Ae. columnaris	UM	107,794	54,974	52,820	97.8
Ae. biuncialis	UM	88,423	34,670	53,753	86.4
Ae. vavilovii	DMS	150,024	50,761	99,263	92.3
Ae. crassa subsp. *crassa* (6x)	DDM	143,429	44,149	99,280	89.4
Ae. juvenalis	DMU	156,779	53,871	102,908	90.7
Ae. neglecta subsp. *recta*	UMN	146,978	55,945	91,033	92.6

Appendix 6 Summary of core, dispensable, and species-specific families and genes in all 25 *Aegilops* species.

Species	Core orthogroup	Core gene	Dispensable orthogroup	Dispensable gene	Species-specific orthogroup	Species-specific gene
Ae. markgrafii	15,809	19,400	17,006	24,913	125	318
Ae. speltoides	15,809	18,587	22,250	26,416	78	191
Ae. umbellulata	15,809	19,762	17,686	26,669	108	294
Ae. uniasristata	15,809	19,858	17,790	27,438	112	259
Ae. comosa	15,809	20,051	17,643	28,162	110	282
Ae. tauschii	15,809	20,936	18,506	39,713	361	2,283
Ae. mutica	15,809	23,816	20,998	47,865	548	2,031
Ae. sharonensis	15,809	21,143	18,409	30,366	16	36
Ae. longissima	15,809	23,242	18,178	30,587	43	91
Ae. bicornis	15,809	20,825	18,105	35,107	20	47
Ae. searsii	15,809	20,769	18,223	36,050	36	77
Ae. neglecta subsp. *neglecta*	15,809	39,178	25,561	55,123	39	81
Ae. cylindrica	15,809	37,110	24,068	44,535	30	78
Ae. ventricosa	15,809	38,767	24,608	54,862	65	144
Ae. triuncialis	15,809	40,084	24,069	64,367	103	321
Ae. peregrina	15,809	36,581	24,028	56,297	49	110
Ae. kotschyi	15,809	39,285	25,618	53,610	42	87
Ae. geniculata	15,809	38,889	24,899	51,883	38	83
Ae. crassa subsp. *crassa* (4x)	15,809	40,055	28,210	95,073	133	338
Ae. columnaris	15,809	38,841	24,530	63,912	79	206
Ae. biuncialis	15,809	41,334	28,405	63,161	181	412
Ae. vavilovii	15,809	51,418	31,588	84,406	158	322
Ae. crassa subsp. *crassa* (6x)	15,809	50,196	30,948	77,689	192	394
Ae. juvenalis	15,809	55,060	32,803	85,495	160	344
Ae. neglecta subsp. *recta*	15,809	55,519	29,722	77,060	151	317

Appendix 7 Summary of synteny and structural variant (SV) counts identified in pairwise genome comparisons between polyploid *Aegilops* subgenomes and their respective diploid progenitors.

Species-progenitor comparison			Synteny		Inversions		Translocations		Duplications		Insertions		Deletion	
Polyploid species	Diploid species	Genome	Count	Genome span %	Count	Genome span %	Count	Genome span %	Count	Genome span %	Count	Genome span %	Count	Genome span %
Ae. neglecta subsp. *neglecta*	*Ae. umbellulata*	U	5,194	26.8	134	8.7	2,546	3.2	12,343	0.5	2,330	0.2	2,442	0.1
Ae. neglecta subsp. *neglecta*	*Ae. comosa*	M	290	2.3	50	21.8	531	0.1	3,742	0.0	28	0.1	19	0.0
Ae. cylindrica	*Ae. tauschii*	D	636	23.1	75	16.7	3,906	3.0	1,137	0.2	675	0.0	837	0.0
Ae. cylindrica	*Ae. markgrafii*	C	1,232	16.1	59	26.3	362	10.9	5,639	2.1	1,000	0.1	1,582	0.1
Ae. ventricosa	*Ae. tauschii*	D	1,981	44.0	144	7.8	4,202	3.1	3,117	0.5	1,230	0.1	1,143	0.1
Ae. ventricosa	*Ae. uniaristata*	N	3,262	31.0	175	17.6	1,927	0.7	5,841	0.2	952	0.0	1,095	0.1
Ae. triuncialis	*Ae. umbellulata*	U	2,723	24.2	96	7.2	3,540	7.4	4,286	0.7	1,035	0.1	1,462	0.1
Ae. triuncialis	*Ae. markgrafii*	C	6,555	28.0	102	13.7	4,484	7.7	10,902	1.4	2,117	0.3	2,344	0.2
Ae. peregrina	*Ae. sharonensis*	S	7,404	36.2	143	6.9	4,329	9.7	15,432	1.7	4,436	0.3	4,723	0.4
Ae. peregrina	*Ae. longissima*	S	6,762	38.6	165	8.2	6,005	13.2	13,860	3.8	4,470	0.3	4,824	0.4
Ae. peregrina	*Ae. umbellulata*	U	17,038	41.4	260	15.7	5,920	6.9	36,853	2.2	5,273	0.6	6,188	0.5
Ae. kotschyi	*Ae. sharonensis*	S^{sh}	1,121	22.6	118	13.5	4,232	1.9	3,129	0.3	776	0.1	881	0.0
Ae. kotschyi	*Ae. longissima*	S^{l}	1,028	20.8	121	13.2	4,805	2.3	3,486	0.4	712	0.0	826	0.0
Ae. kotschyi	*Ae. umbellulata*	U	6,389	38.5	171	8.7	3,015	6.3	15,041	4.5	2,717	0.2	3,267	0.2
Ae. geniculata	*Ae. comosa*	M	6,487	21.7	177	29.8	4,688	2.4	24,612	1.4	1,242	0.1	1,489	0.1
Ae. geniculata	*Ae. umbellulata*	U	3,626	30.6	148	2.6	3,594	3.3	8,324	0.1	1,615	0.1	1,866	0.1
Ae. crassa subsp. *crassa* (4x)	*Ae. tauschii*	D	1,859	5.8	95	19.9	4,362	0.8	7,904	0.4	312	0.0	338	0.0
Ae. crassa subsp. *crassa* (4x)	*Ae. comosa*	M	276	3.7	63	6.3	662	0.1	3,185	0.1	17	0.0	28	0.0
Ae. columnaris	*Ae. umbellulata*	U	1,473	31.1	121	10.2	2,494	1.7	1,864	0.1	487	0.0	869	0.1
Ae. columnaris	*Ae. comosa*	M	567	0.8	123	30.9	910	0.2	9,735	0.3	34	0.0	37	0.0
Ae. biuncialis	*Ae. umbellulata*	U	1,061	20.3	105	20.4	2,944	2.3	3,445	0.3	672	0.0	806	0.0
Ae. biuncialis	*Ae. comosa*	M	1,093	18.7	107	17.7	2,735	1.4	4,623	0.3	462	0.3	628	0.0
Ae. vavilovii	*Ae. tauschii*	D	2,386	15.3	226	18.5	3,101	2.0	5,150	1.0	655	0.0	845	0.0
Ae. vavilovii	*Ae. comosa*	M	170	2.6	54	21.5	292	0.1	2,057	0.0	4	0.0	4	0.0
Ae. vavilovii	*Ae. speltoides*	S	249	3.2	55	8.1	621	0.1	740	0.0	5	0.0	9	0.0

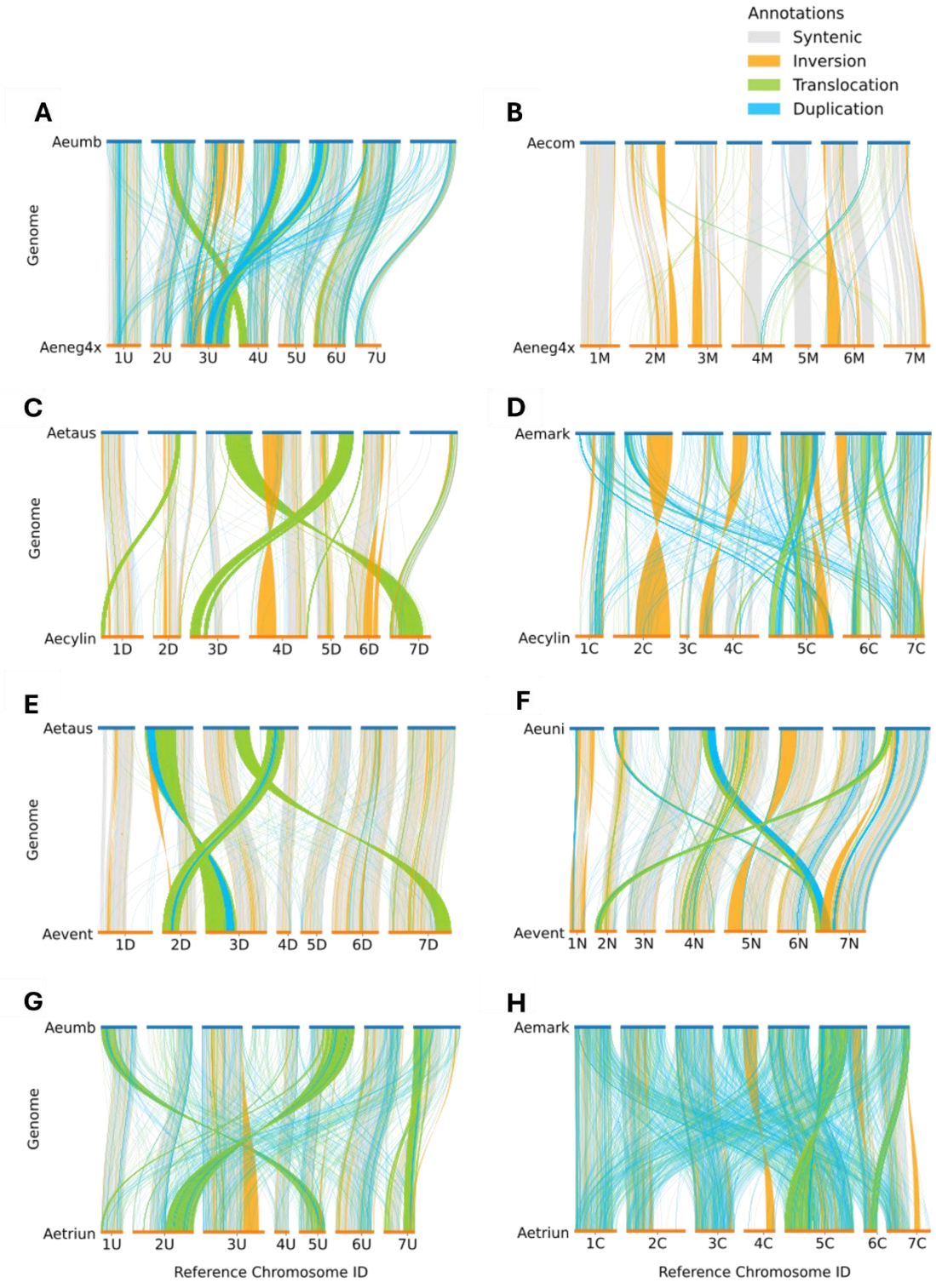
Species-progenitor comparison			Synteny		Inversions		Translocations		Duplications		Insertions		Deletion	
Polyploid species	**Diploid species**	**Genome**	**Count**	**Genome span %**	**Count**	**Genome span %**	**Count**	**Genome span %**	**Count**	**Genome span %**	**Count**	**Genome span %**	**Count**	**Genome span %**
Ae. crassa subsp. *crassa* (6x)	*Ae. tauschii*	D	1,780	12.0	181	12.3	3,566	2.0	4,079	0.4	460	0.0	601	0.0
Ae. crassa subsp. *crassa* (6x)	*Ae. tauschii*	D	1,786	10.5	185	17.1	2,389	1.0	3,470	0.3	340	0.0	427	0.0
Ae. crassa subsp. *crassa* (6x)	*Ae. comosa*	M	110	3.6	42	5.4	343	0.1	1,357	0.0	5	0.0	4	0.0
Ae. juvenalis	*Ae. tauschii*	D	702	20.8	101	7.4	2,206	0.5	933	0.0	260	0.0	258	0.0
Ae. juvenalis	*Ae. comosa*	M	117	3.8	27	5.9	276	0.0	1,210	0.0	2	0.0	4	0.0
Ae. juvenalis	*Ae. umbellulata*	U	6,992	30.9	163	22.9	3,519	8.0	15,452	8.2	3,272	0.2	4,113	0.2
Ae. neglecta subsp. *recta*	*Ae. umbellulata*	U	2,245	17.0	104	18.5	1,694	3.1	12,332	0.5	1,358	0.1	1,401	0.1
Ae. neglecta subsp. *recta*	*Ae. comosa*	M	306	2.4	49	21.3	812	0.3	6,453	0.1	55	0.0	91	0.0
Ae. neglecta subsp. *recta*	*Ae. uniaristata*	N	5,225	28.7	127	3.2	1,640	8.7	14,157	0.5	3,385	0.2	4,288	0.1

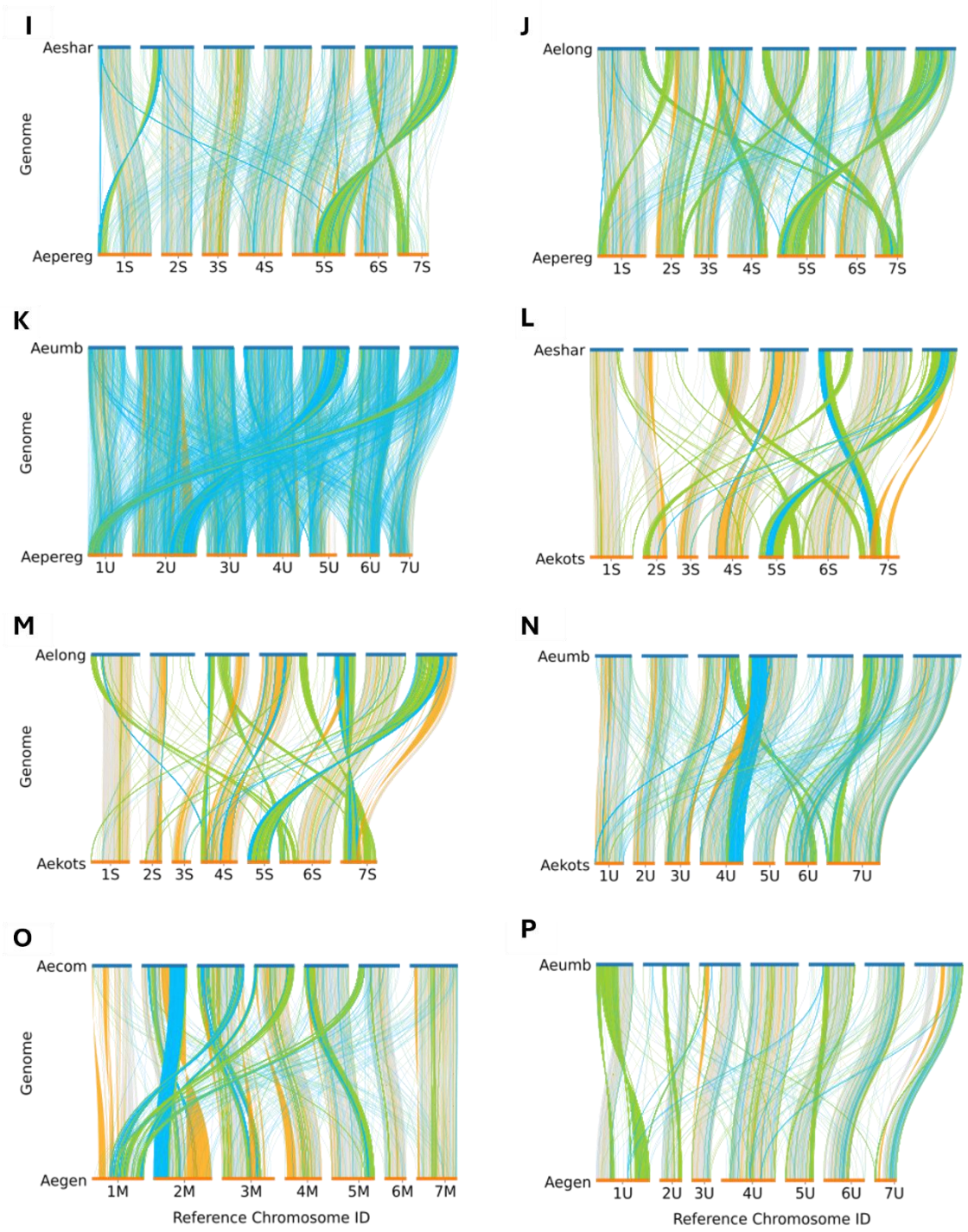
Appendix 8 Size range of each SV type across subgenome-progenitor comparisons.

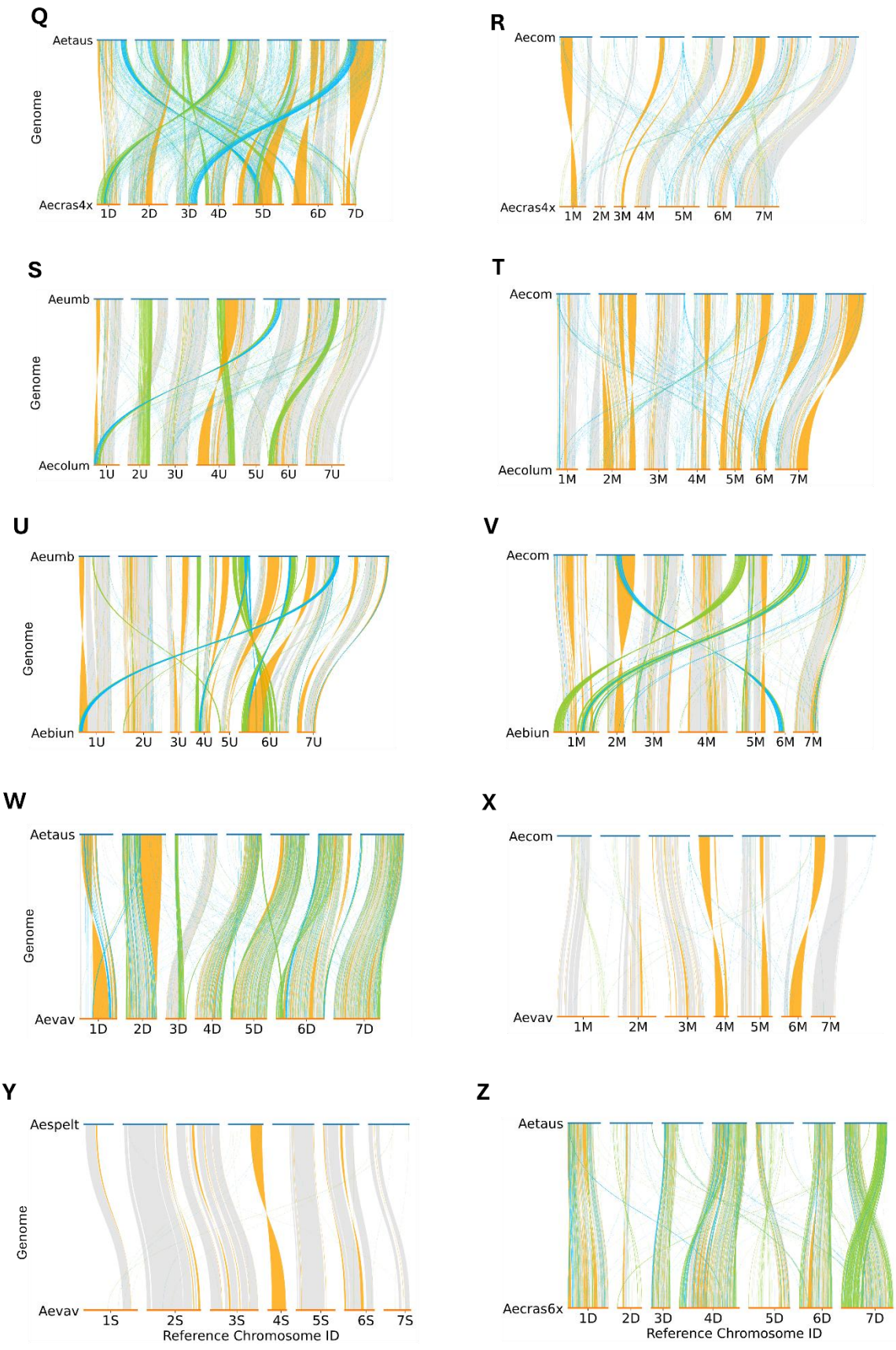
The table reports the maximum and minimum SV lengths (bp) observed.

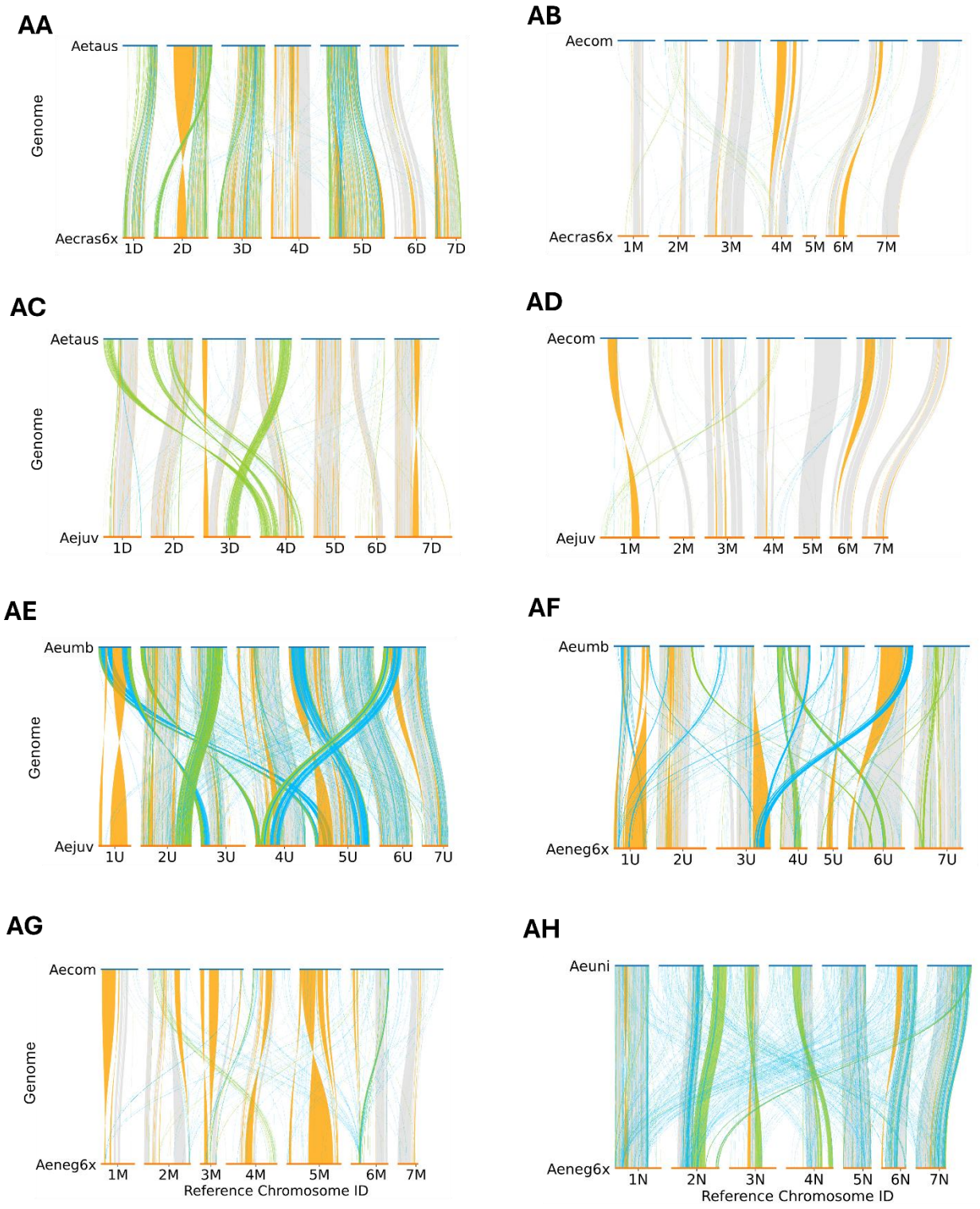
Species-progenitor comparison			Inversions		Translocations		Duplications		Insertions		Deletions	
Polyploid species	**Diploid species**	**Genome**	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Ae. neglecta subsp. *neglecta*	*Ae. umbellulata*	U	540	141,302,232	511	31,696,704	506	100,048,978	51	550,373	51	87,706
Ae. neglecta subsp. *neglecta*	*Ae. comosa*	M	653	131,902,769	682	166,265	500	133,979	54	3,888,173	51	13,526
Ae. cylindrica	*Ae. tauschii*	D	799	274,920,732	551	4,163,652	501	1,374,090	51	82,779	51	34,477
Ae. cylindrica	*Ae. markgrafii*	C	786	309,625,271	760	67,556,364	552	13,932,098	51	87,420	51	340,852
Ae. ventricosa	*Ae. tauschii*	D	1,121	107,789,063	533	682,009	513	521,504	51	141,539	51	36,475
Ae. ventricosa	*Ae. uniaristata*	N	778	243,147,329	621	1,366,416	501	765,854	51	31,283	51	48,716
Ae. triuncialis	*Ae. umbellulata*	U	1,364	20,039,643	1,016	2,420,574	1,044	1,116,248	51	69,676	51	209,933
Ae. triuncialis	*Ae. markgrafii*	C	2,423	177,974,354	1,011	20,552,934	1,009	2,089,853	51	194,567	51	67,028
Ae. peregrina	*Ae. sharonensis*	S^{sh}	917	44,086,214	510	19,449,108	507	11,264,579	51	167,634	51	384,593
Ae. peregrina	*Ae. longissima*	S^{l}	989	78,546,605	511	48,764,705	503	37,805,997	51	363,903	51	640,993
Ae. peregrina	*Ae. umbellulata*	U	630	38,189,621	501	10,682,256	500	4,569,345	51	9,805,656	51	1,365,946
Ae. kotschyi	*Ae. sharonensis*	S^{sh}	1,070	155,596,375	573	1,272,612	501	1,106,516	51	376,606	51	33,338
Ae. kotschyi	*Ae. longissima*	S^{l}	632	209,602,325	516	1,190,102	500	571,366	51	58,932	51	124,207
Ae. kotschyi	*Ae. umbellulata*	U	654	267,174,928	539	24,737,175	500	19,573,756	51	302,323	51	92,278
Ae. geniculata	*Ae. comosa*	M	571	351,610,978	520	1,438,561	500	975,890	51	188,001	51	362,294
Ae. geniculata	*Ae. umbellulata*	U	559	55,123,100	532	2,848,894	500	1,819,588	51	116,231	51	45,934
Ae. crassa subsp. *crassa* (4x)	*Ae. tauschii*	D	1,208	278,646,371	528	110,904	501	216,042	51	246,736	51	64,919
Ae. crassa subsp. *crassa* (4x)	*Ae. comosa*	M	1,005	210,453,381	561	55,256	502	661,271	51	5,179	55	259,244
Ae. columnaris	*Ae. umbellulata*	U	2,442	228,816,081	1,224	4,081,370	1,003	3,705,796	51	61,814	51	2,066,067
Ae. columnaris	*Ae. comosa*	M	697	263,902,654	567	31,392	501	401,437	53	8,639	51	22,478
Ae. biuncialis	*Ae. umbellulata*	U	507	190,283,156	546	416,301	501	313,005	51	31,967	51	45,934
Ae. biuncialis	*Ae. comosa*	M	551	308,666,809	568	239,246	502	309,017	51	11,962,185	51	70,155
Ae. vavilovii	*Ae. tauschii*	D	908	342,978,826	532	3,000,831	517	2,192,992	51	139,051	51	32,524
Ae. vavilovii	*Ae. comosa*	M	539	174,104,239	529	53,941	500	17,554	124	1,915	65	8,868
Ae. vavilovii	*Ae. speltoides*	S	509	171,793,471	439	18,168	397	9,091	62	1,028	76	4,190
Ae. crassa subsp. *crassa* (6x)	*Ae. tauschii*	D	2,118	37,805,518	588	1,883,841	505	530,223	51	283,367	51	200,461
Ae. crassa subsp. *crassa* (6x)	*Ae. tauschii*	D	1,478	331,686,638	554	674,871	508	710,103	51	216,340	51	34,477
Ae. crassa subsp. *crassa* (6x)	*Ae. comosa*	M	573	145,358,307	545	24,973	501	15,952	56	455	63	23,200
Ae. juvenalis	*Ae. tauschii*	D	1,080	80,780,766	542	115,067	516	95,785	51	26,203	52	34,583
Ae. juvenalis	*Ae. comosa*	M	586	153,269,080	516	14,457	500	27,792	353	4,844	63	851
Ae. juvenalis	*Ae. umbellulata*	U	844	246,778,567	505	20,867,252	500	39,127,280	51	255,164	51	120,485

Species-progenitor comparison			Inversions		Translocations		Duplications		Insertions		Deletions	
Polyploid species	**Diploid species**	**Genome**	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Ae. neglecta subsp. *recta*	*Ae. umbellulata*	U	938	314,183,370	588	15,344,076	502	17,796,249	51	220,458	51	75,930
Ae. neglecta subsp. *recta*	*Ae. comosa*	M	560	212,724,891	646	292,326	500	62,433	51	23,047	54	45,683
Ae. neglecta subsp. *recta*	*Ae. uniaristata*	N	856	89,184,307	542	65,365,159	508	6,916,288	51	401,738	51	259,901


Annotations
Syntenic
Inversion
Translocation
Duplication
A
Aeumb
Genome
Aeneg4x
1U 2U 3U 4U 5U 6U 7U
B
Aecom
Aeneg4x
1M 2M 3M 4M 5M 6M 7M
C
Aetaus
Genome
Aecylin
1D 2D 3D 4D 5D 6D 7D
D
Aemark
Aecylin
1C 2C 3C 4C 5C 6C 7C
E
Aetaus
Genome
Aevent
1D 2D 3D 4D 5D 6D 7D
F
Aeuni
Aevent
1N 2N 3N 4N 5N 6N 7N
G
Aeumb
Genome
Aetriun
1U 2U 3U 4U 5U 6U 7U
Reference Chromosome ID
H
Aemark
Aetriun
1C 2C 3C 4C 5C 6C 7C
Reference Chromosome ID

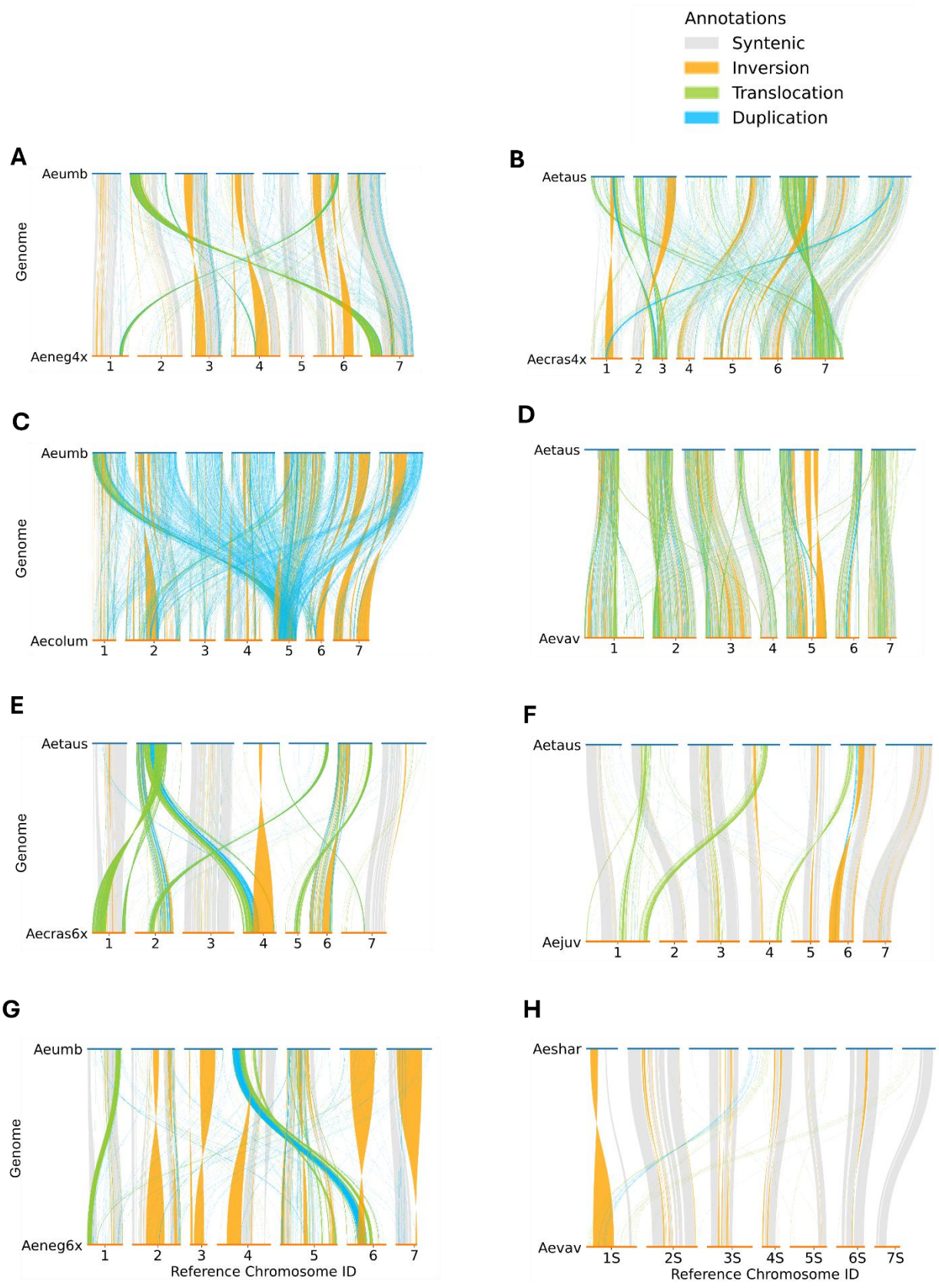

I
Aeshar
Genome
Aepereg
1S 2S 3S 4S 5S 6S 7S
J
Aelong
Aepereg
1S 2S 3S 4S 5S 6S 7S
K
Aeumb
Genome
Aepereg
1U 2U 3U 4U 5U 6U 7U
L
Aeshar
Aekots
1S 2S 3S 4S 5S 6S 7S
M
Aelong
Genome
Aekots
1S 2S 3S 4S 5S 6S 7S
N
Aeumb
Aekots
1U 2U 3U 4U 5U 6U 7U
O
Aecom
Genome
Aegen
1M 2M 3M 4M 5M 6M 7M
Reference Chromosome ID
P
Aeumb
Aegen
1U 2U 3U 4U 5U 6U 7U
Reference Chromosome ID

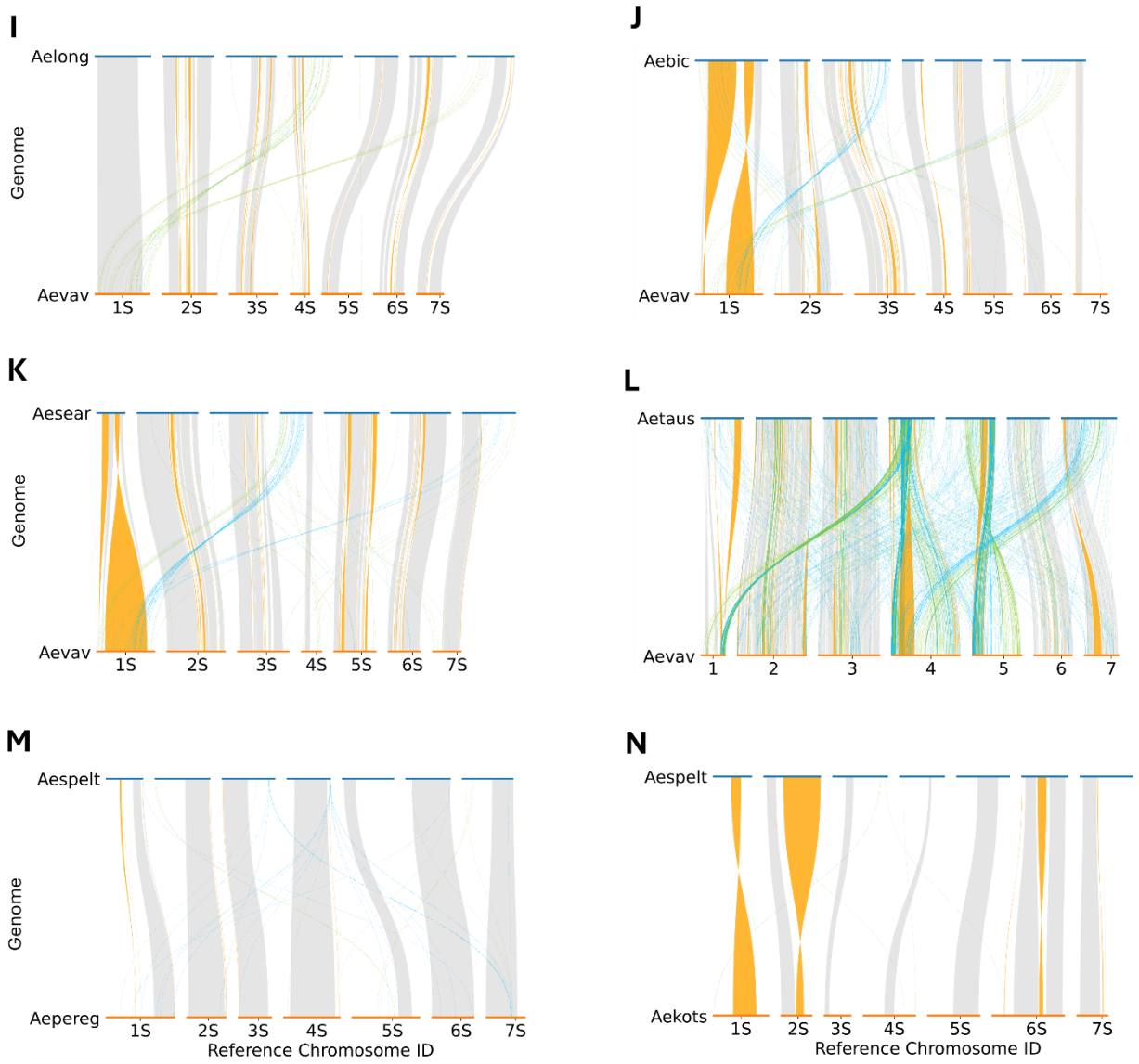

Q
Aetaus
Genome
Aecras4x
1D 2D 3D 4D 5D 6D 7D
R
Aecom
Aecras4x
1M 2M 3M 4M 5M 6M 7M
S
Aeumb
Genome
Aecolum
1U 2U 3U 4U 5U 6U 7U
T
Aecom
Aecolum
1M 2M 3M 4M 5M 6M 7M
U
Aeumb
Genome
Aebiun
1U 2U 3U 4U 5U 6U 7U
V
Aecom
Aebiun
1M 2M 3M 4M 5M 6M 7M
W
Aetaus
Genome
Aevav
1D 2D 3D 4D 5D 6D 7D
X
Aecom
Aevav
1M 2M 3M 4M 5M 6M 7M
Y
Aespelt
Genome
Aevav
1S 2S 3S 4S 5S 6S 7S
Reference Chromosome ID
Z
Aetaus
Aecras6x
1D 2D 3D 4D 5D 6D 7D
Reference Chromosome ID



Appendix 9 Structural variation between polyploid subgenomes and their respective diploid progenitors.

(A-AH) Whole-genome alignments illustrating structural variation between polyploid *Aegilops* subgenomes and their corresponding diploid progenitors: **(A)** U subgenome of *Ae. neglecta* subsp. *neglecta* vs. *Ae. umbellulata*; **(B)** M subgenome of *Ae. neglecta* subsp. *neglecta* vs. *Ae. comosa*; **(C)** D subgenome of *Ae. cylindrica* vs. *Ae. tauschii*; **(D)** C subgenome of *Ae. cylindrica* vs. *Ae. markgrafii*; **(E)** D subgenome of *Ae. ventricosa* vs. *Ae. tauschii*; **(F)** N subgenome of *Ae. ventricosa* vs. *Ae. uniaristata*; **(G)** U subgenome of *Ae. triuncialis* vs. *Ae. umbellulata*; **(H)** C subgenome of *Ae. triuncialis* vs. *Ae. markgrafii*; **(I)** S subgenome of *Ae. peregrina* vs. *Ae. sharonensis*; **(J)** S subgenome of *Ae. peregrina* vs. *Ae. longissima*; **(K)** U subgenome of *Ae. peregrina* vs. *Ae. umbellulata*; **(L)** S subgenome of *Ae. kotschyi* vs. *Ae. sharonensis*; **(M)** S subgenome of *Ae. kotschyi* vs. *Ae. longissima*; **(N)** U subgenome of *Ae. kotschyi* vs. *Ae. umbellulata*; **(O)** M subgenome of *Ae. geniculata* vs. *Ae. comosa*; **(P)** U subgenome of *Ae. geniculata* vs. *Ae. umbellulata*; **(Q)** D subgenome of *Ae. crassa* subsp. *crassa* (4x) vs. *Ae. tauschii*; **(R)** M subgenome of *Ae. crassa* subsp. *crassa* (4x) vs. *Ae. comosa*; **(S)** U subgenome of *Ae. columnaris* vs. *Ae. umbellulata*; **(T)** M subgenome of *Ae. columnaris* vs. *Ae. comosa*; **(U)** U subgenome of *Ae. biuncialis* vs. *Ae. umbellulata*; **(V)** M subgenome of *Ae. biuncialis* vs. *Ae. comosa*; **(W)** D subgenome of *Ae. vavilovii* vs. *Ae. tauschii*; **(X)** M subgenome of *Ae. vavilovii* vs. *Ae. comosa*; **(Y)** S subgenome of *Ae. vavilovii* vs. *Ae. speltoides*; **(Z)** first D subgenome of *Ae. crassa* subsp. *crassa* (6x) vs. *Ae. tauschii*; **(AA)** second D subgenome of *Ae. crassa* subsp. *crassa* (6x) vs. *Ae. tauschii*; **(AB)** M subgenome of *Ae. crassa* subsp. *crassa* (6x) vs. *Ae. comosa*; **(AC)** D subgenome of *Ae. juvenalis* vs. *Ae. tauschii*; **(AD)** M subgenome of *Ae. juvenalis* vs. *Ae. comosa*; **(AE)** U subgenome of *Ae. juvenalis* vs. *Ae. umbellulata*; **(AF)** U subgenome of *Ae. neglecta* subsp. *recta* vs. *Ae. umbellulata*; **(AG)** M subgenome of *Ae. neglecta* subsp. *recta* vs. *Ae. comosa*; and **(AH)** N subgenome of *Ae. neglecta* subsp. *recta* vs. *Ae. uniaristata*. Colored links indicate syntenic regions (grey), inversions (orange), translocations (green), and duplications (blue). White regions represent unaligned or filtered sequences, which may reflect large insertions or deletions, assembly gaps, or structural variants smaller than 10 kb that are not displayed. Species abbreviations: Aeumb, *Ae. umbellulata*; Aecom, *Ae. comosa*; Aetaus, *Ae. tauschii*; Aemark, *Ae. markgrafii*; Aeuni, *Ae. uniaristata*; Aespelt, *Ae. speltoides*; Aeshar, *Ae. sharonensis*; Aelong, *Ae. longissima*; Aeneg4x, *Ae. neglecta* subsp. *neglecta*; Aecylin, *Ae. cylindrica*; Aevent, *Ae. ventricosa*; Aetriun, *Ae. triuncialis*; Aepereg, *Ae. peregrina*; Aekots, *Ae. kotschyi*; Aegen, *Ae. geniculata*; Aecras4x, *Ae. crassa* subsp. *crassa* (4x); Aecolum, *Ae. columnaris*; Aebiun, *Ae. biuncialis*; Aevav, *Ae. vavilovii*; Aecras6x, *Ae. crassa* subsp. *crassa* (6x); Aejuv, *Ae. juvenalis*; Aeneg6x, *Ae. neglecta* subsp. *recta*.


Annotations
Syntenic
Inversion
Translocation
Duplication
A
Aeumb
Genome
Aeneg4x
1 2 3 4 5 6 7
B
Aetaus
Aecras4x
1 2 3 4 5 6 7
C
Aeumb
Genome
Aecolum
1 2 3 4 5 6 7
D
Aetaus
Aevav
1 2 3 4 5 6 7
E
Aetaus
Genome
Aecras6x
1 2 3 4 5 6 7
F
Aetaus
Aejuv
1 2 3 4 5 6 7
G
Aeumb
Genome
Aeneg6x
1 2 3 4 5 6 7
Reference Chromosome ID
H
Aeshar
Aevav
1S 2S 3S 4S 5S 6S 7S
Reference Chromosome ID



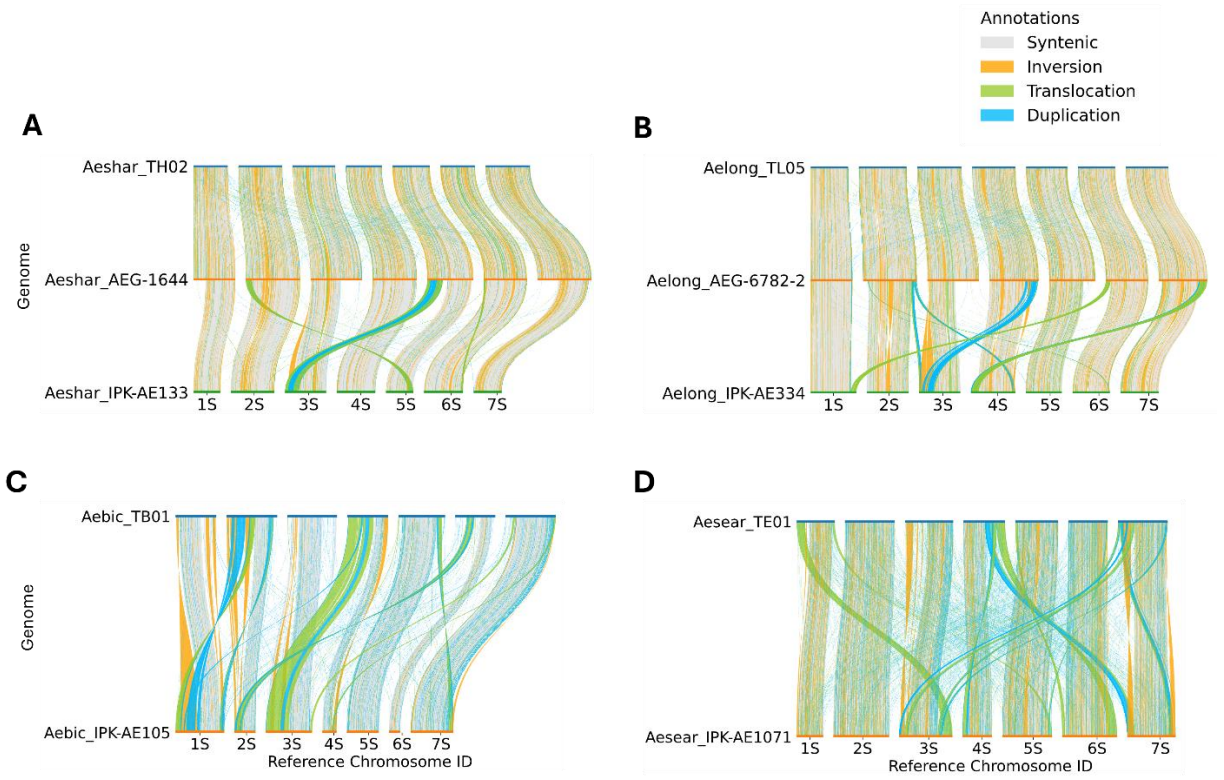
Appendix 10 Structural variation between polyploid subgenomes and references corresponding to their coexisting subgenome counterparts.

(A–N) Pairwise whole-genome alignments illustrating structural variation between polyploid subgenomes and references corresponding to their coexisting subgenome counterparts: **(A)** M subgenome of *Ae. neglecta* subsp. *neglecta* vs. U subgenome (*Ae. umbellulata*); **(B)** M subgenome of *Ae. crassa* subsp. *crassa* (4x) vs. D subgenome (*Ae. tauschii*); **(C)** M subgenome of *Ae. columnaris* vs. U subgenome (*Ae. umbellulata*); **(D)** M subgenome of *Ae. vavilovii* vs. D subgenome (*Ae. tauschii*); **(E)** M subgenome of *Ae. crassa* subsp. *crassa* (6x) vs. D subgenome (*Ae. tauschii*); **(F)** M subgenome of *Ae. juvenalis* vs. D subgenome (*Ae. tauschii*); **(G)** M subgenome of *Ae. neglecta* subsp. *recta* vs. U subgenome (*Ae. umbellulata*); **(H)** S subgenome of

Ae. vavilovii vs. S^{sh} (*Ae. sharonensis*); **(I)** S subgenome of *Ae. vavilovii* vs. S^{l} (*Ae. longissima*); **(J)** S subgenome of *Ae. vavilovii* vs. S^{b} (*Ae. bicornis*); **(K)** S subgenome of *Ae. vavilovii* vs. S^{s} (*Ae. searsii*); **(L)** S subgenome of *Ae. vavilovii* vs. D subgenome (*Ae. tauschii*); **(M)** S subgenome of *Ae. peregrina* vs. S (*Ae. speltoides*); and **(N)** S subgenome of *Ae. kotschyi* vs. S (*Ae. speltoides*). Colored links indicate syntenic regions (grey), inversions (orange), translocations (green), and duplications (blue). White regions represent unaligned or filtered sequences, which may reflect large insertions or deletions, assembly gaps, or structural variants smaller than 10 kb that are not displayed. Species abbreviations: Aeumb, *Ae. umbellulata*; Aeneg4x, *Ae. neglecta* subsp. *neglecta*; Aetaus, *Ae. tauschii*; Aecras4x, *Ae. crassa* subsp. *crassa* (4x); Aecolum, *Ae. columnaris*; Aevav, *Ae. vavilovii*; *Ae. crassa* subsp. *crassa* (6x); Aejuv, *Ae. juvenalis*; Aeneg6x, *Ae. neglecta* subsp. *recta*; Aeshar, *Ae. sharonensis*; Aelong, *Ae. longissima*; Aebic, *Ae. bicornis*; Aesear, *Ae. searsii*; Aespelt, *Ae. speltoides*; Aepereg, *Ae. peregrina*; Aekots, *Ae. kotschyi.*

Appendix 11 Synteny and SV counts for subgenomes compared with the reference genome corresponding to their coexisting subgenome, where applicable.

Species-progenitor comparison									
Polyploid species	**Polyploid genome**	**Diploid species**	**Diploid genome**	**Synteny**	**Inversions**	**Translocations**	**Duplications**	**Insertions**	**Deletions**
Ae. neglecta subsp. *neglecta*	M	*Ae. comosa*	M	290	50	531	3,742	28	19
Ae. neglecta subsp. *neglecta*	M	*Ae. umbellulata*	U	2,165	99	2,565	9,957	962	964
Ae. crassa subsp. *crassa* (4x)	M	*Ae. comosa*	M	276	63	662	3,185	17	28
Ae. crassa subsp. *crassa* (4x)	M	*Ae. tauschii*	D	1,148	98	3,712	5,350	238	239
Ae. columnaris	M	*Ae. comosa*	M	567	123	910	9,735	34	37
Ae. columnaris	M	*Ae. umbellulata*	U	2,129	164	2,520	27,637	773	1,744
Ae. vavilovii	M	*Ae. comosa*	M	170	54	292	2,057	4	4
Ae. vavilovii	M	*Ae. tauschii*	D	2,397	201	3,175	4,472	518	700
Ae. crassa subsp. *crassa* (6x)	M	*Ae. comosa*	M	76	29	315	1,018	5	4
Ae. crassa subsp. *crassa* (6x)	M	*Ae. tauschii*	D	390	58	3,033	2,943	215	285
Ae. juvenalis	M	*Ae. comosa*	M	117	27	276	1,210	2	4
Ae. juvenalis	M	*Ae. tauschii*	D	235	60	938	634	86	111
Ae. neglecta subsp. recta	M	*Ae. comosa*	M	306	49	812	6,453	55	91
Ae. neglecta subsp. recta	M	*Ae. umbellulata*	U	906	68	1,498	8,095	715	772
Ae. vavilovii	S	*Ae. speltoides*	B	249	55	621	740	5	9
Ae. vavilovii	S	*Ae. sharonensis*	S^{sh}	87	41	214	247	6	5
Ae. vavilovii	S	*Ae. longissima*	S^{l}	102	45	306	549	5	4
Ae. vavilovii	S	*Ae. bicornis*	S^{b}	128	39	249	1,139	6	5
Ae. vavilovii	S	*Ae. searsii*	S^{s}	182	52	251	719	10	9
Ae. vavilovii	S	*Ae. tauschii*	D	1,449	114	3,055	13,038	237	291
Ae. peregrina	S	*Ae. speltoides*	S	93	38	274	1,360	5	6
Ae. peregrina	S	*Ae. sharonensis*	S^{sh}	7,404	143	4,329	15,432	4,436	4,723
Ae. peregrina	S	*Ae. longissima*	S^{l}	6,762	165	6,005	13,860	4,470	4,824
Ae. kotschyi	S	*Ae. speltoides*	S	39	16	88	292	0	1
Ae. kotschyi	S	*Ae. sharonensis*	S^{sh}	1,121	118	4,232	3,129	776	881
Ae. kotschyi	S	*Ae. longissima*	S^{l}	1,028	121	4,805	3,486	712	826



Appendix 12 Structural variation identified in intraspecific diploid *Aegilops* genome comparisons.

(A–D) Pairwise genome alignments illustrating structural variation between assemblies of *Ae. sharonensis*, *Ae. longissima*, *Ae. bicornis*, and *Ae. searsii*. Assemblies include previously published genomes from Li et al. (2022): Aeshar_TH02 (*Ae. sharonensis*), Aelong_TL05 (*Ae. longissima*), Aebic_TB01 (*Ae. bicornis*), and Aesear_TE01 (*Ae. searsii*); and Avni et al. (2022): Aeshar_AEG-1644 (*Ae. sharonensis*) and Aelong_AEG-6782-2 (*Ae. longissima*). Assemblies generated in this study include Aeshar_IPK-AE133 (*Ae. sharonensis*), Aelong_IPK-AE334 (*Ae. longissima*), Aebic_IPK-AE105 (*Ae. bicornis*), and Aesear_IPK-AE1071 (*Ae. searsii*). Colored links indicate syntenic regions (grey), inversions (orange), translocations (green), and duplications (blue). White regions represent unaligned or filtered sequences, which may correspond to large insertions or deletions, assembly gaps, or structural variants smaller than 10 kb that are not displayed.

Appendix 13 Synteny and SV counts identified in pairwise diploid *Aegilops* genome comparisons.

Species	Comparison	Genome	Synteny	Inversions	Translocations	Duplications	Insertions	Deletions
Ae. sharonensis	TH02 vs AEG-1644	S^{sh}	18047	837	7183	15485	26427	6901
	AEG-1644 vs IPK-AE133		8477	308	5757	9527	4140	23061
Ae. longissima	TL05 vs AEG-6782-2	S^{l}	17081	927	6952	13814	24117	6897
	AEG-6782-2 vs IPK-AE334		6902	513	4545	7685	4006	17020
Ae. bicornis	TB01 vs IPK-AE105	S^{b}	1253	76	525	4479	518	2562
Ae. searsii	TE01 vs IPK-AE1071	S^{s}	9827	770	2941	10687	5142	4671

Appendix 14 Percentage composition of transposable elements across *Aegilops* genomes.

Class	*Ae. sharonensis*	*Ae. longissima*	*Ae. bicornis*	*Ae. searsii*	*Ae. neglecta* subsp. *neglecta*	*Ae. cylindrica*	*Ae. ventricosa*	*Ae. triuncialis*	*Ae. peregrina*
Class I LTR retrotransposons									
Gypsy	46.84	45.87	45.00	44.52	45.62	41.67	40.60	44.94	45.59
Copia	16.17	16.97	15.86	15.37	17.70	17.71	17.27	18.31	16.42
Unclassified LTRs	7.32	6.75	6.90	7.45	6.66	5.84	7.31	7.14	7.56
Class I non-LTR retrotransposons									
LINE	0.19	0.17	0.27	0.18	0.23	0.25	0.17	0.32	0.22
SINE	0.93	0.92	0.95	0.96	1.17	1.03	1.07	1.17	1.01
Class II DNA transposons									
CACTA	9.37	10.33	11.77	12.34	9.03	11.14	13.41	8.54	9.93
Mutator	0.72	0.79	0.91	0.92	1.12	1.18	0.94	1.00	0.92
PIF_Harbinger	2.04	1.01	1.02	0.98	1.14	1.18	1.17	1.14	0.96
Tc1_Mariner	1.21	1.10	1.33	1.27	1.17	1.18	1.25	1.19	1.39
hAT	0.16	0.16	0.23	0.16	0.24	0.20	0.24	0.26	0.21
Helitrons	2.30	2.59	2.59	2.77	2.55	3.61	3.16	2.65	2.59
Unknown TEs	0.58	0.76	0.54	0.57	0.53	0.45	0.36	0.73	0.35
Total	87.83	87.42	87.37	87.49	87.16	85.44	86.95	87.39	87.15

Appendix 15 Percentage composition of transposable elements across *Aegilops* genomes.

Class	*Ae. kotschyi*	*Ae. geniculata*	*Ae. crassa* subsp. *crassa* (4x)	*Ae. columnaris*	*Ae. biuncialis*	*Ae. vavilovii*	*Ae. crassa* subsp. *crassa* (6x)	*Ae. juvenalis*	*Ae. neglecta* subsp. *recta*
Class I LTR retrotransposons									
Gypsy	46.26	44.54	38.00	45.22	42.69	38.92	39.71	41.25	44.65
Copia	16.81	19.04	13.97	18.06	18.95	14.90	14.72	15.29	18.42
Unclassified LTRs	6.70	4.76	9.25	6.72	6.11	7.31	6.44	6.89	6.74
Class I non-LTR retrotransposons									
LINE	0.19	0.27	0.34	0.28	0.25	0.19	0.17	0.19	0.23
SINE	1.01	1.14	1.10	1.20	1.07	1.02	1.02	1.08	1.12
Class II DNA transposons									
CACTA	10.24	10.15	16.01	9.20	9.93	14.32	15.29	13.80	9.62
Mutator	0.95	1.08	1.00	0.98	1.09	1.92	1.26	1.19	1.05
PIF_Harbinger	1.08	1.19	1.26	1.10	1.20	1.30	1.38	1.28	1.09
Tc1_Mariner	1.09	1.32	1.52	1.30	1.37	2.92	1.44	1.29	1.30
hAT	0.19	0.23	0.20	0.22	0.22	14.32	15.29	13.80	0.25
Helitrons	3.12	2.67	3.46	2.59	3.30	1.92	1.26	1.19	3.19
Unknown TEs	0.46	0.54	0.60	0.51	0.42	1.30	1.38	1.28	0.37
Total	88.10	86.93	86.71	87.38	86.60	2.92	1.44	1.29	88.03

Appendix 16 LTR clade abundance (counts per Gb) in polyploid U subgenomes relative to the diploid U genome

Clade	*Ae. umbellulata*	*Ae. neglecta* subsp. *neglecta*	*Ae. triuncialis*	*Ae. peregrina*	*Ae. kotschyi*	*Ae. geniculata*	*Ae. columnaris*	*Ae. biuncialis*	*Ae. juvenalis*	*Ae. neglecta* subsp. *recta*
Angela	14090	15334	14857	14945	14891	17269	16531	18098	13330	16955
SIRE	6148	6101	6024	6671	6486	6032	6435	5276	5520	5539
Ale	2205	2415	2295	2173	2251	2840	2163	2957	2349	2664
TAR	745	801	766	660	725	803	793	802	841	796
Retand	25962	26888	24780	27345	26858	22989	27861	20364	22542	23790
Athila	9491	10259	9819	10086	9893	9773	11357	9420	11306	10305
Tekay	5238	5515	5048	5517	5383	4617	5490	4546	5082	5843
CRM	6185	6138	4929	5761	4804	3489	5654	2881	6248	5635
Reina	1067	1145	1134	1178	1122	1251	1133	1163	1091	1143
Total	71131	74595	69652	74337	72413	69065	77419	65507	68309	72671

Appendix 17 LTR clade abundance (counts per Gb) in polyploid M subgenomes relative to the diploid M genome.

Clade	*Ae. comosa*	*Ae. neglecta* subsp. *neglecta*	*Ae. geniculata*	*Ae. crassa* subsp. *crassa* (4x)	*Ae. columnaris*	*Ae. biuncialis*	*Ae. vavilovii*	*Ae. crassa* subsp. *crassa* (6x)	*Ae. juvenalis*	*Ae. neglecta* subsp. *recta*
Angela	22238	17990	18300	13218	17045	17055	14872	14057	13987	16944
SIRE	5224	5434	6173	4981	5754	5886	4739	4331	5506	6577
Ale	2081	2398	1810	2404	2697	1821	2444	2800	2032	1885
TAR	647	798	607	1083	854	537	980	1026	896	723
Retand	20210	26708	25252	18400	26516	24146	18103	16125	21510	26791
Athila	9646	10788	11138	12757	10128	10465	11821	11507	12072	11288
Tekay	6121	5667	6264	4577	5106	5876	5354	4442	4806	5954
CRM	3866	4683	6864	3591	4969	6255	4871	2734	4270	4562
Reina	1079	1060	1133	1077	1190	1008	1133	1035	974	1003
Total	71113	75526	77543	62088	74259	73050	64316	58057	66055	75726

Appendix 18 LTR clade abundance (counts per Gb) in polyploid D subgenomes relative to the diploid D genome.

Clade	*Ae. tauschii*	*Ae. cylindrica*	*Ae. ventricosa*	*Ae. crassa* subsp. *crassa* (4x)	*Ae. vavilovii*	*Ae. crassa* subsp. *crassa* (6x, D1)	*Ae. crassa* subsp. *crassa* (6x, D2)	*Ae. juvenalis*
Angela	16,598	14,225	16,880	13,069	14,812	16,907	13,762	14,232
SIRE	4,483	4,572	4,958	4,779	4,619	5,103	4,701	4,956
Ale	2,100	2,167	2,489	2,623	2,624	2,913	2,590	2,724
TAR	1,034	695	935	1,138	1,009	1,104	947	1,046
Retand	17,325	17,535	18,919	17,837	17,176	18,515	17,441	18,549
Athila	10,969	9,919	11,486	12,340	11,837	12,811	11,734	11,414
Tekay	6,538	5,164	5,942	4,384	4,898	5,386	5,015	4,432
CRM	3,203	3,724	3,914	4,610	4,144	4,739	3,909	3,133
Reina	684	988	1,074	1,132	1,094	1,248	1,026	1,125
Total	62,935	58,988	66,598	61,912	62,214	68,727	61,126	61,611

Appendix 19 LTR clade abundance (counts per Gb) in polyploid S subgenomes relative to the diploid S genome.

Clade	*Ae. sharonensis*	*Ae. longissima*	*Ae. bicornis*	*Ae. searsii*	*Ae. peregrina*	*Ae. kotschyi*	*Ae. speltoides*	*Ae. vavilovii*
Angela	14,353	15,352	13,486	13,554	14,013	15,076	15,776	13,935
SIRE	6,110	5,761	6,336	5,814	6,103	6,175	7,836	4,801
Ale	2,140	2,261	2,176	2,183	2,063	2,200	2,859	2,387
TAR	834	839	900	931	739	843	804	988
Retand	23,195	21,713	24,611	24,390	23,802	23,076	26,439	18,460
Athila	9,881	9,447	11,100	11,754	10,048	9,877	10,129	11,950
Tekay	8,261	7,959	6,254	5,253	8,187	7,501	4,224	5,026
CRM	5,035	4,902	3,808	4,499	4,988	3,275	5,961	3,868
Reina	999	1,004	1,127	1,059	1,103	1,046	1,380	980
Total	70,808	69,238	69,797	69,437	71,046	69,069	75,409	62,396

Appendix 20 LTR clade abundance (counts per Gb) in polyploid C subgenomes relative to the diploid C genome.

Clade	*Ae. markgrafii*	*Ae. cylindrica*	*Ae. triuncialis*
Angela	16,879	17,133	15,367
SIRE	6,392	5,167	6,024
Ale	2,751	2,607	2,416
TAR	837	891	703
Retand	25,044	19,374	23,324
Athila	13,453	11,384	11,168
Tekay	5,483	5,238	5,223
CRM	4,827	3,400	4,569
Reina	1,294	1,113	1,252
Total	76,961	66,304	70,047

Appendix 21 LTR clade abundance (counts per Gb) in polyploid N subgenomes relative to the diploid N genome

Clade	*Ae. uniaristata*	*Ae. ventricosa*	*Ae. neglecta* subsp. *recta*
Angela	17,878	17,196	15,471
SIRE	5,443	5,624	4,723
Ale	2,242	2,528	2,162
TAR	784	809	684
Retand	19,759	21,359	19,529
Athila	10,905	12,088	9,466
Tekay	5,990	5,945	4,848
CRM	3,088	3,087	3,122
Reina	1,068	1,102	978
Total	67,155	69,736	60,983

Appendix 22 LTR clade abundance (counts per Gb) across subgenomes within tetraploid *Aegilops* species.

Clade	***Ae. neglecta* subsp. *neglecta***		***Ae. cylindrica***		***Ae. ventricosa***		***Ae. triuncialis***		***Ae. peregrina***	
	U	M	D	C	D	N	U	C	S	U
Angela	15,334	17,990	14,225	17,133	16,880	17,196	14,857	15,367	14,013	14,945
SIRE	6,101	5,434	4,572	5,167	4,958	5,624	6,024	6,024	6,103	6,671
Ale	2,415	2,398	2,167	2,607	2,489	2,528	2,295	2,416	2,063	2,173
TAR	801	798	695	891	935	809	766	703	739	660
Retand	26,888	26,708	17,535	19,374	18,919	21,359	24,780	23,324	23,802	27,345
Athila	10,259	10,788	9,919	11,384	11,486	12,088	9,819	11,168	10,048	10,086
Tekay	5,515	5,667	5,164	5,238	5,942	5,945	5,048	5,223	8,187	5,517
CRM	6,138	4,683	3,724	3,400	3,914	3,087	4,929	4,569	4,988	5,761
Reina	1,145	1,060	988	1,113	1,074	1,102	1,134	1,252	1,103	1,178
Total	74,595	75,526	58,988	66,304	66,598	69,736	69,652	70,047	71,046	74,337

Appendix 23 LTR clade abundance (counts per Gb) across subgenomes within tetraploid *Aegilops* species.

Clade	***Ae. kotschyi***		***Ae. geniculata***		***Ae. crassa* subsp. *crassa* (4x)**		***Ae. columnaris***		***Ae. biuncialis***	
	S	U	M	U	D	M	U	M	U	M
Angela	15,076	14,891	18,300	17,269	13,069	14,057	16,531	17,045	18,098	17,055
SIRE	6,175	6,486	6,173	6,032	4,779	4,331	6,435	5,754	5,276	5,886
Ale	2,200	2,251	1,810	2,840	2,623	2,800	2,163	2,697	2,957	1,821
TAR	843	725	607	803	1,138	1,026	793	854	802	537
Retand	23,076	26,858	25,252	22,989	17,837	16,125	27,861	26,516	20,364	24,146
Athila	9,877	9,893	11,138	9,773	12,340	11,507	11,357	10,128	9,420	10,465
Tekay	7,501	5,383	6,264	4,617	4,384	4,442	5,490	5,106	4,546	5,876
CRM	3,275	4,804	6,864	3,489	4,610	2,734	5,654	4,969	2,881	6,255
Reina	1,046	1,122	1,133	1,251	1,132	1,035	1,133	1,190	1,163	1,008
Total	69,069	72,413	77,543	69,065	61,912	58,057	77,419	74,259	65,507	73,050

Appendix 24 LTR clade abundance (counts per Gb) across subgenomes within hexaploid *Aegilops* species.

Clade	*Ae. vavilovii*			*Ae. crassa* subsp. *crassa (6x)*			*Ae. juvenalis*			*Ae. neglecta* subsp. *recta*		
	D	M	S	D1	D2	M	D	M	U	U	M	N
Angela	14,812	14,872	13,935	16,907	13,762	14,057	14,232	13,987	13,330	16,955	16,944	15,471
SIRE	4,619	4,739	4,801	5,103	4,701	4,331	4,956	5,506	5,520	5,539	6,577	4,723
Ale	2,624	2,444	2,387	2,913	2,590	2,800	2,724	2,032	2,349	2,664	1,885	2,162
TAR	1,009	980	988	1,104	947	1,026	1,046	896	841	796	723	684
Retand	17,176	18,103	18,460	18,515	17,441	16,125	18,549	21,510	22,542	23,790	26,791	19,529
Athila	11,837	11,821	11,950	12,811	11,734	11,507	11,414	12,072	11,306	10,305	11,288	9,466
Tekay	4,898	5,354	5,026	5,386	5,015	4,442	4,432	4,806	5,082	5,843	5,954	4,848
CRM	4,144	4,871	3,868	4,739	3,909	2,734	3,133	4,270	6,248	5,635	4,562	3,122
Reina	1,094	1,133	980	1,248	1,026	1,035	1,125	974	1,091	1,143	1,003	978
Total	62,214	64,316	62,396	68,727	61,126	58,057	61,611	66,055	68,309	72,671	75,726	60,983

Appendix 25 Peak insertion times and insertion counts of *Copia*, *Gypsy*, and unknown LTR retrotransposons across U subgenomes.

Species	*Copia*		*Gypsy*		Unknown	
	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)
Ae. umbellulata	0.524	5,656	0.904	6,515	0.76	1,187
Ae. neglecta subsp. *neglecta*	0.745	1,518	0.925	1,307	0.872	745
Ae. triuncialis	0.164	6,596	0.677	5,530	0.194	2,418
Ae. peregrina	0.605	3,006	0.961	3,907	0.864	1,471
Ae. kotschyi	0.555	1,411	1.023	1,652	1.224	972
Ae. geniculata	0.741	2,916	1.187	2,567	1.339	786
Ae. columnaris	0.517	5,819	0.922	5,057	0.6	1,705
Ae. biuncialis	1.027	593	1.445	287	1.502	398
Ae. juvenalis	0.776	1,846	1.12	2,453	1.205	1,134
Ae. neglecta subsp. *recta*	0.776	2,425	1.168	2,526	1.033	769

Appendix 26 Peak insertion times and insertion counts of *Copia*, *Gypsy*, and unknown LTR retrotransposons across M subgenomes.

Species	*Copia*		*Gypsy*		Unknown	
	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)
Ae. comosa	0.303	10,673	0.102	7,599	0.317	2,112
Ae. neglecta subsp. *neglecta*	0.774	3,099	0.997	3,430	0.863	1,838
Ae. geniculata	0.744	2,569	1.128	2,427	1.259	774
Ae. crassa subsp. *crassa* (4x)	0.751	2,636	1.114	3,513	0.938	3,056
Ae. columnaris	0.488	5,071	0.640	5,689	0.619	1,974
Ae. biuncialis	1.055	830	1.468	503	1.45	748
Ae. vavilovii	0.856	1,136	1.178	1,504	1.2	902
Ae. crassa subsp. *crassa* (6x)	0.933	732	1.165	1,017	1.22	653
Ae. juvenalis	0.830	1,337	1.175	1,942	1.18	1,002
Ae. neglecta subsp. *recta*	0.808	1,728	1.193	1,984	0.886	592

Appendix 27 Peak insertion times and insertion counts of *Copia*, *Gypsy*, and unknown LTR retrotransposons across D subgenomes.

Species	*Copia*		*Gypsy*		Unknown	
	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)
Ae. tauschii	0.788	2,686	0.993	3,195	1.098	2,013
Ae. cylindrica	0.913	635	1.195	441	1.241	753
Ae. ventricosa	0.865	2,141	1.161	2,473	1.268	1,830
Ae. crassa subsp. *crassa* (4x)	0.739	2,956	1.115	4,085	1.03	3,197
Ae. vavilovii	0.863	1,751	1.265	2,326	1.19	1,384
Ae. crassa subsp. *crassa* (6x, D1)	0.888	1,287	1.249	1,683	1.203	948
Ae. crassa subsp. *crassa* (6x, D2)	0.96	1,219	1.232	1,681	1.241	1,193
Ae. juvenalis	0.838	1,606	1.232	2,304	1.258	1,282

Appendix 28 Peak insertion times and insertion counts of *Copia*, *Gypsy*, and unknown LTR retrotransposons across S^x subgenomes.

Species	*Copia*		*Gypsy*		Unknown	
	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)
Ae. sharonensis	0.327	3,113	0.747	3,737	1.225	1,793
Ae. speltoides	1.041	194	0.781	746	0.644	138
Ae. longissima	0.332	2,570	0.849	2,663	1.137	1,502
Ae. bicornis	0.188	3,461	0.969	4,303	1.203	2,038
Ae. searsii	0.544	2,925	0.934	4,651	1.015	2,445
Ae. peregrina	0.213	4,160	0.65	5,191	1.065	2,014
Ae. kotschyi	0.547	2,179	1.013	2,140	1.122	1,366
Ae. vavilovii	0.882	1,760	1.13	2,627	1.143	1,630

Appendix 29 Peak insertion times and insertion counts of *Copia*, *Gypsy*, and unknown LTR retrotransposons across C subgenomes.

Species	*Copia*		*Gypsy*		Unknown	
	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)
Ae. markgrafii	0.15	8,759	0.241	9,657	0.182	3,135
Ae. cylindrica	0.913	569	1.068	310	1.205	606
Ae. triuncialis	0.136	6,543	0.57	5,738	0.172	2,758

Appendix 30 Peak insertion times and insertion counts of *Copia*, *Gypsy*, and unknown LTR retrotransposons across N subgenomes.

Species	*Copia*		*Gypsy*		Unknown	
	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)	Insertion time at peak (Myr)	Number of insertions at peak (n)
Ae. uniaristata	0.249	5,481	1.142	4,245	1.288	1,467
Ae. ventricosa	0.856	1,670	1.158	1,706	1.224	1,201
Ae. neglecta subsp. recta	0.771	2,270	1.1	2,096	1.058	714

Appendix 31 Genome size estimates for *Aegilops* species.

Species	Genome	Genome size[1] (Gb)	Genome size[2] (Gb)
Ae. sharonensis	S^{sh}	7.52	7.08 ± 0.06
Ae. longissima	S^{l}	7.48	7.14 ± 0.10
Ae. bicornis	S^{b}	6.84	6.50 ± 0.04
Ae. searsii	S^{s}	6.65	6.36 ± 0.03
Ae. neglecta subsp. *neglecta*	UM	10.64	10.60 ± 0.31
Ae. cylindrica	DC	9.59	8.83 ± 0.03
Ae. ventricosa	DN	10.64	9.98 ± 0.10
Ae. triuncialis	UC	9.93	9.12 ± 0.02
Ae. peregrina	SU	12.52	12.01 ± 0.06
Ae. kotschyi	SU	12.64	11.73 ± 0.13
Ae. geniculata	MU	10.29	9.43 ± 0.07
Ae. crassa subsp. *crassa* (4x)	DM	10.86	10.39 ± 0.04
Ae. columnaris	UM	10.86	10.58 ± 0.02
Ae. biuncialis	UM	10.37	9.76 ± 0.02
Ae. vavilovii	DMS	17.13	15.12 ± 0.15
Ae. crassa subsp. *crassa* (6x)	DDM	nd	15.02 ± 0.02
Ae. juvenalis	DMU	nd	15.28 ± 0.11
Ae. neglecta subsp. *recta*	UMN	16.22	15.95 ± 0.18

[1] Eilam et al. (2007) and Eilam et al. (2008); nd = not determined; [2] Our estimate based on flow cytometry